Genome-Wide Association Study Reveals a Causal Relationship Between Allergic Rhinitis and Hazelnut Allergy

Sun, Yidan; Vonk, Judith M; Kersten, Elin T G; Qi, Cancan; Sprikkelman, Aline B; Koppelman, Gerard H

Published in:
Allergy

DOI:
10.1111/all.16411

Publication date:
2025

Allergy

WILEY

Allergy EUROPEAN JOURNAL OF ALLERGY AND CLINICAL IMMUNOLOGY EAACI

ORIGINAL ARTICLE OPEN ACCESS

Epidemiology and Genetics

Genome-Wide Association Study Reveals a Causal Relationship Between Allergic Rhinitis and Hazelnut Allergy

Yidan Sun[1,2] | Judith M. Vonk[2,3] | Elin T. G. Kersten[1,2] | Cancan Qi[4] | Aline B. Sprikkelman[1,2] | Gerard H. Koppelman[1,2]

[1]Department of Pediatric Pulmonology and Pediatric Allergy, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands | [2]University Medical Center Groningen, GRIAC Research Institute, University of Groningen, Groningen, The Netherlands | [3]Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands | [4]Division of Laboratory Medicine, Microbiome Medicine Center, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong, China

Correspondence: Gerard H. Koppelman (g.h.koppelman@umcg.nl)

Received: 20 June 2024 | **Revised:** 15 October 2024 | **Accepted:** 23 October 2024

Funding: The authors received no specific funding for this work.

Keywords: atopy | food allergy | genetics | genomics | GWAS

ABSTRACT

Background: Little is known about the genetics of food allergy (FA) to various allergens and the heterogeneity of FA in adults.

Objective: We aimed to investigate genetic susceptibility to FA in an adult population and to assess the association between secondary FA and allergic rhinitis (AR).

Methods: FA and allergen-specific FA were defined based on in-depth questionnaires and a previously published FA algorithm in the Lifelines. We performed a series of genome-wide association studies (GWAS) on FA and nine allergen-specific (e.g., hazelnut) FA in 21,353 adults in Lifelines. Single nucleotide polymorphisms (SNPs) ($p < 1E\text{-}5$) were replicated in a second independent set of 15,518 adults participating in the Lifelines followed by meta-analysis of the results of the two datasets. We subsequently investigated the causal relationship of AR to FA using Mendelian randomization (MR) analysis.

Results: We observed co-occurrence of tree nuts and apple FA, with over 80% of this group also reporting AR. After meta-analysis, we identified one genome-wide significant locus near *HLA-DPA1* associated with self-reported hazelnut allergy (hazelnutFA), of which the top SNP is rs5025825 ($p = 2.51E\text{-}9$, OR = 1.43). Two-sample MR indicated that AR is a significant causal risk factor for hazelnutFA ($p_{\text{-IVW}} = 5.27E\text{-}10$, $\beta = 5.90$, $p_{\text{-pleiotropy}} = 0.46$).

Conclusion: Our questionnaire enabled a large GWAS on self-reported FA in Dutch adults. We report one novel locus in the human leukocyte antigens (HLA) region associated with hazelnutFA, implying an association with antigen recognition. Our findings genetically link secondary FA to AR in adults.

Abbreviations: AR, allergic rhinitis; DBPCFC, double-blind placebo-controlled food challenges; dd-AR, doctor-diagnosed allergic rhinitis; EA, effect allele; eQTL, expression quantitative trait locus; FA, food allergy; GWAS, genome-wide association studies; HLA, human leukocyte antigens; IVs, instrumental variables; IVW, inverse-variance weighted; LD, linkage disequilibrium; LDSC, linkage disequilibrium score regression; likelyFA, likely food allergic; MHC, major histocompatibility complexes; MR, Mendelian randomization; noFA, not food allergic; OFC, oral food challenges; PCs, principal components; SNP, single nucleotide polymorphism; SPT, skin prick tests; sr-AR, self-reported allergic rhinitis; UKB, UK Biobank.

1 | Introduction

Food allergy (FA) is a heterogeneous, immunologic, mostly IgE-mediated disease triggered by specific food proteins that significantly affects quality of life for many children and adults, and even poses a potential threat of fatality. Oral food challenge-diagnosed FA has been reported to have a prevalence as high as 10% in Western countries, China and Africa, affecting both pediatric and adult populations [1]. The etiology of FA involves a complex interplay of genetic and environmental factors [2]. Twin studies indicate a heritability range of 51%–82% for FA [3, 4]. Previous genome-wide association studies (GWAS) have identified various single-nucleotide polymorphisms (SNPs) and genes associated with FA, some of which are specific to individual foods, while others are shared across different FA or overlap with other allergic diseases such as allergic rhinitis (AR), asthma, and eczema [5]. These SNPs associated with genes related to barrier and immune function. However, previous relatively large-scale studies have predominantly focused on FA in European children [5]. In contrast, existing research on adults has been limited to specific allergies and was primarily conducted within East Asian populations [6, 7], often restricted by smaller sample sizes. A comprehensive study on FA in adults is needed to further understand the genetics of FA in this group.

One of the challenges of previous GWAS on FA is heterogeneity in case definition [5]. Patients with FA may have symptoms which are not very specific and can affect multiple systems in the body [8]. To conduct GWAS on a larger scale, it is required to define FA precisely and efficiently. Although skin prick tests (SPT), serum IgE tests, and oral food challenges (OFC) are recognized for supporting a more accurate diagnosis [9], these are not feasible in studies of over tens of thousands of participants. SPT and serum IgE tests are invasive, and their positive results do not invariably indicate the presence of FA in clinical settings [10]. OFC, which is so called "gold standard" for diagnosing FA, is expensive, time-consuming, and pose potential risks to participants [11]. However, self-reported FA may strongly overestimate the prevalence of FA [12], reducing the power of genetic studies due to misclassification of cases. Therefore, our research group developed a questionnaire and algorithm aimed at mitigating misclassification when identifying FA based on the use of questionnaires only [13], enabling the recruitment of a larger sample size necessary for GWAS. Here, we only label a participant to be "likely food allergic (likelyFA)" if a specific reaction to a common allergen within a limited timeframe is reported [13], thereby contributing to an improved diagnosis of FA in large-scaled population-based studies.

The heterogeneity of FA also adds to the complexity. In addition to directly triggering allergic reactions through food consumption, cross-reactivity between certain foods and inhaled aeroallergens can lead to allergic reactions [14]. The categorization of FA into primary and secondary types is based on the mechanism triggering allergic reactions: primary FA directly results from food allergens, while secondary FA (also called pollen food allergy syndrome) results from a cross-reactivity between tree/grass pollen and food allergens [15]. Currently, it remains uncertain whether the genetic factors associated with primary and secondary FA are identical, as secondary FA is more prevalent in adults [16] than in children.

Therefore, we aimed to investigate the genetic risk factors related to likelyFA as well as allergen-specific FA in a large population-based adult Dutch cohort; and to assess the (causal) relation between AR and FA.

2 | Methods

Study subjects were selected from the Lifelines. Lifelines is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviors 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, behavioral, physical and psychological factors which contribute to health and disease of the general population, with a special focus on multi-morbidity and complex genetics [17]. 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 [13].

