

Supplementary Note

Table of contents

1.	BAMSE	2
2.	Copenhagen Prospective Studies on Asthma in Childhood (COPSAC)	4
3.	Estonian Biobank (EBB)	6
4.	EGEA2	7
5.	Generation R	10
6.	GINIplus/LISA north and south	12
7.	IDEFICS/I.Family	16
8.	INfancia y Medio Ambiente – Environment and Childhood (INMA)	18
9.	Isle of Wight Birth Cohort (IOWBC)	20
10.	Cooperative Health Research in the Region of Augsburg (KORA)	21
11.	Lifelines	23
12.	PASTURE	24
13.	PIAMA	25
14.	SHIP	27
15.	The Raine Study	29
16.	The Cardiovascular Risk in Young Finns Study (YFS)	31
17.	CLARA/CLAUS	33
18.	EXCEED	34
19.	GENEVA	35
20.	German Genetics of Food Allergy Study (GOFA)	36
21.	HealthNuts	37
22.	Manchester Asthma and Allergy Study (MAAS)	39
23.	UK Biobank	40

1. BAMSE

Recruitment:

BAMSE (Swedish abbreviation for Child (Barn), Allergy, Milieu, Stockholm, Epidemiology) is an ongoing population-based birth cohort that recruited 4,089 infants born in 1994-1996 in Stockholm (Sweden) (1). Information related to environmental exposures and allergy-related symptoms was obtained from parents through questionnaires two months after the birth of their children (baseline). Participants were then followed up during childhood and adolescence at the ages of 1, 2, 4, 8, 12, 16, and 24 years with a thorough data collection regarding environmental exposures, lifestyle, socioeconomic status, allergy-related, eczema, rhinitis, and respiratory symptoms, and medication use. An extensive clinical examination was performed by trained personnel when participants were 4, 8, 16, and 24 years old. This included spirometry tests and the collection of blood samples (2-4). At 4 and 24 years of age, urine was also collected, as well as nasal brushing was carried out at the 24-year-old follow-up visit. Fractional exhaled nitric oxide (FeNO) was measured at 16 and 24 years of age. Moreover, allergen-specific immunoglobulin (IgE) antibodies were analyzed using the ImmunoCAP fx5® System (Thermo Fisher Scientific, Uppsala, Sweden) following the manufacturer's recommendations. IgE reactivity to a mix of five common food allergens (including egg white, milk, codfish, wheat, peanut, and soybean) was assessed using serum obtained from blood samples collected at the 4, 8, 16, and 24-year-old follow-up visits (5-9).

This study was approved by the Regional Ethics Committee at Karolinska Institutet in Stockholm, Sweden (Dnr 2016/1380-31/2). All participants or their legal guardians provided written informed consent under the Declaration of Helsinki.

Case/control definition:

At 1 and 2 years of age, parent-reported food allergy was considered for any experience of symptoms (such as vomiting, diarrhea, eczema, urticaria, itch or swelling of the lips and/or eyelids, runny nose, and/or asthma) in response to food or drinks. Nonetheless, reported food allergy was defined as symptoms in response to at least one specific food item (milk, egg, fish, tree nuts, peanuts, peas, soy, wheat, and/or fruit with pits/pips) at the 4, 8, 12, and 16-year-old data collections. At 24 years of age, it was considered as any self-reported reaction experienced from eating a type of food (10). Reports of a doctor's diagnosis of food allergy were collected at all follow-up visits. On the other hand, positive sensitization to food allergens at 4, 8, 16, and 24 years of age was defined as IgE levels ≥ 0.35 kUA/l resulting from the ImmunoCAP fx5® assay (1).

The different food allergy-related phenotypes were separately defined in children, using information from when participants were aged up to 16 years, and adults (24 years of age). For self-reported food allergy and doctors-diagnosed food allergy phenotype, cases were

defined as parent/self-reported or doctor's diagnosis of food allergy, respectively, at least one time point, whereas controls never provided any positive report of food allergy. Participants with positive allergic sensitization to food allergens ($\text{IgE} \geq 0.35 \text{ kUA/l}$) at any of the time points assessed were considered cases for food-specific sensitization phenotype. Food allergy plus food-specific sensitization cases were defined as subjects with a parent/self-reported or doctor's diagnosis of food allergy at least one time point, and food-allergen positive sensitization.

Genotyping and imputation:

DNA was extracted from blood samples collected at the age of 8 and 16 years (11). Genotyping was carried out in two subsets of participants, hereon referred to as "Waves" with the Illumina Human 610-quad array (Illumina, Inc.) (Wave 1, $n=505$) and the Illumina Infinium Global Screening Array-24 v1.0 BeadChip (Illumina, Inc.) (Wave 2, $n=2,387$) (12, 13).

Data from each Wave were analyzed independently given the differences in the genotyping platform used.

Quality control (QC) analyses on subjects and genetic variants were conducted in each genotyping Wave using PLINK 2.0 (14, 15). Participants with call rate (CR) $<98\%$, heterozygosity rates higher or lower than 4 standard deviations of the population mean, sex discordances, and/or evidence of relatedness for at least second-degree relatives ($\text{PIHAT} \geq 0.2$) were discarded. Duplicated samples included in both genotyping Waves were excluded as well, retaining them only in the Wave where they showed higher CR. Subjects identified as outliers of the distribution of Europeans through the comparison of their genetic ancestry with populations included in the 1,000 Genome Project reference panel (16) through a Principal Component (PC) analysis were excluded. As a result, 2,636 samples passed QC procedures (Wave 1, $n=463$; Wave 2, $n=2,173$). On the other hand, genetic variants with CR $<95\%$, deviations from the Hardy-Weinberg Equilibrium ($p < 1 \times 10^{-6}$), and minor allele frequency (MAF) <0.01 were discarded from further analyses using the Rapid Imputation for COnsortias PIpeLLine (RICOPILI) (12, 17). The QC procedures described above were carried out using the high-performance UPPMAX computational facility (Uppsala Universitet, Sweden).

Genetic variants were imputed in each Wave using the Positional Burrows-Wheeler Transform (PBWT) imputation software (18) and the Haplotype Reference Consortium (HRC) r1.1 reference panel (19) through the Sanger Imputation Server. EAGLE2 (20) was used for haplotype phasing.

Statistical analyses:

The association of high-quality single nucleotide polymorphisms (SNPs) (MAF ≥ 0.01 ; imputation quality INFO ≥ 0.5) located at autosomal chromosomes with each food-allergy-related phenotype was separately evaluated in children and adults. Subjects with available genome-wide genotype data and the information necessary to define each phenotype were included in this study. Logistic regressions were independently conducted in each genotyping Wave using PLINK 2.0. Sex and three PCs of genetic ancestry were included as covariates. Additional adjustment by hay fever was also applied, which was defined following the same approach as for the different food allergy-related phenotypes described above.

Acknowledgements and Funding:

We thank all BAMSE participants and their parents, the nurses, and all of the staff involved in the study. Reagents for the allergen-specific IgE analyses were kindly provided by Thermo Fisher Scientific. The genome-wide association study (GWAS) quality control procedures were enabled by resources in the project sens2017589 provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at UPPMAX, funded by the Swedish Research Council through grant agreement no. 2022-06725.

This study was supported by grants from the Swedish Research Council (2016-03086, 2018-02524, 2019-01060, 2020-02170); the Swedish Research Council for Health, Working Life and Welfare (2017-00526); Formas (2016-01646); the Swedish Heart-Lung Foundation; Strategic Research Area (SFO) Epidemiology; Karolinska Institutet and Region Stockholm (ALF project, and for cohort and database maintenance); Swedish Asthma and Allergy Association's Research Foundation; and The Cancer and Allergy Foundation. NH-P was supported by a Medium-Term Research Fellowship by the European Academy of Allergy and Clinical Immunology (EAACI) and a Long-Term Research Fellowship by the European Respiratory Society (ERS) (LTRF202101-00861).

2. Copenhagen Prospective Studies on Asthma in Childhood (COPSAC)

Recruitment:

The COPSAC study comprise in total 918 children with complete genotype information from the COPSAC2000 and COPSAC2010 birth cohorts. COPSAC2000 is a mother–child cohort where all the mothers had a history of a doctor's diagnosis of asthma after 7 years of age and thus this is a high-risk asthma cohort comprising 411 children. Newborns were enrolled in the first month of life, and details on the cohort have been described previously (21). COPSAC2010 is a mother–child cohort comprising 700 children born to unselected mothers from Denmark as described previously in detail (22).

Case/control definition:

Self-reported FA: Case status = Individuals who ever reported self-reported food allergy until 6/12 years (C2010/C2000)

Control status= Individuals who never reported self-reported food allergy until 6/12 years of age. OBS: this also includes reported allergy to fruits that may only be skin reaction.

Sensitization: Case status = Individuals who have a food specific allergic sensitization at 6 years; Control status = Individuals who do not have a food specific allergic sensitization at 6 years. Includes SPT ≥ 3 mm and/or sIgE ≥ 0.35 kU/L.

Self-reported FA + sensitization: Case status = Individuals who reported self-reported food allergy up to 6/12 years (COPSAC2010 / COPSAC2000) and have a food specific allergic sensitization at 6 years; Control status = Individuals who did not report self-reported food allergy at 6/12 years (COPSAC2010 / COPSAC2000) and do not have a food specific allergic sensitization at 6 years. 2=participants with inconcurrent reports and sensitization (neither 1 nor 0, but not missing either).

Genotyping and imputation:

See details in Supplementary Data.

Statistical analyses:

We conducted all phenotype associations using logistic regression available within Plink 2.0 software (--glm) using a minor allele frequency threshold of 1% (--maf 0.01). In the basic adjustment model we included sex, age and five principal components as covariates. In the 'hayfever model' we adjusted for symptoms of allergic rhinitis and matching inhalant allergic sensitization (SPT and/or sIgE) at 6-7 years of age.

Acknowledgements and Funding:

We express our deepest gratitude to the children and families of the COPSAC2000 and COPSAC2010 cohort studies for all their support and commitment. We also acknowledge and the unique efforts carried out by the whole COPSAC research team. All funding received by COPSAC is listed on www.copsac.com. The Lundbeck Foundation (Grant no R16-A1694); The Danish Ministry of Health (Grant no 903516); Danish Council for Strategic Research (Grant no 0603-00280B) and The Capital Region Research Foundation have provided core support to the COPSAC research center.

Ethical approval:

All human research was approved by the relevant institutional review boards and ethical committees and conducted according to the Declaration of Helsinki. All participants and/or their parents in the clinical studies provided written and oral informed consent. COPSAC2000 and COPSAC2010 were approved by the Danish Scientific Ethics Committee, Region H.

3. Estonian Biobank (EBB)

Recruitment:

The Estonian Biobank is a population-based biobank with 212,955 participants in the current data freeze (2023v1). All biobank participants have signed a broad informed consent form and information on ICD-10 codes is obtained via regular linking with the national Health Insurance Fund and other relevant databases, with majority of the electronic health records having been collected since 2004 (23).

Case/control definition:

The diagnosis of “Anaphylactic reaction due to food” was based on ICD-code T78.0 (n=105), and participants without the diagnosis in the biobank records were used as controls (n=205,317).

Genotyping and imputation:

All EstBB participants have been genotyped at the Core Genotyping Lab of the Institute of Genomics, University of Tartu, using Illumina Global Screening Array v3.0_EST. Samples were genotyped and PLINK format files were created using Illumina GenomeStudio v2.0.4. Individuals were excluded from the analysis if their call-rate was < 95%, if they were outliers of the absolute value of heterozygosity (> 3SD from the mean) or if sex defined based on heterozygosity of X chromosome did not match sex in phenotype data [PMID:28401899]. Before imputation, variants were filtered by call-rate < 95%, HWE p-value < 1e-4 (autosomal variants only), and minor allele frequency < 1%. Genotyped variant positions were in build 37. Phasing was performed using the Beagle v5.4 software (24). Imputation was performed with Beagle v5.4 software (beagle.22Jul22.46e.jar) and default settings. Dataset was split into batches 5,000. A population specific reference panel consisting of 2,695 WGS samples was utilized for imputation and standard Beagle hg38 recombination maps were used. Based on principal component analysis, samples who were not of European ancestry were removed. Duplicate and monozygous twin detection was performed with KING 2.2.7 (25), and one sample was removed out of the pair of duplicates.

Statistical analyses:

Association analysis in Estonian biobank was carried out for all variants with an INFO score > 0.4 using the additive model as implemented in REGENIE v3.0.3 with standard binary trait settings (26). Logistic regression was carried out with adjustment for current age, age², sex and 10 PCs as covariates, analyzing only variants with a minimum minor allele count of 2.

Acknowledgements and Funding:

We want to acknowledge the participants of the Estonian Biobank for their contributions. The Estonian Genome Center analyses were partially carried out in the High Performance Computing Center, University of Tartu.

