DR. NATHANAËL LEMONNIER (Orcid ID : 0000-0002-3994-8697)
DR. ERICK FORNO (Orcid ID : 0000-0001-6497-9885)
DR. INGER KULL (Orcid ID : 0000-0001-6096-3771)
PROF. CEZMI AKDIS (Orcid ID : 0000-0001-8020-019X)
MRS. MÜBECCEL AKDIS (Orcid ID : 0000-0002-9228-4594)
DR. VALÉRIE SIROUX (Orcid ID : 0000-0001-7329-7237)
DR. CHENG-JIAN XU (Orcid ID : 0000-0003-1586-4672)
DR. JUAN C. CELEDON (Orcid ID : 0000-0002-6139-5320)

Article type : Original Article: Epidemiology and Genetics

Title

A novel whole blood gene expression signature for asthma, dermatitis and rhinitis multimorbidity in children and adolescents

Short title

Multimorbidity in asthma, dermatitis and rhinitis

Authors

Nathanaël Lemonnier[1*†], PhD, Erik Melén[2,3*], MD, PhD, Yale Jiang[4,5*], MDc, Stéphane Joly[6], MSc, Camille Ménard[6], MSc, Daniel Aguilar[7], PhD, Edna Acosta-Perez[8], PhD, Anna Bergström[3], PhD, Nadia Boutaoui[4], PhD, Mariona Bustamante[9-11], PhD, Glorisa Canino[8], PhD, Erick Forno[4], MD, MPH, Juan Ramon González[9], PhD, Judith Garcia-Aymerich[9], MD, PhD, Olena Gruzieva[3], MD, Stefano Guerra[9,12], MD, PhD, Joachim Heinrich[13,14], PhD, Inger Kull[15], PhD, Jesús Ibarluzea Maurolagoitia[11,16], PhD, Loreto Santa-Marina Rodriguez[11], PhD, Elisabeth Thiering[13,17], PhD, Magnus Wickman[18], MD, Cezmi Akdis[19], MD, Mübeccel Akdis[19], MD, PhD, Wei Chen[4,20], PhD, Thomas Keil[21], MD, Gerard H. Koppelman[22,23], MD, PhD, Valérie Siroux[1], PhD, Cheng-Jian

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.14314

This article is protected by copyright. All rights reserved

Accepted Article

Xu[22,23], PhD, Pierre Hainaut[1], PhD, Marie Standl[13], PhD, Jordi Sunyer[9], MD, PhD, Juan C. Celedón[4]*, MD, DrPH, Josep Maria Antó[9]*, MD, PhD, Jean Bousquet[24]*, MD, PhD, and the MeDALL Study Group.

** authors contributed equally*

† corresponding author: nathanael.lemonnier@univ-grenoble-alpes.fr (ORCID 0000-0002-3994-8697)

Affiliations

[1] Institute for Advanced Biosciences, UGA-INSERM U1209-CNRS UMR5309, Site Santé, Allée des Alpes, 38700 La Tronche, France

[2] Sachs' Children's Hospital and Department of Clinical Science and Education Södersjukhuset, Karolinska Institutet, Stockholm, Sweden

[3] Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

[4] Division of Pediatric Pulmonary Medicine, UPMC Children's Hospital of Pittsburgh, University of Pittsburgh, Pittsburgh, PA

[5] School of Medicine, Tsinghua University, Beijing, China

[6] CIRI, International Center for Infectiology Research, Inserm, U1111, Lyon, France

[7] Biomedical Research Networking Center in Hepatic and Digestive Diseases (CIBEREHD), Barcelona, Spain

[8] Behavioral Sciences Research Institute, University of Puerto Rico, San Juan, Puerto Rico

[9] ISGlobal, Barcelona Institute for Global Health, Barcelona, Spain

[10] Universitat Pompeu Fabra (UPF), Barcelona, Spain

[11] Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Instituto de Salud Carlos III, C/Monforte de Lemos 3-5 28029 Madrid, Spain

[12] Asthma and Airway Disease Research Center, University of Arizona, Tucson, AZ, USA

[13] Institute of Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Munich, Germany

[14] Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Inner City Clinic, University Hospital of Munich (LMU), Munich, Germany

[15] Department of Clinical Science and Education, Sodersjukhuset, Karolinska Institute, 118 83 Stockholm, Sweden

This article is protected by copyright. All rights reserved

[16] BIODONOSTIA. Instituto de Investigación Sanitaria, Donostia-San Sebastián, Spain
[17] Ludwig-Maximilians-University of Munich, Dr. von Hauner Children's Hospital, Division of Metabolic Diseases and Nutritional Medicine, Munich, Germany
[18] Centre for Clinical Research Sörmland, Uppsala University, Eskilstuna Sweden
[19] Swiss Institute of Allergy and Asthma Research (SIAF), University of Zurich, Davos, Switzerland
[20] Department of Biostatistics, University of Pittsburgh, Pittsburgh, USA
[21] Institute of Social Medicine, Epidemiology and Health Economics, Charité – Universitätsmedizin Berlin, Berlin; Institute for Clinical Epidemiology and Biometry, University of Würzburg, Würzburg; and Institute for Health Resort Medicine and Health Promotion, Bavarian Health and Food Safety Authority, Bad Kissingen, Germany
[22] University of Groningen, University Medical Center Groningen, Beatrix Children's Hospital, Dpt of Pediatric Pulmonology and Pediatric Allergology
[23] University of Groningen, University Medical Center Groningen, GRIAC Research Institute
[24] MACVIA-France, Fondation partenariale FMC VIA-LR, Montpellier, France, Ageing and chronic diseases. Epidemiological and public health approaches, Villejuif, Université Versailles St-Quentin-en-Yvelines, UMR-S 1168, Montigny le Bretonneux, France, Euforea, Brussels, Belgium, and Charité - Universitätsmedizin Berlin, Comprehensive Allergy Center, Department of Dermatology and Allergy, Berlin, Germany

Acknowledgements

We thank all other investigators of the MeDALL Study Group for their advices and support: Annesi-Maesano Isabella, Baiz Nour, Bedbrook Anna, Cambon-Thomsen Anne, Jacquemin Benedicte, Kauffmann Francine, Pin Isabelle, Rial-Sebbag Emmanuelle, Nadif Rachel, Basagna Xavier, Benet Mora Marta, Kogevinas Manolis, Lavi Iris, Mestre Jordi, Pinart Mariona, Colli Matthias, Hettler-Chen Chih-Mei, Hohmann Cynthia, Keller Teresa, Lau Susanne, Schietinger Andrea, van Hofmann Ingrid, Worm Margitta, Zuberbier Torsten, Pison Christophe, Kerkhof Marjan, Nawijn Martijn C., van Oosterhout Antoon J. M., Wijmenga Cisca, Bachert Claus, Coquet Jonathan, Hammad Hamida, Lambrecht Bart, Saeys Yvan, Haahtela Tari, Hanninen Sinikka, Makela Mika, Reitamo Sakari, von Hertzen Leena, Klimek Magdalena, Kowalski Marek, Carlsen Kai Hakon, Lodrup-Carlsen Karin C., Baar Alexandra, Lupinek Christian, Pahr Sandra, Valenta

This article is protected by copyright. All rights reserved

Accepted Article

Rudolf, van de Veen Willem, Andersson Niklas, Ballardini Natalia, Johansson SGO, Kumar Ashish, Merid Simon Kebede, Thacher Jesse, van Hage Marianne, Westman Marit, Yazdanbakhsh Maria, Tischer Christina, Brunekreef Bert, Gehring Ulrike, Smit Henriette A, Le Naour Stéphanie, Smagghe Delphine, Albang Richard, Arno Albert, Mascaro Angels, Roda Xavier, Sanchez Cristina, Vega Mireia, Baumgartner Ursula, Neubauer Angela, Stolz Franck, McEachan Rosie, Oddie Sam, Petherick Emily, Raynor Pauline, Waiblinger Dagmar, Wright John, Martinez Fernando D., De Carlo Giuseppe, Palkonen Susanna, Salvi Roberta, Wecksell Per-Ake, Bindslev-Jensen Carsten, Eller Esben, Steensen Jens Peter, Forestiere Francesco, Narduzzi Silvia, Porta Daniela.

We acknowledge IRT BIOASTER for hosting MeDALL data production team: Alain Troesch and Nathalie Garçon ; Vincent Lotteau from INSERM for kindly hosting part of MeDALL data production ; and members of the CNRS USR3010: Charles Auffray, Stéphane Ballereau and Johann Pellet, plus Bertrand De Meulder and Diane Lefaudeux from U-BIOPRED project for the active discussions on sample selection and analyses. We thank Dieter Maier for the knowledge management platform and the data sharing between partners.

The authors thank all children and parents participating in the BAMSE cohort, the nurses, and other staff working with the BAMSE project.

The authors thank all the families for their participation in the GINIplus study. Furthermore, we thank all members of the GINIplus Study Group for their excellent work. The GINIplus Study group consists of the following: Institute of Epidemiology I, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg (Heinrich J, Brüske I, Schulz H, Flexeder C, Zeller C, Standl M, Schnappinger M, Ferland M, Thiering E, Tiesler C); Department of Pediatrics, Marien-Hospital, Wesel (Berdel D, von Berg A); Ludwig-Maximilians-University of Munich, Dr von Hauner Children's Hospital (Koletzko S); Child and Adolescent Medicine, University Hospital rechts der Isar of the Technical University Munich (Bauer CP, Hoffmann U); IUF- Environmental Health Research Institute, Düsseldorf (Schikowski T, Link E, Klümper C).

The authors thank all the families for their participation in the INMA project. A full list of INMA researchers can be found at http://www.proyectoinma.org/presentacion-inma/listado-investigadores/listado-investigadores.html.

This article is protected by copyright. All rights reserved

Accepted Article

Funding

This work was supported by the Mechanisms of the Development of ALLergy (MeDALL) project, European framework programme 7, Project No: 261357; 2010-2015.

