The Expression of ‘GSE40274\_XBP1\_VS\_FOXP3\_AND\_XBP1\_TRANSDUCED\_ACTIVATED\_CD4\_TCELL\_DN’ Genes is Associated with Asthma Severity.

Hans Owen

Student ID: 31090613

# Introduction

Research into genetic associations with certain diseases is currently a widespread goal in the field of bioinformatics. Asthma is a respiratory condition that causes the airways and lungs to swell and narrow, additionally increased mucus is produced. Symptoms include shortness of breath, wheezing, and coughing with varying severity (WHO, 2020). The data used to investigate the relationship between asthma and genetics comes from the U-BIOPRED study, which is a 5-year, European-wide research project that is dedicated to understanding different types of asthma and improving asthma diagnosis and treatment, through the use of information and samples collected from adults and children (European Lung Foundation, 2008). To compare genetic correlations with asthmatics, blood gene expression profiling was used. This is the analysis of peripheral human blood samples, which provide useful transcriptional information for human diseases (Vartanian, et al., 2009); by measuring the activity of the genes within the blood, unique gene activity in severe asthmatics can be identified as possible associative factors for asthma severity. The gene pathway was identified as enriched in differentially expressed genes between cohorts by microarray analysis, where fluorescent markers are attached to DNA fragments and these react with DNA probes, the remaining DNA is removed and the target DNA is then identified by fluorescent emission (Govindarajan, et al., 2012). The genetic pathway being investigated in this essay is ‘GSE40274\_XBP1\_VS\_FOXP3\_AND\_XBP1\_TRANSDUCED\_ACTIVATED\_CD4\_TCELL\_DN’. This gene pathway is involved in immunologic signatures and is a combination of genes that are involved with the specification and function of T regulatory cells (Fu, W., et al., 2012). This investigation explores whether the gene expression of this gene pathway is associated with asthma severity.

**[250 Words]**

# Method

## GEO2R Analysis

This investigation involved gathering and manipulating data from databases using Excel and then analysing a logistic regression. The data from the expression profiling in blood from severe asthmatics, moderate asthmatics, and non-asthmatics collected in the U-BIOPRED study were analysed with GEO2R from the GEO entry (GSE69683) (Bigler, et al., 2016). Two cohorts were created: (non-smoking) moderate and severe asthmatics.

## GSEA Analysis

The top 3000 genetic data with the lowest adjusted p-value from GEO2R analysis were copied into gene set enrichment analysis (GSEA). The top 100 gene sets that overlapped between the molecular signature database (MSigDB v7.2) C7 immunologic signature gene sets and the GEO entry were taken. The complete gene expression data was downloaded separately from the GEO entry and the gene name list replaced the list of Affymetrix microarray gene probes, creating a patient and gene expression matrix with gene names. The genes in the pathway GSE40274\_XBP1\_VS\_FOXP3\_AND\_XBP1\_TRANSDUCED\_ACTIVATED\_CD4\_TCELL\_DN were copied to the patient and gene expression matrix and the expression data for the gene pathway were identified.

## Python ORANGE

In Python ORANGE five operations were carried out using the gene expression data. Firstly, a logistic regression classified a patient into either cohort using probability, based upon their gene expression data. Secondly, a test and score evaluated the performance of the logistic regression. Thirdly, a receiver operating characteristic (ROC) curve graphed the true positives (y-axis) against false positives (x-axis) for all probability discrimination thresholds of the logistic regression (0 – 1) (Hoffman, 2019). Then a scatter plot illustrated two data subsections on an x, y-axis. Finally, a confusion matrix returned the number of true/false negatives and positives predicted by the logistic regression (Diez, 2018).

**[250 Words]**

# Results

## ROC Analysis of Logarithmic Regression of the Moderate Non-Smoking Asthmatics Cohort from U-BIOPRED Study (GSE69683) Based Upon Gene Expression in GSE40274\_XBP1\_VS\_FOXP3\_AND\_XBP1\_TRANSDUCED\_ACTIVATED\_CD4\_TCELL\_DN Pathway

Chart, line chart

Description automatically generated

Figure : Receiver operating characteristic (ROC) curve for the logarithmic regression of the gene expression data of the pathway: GSE40274\_XBP1\_VS\_FOXP3\_AND\_XBP1\_TRANSDUCED\_ACTIVATED\_CD4\_TCELL\_DN from the non-smoking moderate asthmatics cohort of the U-BIOPRED study: Expression profiling in blood from subjects with severe asthma, moderate asthma, and non-asthmatics (GSE69683).

