

Machine Learning Models Predict Renal Cell Carcinoma Status from Multiplatform Urine-based Metabolomics.

ASMS Conference,
June 2020

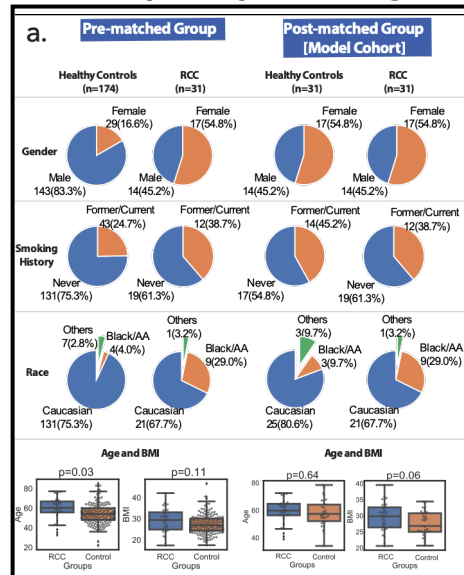
Olatomiwa O. Bifarin^{1,2}, David A. Gaul⁵, Rebecca Arnold³, John A. Petros^{3,4}, Facundo M. Fernández⁵, Arthur S. Edison^{1,2}

¹Complex Carbohydrate Research Center, University of Georgia, Athens, GA, 30602 ²Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, 30602 ³Department of Urology, Emory University, Atlanta, GA 30342 ⁴Atlanta VA Medical Center, Atlanta, GA 30033 ⁵School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta GA 30332 (USA)

1. Background

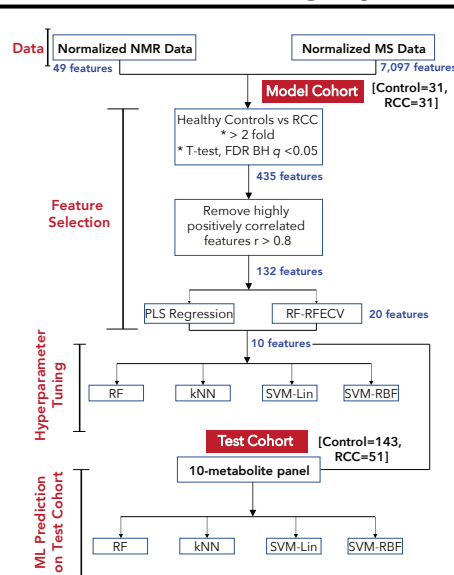
- The best chance of surviving Renal Cell Carcinoma (RCC) is through early diagnosis, however the disease is characterized by asymptomatic progression. [1]
- Currently, RCC is identified through cross-sectional imaging and biopsies, the latter being highly invasive and riddled with sampling errors. [2][3]
- We aim to predict RCC status from urine metabolic profile using mass spectrometry and nuclear magnetic resonance.

4. Propensity Matching



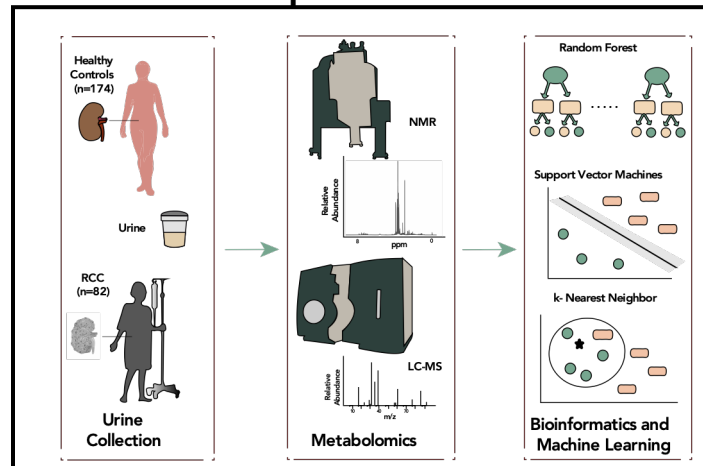
Propensity matching and model cohort characteristics. Study cohort characteristics before matching and after matching (model cohort). *p*-values were calculated using the t-test. For unequal and equal sample sizes, Welch and Student t-test were used respectively.

