Machine Learning Unravels Sex-Specific Biomarkers for Atopic Dermatitis

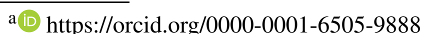
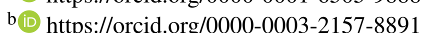
Ana Duarte[a] and Orlando Belo[b]
Algoritmi R&D Centre / LASI, University of Minho, Campus of Gualtar, 4710-057 Braga, Portugal

Keywords: Atopic Dermatitis, Machine Learning, Gene Signature, Sex-Specific Biomarker, Precision Medicine.

Abstract: The prevalence of atopic dermatitis is significantly higher in women than in men. Understanding the differences in the manifestation of the disease between males and females can contribute to more tailored and effective treatments. Our goal in this paper was to discover sex-specific biomarkers that can be used to differentiate between lesional and non-lesional skin in atopic dermatitis patients. Using transcriptomic datasets, we first identified the genes with the highest expression difference. Subsequently, several feature selection methods and machine learning models were employed to select the most relevant genes and identify potential candidates for sex-specific biomarkers. Based on backward feature elimination, we obtained a male-specific signature with 11 genes and a female-specific signature with 10 genes. Both candidate signatures were properly evaluated by an ensemble classifier using an independent test. The obtained AUC and accuracy values for the male signature were 0.839 and 0.7222, respectively, and 0.65 and 0.6667 for the female signature. Finally, we tested the male signature on female data and the female signature on male data. As expected, the analysed metrics decreased considerably in these scenarios. These results suggest that we have identified two promising sex-specific gene signatures, and support that sex affects the ability to distinguish lesions in patients with eczema.

1 INTRODUCTION

Atopic dermatitis (AD) is a chronic and highly complex inflammatory skin condition that significantly reduces the quality of life of patients. In Europe and the USA, one in five children and 10% of adults are diagnosed with AD (Bylund et al., 2020; Laughter et al., 2021). Probably due to socioeconomic and environmental changes, including the growing levels of urbanisation and industrialisation, the prevalence of AD is increasing worldwide, having particularly alarming rates in low-income countries (Nutten, 2015; Schuler et al., 2023; Skevaki et al., 2021; Tsai et al., 2019). The incidence of the disease varies significantly between different geographical regions and cultures, age, sex, and ethnicity (Mesjasz et al., 2023; Nutten, 2015; Schuler et al., 2023). The reasons for this heterogeneity are not clearly understood. They require further investigation and a deeper understanding of how the combination of the multiple factors involved affects the susceptibility, development, progression, and treatment of the disease.


[a] https://orcid.org/0000-0001-6505-9888

[b] https://orcid.org/0000-0003-2157-8891

Numerous diseases, including AD, present a broad spectrum of clinical manifestations, unpredictable courses, and variable responses to therapy. In this regard, the effective management of these complex diseases associated with multiple phenotypes and endotypes requires a precision medicine-based strategy. Concretely, AD is among the disorders that can benefit most from more personalised and targeted interventions (Muraro et al., 2016). However, in contrast to other diseases, precision medicine in AD is still in its early stages. Despite the progress made in recent decades, the clinical reality is still not based on a multifactorial approach tailored to the needs of each patient (Mesjasz et al., 2023; Muraro et al., 2016). Currently, the main challenges for the development of precision medicine in AD are the discovery of new effective therapies with few side effects, taking into account the profile of each patient. Prescribing the most appropriate therapies for each profile presupposes the identification of the pathophysiological mechanisms associated with the disease (Arkwright and Koplin, 2023; Leung, 2024). Particularly, the identification of biomarkers capable of distinguishing lesional from non-lesional skin in AD patients may facilitate the

Duarte, A. and Belo, O.
Machine Learning Unravels Sex-Specific Biomarkers for Atopic Dermatitis.
DOI: 10.5220/0012890700003838
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 1: KDIR*, pages 27-35
ISBN: 978-989-758-716-0; ISSN: 2184-3228
Proceedings Copyright © 2024 by SCITEPRESS – Science and Technology Publications, Lda.

understanding of the mechanisms involved in the development of lesions. A thorough comprehension of these mechanisms is essential to intervene with targeted therapies in the critical pathways. Gene expression profiling based on transcriptomic data can be used to compare lesional and non-lesional skin tissues and identify the key biomarkers involved. Given the multiplicity of factors that influence the disease, the inclusion of supplementary patient characteristics such as sex, age, or existing comorbidities can lead to more accurate biomarker detection. Hence, a holistic approach with the creation of more complex profiles can be tremendously valuable for the management and control of the disease, as well as for enhancing its treatment (Muraro et al., 2016).

