Molecular Subgroups Identified by Machine Learning Enable Accurate Prognostication of Papillary Thyroid Cancer

Steven Craig, MD[1, 2, *], Cynthia Stretch, PhD[3, *], Farshad Farshidfar, MD, PhD[3], Dropen Sheka[4], Nikolay Alabi[4], Adrian Harvey, MD[5], Karen Kopciuk, PhD[6], Young Joo Park, MD, PhD[7], Oliver F. Bathe, MD, MSc[3,5]

[1] Department of Surgery, Illawarra Shoalhaven Local Health District, NSW, Australia, [2] Graduate School of Medicine, University of Wollongong, NSW, Australia, [3] Department of Oncology, University of Calgary, Calgary, AB, Canada, [4] Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, Alberta, Canada, [5] Department of Surgery, University of Calgary, Calgary, AB, Canada, [6] Population Health Research, Alberta Health Services, Calgary, AB, Canada, [7] Department of Internal Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea

BACKGROUND

- Clinical management of papillary thyroid cancer (PTC) depends on an accurate estimate of prognosis
- Prognostication is currently based on clinicopathological variables which are sub-optimal and can only be fully implemented after surgery.
- **Objective: To derive an accurate molecular prognostic tool that can be implemented prior to surgery.**

METHODS

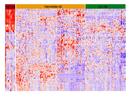
Machine learning algorithm created to identify features associated with recurrence (HighLifeR™)

TCGA RNASeq

Applied to PTC annotated by The Cancer Genome Atlas (1) (N=501)

82 Genes Identified

Identified genes most closely associated with recurrence



Genes subjected to unsupervised clustering to identify prognostic groups

VALIDATED

Prognostic signature validated
- internal TCGA cohort (N=167)
- external Korean cohort (N=124)

RESULTS

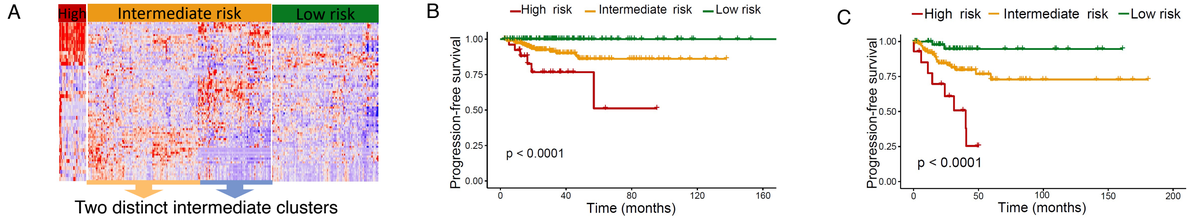


Figure 1. Unsupervised clustering of prognostic genes identified by HighLifeR™ resulted in four distinct clusters

Heatmap **(A)** showing the differences in expression in the 82 genes in the four clusters identified, including two intermediate-risk clusters. Kaplan Meier curves showing the differences in progression-free survival between the four prognostic subgroups in the discovery cohort **(B)** (n=334) and validation dataset **(C)** (n=167)

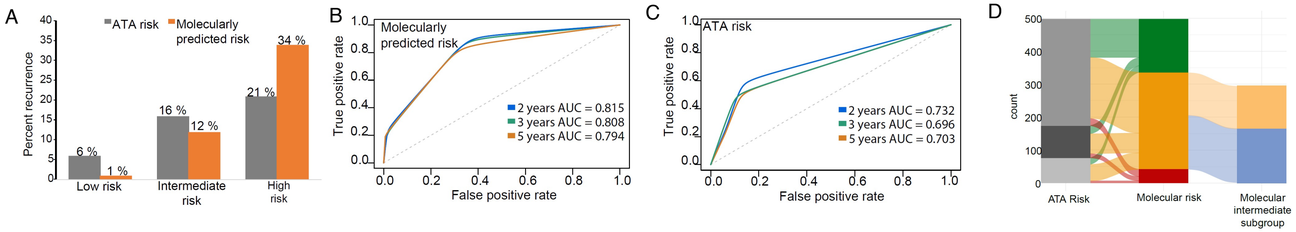


Figure 2. Performance of mRNA-based prognostic risk versus ATA risk stratification for recurrence

