Report OMICS Project

Iris, Jacqualine and Sudip

2019-10-20

Contents

List of Figures

List of Tables

1. Background

Atopic Dermatitis (AD) ia a very common skin disease.

2. Methodology

2.1 Data Management Plan

What we did:: We performed following tasks in the given data set We removed the duplicate MAARS_identifier from the data set which are unique patient identifier. We also checked for normalization of the data set by plotting the histogram and box-plot of the normalized gene expression data set. Removed samples that did not have transcriptome data. We categorized the SCORAD score to mild (<25), moderate (25-50) and severe (>50) based on their value.

2.2 Statistical Analysis Plan

Analysis Objective:

To predict the severity of atopic eczema (atopic dermatitis, AD), as measured by the SCORAD Score, by the transcriptome.

To find out the relation of other clinical characteristics to the severity of AD.

Analysis Sets/ Population/Subsets

The study population consists of adult patients aged 18-70 years with mild to severe AD, plaque-type PSO (PASI score > 7) and healthy volunteers (HV). The patients who were on systematic antibiotic within two weeks, on systematic immunosuppressive therapy or systematic phototherapy or biological agents within the previous 12 weeks prior to screening and patient with concomitant autoimmune disease (eg: diabetes, rheumatoid arthritis) were excluded from the study.

End Points and covariates

The endpoint of the study is severity of AD (given by SCORAD score). We plan to conduct two analyses, one for SCORAD score as continuous endpoint and one for SCORAD classified into categories mild, moderate, and severe.

Handling of Missing values and other data convention

There are no missing values in the dataset.

Statistical Methodology

Statistical Procedure:

What we did::

First of all we performed some descriptive analysis to understand the demographic characteristics of the people in the study. Perform hierarchical clustering analysis to understand hierarchical relationship between the genes

What we are planning::

To look at the other clinical variables and their interaction on gene expression using sPLS method. To predict the specific genes associated with different lesional status of AD (i.e. comparison of lesional and non-lesional skin samples), and genes associated with AD in general (i.e. comparison between AD patients and controls) with R packages limma and edgeR {same procedure for psoriasis patients and control groups} Genes with an adjusted p-value of 0.05 and fold-change > 1.5 are selected We are also planning to select genes associated with AD using sPLS and WGCNA Final selected genes will be analyzed for their biological function using tools such as GO, GSAn, David Measures to adjust for Multiplicity, confounders, Heterogeneity etc: For multiple testing corrections, Benjamini-Hochberg method will be used, which will decrease the false discovery rate. For comparison also Holm correction

Sensitivity analysis::

Compare the results obtained by using limma package and edgeR package of Bioconductor R. Compare different test corrections, and the results obtained by SCORAD as continuous and categorized variable Different test multiplicity adjustment methods and fold-change cutoffs

Validation plans::

Cross validation, bootstrapping