The presence of FA was determined according to the algorithm that was developed by Westerlaken-van Ginkel et al. [13]. Participants with FA were classified as likelyFA, when they reported to exhibit typical reactions (Table S1) to common, well-known allergens (Table S2). Subjects were further excluded from the likelyFA group (1) if they were solely diagnosed by an alternative health practitioner, (2) if the onset of symptoms after ingesting the food is after a day or more, (3) if the duration of symptoms exceeded 1 week, or (4) if they reported a negative result in a double-blind placebo-controlled food challenge (DBPCFC) test. Allergen-specific FA cases met both the likelyFA criteria and reported to a specific allergen. Within the likelyFA group, we defined nine groups of allergen-specific cases based on allergens reported by more than 1% of the participants. Participants without FA were classified as not food allergic (noFA), consisting of individuals that indicated "I don't have food allergy" in the questionnaire. The remaining subjects were classified as indeterminate and excluded from further analysis. In Lifelines, self-reported allergic rhinitis (sr-AR) was determined through questionnaire responses regarding a history of nasal allergy including hay fever [18]. The definition of asthma and eczema in Lifelines was previously described [18].

The genetic data was collected in two sub-cohorts of Lifelines. In 2019, the first batch of 38,030 Lifelines participants was genotyped using the Infinium Global Screening Array (GSA) MultiEthnic Disease Version 1.0. In 2023, a second batch of 29,166 Lifelines participants was genotyped using the FinnGen Thermo Fisher Axiom custom array. The quality control and genetic imputation of these two batches were performed separately. Details on quality control and genetic imputation can be found in the supporting data [19].

This GWAS on FA in adults included individuals of 18 years or older, who completed the Food Allergy Questionnaire, and whose DNA was genotyped. Cases were individuals who are classified as likelyFA or allergen-specific FA. Controls were

individuals who are classified as noFA. We used a two-stage GWAS strategy: discovery and replication (Figure 1).

We conducted a series of GWAS in the discovery set (21,353) which consisted of the first batch of genotyped Lifelines participants. SAIGEgds (1.12.2) [20–22] was used for association analysis on chromosome 1–22, incorporating age, sex, and genetic principal components (PCs) 1–10 as covariates. SAIGEgds was employed to adjust for potential family structure and to address the imbalance in case–control ratios within our dataset. R (4.2.1), qqman (0.1.9) [23], and LocusZoom [24] was used for statistical analysis and graphing. Our genome-wide significance threshold was set at $p < 5E\text{-}8$. In the FA with significant SNPs, we extended the threshold and created a list of lead SNPs with $p < 1E\text{-}5$ for replication; FA without significant results were not followed-up in replication analyses. Finally, previously reported FA SNPs [6, 7, 25–28] and proxy SNPs which in linkage disequilibrium (LD) with FA SNPs were also tested in the discovery set (See Data S1).

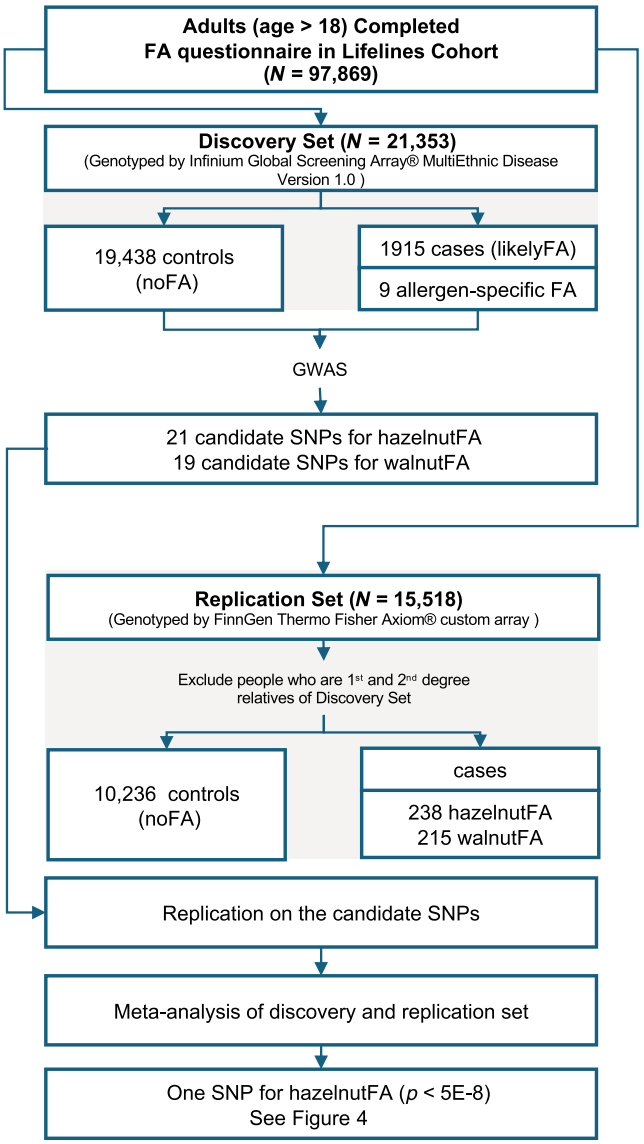

FIGURE 1 | Study design. The discovery cohort consisted of 21,353 individuals. GWAS on 10 phenotypes (likelyFA and nine allergen-specific FA) were conducted. Phenotypes with SNPs with $p < 5E\text{-}8$ were selected for replication. A looser threshold of $p < 1E\text{-}5$ was used to select SNPs in independent loci for replication. Following the replication phase and subsequent meta-analysis, one SNP was identified as being associated with hazelnutFA.

In the replication set, which encompasses 15,518 individuals genotyped in the second batch of Lifelines participants, we removed samples found to be first or second-degree relatives of those in the discovery set (PI_HAT > 0.1875). The case and control definitions and were the same as in the discovery set. Subsequently, we conducted an association analysis on the SNP list generated in the discovery set using SAIGEgds [20–22], employing the same model as that used in the GWAS. Bonferroni adjustment was used to correct for multiple testing, using the number of independent SNPs as the number of independent tests. The significance threshold was calculated as $p < (0.05$ divided by the number of SNPs). We meta-analyzed these SNPs across the discovery and replication sets using effect size estimates and standard errors based on the Inverse Variance Weighted (IVW) model, implemented in the METAL software [29].

To determine if the SNPs had any effect on gene expression, we search the SNPs in the GTEx portal and visualized the colocalization using LocusCompare [30]. Human leukocyte antigens (HLA) imputation was performed using the HIBAG R packages (v1.38.0) [31, 32] to predict HLA alleles from nearby genetic markers according to LD in both the discovery and replication set and obtained 183 variants in the major histocompatibility complexes (MHC) region in the discovery set. The association analysis in HLA region and meta-analysis was conducted using the same method as in the GWAS. In the discovery set, p-values of the association of HLA alleles with FA were corrected for the number of alleles/haplotypes tested (shown as p-corrected $[pc] = p^*$(number of test), applying the significance threshold of $pc < 0.05$ [33] for replication).

