Genomewide association study of peanut allergy reproduces association with amino acid polymorphisms in HLA-DRB1

Martino D.J., Ashley S., Koplin J., Ellis J., Saffery R., Dharmage S.C., Gurrin L., Matheson M.C., Kalb B., Marenholz I., Beyer K., Lee Y.A., Hong X., Wang X., Vukevic D., Motyer A., Leslie S., Allen K.J., Ferreira M.A.R.

This is the final version of the accepted manuscript, which was published in final edited form in:

Clinical and Experimental Allergy
2017 FEB; 47(2): 217-223
2016 NOV 24 (first published online)
doi: 10.1111/cea.12863

Publisher: Wiley-Blackwell

Received Date : 27-Jun-2016
Revised Date : 06-Nov-2016
Accepted Date : 07-Nov-2016
Article type : Original Article-Basic Mechanisms in Allergic Disease

Clinical and Experimental Allergy

November 7, 2016

Genome-wide association study of peanut allergy reproduces association with amino acid polymorphisms in *HLA-DRB1*

David J Martino [1 *], Sarah Ashley [1,4, *], Jennifer Koplin [1,2], Justine Ellis [1], Richard Saffery [1], Shyamali C Dharmage [2], Lyle Gurrin [2], Melanie C Matheson [2], Birgit Kalb [5,6,7], Ingo Marenholz [6,7], Kirsten Beyer [5], Young-Ae Lee [6,7], Xiumei Hong [8], Xiaobin Wang [8], Damjan Vukevic [1,3], Allan Motyer[1,3], Stephen Leslie [1,3], Katrina J. Allen [1] and Manuel A.R. Ferreira [9].

[1]*Murdoch Childrens Research Institute, The University of Melbourne Department of Paediatrics, The Royal Children's Hospital, Melbourne, Australia.*

[2] *The University of Melbourne, School of Population and Global Health, Australia*

[3] *The University of Melbourne, Centre for Systems Genomics, Schools of Mathematics and Statistics and Biosciences, Australia*

[4] *Hudson Institute of Medical Research, Australia*

[5] *Pediatric Pneumology and Immunology, Charité Universitätsmedizin Berlin, Germany*

[6] *Clinic for Pediatric Allergy, Experimental and Clinical Research Center of MDC and Charité, Germany*

[7] *Max-Delbrück- Center for Molecular Medicine (MDC), Germany*

[8] *Center on the Early Life Origins of Disease, Department of Population, Family and Reproductive Health, Johns Hopkins University Bloomberg School of Public Health*

[9] *QIMR Berghofer Medical Research Institute, Brisbane, Australia*

* *Denotes equal contribution*

Key words: Food Allergy, genetics, GWAS, hypersensitivity, peanut allergy,

Corresponding author: Professor Katrina Allen
Gastro and Food Allergy
Murdoch Childrens Research Institute
Flemington Road, Parkville
Melbourne, Australia
Phone: 61 3 9936 6585 Fax: 61 3 9348 1391
Email: Katie.Allen@rch.org.au

Abstract

Background: Genetic variants for IgE-mediated peanut allergy are yet to be fully characterized and to date only one genome-wide association study (GWAS) has been published.

Objective: To identify genetic variants associated with challenge proven peanut allergy.

Methods: We carried out a GWAS comparing 73 infants with challenge-proven IgE-mediated peanut allergy against 148 non-allergic infants (all ~ 1 year old). We tested a total of 3.8 million single nucleotide polymorphism (SNPs), as well as imputed HLA alleles and amino acids. Replication was assessed by *de novo* genotyping in a panel of additional 117 cases and 380 controls, and *in silico* testing in two independent GWAS cohorts.

Results: We identified 21 independent associations at $P \leq 5x10^{-5}$ but were unable to replicate these. The most significant HLA association was the previously reported amino acid variant located at position 71, within the peptide-binding groove of HLA-DRB1 ($P = 2x10^{-4}$). Our study therefore reproduced previous findings for the association between peanut allergy and HLA-DRB1 in this Australian population.

Conclusions & Clinical Relevance: Genetic determinants for challenge proven peanut allergy include alleles at the HLA-DRB1 locus.

