B cell repertoire in children with skin barrier dysfunction supports altered IgE maturation associated with allergic food sensitization

Authors: Kirandeep Gill[1], Carolina Moore[1], Onyekachi Nwogu[1], John W. Kroner[2], Wan-Chi Chang[2], Mariana L. Stevens[2], Asel Baatyrbek kyzy[2], Jocelyn M. Biagini[28], Ashley L. Devonshire[58], Leah Kottyan[3578], Justin T. Schwartz[58], Amal H. Assa'ad[58], Lisa J. Martin[38], Sandra Andorf[1568], Gurjit K. Khurana Hershey[28], Krishna M. Roskin[148]*

Affiliations:

[1]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center; Cincinnati, Ohio, USA.

[2]Division of Asthma Research, Cincinnati Children's Hospital Medical Center; Cincinnati, Ohio, USA.

[3]Division of Human Genetics, Cincinnati Children's Hospital Medical Center; Cincinnati, Ohio, USA.

[4]Division of Immunobiology, Cincinnati Children's Hospital Medical Center; Cincinnati, Ohio, USA.

[5]Division of Allergy and Immunology, Cincinnati Children's Hospital Medical Center; Cincinnati, Ohio, USA.

[6]Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center; Cincinnati, Ohio, USA.

[7]Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center; Cincinnati, Ohio, USA.

[8]Department of Pediatrics, University of Cincinnati, College of Medicine; Cincinnati, Ohio, USA.

*Corresponding author. Email: krishna.roskin@cchmc.org.

One Sentence Summary: Food allergic sensitization is associated with altered B cell development in children with skin barrier dysfunction.

Abstract: The skin is a major immune organ and skin barrier dysfunction is a major risk factor for the development of the inappropriate immune response seen in allergic disease. Skin barrier disruption alters the landscape of antigens experienced by the immune system and the downstream impacts on the antibody repertoire remain poorly characterized, particularly for the IgE isotype responsible for allergic specificity and in early life, when allergic disease is developing. In this study, we sequenced antibody gene repertoires from a large and well-characterized cohort of children with atopic dermatitis and found that food sensitization was associated with lower mutation frequencies in the IgE compartment. This trend was abrogated in children living with pets during the first year of life. These results elucidate potential molecular mechanisms underlying the protective effects of pet ownership and non-antiseptic environs reported for allergic disease, and the hygiene hypothesis more broadly. We also observed increased IgE diversity and increased isotype-switching to the IgE isotype, suggesting that B cell development, particularly isotype-switching, is heavily altered in the those with food allergen sensitizations relative to those without food allergen sensitizations. Unlike for food antigens,

aeroallergen sensitization exhibited no effect on IgE mutation or diversity. Consistent patterns of antibody rearrangement were associated with food allergen sensitization in subjects with atopic dermatitis. Thus, we propose the Immune Repertoire in Atopic Disease (IRAD) score, to quantify this repertoire shift and to aid clinically in patient diagnosis and risk stratification.

Main Text:

INTRODUCTION

Atopic dermatitis (AD) is a common chronic inflammatory skin condition. This skin barrier dysfunction contributes to the inappropriate immune response that develops in atopic diseases, including asthma and food allergy *(1)*. While AD typically presents as dry skin and erythematous pruritic lesions, nonlesional skin is also involved and likely promotes development and severity of atopic disease via subclinical barrier dysfunction and inflammation *(2, 3)*. Furthermore, mouse models suggest that immune dysregulation in the skin can also lead to inappropriate immune responses at other epithelial surfaces, such as the gut *(4)*. Taken together, these results suggest that barrier dysfunction is associated with wide-scale immune dysregulation. It is out of this immune milieu that the antibody producing B cells develop and mature. The antibodies secreted by IgE-positive B cells are directly involved in allergy specificity and atopic disease *(5)*. Thus, we set out to study alterations in the B cell repertoire and its association with allergic sensitization in children with AD.

When exposed to antigen or allergen, B cells clonally expand and activate somatic hypermutation (SHM) targeting B cell receptor (BCR) genes. While SHM correlates with greater antigen affinity of secreted antibodies, the role of SHM and affinity maturation in atopic responses has been unclear *(6, 7)*. Previous work characterized the BCR repertoires of children in Stanford's Outcomes Research in Kids (STORK) birth cohort and found increased SHM levels in children with AD. However, only 14 (out of 51) STORK children had AD, making the generalizability of those finding to atopic disease development in the context of skin barrier dysfunction unclear, especially considering the wide spectrum of AD and atopic disease severity *(8)*. To further elucidate the immunological effects of the skin barrier dysfunction seen in AD on the BCR repertoire, we performed high-throughput sequencing of rearranged immunoglobulin heavy-chain (IgH) genes from peripheral B cells of a subset of 162 children participating in the Mechanisms of Progression of AD to Asthma in Children (MPAACH) study *(9)*. MPAACH is the first US-based prospective early life cohort of children with AD. The study conducted extensive participant phenotyping including several measures of AD severity and atopic disease progression.

Our data reveal changes in SHM levels and BCR diversity in children with AD who have food allergen sensitizations as compared to children with AD lacking food allergen sensitization. These finding suggest major alterations in the maturation pathway of IgE B cells that are associated with food allergen sensitizations. By contrast, sensitization to aeroallergens had no effect on the repertoire (Fig. 4), suggesting that a different immune pathway leading to environmental allergy development. We observe a compensatory response in children living with a pet in the first year of life, thus providing a possible molecular mechanism underling the protective effects of pet ownership and non-antiseptic environs reported for atopic disease and the hygiene hypothesis more broadly *(10)*.

We also propose the Immune Repertoire in Atopic Disease (IRAD) score that captures these B cell repertoire changes and validate its ability to distinguish food allergen sensitized

subjects from non-food allergen sensitized subjects using an independent cohort. Current diagnostic tests for food allergy either have poor performance (skin prick testing, allergen-aspecific IgE testing), particularly in children, or carry considerable risk and cost (oral food challenge) *(11, 12)*. The IRAD score could be used to aid in atopic disease diagnosis and risk stratification.

RESULTS

Antibody SHM increases with age across all isotypes, including IgE

As seen in other studies, the early childhood antibody repertoire shows a progressive increase in SHM over time (Fig. 1A and fig. S1A), which was statistically significant for all isotypes (Table S2) *(8)*. Because of the size of our main cohort (*n*=162, Table S1), we were able to detect age-associated increased SHM in the IgE compartment over early life (Fig. 1A). Prior repertoire analyses either excluded IgE data or noted high variability in the IgE compartment preventing the detection of an age-associated increase in IgE SHM *(8, 13)*. This study is the first to detect an age-associated increase in SHM in the IgE compartment over the first several years of life.

Thus, SHM increases progressively as the child immune system matures, presenting challenges for cross-sectional antibody repertoire analyses. Studies of progressive diseases or diseases that are only diagnosed after a certain age, such as atopic diseases, must be careful to control for age effects. For example, in this cohort, children with food sensitization, high transepidermal water loss (TEWL), or low skin expression of filaggrin (FLG) tend to be younger (fig. S1B) *(2)*. Thus, age and its relationships with SHM must be accounted for when looking for associations with diseases that have an age-dependent phenotype or severity.

We control for the effect of age by modeling the age-SHM relationship using a generalized linear model (GLM). For each isotype, a GLM was created to model the relationship between the age of a subject and their average percent SHM (Fig. 1A, fig. S1A, Table S2). To control for age, we adjusted each subject to their estimated SHM level at 3 years of age, the middle of the age range of this cohort.

