Sputum as a Diagnostic Matrix for Respiratory Disease Screening

Subasa Chandra Bishwal¹, Ranjan Kumar Nanda², Rajendra Kumar Behera³

¹Research Fellow, Translational Health Group, International Centre for Genetic Engineering and Biotechnology, New Delhi, India and School of Life Sciences, Sambalpur University, Jyoti Vihar, Sambalpur, Odisha, India.
²Group Leader, Translational Health Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi.
³Associate Professor, School of Life Sciences, Sambalpur University, Jyoti Vihar, Sambalpur, Odisha, India.

Corresponding Author: Rajendra Kumar Behera

ABSTRACT

Mucus hyper-secretion is a common feature in most of respiratory diseases. Sputum contains white blood cells, cellular debris, dead tissue, serous fluid and offers an easy, inexpensive and non-invasive method of disease diagnosis. Sputum is rich in lipids, glycoconjugates and proteins. These molecules could be secreted from respiratory tract, lungs and infected sites as well as of pathogen origin. Complete sputum proteome profiling could be useful by adopting mass spectrometry methods to discover suitable clinical markers that could explain the disease and could be used for development of easy to use cost effective diagnostic test. The advancement of proteomics tools will be useful to carry out such biomarker discovery work. In this article, we will be discussing about the sputum proteome analysis in different respiratory disease conditions. Successful application of proteomics tool for understanding perturbed pulmonary disease conditions using sputum as a diagnostic matrix will be discussed.

Key words: Diagnosis, Noninvasive, Proteomics, Respiratory, Sputum.

INTRODUCTION

Worldwide more than 1 billion people are reported to be affected by respiratory diseases every year. The main risk factors for respiratory diseases involving lower respiratory tract infection are use of tobacco, air pollution, occupational dust and chemical exposure. Most common respiratory diseases are chronic obstructive pulmonary disease (COPD), asthma, Cystic fibrosis, Lung Cancer, pulmonary hypertension and tuberculosis (TB). Excessive mucus production in airways is a common feature of respiratory diseases. Expectoration is the act of coughing up or spitting out the material produced in the respiratory tract. When mucus is coughed up from the lower airways is called sputum. Primary role of mucus is to defend the respiratory tract from the noxious substance and to maintain airways from external damages. Secretion of mucus in airways is a homeostatic mechanism. It protects the respiratory tract from external particles and internal damages. Mucus is highly hydrated fluid layer that cover the mucosal surface. It is rich in secreted mucin and other molecules involved in host defense against infection. Mucin is secreted by goblet cells in epithelium and submucosal glands. Mucin is 10-40 MDa in size and 3-10 nm in diameter, which is dense, with 25-30 carbohydrate chains per 100 amino acid
residues and constitute up to 80% of the dry weight. (5) Sputum is rich in mucin and commonly used for microbiological investigations for different respiratory disease conditions.

**GENE INVOLVED IN SPUTUM PRODUCTION**

Almost twenty genes, found in chromosomes 1, 3, 4, 7, 11, 12 and 19 are reported to be responsible for mucin production in human. Nice (MUC1, MUC2, MUC4, MUC5AC, MUC5B, MUC7, MUC8, MUC11, MUC13) out these twenty genes are expressed in respiratory tract. These mucin genes are further grouped to membrane bound (MUC1 and MUC4), secreted (MUC2, MUC5AC, MUC5B, MUC7) and rest three (MUC3, MUC6, MUC8) are yet to be categorized. [6-11]

**MECHANISM OF SPUTUM PRODUCTION**

Over expression of mucin genes, excess production secondary to mucus cell hyperplasia, hypertrophy or even metaplasia or goblet cells or gland hyper secretion of formed and stored mucin in the airways might contribute in higher sputum production. Altered eicosanoids and lipid mediators in disease conditions, inflammatory mediators, environmental agents/pollutant, bacterial derived products, reactive oxygen and nitrogen species, ATP and UTP, cytokines might induce over expression of mucin genes. [12-20]