With this paper, we expect to contribute to the advancement of precision medicine in AD. Specifically, we aim to discover candidate biomarkers for AD by applying machine learning (ML) algorithms to gene expression data of two distinct patient profiles. ML and classical statistical methods have demonstrated great potential in biomedicine, namely in the processing of high-dimensional datasets such as those used in the identification of gene signatures (Karthik and Sudha, 2018; Liu et al., 2022). We also hypothesise that sex plays a role in identifying a reliable gene signature to differentiate lesional from non-lesional skin in AD patients. For this reason, we speculate that males and females have different molecular mechanisms involved in the manifestation of AD and, consequently, a separate analysis is required.

The remaining part of this paper is organised as follows. Section 2 reports some of the literature that addresses the application of ML techniques to the discovery of gene signatures, and the influence of sex on the manifestation of AD and other diseases. Section 3 summarises the methodology followed to identify the gene signatures and section 4 presents and discusses the results obtained. Finally, Section 5 concludes with an overview of the main findings and possible future research directions.

2 GENE SIGNATURES FOR AD: OPPORTUNITIES AND CHALLENGES

Despite the urgent need to find candidate gene signatures that could revolutionise current dermatological care, very little research has been conducted at the AD level using ML algorithms. In fact, most research applying ML to advance precision medicine focuses on the most fatal diseases, such as cancer. Nevertheless, a few ML-based studies have explored the identification of biomarker candidate genes for AD. One such example is the work developed by Zhong et al. (Zhong et al., 2021). Based on a bioinformatics approach and using LASSO, the authors used transcriptomic datasets and identified GZMB, CXCL1 and CD274 as potential biomarkers to distinguish AD lesions from non-lesions. On the other side, Möbus and colleagues (Möbus et al., 2022), also based on transcriptomic datasets, observed two distinct endotypes for AD associated with notable clinical differences, allowing patients to be stratified into eosinophil-high and eosinophil-low groups. Implementing the Boruta algorithm and a random forest model, the authors identified the most relevant genes to predict the clusters to which the patients belong. Both investigations demonstrate that some key genes are promising candidate biomarkers that have a major impact on the diagnosis and management of AD. However, these studies do not explore the different phenotypes of the patients, such as age, sex, or ethnicity, which are known to play a preponderant role in the manifestation of the disease.

In contrast, other researchers have analysed the discovery of differentiated biomarkers depending on the phenotypic characteristics of the patients, namely sex. For example, Moon et al. (Moon et al., 2013) proposed a procedure to find sex-specific biomarkers based on three datasets from patients with acute myeloid leukaemia, chronic lymphocytic leukaemia, and cutaneous melanoma. The considered methodology consisted of an algorithm based on the importance of each feature in order to extract the top-ranked genes for male and female patients. The selected genes were properly tested and the results obtained revealed high accuracy values, confirming the validity of the strategy followed and underlining the relevance of sex-specific biomarkers for enhancing prognosis prediction. Further experiments have also been conducted to unveil sex-specific biomarkers for different diseases. For instance, some papers exploit the use of ML methods to find sex-specific biomarkers for Alzheimer's disease, emphasising the importance of considering clinical features in addition to genes for a more thorough and sensitive analysis (Bourquard et al., 2023; Ji et al., 2022).

As far as we know, no previous research has addressed the identification of sex-specific genes for AD. However, many studies have suggested that sex has a significant contribution to the prevalence and severity of the disease. For example, Johansson et al. examined the distribution and characteristics of AD in a Swedish population and concluded that the disease is more common in females among young adults (Jo-

hansson et al., 2022). Similarly, Kiiski et al. noted that in a Finnish cohort, the prevalence of AD was higher among women aged 30 to 49 years than among men of the same age (Kiiski et al., 2022). These results are consistent with other investigations conducted in other countries, such as Italy (Pesce et al., 2015) or the USA (Silverberg and Hanifin, 2013), where the female sex also appears to be a risk factor for AD. Therefore, we argue that the process of searching for suitable candidate gene signatures for AD requires a sex-separated analysis. Accordingly, this paper presents a novel contribution to the identification of sex-specific biomarkers for AD based on a ML approach.

