



## ORIGINAL ARTICLE

## Epidemiology and Genetics



WILEY

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

Nathanaël Lemonnier<sup>1</sup> | Erik Melén<sup>2,3</sup> | Yale Jiang<sup>4,5</sup> | Stéphane Joly<sup>6</sup> | Camille Ménard<sup>6</sup> | Daniel Aguilar<sup>7</sup> | Edna Acosta-Perez<sup>8</sup> | Anna Bergström<sup>3</sup> | Nadia Boutaoui<sup>4</sup> | Mariona Bustamante<sup>9,10,11</sup> | Glorisa Canino<sup>8</sup> | Erick Forno<sup>4</sup> | Juan Ramon González<sup>9</sup> | Judith Garcia-Aymerich<sup>9</sup> | Olena Gruzieva<sup>3</sup> | Stefano Guerra<sup>9,12</sup> | Joachim Heinrich<sup>13,14</sup> | Inger Kull<sup>15</sup> | Jesús Ibarluzea Maurologoitia<sup>11,16</sup> | Loreto Santa-Marina Rodriguez<sup>11</sup> | Elisabeth Thiering<sup>13,17</sup> | Magnus Wickman<sup>18</sup> | Cezmi Akdis<sup>19</sup> | Mübeccel Akdis<sup>19</sup> | Wei Chen<sup>4,20</sup> | Thomas Keil<sup>21,22,23</sup> | Gerard H. Koppelman<sup>24,25</sup> | Valérie Siroux<sup>1</sup> | Cheng-Jian Xu<sup>24,25</sup> | Pierre Hainaut<sup>1</sup> | Marie Standl<sup>13</sup> | Jordi Sunyer<sup>9</sup> | Juan C. Celedón<sup>4</sup> | Josep Maria Antó<sup>9</sup> | Jean Bousquet<sup>26,27,28,29</sup>

<sup>1</sup>Institute for Advanced Biosciences, UGA-INSERM U1209-CNRS UMR5309, Allée des Alpes, France

<sup>2</sup>Department of Clinical Science and Education Södersjukhuset, Karolinska Institutet, Sachs' Children's Hospital, Stockholm, Sweden

<sup>3</sup>Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

<sup>4</sup>Division of Pediatric Pulmonary Medicine, UPMC Children's Hospital of Pittsburgh, University of Pittsburgh, Pittsburgh, PA, USA

<sup>5</sup>School of Medicine, Tsinghua University, Beijing, China

<sup>6</sup>CIRI, International Center for Infectiology Research, Inserm, Lyon, France

<sup>7</sup>Biomedical Research Networking Center in Hepatic and Digestive Diseases (CIBEREHD), Barcelona, Spain

<sup>8</sup>Behavioral Sciences Research Institute, University of Puerto Rico, San Juan, Puerto Rico

<sup>9</sup>ISGlobal, Barcelona Institute for Global Health, Barcelona, Spain

<sup>10</sup>Universitat Pompeu Fabra (UPF), Barcelona, Spain

<sup>11</sup>Instituto de Salud Carlos III, Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Madrid, Spain

<sup>12</sup>Asthma and Airway Disease Research Center, University of Arizona, Tucson, AZ, USA

<sup>13</sup>Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany

<sup>14</sup>Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Inner City Clinic, University Hospital of Munich (LMU), Munich, Germany

<sup>15</sup>Department of Clinical Science and Education, Södersjukhuset, Karolinska Institute, Stockholm, Sweden

<sup>16</sup>BIODONOSTIA, Instituto de Investigación Sanitaria, Donostia-San Sebastián, Spain

<sup>17</sup>Division of Metabolic Diseases and Nutritional Medicine, Dr. von Hauner Children's Hospital, Ludwig-Maximilians-University of Munich, Munich, Germany

<sup>18</sup>Centre for Clinical Research Sörmland, Uppsala University, Eskilstuna, Sweden

<sup>19</sup>Swiss Institute of Allergy and Asthma Research (SIAF), University of Zurich, Davos, Switzerland

**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; PPI, Protein-Protein Interaction Network; R, allergic Rhinitis; RT-qPCR, Reverse Transcription quantitative Polymerase Chain Reaction; WGCNA, Weighted Gene Co-Expression Network Analysis.

Lemonnier, Melén, Jiang, Celedón, Antó and Bousquet are contributed equally.

