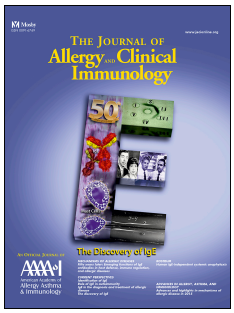
Accepted Manuscript



Evolution of IgE responses to multiple allergen components throughout childhood

Rebecca Howard, MSc, Danielle Belgrave, PhD, Panagiotis Papastamoulis, PhD, Angela Simpson, MD PhD, Magnus Rattray, PhD, Adnan Custovic, MD PhD

PII: S0091-6749(18)30216-1

DOI: 10.1016/j.jaci.2017.11.064

Reference: YMAI 13285

To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 13 January 2017

Revised Date: 4 November 2017

Accepted Date: 13 November 2017

Please cite this article as: Howard R, Belgrave D, Papastamoulis P, Simpson A, Rattray M, Custovic A, Evolution of IgE responses to multiple allergen components throughout childhood, *Journal of Allergy and Clinical Immunology* (2018), doi: 10.1016/j.jaci.2017.11.064.

Evolution of IgE responses to multiple allergen components throughout childhood

Rebecca Howard MSc[1], Danielle Belgrave PhD[2], Panagiotis Papastamoulis PhD[1], Angela Simpson MD PhD*[3], Magnus Rattray PhD*[1], Adnan Custovic MD PhD*[2]

1. Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK.

2. Section of Paediatrics, Department of Medicine, Imperial College London, UK.

3. Division of Infection, Immunity and Respiratory Medicine, University of Manchester and University Hospital of South Manchester, Manchester Academic Health Sciences Centre, Manchester, United Kingdom.

*Joint senior authors

Correspondence and requests for reprints:
Adnan Custovic MD PhD FAAAI
Imperial College London, UK, Department of Paediatrics
St Mary's Campus Medical School, Room 244
Norfolk Place, London W2 1PG, UK
Tel: +44 20 7594 3274, Fax: +44 20 7594 3984
Email: a.custovic@imperial.ac.uk

Declaration of sources of funding: Supported by the UK Medical Research Council (MRC) Grants MR/K002449/1 and MR/LO12693/1, and the MRC Health eResearch Centre (HeRC) grant MR/K006665/1; DB is supported by the MRC Career development grant MRC MR/M015181/1

Word count: 3382
Abstract word count: 250

ABSTRACT

Background: There is a paucity of information about longitudinal patterns of IgE responses to allergenic proteins (components) from multiple sources.

Objective: To investigate temporal patterns of component-specific IgE responses from infancy to adolescence, and their relationship with allergic diseases.

Methods: In a population-based birth cohort, we measured IgE to 112 components at 6 follow-ups during childhood. We used a Bayesian method to discover cross-sectional sensitization patterns and their longitudinal trajectories, and related these patterns to asthma and rhinitis in adolescence.

Results: We identified one sensitization cluster at age one, 3 at age three, 4 at ages five and eight, 5 at age 11, and six at age 16 years. “Broad” cluster was the only cluster present at every follow-up, comprising of components from multiple sources. “Dust mite” cluster formed at age three and remained unchanged to adolescence. At age three, a single-component “Grass” cluster emerged, which at age five absorbed additional grass components and Fel d 1 to form the “Grass/cat” cluster. Two new clusters formed at age 11: “Cat” cluster and “PR-10/profilin” (which divided at age 16 into “PR-10” and “Profilin”). The strongest contemporaneous associate of asthma at age 16 years was sensitization to “Dust mite” cluster (OR [95% CI]: 2.6 [1.2-6.1], $P<0.05$), but the strongest early-life predictor of subsequent asthma was sensitization to “Grass/cat” cluster (3.5 [1.6–7.4], $P<0.01$).

Conclusions: We describe the architecture of the evolution of IgE responses to multiple allergen components throughout childhood, which may facilitate development of better diagnostic and prognostic biomarkers for allergic diseases.

CLINICAL IMPLICATIONS

Development of different clinical phenotypes of allergic diseases may be predicted by the distinct patterns of IgE responses to multiple allergenic proteins.

CAPSULE SUMMARY

We described the architecture of the evolution of IgE responses to multiple allergen components throughout childhood. Understanding this structure in the developmental pathways of IgE responses may facilitate development of better diagnostic and prognostic algorithms for allergic diseases.

KEY WORDS

IgE, childhood, component-resolved diagnostics, machine learning, allergens, asthma, rhinitis

ABBREVIATIONS:

Skin prick test: SPT

Immunoglobulin E: IgE

Component-resolved diagnostics: CRD

Immuno Solid-phase Allergen Chip: ISAC

House dust mite: HDM

Manchester Asthma and Allergy Study: MAAS

ISAC standardized units: ISU

Bernoulli Mixture Model: BMM

Markov chain Monte Carlo algorithm: MCMC

Odds ratio [95% confidence interval]: OR [95% CI]

Pathogenesis-related: PR

INTRODUCTION

Allergic sensitization is a risk factor for asthma and rhinitis[1-3], but the strength of this association is inconsistent[4,5]. A patient is typically deemed to be sensitized based on a positive skin prick test (SPT) or a blood test measuring specific IgE to a range of common inhalant and food allergens[6,7]. However, both these tests can be positive without the patient having any symptoms[8], and neither positive SPT nor IgE can confirm the expression of symptoms upon allergen exposure[8,9]. This is partly because the natural sources which are used to prepare the whole-allergen extracts for skin or blood testing contain multiple allergenic proteins (components), with each component potentially containing multiple epitopes for binding IgE[10]. There is increasing evidence that sensitization to some, but not all of these proteins is important for the expression of allergic disease[9,11,12]. Also, homologous proteins present in different allergen sources may be cross-reactive (e.g. profilins and PR-10 proteins in various plants, or tropomyosin present in mites, insects and crustaceans), and a positive SPT or IgE to whole allergen extract may reflect sensitization to a cross-reactive component[12,13]. Recent evidence suggests that assessing sensitization to allergen components (component-resolved diagnostics [CRD]) may more be informative than standard tests using whole allergen extracts[14].

Current multiplex CRD platforms such as the Immuno Solid-phase Allergen Chip (ImmunoCAP ISAC) allow testing of small volumes of serum for component-specific IgE to more than 100 allergen components in a single assay[13,15], with robust and reproducible results[16]. We have previously shown that patterns of component-specific IgE responses in this multiplex chip-based assay have reasonable discrimination ability for asthma and rhinoconjuinctivitis[17]. In a further study using latent variable modelling, we identified three cross-sectional clusters of IgE responses to different protein families at age 11 years, and each of these patterns was associated with different clinical symptoms[18]. Our subsequent study has indicated that longitudinal trajectories of the cross-sectional sensitization patterns to a limited number of grass

and house dust mite (HDM) allergens during childhood had different associations with clinical outcomes, suggesting that the time of onset of specific patterns of IgE response was critically important[19]. Posa *et al* have recently shown that IgE polysensitization to several HDM molecules predicts current rhinitis, and both current and future asthma[20].

Capturing the heterogeneity in longitudinal patterns of responses to multiple components from different sources is challenging, and the conventional analyses may over-aggregate the underlying complexity[21]. Cluster-based sensitization profiles may provide a methodological framework within which to address this issue[22,23]. We hypothesized that there are distinct developmental patterns of component-specific IgE responses to allergenic molecules from different sources, and that response patterns in early childhood may aid the prediction of clinical outcomes at a later date. To address our hypotheses, we used data from a well-characterised population-based birth cohort in which IgE responses to 112 allergen components were measured at six points from infancy to adolescence. We clustered allergen components based on component-specific IgE response profiles across subjects to identify cross-sectional sets of closely associated components at each age. We then determined the trajectories of these component clusters over time to investigate the evolution of sensitization patterns, and examined their relationship with disease outcomes.

METHODS

Study design, setting and participants

Manchester Asthma and Allergy Study is an unselected birth cohort; participants were recruited prenatally and followed prospectively[24,25]. The study was approved by the Research Ethics Committee; parents gave written informed consent. Participants attended review clinics at ages 1, 3, 5, 8, 11 and 16 years. Validated questionnaires were interviewer-administered to collect information on parentally-reported symptoms, physician-diagnosed diseases and treatments received. Blood samples were collected from participants who gave assent.

Detection and annotation of component-specific IgE antibodies

We measured IgE to 112 components from 51 sources using ImmunoCAP ISAC (Thermo Fisher, Uppsala, Sweden) at all six follow-ups. Levels of component-specific IgE antibodies were reported in ISAC standardized units (ISU). We discretized IgE data using a binary threshold (positive ≥0.30 ISU)[17]. We used the following annotations for component-specific IgE antibody responses:

Active components- We considered components to be active if at least three participants had a positive IgE response at each time point[18].

Components which "drop-out"- Components which become inactive after having been active at an earlier time point.

Definition of clinical outcomes at age 16 years

Current asthma: Any two of the following three features: (1) Current wheeze (positive answer to the question "Has your child had wheezing or whistling in the chest in the last 12 months?"); (2) Current use of asthma medication; (3) Physician-diagnosed asthma ever[26].

Current rhinitis: Positive answer to the question "In the past 12 months, has your child had a problem with sneezing or a runny or blocked nose when he/she did not have a cold or the flu?".

Statistical Analysis

Statistical grouping of allergen components: For each time point, we analysed the data for participants who had at least one positive IgE component response, and for the active allergen components[18]; we thus restricted our analysis to 10, 26, 63, 68, 71 and 72 active components at ages 1, 3, 5, 8, 11 and 16 respectively.

At each age, we inferred component clusters by clustering the data through Bayesian estimation of a mixture of Bernoulli distributions (Bernoulli Mixture Model–BMM). We inferred the model parameters, cluster membership and number of clusters using an allocation sampler with an unknown number of mixture components (representing clusters in our terminology). This sampler is embedded in a Metropolis-coupled Markov chain Monte Carlo (MCMC) algorithm (details in the supplementary material)[27-29]. The generated MCMC samples were post-processed using the ECR algorithm to overcome identifiability issues due to the label-switching problem[30-33]. The model, sampler and the means to post-process the results have been designed and implemented in R (http://www.r-project.org) by this paper's authors, and are published packages on CRAN, available as bayesBinMix() and label.switching() respectively[34,35]. Once the optimal number of clusters, *K,* was inferred at each age, the cluster membership was inferred conditional on that value.

Associations with clinical outcomes: CRD data from ages 1 and 3 years were sparse; we therefore evaluated the association between component clusters at ages 5 and 16 years with asthma and rhinitis at age 16 years. Children who did not respond to any active component were *a priori* assigned to a "Non-sensitized" group. A child was classed as being sensitized to a component cluster if he/she responded to at least 1 component within the cluster. We examined the association between sensitization to component clusters and clinical outcomes (asthma, wheeze and rhinitis) through logistic regression analyses (univariable and multiple); results are reported as odds ratios (OR) with 95% confidence intervals (CI).

RESULTS

Participant flow and demographic data

Of 1184 children born into the cohort, CRD data were available for at least one time point for 922 children. Participant flow is shown in Figure S1. Number of children with CRD data at each follow-up, and the proportion with at least one positive active component response are listed in Table S1. Demographic and clinical characteristics are summarised in Table 1; we observed some minor differences between children included and those excluded from this analysis, none of which were consistent across different ages.

Component-specific IgE responses across childhood

Active, inactive, and components which dropped out: A total of 86 components were active for at least one time-point. Components which were inactive at all ages (*n*=26) are listed in Table S2; note, one or two children had positive IgE to some of these components, and for 3 components (Asp f 1, Bla g 5, Hev b 5) there was no positive response in any subject at any age. Inactive components at each age are listed in Table S3.

Table S4 shows 24 components which dropped out (not necessarily permanently), and number of children who were sensitized to these components. Figure S2 shows detailed longitudinal response profiles of each component that ever becomes inactive after first becoming active, for each child who has ever responded; for 12 components, we linked their drop-out to the resolution of sensitization (Figure S2a), and for the remaining 12 to the absence at subsequent follow-up of previously sensitized subjects (Figure S2b).

