

ECLIPSE analysis proposal

ECLIPSE

Chronic Obstructive Pulmonary Disorder (COPD) is a preventable and treatable disease state characterised by airflow limitation that is not fully reversible. The disease is most commonly caused by smoking. The ECLIPSE (Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints) study is a non-interventional, observational, multicentre, three-year study in 3000 people with COPD. Samples have been collected at baseline and a further seven times during the 3 year timespan. Blood samples were collected for all 3000 individuals and sputum samples for approximately 500 individuals. A subset of these samples have been studied using microarray analysis and are known as the 'transcriptomics cohort'. The transcriptomics cohort is the focus of this proposal.

Website: <http://www.eclipse-copd.com/>

Transcriptomics Cohort from ECLIPSE

At baseline, blood and sputum samples were taken from a study cohort of 222 age and gender matched individuals. The cohort comprised three groups: 74 never smoking controls, 74 former smokers COPD GOLD stage 2, and 74 former smokers COPD GOLD stage 3(60) and 4(14). Blood samples were taken from all three groups, sputum samples were taken from only GOLD2 and GOLD3/4 groups. RNA was isolated from these 370 samples and, following labelling, were hybridised to Affymetrix plus 2 transcriptomics arrays. Clinical data with 2300 variables was also collected for this cohort.

Proposal

A 12 week study to analyse the combined transcriptomics and clinical dataset. The aim of the project will be to identify clinical and transcriptional signatures that characterise each of the three groups and may identify additional subgroups. The project will also investigate alternative groupings that may be generated based on the dataset. A critical component of the study will be to determine if there are signatures (transcriptomic and/or clinical) that predict elevated risk of exacerbation events. Time permitting, the student will follow up with pathway analysis and other datamining approaches to investigate potential biological mechanisms/processes that may underpin these groups and which are reflected in the clinical and transcriptomics features that characterise them.

It is envisaged that the student will consider and use a variety of supervised and unsupervised clustering approaches, which may include machine learning, SVM, etc. The student will therefore need to be proficient with at least some of these approaches, and have access to the methods/tools through the public domain. The dataset that will be utilised will be deposited in the public domain at an unknown time in the future but at present remains proprietary to GSK. Any manuscript or report written that utilises these data must be approved by GSK in advance. In the unlikely event that any IP issues arise at the time the student must submit their report, GSK will ensure that the examiners can gain access to the report, if necessary through a confidentiality agreement. As the data is ultimately going into the public domain, if any findings from the study merit publication, GSK will ensure authorship for the student.