

Predicting Colorectal Cancer Recurrence by Utilizing
Multiple-View Multiple-Learner Supervised Learning

By

Jason Alfred Castellanos

Thesis

Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

in

Biomedical Informatics

August, 2016

Nashville, Tennessee

Approved:

Bing Zhang, Ph.D.

R. Daniel Beauchamp, M.D.

Qi Liu, Ph.D.

To Emily, Iris and Isaac. Thank you for your love, support, and patience.

ACKNOWLEDGEMENTS

This work would not have been possible without the generous financial support of the Vanderbilt University Medical Center Section of Surgical Sciences. I am especially indebted to Dr. R. Daniel Beauchamp, Chairman of the Section of Surgical Sciences, and Dr. Kyla Terhune, Program Director of the General Surgery Residency, who have been supportive of my career goals and who worked actively to provide me with the protected academic time to pursue those goals.

I am also grateful to the members of Dr. Bing Zhang's lab and the members of my Thesis Committee. They have all provided me with the personal and professional guidance that have enabled me to learn a great deal about scientific research. I would especially like to thank Dr. Bing Zhang, the chairman of my committee, for his patience, example, and mentorship. Finally, I would like to thank my parents, my wife Emily, her parents Lynda and Jeremy, and my children, Iris and Isaac, who provided me with the inspiration and support required to complete this work.

TABLE OF CONTENTS

	Page
DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
 Chapter	
I. Introduction	1
Adjuvant therapy for stage II colorectal cancer	2
Clinicopathologic and molecular risk stratification	3
Ensemble learning	6
II. Creating an Ensemble Framework	9
Introduction	9
Data curation and normalization	9
Feature selection methodology	12
Predictive power of clinical vs. microarray data	13
III. Creating a Multiple-View Multiple-Learner Framework	19
Introduction	19
Discretized view	19
Gene set view	20
Network propagation view	20
Network-Expression-Mutation signature view	21
Performance of the multiple-learner multiple-view framework	22
IV. Summary	31
Conclusions	31
 Appendix	
A. Base Learners	33
B. Code to Generate Datasets	35

C. Features Selection and Network Propagation	39
D. Base Learner Training	42
E. Ensemble Training	45
F. Signature Overlap.....	48
REFERENCES	49

LIST OF TABLES

Table	Page
1. Definitions of high-risk stage II CRC by expert group.....	4
2. Patient characteristics by GEO dataset	10
3. Training and validation set patient characteristics.....	11
4. Clinical data features.....	16
5. Ensemble vs. selectBest performance on clinical data	17
6. Ensemble vs. selectBest performance on microarray data.....	18
7. Ensemble vs. selectBest performance on testing data	24
8. Detailed performance metrics of ensemble vs. selectBest methods on testing set	27
9. Detailed performance metrics of ensemble vs. selectBest methods on validation set.....	28

LIST OF FIGURES

Figure	Page
1. Overlap in gene predictors between commercialized recurrence prediction assays.....	6
2. Framework for stacked generalization.....	8
3. Disease free survival by stage.....	10
4. Data preprocessing pipeline.....	12
5. Stacked generalization training diagram.....	14
6. Training performance of clinical vs. microarray models.....	15
7. Performance of individual base learners on original view data.....	15
8. Training performance of molecular view models.....	23
9. Training performance of base learners on molecular view data.....	23
10. Lift curve for ensemble methods.....	25
11. Probability calibration plot for ensemble models.....	26
12. DFS by predictive model compared to stage.....	28
13. DFS by Oncotype Dx recurrence score.....	30
14. Relative contribution of models by molecular view.....	30

CHAPTER 1

INTRODUCTION

Colorectal cancer (CRC) remains the third leading cause of cancer incidence and death in the USA, accounting for 49,700 deaths in 2015.¹ A key therapeutic dilemma in the treatment of CRC is whether patients with stage II and stage III disease require adjuvant chemotherapy after surgical resection. While patients with stage II CRC have an overall survival (OS) of 70 – 80% after surgical resection, 20% will experience disease recurrence.² Adjuvant chemotherapy significantly improves survival for patients with stage III disease, although historical data indicates that 50 – 60% of these treated patients would not experience disease recurrence after surgical resection alone.³⁻⁵

These statistics demonstrate that current treatment strategies both under and over treat patients with CRC. Equivocal results from clinical trials exploring this question underscore the need for improved prognostic models to aid clinicians in the decision to use adjuvant therapy in stage II and stage III CRC.⁶⁻⁸ Attempts to identify molecular signatures predictive of recurrence in patients with stage II CRC resulted in 5 commercialized gene expression assays, none of which are recommended in practice guidelines.⁹⁻¹⁶

First, a historical perspective on the use of adjuvant therapy in stage II and stage III CRC is provided. Next, the concept of high-risk disease is reviewed. Then, novel approaches to prediction are discussed. Finally, our research questions are presented.

Adjuvant therapy for stage II disease

The management of stage II CRC has been among the most challenging and controversial problems in oncology over the past 20 years. Efforts by the Gastrointestinal Intergroup (INT) and National Surgical Adjuvant Breast and Bowel Project (NSABP) to resolve this controversy in the 1990s led to conflicting results. Two INT trials – INT-0035 and INT-0089 – provided evidence that OS was similar with or without adjuvant chemotherapy, while data pooled from four NSABP trials – C-01, C-02, C-03, and C-04 – demonstrated a 30% reduction in overall mortality with adjuvant chemotherapy.^{3,5,17-21} A meta-analysis performed by the International Multicentre Pooled Analysis of B2 Colon Cancer (IMPACT B2) group failed to demonstrate significant increases in disease-free survival (DFS) or OS.^{4,22,23} While these studies failed to clarify the treatment for stage II CRC, they did clearly demonstrate that patients with stage III CRC benefitted significantly from adjuvant chemotherapy; as a result, 5-Fluoruracil (5-FU) and leucovorin became the standard treatment for stage III CRC.

A study of the treatment patterns of oncologists during this period revealed that 27% of Medicare patients with stage II CRC received adjuvant chemotherapy. In this population treatment led to a 5-year OS of 78% vs. 75% with surgical resection alone, a marginal survival benefit that was not statistically significant.²⁴ The Cancer Care Ontario Practice Guideline Initiative Gastrointestinal Cancer Disease Site Group (CCPOGI) performed a meta-analysis that included 37 randomized, controlled trials and 11 meta-analyses. This study determined that adjuvant therapy did improve DFS 5-10%, but this decrease in recurrence rate was not associated with any improvement in OS. Additionally, adjuvant chemotherapy was found to have a mortality ratio of 0.87.^{25,26} Based on CCPOGI study, the American Society of Clinical Oncology (ASCO) recommended against routine adjuvant chemotherapy for stage II CRC.²⁶ Instead, they

advised that disease with high-risk features should prompt a discussion between the patient and oncologist to consider the use of adjuvant therapy.

The Multi-center International Study of Oxaliplatin/5-fluoruracil/Leucovorin in the Adjuvant Treatment of Colon Cancer (MOSAIC) trial demonstrated that the addition of oxaliplatin to 5-FU/leucovorin based adjuvant chemotherapy significantly improved 3-year DFS compared to control among patients with stage III CRC (72.2% vs. 65.3%), but not for patients with stage II CRC (87.0% vs. 84.3%).⁴ As a result, 5-FU/Leucovorin/Oxaliplatin (FOLFOX) became the standard of care for patients with stage III CRC. Updated data from this trial confirmed the lack of survival benefit for patients with stage II disease (78% vs. 79% 10-year OS).²⁷ The Quick and Simple and Reliable (QUASAR) trial compared adjuvant chemotherapy against observation in patients with “uncertain indication for adjuvant therapy”. Patients with stage II disease (91% of the study population) demonstrated only a trend towards improved OS (hazard ratio 0.86, 95% CI -.54-1.19, 5-year OS 83.9% vs. 81.5%), resulting in an approximately 3% survival benefit in the face of toxicity that resulted in deaths in approximately 0.5% of patients.²⁸

Taken together, we see that the decision to administer adjuvant chemotherapy requires a patient and physician to weigh complex information in the face of a great deal of uncertainty. While there is clear evidence demonstrating a survival benefit for adjuvant chemotherapy in stage III cancer, a review of practice patterns demonstrated that only 57% of Medicare patients with stage III disease received adjuvant treatment.²⁹ While some of these trials have shown a trend towards improved OS in stage II disease, other trials that specifically looked at the role of adjuvant chemotherapy in stage II disease failed to demonstrate a significant benefit. Decision

making in the treatment of stage II and III CRC remains complex, and so improving the ability of physicians to perform risk stratification remains a high priority.

Clinicopathologic and molecular risk stratification

Current guidelines recommend consideration of adjuvant chemotherapy to treat stage II CRC if high-risk features are present. However, there is no clear evidence that this strategy improves survival, and even the definition of high-risk features is inconsistent between ASCO, the National Comprehensive Cancer Network (NCCN), or the European Society for Medical Oncology (ESMO) (Table 1). This has resulted in wide practice variation in the treatment of stage II disease.^{30,31}

	ASCO	NCCN	ESMO
T4 Primary Tumor	+	+	+
Inadequately sampled lymph nodes	+	+	+
Poorly differentiated histology	+	+	+
Bowel perforation	+	+	+
Bowel obstruction		+	+
Lymphovascular invasion	+	+	+
Perineural invasion	+	+	+
Close/positive margins		+	

Table 1: Definitions of high-risk stage II CRC by expert group

While patients with high-risk disease may theoretically benefit from adjuvant chemotherapy, patients with microsatellite instability (MSI) clearly do not.^{32,33} Even patients with high-risk clinicopathologic features (e.g. poorly differentiated histology) have a favorable prognosis if their tumors test positively for MSI or deficient mismatch repair (dMMR) status. While characterizing MSI status helps guide patients and clinicians towards observation over

chemotherapy, this cohort only represents approximately 15% of patients with stage II disease. This finding does demonstrate, however, that molecular features may potentially stratify patients at higher risk for recurrence with greater precision than previously possible with only clinicopathologic data.

The advent of accessible transcriptomic sequencing resulted in the creation of a multitude of gene expression signatures focused on identifying patients with stage II CRC that are at high-risk for recurrence.^{9-16,34-38} Five of these prognostic signatures were commercialized in an attempt to allow clinicians to utilize molecular information while making treatment decisions: Oncotype Dx Colon Cancer (Genomic Health, Inc.), ColonPRS (Signal Genetics LLC), ColoPrint (Agendia NV), GeneFx Colon (Precision Therapeutics, Inc), and OncoDefender-CRC (Everist Genomics, Inc.). While the overlap between genes used in each of these signatures is minimal (Figure 1), many of these signatures have demonstrated promising results. However, only two of these tests have been validated externally, and none are able to predict which patients with stage II CRC benefit from adjuvant chemotherapy.³⁹ In fact, the prediction accuracy of the two validated assays – Oncotype Dx and ColoPrint – has been shown to be very poor for identifying high-risk patients (22% and 22-26%, respectively).^{10,40,41} As a result, none of these tests are recommended for clinical use by the NCCN or ESMO clinical practice guidelines, and none are approved by the US Food and Drug Administration.

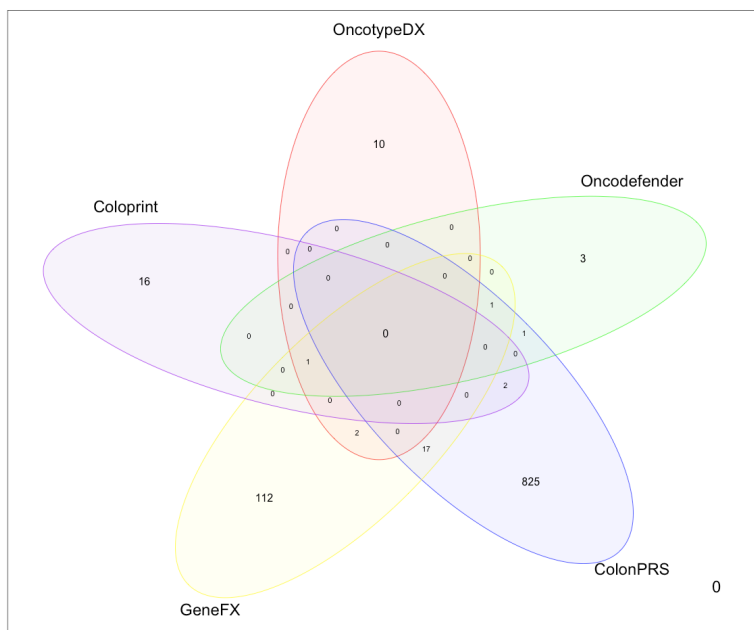


Figure 1. Overlap in gene predictors between commercialized recurrence prediction assays.