NL is a recipient of a postdoctoral fellowship from the French National Research Agency in the framework of the 'Investissements d'avenir'' program (ANR-15-IDEX-02).

The work was also supported by PH under the Chair of Excellence in Translational Research, funded by the Canceropole Lyon Auvergne Rhône-Alpes (CLARA), the Centre Hospitalier Universitaire Grenoble Alpes (CHUGA), and the University Grenoble Alpes (UGA).

The BAMSE study was supported by The Swedish Research Council, The Swedish Heart and Lung Foundation, The Swedish Research Council for Working Life and Social Welfare, the Swedish Asthma and Allergy Association Research Foundation, The Swedish Research Council Formas, and Stockholm County Council (ALF). EM is supported by a grant from the European Research Council (n° 757919).

The GINIplus study was mainly supported for the first 3 years of the Federal Ministry for Education, Science, Research and Technology (interventional arm) and Helmholtz Zentrum Munich (former GSF) (observational arm). The follow-up examinations 4 years, 6 years, 10 years and 15 years of the GINIplus study were covered from the respective budgets of the 5 study centres (Helmholtz Zentrum Munich (former GSF), Research Institute at Marien-Hospital Wesel, LMU Munich, TU Munich and from 6 years onwards also from IUF - Leibniz Research-Institute for Environmental Medicine at the University of Düsseldorf) and a grant from the Federal Ministry for Environment (IUF Düsseldorf, FKZ 20462296). Further, the 15 years follow-up examination of the GINIplus study was supported by the Commission of the European Community, the 7th Framework Program: MeDALL project, and as well by the companies Mead Johnson and Nestlé.

INMA project was mainly supported by the Commission of the European Community, the 7th Framework Program: MeDALL project and by grants from Instituto de Salud Carlos III (Red INMA G03/176; CB06/02/0041; PI041436; PI081151 incl. FEDER funds), and Generalitat de Catalunya-CIRIT 1999SGR 00241.

This article is protected by copyright. All rights reserved

Accepted Article

The EVA-PR Study was supported by grants HL079966 and HL117191 from the U.S. National Institutes of Health (NIH). The contributions of J.C.C. and W.C. were additionally supported by grant MD011764 from the U.S. NIH, and RNA Sequencing was funded by the Department of Pediatrics of UPMC Children's Hospital of Pittsburgh. Research infrastructure for EVA-PR was additionally supported by grant U54MD007587 from the U.S. NIH. The content is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. NIH.

Accepted Article

This article is protected by copyright. All rights reserved

Abstract

Background. Allergic diseases often occur in combination (multimorbidity). Human blood transcriptome studies have not addressed multimorbidity. Large-scale gene expression data were combined to retrieve biomarkers and signaling pathways to disentangle allergic multimorbidity phenotypes.

Methods. Integrated transcriptomic analysis was conducted in 1,233 participants with a discovery phase using gene expression data (Human Transcriptome Array 2.0) from whole blood of 786 children from three European birth cohorts (MeDALL), and a replication phase using RNA-Seq data from an independent cohort (EVA-PR, n=447). Allergic diseases (asthma, atopic dermatitis, rhinitis) were considered as single disease or multimorbidity (at least two diseases), and compared with no disease.

Results. Fifty genes were differentially expressed in allergic diseases. Thirty-two were not previously described in allergy. Eight genes were consistently overexpressed in all types of multimorbidity for asthma, dermatitis and rhinitis (*CLC*, *EMR4P, IL5RA, FRRS1, HRH4, SLC29A1*, *SIGLEC8, IL1RL1*). All genes were replicated the in EVA-PR cohort. RT-qPCR validated the overexpression of selected genes. In MeDALL, 27 genes were differentially expressed in rhinitis alone, but none was significant for asthma or dermatitis alone. The multimorbidity signature was enriched in eosinophil-associated immune response and signal transduction. Protein-protein interaction network analysis identified IL5/JAK/STAT and IL33/ST2/IRAK/TRAF as key signaling pathways in multimorbid diseases. Synergistic effect of multimorbidity on gene expression levels was found.

Conclusion. A signature of eight genes identifies multimorbidity for asthma, rhinitis and dermatitis. Our results have clinical and mechanistic implications, and suggest that multimorbidity should be considered differently than allergic diseases occurring alone.

Keywords

Asthma, Atopic Dermatitis, Multimorbidity, Rhinitis, Transcriptomics.

This article is protected by copyright. All rights reserved

Accepted Article

Abbreviations

A: Asthma

BAMSE: Barn Allergi Miljö Stockholm Epidemiologi Projektet (Children Allergy, Milieu, Stockholm, an Epidemiological study)

D: Atopic Dermatitis

DEGs: Differentially Expressed Genes

EVA-PR: Epigenetic Variation and Childhood Asthma in Puerto Ricans

FC: Fold Change

FDR: False Discovery Rate

GINIplus: German Infant Study on the influence of Nutrition Intervention PLUS environmental and genetic influences on allergy development

HTA: Affymetrix Human Transcriptome Array 2.0

INMA: INfancia y Medio Ambiente (Environment and Childhood)

MeDALL: Mechanisms of the Development of ALLergy

PPIN: Protein-Protein Interaction Network

R: allergic Rhinitis

RT-qPCR: Reverse Transcription quantitative Polymerase Chain Reaction

WGCNA : Weighted Gene Co-Expression Network Analysis

Word count: 3,500 words

Accepted Article

This article is protected by copyright. All rights reserved

Accepted Article

Introduction

Allergic diseases (asthma: A, dermatitis: D, rhinitis: R) are complex and often coexist in the same patient (multimorbidity).[1] In children and adolescents, over 85% of asthmatics patients have rhinitis, suggesting common gene pathways, while only 20–30% of rhinitis patients have asthma suggesting rhinitis-specific genes.[2] Atopic dermatitis is characterized by a complex clinical phenotype varying during the life cycle.[3]

MeDALL (Mechanisms of the Development of ALLergy) found that multimorbidity was more common than expected by chance, suggesting genes underlying multimorbidity.[1,4] MeDALL proposed that type 2 patterns are associated with multimorbidity,[5] as suggested through an *in silico* model.[6] Risk variants shared between asthma, dermatitis and rhinitis were identified in genome wide association studies, without considering the impact of multimorbidity.[7,8] No study assessed multimorbidity using transcriptomics. Genes and pathways for atopy and atopic asthma in children and adolescents were found using transcriptome-wide data from peripheral blood.[9]

We used a subset of MeDALL samples as a discovery population to compare gene expression patterns between participants with and without asthma, dermatitis or rhinitis. We examined gene expression of single diseases and multimorbidity. We replicated the results in the Epigenetic Variation and Childhood Asthma in Puerto Ricans (EVA-PR) cohort.[9] To our knowledge, the present study constitutes to date the largest comprehensive whole blood gene expression dataset on allergy and the first assessment of multimorbidity molecular signatures.

This article is protected by copyright. All rights reserved

Methods

Study design. A cross-sectional analysis was carried out in participants from four cohorts using whole blood: three MeDALL cohorts (discovery study), and EVA-PR cohort (replication study). We compared participants with seven phenotypes: A, D, R, A+D, A+R, D+R, or A+D+R to those without asthma, dermatitis or rhinitis (non-allergic participants, Table 1, Figure S1). We analysed samples from the discovery study using Affymetrix Human Transcriptome Array 2.0 (HTA) and detected differentially expressed genes (DEGs). Relative expression of a randomly selected subset of eight genes was assessed by reverse transcription quantitative polymerase chain reaction (RT-qPCR). The replication arm was performed for DEGs in any multimorbidity, rhinitis alone and asthma alone in the EVA-PR cohort,[9] using RNA-Sequencing. We retrieved co-expressed genes with weighted gene co-expression network analysis (WGCNA). Biological interpretation involved gene, disease ontology and pathway analyses, protein-protein interaction network (PPIN), and literature review (Figure 1).

Methods and references are available as online supplement.

Discovery Study.

Setting and population. Whole blood PAXgene tubes were available for 824 participants ; 786 were included after quality control and outlier detection (Table 1): 256 from BAMSE,[10] Sweden (GSE141623, mean age: 16.7 ± 0.4 years), 329 from GINIplus,[11] Germany (15.2 ± 0.3 years), and 201 from INMA,[12] Spain (GSE141631, 4.2 ± 0.3 years). Over 95% of children were of European ancestry (Figure S1). Samples were enriched for allergic cases.

Disease definition. Definition of current asthma, dermatitis or rhinitis was agreed by a panel of experts and used in all MeDALL studies.[1] Allergic diseases were defined by questionnaire-based parent- or self-reported doctor diagnosis and related symptoms.

Data production. RNA was extracted, quality-checked (RNA Integrity Number > 6), transformed into labelled cDNA and hybridized to HTA (Affymetrix, UK).

Statistical methods – Data cleaning and differential expression. Expression levels were normalized with the robust multi-array average algorithm,[13] using *oligo* in *R v3.5.3*. Multivariable models were adjusted for covariables (sex, cohort, batch, age) and fit with *limma* to identify DEGs

This article is protected by copyright. All rights reserved

Accepted Article

(NetAffx v36 annotation). Surrogate variable analysis was used to capture cell blood heterogeneity (Figure S2).[14] P-values were corrected for multiple testing (Benjamini-Hochberg false discovery rate, FDR),[15] with threshold < 5%. No threshold was set for fold change (log_2FC). Heatmap clustering was computed with ward method on Euclidean distance. We used *G*Power v3.1.9.3* to assess power (post-hoc mode with $log_2FC=0.2$ as the effect size, achieved power for DEGs with their actual log_2FC). Sensitivity analysis was performed with adjustment for blood cell counts (neutrophils, eosinophils, lymphocytes, monocytes), or inhaled corticosteroids (BAMSE), or grouping cohorts by age. Synergistic effect of multimorbidity on gene expression was assessed from results of linear models (based on phenotype or disease number) including all samples, by computing the coefficient of linearity σ based on the difference in normalized FC between 1 and 2 diseases, and 2 and 3 diseases.