Component clusters at each time point and their longitudinal flow

Table 2 shows the number of component clusters inferred at each time point, and their posterior probabilities determined using Bayesian inference. The optimal solution identified one sensitization cluster at age one, 3 at age three, 4 at ages five and eight, 5 at age 11, and 6 at

age 16 years. The posterior probabilities for the most probable number of clusters were at least 0.87 for the first five time points, and remained above 0.70 at age 16 years. Tables S5-S10 list components in each cluster at each time point.

We qualitatively labelled clusters at each age based on the profile of allergen components to which sensitization occurred. Figure 1 shows the number of active components contained within each cluster for each time point (red), how many components were inactive (blue), and how many components were shared between clusters at adjacent time points.

The “Broad” cluster comprising of components originating from multiple sources was the only cluster identified at every time point. Components forming this cluster differed at different ages; Table S11 shows 24 components which were only ever assigned to the “Broad” cluster.

From age three onwards, the “HDM” cluster formed and remained unchanged by age 16, consisting of four mite components (Der p 1-2, Der f 1-2). Also at age three, the “Grass” cluster emerged, consisting of a single component (Phl p 1; Table S6). This cluster absorbed an additional 3 grass components, as well as cat component Fel d 1 to form the “Grass/cat” cluster at age five (Table S7). The membership of this cluster remained unchanged at age eight, although Fel d 1 assignment probability was reduced from >0.95 at age five to 0.70 (Table S8). A further cluster that was shared across ages five and eight was the “*Alternaria*” cluster, comprising of only Alt a 1. At age 11, this component was reabsorbed by the “Broad” cluster, the only component to do so throughout this flow (Figure 1).

Two new clusters formed at age 11 years: the “Cat” cluster (comprising of Fel d 1) and the “PR-10/profilin” cluster (Table S9). The latter was composed solely of components which have moved from the “Broad” cluster at age 8. Additional grass components were absorbed from the “Broad” into the “Grass” cluster at age 11 years (Phl p 2 and Phl p 6). This cluster divided at age 16 into two: “PR-10” and “Profilin” (Table S10); other clusters remained unchanged at age 16.

Figure 2 shows the change of activity across all components, and their cluster membership during childhood. Not all 86 components that were ever active across the six time points were active at every point. The inactive components populate the nodes in the left-hand pathway of Figure 1. All 24 components which dropped-out (Table S4) were assigned only to the "Broad" cluster. Components from all other clusters remained active once they first became so.

Sensitization to component clusters and clinical outcomes

The frequencies of component cluster sensitization profiles at ages 5 and 16 are shown in Table S12. For children who were sensitized to at least one cluster at age 5, the most common response (*n*=42) was to the "Grass" cluster only. The confusion matrix in Table S13 displays the number of children who shared sensitization to the clusters at ages 5 and 16, for 255 children who had CRD data at both follow-ups. Of 62 children who were sensitized to "Broad" cluster at age 5, 53 went on to respond to "Grass" cluster at age 16, with 51 remaining sensitized to the "Broad" cluster as well.

Univariable analyses: Sensitization to any of the component clusters at ages 5 and 16 years was associated with a significantly higher risk of asthma, wheeze and rhinitis at age 16 (Figure S3). However, the associations differed at different ages. At age 16 years, we observed the highest risk of asthma in relation to contemporaneous sensitization to the "HDM" cluster (OR [95% CI]: 12.4 [4.2–36.8], *P*<0.001; Figure S3a), but the strongest associate of asthma in adolescence in relation to sensitization at age 5 years was conferred by sensitization to the "Grass/cat" cluster (OR [95% CI]: 10.0 [4.6–21.7], *P*<0.001; Figure S3b). Similarly, the risk of rhinitis was greatest for those sensitized to the "Profilin" cluster at age 16 (OR [95% CI]: 30.6 [14.9-62.9], *P*<0.001), but at age 5 years, the strongest associate of subsequent rhinitis was sensitization to the "Broad" cluster (OR [95% CI]: 7.0 [2.9–11.4], *P*<0.001).

Multiple logistic regression (Figure 3): In the analysis which evaluated the association between sensitization to component clusters at age 16 years with contemporaneous allergic diseases

(Figure 3a), only sensitization to the “HDM” cluster was associated with the increased risk of asthma and wheeze (OR [95% CI]: 2.6 [1.2–6.1], *P*<0.05, and 3.1 [1.5–6.5], *P*<0.01, respectively). When we extended the time frame to investigate the relationship between cluster sensitization at age 5 years and clinical outcomes at age 16 (Figure 3b, Table S14), there was no significant association between asthma and sensitization to “Broad” and “HDM” clusters, and the strongest risk of subsequent asthma was conferred by sensitization to the “Grass/cat” and “Alternaria” clusters (OR [95% CI]: 3.5 [1.6–7.4], *P*<0.01, and 3.1 [1.4-6.8], P=0.005, respectively). Similarly, the magnitude of risk for contemporaneous rhinitis was greatest among children sensitized to the “Profilin” cluster (OR [95% CI]: 5.0 [2.3-11.2], *P*<0.001), but at age 5 years, the strongest predictor of subsequent rhinitis was sensitization to the “Broad” cluster (OR [95% CI]: 4.2 [2.4–7.4], P<0.001).

DISCUSSION

We describe the architecture of the evolution of IgE responses to multiple allergen components throughout childhood, taking into account responses to more than 100 allergenic molecules. By applying novel machine learning techniques to CRD sensitization data from infancy to adolescence among children from a population-based birth cohort, we identified latent structure in the diversification of the IgE responses during childhood (Figures 1 and 2). Our comprehensive description of the patterns of IgE responses to multiple components from infancy to adolescence demonstrated that the timing of onset of specific patterns of sensitization may be one of the important indicators of the subsequent risk of allergic disease (Figure 3). While children were frequently sensitized to more than one cluster, sensitization to distinct clusters was associated with different clinical presentations, indicating that some sensitization patterns pose greater risk for the development of specific clinical symptoms than others.

One of the limitations of our study includes the lack of potentially important components which are not included on the ISAC chip, such as those from HDM and fungi (e.g. ISAC has 6/109 fungal allergens identified in IUIS). This may be one of the reasons why the "*Alternaria*" cluster contained only one component (Alt a 1). Of note, sensitization to this small cluster at age 5 years conferred a strong risk for asthma in later life. This is also of relevance to the "HDM" cluster, which was the only cluster to remain unchanged once it had formed at age three, with Der p 1 being the dominant component. A recent study which measured IgE response to a broader range of HDM allergens has shown that sensitization increases in breadth with respect to the number of recognized allergenic molecules during the first decade of life[20]. It is possible that we would have observed similar "epitope spreading" if we measured IgE to a greater number of HDM allergens.

We acknowledge that the number of sensitized children in early life was small (only 10/226 at age 1 year), and we cannot exclude the possibility that this may have introduced bias in our

277 analyses. However, we believe that presenting data at all ages is important to ascertain the life-
278 course perspective.

279 We were unable to determine the effect of partial or complete sensitization to each cluster, and
280 the relative importance of sensitization to specific "lead" component(s) compared the number of
281 components within each cluster. This question will need to be addressed in future studies. We
282 also acknowledge that our study population comes from a specific geographical area, and that
283 different component clusters may arise in areas with different patterns of allergen exposure, or
284 by using a more comprehensive allergen panel. Thus, different components may be informative
285 in a different geographical or analytical context.

286 Allergen-specific IgG may be important in modulating the consequences of Th2 immunity in IgE-
287 sensitized children[36,37]. However, exploring IgG responses and IgG/IgE ratios was beyond the
288 scope of the current study.

289 Our method identified cross-sectional sensitization patterns and their longitudinal trajectories. It
290 is of note that despite the increasing number of active components, the varying number of
291 participants, and the derivation of our clusters being independent at different time points, the
292 components allocated to clusters were strikingly consistent across time, and the assignment
293 probabilities were very high. Our finding that IgE reactivity diversifies in molecular heterogeneity,
294 and that component-specific IgE responses are assigned to a steadily diversifying set of
295 clusters, is consistent with the "molecular spreading" hypothesis[38], and indirectly supports our
296 findings which suggested the existence of multiple subtypes of allergic sensitization[39,40]. The
297 increasing number of component-specific IgEs to which individual patients are responding in
298 later childhood (polysensitization) is associated with increasing severity of allergic disease[18], but
299 may also indicate that the sensitization process has started earlier. Our data extend the
300 relatively broad concepts of "polysensitization" and "early sensitization" to demonstrate that for a
301 more precise ascertainment of future and current risk of allergic diseases, we need accurate

information about the specific patterns of sensitization to unique sets of allergenic molecule, as well as the timing of onset of sensitization.

Our results suggest that the timing of onset of specific sensitization patterns may be a key indicator of future risk, and that apparently similar cross-sectional profiles of component-specific IgE responses may have different clinical associations depending on the age at which they emerge. This expands upon our previous study in which we used a limited number of Timothy grass and HDM components, which described two grass pollen IgE trajectories ("Late onset" and "Early onset")[19]. Although the progression of IgE component responses over time was identical in the two trajectories, following the sequence of Phl p 1/5→Phl p 2/4/6→Phl p 7/11/12, their clinical associations were different. The "Early onset" trajectory (in which Phl p 1/5 IgE responses emerged in preschool age) was associated with asthma and multimorbidity, while the "Late onset" trajectory (in which the same component-specific IgE responses were first observed in the school-age) was associated with rhinitis[19]. At the time when we conducted previous analyses, limitations including computing power and available methodologies precluded us from investigating the developmental pathways across all 112 components. In the current study, a more complex structure emerged. This is highlighted by the emergence of "Grass/cat" cluster at age 5 years, in which allergenic proteins from diverse sources, and with a fundamentally different function, clustered together. Although it may appear counterintuitive that Fel d 1 should be in the same cluster as the Timothy and Bermuda grass components, the assignment probability for the cat component belonging to this cluster was very high (0.97). The response to this cluster was strongly associated with asthma at age 16 years (3.5-fold increase in risk). This may suggest that the latent structure of IgE component clusters is not only a reflection of the source of allergens, or the function of allergenic molecules (as suggested by one of our previous studies)[18], but that it may also be a marker of the underlying pathophysiological processes leading to the development of distinct clinical phenotypes. Thus,

one possible reason why cat and grass components clustered together in 5-year old children from our area may be due to the IgE responses to these components foreshadowing the pathophysiological pathway leading to asthma (although we acknowledge that these IgE responses do not necessarily have to be causal).

In conclusion, different patterns of IgE responses to multiple allergen components evolve throughout childhood, and can be uncovered using machine learning. Specific sensitization patterns in early childhood are predictive of distinct allergic phenotypes in adolescence. Better resolution of longitudinal patterns may contribute to a better understanding of the pathophysiological processes giving raise to different allergic diseases, and may facilitate the development of diagnostic algorithms, which can be used for the prediction of current and future risk.

REFERENCES

1. Arshad SH, Kurukulaaratchy RJ, Fenn M, Matthews S. Early life risk factors for current wheeze, asthma, and bronchial hyperresponsiveness at 10 years of age. *Chest* 2005; **127**(2): 502-8.

2. Simpson BM, Custovic A, Simpson A, et al. NAC Manchester Asthma and Allergy Study (NACMAAS): risk factors for asthma and allergic disorders in adults. *Clin Exp Allergy* 2001; **31**(3): 391-9.

3. Sly PD, Boner AL, Bjorksten B, et al. Early identification of atopy in the prediction of persistent asthma in children. *Lancet* 2008; **372**(9643): 1100-6.

4. Carroll WD, Lenney W, Child F, et al. Asthma severity and atopy: how clear is the relationship? *Arch Dis Child* 2006; **91**(5): 405-9.

5. Weinmayr G, Weiland SK, Bjorksten B, et al. Atopic sensitization and the international variation of asthma symptom prevalence in children. *Am J Respir Crit Care Med* 2007; **176**(6): 565-74.

6. Dreborg S, Frew A. Position Paper - Allergen Standardization and Skin-Tests. *Allergy* 1993; **48**(14): 49-82.

7. Johansson SG, Bieber T, Dahl R, et al. Revised nomenclature for allergy for global use: Report of the Nomenclature Review Committee of the World Allergy Organization, October 2003. *J Allergy Clin Immunol* 2004; **113**(5): 832-6.

8. Custovic A, Lazic N, Simpson A. Pediatric asthma and development of atopy. *Curr Opin Allergy Clin Immunol* 2013; **13**(2): 173-80.

