

DR. VASSILI SOUMELIS (Orcid ID : 0000-0003-1849-9834)

MR. PÉTER OLÁH (Orcid ID : 0000-0002-3184-8376)

Article type : Original article

TITLE

Transcriptome-based identification of novel endotypes in adult atopic dermatitis

SHORT TITLE

Novel atopic dermatitis transcriptomic endotypes

AUTHORS

Alain Lefèvre-Utile^{1,2,3}, Melissa Saichi¹, Péter Oláh^{4,5}, Marc Delord⁶, Bernhard Homey⁴, Vassili Soumelis^{1,7}; MAARS consortium

MAARS Consortium

Juha Kere^{a,b}, Francesca Levi-Schaffer^c, Dario Greco^{d,e}, Noora Ottman^f, Jonathan Baker^g, Björn Andersson^h, Mauricio Barrientos-Somarrivas^h, Stefanie Prast-Nielsenⁱ, Lukas Wisgrill^j, Sophia Tsoka^k, Nanna Fyhrquist^{fl}, Harri Alenius^{fl}, Helen Alexander^g, Jens M. Schröder^m, Frank O. Nestle^g, Antti Lauermaⁿ, Philippe Hupé^o, Annamari Rankiⁿ

Manuscript Acceptance Date : 18-Oct-2021

AFFILIATIONS

¹ Université de Paris, Inserm, U976 HIPI Unit, Institut de Recherche Saint-Louis, F-75006, Paris, France

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/ALL.15150](https://doi.org/10.1111/ALL.15150)

This article is protected by copyright. All rights reserved

² Assistance Publique-Hôpitaux de Paris (APHP), General Pediatrics and Pediatric Emergency Department, Jean Verdier Hospital, Bondy, France

³ Université Paris Saclay

⁴ Department of Dermatology, University of Duesseldorf, Duesseldorf, Germany

⁵ Department of Dermatology, Venereology, and Oncodermatology, Medical Faculty, University of Pécs, Hungary.

⁶ Clinical Research Center, Centre Hospitalier de Versailles, Le Chesnay, France

⁷ Assistance Publique-Hôpitaux de Paris (AP-HP), Laboratoire d'Immunologie et histocompatibilité, Hôpital Saint-Louis, F-75010, Paris, France

MAARS Consortium members' affiliations

^a Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden

^b Stem Cells and Metabolism Research Program, Folkhälsan Research Institute, University of Helsinki, Helsinki, Finland

^c Pharmacology and Experimental Therapeutics Unit, Institute for Drug Research, School of Pharmacy, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel

^d Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

^e Institute of Biotechnology, University of Helsinki, Helsinki, Finland

^f Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

^g St John's Institute of Dermatology, Faculty of Medicine and Life Sciences, Kings College London, London, United Kingdom

^h Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden

ⁱ Department of Microbiology, Tumour and Cell Biology, Karolinska Institutet, Stockholm, Sweden

^j Division of Neonatology, Pediatric Intensive Care and Neuropediatrics, Comprehensive Center for Pediatrics, Department of Pediatrics and Adolescent Medicine; Medical University of Vienna, Austria

^k Department of Informatics, Faculty of Natural and Mathematical Sciences, Kings College London, London, United Kingdom

^l Department of Bacteriology and Immunology, Medicum, University of Helsinki, Helsinki, Finland

^m Department of Dermatology, University Hospital Schleswig-Holstein, Kiel, Germany

ⁿ Department of Dermatology, Allergology, and Venerology, University of Helsinki, Helsinki University Hospital, Inflammation Centre, Helsinki, Finland

^o INSERM U900, CNRS UMR144, Institut Curie, Mine Paris Tech, Paris, France

CORRESPONDANCE TO

Vassili Soumelis, MD, PhD. Inserm U976 HIPI, Institut de recherche St Louis, Hôpital St Louis APHP, 1 Avenue Claude Vellefaux, 75010 Paris. vassili.soumelis@aphp.fr.

CONFLICT OF INTEREST

ALU, MS, PO and MD have nothing to disclose. BH reports honoraria from Abbvie, Galderma, EADV, Janssen-Cilag, Lilly, RG-Gesellschaft, RWDG, Streamdub GmbH, and VK-Verband Kölner Hautärzte for educational event or manuscript writing. BH reports honoraria from Boehringer-Ingelheim, Galderma, Sanofi-Aventis, Union therapeutics and Milteny for participation to advisory board. BH reports to be a member of ESDR. VS reports honoraria from Leo pharma for educational event, and from Roche for attending meetings.

ACKNOWLEDGMENTS

This work was supported by the *Fondation pour la Recherche Médicale*, grant FDM201806006187, to Alain Lefèvre-Utile. The MAARS Consortium has received funding from the FP7/2007–2013 (Grant 261366). Thanks to Nuala Mooney, Andrei Zinovyev, and Elise Amblard for their critical reviews. Thanks to Cristina Ghirelli for the scientific editing.

ABSTRACT

Atopic dermatitis (AD) is a frequent and heterogeneous inflammatory skin disease, for which personalized medicine remains a challenge. High-throughput approaches have improved understanding of the complex pathophysiology of AD. However, a purely data-driven AD classification is still lacking. To address this question, we applied an original unsupervised approach on the largest available transcriptome dataset of AD lesional (n=82) and healthy (n=213) skin biopsies. Taking into account pathological and physiological state, a variance-based filtering revealed 222 AD-specific hyper-variable genes that efficiently classified the AD samples into 4 clusters that turned out to be clinically and biologically distinct. Comparison of gene expression between clusters identified 3 sets of upregulated genes used to derive metagenes (MG): MG-I (19 genes) was associated with IL-1 family signaling (including IL-36A and 36G) & skin remodeling, MG-II (23 genes) with negative immune regulation (including IL-34 and 37) & skin architecture, and MG-III (17 genes) with B lymphocyte immunity. Sample clusters differed in terms of disease severity ($p=0.02$) and *S. aureus* (SA) colonization ($p=0.02$). Cluster 1 contained the most severe AD, highest SA colonization, and overexpressed MG-I. Cluster 2 was characterized by less severe AD, low SA colonization, and high MG-II expression. Cluster 3 included mild AD, mild SA colonization, and mild expression of all MGs. Cluster 4 had the same clinical features as cluster 3 but had hyper-expression of MG-III. Last, we successfully validated our method and results in an independent cohort. Our study revealed unrecognized AD endotypes with specific underlying biological pathways, highlighting novel pathophysiological mechanisms. These data could provide new insights into personalized treatment strategies.

KEYWORDS

Atopic dermatitis

Transcriptome

Endotype

Clustering

ABBREVIATIONS

89ADGES: 89 Atopic Dermatitis Gene Expression Signatures

AD: Atopic dermatitis

AKR1B: Aldo-Keto Reductase Family 1 Member

BTC: Betacellulin

CLEC2A: C-Type Lectin Domain Family 2 Member A

DEFB4A: Defensin Beta 4A

DGAT2L6: Diacylglycerol O-Acyltransferase 2 Like 6

IL: Interleukin

EPGN: Epithelial Mitogen

FLG: Filaggrin

GJB4: Gap Junction Protein Beta 4

GO: Gene Ontology

GSTA3: Glutathione S-Transferase Alpha 3

IGH: Immunoglobulin Heavy Locus

IGK: Immunoglobulin Kappa Locus

IGL: Immunoglobulin Lambda Locus

k-NN: *k*-Nearest Neighbors

KRT: Keratin

KRTAP: Keratin Associated Protein

LOR: Loricrin

MAARS: Microbes in Allergy and Autoimmunity Related to the Skin

MADAD: Meta-Analyses Derived Atopic Dermatitis

MG: Metagene

MST: Minimal Spanning Tree

OGN: Osteoglycin

PCA: Principal component analysis

PI: Peptidase Inhibitor

PILAR: Proliferation-Induced Lymphocyte-Associated Receptor

PLA2G4D: Phospholipase A2 Group IVD

PM20D1: Peptidase M20 Domain Containing 1

PRSS22: Serine Protease 22

S100A: S100 Calcium Binding Protein A

SA: *Staphylococcus aureus*

SCORAD: Score Atopic Dermatitis

SLC46A2: Solute Carrier Family 46 Member 2

SPRR: Small Proline-Rich Protein

UGT3A2: UDP Glycosyltransferase Family 3 Member A2

XIST: X Inactive Specific Transcript

INTRODUCTION

Atopic dermatitis (AD), i.e. atopic eczema, is an inflammatory skin condition characterized by chronic and recurrent episodes of pruriginous, and erythematous lesions of varying severity¹. Depending on the geographical region, AD can affect up to 20% of children and 5% of adults^{2,3}. AD is clinically very heterogeneous in terms of symptoms, severity, comorbidities, and response to therapy.

AD pathophysiology is multifaceted, involving several biological and organizational levels, such as dermo-epidermal architecture, water and lipid metabolism, and skin immunity⁴. These interact with the patient's genetic background as well as multiple environmental factors like the microbiome, radiations, or allergens⁵. AD has been typically described as a Th2 disease⁶, but the underlying cytokine networks appear more complex with inter-relationships of Th1, Th17, Th22, or Tfh polarizations^{7,8}.

The marked clinical and biological heterogeneity of AD has suggested the existence of different disease endotypes, defined as patient subtypes with shared pathobiological mechanisms⁹. Since the description of extrinsic and intrinsic AD^{6,10}, several other endotypes have been identified, mainly based on the age of disease onset, duration of symptoms, and patient ethnicity¹¹. Selected biomarkers, including serum IgE levels, blood eosinophils, or panels of pre-defined cytokines, have been applied to design supervised approaches allowing stratification of patient groups^{12,13} and prediction of treatment response¹⁴ without reaching a consensus regarding their clinical applications.

Large-scale and omics approaches are increasingly used because they offer the possibility of characterizing complex diseases based on a diversity of biological variables and pathways. Previous AD studies have identified diagnostic molecular signatures from skin transcriptome^{15,16}, and endotypes from a knowledge-driven set of blood biomarkers^{13,17}. However, a purely data-driven skin transcriptomic classification of AD is still lacking.

In this study, we sought to establish a systematic, unbiased strategy for patient classification and characterization of AD heterogeneity. We exploited the largest available AD and normal skin transcriptomic datasets, and validate our results on an independent cohort. Our work provides new insights into AD heterogeneity and suggests how unprecedented potential strategies for personalized treatments could be designed.

MATERIAL AND METHOD

Study cohort

The data were obtained from the MAARS Consortium^{18,19}, whose dataset is publicly available on the Array Express interface (E-MTAB-8149). Patient recruitment and data generation methodologies are comprehensively described in Fyhrquist et al publication¹⁸. Briefly, AD patients and healthy controls have been recruited in three European Dermatology departments. All AD patients met the Hanifin and Rajka criteria²⁰. Sampling and data generation occurred between 2012 and 2013. A vast amount of clinical features was collected, including disease severity AD scoring (SCORAD)²¹. A 6 mm punch biopsy was performed in lesional skin of AD patients and at two different sites for healthy controls. Skin biopsy localizations were well balanced within the two cohorts with back (44%), or thigh (54%) for AD, and back (48%), or thigh (46%) for controls. Bulk transcriptomic analysis was performed after mRNA extraction with Affymetrix GeneChip® Whole Transcript Expression Arrays. SA colonization was determined by 16s rRNA amplification and sequencing, as a fraction of next-generation total microbiome, from a non-invasive skin swab made on the biopsy site.

Validation cohort

To assess the reproducibility and robustness of both our unsupervised approach and the main findings of our study, we applied the same clustering on an independent dataset. To reduce technological and technical biases, we sourced independent cohorts using a comparable transcriptomic technology on skin biopsies, with available annotation on disease severity. Among the total number of pre-selected transcriptomic cohorts (n=48), only six studies met the above criteria (Suppl 9). We selected the largest one, with n=51 lesional skin sample whose methodology is described in Guttman *et al* study²². Bulk transcriptomic data were generated using Affymetrix Human U133Plus 2.0® gene arrays. The expression matrix was downloaded through the Gene expression omnibus (GEO) interface (GSE130588) using the *GEOquery* package²³ (version 2.51.1).

Expression array preprocessing

The MAARS dataset was loaded and analyzed with *R* language (version 3.6.0) on the *R Studio* interface (version 1.2.1335). As an exploratory step, we projected the dataset using Principal Component Analysis and performed several clustering methods. We discarded the sample *MAARS_3_070_03* as a potential mislabeled outlier. After quality control steps, 213 samples from the 113 healthy controls were pooled with AD samples. Thus, 82 AD lesional and 213 healthy skin samples were used for further analyses

Variance-based feature selection

Expression variance was computed for each gene across AD lesional and healthy control samples. Subsequently, a ratio of the corresponding variance between the two populations was calculated for each gene. We included genes that were highly variable and specific of AD lesional skin: variance >0.5 (corresponding approximatively to the 5th percentile of all variances), and ratio >2 . We further term the selected genes from this step as *AD-specific hyper-variable genes* and use them for subsequent steps.