Compared to children without AD, those with AD have higher SHM across all isotypes, except for IgE, where they exhibit lower SHM, particularly in those with food sensitizations

After age adjustment, children diagnosed with AD had higher SHM relative to children without an AD diagnosis for all isotypes except for IgE, associations that reached statistical significance for IgA2 and IgG2 (Fig. 1B). The IgE compartment did not follow this trend: AD subjects had slightly lower SHM levels than subjects without AD, though this trend was not significant. However, upon stratification of AD children into those with food allergen sensitization (positive SPT results to at least one of the thirteen food allergens tested, see Methods) and those without food allergen sensitization (negative SPT to all food allergens tested), the reduction in age-adjusted IgE SHM was statistically significant and much more pronounced in subjects with sensitization to food allergens (Fig. 1C). Thus, counter to our expectations, IgE SHM is decreased and potentially suppressed in children with food allergen sensitizations compared to those with no food allergen sensitizations. This patten of lower IgE SHM in food allergen sensitized subjects held robustly across the age range of our cohort, with possible resolution by 6 years of age (Fig. 1D). Other isotypes exhibited higher SHM in food sensitized AD subjects versus non-food sensitized AD subjects, but the difference was only significant for IgD and IgG3 (fig. S1C). No significant differences were seen for age-adjusted SHM levels when stratified by

several measures of AD severity including assessment of skin barrier integrity (TEWL, FLG expression) and clinical assessment (SCORing Atopic Dermatitis, SCORAD) (fig. S2).

The presence of a household dog is associated with higher IgE SHM levels and counteracts the suppressive effect associated with food allergen sensitization

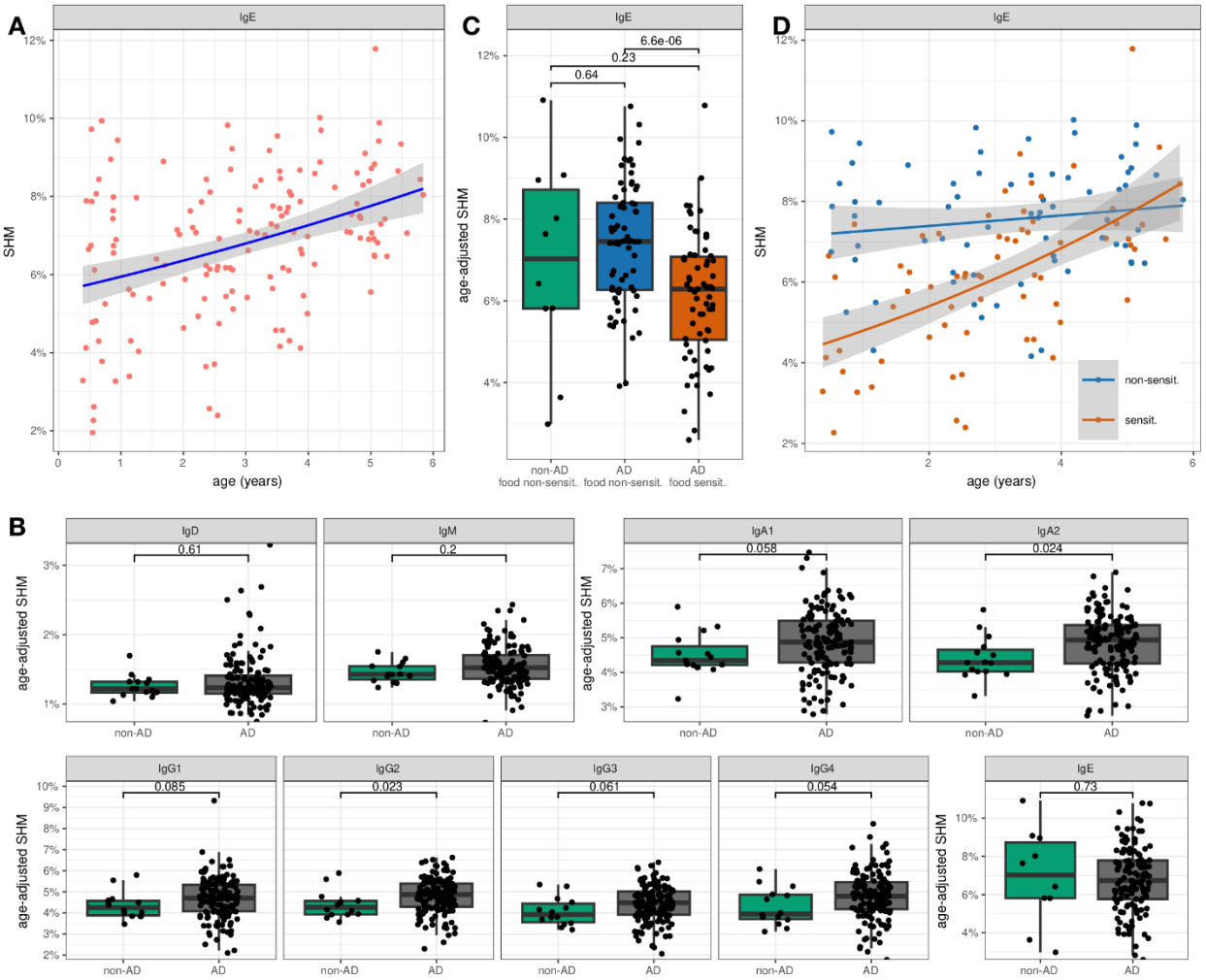

Fig. 1. Modeling of somatic hypermutation (SHM) with age, AD, and food allergen sensitization status. (A) Logistic regression of per-subject mean IgE SHM with age. A statistically significant correlation of SHM with age was seen in all isotypes and subtypes including IgE (fig. S1, Table S2A). **(B)** Age-adjusted SHM levels were calculated using the models in (A) and fig. S1A by taking the expected SHM level of all subjects at 3 years of age, the middle of the age range. Median age-adjusted SHM was higher in subjects with AD as compared to non-AD subject for all isotypes except for IgE, but the difference was only statistically significant for IgA2 and IgG2. In the IgE compartment, no increase was observed in AD subjects, and even a slight decrease was seen. **(C)** When the data was analyzed based on food allergen sensitization status (as measured by skin prick testing, SPT), AD subjects sensitized to food allergens had significantly lower age-adjusted IgE SHM than AD subjects with no food allergen sensitization. Subjects without AD, who also had no food sensitization in this cohort, were comparable to AD subjects with no food allergen sensitization. Other isotypes shown in fig. S1C. **(D)** Regression modeling of IgE SHM with age as a function of food allergen sensitization status. The suppression of age-adjusted IgE SHM seen in (C) for food allergen sensitized subjects occurs across all age ranges but seem to resolve by age 6.

The presence of a household pet, particularly dogs, has been proposed to attenuate the development of atopic disease in infants *(14, 15)*. We found that MPAACH subjects living with a pet dog in the first year of life had significantly higher levels of age-adjusted IgE SHM (Fig. 2A), approaching the levels seen in non-AD subjects. Furthermore, the lower IgE SHM levels seen in subjects with food allergen sensitization (Fig. 1C) was not seen in subjects with a pet dog (Fig. 2B), suggesting that the presence of a household dog may promote normal IgE SHM levels. The overall pattern also holds for cat ownership (fig. S3), however, the rarity of cat ownership in the food sensitized group (*n*=3) makes the validity of the results uncertain.