<sup>20</sup>Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, USA

<sup>21</sup>Institute of Social Medicine, Epidemiology and Health Economics, Charité – Universitätsmedizin Berlin, Berlin, Germany

<sup>22</sup>Institute for Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany

<sup>23</sup>Institute for Health Resort Medicine and Health Promotion, Bavarian Health and Food Safety Authority, Bad Kissingen, Germany

<sup>24</sup>Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

<sup>25</sup>GRIAC Research Institute, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

<sup>26</sup>MACVIA-France, Fondation partenariale FMC VIA-LR, Montpellier, France

<sup>27</sup>Ageing and Chronic Diseases, Epidemiological and Public Health Approaches, Université Versailles St-Quentin-en-Yvelines, UMR-S 1168, Montigny le Bretonneux, France

<sup>28</sup>Euforea, Brussels, Belgium

<sup>29</sup>Department of Dermatology and Allergy, Charité – Universitätsmedizin Berlin, Comprehensive Allergy Center, Berlin, Germany

## Correspondence

Nathanaël Lemonnier, Institute for Advanced Biosciences, UGA-INSERM U1209-CNRS UMR5309, Site Santé, Allée des Alpes, 38700 La Tronche, France.  
Email: nathanael.lemonnier@univ-grenoble-alpes.fr

## Funding information

NIH, Grant/Award Number: HL079966, HL117191, MD011764 and U54MD007587; ANR, Grant/Award Number: ANR-15-IDEX-02; European Research Council, Grant/Award Number: 757919; Federal Ministry for Education, Science, Research and Technology – Helmholtz Zentrum Munich – Research Institute at Marien-Hospital Wesel – LMU Munich – TU Munich – IUF – Leibniz Research-Institute for Environmental Medicine, University of Düsseldorf; CLARA – CHUGA – UGA, Grant/Award Number: Chair of Excellence in Translational Research; Instituto de Salud Carlos III, Grant/Award Number: CB06/02/0041, G03/176, PI041436 and PI081151; Federal Ministry for Environment, Grant/Award Number: FKZ 20462296; 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 – Stockholm County Council; Generalitat de Catalunya, Grant/Award Number: CIRIT 1999SGR 00241; Mechanisms of the Development of ALLergy (MeDALL) project, European framework programme 7, Grant/Award Number: 261357; Bart

## 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 1233 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 Sequencing 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 in the 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 multimorbidity 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

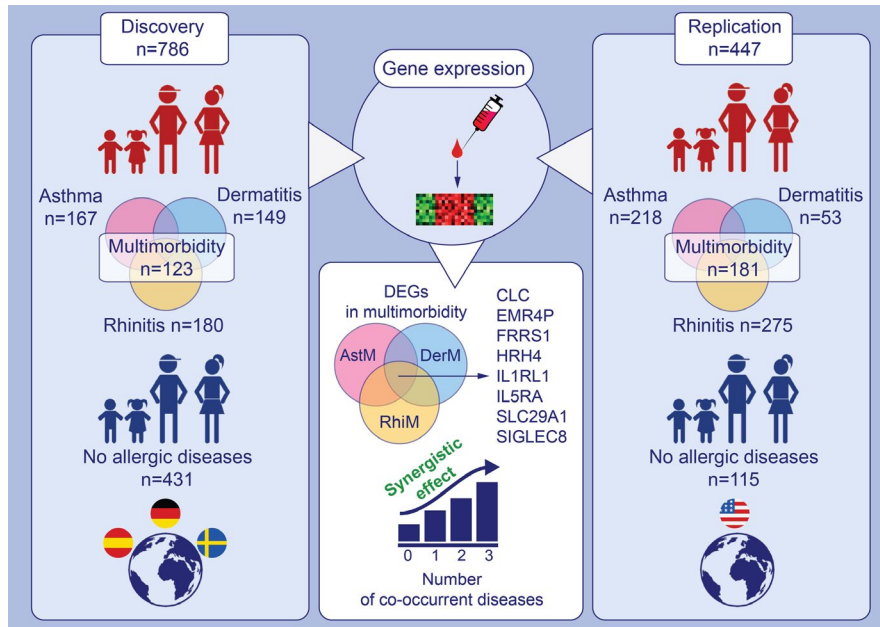
asthma, atopic dermatitis, multimorbidity, rhinitis, transcriptomics