9. Nicolaou N, Custovic A. Molecular diagnosis of peanut and legume allergy. *Curr Opin Allergy Clin Immunol* 2011; **11**(3): 222-8.

10. Gould HJ, Sutton BJ. IgE in allergy and asthma today. *Nat Rev Immunol* 2008; **8**(3): 205-17.

11. Nicolaou N, Poorafshar M, Murray C, et al. Allergy or tolerance in children sensitized to peanut: prevalence and differentiation using component-resolved diagnostics. *J Allergy Clin Immunol* 2010; **125**(1): 191-7 e1-13.

12. Nicolaou N, Murray C, Belgrave D, Poorafshar M, Simpson A, Custovic A. Quantification of specific IgE to whole peanut extract and peanut components in prediction of peanut allergy. *J Allergy Clin Immunol* 2011; **127**(3): 684-5.

13. Treudler R, Simon JC. Overview of component resolved diagnostics. *Curr Allergy Asthma Rep* 2013; **13**(1): 110-7.

14. Patelis A, Borres MP, Kober A, Berthold M. Multiplex component-based allergen microarray in recent clinical studies. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* 2016; **46**(8): 1022-32.

15. Valenta R, Lidholm J, Niederberger V, Hayek B, Kraft D, Gronlund H. The recombinant allergen-based concept of component-resolved diagnostics and immunotherapy (CRD and CRIT). *Clin Exp Allergy* 1999; **29**(7): 896-904.

16. Melioli G, Bonifazi F, Bonini S, et al. The ImmunoCAP ISAC molecular allergology approach in adult multi-sensitized Italian patients with respiratory symptoms. *Clin Biochem* 2011; **44**(12): 1005-11.

17. Prosperi MC, Belgrave D, Buchan I, Simpson A, Custovic A. Challenges in interpreting allergen microarrays in relation to clinical symptoms: a machine learning approach. *Pediatr Allergy Immunol* 2014; **25**(1): 71-9.

18. Simpson A, Lazic N, Belgrave DC, et al. Patterns of IgE responses to multiple allergen components and clinical symptoms at age 11 years. *J Allergy Clin Immunol* 2015; **136**(5): 1224-31.

19. Custovic A, Sonntag HJ, Buchan IE, Belgrave D, Simpson A, Prosperi MC. Evolution pathways of IgE responses to grass and mite allergens throughout childhood. *J Allergy Clin Immunol* 2015; **136**(6): 1645-52 e1-8.

20. Posa D, Perna S, Resch Y, et al. Evolution and predictive value of IgE responses toward a comprehensive panel of house dust mite allergens during the first 2 decades of life. *J Allergy Clin Immunol* 2016.

21. Bishop CM. Model-based machine learning. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2013; **371**(1984).

22. Prosperi MC, Sahiner UM, Belgrave D, et al. Challenges in identifying asthma subgroups using unsupervised statistical learning techniques. *Am J Respir Crit Care Med* 2013; **188**(11): 1303-12.

23. Belgrave D, Henderson J, Simpson A, Buchan I, Bishop C, Custovic A. Disaggregating Asthma: Big Investigation vs. Big Data. *J Allergy Clin Immunol* 2016.

24. Custovic A, Simpson BM, Murray CS, et al. The National Asthma Campaign Manchester Asthma and Allergy Study. *Pediatr Allergy Immunol* 2002; **13 Suppl 15**: 32-7.

25. Belgrave DC, Simpson A, Semic-Jusufagic A, et al. Joint modeling of parentally reported and physician-confirmed wheeze identifies children with persistent troublesome wheezing. *J Allergy Clin Immunol* 2013; **132**(3): 575-83 e12.

26. Carlsen KCL, Roll S, Carlsen KH, et al. Does Pet Ownership in Infancy Lead to Asthma or Allergy at School Age? Pooled Analysis of Individual Participant Data from 11 European Birth Cohorts. *Plos One* 2012; **7**(8).

27. Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* 2004; **20**(3): 407-15.

28. Geyer CJ. Markov-Chain Monte-Carlo Maximum-Likelihood. *Computing Science and Statistics* 1991: 156-63.

29. Geyer CJ, Thompson EA. Annealing Markov-Chain Monte-Carlo with Applications to Ancestral Inference. *J Am Stat Assoc* 1995; **90**(431): 909-20.

30. Papastamoulis P, Iliopoulos G. An Artificial Allocations Based Solution to the Label Switching Problem in Bayesian Analysis of Mixtures of Distributions. *Journal of Computational and Graphical Statistics* 2010; **19**(2): 313-31.

31. Papastamoulis P, Iliopoulos G. On the Convergence Rate of Random Permutation Sampler and ECR Algorithm in Missing Data Models. *Methodology and Computing in Applied Probability* 2013; **15**(2): 293-304.

32. Papastamoulis P. Handling the Label Switching Problem in Latent Class Models Via the ECR Algorithm. *Communications in Statistics - Simulation and Computation* 2014; **43**(4): 913-27.

33. Redner RA, Walker HF. Mixture Densities, Maximum-Likelihood and the Em Algorithm. *Siam Rev* 1984; **26**(2): 195-237.

34. Papastamoulis P. label.switching: An R Package for Dealing with the Label Switching Problem in MCMC Outputs. *Journal of Statistical Software* 2016; **69**(C1): 1-24.

35. Papastamoulis P, Rattray M. BayesBinMix: an R package for Model Based Clustering of Multivariate Binary Data. *The R Journal* 2017**: 9**(1): 403-420.**.**

36. Holt PG, Strickland D, Bosco A, et al. Distinguishing benign from pathologic TH2 immunity in atopic children. *J Allergy Clin Immunol* 2016; **137**(2): 379-87.

37. Schwarz A, Panetta V, Cappella A, et al. IgG and IgG4 to 91 allergenic molecules in early childhood by route of exposure and current and future IgE sensitization: Results from the Multicentre Allergy Study birth cohort. *J Allergy Clin Immunol* 2016; **138**(5): 1426-33 e12.

38. Hatzler L, Panetta V, Lau S, et al. Molecular spreading and predictive value of preclinical IgE response to Phleum pratense in children with hay fever. *J Allergy Clin Immunol* 2012; **130**(4): 894-901 e5.

39. Lazic N, Roberts G, Custovic A, et al. Multiple atopy phenotypes and their associations with asthma: similar findings from two birth cohorts. *Allergy* 2013; **68**(6): 764-70.

40. Simpson A, Tan VY, Winn J, et al. Beyond atopy: multiple patterns of sensitization in relation to asthma in a birth cohort study. *Am J Respir Crit Care Med* 2010; **181**(11): 1200-6.

LEGEND FOR FIGURES

Figure 1. Clustering active IgE components throughout childhood. Cluster membership was determined using a Bernoulli Mixture Model applied to binarized sensitization data from all subjects.

Figure 2. The change of activity across all components, and their cluster membership during childhood

a) Individual allergen component activity at each age, black for active, grey for inactive;

b) Colour-coded by cluster membership; blank if inactive at a time point. Allergen components are sorted according to the time point of first activity, then by total number of time points active at, then by cluster membership, and finally based on persistence i.e. do the components remain active after first becoming so. Exceptions are the components that are active at only one time point which appear at the bottom.

Figure 3. Odds ratios and 95% CIs from multiple logistic regression, for asthma and rhinitis at age 16 based on subjects' reduced responses to (a) component clusters at age 16; (b) component clusters at age five.

LEGEND FOR TABLES

Table1. Demographic characteristics of the study population at each time point, and differences between children included and excluded from the analysis

Table 2. Inference of the number of component clusters at each time point. The posterior probability of the number of clusters, K, was determined through Bayesian inference with a Bernoulli Mixture Model applied to binarized sensitization data from all subjects.

The most probable K for each time point is highlighted in bold.

Table 1

Clinical Variable		CRD data for ages: 1 n (%)	3 n (%)	5 n (%)	8 n (%)	11 n (%)	16 n (%)
Overall	**Included**	226/1184 (19.09)	248/1184 (20.95)	588/1184 (49.66)	543/1184 (45.86)	461/1184 (38.94)	361/1184 (30.49)
	Excluded	958/1184 (80.91)	936/1184 (79.05)	596/1184 (50.34)	641/1884 (54.14)	723/1184 (61.06)	823/1184 (69.51)
Gender (Male)	**Included**	120/225 (53.33)	140/248 (56.45)	321/582 (55.15)	285/539 (52.88)	255/461 (55.31)	229/361 (63.43)
	Excluded	522/959 (54.43)	502/936 (53.63)	321/602 (53.32)	357/645 (55.35)	387/723 (53.53)	413/823 (50.18)
	p-value	0.82	0.47	0.57	0.43	0.59	**<0.001**
Older siblings	**Included**	114/204 (55.88)	136/239 (56.90)	332/572 (58.04)	304/539 (56.40)	254/458 (55.46)	191/359 (53.20)
	Excluded	485/872 (55.62)	463/837 (55.32)	267/504 (52.98)	295/537 (54.93)	345/618 (55.83)	408/717 (56.90)
	p-value	1	0.72	0.11	0.67	0.95	0.28
Maternal Asthma	**Included**	47/225 (20.89)	43/248 (17.34)	82/581 (14.11)	73/537 (13.59)	55/461 (11.93)	49/361 (13.57)
	Excluded	125/954 (13.10)	129/931 (13.86)	90/598 (15.05)	99/642(15.42)	117/718 (16.30)	123/818 (15.04)
	p-value	**0.004**	0.2	0.71	0.42	**0.047**	0.57
Maternal Smoking (during pregnancy)	**Included**	34/225 (15.11)	38/248 (15.32)	84/579 (14.51)	69/537 (12.85)	59/460 (12.83)	36/359 (10.03)
	Excluded	140/952 (14.71)	136/929 (14.64)	90/598 (15.05)	105/640 (16.41)	115/717 (16.04)	138/818 (16.87)
	p-value	0.96	0.87	0.86	0.1	0.15	**0.003**
Paternal Asthma	**Included**	20/225 (8.89)	31/248 (12.50)	43/579 (7.43)	43/537 (8.01)	41/458 (8.95)	26/358 (7.26)
	Excluded	65/951 (6.83)	54/928 (5.82)	42/597 (7.04)	42/639 (6.57)	44/718 (6.13)	59/818 (7.21)
	p-value	0.35	**<0.001**	0.88	0.4	0.09	1
Maternal Atopy	**Included**	174/225 (77.33)	190/248 (76.61)	330/566 (58.30)	295/526 (56.08)	256/450 (56.89)	211/352 (59.94)
	Excluded	508/921 (55.16)	492/898 (54.79)	352/580 (60.69)	387/620 (62.42)	426/696 (61.21)	471/794 (59.32)
	p-value	**<0.001**	**<0.001**	0.45	**0.034**	0.16	0.89
Paternal Atopy	**Included**	174/225 (77.33)	190/248 (76.61)	358/562 (63.70)	331/522 (63.41)	277/448 (61.83)	222/352 (63.07)
	Excluded	542/912 (59.43)	526/889 (59.17)	358/575 (62.26)	385/615 (62.60)	439/689 (63.72)	494/785 (62.93)
	p-value	**<0.001**	**<0.001**	0.66	0.83	0.56	1
Current Asthma (Age 16)	**Included**	19/151 (12.58)	23/168 (13.69)	50/413 (12.11)	42/407 (10.32)	43/377 (11.41)	41/351 (11.68)
	Excluded	71/585 (12.14)	67/568 (11.80)	40/323 (12.38)	48/329 (14.59)	47/359 (13.09)	49/385 (12.73)
	p-value	0.99	0.60	1.00	0.10	0.56	0.75
Current Wheeze (Age 16)	**Included**	25/149 (16.78)	31/167 (18.56)	71/413 (17.19)	62/405 (15.31)	66/382 (17.28)	55/354 (15.54)
	Excluded	102/590 (17.29)	96/572 (16.78)	56/326 (17.18)	65/334 (19.46)	61/357 (17.09)	72/385 (18.70)
	p-value	0.98	0.67	1	0.16	1	0.30
Current Rhinitis (Age 16)	**Included**	67/150 (44.67)	79/168 (47.02)	169/417 (40.53)	155/405 (38.27)	154/383 (40.21)	146/357 (40.90)
	Excluded	242/594 (40.74)	230/576 (39.93)	140/327 (42.81)	154/339 (45.43)	155/361 (42.94)	163/387 (42.12)
	p-value	0.44	0.12	0.58	0.06	0.50	0.79
Asthma Medication (Age 16)	**Included**	24/151 (15.89)	27/169 (15.98)	74/420 (17.62)	62/411 (15.09)	63/386 (16.32)	56/359 (15.60)
	Excluded	104/600 (17.33)	101/582 (17.35)	54/331 (16.31)	66/340 (19.41)	65/365 (17.81)	72/392 (18.37)
	p-value	0.76	0.76	0.71	0.14	0.66	0.36
Asthma Ever (Age 16)	**Included**	37/147 (25.17)	44/165 (26.67)	124/411 (30.17)	109/398 (27.39)	113/376 (30.05)	102/348 (29.31)
	Excluded	183/584 (31.34)	176/566 (31.10)	96/320 (30.00)	111/333 (33.33)	107/355 (30.14)	118/383 (30.81)
	p-value	0.18	0.32	1	0.10	1	0.72
FEV1/FVC Ratio (Age 16)	**Included**	88.41 (n = 131)	88.39 (n = 150)	88.04 (n = 372)	87.87 (n = 355)	88.16 (n = 356)	88.01 (n = 355)
	Excluded	88.04 (n = 498)	88.03 (n = 479)	88.23 (n = 257)	88.43 (n = 274)	88.06 (n = 273)	88.25 (n = 274)
	p-value	0.59	0.59	0.75	0.77	0.51	0.44
Sensitization (skin prick test)	**Included**	23/222 (10.36)	64/245 (26.12)	177/572 (30.94)	155/531 (29.19)	164/455 (36.04)	188/341 (55.13)
	Excluded	33/282 (11.70)	161/738 (21.82)	117/391 (29.92)	159/396 (40.15)	116/340 (34.12)	134/259 (51.74)
	p-value	0.74	0.19	0.79	**<0.001**	0.63	0.46
Taken at the respective ages							