Biological interpretation. Functional enrichment included gene ontology, pathways (KEGG, REACTOME) with g:Profiler and disease ontology (*DOSE*). Treemap visualization was performed with REVIGO. PPIN was retrieved from *BIOGRID v3.5* and visualized with *Cytoscape v3.7.1* (network metrics from *NetworkAnalyzer*).

Validation. Expression levels were assessed with RT-qPCR (LightCycler 480 II, Roche, QuantiFAST SyBR kit, Qiagen) on a random subset of 8 genes among the significant DEGs in multimorbidity, using RNA from 45 participants.

Weighted Gene Co-expression Network Analysis. Modules of co-expressed genes were retrieved on discovery cohort expression values with WGCNA, discarding probesets with low variance, and using soft power thresholds of 3 and module eigengene dissimilarity threshold of 0.1 to 0.9. Genes within the same modules are likely to behave similarly and process within the same biological processes. Modules were correlated to clinical traits: presence of any multimorbidity, asthma multimorbidity, dermatitis multimorbidity, rhinitis multimorbidity, sex, age and cohort. Gene significance to multimorbidity and gene-module membership were calculated.

Replication Study.

Setting and population. RNA sequencing data was available from white blood cells of 447 participants (15.5 years, range 10-20, San Juan, Puerto Rico) from EVA-PR cohort, without

This article is protected by copyright. All rights reserved

missing data on variables and outcomes of interest, including 218 participants with asthma, 275 with rhinitis, 53 with dermatitis and 115 without allergic disease (Table 1).

Disease definition. Asthma: doctor's diagnosis of asthma and wheeze in the previous year. Rhinitis: physician-diagnosed allergic rhinitis and symptoms (at a time when the child did not have a cold or flu) in the previous 12 months. Dermatitis: symptoms and at least one positive allergen-specific IgE.

Data production. RNA-Sequencing was performed using 350 ng of RNA extracted from white blood cells after removing haemoglobin (RNA Integrity Number > 7). Library preparation was done using TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold High Throughput kit to remove cytoplasmic and mitochondrial ribosomal RNA (Illumina Inc., San Diego, CA). Libraries were run on the NextSeq 500 using NextSeq® 500/550 High Output Kit v2, using paired-end reads at 75 cycles and with 80 million reads per sample. Quality control for raw RNA-Seq fastq files was performed using FastQC. Low quality reads and 3' adapters were trimmed with Trim Galore! and Cutadapt. Trimmed reads were aligned to reference human genome (hg19) with STAR. Subsequently, the count of uniquely mapped read fragments was quantified for each sample as proxy for the expression level of each of the 26,335 human genes annotated in the Illumina iGenomes database using RSEM. Samples with low alignment percentage were removed from downstream analyses. Independent filtering of low-expressed genes (mean count < 2) was performed. After preprocessing, 16,880 genes and 447 samples were retained in the final analysis.

Statistical methods – Differential expression. DEGs were analyzed based on raw counts using DESeq2.[16] Multivariable models were adjusted for age, sex, batch, and cell type proportions of eosinophils, neutrophils, monocytes, lymphocytes in white blood cells. Subjects without any allergic disease were used as controls. A zero-mean normal prior was put on the non-intercept coefficients, to ensure that fold changes are independent of the choice of reference level. The Benjamini-Hochberg procedure was applied to adjust for multiple testing.[15]

Genes selected for replication. All the DEGs significant in any multimorbidity versus controls and in rhinitis versus controls were tested in the replication study. Moreover, a selection of genes significant in multimorbidity was tested in asthma alone.

This article is protected by copyright. All rights reserved

Results

Clinical characteristics of discovery and replication population.

In MeDALL, 46% of allergic participants had asthma (61% with multimorbidity), 42% dermatitis (48%), 51% rhinitis (49%). Asthma was more common in BAMSE (63%), and rhinitis in GINIplus (79%). Dermatitis in pre-school INMA participants was predominant (74%). Fifty-five percent of the participants had no allergic disease (BAMSE 39%, GINIplus 67%, INMA 54%) (Table 1). In EVA-PR, 66% of allergic participants had asthma (78% with multimorbidity), 16% dermatitis (87%), 83% rhinitis (65%) (Table 1). Inhaled corticosteroids in the past year were administered to 25.5% of asthmatic participants in BAMSE, 21.2% in GINIplus, 25.0% in INMA, and 25.2% in EVA-PR.

Identification and validation of differentially expressed genes in discovery study.

Overall, in whole blood, FDR was reached for 50 genes comparing participants with either any multimorbidity, or asthma, dermatitis or rhinitis multimorbidity, or rhinitis alone, to controls with no allergic disease (Table 2). Their chromosome localization is shown in Table S2 and Figure S3. Thirty-two genes were not previously reported in allergic diseases, of which 17 genes were linked to immune processes (Table 2, online supplement).

No DEG was associated with asthma alone or dermatitis alone but 27 DEGs were found in rhinitis alone (Figure S4.2), of which 18 were not previously described in relation to rhinitis.

FDR was reached for 13 genes when pooling participants with any multimorbidity (Figure S4.1). A multimorbidity signature of 8 genes was present in all types of multimorbidity for asthma, dermatitis and rhinitis (Table 3, Figure 2A–C): *CLC*, *SLC29A1*, *FRRS1, IL5RA, HRH4, SIGLEC8, EMR4P, IL1RL1*. None of these multimorbidity genes was found in rhinitis alone.

These 8 genes are overexpressed in participants with multimorbidity (Figure S5), with relatively limited fold change magnitude. *CLC* has the highest $\log_2$FC value (0.44–0.48), followed by *IL5RA* (0.28–0.32), *EMR4P* (0.23–0.29), *IL1RL1* (0.20–0.24), *HRH4* (0.20–0.24).

Power calculation and list of significant DEGs are reported in Table S1. Power in any single disease vs no disease was estimated to 79%, reached 62.1% in any multimorbidity, and was

This article is protected by copyright. All rights reserved

comparable between asthma multimorbidity, dermatitis multimorbidity, or rhinitis multimorbidity (47.3%–56.7%), as compared to asthma alone, dermatitis alone, or rhinitis alone (44.3%–53.5%).

We randomly selected 8 genes for validation with RT-qPCR, which had a higher relative gene expression in multimorbidity (Figure S6, Table S3).

Sensitivity analysis with adjustment for blood cell counts in BAMSE revealed, at most, a 13% decrease of the fold change estimate for *CLC* (Table S4.1). Inhaled corticosteroids did not impact the fold change magnitude (Table S4.2).

Replication of multimorbidity results in EVA-PR cohorts.

Twelve of the 13 DEGs in any multimorbidity had expression values available, and 10 were replicated (83%) in white blood cells from EVA-PR. Among the multimorbidity signature of 8 genes, all were confirmed. The genes with the highest fold changes were *SLC29A1*, *IL5RA*, *CLC*, *FRRS1*, *EMR4P* and *HRH4* ($\log_2$FC=0.12–0.17). Moreover, among the genes tested, none was found significant in participants with asthma alone. One gene (*CEBPE*) and one locus (*LOC101927780*) in rhinitis alone were replicated. Exhaustive results are in Table S5 and Figure S7.

Synergistic effects of multimorbidity on the significant genes.

Using a linear model including all samples (Table S6), all genes had synergistic effects of multimorbidity (σ = 1.5 – 58.4), except LOC101929979 (σ = 1.1). Figure 2D shows the figures for the 12 genes present in at least two of asthma, dermatitis or rhinitis multimorbidity. A linear model based on the number of diseases produced the same results (Table S6.3).

Biological interpretation of differentially expressed genes: Functional enrichment analysis.

We interrogated ontology databases with the DEGs in any multimorbidity (Figure S8, Table S7). We identified biological processes and molecular functions such as ‘‘*transmembrane signaling receptor activity*’’, ‘‘*immune response*’’, ‘‘*cell communication*’’, or ‘‘*interleukin-5 production*’’ (FDR p=1.2e-03 to p=3.0e-03), as well as REACTOME pathways such as ‘‘*Signal Transduction*’’, ‘‘*Signaling by Interleukins*’’ and ‘‘*Interleukin-33 signaling*’’ (FDR p=1.2e-03 to p=1.5e-02). Leukocyte activation was involved (FDR p=3.5e-03). Similar ontology was identified

This article is protected by copyright. All rights reserved

with the DEGs found in adolescents with any multimorbidity (Table S7.7). Disease ontology retrieved clusters of respiratory and skin diseases (Figure S9, Table S8).

Biological interpretation of differentially expressed genes: Protein-protein interaction network.

We retrieved 37 proteins interacting with the proteins encoded by the 8 DEGs shared in multimorbidity. *CLC*, *SLC29A1*, *IL5RA*, *SIGLEC8*, and *IL1RL1*, but not *HRH4*, *EMR4P*, and *FRRS1*, were retrieved in BIOGRID (Figure 2F). The PPIN had an average clustering coefficient of 20.7 %, and centralization of 49.8 %. Nine of the 37 interactors were involved in IL5/JAK/STAT and IL33/ST2/IRAK/TRAF pathways, and 24 in allergy or inflammatory processes (online supplement). IL5RA and IL1RL1 (ST2) belong to IL5/JAK/STAT and IL33/ST2/IRAK/TRAF pathways respectively. Pathway analysis on the interactors (Table S9) retrieved 18 members of the REACTOME pathway "Immune System" (R-HSA-168256, p=6.6e-14), and 13 members of "Signaling by Interleukins" (R-HSA-449147, p=2.1e-15).

Co-expression modules.