The central line represents a random sample and the curve to the left indicates that the logistic regression predicts more true positives than false positives based on the gene expression data. The ROC curve has an area under the curve (AUC) of 0.76, suggesting the logistic regression machine learning algorithm has fair predictive power for asthma severity in patients of the UBIOPRED study based upon their gene expression data of the GSE40274\_XBP1\_VS\_FOXP3\_AND\_XBP1\_TRANSDUCED\_ACTIVATED\_CD4\_TCELL\_DN pathway.

## Scatter Plot Comparing the Gene Expression Data of the Gene ZNHIT6 and the Logistic Regression Predicting for Severe Asthma in the Non-Smoking Moderate and Severe Asthmatic Cohorts



Figure : Scatter plot mapping patients from the moderate asthma non-smoking cohort and the severe asthma non-smoking cohort from the UBIOPRED study (GSE69683) based upon the gene expression data of the ZNHIT6(1) gene from the gene pathway: GSE40274\_XBP1\_VS\_FOXP3\_AND\_XBP1\_TRANSDUCED\_ACTIVATED\_CD4\_TCELL\_DN against the logistic regression classifying for severe asthma.

The gene within the gene pathway GSE40274\_XBP1\_VS\_FOXP3\_AND\_XBP1\_TRANSDUCED\_ACTIVATED\_CD4\_TCELL\_DN that shows the greatest association with asthma severity according to the rank-ordered scattered plots in python orange is ZNHIT6(1). Those who had moderate asthma typically showed higher levels of ZNHIT6(1) gene expression (see fig.2). Moderate asthmatics expressed the ZNHIT6(1) between 9.8-8.2, whereas severe asthmatics expressed generally lowermost from 9.4-7.6. However, a high degree of overlap between the two cohorts is observed and therefore this gene is not a causal factor for asthma severity.

## Confusion Matrix for the Logistic Regression Classifying U-BIOPRED Patients into Severe and Moderate Asthmatic Cohorts

A picture containing graphical user interface

Description automatically generated

Figure : Confusion matrix for the logistic regression classifying severe and moderate non-smoking asthmatic cohorts based upon their gene expression data of the GSE40274\_XBP1\_VS\_FOXP3\_AND\_XBP1\_TRANSDUCED\_ACTIVATED\_CD4\_TCELL\_DN gene pathway from patients in the U-BIOPRED study (GSE69683).

The confusion matrix shows that the logistic regression classified 24 moderate asthmatics correctly and 35 severe asthmatics as moderate asthmatics incorrectly. The logistic regression classified 211 severe asthmatics correctly and 53 moderate asthmatics as severe asthmatics incorrectly. This suggests the logistic regression can differentiate between moderate and severe asthmatics as many true positives for severe asthmatics was predicted. The logistic regression’s ability to predict moderate asthma appears poor, however this could be attributed to the lower sample size of moderate asthmatics and the large degree of overlap.

**[257 Words]**

# Discussion

T regulatory cells (Tregs) are crucial in controlling allergic diseases, including asthma. These cells are responsible for regulating the levels of T helper cell type 2 (Th2), which causes asthmatic inflammation of the airway due to an excessive immune response (Zhao & Wang, 2018). The genetic pathway GSE40274\_XBP1\_VS\_FOXP3\_AND\_XBP1\_TRANSDUCED\_ACTIVATED\_CD4\_TCELL\_DN is partially responsible for the transcription signature of Tregs through the transcription factor protein FoxP3 (Fu, et al., 2012). Based upon the logistic regression’s fair predictive power between moderate and severe asthmatics derived from their gene expression data of the gene pathway GSE40274\_XBP1\_VS\_FOXP3\_AND\_XBP1\_TRANSDUCED\_ACTIVATED\_CD4\_TCELL\_DN, this gene pathway has a fair degree of association with asthma severity in non-smoking asthmatic patients from the U-BIOPRED study. This could be attributed to the FoxP3 transcription factor protein being non-essential for Treg expression as other transcription factors, such as Eos, Irf4, Satb1, Lef1, and Gata1, elicit the same effect as FoxP3. A synergic combination of these transcription factors is responsible for the majority of Treg expression. Moreover, the Treg transcription signature remains stable after inactivation of any single cofactor (Fu, et al., 2012). From this information, we can conclude that solely the gene expression data of the gene pathway GSE40274\_XBP1\_VS\_FOXP3\_AND\_XBP1\_TRANSDUCED\_ACTIVATED\_CD4\_TCELL\_DN is not reliable enough for clinical diagnosis of asthma severity. Moreover, this gene pathway does not provide an obvious biological mechanism for the treatment of asthma severity.