To better understand the secondary FA caused by cross-reacting allergens found in both pollen and food, we used AR as a proxy for pollen allergy to assess the causal relation between pollen allergy and FA. First, we investigated the associations between FA-specific SNPs and sr-AR within the Lifelines dataset. Then we incorporated GWAS summary statistics for doctor diagnosed allergic rhinitis (dd-AR), obtained from public GWAS studies conducted within the UK Biobank (UKB) to indicate the genetics of AR [34, 35]. We assessed the genetic overlap and potential causal relationships between FA outcomes and dd-AR using LDSC (Linkage Disequilibrium Score Regression) [36, 37], and TwoSampleMR (0.5.7) [35, 38]. IVW multivariable Mendelian randomization (MR) analysis and MR Egger were used as the MR method, using all independent dd-AR SNPs ($p < 5E\text{-}8$) obtained from GWAS in UKB [34, 35] as instrumental variables (IVs) after sensitivity analysis.

3 | Results

3.1 | Study Population

The discovery set consisted of 1915 likelyFA cases (8.67% of the total), which were compared to 19,438 control subjects. In likelyFA, nine allergens had a prevalence higher than 1%, including

apple (appleFA, $N=577$), hazelnut (hazelnutFA, $N=489$), cow milk (cowsmilkFA, $N=436$), walnut (walnutFA, $N=395$), wheat (wheatFA, $N=271$), shellfish (shellfishFA, $N=225$), peanut (peanutFA, $N=223$), almond (almondFA, $N=221$), and kiwi (kiwiFA, $N=219$). Demographic characteristics of the subjects including age, sex, and other allergic diseases are shown in Table 1.

The co-occurrence of likelyFA, allergen-specific FA, as well as with other allergic diseases (asthma, sr-AR, and eczema), can be observed in Figure 2. We observed co-occurrence of reported allergies between tree nuts and apples. Among individuals reporting an allergy to apple, hazelnut, walnut, or almond, more than 80% also reported having sr-AR (Table 1 and Figure 2). We also observed co-occurrence of reported

TABLE 1 | Demographics of the discovery and replication sets.

	Age in years	Sex	Allergic rhinitis	Eczema	Asthma
	Mean (SD)	Female (%)	Affected individuals (%)	Affected individuals (%)	Affected individuals (%)
Discovery					
Controls					
NoFA ($N=19,438$)	48.6 (14.3)	11,419 (58.7%)	4348 (22.8%)	2638 (14.0%)	1400 (7.4%)
Cases					
LikelyFA ($N=1915$)	46.4 (13.5)[a]	1419 (74.1%)[a]	1146 (60.9%)[a]	511 (27.5%)[a]	327 (17.4%)[a]
Top 9 reported allergies in likelyFA					
AppleFA ($N=577$)	45.0 (13.1)[a]	402 (69.7%)[a]	503 (88.2%)[a]	157 (27.9%)[a]	108 (19.0%)[a]
HazelnutFA ($N=489$)	45.3 (13.3)[a]	341 (69.7%)[a]	437 (90.3%)[a]	151 (31.7%)[a]	123 (25.5%)[a]
CowsmilkFA ($N=436$)	45.3 (13.7)[a]	356 (81.7%)[a]	207 (48.9%)[a]	126 (30.1%)[a]	77 (18.3%)[a]
WalnutFA ($N=395$)	44.3 (13.1)[a]	287 (72.7%)[a]	329 (83.9%)[a]	134 (34.8%)[a]	99 (25.3%)[a]
WheatFA ($N=271$)	45.5 (13.4)[a]	230 (84.9%)[a]	119 (45.2%)[a]	70 (27.0%)[a]	46 (17.6%)[a]
ShellfishFA ($N=225$)	52.1 (12.9)[a]	159 (70.7%)[a]	92 (41.4%)[a]	44 (20.4%)	33 (14.9%)[a]
PeanutFA ($N=223$)	44.3 (15.8)[a]	156 (70.0%)[a]	153 (69.9%)[a]	87 (40.7%)[a]	66 (30.3%)[a]
AlmondFA ($N=221$)	43.6 (13.7)[a]	166 (75.1%)[a]	181 (82.6%)[a]	80 (37.4%)[a]	72 (32.9%)[a]
KiwiFA ($N=219$)	45.8 (13.0)[b]	186 (84.9%)[a]	122 (57.0%)[a]	73 (34.3%)[a]	37 (17.3%)[a]
Replication					
Controls					
NoFA ($N=10,236$)	49.1 (12.8)	6350 (62.0%)	2372 (23.3%)	1445 (14.3%)	802 (7.9%)
Cases					
HazelnutFA ($N=238$)	45.9 (12.7)[a]	177 (74.4%)[a]	196 (83.1%)[a]	78 (33.2%)[a]	57 (24.3%)[a]
WallnutFA ($N=215$)	45.1 (13.1)[a]	150 (69.8%)[c]	159 (74.6%)[a]	68 (32.1%)[a]	50 (23.6%)[a]

[a]Represents statistical significance compared to the noFA ($p<0.001$).
[b]Represents statistical significance compared to the noFA ($p<0.01$).
[c]Represents statistical significance compared to the noFA ($p<0.05$).

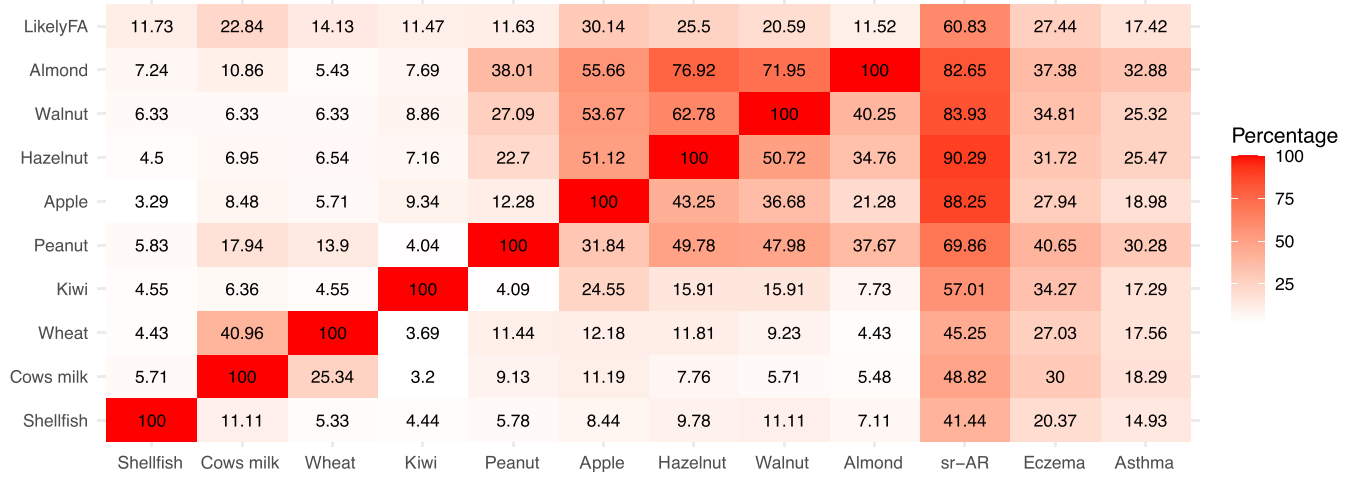

FIGURE 2 | Co-occurrence of different allergies and allergic diseases in discovery set. Heatmap to show the percentage of overlapping individuals who self-reported one FA (likelyFA and allergen specific FA) on the *y*-axis that also reported FA or allergic diseases on the *x*-axis. The color gradient toward red on the plot signifies that a greater proportion of individuals exhibiting the phenotype represented on the *y*-axis also report the phenotype on the *x*-axis.

allergies between wheat and cow's milk. In addition, individuals reporting allergies to kiwi, wheat, cow's milk, and shellfish are less likely to suffer from AR compared to those with tree nuts or apple allergies. It was shown that peanut allergy is more frequently associated with eczema than other FA (Table 1 and Figure 2).

