EBioMedicine xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

# A Predictive 7-Gene Assay and Prognostic Protein Biomarkers for Non-small Cell Lung Cancer

Nancy Lan Guo [a,*], Afshin Dowlati [b], Rebecca A. Raese [a], Chunlin Dong [a], Guoan Chen [c], David G. Beer [c], Justine Shaffer [a], Salvi Singh [a], Ujala Bokhary [d], Lin Liu [d], John Howington [d], Thomas Hensing [d], Yong Qian [e]

[a] West Virginia University Cancer Institute, Mary Babb Randolph Cancer Center, West Virginia University, Morgantown, WV 26506-9300, United States
[b] Case Comprehensive Cancer Center, Case Western Reserve University, 10900 Euclid Ave., Cleveland, OH 44106, United States
[c] Comprehensive Cancer Center, University of Michigan, 1500 East Medical Center Drive, Ann Arbor, MI 48109-0944, United States
[d] Kellogg Cancer Center, NorthShore University HealthSystem, 2650 Ridge Avenue, Evanston, IL 60201, United States
[e] National Institute of Occupational Safety and Health, 1095 Willowdale Road, Morgantown, WV 26505, United States

ARTICLE INFO

*Article history:*
Received 17 April 2018
Received in revised form 9 May 2018
Accepted 21 May 2018
Available online xxxx

*Keywords:*
Predictive gene assay
Prognostic protein biomarkers
Non-small cell lung cancer (NSCLC)
Chemotherapy
Immunotherapy
Precision therapy

ABSTRACT

*Purpose:* This study aims to develop a multi-gene assay predictive of the clinical benefits of chemotherapy in non-small cell lung cancer (NSCLC) patients, and substantiate their protein expression as potential therapeutic targets.
*Patients and methods:* The mRNA expression of 160 genes identified from microarray was analyzed in qRT-PCR assays of independent 337 snap-frozen NSCLC tumors to develop a predictive signature. A clinical trial JBR.10 was included in the validation. Hazard ratio was used to select genes, and decision-trees were used to construct the predictive model. Protein expression was quantified with AQUA in 500 FFPE NSCLC samples.
*Results:* A 7-gene signature was identified from training cohort ($n = 83$) with accurate patient stratification ($P = 0.0043$) and was validated in independent patient cohorts ($n = 248$, $P < 0.0001$) in Kaplan-Meier analyses. In the predicted benefit group, there was a significantly better disease-specific survival in patients receiving adjuvant chemotherapy in both training ($P = 0.035$) and validation ($P = 0.0049$) sets. In the predicted non-benefit group, there was no survival benefit in patients receiving chemotherapy in either set. The protein expression of ZNF71 quantified with AQUA scores produced robust patient stratification in separate training ($P = 0.021$) and validation ($P = 0.047$) NSCLC cohorts. The protein expression of CD27 quantified with ELISA had a strong correlation with its mRNA expression in NSCLC tumors (Spearman coefficient = 0.494, $P < 0.0088$). Multiple signature genes had concordant DNA copy number variation, mRNA and protein expression in NSCLC progression.
*Conclusions:* This study presents a predictive multi-gene assay and prognostic protein biomarkers clinically applicable for improving NSCLC treatment, with important implications in lung cancer chemotherapy and immunotherapy.


## 1. Introduction

Lung cancer is the leading cause of cancer-related deaths in the world, and non-small cell lung cancer (NSCLC) accounts for almost 80% of lung cancer deaths [1]. The heterogeneous nature of lung cancer makes it a very difficult disease to treat. Major histology of NSCLC includes lung adenocarcinoma and squamous cell lung carcinoma. Surgical resection is the major treatment for early stage NSCLC. However, about 22–38% of stage I NSCLC patients will develop tumor recurrence within five years following the surgery [2]. It is therefore important to select early stage NSCLC patients for more aggressive treatment. While adjuvant chemotherapy of stage II and stage III disease has resulted in 10–15% increased overall survival [3], the prognosis for early stage NSCLC remains poor [4]. Currently, there are no clinically available molecular assays to predict the risk for tumor recurrence and the clinical benefits of chemotherapy in NSCLC patients.

