Figures and figure supplements

Trajectories of childhood immune development and respiratory health relevant to asthma and allergy

Howard HF Tang et al
Figure 1. Non-parametric mixture-model-based clustering of CAS dataset, based on 174 features. SPT = skin prick test. White spaces within the heatmap indicate missing data. Rows represent individuals; columns represent clustering features with general categories as labelled on grey background. Variables with grey background are clustering features ordered by category or type of variable first (e.g. all HDM IgE-related variables grouped together), then by timepoint (earlier to later, from left to right). Variables with lilac background indicate resultant cluster membership and outcome variable (age-5 wheeze). Heatmap values are scaled relative to range and median values for each feature; the median is coloured beige-yellow, the median +range red, and median – range blue. For sex, −1/blue = female, 0/yellow (median) = male.

DOI: https://doi.org/10.7554/eLife.35856.003
Figure 1—figure supplement 1. Scatterplot of principal components analysis (PCA) of the complete-case CAS dataset (N = 186), with points coloured by npEM clusters. Each point represents an individual. The first two PCs (shown) account for 16.7% of the total variance.

DOI: https://doi.org/10.7554/eLife.35856.004
Figure 1—figure supplement 2. Silhouette widths of clusters generated by npEM. $j = \text{cluster number}$; $n_j = \text{cluster size}$; $\text{ave}_{i \in C_j} s_i = \text{average silhouette width among members } i \text{ of cluster } C_j$. Overall average silhouette width across all clusters is also given.

DOI: https://doi.org/10.7554/eLife.35856.005
Figure 1—figure supplement 3. Overview of study methodology. Dashed arrows indicate non-critical elements of our method.
DOI: https://doi.org/10.7554/eLife.35856.006
**Figure 2.** Graphical summary of proposed clusters. “Early” specifically refers to “within the first 6 months of life”.
DOI: https://doi.org/10.7554/eLife.35856.007
Figure 3. Incidence of multiple phenotypes, including parent-reported wheeze. (A) Physician-diagnosed asthma (B) defined wheeze phenotypes (C) in relation to food and inhalant sensitisation (D) stratified by cluster and time in the CAS dataset. Points indicate observed proportion; bars indicate 95% CI (binomial distribution). Wheeze phenotypes defined as: no wheeze = no wheeze at ages 1 to 3, or age 5; transient wheeze = any wheeze at ages 1 to 3, but not age 5; late wheeze = wheeze at age 5, but not ages 1 to 3; persistent wheeze = any wheeze at both ages 1 to 3 and age 5. Food sensitisation defined as peanut IgE ≥0.35 kU/L at any age, or cow’s milk, egg white, peanut SPT > 2 or 3 mm for age ≤2 or >2 respectively. Inhalant sensitisation defined as HDM, cat, couchgrass, ryegrass, mould or Phadiatop IgE ≥0.35 kU/L at any age, or mould SPT (Alternaria or Aspergillus spp.)>2 or 3 mm for age ≤2 or >2, respectively.

DOI: https://doi.org/10.7554/eLife.35856.008
Figure 3—figure supplement 1. Relationship of clusters to food sensitisation, eczema and wheeze. Percentages denote proportion of cluster displaying phenotype (numbers in brackets denote actual sample numbers). Food sensitisation defined as peanut IgE ≥0.35 kU/L at any age, or cow’s milk IgE ≥0.35 kU/L at 3, 4, or 5 years of age.

A: Cluster and food sensitisation at age five

B: Cluster and wheeze status at age five
milk, egg white, peanut SPT > 2 or 3 mm for age ≤2 or >2, respectively. Subphenotypes defined for food sensitisation, eczema and wheeze as: no phenotype = phenotype absent at all ages; transient = any incidence of phenotype at the earlier ages (1 to 3 for wheeze, 6 m to three for eczema, 6 m to two for sensitisation), but not age 5; late = phenotype at age 5, but not the earlier ages; persistent = any incidence of phenotype at both earlier ages and age 5.

DOI: https://doi.org/10.7554/eLife.35856.009
Figure 4. HDM IgE (A), IgG (B) and IgG4 (C), and peanut IgE (D) and IgG4 (E) stratified by cluster and time, in the CAS dataset. Points indicate means; bars indicate 95% CI (t-distribution).

DOI: https://doi.org/10.7554/eLife.35856.010
Figure 4——figure supplement 1. Correlation patterns between IgE vs IgG4 (A) and IgE vs IgG (B) at age five *p<0.05 for Spearman correlation with Holm correction for multiple testing. Note the slightly stronger heat along the main diagonals of both heatmaps, especially for HDM.