The work of the Estonian Genome Center, University of Tartu was funded by the European Union through Horizon 2020 research and innovation program under grants no. 810645 and 894987, through the European Regional Development Fund projects GENTRANSMED (2014-2020.4.01.15-0012), MOBEC008, MOBERA21 and Estonian Research Council Grant PRG1291.

4. EGEA2

Recruitment:

The French Epidemiological study on the Genetics and Environment of Asthma (EGEA) cohort combines an initial group of asthma cases recruited in chest clinics with their first-degree relatives (388 families, n=1,705), and a group of population-based participants followed-up over 20 years (n=415). All subjects were born in France and were of European ancestry. The protocol of this study has been described elsewhere (27, 28). The present analysis uses 927 participants with specific IgE measurements at EGEA2 and genotyped data (EGEA2: 2003-2007) (29, 30).

Written informed consent was obtained from all adult participants and child's legal guardians at both surveys. Ethical approval was obtained from the relevant institutional review board committees (Cochin Port-Royal Hospital and Necker-Enfants Malades Hospital, Paris: n° 01-07-07, 04-05-03, 04-11-13 and 04-11-18).

Case/control definition:

Self-reported FA: Self-reported food allergy was defined by a positive answer to “Are you allergic to any food?”

Sensitization: IgE reactivity to micro-arrayed allergens was determined by the MeDALL-chip (multiplex micro-array), containing 176 aero- and food-allergen components (31). For the purpose of this analysis, IgE responses to 75 food-allergens components were considered (Fruits: 7; nuts: 26; vegetables: 4; flour: 17; cow's milk: 12; egg white: 3; egg yolk/chicken: 1;

seafood: 5). Positivity of each food allergen-specific IgE was defined with a threshold of 0.3 ISU. Participants with at least one positive food allergen-specific IgE were defined as cases for the food-specific allergic sensitization phenotype whereas those without any positive food allergen-specific IgE were defined as controls.

Self-reported FA + sensitization: For this phenotype, participants who both ever self-reported food allergy and were with positive food-specific allergic sensitization were defined as cases whereas participants who never reported food allergy and do not have a food-specific allergic sensitization were defined as control.

Allergic rhinitis ever was defined by a positive answer to “Have you ever had allergic rhinitis?” or “Have you ever had hay fever?”.

Genotyping and imputation:

EGEA samples were genotyped using the Illumina 610-Quad array (Illumina, San Diego, CA), as part as of the European Gabriel asthma consortium GWAS. Stringent quality criteria (QC) were applied to select both individuals and SNPs and have been previously detailed (32). In summary, subjects were excluded if they had more than 3% of missing genotypes and were of putative non-European descent based on deviation from the mean ($>\pm 3$ standard deviations (SD)) of the first two ancestry-informative principal components. Cryptic relatedness was detected by identity-by-state analysis and known family relationships in EGEA were confirmed. Subjects with sex inconsistencies between genetic and reported data, heterozygosity rate less than 0.30 or greater than 0.33, and duplicates were also excluded. SNP inclusion criteria were: (1) genotype missing rate $< 3\%$ in both cases and controls; (2) minor allele frequency (MAF) $\geq 5\%$ in controls; (3) consistency with Hardy-Weinberg equilibrium by a 1 degree-of-freedom goodness-of-fit test in controls ($P > 10^{-4}$). After this QC, a total of 500,978 genotyped SNPs were available in EGEA dataset.

Prior to the imputation process, we applied an additional filtering by excluding the ambiguous SNPs (A/T or C/G), and the SNPs with different alleles or with a difference in allele frequencies greater than 0.2 compared to the reference panel. We used Minimac4 software and the 1000 Genomes reference panel (European samples - EUR) to perform genome-wide genotype imputation. We kept for association analyses bi-allelic SNPs with imputation quality score (Rs_q) ≥ 0.5 and minor allele frequency (MAF) $\geq 1\%$, making a total of 8 million SNPs.

Statistical analyses:

We performed a univariate association analysis between individual SNPs and each phenotype. This analysis was based on a logistic regression model assuming an additive model for SNP effect. This model was adjusted for age, sex and the first four genetic principal components computed using EIGENSTRAT2.0 software to account for population structure. We took into

account familial dependencies using the cluster and robust options of the logit function in Stata® V14.1. Test of SNP effect was based on a Wald-test.

Acknowledgements and Funding:

The authors thank all those who participated to the setting of the study and on the various aspects of the examinations of the individuals. The authors are grateful to the three CIC-Inserm of Necker, Grenoble and Marseille who supported the EGEA study and in which subjects were examined. They are also grateful to the biobanks in Lille (CIC-Inserm) and at Annemasse (Etablissement français du sang) where EGEA biological samples were/are stored. They are indebted to all the individuals who participated, without whom the study would not have been possible.

We thank the EGEA cooperative group members: Coordination: V Siroux (epidemiology, PI since 2013); F Demenais (genetics); I Pin (clinical aspects); R Nadif (biology); F Kauffmann (PI 1992-2012). Respiratory epidemiology: Inserm ex-U 700, Paris: M Korobaeff (Egea1), F Neukirch (Egea1); Inserm ex-U 707, Paris: I Annesi-Maesano (Egea1-2); Inserm U 1018, Villejuif: O Dumas, F Kauffmann, N Le Moual, R Nadif, MP Oryszczyn (Egea1-2), R Varraso; Inserm U 1209 Grenoble: J Lepeule, V Siroux. Genetics: Inserm ex-U 393, Paris: J Feingold; Inserm UMR 1124, Paris: E Bouzigon, MH Dizier, F Demenais; CNG, Evry: I Gut (now CNAG, Barcelona, Spain), M Lathrop (now Univ McGill, Montreal, Canada). Clinical centers: Grenoble: I Pin, C Pison; Lyon: D Ecochard (Egea1), F Gormand, Y Pacheco; Marseille: D Charpin (Egea1), D Vervloet (Egea1-2); Montpellier: J Bousquet; Paris Cochin: A Lockhart (Egea1), R Matran (now in Lille); Paris Necker: E Paty (Egea1-2), P Scheinmann (Egea1-2); Paris-Trousseau: A Grimfeld (Egea1-2), J Just. Data and quality management: Inserm ex-U155, Paris: J Hochez (Egea1); Inserm U 1018, Villejuif: N Le Moual, L Orsi; Inserm ex-U780, Villejuif: C Ravault (Egea1-2); Inserm ex-U794, Evry: N Chateigner (Egea1-2); Inserm UMR 1124, Paris: H Mohamdi; Inserm U1209, Grenoble: A Boudier, J Quentin (Egea1-2).

This work was supported by the French National Agency for Research (ANR-19-CE36-0005 Nirvana, ANR-20-CE36-0009 GenCAST), Université Paris Cité doctoral fellowship. Genotyping was supported by grants from the European Commission (No. LSHB-CT-2006-018996-GABRIEL) and allergen-specific IgE measurements were supported by grants from Inserm Aviesan Itmo sante publique, the Scientific committee “AGIR for chronic diseases,” and ANR-PRSP 2009.

Ethical approval

Appropriate ethical approval was obtained for all cohorts by their ethics committees as detailed in the individual cohort methods.

5. Generation R

Recruitment:

The Generation R Study is a population-based prospective cohort study from fetal life until adulthood. The study is designed to identify early environmental and genetic causes and causal pathways leading to normal and abnormal growth, development and health from fetal life, childhood and young adulthood onwards (33).

Case/control definition:

Information on food allergy was obtained from questionnaires during follow-up visits from birth until age 9 years. We defined three different food allergy phenotypes:

1. Participants whose parents have reported a doctor diagnosed any food allergy by having an affirmative response to “Is your child allergic to any food – Yes, diagnosed by doctor.”
2. Food specific allergic sensitization was present if the skin prick test was positive for any food allergen.
3. Participants whose parents have reported either a doctor diagnosed or self-reported food allergy and have food-specific allergic sensitization. This phenotype was a combination of both 1 and 2, where 1 could be also positive when the food allergy was not diagnosed by a doctor.

Occurrence of hay fever was defined as an affirmative response to the question “Is your child allergic to pollen (hay fever) – Yes, diagnosed by doctor?”.

Genotyping and imputation:

The Generation R Study was genotyped in two waves in the Genetic Laboratory of the Department of Internal Medicine (www.glimdna.org). The first and largest subset of participants was genotyped using Human610K and Human 660W genotyping arrays whereas the second was genotyped using the GSA-MD v2 genotyping arrays (Illumina, Inc., San Diego, CA, USA). Phasing, imputation, and quality control steps for the first subset (GENR3) are described in detail elsewhere (34). The second subset phasing, imputation, and quality control steps are detailed in a recent publication (35). Overall, the GWAS dataset underwent a stringent QC process whereby variants with call rate below 97.5 % and Hardy-Weinberg equilibrium (HWE) deviation ($p\text{-value} < 1 \times 10^{-7}$) were excluded. Also, individuals with low call rates ($< 97.5\%$), excess of heterozygosity (> 4 standard deviations from the mean), sex mismatches and/or genetic duplicates were excluded. Reported familial relationships were also leveraged to identify and exclude sample swaps. For GENR3, we used Eagle/MACH (version 1.0.15) software to phase and impute genotypes to the HRC v1.1 reference panel in the Michigan Server. For GENR4, an in-house pipeline based on minimac4/MACH was used to phase and impute the data to the 1000 Genomes Phase 3 v 5 panel. The first 20 genomic principal

components (PCs) were calculated for all participants of both datasets simultaneously. Roughly 27,000 independent and common single nucleotide polymorphisms (SNPs) (minor allele frequency > 0.05) that were in common between both subsets, and the 1000 Genomes Project panel were used in the calculation. Individuals were classified as of European background if the value for their first four PCs lied within 4 standard deviations of the mean value for all European individuals in the reference panel (35).

Statistical analyses:

We conducted GWAS in participants of European ancestry using the PLINK2 software (v2.00a3LM, 2 March 2020 release). Logistic regression models were fitted assuming an additive genetic model, using imputed genotype dosages derived from genome-wide imputation. We restricted the analysis to variants with a minor allele count of at least 20 and applied the firth-fallback option to address potential issues with complete or quasi-complete separation. Variants on autosomal chromosomes 1 through 22 were analyzed separately and results were combined in a single output.

Analyses were adjusted for the top ten PCs and additional covariates depending on the model specification. Covariates were variance-standardized where applicable. We performed two separate GWAS per phenotype:

1. Sex- and age adjusted model: including sex, age of the child, and the top ten PCs.
2. Sex-, age-, and hay fever-adjusted model: additionally adjusting for parent-reported hay fever.

All models included standard quality control measures, and analyses were limited to unrelated participants of European ancestry to minimize confounding due to population stratification. Genome-wide significance was determined using a threshold of $p < 5 \times 10^{-8}$.

Acknowledgements and Funding:

The Generation R Study is conducted by Erasmus MC, University Medical Center Rotterdam in close collaboration with the School of Law and Faculty of Social Sciences of the Erasmus University Rotterdam, the Municipal Health Service Rotterdam area, Rotterdam, the Rotterdam Homecare Foundation, Rotterdam and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond (STAR-MDC), Rotterdam. We gratefully acknowledge the contribution of children and parents, general practitioners, hospitals, midwives and pharmacies in Rotterdam. The generation and management of GWAS genotype data for the Generation R Study was done at the Human Genomics Facility, HuGe-F, housed within the Laboratory for Population Genomics of the Department of Internal Medicine at Erasmus MC. Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, The Netherlands. We thank Pascal Arp, Gaby

van Dijk, Marijn Verkerk, Samuel Ghatan, Dr. Carolina Medina-Gomez, Mila Jhamai, Lizbeth Herrera, Dr. Marjolein Peters, Dr. Linda Broer and Jard de Vries for their help in creating, managing and quality controlling the GWAS database.

The general design of the Generation R Study is made possible by financial support from the Erasmus MC, Erasmus University Rotterdam, the Netherlands Organization for Health Research and Development and the Ministry of Health, Welfare and Sport. LD received funding from the European Union's Horizon 2020 research and innovation programme (LIFECYCLE, grant agreement No 733206, 2016 EUCAN-Connect grant agreement No 824989 ATHLETE, grant agreement No 874583, LongITools, grant agreement No 874739). HL received personal funding for a PhD fellowship from the China Scholarship Council (grant number: 202004910412).

Ethical approval:

The general design, all research aims and the specific measurements in the Generation R Study have been approved by the Medical Ethical Committee of Erasmus MC, University Medical Center Rotterdam. New measurements are only introduced into the study after approval of the Medical Ethical Committee. Participants need to give written informed consent for each phase of the study (fetal, preschool, childhood and adolescence period). From the age of 12 years onwards, children must sign their own consent form, in accordance with Dutch Law. At the start of each phase, children and their parents receive written and oral information about the study. Even with consent, when the child or the parents are not willing to participate actively, specific measurements are skipped or no measurements at all are performed.