## 1 | INTRODUCTION

Allergic diseases (asthma: A, dermatitis: D, rhinitis: R) are complex and often coexist in the same patient (multimorbidity).<sup>1</sup> In children and adolescents, over 85% of asthmatic patients have rhinitis, suggesting

common gene pathways, while only 20%-30% of rhinitis patients have asthma suggesting rhinitis-specific genes.<sup>2</sup> Atopic dermatitis is characterized by a complex clinical phenotype varying during the life cycle.<sup>3</sup>

Mechanisms of the Development of ALLergy (MeDALL) found that multimorbidity was more common than expected by chance,



## GRAPHICAL ABSTRACT

This study compares gene expression from whole blood of European children (4-16 years) with asthma and/or dermatitis and/or rhinitis to controls without allergy. Eight genes are overlapping among DEGs found in multimorbidity for asthma, dermatitis and rhinitis, which had synergistic effects along the number of co-occurring diseases. Results were replicated in North American cohort with similar features.

suggesting genes underlying multimorbidity.<sup>1,4</sup> MeDALL proposed that type 2 patterns are associated with multimorbidity,<sup>5</sup> as suggested through an in silico model.<sup>6</sup> Risk variants shared between asthma, dermatitis, and rhinitis were identified in genome-wide association studies, without considering the impact of multimorbidity.<sup>7,8</sup> 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.<sup>9</sup>

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.<sup>9</sup> 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.

## 2 | METHODS

### 2.1 | 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 (nonallergic participants, Table 1, Figure S1). We analyzed 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,<sup>9</sup> 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.

### 2.2 | Discovery study

#### 2.2.1 | 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,<sup>10</sup> Sweden (GSE141623, mean age:  $16.7 \pm 0.4$  years), 329 from GINIplus,<sup>11</sup> Germany ( $15.2 \pm 0.3$  years), and 201 from INMA,<sup>12</sup> Spain (GSE141631,  $4.2 \pm 0.3$  years). Over 95% of children were of European ancestry (Figure S1). Samples were enriched for allergic cases.

**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, % <sup>a</sup> )	143 (55.9)	160 (48.6)	99 (49.2)	402 (51.1)	177 (53.3)
female (n, % <sup>a</sup> )	113 (44.1)	169 (51.4)	102 (50.8)	384 (48.9)	155 (46.7)
controls (n, % <sup>a</sup> )	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, % <sup>b</sup> )	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)

Note: Characteristics of discovery and replication population. A: asthma, D: dermatitis, R: rhinitis.

<sup>a</sup>% of total.

<sup>b</sup>% of cases.

## 2.2.2 | Disease definition

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

## 2.2.3 | Data production

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

## 2.2.4 | Statistical methods—data cleaning and differential expression

Expression levels were normalized with the robust multiarray average algorithm,<sup>13</sup> using *oligo* in R v3.5.3. Multivariable models were adjusted for covariables (sex, cohort, batch, age) and fit with *limma* to identify DEGs (NetAffx v36 annotation). Surrogate variable analysis was used to capture cell blood heterogeneity (Figure S2).<sup>14</sup> *P*-values were corrected for multiple testing (Benjamini-Hochberg false discovery rate, FDR),<sup>15</sup> 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  $\sigma$  based on the difference in normalized FC between 1 and 2 diseases, and 2 and 3 diseases.