3 MATERIALS AND METHODS

From a broader perspective, the methodology considered in our proposal can be divided into two major stages. Each of these stages must be performed independently for the male and female data. First, two datasets sharing the same platform were used to identify the differentially expressed genes (DEGs). Subsequently, based on the selected DEGs, an extra dataset was also taken into account to determine gene signatures for AD. A more detailed description of the sequential processes considered at each stage is provided in Figure 1.

3.1 Data Gathering and Exploration

The datasets used to conduct this project were obtained from the Gene Expression Omnibus[1] (GEO), a public repository containing experimental gene expression data. A preliminary search was performed in order to obtain transcriptomic datasets for AD with information about the sex of the patients. With this objective in mind, we selected the datasets GSE130588, GSE58558, and GSE150797. Since some academics state that datasets from different manufacturers should not be combined to avoid potential bias in the data, all these data sources were generated by the Affymetrix manufacturer (Liu et al., 2021; Serio, 2023). These datasets contain normalised microarray data from skin samples collected from AD patients. As each of these datasets was designed with the purpose of assessing the patients' response to treatments, only the samples that referred to the start of the therapies were considered. Furthermore, we took into account only patients with lesional (AD-L) or non-lesional (AD-NL) AD – Table 1 indicates the main characteristics of the selected data.

Combining the three datasets, the number of samples by sex is balanced. In total, 87 samples correspond to male patients and 88 samples to female patients. Particularly, 51 samples from males refer to AD-L, while 36 refer to AD-NL. Conversely, 49 samples from females correspond to AD-L and 39 to AD-NL.

3.2 Differential Expressed Genes

Since the identified DEGs may have a significant impact on the overall results, we need to plan the differential gene expression analysis in order to avoid the exclusion of important genes. Even if they are from the same manufacturer, datasets from distinct platforms are more likely to introduce bias into the data (Campain and Yang, 2010). For this reason, we decided to perform the differential gene expression analysis using exclusively the datasets produced on the same platform, i.e., datasets GSE130588 and GSE58558.

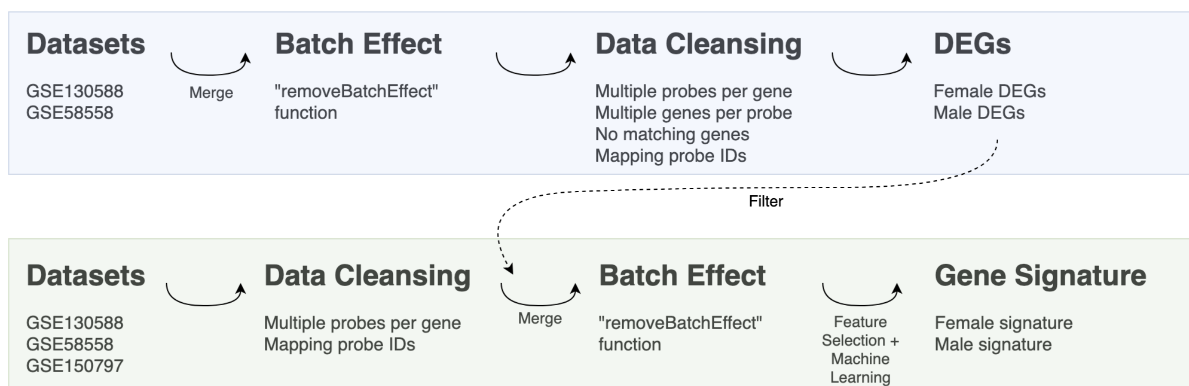


Figure 1: Sequential processes for obtaining candidate gene signatures for AD.

[1] https://www.ncbi.nlm.nih.gov/geo/

Table 1: Summary of the properties of the datasets used.

Dataset	Manufacturer	Platform	Requirements	Sex	Samples AD-L	Samples AD-NL
GSE130588	Affymetrix	GPL570	Time: week 0	Female	22	21
			Tissue: LS or NL	Male	29	21
GSE58558	Affymetrix	GPL570	Time: day 1	Female	6	7
				Male	12	10
GSE150797	Affymetrix	GPL23159	Treatment:	Female	21	11
			untreated	Male	10	5