Clustering method and optimal number of clusters estimation

Unsupervised clustering algorithms, such as *k*-means, hierarchical clustering, *k*-Nearest Neighbors (*k*-NN) coupled with a Minimal Spanning Tree (MST) partitioning, were applied on the expression matrix, including only AD samples ($n=82$). The optimal number of clusters to consider for downstream analysis was determined by the elbow method, average silhouette width, and MST. To assess the clustering efficiency, we calculated the clustering purity with the Dunn index, Rand index, and Jaccard similarity coefficient, and chose the method which scored the best in the different metrics.

Metagene construction

Differential analysis, using Student test and Bonferroni correction, was computed between patient clusters to identify the set of genes characteristic of each sample group. Unique and upregulated genes were further pooled to construct *Metagenes* (MG) using pairwise correlation.

MG expressions correspond to the expression mean of their respective genes. Clinical and biological parameters were compared using the Kruskal-Wallis test for continuous variables and the Fisher test for binary variables. Sample clusters were then characterized using the most significant features and metagene expressions. Inter clusters metagene and gene expression comparison were performed using ANOVA test. Correlation analyses between MG, features, and AD clusters were performed with non-parametric Spearman or parametric Pearson statistics, depending on the nature of the data.

Functional enrichment

Pathway enrichment analysis and functional annotation of the metagenes were performed first using g:Profiler web server (<https://biit.cs.ut.ee/gprofiler/gost>) with Kyoto Encyclopedia of Genes and Genomes database (KEGG)²⁴ and then Cytoscape (version 3.8.0)²⁵ with ClueGO extension (version 2.5.6)²⁶ using gene ontology (GO) biological process. An extensive manual curation of the literature through PubMed search was carried out to ensure a complete functional annotation of identified genes.

Statistics and data visualization

Statistics were conducted with *stats* (version 3.6.0) package and *Rquery* function. Figures were generated with *ggplot2* package (version 3.3.1). Statistical differences in clinical and transcriptomic data were considered significant at an adjusted p-value (adj.p) <0.05. For correlation analysis, $r > 0.3$ and $p < 0.05$ were considered significant.

RESULTS

Variance-based gene selection revealed pathways relevant to AD pathophysiology

To identify clinically and biologically relevant AD endotypes, we exploited a large and multicentric transcriptomic dataset containing 82 AD samples and 213 healthy skin control samples from 113 healthy volunteers (Suppl 1a)¹⁸. The control cohort was composed of 69

females and 44 males with a mean age of 35 years [min-max: 19-77]. We have chosen to keep all healthy samples to maximize statistical power. AD cohort contained different levels of AD severity since it included mild (n=6), moderate (n=39), and severe (n=37) AD, with a mean SCORAD (Score Atopic dermatitis) of 52 [min-max: 17-89]. It included 36 females and 46 males. The mean age was 44 years [min-max: 20-83]. More than 90% of AD patients and controls were Caucasian. All patients were recruited after a therapeutic wash-out of 3 months for systemic immunosuppressive drugs, and 2 weeks for topical steroids.

Transcriptomic profiles were obtained using Affymetrix protocols, which enabled the detection of 22,635 coding genes. We designed a feature selection strategy based on the premise that hyper-variable genes in AD samples would contain the most discriminatory information for the subsequent definition of patient clusters (Fig 1a). We applied a cut-off on variance values (>0.5) to filter out low variable genes that resulted in a selection of 575 “hyper-variable genes”. Next, we considered that hyper-variable genes in both AD and healthy states could represent physiologically hyper-variable genes. Such genes would also vary across healthy skin and their corresponding ratio would oscillate around 1, such as gender-related genes. Hence, we applied an additional filter based on a variance ratio >2 , to select highly variable genes solely in the AD state, that would represent “AD-specific hyper-variable genes”. As expected, XIST and DGAT2L6, which are involved in gender representation, are two of the most variable genes in AD lesional and control skin with respective variances of 11.6 vs 11.3, and 7.2 vs 6.7. Genes not associated with sex chromosomes may also vary among healthy skin samples for physiological reasons. This was the case for KRTAP4-12 and PM20D1 with respective variances of 5.4 vs 8.6, and 6.7 vs 6.6. We reasoned that such genes would not inform on the disease-specific variability that should form the basis for characterizing putative AD clusters. Overall, the application of these two filters on the initial dataset identified 222 AD-specific hyper-variable genes (Fig 1b, Suppl 2).

We exploited several databases such as the GO biological process (Fig 1c) and the KEGG pathways (Suppl 1b) for functional interpretation and biological relevance of these 222 genes. AD-specific hyper-variable genes were enriched in functions such as “keratinocyte differentiation”, “type-I interferon signaling pathways”, “lymphocyte-mediated immunity” (Fig 1c). Most of these pathways were consistent with previously described mechanisms of AD onset and development, confirming that the selected genes were representative of AD pathophysiology. This was a first step that revealed the biological relevance of the AD-specific hyper-variable genes.

AD-specific hyper-variable genes identified AD clusters with distinct clinical features

The 222 AD-specific hyper-variable genes were used to cluster the AD samples (Fig 2a). First, we estimated the optimal number of clusters and performed clustering through different methods like the elbow method coupled with *k*-means (Fig 2b), MST partitioning coupled with *k*-NN (Suppl 3a), and the average silhouette width coupled with hierarchical clustering (Suppl 3b-c). All approaches indicated four clusters. We compared the different clustering methods based on the intra-cluster metric Dunn index, and the inter-cluster metrics Rand index and Jaccard similarity coefficient. We selected *k*-means clustering because of its highest score in all metrics (Suppl 4). *k*-means clustering classified samples in four AD clusters of $n=12$, 23, 29, and 18 samples, respectively. Principal component analysis (PCA) showed the embedding of previously defined clusters (Fig 2b).

Second, we asked whether clinical, biological, or bacteriological features were different between clusters (Suppl 5a). We found that the SCORAD²¹ and the 16s rRNA quantification of skin SA colonization differed in at least one cluster compared to the others (adj.p=0.02, and adj.p=0.02, respectively) (Fig 2c). Other variables, such as personal history of other allergic diseases (adj.p=0.95), gender (adj.p=0.98), pediatric age at disease onset (adj.p=0.44), age (adj.p=0.98) and raised IgE levels (adj.p=0.84) (Suppl 5a,b) did not show significant variation among clusters.

Construction of metagenes and identification of their specific biological functions

Considering the 222 AD-specific hyper-variable genes, and the 4 AD clusters, we sought to identify the genes with the best ability to discriminate between clusters (Fig 3a).

We performed differential expression analyses of one AD cluster versus all the others. It revealed 3 sets of genes since AD cluster 3 did not have any differentially expressed gene (Suppl 6a). In total, 62 differentially expressed genes, and some were shared between several gene sets (Fig 3b). To get the strongest positive signal, we considered only the 59 specific to one cluster and upregulated genes. We then carried out a correlation analysis, and grouped genes that had the same behavior into metagenes (Fig 3c). We thereby defined three metagenes: MG-I, MG-II, and MG-III containing respectively 19, 23, and 17 genes (Suppl 6b).

Next, we wanted to functionally characterize those three metagenes. GO biological process analysis showed enrichment in *keratinization* and *innate immune pathways* for MG-I, *complement activation*, and *humoral response* for MG-III. Probably due to the low number of genes and their apparent heterogeneity, it did not identify a significant association for MG-II. Hence, we turned to manual curation and functional annotation through an extensive PubMed search for each gene in each metagene (Suppl 7). This approach revealed more specific information about each gene and allowed grouping into relevant biological families. MG-I had a strong enrichment in IL-1 family signaling (DEFB4A, IL36A, IL36G, PI3, PLA2G4D, S100A7A, S100A12), and epidermal proliferation and differentiation (EPGN, KRT6A, KRT6C, AKR1B10, SPRR2B). MG-II was enriched in negative immune regulation (CLEC2A (=PILAR), IL34, IL37, SLC46A2), skin architecture (FLG2, LOR, OGN), and epidermal homeostasis (BTC, GJB4, GSTA3, KRT77, UGT3A2). Finally, MG-III was entirely composed of B lymphocyte immunity-related genes (IGLs, IGHs, IGKs). In total, we could characterize the predominant molecular functions of the 59 most important genes in our AD clusters.

Multilayer characterization of AD clusters

We compared the metagene expression distribution across AD clusters (Fig 3d). Hyper-expression of MG-I was associated with AD cluster 1 ($p < 2.0 \times 10^{-16}$), MG-II with AD cluster 2 ($p < 2.0 \times 10^{-16}$), and MG-III with AD cluster 4 ($p = 1.4 \times 10^{-14}$). AD cluster 3 was the only one with a mild expression of all metagenes. Differential expression confirmed that metagenes could act as discriminating variables allowing the classification of AD samples and characterization of AD clusters at the biological level. We checked for differential expression of AD reference pathways such as Th2 (IL4R, IL5RA, IL13RA1, IL13RA2, IL31RA, IL4R, DPP4, CCL13, CCL17, CCL18, CCL22, CCL26, TSLP), Th17/22 (IL17A, IL17B, IL17C, IL17RA, IL17RC, IL17RE, IL17REL, IL19, IL22, IL22RA1, CXCL1, LCN2), and skin barrier (CLDND1, CLDN5, CLDN18, IVL, FLG), to evaluate how knowledge-based gene selection characterize the AD clusters (Suppl 8 a,b,c). While some genes were differentially expressed between clusters, they did not behave homogeneously within clusters at the signature level, thus had a low potential to separate AD clusters.

Then we correlated metagene expression levels with the clinical and biological features that differed between groups (Fig 4a). This revealed a positive correlation of MG-I with the SCORAD

($R=0.48$, $p=4.7e-06$), and with SA colonization ($R=0.45$, $p=3.1e-05$), and an inverse-correlation between MG-II and those 2 parameters ($R=-0.88$, $p=3.5e-06$ and $R=-0.46$, $p=1.8e-05$, respectively). In contrast, MG-III was not correlated with any of these parameters (Fig 4b). In conclusion, the expression level of *IL-1 family signaling & skin remodeling* MG-I was associated with greater clinical severity and SA colonization, whereas the *Negative immune regulation & skin architecture* MG-II was associated with milder forms of the disease.

Validation on an independent cohort

We set out to validate the contrasting association of the *IL-1 family signaling & skin remodeling* MG-I, and the *Negative immune regulation and skin architecture* MG-II with disease severity. To do so, we screened independent datasets from Affymetrix® microarray AD skin biopsy studies that included annotations about disease severity. Six cohorts with 12 to 51 lesional AD skin samples met these criteria in the GEO database (Suppl 9). We selected the largest dataset²² containing 51 mild-to-severe adult Caucasian AD samples. We used our reference AD-specific hyper-variable genes (Suppl 2) with the same analysis pipeline by clustering the patient samples in four groups using the *k*-means method. We then built metagenes with the same genes as in our study (Suppl 6b) to compare the metagene expression and the AD severity among clusters (Fig 4c).

SCORAD significantly differed between AD clusters ($p=0.04$), as well as for the 3 metagene expressions (with respectively $p=5.9e-06$, $p=1.1e-06$, $p=0.004$). As in our cohort, the most severe AD cluster was associated with MG-I hyper-expression, and the less severe AD cluster was associated with the hyper-expression of MG-II. Finally, we calculated the correlation of metagene expression and disease severity (Fig 4d) and confirmed the significant positive association of *IL-1 family signaling & skin remodeling* MG-I with disease severity ($R=0.35$, $p=0.013$) and the significant negative association of *Negative immune regulation and skin architecture* MG-II with disease severity ($R=-0.42$, $p=0.0019$).

This analysis, based on the largest independent dataset available, validates the existence of four AD clusters characterized by different disease severity and underlying mechanisms.

DISCUSSION

The question of *how to evaluate and interpret human disease heterogeneity?* is difficult to assess. Since the beginning of the high-throughput data era, the duality between supervised and unsupervised approaches prevails. Practically, both strategies are complementary. Because they are less difficult to design, supervised approaches are mainly used. The idea is to compare different patient groups with pre-defined clinical or biological characteristics with the scope of identifying specific pathological mechanisms. Thus, AD disease signatures as MADAD (Meta-Analyses Derived Atopic Dermatitis)¹⁵ and 89ADGES (89 Atopic Dermatitis Gene Expression Signature)¹⁶, have been robustly designed comparing AD patients to healthy skin transcriptomes. In AD and other complex diseases, such as systemic lupus erythematosus, supervised approaches on disease severity score were able to highlight particular pathways^{27–29}. Unsupervised approaches are more rarely used and consist in using data structure to generate new hypotheses independently of clinical and biological annotations. They require a higher number of samples compared to supervised methods to maximize their statistical robustness. We decided to design a purely data-driven unsupervised strategy using the largest AD and controls microarray cohort. Our method is novel and therefore able to reveal original findings. It also consolidates previous discoveries because of its complementarity to published studies^{30–32}.

Feature selection is an essential step in complex data analysis. It consists in reducing data dimension and complexity by removing redundant and irrelevant features to reduce noise and improve classification performance. Differential expression and machine learning wrapping are the most commonly used method in the supervised approach³³. As genes with high expression variance could drive phenotypical diversity, we hypothesized that their study is important for the success of unsupervised approaches. The variance could be used directly to select genes^{34,35} or through variance-based methods like PCA³⁶. Considering that important gene expression differences have been reported between healthy individuals³⁷ we elaborate an original strategy of gene selection using their variances in AD versus physiological context.