3.2 | Discovery Analysis

No genome-wide significant loci were detected in the GWAS on likelyFA (Figure S1). A locus in the intergenic region between *HLA-DOA* and *HLA-DPA1* on chromosome 6 (rs5025825, $p = 1.26E\text{-}8$, effect allele (EA) = A, OR = 1.52, Figure 3) was significantly associated with hazelnutFA in adults. A locus in the intergenic region between the *MEAT6* and *LOC124901500* on chromosome 6 (rs117169933, $p = 4.94E\text{-}08$, EA = T, OR = 4.96, Figure S1) was associated with walnutFA. Conditional analysis did not reveal additional variants within the corresponding locus that associated with hazelnutFA or walnutFA. When setting threshold *p* value for replication to < 1E-5, 21 independent SNPs were selected with suggestive associations with hazelnutFA (Table S3) and 19 independent SNPs with suggestive associations ($p < 1E\text{-}5$) with walnutFA (Table S4). No genome-wide significant loci were detected from the other seven allergen-specific FA GWAS, which therefore were not followed up in replication (Figure S1). Finally, we observed that several proxy SNPs that have LD with previous FA SNPs had a *p*-value < 0.05 in our data (Table S5).

3.3 | Replication and Meta-Analysis

In the replication study, we investigated 21 hazelnutFA SNPs in 258 hazelnutFA case subjects and 11,475 control subjects and found that rs5025825 was associated with hazelnutFA ($p = 0.02$, EA = A, OR = 1.27). Meta-analysis showed a genome-wide significant association between rs5025825 and hazelnutFA across the discovery and replication sets ($p = 2.51E\text{-}09$, OR = 1.43) (Figure 4, Figure S2, Table S3). Twenty other SNPs associated with hazelnutFA at a looser *p*-value threshold ($5E\text{-}8 < p < 1E\text{-}5$) that did not reach genome-wide significance in either the replication set or the combined meta-analysis (Figure S2, Table S3). The analysis of 19 SNPs ($p < 1E\text{-}5$) associated with walnutFA showed no significant associations in either the replication set (investigated in 235 cases and 11,475 controls) or the combined meta-analysis (Figure S3, Table S4).

3.4 | Association of HLA Region Genes With hazelnutFA

The nearest gene of rs5025825 is *HLA-DPA1* (Figure 3C). This SNP serves as an expression quantitative trait locus (eQTL) of multiple nearby genes across multiple tissues including esophagus (*HLA-DPA1*, $p = 3.3E\text{-}7$), lung (*HLA-DPA1*, $p = 1.53E\text{-}5$), and skin (*HLA-DPA1*, $p = 6.03E\text{-}19$) (Table S6). It was implied that rs5025825 (risk allele = A) is associated with a lower expression of *HLA-DPA1* in esophagus, skin, and lung (Table S6, Figure S4). Association tests on HLA genotype indicated that there were six HLA alleles associated with hazelnutFA ($pc < 0.05$), and *HLA-DPB1*04:01* is the most significant ($pc = 1.89E\text{-}05$, OR = 0.73). Four of six HLA alleles were imputed in the replication set and *HLA-DQB1 05:01* was replicated (Table S7).

3.5 | The hazelnutFA Variant is Associated With likelyFA and Other Allergen-Specific FA

We next examined the top hazelnutFA SNP (rs5025825, risk allele = A) for associations with likelyFA and other top allergens respectively in the discovery set. The rs5025825 was associated with appleFA ($p = 5.93E\text{-}07$, OR = 1.40), almondFA ($p = 0.001$, OR = 1.40), walnutFA ($p = 0.002$, OR = 1.29),