The first step in determining the DEGs was to divide each of the datasets into two groups, each corresponding to a profile (male or female). Each profile was analysed separately in order to find specific DEGs for the group of male patients and specific DEGs for the group of female patients. All the necessary processes were performed in R (version 4.3.1) using the limma package. One of the first data treatments was to merge the datasets associated with the same profile, i.e., the male samples in the GSE130588 dataset were combined with the male samples in the GSE58558 dataset, and the same was done for the female samples. As a result, we obtained a specific dataset for males with 41 AD-L and 31 AD-NL samples and a dataset for females with 28 AD-L and 28 AD-NL records. Since the merged data came from different experiments, we corrected the batch effect using the "removeBatchEffect" function from the limma package. Moreover, because we were working with microarray data, some additional cleansing tasks were also required for the following encountered situations:

- **Multiple probes corresponding to the same gene –** only the probe with the highest average expression was kept, and the other probes were discarded. Ties in average counts were resolved by choosing one of the probes and eliminating the rest (Miller et al., 2011).
- **Probes associated with various genes –** these records were removed (Hu et al., 2023).
- **Probes that did not match a specific gene –** these records were removed (Wang and Yu, 2023).

To finalise the data treatment, the probe IDs were converted into their corresponding gene symbols. Finally, genes with an absolute $\log_2(\text{fold change}) \geq 1$ and $p_{\text{adj}} < 0.05$ were identified as DEGs and saved in text files.

3.3 Gene Selection Strategy

After identifying the DEGs for each profile, we prepared the datasets GSE130588, GSE58558, and GSE150797 for feature selection and ML modelling. As with the differential expression analysis, this process was also performed using R, analysing each patient profile separately. Each of the three datasets was therefore split into male and female groups, resulting in a total of six subsets. For each subset, we checked the existence of multiple probes corresponding to the same gene and addressed the issue with a similar approach to the one we used to identify the DEGs. Probe IDs were also transformed into the corresponding gene symbols. By importing the text file containing the determined DEGs, we filtered the data in order to have only DEGs. For each profile, the corresponding subsets were merged, and the batch effect was removed. The resulting male and female datasets were then ready to support the next steps of the work.

The following tasks, including additional data processing, feature selection, and construction of ML models, were performed in Python 3.6 using the scikit-learn library. All these tasks were applied in parallel to the male and female datasets. For each dataset, we divided the data into train (80%) and test (20%), using a stratified strategy to maintain the proportion of AD-L and AD-NL samples in both sets. Only the training data were used for feature selection and for the construction of ML classifiers to identify potential gene signatures. For ML modelling, we considered a shuffled and stratified 5-fold cross-validation. The optimal hyperparameters for each algorithm were found using BayesSearchCV with 30 iterations.

Although the identification of DEGs helps to reduce the dimensionality of the data, the resulting high number of genes is still a drawback for efficient processing by ML methods. Feature selection is a common strategy used to minimise these gaps. Therefore, before building ML algorithms, we conducted a feature selection approach using Boruta, Support Vector Machine Recursive Feature Elimination (SVM-RFE), and Least Absolute Shrinkage and Selection Operator (LASSO). These three methods reduced the number of genes differently and thus originated three distinct gene sets. The Random Forest (RF), XGBoost, Ad-