5. Machine Learning Pipeline

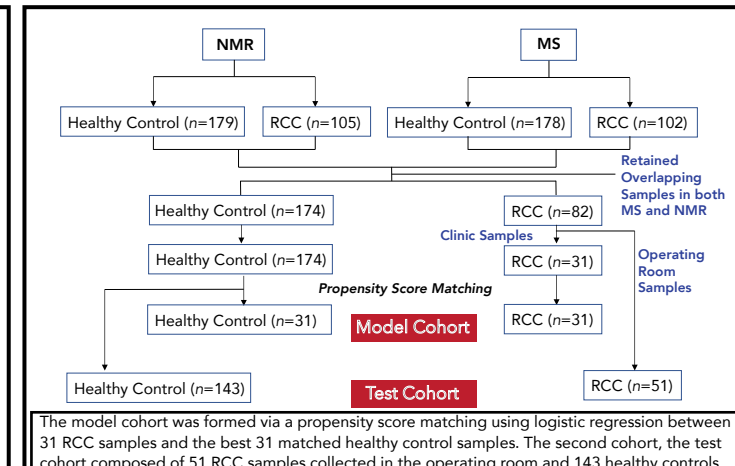


Using the model cohort, all NMR and MS features were subjected to a hybrid method of feature selection culminating in 10 selected metabolites. Hyperparameters for four different machine learning models were tuned using the model cohort and the 10-metabolite panel. Final predictions were made with using the test cohort under cross-validated conditions. PLS: Partial Least Squares. RF-RFECV: Random Forest Recursive Feature Elimination – Cross Validation FDR-BH: False Discovery Rate Benjamini Hochberg procedure k-NN: k-nearest neighbors. SVM: Support Vector Machines (Lin: linear, RBF: Radial Basis Function)

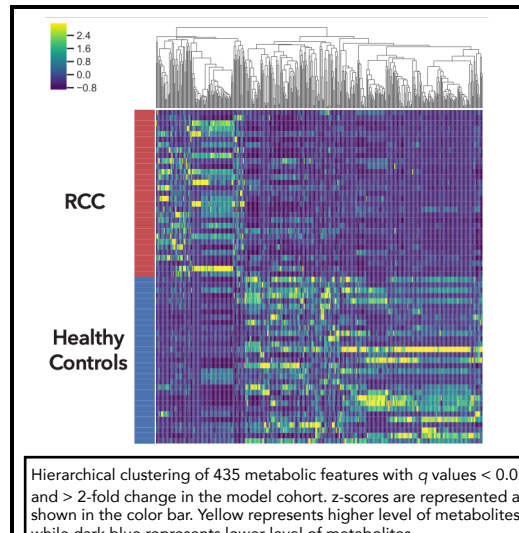
2. Graphical Abstract



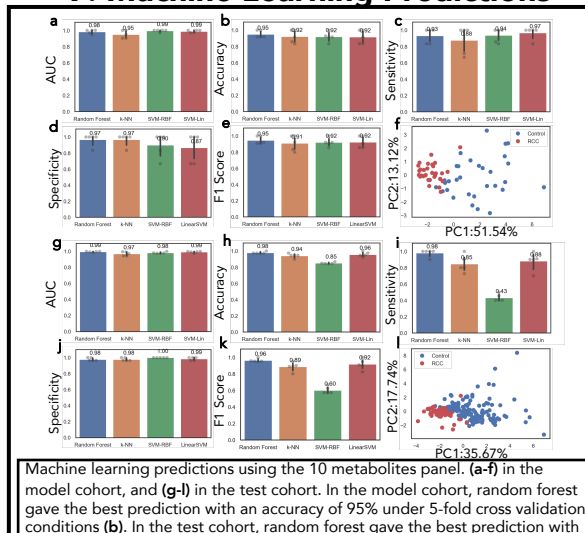
3. Patient Selection



6. Differential Metabolites



7. Machine Learning Predictions



8. Future Directions

ID	Mode	RT [min]	Formula
720	positive	2.562	C8 H9 N O
1481	positive	6.290	NaN
2102	positive	3.449	C8 H19 N
3141	positive	1.133	C7 H18 N8 O6 S
3675	positive	1.184	NaN
3804	positive	2.595	C4 H12 N O6 P
3872	positive	4.049	NaN
4080	positive	0.821	C10 H21 N3 O8 P2 S
6261	negative	2.591	C9 H18 N9 O2 P
6262	negative	2.667	C10 H20 N9 O5 P

MS/MS experiments for identification of the 10 metabolite panel

9. References

- Ann. Oncol. 2019, 30 (5), 706-720.
- Radiol Clin North Am 2017, 55 (6), 1235-1250.
- Curr Urol Rep 2017, 18 (4), 28.

Acknowledgements

