Poster #1400

Novel expression biomarkers via prediction of response to FOLFIRINOX (FFX) treatment for PDAC

Hossein Asghari[1], Ehsan Haghshenas[†], Roby Thomas[2], Eric Schultz[1], Rob Patro[1], Stan Skrzypczak[1] and Carl Kingsford[1*]

[1]Ocean Genomics, Inc. Pittsburgh, PA
[2]Hillman Cancer Center, University of Pittsburgh Medical Center, Pittsburgh, PA
[†]Work done while at Ocean Genomics, Inc.
*Corresponding author.

OCEAN GENOMICS

Overview and significance

- We present one of the first multi-gene expression biomarkers for efficacy of FOLFIRINOX (FFX) in pancreatic ductal adenocarcinoma (PDAC).
- We present a robust set of 5 genes associated with response to FFX treatment by a novel algorithm employing bulk RNA-Seq data.
- FFX is a chemotherapy regimen for treating PDAC and is used as first-line therapy in a cohort of patients. Patients in this cohort received modified FOLFIRINOX (m-FOLFIRINOX) which includes 2 weekly intravenous leucovorin 400 mg/m2, 5-fluorouracil 2,400 mg/m2 given over 46 hours, irinotecan 150 mg/m2, and oxaliplatin 85 mg/m2. The response to treatment was assessed every 8 weeks using MRI and RECIST 1.1.
- Ocean Genomics DiscoverAI™ learned a predictor for response using machine learning techniques. The 5 most important genes are identified by permutation feature importance which improves the performance of the machine learning model (measured by area under the receiver operating characteristic curve- AUROC) by 7% in training data and 19% in test data.

Experimental questions

- What genes are expression biomarkers for predicting the response to FFX treatment in PDAC patients?
- Does feature selection process improve the prediction performance for response to FFX in terms of prediction AUROC?
- Do the identified biomarkers outperform previously identified biomarkers in PDAC datasets?

Datasets

Subset of samples from COMPASS study* treated by FFX:

- 103 whole-transcriptome RNA-Seq samples
- Obtained under agreement with University Hospital Network, Toronto
- RECIST 1.1 response labels:
  - ❑ Responders (CR/PR): 23 samples
  - ❑ Non-responders (SD/PD/NE): 80 samples

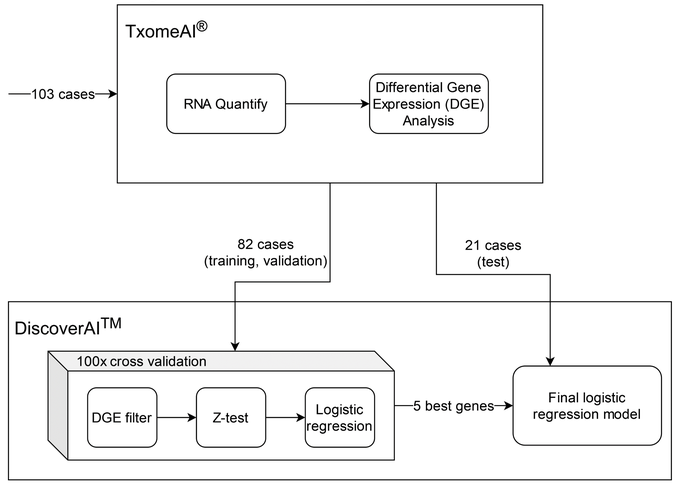
* Aung, Kyaw L et al. “Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial.” *Clinical cancer research : an official journal of the American Association for Cancer Research* vol. 24,6 (2018)

Methods

- Ocean Genomics TxomeAI® and DiscoverAI™ computational platforms were used to process bulk RNA-Seq samples and learn predictors respectively.
- TxomeAI® quantified expression of every gene using GENCODE v31 reference and genes with 0 expression in all samples were discarded.
- DiscoverAI™ learned predictors based on the processed data from TxomeAI®.
- Data split:
  - ❑ Training data: 66 cases
  - ❑ Validation data: 16 cases
  - ❑ Held out test data: 21 cases
- 82 (train + validation) cases were randomly split into 80/20 training/validation sets a 100 times.