Table 2

		p(K), where K = …							
Age	**K_{max}**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	9	**0.8958**	0.0932	0.0106	0.0004	0	0	0	0
3	25	0.0004	0.0208	**0.8784**	0.0942	0.0062	0	0	0
5	25	0.0004	0	0.0012	**0.9548**	0.0426	0.0010	0	0
8	25	0.0004	0	0	**0.9440**	0.0532	0.0024	0	0
11	25	0	0.0004	0.0012	0.0032	**0.9416**	0.0516	0.0012	0.0008
16	25	0	0.0004	0	0.0000	0.2536	**0.7066**	0.0358	0.0036

Figure 1

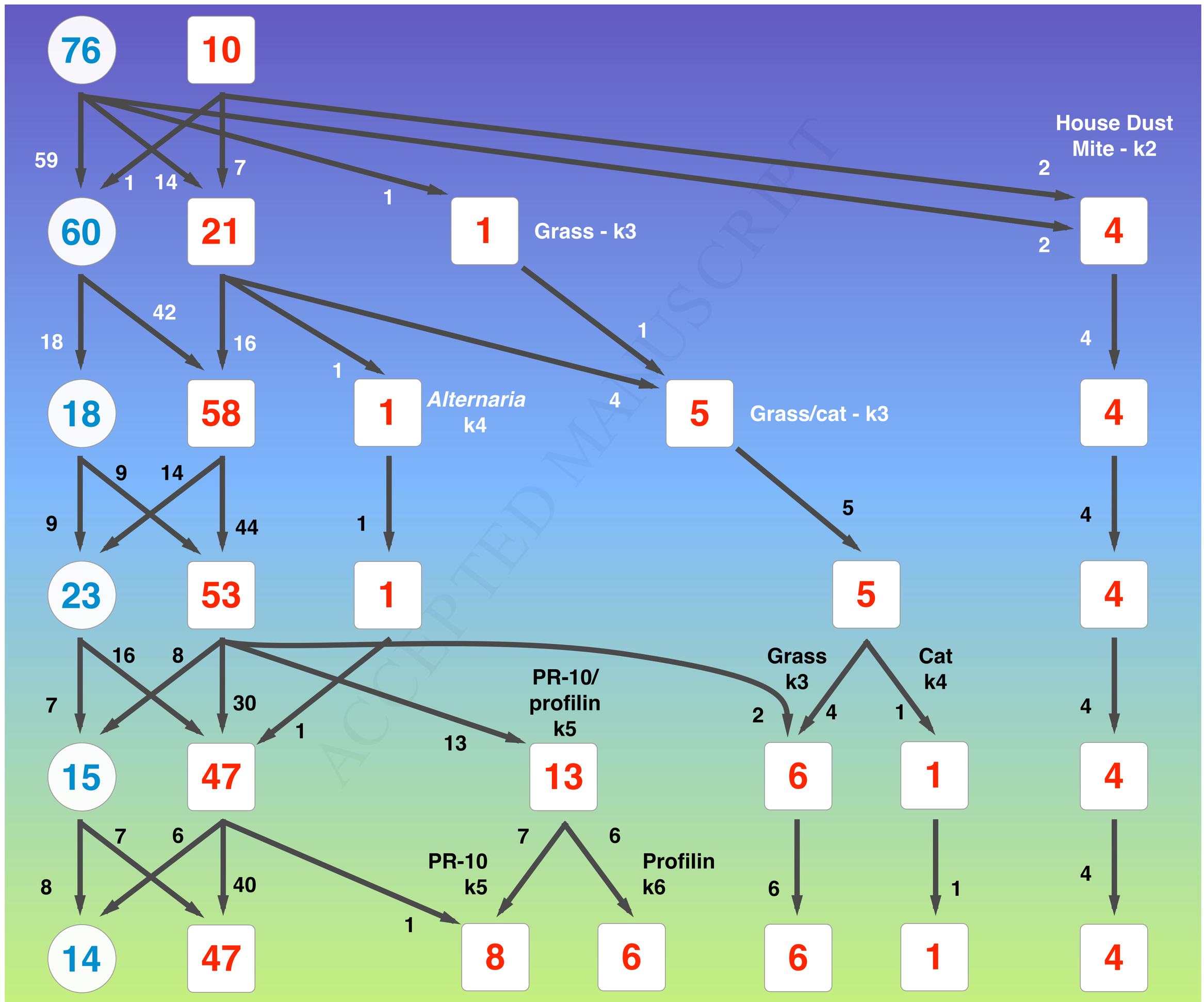


Figure 2

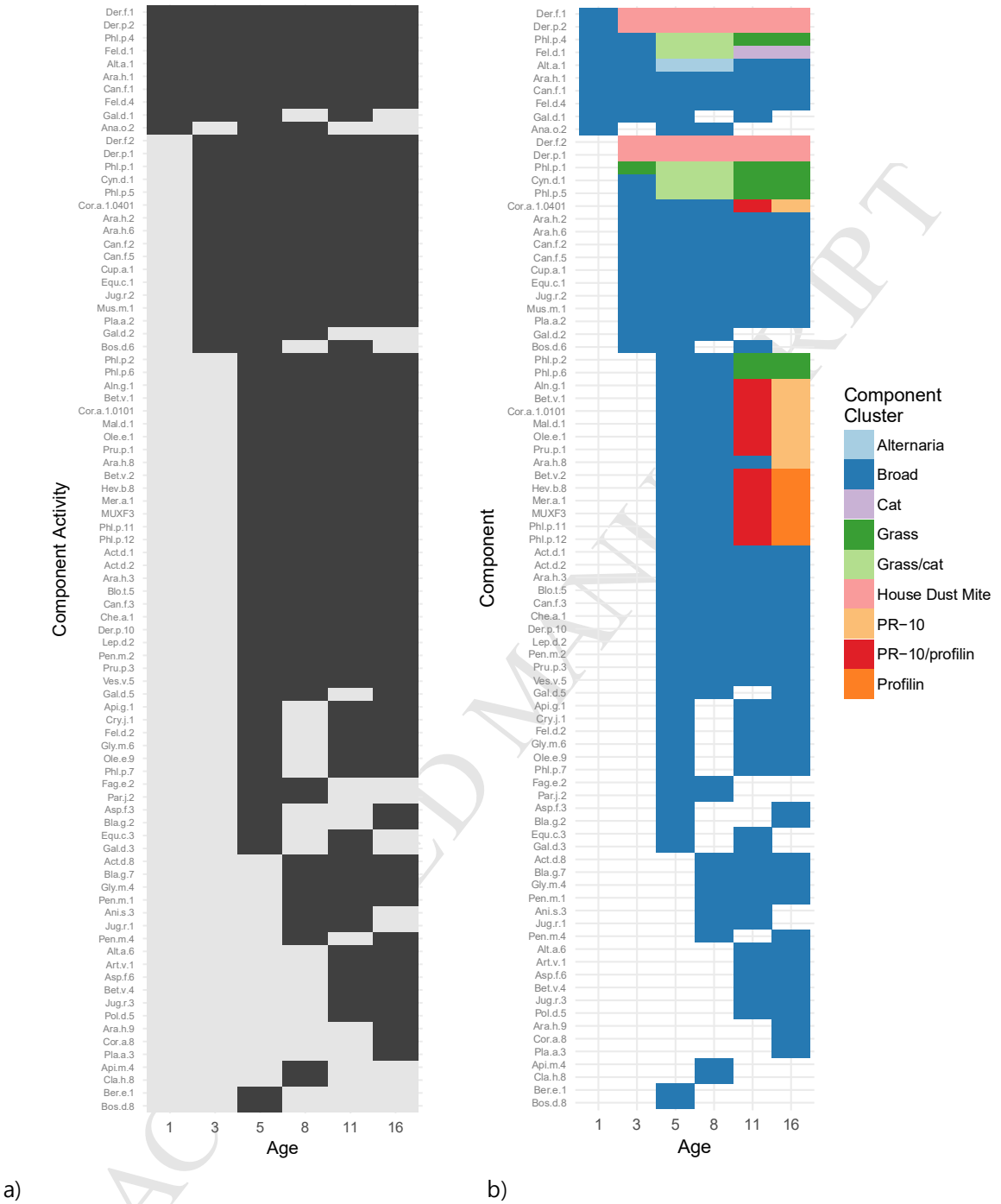
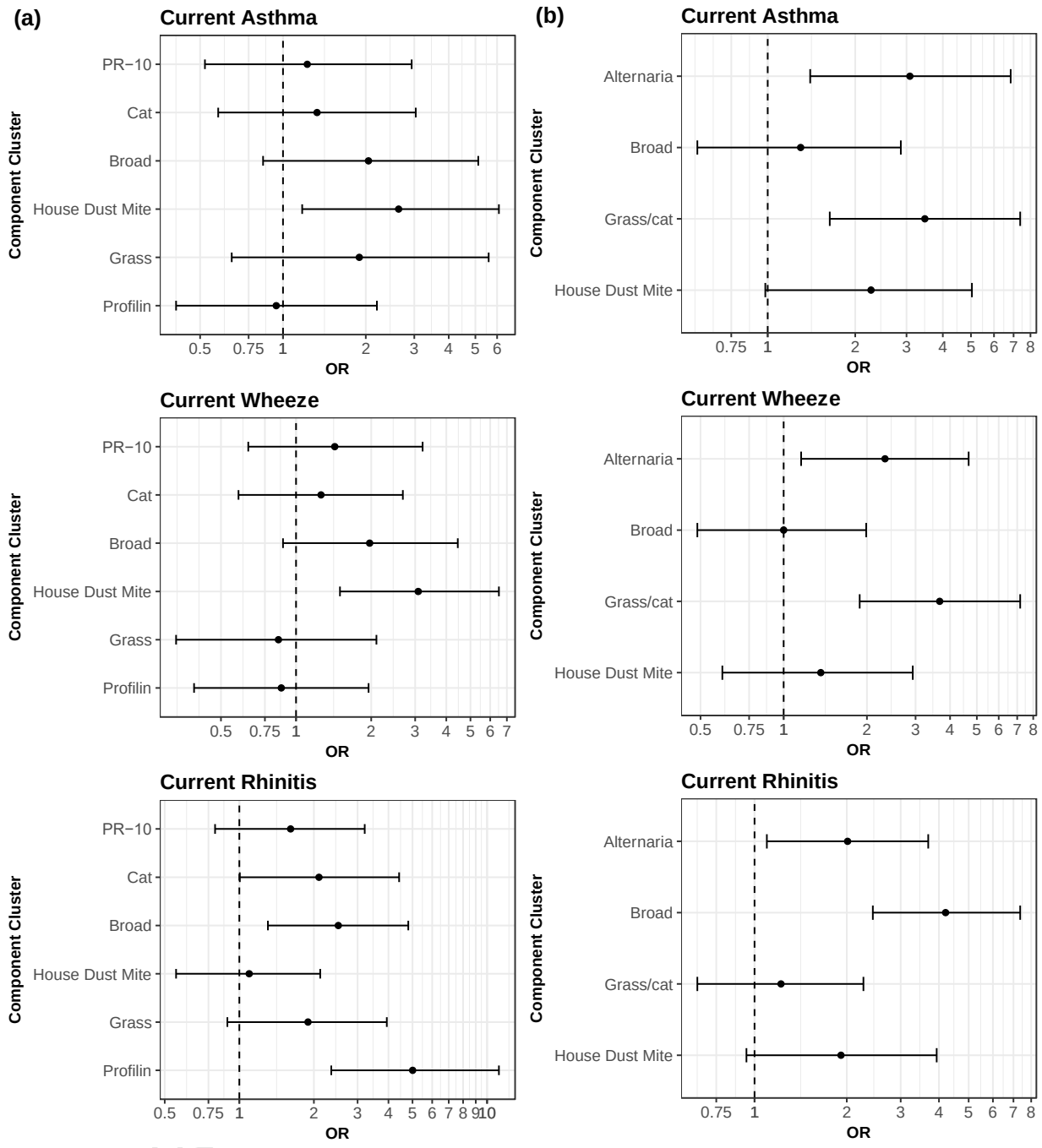