A module of 60 co-expressed genes ("paleturquoise") was detected with stringent clustering height (h=0.9, Figure S10.1). Lower height cutoff did not split this module (h=0.6, Figure S10.2). It includes the 8 genes shared in multimorbidity. *CLC* scored the highest gene significance for multimorbidity (GS.multimorbidity=0.16, p=4.8e-06), and *IL5RA* scored the highest module membership (MM.paleturquoise=0.94, p<1.8e-311). The module was positively correlated to any multimorbidity (r2=0.14, p=5e-05), asthma multimorbidity, dermatitis multimorbidity and rhinitis multimorbidity (r2=0.13, p-value from 3e-04 to 2e-04). Gene ontology for the 60 co-expressed genes in "paleturquoise" (Table S10) retrieved an immune system and signal transduction signature (FDR from p=4.5e-11). The implication of leukocyte (GO:0045321) was detected with 10 genes (p=4e-06). Disease ontology retrieved clusters of respiratory and skin diseases.

This article is protected by copyright. All rights reserved

Accepted Article

Discussion

For the first time, this study shows a notable difference between multimorbid and single allergic diseases. With a multimorbid approach, we identified important genomic features of allergic diseases that would have been missed using a single disease approach. No gene was associated to asthma or dermatitis alone but 27 were found in rhinitis alone in MeDALL. Thirteen DEGs were found in participants with any multimorbidity as compared to controls. Eight genes were common to asthma, dermatitis and rhinitis multimorbidity. Our key results were highly replicated in the EVA-PR cohort, with 10 of the 12 tested multimorbidity genes being significant. RT-qPCR in MeDALL validated overexpression of selected genes in multimorbidity, which were found associated to immune and signalling pathway. Unsupervised analysis retrieved our candidate genes co-expressed in a same module, validating their potential involvement in related biological processes. Moreover, fold changes of our candidate genes increased along with the number of co-occurrent diseases. This synergistic effect launches new insights on the mechanisms of multimorbidity for allergic diseases. Overall, eosinophil- and T2-associated genes were identified, confirming the MeDALL hypothesis that type 2 immunity has an important role in the origin of multimorbidity.[5]

Strengths and limitations.

We used HTA in the discovery study because of its genome-wide coverage and high number of probesets. At the time of study, the reduced costs compared to RNA sequencing allowed the analysis of a greater number of samples, thus an increased power to dig into gene expression specificities of multimorbidity. The HTA results were highly replicated and comparable with the RNA-sequencing results in EVA-PR.

Since allergic diseases are systemic and we assessed multimorbidity, we did not use specific tissues from the skin, nose and bronchi but whole blood. We detected genes associated with eosinophils and immune response, which partly overlap with the findings from our recent MeDALL epigenome-wide study on asthma,[17] where *CLC*, *IL5RA*, *SLC29A1* and *SIGLEC8* still reach significance after eosinophil correction. Since blood samples were not drawn after allergen challenge, which may increase the signal, we expected to find low effect sizes (small $\log_2$FC values). The limited magnitude of fold change values, observed in the literature for asthma[18], does

This article is protected by copyright. All rights reserved

not prevent biological significance.[19] Moreover, we validated the biological relevance of our genes with RT-qPCR, ontology and network-based analyses. We retrieved 13 DEGs comparing 123 participants with any multimorbidity to 431 participants without allergic disease. The pooling approach of MeDALL allowed us to increase FDR significance.

Power was anticipated with a low effect size (d=0.2). In the present study, we retrieved more DEGs in 123 participants with any multimorbidity (13 DEGs with power=62%) than in 232 participants with single disease (no DEGs, power=79%, Table S1.11). Each discovery cohort alone had an insufficient power, validating the pooling approach of MeDALL.

Our discovery study included two cohorts of adolescents and one of children, thus limiting our capacity to assess the role of age.[20] Nevertheless, FDR was reached for 9 genes grouping adolescents with any multimorbidity, including *CLC*, *FRRS1*, *HRH4*, *IL5RA*, *EMR4P* and *IL1RL1*, shared between asthma, dermatitis and rhinitis multimorbidity (Table S4.3).

There were some differences between the European and Puerto Rican cohorts, for example in age distribution and disease prevalences, that may be explained by the different study designs, populations, and/or lifestyle factors. A complete harmonization of disease definitions was not available between discovery and replication cohorts. The resulting dissimilarities (e.g. inclusion of positive IgE sensitization, or symptoms-based definition versus self-reported doctor diagnosis) could have led to bias in over-selection of IgE mechanism associated genes, or differences in disease prevalence (over selection of more severe cases), but the reproducibility of the results showed the generalisability of the findings.

Replication analyses of multimorbidity.

The results for multimorbidity were consistent in particular since the replication cohort was from a different environment (Puerto Rico), and RNA-sequencing on white blood cells was used.[9,21] For example, the $\log_2$FC value for *CLC* in discovery is 0.44 in any multimorbidity (expression increased by 36% in allergic participants compared to controls), and 0.14 in replication (10% increase). The inter-platform difference is thus 26% for *CLC*, and less than 10% in the other multimorbidity genes. Thus, although some differences in effect sizes were noted between the discovery and replication cohorts, our analyses identify robust multimorbidity genes such as *CLC*.

This article is protected by copyright. All rights reserved

In children and adolescents, asthma or dermatitis alone were not associated with any DEG whereas rhinitis alone was associated with 27 DEGs, in line with the clinical knowledge that 85% of asthma patients have also rhinitis, whereas only a third of rhinitis patients have asthma.[2] This study suggests that mechanistic studies in asthma could also benefit from considering the presence of rhinitis as multimorbidity. However, despite a distinct signature of rhinitis alone in the discovery cohort, the limited replication suggests that further studies need to be done.

Candidate genes function.

Of the eight multimorbidity genes, five were eosinophil-specific genes. *CLC* (Galectin-10),[22] which is involved in regulation of T-cell proliferation, and predicts oral corticosteroid response in asthma (see supplementary bibliography), featured the highest log_2FC and was significant in any type of multimorbidity. *IL5RA* has been involved in dermatitis[23] and asthma,[24] *IL1RL1* in asthma,[25] and was found associated with multimorbidity in our computational study of the protein interaction network.[5] *SIGLEC8* is a susceptibility locus for asthma.[26] Expression levels of *SIGLEC8* in our RT-qPCR validation were comparable in asthma multimorbidity and asthma alone, highlighting the importance of *SIGLEC8* in asthma. *EMR4P* may be involved in leukocyte adhesion and migration (supplementary bibliography). Among the three other genes, *HRH4* was related to allergic diseases and immune response (Table 3).[27] *FRRS1* is a member of cytochrome b561, and *SLC29A1* mediates the cellular uptake of nucleosides from the surrounding medium (supplementary bibliography). *EMR4P*, *FRRS1* and *SLC29A1* were not previously described in allergic diseases.

Moreover, among the genes found with any multimorbidity, or multimorbidity for asthma, dermatitis or rhinitis, eight were linked to allergic inflammation and immune response (Table 2 and online supplement). In addition to genes cited above, *SEMA7A* was eosinophil-specific and may play a role in airway remodelling of asthma.[28] *CLC*, *IL5RA*, *SIGLEC8*, *SEMA7A* correlated strongly to differentially methylated CpG sites in asthma.[17] In addition to *HRH4*, *ALOX15*[29] and *CYSLTR2*[30] were related to allergic diseases and immune response.

Among genes found in rhinitis alone, *SENP1* is important for the epithelial-mesenchymal trophic unit, *AK2*, *CEBPE, GPR65*, are involved in white blood cells differentiation, and *BCAP29*

This article is protected by copyright. All rights reserved

influences the traffic of MHC class I molecules (online supplement). *TNFSF14* was known in allergy with no known role in inflammation.[31]

Thirty-two genes were potential novel targets, notably *EMR4P*, *FRRS1* and *SLC29A1*. Nine genes had known roles in inflammatory processes, and nine uncharacterized genes corresponded to long intergenic non-protein coding RNA, or pseudogenes.

These results enforce the role of eosinophil activation in childhood allergic diseases through DNA methylation and gene expression processes. Without exhaustive information on blood cell composition in MeDALL, we used surrogate variable analysis to detect and adjust for unknown surrogate variables, which demonstrated the best performance for cell-type mixture.[14] Moreover, we validated our approach with a sensitivity analysis in one discovery cohort (Table S4), and found no major overall influence on the association results. All analyses in EVA-PR were adjusted for white blood cell type. Thus, our overall findings should not be confounded by higher eosinophil counts in allergic patients *per se*. Inhaled corticosteroids and ß2 agonists did not impact the observed effect (Table S4.2).

Multimorbidity related pathways and unsupervised analysis.

PPIN retrieved IL5/JAK/STAT pathway, well-known in IL-5 inflammation, and IL33/ST2/IRAK/TRAF pathway, involved in IL-33 and IL1RL1 inflammation (online supplement). *IL1RL1* and *IL5RA* were shared between in multimorbidity. Fifteen other interactors were related to allergy, immune or inflammatory processes (online supplement).

We validated the signature with WGCNA, and retrieved, in an unsupervised manner, a module of co-expressed genes with a high clustering coefficient, which survived a stringent height cut-off, indicating biological relationship and strong expression similarities with other genes of the module. All 8 genes shared in multimorbidity were co-expressed in this module. The module correlation to multimorbidity was significant. Fourteen co-expressed genes were involved in allergy and immune processes (online supplement).

Conclusion

Among the 50 genes differentially expressed in participants with any allergic disease, 8 were common to multimorbidity, and 4 associated with eosinophil expression. All had synergistic effect

This article is protected by copyright. All rights reserved

associated to multimorbidity. This genetic signature of multimorbidity can provide new diagnostic and therapeutic opportunities. Thirty-two genes previously unknown in allergic diseases were discovered. Our results support the need to assess the presence of multimorbidity in clinical and genetic investigations of allergic diseases in children and adolescents.