(A) Percent recurrences in patients assigned into the low-, intermediate-, and high-risk groups by the ATA and the mRNA-based prognostic risk, **(B)** and **(C)** Time-dependent area under the receiver operating characteristic curves for the mRNA-based predicted risk and ATA-predicted risk. **(D)** Alluvial plot showing how patients originally assigned into risk groups by ATA are reclassified by the mRNA-based prognostic risk classifiers.

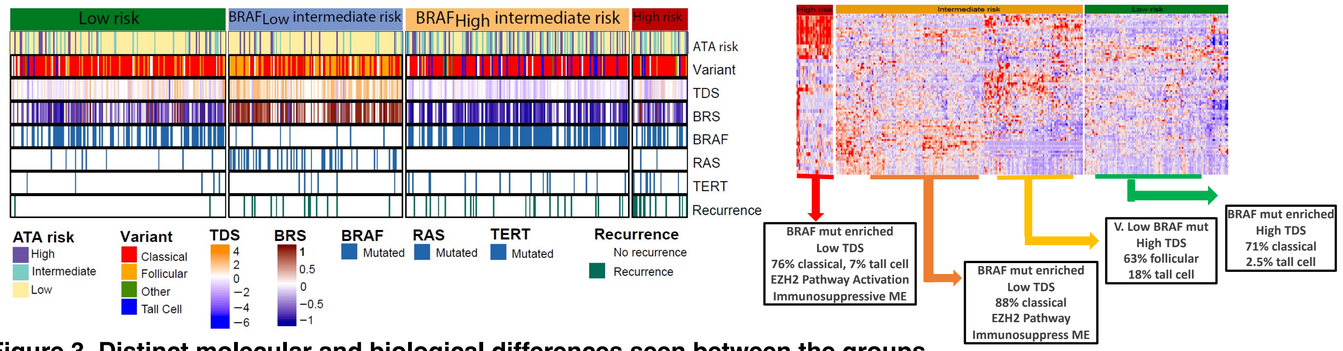


Figure 3. Distinct molecular and biological differences seen between the groups

The most striking differences were in variant type, thyroid differentiation score (TDS), BRAF-RAS score (BRS), and number of mutations in BRAF, RAS and TERT genes.

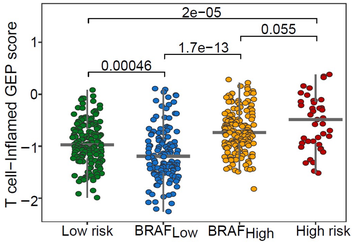


Figure 4. Biological differences between groups may guide therapeutic approaches.

T-cell inflamed gene expression scores for each of the mRNA-based risk groups are shown. High scores are prerequisites for clinical benefit from PD-1 blockade (2). High-risk group was characterized by an immunosuppressive microenvironment.

CONCLUSIONS

- **Our mRNA-based prognostic risk method outperforms clinicopathological risk stratification**
- mRNA-based prognostic risk can be applied preoperatively, thereby potentially directing surgical management
- Biological features of the four molecular subgroups could direct future therapies including immune checkpoint inhibitors and re-differentiation strategies using EZH2 inhibitors.

REFERENCES

1. Cancer Genome Atlas Research, N. (2014). Integrated genomic characterization of papillary thyroid carcinoma. Cell *159*, 676-690.
2. Cristescu, R., Mogg, R., Ayers, M., Albright, A., Murphy, E., Yearley, J., Sher, X., Liu, X.Q., Lu, H., Nebozhyn, M., et al. (2018). Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. Science *362*.

UNIVERSITY OF WOLLONGONG AUSTRALIA

UNIVERSITY OF CALGARY