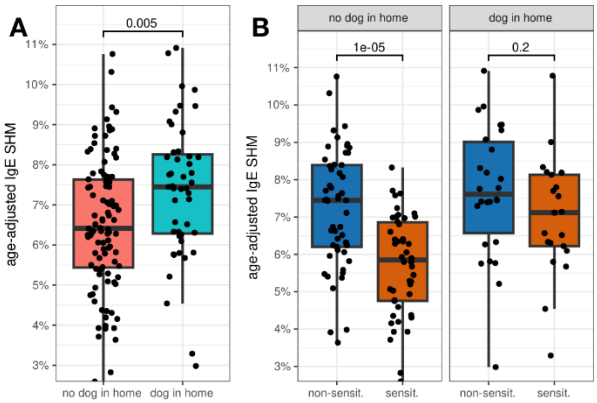

Fig. 2. Pet dog ownership counteracts the decrease in IgE SHM levels associated with food allergen sensitization. **(A)** AD subjects who lived with a pet dog in the first year of life had significantly higher IgE SHM than those without, nearing levels seen in non-AD subjects. **(B)** The pronounced decrease in IgE SHM levels seen in food allergen sensitized AD subjects was excusive to dog-free households (left). For those who lived with a dog during the first year of life, food allergen sensitization had no significant effect on IgE SHM (right).

Repertoire diversity remains constant during early life while IgE diversity is greatly increased in food allergen sensitized subjects

Given the correlation between SHM and age (Fig. 1A), we next examined if repertoire diversity also changes during early life. We measured repertoire diversity controlling for SHM (see Methods). Unlike average SHM, repertoire diversity did not change significantly with age in early childhood (fig. S4, Table S3).

Food allergen sensitized children with AD, relative to children with AD without food allergen sensitization, had increased diversity in all class-switched B cell isotypes (Fig. 3A). The IgE compartment, the isotype directly involved in allergy and atopic disease, had the greatest increase in and range of diversity. IgE-positive B cells are relatively rare; thus, the baseline for IgE diversity is drastically lower than for the other isotypes (bits, the units of diversity in Fig 3A, are logarithmic). It is therefore striking that the median IgE diversity of food allergen sensitized subjects begins to approach the diversity observed in the other isotypes. In contrast to the class-switched isotypes, the diversities of naïve, non-class switched isotypes (IgM, IgD) slightly

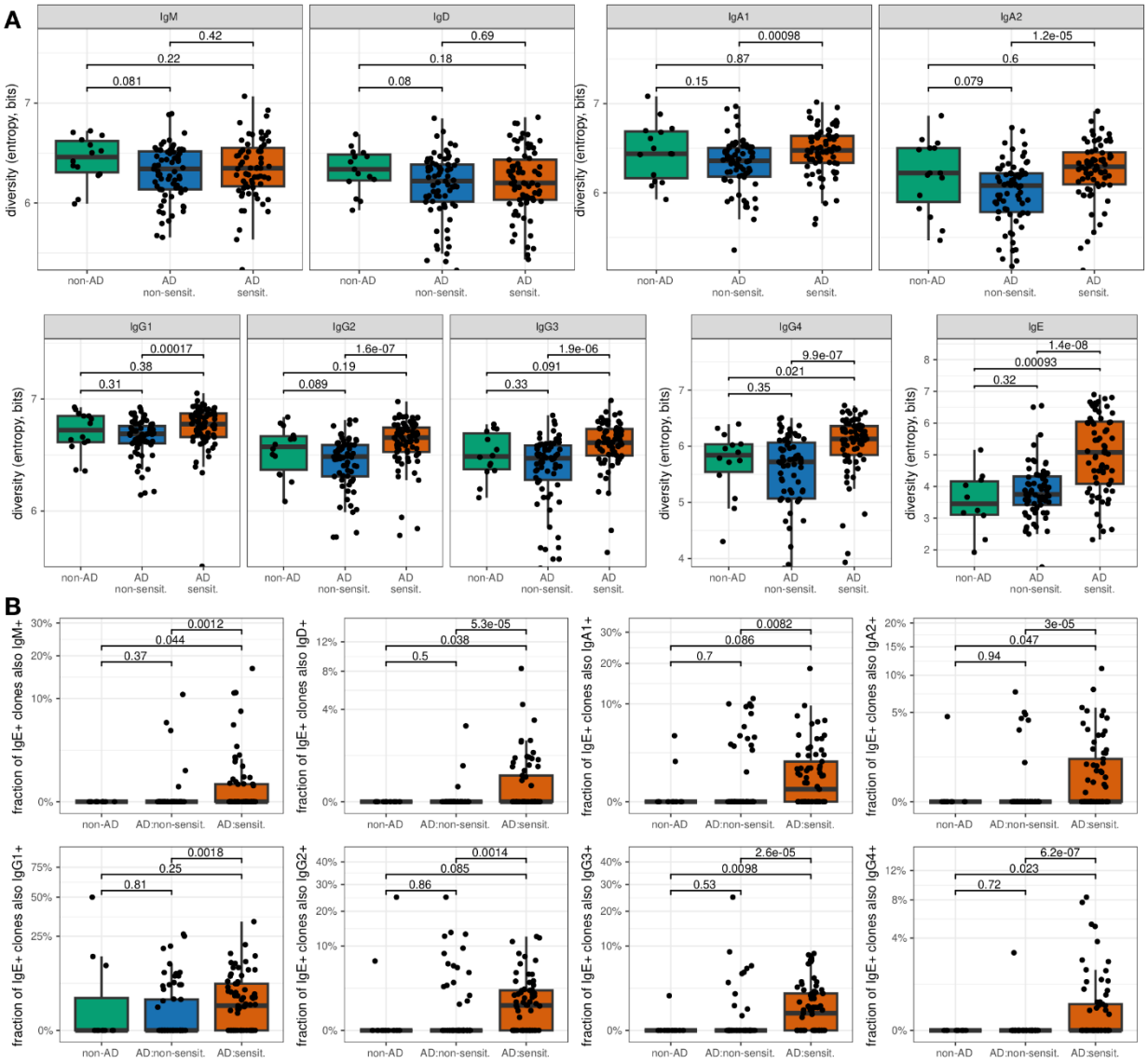

Fig. 3. Repertoire diversity and clonal overlap in non-AD and AD subjects with and without food sensitizations. (A) Non-isotype switched B cells showed a small and not statistically significant decrease in BCR diversity in AD subjects vs. non-AD subjects. By contrast, the switched isotypes and -subtypes showed statistically significant increases in diversity in AD subjects with food allergen sensitizations as compared to those with no food allergen sensitizations. This increase was particularly strong in the IgE compartment. Diversity was measured as the Shannon entropy of the distribution of V- and J-segment usage of the B cell clones expressing the given isotype and -subtype. Since this measures repertoire combinatorial diversity, these results are independent of the SHM differences described in Fig. 1 and 2 *(45, 46)*. **(B)** Clonal overlap of IgE with other isotypes. Fraction of IgE-positive clones that also contain clonally related sequences of other isotypes. AD subjects with food allergen sensitizations as compared with AD subjects with no food allergen sensitizations exhibited more overlap of the IgE compartment with all other compartments and this difference was statistically significant for all isotypes and -subtypes.

decrease from non-AD to AD patients; with no significant differences within the AD group between those with and without food allergen sensitization. Overall, these changes in diversity are consistent with accelerated class switching from naïve IgM and IgD to IgA, IgG, and IgE, perhaps due to increased antigen stimulation or altered B cell developmental pathways.