Figure No. 3



Evolution of IgE responses to multiple allergen components throughout childhood

Rebecca Howard MSc[1], Danielle Belgrave PhD[2], Panagiotis Papastamoulis PhD[1], Angela Simpson MD PhD[3], Magnus Rattray PhD[1], Adnan Custovic MD PhD[2]

1. Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK.

2. Section of Paediatrics, Department of Medicine, Imperial College London, UK.

3. Division of Infection, Immunity and Respiratory Medicine, University of Manchester and University Hospital of South Manchester, Manchester Academic Health Sciences Centre, Manchester, United Kingdom.

ONLINE DATA SUPPLEMENT

METHODS

Screening & Recruitment

All pregnant women were screened for eligibility at antenatal visits (8^{th}-10^{th} week of pregnancy). Of the 1499 couples who met the inclusion criteria ($\leq$10 weeks of pregnancy, maternal age $\geq$18 years), 288 declined to take part and 27 were lost to follow-up between recruitment and birth of a child. A total of 1184 participants had some evaluable data.

Follow-up

Children have been followed prospectively, and attended review clinics at ages 1, 3, 5, 8, 11 and 16 years. At age 1 year, only children with either both atopic parents, or no atopic parents were invited to attend for clinical follow up. At all other time points for all other measures all children were invited to participate.

Statistical grouping of allergen components

We assumed that there exist clusters of components to which subjects have similar IgE responses (i.e., either being sensitized or not to most of the components within the same cluster). At each age, we inferred component clusters by clustering the data through Bayesian estimation of a mixture of Bernoulli distributions (Bernoulli Mixture Model–BMM). The BMM method provides a fully Bayesian method which can effectively deal with missing data and an unknown number of clusters. This method was shown to achieve a much better estimation of the number of clusters compared to the EM implementation contained within the FlexMix() R package (for detailed description of the methodology and benchmarking, please see Panagiotis Papastamoulis and Magnus Rattray , The R Journal (2017) 9:1, pages 403-420.; (URL: https://journal.r-project.org/archive/2017/RJ-2017-022/index.html). Specifically, once the total number of clusters exceeds 4, our Bayesian method does not underestimate the correct number of clusters, unlike the EM implementation.

The observed likelihood for a binary data matrix ***x*** under the BMM model is given by:

$$L_K(\boldsymbol{p}, \boldsymbol{\theta}; \boldsymbol{x}) = \prod_{i=1}^{n} \sum_{k=1}^{K} p_k \prod_{j=1}^{d} \theta_{kj}^{x_{ij}} \left(1 - \theta_{kj}\right)^{1-x_{ij}}$$

where θ_{jk} is between 0 and 1 and represents the frequency of sensitisation to component *j* for subjects in cluster *k* and p_k represents the weight of cluster *k* which is the prior probability that a subject belongs to that cluster.

Associations with clinical outcomes

The relationships between a subject's responses to the BMM's cluster output and their disease outcomes were assessed using univariable and multiple logistic regression analyses (adjusting for sensitization to each of the component clusters, and the sex of the child). In addition to frequentist intervals, we calculated Bayesian posterior credible regions.

RESULTS

Table S1. Number of children with component-resolved diagnostics data and proportion of those with at least one positive allergen component response at each follow-up

Table S2. The list of 26 components which were labelled inactive at all 6 time points

Table S3. Components labelled inactive for ages a) 1, b) 3, c) 5, d) 8, e) 11, and f) 16 years.

Table S4. Subject response totals for each of the allergen components which "dropped-out" (i.e. become inactive after first being active). Italics indicates when that component is "Active"; bold if "inactive" but 1 or 2 subjects have positively responded to that component at that time point; otherwise if "inactive" and no subjects have positively responded at that time.

Table S5. Age 1's component cluster members and the number of children that respond to each member.

Table S6. Age 3's component cluster members, number of children that respond to each member, and the assignment probability to the most probable cluster.

Table S7: Age 5's component cluster members, number of children that respond to each member, and the assignment probability to the most probable cluster.

Table S8: Age 8's component cluster members, number of children that respond to each member, and the assignment probability to the most probable cluster.

Table S9: Age 11's component cluster members, number of children that respond to each member, and the assignment probability to the most probable cluster.

Table S10: Age 16's component cluster members, number of children that respond to each member, and the assignment probability to the most probable cluster; Age 16, K = 6

Table S11. Components which were only ever assigned to the "Broad" cluster.

Table S12. Frequencies for each subject's reduced response to the component clusters found at **(a)** age 5, and **(b)** age 16.

Table S13. Confusion matrix for the reduced response frequencies of the 255 children that had ISAC data for both ages 5 and 16. The clusters from age 5: the left, the clusters from age 16: the top. Note that rows and columns do not sum to the totals, as responses to the clusters are not mutually exclusive. Note the relatively small proportion of children that have reduced responses to each of the clusters at each of these ages (but particularly at age 5), acting as a main source for a wide range in confidence intervals for associations with clinical outcomes.

Table S14. C-statistic reported for each of the multivariate logistic regression models applied to both age 5 and age 16's cluster response data, with relation to rhinitis and asthma-related clinical outcomes at age 16.

Figure S1. CONSORT diagram for participant flow

Figure S2. Response profiles for each component that ever becomes inactive after first becoming active, for each child that ever responds to each of these components. Darker blue fill for when the component is active at that time point i.e. at least three subjects had a positive response to that component at that time point. Points represent when ISAC data is available for that child. Response: positive (red) or negative (blue).

a) Components whose drop-out can be ascribed to desensitisation *i.e.* if all children who were positive to a component at the time point preceding the drop-out were still positive.
b) The remaining 12 components, all of whose drop-outs can be explained by the subject loss to follow-up.

Figure S3. Odds ratios and 95% CIs from univariable logistic regression for asthma and rhinitis at age 16, based on subjects' reduced responses to component clusters at **(a)** age 16; **(b)** age 5. Bayesian posterior credible regions were also computed, and agreed closely with the 95% CIs shown.

Age (y)	Number of children tested	Children responding to ≥1 active components (%)
1	226	43 (19.0%)
3	248	88 (35.5%)
5	588	253 (43.0%)
8	543	256 (47.1%)
11	461	220 (47.7%)
16	361	207 (57.3%)

Table S2

Perennially inactive components
Act.d.5
Amb.a.1
Ani.s.1
Api.m.1
Art.v.3
Asp.f.1
Bla.g.1
Bla.g.5
Bos.d.4
Bos.d.5
Bos.d.lactoferrin
Cor.a.9
Gad.c.1
Gly.m.5
Hev.b.1
Hev.b.3
Hev.b.5
Hev.b.6.01
Ole.e.7
Pla.a.1
Pla.l.1
Sal.k.1
Ses.i.1
Tri.a.14
Tri.a.19.0101
Tri.a.aA_TI

Table S3a

Age 1				
	No children positive		**1 child positive**	**2 children positive**
Act.d.1	Bos.d.8	Mer.a.1	Ara.h.6	Ara.h.2
Act.d.2	Bos.d.lactoferrin	Mus.m.1	Bos.d.6	Bla.g.7
Act.d.5	Can.f.3	MUXF3	Can.f.2	Der.f.2
Act.d.8	Can.f.5	Ole.e.1	Cor.a.1.0401	Der.p.1
Aln.g.1	Che.a.1	Ole.e.7	Gal.d.3	Gal.d.2
Alt.a.6	Cla.h.8	Ole.e.9	Gly.m.5	Gly.m.6
Amb.a.1	Cor.a.1.0101	Par.j.2	Pen.m.1	
Ani.s.1	Cor.a.8	Pen.m.2		
Ani.s.3	Cor.a.9	Pen.m.4		
Api.g.1	Cry.j.1	Phl.p.1		
Api.m.1	Cup.a.1	Phl.p.11		
Api.m.4	Cyn.d.1	Phl.p.12		
Ara.h.3	Der.p.10	Phl.p.2		
Ara.h.8	Equ.c.1	Phl.p.5		
Ara.h.9	Equ.c.3	Phl.p.6		
Art.v.1	Fag.e.2	Phl.p.7		
Art.v.3	Fel.d.2	Pla.a.1		
Asp.f.1	Gad.c.1	Pla.a.2		
Asp.f.3	Gal.d.5	Pla.a.3		
Asp.f.6	Gly.m.4	Pla.l.1		
Ber.e.1	Hev.b.1	Pol.d.5		
Bet.v.1	Hev.b.3	Pru.p.1		
Bet.v.2	Hev.b.5	Pru.p.3		
Bet.v.4	Hev.b.6.01	Sal.k.1		
Bla.g.1	Hev.b.8	Ses.i.1		
Bla.g.2	Jug.r.1	Tri.a.14		
Bla.g.5	Jug.r.2	Tri.a.19.0101		
Blo.t.5	Jug.r.3	Tri.a.aA_TI		
Bos.d.4	Lep.d.2	Ves.v.5		
Bos.d.5	Mal.d.1			

Table S3b

Age 3			
No children positive		**1 child positive**	**2 children positive**
Act.d.2	Der.p.10	Act.d.1	Bet.v.1
Act.d.5	Gad.c.1	Aln.g.1	Bos.d.lactoferrin
Act.d.8	Gal.d.5	Ana.o.2	Can.f.3
Alt.a.6	Gly.m.4	Ani.s.1	Cor.a.9
Amb.a.1	Hev.b.1	Ara.h.3	Cry.j.1
Ani.s.3	Hev.b.3	Bla.g.2	Gal.d.3
Api.g.1	Hev.b.5	Bos.d.4	Hev.b.8
Api.m.1	Hev.b.6.01	Bos.d.5	MUXF3
Api.m.4	Jug.r.3	Bos.d.8	Phl.p.11
Ara.h.8	Lep.d.2	Equ.c.3	Phl.p.6
Ara.h.9	Mer.a.1	Fag.e.2	Pol.d.5
Art.v.1	Ole.e.1	Fel.d.2	
Art.v.3	Ole.e.7	Gly.m.5	
Asp.f.1	Par.j.2	Gly.m.6	
Asp.f.3	Pen.m.1	Jug.r.1	
Asp.f.6	Phl.p.12	Mal.d.1	
Ber.e.1	Phl.p.7	Ole.e.9	
Bet.v.2	Pla.a.1	Pen.m.2	
Bet.v.4	Pla.a.3	Pen.m.4	
Bla.g.1	Pla.l.1	Phl.p.2	
Bla.g.5	Pru.p.3	Pru.p.1	
Bla.g.7	Sal.k.1		
Blo.t.5	Ses.i.1		
Che.a.1	Tri.a.14		
Cla.h.8	Tri.a.19.0101		
Cor.a.1.0101	Tri.a.aA_TI		
Cor.a.8	Ves.v.5		

