SYNOPSIS

Introduction

Asthma is a complex chronic inflammatory disease of the airways, characterized by the activation of many inflammatory and structural cells, all of which orchestrate the typical pathophysiological changes in asthma by the release of inflammatory mediators. According to the Global Initiative for Asthma Guidelines (GINA 2002), asthma is defined as a chronic inflammatory disorder of the airways and a plethora of genetic and environmental factors contribute to its pathogenesis. These conditions are variable and usually reversible in asthma. Each of the mechanisms of Asthma has multiple underlying potential causes. Therefore a fixed modus operandi for managing asthma leads to many patients being insufficiently diagnosed and wrongly treated, as many of the underlying mechanisms are left undiscovered or unaddressed. This has led to the suggestion that asthma is a syndrome comprising different types of sub-types that share a common general phenotype of episodic reversible bronchoconstriction, but may be mechanistically different.

According to the report of Global Burden of Asthma, it is now one of the world’s most common long-term conditions. Depending upon how the diagnosis of asthma is defined, prevalence varies from one region to another. 300 million individuals are estimated to suffer from asthma worldwide and by 2025; there may be an addition of 100 million more to this number. There are worldwide reports on its increasing prevalence, mortality, incidence and severity.

Several tools have been developed for diagnosing, monitoring and evaluating the lung status. Lung function tests are tools to provide insights into changes in airway calibre, flow, lung volumes and gas exchange. Imaging techniques, especially high resolution computed tomographic scanning, provide insight into changes in lung tissue composition. Both these methods are indirect measurements of adverse effects occurring in a biological system as a long-term consequence of exposure to toxic or noxious agents. In last decades, there has been an increased application of biomarkers with potential utility in the diagnosis and prognosis of asthma and other airway diseases...
like COPD. They have been used in monitoring the natural history of these diseases and the effect of therapeutic interventions. A biomarker has been defined as any measurement that predicts a patient’s disease state (a diagnostic or prognostic marker) or response to treatment (a clinical end point).

An important biomarker in asthma is eosinophilic inflammation which is also well recognized to be a major feature in the pathogenesis of asthma. In asthma, the level of Eosinophil peroxidase is seen to increase while Myeloperoxidase, which is expressed in neutrophils and monocytes, is goes up in COPD. The levels of 8-isoprostrane have been observed to be high in COPD and also in other respiratory diseases (cystic fibrosis, interstitial lung disease and asthma). Therefore it has limited specificity and sensitivity making it unreliable for assessing human subjects. Other biomarkers like H$_2$O$_2$, nitrotyrosine, leukotriene, serotonin, and cytokines (IL-1, sIL-2R, TNF-α) are also studied. Although the levels of many physiologic and inflammatory markers correlate with various clinical features of asthma and COPD, evidence supporting the use of any of these as valid targets for asthma and COPD management is limited. One of the reasons behind this can be that allergen induced bronchoconstriction is mostly characterized by biphasic responses known as early and late asthmatic responses (EAR and LAR respectively). Other than clinical markers of asthma severity such as symptoms and lung function, there is little else to guide clinicians as to when long acting β$_2$ agonist or other second line agent such as leukotriene receptor antagonist should be started, as these clinical markers are not sufficient to discriminate the early and the late state of disease.

The application of complementary approaches, including liquid and gas chromatography (LC and GC), protein microarray, nuclear magnetic resonance (NMR), mass spectrometry-based proteomic techniques on sputum, bronchoalveolar lavage, blood, and exhaled breath, may provide a better understanding of the proteome and/or metabolome differentially expressed by asthma patients in the course of the disease. Thus the identification of appropriate and reliable biomarkers is an essential step for the diagnostics and treatment. Also these hold the potential for identifying subtypes of asthma and providing insight into mechanisms.
The cellular mediators of the airways in health and disease can be sampled by Bronchoalveolar Lavage (BAL), Bronchial washing (BW), Nasal Scraping (NS) or bronchial biopsies. However these methods are invasive, cause discomfort to volunteer subjects, and are expensive and time consuming, making it difficult to obtain repeated samples. So the use of exhaled breath condensate (EBC), which samples the airway lining fluid, can be of great help in identifying biomarkers such as; lipid mediator, leukotriene and small molecules.