IgE-containing lineages have more overlap with switched isotypes in food allergen sensitized subjects relative to subjects without food allergen sensitizations

The increased diversity seen in the IgE compartment suggests an increase in clonal lineages isotype-switching to IgE. We looked for evidence of this by measuring the fraction of IgE containing clones that also contain another isotype. Compared to subjects with no food allergen sensitizations, food allergen sensitized subjects showed an increased frequency of overlap of IgE-positive clones with the other isotypes (Fig. 3B).

Aeroallergen sensitization status is not associated with changes in IgE SHM or diversity

The changes in age-adjusted IgE SHM and diversity based on food allergen sensitization status we observe were not recapitulated for aeroallergen sensitizations. Unlike for food, aeroallergen sensitization status has no effect on age-adjusted IgE SHM (Fig. 4A) or on B cell diversity (Fig. 4B) alone or in combination with food allergen sensitization.

Food allergen sensitized subjects exhibit a skewed repertoire

Differences in SHM, diversity, and isotype overlap in AD subjects with food allergen sensitizations underlie complex, systemic changes in the immune state. Sequencing of the B cell receptor repertoire provides an opportunity to measure these changes. These measures could be used to categorize of atopic disease endotypes. We examined the V-segment usage frequencies across each isotype of the subjects in our cohort. With 57 functional V-segments and 5 isotypes, this is a high-dimensional space *(16)*. We used principal component analysis (PCA) to linearly reduce this space to two dimensions which capture the largest trends in our data (Fig. 5A). Although PCA does not take phenotype into account, the axis of highest variance (PC1) was associated with food allergen sensitization and statistically separated food allergen sensitized AD subjects from AD subjects with no food allergen sensitizations (Fig. 5B). To validate the ability of this PC1-based score to discriminate subjects with and without food allergen sensitization, we collected samples from a validation cohort (*n*=20) in collaboration with the Allergy Clinic at Cincinnati Children's Hospital Medical Center (Table S4). The samples were subjected to

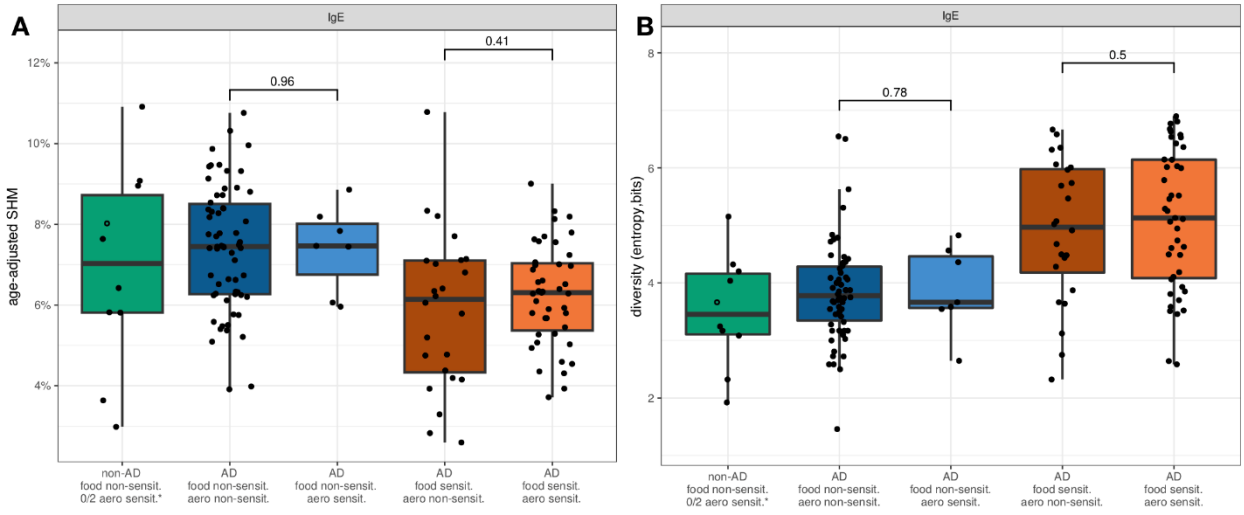

Fig. 4. Aeroallergen sensitization status has no effect on age-adjusted IgE SHM or diversity. (A) When food allergen sensitization status is further subdivided based on sensitization to aeroallergens, no statistically significant changes in age-adjusted IgE SHM is observed between those with and without aeroallergen sensitizations. **(B)** IgE diversity is also unaffected by aeroallergen sensitization regardless of food allergen sensitization status. *The non-AD group included a single subject, shown as an open circle, that was SPT positive to two aeroallergens.

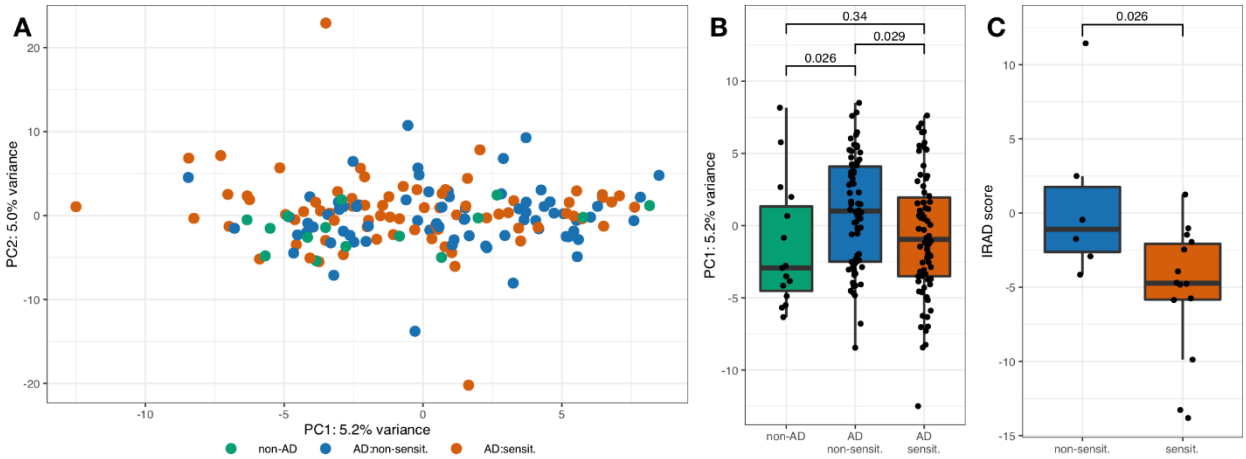

Fig. 5. Principal Component Analysis (PCA) of V-segment usage. (A) Two-dimensional PCA projection of subject V-segment usage in each isotype. Each dot is colored according to subject phenotype. **(B)** Quantification of how well the component of highest variance (PC1) distinguishes the three phenotypes. By this measure, AD subjects with food allergen sensitization scored lower than AD subjects with no food allergen sensitizations and this difference was statistically significant. **(C)** This score, which we call the Immune Repertoire in Atopic Disease (IRAD) score, applied to an independence cohort. In this validation cohort, food allergen sensitized subjects had statistically significantly lower scores than subjects without food allergen sensitizations.

antibody gene repertoires sequencing using the same protocol used for the discovery cohort. We applied the rotations of the above PCA analysis to calculate the PC1 score on data from the validation cohort (fig. S5). This score, which we call the Immune Repertoire in Atopic Disease (IRAD) score, was able to statistically separate food allergen sensitized subjects from non-food allergen sensitized subject in this independent cohort (Fig. 5C).