## 2.2.5 | 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*).

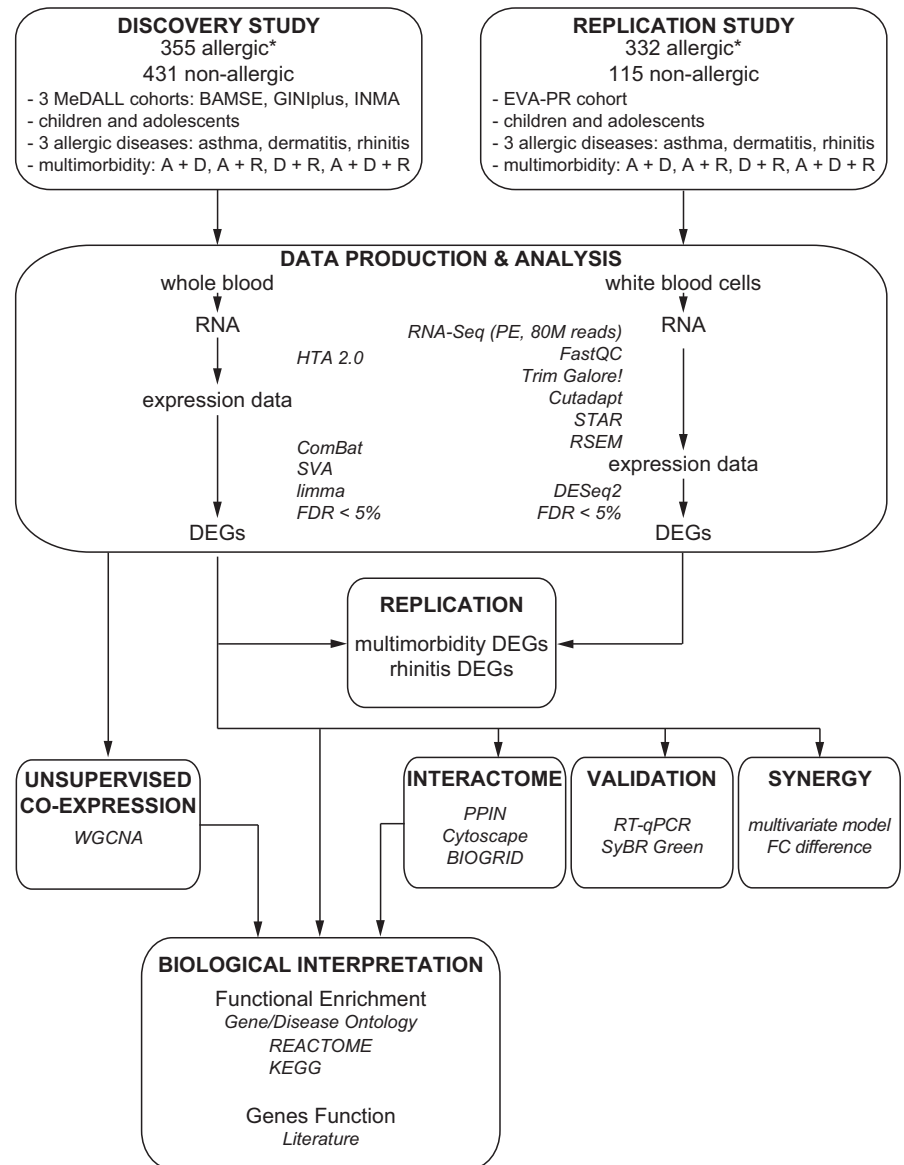
## 2.2.6 | Validation

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.

## 2.2.7 | 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

**FIGURE 1** Data analysis workflow. Gene expression levels of allergic participants were compared to nonallergic 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-occurrence: 1 disease, or 2 or more diseases (multimorbidity), and/or based on the nature of the disease: asthma, dermatitis or rhinitis, and compared to nonallergic participants



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.

## 2.3 | Replication study

### 2.3.1 | Setting and population

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

### 2.3.2 | 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.

### 2.3.3 | Data production

RNA Sequencing was performed using 350 ng of RNA extracted from white blood cells after removing hemoglobin (RNA Integrity

**TABLE 2** Differentially expressed genes in MeDALL and replication in EVA-PR

Gene identifier	Known in allergy	Known in immune process	MeDALL			EVA-PR		
			Any multimorbidity (62.1%)			Any multimorbidity (51.1%)		
			log <sub>2</sub> FC	P-value	FDR	log <sub>2</sub> FC	P-value	FDR
<b>A. Multimorbidity</b>								
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 <sup>a</sup>	N/A	N/A	0.32	5.7E-07	7.1E-03	Null <sup>b</sup>	Null <sup>b</sup>	Null <sup>b</sup>
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
Gene Identifier	Known in allergy	Known in immune process	MeDALL			EVA-PR		
			Rhinitis alone (53.4%)			Rhinitis alone (42.1%)		
			log <sub>2</sub> FC	P-value	FDR	log <sub>2</sub> FC	P-value	FDR
<b>B. Rhinitis alone</b>								
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 <sup>a</sup>	N/A	N/A	-0.08	1.3E-05	1.6E-02	Null <sup>b</sup>	Null <sup>b</sup>	Null <sup>b</sup>
MAGEA9B	No	No	-0.05	1.5E-05	1.7E-02	Null <sup>b</sup>	Null <sup>b</sup>	Null <sup>b</sup>
RP1-149A16.12 <sup>a</sup>	N/A	N/A	-0.07	1.6E-05	1.8E-02	Null <sup>b</sup>	Null <sup>b</sup>	Null <sup>b</sup>
RNA5SP335 <sup>a</sup>	N/A	N/A	0.10	1.7E-05	1.8E-02	Null <sup>b</sup>	Null <sup>b</sup>	Null <sup>b</sup>
LOC101928812 <sup>a</sup>	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
NPIP5	No	No	-0.08	3.2E-05	2.6E-02	0.05	>5E-02	>5E-02
LA16c-395F10.1 <sup>a</sup>	N/A	N/A	0.09	3.4E-05	2.7E-02	Null <sup>b</sup>	Null <sup>b</sup>	Null <sup>b</sup>
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 <sup>b</sup>	Null <sup>b</sup>	Null <sup>b</sup>
MIR539	No	No	-0.11	4.2E-05	3.2E-02	Null <sup>b</sup>	Null <sup>b</sup>	Null <sup>b</sup>
LOC101927780 <sup>a</sup>	N/A	N/A	0.08	4.2E-05	3.2E-02	0.18	1.3E-04	2.3E-03
RNA5SP440 <sup>a</sup>	N/A	N/A	-0.09	4.4E-05	3.3E-02	Null <sup>b</sup>	Null <sup>b</sup>	Null <sup>b</sup>
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