Immunotherapy has rapidly gained attention of oncologists as an effective and less toxic treatment than chemotherapy in patients with advanced lung cancers [5–8]. A recent study used paired single cell analysis to compare normal lung tissue and blood with tumor tissue in stage I NSCLC, and found that early-stage tumors had already begun to alter the immune cells in their microenvironment [8]. These results suggest that immunotherapy could potentially be used to treat early stage lung cancer patients. However, predictive biomarkers of immunotherapy are not well established except PD-1 or PD-L1, and it is unlikely that a single marker is sufficient.

* Corresponding author at: 2816 HSS, West Virginia University Cancer Institute, Morgantown, WV 26506-9300, United States.
*E-mail address:* lguo@hsc.wvu.edu (N.L. Guo).

https://doi.org/10.1016/j.ebiom.2018.05.025

High-throughput technologies, such as microarray and RNA-seq, promise the discovery of novel biomarkers from genome-scale studies. The FDA conducted a systematic evaluation and suggested continued usefulness of legacy microarray data and established microarray biomarkers and predictive models in the forthcoming RNA-seq era [9]. However, several disadvantages have limited the application of high-throughput techniques in routine clinical tests, including costs, reproducibility, and data analyses [10]. Compared with microarray and RNA-seq, quantitative real-time RT-PCR (qRT-PCR) is more efficient, consistent, and able to measure gene expression over a greater dynamic range. It requires only a small sample and can be modified to quantify gene expression in formalin-fixed and paraffin-embedded (FFPE) tissues [11]. The combined use of real-time qRT-PCR with high-throughput analysis can overcome the inherent biases of the high-throughput techniques and is emerging as the optimal method of choice to translate genome research into clinical practice [12]. The protein expression validation of the identified mRNA biomarkers could substantiate their ultimate functional involvements in disease, and may lead to the discovery of potential proteomic biomarkers in abundant FFPE samples for broader applications in community hospitals.

DNA microarray-based studies identified gene expression-based NSCLC prognostic [13] and predictive biomarkers [14,15]. A qRT-PCR based 14-gene assay by Kratz et al. [16] is prognostic of non-squamous NSCLC outcome in FFPE tissues and is ready for wide-spread clinical applications. However, this 14-gene assay is limited to non-squamous NSCLC and is not shown to be predictive of the clinical benefits of chemotherapy.

In this study, a combined analysis of genome-wide transcriptional profiles and qRT-PCR was utilized to develop a multi-gene assay both prognostic of NSCLC outcome and predictive of the benefits of chemotherapy. Patient cohorts from multiple hospitals in the US and JBR.10 data [14] were used to validate this multi-gene assay. Protein expression of the identified biomarkers was also evaluated in patient tissue samples and correlated with the mRNA expression and DNA copy number variation to substantiate their functional involvement and potential as therapeutic targets in chemotherapy and immunotherapy, in addition to companion tests.

## 2. Materials and Methods

### 2.1. Patient Samples

Clinical characteristics of patient cohorts used in qRT-PCR assays is summarized in Table 1. All NSCLC patients were staged I, II, or IIIA at the time of diagnosis. Tumor tissues were collected in surgical resections and were snap-frozen at −80 °C until used for RNA extraction. Tumor cell content was above 50% for qRT-PCR assays. Those with missing AJCC staging information, missing histology, death within 30 days of resection or from other disease conditions were excluded from further analysis. A total of 122 NSCLC patient samples were obtained from Case Western Reserve University (CWRU) Comprehensive Cancer Center. Total RNA of good quality was extracted from 89 tumor specimens. Good quality RNA from 101 lung adenocarcinoma tumor specimens was obtained from University of Michigan (UM) Comprehensive Cancer Center, with detailed description of patients, tissue specimens and mRNA quality check provided in [17]. A total of 65 NSCLC tumor specimens from NorthShore University HealthSystem Kellogg Cancer Center and 49 specimens from West Virginia University Cancer Institute [Mary Babb Randolph Cancer Center (MBRCC)] generated good quality mRNA. The tissue collection in this study was approved by an Institutional Review Board (IRB) at each institution.