6. GINIplus/LISA north and south

Recruitment:

The influence of Life-style factors on the development of the Immune System and Allergies in East and West Germany (LISA) study is a population-based birth cohort study. A total of 3,094 healthy, full-term neonates were recruited between 1997 and 1999 in Munich, Leipzig, Wesel and Bad Honnef. The participants were not pre-selected based on family history of allergic diseases (36). A total of 5,991 mothers and their newborns were recruited into the German infant study on the influence of Nutrition intervention PLUS environmental and genetic influences on allergy development (GINIplus) between September 1995 and June 1998 in Munich and Wesel. Infants with at least one allergic parent and/or sibling were allocated to the interventional study arm investigating the effect of different hydrolysed formulas for allergy prevention in the first year of life (37). All children without a family history of allergic diseases and children whose parents did not give consent for the intervention were allocated to the non-interventional arm. Detailed descriptions of the LISA and GINIplus

studies have been published elsewhere (36, 37). DNA was collected at the age 6 and 10 years. Food-specific sensitization was measured at age 6, 10 and 15 using ImmunoCAP Fx5 food mix. For both studies, approval by the local Ethics Committees and written consent from participant's families were obtained. For the GINIplus north and LISA north cohort, which have nearly identical study designs and outcome definitions, data were pooled.

Case/control definition:

Information on ever parents-/self-report of a doctors-diagnosed food allergy was assessed via questionnaire during each follow-up with questions corresponding to each year of life since the previous follow-up covering birth up to age 17 years. Cases were defined as participants who ever reported a doctors-diagnosed food allergy at any time point. Controls were defined as individuals who never reported a doctors-diagnosed food allergy and have less than 50% missing values across the follow-ups. This leads to 150 cases and 603 controls for GINIplus/LISA north and 330 cases and 1140 controls for GINIplus/LISA south.

Food-specific allergic sensitization was measured by using the Fx5 food mix (hen's egg, milk protein, cod, soy, peanut, wheat) at 6, 10 and 15 years of age in LISA and GINIplus. Cases were defined as participants who ever had a food-specific allergic sensitization (sIgE > 0.35 kU/L) and controls were those who never had positive food-specific sensitization. This leads to 140 cases and 649 controls for GINIplus/LISA north and 351 cases and 1082 controls for GINIplus/LISA south.

Cases for phenotype 4 (doctors-diagnosed food allergy AND food-specific allergic sensitization) were defined as participants who ever reported a doctors-diagnosed food allergy and had ever had a food-specific allergic sensitization (fx5 food mix). Controls were those who never reported a doctors-diagnosed food allergy and less than 50% missing values across the follow-ups and without food-specific allergic sensitization. This leads to 38 cases and 501 controls for GINIplus/LISA north and 125 cases and 853 for GINIplus/LISA south.

Information on doctors-diagnosed hay fever was assessed using parents-/self-administered questionnaires corresponding to each year of life since the previous follow-up covering birth up to age 17 years. A hay fever ever variable was created where all participants who ever reported hay fever were considered as cases and participants who never reported hay fever are controls.

Genotyping and imputation:

Genome-wide genotyping provided by the German Research Centre for Environmental Health at the Helmholtz Zentrum München in April 2019 was performed in 883 blood samples from GINIplus/LISA north (collected in follow-up examination at age 10 years) using

the Infinium Global Screening Array GSA v2 MD (GRCh37/hg19) with GenomeStudio Version 2.0 resulting in 712,189 variants. 1,511 children from Munich were genotyped (835 (55%) children from the GINIplus study and 676 (45%) children from the LISA study). 1423 individuals (835 from the GINIplus study and 588 from the LISA study) were analysed using the Affymetrix Human SNP Array 5.0 and 88 individuals from the LISA study were analysed using Affymetrix Human SNP Array 6.0. Genotypes were called using BRLMM-P algorithm (5.0), respectively BIRDSEED V2 algorithm (6.0). In the pre-imputation quality control (38) variants on chromosome 0, insert/deletion variants, variants with low minor allele frequency (<0.01), and low call rates (<0.95) were excluded. After that, duplicated individuals, individuals with sex-mismatch, with low call rates (<0.95), with minimal heterozygosity (inbreeding coefficient 0.1), highly related individuals (identity-by-descent analysis with $Id.tresh=0.2$ and $kin.tresh=0.1$), as well as individuals belonging to non-European ancestry group (Tukey's rule based on the 1-10 eigenvectors from Principal Component Analysis), and violations of Hardy-Weinberg ($p<10^{-6}$) were removed. Finally, SNPs that deviate from Hardy-Weinberg equilibrium ($p<10^{-6}$) were removed. 792 individuals and 479,023 SNPs passed these quality control filters. To find haplotype segments that are shared by study individuals and the HRC r1.1 2016 (GRCh37/hg19), we did a genotype imputation with minimac4 1.5.7 using the Michigan Imputation Server (39). With the help of imputation, information for 792 individuals were calculated on 39,131,600 variants based on 469,304 SNPs that passed the quality control of Michigan Imputation Server.

In post-imputation processing multi-allelic markers and markers that deviate from Hardy-Weinberg equilibrium ($p<1\times 10^{-12}$ in controls) were excluded using vcftools.

Statistical analyses:

Genome-wide association testing was performed using SNPTEST (40-42). Allele effects were estimated from additive model ($-frequentist\ 1$) using expected genotype counts ($-method\ expected$). Sex was included as a covariate.

Acknowledgements and Funding:

The authors thank all the families for their participation in the GINIplus study. Furthermore, we thank all members of the GINIplus Study Group for their excellent work. The GINIplus Study group consists of the following: Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg (Heinrich J, Bröske I, Schulz H, Flexeder C, Zeller C, Standl M, Schnappinger M, Ferland M, Thiering E, Tiesler C); Department of Pediatrics, Marien-Hospital, Wesel (Berdel D, von Berg A); Ludwig-Maximilians-University of Munich, Dr von Hauner Children's Hospital (Koletzko S); Child and Adolescent Medicine, University Hospital rechts der Isar of the Technical University Munich

(Bauer CP, Hoffmann U); IUF- Environmental Health Research Institute, Düsseldorf (Schikowski T, Link E, Klümper C, Krämer U, Sugiri D). The authors thank all the families for their participation in the LISA study. Furthermore, we thank all members of the LISA Study Group for their excellent work. The LISA Study group consists of the following: Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Epidemiology, Munich (Heinrich J, Schnappinger M, Brüske I, Ferland M, Schulz H, Zeller C, Standl M, Thiering E, Tiesler C, Flexeder C); Department of Pediatrics, Municipal Hospital “St. Georg”, Leipzig (Borte M, Diez U, Dorn C, Braun E); Marien Hospital Wesel, Department of Pediatrics, Wesel (von Berg A, Berdel D, Stiers G, Maas B); Pediatric Practice, Bad Honnef (Schaaf B); Helmholtz Centre of Environmental Research – UFZ, Department of Environmental Immunology/Core Facility Studies, Leipzig (Lehmann I, Bauer M, Röder S, Schilde M, Nowak M, Herberth G, Müller J); Technical University Munich, Department of Pediatrics, Munich (Hoffmann U, Paschke M, Marra S); Clinical Research Group Molecular Dermatology, Department of Dermatology and Allergy, Technische Universität München (TUM), Munich (Ollert M, J. Grosch). The GINIplus study was mainly supported for the first 3 years of the Federal Ministry for Education, Science, Research and Technology (interventional arm) and Helmholtz Zentrum Munich (former GSF) (observational arm). The 4 year, 6 year, 10 year and 15 year follow-up examinations of the GINIplus study were covered from the respective budgets of the 5 study centres (Helmholtz Zentrum Munich (former GSF), Research Institute at Marien-Hospital Wesel, LMU Munich, TU Munich and from 6 years onwards also from IUF - Leibniz Research-Institute for Environmental Medicine at the University of Düsseldorf) and a grant from the Federal Ministry for Environment (IUF Düsseldorf, FKZ 20462296). Further, the 15 year follow-up examination of the GINIplus study was supported by the Commission of the European Communities, the 7th Framework Program: MeDALL project, and as well by the companies Mead Johnson and Nestlé. The LISA study was mainly supported by grants from the Federal Ministry for Education, Science, Research and Technology and in addition from Helmholtz Zentrum Munich (former GSF), Helmholtz Centre for Environmental Research - UFZ, Leipzig, Research Institute at Marien-Hospital Wesel, Pediatric Practice, Bad Honnef for the first 2 years. The 4 year, 6 year, 10 year and 15 year follow-up examinations of the LISA study were covered from the respective budgets of the involved partners (Helmholtz Zentrum Munich (former GSF), Helmholtz Centre for Environmental Research - UFZ, Leipzig, Research Institute at Marien-Hospital Wesel, Pediatric Practice, Bad Honnef, IUF – Leibniz Research Institute for Environmental Medicine at the University of Düsseldorf) and in addition by a grant from the Federal Ministry for Environment (IUF Düsseldorf, FKZ 20462296). Further, the 15-year follow-up examination of the LISA study was supported by the Commission of the European Communities, the 7th Framework Program: MeDALL project. The research leading to the ESCAPE results has

received funding from the European Community's Seventh Framework Program (FP7/2007-2011) under grant agreement number: 211250.

7. IDEFICS/I.Family

Recruitment:

The pan-European IDEFICS/I.Family cohort (43, 44) is a multi-center, prospective study investigating the impact of social, environmental and behavioral factors on children's health. Children were recruited through kindergarten or school settings in Belgium, Cyprus, Estonia, Germany, Hungary, Italy, Spain and Sweden. The baseline survey of the IDEFICS study (2007/2008) included 16,229 children aged between 2 and 9.9 years. Follow-up assessments were conducted as part of the IDEFICS study after two years (FU1, N = 11,043 plus 2,543 newly enrolled children) and after six years in course of the follow-up study I.Family (FU2, N = 7,117 plus 2,512 newly enrolled children). Standardized physical examinations covered a wide range of health parameters. Parents completed any questionnaires, e.g., on dietary behavior for children under 12, while adolescents aged 12 and older self-reported in FU2.

Case/control definition:

Self-reported food allergy was assessed on the basis of the medical health questionnaire at FU2 (I.Family study), where the mean (SD) age of the children was 10.9 (2.9) years. Children were categorized as having a food allergy if their parents or legal guardians stated that their child had been diagnosed as having an allergy to soya, egg white, milk protein, nuts, fish/shellfish or fruit/vegetables.

Genotyping and imputation:

DNA was extracted from saliva or blood samples following established protocols. Genotyping of 3,330 children was carried out using the UK Biobank Axiom array (Santa Clara, USA) in two separate batches (2015 and 2017). Quality control measures were applied to both sample and genotype data according to recommendations of Weale et al. (45). After visually inspecting the cumulative SNP coverage distribution, SNPs with call rates below 97%, a minor allele frequency (MAF) of less than 0.01, and those deviating from Hardy-Weinberg equilibrium ($p\text{-value} < 10^{-6}$) were excluded. Additionally, children with call rates below 97%, an unusually high heterozygosity (absolute value $> 3SD$ from the mean), with sex mismatch, and duplicates were removed from the dataset. Principal components analysis (PCA) was performed using the R package SNPRelate to identify population outliers who were subsequently removed. Children from Cyprus were not included in this initial genotyping to minimize population stratification. Genotyped variant positions were in build 37. After filtering, a total of 3,153 children and 549,039 autosomal variants remained for further analysis.

Genome-wide imputation was conducted with Minimac4, using reference haplotypes from unrelated individuals in the 1000 Genomes Project phase III v5. Related participants with a kinship coefficient > 0.0625 were removed as well as participants without phenotype information. Ultimately, 2,810,036 imputed variants from 2,332 children and adolescents were included in the analyses.

Statistical analyses:

Logistic regression (--glm) was performed using PLINK v2.00a3 SSE4.2 (18 Feb 2022) (15) adjusting for age, sex, region, and the first five principal components to account for population stratification.

Ethical approval:

The study was conducted in agreement with the Declaration of Helsinki; all procedures were approved by the local ethics committees and written and oral informed consents were obtained from the parents, their children and adolescents, respectively, as applicable. The pan-European IDEFICS/I.Family children cohort has been registered at the ISRCTN clinical trials registry (ISRCTN62310987). Children were selected for a whole-genome scan based on their participation in the individual study modules.

Funding:

The IDEFICS study was financially supported by the European Commission within the Sixth RTD Framework Programme Contract No. 016181 (FOOD); the I.Family study was funded by the European Commission within the Seventh RTD Framework Programme Contract No.266044 (KBBE 2010-14). Participating partners have contributed their own resources to the genotyping of children.