Ensemble learning

While the generation of molecular data through genomic, transcriptomic, epigenomic, and proteomic sequencing becomes more feasible and affordable, the role of this data in prognostic prediction remains unrealized. Additionally, previously constructed and validated predictive models may rely upon technology that becomes outdated as newer modalities arise. For example, microarray sequencing has largely been replaced with whole transcriptome sequencing (RNA-seq), yet all the commercialized recurrence prediction assays discussed above are based on microarray or qRT-PCR expression levels. Previous work has demonstrated that integrating clinical and molecular data can significantly improve prediction performance.⁴² Despite the additive effect of integrating molecular data with clinicopathologic data, Yuan et al. found that gains in performance were limited, with clinicopathologic characteristics remaining the most informative features.

The ability to individualize prognosis and treatment for patients hinges on the effective use of genomic data. To promote progress towards the development of precision medicine, The National Cancer Institute (NCI) and the Dialogue on Reverse Engineering Assessment and Methods (DREAM) project held a competition to predict drug sensitivity across 53 breast cancer cell lines using multi-omics data.⁴³ The two top performing entries utilized nonlinear models and integrated data through multi-view learning and ensemble learning, respectively. In multi-view learning, the original data sets (e.g. gene expression microarray) can be turned into multiple knowledge-enhanced data representations, also known as views, and integrated into a single model. Ensemble learning also allows the integration of prediction models, and this strategy was shown to enhance prediction performance in a previous DREAM project effort.⁴⁴

An ensemble is a collection of prediction models that are combined in some manner (e.g. majority vote or weighted averaging). Different ensemble learning techniques are differentiated by how the ensemble members, or base learners, are selected and how training data is partitioned to train these base learners. One of these methods – stacked generalization - was famously used in the two top performing entries of the Netflix Prize and it remains a successful technique in many winning data science competition entries. In stacked generalization, also known as stacking, a diverse set of base learners is trained using cross-validation.⁴⁵ The predictions of the hold-out sets are integrated into a second training dataset that is then used to train a new classifier that is called a meta-learner (Figure 2). As long as the number of base learners is polynomial with regard to the sample size, this framework guarantees that the ensemble will perform at least as well as the top performing base learner.⁴⁶

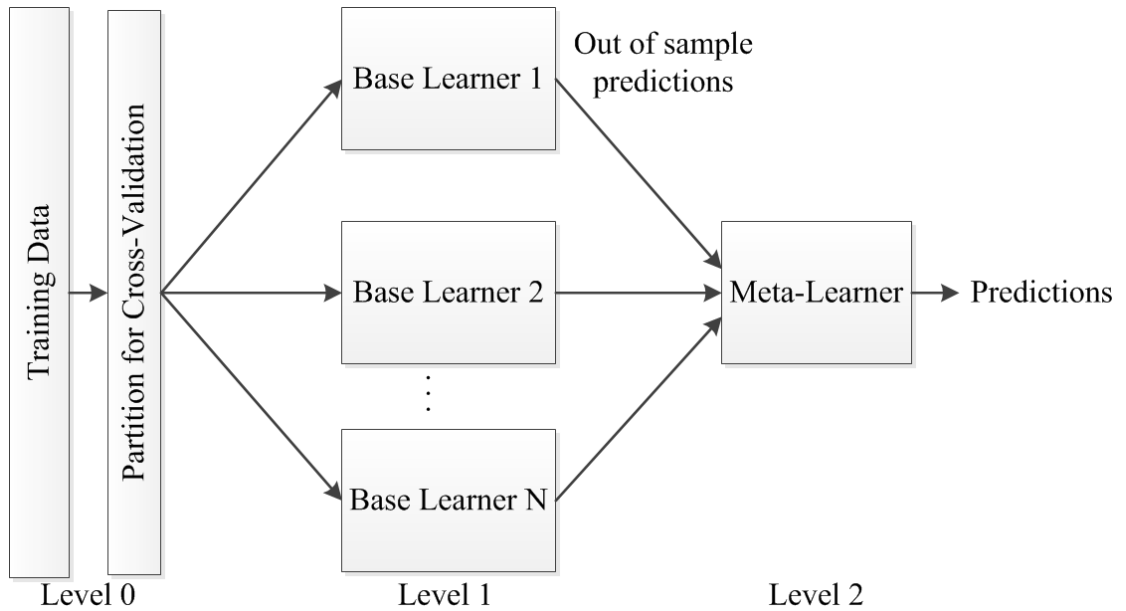


Figure 2. Framework for stacking.

Based on the insights from data science competitions and the machine learning literature, we propose a multiple-view multiple-learner framework to improve recurrence prediction in stage II/III CRC. The problem is formulated as a binary classification problem, and we use a supervised learning approach. Our hypothesis is that integration of models and multiple data views (clinical, transcriptomic, discretized, gene set, and network-based views) through stacking will generate a robust predictive framework that enhances therapeutic decision-making in CRC. We first determine whether predictions made using clinical and microarray data can be improved through stacking. Next, we generate multiple views of the molecular data in order to encompass prior biologic knowledge. After the views are generated we evaluate whether these additional views improve the prediction performance of the stacking framework and systematically evaluate model performance.

CHAPTER 2

CREATING AN ENSEMBLE FRAMEWORK

Introduction

To determine if an ensemble framework provides an additive benefit in the prediction of CRC recurrence we assembled a large dataset from publically available microarray data. We utilize a normalization methodology that allows for individual accrual of samples, and employ feature selection methodology that is informed by results from the Neural Information Processing Systems (NIPS) feature selection challenge. Finally, we curate a diverse set of base learners using stacked generalization and assess the effect on prediction performance when trained on clinical vs. microarray vs. integrated clinical and microarray datasets.

Data curation and normalization

Data from six studies were curated from the Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>): GSE14333, GSE17538, GSE26906, GSE33113, GSE37892, and GSE39582.^{16,38,47-50} Studies were included if they contained patients with stage II and/or stage III CRC with clinical annotation that included the following features: age, sex, stage, recurrence status, and recurrence time. This set was further limited by microarray platform to include only those studies that utilized Affymetrix HG-U133 Plus 2.0 Array (Affymetrix, Santa Clara, CA). Previous studies indicate that 3-year DFS status is an effective surrogate for 5-year OS, so recurrence status at 3 years was chosen as the binary outcome of the classification task.⁵¹ Disease free survival for all stage II and stage III patients prior to selection is shown in figure 3.

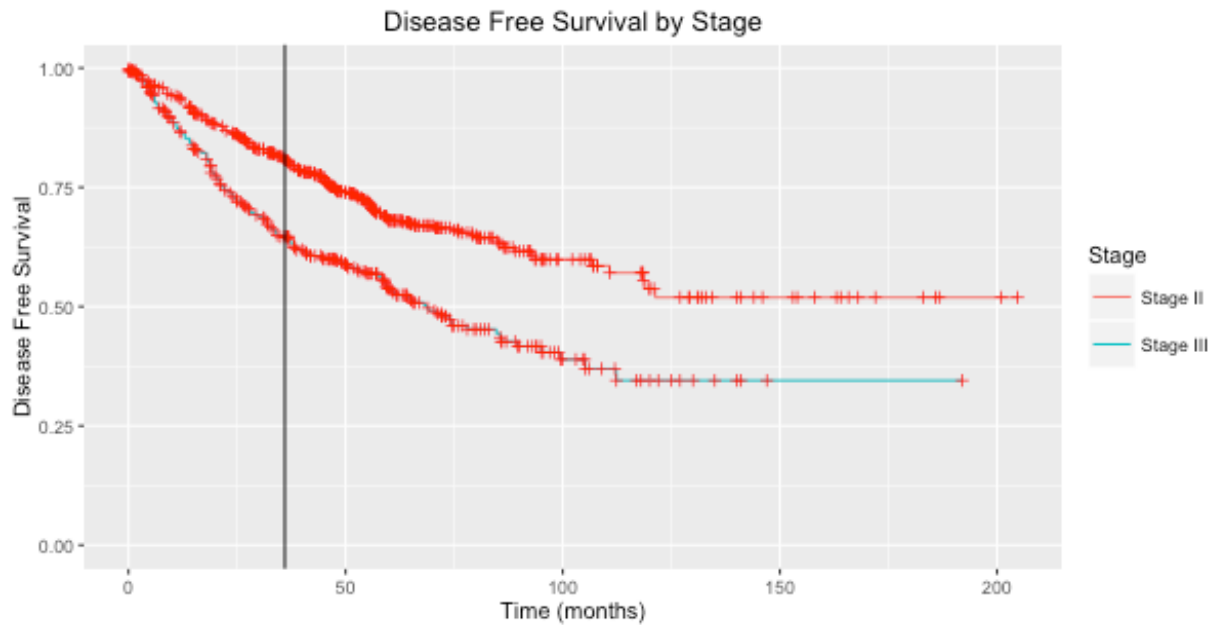


Figure 3. Disease free survival by stage. Vertical black line indicates 36 months.

Patients were included if they met the following criteria: no recurrence with length of follow up of at least 3 years, or recurrence between 1 month and 36 months after resection. Patients were excluded if they recurred within 1 month as that was an event that could include misdiagnosis/incorrect staging. Table 2 contains a summary of patient characteristics for each dataset. While patient characteristics were similar across data sets for the most part, GSE14333 contained significantly more recurrence events.

	GSE14333	GSE17538	GSE26906	GSE33113	GSE37892	GSE39582
N	58	101	72	68	109	370
% Male	53.4%	49.5%	56.9%	54.4%	47.7%	46.8%
Mean age (SD)	68.6 (12.9)	63.2 (13.4)	67.7 (12.5)	67.9 (12.8)	67.9 (13.2)	67.3 (13.2)
Stage 2	32	44	72	68	60	199
Stage 3	26	57	0	0	49	171
Recurrence Events (Stage II/ Stage III)	49 (30/19)	27 (9/18)	7 (7/0)	18 (18/0)	32 (6/26)	114 (42/72)

Table 2. Patient characteristics by GEO dataset

Training and validation sets were partitioned geographically in an attempt to increase the ability to assess the generalizability of the resulting predictive framework. Specifically, the samples were partitioned in an approximately 80:20 stratified split such that both sets contained data generated in the USA as well as Europe or Australia. The training set (n=624) was comprised of GSE39582, GSE33113, GSE37892, and the Moffitt Cancer Center patients from GSE17538, while the validation set (n=154) was comprised of GSE14333, GSE26906 and the Vanderbilt University Medical Center patients from GSE17538 (table 3). Recurrence events affected 30.4% of patients in the training set and 37.0% of patients in the validation set.

	Training set	Validation set
N	624	154
% Male	48.2%	53.9%
Mean age (SD)	67.0 (13.3)	67.2 (12.5)
Stage 2	362	113
Stage 3	262	41
Recurrence Events (Stage II/Stage III)	190 (75/115)	57 (37/20)

Table 3. Training and validation set patient characteristics.

A principle drawback of many published molecular recurrence signatures is that the validation occurs on an entire cohort rather than on individual patients. This means that for an individual to be evaluated, their molecular data would need to be normalized within a cohort of patients. In order to predict the recurrence status of an individual patient, we utilize frozen robust multiarray analysis (fRMA). Frozen RMA is a novel method of microarray preprocessing that allows an individual array to be processed instead of being dependent on multiple arrays being analyzed simultaneously.⁵²⁻⁵⁴ By using fRMA we can generate the gene expression features for each individual sample in both the training and the validation sets individually instead of by batch. McCall et al. determined that fRMA outperformed RMA in the context of analyzing

multiple batches, and so we utilized this technique without further batch normalization methods.

Data preprocessing steps are summarized in figure 4.

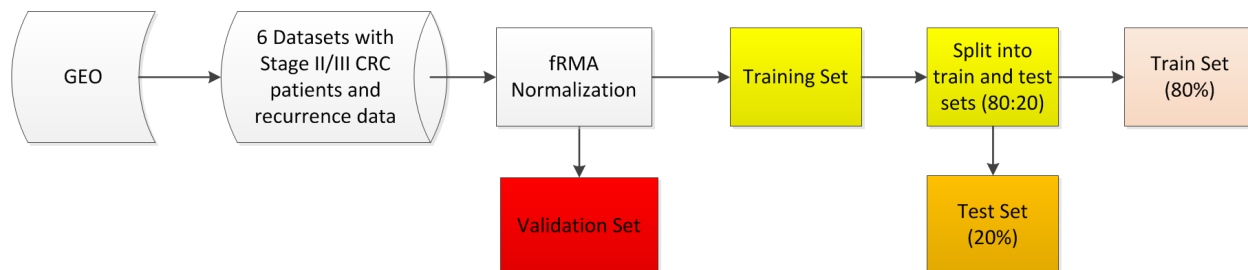


Figure 4. Data preprocessing pipeline.