### 2.2. RNA Extraction, and Quality and Concentration Assessments

Total RNA was extracted from snap-frozen tumor tissues using a RNeasy mini kit according the manufacturer's protocol (Qiagen, USA),

**Table 1**
Clinical information of non-small cell lung cancer patient cohorts collected for the qRT-PCR analysis.

| | | CWRU ($n = 89$) | MBRCC ($n = 49$) | UM ($n = 101$) | NorthShore ($n = 65$) | |
|---|---|---|---|---|---|---|
| Age | Mean (Std error) | 70.11 (0.94) | 66.70 (1.25) | 67.04 (0.96) | 69.64 (1.02) | |
| | <60 | 15 (15.15%) | 7 (14.29%) | 28 (27.72%) | 7 (10.77%) | |
| | ≥60 | 84 (84.85%) | 39 (79.59%) | 73 (72.28%) | 48 (73.85%) | |
| | Missing | | 3 (6.12%) | | 10 (15.38%) | |
| Sex | F | 52 (52.53%) | 23 (46.94%) | 53 (52.48%) | 34 (52.31%) | |
| | M | 47 (47.47%) | 26 (53.06%) | 48 (47.52%) | 21 (32.31%) | |
| | Missing | | | | 10 (15.38%) | |
| Smoking | Current | 43 (43.43%) | 1 (2.04%) | | Yes | 60 (92.31%) |
| | Former | 40 (40.40%) | 3 (6.12%) | | | |
| | Never | 8 (8.08%) | | | No | 5 (7.69%) |
| | Passive | 1 (1.01%) | | | | |
| | Other | 1 (1.01%) | | | | |
| | Missing | 6 (6.06%) | 45 (91.48%) | | | |
| AJCC stage | I | 46 (46.46%) | 27 (55.10%) | 59 (58.42%) | 46 (70.77%) | |
| | II | 46 (46.46%) | 16 (32.65%) | 16 (15.84%) | 15 (23.08%) | |
| | III | 6 (6.06%) | 6 (12.25%) | 26 (25.74%) | 4 (6.15%) | |
| | Missing | 1 (1.01%) | | | | |
| Chemotherapy | Yes | 29 (29.29%) | 27 (55.10%) | 24 (23.76%) | 28 (40.03%) | |
| | No | 52 (52.53%) | 20 (40.82%) | 77 (76.24%) | 36 (55.38%) | |
| | Missing | 13 (13.13%) | 2 (4.08%) | | 1 (1.54%) | |
| Histology | Adenocarcinoma | 65 (65.66%) | 27 (55.10%) | 101 (100%) | 43 (66.15%) | |
| | Squamous | 27 (27.27%) | 14 (28.57%) | | 11 (16.92%) | |
| | Other | 7 (7.07%) | 8 (16.33%) | | 6 (9.23%) | |
| | Missing | 5 (5.05%) | | | 5 (7.69%) | |
| Differentiation | Well | 5 (5.05%) | | 28 (27.72%) | 20 (30.77%) | |
| | Moderate | 44 (44.44%) | | | 4 (6.15%) | |
| | Moderate to Poorly | 4 (4.04%) | | 39 (38.61%) | 22 (33.85%) | |
| | Poorly | 35 (35.35%) | | 34 (33.66%) | 17 (26.15%) | |
| | Missing | 11 (11.11%) | | | 2 (3.08%) | |
| Tumor Grade | 1 | 5 (5.05%) | 3 (6.12%) | | 20 (30.77%) | |
| | 2 | 44 (44.44%) | 18 (36.73%) | | 19 (29.23%) | |
| | 3 | 36 (36.36%) | 22 (4.90%) | | 21 (32.31%) | |
| | Other | 3 (3.03%) | | | | |
| | Missing | 11 (11.11%) | 6 (12.25%) | | 5 (7.69%) | |

Please cite this article as: Guo NL, et al, A Predictive 7-Gene Assay and Prognostic Protein Biomarkers for Non-small Cell Lung Cancer, EBioMedicine (2018), https://doi.org/10.1016/j.ebiom.2018.05.025