EBC is a relatively unexplored biological medium that could provide an assessment of lung pathobiology. This biological fluid is a convenient and relevant matrix in which biomarkers may be identified. Recently, there has been an increasing interest in using exhaled breath as a simple non invasive means to detect non volatile macromolecules (protein, lipids, oxidants, and nucleotide) and volatile molecules (acetic acid, formic acid, ammonia etc), which may represent biomarkers of various pathological processes in the lung. Highly sensitive and high-throughput techniques such as NMR spectroscopy and mass spectroscopy coupled GC and/or LC etc might help us provide insights into the metabolome. Metabolomic analysis (i.e. measurement of multiparametric response of metabolome in living systems to pathophysiological stimuli or genetic modification) can provide new biochemical profiles for the assessment of low molecular weight metabolites in EBC.

Specific and sensitive reporters of disease, (biomarkers) are of immense importance. The presence of disease or diagnosis of an otherwise clinically indistinguishable disease from related disease processes can be potentially possible with their help. So this treasure trove of diagnostic information can be a milestone for asthma and allergy.

An –omics approach involves studying different disciplines such as metabolomics, proteomics, transcriptomics, genomics, phenomics etc. It gives a wider perspective of looking into disease biology rather than studying individual disciplines.

Metabolomics is the study of different metabolites present in a system. It represents the multi parametric metabolic changes occurring as a response to pathophysiological stimuli.
Transcriptome deals with all the RNA species (miRNAs, siRNAs, snoRNAs, piRNAs etc) concerned with the biochemical and metabolic pathways in a cell. It marks a major event as there is direct transfer of information from the DNA to the RNA. Tapping the transcriptomic resources is extremely informative and provides insight into the major pathways involved in health and disease.

On the other hand Genomics is an approach that involves looking for genetic associations (both at SNP and CNV level) across the complete sets of DNA, or genomes, to find genetic variations associated with a particular disease. Once new genetic associations are identified, researchers can use the information to develop better strategies to detect, treat and prevent the disease. Health professionals will be able to use such tools to provide patients with individualized information about their risks of developing certain diseases. The information will enable health professionals to tailor prevention programs to each person's unique genetic makeup. In addition, if a patient does become ill, the information can be used to select the treatments most likely to be effective and least likely to cause adverse reactions in that particular patient. Thus looking from the genome wide perspective into the complexity of asthma might provide us with newer and broader avenues of developing a better understanding of the disease and preventing it.

Gathering of information from different sources does not help much in the understanding of complex biology. It is imperative to collate information from different sources and integrate them to make sense of the biology. Integration and modeling of these different disciplines to study the functional behavior is termed as ‘Systems Biology’ and such an approach could provide researchers with information hitherto unknown. Machine learning is best suited to make sense of large multivariate data in the understanding of complex biology.

Asthma, is a complex heterogeneous trait with multiple factors contributing towards its pathogenesis. There are endo-phenotypes of asthma with poor clinical diagnosis. Differentiating between cases and controls on the basis of markers which are easy to score and non-invasive is the need of the hour. A model that would be able to predict disease occurrence and exacerbations is highly valuable.
Identification of these different sub-phenotypes with distinct pathophysiology and clinical outcomes is imperative to personalized therapeutics and clinical management.

**Rationale and Significance of the Study**

With the advances in the high throughput technology large data sets are being generated from different platforms every day. Proteomics (Mass Spectrometry etc.), Metabolomics (Nuclear Magnetic Resonance Spectroscopy etc.), Phenomics, Transcriptomics and several other disciplines seem to add multifarious dimensions into the understanding of complex diseases. Teasing out all these different disciplines ultimately to make some sense would be ideal into unraveling complex pathophysiology.

In order to achieve such a holistic viewpoint and knowledge it is imperative that the multi-dimensional data from multiple ends be integrated and used together to make some sense rather than studying them independently.

It is believed that a complex system involves a close interplay of all these layering of disciplines and it is essential to understand the complex network that exists to generate them. Hence the requirement for integration of all these data using machine learning approaches. Such a systems approach would help us unravel the complexity of asthma and also assist in finding novel sub-phenotypes. This might also help us find some novel biomarkers that would help us predict the disease in patients beforehand.
Interaction of different –omes constitutes the specific physiome or functional entity of every individual groups.