Feature selection methodology

The high dimensionality of molecular data makes prediction of clinical outcomes difficult, as the number of features dramatically exceeds the number of samples. In this study, we have curated a large data set comprised of 778 patients, yet an Affymetrix HG-U133 Plus 2.0 Array contains over 50,000 probesets. Reducing the feature space theoretically decreases the risk of over fitting the data while also making the task computationally manageable, but there are myriad ways to do so. The 2003 NIPS feature selection challenge addressed this issue by organizing a competition to improve approaches to classification problems with high dimensional data.⁵⁵ The top performing entry utilized simple univariate significance tests to reduce the feature space to a few hundred and then performed classification using methods based on Bayesian neural network learning. This approach demonstrated that simple feature selection methods work well with sophisticated classification approaches, and motivated our choice of feature selection strategy.

To reduce the feature space to a manageable size we first mapped probesets to genes. When multiple probesets represented the same gene the average of the probesets was used as the

feature value, and the resulting dataset had greater than 20,000 features. Those probesets that represented multiple genes were excluded. A univariate logistic regression model was constructed for each feature in the training test, and those features with significant Wald test ($p < 0.05$) were retained. Next, the pre-selected features were limited to a feature set that was equivalent in size to the number of recurrence events in the training set by selecting those with lowest p values. Finally, regularization was performed using L1 penalized log partial likelihood (LASSO) to determine the final set of features for each molecular view.⁵⁶ The R package “glmnet” was used to perform LASSO, with the penalty parameter λ chosen after fivefold cross-validation on the training set.⁵⁷

Predictive power of clinical vs. microarray data

A diverse set of base learners was trained on each data set using ten repeats of tenfold cross-validation. Base learners were curated for diversity by underlying algorithm (e.g. linear vs. nonlinear) and tuning parameters, and models were trained using the “caret” R package.⁵⁸ Model parameters were tuned using grid search over 10 random values for each model parameter (See Appendix A for the full list of models and parameters). In total, 27 base learners were trained on clinical and microarray data. Prior to training each model, the training dataset was partitioned into 10 disjoint sets, such that a sample in one partition is not in any other. Each of the base learners is then trained on 9 of those blocks, and predictions are made in the 1 block not used in training. At the end of model training and parameter selection, a dataset of predictions made on the 10 hold out sets comprises the training set for our meta-learner. While we repeat cross validation ten times, we use only the first repetition to create this new meta-learner training set. Since recurrences occur only in 1/3 of patients in this dataset, we optimize models on a metric

that is suited to class-imbalanced data: Area Under the ROC Curve (AUC). An AUC of 1 indicates perfect prediction, whereas an AUC of 0.5 indicates that test results are equivalent to a random guess. To determine the performance of a model, predictions will be made using the testing dataset and then an ROC curve plot will be created. Classification thresholds will be determined for each model by calculating Youden’s index and selecting the threshold with maximum difference between sensitivity and specificity.⁵⁹

The hold-out set predictions for each base learner are then utilized to train the meta-learner. Logistic regression is a typical meta-learner in a stacking framework, although non-linear models can also be used. We selected and evaluated a diverse group of meta-learner models to compare against the typical classifier used. After evaluation on the testing set, the base learners are then trained on all training data (training and testing datasets) and performance evaluated on the validation set (Figure 5).

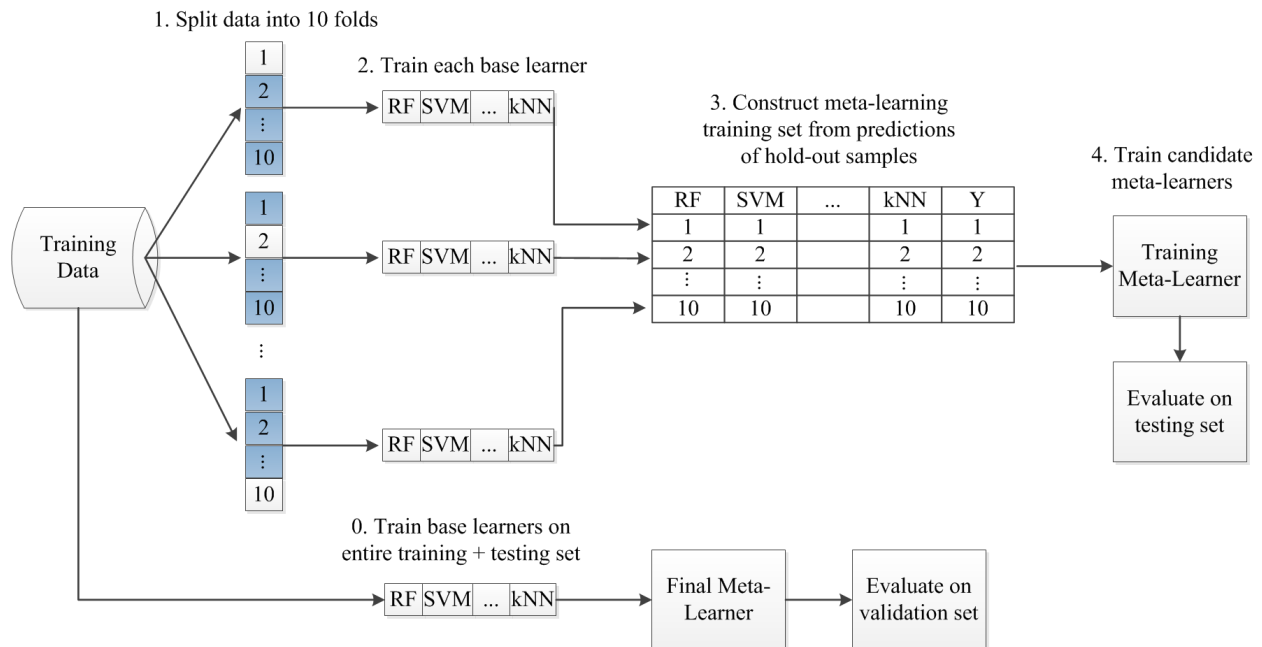


Figure 5. Stacked generalization training diagram. The training (1 – 4) and validation (0) processes are outlined in this diagram. RF = random forest, SVM = support vector machines, kNN = k-Nearest Neighbors, Y = binary outcomes of the data.

In our study, the training performance of microarray trained models was significantly better compared to models trained on clinical data (Paired Wilcoxon signed rank test, $p = 1.49 \times 10^{-8}$) (Figures 6 and 7).

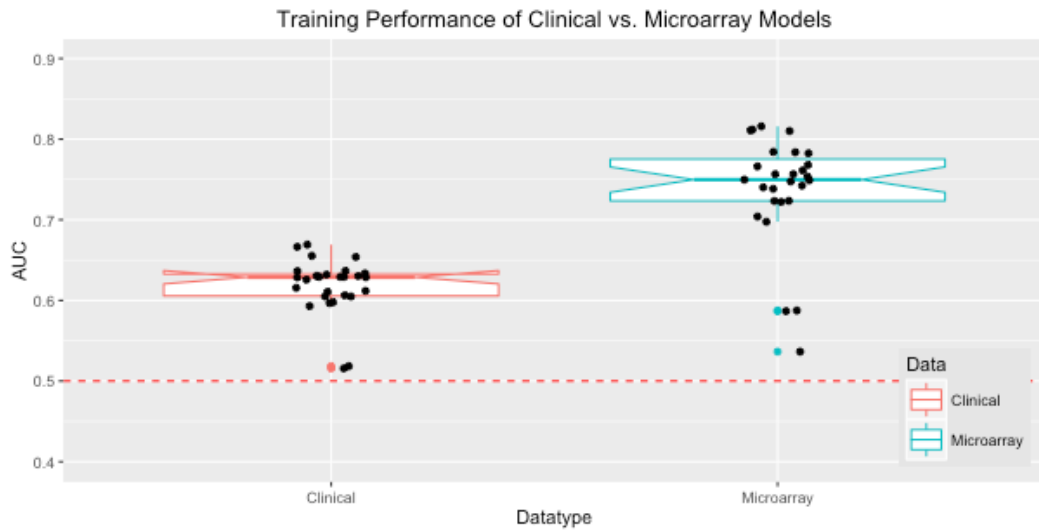


Figure 6. Training performance of clinical vs. microarray models

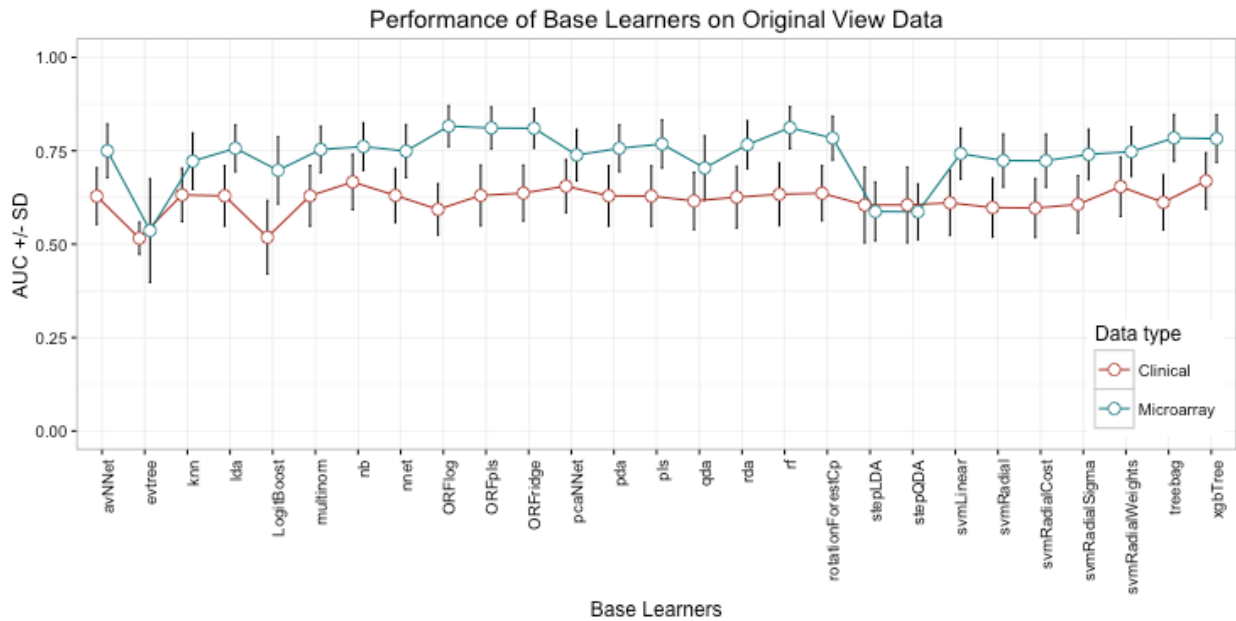


Figure 7. Performance of individual base learners on original view data. SD = standard deviation. The median training AUC for predictive models was 0.629 (SD 0.035) for those trained on clinical data vs. 0.75 (SD 0.068) for those trained on microarray data after feature selection.

Model performance was more homogenous between clinical models than between microarray models, as evidenced by the smaller standard deviation (0.035 vs. 0.068). The top performing clinical model was trained using the eXtreme Gradient Boosting (xgbTree) algorithm and had a training performance of 0.67. For the microarray data, models trained on random forest (rf) based models performed markedly better than others. Review of the model training performances confirmed that non-linear classifiers (e.g. rf, support vector machines with radial kernels) often outperform linear classifiers (e.g. linear discriminant analysis, partial least squares), and that ensemble methods (e.g. rf) can also enhance performance.

These findings contrast with those of Yuan et al., who found that clinical variables were the most informative data source across four cancer cohorts (renal clear cell carcinoma, ovarian serous cystadenocarcinoma, glioblastoma multiforme, and lung squamous cell carcinoma).⁴² The difference in conclusion is most likely not attributable to differences in the clinical features between the datasets, as the data used by Yuan et al. was very similar, and instead may indicate that the current staging paradigm in CRC is poorly suited to determine risk of recurrence (Table 4). Additionally, these findings also contrast with the Cancer Genome Atlas (TCGA) analysis demonstrating that molecular subtypes of CRC cohorts did not correlate with clinical phenotypes, including patient survival.⁶⁰

	Age	Sex	Stage	Grade	Performance Status
CRC	+	+	+		
KIRC	+	+	+	+	
OV	+		+	+	
GBM	+	+			+
LUSC	+	+	+		

Table 4. Clinical data features. KIRC = renal clear cell carcinoma, OV = ovarian serous cystadenocarcinoma, GBM = glioblastoma multiforme, LUSC = lung squamous cell carcinoma

Next, we determined if performance could be improved by stacking clinical or microarray models. First, we eliminated the worse performing models for clinical data (evTree, LogitBoost) and for microarray data (evtree, stepLDA, stepQDA), as they did not consistently have an AUC greater than 0.5. Next, we trained the level 2 stacker using a curated list of meta-models. Logistic regression stackers were initially considered for the level 2 meta-learners, but due to poor performance we focused on four methods with good performance on training data: naïve bayes (nb), rf, k-Nearest Neighbors (knn), and partial least squares (pls). We compared all results against the best performing base learner from level 1, a strategy also known as selectBest. To perform ensemble model building we utilized the R package “caretEnsemble” to build level 2 meta-learners shown in figure 2. The results of these ensembles on the clinical data are shown in table 5.

