Diagnosis of Aortic Aneurysm from Gene Expression Profiling of Peripheral Blood

Catalin Barbacioru[1],Yulei Wang[1], Dov Shiffman[2], Olga Iakoubova[2], Sriram Balasubramanian[3], Julie Blake[1], John Elefteriades[4] and Raymond Samaha[1],

[1] Applied Biosystems, 850 Lincoln Centre Dr., Foster City, CA 94404, [2] Celera Diagnostics, 1401 Harbor Bay Pkwy, Alameda, California 94502, [3] Celera Genomics, 180 Kimball Way, South San, Francisco, CA 94080, [4] Department of Cardiothoracic Surgery, Yale University School of Medicine, 333 Cedar St (121 FMB), New Haven, CT 06510

ABSTRACT

We report in this study that gene expression profiles of peripheral blood cells may allow early detection and diagnosis of aortic aneurysm. Gene expression profiles of peripheral blood samples collected from 58 individuals diagnosed with thoracic aortic aneurysm (cases) and 36 normal individuals (controls) were analyzed using the Applied Biosystems Expression Array Systems and Human Genome Survey Microarrays. Prediction models constructed from bootstrapped samplings on a training set containing 61 samples (36 cases, 25 controls), and containing 41 genes are identified and used to predict the remaining 33 testing samples (22 cases and 11 controls) unused in the training phase. The prediction accuracy for this set of samples is 81% while the 95% confidence interval estimated from the training samples is [0.75, 0.85]. Our study provides a comprehensive gene expression profile of peripheral blood cells from thoracic aortic aneurysm patients. The biological pathways associated with the aneurysm patients in this study may provide further insights into the molecular mechanisms attributed to the pathogenesis of this disease. Our results also demonstrated that a blood-based gene expression test may facilitate the diagnosis of aortic aneurysm disease.

INTRODUCTION

Aortic Aneurysms is the 17th leading cause of death in the United States, accounting for more than 15,000 deaths each year. Aortic aneurysm is often called a "silent killer" because it is usually asymptomatic until it ruptures, which often results in death. These deaths can be avoided by early detection and treatment. Because aortic aneurysm displays various immune responses and is frequently associated with atherosclerosis, a chronic inflammatory disorder of the vascular wall, we hypothesized that gene expression patterns of peripheral blood cells from subjects with aneurism could reflect their disease status. We report in this study that gene expression profiles of peripheral blood cells may allow early detection and diagnosis of aortic aneurysm.

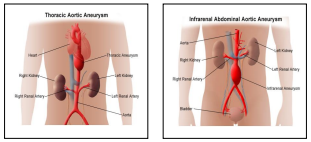

Figure. 1. Aortic aneurysm is a vascular disease with a bulge in the wall of the aorta (the largest artery in the body). Aneurysm can occur in the abdomen below the kidneys (abdominal aortic aneurysm, right), or occur in the chest cavity (thoracic aortic aneurysm, left).

RESULTS

Total RNA was extracted from the whole blood samples and gene expression profiles of 29,098 individual human genes were analyzed using the Applied Biosystems Expression Array Systems and Human Genome Survey Microarrays.
A 61-sample training set containing 36 TAA patients and 25 controls were used to select predictor genes and construct prediction models. The first step in this analysis is to select a list of genes capable of differentiating between cases and controls. Because of the heterogeneity of the blood samples, we used bootstrap resampling procedure to identify candidates for predictor genes. Bootstrapping acknowledges the complex relation between genes as well as the variability between samples. The t statistic is used as a measure of discrimination between the two phenotypes. Relative ranking of genes according to absolute value of the t statistic for each bootstrap step is used to estimate the selection probability of each gene [1]. This method is based on 10-fold cross validation (of the 61 training samples) for sample classification to the nearest shrunken centroids (PAM) [2]. Variability of this algorithm in gene selection step as well as estimators for accuracy of class prediction are presented for this study. A 41 prediction gene model is identified and performance on both training (CV) and testing samples is presented. TaqMan assays confirmed the results found using Applied Biosystems Expression Array System.