Recent studies discovered¹⁷ and validated¹³ prospectively the existence of four AD endotypes using unsupervised clustering based on a knowledge-driven set of blood biomarkers. These are efficient strategies for endotyping patients using low invasive and cost-effective procedures. Our

study differs from the above ones on two main points. First, we used skin transcriptomic data considering that the skin microenvironment would be more representative of AD pathophysiology. Second, we did not follow preconceived ideas for selecting the genes of interest, in order to be able to identify original genes and new mechanisms. This may explain why important genes in AD pathophysiology were not retained in our analyses. Low variance genes were excluded (e.g. IL4R = 0.24, TSLP = 0.05, IL17A = 0.03, IL22RA1 = 0.096, FLG = 0.34, CLDN1 = 0.12), and only upregulated genes in one cluster versus others were used for functional signature (which excluded IL19 and IL22, as an example).

While down-regulation of IL-34 and IL-37 expressions have already been described in AD in comparison to controls^{31,32}, the role of IL-37 as a negative immune regulator³⁸, especially in a Th2 context has just been reported³⁹. Also, IL-34 can lead to a pro- or anti-inflammatory state, depending on the context⁴⁰, and has been recently described as inversely-correlated with AD severity²⁸. FLG2 and LOR are part of the epidermal differentiation complex, they act as important mediators of skin architecture. As a reflection of skin reconstruction defects, they are also downregulated in comparison to healthy skin⁴¹. In our study, IL-34 and IL-37 are part of MG-II that groups skin architecture genes such as FLG2, LOR, OGN, and skin homeostasis genes. We showed that MG-II expression is strongly inversely-correlated with AD severity and is the functional signature of the less severe endotype. This supports the notion that co-expression of anti-inflammatory and skin reconstruction signals may be responsible for a protective effect.

The more severe SCORAD and SA colonized endotype was characterized by the up-regulation of MG-I that was enriched in IL-1 family signaling, epidermal proliferation, and differentiation. This metagene shows a high level of overlap with the molecular signature of psoriasis, in particular Th17 immunity⁴². Although it is often described as downregulated in AD compared with psoriasis⁴³, it could be upregulated according to the endotype⁴⁴ and participate in AD development⁴⁵. As part of Th17 polarization, upregulated IL-36 pathway is responsible for mendelian monogenic psoriasis⁴⁶, besides IL-36G and IL-36A have been described as upregulated in AD³⁰. In contrast with MG-II, pro-inflammatory IL-1 signaling is grouped with genes of keratinocyte proliferation (such as KRT6A and C) and proteases (such as PRSS22 and PI15) which could reveal a disorganized tissue response. As in asthma, where a Th17 endotype has been described⁴⁷, our findings could suggest the IL-36 blockers use in the AD therapeutic arsenal.

Unexpectedly, one endotype was strongly enriched in B cell immunity (MG-III) whose role in AD is not well delineated. While circulating B lymphocytes do not differ in AD patients compared to controls, they do differ quantitatively and qualitatively at the skin level^{48,49}. In addition, B-cell-targeted therapeutics showed disappointing results in AD⁵⁰. Association with allergic comorbidities or high IgE levels could have been expected, but MG-III expression was not associated with any clinical or biological features. In further studies, a higher number of samples and a deeper clinical characterization may specify the role of B cell immunity in AD endotypes. These preliminary findings could suggest the use of B cell-targeting therapies for a minority of AD patients.

Application of patient stratification based on transcriptomic data in daily care is still very rare. The validation step of those data is critical and should be carried out in each study, at least on independent and ideally prospective cohorts. However, validation of a transcriptomic signature could be a long process. Breast cancer is currently the only situation in which a transcriptomic signature is routinely used in patient management^{51,52}. In our study, we decided to validate our main biological findings taking advantage of a large independent skin biopsy dataset, with a comparable transcriptomic technology. However, the question of technologies and the observation scales remains. While RNA sequencing offers a better resolution than microarray, the latter is still commonly used in the largest skin biopsy AD cohorts and remains robust for bulk transcriptomic analyses⁵³. In addition, classification studies require large cohorts. Those are limited by the fact that sampling AD patients and control volunteers involve invasive skin biopsies that are not common practice in AD management, especially in children. This could be addressed by implementing recent non-invasive and validated methods such as skin tape strippings^{28,31,54}. In addition to innovative sampling techniques, newer high-throughput approaches could allow the detection of gene expression with higher resolution. Indeed, the first studies using single-cell RNA sequencing have been recently published^{55–58} and have begun dissecting AD heterogeneity at the cellular level.

PERSPECTIVES

Our logical and intuitive method selected biologically relevant genes from a large amount of data. Combined with a data-driven unsupervised clustering, it enabled reproducible AD endotype identification. This pipeline could be applied in other complex diseases for which classification remains challenging. Regarding AD, endotypes were characterized by distinct disease severities and conflicting biological mechanisms. This could help to better understand AD heterogeneity as well as develop new and personalized therapeutic approaches.

World count: 3832

REFERENCES

1. Langan SM, Irvine AD, Weidinger S. Atopic dermatitis. *The Lancet*. 2020;396(10247):345-360. doi:10.1016/S0140-6736(20)31286-1
2. Deckers IAG, McLean S, Linssen S, Mommers M, van Schayck CP, Sheikh A. Investigating International Time Trends in the Incidence and Prevalence of Atopic Eczema 1990–2010: A Systematic Review of Epidemiological Studies. *PLoS ONE*. 2012;7(7):28.
3. Barbarot S, Auziere S, Gadkari A, et al. Epidemiology of atopic dermatitis in adults: Results from an international survey. *Allergy*. Published online February 13, 2018. doi:10.1111/all.13401
4. Weidinger S, Beck LA, Bieber T, Kabashima K, Irvine AD. Atopic dermatitis. *Nat Rev Dis Primers*. 2018;4(1):1. doi:10.1038/s41572-018-0001-z
5. Stefanovic N, Flohr C, Irvine AD. The exposome in atopic dermatitis. *Allergy*. 2020;75(1):63-74. doi:10.1111/all.13946
6. Reinhold U, Kukel S, Goeden B, Neumann U, Kreysel HW. Functional characterization of skin-infiltrating lymphocytes in atopic dermatitis. *Clinical & Experimental Immunology*. 1991;86(3):444-448. doi:10.1111/j.1365-2249.1991.tb02951.x
7. Carmi-Levy I, Homey B, Soumelis V. A modular view of cytokine networks in atopic dermatitis. *Clin Rev Allergy Immunol*. 2011;41(3):245-253. doi:10.1007/s12016-010-8239-6
8. Szabó K, Gáspár K, Dajnoki Z, et al. Expansion of circulating follicular T helper cells associates with disease severity in childhood atopic dermatitis. *Immunology Letters*. 2017;189:101-108. doi:10.1016/j.imlet.2017.04.010
9. Russell CD, Baillie JK. Treatable traits and therapeutic targets: Goals for systems biology in infectious disease. *Current Opinion in Systems Biology*. 2017;2:140-146. doi:10.1016/j.coisb.2017.04.003
10. Wüthrich B. Atopic neurodermatitis. *Wien Med Wochenschr*. 1989;139(6-7):156-165.

- Accepted Article
11. Czarnowicki T, He H, Krueger JG, Guttman-Yassky E. Atopic dermatitis endotypes and implications for targeted therapeutics. *Journal of Allergy and Clinical Immunology*. 2019;143(1):1-11. doi:10.1016/j.jaci.2018.10.032
 12. Thijs JL, de Bruin-Weller MS, Hijnen D. Current and Future Biomarkers in Atopic Dermatitis. *Immunology and Allergy Clinics of North America*. 2017;37(1):51-61. doi:10.1016/j.iac.2016.08.008
 13. Bakker DS, Nierkens S, Knol EF, et al. Confirmation of multiple endotypes in atopic dermatitis based on serum biomarkers. *Journal of Allergy and Clinical Immunology*. Published online June 2020:S0091674920308010. doi:10.1016/j.jaci.2020.04.062
 14. Bakker DS, Drylewicz J, Nierkens S, et al. Early identification of atopic dermatitis patients in need of systemic immunosuppressive treatment. *Clin Exp Allergy*. 2019;49(12):1641-1644. doi:10.1111/cea.13495
 15. Ewald DA, Malajian D, Krueger JG, et al. Meta-analysis derived atopic dermatitis (MADAD) transcriptome defines a robust AD signature highlighting the involvement of atherosclerosis and lipid metabolism pathways. *BMC Medical Genomics*. 2015;8(1). doi:10.1186/s12920-015-0133-x
 16. Ghosh D, Ding L, Sivaprasad U, et al. Multiple Transcriptome Data Analysis Reveals Biologically Relevant Atopic Dermatitis Signature Genes and Pathways. Simon M, ed. *PLOS ONE*. 2015;10(12):e0144316. doi:10.1371/journal.pone.0144316
 17. Thijs JL, Strickland I, Bruijnzeel-Koomen CAFM, et al. Moving toward endotypes in atopic dermatitis: Identification of patient clusters based on serum biomarker analysis. *Journal of Allergy and Clinical Immunology*. 2017;140(3):730-737. doi:10.1016/j.jaci.2017.03.023
 18. Fyhrquist N, Muirhead G, Prast-Nielsen S, et al. Microbe-host interplay in atopic dermatitis and psoriasis. *Nat Commun*. 2019;10(1):4703. doi:10.1038/s41467-019-12253-y
 19. Ottman N, Barrientos-Somarribas M, Fyhrquist N, et al. Microbial and transcriptional differences elucidate atopic dermatitis heterogeneity across skin sites. *Allergy*. Published online October 14, 2020:all.14606. doi:10.1111/all.14606

20. Hanifin, J.M. R G. Diagnostic Features of Atopic Dermatitis. *Acta Derm Venereol*. Published online 1980.
21. European Task Force on Atopic Dermatitis. Severity scoring of atopic dermatitis: the SCORAD index. Consensus Report of the European Task Force on Atopic Dermatitis. *Dermatology (Basel)*. 1993;186(1):23-31. doi:10.1159/000247298
22. Guttman-Yassky E, Bissonnette R, Ungar B, et al. Dupilumab progressively improves systemic and cutaneous abnormalities in patients with atopic dermatitis. *Journal of Allergy and Clinical Immunology*. 2019;143(1):155-172. doi:10.1016/j.jaci.2018.08.022
23. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*. 2007;23(14):1846-1847. doi:10.1093/bioinformatics/btm254
24. Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. *Nucleic Acids Res*. 2002;30(1):42-46. doi:10.1093/nar/30.1.42
25. Shannon P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*. 2003;13(11):2498-2504. doi:10.1101/gr.1239303
26. Bindea G, Mlecnik B, Hackl H, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009;25(8):1091-1093. doi:10.1093/bioinformatics/btp101
27. Alarcón-Riquelme ME. New Attempts to Define and Clarify Lupus. *Curr Rheumatol Rep*. 2019;21(4):11. doi:10.1007/s11926-019-0810-4
28. He H, Bissonnette R, Wu J, et al. Tape strips detect distinct immune and barrier profiles in atopic dermatitis and psoriasis. *Journal of Allergy and Clinical Immunology*. Published online July 2020:S0091674920308241. doi:10.1016/j.jaci.2020.05.048
29. Toro-Dom D, Carmona-S P. Stratification of Systemic Lupus Erythematosus Patients Into Three Groups of Disease Activity Progression According to Longitudinal Gene Expression. :11.