	Training AUC (SD)	Testing AUC
selectBest (level 1)	0.6692 (0.075)	0.6034
nb meta-learner (level 2)	0.6619 (0.08)	0.6613
rf meta-learner (level 2)	0.6471 (0.11)	0.6265
knn meta-learner (level 2)	0.7077 (0.085)	0.6212
pls meta-learner (level 2)	0.6619 (0.092)	0.6095

Table 5. Ensemble vs. selectBest performance on clinical data. glm = logistic regression.

These results show that while training performance is similar between learners at all levels, the performance on testing data is notably better with level 2 meta-learners.

For molecular data, the performance is somewhat less impressive (Table 6). The best level 1 base learner (Oblique random forest with logistic regression discriminative node model - ORFlog) does have a higher AUC on the testing data. Comparison of the ROC curves between

the best base learner and the rf meta-learner reveals no significant difference ($p = 0.1602$).⁶¹ Thus, while models trained on molecular data perform markedly better than those trained on clinical data, the use of an ensemble framework recapitulates the performance of the best base learner without necessarily exceeding it.

	Training AUC (SD)	Testing AUC
selectBest (level 1)	0.8159 (0.055)	0.8032
nb meta-learner (level 2)	0.7830 (0.059)	0.7677
rf meta-learner (level 2)	0.8216 (0.043)	0.7690
knn meta-learner (level 2)	0.8061 (0.055)	0.7341
pls meta-learner (level 2)	0.8384 (0.061)	0.7616

Table 6. Ensemble vs. selectBest performance on microarray data.

In this chapter we have described how the data was curated and normalized, how features were selected in the microarray dataset, and how model training was conducted. We have demonstrated clearly that molecular data is superior to clinical data in the prediction of CRC recurrence. Also, we demonstrated that stacking leads to better predictions from clinical data and equivalent predictions with microarray data. In the next chapter, we will explore different methods for incorporating biological knowledge into this predictive framework.

CHAPTER 3

CREATING A MULTIPLE-VIEW MULTIPLE-LEARNER FRAMEWORK

Introduction

While molecular data clearly outperforms clinical data in recurrence prediction, the original expression values do not leverage any expert knowledge or biological insights into the mechanisms of tumorigenesis and recurrence. One of the key insights from the drug sensitivity prediction DREAM challenge was that the application of prior knowledge (e.g. biological pathway knowledge) improved prediction performance.⁴³ To build upon this insight we engineered four molecular views that integrate outside information, and then incorporated the models trained on these views into our ensemble framework.

Discretized view

To determine if the relationship between tumor and non-tumor (normal adjacent or adenomatous tissue) samples could aid in recurrence prediction, 31 samples from non-tumor tissue (GSE33113 – $n = 6$, GSE17538 – $n = 6$, GSE39582 – $n = 19$) were fRMA normalized. Each value in the training, testing and validation sets was then compared against the distribution of non-tumor values for each gene and given one of three values based on the following algorithm:

Algorithm *Discretize*(n, k, M)

Input: n samples, k genes, $n \times k$ matrix of expression values M

Output: $n \times k$ matrix of discretized values D

1. For each sample i
2. For each gene j
3. Compare M_{ij} against distribution of non-tumor sample expression values

4. If $M_{ij} \leq 1$ SD below mean of non-tumor expression, then $D_{ij} = -1$
5. If $M_{ij} \geq 1$ SD above mean of non-tumor expression, then $D_{ij} = 1$
6. Else, $D_{ij} = 0$
7. **endfor**
8. **endfor**
9. **return** D_{ij}

Gene set view

Compiled gene sets from the Molecular Signatures Database (MSigDB) were utilized to incorporate biological pathway knowledge into the recurrence prediction framework. The MSigDB canonical pathways (CP) collection contains 1330 gene sets curated from nine online databases (http://software.broadinstitute.org/gsea/msigdb/collection_details.jsp#CP). To construct this view, the following algorithm was used:

Algorithm: *Geneset*(n, k, j, M)

Input: n samples, k genes, j gene sets, $n \times k$ matrix of expression values M

Output: $n \times j$ matrix of gene set values G

1. For each sample i
2. For each gene set j
3. G_{ij} equals the average of expression values for each gene in gene set j
4. **endfor**
5. **endfor**
6. **return** G_{ij}

Network propagation view

Network-based stratification has been used to effectively subtype multiple tumor types using mutation data from TCGA patients.⁶² This approach successfully created subtypes predictive of clinical outcomes, and so we have adapted it to utilize gene expression data. The overall goal is to propagate an individual patient's gene expression values onto a protein-protein interaction (PPI) network through the use of network propagation. The PPI was curated using iRefWeb, and genes were limited to those that were included in the gene expression dataset.⁶³

The start probabilities are defined by normalized gene expression values. First, the gene expression data is z-score normalized by gene, and then it is rescaled so that the values of each gene range from 0 to 1. Each sample is then scaled so that the sum of all gene expression values for a given patient is equal to 1. Finally, a vector of gene weights is calculated through the use of random walk with restart, which is of the form:

$$p^{t+1} = (1 - r)Wp^t + rp^0.$$

In the random walk with restart equation, r is restart probability, W is the column-normalized adjacency matrix derived from the PPI, and p^t is a vector of gene weights such that the i^{th} value is the probability of being at gene i at time t . The start probability p^0 is initialized through the normalization steps outlined above, and we set r to 0.5. This algorithm is run iteratively until convergence, which is defined as

$$\sum_{i=1}^n |p_i^{t+1} - p_i^t| \leq 10^{-6},$$

where p_i^t is the probability for gene i at the t^{th} iteration. After convergence of all samples, the rows of the resulting matrix are quantile normalized.

Network-Expression-Mutation signature view

Thus far the molecular views have leveraged knowledge of non-tumor tissue expression patterns as well as gene set and PPI structure. To incorporate knowledge gained in previous signatures and also in the discovery of cancer driver mutations in CRC we modified our previous Network, Expression, and Mutation (NEM) signature. To create the updated NEM signature we utilize random walk with restart as defined above, but with start probabilities defined as follows:

$$p_i^0 = \frac{s_i}{2 \sum_{i=1}^n s_i} + \frac{m_i}{2 \sum_{i=1}^n m_i}$$

where s_i is defined as the number of CRC gene expression signatures that include gene i , and m_i is defined as the number of times a mutation is identified as cancer driver or tumor suppressor mutation in CRC.⁶⁴ Seven of the eight signatures used to create the previous NEM signature were utilized, and the signatures from the five commercialized assays were included as well. The signature from Smith et al. was removed since that dataset was used to train and validate the models. Mutations were curated from intOGene (<https://www.intogen.org/>), COSMIC (<http://cancer.sanger.ac.uk>), and MutSigCV mutation analysis.⁶⁵⁻⁶⁷

After convergence was reached, individual gene scores were assessed by construction of 1000 sets of randomly permuted start probabilities and 1000 sets of random scores. Local p values were obtained for each gene by comparing the actual score to the random scores. Global p values were obtained by comparing the actual score to random scores for all genes. Genes were selected if both the local and global p values were significant. The resulting updated NEM signature contained 547 genes compared to the 487 genes in the original NEM, with 122 genes overlapping.

Performance of the multiple-learner multiple-view framework

After views were generated for the training and testing data sets, feature selection was applied to each view as outlined in chapter 1. The resulting performance is shown in figures 8 and 9. Pairwise comparisons using a paired Wilcoxon signed rank test with a bonferroni correction (threshold $p < 0.005$) demonstrate that three views were superior to the microarray view during training: discretized ($p = 6.35 \times 10^{-5}$), network propagation ($p < 0.003$), and NEM ($p = 4.75 \times 10^{-5}$). The gene set view performed worse than microarray data ($p = 1.49 \times 10^{-7}$) as well as the remaining three views. The training profile of individual base learners revealed that

similarly to the microarray view, models using the evtree, stepLDA and stepQDA algorithms performed poorly. Additionally, the network view trained xgbTree model performed significantly worse compared to other xgbTree models. These poor performing models were not used for ensemble building as their AUCs were not consistently greater than 0.5.

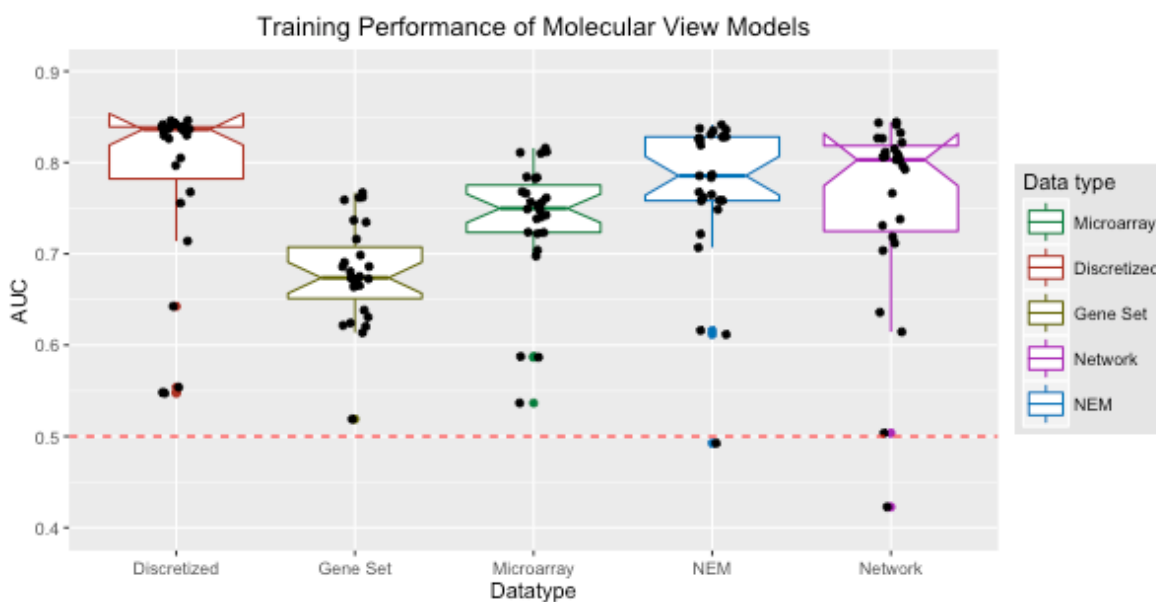


Figure 8. Training performance of molecular view models

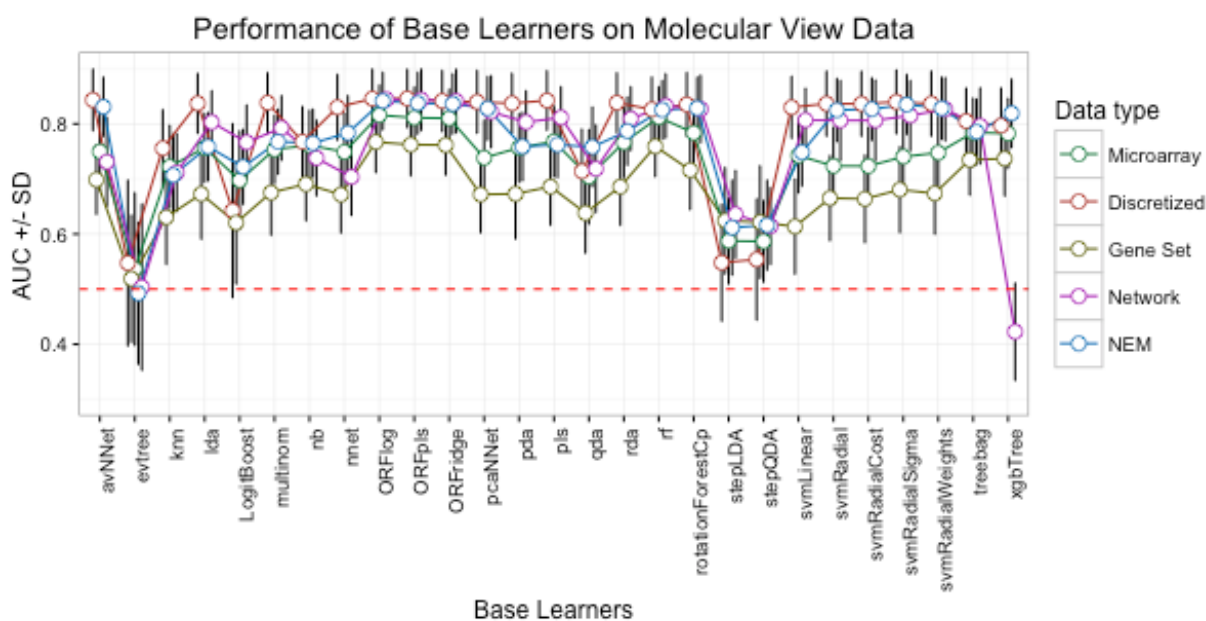


Figure 9. Training performance of base learners on molecular view data

Just as the worst performing models for the molecular views were similar to those that performed poorly with microarray data, the best performing models were all random forest based. In fact, many of the top performing models during training were variants of oblique random forests. While random forests are decision tree ensembles generated from orthogonal trees, oblique random forests are built from trees that split using linear discriminative models, such as lda, ridge regression and logistic regression.⁶⁸ The success of these models confirms the importance of including non-linear models in the prediction framework.