Acknowledgements:

This research was done on behalf of the IDEFICS/I.Family consortia. The authors wish to thank the children and their parents for participating in this extensive examination. We are grateful for the support of school boards, head teachers and communities, and for the effort of the study nurses, interviewers, laboratory technicians and data managers, especially Claudia Brünings-Kuppe and Kathrin Günther.

8. Infancia y Medio Ambiente – Environment and Childhood (INMA)

Recruitment:

The INfancia y Medio Ambiente – Environment and Childhood (INMA) (46) study is a prospective, population-based cohort study conducted in Spain. It aims to investigate the impact of environmental pollutants in air, water, and diet during pregnancy and early childhood on child growth, health, and development. Further details about the INMA project can be found on our website <http://www.proyectoinma.org/>. This particular analysis uses the cohorts of INMA Sabadell, Valencia, and Gipuzkoa. The study received approval from the ethical committees of the participating centers, and all participants provided written informed consent.

Case/control definition:

Food allergy was evaluated through questionnaires during follow-up visits from birth until 9 years of age. Children classified as food allergy cases were those who had experienced at least one instance of food allergy during any of the follow-up visits. Specifically, occurrence of food allergy was defined as an affirmative response to any of the following questions: “Has your child ever had a reaction to any food?”, “Has your child had any reactions with any food since last interview?”, “Has your child ever had one or several abnormal reactions after eating a particular food?”.

Occurrence of hay fever was defined as an affirmative response to the question “Has your child ever diagnosed with having allergic rhinitis or hay fever?”.

Genotyping and imputation:

Genome-wide genotyping was conducted using different versions of the Illumina GSA Beadchip and the Illumina HumanOmni1-Quad Beadchip at two locations: the Human Genotyping Facility (HuGeF) in the Department of Internal Medicine at Erasmus MC, The Netherlands, and the Spanish National Genotyping Center (CEGEN-Madrid), respectively. Genotype calling was conducted using the GeneTrain2.0 algorithm, based on HapMap clusters integrated into the GenomeStudio software.

Quality control was conducted using the PLINK 1.9 program. Variants with a call rate below 95%, a minor allele frequency below 1% and a Hardy-Weinberg exact test P-value below $1e-06$ were excluded in both GSA and OMNI datasets. Samples displaying average heterozygosity values deviating beyond 4 standard deviations from the mean, those with discordant sex, PI-HAT estimate higher than 0.18, or a missing genotype rate exceeding 3% and 2% in GSA and OMNI datasets, respectively, were removed.

Genotype imputation was performed using the Michigan Imputation Server with the Haplotype Reference Consortium reference panel, Version r1.1 2016. Haplotype phasing was carried out with Eagle v2.4 and imputation was performed with Minimac4.

Statistical analyses:

GWAS analyses were conducted employing SNPTEST software using a logistic regression model assuming an additive effect of the allelic variants and evaluating the imputed dosages generated from imputation. All GWAS analyses incorporated the top ten principal components and sex as covariates. Specifically, we performed three different GWAS:

-GWAS using the GSA dataset with the basic adjustment for sex and 10 principal components.

-GWAS using the OMNI dataset with the basic adjustment for sex and 10 principal components.

GWAS using the OMNI dataset with an adjustment for sex, 10 principal components and hay fever.

Acknowledgements and Funding:

The INMA researchers would like to acknowledge to all participants for their generous collaboration. A full roster of the INMA Project Investigators can be found at <https://www.proyectoinma.org/>.

The INMA-GIP cohort was funded by the Spanish Biomedical Research Center in Epidemiology and Public Health (CIBERESP) and by grants from the Instituto de Salud Carlos III (PI06/0867, PI09/00090, PI13/02187, PI18/01142, PI21/01491) cofunded by ERDF, “A way to make Europe”, the Basque Department of Health (Projects GV-SAN2005111093, GV-SAN2009111069, GV-SAN2013111089, GV-SAN2015111065 and GV-SAN2018111086, GV-SAN2019111085), the Provincial Government of Gipuzkoa (Projects DFG06/002, DFG08/001 and DFG15/221) and annual agreements with the municipalities of the study area (Zumarraga, Urretxu, Legazpi, Azkoitia, Azpeitia and Beasain). The INMA Valencia (VAL) and Sabadell (SAB) cohorts were funded by grants from the Instituto de Salud Carlos III (CB06/02/0041, G03/176, FIS PI041436, PI081151, PI041705, and PS09/00432, FIS-FEDER 03/1615, 04/1509, 04/1112, 04/1931, 05/1079, 05/1052, 06/1213, 07/0314, 09/02647 and 23/1578), Fundació La Marató de TV3 (090430), Generalitat de Catalunya-CIRIT (1999SGR 00241), Conselleria de Sanitat Generalitat Valenciana, and Fundación Roger Torné. ISGlobal is supported by the CERCA Programme, Generalitat de Catalunya.

9. Isle of Wight Birth Cohort (IOWBC)

Recruitment:

Isle of Wight Birth Cohort (IOWBC) was described in Arshad et al 2018 (47). Briefly, the IOWBC is an unselected birth cohort study established in 1989 on the IOW, UK. After the exclusion of adoptions, perinatal deaths, and refusal for follow-up, written informed consent was obtained from parents to enrol 1456 new-borns born between 1 January 1989 and 28 February 1990. Follow-up assessments were conducted to 26 years of age to prospectively study the development of asthma and allergic diseases. At each follow-up, validated questionnaires were completed by the parent or subject. Additionally, skin prick tests to common allergens were performed on 980, 1036, and 853 participants at 4, 10, and 18 years of age to check allergic reactions. At 10, 18 and 26, spirometry and methacholine challenge tests were performed.

Case/control definition:

Food allergen sensitization was defined as +ve skin prick response (≥ 3 mm above control) for the following allergens: cow's milk, hen's egg, cod, wheat, soya and peanut (food panel), additional allergens were used depending on clinical history (extracts from ALK-Abello) as described in as described in Fong et al. (48)

Clinical diagnosis of food allergy was defined using physician review of patients reported symptoms following reaction to food, particularly time of onset as previously described (48, 49). Briefly, for FA, three criteria were defined a priori, and all three needed to be met for the diagnosis of food allergy. These included (1) a reaction to a recognized food allergen as defined by the European Union and the Committee on Toxicity of Chemicals in Foods, Consumer Products and the Environment (2) the report of recognized allergic symptoms and (3) symptoms developing within 4 h of food ingestion. Food challenges were not performed in our study.

Allergic rhinitis was defined based on information provided by participants on respiratory, nasal, and dermatologic symptoms. Study-specific plus International Study of Asthma and Allergies in Childhood (50) questionnaires were used. Rhinitis at 10 and 18 years of age was defined by the question "Have you ever had a problem with sneezing, runny or blocked nose in the absence of cold or flu?" plus the presence of "symptoms in the last 12 months."

Genotyping and imputation:

SNPs were genotyped using the illumina Infinium Omni2.5–8 v1.3 BeadChip genotyping platform (Illumina Inc, San Diego, CA, USA). Genotypes were imputed using Sanger imputation server using EAGLE2 (20) and Haplotype Reference Consortium reference panel with genome build GRCh37 (the quality control information was showed in Granell et al.

2023 (51)). We used SNPTEST v2.5.2 (42) with frequentist additive model with expected method to investigate the association between SNPs and food allergy phenotypes with sex as covariate and Hayfever/Rhinitis as sensitivity.

Funding:

This work was funded by the National Institutes of Health [grant numbers R01 AI061471, R01 HL082925, R01 AI091905, R01 AI121226 and R01 HL132321], the Medical Research Council UK MR/S025340/1, Asthma UK for 10-year assessment (grant no. 364. PI: S.H. Arshad) and the David Hide Asthma and Allergy Research Trust. The funders did not contribute to study design, data collection, analysis, interpretation of data or in the preparation of the manuscript.

Acknowledgments:

The authors are thankful to all the IOWBC participants and greatly appreciate the support of the IRIDIS High-Performance Computing facility at the University of Southampton. The authors are thankful to the nurses and staff at the David Hide Asthma & Allergy Research Centre, Isle of Wight, UK, for their help in recruitment and sample collections, and are thankful to all the cohort participants.

Ethical approval:

Ethics approval for the IoW cohort was originally given by the Isle of Wight local research ethics committee in 1989 and at each subsequent follow up (1, 2 and 4 years). At the age 10 follow up (including DNA and genotyping): Isle of Wight Health Authority Local Research Ethics Committee 18/98. Age 18 Follow up (including DNA and genotyping): Isle of Wight, Portsmouth & Southeast Hampshire Research Ethics Committee 06/Q1701/34.

10. Cooperative Health Research in the Region of Augsburg (KORA)

Recruitment:

The KORA study is a series of independent population-based epidemiological surveys of participants living in the region of Augsburg, Southern Germany (52). All survey participants are residents of German nationality identified through the registration office and informed consent has been given by all participants. Baseline examinations were conducted in 1999-2001 (KORA S4) and follow-up examinations were carried out in 2006-2008 (KORA F4), and 2013-2014 (KORA FF4).

Case/control definition:

Self-reported food allergy was assessed via questionnaire at the initial S4 survey (1991-2001), FF4 follow-up (2013-2014) and during the GEFU5 health survey (2021). Self-reported

food allergy cases were defined as participants who ever reported a food allergy in any of the three follow-ups. Controls include all participants who never reported a food allergy and for whom information was provided in at least two of the three questionnaires mentioned above. This leads to 735 cases and 1505 controls.

Information of a doctors-diagnosed food allergy was assessed at the FF4 and GEFU5 surveys. Doctors-diagnosed cases include participants who ever reported a doctors-diagnosed food allergy. Controls are participants who never indicated a doctors-diagnosed food allergy and with available information from both questionnaires. This leads to 223 cases and 867 controls.

Food-specific allergic sensitization was assessed in a subsample of the KORA FF4 cohort (N=780) by using the ImmunoCAP ISAC 112 chip. Food-specific allergic sensitization was defined as allergic sensitization to any food allergen > 0.35 kU/L. The following food allergens were included: Act d 1, Act d 2, Act d 5, Act d 8 (kiwi), Ana o 2, Ana o 3 (cashew nut), Api g 1 (celery), Cor a 14, Cor a 9, Cor a 8, Cor a 1.0401 (hazelnut), Ara h 1, Ara h 2, Ara h 3, Ara h 6, Ara h 8, Ara h 9 (peanut), Ber e 1 (brazil nut), Bos d 4, Bos d 5, Bos d 6, Bos d 8, Bos d Lactofe (cow's milk/beef), Fag e 2 (buckwheat), Gad c 1 (cod), Gal d 1, Gal d 2, Gal d 3, Gal d 5 (hen's egg/chicken), Gly m 4, Gly m 5, Gly m 6 (soy bean), Jug r 1, Jug r 3 (walnut), Mal d 1 (apple), Pen m 1, Pen m 2, Pen m 4 (shrimps), Alpha Gal (red meat), Pru p 1, Pru p 3 (peach), Ses i 1 (sesame), Tri a 14, Tri a 19.0101, Tri a A_T1 (wheat). Controls are participants without any sensitization against any food allergen. This leads to 181 cases and 599 controls.

Cases for doctors-diagnosed food allergy plus food-specific sensitization were defined as participants who ever reported a doctors-diagnosed food allergy and have a food-specific allergic sensitization (sIgE > 0.35 kU/L). Controls are participants who never reported a doctors-diagnosed food allergy and who did not show any positive sensitization against any food allergen. This leads to 68 cases and 382 controls.

For sensitivity analyses, an ever-reported hay fever variable was created. Information on self-reported hay fever was available the initial S4 survey (1991-2001). Doctors-diagnosed hay fever was assessed at the F4 follow-up (2006-2008) and the FF4 follow-up (2013-2014). Participants were defined as hay fever cases if they at least once indicated having hay fever. Controls were defined as participants who never indicated having hay fever at any of the three time points (only one missing is allowed for controls).

Genotyping and imputation:

Genotyping of 3,775 participants was done by using Affymetrix Axiom array. Genotype calling was conducted using the Affymetrix Software algorithm with NCBI build 37.

During the quality control on subjects, participants were excluded based on the following criteria: sex discordance, call rate < 97%, non-European ancestry, population outliers, heterozygosity rates higher or lower than 5 standard deviations of the population mean. Genetic variants with call rate <98%, Hardy-Weinberg Equilibrium p-value <5x10⁻¹⁰, minor allele frequency <0.02 were discarded from further analyses.

Genetic variants were imputed using the minimac3 imputation software and the Haplotype Reference Consortium (r1.1, April 2016) panel through the Michigan Imputation Server. SHAPEIT (v2) was used for haplotype phasing.