Methods

- For each cross-validation (CV) fold, genes were selected by
  - ❑ Identifying statistically significant differentially expressed genes (DGE) between responders and non-responders.
    - Criteria: $p < 0.01$, $|\text{Log2 fold change}| > 0.5$
  - ❑ For each DGE gene, the difference in response rate was calculated and Z-test for proportions was performed between low and high expression cases using cutoffs obtained via maximal chi-squared statistics.
  - ❑ 30 genes with Z-test $p < 0.01$ and largest difference in response rate were fed into permutation feature importance based on logistic regression to identify 5 most important genes.
  - ❑ A logistic regression model was trained on the fold training set and using the 5 selected genes instead of all the genes.



- Genes from the best-performing fold were selected to train a final model on 82-sample training set and tested on the held out test set.

Biomarker discovery results

- Ocean Genomics DiscoverAI™ machine learning model based on logistic regression selected 5 genes with the highest permutation feature importance for each fold to train a final model on a training set.
- Genes selected by the final model are presented below. These genes were among the top-7 most frequently selected genes across folds.
- Three of the selected gene biomarkers are immunoglobulin related and immune-complex-bound proteins are predictive of response to chemotherapy†.

Gene Symbol	Gene Name	Selected in #folds
IGHG2	Immunoglobulin heavy constant gamma 2	23/100
IGKV3-20	Immunoglobulin kappa variable 3-20	48/100
IGLL5	Immunoglobulin lambda like polypeptide 5	32/100
WASH8P	Wiskott-Aldrich syndrome related pseudogene	43/100
HBB	Hemoglobin subunit beta	37/100

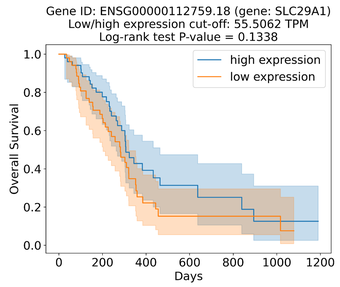
† Mandili, Giorgia et al. “Immune-Complexome Analysis Identifies Immunoglobulin-Bound Biomarkers That Predict the Response to Chemotherapy of Pancreatic Cancer Patients.” *Cancers* vol. 12,3 746. (2020)

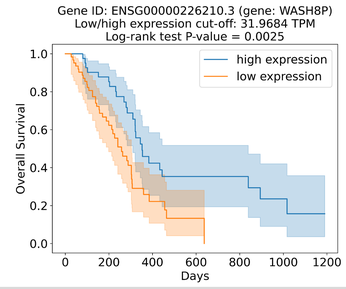
Biomarker discovery results

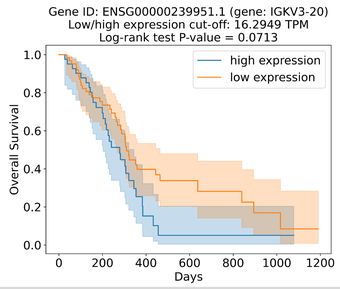
- Performance of the machine learning models using area under the receiver operating characteristic curve (AUROC).
- Logistic regression model was trained with and without prior feature selection.

Feature selection	CV AUROC	Test AUROC
5 selected genes from the best-performing fold	0.68	0.63
All genes (no feature selection)	0.61	0.44

- *hENT1* (aka SLC29A1) is an existing biomarker for Gemcitabine (GA) effectiveness in PDAC. It shows separation in the survival curves (Kaplan-Meier estimate) on this dataset with high expression cases having better life expectancy.
- Survival curves of two selected biomarkers in this study (*WASH8P* and *IGKV3-20*) are shown below, which have better separation in survival curves.







Conclusions

- Resulted in a robust candidate set of 5 genes (predictors) from a large PDAC, FFX-treated cohort.
- Each predictor is individually statistically significantly associated with response to treatment.
- The separation of the survival curves is larger in several of these genes, compared with *hENT1* which is a previous biomarker of Gemcitabine (GA) effectiveness in PDAC.
- All 5 candidate genes together are more predictive of response to FFX in this cohort compared to the all-gene model.
- Additional biological validation and additional computational variations of the study design are required to confirm the predictors.

Contact information

Hossein Asghari, Computational Biologist
hossein@oceangenomics.com

Ocean Genomics, Inc
4620 Henry Street Pittsburgh, PA 15213
contact@oceangenomics.com
@Ocean_Genomics