Using ‘Big Data’ approaches to gain new insights into asthma

By combining fibre optic camera technology with the mathematics of ‘Big Data’ from Silicon Valley, researchers at Southampton have been with gaining new insights into the nature of the world’s commonest long-term lung disease: asthma. Up to 300 million people are affected worldwide of which 1 in 20 suffer severe disease which responds poorly to current treatments. There remains no cure, and a major barrier to research has been a poor understanding of the different forms of asthma, and of the specific problems with the immune system which underlie each form.

Fig. 1. A Bayesian Belief Network showing the strongest interactions between different clinical and immunological measurements. Line thickness represents the strength of interaction (Euclidean distance). Line colors: green, positive associations; red, negative associations; black, non-linear associations. Abbreviations: BMI, body mass index; ICS, inhaled corticosteroids; IL, interleukin; MAIT, mucosal associated invariant T-cell; Tc1, CD8+IFN-?+ T cell; Tc2 CD8+IL-13+ T cell; Th, T helper cell; Treg, regulatory T-cell.

Over the last 4 years researchers at the Southampton Biomedical Research Unit studied a group of 84 volunteers who suffered from asthma or who were healthy. They examined the immune system of their lungs in great detail using tests of their lung function, of their allergies and by taking samples from their airways. They did this by taking samples of sputum from volunteers who had breathed in a fine mist of very salty water, and by using a flexible fibreoptic camera to take tiny 2mm samples of the airways deep inside their lungs. The samples were then broken down using enzymes and stained with multi-coloured antibodies to be analysed by a cytometer that uses three different coloured lasers to count different cells at a rate of 10000 per second. Having collected over 100 different measurements on each person the challenge was how to make sense of the data.

The team pioneered a new approach to this analysis. Using ideas originally developed by the 18th century Presbyterian minister Thomas Bayes, the complex interactions between different measurements can now be modelled as a 'Bayesian Network' (Fig. 1) similarly to how financial analysts model fluctuating share prices in stock markets. The study revealed the extent to which the symptoms of asthma depended on the interplay of a wide range of different factors including 'clinical' factors – tobacco use, nasal polyps, hay fever, body mass – and immunological factors – particularly underlying the importance of the somewhat forgotten 'mast cells' and two novel classes of white blood cell called T helper-17 (Th17) cells and mucosal associated invariant T (MAIT) cells. MAIT cells were unexpectedly low in severe asthma.

By combining this information with samples from when the volunteers were suffering from an asthma attack their findings overturned a widely held theory that one of these components – called 'Th17 cells' – play a major role in asthma. This knowledge may help avoid conducting costly trials of medicines targeting this immune component. It also underscores the need to develop better techniques to study asthma in the laboratory using lung samples from humans rather than animals alone.

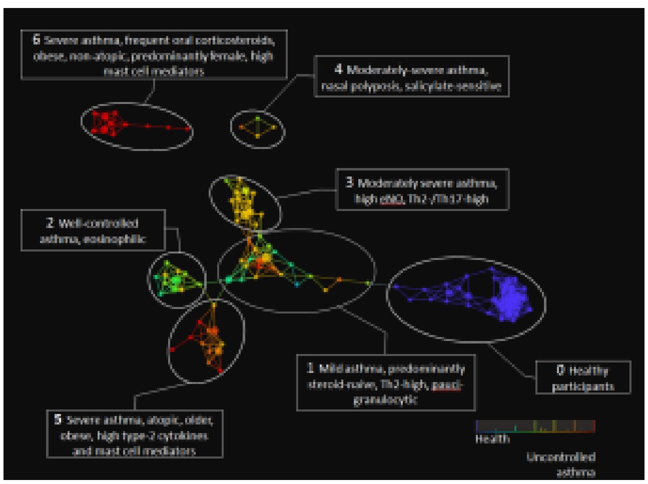

Fig. 2. Topological data analysis of clinical and immunological features generates one healthy (blue) and six distinct multi-dimensional clusters in. The network is colored by disease severity (GINA classification), with most severe subjects in red and the other, milder forms in varying shades of orange, yellow and green.

To tackle afresh the problem of working out what different sorts of asthma exist the researchers teamed up with a company from Palo Alto (Ayasdi Inc) to pioneer the novel approach of 'Topological Data Analysis'. Instead of asking a string of questions of a dataset, this new approach uses machine learning algorithms to plot the data as a three dimensional 'network' (Fig. 2), analysing potentially thousands of measurements simultaneously and allows features within the data to emerge, letting the data speak for itself. In this case six different forms of asthma became apparent. Some were previously well recognised (groups 1,3 and 4) but others had not been defined previously and showed two different forms of severe asthma. ‘Mast cells’ play a role in both of these, but type-2 cytokines – whose importance in asthma is well recognised – were only increased in one of these two groups.

This approach paves the way for larger studies, now ongoing, to define the precise nature of distinct forms of asthma.

This work was funded by the Wellcome Trust

Publications

Multidimensional endotypes of asthma: topological data analysis of cross-sectional clinical, pathological, and immunological data.
Hinks T, Zhou X, Staples K, Dimitrov B, Manta A, Petrossian T, Lum P, Smith C, Ward J, Howarth P, Walls A, Gadola SD, Djukanovi? R
Lancet. 2015 Feb 26

Innate and adaptive T cells in asthmatic patients: Relationship to severity and disease mechanisms.
Hinks TS, Zhou X, Staples K, Dimitrov B, Manta A, Petrossian T, Lum P, Smith C, Ward J, Howarth P, Walls A, Gadola SD, Djukanovi? R.
J Allergy Clin Immunol 2015