DISCUSSION

Previous studies have noted low levels of SHM during early life, extending as far back as gestational week 27 *(13, 17)*. The first major IgH repertoire study of infants up to 2 years of age noted a lack of evidence for increasing SHM in the IgE compartment *(18)*. This raised the question of how and when does the developing immune system gain the additional 3-4% SHM level seen in healthy adults *(8)*. This current larger study, which extends to 6 years of age, demonstrates an increase in IgE SHM and provides a possible confounding variable that obscured detection of this trend in previous reports. In this AD cohort, food allergen sensitized subjects showed reduced IgE SHM levels when compared to subjects with no food allergen sensitizations (Fig. 1C). We hypothesize that this decrease coupled with the age-associated increase in SHM (Fig. 1A and fig. S1A) resulted in the flat SHM levels previously seen in the first 2 years of life *(8)*.

A previous study reported increased SHM associated with atopic disease *(8)*. This presents a paradox for the *hygiene hypothesis*, which posits a protective role for increased antigen exposure *(19)*. If this “protective” antigen exposure results in increased SHM, how is increased SHM also associated with atopic disease? With a larger cohort, greater statistical power, and methodological improvements, our findings provide a potential resolution of this contradiction: we show that food allergen sensitization is associated with an immune-repressed IgE milieu instead of a stimulated one. Further studies will be needed to determine if this relationship holds for food allergen sensitized, non-AD subjects and other demographic groups. Our data also provides an explanation for the protective effects of dog ownership seen in epidemiologic studies, as pet ownership should yield increased antigen load *(15)*. The presence

of a household dog attenuates the reduction of IgE SHM levels associated with food allergen sensitization, thus helping maintain "normal" IgE SHM levels (Fig. 2B).

For food allergen sensitized subjects, the increased repertoire diversity in IgE (Fig. 3A) together with that compartment's lower SHM levels (Fig. 1C) suggests that B cells are isotype switching to IgE more readily in these subjects, before they can accumulate SHM. This is consistent with the increased percentage of IgE-positive clones that also contain members of other isotypes seen in food allergen sensitized subjects (Fig. 3B).

The influx of diverse B cells into the IgE compartment could be due to broad nonspecific, immune stimulation from a superantigen, such as those expressed by *Staphylococcus aureus* and implicated in AD *(20)*. While we do observe a broad system immune response, we also observe specific V-segment usage patterns that are predictive of food allergen sensitization (Fig. 5). Selection for these stereotypic rearrangements suggests that antigen selection *is* contributing to this response. This is consistent with epidemiological studies that find the protective effects of farming environments on atopic disease is restricted to specific kinds of livestock and farming activities and thus likely has some antigen-specific component *(21)*.

Principle component analysis (PCA) of the repertoire shows that the largest trend, captured by the first principal component (PC1), accounts for ~5% of the variance in the repertoire and separates food allergen sensitized subjects from those without food allergen sensitizations (Fig. 5). The dominance of sensitization-dependent repertoire shifts recovered from this unbiased analysis strongly suggests the level to which the peripheral B cell repertoire is involved in the sensitization response.

The B cell repertoire changes in atopic disease described here primarily occur based on food allergen sensitization status. Sensitization to aero allergens appear to have no effect on these peripheral blood derived data (Fig. 4). B cell receptor repertoire sequencing of other B cells, for example lung-resident B cells, might be needed to measure the repertoire effects of sensitization to aeroallergens.

Several scales of AD severity have been proposed that measure clinical severity and skin barrier function. These clinical measures are not associated with changes in IgE SHM (fig. S2). Thus, a measure of these immunological changes based on sequencing antibody genes would provide novel information that could be used to uncover atopic disease endotypes. To that end, we propose the IRAD score (Fig. 5) and used an independent cohort, with mixed AD status, to validate the ability of this score to statistically separate food allergen sensitized subjects from non-food allergen sensitized subjects. Further work is needed to study the performance of the IRAD score in a larger cohort and to pick an optimal score cutoff for prediction. It will also be important to understand how this immunological measure of food allergen sensitization correlates with other clinical phenotypes, demographics, and how it can be combined with other predictive features and used in the clinical setting in the future.

A limitation of the current study is the small numbers of non-AD subjects which primarily serves to establish directionally. The repertoire changes described above might be specific to this high-risk cohort. A larger study with more non-AD subjects, particularly non-AD subjects with food allergen sensitizations, will be needed to better explore these clinical phenotypes and to determine how these results are generalizable to subjects without AD.

The differences in IgE SHM described above between food allergen sensitized subjects and subjects with no food allergen sensitizations decreases with age and appears to dissipate around 6 years of age (Fig. 1D). This is consistent with natural history studies of peanut, milk, and egg sensitization, and food allergen sensitization and food allergy in general *(22–26)*. The repertoire data in this study are from a single timepoint. Incorporation of data from additional timepoints could further elucidate the dynamics of this process and help determine if the immunophenotypes observed here are durable, or wax and wane with allergen sensitization.

The hypothesis that IgE undergoes different affinity selection than other isotypes has been proposed previously. Dahlke and coauthors analyzed mutation patterns in IgE sequences and argue that IgE B cells undergo reduced affinity selection as compared to IgG B cells *(27)*. They hypothesize this is the result of an extrafollicular origin for these B cells. It has been suggested, that in AD, CD27-negative IgE-positive B cells are derived in a T cell-independent manner resulting in low SHM levels *(28)*. An increase in extrafollicular IgE B cells in food allergen sensitized subjects could underlie the IgE SHM differences seen here. Further studies that link the B cell phenotype of these cells with their isotype and SHM status will be needed to verify this hypothesis. Such studies could also identify druggable vulnerabilities in this potentially pathogenic cell population. If this pathogenic population could be altered, it could positively impact atopic disease development.

In summary, our analysis of the BCR repertoire in a cohort of subjects with AD finds large immune perturbations associated with food allergen sensitization. We observe lower IgE SHM in subjects with food allergen sensitizations compared to those without food allergen sensitizations. While this result was unexpected, it does provide an explanation for the protection against atopic diseases conferred by pet ownership and non-antiseptic living that has been described epidemiologically. In food allergen sensitized subjects, we also see increased diversity in the IgE compartment and increase overlap between this compartment and other isotypes. This suggests that B cells are isotype switching to IgE at a faster rate in these subjects. Finally, we propose a measure that quantifies these immune changes that could be deployed in the future to aid in diagnosis and risk stratification of allergic disease.

MATERIALS AND METHODS

Study design

The Mechanisms of Progression of AD to Asthma in Children (MPAACH) is a prospective early life cohort of children with AD who are followed with yearly visits *(2)*. Inclusion criteria were (1) age ≤2 years at enrollment, (2) gestation of greater of at least 36 weeks and (3) a diagnosis of AD (based on the Hanifin and Rajka criteria for AD*(29)*) or the parent(s)/legal authorized representative indicates a positive response to each of the 3 questions from the Children's Eczema Questionnaire*(30)*. Exclusions criteria include (1) a comorbid lung condition including cystic fibrosis, congenital anomaly, or bronchopulmonary dysplasia, (2) dependence on immunosuppression or oral steroids for a medical condition other than asthma, (3) condition that precludes sampling of the proposed biologic samples or completion of spirometry, and (4) a bleeding diathesis. This study was approved by the Institutional Review Board at Cincinnati Children's Hospital Medical Center (CCHMC). All subjects or their legal representative signed informed consent documents prior to participation. The study population in this analysis is a subset of size n=162 of MPAACH enriched for those sensitized to peanut, hen egg, or cow milk. The characteristics of this sub-cohort are shown in Table S1.