Statistical analyses:

Genome-wide association testing was performed using SNPTTEST software version 2.5.6. Allele effects were estimated from additive model (-frequentist 1) using expected genotype counts (-method expected). Sex and age were included as covariates.

11. Lifelines

Case/control definition:

Lifelines is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviours of 167,729 persons living in the North of the Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioural, physical and psychological factors which contribute to health and disease of the general population, with a special focus on multi-morbidity and complex genetics (53). The Lifelines protocol has been approved by the UMCG Medical ethical committee under number 2007/152 and all participants provided informed consent. Participants were invited for the follow-up visits which are scheduled every 5 years. FA information was collected through a 'Food Allergy Questionnaire' during the second follow-up assessment (54). Self-reported FA were defined as cases where participants reported having allergic reactions to specific foods. Doctor-diagnosed FA were defined as cases where a participant's self-reported food allergies were confirmed by a doctor. Healthy controls consisted of participants who reported no symptoms of FA.

Genotyping and imputation

The genetic data was collected in two sub-cohorts of the Lifelines cohort using the following criteria: availability of isolated DNA-samples of adequate volume and concentration; Caucasian-ancestry samples. In 2019, the first batch of 38,030 Lifelines participants was genotyped using the Infinium Global Screening Array® (GSA) MultiEthnic Disease Version 1.0. After quality control, 36,339 samples were imputed with the Sanger imputation service (EAGLE2 v2.0.5 and PBWT) using the Haplotype Reference Consortium Release 1.1.

Details on quality control and genetic imputation can be found in the supporting data (<https://wiki-lifelines.web.rug.nl/doku.php?id=ugli>).

Statistical analyses

to assess whether there are associations between SNPs and FA we conducted GWAS using SAIGEgds (scalable generalized mixed model) (55-57) adjusting for potential confounders in the discovery set. The model was formulated as: $FA \sim SNP + age + sex + PCs_{1-10}$.

Funding:

The Lifelines initiative has been made possible by subsidy from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the University Medical Center Groningen (UMCG), Groningen University and the Provinces in the North of the Netherlands (Drenthe, Friesland, Groningen).

12. PASTURE

Study design and population:

The PASTURE study is a prospective birth cohort study started in 2002 and is conducted in children from rural areas of 5 European countries (Austria, Finland, France, Germany, and Switzerland) (58). The study was designed to evaluate risk and preventive factors for atopic diseases and was approved by the local research ethics committees in each country. The written informed consents were obtained from the children's parents. Briefly, pregnant women were invited to participate during their third trimester of pregnancy and the children from the participating women were recruited at birth. Children of mothers living on family-run livestock farms at birth of the child was assigned to the farm group. The nonfarm group included children of mothers from the same rural areas but not living on a farm. Information was obtained through questionnaires in interviews or self-administered questionnaires from mothers.

Definitions of cases/controls

Self-reported food allergy cases were assessed using questionnaires at 1, 5 or 6 years. Children were classified as cases if they answered yes at least once to any of the following questions: "Has your child reacted to any foods (milk, dairy products, chicken eggs, wheat flour, other cereal items, or nuts) with gastrointestinal complaints (nausea, vomiting, stomach or abdominal pain, diarrhea or soft stools) in the last 12 months?" or "Has your child reacted to any foods (milk, dairy products, chicken eggs, wheat flour, other cereal items, or nuts) in the last 12 months with skin changes (hives or the appearance or worsening of

neurodermatitis)?". Controls include all children who never reported a food allergy at all the three time points. Hay fever was defined by parent reported symptoms (itchy, runny, or blocked nose without a cold accompanied by red itchy eyes) and/or a physician's diagnosis of hay fever ever using questionnaires at 1, 5, or 6 years.

Genotyping and imputation:

Genotyping was done using Infinium-Global-Screening Array-v 3.0. The quality control included checks for sex discrepancy, allele frequency (MAF>1%), call-rate by snp > 0.95, call-rate by subject >0.95, and p-value for HWE > 10⁻⁶. The genotypes were phased and imputed (59, 60) with cosmopolitan reference (61), to SNPs with a minor-allele-frequency >1% in the European reference panel.

Statistical analyses:

Genome-wide association test was done using SNPTEST version 2.5.4 with the method expected and frequentist association test in the additive model. The BASIC model was adjusted for age (in months), gender, center, and farming-status. The sensitivity model was additionally adjusted for hay fever.

Acknowledgements and Funding:

We are thankful to all the participants and parents of the PASTURE Study. The PASTURE study was supported by the Federal Ministry of Education and Research (BMBF) under the ABROGATE project (grant agreement No. 01EA2106B) supporting the current analyses, the European Commission (research grants QLK4-CT-2001-00250, FOOD-CT-2006-31708 and KBBE-2007-2-2-06), the European Research Council (grant 250268), Deutsche Forschungsgemeinschaft, German Center for Lung Research (DZL), Kühne Foundation, French National Programme for Hospital Research, EVO- and VTR-funding, the Academy of Finland (grant 139021, 287675, 338679, 339666), the Yrjö Jahnsson Foundation, and EU, Academy of Finland and smaller grants from Finnish foundations supporting medical research (Juho Vainio Foundation, Päivikki and Sakari Sohlberg Foundation, Finnish Cultural Foundation).

13. PIAMA

Case/control definition:

The PIAMA study protocol has been described elsewhere (62). Mothers were recruited from the general population through 52 antenatal clinics in three different regions of the Netherlands. In short, this cohort includes 3963 children born in 1996-1997, with follow-up visits and questionnaires from birth onwards. The study protocol was approved by the

medical ethics committees of the participating institutions and all parents gave written informed consent. FA information was collected through a 'Food Allergy Questionnaire' at the age of 23 years, self-reported food allergy, Doctor-diagnosed FA and health controls were defined based on the answer of participants. Specific IgE levels for the foods was measured in serum and classified as positive if ≥ 0.35 IU/ml at the age of 12.

Genotyping and imputation

DNA was extracted from blood or buccal swabs collected at the age of 4 or 8 years. Samples were genotyped using four different platforms: Illumina Human610 quad array, Illumina HumanOmniExpress array, Illumina Human Omni Express Exome Array, Illumina Infinium Global Screening Array. SNPs were harmonized by base pair position annotation to the GRCh37/hg19 build of the human genome, genetic variant identifier, and strand for each platform. The removed individuals were 1) sex-mismatch, 2) heterozygous outliers (deviate $\pm 3SD$ from the sample heterozygosity mean), 3) duplicated and highly related (IBD score < 0.1875), 4) ethnic outliers from PCA plot. After QC procedures, 2,075 samples remained and data from the four platforms were merged together. Genetic variants with CR $\leq 95\%$, MAF $\leq 1\%$, and HWE $p < 1 \cdot 10 \times 10^{-6}$ were filtered out. Then, imputation was performed using Michigan server with a reference panel of HRC.r1.

Statistical analyses

To assess whether there are associations between SNPs and FA we conducted GWAS using SNPTTEST (v2.5.1), adjusting for potential confounders in the discovery set. The model was formulated as: FA $\sim$ SNP + sex + PCs1-10.

Funding:

The PIAMA study is supported by The Netherlands Organization for Health Research and Development; The Netherlands Organization for Scientific Research; BBMRI-NL, The Netherlands Ministry of Spatial Planning, Housing, and the Environment; and The Netherlands Ministry of Health, Welfare, and Sport.

Genotyping and imputation:

DNA was extracted from blood or buccal swabs collected at the age of 4 or 8 years. Samples were genotyped using four different platforms: Illumina Human610 quad array, Illumina HumanOmniExpress array, Illumina Human Omni Express Exome Array, Illumina Infinium Global Screening Array. SNPs were harmonized by base pair position annotation to the GRCh37/hg19 build of the human genome, genetic variant identifier, and strand for each platform. The removed individuals were 1) sex-mismatch, 2) heterozygous outliers (deviate $\pm 3SD$ from the sample heterozygosity mean), 3) duplicated and highly related (IBD score <0.1875), 4) ethnic outliers from PCA plot. After QC procedures, 2,075 samples remained and data from the four platforms were merged together. Genetic variants with CR $\leq 95\%$, MAF $\leq 1\%$, and HWE $p < 1 \cdot 10 \times 10^{-6}$ were filtered out. Then, imputation was performed using Michigan server with a reference panel of HRC.r1.

Statistical analyses:

To assess whether there are associations between SNPs and FA we conducted GWAS using SNPTEST (v2.5.1), adjusting for potential confounders in the discovery set. The model was formulated as: $FA \sim SNP + sex + PCs_{1-10}$.

14. SHIP

Recruitment:

The Study of Health In Pomerania (SHIP) is a prospective longitudinal population-based cohort study in Mecklenburg-Western Pomerania assessing the prevalence and incidence of common diseases and their risk factors (63, 64). SHIP encompasses the two independent cohorts SHIP-START and SHIP-TREND. Participants aged 20 to 79 with German citizenship and principal residency in the study area were recruited from a random sample of residents living in the three local cities, 12 towns as well as 17 randomly selected smaller towns. Individuals were randomly selected stratified by age and sex in proportion to population size of the city, town or small towns, respectively. A total of 4,308 participants were recruited between 1997 and 2001 in the SHIP-START cohort. Between 2008 and 2012 a total of 4,420 participants were recruited in the SHIP-TREND cohort. Individuals were invited to the SHIP study center for a computer-assisted personal interviews and extensive physical examinations.

Case/control definition:

Self-reported food allergy was evaluated during the baseline survey in the recruitment periods for SHIP-START and SHIP-TREND using a computer-assisted interview. Cases were defined as participants who were known to have a food allergy or had been diagnosed

with a food allergy by a physician. Controls were defined as participants who answered no to the question about known or diagnosed food allergies.

Genotyping and Imputation:

Genome-wide SNP-typing was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 (SHIP-START samples), the Illumina Infinium HumanOmni2.5 BeadChip, or the Illumina Infinium Global Screening Array (SHIP-TREND samples). Array processing was carried out in accordance with the manufacturer's standard recommendations. Genotypes were determined using the Birdseed2 clustering algorithm for SHIP-START, GenomeStudio Genotyping Module v1.0, and GenomeStudio 2.0 Genotyping Module (GenCall) for SHIP-TREND. The genotypes were imputed with the Michigan Imputation Server (Eagle v2.3 and Minimac3) using the reference panel of the Haplotype Reference Consortium (HRC r1.1 2016) to the genome build GRCh37.

Statistical analysis:

Genome-wide association testing was performed using SNPTTEST v2.5.2. Allele effects were estimated from additive model (-frequentist 1) using expected genotype counts (-method expected). Age, sex and the first ten genetic principal components were included as covariates. In a sensitivity analysis, hay fever was included as an additional covariate.

Acknowledgements and funding:

SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald Approach to Individualized Medicine (GANI_MED)' funded by the Federal Ministry of Education and Research (grant 03IS2061A). Genome-wide data were supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg- West Pomerania.

Ethics:

The study protocol was approved by the medical ethics committee of the University of Greifswald. Oral and written informed consent was obtained from each of the study participants.

15. The Raine Study

Recruitment:

The Raine Study is a prospective pregnancy cohort where 2,900 mothers (Gen1) were recruited between 1989 and 1991 (65, 66). Recruitment took place at Western Australia's major perinatal centre, King Edward Memorial Hospital, and nearby private practices.

Women who had sufficient English language skills, an expectation to deliver at King Edward Memorial Hospital, and an intention to reside in Western Australia to allow for future follow-up of their child (Gen2) were eligible for the study.

The Raine Study is known as one of the largest successful prospective cohorts richly phenotyped at multiple time points over pregnancy, infancy, childhood, adolescence, and young adulthood. The mothers (Gen1) completed questionnaires regarding their children (Gen2), and the children (Gen2) had physical examinations at ages 1, 2, 3, 6, 8, 10, 14, 17, 20, 22, 27, and 28 years.

Case/control definition:

Food allergens in the Raine Study included cow's milk, soy products, eggs, peanuts, other foods, and food additives/preservatives/colouring.

Self-reported cases included study participants who were informed by a child health nurse/school nurse, a dietitian, a naturopath, a pharmacist, a relative, friend, teacher or self-diagnosed.

Self-reported controls included study participants who did not report food allergies by child health nurse/school nurse, a dietitian, a naturopath, a pharmacist, a relative, friend, teacher or self-diagnosed, at any of the follow-ups attended.

Doctors-diagnosed cases included study participants who were informed by a paediatrician/specialist or a doctor.

Doctors-diagnosed controls included study participants who did not report a paediatrician/specialist or a doctor diagnosed food allergy.

Childhood outcomes included the Raine Study Gen2-3,-5, -8, -14 and -17 year follow-ups, whilst adult outcomes included from the Raine Study Gen2-20 and -22 year follow-ups. If a participant did not attend any of the follow-ups, their outcomes was coded to missing.