Accepted Article

This article is protected by copyright. All rights reserved

Accepted Article

References

1. Pinart M, Benet M, Annesi-Maesano I, et al. Comorbidity of eczema, rhinitis, and asthma in IgE-sensitised and non-IgE-sensitised children in MeDALL: a population-based cohort study. *Lancet Respir Med* 2014; **2**: 131–40
2. Bousquet J, Khaltaev N, Cruz AA, et al. Allergic Rhinitis and its Impact on Asthma (ARIA) 2008 update (in collaboration with the World Health Organization, GA(2)LEN and AllerGen). *Allergy* 2008; **63** Suppl 86:8–160
3. Bieber T. Atopic dermatitis. *N Engl J Med* 2008; **358**: 1483–94
4. Anto JM, Bousquet J, Akdis M, et al. Mechanisms of the Development of Allergy (MeDALL): Introducing novel concepts in allergy phenotypes. *J Allergy Clin Immunol* 2017; **139**: 388–99
5. Bousquet J, Anto JM, Wickman M, et al. Are allergic multimorbidities and IgE polysensitization associated with the persistence or re-occurrence of foetal type 2 signalling? The MeDALL hypothesis. *Allergy* 2015; **70**: 1062–78
6. Aguilar D, Pinart M, Koppelman GH, et al. Computational analysis of multimorbidity between asthma, eczema and rhinitis. *PLoS One* 2017; **12**: e0179125
7. Demenais F, Margaritte-Jeannin P, Barnes KC et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat Genet*. 2018; **50**: 42-53.
8. Ferreira MA, Vonk JM, Baurecht H, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat Genet* 2017; **49**: 1752–57
9. Jiang Y, Gruzieva O, Wang T et al. Transcriptomics of atopy and atopic asthma in white blood cells from children and adolescents. *Eur Respir J.* 2019. pii: 1900102. doi: 10.1183/13993003.00102-2019
10. Gref A, Merid SK, Gruzieva O, et al. Genome-Wide Interaction Analysis of Air Pollution Exposure and Childhood Asthma with Functional Follow-up. *Am J Respir Crit Care Med* 2017; **195**: 1373–83
11. Von Berg A, Kramer U, Link E, et al. Impact of early feeding on childhood eczema: development after nutritional intervention compared with the natural course–the GINIplus study up to the age of 6 years. *Clin Exp Allergy* 2010; **40**: 627–36

This article is protected by copyright. All rights reserved

Accepted Article

12. Guxens M, Ballester F, Espada M, et al. Cohort Profile: the INMA-INfancia y Medio Ambiente-(Environment and Childhood) Project. *Int J Epidemiol* 2012; **41**: 930–40
13. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; **4**: 249–64
14. McGregor K, Bernatsky S, Colmegna I, et al. An evaluation of methods correcting for cell-type heterogeneity in DNA methylation studies. *Genome Biol* 2016; **17**: 84
15. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B 1995; 57: 289–300
16. Love, M.I., W. Huber, and S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology, 2014. 15(12): p. 550.
17. Xu CJ, Söderhäll C, Bustamante M, et al. DNA methylation in childhood asthma: an epigenome-wide meta-analysis. Lancet Respir Med 2018; 6: 379–88
18. Morrow JD, Qiu W, Chhabra D, et al. Identifying a gene expression signature of frequent COPD exacerbations in peripheral blood using network methods. *BMC Med Genomics* 2015; **8**: 1
19. St Laurent G, Shtokalo D, Tackett MR, et al. On the importance of small changes in RNA expression. *Methods*. 2013; **63**(1):18-24
20. Belgrave CD, Granell R, Simpson A, et al. Developmental profiles of eczema, wheeze, and rhinitis: two population-based birth cohort studies. *PLoS Med* 2014; **11**: e1001748
21. Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLoS One* 2014; **9**(1): e78644
22. Su J. A Brief History of Charcot-Leyden Crystal Protein/Galectin-10 Research. *Molecules* 2018; **23**: 2931
23. Semic-Jusufagic A, Gevaert P, Bachert C, Murray C, Simpson A, Custovic A. Increased serum-soluble interleukin-5 receptor alpha level precedes the development of eczema in children. *Pediatr Allergy Immunol* 2010; **21**: 1052–58
24. Yasruel Z, Humbert M, Kotsimbos TC, et al. Membrane-bound and soluble alpha IL-5 receptor mRNA in the bronchial mucosa of atopic and nonatopic asthmatics. *Am J Respir Crit Care Med* 1997; **155**: 1413–18
25. Dijk FN, Xu C, Melén E et al. Genetic regulation of IL1RL1 methylation and IL1RL1-a protein levels in asthma. *Eur Respir J* 2018; **51**: 1701377

This article is protected by copyright. All rights reserved

Accepted Article

26. Gao PS, Shimizu K, Grant AV, et al. Polymorphisms in the sialic acid-binding immunoglobulin-like lectin-8 (Siglec-8) gene are associated with susceptibility to asthma. *Eur J Hum Genet* 2010; **18**: 713–19
27. Walter M, Kottke T, Stark H. The histamine H4 receptor: Targeting inflammatory disorders. *Eur J Pharmacol* 2011; **668**: 1–5
28. Mizutani N, Nabe T, Yoshino S. Semaphorin 7A plays a critical role in IgE-mediated airway inflammation in mice. *Eur J Pharmacol* 2015; **764**: 149–56
29. Wen Y, Gu J, Chakrabarti SK, et al. The role of 12/15-lipoxygenase in the expression of interleukin-6 and tumor necrosis factor-alpha in macrophages. *Endocrinol* 2007; **148**: 1313–22
30. Thompson MD, Capra V, Clunes MT et al. Cysteinyl Leukotrienes Pathway Genes, Atopic Asthma and Drug Response: From Population Isolates to Large Genome-Wide Association Studies. *Front Pharmacol* 2016 Dec 1; **7**: 299
31. Shi F, Xiong Y, Zhang Y, et al. The Role of TNF Family Molecules Light in Cellular Interaction Between Airway Smooth Muscle Cells and T Cells During Chronic Allergic Inflammation. *Inflammation* 2018; **41**:1021–31

This article is protected by copyright. All rights reserved

Accepted Article

Tables

Table 1: Characteristics of the MeDALL discovery population.

	BAMSE	GINIplus	INMA	Total (Discovery)	EVA-PR (Replication)
included in the analysis	**256**	**329**	**201**	**786**	**447**
male (n,%†)	*143 (55.9)*	*160 (48.6)*	*99 (49.2)*	*402 (51.1)*	*177 (53.3)*
female (n,%†)	*113 (44.1)*	*169 (51.4)*	*102 (50.8)*	*384 (48.9)*	*155 (46.7)*
controls (n,%†)	*100 (39.1)*	*222 (67.5)*	*109 (54.2)*	*431 (54.8)*	*115 (25.7)*
A alone (n)	*32*	*15*	*18*	*65*	*47*
D alone (n)	*22*	*6*	*48*	*76*	*7*
R alone (n)	*23*	*64*	*4*	*91*	*97*
A + D (n)	*17*	*1*	*16*	*34*	*3*
A + R (n)	*33*	*15*	*2*	*50*	*135*
D + R (n)	*13*	*4*	*4*	*21*	*10*
A + D + R (n)	*16*	*2*	*0*	*18*	*33*
any multimorbidity (n,%‡)	*79 (50.6)*	*22 (20.6)*	*22 (23.9)*	*123 (34.6)*	*181 (54.5)*
age (mean years ± sd)	*16.7 ± 0.4*	15.0 ± 0.2	4.2 ± 0.3	12.8 ± 5.1	*15.4 (10-20)*

Characteristics of discovery and replication population. A: asthma, D: dermatitis, R: rhinitis ; †: % of total ; ‡: % of cases.

This article is protected by copyright. All rights reserved

Table 2: Differentially expressed genes in MeDALL and replication in EVA-PR

A. Multimorbidity

			MeDALL			EVA-PR		
			any multimorbidity (62.1 %)			any multimorbidity (51.1 %)		
gene identifier	known in allergy	known in immune process	log2FC	p-value	FDR	log2FC	p-value	FDR
CLC	yes	yes	**0.44**	**1.3E-08**	**8.8E-04**	**0.14**	**1.4E-04**	**4.4E-04**
SLC29A1	no	no	**0.09**	**3.1E-07**	**7.1E-03**	**0.17**	**2.6E-06**	**1.9E-05**
HRH4	yes	yes	**0.20**	**4.7E-07**	**7.1E-03**	**0.12**	**1.1E-03**	**2.7E-03**
FRRS1	no	no	**0.13**	**5.1E-07**	**7.1E-03**	**0.14**	**7.2E-05**	**2.6E-04**
LOC101929979 †	n/a	n/a	**0.32**	**5.7E-07**	**7.1E-03**	null‡	null‡	null‡
IL5RA	yes	yes	**0.28**	**6.3E-07**	**7.1E-03**	**0.15**	**4.2E-05**	**1.8E-04**
CACNA1D	no	no	**0.04**	**3.3E-06**	**2.5E-02**	**0.07**	**3.9E-02**	>5E-02
IL1RL1	yes	yes	**0.20**	**3.3E-06**	**2.5E-02**	**0.09**	**1.2E-02**	**2.2E-02**
TTC7B	no	no	**0.08**	**4.6E-06**	**2.8E-02**	0.03	3.3E-01	>5E-02
CYSLTR2	yes	yes	**0.16**	**4.6E-06**	**2.8E-02**	**0.07**	**3.1E-02**	>5E-02
EMR4P	no	yes	**0.23**	**5.7E-06**	**3.0E-02**	**0.13**	**5.0E-04**	**1.4E-03**
PIK3R6	no	yes	**0.10**	**7.4E-06**	**3.6E-02**	**0.09**	**5.4E-03**	**1.1E-02**
ZSCAN29	no	no	**0.06**	**9.6E-06**	**4.0E-02**	0.03	>5E-02	>5E-02