30. Tsoi LC, Rodriguez E, Degenhardt F, et al. Atopic Dermatitis Is an IL-13–Dominant Disease with Greater Molecular Heterogeneity Compared to Psoriasis. *Journal of Investigative Dermatology*. 2019;139(7):1480-1489. doi:10.1016/j.jid.2018.12.018
31. Guttman-Yassky E, Diaz A, Pavel AB, et al. Use of Tape Strips to Detect Immune and Barrier Abnormalities in the Skin of Children With Early-Onset Atopic Dermatitis. *JAMA Dermatol*. 2019;155(12):1358. doi:10.1001/jamadermatol.2019.2983
32. Esaki H, Ewald DA, Ungar B, et al. Identification of novel immune and barrier genes in atopic dermatitis by means of laser capture microdissection. *Journal of Allergy and Clinical Immunology*. 2015;135(1):153-163. doi:10.1016/j.jaci.2014.10.037
33. Hira ZM, Gillies DF. A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. *Advances in Bioinformatics*.:13.
34. Sørlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences*. 2001;98(19):10869-10874.
35. Dozmorov I. Hypervariable genes--experimental error or hidden dynamics. *Nucleic Acids Research*. 2004;32(19):e147-e147. doi:10.1093/nar/gnh146
36. Thijs JL, Strickland I, Bruijnzeel-Koomen CAFM, et al. Serum biomarker profiles suggest that atopic dermatitis is a systemic disease. *Journal of Allergy and Clinical Immunology*. 2018;141(4):1523-1526. doi:10.1016/j.jaci.2017.12.991
37. Wright FA, Sullivan PF, Brooks AI, et al. Heritability and genomics of gene expression in peripheral blood. *Nat Genet*. 2014;46(5):430-437. doi:10.1038/ng.2951
38. Wang L, Quan Y, Yue Y, Heng X, Che F. Interleukin-37: A crucial cytokine with multiple roles in disease and potentially clinical therapy (Review). *Oncol Lett*. Published online February 7, 2018. doi:10.3892/ol.2018.7982
39. Li W, Ding F, Zhai Y, et al. IL-37 is protective in allergic contact dermatitis through mast cell inhibition. *International Immunopharmacology*. 2020;83:106476. doi:10.1016/j.intimp.2020.106476

40. Guillon C, Bézie S, Anegón I. Immunoregulatory properties of the cytokine IL-34. *Cell Mol Life Sci.* 2017;74(14):2569-2586. doi:10.1007/s00018-017-2482-4
41. Guttman-Yassky E, Suárez-Fariñas M, Chiricozzi A, et al. Broad defects in epidermal cornification in atopic dermatitis identified through genomic analysis. *Journal of Allergy and Clinical Immunology.* 2009;124(6):1235-1244.e58. doi:10.1016/j.jaci.2009.09.031
42. Li J, Chen X, Liu Z, Yue Q, Liu H. Expression of Th17 cytokines in skin lesions of patients with psoriasis. *J Huazhong Univ Sci Technolog Med Sci.* 2007;27(3):330-332. doi:10.1007/s11596-007-0329-1
43. Guttman-Yassky E, Lowes MA, Fuentes-Duculan J, et al. Low Expression of the IL-23/Th17 Pathway in Atopic Dermatitis Compared to Psoriasis. *J Immunol.* 2008;181(10):7420-7427. doi:10.4049/jimmunol.181.10.7420
44. Suárez-Fariñas M, Dhingra N, Gittler J, et al. Intrinsic atopic dermatitis shows similar TH2 and higher TH17 immune activation compared with extrinsic atopic dermatitis. *Journal of Allergy and Clinical Immunology.* 2013;132(2):361-370. doi:10.1016/j.jaci.2013.04.046
45. Sugaya M. The Role of Th17-Related Cytokines in Atopic Dermatitis. *IJMS.* 2020;21(4):1314. doi:10.3390/ijms21041314
46. Akiyama M, Takeichi T, McGrath JA, Sugiura K. Autoinflammatory keratinization diseases. *J Allergy Clin Immunol.* 2017;140(6):1545-1547. doi:10.1016/j.jaci.2017.05.019
47. Östling J, van Geest M, Schofield JPR, et al. IL-17-high asthma with features of a psoriasis immunophenotype. *Journal of Allergy and Clinical Immunology.* 2019;144(5):1198-1213. doi:10.1016/j.jaci.2019.03.027
48. Czarnowicki T, Esaki H, Gonzalez J, et al. Alterations in B-cell subsets in pediatric patients with early atopic dermatitis. *J Allergy Clin Immunol.* 2017;140(1):134-144.e9. doi:10.1016/j.jaci.2016.09.060
49. Czarnowicki T, Gonzalez J, Bonifacio KM, et al. Diverse activation and differentiation of multiple B-cell subsets in patients with atopic dermatitis but not in patients with psoriasis. *J Allergy Clin Immunol.* 2016;137(1):118-129.e5. doi:10.1016/j.jaci.2015.08.027

50. Wang W-M, Guo L, Jin H-Z. Role of B cells in immune-mediated dermatoses. *Mol Immunol*. 2020;126:95-100. doi:10.1016/j.molimm.2020.07.016
51. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347(25):1999-2009. doi:10.1056/NEJMoa021967
52. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med*. 2016;375(8):717-729. doi:10.1056/NEJMoa1602253
53. Suárez-Fariñas M, Ungar B, Correa da Rosa J, et al. RNA sequencing atopic dermatitis transcriptome profiling provides insights into novel disease mechanisms with potential therapeutic implications. *Journal of Allergy and Clinical Immunology*. 2015;135(5):1218-1227. doi:10.1016/j.jaci.2015.03.003
54. Hulshof L, Hack DP, Hasnoe QCJ, et al. A minimally invasive tool to study immune response and skin barrier in children with atopic dermatitis. *Br J Dermatol*. 2019;180(3):621-630. doi:10.1111/bjd.16994
55. He H, Suryawanshi H, Morozov P, et al. Single-cell transcriptome analysis of human skin identifies novel fibroblast subpopulation and enrichment of immune subsets in atopic dermatitis. *Journal of Allergy and Clinical Immunology*. 2020;145(6):1615-1628. doi:10.1016/j.jaci.2020.01.042
56. Rojahn TB, Vorstandlechner V, Krausgruber T, et al. Single-cell transcriptomics combined with interstitial fluid proteomics defines cell type-specific immune regulation in atopic dermatitis. *Journal of Allergy and Clinical Immunology*. Published online April 2020:S009167492030556X. doi:10.1016/j.jaci.2020.03.041
57. Bangert C, Rindler K, Krausgruber T, et al. Persistence of mature dendritic cells, TH2A, and Tc2 cells characterize clinically resolved atopic dermatitis under IL-4R α blockade. *Sci Immunol*. 2021;6(55). doi:10.1126/sciimmunol.abe2749

58. Kalinina P, Vorstandlechner V, Buchberger M, et al. The Whey Acidic Protein WFDC12 Is Specifically Expressed in Terminally Differentiated Keratinocytes and Regulates Epidermal Serine Protease Activity. *J Invest Dermatol*. Published online November 4, 2020. doi:10.1016/j.jid.2020.09.025

FIGURE LEGENDS

Figure 1: AD-specific hyper-variable genes are representative of known AD mechanisms.

(a) Workflow of feature selection and functional annotation. First, we selected only hyper-variable genes in AD lesional skin (variance > 0.5). Second, we kept genes that were twice more variable than in healthy skin (ratio: variance AD / Control skin > 2) (Fig 1b). From the 222 AD-specific hyper-variable genes, we performed several functional enrichments that highlighted known AD mechanisms (Fig 1c, Suppl 1b).

(b) AD-specific hyper-variable gene selection. Among AD lesional skin, $n = 575$ genes were highly variable (variance > 0.5). Among them, $n = 222$ were twice more variable than in healthy skin and called AD-specific hyper-variable genes.

(c) Functional enrichment of all AD-specific hyper-variable genes with GO biological process. It confirmed the AD-specific hyper-variable gene role as representative of AD known pathophysiology. Network representation of significative terms (adj.p >0.05). Nodes represent main GO terms. Edges symbolize interactions between nodes.

AD: atopic dermatitis, GO: gene ontology

Figure 2: Unsupervised clustering of samples in four AD clusters with distinct clinical features, using AD-specific hyper-variable genes.

(a) Workflow of sample clustering and clinical characterization. The 222 AD-specific hyper-variable genes have been used to cluster samples. The optimal number of clusters was estimated at four by all the different methods (Fig 2b and Suppl 3). The different clustering methods have been compared to choose the most consensual and consistent (Suppl 4).

(b) *k*-means clustering using 4 as the optimal number of clusters according to the elbow method. *K*-means clusters are represented in a dimension reduction PCA plot.

(c) Variable distribution comparison between clusters: disease severity (SCORAD) and SA colonization are statistically different. Student tests have been used for inter-cluster comparisons.

AD: atopic dermatitis, PCA: principal component analysis, MST: Minimal Spanning Tree, SA: Staphylococcus aureus, SCORAD: score AD

Figure 3: Metagene expression in sample clusters reveal exciting molecular and functional signatures.

(a) Workflow of discriminant genes final selection, grouping, and functional annotation to build molecular signature of AD clusters. We used the four AD clusters made from the 222 AD-specific hyper-variable genes. We compare gene expression in one AD cluster versus the others (Fig 3b). Significantly upregulated genes were grouped in metagenes (Fig 3c) that acted as AD cluster molecular signatures (Fig 3d).

(b) Heatmap representation of differentially expressed genes expression among samples and AD clusters. Genes as part of a unique metagene had the same behavior according to sample clusters.

(c) Metagenes construction by correlation analysis using Pearson statistic. We select unique and up-regulated genes from the differential expression analysis and grouped correlated genes into their respective metagenes.

(d) Metagene expression among sample clusters. ANOVA test showed differential expression of metagenes between clusters.

AD: atopic dermatitis, ANOVA: Analysis of variance, MG: Metagene

Figure 4: Data integration for multiscale AD cluster characterization and external validation.

(a) Workflow of the data integration step. The different phases were independent of each other.

(b) Correlation between metagene expressions and clinical features using Spearman statistics. Significant positive correlations were shown between MG-I, SCORAD, and SA colonization. Oppositely, significant negative correlations were found between MG-II, SCORAD, SA colonization. No significant results were found between metagenes expression and raised IgE levels (binary variable) or age (continuous variable).

(c) Metagene expression and disease severity disparities among clusters in an independent dataset (GSE130588). AD cluster colors have been chosen relatively to their mean severity score. ANOVA test showed metagenes as differentially expressed between clusters. As in our cohort, the AD cluster with the more severe SCORAD had an upregulation of MG-I, and the AD cluster with the less severe SCORAD had an upregulation of MG-II.

(d) Correlation of metagene and disease severity in an independent dataset (GSE130588). Spearman statistics confirmed the positive correlation between MG-I expression and disease severity as well as the negative correlation between MG-II expression and disease severity.

AD: atopic dermatitis, ANOVA: Analysis of variance, MG: metagene, SA: Staphylococcus aureus, SCORAD: score AD

SUPPLEMENTARY FIGURE LEGENDS

Supplementary 1:

- (a) Demographical data of AD and control cohorts.
- (b) Functional enrichment of all AD-specific hyper-variable genes based on the KEGG database. The top 5 statistically significant biological processes are ranked by adjusted p-value.

Supplementary 2: The 222 AD specific hyper-variable genes with their respective variance in the pathological and physiological states. Genes are ranked by ratio.

Supplementary 3: Cluster number optimization.

- (a) Sample k-NN and MST clustering in four clusters. The number written within each circle refers to the sample identity.
- (b) The highest average silhouette width corresponded to four sample clusters.
- (c) Hierarchical clustering: Euclidian distance has been used to constitute four clusters.

Supplementary 4: Clustering methods comparison. *k*-means clustering method has been chosen for sample clustering because of its highest intra-cluster (Dunn index) and inter-cluster (Rand index and Jaccard similarity coefficient) metrics.

Supplementary 5: Clinical features representation within AD clusters.

- (a) Table of clinical, biological, and demographical variables among clusters. The statistically different clinical parameters were: SCORAD, and SA colonization (adj.p<0.05). Kruskal Wallis test has been used for continuous, and Fisher test for Boolean variables. Medians and interquartile ranges are provided for quantitative variables, except for SA colonization for which the mean is provided.
- (b) Important non-significant clinical features: early age of disease onset, IgE concentration, gender, age.

OTU: Operational taxonomic unit, SA: Staphylococcus aureus, SCORAD: score AD

Supplementary 6: AD clusters transcriptomic signature.

- (a) Differentially expressed genes in one cluster versus the others.

(b) Unique and upregulated differentially expressed genes grouped in their metagenes.

AD: atopic dermatitis

Supplementary 7: Literature mining manually curated signature revealed interesting and distinct metagene functions. Computational approaches were not suitable for this analysis due to the low number of genes per metagene. Thus, a manually curated approach was favored. Genes belonging to one metagene shared common biological functions.

Supplementary 8: Expression comparisons, between AD clusters, of important genes already described in AD studies. Classical functional groups such as Th2, Th17 polarizations, and skin barrier did not show homogeneous expression profiles within AD clusters.

(a) Th2 immunity genes.

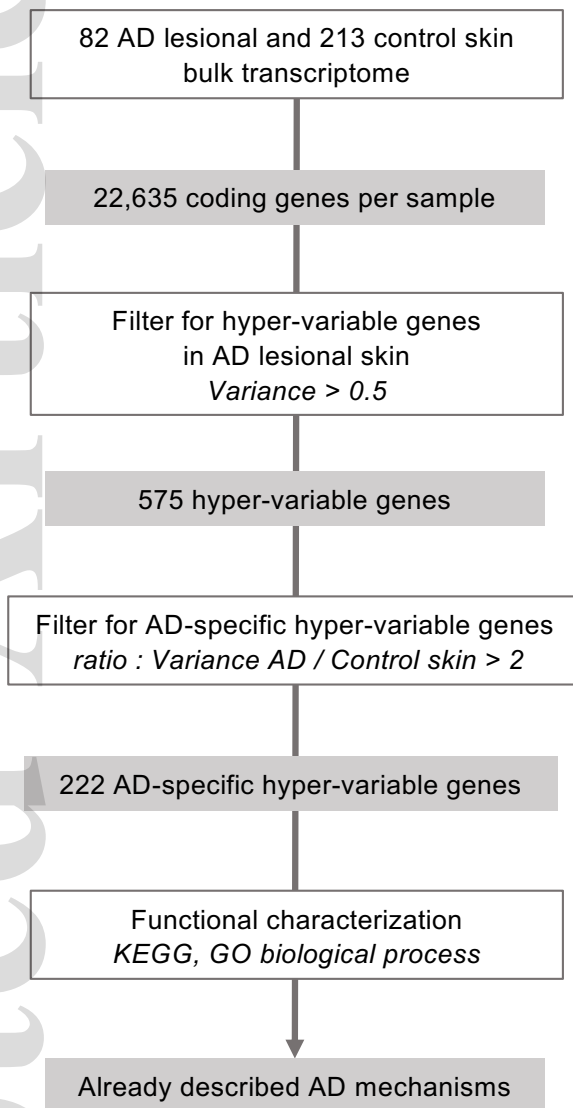
(b) Th17 immunity genes.