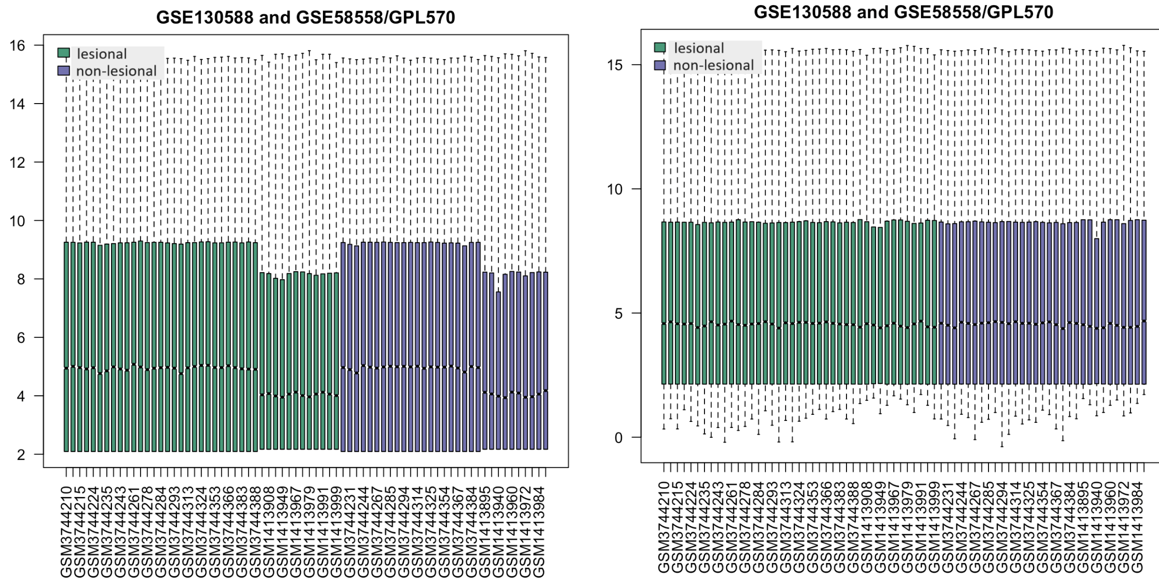


Figure 2: Before (left) and after (right) batch effect correction in the male profile.

aBoost, linear Support Vector Machine (SVM), and Logistic Regression (LR) methods were applied to each of these sets. For each ML method, we extracted the importance of each gene and used the Min-Max method to normalise the obtained value between 0 and 1. Thus, for each of the three gene sets, we obtained the normalised importance values of each gene for each algorithm, which we then summed to generate a gene score. The top genes with significantly higher scores were selected as candidate genes for the creation of a gene signature.

From the selected set of candidate genes, we conceived new ML models using the same five methods and constructed a soft-voting classifier that combines the predictions of the models. This ensemble method was used to evaluate the predictive power of the models, considering both AUC and accuracy values. Adopting a backward feature elimination strategy, we discarded the least important gene and created new classifiers to compare the AUC and accuracy values with the prior solution. These steps were performed iteratively until the performance metrics were worse than the values obtained in the previous step. Thus, the genes that were not eliminated by this iterative process became part of our proposed gene signature. In the end, we obtained one candidate gene signature for males and another for females.

To test our initial hypothesis that sex must be taken into account when establishing gene signatures for AD, we used the independent test to compare the AUC and accuracy values obtained when the male signature is applied to the male and female datasets and vice versa. The additional step of testing the male signature against the female dataset and vice versa was necessary to determine whether the identified genes can be considered sex-specific candidates.

4 RESULTS

4.1 DEGs Identification and Initial Gene Selection

At the beginning of the differential gene expression analysis, two distinct groups were identified after merging the datasets GSE130588 and GSE58558. In both profiles, the samples belonging to the same datasets had similar expression values, but these values were significantly different from the expression values of the samples in the other dataset. Batch effect correction removed these technical differences. Figure 2 shows the box plots obtained before and after removing the batch effect in the male profile. For females, the batch effect correction led to similar results. After processing the data, we identified 188 and 764 DEGs for the male and female datasets, respectively.

Once the DEGs were determined, we proceeded to the individual treatment of the three datasets, which involved filtering the genes according to the DEGs found. As the GSE150797 dataset was generated using a different platform, some DEGs did not match the probe IDs. Consequently, after merging the datasets, the number of DEGs was reduced to 172 in the male profile and 700 in the female scenario. Moreover, we also observed that there were larger differences be-

tween the GSE150797 dataset (GPL23159 platform) and the datasets from the GPL570 platform before batch effect correction. After removing the batch effect, the existing differences were successfully minimized. Boruta, SVM-RFE and LASSO yielded the three gene subsets indicated in Table 2.

Table 2: Number of selected genes by each feature selection approach.

Profile	Boruta	SVM-RFE	LASSO
Male	12	90	62
Female	23	75	92

4.2 Determination of Sex-Specific Gene Signatures

Since the number of DEGs in the male dataset is relatively small, in this case, we decided to include an additional scenario corresponding to the training and validation of the ML algorithms without using any feature selection method. Table 3 lists the top genes obtained after implementing the ML process and determining the scores. In general, the different feature selection strategies identified the top genes consistently for the same profiles.