Introduction

IgE-mediated peanut allergy is a common food allergy that frequently manifests in early life (1). It

affects 1-3% of infants world-wide (2), is often a persistent type of food allergy, more likely to be associated with severe reactions and potentially fatal anaphylaxis (3). Twin and family studies provide compelling evidence that genetic factors contribute to peanut allergy, with estimated heritability of around 80% (4),(5) for peanut allergy and 15 – 35% for IgE sensitization to different foods (6). Despite a significant heritability, only a small number of genetic studies of food allergy have been previously published. Candidate gene studies have identified a number of genetic associations in genes involved in immune function, or skin barrier integrity. For the latter, the filaggrin loss-of-function mutations have been consistently reported as a risk factor for peanut sensitization (7), and more recently peanut allergy (8,9), potentially by increasing trans-cutaneous sensitization to peanut in childhood (10-12). Variants in the Serine protease inhibitor Karzal type 5 (SPINK5) (13) gene have been associated with atopic eczema in Japanese children, and may increase the risk for food allergy. Polymorphisms in immunological genes associated with food allergy include Interleukin-10 (IL-10) (14), interleukin-13 (IL-13) (15), Cluster of differentiation 14 (CD14) (16), Forkhead box P3 (FOXP3) (17) and Signal transducer and activator 6 (STAT6) (18). To date, only one genome-wide association study (GWAS) has been conducted for peanut allergy (19), although case-control status in that study was not defined using challenge-proven outcomes. In that GWAS, the single nucleotide polymorphisms (SNP) rs7192 and rs9275596, located in the major histocompatibility locus (MHC), and the amino acid polymorphism at position 71 of *HLA-DRB1* were associated with peanut allergy. HLA class II variants have also been associated with peanut allergy in candidate gene studies (20-22), although this has not always been a consistent finding (23). Thus, most genetic determinants of peanut allergy remain to be identified. The aim of this study was to carry out a GWAS on a well-characterized population of challenge-proven paediatric peanut allergic cases and controls to identify SNPs with a large effect on disease risk using an agnostic genome-wide approach, and to perform a comprehensive assessment of the MHC.

Methods

Statement of ethics

This study was approved by the Office for Children HREC (ref. no. CDF/07/492), Department of Human Services HREC (ref. no. 10/07), Royal Children's Hospital HREC (ref. no.27047) and the QIMR Berghofer Medical Research Institute (ref. no. P710).

GWAS study population

All participants included in the GWAS are part of the HealthNuts study, which utilized a population-based sampling frame and diagnostic food challenges. All participants included in the GWAS are part of the HealthNuts population-based cohort (24). Briefly, 5,300 infants aged 11-15 months inclusive were recruited from council-run immunization sessions across Melbourne, Australia, between June 2008 and August 2011. Skin prick testing to egg and peanut was carried out at immunization sessions and those testing positive (~1000) were invited for allergy assessment at the Royal Children's Hospital Melbourne (24). Clinical allergy assessment was carried out by a registered nurse, under hospital supervision using skin prick testing and open peanut challenges using predetermined objective criteria (25)). Non-allergic controls were skin prick test negative to a panel of common food allergens (egg white, peanut, sesame, shrimp or cow's milk, cashew, almond, hazelnut, soy and wheat) and safely tolerated peanut during oral peanut challenge. Blood samples were collected from participants after peanut challenge with guardian's written consent. After quality control (summarized below and described in detail in the **Supplementary Methods**), genotyping data were available for 73 cases and 148 controls. Cases were defined as having a skin prick test greater or equal to 2 mm above the negative control, or specific IgE > 0.35 kU/L to peanut on the day of challenge, with evidence of clinical reactivity during oral peanut challenge. Clinical characteristics of the 221 genotyped infants are given in **Supplementary Table 1.**

Peanut food challenges

Peanut oral food challenges (OFC) were conducted at the Royal Children's Hospital under medical supervision. Positive oral food challenge was defined as 1 or more than 1 of the following: 3 or more concurrent noncontact urticaria lasting 5 minutes or more; perioral/periorbital angioedema; vomiting; or circulatory or respiratory compromise within 2 hours of ingestion of a challenge dose. Infants underwent OFC irrespective of their history of ingestion or SPT wheal size, unless there was a parent report of a clear history of an immediate reaction to peanut within the past 2 months (as per the HealthNuts challenge criteria (Koplin 2012)). Three infants were deemed peanut allergic based on the latter definition.