DOI: https://doi.org/10.7554/eLife.35856.011
Distinct biological signals of HDM IgE, IgG4, SPT, and Th2 cytokine (IL-13). (A) Th2 cytokine (IL-13) to HDM stimulation of PBMCs in vitro, vs. HDM IgE responses in vivo, stratified by age of testing, and SPT result (positive denoted by ≥2 mm at age <2 or 3 mm at age 5). Dotted line represents traditional threshold for HDM IgE-positive result (0.35 kU/L). Note the significant number of individuals on either side of the dotted line for both HDM SPT-positive and negative subgroups. (B) Th2 cytokine (IL-13) to HDM stimulation of PBMCs in vitro, vs. HDM IgG4 responses in vivo, stratified by age of testing and HDM IgE result (positive denoted by ≥0.35 kU/L).

DOI: https://doi.org/10.7554/eLife.35856.012
**Figure 5.** LRI frequency (A), wheezy LRI (wLRI) frequency (B), and HDM IgE (C), stratified by age-5 wheeze status, cluster and time, in the CAS dataset. Points indicate means; bars indicate 95% CI (t-distribution). #p<0.05 for repeated-measures ANOVA across timepoints from the first 3 years of life (see Table 4). *p<0.05 for Mann-Whitney-Wilcoxon comparison within each timepoint.

DOI: https://doi.org/10.7554/eLife.35856.015
Figure 5—figure supplement 1. A ‘simple’ decision tree generated by recursive partitioning from CAS data, with breakdown of tree clusters by actual CAS npEM-derived clusters. (A); scatterplot showing separation of CAS clusters by decision split thresholds (B). Percentages in Panel A may not sum up.

Figure 5—figure supplement 1 continued on next page.
Figure 5—figure supplement 1 continued

to 100% because some individuals have missing values for decision node variables, hence making them impossible to classify. In Panel B, note that leftmost column of points represent values of HDM IgG4 that were less than the limit-of-detection (LOD) for that assay (0.0003 μg/L) and were subsequently assigned to half the LOD (0.00015 μg/L). Most of these points belonged to individuals from CAS1.

DOI: https://doi.org/10.7554/eLife.35856.016
Figure 5—figure supplement 2. Decision tree generated by recursive partitioning from CAS data, excluding Phadiatop assay variables.
DOI: https://doi.org/10.7554/eLife.35856.017
Figure 5—figure supplement 3. A ‘comprehensive’ decision tree generated by recursive partitioning from CAS data, given CAS npEM-derived clusters and age-5 wheezing status. CAS1.x, 2.x, and 3.x, and tree cluster 1.x, 2.x and 3.x, refer to the intersection of npEM-generated clusters and age-5 wheeze status, and their analogous decision tree cluster, respectively. The second digit (x.0 or x.1) refers to age-5 wheeze status, with ‘1’—present wheeze and ‘0’—no wheeze. Boxes, bars and digits with red outline indicate those with predicted (tree cluster) or actual (CAS npEM cluster) age-5 wheeze. Note that the tree did not predict for non-wheezing CAS3, so there is no tree cluster 3.0, and all CAS3 individuals were automatically assigned to a wheezing tree cluster.

DOI: https://doi.org/10.7554/eLife.35856.018
Figure 5—figure supplement 4. Comparison of predictors for age-5 wheeze in CAS and COAST clusters. Colour coding and numbers in cells indicate odds ratio (OR) of predictor for age-5 wheeze in GLM, with sex, maternal and paternal history of asthma, and (for CAS) BMI as covariates. Non-grey cell with number indicates statistically significant association (p<0.05). Grey cell with number indicates non-significant (p>0.05); grey non-numbered cell indicates test not done due to lack of data.

DOI: https://doi.org/10.7554/eLife.35856.019
Figure 6. PBMC expression of IL-5 (A) and IL-4 mRNA (B), as well as IL-13 protein (C), in response to stimulation HDM, stratified by cluster and time (CAS) Cord = cord blood sample collected at birth. Points indicate means; bars indicate 95% CI (t-distribution).

DOI: https://doi.org/10.7554/eLife.35856.022
Figure 7. Description of npEM-derived clusters in external cohorts: in MAAS, incidence of wheeze (A), asthma diagnosis (B), and HDM IgE levels (C); in COAST, incidence of asthma diagnosis (D), proportion of individuals with detectable aeroallergen-specific IgE levels (E), and PBMC protein expression (F).
of IL-13 following HDM stimulation above unstimulated control (F) MAAS cohort (N = 934) was classified using npEM model from CAS, into MAAS1 (N = 199, 21%), MAAS2 (N = 692, 74%) and MAAS3 (N = 43, 5%); these correspond to CAS clusters CAS1, 2 and 3, respectively. COAST cohort (N = 285) was similarly classified into COAST1 (N = 105, 37%), COAST2 (N = 151, 53%) and COAST3 (N = 29, 10%).

DOI: https://doi.org/10.7554/eLife.35856.023