However, studies do show reduced expression of the FoxP3 gene in asthmatics and those with allergies relative to healthy controls (Marques, et al., 2015). Therefore, the gene pathway still has some responsibility for asthma severity despite the many influencing factors. Based on this evidence, further investigation into the gene expression data for multiple gene pathways responsible for the transcription factor proteins that influence the expression of Tregs could lead to a more comprehensive machine learning algorithm that could accurately classify moderate and severe asthmatics based upon their gene expression data. If such an algorithm proved to be reliable in its classifications it could be applied to clinical diagnosis and possibly identify specific biological mechanisms that could enhance our understanding of asthma severity treatments. Evidence from studies indicates environmental factors and epigenetic influences play a role in the expression of FoxP3. A study conducted by Nadeau et al. investigated the epigenetic factors that influence the FoxP3 transcription factor, diesel exhaust fumes and second-hand exposure to smoke increased methylation of the FoxP3 gene and reduced the expression of the FoxP3 protein. The hypermethylation of the FoxP3 gene correlates with increased asthma and wheezing in early childhood (Marques, et al., 2015). This supports the idea that the gene pathway GSE40274\_XBP1\_VS\_FOXP3\_AND\_XBP1\_TRANSDUCED\_ACTIVATED\_CD4\_TCELL\_DN has an association with asthma and also that the expression of this gene pathway can be heavily influenced by epigenetic factors. An investigation into reducing the hypermethylation of these genes in this gene pathway could present a possible treatment for asthma severity, as it would allow for increased expression of the FoxP3 transcription factor, which in turn would increase the expression for Tregs that counteract the Th2 cells responsible for asthmatic inflammation of the airways.

**[500 Words]**

# References

Begin, P. & Nadeau, K. C., 2014. Epigenetic regulation of asthma and allergic disease. *Allergy, Asthma and Clinical Immunology,* 10(1), p. 27.

Bigler, J., Timour, M. & M, B., 2016. *Expression profiling in blood from subjects with severe asthma, moderate asthma, and non-asthmatics collected in the U-BIOPRED study,* Seattlw: NCBI Gene Expression Omnibus (GEO).

Diez, P., 2018. *Smart Wheelchairs and Brain-Computer Interfaces.* 1st ed. s.l.:Academic Press.

European Lung Foundation, 2008. *U-BIOPRED What is the Project?.* [Online]   
Available at: https://www.europeanlung.org/en/projects-and-research/projects/u-biopred/what-is-the-project/  
[Accessed 28 November 2020].

Fu, W. et al., 2012. A multiply redundant genetic switch 'locks in' the transcriptional signature of regulatory T cells. *Nature Immunology,* Volume 13, pp. 972-980.

Govindarajan, R., Duraiyan, J., Kaliyappan, K. & Palanisamy, M., 2012. Microarray and its applications. *Journal of Pharmacy and BioAllied Sciences,* 4(6), pp. 310-312.

Hoffman, J. I., 2019. *Basic Biostatistics for Medical and Biomedical Practitioners.* 2 ed. s.l.:Academic Press.

Marques, C. R. et al., 2015. Genetic and epigenetic studies of FOXP3 in asthma and allergy. *Asthma research and practice,* 1(10).

Vartanian, K. et al., 2009. Gene expression profiling of whole blood: Comparison of target preparation methods for accurate reproducible microarray analysis. *BMC Genomics,* Volume 10, p. 2.

WHO, 2020. *Asthma.* [Online]   
Available at: https://www.who.int/news-room/fact-sheets/detail/asthma  
[Accessed 28 November 2020].

Zhao, S. T. & Wang, C. Z., 2018. Regulatory T cells and asthma. *Journal of Zhejiang University. Science. B,* 19(9), pp. 663-673.