(Continues)

TABLE 1 (Continued)

Gene Identifier	Known in allergy	Known in immune process	MeDALL			EVA-PR		
			Rhinitis alone (53.4%)			Rhinitis alone (42.1%)		
			log <sub>2</sub> FC	P-value	FDR	log <sub>2</sub> FC	P-value	FDR
EXOC5	No	No	<b>0.04</b>	<b>5.8E-05</b>	<b>3.8E-02</b>	0.01	>5E-02	>5E-02
HMGA1P4	No	No	<b>0.10</b>	<b>7.3E-05</b>	<b>4.4E-02</b>	Null <sup>b</sup>	Null <sup>b</sup>	Null <sup>b</sup>
LOC101929526 <sup>a</sup>	N/A	N/A	<b>0.09</b>	<b>8.2E-05</b>	<b>4.8E-02</b>	Null <sup>b</sup>	Null <sup>b</sup>	Null <sup>b</sup>

Note: Differentially expressed genes with log<sub>2</sub>FC 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 < .05 are in bold.

<sup>a</sup>Long intergenic nonprotein coding RNA, pseudogenes or ribosomal pseudogenes, their role in allergy or immune processes could not be determined (N/A: not applicable)

<sup>b</sup>null: not tested due to QC and RNA Sequencing issues. Genes that were specific for asthma, dermatitis, and rhinitis multimorbidity are not shown (ACSM3, SIGLEC8, ALOX15, ALOX15P1, BACE2-IT1, LGALS12, PNPLA6, SEMA7A, SMPD3, and SORD).

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 Sequencing 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.

### 2.3.4 | Statistical methods—Differential expression

DEGs were analyzed based on raw counts using DESeq2.<sup>16</sup> Multivariable models were adjusted for age, sex, batch, and cell type proportions of eosinophils, neutrophils, monocytes, and lymphocytes in white blood cells. Subjects without any allergic disease were used as controls. A zero-mean normal prior was put on the nonintercept 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.<sup>15</sup>

### 2.3.5 | 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.

## 3 | RESULTS

### 3.1 | Clinical characteristics of discovery and replication population

In MeDALL, 46% of allergic participants had asthma (61% with multimorbidity), 42% dermatitis (48%), and 51% rhinitis (49%). Asthma was more common in BAMSE (63%), and rhinitis in GINIplus (79%). Dermatitis in preschool 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%), and 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.

### 3.2 | 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 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).

### 3.3 | 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.

### 3.4 | Synergistic effects of multimorbidity on the significant genes

Using a linear model including all samples (Table S6), all genes had synergistic effects of multimorbidity ( $\sigma = 1.5$ -58.4), except *LOC101929979* ( $\sigma = 1.1$ ). 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).

### 3.5 | 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 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).

### 3.6 | 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,  $P = 2.1e-15$ ).

### 3.7 | 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 ( $r^2 = .14$ ,  $P = 5e-05$ ), asthma multimorbidity, dermatitis multimorbidity, and rhinitis multimorbidity ( $r^2 = 0.13$ ,  $P$ -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.

## 4 | 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



**TABLE 3** DEGs shared between multimorbidity for asthma, dermatitis, or rhinitis

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

Note: Name, description and previous identification of genes shared between asthma, dermatitis, and rhinitis multimorbidity.