(c) Skin barrier genes.

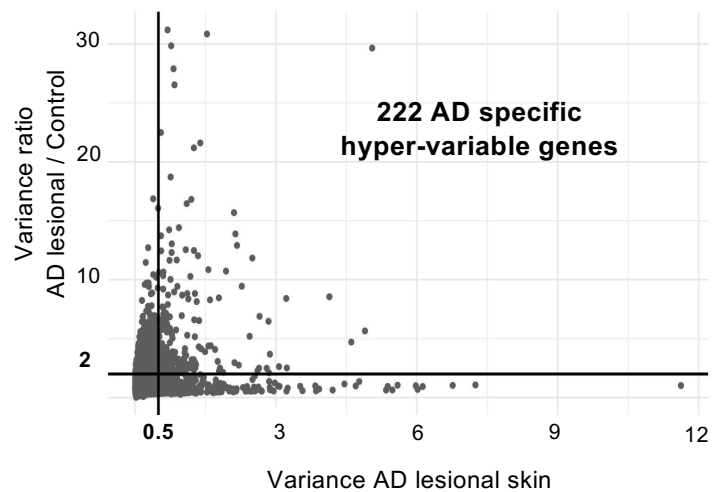
AD: atopic dermatitis, Th: T-lymphocyte helper

Supplementary 9: Comprehensive list of independent skin AD transcriptomes with available clinical annotations on disease severity. The largest expression matrix (GSE130588) was used to validate our main findings.

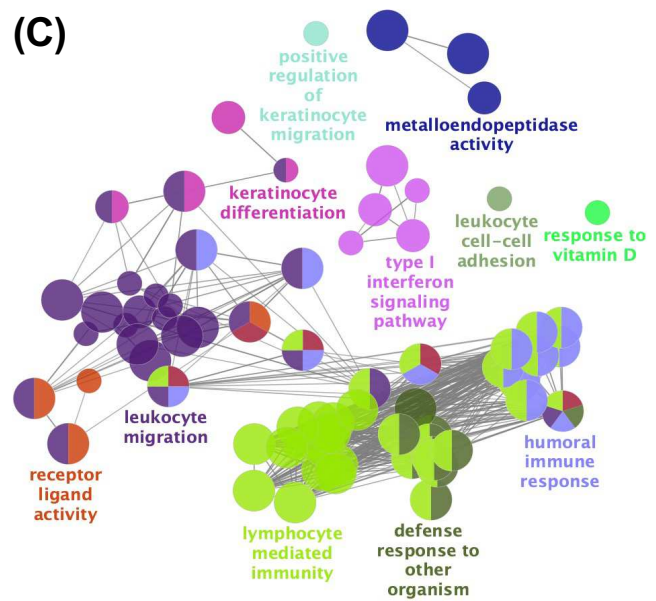
(A)

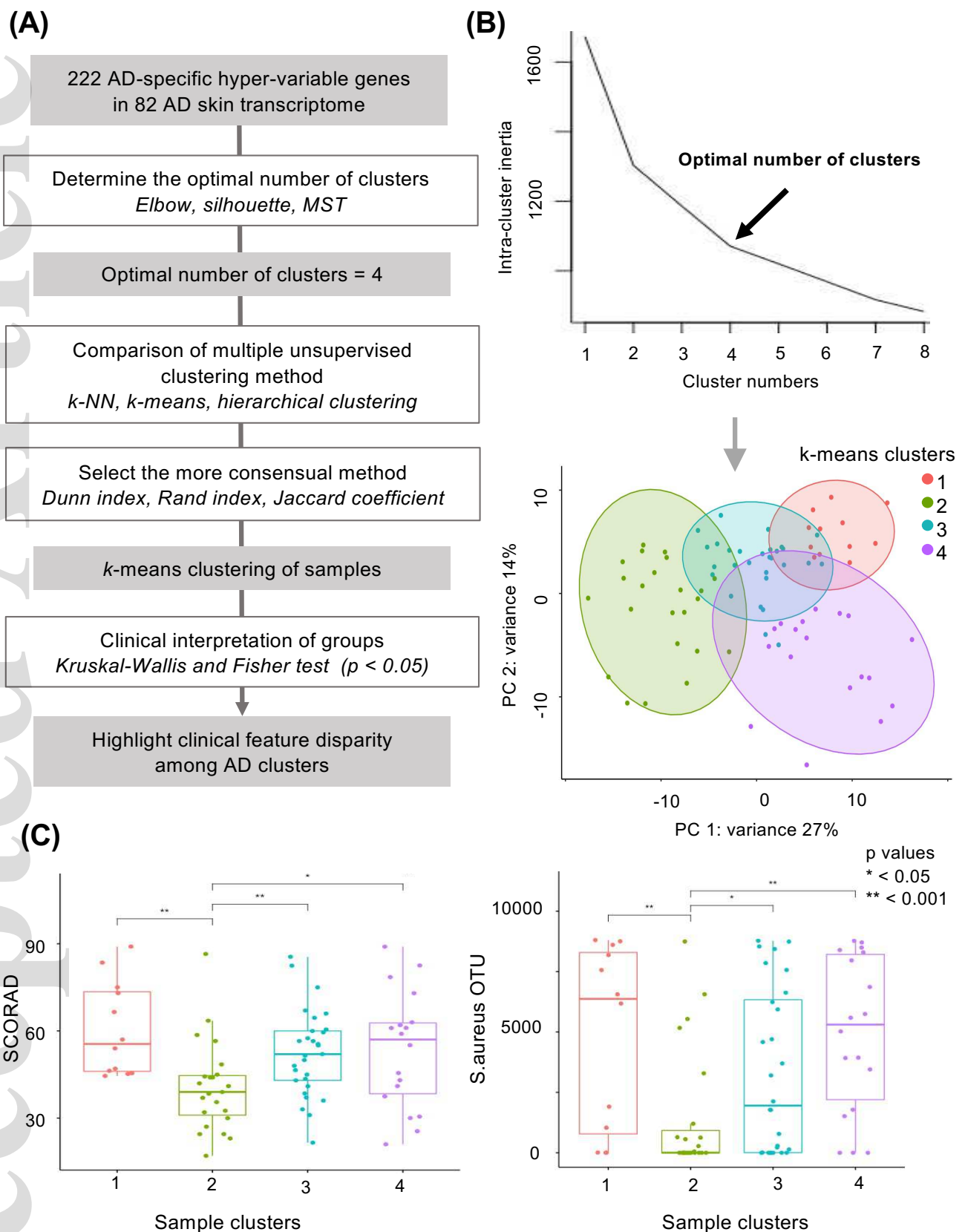


(B)

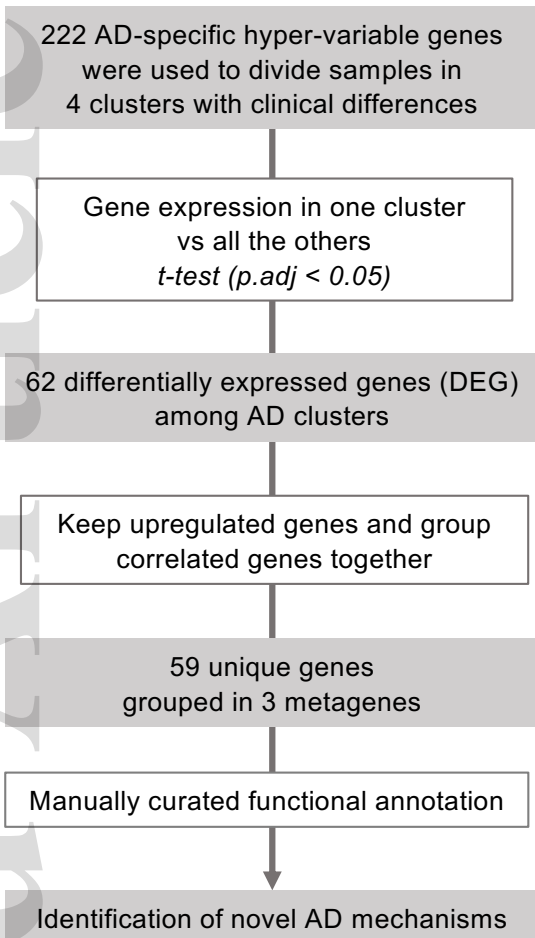


(C)

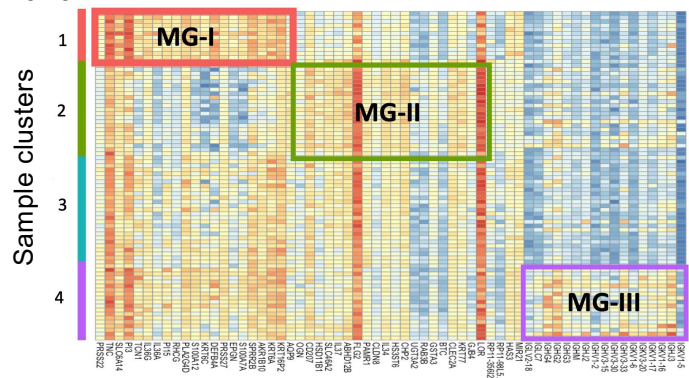




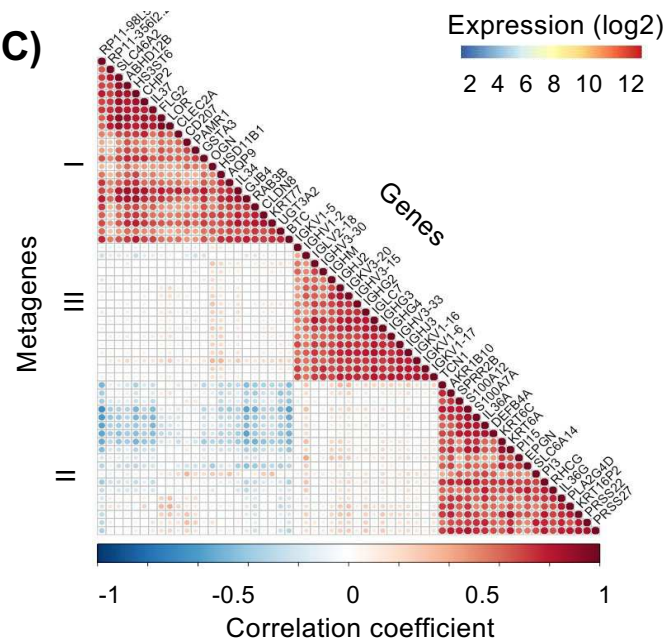
(A)



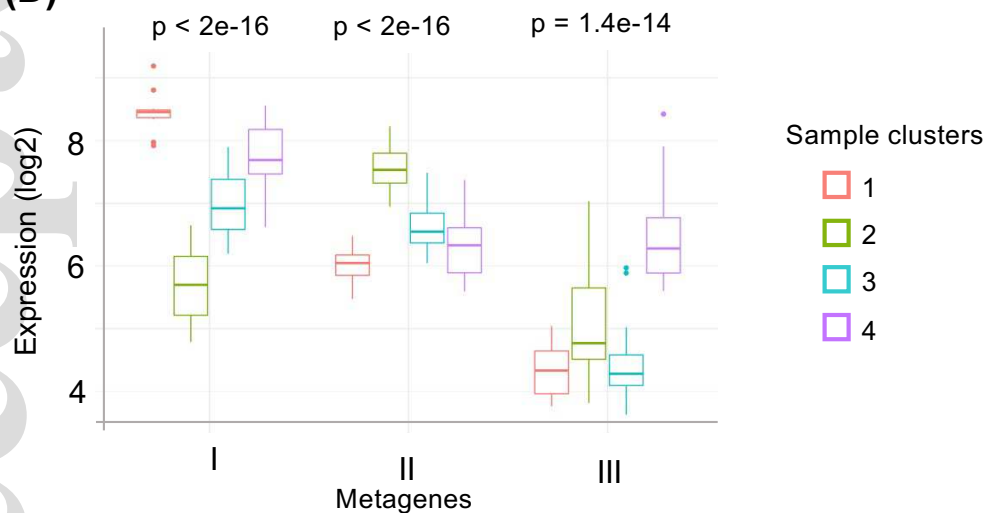
(B)

