Afirma®
Thyroid FNA Analysis

Poster #180

Towards a Next-Generation GEC with Improved Specificity: Feasibility Analysis Using Machine Learning on a Combination of Gene Expression, Variants and Fusions from a Single Novel Sequencing Platform

86th Annual Meeting of the American Thyroid Association

September 21-25, 2016
Denver, Colorado

veracyte®

6000 Shoreline Court, Suite 300
South San Francisco, CA 94080
www.veracyte.com
1.888.9AFIRMA (888.923.4642)

Towards a Next-Generation GEC with Improved Specificity: Feasibility Analysis Using Machine Learning on a Combination of Gene Expression, Variants and Fusions from a Single Novel Sequencing Platform

Kennedy, Giulia C.;[1] Huang, Jing;[1] Whitney, Duncan;[1] Kim, Su Yeon;[1] Imtiaz, Urooj;[1] Tom, Ed;[1]
Wong, Mei;[1] Hao, Yangyang;[1] Walsh, Sean;[1] Barth, Neil;[1] LiVolsi, Virginia;[2] Kloos, Richard[1]

[1] Veracyte, Inc. [2] University of Pennsylvania

INTRODUCTION

Advances in genomic technology have enabled evaluation of indeterminate thyroid nodules for a large number of cancer-associated DNA mutations and fusions. Likewise, genome-wide transcriptome analysis on RNA has enabled gene expression "signatures" to be developed with high sensitivity and modest specificity. However, Next-Generation Sequencing (NGS) has yet to be exploited as a platform for combining the knowledge from both RNA expression and DNA variation to build stronger classifiers that more accurately diagnose thyroid nodules preoperatively from fine needle aspirates (FNA).
We developed a robust pipeline for capturing transcriptional data, mutations, variants and fusions all from the same RNA sample. Our goal was to determine the feasibility of adding richer genomic content to train a genomic classifier to improve the specificity of diagnosing benign nodules while maintaining high sensitivity.

METHODS

FNA biopsy samples from 88 patients were collected preoperatively and nucleic acids were isolated. The patients underwent thyroidectomies and the surgical tissue was diagnosed by a panel of histopathology experts. The cohort was balanced with 44 malignant (PTC, HCC, FC, MTC, and WDC-NOS) and 44 benign nodules (BFN, FA, HCA, LCT, NHP, and HTA). Training (n=58) and testing (n=30) sets were defined by carefully balancing cytology and histology, and classifier training was conducted in a blinded manner (Table 1). Samples were subjected to NGS with 15ng of RNA input. Classification models were evaluated within the training set in cross-validation according to overall performance (Figure 1). The best model was then selected to analyze the test set.

RESULTS

The top 2000 differentially expressed genes, 1402 sequence variants and nine fusion-pairs were used to develop several models (Figure 2a&b). The best model uses an ensemble score (median probability) from three models (SVM, LASSO, Random Forest). In the test set, this classifier yielded an overall AUC of 0.88, with a sensitivity and specificity of 93% and 80% (Table 2, Figure 3).

CONCLUSIONS

Classifiers with high sensitivity and improved specificity can be developed from a combination of features generated using our NGS assay. Although this feasibility study is based on a relatively small data set, the principles of how counts, variants and fusions can be effectively combined has been demonstrated. Efforts are underway to apply this approach to a larger cohort.

TABLE 1.

Feasibility Cohort Characteristics

	Train	Test
Gender		
Male	12 (21%)	4 (13%)
Female	42 (72%)	25 (83%)
Unknown	4 (7%)	1 (3%)
Cytology		
Benign	18 (31%)	12 (40%)
AUS/FLUS	8 (14%)	4 (13%)
FN/SFN	7 (12%)	3 (10%)
SFM	9 (16%)	3 (10%)
Malignant	16 (28%)	8 (27%)
Histopathology		
Benign	29	15
Malignant	29	15

TABLE 2.

Performance Summary

Performance Matrix	Point Est. (95 CI)
AUC	0.88 (0.74 - 1.0)
Sensitivity	93% (68-100%)
Specificity	80% (52-96%)
NPV (at 24% ROM)	97% (85-100%)
PPV (at 24% ROM)	60% (35-80%)

FIGURE 1.