Table S3c

Age 5		
No children positive	**1 child positive**	**2 children positive**
Act.d.5	Act.d.8	Ara.h.9
Alt.a.6	Ani.s.1	Bet.v.4
Amb.a.1	Ani.s.3	Bos.d.5
Api.m.1	Api.m.4	Gad.c.1
Art.v.1	Art.v.3	Gly.m.5
Asp.f.1	Asp.f.6	Jug.r.1
Bla.g.1	Bos.d.4	Pen.m.1
Bla.g.5	Bos.d.lactoferrin	Pen.m.4
Bla.g.7	Cla.h.8	Pol.d.5
Hev.b.1	Cor.a.8	
Hev.b.3	Cor.a.9	
Hev.b.5	Gly.m.4	
Hev.b.6.01	Jug.r.3	
Ole.e.7	Sal.k.1	
Pla.a.1	Tri.a.aA_TI	
Pla.a.3		
Pla.l.1		
Ses.i.1		
Tri.a.14		
Tri.a.19.0101		

Table S3d

Age 8		
No children positive	**1 child positive**	**2 children positive**
Act.d.5	Amb.a.1	Alt.a.6
Ani.s.1	Art.v.1	Api.g.1
Api.m.1	Ber.e.1	Ara.h.9
Art.v.3	Bla.g.1	Asp.f.3
Asp.f.1	Bos.d.4	Asp.f.6
Bla.g.5	Bos.d.6	Bet.v.4
Bos.d.5	Bos.d.8	Bla.g.2
Bos.d.lactoferrin	Cry.j.1	Cor.a.9
Cor.a.8	Equ.c.3	Gal.d.1
Gly.m.5	Fel.d.2	Gly.m.6
Hev.b.1	Gad.c.1	Ole.e.9
Hev.b.3	Gal.d.3	Phl.p.7
Hev.b.5	Jug.r.3	Sal.k.1
Hev.b.6.01	Pla.a.1	
Ole.e.7	Pla.a.3	
Ses.i.1	Pla.l.1	
Tri.a.19.0101	Pol.d.5	
Tri.a.aA_TI	Tri.a.14	

Table S3e

Age 11		
No children positive	**1 child positive**	**2 children positive**
Act.d.5	Ani.s.1	Ara.h.9
Amb.a.1	Asp.f.3	Art.v.3
Ana.o.2	Ber.e.1	Bos.d.4
Api.m.1	Bla.g.1	Bos.d.5
Api.m.4	Bos.d.lactoferrin	Bos.d.8
Asp.f.1	Cor.a.8	Gal.d.2
Bla.g.2	Cor.a.9	Gly.m.5
Bla.g.5	Gad.c.1	Hev.b.3
Cla.h.8	Hev.b.1	Hev.b.6.01
Fag.e.2	Ole.e.7	Pla.a.1
Gal.d.5	Par.j.2	Pla.a.3
Hev.b.5	Pen.m.4	Pla.l.1
Sal.k.1	Ses.i.1	Tri.a.19.0101
Tri.a.14	Tri.a.aA_TI	

Table S3f

Age 16		
No children positive	**1 child positive**	**2 children positive**
Amb.a.1	Act.d.5	Ani.s.3
Ani.s.1	Ana.o.2	Api.m.4
Asp.f.1	Api.m.1	Bos.d.4
Bla.g.5	Art.v.3	Bos.d.6
Bos.d.8	Ber.e.1	Cla.h.8
Fag.e.2	Bla.g.1	Cor.a.9
Hev.b.1	Bos.d.5	Equ.c.3
Hev.b.3	Bos.d.lactoferrin	Gal.d.1
Hev.b.5	Gad.c.1	Gal.d.3
Hev.b.6.01	Gal.d.2	Gly.m.5
Tri.a.14	Ole.e.7	Jug.r.1
	Sal.k.1	Par.j.2
	Ses.i.1	Pla.a.1
	Tri.a.19.0101	Pla.l.1
	Tri.a.aA_TI	

Table S4

	Age					
Component	**1**	**3**	**5**	**8**	**11**	**16**
Bos.d.8	0	**1**	*3*	**1**	**2**	0
Ber.e.1	0	0	*4*	**1**	**1**	**1**
Cla.h.8	0	0	**1**	*5*	0	**2**
Api.m.4	0	0	**1**	*3*	0	**2**
Pen.m.4	0	**1**	**2**	*10*	**1**	*3*
Jug.r.1	0	**1**	**2**	*3*	*4*	**2**
Ani.s.3	0	0	**1**	*3*	*7*	**2**
Gal.d.3	**1**	**2**	*4*	**1**	*13*	**2**
Equ.c.3	0	**1**	*3*	**1**	*5*	**2**
Bla.g.2	0	**1**	*3*	**2**	0	*3*
Asp.f.3	0	0	*3*	**2**	**1**	*4*
Par.j.2	0	0	*3*	*3*	**1**	**2**
Fag.e.2	0	**1**	*3*	*8*	0	0
Phl.p.7	0	0	*4*	**2**	*9*	*8*
Ole.e.9	0	**1**	*5*	**2**	*3*	*3*
Gly.m.6	**2**	**1**	*5*	**2**	*4*	*4*
Fel.d.2	0	**1**	*6*	**1**	*10*	*6*
Cry.j.1	0	**2**	*5*	**1**	*12*	*15*
Api.g.1	0	0	*7*	**2**	*7*	*10*
Gal.d.5	0	0	*9*	*8*	0	*7*
Bos.d.6	**1**	*3*	*3*	**1**	*4*	**2**
Gal.d.2	**2**	*3*	*6*	*4*	**2**	**1**
Ana.o.2	*3*	**1**	*8*	*5*	0	**1**
Gal.d.1	*6*	*4*	*12*	**2**	*3*	**2**

Table S5

k = 1 Broad	
Component	**Frequency**
Alt.a.1	11
Ana.o.2	3
Ara.h.1	5
Can.f.1	7
Der.f.1	3
Der.p.2	3
Fel.d.1	12
Fel.d.4	3
Gal.d.1	6
Phl.p.4	5

Table S6

k = 1 Broad					
Component	Frequency	Assignment Probability	Component	Frequency	Assignment Probability
Alt.a.1	18	1	Equ.c.1	7	1
Ara.h.1	6	1	Fel.d.1	24	0.999
Ara.h.2	8	1	Fel.d.4	4	1
Ara.h.6	7	1	Gal.d.1	4	1
Bos.d.6	3	1	Gal.d.2	3	1
Can.f.1	12	1	Jug.r.2	3	1
Can.f.2	4	1	Mus.m.1	3	1
Can.f.5	7	1	Phl.p.4	20	0.992
Cor.a.1.0401	5	1	Phl.p.5	7	1
Cup.a.1	3	1	Pla.a.2	7	1
Cyn.d.1	10	1			

k = 2 House Dust Mite			k = 3 Grass		
Component	Frequency	Assignment Probability	Component	Frequency	Assignment Probability
Der.f.1	22	1	Phl.p.1	29	0.999
Der.f.2	19	1			
Der.p.1	27	1			
Der.p.2	19	1			

Table S7

k = 1

Broad

Component	Frequency	Assignment Probability	Component	Frequency	Assignment Probability	Component	Frequency	Assignment Probability
Act.d.1	4	1	Can.f.3	6	1	Lep.d.2	18	1
Act.d.2	3	1	Can.f.5	9	1	Mal.d.1	9	1
Aln.g.1	16	1	Che.a.1	8	1	Mer.a.1	4	1
Ana.o.2	8	1	Cor.a.1.0101	7	1	Mus.m.1	6	1
Api.g.1	7	1	Cor.a.1.0401	32	1	MUXF3	13	1
Ara.h.1	14	1	Cry.j.1	5	1	Ole.e.1	10	1
Ara.h.2	11	1	Cup.a.1	10	1	Ole.e.9	5	1
Ara.h.3	4	1	Der.p.10	3	1	Par.j.2	3	1
Ara.h.6	16	1	Equ.c.1	20	1	Pen.m.2	4	1
Ara.h.8	4	1	Equ.c.3	3	1	Phl.p.11	8	1
Asp.f.3	3	1	Fag.e.2	3	1	Phl.p.12	4	1
Ber.e.1	4	1	Fel.d.2	6	1	Phl.p.2	27	1
Bet.v.1	24	1	Fel.d.4	18	1	Phl.p.6	34	1
Bet.v.2	5	1	Gal.d.1	12	1	Phl.p.7	4	1
Bla.g.2	3	1	Gal.d.2	6	1	Pla.a.2	15	1
Blo.t.5	5	1	Gal.d.3	4	1	Pru.p.1	7	1
Bos.d.6	3	1	Gal.d.5	9	1	Pru.p.3	5	1
Bos.d.8	3	1	Gly.m.6	5	1	Ves.v.5	11	1
Can.f.1	25	1	Hev.b.8	9	1			
Can.f.2	4	1	Jug.r.2	17	1			

k = 2

House Dust Mite

Component	Frequency	Assignment Probability
Der.f.1	75	1
Der.f.2	72	1
Der.p.1	82	1
Der.p.2	73	1

k = 3

Grass/cat

Component	Frequency	Assignment Probability
Cyn.d.1	66	1
Fel.d.1	83	0.971
Phl.p.1	104	1
Phl.p.4	89	1
Phl.p.5	71	1

k = 4

Alternaria

Component	Frequency	Assignment Probability
Alt.a.1	72	1

Table S8

k = 1 Broad								
Component	Frequency	Assignment Probability	Component	Frequency	Assignment Probability	Component	Frequency	Assignment Probability
Act.d.1	4	1	Can.f.3	5	1	Mal.d.1	19	1
Act.d.2	6	1	Can.f.5	6	1	Mer.a.1	9	1
Act.d.8	3	1	Che.a.1	3	1	Mus.m.1	4	1
Aln.g.1	23	1	Cla.h.8	5	1	MUXF3	5	1
Ana.o.2	5	1	Cor.a.1.0101	12	1	Ole.e.1	32	1
Ani.s.3	3	1	Cor.a.1.0401	35	1	Par.j.2	3	1
Api.m.4	3	1	Cup.a.1	7	1	Pen.m.1	3	1
Ara.h.1	8	1	Der.p.10	4	1	Pen.m.2	4	1
Ara.h.2	10	1	Equ.c.1	13	1	Pen.m.4	10	1
Ara.h.3	4	1	Fag.e.2	8	1	Phl.p.11	15	1
Ara.h.6	11	1	Fel.d.4	8	1	Phl.p.12	9	1
Ara.h.8	7	1	Gal.d.2	4	1	Phl.p.2	37	0.998
Bet.v.1	39	1	Gal.d.5	8	1	Phl.p.6	38	0.994
Bet.v.2	9	1	Gly.m.4	3	1	Pla.a.2	7	1
Bla.g.7	3	1	Hev.b.8	12	1	Pru.p.1	9	1
Blo.t.5	4	1	Jug.r.1	3	1	Pru.p.3	3	1
Can.f.1	31	1	Jug.r.2	8	1	Ves.v.5	7	1
Can.f.2	10	1	Lep.d.2	18	1			

k = 2 House Dust Mite			k = 3 Grass/cat			k = 4 *Alternaria*		
Component	Frequency	Assignment Probability	Component	Frequency	Assignment Probability	Component	Frequency	Assignment Probability
Der.f.1	75	1	Cyn.d.1	80	1	Alt.a.1	63	1
Der.f.2	73	1	Fel.d.1	71	0.696			
Der.p.1	80	1	Phl.p.1	118	1			
Der.p.2	77	1	Phl.p.4	95	1			
			Phl.p.5	87	1			