<sup>a</sup>See online Supplement.

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 signaling 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.<sup>5</sup>

#### 4.1 | 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,<sup>17</sup> 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,<sup>18</sup> does not prevent biological significance.<sup>19</sup> 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.

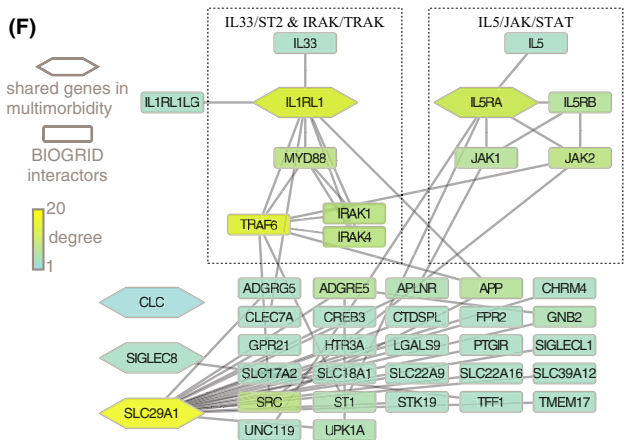
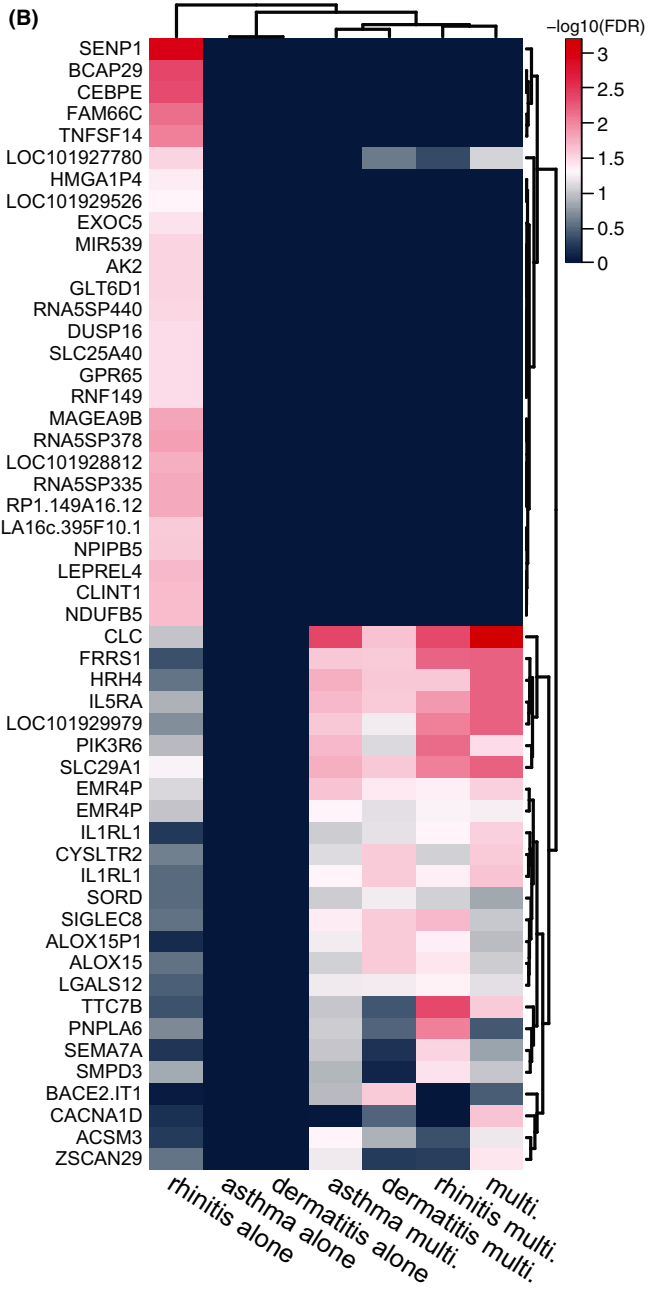
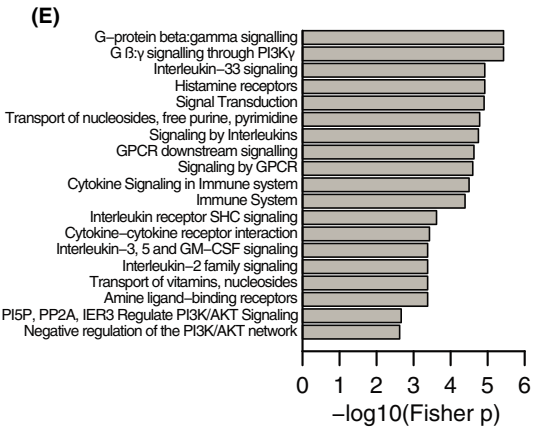
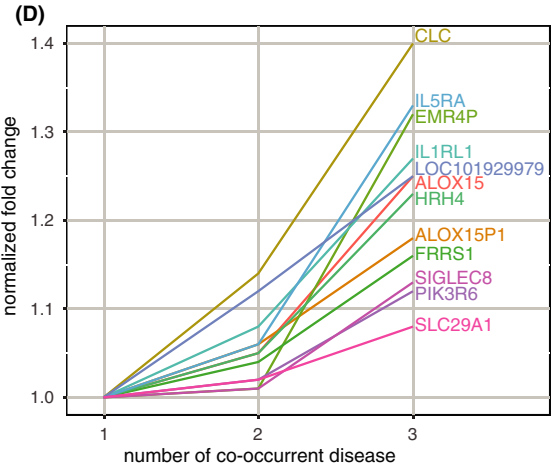
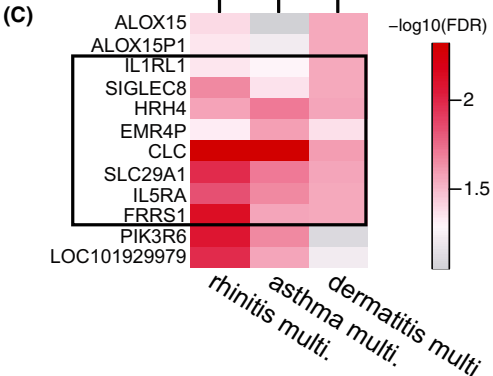
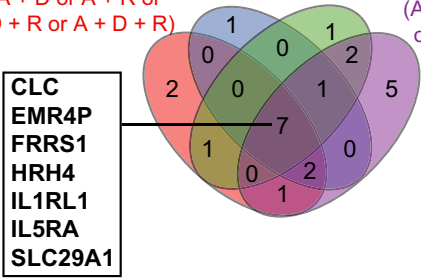
Our discovery study included two cohorts of adolescents and one of children, thus limiting our capacity to assess the role of age.<sup>20</sup> Nevertheless, FDR was reached for 9 genes grouping adolescents with any multimorbidity, including *CLC*, *FRRS1*, *HRH4*, *IL5RA*, *EMR4P*, and *IL1RL1*, shared 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 (eg, inclusion of positive IgE sensitization, or symptom-based definition versus self-reported doctor diagnosis) could have led to bias in overselection of IgE mechanism-associated genes, or differences in disease prevalence (overselection of more severe cases), but the reproducibility of the results showed the generalizability of the findings.

