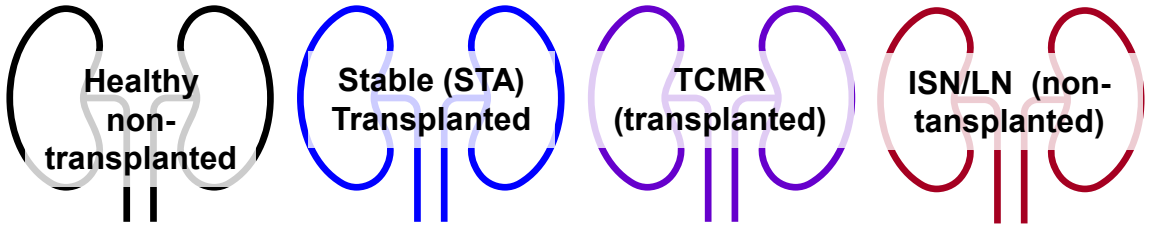
Caveats of Machine Learning Algorithms for Predicting Disease States in the Transplanted Kidney

Mitchell Ellison PhD, Yuchen Huang, and Parmjeet Randhawa MD
UPMC Department of Pathology

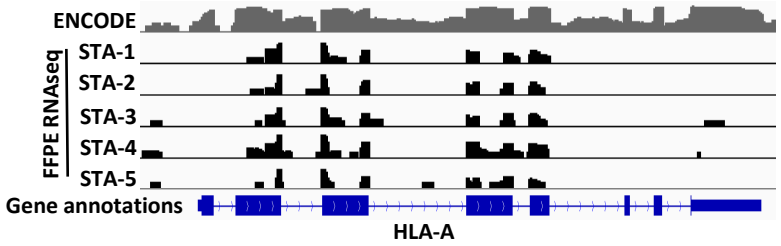
Background

- Kidney transplant is a successful medical procedure that has improved the quality of life in hundreds of thousands of patients with end-stage kidney disease.
- Transplanted kidney grafts must be monitored post-transplant for signs of immunological rejection such as T-cell mediated rejection (TCMR) or antibody mediated rejection (ABMR)
- Biopsy plays a key role in defining disease states and planning treatment but can only interrogate tissue at a structural level.
- Molecular methods for diagnosing TCMR have been described in the literature (MMDX^{R}), but the specificity of the underlying algorithms has not yet been rigorously examined.
- It is likely that molecular signatures in different disease states share common pathogenetic mechanisms, just as one histologic lesion can be associated with more than one pathologic diagnosis.
- Interstitial nephritis (ISN) and lupus nephritis (LN) are conditions that cause inflammation of the kidney, which may be diagnosed as TCMR by MMDX-like algorithms.



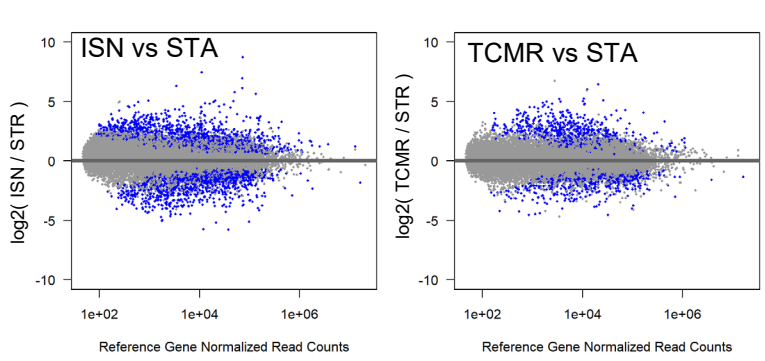
Methods

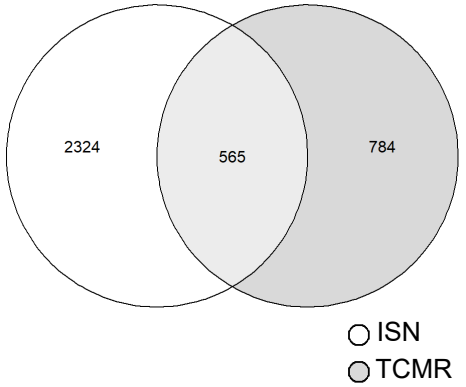
- Align RNAseq data to human genome and obtain per-gene read counts (STAR via HTC cluster).
- Perform QC comparing FFPE RNAseq to standard RNAseq data from ENCODE.
- Perform replicate QC.
- Perform differential expression analysis (DEseq2).



- Train binary classifiers to diagnose TCMR using the caret R package.
- Test predictive algorithms on previously published datasets.