Atopic dermatitis phenotyping

AD severity was evaluate clinically using SCORing Atopic Dermatitis (SCORAD) *(31)*. Skin barrier integrity was quantified using transepidermal water loss (TEWL) as measured on lesional skin using the DermaLab TEWL probe (Cortex Technology, Hadsund, Denmark). Skin barrier dysfunction was quantified by expression of filaggrin (FLG) from lesional keratinocytes normalized to the expression of the 18S rRNA, as described previously *(2, 32)*.

Allergic sensitization

Allergic sensitization was measured on the day of blood draw by skin prick testing (SPT) with a panel of 13 foods (milk, peanut, egg white, egg yolk, cashew, almond, walnut, pistachio, pecan, hazelnut, Brazil nut, wheat, soy) foods and 11 aeroallergens (two tree mixes, dog, ragweed, cat, two mold mixes, grass, cockroach, mites, weeds) *(2)*. A positive SPT result was a wheal diameter >3 mm larger than the diluent control. The replication cohort had food allergy sensitization determined by history of immediate allergic reaction to a food, a serum IgE that is predictive of clinical reactivity and/or a positive SPT to the food in question with a wheal diameter >3 mm above the negative control.

IgH sequencing libraries

Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood using the SepMate column system (StemCell Technologies) with Histopaque-1077. AllPrep column purification (Qiagen, Valencia, CA) was used to isolate RNA. Complementary DNA (cDNA) was generated from total RNA using SuperScript III (Invitrogen, Waltham, MA) and random hexamers. PCR amplification of IGH rearrangements from the cDNA template was carried out according to ref. *(8)*. In brief, template was amplified using multiplexed primers targeting IGHV gene segments using the BIOMED-2 first framework primers and isotype-specific primers located in the CH1*(33, 34)*. This first round PCR primers also included half of the Illumina P5 and P7 adapter sequences. The first round PCR used Platinum PCR Master Mix (Applied Biosystems, Waltham, MA) according manufacture instructions, and the following program: 95°C for 2 min, 35 cycles of (95°C for 30 sec, 60°C for 90 sec, 72°C for 60 sec), and a final extension at 72°C for 10 min. Illumina adapters were completed by a second PCR carried out with the Qiagen Multiplex PCR kit (Qiagen, Valencia, CA), using 1 µl of the first PCR product as the template in a 20 µl reaction with the following program: 94°C for 15 min, 12 cycles of (94°C for 30 sec, 60°C for 45 sec, 72°C for 90 sec), and a final extension at 72°C for 10 min. Each isotype was amplified separately to decrease chimeric product generation. PCR reactions for all samples were pooled and purified by agarose gel electrophoresis and gel extracted using the QIAquick kit (Qiagen, Valencia, CA). Sequencing of the final libraries was performed on the Illumina MiSeq instrument using 600-cycle kits by the CCHMC DNA Core.

IgH sequence annotation

High-throughput sequencing data was processed as previously described *(8, 35)*. In short, 300bp paired-end reads were merged using FLASH *(36)*. Reads were mapped to samples using barcode sequences. The V-, D-, and J-segments and framework and CDR regions were identified using IgBLAST (version 1.16.0) *(37)*. Sequences were quality filtered to include only productive reads with a CDR-H3 region, minimum V-segment alignment score of 70, and minimum J-segment alignment score of 26. The isotype of each transcript was determined by exact matching to a database of constant region gene sequences upstream from the primer *(16)*. Subjects were

required to have at least 5 clones of a given isotypes or were excluded from the analysis due to poor estimates of per-subject mean SHM, diversity, or V-segment usage.

Clonal inference

Clonal relationships were assigned as previously described, using mmseqs2 (Sep. 2019 build) *(38)* to cluster the sequences into clones using the same V- and J-segments (without considering the allele), equal CDR-H3 length, and at least 90% CDR-H3 nucleotide identity *(34)*. Mutation levels and V- and J-usage frequencies are calculated per clone using this inference of clonal grouping.

SHM calculation

The per-read SHM level was calculated as the percentage of V-segment bases that did not match the inferred germline sequence, excluding the region targeted by primer sequences and the part of CDR3 encoded by the end of the V-segment. For each isotype, the clonal SHM level was calculated as the median SHM level of all reads of that isotype in clonal lineage. For each isotype, the subject SHM level was calculated as the mean SHM of each lineage of that isotype.

Age-adjustment of SHM levels

Regression of SHM level with age was performed using a Gaussian error distribution and a logit link function. Each subject's isotype-specific SHM level was age-adjusted by taking the residual after regression and adding the SHM level expected for a subject 3 years old, the middle of the age range of this cohort.

Diversity analysis

For each isotype and subject, the isotype-specific usage frequency of V- and J-segment pairs of each clone was calculated. From this distribution, Shannon entropy was calculated, in bits, to measure the diversity *(39)*. Since only extreme levels of SHM would cause IgBLAST to misclassify a V- or J-segment artificially increasing diversity, this measure is independent of SHM.

Statistical analysis

Statistical analysis and graphs were generated using the R statistical language (version 4.2.0), in RStudio (build 485) *(40, 41)*. Regression analysis of SHM levels was performed using the glm R function with a Gaussian error distribution and the logit link function. Regression analysis of diversity with ages was performed using the lm function. The statistical significance of differences in SHM, age, diversity, repertoire overlap, principal components, and IRAD score between groups were tested using the two-side Wilcoxon test as implemented by the wilcox.test function in R. A p-value of 0.05 was considered significant. Plots were generated using the ggplot2 (version 3.3.6) and ggpubr (version 0.4.0) packages *(42, 43)*. Box-whisker plots show median (horizontal line), interquartile range (box), and 1.5 times the interquartile range (whiskers). Confidence intervals around regression curves show 95% confidence intervals and were calculated using the stat_smooth ggplot2 function. Principle Component Analysis (PCA) was done using the pca function of the PCAtools R package (version 2.8.0) *(44)*. PCA analysis was performed after removing one-sixth of the features with the lowest variance.

