ABSTRACT
The use of tree harvesting can enable the analysis of gene expression data and the discovery of intriguing clusters of genes that require further investigation.

INTRODUCTION
The purpose of this work is to provide an introduction to the topic of supervised harvesting of expression trees (SATs). In this paper we present the results of the first successful example of a SAT that was successfully harvested from a large group of human subjects (n = 35). For the purposes of this paper we will refer to the subject as the participant. The subjects were recruited from the University Hospitals of Manchester (UK) and St. Mary's Hospital (UK) NHS Trusts (the UK and UK). The subject's participation was voluntary and the study was approved by the University's Research Ethics Committee. The subject's first assignment was to take part in a series of experiments designed to generate a sample of 1,000 SATs. Subjects were instructed that the experimental procedure was intended to generate large numbers of SATs and that the samples were to be collected in The paper introduces 'tree harvesting' - a general method of supervised learning from gene expression data. It consists of taking real-valued expression measurements for thousands of genes over x set of samples (typically 50 or 100), and measuring the outcome according to varying models such as survival time or cancer class). The task of predicting an outcome measure from a set of features is known as'supervised learning'. If the outcome is quantitative, it is called regression, while if it was categorical, -classification' is used. There are several available methods for supervised learned techniques, such as linear regression, discriminant analysis, neural networks, support vector machines, and boosting, but they do not work well... due to the high complexity of processing expression data with large numbers of inputs (genes) and their high correlation. Initially, we cluster the genes using hierarchical clustering and then use the average expression profiles from all the clustered genes as inputs to our prediction model. This approach has two benefits: firstly, it is a common practice in interpretation of expression data, which means that by harvesting these clusters, the components of our predictive model can be conveniently interpreted; and secondly because we are able to bias the inputses towards correlated sets of genes, thus reducing the rate at which the model is overfitted. In particular they are larger cluster size. Continuing below, we describe the fundamental technique with a quantitative output and squared error, then expand upon other conditions such as survival data and qualitative responses. Two real examples illustrate tree harvesting, while supplementary information on simulation studies is presented to explore the effectiveness of the method. Finally, nonlinear expression effects are permitted in tree extracting.

CONCLUSION
Remarkable conclusions A general and promising approach to learning from gene expression data is the tree harvest procedure, which involves discovering the additive and interaction structure of clusters of genes in relation to an outcome measure. However, this method requires significant sample sizes, and as such, we anticipate experimenting with tree extracting on larger datasets. We decided to use a forward stepwise strategy involving the sum AND products of the average gene Expression of chosen cluster(s) to produce interpretable, biologically plausible models. Other models could be constructed from the combined gene-expression of

other cluster model using methods like Tree-based methods or enhancing Further details are available. The online version of this article contains supplementary data, such as clusters from the harvest model applied to lymphoma data.