Genotyping and quality control

DNA was extracted from peripheral blood using the Qiagen FlexiGene protocol and stored at -20

degrees prior to genotyping. Genotyping was carried out by the Australian Genome Research Facility using the Illumina HumanOmni 2.5-8 SNP array. DNA was submitted as a single batch comprising 9 plates with subjects randomised across these, and raw data were processed and analysed together. Only unrelated individuals with call rate > 95% were included in the analysis. SNPs were excluded if MAF <1%, call rate <95% and/or with Hardy Weinberg Equilibrium (HWE) test P-value < 10^{-6}, and/or had significant (P<0.001) MAF differences when compared against the 1000 Genomes Project samples (Europeans only), leaving 389,427 high quality directly genotyped variants. Ancestry was inferred from visual inspection of results of a multidimensional 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) (27). Full details of quality control are provided in **Supplementary Methods.**

Imputation of variants from the 1000 Genomes Project and association analyses

Genotype data for 389,427 directly genotyped variants from 364 individuals was 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. The association between case-control status and allelic dosage was tested using SNPTEST (29) with sex and ancestry strata included as covariates. A P-value of $5x10^{-8}$ was considered the threshold for genome-wide significance, and variants achieving an a priori determined P-value ≤ 10^{-5} were selected for subsequent replication studies. 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 (|beta| >3 and/or SE>1.5), or MAF < 1%, results were available for 3,814,967 variants.

Imputation of HLA alleles and amino acid variants

HLA alleles for 8 loci (*HLA-A*, *HLA-B*, *HLA-C*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DRB1*) were imputed from genotyped SNPs using the HLA*IMP:02 method (30). These imputed alleles were then mapped to amino acid alleles and SNP alleles within the HLA genes, using the same method as was done to build the reference panel for SNP2HLA (31). Association testing for the HLA

variants was carried out using logistic regression adjusted for sex and ancestry strata using PLINK (v1.07) (27). Conditional analysis was carried out using forward step-wise logistic regression adjusted for sex and ancestry strata.

Replication study – de novo genotyping

The replication study was conducted in two phases: *de novo* genotyping in a replication sample, and in silico replication in two independent GWAS studies of paediatric food allergy (**Supplementary Figure 1**). A replication panel was assembled to include 117 peanut allergic cases and 380 non-allergic controls from HealthNuts, the Barwon Infant Study (BIS) (32), the Peanut Allergen Threshold study (PAT) (33) and the Peanut and Probiotic Oral Immunotherapy Trial (PPOIT) (34), all Melbourne-based paediatric allergy studies with identical phenotyping criteria including diagnostic peanut challenges using the same protocol. Clinical characteristics are reported in **Supplementary Table 2.** Case and control definitions were synonymous with the HealthNuts discovery samples. DNA was extracted from cryopreserved peripheral blood using Qiagen AllPrep DNA/RNA mini method at the Murdoch Childrens Research Institute laboratories and stored at -30 degrees prior to genotyping. Targeted genotyping was carried out as a single batch using the MassARAY platform and IPLEX chemistry (Agena Bioscience) for 19 selected SNPs and a panel of ancestry informative markers (35). Full details are provided in the S**upplementary Methods.**

In silico replication population (American and German studies)

In silico replication was carried out using data from 316 peanut allergy cases, 144 non-allergic controls, and 1,737 controls of uncertain phenotypes (589 children and 1,148 parents) from the Chicago Food Allergy study (19), tested with MQLS (36). Cases were defined as having a history of a clinical allergic reaction on ingestion of peanut and evidence of sensitization to peanut defined as having detectable peanut-specific IgE (>= 0.1 kUl-1 $^{-1}$) and/or positive skin prick test (wheal diameter >3mm) to peanut. Non-allergic non-sensitized controls had no evidence of clinical reactivity by food challenge or sensitization to a panel of allergens tested.

In silico replication was also assessed using data from a Caucasian population of 205 peanut allergy cases and 2,387 controls from the German Understanding Food Allergy (UFA) study. Peanut allergic cases were defined by double-blind placebo-controlled food challenge or a history of a severe allergic reaction to peanuts plus specific sensitization to peanut protein (>0.35 kUl-1 $^{-1}$). Controls were

unrelated individuals from the German population based Heinz Nixdorf Recall Study (37). Full details in S**upplementary Methods.**