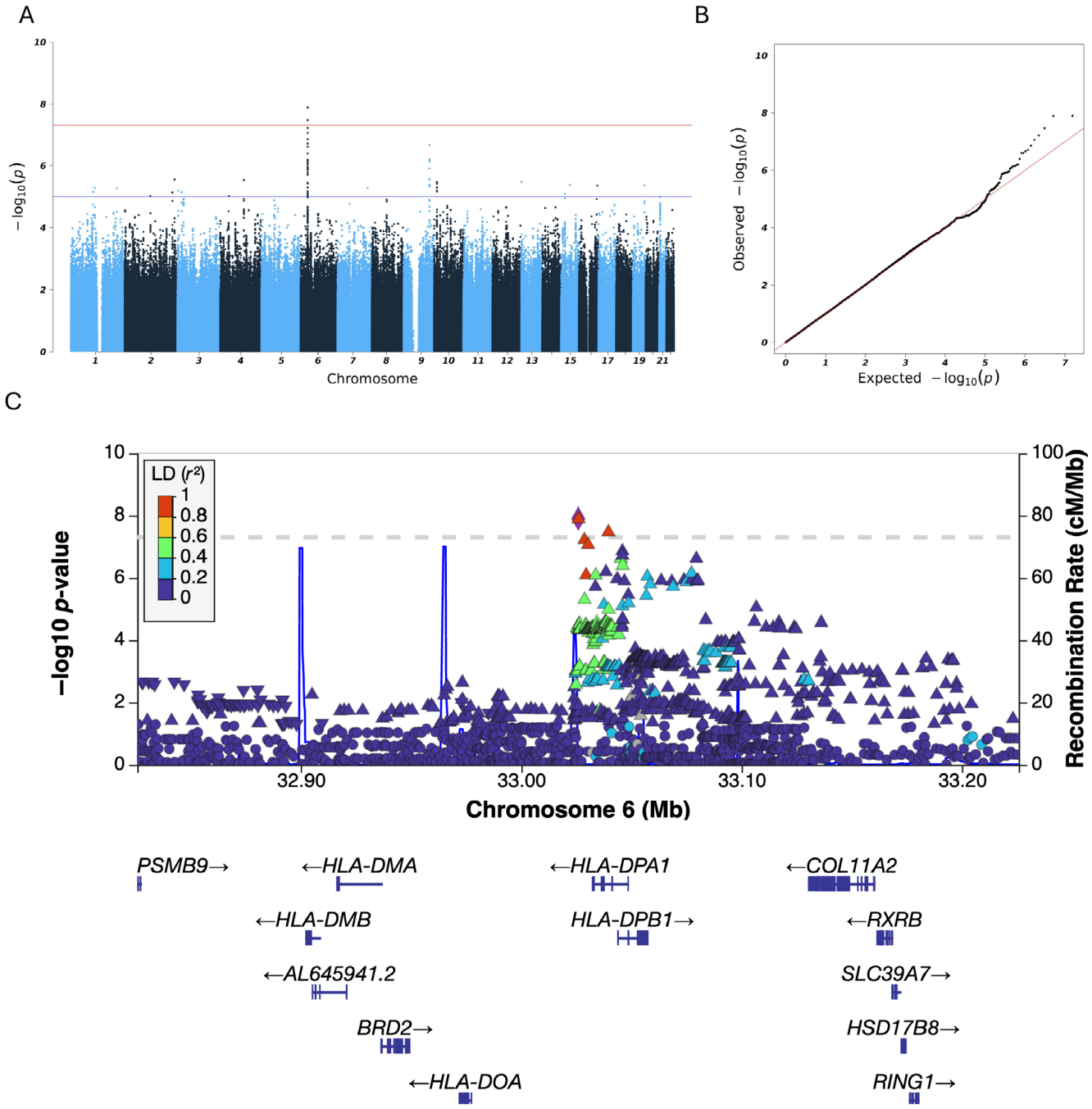

FIGURE 3 | GWAS on hazelnutFA in adults shows an association at the HLA region (discovery, $N = 19{,}927$). (A) Manhattan plot of GWAS on hazelnutFA. The red horizontal line represents the genome-wide significant threshold ($p < 5E\text{-}8$), and the blue horizontal line represents the suggestive threshold ($p < 1E\text{-}5$). (B) QQ-plot of GWAS on hazelnutFA ($\lambda = 1.01$). The red line represents the expected distribution of the p-value. (C) Regional plot of hazelnutFA association signal. The gray dotted horizontal line represents the genome-wide significant threshold ($p < 5E\text{-}8$). Points colored by LD (r^2) to the top SNP (rs5025825, the purple rhombus).

likelyFA ($p = 0.003$, OR = 1.12), and peanutFA ($p = 0.038$, OR = 1.25) (Figure S5).

3.6 | The Association Between hazelnutFA and AR

We used AR as a proxy for grass and tree pollen allergy and next investigated the association of hazelnutFA with sr-AR in Lifelines. It was observed that 90.3% of the participants in the discovery cohort and 83.1% in the replication cohort with hazelnutFA also had sr-AR (Table 1). Rs5025825 (risk allele = A) was modestly associated to sr-AR in Lifelines ($p = 0.02$, OR = 1.04). When we adjusted the genetic association between rs5025825 (risk allele = A) and hazelnutFA for sr-AR, it became less significant ($p = 7.14E\text{-}08$, OR = 1.38).

To further understand the genetic correlation between hazelnutFA in adults and AR, we investigated the relationship between hazelnutFA and AR, and found a significant genetic correlation between hazelnutFA (Lifelines) and dd-AR (UKB)

[34, 35] (rg = 0.87, $p = 0.04$). We investigated a potential causal relationship between hazelnutFA and dd-AR by performing a two-sample MR. We analyzed a causal effect of dd-AR on hazelnutFA using 31 IVs obtained from GWAS in UKB [34, 35] after sensitive analysis (Figures S6–S8). The IVW MR estimate for hazelnutFA was statistically significant ($p = 5.27E\text{-}10$, $\beta = 5.90$) without pleiotropy ($p_{\text{-MR Egger intercept}} = 0.21$) or heterogeneity ($p_{\text{-Q-IVW}} = 0.425$), suggesting a causal relationship of dd-AR on hazelnutFA in adults (Figure 5). A comprehensive sensitivity analysis confirmed the consistency of the effect sizes estimated by individual genetic variants (Figure S6). Furthermore, no single variant was found to disproportionately influence the MR results (Figure S7), and there were no indications of potential biases in our analysis (Figure S8).

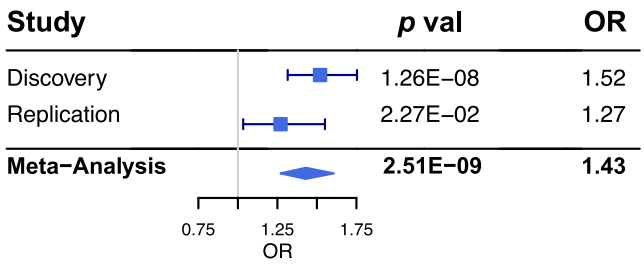

FIGURE 4 | Forest plots for lead SNP associated with hazelnutFA. Meta-analysis between the discovery and replication sets for the lead SNP (rs5025825) showed genome-wide significance. OR and 95% confidence interval are displayed on the *x*-axis.

4 | Discussion

This large GWAS on any FA and nine allergen-specific FA in Dutch adults identified a genetic variant near the *HLA-DPA1* locus on chromosome 6 as a risk factor for self-reported history of likely hazelnut FA at genome wide significance after meta-analysis. This locus also showed evidence for association with other FA, specifically apple, and with sr-AR. Using MR, we characterized AR as a causal risk factor of hazelnutFA. A strong association of many tree-nut allergy and apple allergy in adults was observed with sr-AR, and to a smaller extent with eczema and asthma (Figure 2). The co-occurrence suggests a differential pattern of cross-reactivity among food and environmental allergens. These results implied that information on allergen-specific FA may assist us in further subclassifying FA in adults; specifically, hazelnut allergy may serve as a representative marker to define this subgroup of FA.

Currently, multiple HLA loci related to FA have been discovered [39]. HLA genes which are located within the major MHC on chromosome 6p21.3 are divided into three classes in humans. Classes I and II encode a family of cell-surface proteins that play a key role in antigen presentation. There are many alleles of HLA genes: more than 25,000 HLA genes have been identified so far [40], and they are usually named by serological antigen types. We identified associations between the *HLA-DP* (HLA Class II) locus and hazelnutFA in our meta-analysis. The most significant SNP (rs5025825, allele A) which is near *HLA-DPA1* was shown to be associated with