Age for cases is age at onset, age for controls is age at most recent follow-up attended by the study participant.

Genotyping and imputation:

The Raine Study Gen2 participants were genotyped on an Illumina 660 W Quad Array at the Centre for Applied Genomics, Toronto, Canada. Quality control (QC) of the Genome-Wide-Association-Study (GWAS) genotyped data were performed as per standard protocol. In brief, a total of 1593 Raine Study Gen2 participants were genotyped on an Illumina 660 Quad Array, which included 657,366 genetic variants, consisting of ~ 560,000 single-nucleotide-polymorphisms (SNPs) and ~ 95,000 copy number variants (CNVs), at the Centre for Applied Genomics, Toronto, Canada. Plate controls and replicates with a higher proportion of missing data were excluded before individuals were assessed for low genotyping success (> 3% missing), excessive heterozygosity, and gender discrepancies between the core data and genotyped data. Cryptic relatedness ($\pi > 0.1875$, in between second- and third-degree relatives—e.g. between half-siblings and cousins) were identified and excluded prior to statistical analyses. At the SNP level, the SNP data were cleaned using plink (14) following the Wellcome Trust Case—Control Consortium protocol (41). The exclusion criteria for SNPs included: Hardy–Weinberg-Equilibrium $p < 1E-06$; minor-allele frequency (MAF) < 1%; call-rate < 95% for SNPs with MAF >5%; call-rate < 99% for SNPs with MAF <5%; SNPs with possible strand ambiguity (i.e. A/T and C/G SNPs); SNPs with no match to the HRC reference panel; SNPs with allele frequency deviation > 0.2 when compared against the HRC CEU population; palindromic SNPs with frequency > 0.4; SNPs with non-matching allele; and duplicated SNPs. The cleaned GWAS data were imputed using the Michigan Imputation Server against the Haplotype Reference Consortium reference panel (ref). A total of 1559 individuals with 518,288 autosome SNPs and 11781 chrX SNPs remained after genotype QC; imputation resulted in 39,117,083 and 1,228,034 SNPs across the 22-autosomes and X-chromosome, respectively. A total of 1494 individuals remain after applying cryptic relatedness identified during the genotype QC protocol. Principal components (PCs) analysis was carried out, using SMARTPCA from v.3.0 of EIGENSOFT (67), on the cleaned genotyped data, where PCs were generated for purposes of adjusting for population stratification in all genetic analyses.

Statistical analyses:

Statistical analyses were performed using PLINK v2.00a3.7LM AVX2 Intel (24 Oct 2022) (15) using the `–glm` and `–covar-variance-standardize` flags, and adjusting for age, sex, and the first two principal components to account for population stratification.

Acknowledgements and Funding:

The Raine Study was supported by the National Health and Medical Research Council of Australia [grant numbers 572613, 403981 and 1059711, 211912, 003209, 353514] and the

Canadian Institutes of Health Research [grant number MOP-82893]. The authors are grateful to the Raine Study participants and their families for their ongoing participation in the study. The authors also acknowledge the research staff from the Raine Study, the Lions Eye Institute (Raine Study Gen2-20-year follow-up), and The University of Western Australia Centre for Sleep Science for utilisation of the facility (Raine Study Gen2-22-year follow-up) for study co-ordination and data collection. The authors gratefully acknowledge the NHMRC and the Raine Medical Research Foundation for their support. The core management of the Raine Study is funded by The University of Western Australia (UWA), Curtin University, The Kids Institute Australia, Women and Infants Research Foundation, Edith Cowan University, Murdoch University, and The University of Notre Dame (Australia). The authors gratefully acknowledge the assistance of the Western Australian DNA Bank (National Health and Medical Research Council of Australia National Enabling Facility). This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and Government of Western Australia.

Ethical approval:

Ethics approval for the original pregnancy cohort and subsequent follow-ups were granted by the Human Research Ethics Committee of King Edward Memorial Hospital, Princess Margaret Hospital, the University of Western Australia, and the Health Department of Western Australia.

16. The Cardiovascular Risk in Young Finns Study (YFS)

Recruitment:

The Cardiovascular Risk in Young Finns Study (YFS) is a prospective multicenter study from Finland initiated in the late 70s. The first large baseline examination was conducted in 1980 (baseline age, 3–18 years) (68). Children aged 3, 6, 9, 12, 15, and 18 years were haphazardly chosen from the population register from the 5 Finnish university cities and their surrounding rural communities. A total of 4320 subjects were invited, and 3596 of them participated in the baseline measurement. The follow-up visits or surveys have been conducted in 1983, 1986, 1989, 1992, 2001, 2007, 2011 and 2018-2020. The study design has been approved by the ethical committees of all Finnish Universities with a medical faculty. All the participants or their parents (participants aged <18 years) provided informed consent before participation. The YFS has been carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Case/control definition:

Case childhood definition: Participants who reported special diet due to food allergy when <18 years old at assessment (YFS follow-up visits 1980, 1983, 1986 and 1992). Case

adulthood definition: Participants who self-reported food allergy-constrained diet when ≥ 18 years old at assessment (YFS follow-up visit 2018-2020). Control definition: Participants who never self-reported a food allergy-constrained diet.

Genotyping and imputation:

Genotyping was done for 2556 samples using a custom-build Illumina Human 670K BeadChip at Wellcome Trust Sanger Institute. Sample call rate < 0.95 , excess heterozygosity, sex mismatch, cryptic relatedness ($\pi\text{-hat} > 0.2$), SNP call rate < 0.95 , MAF < 0.01 , and HWE $p\text{-value} < 1e-6$ were used as quality control filters. After the quality control, there were 2442 samples and 546,677 genotyped SNPs available for further analysis (69). Genotype Imputation to Haplotype Reference Consortium (HRC) reference was performed using Eagle2 for haplotype phasing and Minimac3 and HRC version r1.1 haplotypes for genotype imputation.

Statistical analyses:

Genome-wide association testing was performed using SNPTEST software version 2.5.2. Allele effects were estimated from additive model (-frequentist 1) using expected genotype counts (-method expected). Models for childhood and adulthood food allergy were adjusted by sex, age and first ten genetic principal components. Additionally for adulthood food allergy, a model adjusted by self-reported allergic rhinitis, sex, age and first ten genetic principal components was estimated.

Acknowledgements and Funding:

The Young Finns Study has been financially supported by the Academy of Finland: grants 356405, 322098, 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117797 (Gendi), and 141071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; EU Horizon 2020 (grant 755320 for TAXINOMISIS and grant 848146 for To Aition); European Research Council (grant 742927 for MULTIEPIGEN project); Tampere University Hospital Supporting Foundation; Finnish Society of Clinical Chemistry; the Cancer Foundation Finland; pBETTER4U_EU (Preventing obesity through Biologically and bEhaviorally Tailored inTERventions for you; project number: 101080117); and the Jane and Aatos Erkko Foundation.

Replication studies:

17. CLARA/CLAUS

Recruitment:

Cross-sectional study registered at the DRKS German study registry (no. DRKS00004635; http://drks-neu.uniklinikfreiburg.de/drks_web/). Parents completed a detailed questionnaire assessing health data on allergy, asthma, and socioeconomic factors. The study population contained healthy control subjects (HCs) and patients with mild-to-moderate AA and NA (70, 71). Informed oral and written consent was obtained from the parents for participation and blood collection. Approval was obtained from the local ethics board, LMU Munich, Germany (Nr. 37908; 4th of August 2008). This study does conform to the Declaration of Helsinki.

Case/control definition:

Children with available sensitization data to 7 different food allergens, available genotype data and European ancestry were included (N = 492). Cases were defined as children with a positive sensitization value to at least one food allergen. Controls were defined as children with sIgE values lower than 0.35 IU/ml to all food allergens. A positive specific IgE to an allergen was defined as a positive reaction ≥ 0.35 IU/mL to a specific food allergen. This food allergens were measured within a panel of 20 common allergens by specific IgE testing (Mediwiss Analytic, Moers, Germany; house dust mite: dermatophagoides pteronyssinus, dermatophagoides farinae; tree pollen: birch, hazel; grass pollen: grass-mix, timothy, mugwort, plantain; animals: cat epithelium, horse epithelium, dog epithelium; mould: alternaria alternata; food allergens: egg white, cow's milk, peanut, soybean, hazelnut, carrot, wheat; other: latex)

Genotyping and imputation:

DNA was extracted from EDTA-blood by using the QIAamp DNA blood kit (Qiagen, Hilden, Germany), concentrations and quality was assessed by NanoDrop. SNPs were genotyped using Infinium-Global-ScreeningArray-v 2.0/3.0 and quality was ensured by plausibility checks for sex and allele-frequencies; subject-and SNPwise call-rates $>95\%$, P-value of Chi-square test for deviation from HWE $> 10^{-6}$. Genotypes were phased and imputed (59, 60) with cosmopolitan reference (61), to SNPs with a minor-allele-frequency $>1\%$ in the European reference panel.

Statistical analyses:

The association between the genotypes and the phenotype 3 was assessed using multiple logistic regression. An additive model was assumed for the allele effects. The first model was adjusted for sex, age and the first five PC's from the genotype data, a second model was

fitted including the diagnosis of hay fever. All statistical analyses were conducted using the R programming language (base::glm).

Funding and acknowledgments:

We thank all CLARA/CLAUS participants and their parents, the nurses, the recruitment team and study physicians, in particularly Kathrin Urner, Franziska Sattler, Kathrin Kalkbrenner.

Supported by the German Research Foundation (DFG SCHA 997/9-1; SCHA 997/10-1, SCH 997/11-1, SFB-TR22, Federal Ministry of Education and Research (BMBF-CHAMP) 01GL1742A, the German Centre for Lung Research (DZL) as part of the Comprehensive Pneumology Centre in Munich (CPC-M). DZL german lung center, CPC-M, Munich, 82DZL033C2, 01GL2406A and the German Center for Child and Adolescent Health (DZKJ; LMU/LMU Klinikum: 01GL2406A.

18. EXCEED

Recruitment:

EXCEED is a longitudinal population-based cohort which facilitates investigation of genetic, environmental and lifestyle-related determinants of a broad range of diseases and of multiple morbidity, through data collected at baseline and via electronic health care record linkage. Recruitment of participants aged 30-69 has taken place since 2013, primarily from the general population through general practices in Leicester City, Leicestershire and Rutland. Participants have consented to follow-up for up to 25 years through electronic health records.

Case/control definition:

Primary care electronic health records were used to define cases and controls in EXCEED.

Cases were defined as participants with a diagnosis of food-related allergy after age 18 years based on the presence of at least one of the following CTV3 codes: Xa1aX (Food allergy), XabWn (Allergy to banana), Xaa0d (Allergy to soya), Xa7II (Allergy to strawberries), Xad15 (Allergy to tomato), Xa7IP (Cheese allergy), Xa7IN (Chocolate allergy), Xa1nn (Cow's milk allergy), Xa1nm (Egg allergy), SN581 (Egg protein allergy), XaOYT (Wheat allergy), Y06ac (Wheat allergy), Y2199 (Lentil Allergy), Xa7IJ (Nut allergy), Xa1no (Peanut allergy), Xa1np (Seafood allergy), Xa7IO (Shellfish allergy), J432. (Cow's milk allergy), XE0ot (Milk-free diet - allergy), XE0ou (Egg-free diet - allergy), XaQaY (Fish allergy). Controls were defined as participants without any food-related allergy diagnosis.

Hay fever was defined as diagnosis in primary care records, as per the following CTV3 codes: Xa0IX (seasonal allergic rhinitis), XE2QI (allergic rhinitis due to pollen), H170 (Pollinosis (& allergic rhinitis due to pollens), Hyu20 (other seasonal allergic rhinitis).

Genotyping and imputation:

Genotyping of EXCEED participants was conducted using the UK Biobank Axiom array. Sample-level and variant-level quality control was performed on genotype data prior to imputation to the TopMed reference panel. SNPs with a missing rate >5%, HWE p -value $<1e^{-15}$ and MAF <1% were excluded. We excluded duplicate samples, samples with a heterozygosity rate higher or lower than 6 standard deviations of the population mean, and samples with >5% missingness and >20% sex mismatch. Additional pre-imputation checks were performed using the Will Rayner toolbox. Genotypes were imputed to the TOPMed-r3 reference panel using the TOPMed imputation server.

Statistical analyses:

Statistical analyses were performed using REGENIE v3.3. Association testing was conducted using imputed dosages, assuming an additive genetic model and adjusted for age, sex, 10 principal components of genetic ancestry and genotyping batch. A further analysis was performed including hay fever as an additional covariate.