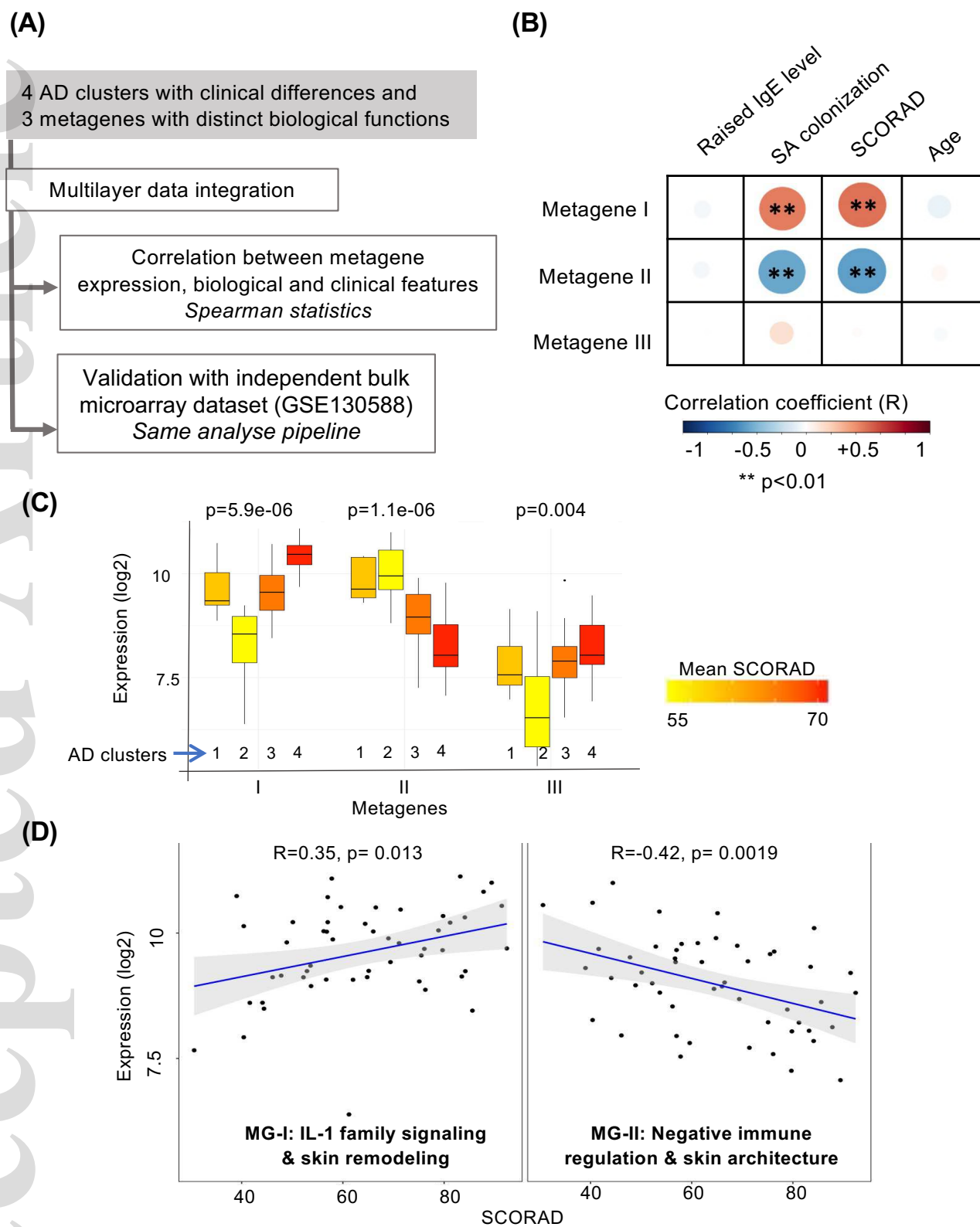


(C)



(D)





Supplementary 1

a

Demographical data of AD and control cohort

	AD patients	Controls
Individual (nb)	82	113
Sample (nb)	82	213
Age mean (yo) [min-max]	44 [20-83]	35 [19-77]
Gender (F/M)	36/46	69/44
Ethnicity	Caucasian 91%	Caucasian 88%
SCORAD	52 [17-89]	-

b

Functional enrichment of all AD-specific hyper-variable genes with KEGG database

Term name	Terme ID	adj.p
Cytokine-cytokine receptor interaction	KEGG:04060	3.4 e-9
Viral protein interaction with cytokine	KEGG:04061	1.4 e-8
IL-17 signaling pathway	KEGG:04657	2.8 e-4
Influenza A	KEGG:05164	3.3 e-3
AGE-RAGE signalling pathway in diabetic complications	KEGG:04933	3.9 e-3

Supplementary 2

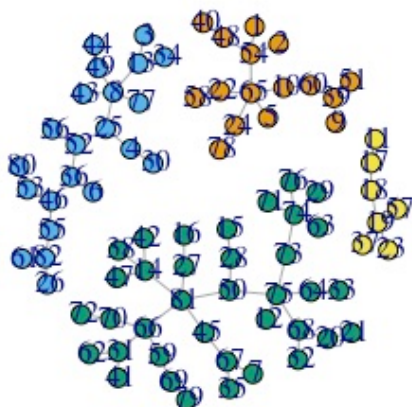
	GeneSymbol	Variance in AD	Variance in controls	Ratio	ENSG00000211598	IGKV4-1	0.565	0.098	5.8
ENSG00000175426	PCSK1	0.695	0.022	31.2	ENSG00000226145	AC022596.6	0.876	0.153	5.7
ENSG00000197915	HRNR	1.534	0.05	30.8	ENSG00000171711	DEFB4A	4.897	0.867	5.7
ENSG00000127318	IL22	0.772	0.026	29.8	ENSG00000131969	ABHD12B	0.755	0.134	5.6
ENSG00000196611	MMP1	5.05	0.17	29.7	ENSG00000119508	NR4A3	0.514	0.092	5.6
ENSG00000162892	IL24	0.822	0.029	27.9	ENSG00000159337	PLA2G4D	0.679	0.122	5.6
ENSG00000142224	IL19	0.842	0.032	26.5	ENSG00000137440	FGFBP1	0.755	0.136	5.5
ENSG00000254651	RP11-430H10.3	0.554	0.025	22.5	ENSG00000169213	RAB3B	0.788	0.148	5.3
ENSG00000244057	LCE3C	1.391	0.064	21.6	ENSG00000169248	CXCL11	1.113	0.211	5.3
ENSG00000108702	CCL1	1.255	0.059	21.2	ENSG00000200972	RNU5A-8P	2.442	0.47	5.2
ENSG00000105641	SLC5A5	0.76	0.041	18.7	ENSG00000094804	CDC6	0.523	0.101	5.2
ENSG00000169429	IL8	1.201	0.071	16.8	ENSG00000125571	IL37	1.276	0.247	5.2
ENSG00000006074	CCL18	1.102	0.067	16.5	ENSG00000137648	TMPRSS4	0.8	0.159	5.0
ENSG00000149968	MMP3	2.109	0.134	15.7	ENSG00000211900	IGHJ6	4.605	0.979	4.7
ENSG00000211936	IGHV4-4	0.936	0.065	14.4	ENSG00000189182	KRT77	0.743	0.16	4.7
ENSG00000143520	FLG2	0.729	0.051	14.2	ENSG00000211838	TRAJ52	0.578	0.126	4.6
ENSG00000185962	LCE3A	2.139	0.154	13.9	ENSG00000112984	KIF20A	0.527	0.117	4.5
ENSG00000162891	IL20	0.554	0.04	13.7	ENSG00000168671	UGT3A2	1.614	0.367	4.4
ENSG00000123496	IL13RA2	0.784	0.06	13.0	ENSG00000239855	IGKV1-6	1.566	0.358	4.4
ENSG00000182585	EPGN	2.173	0.168	12.9	ENSG00000149090	PAMR1	0.516	0.119	4.3
ENSG00000103569	AQP9	1.082	0.086	12.5	ENSG00000117594	HSD11B1	1.358	0.316	4.3
ENSG00000134028	ADAMDEC1	1.256	0.101	12.5	ENSG00000138642	HERC6	0.65	0.153	4.2
ENSG00000166509	CLEC3A	0.559	0.045	12.4	ENSG00000157601	MX1	0.669	0.159	4.2
ENSG00000212556	Y_RNA	0.781	0.063	12.3	ENSG00000126787	DLGAP5	0.56	0.133	4.2
ENSG00000153802	TMPRSS11D	1.343	0.112	12.0	ENSG00000172382	PRSS27	0.673	0.16	4.2
ENSG00000136694	IL36A	2.499	0.211	11.8	ENSG00000231475	IGHV4-31	1.065	0.257	4.2
ENSG00000163661	PTX3	0.89	0.076	11.7	ENSG00000177257	DEFB4B	1.41	0.341	4.1
ENSG00000182566	CLEC4G	0.736	0.063	11.6	ENSG00000187498	COL4A1	0.542	0.132	4.1
ENSG00000175592	FOSL1	1.564	0.144	10.8	ENSG00000105664	COMP	1.706	0.418	4.1
ENSG00000188293	IGFL1	1.939	0.181	10.7	ENSG00000137558	PI15	1.494	0.384	3.9
ENSG00000181617	FDCSP	0.606	0.057	10.6	ENSG00000106366	SERPINE1	0.803	0.206	3.9
ENSG00000236481	AC002331.1	0.51	0.048	10.6	ENSG00000159167	STC1	0.856	0.223	3.8
ENSG00000158859	ADAMTS4	0.6	0.058	10.3	ENSG00000006327	TNFRSF12A	0.7	0.183	3.8
ENSG00000206384	COL6A6	1.181	0.115	10.3	ENSG00000236543	RP11-98L5.5	1.059	0.278	3.8
ENSG00000203782	LOR	0.757	0.076	10.0	ENSG00000211753	TRBV28	0.618	0.163	3.8
ENSG00000110347	MMP12	2.275	0.241	9.4	ENSG00000134321	RSAD2	0.572	0.152	3.8
ENSG00000171889	MIR31HG	0.904	0.096	9.4	ENSG000000012223	LTF	2.873	0.78	3.7
ENSG00000223872	AC006372.5	0.601	0.065	9.2	ENSG00000151006	PRSS53	0.617	0.17	3.6
ENSG00000189433	GJB4	0.836	0.093	9.0	ENSG00000184613	NELL2	0.834	0.232	3.6
ENSG00000248329	RP11-366M4.3	1.114	0.126	8.8	ENSG00000039987	BEST2	0.643	0.179	3.6
ENSG00000198074	AKR1B10	1.277	0.145	8.8	ENSG00000211904	IGHJ2	1.09	0.304	3.6
ENSG00000162040	HS3ST6	0.706	0.081	8.7	ENSG00000005001	PRSS22	0.512	0.143	3.6
ENSG00000204936	CD177	1.006	0.116	8.7	ENSG00000187116	LILRA5	0.688	0.193	3.6
ENSG00000170465	KRT6C	4.139	0.484	8.6	ENSG00000231331	AC103563.2	0.74	0.209	3.5
ENSG00000102837	OLFM4	1.787	0.211	8.5	ENSG00000111335	OAS2	0.711	0.203	3.5
ENSG00000184330	S100A7A	3.221	0.383	8.4	ENSG00000211866	TRAJ23	0.721	0.206	3.5
ENSG00000205362	MT1A	1.137	0.136	8.4	ENSG00000211952	IGHV4-28	0.754	0.216	3.5
ENSG00000105205	CLC	1.598	0.193	8.3	ENSG00000211880	TRAJ9	0.508	0.148	3.4
ENSG00000213886	UBD	1.315	0.161	8.1	ENSG00000137959	IFI44L	1.052	0.314	3.4
ENSG00000227471	AKR1B15	0.699	0.089	7.9	ENSG00000161905	ALOX15	1.047	0.314	3.3
ENSG00000115758	ODC1	0.584	0.076	7.7	ENSG00000152268	SPON1	0.552	0.166	3.3
ENSG00000166736	HTR3A	0.603	0.079	7.6	ENSG00000211950	IGHV1-24	0.616	0.186	3.3
ENSG00000104368	PLAT	0.663	0.093	7.2	ENSG00000211956	IGHV4-34	0.683	0.207	3.3
ENSG00000103044	HAS3	0.605	0.085	7.1	ENSG00000166948	TGM6	0.563	0.171	3.3
ENSG00000211966	IGHV5-51	0.997	0.143	7.0	ENSG00000158485	CD1B	0.923	0.282	3.3
ENSG00000163221	S100A12	2.652	0.385	6.9	ENSG00000184557	SOC3	0.656	0.201	3.3
ENSG00000138135	CH25H	0.743	0.108	6.9	ENSG00000183813	CCR4	0.515	0.158	3.3
ENSG00000163600	ICOS	0.57	0.086	6.6	ENSG00000162739	SLAMF6	0.62	0.191	3.2
ENSG00000211791	TRAV13-2	0.638	0.096	6.6	ENSG00000211949	IGHV3-23	0.718	0.223	3.2
ENSG00000157368	IL34	0.543	0.082	6.6	ENSG00000111012	CYP27B1	0.612	0.19	3.2
ENSG00000243466	IGKV1-5	1.263	0.191	6.6	ENSG00000140519	RHCG	1.252	0.389	3.2
ENSG00000211653	IGLV1-40	1.367	0.209	6.5	ENSG00000041982	TNC	0.654	0.204	3.2
ENSG00000134827	TCN1	2.844	0.44	6.5	ENSG00000205678	TECRL	0.763	0.238	3.2
ENSG00000163638	ADAMTS9	0.525	0.081	6.5	ENSG00000119125	GDA	1.207	0.377	3.2
ENSG00000103888	KIAA1199	0.661	0.102	6.5	ENSG00000162692	VCAM1	0.527	0.167	3.1
ENSG00000102970	CCL17	0.509	0.08	6.4	ENSG00000116031	CD207	0.63	0.2	3.1
ENSG00000227300	KRT16P2	0.638	0.102	6.3	ENSG00000136688	IL36G	0.73	0.232	3.1
ENSG00000166869	CHP2	0.681	0.109	6.3	ENSG00000211875	TRAJ14	0.542	0.173	3.1
ENSG00000169385	RNASE2	0.681	0.109	6.3	ENSG00000205364	MT1M	0.581	0.186	3.1
ENSG00000188404	SELL	0.598	0.099	6.1	ENSG00000211853	TRAJ36	0.539	0.175	3.1
ENSG00000119457	SLC46A2	0.744	0.126	5.9	ENSG00000240382	IGKV1-17	1.761	0.574	3.1
ENSG00000124731	TREM1	0.669	0.116	5.8	ENSG00000241244	IGKV1D-16	0.546	0.179	3.0