Next, we performed stacked generalization on the models within each view to determine if this strategy could enhance prediction performance. Additionally, we combined all views together and performed stacked generalization on this combined view. Performance results from predictions on the testing set data indicates that stacking can improve AUC compared to selectBest for the discretized and network views, but not the gene set and NEM views (Table 7). When the predictions from all models from all views are utilized, however, the performance is improved, with the random forest level 2 meta-learner exhibiting the best overall AUC.

	Discretized View	Gene set View	Network View	NEM View	Combined View
selectBest (level 1)	0.7824	0.8034	0.7394	0.8022	
nb meta-learner (level 2)	0.7739	0.7889	0.7659	0.7910	0.7913
rf meta-learner (level 2)	0.7829	0.7834	0.7154	0.7909	0.8146
knn meta-learner (level 2)	0.7740	0.7601	0.7550	0.7853	0.7976
pls meta-learner (level 2)	0.7965	0.7561	0.7228	0.7818	0.7521
glm meta-learner (level 3)	0.7904	0.7748	0.7243	0.7864	0.7913

Table 7. Ensemble vs. selectBest performance on testing data.

For the molecular views, we also performed stacked generalization on the level 2 meta-learners by training a logistic regression classifier on the predictions from the meta-learners. The performance of these level 3 meta-learners was similar to the level 2 meta-learners.

To further characterize the performance of these models we constructed a lift curve of the combined view ensemble models (Figure 10). This plot shows the number of events captured (y-axis) as a function of the number of samples tested (x-axis). The left side of the gray shaded region represents the ideal model, while the right side of the region is a line going from bottom left to top right that is equivalent to random guessing. While performance for most of the stackers is equivalent, we see that the pls and especially the level 3 glm stacker (called final in the plot) are worse.

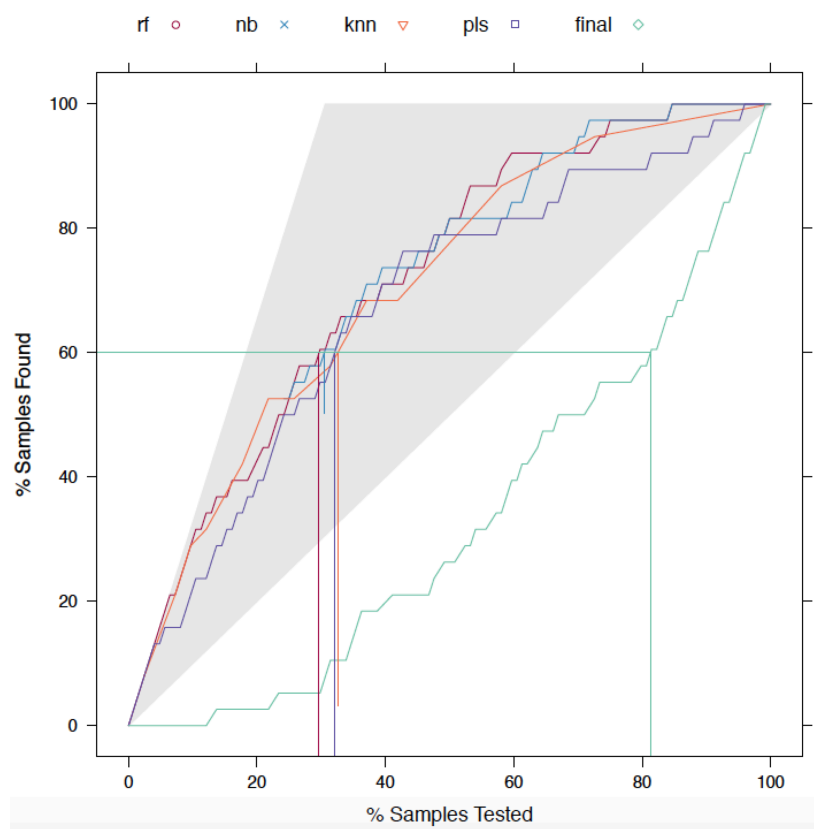


Figure 10. Lift curve for ensemble methods. The level 3 meta-learner “final” is a glm that stacks all 4 level 2 meta-learners.

We analyzed the probability calibration by constructing a calibration curve for the three models that appeared to perform the best based on the lift curve (Figure 11). This plot allows for visualization of the relationship between probability predictions and observed events. A properly calibrated model would track the diagonal line coursing from the bottom left to the top right of the plot. This visualization reveals that the nb meta-learner is poorly calibrated, and knn stacker appears to have the best calibration. Detailed comparison of performance metrics between the rf and knn stackers revealed the best overall performance with rf (Table 8). Of note, the selectBest model was the base learner that had the highest AUC on the testing data, rather than the training data, compared to all other base learners.

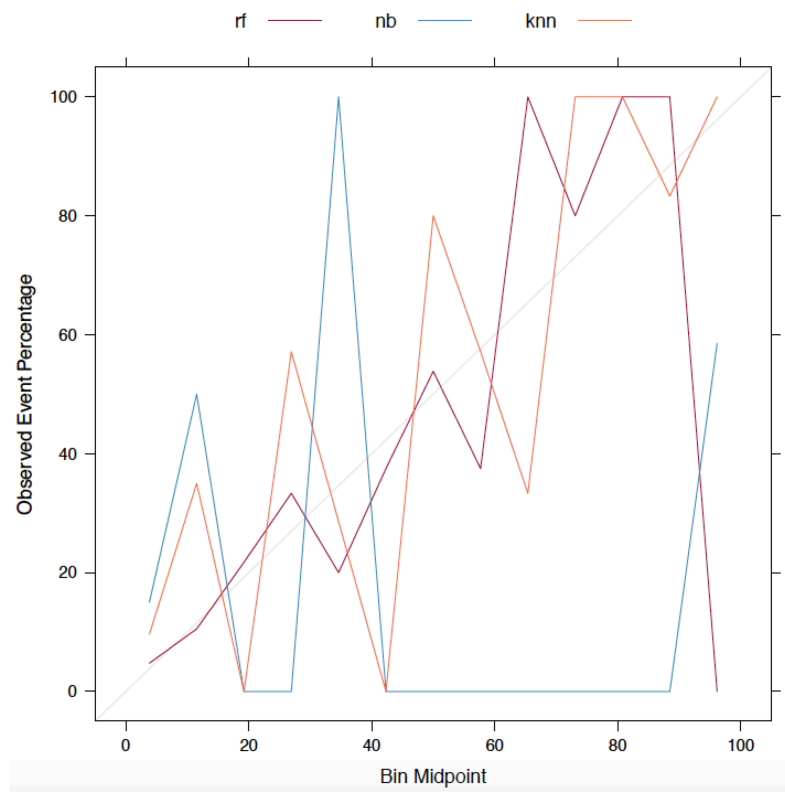


Figure 11. Probability calibration plot for ensemble models.

	rf	knn	selectBest
AUC	0.8146	0.7976	0.8161
Sensitivity	0.9211	0.6842	0.7105
Specificity	0.5698	0.7674	0.7791
Accuracy	0.6774	0.7419	0.7581
Positive Predictive Value	0.4861	0.5652	0.5870
Negative Predictive Value	0.9423	0.8462	0.8590
F Measure	0.6364	0.6190	0.6429

Table 8. Detailed performance metrics of ensemble vs. selectBest methods on testing set. selectBest represents the top performing base learner on testing data: rf trained on microarray data.

Detailed metrics reveal that the rf meta-learner performs similarly to selectBest, although accuracy is worse. The F measure and AUC are similar to selectBest, however. The F measure is the harmonic mean of the model precision (positive predictive value) and recall (sensitivity), and it is an important metric in binary classification. Based on these findings, we hypothesize that the ideal framework to predict recurrence status is to utilize a rf stacker to integrate the base learner predictions from all view models.

Ensemble performance on the validation set was diminished in terms of AUC, although the best base learner did have a higher AUC than in the testing set (Table 9). Interestingly, while the best base learner on the testing data was a microarray data trained rf model, the best base learner on the validation set was a bagged classification and regression tree (treebag) model trained on NEM view data. The results in table 9 demonstrate that the rf ensemble performed slightly superiorly or equivalently to the best base learner in all metrics aside from AUC, and actually exceeded the performance demonstrated in the test set in terms of F measure and accuracy.

	rf	knn	selectBest
AUC	0.7860	0.7502	0.8388
Sensitivity	0.9474	0.7719	0.9298
Specificity	0.6495	0.6804	0.6598
Accuracy	0.7597	0.7143	0.7597
Positive Predictive Value	0.6136	0.5867	0.6163
Negative Predictive Value	0.9545	0.8354	0.9412
F Measure	0.7448	0.6667	0.7393

Table 9. Detailed performance metrics of ensemble vs. selectBest methods on validation set. selectBest represents the top performing base learner on validation data: treebag trained on NEM data.

Classifications made by each of these three methods were used to plot DFS for the validation set patients, and compared to stratification by stage (Figure 12).

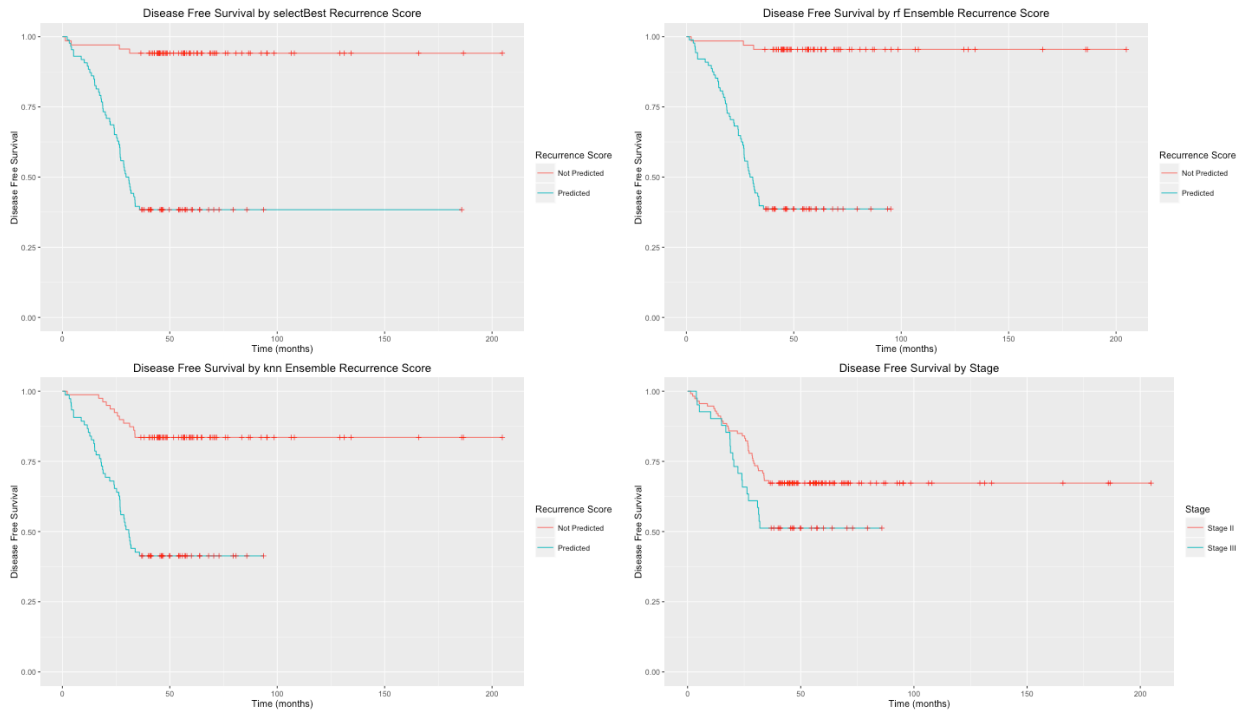


Figure 12. DFS by predictive model compared to stage.