Acknowledgements and Funding:

EXCEED is supported by the University of Leicester, the NIHR Leicester Respiratory Biomedical Research Centre, by Wellcome [202849, <https://doi.org/10.35802/202849>], by BREATHE - The Health Data Research Hub for Respiratory Health [UKR_PC_19004] in partnership with SAIL Databank, and by Cohort Access fees from studies funded by the Medical Research Council (MRC), BBRSC, NIHR, the UK Space Agency, and GSK. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. It was previously supported by MRC grant G0902313. Our analyses of EXCEED used the ALICE High Performance Computing Facility at the University of Leicester. The EXCEED study gratefully acknowledges the support of all participants and staff who have contributed to the study.

19. GENEVA

Case/control definition:

The GENEVA cohort consisted of 613 trios (parents and child) where the child had a double-blind, placebo-controlled food challenge (DBPCFC) as part of routine tertiary pediatric allergy care due to a history indicating an IgE-mediated reaction following food ingestion. A subset of the GENEVA cohort had been previously described (72-74). Recruitment started at the University Medical Center Groningen in 2005. The study received ethical approval (METc 2004-146), and written informed consent was obtained from the parents.

All DBPCFCs, regardless of positive or negative outcomes, were included. The DBPCFCs were conducted as detailed in a prior study (75), and specific IgE levels for the foods tested in the DBPCFC were measured using CAP-FEIA (ImmunoDiagnostics, Uppsala, Sweden).

Genotyping and imputation:

Genotyping of the GENEVA cohort was carried out at the Human Genotyping Facility of the Genetic laboratory of the department of Internal Medicine at Erasmus MC using the Illumina GSA beadchip. Raw data was processed and analyzed together using plink (v1.07).

Individuals with a call rate < 95% or discordant sex information were excluded from analysis. SNPs with a MAF < 1%, call rate < 90% and/or with Hardy–Weinberg equilibrium (HWE) test P-value < 10E-6 were excluded, leaving 489,864 high-quality directly genotyped variants in 608 children of which 428 with parents. The genotyped variants (489,864 in 1,464 samples) were used to impute unmeasured variants, using the Michigan imputation server with the 1 HRC.r1 as reference panel and pre-phasing using Michigan Imputation Server 2.

Statistical analysis:

To valid whether there are associations between candidate SNPs and FA defined by OFC, we conducted a family-based study, calculating transmission disequilibrium of candidate SNPs using plink v1.90.

Funding:

G.H.K. received an unrestricted grant from the Nutricia Research Foundation to complete the Dutch GENEVA cohort

20. German Genetics of Food Allergy Study (GOFA)

Case/control definition:

Children of the GOFA study were recruited at clinical centers in Berlin, Wangen, Hannover, and Oldenburg, Germany. In most of the cases food allergy was diagnosed by DBPCFC. Additional cases (14%) had a convincing history of an immediate, severe allergic reaction plus specific sensitization to the same food (IgE > 0.35 kU/L). They were included without oral food challenge as it was contraindicated due to the risk of severe anaphylactic reactions. Control individuals without information on FA were derived from two population-based cohorts from Germany, the Heinz Nixdorf Recall Study and the Study of Health in Pomerania. The study population combines set 1 and set 2 of our previous GWAS on food allergy (76) which describes GOFA in more detail. The study was approved by the local ethics committees, and written informed consent was given by all participants or their legal guardians.

Genotyping and imputation:

Samples were genotyped on Illumina's HumanOmniExpressExome-8 v1.2, HumanOmniExpress-12 v1.0, or HumanOmni1-Quad v1 arrays. For imputation only those SNPs were used which were genotyped on all arrays. Quality control criteria were the same as described previously (76). After QC, the study population included 847 cases with FA and 3358 controls.

Statistical analyses:

Association analyses between the candidate variants and disease were conducted with mach2dat (v1.0.24) using allele dosages and assuming an additive model. Sex and PCs 1-3 were included as covariates.

21. HealthNuts

Recruitment:

HealthNuts recruited n=5276 12-month-old infants from council-run immunization sessions in Melbourne Victoria, Australia between 2007-2011. Infants underwent SPT to 4 common food allergens (egg, peanut, sesame and either cow's milk or shrimp) and OFC if SPT was 1mm or greater (egg, peanut and sesame only). Blood samples were collected at the OFC visit.

Case/control definition:

Food allergy was defined as a positive OFC in conjunction with a positive test of sensitization (SPT ≥ 2 mm and/or sIgE ≥ 0.35 ku/l). Infants who had a reaction that met OFC stopping criteria in the previous month for egg, or 2 months for peanut and sesame were deemed allergic without challenge. Food tolerance was defined as a negative OFC or SPT 0mm.

Genotyping and imputation:

For genome-wide genotyping, DNA samples were thawed and randomly assigned into 8 plates and submitted to the Australian Genome Research Facility for genotyping on the Illumina HumanOmni 2.5-8 SNP array. Plates 1 and 2 were submitted as a single batch, plates 3-8 as a second batch. Individual SNP genotype calls with a GenCall score < 0.7 were set to missing. SNPs were excluded if MAF $< 1\%$, call rate $< 95\%$ and/or with Hardy Weinberg Equilibrium (HWE) test P-value $< 10^{-6}$, estimated on a per-plate and per-batch basis, leaving 402,866 SNPs common to all plates. SNPs with significant ($P < 0.001$) MAF differences between plates were excluded ($n = 15,069$). SNPs with MAF $< 1\%$, call rate $< 95\%$ and/or with HWE test P-value $< 10^{-6}$ in the combined dataset were excluded, leaving 400,993 SNPs for analysis. SNPs were excluded if not present in the 1000 Genomes project, did not have a one-to-one rs# match, had different alleles, and/or significant ($P < 0.001$) MAF differences between the HealthNuts samples and the 1000 Genomes Project samples (Europeans only), leaving 389,427 variants. Samples were excluded if genotyping call rates were below 95%

(based on a subset of SNPs with a minor allele frequency [MAF] >0.2), and checked for consistency between self-reported and genotype-inferred sex. Individuals with high genome-wide identity-by-descent (IBD, >0.2) with other individuals were excluded. In total 272 individuals were removed during QC. In the final data set 221 post-qc individuals met our case definition for peanut allergy or our control criteria. Ancestry was inferred from visual inspection of results of a multi-dimensional scaling (MDS) analysis of identity-by-state (IBS) distance between all individuals and samples of known ancestry from the 1000 Genomes Project 26, using PLINK (v1.07).

Genotype data for 389 427 directly genotyped variants from 364 individuals were used to impute unmeasured variants, using Impute2 28 with default options and the 1000 Genomes Project March 2012 release of reference haplotypes available through the Impute2 website (files ALL_1000G_phase1integrated_v3_chr*_impute.*). Variants in the X chromosome were imputed using the same approach.

Statistical analyses:

The association between case–control status and allelic dosage was tested using SNPTTEST with sex and ancestry strata included as covariates. A P-value of 5×10^{-8} was considered the threshold for genomewide significance. The options used for the association test of autosomes were –frequentist 1 –method expected. For the X chromosome, options were –method newml and –assume_chromosome X. After excluding variants with imputation info < 0.95, HWE test P-value < 10^{-6} , outlier beta estimates ($|\text{beta}| > 3$ and/or $\text{SE} > 1.5$), or MAF < 1%, results were available for 3 814 967 variants.

Acknowledgements and Funding:

Genotyping, QC, imputation and initial HealthNuts GWAS analysis was carried out by M. A. R. Ferreira.. The genomewide association in this study was funded by the US Department of Defense (grant # W81XWH-10-1-0487). The HealthNuts cohort is funded by the Australian National Health and Medical Research Council. This work was supported by the Australian National Health and Medical Research Council (NHMRC). Research at the Murdoch Childrens Research Institute was supported by the Victorian Government's Operational Infrastructure Support Program. The HealthNuts study investigators include Katie Allen, Jennifer Koplin, Mimi Tang, Shyamali Dharmage, Lyle Gurrin, Jeffrey Craig, Justine Ellis, Manuel Ferreira, Adrian Lowe, Melanie Matheson, Nicholas Osborne, Richard Saffery.

22. Manchester Asthma and Allergy Study (MAAS)

Recruitment:

Recruitment of the MAAS cohort is previously described in detail by Granell et al. (51).

Case/control definition:

OFC-proven food allergy: Cases = Positive OFC + Considered peanut allergic by study criteria age 8, not challenged + Peanut allergy confirmed at age 11.

Controls = not allergic to peanut

Hayfever: Case: Answered yes to the question “Has a doctor ever told you your child had hayfever?” at any timepoint at age 5, 8, 11. Control: Did not answer yes to the question “Has a doctor ever told you your child had hayfever?” at any timepoint at age 5, 8, 11.

Genotyping and imputation:

Genotyping and imputation of the MAAS cohort is previously described in detail by Granell et al. (51).

Statistical analyses:

Genome-wide association testing was performed using SNPTTEST version 2.5.6. Allele effects were estimated using the additive model (-frequentist add) with expected genotype counts (-method expected) on the genotype call probability (-genotype_field GP). Models for food allergy were adjusted by sex, and the first three genetic principal components. Additionally, a model adjusted by self-reported allergic rhinitis, sex, and the first three genetic principal components was estimated.

Acknowledgements and Funding:

The authors would like to thank the MAAS study participants and their parents for their continued support and enthusiasm. MAAS was supported by the Asthma UK Grants No 301 (1995-1998), No 362 (1998-2001), No 01/012 (2001-2004), No 04/014 (2004-2007), BMA James Trust (2005) and The JP Moulton Charitable Foundation (2004-current), The North west Lung Centre Charity (1997-current) and the Medical Research Council (MRC) G0601361 (2007-2012), MR/K002449/1 (2013-2014) and MR/L012693/1 (2014-2018), and MR/S025340/1 UNICORN (Unified Cohorts Research Network): Disaggregating asthma (2020-2024) This study has been delivered through the National Institute for Health and Care Research (NIHR) Manchester Biomedical Research Centre (BRC) (NIHR203308). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care

23. UK Biobank

Case/control definition

GWAS summary statistics provided by Neale Lab version 3 were used for the replication.

Self-reported food allergy cases and controls were defined according to self-reported allergy or anaphylactic reaction to food.

More details are provided on

<https://docs.google.com/spreadsheets/d/1kvPoupSzsSFBNSztMzl04xMoSC3Kcx3CrjVf4yBmESU/edit?gid=227859291#gid=227859291> and

https://github.com/Nealelab/UK_Biobank_GWAS?tab=readme-ov-file#imputed-v3-phenotypes.

References:

1. Melén E, Bergström A, Kull I, Almqvist C, Andersson N, Asarnej A, et al. Male sex is strongly associated with IgE-sensitization to airborne but not food allergens: results up to age 24 years from the BAMSE birth cohort. *Clin Transl Allergy*. 2020;10:15.
2. Ödling M, Andersson N, Hallberg J, Almqvist C, Janson C, Bergström A, et al. A Gap Between Asthma Guidelines and Management for Adolescents and Young Adults. *J Allergy Clin Immunol Pract*. 2020;8(9):3056-65.e2.
3. Mitselou N, Andersson N, Bergström A, Kull I, Georgelis A, van Hage M, et al. Preterm birth reduces the risk of IgE sensitization up to early adulthood: A population-based birth cohort study. *Allergy*. 2022;77(5):1570-82.
4. Wang G, Hallberg J, Faner R, Koefoed HJ, Kebede Merid S, Klevebro S, et al. Plasticity of Individual Lung Function States from Childhood to Adulthood. *Am J Respir Crit Care Med*. 2023;207(4):406-15.
5. Ödling M, Wang G, Andersson N, Hallberg J, Janson C, Bergström A, et al. Characterization of Asthma Trajectories from Infancy to Young Adulthood. *J Allergy Clin Immunol Pract*. 2021;9(6):2368-76.e3.
6. Mogensen I, Hallberg J, Björkander S, Du L, Zuo F, Hammarström L, et al. Lung function before and after COVID-19 in young adults: A population-based study. *J Allergy Clin Immunol Glob*. 2022;1(2):37-42.
7. Kere M, Klevebro S, Hernandez-Pacheco N, Ödling M, Ekström S, Mogensen I, et al. Exploring proteomic plasma biomarkers in eosinophilic and neutrophilic asthma. *Clin Exp Allergy*. 2023;53(2):186-97.
8. Björkander S, Klevebro S, Hernandez-Pacheco N, Kere M, Ekström S, Sparreman Mikus M, et al. Obese asthma phenotypes display distinct plasma biomarker profiles. *Clin Transl Allergy*. 2023;13(3):e12238.
9. Hallberg J, Ballardini N, Almqvist C, Westman M, van Hage M, Lilja G, et al. Impact of IgE sensitization and rhinitis on inflammatory biomarkers and lung function in adolescents with and without asthma. *Pediatr Allergy Immunol*. 2019;30(1):74-80.
10. Protudjer JL, Vetander M, Kull I, Hedlin G, van Hage M, Wickman M, et al. Food-Related Symptoms and Food Allergy in Swedish Children from Early Life to Adolescence. *PLoS One*. 2016;11(11):e0166347.
11. Gref A, Merid SK, Gruzieva O, Ballereau S, Becker A, Bellander T, et al. Genome-Wide Interaction Analysis of Air Pollution Exposure and Childhood Asthma with Functional Follow-up. *Am J Respir Crit Care Med*. 2017;195(10):1373-83.
12. Hernandez-Pacheco N, Kilanowski A, Kumar A, Curtin JA, Olvera N, Kress S, et al. Exploring the genetics of airflow limitation in lung function across the lifespan – a polygenic risk score study. *eClinicalMedicine*. 2024;75.
13. Kilanowski A, Thiering E, Wang G, Kumar A, Kress S, Flexeder C, et al. Allergic disease trajectories up to adolescence: Characteristics, early-life, and genetic determinants. *Allergy*. 2023;78(3):836-50.
14. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-75.
15. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
16. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65.
17. Lam M, Awasthi S, Watson HJ, Goldstein J, Panagiotaropoulou G, Trubetskoy V, et al. RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics*. 2020;36(3):930-3.
18. Durbin R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics*. 2014;30(9):1266-72.

19. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016;48(10):1279-83.
20. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Y AR, H KF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet.* 2016;48(11):1443-8.
21. Bisgaard H. The Copenhagen Prospective Study on Asthma in Childhood (COPSAC): design, rationale, and baseline data from a longitudinal birth cohort study. *Ann Allergy Asthma Immunol.* 2004;93(4):381-9.
22. Bisgaard H, Vissing NH, Carson CG, Bischoff AL, Følsgaard NV, Kreiner-Møller E, et al. Deep phenotyping of the unselected COPSAC2010 birth cohort study. *Clin Exp Allergy.* 2013;43(12):1384-94.
23. Leitsalu L, Haller T, Esko T, Tammesoo ML, Alavere H, Snieder H, et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int J Epidemiol.* 2015;44(4):1137-47.
24. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. *Am J Hum Genet.* 2021;108(10):1880-90.
25. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics.* 2010;26(22):2867-73.
26. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet.* 2021;53(7):1097-103.
27. Kauffmann F, Dizier MH, Pin I, Paty E, Gormand F, Vervloet D, et al. Epidemiological study of the genetics and environment of asthma, bronchial hyperresponsiveness, and atopy: phenotype issues. *Am J Respir Crit Care Med.* 1997;156(4 Pt 2):S123-9.
28. Bouzigon E, Nadif R, Le Moual N, Dizier MH, Aschard H, Boudier A, et al. [Genetic and environmental factors of asthma and allergy: Results of the EGEA study]. *Rev Mal Respir.* 2015;32(8):822-40.
29. Gheerbrant H, Guillien A, Vernet R, Lupinek C, Pison C, Pin I, et al. Associations between specific IgE sensitization to 26 respiratory allergen molecules and HLA class II alleles in the EGEA cohort. *Allergy.* 2021;76(8):2575-86.
30. Siroux V, Ballardini N, Soler M, Lupinek C, Boudier A, Pin I, et al. The asthma-rhinitis multimorbidity is associated with IgE polysensitization in adolescents and adults. *Allergy.* 2018;73(7):1447-58.
31. Lupinek C, Wollmann E, Baar A, Banerjee S, Breiteneder H, Broecker BM, et al. Advances in allergen-microarray technology for diagnosis and monitoring of allergy: the MeDALL allergen-chip. *Methods.* 2014;66(1):106-19.
32. Vernet R, Matran R, Zerimech F, Madore AM, Lavoie ME, Gagnon PA, et al. Identification of novel genes influencing eosinophil-specific protein levels in asthma families. *J Allergy Clin Immunol.* 2022;150(5):1168-77.
33. Kooijman MN, Kruithof CJ, van Duijn CM, Duijts L, Franco OH, van IMH, et al. The Generation R Study: design and cohort update 2017. *Eur J Epidemiol.* 2016;31(12):1243-64.
34. Medina-Gomez C, Felix JF, Estrada K, Peters MJ, Herrera L, Kruithof CJ, et al. Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. *Eur J Epidemiol.* 2015;30(4):317-30.
35. Prijatelj V, Grgic O, Uitterlinden AG, Wolvius EB, Rivadeneira F, Medina-Gomez C. Bone health index in the assessment of bone health: The Generation R Study. *Bone.* 2024;182:117070.
36. Heinrich J, Bolte G, Hölscher B, Douwes J, Lehmann I, Fahlbusch B, et al. Allergens and endotoxin on mothers' mattresses and total immunoglobulin E in cord blood of neonates. *Eur Respir J.* 2002;20(3):617-23.
37. Berg A, Krämer U, Link E, Bollrath C, Heinrich J, Brockow I, et al. Impact of early feeding on childhood eczema: development after nutritional intervention compared with the natural course - the GINIplus study up to the age of 6 years. *Clin Exp Allergy.* 2010;40(4):627-36.
38. Reed E, Nunez S, Kulp D, Qian J, Reilly MP, Foulkes AS. A guide to genome-wide association analysis and post-analytic interrogation. *Stat Med.* 2015;34(28):3769-92.

39. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016;48(10):1284-7.
40. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007;39(7):906-13.
41. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007;447(7145):661-78.
42. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet.* 2010;11(7):499-511.
43. Ahrens W, Siani A, Adan R, De Henauw S, Eiben G, Gwozdz W, et al. Cohort Profile: The transition from childhood to adolescence in European children-how I.Family extends the IDEFICS cohort. *Int J Epidemiol.* 2017;46(5):1394-5j.
44. Ahrens W, Bammann K, Siani A, Buchecker K, De Henauw S, Iacoviello L, et al. The IDEFICS cohort: design, characteristics and participation in the baseline survey. *Int J Obes (Lond).* 2011;35 Suppl 1:S3-15.
45. Weale ME. Quality control for genome-wide association studies. *Methods Mol Biol.* 2010;628:341-72.
46. Guxens M, Ballester F, Espada M, Fernández MF, Grimalt JO, Ibarluzea J, et al. Cohort Profile: the INMA--INfancia y Medio Ambiente--(Environment and Childhood) Project. *Int J Epidemiol.* 2012;41(4):930-40.
47. Arshad SH, Holloway JW, Karmaus W, Zhang H, Ewart S, Mansfield L, et al. Cohort Profile: The Isle Of Wight Whole Population Birth Cohort (IOWBC). *Int J Epidemiol.* 2018;47(4):1043-4i.
48. Fong WCG, Chan A, Zhang H, Holloway JW, Roberts G, Kurukulaarachy R, et al. Childhood food allergy and food allergen sensitisation are associated with adult airways disease: A birth cohort study. *Pediatr Allergy Immunol.* 2021;32(8):1764-72.
49. Venkataraman D, Erlewyn-Lajeunesse M, Kurukulaarachy RJ, Potter S, Roberts G, Matthews S, et al. Prevalence and longitudinal trends of food allergy during childhood and adolescence: Results of the Isle of Wight Birth Cohort study. *Clin Exp Allergy.* 2018;48(4):394-402.
50. Kurukulaarachy RJ, Zhang H, Patil V, Raza A, Karmaus W, Ewart S, et al. Identifying the heterogeneity of young adult rhinitis through cluster analysis in the Isle of Wight birth cohort. *J Allergy Clin Immunol.* 2015;135(1):143-50.
51. Granell R, Curtin JA, Haider S, Kitaba NT, Mathie SA, Gregory LG, et al. A meta-analysis of genome-wide association studies of childhood wheezing phenotypes identifies ANXA1 as a susceptibility locus for persistent wheezing. *Elife.* 2023;12.
52. Holle R, Happich M, Löwel H, Wichmann HE. KORA--a research platform for population based health research. *Gesundheitswesen.* 2005;67 Suppl 1:S19-25.
53. Zijlema WL, Smidt N, Klijs B, Morley DW, Gulliver J, de Hoogh K, et al. The LifeLines Cohort Study: a resource providing new opportunities for environmental epidemiology. *Arch Public Health.* 2016;74:32.
54. Westerlaken-van Ginkel CD, Vonk JM, Flokstra-de Blok BMJ, Sprickelman AB, Koppelman GH, Dubois AEJ. Likely questionnaire-diagnosed food allergy in 78, 890 adults from the northern Netherlands. *PLoS One.* 2020;15(5):e0231818.
55. Zheng X, Davis JW. SAIGEgds-an efficient statistical tool for large-scale PheWAS with mixed models. *Bioinformatics.* 2021;37(5):728-30.
56. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wofford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet.* 2018;50(9):1335-41.
57. Zheng X, Gogarten SM, Lawrence M, Stilp A, Conomos MP, Weir BS, et al. SeqArray-a storage-efficient high-performance data format for WGS variant calls. *Bioinformatics.* 2017;33(15):2251-7.
58. von Mutius E, Schmid S, Group PS. The PASTURE project: EU support for the improvement of knowledge about risk factors and preventive factors for atopy in Europe. *Allergy.* 2006;61(4):407-13.
59. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods.* 2011;9(2):179-81.

60. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012;44(8):955-9.
61. Fairley S, Lowy-Gallego E, Perry E, Flicek P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res.* 2020;48(D1):D941-d7.
62. Wijga AH, Kerkhof M, Gehring U, de Jongste JC, Postma DS, Aalberse RC, et al. Cohort profile: the prevention and incidence of asthma and mite allergy (PIAMA) birth cohort. *Int J Epidemiol.* 2014;43(2):527-35.
63. Völzke H, Alte D, Schmidt CO, Radke D, Lorbeer R, Friedrich N, et al. Cohort profile: the study of health in Pomerania. *Int J Epidemiol.* 2011;40(2):294-307.
64. Völzke H, Schössow J, Schmidt CO, Jürgens C, Richter A, Werner A, et al. Cohort Profile Update: The Study of Health in Pomerania (SHIP). *Int J Epidemiol.* 2022;51(6):e372-e83.
65. Newnham JP, Evans SF, Michael CA, Stanley FJ, Landau LI. Effects of frequent ultrasound during pregnancy: a randomised controlled trial. *Lancet.* 1993;342(8876):887-91.
66. Straker L, Mountain J, Jacques A, White S, Smith A, Landau L, et al. Cohort Profile: The Western Australian Pregnancy Cohort (Raine) Study-Generation 2. *Int J Epidemiol.* 2017;46(5):1384-5j.
67. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38(8):904-9.
68. Raitakari OT, Juonala M, Rönnemaa T, Keltikangas-Järvinen L, Räsänen L, Pietikäinen M, et al. Cohort profile: the cardiovascular risk in Young Finns Study. *Int J Epidemiol.* 2008;37(6):1220-6.
69. Smith EN, Chen W, Kähönen M, Kettunen J, Lehtimäki T, Peltonen L, et al. Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. *PLoS Genet.* 2010;6(9):e1001094.
70. Raedler D, Ballenberger N, Klucker E, Böck A, Otto R, Prazeres da Costa O, et al. Identification of novel immune phenotypes for allergic and nonallergic childhood asthma. *J Allergy Clin Immunol.* 2015;135(1):81-91.
71. Salvermoser M, Zeber K, Boeck A, Klucker E, Schaub B. Childhood asthma: Novel endotyping by cytokines, validated through sensitization profiles and clinical characteristics. *Clin Exp Allergy.* 2021;51(5):654-65.
72. van Ginkel CD, Flokstra-de Blok BM, Kollen BJ, Kukler J, Koppelman GH, Dubois AE. Loss-of-function variants of the filaggrin gene are associated with clinical reactivity to foods. *Allergy.* 2015;70(4):461-4.
73. Asai Y, Eslami A, van Ginkel CD, Akhabir L, Wan M, Ellis G, et al. Genome-wide association study and meta-analysis in multiple populations identifies new loci for peanut allergy and establishes C11orf30/EMSY as a genetic risk factor for food allergy. *J Allergy Clin Immunol.* 2018;141(3):991-1001.
74. van Ginkel CD, Pettersson ME, Dubois AEJ, Koppelman GH. Association of STAT6 gene variants with food allergy diagnosed by double-blind placebo-controlled food challenges. *Allergy.* 2018;73(6):1337-41.
75. Vlieg-Boerstra BJ, Bijleveld CM, van der Heide S, Beusekamp BJ, Wolt-Plompen SA, Kukler J, et al. Development and validation of challenge materials for double-blind, placebo-controlled food challenges in children. *J Allergy Clin Immunol.* 2004;113(2):341-6.
76. Marenholz I, Grosche S, Kalb B, Rüschenhoff F, Blümchen K, Schlögl R, et al. Genome-wide association study identifies the SERPINB gene cluster as a susceptibility locus for food allergy. *Nat Commun.* 2017;8(1):1056.