List of Supplementary Materials

References and Notes

1. H. A. Brough, A. H. Liu, S. Sicherer, K. Makinson, A. Douiri, S. J. Brown, A. C. Stephens, W. H. Irwin McLean, V. Turcanu, R. A. Wood, S. M. Jones, W. Burks, P. Dawson, D. Stablein, H. Sampson, G. Lack, Atopic dermatitis increases the effect of exposure to peanut antigen in dust on peanut sensitization and likely peanut allergy. *Journal of Allergy and Clinical Immunology* **135**, 164-170.e4 (2015).
2. J. M. Biagini Myers, M. G. Sherenian, A. Baatyrbek Kyzy, R. Alarcon, A. An, Z. Flege, D. Morgan, T. Gonzalez, M. L. Stevens, H. He, J. W. Kroner, D. Spagna, B. Grashel, L. J. Martin, A. B. Herr, G. K. Khurana Hershey, Events in Normal Skin Promote Early-Life Atopic Dermatitis—The MPAACH Cohort. *J Allergy Clin Immunol Pract* **8**, 2285-2293.e6 (2020).
3. Q. Hamid, M. Boguniewicz, D. Y. M. Leung, Differential in situ cytokine gene expression in acute versus chronic atopic dermatitis. *J Clin Invest* **94**, 870–876 (1994).
4. L. M. Bartnikas, M. F. Gurish, O. T. Burton, S. Leisten, E. Janssen, H. C. Oettgen, J. Beaupré, C. N. Lewis, K. F. Austen, S. Schulte, J. L. Hornick, R. S. Geha, M. K. Oyoshi, Epicutaneous sensitization results in IgE-dependent intestinal mast cell expansion and food-induced anaphylaxis. *Journal of Allergy and Clinical Immunology* **131**, 451-460.e6 (2013).
5. S. J. Galli, M. Tsai, IgE and mast cells in allergic disease. *Nature Medicine 2012 18:5* **18**, 693–704 (2012).
6. K. J. L. Jackson, Y. Wang, A. M. Collins, Human immunoglobulin classes and subclasses show variability in VDJ gene mutation levels. *Immunol Cell Biol* **92**, 729–733 (2014).
7. Y. Wang, K. J. L. Jackson, J. Davies, Z. Chen, B. A. Gaeta, J. Rimmer, W. A. Sewell, A. M. Collins, IgE-Associated IGHV Genes from Venom and Peanut Allergic Individuals Lack Mutational Evidence of Antigen Selection. *PLoS One* **9**, e89730 (2014).
8. S. C. A. Nielsen, K. M. Roskin, K. J. L. Jackson, S. A. Joshi, P. Nejad, J. Y. Lee, L. E. Wagar, T. D. Pham, R. A. Hoh, K. D. Nguyen, H. Y. Tsunemoto, S. B. Patel, R. Tibshirani, C. Ley, M. M. Davis, J. Parsonnet, S. D. Boyd, Shaping of infant B cell receptor repertoires by environmental factors and infectious disease. *Sci Transl Med* **11**, eaat2004 (2019).
9. J. M. Biagini Myers, M. G. Sherenian, A. Baatyrbek Kyzy, R. Alarcon, A. An, Z. Flege, D. Morgan, T. Gonzalez, M. L. Stevens, H. He, J. W. Kroner, D. Spagna, B. Grashel, L. J. Martin, A. B. Herr, G. K. Khurana Hershey, Events in Normal Skin Promote Early-Life Atopic Dermatitis—The MPAACH Cohort. **8**, 2285-2293.e6 (2020).
10. J. F. Bach, The hygiene hypothesis in autoimmunity: the role of pathogens and commensals. *Nature Reviews Immunology 2017 18:2* **18**, 105–120 (2017).
11. J. J. S. Chafen, S. J. Newberry, M. A. Riedl, D. M. Bravata, M. Maglione, M. J. Suttorp, V. Sundaram, N. M. Paige, A. Towfigh, B. J. Hulley, P. G. Shekelle, Diagnosing and Managing Common Food Allergies. *JAMA* **303**, 1848 (2010).
12. R. S. Gupta, M. M. Walkner, M. Greenhawt, C. H. Lau, D. Caruso, X. Wang, J. A. Pongracic, B. Smith, Food Allergy Sensitization and Presentation in Siblings of Food Allergic Children. *J Allergy Clin Immunol Pract* **4**, 956–962 (2016).
13. M. Ghraichy, J. D. Galson, A. Kovaltsuk, V. von Niederhäusern, J. Pachlopnik Schmid, M. Recher, A. J. Jauch, E. Miho, D. F. Kelly, C. M. Deane, J. Trück, Maturation of the Human Immunoglobulin Heavy Chain Repertoire With Age. *Front Immunol* **11**, 1734 (2020).

14. J. M. Biagini Myers, N. Wang, G. K. Lemasters, D. I. Bernstein, T. G. Epstein, M. A. Lindsey, M. B. Ericksen, R. Chakraborty, P. H. Ryan, M. S. Villareal, J. W. Burkle, J. E. Lockey, T. Reponen, G. K. Khurana Hershey, Genetic and environmental risk factors for childhood eczema development and allergic sensitization in the CCAAPS cohort. *J Invest Dermatol* **130**, 430–437 (2010).
15. J. E. Gern, C. L. Reardon, S. Hoffjan, D. Nicolae, Z. Li, K. A. Roberg, W. A. Neaville, K. Carlson-Dakes, K. Adler, R. Hamilton, E. Anderson, S. Gilbertson-White, C. Tisler, D. DaSilva, K. Anklam, L. D. Mikus, L. A. Rosenthal, C. Ober, R. Gangnon, R. F. Lemanske, Effects of dog ownership and genotype on immune development and atopy in infancy. *Journal of Allergy and Clinical Immunology* **113**, 307–314 (2004).
16. M. P. Lefranc, V. Giudicelli, P. Duroux, J. Jabado-Michaloud, G. Folch, S. Aouinti, E. Carillon, H. Duvergey, A. Houles, T. Paysan-Lafosse, S. Hadi-Saljoqi, S. Sasorith, G. Lefranc, S. Kossida, IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Res* **43**, D413–D422 (2015).
17. T. Rogosch, S. Kerzel, K. Hoß, G. Hoersch, C. Zemlin, M. Heckmann, C. Berek, H. W. Schroeder, R. F. Maier, M. Zemlin, IgA Response in Preterm Neonates Shows Little Evidence of Antigen-Driven Selection. *The Journal of Immunology* **189**, 5449–5456 (2012).
18. J. Ridings, L. Dinan, R. Williams, D. Roberton, H. Zola, Somatic mutation of immunoglobulin VH6 genes in human infants. *Clin Exp Immunol* **114**, 33 (1998).
19. D. P. Strachan, Hay fever, hygiene, and household size. *BMJ : British Medical Journal* **299**, 1259 (1989).
20. J. A. Geoghegan, A. D. Irvine, T. J. Foster, Staphylococcus aureus and Atopic Dermatitis: A Complex and Evolving Relationship. *Trends Microbiol* **26**, 484–497 (2018).
21. M. J. Ege, R. Frei, C. Bieli, D. Schram-Bijkerk, M. Waser, M. R. Benz, G. Weiss, F. Nyberg, M. van Hage, G. Pershagen, B. Brunekreef, J. Riedler, R. Lauener, C. Braun-Fahrländer, E. von Mutius, Not all farming environments protect against the development of asthma and wheeze in children. *Journal of Allergy and Clinical Immunology* **119**, 1140–1147 (2007).
22. R. L. Peters, K. J. Allen, S. C. Dharmage, J. J. Koplin, T. Dang, K. P. Tilbrook, A. Lowe, M. L. K. Tang, L. C. Gurrin, Natural history of peanut allergy and predictors of resolution in the first 4 years of life: A population-based assessment. *Journal of Allergy and Clinical Immunology* **135**, 1257-1266.e2 (2015).
23. R. A. Wood, S. H. Sicherer, B. P. Vickery, S. M. Jones, A. H. Liu, D. M. Fleischer, A. K. Henning, L. Mayer, A. W. Burks, A. Grishin, D. Stablein, H. A. Sampson, The natural history of milk allergy in an observational cohort. *Journal of Allergy and Clinical Immunology* **131**, 805-812.e4 (2013).
24. S. H. Sicherer, R. A. Wood, B. P. Vickery, S. M. Jones, A. H. Liu, D. M. Fleischer, P. Dawson, L. Mayer, A. W. Burks, A. Grishin, D. Stablein, H. A. Sampson, The natural history of egg allergy in an observational cohort. *Journal of Allergy and Clinical Immunology* **133**, 492-499.e8 (2014).
25. D. Venkataraman, M. Erlewyn-Lajeunesse, R. J. Kurukulaaratchy, S. Potter, G. Roberts, S. Matthews, S. H. Arshad, Prevalence and longitudinal trends of food allergy during childhood and adolescence: Results of the Isle of Wight Birth Cohort study. *Clinical & Experimental Allergy* **48**, 394–402 (2018).
26. H. Dai, F. Wang, L. Wang, J. Wan, Q. Xiang, H. Zhang, W. Zhao, W. Zhang, An epidemiological investigation of food allergy among children aged 3 to 6 in an urban area of Wenzhou, China. *BMC Pediatr* **20**, 1–8 (2020).
27. I. Dahlke, D. J. Nott, J. Ruhno, W. A. Sewell, A. M. Collins, Antigen selection in the IgE response of allergic and nonallergic individuals. *Journal of Allergy and Clinical Immunology* **117**, 1477–1483 (2006).
28. M. A. Berkowska, J. J. Heeringa, E. Hajdarbegovic, M. van der Burg, H. B. Thio, P. M. van Hagen, L. Boon, A. Orfao, J. J. M. van Dongen, M. C. van Zelm, Human IgE+ B cells are derived from T cell–dependent and T cell–independent pathways. *Journal of Allergy and Clinical Immunology* **134**, 688-697.e6 (2014).
29. J. M. Hanifin, G. Rajka, Diagnostic features of atopic dermatitis. *Acta Derm. Venereol.* **92**, 44–47 (1980).
30. L. B. von Kobyletzki, A. Berner, F. Carlstedt, M. Hasselgren, C. G. Bornehag, A. Svensson, Validation of a Parental Questionnaire to Identify Atopic Dermatitis in a Population-Based Sample of Children up to 2 Years of Age. *Dermatology* **226**, 222–226 (2013).
31. J. F. Stalder, A. Taïeb, D. J. Atherton, P. Bieber, E. Bonifazi, A. Broberg, A. Calza, R. Coleman, Y. de Prost, J. F. Stalder, C. Gelmetti, A. Cuannetti, J. Harper, B. Künz, J. M. Lachapelle, T. Langeland, R. Lever, A. P. Oranje, C. Oueille-Roussel, J. Revuz, J. Ring, J. C. Roujeau, J. H. Saurat, M. Song, D. Tennstedt, D. van Neste, D. Vieluf, M. Poncet, Severity scoring of atopic dermatitis: The SCORAD index: Consensus report of the european task force on atopic dermatitis. *Dermatology* **186**, 23–31 (1993).
32. M. L. Stevens, T. Gonzalez, E. Schauberger, A. Baatyrbek kyzy, H. Andersen, D. Spagna, M. K. Kalra, L. J. Martin, D. Haslam, A. B. Herr, J. M. Biagini Myers, G. K. Khurana Hershey, Simultaneous skin biome and keratinocyte genomic capture reveals microbiome differences by depth of sampling. *Journal of Allergy and Clinical Immunology* **146**, 1442–1445 (2020).