The rf ensemble, knn ensemble and selectBest model all significantly separated a high and low risk recurrence population (log-rank test, $p < 0.001$ for each model), while stage failed to

discriminate against these two populations ($p = 0.0529$). The rf stacker performed similarly to the selectBest model, as expected. To further determine the utility of this prediction framework, we calculated Oncotype Dx scores by modifying the algorithm from Clark-Langone et al. to utilize microarray expression data.⁶⁹

Algorithm *Oncotype*(n, k, M)

Input: n samples, k genes, $n \times k$ matrix of expression values M

Output: n normalized recurrence scores

1. For each sample i
2. Calculate mean of reference genes: *UBB, ATP5E, PGK1, GPX1, VDAC2*
3. For genes *BGN, MYC, FAP, GADD45B, INHBA, MKI67, MYBL2*
4. Normalized value \leftarrow Divide gene value by mean of reference genes
5. Gene value \leftarrow mean of reference genes – normalized value + 10
6. **endfor**
7. **endfor**
8. Subtract the lowest gene value from every gene value for every sample
9. unscaled recurrence score $\leftarrow 0.1263 * (\text{mean of gene values for } BGN, FAP, INHBA) -$
 $.3158 * (\text{mean of gene values for } MYBL2, MKI67, MYC) + .3406 * \text{gene value for } GADD45B$
10. normalized recurrence score $\leftarrow (\text{unscaled recurrence score} + 0.3) * 44.16$
11. Subtract the lowest recurrence score from the score for each sample
12. **return** normalized recurrence scores

After applying this algorithm to the microarray validation dataset, scores ranged from 0 to 77.84. According to the Oncotype Dx algorithm, patients are stratified into three risk groups by score: low risk if score < 30 , intermediate risk if score is 30 to 40, and high risk if score is > 40 . To compare these scores to the performance of our classification models, the recurrence score was turned into a binary predictor such that low risk corresponded to 0 and intermediate or high risk corresponded to 1. While a previous external validation of CRC prognostic genomic predictors found that Oncotype Dx demonstrated good performance in determining patients with poor prognosis, these findings are not reproduced in our validation cohort (Figure 13).⁷⁰ Analysis of just the stage II patients also failed to demonstrate any significant discrimination by recurrence score.

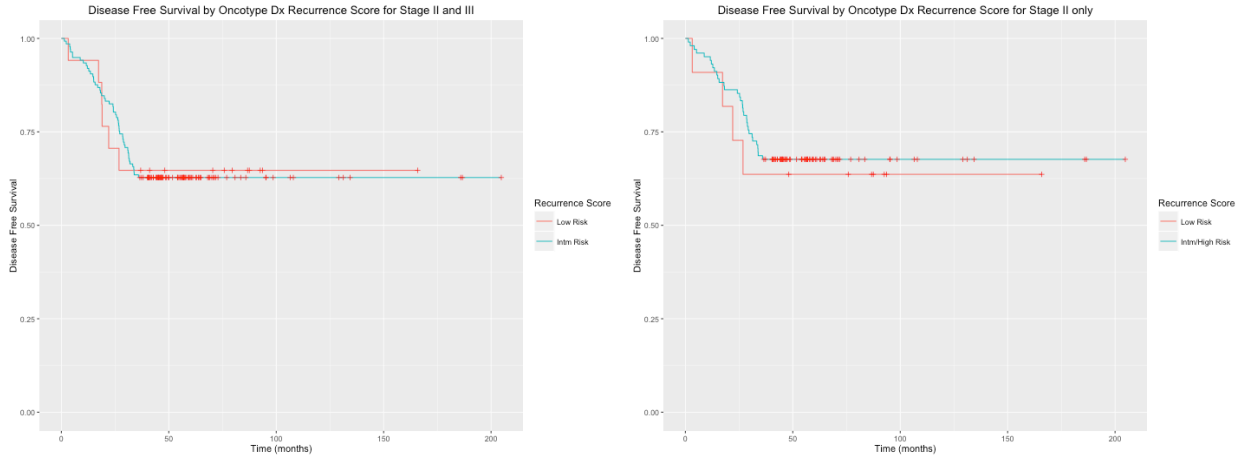


Figure 13. DFS by Oncotype Dx recurrence score. The plot on the left contains patients from the validation set with stage II and stage III disease, while the plot on the right contains only patients with stage II disease.

We have shown that an ensemble framework built from multiple views of molecular data using a diverse set of base learners can improve the performance of a predictive framework in comparison to the single best base learner. The importance of the molecular views is evident when the variable importance of the two top ensemble methods is analyzed (Figure 14). We see that the NEM view was particularly important for the both ensembles, and that other views, such as the discretized view, contributed more to prediction than either of the original views.

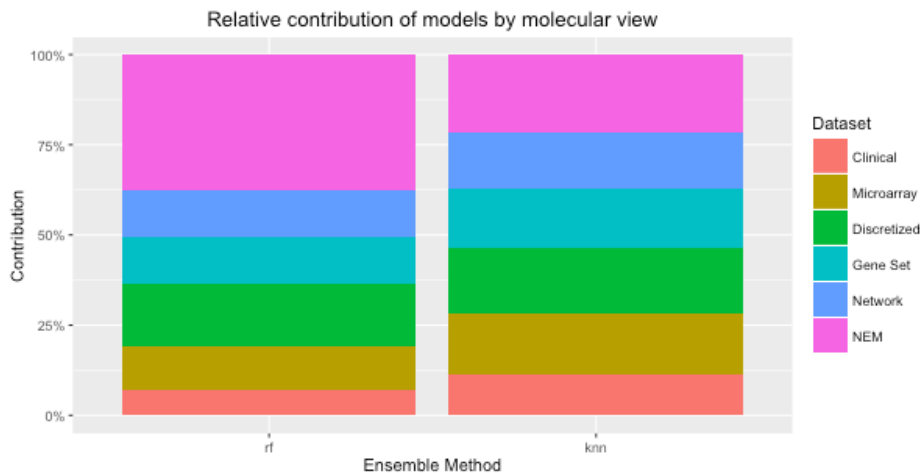


Figure 14. Relative contribution of models by molecular view.

CHAPTER 4

SUMMARY

Conclusions

This work focuses on the construction of an ensemble framework to predict CRC recurrence. After demonstrating that molecular data is more useful for predicting CRC recurrence, we proposed and implemented a multiple-view multiple-learner framework. The benefits of this approach include:

Incorporation of prior knowledge: While molecular data can improve prediction over clinicopathologic features, this improvement has not been translated into better predictions in the clinical setting. We demonstrate the views incorporating information from non-tumor tissue gene expression patterns, gene set structure, PPI structure, previously curated molecular signatures, and identified tumor suppressor/driver mutations improves prediction.

Use of non-linear models: The top performing base learners and ensemble models were overwhelmingly non-linear, with random forest and related models performing particularly well. This confirms findings from the data science competitions that partially motivated this work.

Feasible integration of multiple data types: The stacking framework we utilized can easily accommodate other types of molecular data (e.g. proteomic or methylation) data, and so we have provided a scalable and flexible framework.

Finally, this work represents the first effort to use ensemble learning to predict CRC recurrence. The ability of ensemble learning to improve predictions, or at least perform as well as the single best base learner has been shown to apply to the problem of CRC recurrence

prediction. Going forward, it is our hope that the community of researchers focused on this task shares details of their predictive models and collaborates to see if ensemble learning can be used to further improve prediction performance. These results also highlight the importance of data sharing, as the prediction performance still needs much improvement if the goals of precision medicine are to be realized. Finally, while our classification system can outperform staging, Oncotype Dx, we have not demonstrated that the high-risk population will benefit from the standard adjuvant chemotherapy regimens. Further improvements in predicting patients at high-risk for recurrence will lead to the need to accurately identify the optimal therapeutic regimen.

APPENDIX A

BASE LEARNERS

Model	Shorthand	Type	Packages	Parameters
Bagged CART	treebag	Dual Use	ipred, plyr,, e1071	none
Boosted Logistic Regression	LogitBoost	Classification	caTools	nIter
Model Averaged Neural Network	avNNet	Dual Use	nnet	size, decay, bag
Linear Discriminant Analysis	lda	Classification	MASS	None
Linear Discriminant Analysis with Stepwise Feature Selection	stepLDA	Classification	klaR, MASS	maxvar, direction
Naive Bayes	nb	Classification	klaR	fL, usekernel, adjust
Neural Network	nnet	Dual Use	nnet	size, decay
Neural Networks with Feature Extraction	pcaNNet	Dual Use	nnet	size, decay
Oblique Random Forest	ORFlog	Classification	obliqueRF	mtry
Oblique Random Forest	ORFpls	Classification	obliqueRF	mtry
Oblique Random Forest	ORFridge	Classification	obliqueRF	mtry
Partial Least Squares	pls	Dual Use	pls	ncomp
Penalized Discriminant Analysis	pda	Classification	mda	lambda
Penalized Multinomial Regression	multinom	Classification	nnet	decay
Quadratic Discriminant Analysis	qda	Classification	MASS	None
Quadratic Discriminant Analysis with Stepwise Feature Selection	stepQDA	Classification	klaR, MASS	maxvar, direction
Random Forest	rf	Dual Use	randomForest	mtry
Rotation Forest	rotationForestCp	Classification	rpart, plyr, rotationForest	K, L, cp
Support Vector Machines with Linear Kernel	svmLinear	Dual Use	kernlab	C
Support Vector Machines with Radial Basis Function Kernel	svmRadial, svmRadialCost	Dual Use	kernlab	sigma, C

Model	Shorthand	Type	Packages	Parameters
Support Vector Machines with Class Weights	svmRadialWeights	Classification	kernlab	sigma, C, Weight
Tree Models from Genetic Algorithms	evtree	Dual Use	evtree	alpha
eXtreme Gradient Boosting	xgbTree	Dual Use	xgboost, plyr	nrounds, max_depth, eta, gamma, colsample_bytr ee, min_child_wei ght
Regularized Discriminant Analysis	rda	Classification	klaR	gamma, lambda

Base learners are listed in the tables above. Shorthand indicates the variable name for the model in the “caret” package. Type indicates the type of problem a model can be used for: classification, regression or both (Dual Use). The “caret” package acts as a wrapper around models from many different packages, and the specific package for each model is listed in the Packages column. Finally, the model parameters are listed in the Parameters column.

APPENDIX B

The following R code was used to generate the training, testing and validation data sets.

```
rm(list=ls())
setwd("/scratch/castelja/CEL/")
library(glmnet)
library(foreach)
library(aod)
library(pROC)
library(caret)
library("igraph")
library("preprocessCore")
library(randomForest)
library("gsubfn")
source("/dors/bioinfo/castelja/functions.R")

#Processing GSE33113, GSE14333, GSE17538
#GSE39582, GSE37892, and GSE26906 omitted
#Aggregate into one table
Complete<-
  rbind(GSE14333,GSE17538,GSE26906,GSE33113,GSE37892,GSE39582)

#Filter out based on selection criteria
ones<-grep(TRUE,Complete[, "status"]=="1")
zeroes<-grep(TRUE,Complete[, "status"]=="0")
surv_ones<-grep(TRUE,as.numeric(Complete[ones, "time"])<=36)
surv_zeroes<-grep(TRUE,as.numeric(Complete[zeroes, "time"])>=36)
indices<-sort(c(ones[surv_ones],zeroes[surv_zeroes]))
Complete<-Complete[indices,]

# Load matrices with previously frma normalize expressing data
load("/scratch/castelja/CEL/test_expression.RData")
load("/scratch/castelja/CEL/discretized.Rdata")
#Identify samples for validation set
trainIndex<-
  c(grep("Ludwig",Complete[,6]),grep("Vanderbilt",Complete[,6]),
    grep("STA",Complete[,6]),grep("IPC",Complete[,6]),
    grep("CAL",Complete[,6]),grep("LAR",Complete[,6]),
    grep("BER",Complete[,6]))
```

```

trainclass<-Complete[-trainIndex,4]
testclass<-Complete[trainIndex,4]

#Clinical data sets
clin.train<-Complete[-trainIndex,]
clin.val<-Complete[trainIndex,]

#Microarray data sets
micro.train<-Test_expression[-trainIndex,]
micro.val<-Test_expression[trainIndex,]

#Geneset data sets
#create geneset score data sets based on average value of gene
#set expression in the canonical pathways gene set in MSigDB
cp<-read.csv("/dors/bioinfo/castelja/CRC/
  transposed_cp.csv",header=TRUE)
cp<-as.matrix(cp)
cp<-cp[-1,]

geneset.train<-geneset(cp,micro.train)
geneset.val<-geneset(cp,micro.val)

#Discretized data sets
micro.train.binary<-discretized[-trainIndex,]
micro.val.binary<-discretized[trainIndex,]

#Network propagation data sets
#Create W - column normalized adjacency matrix
library(igraph)
iRef <- read.delim("/dors/bioinfo/castelja/CRC/iRef.net",
  header=FALSE)
network<-graph.data.frame(iRef,directed=FALSE)
adjMatrix <- get.adjacency(network)
adjMatrix <- as.matrix(adjMatrix)
de <- apply(adjMatrix,2,sum)
#normalizing by column
W <- t(t(adjMatrix)/de)

#Micro - convert to include only nodes contained in ppi
common<-intersect(colnames(W),colnames(micro.train))