Table S9

k = 1
Broad

Component	Frequency	Assignment Probability	Component	Frequency	Assignment Probability	Component	Frequency	Assignment Probability
Act.d.1	5	1	Blo.t.5	9	1	Gly.m.4	14	1
Act.d.2	7	1	Bos.d.6	4	1	Gly.m.6	4	1
Act.d.8	3	1	Can.f.1	37	1	Jug.r.1	4	1
Alt.a.1	21	1	Can.f.2	9	1	Jug.r.2	16	1
Alt.a.6	3	1	Can.f.3	8	1	Jug.r.3	3	1
Ani.s.3	7	1	Can.f.5	23	1	Lep.d.2	35	1
Api.g.1	7	1	Che.a.1	16	1	Mus.m.1	11	1
Ara.h.1	14	1	Cry.j.1	12	1	Ole.e.9	3	1
Ara.h.2	20	1	Cup.a.1	26	0.870	Pen.m.1	7	1
Ara.h.3	9	1	Der.p.10	7	1	Pen.m.2	5	1
Ara.h.6	19	1	Equ.c.1	19	1	Phl.p.7	9	1
Ara.h.8	19	0.817	Equ.c.3	5	1	Pla.a.2	15	1
Art.v.1	4	1	Fel.d.2	10	1	Pol.d.5	3	1
Asp.f.6	3	1	Fel.d.4	20	1	Pru.p.3	4	1
Bet.v.4	5	1	Gal.d.1	3	1	Ves.v.5	5	1
Bla.g.7	7	1	Gal.d.3	13	1			

k = 2
House Dust Mite

Component	Frequency	Assignment Probability
Der.f.1	92	1
Der.f.2	99	1
Der.p.1	97	1
Der.p.2	95	1

k = 3
Grass

Component	Frequency	Assignment Probability
Cyn.d.1	115	1
Phl.p.1	150	1
Phl.p.2	83	1
Phl.p.4	106	1
Phl.p.5	125	1
Phl.p.6	74	1

k = 4
Cat

Component	Frequency	Assignment Probability
Fel.d.1	80	1

k = 5
PR-10/profilin

Component	Frequency	Assignment Probability
Aln.g.1	43	1
Bet.v.1	70	1
Bet.v.2	32	1
Cor.a.1.0101	30	1
Cor.a.1.0401	49	1
Hev.b.8	38	1
Mal.d.1	39	1
Mer.a.1	38	1
MUXF3	27	0.808
Ole.e.1	54	1
Phl.p.11	46	1
Phl.p.12	25	1
Pru.p.1	36	1

Table S10

k = 1 Broad

Component	Frequency	Assignment Probability	Component	Frequency	Assignment Probability	Component	Frequency	Assignment Probability
Act.d.1	4	1	Bla.g.7	3	1	Gly.m.6	4	1
Act.d.2	7	1	Blo.t.5	13	1	Jug.r.2	13	1
Act.d.8	15	1	Can.f.1	38	1	Jug.r.3	5	1
Alt.a.1	24	1	Can.f.2	9	1	Lep.d.2	30	1
Alt.a.6	4	1	Can.f.3	8	1	Mus.m.1	5	1
Api.g.1	10	1	Can.f.5	32	1	Ole.e.9	3	1
Ara.h.1	8	1	Che.a.1	4	1	Pen.m.1	6	1
Ara.h.2	11	1	Cor.a.8	4	1	Pen.m.2	5	1
Ara.h.3	8	1	Cry.j.1	15	1	Pen.m.4	3	1
Ara.h.6	12	1	Cup.a.1	20	1	Phl.p.7	8	1
Ara.h.9	4	1	Der.p.10	6	1	Pla.a.2	16	1
Art.v.1	7	1	Equ.c.1	24	1	Pla.a.3	4	1
Asp.f.3	4	1	Fel.d.2	6	1	Pol.d.5	3	1
Asp.f.6	8	1	Fel.d.4	22	1	Pru.p.3	4	1
Bet.v.4	5	1	Gal.d.5	7	1	Ves.v.5	4	1
Bla.g.2	3	1	Gly.m.4	16	1			

k = 2 House Dust Mite

Component	Frequency	Assignment Probability
Der.f.1	84	1
Der.f.2	79	1
Der.p.1	87	1
Der.p.2	87	1

k = 3 Grass

Component	Frequency	Assignment Probability
Cyn.d.1	109	1
Phl.p.1	154	1
Phl.p.2	85	1
Phl.p.4	95	1
Phl.p.5	128	1
Phl.p.6	91	1

k = 4 Cat

Component	Frequency	Assignment Probability
Fel.d.1	84	0.985

k = 5 PR-10

Component	Frequency	Assignment Probability
Aln.g.1	57	1
Ara.h.8	39	1
Bet.v.1	86	1
Cor.a.1.0101	45	1
Cor.a.1.0401	73	1
Mal.d.1	65	1
Ole.e.1	68	1
Pru.p.1	49	1

k = 6 Profilin

Component	Frequency	Assignment Probability
Bet.v.2	41	1
Hev.b.8	49	1
Mer.a.1	50	1
MUXF3	34	0.798
Phl.p.11	41	0.798
Phl.p.12	41	1

Table S11

Component
Act.d.1
Act.d.2
Ara.h.1
Ara.h.2
Ara.h.3
Ara.h.6
Blo.t.5
Can.f.1
Can.f.2
Can.f.3
Can.f.5
Che.a.1
Cup.a.1
Der.p.10
Equ.c.1
Fel.d.4
Gal.d.1
Jug.r.2
Lep.d.2
Mus.m.1
Pen.m.2
Pla.a.2
Pru.p.3
Ves.v.5

Table S12a

Frequency	Broad	Alternaria	Grass	HDM
332	0	0	0	0
42	0	0	1	0
37	1	0	1	1
35	1	0	1	0
25	0	1	0	0
23	0	0	1	1
20	0	0	0	1
17	1	0	0	0
12	1	1	1	1
11	1	1	0	0
10	1	1	1	0
5	0	1	1	0
5	1	0	0	1
4	1	1	0	1
2	0	1	0	1
2	0	1	1	1

Table S12b

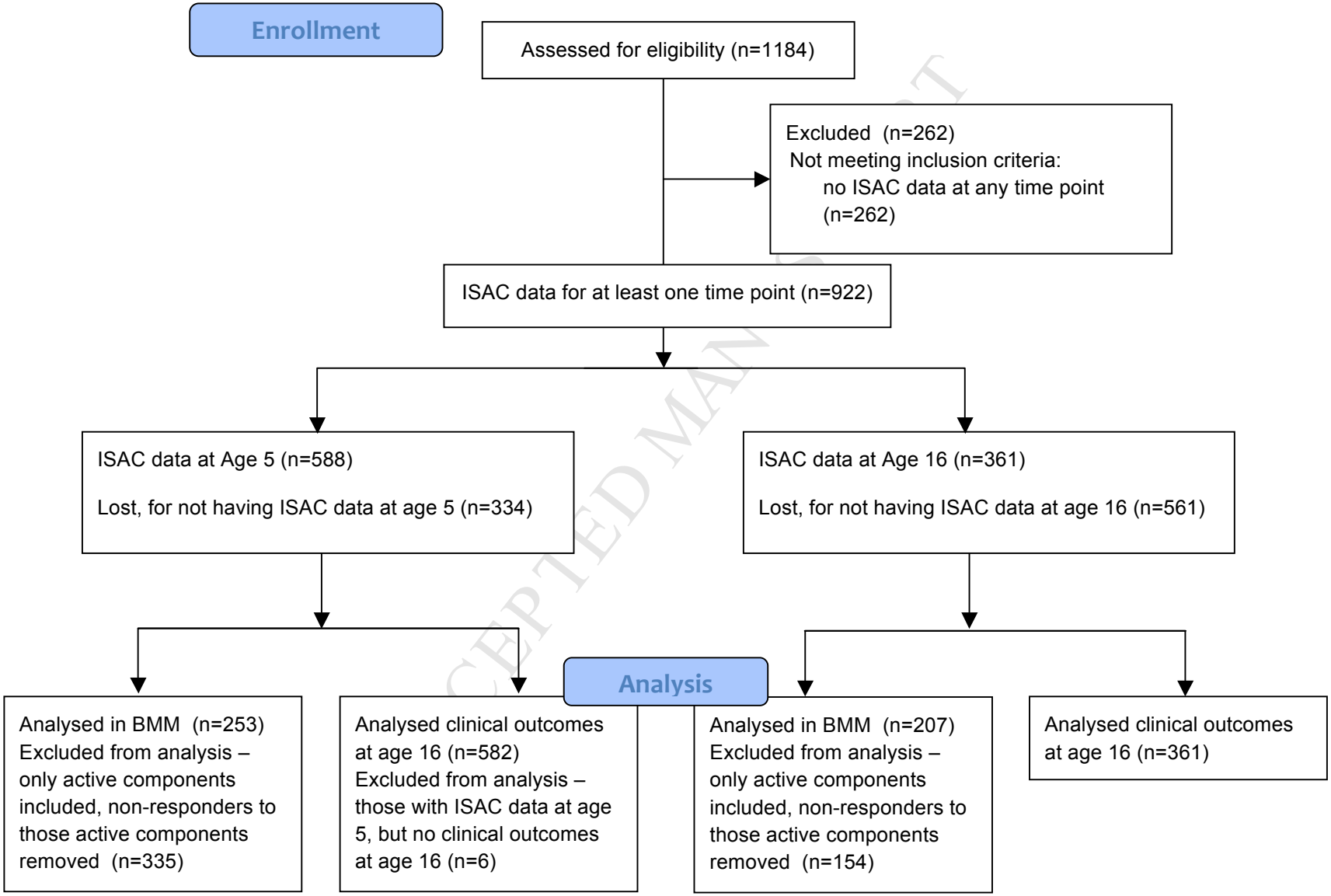
Frequency	Broad	PR-10	Profilin	Grass	Cat	HDM
154	0	0	0	0	0	0
17	1	1	1	1	1	1
14	0	1	1	1	1	1
12	1	1	1	0	1	1
10	0	0	0	0	1	0
10	1	0	0	0	1	0
9	1	0	0	0	1	1
9	1	1	0	0	1	1
9	0	0	1	0	0	0
8	1	1	0	1	1	1
7	0	0	0	0	0	1
7	0	0	1	0	1	0
7	0	1	1	0	1	1
6	0	0	0	0	1	1
6	1	1	0	0	1	0
6	0	1	0	1	1	1
6	0	1	1	1	1	0
5	0	1	0	0	0	0
5	0	0	1	1	1	1
4	0	1	0	0	1	0
4	0	0	0	1	0	0
4	1	1	0	1	1	0
4	0	0	1	0	1	1
4	1	0	1	1	1	1
3	0	1	0	0	1	1
3	0	0	0	1	1	1
3	0	0	1	0	0	1
3	1	0	1	0	1	0
3	1	1	1	0	1	0
3	0	1	1	1	0	1
2	0	1	0	0	0	1
2	1	0	1	0	1	1
2	0	0	1	1	0	0
2	0	0	1	1	0	1
1	0	0	0	1	0	1
1	0	0	0	1	1	0
1	0	1	0	1	1	0
1	1	0	0	1	1	1
1	0	1	1	0	0	1
1	0	1	1	0	1	0
1	0	0	1	1	1	0
1	1	0	1	1	1	0