B. Rhinitis alone

			MeDALL			EVA-PR		
			rhinitis alone (53.4 %)			rhinitis alone (42.1 %)		
gene identifier	known in allergy	known in immune process	log2FC	p-value	FDR	log2FC	p-value	FDR
SENP1	no	no	**0.06**	**7.4E-07**	**1.4E-03**	0.00	>5E-02	>5E-02
BCAP29	no	yes	**0.07**	**3.3E-06**	**5.0E-03**	-0.01	>5E-02	>5E-02
CEBPE	no	yes	**0.11**	**3.6E-06**	**5.3E-03**	**0.11**	**1.6E-02**	>5E-02
FAM66C	no	yes	**-0.03**	**6.5E-06**	**8.6E-03**	0.07	>5E-02	>5E-02
TNFSF14	yes	no	**0.15**	**7.9E-06**	**1.0E-02**	-0.08	>5E-02	>5E-02
RNA5SP378 †	n/a	n/a	**-0.08**	**1.3E-05**	**1.6E-02**	null‡	null‡	null‡
MAGEA9B	no	no	**-0.05**	**1.5E-05**	**1.7E-02**	null‡	null‡	null‡

This article is protected by copyright. All rights reserved

RP1-149A16.12 †	n/a	n/a	**-0.07**	**1.6E-05**	**1.8E-02**	null‡	null‡	null‡
RNA5SP335 †	n/a	n/a	**0.10**	**1.7E-05**	**1.8E-02**	null‡	null‡	null‡
LOC101928812 †	n/a	n/a	**0.09**	**2.0E-05**	**1.9E-02**	-0.05	>5E-02	>5E-02
LEPREL4	no	no	**-0.03**	**2.5E-05**	**2.2E-02**	0.01	>5E-02	>5E-02
CLINT1	no	no	**0.05**	**2.7E-05**	**2.3E-02**	0.00	>5E-02	>5E-02
NDUFB5	no	no	**0.05**	**2.7E-05**	**2.3E-02**	0.00	>5E-02	>5E-02
NPIPB5	no	no	**-0.08**	**3.2E-05**	**2.6E-02**	0.05	>5E-02	>5E-02
LA16c-395F10.1 †	n/a	n/a	**0.09**	**3.4E-05**	**2.7E-02**	null‡	null‡	null‡
AK2	no	yes	**0.04**	**3.9E-05**	**3.1E-02**	0.01	>5E-02	>5E-02
GLT6D1	no	no	**-0.08**	**3.9E-05**	**3.1E-02**	null‡	null‡	null‡
MIR539	no	no	**-0.11**	**4.2E-05**	**3.2E-02**	null‡	null‡	null‡
LOC101927780 †	n/a	n/a	**0.08**	**4.2E-05**	**3.2E-02**	**0.18**	**1.3E-04**	**2.3E-03**
RNA5SP440 †	n/a	n/a	**-0.09**	**4.4E-05**	**3.3E-02**	null‡	null‡	null‡
DUSP16	no	no	**0.10**	**5.0E-05**	**3.5E-02**	-0.01	>5E-02	>5E-02
GPR65	no	yes	**0.09**	**5.1E-05**	**3.5E-02**	0.03	>5E-02	>5E-02
RNF149	no	no	**-0.05**	**5.2E-05**	**3.5E-02**	-0.03	>5E-02	>5E-02
SLC25A40	no	no	**0.06**	**5.2E-05**	**3.5E-02**	-0.02	>5E-02	>5E-02
EXOC5	no	no	**0.04**	**5.8E-05**	**3.8E-02**	0.01	>5E-02	>5E-02
HMGA1P4	no	no	**0.10**	**7.3E-05**	**4.4E-02**	null‡	null‡	null‡
LOC101929526 †	n/a	n/a	**0.09**	**8.2E-05**	**4.8E-02**	null‡	null‡	null‡

Differentially expressed genes with log_2FC and FDR p-values in **A.** any multimorbidity (sorted by increasing FDR, or **B.** rhinitis alone for MeDALL and EVA-PR participants. Power of each group in parenthesis (%). P-values < 0.05 are in bold, †: long intergenic non-protein coding RNA, pseudogenes or ribosomal pseudogenes, their role in allergy or immune processes could not be determined (*n/a*: not applicable) ; *null‡*: not tested due to QC and RNA-sequencing issues. Genes that specific for asthma-, dermatitis-, and rhinitis- multimorbidity are not shown (*ACSM3, SIGLEC8, ALOX15, ALOX15P1, BACE2-IT1, LGALS12, PNPLA6, SEMA7A, SMPD3* and *SORD*).

This article is protected by copyright. All rights reserved

Table 3: DEGs shared between multimorbidity for asthma, dermatitis or rhinitis.

Gene identifier	Chromosome	Name	Description	Previous identification†
CLC	19q13.2	Charcot-Leyden crystal galectin	Eosinophil	asthma, rhinitis
SLC29A1	6p21.1	Solute carrier family 29 member 1	Nucleoside transporter	no
FRRS1	1p21.2	Ferric chelate reductase 1	Cytochrome b561	no
IL5RA	3p26.2	Interleukin 5 receptor subunit alpha	Eosinophil	asthma, dermatitis, rhinitis
HRH4	18q11.2	Histamine receptor H4	Histamine	asthma, dermatitis, rhinitis
SIGLEC8	19q13.41	Sialic acid binding Ig like lectin 8	Eosinophil	asthma, rhinitis
EMR4P	19p13.2	Adhesion G protein-coupled receptor E4	Leukocyte adhesion	no
IL1RL1	2q12.1	Interleukin 1 receptor like 1	Eosinophil	asthma, rhinitis

Name, description and previous identification of genes shared between asthma, dermatitis and rhinitis multimorbidity. †: see online supplement.

Accepted Article

This article is protected by copyright. All rights reserved

Accepted Article

Figures

Figure 1: Data analysis workflow

Gene expression levels of allergic participants were compared to non-allergic controls. Differentially expressed genes (DEGs) with FDR < 5% were detected. Expression levels of an arbitrary subset of 8 genes were validated with RT-qPCR. Subset of genes was replicated in EVA-PR cohort. Co-expression was assessed with Weighted Gene Co-expression Network Analysis (WGCNA). Biological role of genes was interpreted with gene and disease ontology functional enrichment, protein-protein network analysis (PPIN) and review of literature. *: Allergy was defined as the presence of at least one disease among asthma, dermatitis, or rhinitis ; allergic participants were grouped based on allergic disease co-occurence: 1 disease, or 2 or more diseases (multimorbidity), and/or based on the nature of the disease: asthma, dermatitis or rhinitis, and compared to non-allergic participants.

Figure 2: Shared genes in multimorbidity

A. Number of DEGs shared between any multimorbidity, asthma, dermatitis or rhinitis multimorbidity. The text box lists the 7 genes shared by all groups. **B.** Heatmap of $-\log_{10}$(FDR p-values) for the 50 genes and each group, with blue-white color scale on top right panel. Clustering on genes and groups was done with ward method on euclidean distance. *EMR4P* and *IL1RL1* are targeted by two probesets. multi.: multimorbidity. **C.** Heatmap of $-\log_{10}$(FDR p-values) for each gene that is shared between at least two of asthma multimorbidity, dermatitis multimorbidity or rhinitis multimorbidity, with blue-white color scale on top right panel. The 8 genes shared between asthma, dermatitis, or rhinitis multimorbidity are in bold. Clustering on genes and groups was done with ward method on euclidean distance. **D.** Normalized fold change variation across increasing number of co-occurrent disease, for the 12 genes in **C.**, using a linear multivariate model including all allergic phenotypes. **E.** KEGG and REACTOME pathways significant (FDR < 5%) with DEGs in asthma, dermatitis or rhinitis multimorbidity versus controls (Table S7.1). Terms are sorted by decreasing significance. **F.** BIOGRID PPIN from DEGs shared between asthma, dermatitis, and rhinitis multimorbidity. Octogons: proteins encoded by our candidate genes ; rounded rectangle: interactors, alphabetically-sorted in grid layout on the bottom right. All types of interactions were considered.

This article is protected by copyright. All rights reserved

Accepted Article

Statement of contribution

Nathanaël Lemonnier performed the transcriptomic study, led the data production, led data pre-processing and data quality checking, contributed to data analysis plan, made the analyses, contributed to the replication plan and wrote the paper.

Erik Melén is the leader of the BAMSE cohort, contributed to sample selection, made significant comments on data analysis design and results, and led the replication plan.

Yale Jiang contributed to the EVA-PR cohort and made significant contribution to the replication analysis.

Stéphane Joly and Camille Ménard performed the data production, quality checking, RT-qPCR validation.

Daniel Aguilar, Judith Garcia-Aymerich and Stefano Guerra discussed the analysis and the paper.

Edna Acosta-Perez, Nadia Boutaoui, Glorisa Canino and Erick Forno contributed to the EVA-PR cohort and the replication study.

Anna Bergström, Olena Gruzieva, Inger Kull and Magnus Wickman contributed to the BAMSE cohort.

Mariona Bustamante, Juan Ramon González, Jesús Ibarluzea Maurolagoitia and Loreto Santa-Marina Rodriguez contributed to the INMA cohort.

Joachim Heinrich and Elisabeth Thiering contributed to the GINIplus cohort.

Cezmi Akdis, Mübeccel Akdi, Thomas Keil, Gerard H. Koppelman, Valérie Siroux and Cheng-Jian Xu participated to the discussion of the new data analysis plan and made significant comments on the analysis

Wei Chen is a principal investigator of the EVA-PR cohort and the replication analysis.