Objectives of the Study

A. To differentiate between normals and asthmatics and to identify different sub-phenotypes of asthma based on omics approach:
   - Metabolomics
   - Proteomics
   - Transcriptomics

B. To investigate genetic differences and/or other factors that may contribute towards asthma pathogenesis.

Techniques and Experimental Strategies to be Used

Study Design

A prospective cohort is being built up in due collaboration with AIIMS with patients controls followed up on an every 3- month basis over a period of two years. A complete detailed clinical and molecular profiling is being performed for each patient which includes the following:

- Complete clinical history and data with every 3 monthly follow up.
- Immune profiling at baseline and during exacerbation.
- Blood for isolation of DNA for genomics
- Exhaled Breath metabolomics
- Nasal Scrapes transcriptomics at baseline and during exacerbation.
- Nasal Aspirate
- Skin Prick Test
- Spirometry andIOS
- Exhaled Nitric Oxide
- Urine at baseline and during exacerbation.
Experiments with the Exhaled Breath Condensate (EBC)

Exhaled breath condensate (EBC) is condensed from water but also contains droplets that carry solute from the lower respiratory tract. The collection of EBC by cooling exhaled air has several advantages compared to many other methods of sampling the airspaces. The main advantage of working with exhaled breath is that it is non-invasive, relatively inexpensive and could be collected multiple times from subjects without discomfort. So far, it has no reported role on airway inflammation or any significant influence on airway function, making repeated sampling possible. Because it is effort-independent, children and even mechanically ventilated patients can participate in the collection. Very little is known actually regarding the utility and importance of EBC and Nasal scrape as matrices for studying asthma. Small molecule and protein signatures obtained from these matrixes in patients and controls would help us gain insights into the understanding of the disease.

Evidences suggest that the degree of asthma and COPD severity can be reflected by the abnormalities in biomarkers present in EBC, even better than what can be assessed by either spirometry or symptoms. The major drawbacks of EBC, in spite of reflecting the composition of the extracellular lining fluid (ELF) are contamination of saliva, anomalies due to dilution, and variability. There exists a lack of evidence of the primary origin of the aerosol particles that carry the droplets from the lower respiratory tract. The plethora of potential markers that cannot be consistently or easily analyzed in these highly diluted samples pose major limitation in the use of EBC.

How it can add to Existing Knowledge Related to Human Health

- Identification of novel biomarkers can help in detection of pre-clinical disease and enable better patient management.
- Asthma is a heterogeneous and multifactorial syndrome with many ill-defined endo-phenotypes. Hence identification of these different sub-phenotypes with distinct pathophysiology and clinical outcomes is imperative to tailored therapeutics and clinical management.
- Identification of an indicator of severity and monitoring state of health.
Collection of EBC

EBC was collected using a commercial EBC collection device the RTube.

A. To Differentiate between Normals and Asthmatics and to Identify different Sub-phenotypes of Asthma based on Omics Approach

- **Metabolomics**

  **Nuclear Magnetic Resonance (NMR)**

  1-D H-NMR experiments were performed using Exhaled breath Condensates from asthmatics and healthy subjects.

  NMR spectra was obtained using FID files in iNMR software and analysed manually for presence of peaks.

  Routine peak fitting was done using Chenomx software.

- **Transcriptomics**

  Exhaled breath miRNA signatures were looked into in patients with asthma and healthy volunteers. Profiling of about 1100 miRNAs were done using miRNome array (System Biosciences).

**Nasal Scrape Transcriptome**

Nasal scrape was isolated from inferior turbinate of nose from subjects. RNA was isolated and c-DNA micro array was performed using Illumina platform. Control subjects and asthmatics were compared.
Phenomics

An array of different tests starting from anthropometric measures, blood pressure, current symptoms such as breathlessness, wheezing etc., patient history, skin disease, skin prick test etc. were performed and recorded. These phenomic data was used in the machine learning analyses to predict for sub-phenotypes of asthma.

Bioinformatics (Maching Learning)

Use of Random Forest etc, state of the art machine learning (ML) techniques to differentiate between cases and controls. ML was used to differentiate the different sub-phenotypes of asthma and also to find out an array of clinical tests that could be easily done at clinic settings that could reflect similar resolution.