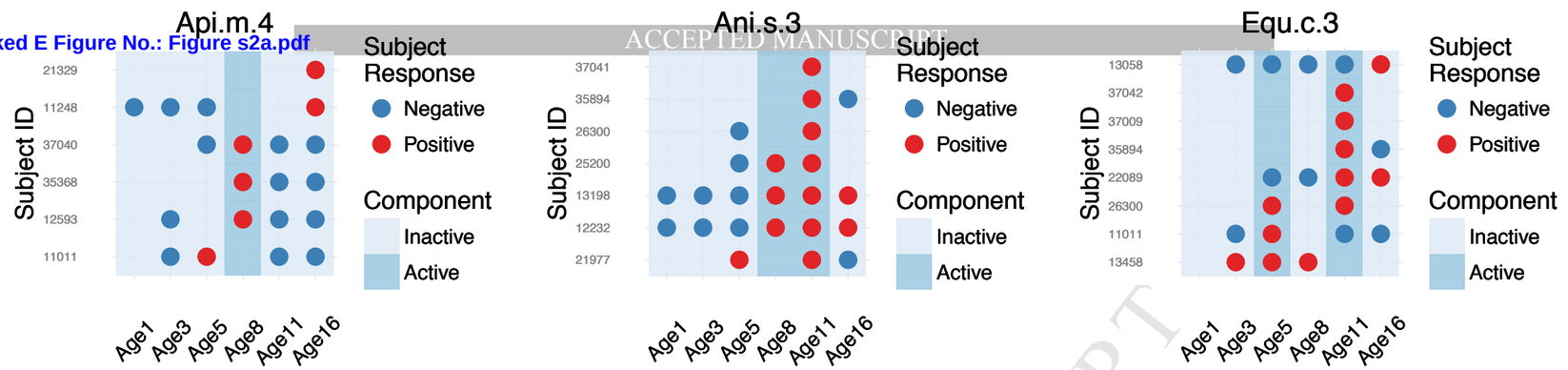
Table S13

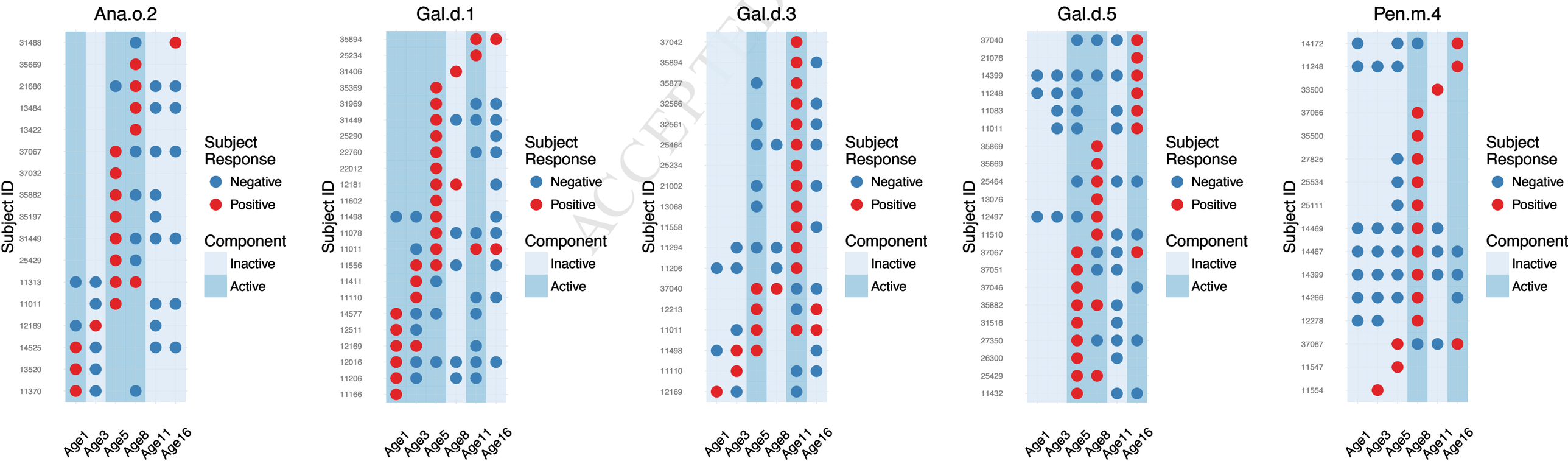
Age 5 Clusters	Age 16 Clusters						
	Broad	**HDM**	**Grass**	**Cat**	**PR-10**	**Profilin**	**Total**
Broad	51	37	53	32	46	30	62
HDM	34	44	37	26	31	20	47
Grass/cat	58	41	71	39	53	41	78
Alternaria	14	14	16	5	11	10	27
Total	92	79	121	58	82	64	

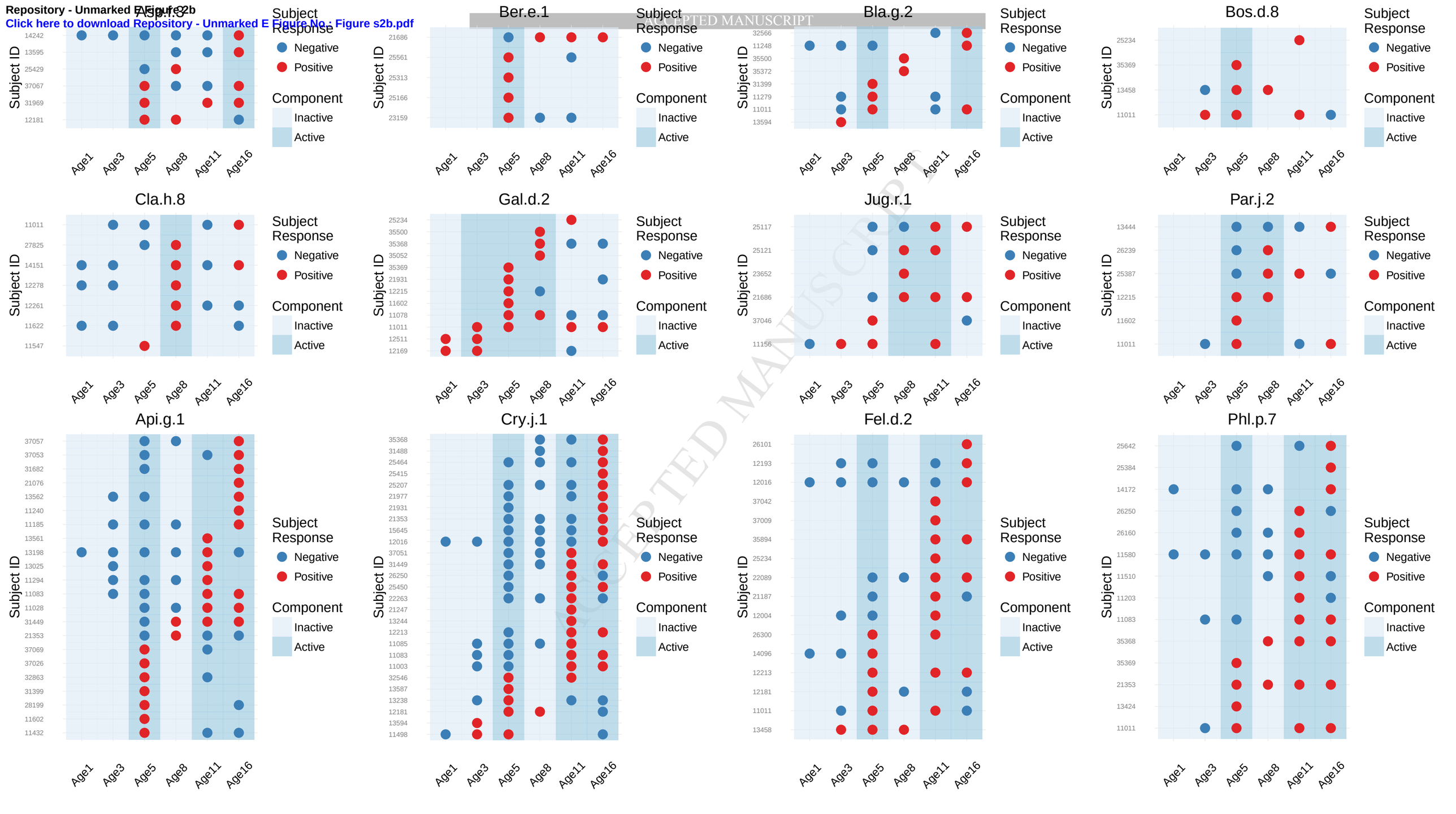
Table S14

	Cluster Responses	
	Age 5's	**Age 16's**
Current Asthma	0.73	0.76
Current Wheeze	0.63	0.70
Current Rhinitis	0.65	0.81

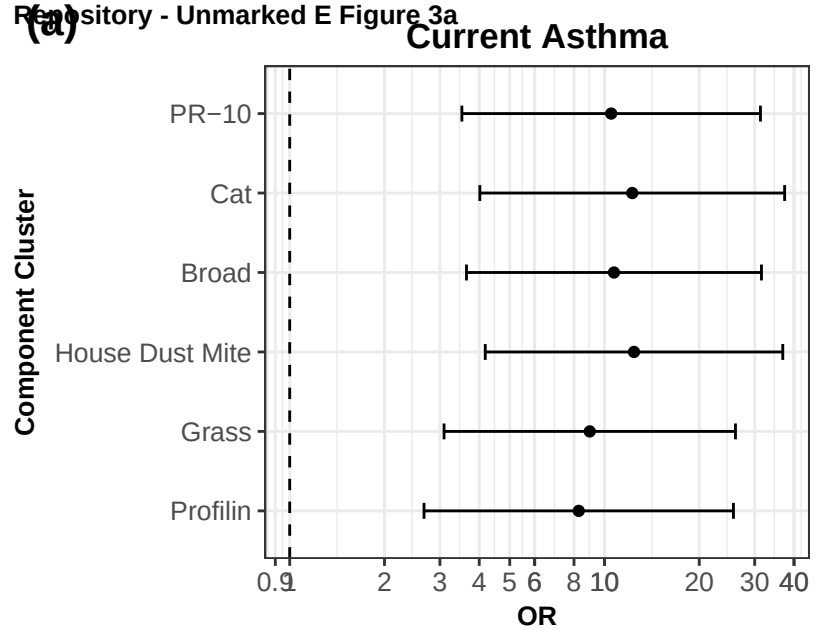


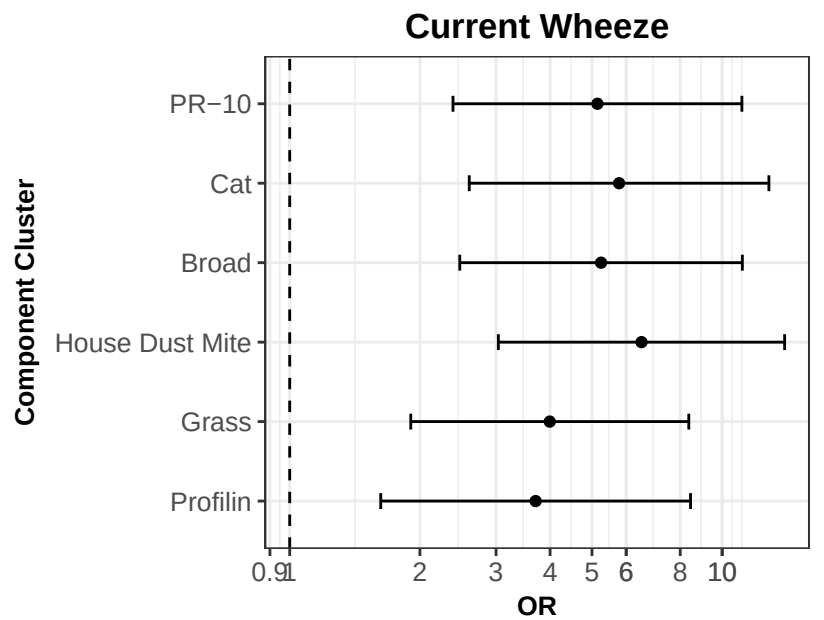


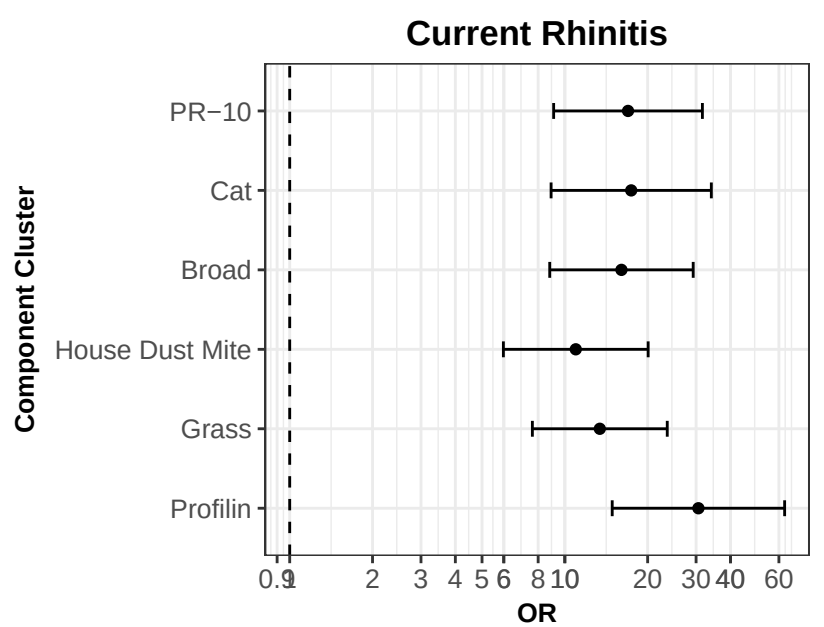




(a)







(b)