**SPUTUM AS A DIAGNOSTIC MATRIX**

Due to its source, rich molecular details and primarily due to its noninvasive method of collection, sputum is used as the most common diagnostic matrix in many respiratory disease conditions. In fact, sputum color chart is used to find our presence of microorganisms like rusty color indicates presence of pneumococcal bacteria. Green color of sputum is caused by neutrophil myeloperoxidase, foamy white may come from obstruction or even edema, frothy pink in pulmonary edema. [4] Sputum smear observed under microscope could help identification of many bacterial and fungal diseases. [4,6] Sputum culture provides the most sensitive and specific test results for multiple bacterial and fungal disease conditions like tuberculosis and pneumonia. Effect of therapeutic intervention may also be monitored by change in sputum color. Purulent containing pus, yellow-greenish (mucopurulent) color suggests that treatment with antibiotics can reduce symptoms. A white, milky or opaque (mucoid) appearance often means that antibiotics may be ineffective in treating symptoms.

To find disease specific early biomarkers for important health conditions, before getting it complicated is important. Sputum as a diagnostic matrix has several advantages to other commonly used diagnostic matrix like serum or plasma due to their non-invasive method of collection. So a noninvasive method of diagnostic matrix collection with minimum anxiety, discomfort gains higher acceptance to patient to undergo health inspection. [21-23] Disease onset, progression and treatment outcome may be monitored by sputum biomarkers. [24,25] Sputum analysis can indicate the presence of microbes and the degree and type of inflammation in the airways. [26] This makes sputum as a preferred and important matrix for diagnosis of various respiratory diseases.

**SPUTUM PROTEOMICS APPROACH IN DISEASE DETECTION**

Sputum is rich in genetic materials of both host and pathogen origin. Analysis of genetic material by amplifying the signal through polymerase chain reaction pathogen specific primer is important. It involves requirement of cost intensive chemicals, instruments and needs specific training. These methods have low turnaround time to get results and minimize time for initiating treatment. Transcriptomics analysis provides interesting information on disease diagnosis and therapeutic outcomes. Like
analysis of Isocitrate lyase mRNA was found to be correlated well with bacterial load in sputum at time of diagnosis and during therapeutic interventions. Quantitative real time polymerase chain reaction is useful for monitoring important analytes (mRNA) however it is cost intensive and requires high end instruments at temperature-controlled environment. Protein profiling has a potential advantage of discovering suitable clinical biomarkers that are significantly deregulated by disease and receiving treatment \cite{27}. Identification of important proteins provide useful translational potential to develop easily deployable solutions like lateral flow devices that could be used in point of care for disease screening or treatment monitoring. Employing advanced quantitative proteomics tools, deregulated proteins expressed in diverse biofluids like plasma, serum, urine, sputum or saliva in disease conditions could be identified as important marker molecules. These marker molecules, after careful validation in independent test populations of diverse ethnic background could be useful to develop clinical assay for monitoring pathogenesis for respiratory diseases Fig. 1. \cite{28,29}

**Tuberculosis**

Majority of the pulmonary tuberculosis diagnosis is done using a 120 years old sputum acid fast microscopy test. Sputum microscopy test has significantly low sensitivity and specificity but due to its easy availability in resource limited conditions it is the important diagnosis test. Growing the tuberculosis causing microorganisms in sputum provides the highest sensitivity and specificity but suffer limited use due to long turnaround time and resource intensive procedure. A protein-based tuberculosis diagnostic method might be useful to develop an easy to use TB diagnostic and monitoring test to determine the treatment efficacy. \cite{30-32} Sputum proteins from active tuberculosis and control subjects, separated in a two-dimensional electrophoresis unit identified 62 differentially expressed protein spots. These proteins spots were analyzed using matrix-assisted laser desorption ionization time-of-flight/time-of-flight mass spectrometry (MALDI-TOF/TOF) and found to be 47 proteins were down regulated and 15 proteins were up-regulated. Most of these deregulated proteins were involved in acute phase response, signal transduction, cytoskeleton structure and immune response. Increase in abundance of acute phase proteins is to improve survival and restore haemostatic in host system. Among those deregulated proteins, eleven were signal-related proteins, six cytoskeleton-related proteins, five immune response-related proteins, four bacteriostatic proteins, 11 hypothetical human proteins and interestingly, a set of 7 bacterial proteins were identified in the sputum of active TB patients. Out of those 7 bacterial proteins MT3876 is a hypothetical MTB protein, which function is still unknown. Further validation of this MT3876 may be use as a biomarker for tuberculosis. \cite{33}