B. To Investigate Genetic Differences and/or other Factors that may contribute towards Asthma Pathogenesis

Genome Wide Association Study was performed on pediatric cohort with control samples (145 cases and 155 healthy controls) to look for genomic/genetic associations involved in asthma.

Results

NMR Spectroscopy of EBC Identifies Ammonia to be Different in Cases and Controls

1-D NMR spectroscopy of exhaled breath identified a compound (a trident peak) at 7 ppm region that could distinguish asthmatics from healthy subjects with high confidence.

The compound was identified to be ammonia.

Ammonia is produced in airway epithelial cells due to an enzyme called Glutaminase, which acts on Glutamine to convert it into Glutamate and ammonia. This ammonia neutralizes the acid load in the airways.

In healthy subjects the glutaminase levels are high indicating that the stoichiometry is intact, whereas in asthmatic subjects the levels of glutaminase are very low, thereby low ammonia levels and high airway acid load.
High Levels of Blood Glutaminase in Healthy Subjects vs. Asthmatics

Levels of glutaminase were checked in sera of subjects with and without asthma. It was found that the levels of glutaminase varied significantly amongst them (upregulated in healthy subjects compared to asthmatics).

Leads from Transcriptomics

Another interesting study which requires a mention in omics study is transcriptomics. Nasal scrape from inferior turbinate of nose were collected and RNA isolated from them, c-DNA micro array was performed.

Asthma Reflected as one of the most Important Pathways from Nasal Scrape Micro Array

c-DNA micro array from nasal scrape of subjects with asthma and healthy controls revealed many differential genes. When these differentially regulated genes were looked for pathways enrichment using DAVID, Asthma came out as one of the most significant indicating the usefulness of nasal scrape as a matrix.

miRNome of Exhaled Breath

miRNAs are very stable and robust RNA species found in almost all body fluids and is known to modulate cellular processes. The presence of miRNAs could be reported for the first time in exhaled breath using real time PCR analysis.

10 asthmatic subjects and healthy volunteers were included in the study and about 1100 miRNAs were profiled using miRNome profiling kit (System Biosciences).

The miRNAs were normalized and analysed using novel strategies. The case and control cohort differed in terms of differentially expressed miRNAs although modestly.
The differential miRNA list includes the following:

<table>
<thead>
<tr>
<th>Upregulated in Asthma</th>
<th>Downregulated in Asthma</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-2861</td>
<td>hsa-miR-556-5p</td>
</tr>
<tr>
<td>hsa-miR-574-5p</td>
<td>hsa-miR-4256</td>
</tr>
<tr>
<td>hsa-miR-595</td>
<td>hsa-miR-762</td>
</tr>
<tr>
<td>hsa-miR-1264</td>
<td>hsa-miR-516a-5p</td>
</tr>
<tr>
<td>hsa-miR-649</td>
<td></td>
</tr>
<tr>
<td>hsa-miR-624</td>
<td></td>
</tr>
<tr>
<td>hsa-miR-421</td>
<td></td>
</tr>
</tbody>
</table>

Pathway Enrichment of Differential miRNAs

The targets of these miRNAs were predicted using MiRanda and RNA Hybrid. The predicted targets were looked for pathways enrichment. Some of the most significant pathways that were reflected were all known to be involved in allergy, inflammation or asthma.

Exosome Enclosed miRNAs in Breath

The fact that miRNAs are stably expressed in breath indicates some sort of protection being conferred to them. We hypothesized that these miRNAs are enclosed within exosomes.

Presence of exosomes was therefore probed in concentrated EBC samples. Latex beads (singlets) coated with anti-CD-63 antibody (exosome specific) were used to bind exosomes and pelleted to separate the exosomal fraction. We found that exosomes are present in EBC and contain the majority of the miRNA content. This is extremely relevant, for two reasons. EBC miRnome reflects ongoing biological processes since exosome secretion and exosome contents are highly coordinated and regulated. It also signifies that scoring of EBC miRnome will be a robust strategy, as it is robust and does not require care during sample storage and transport.
Use of Machine Learning to Make Sense of Metabolomic Data

Data preprocessing: NMR data was obtained in FID form, converted to frequency domain using Fourier Transformation.

The data was normalized, binned dynamically and used for analyses.