```

```

#find indexes in micro
indices<-find_index(common,colnames(micro.train))
#truncate micro down to size
micro.rw.train <-micro.train[,indices]
micro.rw.train <-micro.rw.train[,sort(colnames(micro.rw.train))]
micro.rw.val <-micro.val[,indices]
micro.rw.val <-micro.rw.val[,sort(colnames(micro.rw.train))]
diff<-setdiff(colnames(W),colnames(micro.rw.train))
#find indexes in W
indices<-find_index(diff,colnames(W))
#truncate W down to size
W<-W[,-indices]
W<-W[-indices,]
W<-W[,sort(colnames(W))]
W<-W[sort(rownames(W)),]

r<-0.5
valRow<-rownames(micro.rw.val)

#Scale the sets by Z score normalization
micro.rw.val<-apply(micro.rw.val,2,scale)
rownames(micro.rw.val)<-valRow

#Rescale to be between 0 and 1
micro.rw.val<-apply(micro.rw.val,2,function(x) ((x-
min(x))/diff(range(x))))
rownames(micro.rw.val)<-valRow

#Ensure each row sums to 1
micro.rw.val<-sum_to_one(micro.rw.val)
network.val<-vary_restart(micro.rw.val,W,r)

#NEM view
nem<-read.csv("new_NEM.csv")
common<-intersect(colnames(micro.train),as.character(nem[,1]))
indices<-find_index(common,colnames(micro.train))
micro.train.NEM<-micro.train[,indices]
micro.val.NEM<-micro.val[,indices]

```

```

save(trainclass,testclass,clin.train,clin.val,micro.train,micro.
val,trainIndex,micro.train.binary,micro.val.binary,geneset.train
,geneset.val,W,micro.rw.train,micro.rw.val,micro.train.NEM,micro
.val.NEM,file="/scratch/castelja/CEL/datasets.RData",
compress=FALSE)

#save network data separately
trainRow<-rownames(micro.rw.train)

#Scale the sets by Z score normalization
micro.rw.train<-apply(micro.rw.train,2,scale)
rownames(micro.rw.train)<-trainRow

#Rescale to be between 0 and 1
micro.rw.train<-apply(micro.rw.train,2,function(x) ((x-
min(x))/diff(range(x))))
rownames(micro.rw.train)<-trainRow

#Ensure each row sums to 1
micro.rw.train<-sum_to_one(micro.rw.train)
network.train<-vary_restart(micro.rw.train,W,r)
save(network.train,network.val,file=
"/scratch/castelja/CEL/network.RData")

```

APPENDIX C

The following R code contains functions for feature selection and network propagation.

```
#Univariate feature selection with logistic regression
feature_select<-function(x.train,y.train){
  pvalue.cutoff=0.05
  top=sum(y.train)
  feature.pvalue <- c()
  feature.name <- c()
  feature.col <- c()
  #for (j in 1:ncol(x.train))
  result <- foreach
(j=1:ncol(x.train),.errorhandling='remove',.combine=rbind)
%dopar%
  {
    feature <- colnames(x.train)[j]
    x <- x.train[,j]
    if (length(which(table(x)> 0.8*length(x)))>0) # discard the
flat values, e.g. zeros for RNAseq and miRNAseq
    {
      stop()
    }
    logistic<-glm(y.train~x,family=binomial(link="logit"))
    p.value <-
as.numeric(unlist(wald.test(b=coef(logistic),Sigma=vcov(logistic
),Terms=2))[[13]]))
    list(p.value, feature, j)
  }

  feature.pvalue <- unlist(result[,1])
  print(paste("Total number of valid features (after removal of
potential flat records):", length(feature.pvalue)))
  feature.name <- unlist(result[,2])
  feature.col <- unlist(result[,3])

  names(feature.pvalue)=c()
  names(feature.name)=c()
  names(feature.col)=c()

  q.value <- p.adjust(feature.pvalue, method="fdr")
}
```

```

for (i in 1:10/10)
{
  print(paste("qvalue <=", i, ":", length(which(q.value<=i))))
}
col.sig <- which( feature.pvalue < pvalue.cutoff)
print(paste("Significant records: p-value < 0.05:",
length(col.sig)))

name.sig <- feature.name[col.sig]
pvalue.sig <- feature.pvalue[col.sig]
#qvalue.sig <- q.value[col.sig]

col.retain <- feature.col[col.sig]
# Only keep the top significant ones if there are too many
if (length(col.retain)> top)
{
  col.retain <- col.retain[head(sort(pvalue.sig,
index.return=T)$ix, n=top)]
}
cols.include<-col.retain
if (length(cols.include)==0)
{
  stop("No feature passed the univariate cox screen: exit.")
}

print(paste("After univariate cox screen, features remain:",
length(cols.include)))
return(cols.include)
}

```

```

#Create network propagation view using vary_restart and
netwalker

```

```

vary_restart<-function(data,W,r){
P<-data
for(i in 1:length(data[,1])){
  print(paste("Sample",i,"of",length(data[,1])))
  P[i,]<-netwalker(data[i,],W,r)
}
#quantile normalize rows to ensure that each patient follows

```

```

same distribution
network<-t(normalize.quantiles(t(P)))
rownames(network)<-rownames(P)
colnames(network)<-colnames(P)
return(network)
}

#Network propagation via random walk with restart
netwalker <- function(p0,W,r){

  #Set up for the first iteration
  pt <- p0
  #First iteration
  pt1 <- (1-r)*(W%%pt)+r*p0

  #Iterate until convergence
  threshold<-1e-6
  iter<-0
  while(sum(abs(pt1-pt))>threshold){
    pt <- pt1
    pt1 <- (1-r)*(W%%pt)+r*p0
    iter<-iter+1
  }
  print(paste("converged in",iter,"iterations"))
  return(pt1)
}

```


APPENDIX D

The following R code is used to train the base learners on training data.

```
#Pass parameters to R script
options(echo=TRUE)
args<-commandArgs(trailingOnly=TRUE)
print(args)
#Subset #
i<-as.numeric(args[1])
#View #
j<-as.numeric(args[2])
#Model #
k<-as.numeric(args[3])

clinfilename<-paste("frma_",i,"_",j,"_",k,"_clin.RData",sep="")
microfilename<-
  paste("frma_",i,"_",j,"_",k,"_micro.RData",sep="")
binfilename<-paste("frma_",i,"_",j,"_",k,"_binary.RData",sep="")
gsfilename<-paste("frma_",i,"_",j,"_",k,"_geneset.RData",sep="")
networkfilename<-
  paste("frma_",i,"_",j,"_",k,"_network.RData",sep="")
NEMfilename<-paste("frma_",i,"_",j,"_",k,"_NEM.RData",sep="")

rule<-"ROC"

#Curated list of 27 base learners
methods<-
  c("treebag","LogitBoost","avNNet","lda","stepLDA","nb","nnet","p
  caNNet","ORFlog","ORFpls",
  "ORFridge","pls","pda","multinom","qda","stepQDA","rf","rda","ro
  tationForestCp","svmRadialWeights",
  "svmLinear","svmRadial","svmRadialCost","svmRadialSigma","evtree
  ","xgbTree","knn")

filename<-
  paste("/dors/bioinfo/castelja/CRC/final/iter_",i,"_frma.RData",s
  ep="")
load(filename)
method<-methods[k]
```

```

#10 x 10 fold cross validation
ctrl<-trainControl(method="repeatedcv",
  number=10,
  repeats=10,
  savePredictions="final",
  classProbs=TRUE,
  index=resamples,
  summaryFunction=twoClassSummary)

if(j==1){
#Clinical View
model_clin<-train(clin.train,trainclass,
  method=method,
  trControl=ctrl,
  #family=binomial,
  #verbose=FALSE,
  tuneLength=10,
  metric=rule)
save(model_clin,file=clinfilename)
} else if(j==2){
#Microarray Original View
model_micro<-train(micro.train.fs,trainclass,
  method=method,
  trControl=ctrl,
  #verbose=FALSE,
  preprocess=c("center","scale"),
  tuneLength=10,
  metric=rule)
save(model_micro,file=microfilename)
} else if(j==3){
#Microarray Discretized View
binary_micro<-train(micro.train.binary,trainclass,
  method=method,
  trControl=ctrl,
  #verbose=FALSE,
  tuneLength=10,
  metric=rule)
save(binary_micro,file=binfilename)
} else if(j==4){
#Microarray Geneset View

```

```

geneset_micro<-train(geneset.train,trainclass,
                    method=method,
                    trControl=ctrl,
                    #verbose=FALSE,
                    preprocess=c("center","scale"),
                    tuneLength=10,
                    metric=rule)
save(geneset_micro,file=gsfilename)
} else if(j==5){
#Microarray Network View
network_micro<-train(network.train,trainclass,
                    method=method,
                    trControl=ctrl,
                    #verbose=FALSE,
                    preprocess=c("center","scale"),
                    tuneLength=10,
                    metric=rule)
save(network_micro,file=networkfilename)
} else{
NEM_micro<-train(micro.train.NEM.fs,trainclass,
                method=method,
                trControl=ctrl,
                #verbose=FALSE,
                preprocess=c("center","scale"),
                tuneLength=10,
                metric=rule)
save(NEM_micro,file=NEMfilename)
}

```

APPENDIX E

The following R code was used to train the ensemble model.

```
create_list<-function(i,j,type){
  methods<-
  c("treebag","LogitBoost","avNNet","lda","stepLDA","nb","nnet","p
  caNNet","ORFlog","ORFpls","ORFridge","pls","pda","multinom","qda
  ","stepQDA","rf","rda","rotationForestCp","svmRadialWeights","sv
  mLinear","svmRadial","svmRadialCost","svmRadialSigma","evtree","
  xgbTree","knn")
  seq<-1:length(methods)
  out_list<-list()
  null_list<-c()
  for(k in 1:length(seq)){
    filename<-
    paste("frma_",i,"_",j,"_",k,"_",type,".RData",sep="")
    print(filename)
    if(file.exists(filename)){
      print("exists")
      e1 = new.env()
      invisible(lapply(filename, load, envir = e1))
      my_list = as.list(e1)
      out_list[[k]]<-my_list[[1]]
      method<-methods[k]
      names(out_list)[k]<-paste(method,type,sep="_")
    } else{
      print("Does not exist")
      null_list<-c(null_list,k)
    }
  }
  if(length(null_list)>0){
    out_list[null_list]<-NULL
    print(paste("These methods did not
  exist:",methods[null_list]))
  } else{
    print("All methods exist")
  }
  class(out_list)<- "caretList"
  return(out_list)
}
```

```

#Set up model lists for each view
clin_list<-create_list(i,1,"clin")
micro_list<-create_list(i,2,"micro")

geneset_list<-create_list(i,4,"geneset")
binary_list<-create_list(i,3,"binary")
network_list<-create_list(i,5,"network")
NEM_list<-create_list(i,6,"NEM")
meta_list<-append(clin_list,micro_list)
meta_list<-append(meta_list,binary_list)
meta_list<-append(meta_list,geneset_list)
meta_list<-append(meta_list,network_list)
meta_list<-append(meta_list,NEM_list)
names(meta_list)<-c(names(clin_list),names(micro_list),
names(binary_list),names(geneset_list),names(network_list),
names(NEM_list))

ens.train<-c()
lapply(meta_list,function(x){
  tmp<-setorderv(x$pred,c("Resample","rowIndex"))
  tmp<-setorderv(tmp[grepl("Rep01",tmp$Resample),],c("rowIndex"))
  ens.train<-cbind(ens.train,tmp$Y)
})
)
colnames(ens.train)<-names(meta_list)
rownames(ens.train)<-rownames(clin.train)

resamples<-fold_generator(trainclass)
ens_control<-trainControl(
  method="cv",
  number=10,
  classProbs=TRUE,
  index=resamples,
  savePredictions="final",
  summaryFunction=twoClassSummary)

library("caretEnsemble")
model_list_big <- caretList(
  ens.train,ens.trainclass,