Pierre Hainaut made significant comments on multimorbidity and chronic diseases and the overall interpretation of the results.

Marie Standl is the leader of the GINIplus cohort.

Jordi Sunyer is the leader of the INMA cohort.

Juan C. Celedón is the leader of the EVA-PR cohort and led the replication analysis.

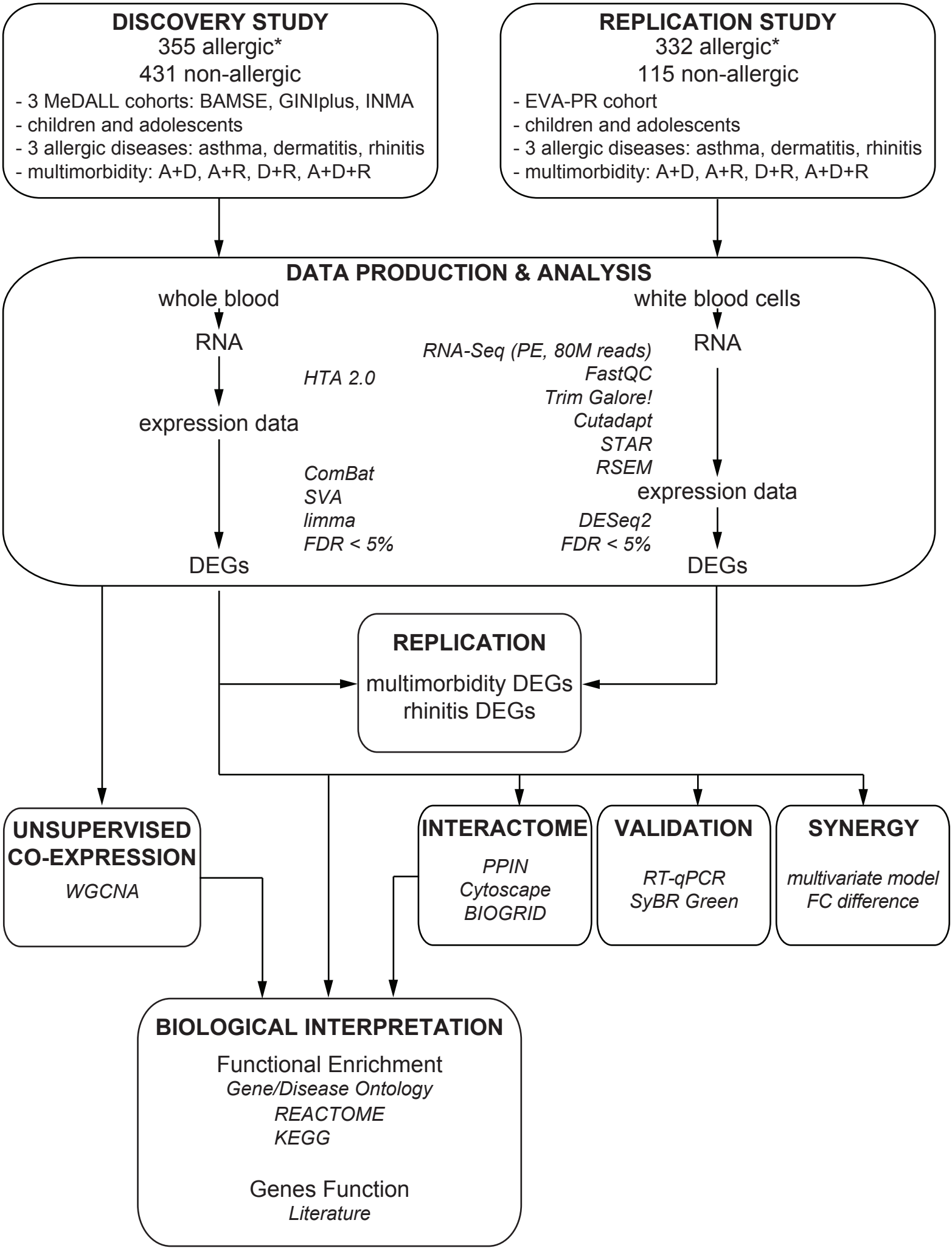
This article is protected by copyright. All rights reserved

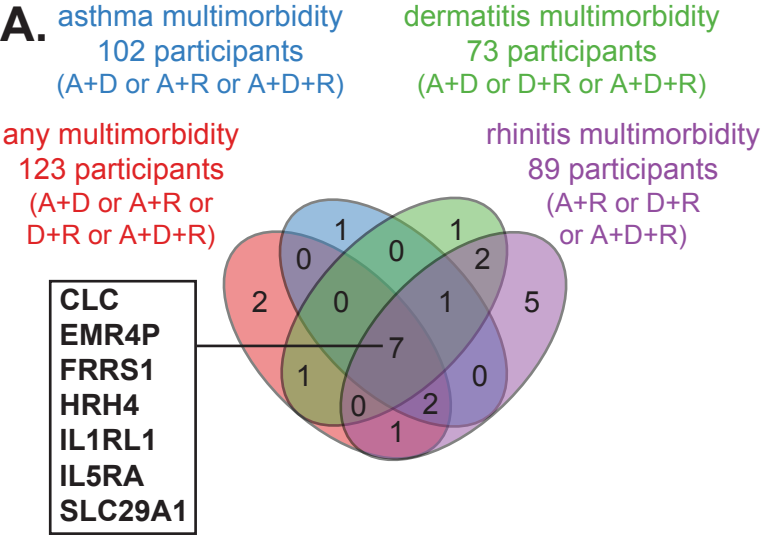
Josep M Antó was co-coordinator of MeDALL proposed the new analysis plan on multimorbidity with Jean Bousquet, led the analysis methodology with JB and NL and the overall interpretation of the results.

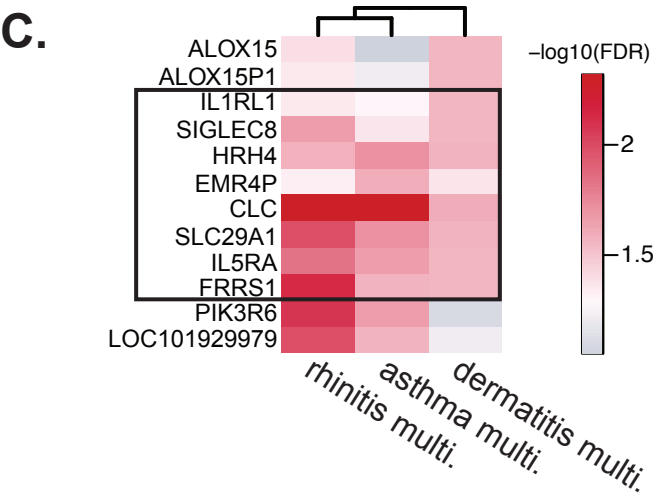
Jean Bousquet was coordinator of MeDALL, proposed the new analysis plan on multimorbidity with Josep M Antó, led the multimorbidity analyses with JMA and NL and wrote the paper with NL.

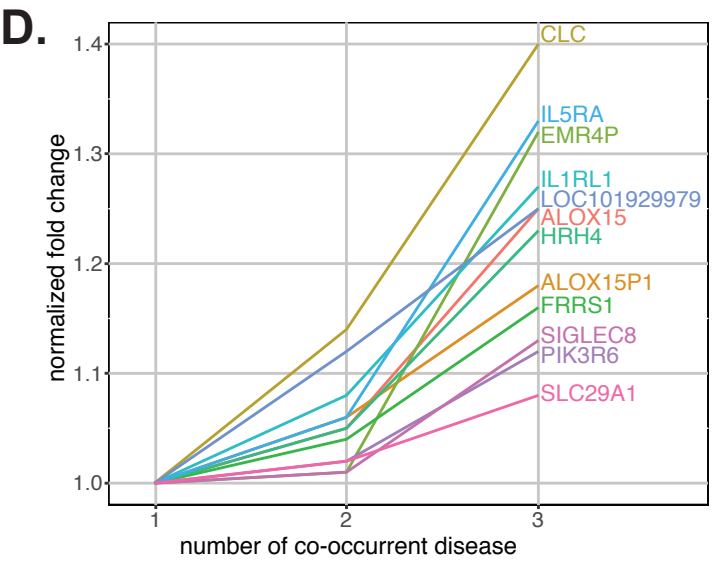
Accepted Article

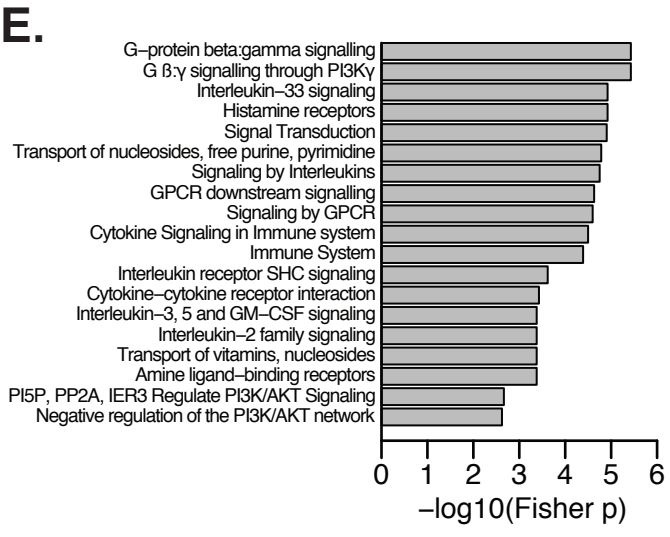
This article is protected by copyright. All rights reserved