Use of Random Forest to Differentiate between Cases and Controls

Random Forest works at 2 levels basically. One being Supervised learning and the other one is Unsupervised learning. In supervised learning the class labels are provided to the computer. The computer partitions the data into 2/3rd and 1/3rd part. The machine builds a model on the 2/3rd data and predicts on the remaining 1/3rd and thereby calculates the efficiency/accuracy of the model. Unsupervised learning is a technique where in class labels are not provided to the computer and it tries to group and separate objects based on similarity and differences amongst them. In sub-phenotype discovery problems unsupervised is the choice of learning algorithm.

89 asthmatics ad 21 controls were taken to build a model using Random Forest algorithm to distinguish health and disease.

The algorithm used 2/3rd data set to predict the model and tested the efficacy of the model using the remaining 1/3rd.

Initially the prediction accuracy was not very high but subject to proper tuning and optimization strategies the error rate decreased. The model could finally predict asthmatics from healthy subjects with ~85% Sensitivity and 81% specificity.

Integrated Use of Multi-dimensional Data to Predict Asthma Endo-phenotypes

Unsupervised clustering algorithm was employed using RF technique to create proximity measures. Clustering was performed using Partitioning Around medoids method using proximity scores.

RF could predict 3 clusters with 80 % accuracy.
Significance of Clusters

The 3 clusters that were predicted majorly differed amongst themselves by the presence of 2 compounds:

- Urea
- N, N Dimethylglycine

Other compounds that differed in between the clusters include:

- Propionate
- Acetate
- Formate
- Methanol etc.

Clinical Correlates of Molecular Sub-phenotypes

Next the entire array of clinical parameters was used to predict clusters in asthma.

From the list of several clinical values some of the not important ones were given by RF that could predict the same clusters as predicted by NMR data.

Thus we could arrive at a set of the few most important clinical parameters that could correlate with molecular data. While there were many parameters, the most understandable clinical correlate was that the clusters corresponded to varying degrees of allergic inflammation (judged by FeNO) and obstruction (judged by forced expiratory volumes), such that one cluster had high allergic inflammation but minimal obstruction, another with obstruction but minimal allergic inflammation, and the third had the highest levels of obstruction as well as allergic inflammation. Such clusters are also seen in clinical practice and understanding the molecular correlates may provide insight into the pathobiology.

<table>
<thead>
<tr>
<th>CLUSTERS</th>
<th>FeNO</th>
<th>POLY</th>
<th>MEF75</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SEM</td>
<td>Mean</td>
</tr>
<tr>
<td>C1</td>
<td>23.15152</td>
<td>±3.524392</td>
<td>56.0625</td>
</tr>
<tr>
<td>C2</td>
<td>12</td>
<td>±1.65328</td>
<td>48.5</td>
</tr>
<tr>
<td>C3</td>
<td>19.6087</td>
<td>±2.141016</td>
<td>59.95652</td>
</tr>
</tbody>
</table>

(C1): Allergic; Limited Airway Disease; (C2): Mild Airway Obstruction; minimal allergy Neutropenic? → Viral Infection; (C3): Allergic; Significant Small Airway Obstruction
The hypothetical clinical clusters C1, C2, C3 based on clinical data which has molecular correlates.

In a separate study, another group of researchers could also identify three sub-phenotypes of asthma based on clinical parameters which correspond to the 3 clusters found in our study based on breath metabolome. This proves the relevance of the clusters identified and speaks volumes of the resolution offered by such molecular approaches.

**Genomics: To Look for Genomic/Genetic Factors Contributing towards Asthma Pathogenesis**

An initial case – control GWAS was piloted on about 145 cases and 155 control subjects using Illumina HumanOmni1-Quad.

Initial results show 8 most important SNPs after multiple comparison corrections (FDR).

Some of them are already known to be involved in asthma and some are known to target genes involved in asthma.

Although interesting, these findings need to be proved and validated in larger cohorts to increase the power of the study and the significance of associations.

**Conclusion**

Modern day medicine involves a “systems approach” to treatment and aims at stratified therapies for patients.

A “systems biology” approach to a complex biological question involves the integration of knowledge from different disciplines to understand the mechanism. Asthma is a similar problem with high complexity and heterogeneity. Hence an –omics approach is best suited for such a cause.

It helps gain resolution into the disease biology and also helps identify endo-phenotypes of the disease with distinct clinical symptoms and pathophysiology.