```

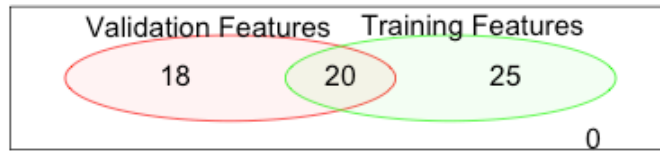
```
trControl=ens_control,  
metric="ROC",  
methodList=c("rf","knn","nb","pls"),  
tuneLength=10  
)  
modelCor(resamples(model_list_big))  
  
greedy_ensemble <- caretEnsemble(  
  model_list_big,  
  metric="ROC",  
  trControl=trainControl(  
    method="cv",  
    number=10,  
    summaryFunction=twoClassSummary,  
    classProbs=TRUE  
  ))  
summary(greedy_ensemble)
```

APPENDIX F

Microarray Features



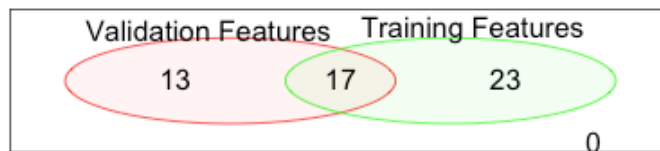
Discretized Features



Gene Set Features



Network Features



NEM Features



Signature overlap between validation and training data sets after univariate feature selection and LASSO.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA: a cancer journal for clinicians* 2015;65:5-29.
2. Nauta R, Stablein DM, Holyoke ED. Survival of patients with stage B2 colon carcinoma. The Gastrointestinal Tumor Study Group experience. *Archives of surgery* 1989;124:180-2.
3. Moertel CG, Fleming TR, Macdonald JS, et al. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *The New England journal of medicine* 1990;322:352-8.
4. Andre T, Boni C, Mounedji-Boudiaf L, et al. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *The New England journal of medicine* 2004;350:2343-51.
5. Moertel CG, Fleming TR, Macdonald JS, et al. Fluorouracil plus levamisole as effective adjuvant therapy after resection of stage III colon carcinoma: a final report. *Annals of internal medicine* 1995;122:321-6.
6. Quasar Collaborative G, Gray R, Barnwell J, et al. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. *Lancet* 2007;370:2020-9.
7. Sobrero A. Should adjuvant chemotherapy become standard treatment for patients with stage II colon cancer? For the proposal. *The Lancet Oncology* 2006;7:515-6.
8. Kohne CH. Should adjuvant chemotherapy become standard treatment for patients with stage II colon cancer? Against the proposal. *The Lancet Oncology* 2006;7:516-7.
9. O'Connell MJ, Lavery I, Yothers G, et al. Relationship between tumor gene expression and recurrence in four independent studies of patients with stage II/III colon cancer treated with surgery alone or surgery plus adjuvant fluorouracil plus leucovorin. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2010;28:3937-44.
10. Salazar R, Roepman P, Capella G, et al. Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2011;29:17-24.
11. Van Laar RK. An online gene expression assay for determining adjuvant therapy eligibility in patients with stage 2 or 3 colon cancer. *British journal of cancer* 2010;103:1852-7.
12. Lenehan PF, Boardman LA, Riegert-Johnson D, et al. Generation and external validation of a tumor-derived 5-gene prognostic signature for recurrence of lymph node-negative, invasive colorectal carcinoma. *Cancer* 2012;118:5234-44.
13. Kennedy RD, Bylesjo M, Kerr P, et al. Development and independent validation of a prognostic assay for stage II colon cancer using formalin-fixed paraffin-embedded tissue. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2011;29:4620-6.
14. Eschrich S, Yang I, Bloom G, et al. Molecular staging for survival prediction of colorectal cancer patients. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2005;23:3526-35.
15. Lin YH, Friederichs J, Black MA, et al. Multiple gene expression classifiers from different array platforms predict poor prognosis of colorectal cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2007;13:498-507.
16. Smith JJ, Deane NG, Wu F, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology* 2010;138:958-68.

17. Haller DG, Catalano PJ, Macdonald JS, et al. Phase III study of fluorouracil, leucovorin, and levamisole in high-risk stage II and III colon cancer: final report of Intergroup 0089. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2005;23:8671-8.
18. Mamounas E, Wieand S, Wolmark N, et al. Comparative efficacy of adjuvant chemotherapy in patients with Dukes' B versus Dukes' C colon cancer: results from four National Surgical Adjuvant Breast and Bowel Project adjuvant studies (C-01, C-02, C-03, and C-04). *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 1999;17:1349-55.
19. Wolmark N, Rockette H, Fisher B, et al. The benefit of leucovorin-modulated fluorouracil as postoperative adjuvant therapy for primary colon cancer: results from National Surgical Adjuvant Breast and Bowel Project protocol C-03. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 1993;11:1879-87.
20. Wolmark N, Rockette H, Mamounas E, et al. Clinical trial to assess the relative efficacy of fluorouracil and leucovorin, fluorouracil and levamisole, and fluorouracil, leucovorin, and levamisole in patients with Dukes' B and C carcinoma of the colon: results from National Surgical Adjuvant Breast and Bowel Project C-04. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 1999;17:3553-9.
21. Wolmark N, Rockette H, Wickerham DL, et al. Adjuvant therapy of Dukes' A, B, and C adenocarcinoma of the colon with portal-vein fluorouracil hepatic infusion: preliminary results of National Surgical Adjuvant Breast and Bowel Project Protocol C-02. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 1990;8:1466-75.
22. Efficacy of adjuvant fluorouracil and folinic acid in B2 colon cancer. International Multicentre Pooled Analysis of B2 Colon Cancer Trials (IMPACT B2) Investigators. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 1999;17:1356-63.
23. Marsoni S, International Multicenter Pooled Analysis of Colon Cancer Trials I. Efficacy of adjuvant fluorouracil and leucovorin in stage B2 and C colon cancer. International Multicenter Pooled Analysis of Colon Cancer Trials Investigators. *Seminars in oncology* 2001;28:14-9.
24. Schrag D, Rifas-Shiman S, Saltz L, Bach PB, Begg CB. Adjuvant chemotherapy use for Medicare beneficiaries with stage II colon cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2002;20:3999-4005.
25. Figueredo A, Charette ML, Maroun J, Brouwers MC, Zuraw L. Adjuvant therapy for stage II colon cancer: a systematic review from the Cancer Care Ontario Program in evidence-based care's gastrointestinal cancer disease site group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2004;22:3395-407.
26. Benson AB, 3rd, Schrag D, Somerfield MR, et al. American Society of Clinical Oncology recommendations on adjuvant chemotherapy for stage II colon cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2004;22:3408-19.
27. Andre T, de Gramont A, Vernerey D, et al. Adjuvant Fluorouracil, Leucovorin, and Oxaliplatin in Stage II to III Colon Cancer: Updated 10-Year Survival and Outcomes According to BRAF Mutation and Mismatch Repair Status of the MOSAIC Study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2015;33:4176-87.

28. Comparison of fluorouracil with additional levamisole, higher-dose folinic acid, or both, as adjuvant chemotherapy for colorectal cancer: a randomised trial. QUASAR Collaborative Group. *Lancet* 2000;355:1588-96.
29. O'Connor ES, Greenblatt DY, LoConte NK, et al. Adjuvant chemotherapy for stage II colon cancer with poor prognostic features. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2011;29:3381-8.
30. Chagpar R, Xing Y, Chiang YJ, et al. Adherence to stage-specific treatment guidelines for patients with colon cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2012;30:972-9.
31. Earle CC, Weiser MR, Ter Veer A, et al. Effect of lymph node retrieval rates on the utilization of adjuvant chemotherapy in stage II colon cancer. *Journal of surgical oncology* 2009;100:525-8.
32. Ribic CM, Sargent DJ, Moore MJ, et al. Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. *The New England journal of medicine* 2003;349:247-57.
33. Sargent DJ, Marsoni S, Monges G, et al. Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2010;28:3219-26.
34. Bandres E, Malumbres R, Cubedo E, et al. A gene signature of 8 genes could identify the risk of recurrence and progression in Dukes' B colon cancer patients. *Oncology reports* 2007;17:1089-94.
35. Barrier A, Roser F, Boelle PY, et al. Prognosis of stage II colon cancer by non-neoplastic mucosa gene expression profiling. *Oncogene* 2007;26:2642-8.
36. Barrier A, Boelle PY, Roser F, et al. Stage II colon cancer prognosis prediction by tumor gene expression profiling. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2006;24:4685-91.
37. Wang Y, Jatko T, Zhang Y, et al. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2004;22:1564-71.
38. Jorissen RN, Gibbs P, Christie M, et al. Metastasis-Associated Gene Expression Changes Predict Poor Outcomes in Patients with Dukes Stage B and C Colorectal Cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2009;15:7642-51.
39. Gao S, Tibiche C, Zou J, et al. Identification and Construction of Combinatory Cancer Hallmark-Based Gene Signature Sets to Predict Recurrence and Chemotherapy Benefit in Stage II Colorectal Cancer. *JAMA oncology* 2016;2:37-45.
40. Gray RG, Quirke P, Handley K, et al. Validation study of a quantitative multigene reverse transcriptase-polymerase chain reaction assay for assessment of recurrence risk in patients with stage II colon cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2011;29:4611-9.
41. Kopetz S, Taberero J, Rosenberg R, et al. Genomic classifier ColoPrint predicts recurrence in stage II colorectal cancer patients more accurately than clinical factors. *The oncologist* 2015;20:127-33.
42. Yuan Y, Van Allen EM, Omberg L, et al. Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nature biotechnology* 2014;32:644-52.

43. Costello JC, Heiser LM, Georgii E, et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nature biotechnology* 2014;32:1202-12.
44. Marbach D, Costello JC, Kuffner R, et al. Wisdom of crowds for robust gene network inference. *Nature methods* 2012;9:796-804.
45. Wolpert DH. Stacked generalization. *Neural Networks* 1992;5:241-59.
46. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Statistical applications in genetics and molecular biology* 2007;6:Article25.
47. Birnbaum DJ, Laibe S, Ferrari A, et al. Expression Profiles in Stage II Colon Cancer According to APC Gene Status. *Translational oncology* 2012;5:72-6.
48. de Sousa EMF, Colak S, Buikhuisen J, et al. Methylation of cancer-stem-cell-associated Wnt target genes predicts poor prognosis in colorectal cancer patients. *Cell stem cell* 2011;9:476-85.
49. Kemper K, Versloot M, Cameron K, et al. Mutations in the Ras-Raf Axis underlie the prognostic value of CD133 in colorectal cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2012;18:3132-41.
50. Marisa L, de Reynies A, Duval A, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS medicine* 2013;10:e1001453.
51. Sargent DJ, Wieand HS, Haller DG, et al. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2005;23:8664-70.
52. McCall MN, Jaffee HA, Irizarry RA. fRMA ST: frozen robust multiarray analysis for Affymetrix Exon and Gene ST arrays. *Bioinformatics* 2012;28:3153-4.
53. McCall MN, Irizarry RA. Thawing Frozen Robust Multi-array Analysis (fRMA). *BMC bioinformatics* 2011;12:369.
54. McCall MN, Bolstad BM, Irizarry RA. Frozen robust multiarray analysis (fRMA). *Biostatistics* 2010;11:242-53.
55. Guyon I GS, Ben-Hur A, and Dror G. *Result Analysis of the NIPS 2003 Feature Selection Challenge*: MIT Press; 2005.
56. R T. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society* 1996;58:267-88.
57. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of statistical software* 2010;33:1-22.
58. Max Kuhn. Contributions from Jed Wing SW, Andre Williams, Chris Keefer and Allan Engelhardt (2012). caret: Classification and Regression Training. R package version 5.15-044., <http://CRAN.R-project.org/package=caret>.
59. DMW P. Evaluation: From Precision, Recall and F-Score to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies* 2011;2:37-63.
60. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;487:330-7.
61. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837-45.
62. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. *Nature methods* 2013;10:1108-15.

63. Turner B, Razick S, Turinsky AL, et al. iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. *Database : the journal of biological databases and curation* 2010;2010:baq023.
64. Shi M, Beauchamp RD, Zhang B. A network-based gene expression signature informs prognosis and treatment for colorectal cancer patients. *PloS one* 2012;7:e41292.
65. Rubio-Perez C, Tamborero D, Schroeder MP, et al. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. *Cancer cell* 2015;27:382-96.
66. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic acids research* 2015;43:D805-11.
67. doi:10.7908/C1G44PPJ BITGDACMAMvBIOMaH.
68. Menze BH KM, Splitthoff DN, Koethe U, Hamprecht FA. *On Oblique Random Forests*: Springer Berlin Heidelberg; 2011.
69. Clark-Langone KM, Sangli C, Krishnakumar J, Watson D. Translating tumor biology into personalized treatment planning: analytical performance characteristics of the Oncotype DX Colon Cancer Assay. *BMC cancer* 2010;10:691.
70. Park YY, Lee SS, Lim JY, et al. Comparison of prognostic genomic predictors in colorectal cancer. *PloS one* 2013;8:e60778.