#### 4.2 | 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.<sup>9,21</sup> 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

(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)



**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  $-\log_{10}(\text{FDR } P\text{-values})$  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  $-\log_{10}(\text{FDR } p\text{-values})$  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 ( $\text{FDR} < 5\%$ ) 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. Octagons: proteins encoded by our candidate genes; rounded rectangle: interactors, alphabetically sorted in grid layout on the bottom right. All types of interactions were considered

were noted between the discovery and replication cohorts, our analyses identify robust multimorbidity genes such as *CLC*.

In children and adolescents, asthma or dermatitis alone was 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.<sup>2</sup> 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.

### 4.3 | Candidate genes function

Of the eight multimorbidity genes, five were eosinophil-specific genes. *CLC* (Galectin-10),<sup>22</sup> which is involved in regulation of T-cell proliferation, and predicts oral corticosteroid response in asthma (see Supplementary Bibliography), featured the highest  $\log_2\text{FC}$  and was significant in any type of multimorbidity. *IL5RA* has been involved in dermatitis<sup>23</sup> and asthma,<sup>24</sup> *IL1RL1* in asthma,<sup>25</sup> and was found associated with multimorbidity in our computational study of the protein interaction network.<sup>5</sup> *SIGLEC8* is a susceptibility locus for asthma.<sup>26</sup> 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),<sup>27</sup> *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 remodeling of asthma.<sup>28</sup> *CLC*, *IL5RA*, *SIGLEC8*, *SEMA7A* correlated strongly to differentially methylated CpG sites in asthma.<sup>17</sup> In addition to *HRH4*, *ALOX15*,<sup>29</sup> and *CYSLTR2*<sup>30</sup> were related to allergic diseases and immune response.