Based on the results, we selected an initial set of 11 candidate genes for males, and for females, we considered the 16 most important genes (Table 3, highlighted in bold). Our backward feature elimination strategy led to the discovery of an optimal 11-gene signature specific to males (KIF2C, AKR1B10, PHYHIP, FOSL1, FPR1, HS3ST3A1, MX1, KANK4, PPARG, BCL2A1, and KLHDC7B) and a 10-gene signature specific to females (CEP126, FCHSD1, C17orf96, IL18RAP, P2RY10, PTAFR, ANKFN1, TBX18, P2RY2, and AEN). Interestingly, none of the genes are common to both signatures. This may indicate that the molecular pathways involved in the development of AD lesions may differ between men and women, suggesting that the sex of the patients should be considered for better disease management and treatment. The genes KANK4, PHYHIP and PPARG from the male profile were downregulated in the lesions, while the rest were upregulated. In the female profile, only ANKFN1, CEP126 and TBX18 were downregulated, while all others were upregulated. There is scientific evidence that some of these genes may have a major impact on AD. For example, the downregulation of the PPARG gene, identified in the male signature, may be associated with inflammation, keratinisation, and sebaceous gland function (Konger et al., 2021). On the other hand, some studies suggest that P2Y receptors, such as the P2RY10 and P2RY2 genes found in the female signature, can be involved in skin inflammation (Pastore et al., 2007).

Table 4 presents the AUC and accuracy values obtained for each candidate signature using the voting classifier. A closer analysis of the results shows that the AUC and accuracy values of the male signature in the independent test are considerably high when applied to the male data (0.839 and 0.7222, respectively). However, when this signature is tested with the female data, the AUC (0.575) and accuracy (0.6111) values deteriorate substantially. Although the difference is not as marked, the same is true for the female signature. The AUC and accuracy values of the female signature when applied to the female data are 0.650 and 0.6667, respectively. These values decrease when the female signature is tested on the male dataset (AUC = 0.552 and accuracy = 0.6111). These findings thus demonstrate that considering sex-specific biomarkers leads to improved gene signatures for distinguishing lesions from non-lesions, reinforcing the benefits of a sex-separate analysis to establish candidate gene signatures for AD.

Table 3: Top genes identified by the feature selection methods for each profile.

Male profile

Boruta	SVM-RFE	LASSO	All DEGs
KIF2C	KIF2C	FOSL1	KIF2C
AKR1B10	FOSL1	KIF2C	FOSL1
PHYHIP	FPR1	MX1	PHYHIP
FOSL1	**MX1**	PHYHIP	AKR1B10
FPR1	**KANK4**	HS3ST3A1	**KLHDC7B**
HS3ST3A1	**PPARG**	KANK4	FPR1
		BCL2A1	HS3ST3A1

Female profile

Boruta	SVM-RFE	LASSO
CEP126	FCHSD1	C17orf96
FCHSD1	C17orf96	IL18RAP
C17orf96	IL18RAP	MS4A14
GNA15	**PTAFR**	STRIP2
IL18RAP	**WIF1**	**TBX18**
P2RY10	**STRIP2**	**P2RY2**
	PLAG1	**AEN**
	MS4A14	GNA15
	ANKFN1	**HSD11B1**

Table 4: AUC and accuracy values of the optimal gene signatures when tested against the male and female datasets.

		Male data		Female data	
		AUC	accuracy	AUC	accuracy
Male signature	train	0.9737	0.8099	0.8543	0.8429
	test	0.839	0.7222	0.575	0.6111
Female signature	train	0.76	0.6659	0.975	0.9286
	test	0.552	0.6111	0.650	0.6667

5 CONCLUSION AND FUTURE WORK

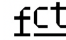
AD presents an unequal distribution between men and women, and its incidence is increasing worldwide. Nevertheless, there are very few studies on this disease to identify biomarkers through ML techniques. Specifically, we have not found any scientific study aimed at finding sex-specific biomarkers for the disease. This could be of particular relevance to gain deeper insights into how AD manifests in men and women. To fill this gap in the literature, we developed a ML approach using transcriptomic datasets and intended to identify male and female biomarkers that distinguish normal from lesional skin in patients with atopic eczema.

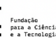
Our research led to the definition of a male-specific gene signature consisting of the KIF2C, AKR1B10, PHYHIP, FOSL1, FPR1, HS3ST3A1, MX1, KANK4, PPARG, BCL2A1, and KLHDC7B genes, and a female-specific gene signature comprising the CEP126, FCHSD1, C17orf96, IL18RAP, P2RY10, PTAFR, ANKFN1, TBX18, P2RY2, and AEN genes. For some of the identified genes, there is evidence in the literature to support their possible influence on the skin. The difference between the genes of the two signatures could indicate that different mechanisms are involved in the manifestation of AD in men and women. A better understanding of these mechanisms could promote the emergence of targeted treatments and contribute to the development of precision medicine in AD.