Assessing overlap between differentially expressed gene sets



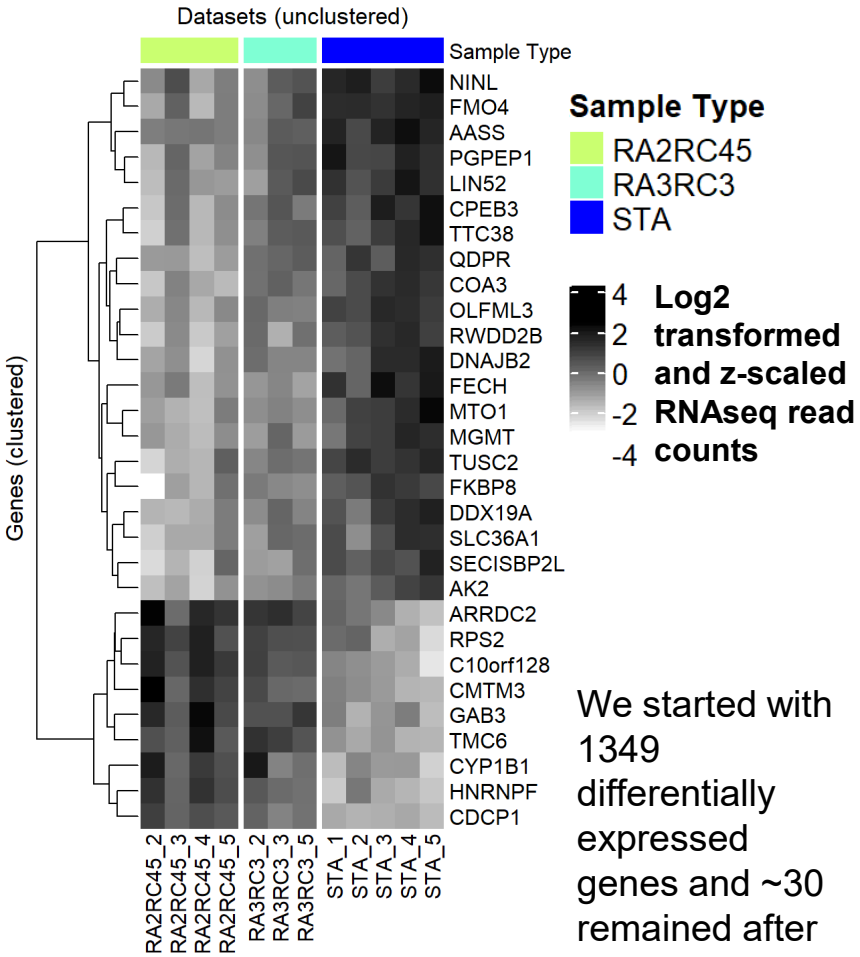


Differential expression analysis results determined using DEseq2 package in R. Blue points are statistically significant

Differentially expressed genes selected for use with TCMR classifier

Differentially expressed genes were filtered for

- FDR adjusted p-value >= 0.001
- log2 fold change standard error of <= 0.05
- log2 fold change of <= 1



We started with 1349 differentially expressed genes and ~30 remained after stringent filtering

Model Predictions

Sample Type	Dataset	Number of Samples	% Predicted as TCMR: NULL	KNN	RF	LDA	SVM
STA	This study	5	100	0	0	0	0
TCMR	This study	7	100	100	100	100	100
ISN	This study	5	100	80	80	100	80
Nephrectomy	GSE36059	8	100	37.5	37.5	75	50
TCMR	GSE36059	35	100	94.28571	97.14286	91.42857	94.28571
ABMR	GSE36059	65	100	86.15385	84.61538	83.07692	87.69231
TCMR+ABMR	GSE36059	22	100	90.90909	95.45455	86.36364	90.90909
Non-rejecting	GSE36059	281	100	61.92171	59.4306	67.97153	64.76868
Healthy	GSE127797	3	100	33.33333	33.33333	66.66667	33.33333
Lupus Nephritis	GSE127797	44	100	75	65.90909	77.27273	77.27273

ISN is misdiagnosed as TCMR in FFPE-RNAseq dataset

ABMR, mixed and non rejecting predicted as TCMR

Sample Type	Dataset	Number of Samples	% Predicted as ISN: NULL	KNN	RF	LDA	SVM
STA	This study	5	100	0	0	0	0
TCMR	This study	7	100	42.85714	42.85714	42.85714	42.85714
ISN	This study	5	100	100	100	100	100

TCMR is misdiagnosed as ISN in FFPE-RNAseq dataset

- Tissue labeled as ABMR in some public datasets is likely to be mixed ABMR-TCMR. Note that our TCMR classifier behaves identical on what the public datasets ABMR and ABMR+TCMR samples.
- Non-rejecting samples are a very heterogenous category and include biopsies with inflammation that is either incipient TCMR or reflects other injuries with a similar inflammatory state.
- Lupus nephritis frequently has a gene expression profile that overlaps TCMR.
- Even nephrectomy samples and healthy samples have some inflammation that is interpreted as TCMR by our algorithms.

Novelty – Algorithms for predicting TCMR had not been rigorously tested against disease states with similar molecular signatures to TCMR, thus their accuracy was overestimated. This study is the first to test TCMR predictors in ISN and LN datasets and demonstrate that these models (the commercial versions of which are extremely expensive) do very poorly in these contexts.

Importance – Identifying that TCMR algorithms have limited diagnostic accuracy for transplant kidney biopsies means that their predictions must be assessed in the context of the larger clinical picture. Meaning we must consider that other diseases that can excite a similar inflammatory response may be present and interpret the predictions accordingly.

Future Work – Building more accurate ML algorithms will require a dataset with larger numbers of non-rejection disease samples in order to identify TCMR-specific predictors for diagnosing TCMR and its mimics.