ENSG00000239951	IGKV3-20	0.922	0.305	3.0
ENSG00000117228	GBP1	0.524	0.173	3.0
ENSG00000211945	IGHV1-18	0.838	0.279	3.0
ENSG00000102962	CCL22	0.961	0.32	3.0
ENSG00000174502	SLC26A9	0.74	0.248	3.0
ENSG00000007908	SELE	1.278	0.429	3.0
ENSG00000235842	RP11-356I2.2	1.052	0.354	3.0
ENSG00000211953	IGHV3-30	2.119	0.715	3.0
ENSG00000140534	TICRR	0.514	0.175	2.9
ENSG00000200648	U6	0.517	0.176	2.9
ENSG00000164687	FABP5	0.507	0.174	2.9
ENSG00000211860	TRAJ29	0.503	0.173	2.9
ENSG00000199004	MIR21	0.696	0.24	2.9
ENSG00000185745	IFIT1	0.515	0.178	2.9
ENSG00000119938	PPP1R3C	0.525	0.184	2.9
ENSG00000107984	DKK1	0.681	0.24	2.8
ENSG00000240834	IGKV1D-12	1.198	0.423	2.8
ENSG00000211934	IGHV1-2	1.102	0.39	2.8
ENSG00000240864	IGKV1-16	1.301	0.464	2.8
ENSG00000156284	CLDN8	0.558	0.199	2.8
ENSG00000172752	COL6A5	0.92	0.33	2.8
ENSG00000127954	STEAP4	0.524	0.188	2.8
ENSG00000168685	IL7R	0.606	0.22	2.8
ENSG00000196805	SPRR2B	2.209	0.802	2.8
ENSG00000183760	PAPL	0.541	0.198	2.7
ENSG00000158488	CD1E	0.601	0.221	2.7
ENSG00000174156	GSTA3	0.579	0.213	2.7
ENSG00000140285	FGF7	0.564	0.208	2.7
ENSG00000211899	IGHM	0.985	0.371	2.7
ENSG00000188393	CLEC2A	0.527	0.199	2.6
ENSG00000211892	IGHG4	3.067	1.167	2.6
ENSG00000251616	RP11-485M7.3	0.583	0.223	2.6
ENSG00000119917	IFIT3	0.547	0.211	2.6
ENSG00000147138	GPR174	0.625	0.242	2.6
ENSG00000211859	TRAJ30	0.521	0.202	2.6
ENSG00000169313	P2RY12	0.524	0.204	2.6
ENSG00000184348	HIST1H2AK	0.676	0.266	2.5
ENSG00000211867	TRAJ22	0.797	0.317	2.5
ENSG00000242887	IGHJ3	3.233	1.285	2.5
ENSG00000199377	RNU5F-1	1.107	0.44	2.5
ENSG00000211943	IGHV3-15	1.81	0.723	2.5
ENSG00000188257	PLA2G2A	1.284	0.513	2.5
ENSG00000211673	IGLV3-1	1.189	0.475	2.5
ENSG00000124102	PI3	2.64	1.058	2.5
ENSG00000211955	IGHV3-33	2.804	1.127	2.5
ENSG00000087916	SLC6A14	0.927	0.374	2.5
ENSG00000211858	TRAJ31	1.01	0.411	2.5
ENSG00000211940	IGHV3-9	0.504	0.206	2.4
ENSG00000137965	IFI44	0.739	0.307	2.4
ENSG00000162654	GBP4	0.518	0.216	2.4
ENSG00000211664	IGLV2-18	1.256	0.529	2.4
ENSG00000174808	BTC	0.6	0.253	2.4
ENSG00000211855	TRAJ34	0.571	0.242	2.4
ENSG00000113070	HBEGF	0.795	0.34	2.3
ENSG00000153234	NR4A2	0.528	0.227	2.3
ENSG00000100985	MMP9	0.503	0.216	2.3
ENSG00000243290	IGKV1-12	0.892	0.383	2.3
ENSG00000211893	IGHG2	2.611	1.147	2.3
ENSG00000211685	IGLC7	1.791	0.793	2.3
ENSG00000211765	TRBJ2-2	0.548	0.245	2.2
ENSG00000170801	HTRA3	0.564	0.254	2.2
ENSG00000211884	TRAJ5	0.545	0.25	2.2
ENSG00000171246	NPTX1	0.585	0.268	2.2
ENSG00000249437	NAIP	0.595	0.274	2.2
ENSG00000211897	IGHG3	1.077	0.498	2.2
ENSG00000211856	TRAJ33	0.62	0.289	2.1
ENSG00000177575	CD163	0.699	0.327	2.1
ENSG00000205420	KRT6A	1.876	0.878	2.1
ENSG00000211873	TRAJ16	0.653	0.308	2.1
ENSG00000090659	CD209	0.618	0.293	2.1
ENSG00000169245	CXCL10	2.853	1.355	2.1
ENSG00000115008	IL1A	0.98	0.466	2.1
ENSG00000211699	TRGV3	0.603	0.288	2.1

Supplementary 3

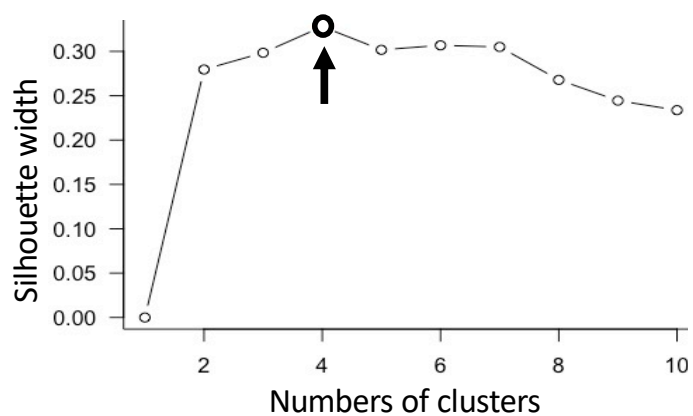
a

***k*-NN/MST sample clustering in four clusters**



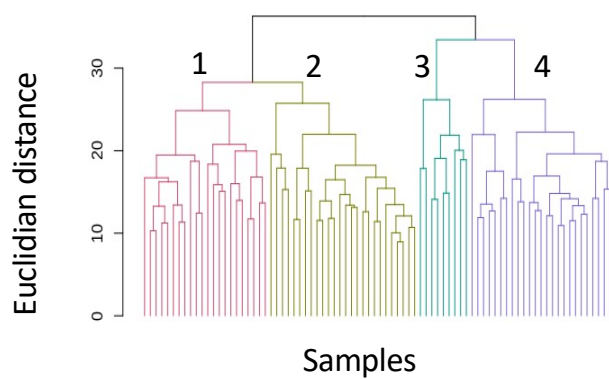
b

Average silhouette width in all samples



c

Hierarchical clustering of samples in four clusters



Supplementary 4

Intra and inter-cluster metrics used to select the more consensual method

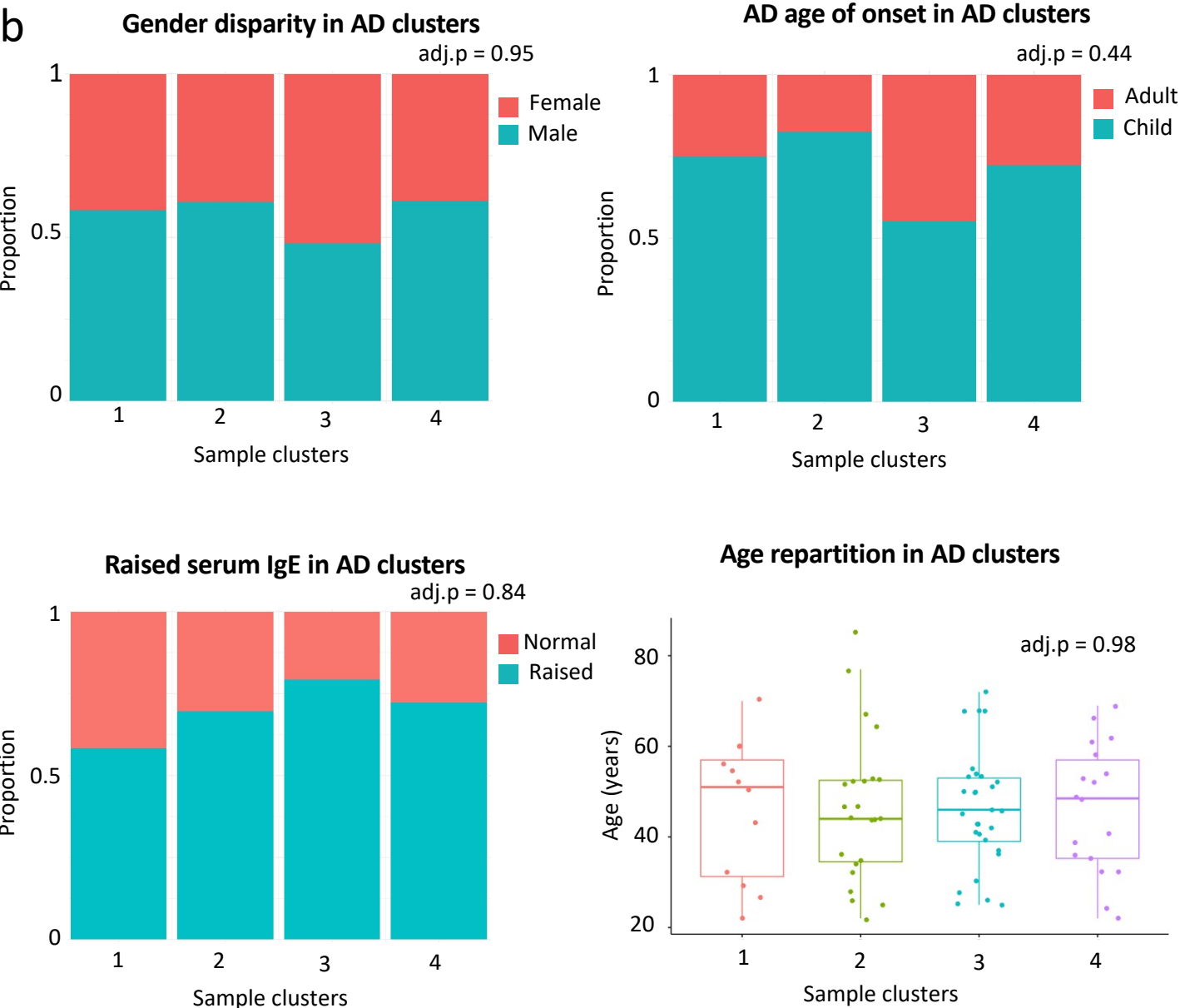
Intra-cluster metric			
Clustering methods	Hierarchical clustering	<i>k</i> -NN clustering	<i>k</i> -means clustering
Dunn Index	0,42	0,34	0,43

Inter-cluster metrics			
Clustering methods	Hierarchical clustering	<i>k</i> -NN clustering	<i>k</i> -means clustering
Rand Index			
Hierarchical clustering		0,68	0,79
<i>k</i> -NN clustering	0,68		0,73
<i>k</i> -means clustering	0,79	0,73	
Jaccard similarity coefficient			
Hierarchical clustering		0,30	0,44
<i>k</i> -NN clustering	0,30		0,38
<i>k</i> -means clustering	0,44	0,38	

Supplementary 5

a

	AD cluster 1	AD cluster 2	AD cluster 3	AD cluster 4	adjusted p
Patient number (/82)	12	23	29	18	-
Personal history					
Age (years)	51 [31-57]	44 [34-52]	46 [39-53]	48.5 [35-57]	0.98
Gender (M F)	7 5	14 9	14 15	11 7	0.95
Early age of AD onset	75%	83%	55%	72%	0.44
Allergies	83%	74%	72%	78%	0.95
Familial history					
Allergies	83%	52%	69%	72%	0.52
Clinical features					
SCORAD (/103)	55 [46-73]	39 [31-45]	52 [43-60]	57 [38-63]	0.02
Biopsy localization (back/tigh)	2 10	14 9	12 17	10 8	0.21
Biological features					
SA colonization (mean OTU)	4801	1420	3286	4914	0.02
Raised serum IgE level	58%	70%	79%	72%	0.84