Although the results obtained emphasise the need to investigate sex-specific biomarkers for AD, our study has certain limitations. The main shortcomings are the limited number of samples and the lack of public databases providing gene expression data in combination with clinical phenotypes. Therefore, new studies on this topic and the availability of new datasets integrating transcriptomic and phenotypic data are currently a priority. It would also be valuable for future research to replicate the proposed methodology to other diseases, particularly those where it is suspected that different molecular mechanisms may be involved depending on the sex. In addition, a thorough investigation of the discovered biomarkers as well as the associated molecular mechanisms is required to gain a comprehensive understanding of how men and women differ in the development of AD lesions. Finally, any research of this nature requires clinical validation.

ACKNOWLEDGEMENTS

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020, and the PhD grant: 2022.12728.BD.





REFERENCES

Arkwright, P. D. and Koplin, J. J. (2023). Challenging best practice of atopic dermatitis. *Journal of Allergy and Clinical Immunology: In Practice*, 11:1391–1393.

Bourquard, T., Lee, K., Al-Ramahi, I., Pham, M., Shapiro, D., Lagisetty, Y., Soleimani, S., Mota, S., Wilhelm, K., Samieinasab, M., Kim, Y. W., Huh, E., Asmussen, J., Katsonis, P., Botas, J., and Lichtarge, O. (2023). Functional variants identify sex-specific genes and pathways in alzheimer's disease. *Nature Communications*, 14.

Bylund, S., Kobyletzki, L. B. V., Svalstedt, M., and Åke Svensson (2020). Prevalence and incidence of atopic dermatitis: A systematic review. *Acta Dermato-Venereologica*, 100:320–329.

Campain, A. and Yang, Y. H. (2010). Comparison study of microarray meta-analysis methods. *BMC Bioinformatics*, 11.

Hu, Y., Chen, X., Mei, X., Luo, Z., Wu, H., Zhang, H., Zeng, Q., Ren, H., and Xu, D. (2023). Identification of diagnostic immune-related gene biomarkers for predicting heart failure after acute myocardial infarction. *Open Medicine*, 18.

Ji, W., An, K., Wang, C., and Wang, S. (2022). Bioinformatics analysis of diagnostic biomarkers for alzheimer's disease in peripheral blood based on sex differences and support vector machine algorithm. *Hereditas*, 159.

Johansson, E. K., Bergström, A., Kull, I., Melén, E., Jonsson, M., Lundin, S., Wahlgren, C. F., and Ballardini, N. (2022). Prevalence and characteristics of atopic dermatitis among young adult females and males - report from the swedish population-based study bamse. *Journal of the European Academy of Dermatology and Venereology*, 36:698–704.

Karthik, S. and Sudha, M. (2018). A survey on machine learning approaches in gene expression classification in modelling computational diagnostic system for complex diseases. *International Journal of Engineering and Advanced Technology (IJEAT)*, 8:182–191.

Kiiski, V., Salava, A., Susitaival, P., Barnhill, S., Remitz, A., and Heliovaara, M. (2022). Atopic dermatitis in adults: a population-based study in finland. *International Journal of Dermatology*, 61:324–330.

Konger, R. L., Derr-Yellin, E., Zimmers, T. A., Katona, T., Xuei, X., Liu, Y., Zhou, H.-M., Simpson, E. R., and Turner, M. J. (2021). Epidermal pparγ is a key homeostatic regulator of cutaneous inflammation and barrier function in mouse skin. *International Journal of Molecular Sciences*, 22.

Laughter, M. R., Maymone, M. B., Mashayekhi, S., Arents, B. W., Karimkhani, C., Langan, S. M., Dellavalle, R. P., and Flohr, C. (2021). The global burden of atopic dermatitis: lessons from the global burden of disease study 1990–2017. *British Journal of Dermatology*, 184:304–309.

Leung, D. Y. (2024). Evolving atopic dermatitis toward precision medicine. *Annals of allergy, asthma & immunology*, 132:107–108.

Liu, J., Liu, L., Antwi, P. A., Luo, Y., and Liang, F. (2022). Identification and validation of the diagnostic characteristic genes of ovarian cancer by bioinformatics and machine learning. *Frontiers in Genetics*, 13.

Liu, L., Wang, T., Huang, D., and Song, D. (2021). Comprehensive analysis of differentially expressed genes in clinically diagnosed irreversible pulpitis by multiplatform data integration using a robust rank aggregation approach. *Journal of Endodontics*, 47:1365–1375.