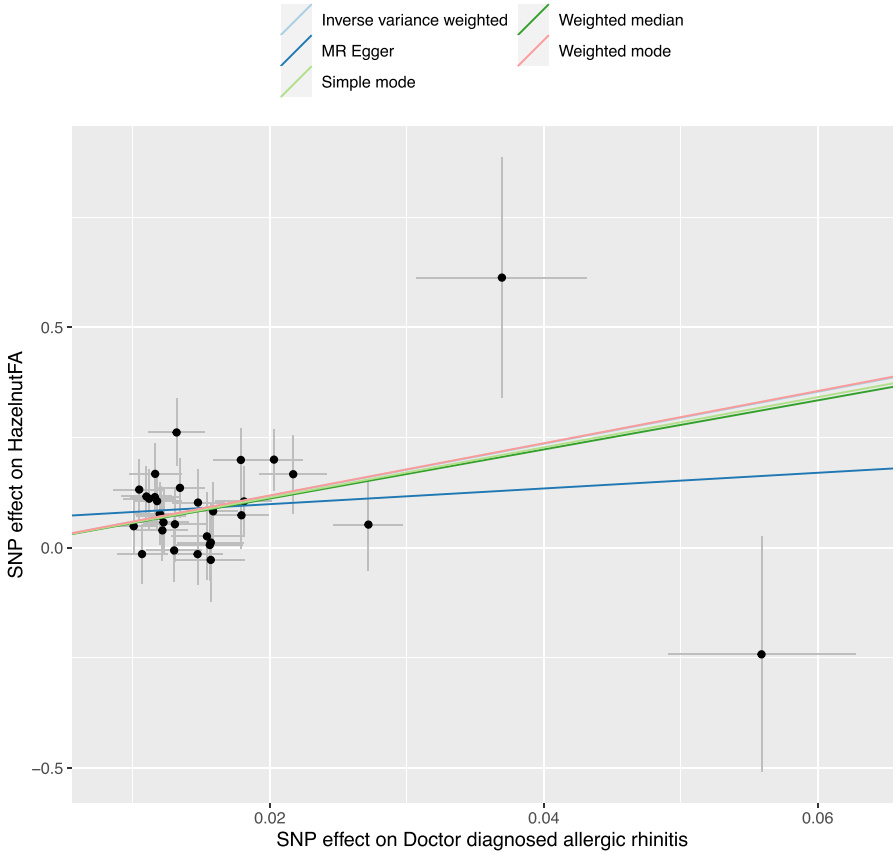

FIGURE 5 | Scatter plot for Mendelian randomization. Mendelian randomization scatter plot with effect sizes ($\beta \pm$ SE) of each variation allergic rhinitis as exposure and hazelnutFA as outcome ($p_{\text{-IVW}} = 5.27E\text{-}10$, $p_{\text{-pleiotropy}} = 0.46$; Figures S6–S8).

lower expression of *HLA-DPA1* in esophagus, skin, and lung (Table S6). Previously one other SNP near *HLA-DPA1* and *HLA-DPB1* was reported to be related to wheat-dependent exercise-induced anaphylaxis in a Japanese cohort [6]. However, the wheat-dependent exercise-induced anaphylaxis related locus seems to be independent from the hazelnutFA locus ($r^2 < 0.01$). To better characterize the HLA locus, we performed HLA imputation. Here, we observed a negative association between hazelnutFA and *HLA-DQB1*05:01* in the discovery set and verified this in the replication set (Table S7). This is in line with a previous study, that showed a significantly lower allele frequency of *HLA-DQB1*05:01* in patients with birch pollen-associated hazelnut allergy [41] and peach allergy [28] compared to non-affected controls. The *HLA-DPB1*04:01* allele displayed the most significant association with hazelnutFA in the discovery set (Table S7). *HLA-DPB1*04:01* was previously reported to protect genetically susceptible children from celiac disease autoimmunity [42]. Furthermore, a positive association between *HLA-DPB1*09:01* and hazelnutFA was observed in the discovery set (Table S7). *HLA-DPB1*09:01* was reported to be an HLA susceptibility risk haplotype for peach allergy [28]. HLA alleles could play an important role in the etiology of FA and further studies are needed to investigate this.

The subgroup of patients with FA represented by tree nuts allergy may represent secondary FA in adults. Secondary FA, also known as pollen-food allergy syndrome, is more prevalent in adults [16] than in children. It is conceivable that the genetic risk factors for primary and secondary FA differ, as well as its overlap with pollen allergy. Apple and tree nuts are recognized allergens with known cross-reactivity with pollen [43]. In our study, we found that individuals with allergies to apple and tree nuts exhibited a higher prevalence of sr-AR, and subsequently genetically confirming the correlation of AR with tree-nut FA, specifically hazelnut allergy. The association between rs5025825 and hazelnutFA is partly driven by sr-AR and rs5025825 showed association with sr-AR in Lifelines, suggesting a shared mechanism of AR and hazelnutFA. We provided evidence for a causal relationship from dd-AR to hazelnutFA in MR analysis, supporting the concept of secondary allergies: pollen triggers pollen food allergy syndrome [44].

Previous GWAS identified several loci related to FA [25, 26], peanut allergy [25–27], egg allergy [26], wheat allergy [6, 7], peach allergy [28], and shrimp allergy [28]. In our discovery set these SNPs did not achieve nominal significance ($p < 0.05$), however, we identified 1 proxy SNP of previous FA SNP that was nominally associated with likelyFA on chromosome 11 near *EMSY*, 2 proxy SNPs were related to peanutFA on chromosome 6 near *HLA-DPA1*and *HLA-DQB1-HLA-DQA1*, and 2 proxy SNPs that were related to wheatFA on chromosome 6 near *HLA-DPA1* and *HLA-DRB1-HLA-DQA* (Table S5). Although we provide support for several previously reported loci, the challenges in achieving replication may stem from variations in definition of FA, population demographics, differences between genetic risk factors for FA presenting in children and adults, and genetic heterogeneity. The genetic risk factors may differ between primary FA and secondary FA. Some previous GWAS studies were performed in populations from East Asia, while our studies were performed in a European white population, for example, rs74995702 risk allele is common in East Asians, while it is rare in Europeans (Table S5). Differences in allele types and frequencies, may make transethnic replication studies difficult [45].

This study has several strengths and limitations that need to be considered when interpreting our data. It is one of the largest GWAS on FA in an adult population utilizing a specific questionnaire to improve FA assessment in large genetic-epidemiological studies. Our questionnaire helps reducing false positive FA cases [13], which enabled us to study FA in large population. This is further supported by replication of previous results. However, questionnaire defined hazelnutFA may still be different from a diagnosis of hazelnut allergy by OFC, and some misclassification may still be present. Although this may reduce power, it is not likely to cause false-positive results. Further validation of the hazelnutFA associated SNP in additional populations with more strict phenotypes such as OFC is warranted. One limitation of our study is the lack of data on the age of onset of FA, which could further facilitate the verification of secondary FA. In addition, we note that the number of cases in our study is still limited for a genome wide association study. The other limitation is that we only performed one direction of MR because enough instrumental variables for the hazelnutFA (or FA) were lacking. These analyses can only be performed if future, well-powered GWAS studies reveal a robust SNP-set for FA. Finally, we restricted our analysis of the Dutch population to mostly European ancestry and were not able to investigate people of other ancestries, reflecting the genetic make-up of the population of the Northern Netherlands.

In conclusion, this study identified a subgroup of FA marked by hazelnut allergy in adults using a questionnaire. Using the MR framework, we have provided evidence for a potential causal effect of AR on this subgroup. We identified an associated variant for Hazelnut allergy in the HLA region, which also was associated with allergies to apple, almond, walnut, or peanut. We propose that our work may help to disentangle mechanisms and perhaps provide biomarkers of primary FA from secondary FA.

Author Contributions

Y.S., J.M.V., E.T.G.K., A.B.S., and G.H.K. contributed to the design of the study. J.M.V., E.T.G.K., A.B.S., and G.H.K. contributed to study procedures. Y.S., J.M.V., and C.Q. contributed to data analysis. All co-authors contributed to results interpretation. Y.S., E.T.G.K., and G.H.K. wrote the initial manuscript draft. All co-authors revised the manuscript draft for important intellectual content and approved the final version for submission.