**COPD**

COPD is a common lung disease affects 8%-10% of adult of the developing World. \cite{34} COPD is a polygenic disease and characterized by the chronic air flow obstruction and a range of pathological changes in the lung, some significant extra pulmonary effects such as cardiovascular, mental, systematic inflammation, anemia and musculoskeletal comorbidities may be contributed to the severities of the disease in individual patients. \cite{35} The disease prevalence, morbidity, and mortality vary across countries and across different groups within countries. COPD is also influenced by other risk factors like deficiency of alpha-1 antitrypsin, a major circulating inhibitor of serine proteases and other direct effect of tobacco smoking, burning of wood and other biomass fuels. \cite{2,36} Sputum proteome of normal (healthy smokers) to chronic bronchitis, chronic obstructive pulmonary disease (COPD), and COPD
with emphysema were analyzed using CapLC-ESI-Q/TOF-MS identified a total of 203 host proteins. These proteins also showed evidence of disease progression from healthy to more advanced stage.

Zinc-α-2-glycoprotein, β-microsemino protein, cystatin S, and transthyretin shows significant variation in Smokers and Mild-to-Moderate COPD. In a separate study, sputum samples processed in a two-dimensional poly acrylamide gel chromatography, 1325 individual spots were identified in COPD and control subjects. Out of which 37 were quantitatively and 3 were qualitatively showed significant difference between these study groups. Fifteen proteins were identified from the 40 protein spots analyzed using tandem mass spectrometry analysis. Seven of these important proteins were further quantified in induced sputum from 97 study individuals. Using this sequential approach, two potential biomarkers i.e., apolipoprotein A1 and lipocalin-1 were reported to be significantly reduced in patients with COPD when compared with healthy smokers. Abundance of these important proteins was correlated with FEV₁/FVC, indicating their relationship to disease severity. In a recent study, 13 significantly deregulated proteins were identified in COPD patients using TandemMass Tag™6-plex (TMTsixplex™) and liquid chromatography tandem mass-spectrometry (LC–MS/MS) using an EASY-nanoLC 1000 instrument connected online to a QExactive™mass-analyzer (Thermo Scientific) . Among which Keratin, type I cytoskeletal 19, UPF0762 protein C6orf58, Metalloproteinase inhibitor 1, BPI fold-containing family B member 1, Peptidylprolyl cis-trans isomerase B are upregulated and Serotransferrin, Alpha-2-HS-glycoprotein, Antithrombin-III, Afamin, Serum albumin, Histidine-rich glycoprotein, Apolipoprotein A-I, Beta-Ala-His dipeptidase are down regulated.

Lung cancer

An estimated 1.8 million new lung cancer cases are reported every year accounting for 13% of all form of cancer diagnosis worldwide. It is the most common cause of cancer deaths. Lung cancer symptoms manifests usually very late or in the advanced stage with symptoms of persistence cough, sputum with blood, chest pain, voice change, shortness of breath and recurrent pneumonia or bronchitis. The prognosis for the patients with lung cancer is strongly correlated to the stage of the disease at the time of diagnosis. Whereas
patients with clinical stage IA, disease have a 5-year survival of about 60%, the clinical stage II-IV disease 5-year survival rate ranges from 40% to < 5%. 41 Over two-thirds of the patients have regional lymph-node involvement or distant disease at the time of presentation. 42 In sputum of lung cancer patients, majority of the protein compositions were from cancer sites. Approximately 85% of lung tumors are non-small cell lung cancers (NSCLCs), and it has two major histological subtypes: squamous cell carcinoma (SCC) and adenocarcinoma (AC). 43 Mortality reduction of Lung cancer depends on the development of noninvasive approaches for early detection of NSCLC followed by suitable treatments. Cytological and molecular studies of exfoliated bronchial epitheliums in sputum were more accurate to identify SCC tumors that largely located in central areas of the lungs. 44-46 Cytological techniques have low sensitivity and specificity for the diagnosis of ACs, which are the most common type in peripheral region of the lungs. However, a cell surface protein ENO1 was identified as a marker by 1D gel electrophoresis and LC/MS-based proteomics in sputum sample of lung cancer patients. Later it was validated using western blotting and ELISA. The upregulated ENO1 in sputum sample has similar sensitivity and specificity for diagnosis of both SCCs and ACs. So, ENO1 could be used as a potential biomarker for detection of SCCs and ACs. 47