Mesjasz, A., Kołkowski, K., Wollenberg, A., and Trzeciak, M. (2023). How to understand personalized medicine in atopic dermatitis nowadays? *International Journal of Molecular Sciences*, 24.

Miller, J. A., Cai, C., Langfelder, P., Geschwind, D. H., Kurian, S. M., Salomon, D. R., and Horvath, S. (2011). Strategies for aggregating gene expression data: The collapserows r function. *BMC Bioinformatics*, 12.

Moon, H., Lopez, K. L., Lin, G. I., and Chen, J. J. (2013). Sex-specific genomic biomarkers for individualized treatment of life-threatening diseases. *Disease Markers*, 35:661–667.

Muraro, A., Lemanske, R. F., Hellings, P. W., Akdis, C. A., Bieber, T., Casale, T. B., Jutel, M., Ong, P. Y., Poulsen, L. K., Schmid-Grendelmeier, P., Simon, H. U., Seys, S. F., and Agache, I. (2016). Precision medicine in patients with allergic diseases: Airway diseases and atopic dermatitis - practall document of the european academy of allergy and clinical immunology and the american academy of allergy, asthma & immunology. *Journal of Allergy and Clinical Immunology*, 137:1347–1358.

Möbus, L., Rodriguez, E., Harder, I., Boraczynski, N., Szymczak, S., Hübenthal, M., Stölzl, D., Gerdes, S., Kleinheinz, A., Abraham, S., Heratizadeh, A., Handrick, C., Haufe, E., Werfel, T., Schmitt, J., and Weidinger, S. (2022). Blood transcriptome profiling identifies 2 candidate endotypes of atopic dermatitis. *Journal of Allergy and Clinical Immunology*, 150:385–395.

Nutten, S. (2015). Atopic dermatitis: Global epidemiology and risk factors. *Annals of Nutrition and Metabolism*, 66:8–16.

Pastore, S., Mascia, F., Gulinelli, S., Forchap, S., Dattilo, C., Adinolfi, E., Girolomoni, G., Virgilio, F. D., and Ferrari, D. (2007). Stimulation of purinergic receptors modulates chemokine expression in human keratinocytes. *Journal of Investigative Dermatology*, 127:660–667.

Pesce, G., Marcon, A., Carosso, A., Antonicelli, L., Cazzoletti, L., Ferrari, M., Fois, A. G., Marchetti, P., Olivieri, M., Pirina, P., Pocetta, G., Tassinari, R., Verlato, G., Villani, S., and Marco, R. D. (2015). Adult eczema in italy: prevalence and associations with environmental factors. *Journal of the European Academy of Dermatology and Venereology*, 29:1180–1187.

Schuler, C. F., Billi, A. C., Maverakis, E., Tsoi, L. C., and Gudjonsson, J. E. (2023). Novel insights into atopic dermatitis. *Journal of Allergy and Clinical Immunology*, 151:1145–1154.

Serio, P. (2023). Gene expression microarray merging. https://rpubs.com/Karksus/1013177.

Silverberg, J. I. and Hanifin, J. M. (2013). Adult eczema prevalence and associations with asthma and other health and demographic factors: A us population-based study. *Journal of Allergy and Clinical Immunology*, 132:1132–1138.

Skevaki, C., Ngocho, J. S., Amour, C., Schmid-Grendelmeier, P., Mmbaga, B. T., and Renz, H. (2021). Epidemiology and management of asthma and atopic dermatitis in sub-saharan africa. *Journal of Allergy and Clinical Immunology*, 148:1378–1386.

Tsai, T.-F., Rajagopalan, M., Chu, C.-Y., Encarnacion, L., Gerber, R. A., Santos-Estrella, P., Llamado, L. J. Q., and Tallman, A. M. (2019). Burden of atopic dermatitis in asia. *Journal of Dermatology*, 46:825–834.

Wang, X. and Yu, G. (2023). Drug discovery in canine pyometra disease identified by text mining and microar-

ray data analysis. *BioMed Research International*, 2023.

Zhong, Y., Qin, K., Li, L., Liu, H., Xie, Z., and Zeng, K. (2021). Identification of immunological biomarkers of atopic dermatitis by integrated analysis to determine molecular targets for diagnosis and therapy. *International Journal of General Medicine*, 14:8193–8209.