Acknowledgments

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). The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: the GTEx Portal on 04/04/2024. Y.S. is supported by a grant from the Chinese Scholarship Council (CSC).

Conflicts of Interest

The authors declare no conflicts of interest. E.T.G.K. reports grant support from the Netherlands Lung Foundation. G.H.K. reports grant support from the Netherlands Lung Foundation, ZON-MW (VICI grant), Health Holland, European Union, TEVA the Netherlands, GSK and Vertex outside the submitted work (Money to Institution). G.H.K. participated in advisory boards or gave lectures supported by AZ, PURE-IMS, Boehringer Ingelheim and Sanofi (money to Institution). A.B.S. reports research grants from Nestle Research, Lausanne and Aimmune, outside the submitted work. Her institution received compensation for her consultancy for Sanofi Netherlands and Nestle Research, Lausanne.

Data Availability Statement

The data that supports the findings of this study are available in the Supporting Information of this article.

References

1. I. Agache, I. Annesi-Maesano, A. Bonertz, et al., "Prioritizing Research Challenges and Funding for Allergy and Asthma and the Need for Translational Research-The European Strategic Forum on Allergic Diseases," *Allergy* 74, no. 11 (2019): 2064–2076, https://doi.org/10.1111/all.13856.

2. S. H. Sicherer and H. A. Sampson, "Food Allergy: A Review and Update on Epidemiology, Pathogenesis, Diagnosis, Prevention, and Management," *Journal of Allergy and Clinical Immunology* 141, no. 1 (2018): 41–58, https://doi.org/10.1016/j.jaci.2017.11.003.

3. X. Liu, S. Zhang, H. J. Tsai, et al., "Genetic and Environmental Contributions to Allergen Sensitization in a Chinese Twin Study," *Clinical and Experimental Allergy* 39, no. 7 (2009): 991–998, https://doi.org/10.1111/j.1365-2222.2009.03228.x.

4. S. H. Sicherer, T. J. Furlong, H. H. Maes, R. J. Desnick, H. A. Sampson, and B. D. Gelb, "Genetics of Peanut Allergy: A Twin Study," *Journal of Allergy and Clinical Immunology* 106, no. 1 Pt 1 (2000): 53–56, https://doi.org/10.1067/mai.2000.108105.

5. K. Kanchan, S. Clay, H. Irizar, S. Bunyavanich, and R. A. Mathias, "Current Insights Into the Genetics of Food Allergy," *Journal of Allergy and Clinical Immunology* 147, no. 1 (2021): 15–28, https://doi.org/10.1016/j.jaci.2020.10.039.

6. K. Fukunaga, Y. Chinuki, Y. Hamada, et al., "Genome-Wide Association Study Reveals an Association Between the HLA-DPB1(*)02:01:02 Allele and Wheat-Dependent Exercise-Induced Anaphylaxis," *American Journal of Human Genetics* 108, no. 8 (2021): 1540–1548, https://doi.org/10.1016/j.ajhg.2021.06.017.

7. E. Noguchi, M. Akiyama, A. Yagami, et al., "HLA-DQ and RBFOX1 as Susceptibility Genes for an Outbreak of Hydrolyzed Wheat Allergy," *Journal of Allergy and Clinical Immunology* 144, no. 5 (2019): 1354–1363, https://doi.org/10.1016/j.jaci.2019.06.034.

8. C. E. Mendonca and D. A. Andreae, "Food Allergy," *Primary Care* 50, no. 2 (2023): 205–220, https://doi.org/10.1016/j.pop.2023.01.002.

9. A. F. Santos, C. Riggioni, I. Agache, et al., "EAACI Guidelines on the Diagnosis of IgE-Mediated Food Allergy," *Allergy* 78, no. 12 (2023): 3057–3076, https://doi.org/10.1111/all.15902.

10. N. A. LaHood and S. U. Patil, "Food Allergy Testing," *Clinics in Laboratory Medicine* 39, no. 4 (2019): 625–642, https://doi.org/10.1016/j.cll.2019.07.009.

11. N. Patel, W. G. Shreffler, A. Custovic, and A. F. Santos, "Will Oral Food Challenges Still be Part of Allergy Care in 10 Years' Time?," *Journal of Allergy and Clinical Immunology. In Practice* 11, no. 4 (2023): 988–996, https://doi.org/10.1016/j.jaip.2023.02.010.

12. B. I. Nwaru, L. Hickstein, S. S. Panesar, et al., "Prevalence of Common Food Allergies in Europe: A Systematic Review and Meta-Analysis," *Allergy* 69, no. 8 (2014): 992–1007, https://doi.org/10.1111/all.12423.

13. C. D. Westerlaken-van Ginkel, J. M. Vonk, B. M. J. Flokstra-de Blok, A. B. Sprikkelman, G. H. Koppelman, and A. E. J. Dubois, "Likely Questionnaire-Diagnosed Food Allergy in 78,890 Adults From the Northern Netherlands," *PLoS One* 15, no. 5 (2020): e0231818, https://doi.org/10.1371/journal.pone.0231818.

14. J. M. Smeekens, K. Bagley, and M. Kulis, "Tree Nut Allergies: Allergen Homology, Cross-Reactivity, and Implications for Therapy," *Clinical and Experimental Allergy* 48, no. 7 (2018): 762–772, https://doi.org/10.1111/cea.13163.

15. M. A. Faber, A. L. Van Gasse, I. I. Decuyper, et al., "Cross-Reactive Aeroallergens: Which Need to Cross Our Mind in Food Allergy Diagnosis?," *Journal of Allergy and Clinical Immunology: In Practice* 6, no. 6 (2018): 1813–1823, https://doi.org/10.1016/j.jaip.2018.08.010.

16. O. I. Iweala, S. K. Choudhary, and S. P. Commins, "Food Allergy," *Current Gastroenterology Reports* 20, no. 5 (2018): 17, https://doi.org/10.1007/s11894-018-0624-y.

17. W. L. Zijlema, N. Smidt, B. Klijs, et al., "The LifeLines Cohort Study: A Resource Providing New Opportunities for Environmental Epidemiology," *Archives of Public Health* 74 (2016): 32, https://doi.org/10.1186/s13690-016-0144-x.

18. M. A. Ferreira, J. M. Vonk, H. Baurecht, et al., "Shared Genetic Origin of Asthma, Hay Fever and Eczema Elucidates Allergic Disease Biology," *Nature Genetics* 49, no. 12 (2017): 1752–1757, https://doi.org/10.1038/ng.3985.

19. LifeLines Subcohort, "UGLI Subcohort," accessed May 6, 2024, https://wiki-lifelines.web.rug.nl/doku.php?id=ugli.

20. X. Zheng and J. W. Davis, "SAIGEgds-An Efficient Statistical Tool for Large-Scale PheWAS With Mixed Models," *Bioinformatics* 37, no. 5 (2021): 728–730, https://doi.org/10.1093/bioinformatics/btaa731.