Cystic Fibrosis

Cystic fibrosis (CF) is a fatal genetic disease with incident of 1 in every 2,500. CF is caused by mutation in gene encoding CF transmembrane conductance regulator (CFTR) with symptoms of persistence coughing with phlegm, wheezing, and shortness of breath. In CF due to chronic bacterial infection, frequent exacerbation the patient die of respiratory failure. 48 Bacterial colonization in airways results submucosal hypertrophy and excessive mucus secretion in lung. 49 CF progression over time is complex and is associated with infectious colonization of lungs, nutritional, environmental and social variables. Therapy depends on the symptoms, lung function and radiological changes which are lag behind the occurrence of established lung pathology. Induce sputum protein profiling from adult CF patients in comparison to healthy human and children with CF were studied using 2DE-PAGE and the protein spots with differential intensity were identified using MALDI-TOF. A set of protein panel including myeloperoxidase, α1-antitrypsin, IgG degradation, and total protein concentration in comparison with IL-8 showed differential expression which can be used as biomarker in lung exacerbation for diagnosis and prognosis of CF. 50

Asthma

In 2017, around 334 million people reported with symptoms of asthma around the world. Combination of genetic and environmental factors is mainly responsible for Asthma. Its diagnosis is usually based on the pattern of symptoms, response to therapy over time, and spirometry. Asthma is classified on the basis of the frequency of symptoms, forced expiratory volume in one second (FEV1), and peak expiratory flow rate. In asthma sputum hypersecretion contributes to the development of viscid mucus plugs that can occlude even the large airways and in chronic bronchitis it contributes to early morning coughing, sputum production and airways obstruction. In shotgun proteomics study 240 important proteins were identified from induce sputum of asthmatic patients. Most of these proteins were involved in defense response, protease inhibitor activity, immunity, response to inflammation and complement activation. Out of these 240 identified proteins, seventeen were differentially expressed between healthy and asthmatic group and involved in defense and stress response proteins from extracellular space. Among them calcium binding proteins S100A9 and S100A8, SERPINA1, SMR3B and
SCGB1A1. [51] Mouthuy et al. showed IgE levels are not related to disease severity but clearly increased in those exhibiting airway eosinophilic inflammation. The role of IgE has been traditionally assigned to allergic reaction towards an aeroallergen in sensitized patients. Humbert et al have drawn attention to the potential role of IgE in non-atopic asthma by showing increased expression of the receptor FceRI in the bronchial mucosa in asthmatics irrespective of the atopic status. [52] Here they found that GRO-α, eotaxin-2, and Pulmonary and activation-regulated chemokine (PARC) were increased significantly in sputum specimens from patients with asthma. In particular, PARC in the airways may play an important role in the eosinophilic inflammation of asthmatic airways. PARC is elevated in sputum specimens from patients with asthma and may play important roles in development of airway eosinophilic inflammation in asthma. [53]

CONCLUSION
Sputum as a diagnostic matrix for several respiratory disease conditions have showed high potential and may also be useful to monitor therapeutic outcome. Due to noninvasive mode of sample collection, it also possesses enormous translational value for clinical application. In this chapter we discussed the importance of different sputum proteins as marker for diagnosis of different respiratory disease conditions. The significance of recent advancements in the potential application of proteomic profile analyses of sputum has been highlighted in many common respiratory disease conditions. Sputum proteomics have been a novel approach in search for protein biomarker from different biofluids including sputum to detect human diseases. Comprehensive analysis and identification of the proteome content of sputum may also contribute to the understanding of the pathophysiology for respiratory disease conditions.

ACKNOWLEDGEMENTS

Competing interests
Authors have declared that no competing interests exist.

Authors’ Contributions
SCB, RKN and RKB wrote the review.

Consent (where ever applicable)
Not applicable

Ethical approval (where ever applicable)
Not applicable

REFERENCES
6. Rogers DF. Physiology of airway mucus secretion and pathophysiology of hypersecretion. Respir Care. 2007; 52(9):1134-49.
32. McNerney R, Daley P. Towards a point-of-care test for active tuberculosis: obstacles and

How to cite this article: Bishwal SC, Nanda RK, Behera RK. Sputum as a diagnostic matrix for respiratory disease screening. Int J Health Sci Res. 2018; 8(6):344-351.

***********