Among genes found in rhinitis alone, *SEN1* is important for the epithelial-mesenchymal trophic unit, *AK2*, *CEBPE*, *GPR65*, are involved in white blood cells differentiation, and *BCAP29* influences the traffic of MHC class I molecules (online Supplement). *TNFSF14* was known in allergy with no known role in inflammation.<sup>31</sup>

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 nonprotein 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.<sup>14</sup> 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  $\beta_2$  agonists did not impact the observed effect (Table S4.2).

### 4.4 | 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 asthma, dermatitis, and rhinitis 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 cutoff, 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).

## 5 | 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 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.

### ACKNOWLEDGMENTS

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 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, Smaghe 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>.

### CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

### AUTHOR'S CONTRIBUTION

Nathanaël Lemonnier performed the transcriptomic study, led the data production, led data preprocessing 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, Mübecce 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. 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.

## ORCID

Nathanaël Lemonnier  <https://orcid.org/0000-0002-3994-8697>  
 Erick Forno  <https://orcid.org/0000-0001-6497-9885>  
 Inger Kull  <https://orcid.org/0000-0001-6096-3771>  
 Cezmi Akdis  <https://orcid.org/0000-0001-8020-019X>  
 Mübeccel Akdis  <https://orcid.org/0000-0003-0554-9943>  
 Valérie Siroux  <https://orcid.org/0000-0001-7329-7237>  
 Cheng-Jian Xu  <https://orcid.org/0000-0003-1586-4672>

## REFERENCES

- Pinart M, Benet M, Annesi-Maesano I, et al. Comorbidity of eczema, rhinitis, and asthma in IgE-sensitized and non-IgE-sensitized children in MeDALL: a population-based cohort study. *Lancet Respir Med*. 2014;2:131-140.
- 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.
- Bieber T. Atopic dermatitis. *N Engl J Med*. 2008;358:1483-1494.
- 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;139:388-399.
- 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;70:1062-1078.
- Aguilar D, Pinart M, Koppelman GH, et al. Computational analysis of multimorbidity between asthma, eczema and rhinitis. *PLoS ONE*. 2017;12:e0179125.
- Deménais 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.
- Ferreira MA, Vonk JM, Baurecht H, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat Genet*. 2017;49:1752-1757.
- Jiang Y, Gruziova O, Wang T, et al. Transcriptomics of atopy and atopic asthma in white blood cells from children and adolescents. *Eur Respir J*. 2019;53(5):1900102.
- Gref A, Merid SK, Gruziova 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;195:1373-1383.
- 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;40:627-636.
- Guxens M, Ballester F, Espada M, et al. Cohort profile: the INMA-Infancia y Medio Ambiente-(Environment and Childhood) project. *Int J Epidemiol*. 2012;41:930-940.
- Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4:249-264.
- 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.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B*. 1995;57:289-300.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
- Xu CJ, Söderhäll C, Bustamante M, et al. DNA methylation in childhood asthma: an epigenome-wide meta-analysis. *Lancet Respir Med*. 2018;6:379-388.
- 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.
- St Laurent G, Shtokalo D, Tackett MR, et al. On the importance of small changes in RNA expression. *Methods*. 2013;63(1):18-24.
- 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.
- 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.
- Su J. A Brief history of charcot-leyden crystal protein/Galectin-10 research. *Molecules*. 2018;23:2931.
- 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;21:1052-1058.
- 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;155:1413-1418.
- 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.
- 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;18:713-719.
- Walter M, Kottke T, Stark H. The histamine H4 receptor: Targeting inflammatory disorders. *Eur J Pharmacol*. 2011;668:1-5.
- Mizutani N, Nabe T, Yoshino S. Semaphorin 7A plays a critical role in IgE-mediated airway inflammation in mice. *Eur J Pharmacol*. 2015;764:149-156.
- 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;148:1313-1322.
- 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;1(7):299.
- 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;41:1021-1031.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Lemonnier N, Melén E, Jiang Y, et al. A novel whole blood gene expression signature for asthma, dermatitis, and rhinitis multimorbidity in children and adolescents. *Allergy*. 2020;00:1-13. <https://doi.org/10.1111/all.14314>