21. W. Zhou, J. B. Nielsen, L. G. Fritsche, et al., "Efficiently Controlling for Case-Control Imbalance and Sample Relatedness in Large-Scale Genetic Association Studies," *Nature Genetics* 50, no. 9 (2018): 1335–1341, https://doi.org/10.1038/s41588-018-0184-y.

22. X. Zheng, S. M. Gogarten, M. Lawrence, et al., "SeqArray-A Storage-Efficient High-Performance Data Format for WGS Variant Calls," *Bioinformatics* 33, no. 15 (2017): 2251–2257, https://doi.org/10.1093/bioinformatics/btx145.

23. S. Turner, "Qqman: An R Package for Visualizing GWAS Results Using QQ and Manhattan Plots," *Journal of Open Source Software* 3 (2018): 731.

24. A. P. Boughton, R. P. Welch, M. Flickinger, et al., "LocusZoom. Js: Interactive and Embeddable Visualization of Genetic Association Study Results," *Bioinformatics* 37, no. 18 (2021): 3017–3018.

25. Y. Asai, A. Eslami, C. D. van Ginkel, 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," *Journal of Allergy and Clinical Immunology* 141, no. 3 (2018): 991–1001, https://doi.org/10.1016/j.jaci.2017.09.015.

26. I. Marenholz, S. Grosche, B. Kalb, et al., "Genome-Wide Association Study Identifies the SERPINB Gene Cluster as a Susceptibility Locus for Food Allergy," *Nature Communications* 8, no. 1 (2017): 1056, https://doi.org/10.1038/s41467-017-01220-0.

27. X. Hong, K. Hao, C. Ladd-Acosta, et al., "Genome-Wide Association Study Identifies Peanut Allergy-Specific Loci and Evidence of Epigenetic Mediation in US Children," *Nature Communications* 6 (2015): 6304, https://doi.org/10.1038/ncomms7304.

28. S. S. Khor, R. Morino, K. Nakazono, et al., "Genome-Wide Association Study of Self-Reported Food Reactions in Japanese Identifies Shrimp and Peach Specific Loci in the HLA-DR/DQ Gene Region," *Scientific Reports* 8, no. 1 (2018): 1069, https://doi.org/10.1038/s41598-017-18241-w.

29. C. J. Willer, Y. Li, and G. R. Abecasis, "METAL: Fast and Efficient Meta-Analysis of Genomewide Association Scans," *Bioinformatics* 26, no. 17 (2010): 2190–2191, https://doi.org/10.1093/bioinformatics/btq340.

30. B. Liu, M. J. Gloudemans, A. S. Rao, E. Ingelsson, and S. B. Montgomery, "Abundant Associations With Gene Expression Complicate GWAS Follow-Up," *Nature Genetics* 51, no. 5 (2019): 768–769, https://doi.org/10.1038/s41588-019-0404-0.

31. X. Zheng, J. Shen, C. Cox, et al., "HIBAG—HLA Genotype Imputation With Attribute Bagging," *Pharmacogenomics Journal* 14, no. 2 (2014): 192–200, https://doi.org/10.1038/tpj.2013.18.

32. X. Zheng, "Imputation-Based HLA Typing With SNPs in GWAS Studies," *Methods in Molecular Biology* 1802 (2018): 163–176, https://doi.org/10.1007/978-1-4939-8546-3_11.

33. S. Ashouri, S. S. Khor, Y. Hitomi, et al., "Genome-Wide Association Study for Chronic Hepatitis B Infection in the Thai Population," *Frontiers in Genetics* 13 (2022): 887121, https://doi.org/10.3389/fgene.2022.887121.

34. B. Elsworth, M. Lyon, T. Alexander, et al., "The MRC IEU OpenGWAS Data Infrastructure," bioRxiv, https://doi.org/10.1101/2020.08.10.244293.

35. G. Hemani, J. Zheng, B. Elsworth, et al., "The MR-Base Platform Supports Systematic Causal Inference Across the Human Phenome," *eLife* 7 (2018): e34408, https://doi.org/10.7554/eLife.34408.

36. B. K. Bulik-Sullivan, P. R. Loh, H. K. Finucane, et al., "LD Score Regression Distinguishes Confounding From Polygenicity in Genome-Wide Association Studies," *Nature Genetics* 47, no. 3 (2015): 291–295, https://doi.org/10.1038/ng.3211.

37. B. Bulik-Sullivan, H. K. Finucane, V. Anttila, et al., "An Atlas of Genetic Correlations Across Human Diseases and Traits," *Nature Genetics* 47, no. 11 (2015): 1236–1241, https://doi.org/10.1038/ng.3406.

38. G. Hemani, K. Tilling, and S. G. Davey, "Orienting the Causal Relationship Between Imprecisely Measured Traits Using GWAS Summary Data," *PLoS Genetics* 13, no. 11 (2017): e1007081, https://doi.org/10.1371/journal.pgen.1007081.

39. M. Kostara, V. Chondrou, A. Sgourou, K. Douros, and S. Tsabouri, "HLA Polymorphisms and Food Allergy Predisposition," *Journal of Pediatric Genetics* 9, no. 2 (2020): 77–86, https://doi.org/10.1055/s-0040-1708521.

40. J. Robinson, D. J. Barker, X. Georgiou, M. A. Cooper, P. Flicek, and S. G. E. Marsh, "IPD-IMGT/HLA Database," *Nucleic Acids Research* 48, no. D1 (2020): D948–D955, https://doi.org/10.1093/nar/gkz950.

41. W. H. Boehncke, C. Loeliger, P. Kuehnl, H. Kalbacher, B. O. Bohm, and H. Gall, "Identification of HLA-DR and -DQ Alleles Conferring Susceptibility to Pollen Allergy and Pollen Associated Food Allergy," *Clinical and Experimental Allergy* 28, no. 4 (1998): 434–441, https://doi.org/10.1046/j.1365-2222.1998.00246.x.

42. D. Hadley, W. Hagopian, E. Liu, et al., "HLA-DPB1*04:01 Protects Genetically Susceptible Children From Celiac Disease Autoimmunity in the TEDDY Study," *American Journal of Gastroenterology* 110, no. 6 (2015): 915–920, https://doi.org/10.1038/ajg.2015.150.

43. N. E. Eriksson, H. Formgren, and E. Svenonius, "Food Hypersensitivity in Patients With Pollen Allergy," *Allergy* 37, no. 6 (1982): 437–443, https://doi.org/10.1111/j.1398-9995.1982.tb02323.x.

44. S. H. Sicherer, C. M. Warren, C. Dant, R. S. Gupta, and K. C. Nadeau, "Food Allergy From Infancy Through Adulthood," *Journal of Allergy and Clinical Immunology. In Practice* 8, no. 6 (2020): 1854–1864, https://doi.org/10.1016/j.jaip.2020.02.010.

45. Y. Sun, E. T. G. Kersten, C. Qi, and G. H. Koppelman, "Asthma Susceptibility: Learning From Genetic Diversity," *Journal of Allergy and Clinical Immunology* 151, no. 4 (2023): 904–906, https://doi.org/10.1016/j.jaci.2023.01.034.

Supporting Information

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