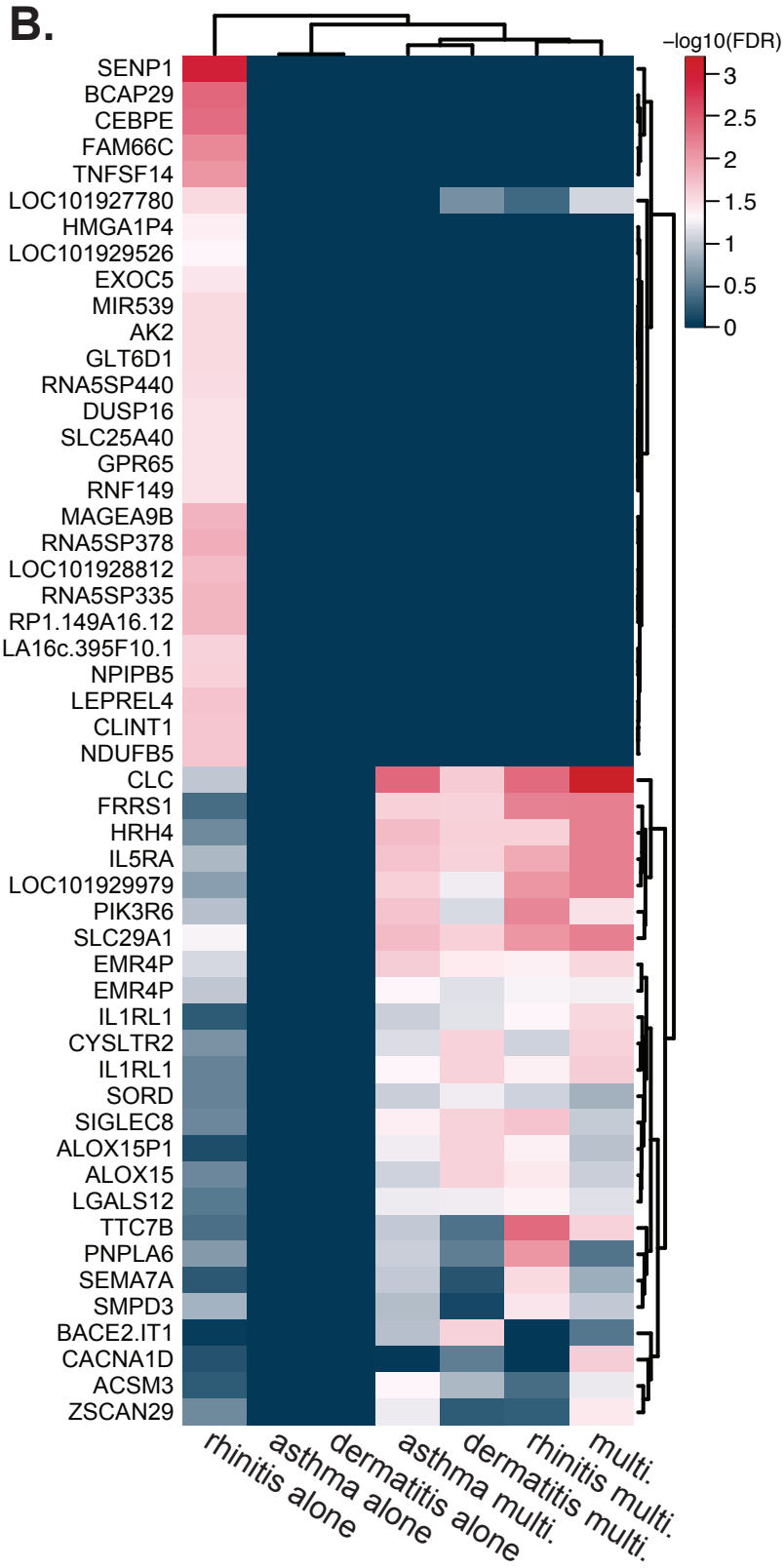

DISCOVERY STUDY
355 allergic*
431 non-allergic
- 3 MeDALL cohorts: BAMSE, GINIplus, INMA
- children and adolescents
- 3 allergic diseases: asthma, dermatitis, rhinitis
- multimorbidity: A+D, A+R, D+R, A+D+R
REPLICATION STUDY
332 allergic*
115 non-allergic
- EVA-PR cohort
- children and adolescents
- 3 allergic diseases: asthma, dermatitis, rhinitis
- multimorbidity: A+D, A+R, D+R, A+D+R
DATA PRODUCTION & ANALYSIS
whole blood
RNA
HTA 2.0
expression data
ComBat
SVA
limma
FDR < 5%
DEGs
white blood cells
RNA
RNA-Seq (PE, 80M reads)
FastQC
Trim Galore!
Cutadapt
STAR
RSEM
expression data
DESeq2
FDR < 5%
DEGs
REPLICATION
multimorbidity DEGs
rhinitis DEGs
UNSUPERVISED CO-EXPRESSION
WGCNA
INTERACTOME
PPIN
Cytoscape
BIOGRID
VALIDATION
RT-qPCR
SyBR Green
SYNERGY
multivariate model
FC difference
BIOLOGICAL INTERPRETATION
Functional Enrichment
Gene/Disease Ontology
REACTOME
KEGG
Genes Function
Literature

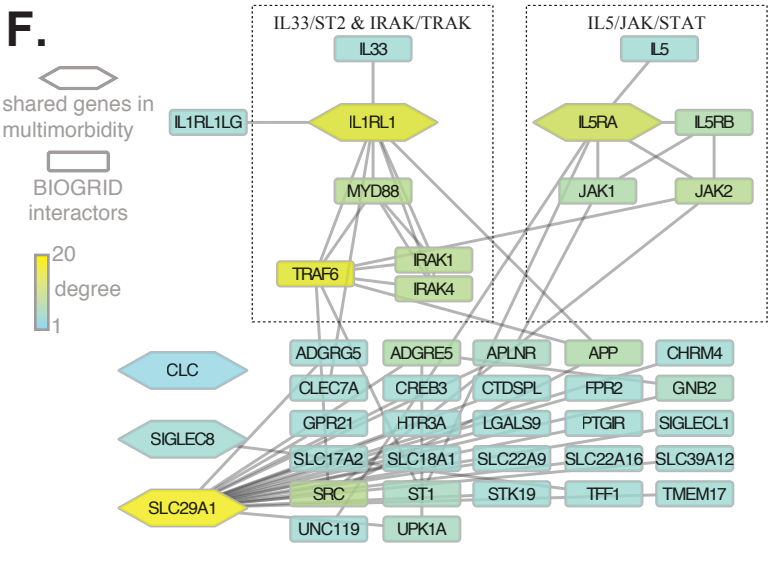

A.
asthma multimorbidity
102 participants
(A+D or A+R or A+D+R)
dermatitis multimorbidity
73 participants
(A+D or D+R or A+D+R)
any multimorbidity
123 participants
(A+D or A+R or
D+R or A+D+R)
rhinitis multimorbidity
89 participants
(A+R or D+R
or A+D+R)
CLC
EMR4P
FRRS1
HRH4
IL1RL1
IL5RA
SLC29A1


C.
ALOX15
ALOX15P1
IL1RL1
SIGLEC8
HRH4
EMR4P
CLC
SLC29A1
IL5RA
FRRS1
PIK3R6
LOC101929979
−log10(FDR)
2
1.5
rhinitis multi.
asthma multi.
dermatitis multi.


D.
normalized fold change
number of co-occurrent disease
CLC
IL5RA
EMR4P
IL1RL1
LOC101929979
ALOX15
HRH4
ALOX15P1
FRRS1
SIGLEC8
PIK3R6
SLC29A1


E.
G−protein beta:gamma signalling
G ß:γ signalling through PI3Kγ
Interleukin−33 signaling
Histamine receptors
Signal Transduction
Transport of nucleosides, free purine, pyrimidine
Signaling by Interleukins
GPCR downstream signalling
Signaling by GPCR
Cytokine Signaling in Immune system
Immune System
Interleukin receptor SHC signaling
Cytokine−cytokine receptor interaction
Interleukin−3, 5 and GM−CSF signaling
Interleukin−2 family signaling
Transport of vitamins, nucleosides
Amine ligand−binding receptors
PI5P, PP2A, IER3 Regulate PI3K/AKT Signaling
Negative regulation of the PI3K/AKT network
−log10(Fisher p)


B.
−log10(FDR)
SENP1
BCAP29
CEBPE
FAM66C
TNFSF14
LOC101927780
HMGA1P4
LOC101929526
EXOC5
MIR539
AK2
GLT6D1
RNA5SP440
DUSP16
SLC25A40
GPR65
RNF149
MAGEA9B
RNA5SP378
LOC101928812
RNA5SP335
RP1.149A16.12
LA16c.395F10.1
NPIPB5
LEPREL4
CLINT1
NDUFB5
CLC
FRRS1
HRH4
IL5RA
LOC101929979
PIK3R6
SLC29A1
EMR4P
EMR4P
IL1RL1
CYSLTR2
IL1RL1
SORD
SIGLEC8
ALOX15P1
ALOX15
LGALS12
TTC7B
PNPLA6
SEMA7A
SMPD3
BACE2.IT1
CACNA1D
ACSM3
ZSCAN29
rhinitis alone
asthma alone
dermatitis alone
asthma multi.
dermatitis multi.
rhinitis multi.
multi.


F.
shared genes in multimorbidity
BIOGRID interactors
20
degree
1
IL33/ST2 & IRAK/TRAK
IL5/JAK/STAT
IL33
IL5
IL1RL1LG
IL1RL1
IL5RA
IL5RB
MYD88
JAK1
JAK2
TRAF6
IRAK1
IRAK4
CLC
SIGLEC8
SLC29A1
ADGRG5
ADGRE5
APLNR
APP
CHRM4
CLEC7A
CREB3
CTDSPL
FPR2
GNB2
GPR21
HTR3A
LGALS9
PTGIR
SIGLECL1
SLC17A2
SLC18A1
SLC22A9
SLC22A16
SLC39A12
SRC
ST1
STK19
TFF1
TMEM17
UNC119
UPK1A