33. J. J. M. van Dongen, A. W. Langerak, M. Brüggemann, P. A. S. Evans, M. Hummel, F. L. Lavender, E. Delabesse, F. Davi, E. Schuuring, R. García-Sanz, J. H. J. M. van Krieken, J. Droese, D. González, C. Bastard, H. E. White, M. Spaargaren, M. González, A. Parreira, J. L. Smith, G. J. Morgan, M. Kneba, E. A. Macintyre, Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* **17**, 2257–2317 (2003).
34. K. M. Roskin, N. Simchoni, Y. Liu, J.-Y. Lee, K. Seo, R. A. Hoh, T. Pham, J. H. Park, D. Furman, C. L. Dekker, M. M. Davis, J. A. James, K. C. Nadeau, C. Cunningham-Rundles, S. D. Boyd, IgH sequences in common variable immune deficiency reveal altered B cell development and selection. *Sci Transl Med* **7**, 302ra135 (2015).
35. K. M. Roskin, K. J. L. Jackson, J. Lee, R. A. Hoh, S. A. Joshi, K. Hwang, M. Bonsignori, I. Pedroza-Pacheco, H. Liao, M. A. Moody, A. Z. Fire, P. Borrow, B. F. Haynes, S. D. Boyd, Aberrant B cell repertoire selection associated with HIV neutralizing antibody breadth. *Nat Immunol* **21**, 199–209 (2020).
36. T. Magoč, S. L. Salzberg, FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
37. J. Ye, N. Ma, T. L. Madden, J. M. Ostell, IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res* **41**, W34–W40 (2013).
38. M. Steinegger, J. Söding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**, 1026–1028 (2017).
39. C. E. Shannon, A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379–423 (1948).
40. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2019).
41. RStudio Team, RStudio: Integrated Development Environment for R (2016) (available at http://www.rstudio.com/).
42. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2009; http://ggplot2.org).
43. A. Kassambara, ggpubr: "ggplot2" Based Publication Ready Plots (2020) (available at https://cran.r-project.org/package=ggpubr).
44. K. B. and A. Lun, PCAtools: PCAtools: Everything Principal Components Analysis (2022).
45. K. J. L. Jackson, M. J. Kidd, Y. Wang, A. M. Collins, The shape of the lymphocyte receptor repertoire: Lessons from the B cell receptor. *Front Immunol* **4**, 263 (2013).
46. S. Tonegawa, Somatic generation of antibody diversity. *Nature* **302**, 575–581 (1983).

Acknowledgments:

The authors thank all the children and their families who participated in the MPAACH cohort, the Schubert Research Clinic of Cincinnati Children's Hospital Medical Center (CCHMC) for assistance with research participants, the CCHMC DNA Sequencing and Genotyping Core in particular Brian Quinn for help optimizing library construction, the CCHMC Information Services for Research (IS4R) group for hosting the data storage and processing infrastructure.

Funding:

National Institutes of Health (NIH) National Institute of Allergy and Infectious Diseases (NIAID) grant U19 AI070235 (GKKH, JBM, LJM, WCC).

NIAID Asthma and Allergic Diseases Cooperative Research Centers (AADCRC) Opportunity Fund (KMR). AADCRC Opportunity Fund (SA).

Cincinnati Children's Hospital Medical Center for Pediatric Genomics (CpG) Pilot Grant (KMR).

Food Allergy Research and Education (FARE) Biobank and Biomarker Discovery Center (BBDC) (AHA).

Author contributions:

Conceptualization: KG, SA, GKKH, KMR
Methodology: KG, CM, ON, MLS, JMB, LJM, SA, KMR
Investigation: KG, CM, ON, JWK, WCC, ABK, JTS, AHA, KMR
Visualization: KG, KMR
Funding acquisition: SA, GKKH, KMR
Project administration: JWK,WCC, MLS, JMB, KMR
Supervision: ALD, LK, JTS, AHA, GKKH, LJM, KMR
Writing, original draft: KG, SA, GKKH, KMR
Writing, review & editing: KG, CM, ON, JWK, WCC, MLS, ABK, JMB, ALD, LK, JTS, AMA, LJM, SA, GKKH, KMR

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data associated with this study will be available from the NIH Short Read Archive (SRA) under BioProject PRJNA857098 on publication. All python and R code is available upon request.