Supplementary 6

a Differentially expressed genes in one cluster vs the others

Cluster 1 vs 2, 3 and 4	Cluster 2 vs 1, 3 and 4	Cluster 3 vs 1, 3 and 4	Cluster 4 vs 1, 2, and 3
AKR1B10 DEFB4A IL36A IL36G EPGN KRT6A KRT6C KRT16P2 PI3 PI15 PLA2G4D PRSS22 PRSS27 RHCG S100A7A S100A12 SLC6A14 SPRR2B TCN1 TNC	ABHD12B AQP9 BTC CD207 CLDN8 CLEC2A CHP2 HS3ST6 HSD11B1 FLG2 GJB4 GSTA3 IL34 IL37 KRT77 LOR OGN PAMR1 RAB3B RP11-356I2.2 RP11-98L5.5 SLC46A2 UGT3A2	No gene	AKR1B10 HAS3 IGLV2-18 IGLC7 IGHG4 IGHG2 IGHG3 IGHM IGHJ2 IGHV1-2 IGHV3-15 IGHV3-30 IGHV3-33 IGKV1-6 IGKV3-20 IGKV1-17 IGKV1-16 IGHJ3 IGKV1-5 KRT6A MIR21 PI3 PRSS27 S100A12 TCN1

b Upregulated differentially expressed genes grouped in corresponding metagenes

Metagenes		
1	2	3
AKR1B10 DEFB4A EPGN IL36A IL36G KRT16P2 KRT6A KRT6C PI15 PI3 PLA2G4D PRSS22 PRSS27 RHCG S100A12 S100A7A SLC6A14 SPRR2B TCN1	ABHD12B AQP9 BTC CD207 CHP2 CLDN8 CLEC2A FLG2 GJB4 GSTA3 HS3ST6 HSD11B1 IL34 IL37 KRT77 LOR OGN PAMR1 RAB3B RP11-356I2.2 RP11-98L5.5 SLC46A2 UGT3A2	IGLV2-18 IGLC7 IGHG4 IGHG2 IGHG3 IGHM IGHJ2 IGHV1-2 IGHV3-15 IGHV3-30 IGHV3-33 IGKV1-6 IGKV3-20 IGKV1-17 IGKV1-16 IGHJ3 IGKV1-5

Supplementary 7

Metagene I (19 genes) – IL-1 family signaling & skin remodeling

Th17 : DEFB4A, IL36A, IL36G, PI3, PLA2G4D, S100A7A, S100A12

Epidermal proliferation and differentiation : EPGN, KRT6A, KRT6C, AKR1B10, SPRR2B

Protease activity: PRSS22, SLC6A14, PI15, TCN1

Poorly described : KRT16P2, PRSS27, RHCG

Metagene II (23 genes) – Negative immune regulation & skin architecture

Negative immune regulation: CLEC2A (=PILAR), IL34, IL37

Skin architecture: FLG2, LOR, OGN

Epidermal homeostasis: BTC, GSTA3, GJB4, KRT77, UGT3A2,

Water metabolism: AQP9, CLDN8

Langerhans cell marker: CD207

Lipase: ABHD12B

Endogenous steroid activation: HSD11B1

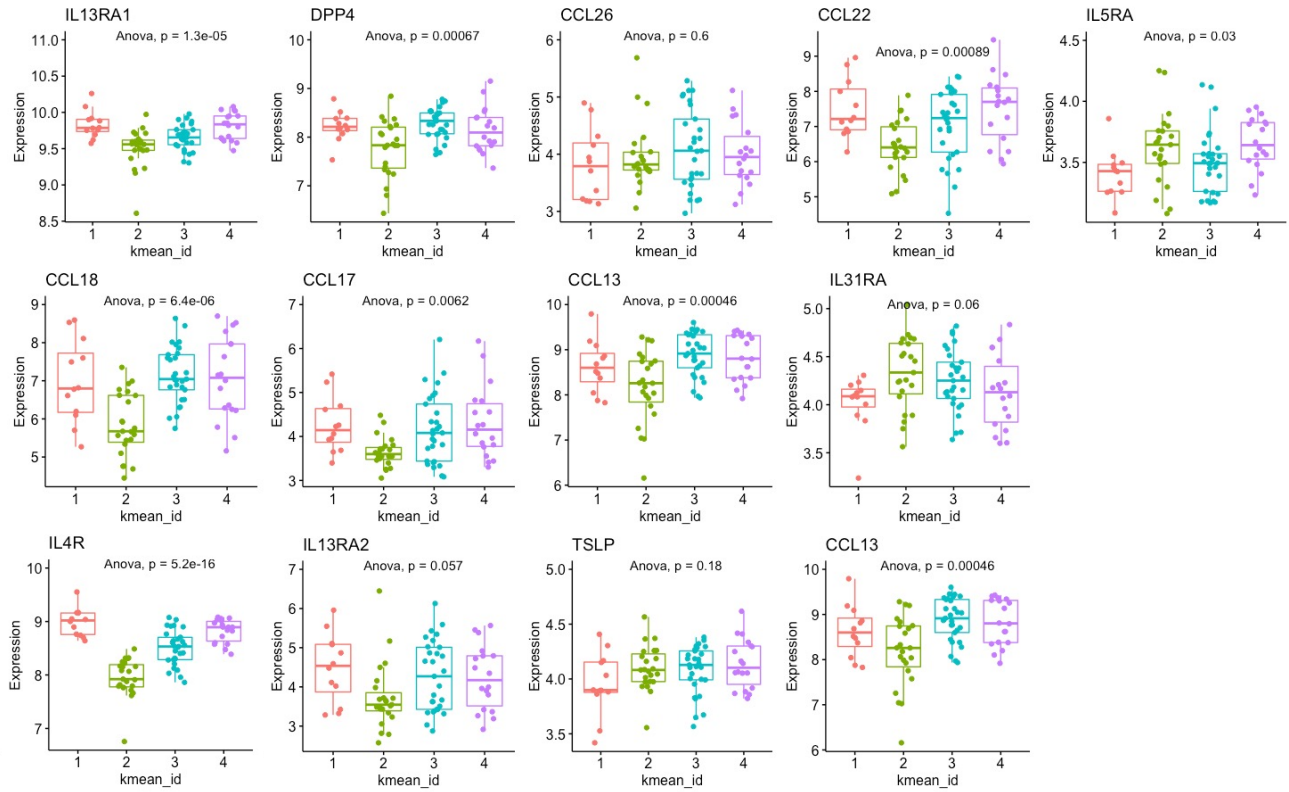
Poorly described : CHP2, HS3ST6, PAMR1, RAB3B, RP11-356I2.2, RP11-98L5.5, SLC46A2

Metagene III (17 genes) – B lymphocyte immunity

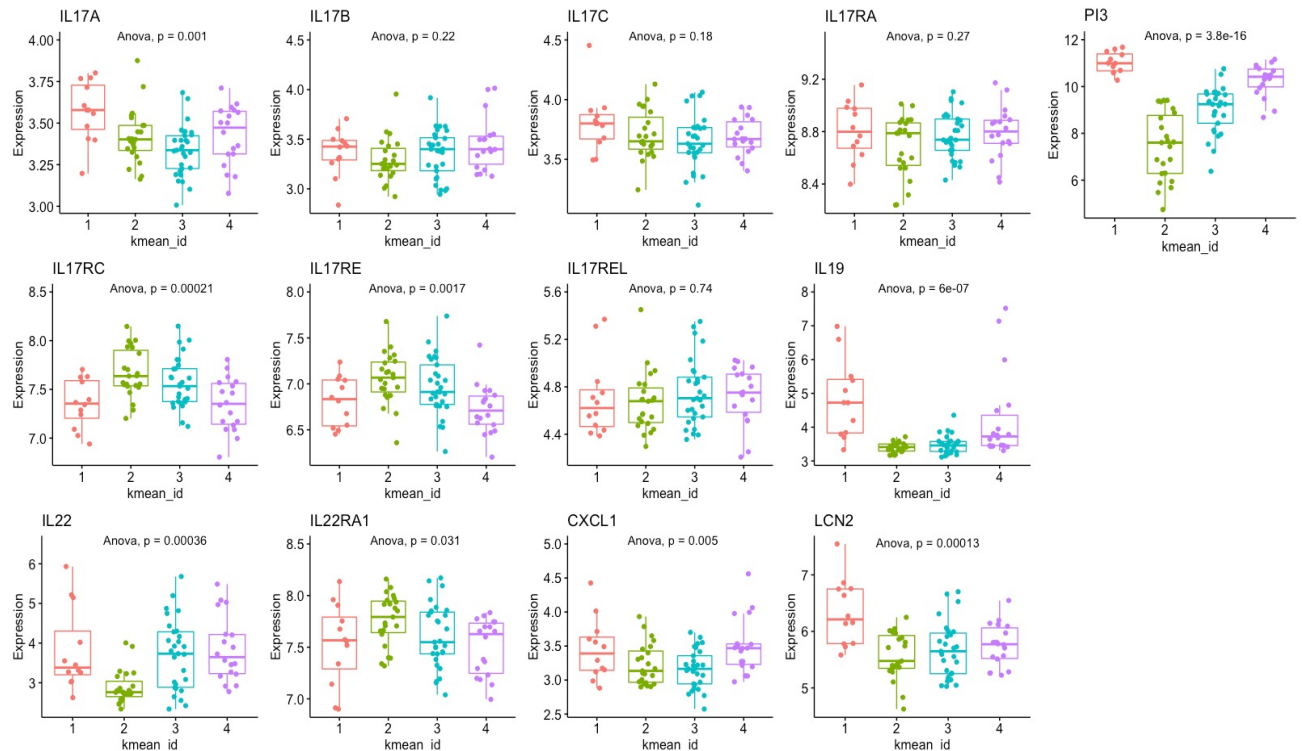
Immunoglobulin recombination: IGLV2-18, IGLC7, IGHG4, IGHG2, IGHG3, IGHM, IGHJ2, IGHV1-2, IGHV3-15, IGHV3-30, IGHV3-33, IGKV1-6, IGKV3-20, IGKV1-17, IGKV1-16, IGHJ3, IGKV1-5

Supplementary 8

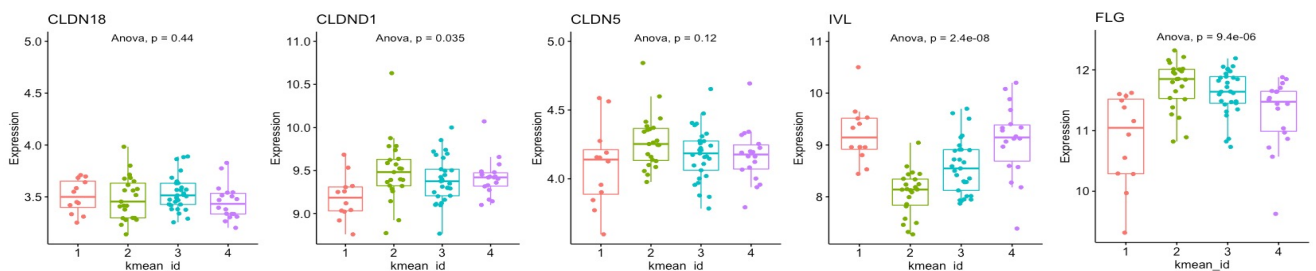
a



b



c



Supplementary 9

TITLE	DOI	GEO	ARRAY EXPRESS	AUTHOR	YEAR	TECH_MICROARRAY	SAMPLES	AGE	SCOR AD
Dupilumab progressively improves systemic and cutaneous abnormalities in patients with atopic dermatitis	10.1016/j.jaci.2018.08.022	GSE130588	NA	E. Guttman-Yasski et al	2018	Affymetrix Human U133Plus 2.0 gene arrays	51	Adult	YES
Cyclosporine in patients with atopic dermatitis modulates activated inflammatory pathways and reverses epidermal pathology	10.1016/j.jaci.2014.03.003	GSE58558	E-GEOD-58558	S. Khattri et al	2014	Affymetrix Human U133Plus 2.0 gene arrays	12	Adult	YES
Dupilumab improves the molecular signature in skin of patients with moderate-to-severe atopic dermatitis	10.1016/j.jaci.2014.10.013	GSE59294	E-GEOD-59294	J.D. Hamilton et al	2014	Affymetrix Human U133Plus 2.0 gene arrays	18	Adult	YES
Progressive Activation of Th2/Th22 characterizes acute and chronic atopic dermatitis	10.1016/j.jaci.2012.07.012	GSE36842	NA	JK. Gittler et al	2012	Affymetrix Human U133Plus 2.0 gene arrays	17	Adult	YES
Reversal of atopic dermatitis with narrow-band UVB phototherapy and biomarkers for therapeutic response	10.1016/j.jaci.2011.05.042	GSE27887	E-GEOD-27887	S Tintle et al	2011	Affymetrix Human U133Plus 2.0 gene arrays	12	Adult	YES
Nonlesional atopic dermatitis skin is characterized by broad terminal differentiation defects and variable immune abnormalities	10.1016/j.jaci.2010.12.1124	GSE32924	NA	M Suárez-Fariña et al	2011	Affymetrix Human U133Plus 2.0 gene arrays	12	Adult	YES