Classifier Development

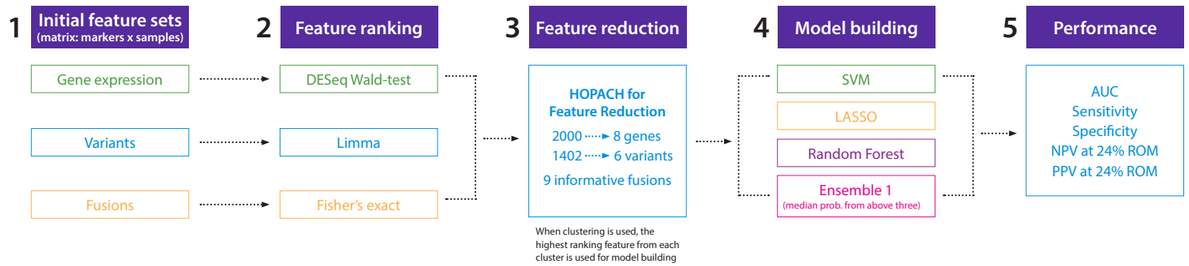
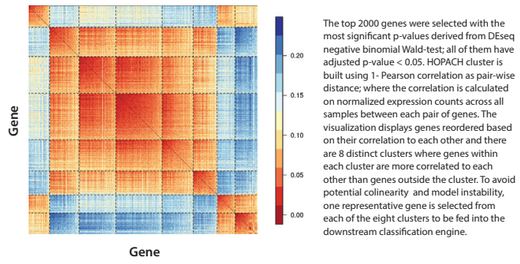


FIGURE 2.

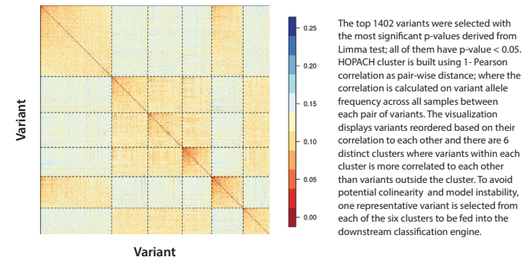
HOPACH Clustering in Training Set

2a. Clustering on Top 2000 Expression Genes



The top 2000 genes were selected with the most significant p-values derived from DEseq negative binomial Wald-test; all of them have adjusted p-value < 0.05. HOPACH cluster is built using 1- Pearson correlation as pair-wise distance; where the correlation is calculated on normalized expression counts across all samples between each pair of genes. The visualization displays genes reordered based on their correlation to each other and there are 8 distinct clusters where genes within each cluster are more correlated to each other than genes outside the cluster. To avoid potential colinearity and model instability, one representative gene is selected from each of the eight clusters to be fed into the downstream classification engine.

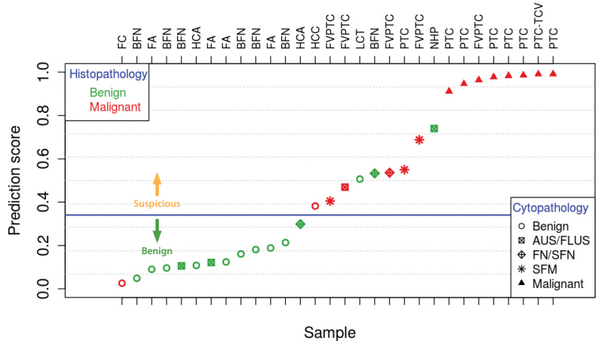
2b. Clustering on Top 1402 Variants



The top 1402 variants were selected with the most significant p-values derived from Limma test; all of them have p-value < 0.05. HOPACH cluster is built using 1- Pearson correlation as pair-wise distance; where the correlation is calculated on variant allele frequency across all samples between each pair of variants. The visualization displays variants reordered based on their correlation to each other and there are 6 distinct clusters where variants within each cluster is more correlated to each other than variants outside the cluster. To avoid potential colinearity and model instability, one representative variant is selected from each of the six clusters to be fed into the downstream classification engine.

FIGURE 3.

Individual Scores



Each point represents one of the 30 validation samples. They are ordered by their classification prediction score (ensemble score) from low to high. The color of the point indicates the true histopathology status of the sample: red corresponds to malignant and green corresponds to benign; the labels provided on the top of the figure are the histopathology subtype of each sample. The shape of the point indicates the Bethesda cytology category of each sample; for more details please refer to the figure legend. The blue horizontal line is the cut-off of the final classifier where samples above the line are classified as suspicious and samples below the line are classified as benign.