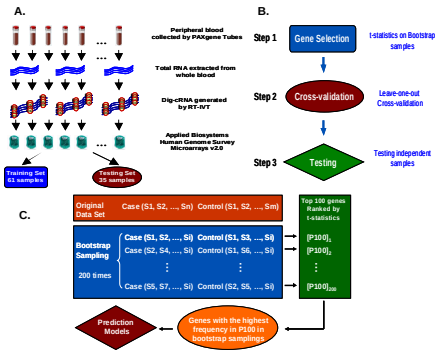

Figure. 2. Experimental Design: (A) Peripheral blood samples were collected from 58 individuals diagnosed with thoracic aortic aneurysm (cases) and 36 normal individuals (controls). Total RNA was extracted from the whole blood samples, Dig-labeled cRNA were generated via RT-IVT reactions and gene expression profiles of 29,098 individual human genes were analyzed using the Applied Biosystems Expression Array Systems and Human Genome Survey Microarrays v2.0. Data from 61 samples were used as the training set for construction of the prediction models, while data from the rest 35 samples were set aside as the independent testing set.. (B) Scheme of construction of prediction models for risk assessment of aneurysm. (C) Gene selection using t-statistics on bootstrapped samples. Only genes detectable (S/N > 3) in at least 70% of the cases or at least 70% of the controls were used for the analysis. For each of the 200 bootstrap samplings, t-statistics was computed for each gene and the top 100 genes (P100) with the smallest p-values were identified. Genes with the highest frequency in P100 among the 200 bootstrap samplings were used to construct the prediction models.

Algorithm

1. Gene selection (bootstrap ranking, training samples):
- Present genes in at least 1 phenotype are ranked according to their t-stat in 500 'bootstrap data sets'
- Genes showing up in top 500 in at least half of the bootstrap rankings and with average ranking > 500 are choose for further analysis (in general about 105-120 genes are selected this way)

2. Number of predictor genes (Cross-validation, training samples):
- Use 'outer' 10-fold cross validation (on training samples): construct gene lists as in previous step, from the training bins, predict the remaining bin. Estimate prediction accuracy as a function of the number of predictor genes (Figure 3)

3. Testing (training and testing samples):
- Use all 61 training samples to identify the final 41 predictor genes as in step 1. Use 10-fold cross validation of the 61 training set (Figure 4 A.)
- Lea … se PAM to pred …

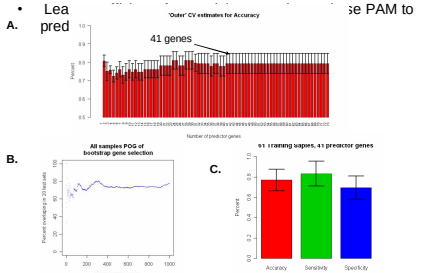

Figure. 3. Number of Predictor Genes and estimates for accuracy, sensitivity and specificity. (A) Cross-validation estimates of accuracy for different number of predictor genes. (B) Percent overlapping between the 10 gene lists in cross validation step. (C) Cross-validation estimates of accuracy, sensitivity and specificity.

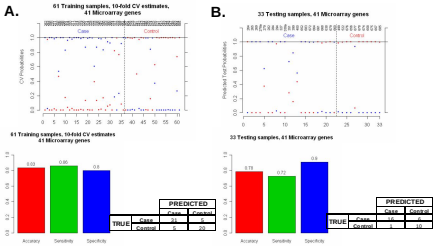

Figure 4. Performance of the prediction of training and testing samples. (A) 10-fold CV of training samples based on 41 predictor genes. (B) Prediction of the 33 testing samples: the 41 predictor genes and their coefficients are determined from the 61 training samples. PAM was used to predict testing samples.

TaqMan validation

TaqMan® Gene Expression Assay for the 41 predictor probes were used in quadruplicates, to screen 82 blood samples (52 from the training set and 30 from the testing set). Coefficients of prediction model were learned from the 52 training samples and the 30 testing samples were predicted using PAM. The performance of the prediction model confirms the results obtained from Applied Biosystems Expression Array System.

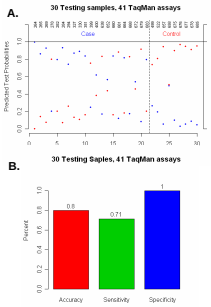

Figure 5. TaqMan validation.

(A) Prediction probabilities of the 30 testing blood samples. Coefficients of prediction model were learned from the 52 training samples and the 30 testing samples were predicted using PAM. (B) and (C) Prediction performance on the 30 testing samples using TaqMan® assays for the 41 predictor genes identified by Applied Biosystems Expression Array System.

C.

		PREDICTED	
		Case	Control
TRUE	Case	15	6
	Control	0	9

CONCLUSIONS

Our study provides a comprehensive gene expression profile of peripheral blood cells from thoracic aortic aneurysm patients. We develop a prediction model for aneurysm diagnostic based on 41 genes, showing good accuracy (80%) and good sensitivity (72%) . Our results also demonstrated that a blood-based gene expression test may facilitate the diagnosis of aortic aneurysm disease.

REFERENCES

[1] Pepe MS, Longton G, Anderson GL, Schummer M: Selecting differentially expressed genes from microarray experiments. *Biometrics* 2003, 59:133-142.
[2] Tibshirani R, Hastie T, Narasimhan B, Chu G: Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* 2002, 99:6567-6572.

ACKNOWLEDGEMENTS

We thank Sergei Labur, Alexandra Fuller, La-Arni Macalik from Applied Biosystems and Stephen Capaldi from Celera South San Francisco, for their technical supports.