RESULTS

GWAS of challenge-proven peanut allergy

Demographics and clinical characteristics of the 221 participants included in the discovery GWAS are summarized in **Supplementary Table 1**. After quality control and imputation of unmeasured variants, a total of 3,814,967 SNPs were tested for association with case-control status. The genomic inflation factor of this analysis was 0.998, consistent with no significant effects of technical artefacts or unaccounted population substructure on the association results. A single variant (rs10018666, OR = 5.9, $P = 4x10^{-8}$) located in the *SLC2A9* gene reached genome-wide significance for association with peanut allergy (**Figure 1**). Multiple variants in linkage disequilibrium (LD) with rs10018666 supported this association, including two variants (rs13129697, imputed, $P = 10^{-5}$; rs10939650, genotyped, $P = 2x10^{-6}$) previously reported to associate with variation in the expression of *SLC2A9* in blood cells (**Supplementary Table 3**). This variant, and an additional 20 that were associated with peanut allergy at $P \leq 5x10^{-5}$ (**Supplementary Table 4**), were selected for a replication analysis and 19 of the 21 SNPs (designated in **Supplementary Table 4**) were incorporated into a multiplexed genotyping assay for the replication stage. For most of these variants, the direction of effect was the same across the three ancestry strata analysed (**Supplementary Table 5**).

Replication of top associations in three independent studies

We directly genotyped the 19 SNPs and 41 ancestry markers in an additional 117 peanut allergy cases and 380 non-allergic non-sensitized controls from Melbourne (HealthNuts, Barwon, Peanut Allergen Threshold, PPOIT studies), who were unrelated to the participants included in the discovery GWAS. Four variants were found to have suggestive associations with peanut allergy (unadjusted $P < 0.05$), with similar effects between the two ancestry populations (**Supplementary Table 6**), but these associations did not survive a correction for multiple testing. To increase power of the replication study, we extracted results for the top 21 SNPs from two additional studies of food allergy – the German UFA (205 peanut allergy cases and 2,387 controls, **Supplementary Table 7**) and Chicago (316 peanut allergy cases, 144 controls, **Supplementary Table 8**) studies – but found no overall consistent evidence (same direction of effect) for association with any of the 21 variants. Therefore, we

conclude that the top 21 SNPs identified in the discovery GWAS likely represent false-positive associations.

Analysis of HLA polymorphisms

We studied in greater detail the association between HLA polymorphisms and food allergy by imputing classical 102 2-digit alleles, 187 4-digit alleles, as well as 2,205 polymorphic amino acid loci, in addition to 2,412 SNPs across the MHC. The most significant associations were observed with amino acid polymorphisms at position 37 (OR = 0.3, P = 9.8 $x10^{-5}$) and 71 (OR = 0.34, P = 1.5 x 10^{-4}) of the *HLA-DRB1* gene (**Table 1**). The direction of effect for these associations was the same as that reported for position 71 previously in the Hong et al GWAS (OR= 0.38, P= 9.8 $x10^{-11}$) (19). Conditioning on position 37 in our study substantially reduced the effect of position 71 (OR = 0.4, P = 0.03), whilst condition on position 71 also reduced the effect of 37 (OR = 0.4, P = 0.02; **Supplementary Figure 2)**. These two variants are in moderate linkage disequilibrium (r^2=0.5) and we could not statistically distinguish the effect of one over the other, suggesting that they are unlikely to be independent. All of the remaining associations were greatly reduced after conditioning on amino acid positions 37 or 71, indicating that the associations at HLA were primarily driven by these polymorphisms.

Discussion

In this study, we report the second GWAS of IgE-mediated peanut allergy, but the first using challenge-proven outcomes. With the sample size for this study, only variants with a very large effect on disease risk would be detected at the genome-wide significance level. In total, we identified 21 SNPs with a suggestive association with peanut allergy in the discovery GWAS but were unable to replicate these in three replication studies, suggesting that they were likely to be false-positive associations. On the other hand, a comprehensive analysis of HLA polymorphism identified an association with amino acid positions 37 and 71 of the *HLA-DRB1* gene which supports the association originally reported with this polymorphism in a US population of children with peanut allergy (19). Although the p-value did not reach genome-wide significance, the association was clearly reproduced in our data set as evidenced by the consistent direction and effect size, and the fact that it was the strongest signal at the MHC. We were unable to assess the HLA associations for rs7192 and rs9275596 also reported by Hong (19) since these SNPs did not pass QC in our data set. Collectively, our data support a role for polymorphisms in the HLA region on the risk of peanut allergy, specifically those

relating to *HLA-DRB1* amino acids. *HLA-DR* molecules comprise the major HLA class II isotypes present on the surface of antigen presenting cells, and *HLA-DRB1* is a highly polymorphic locus. The binding groove of *DRB1* consists of pockets that interact with peptide side chain residues. Polymorphisms in the binding groove can alter the overall binding affinity and specificity of *HLA-DRB1* for peptide ligands. The outcome of these effects can play a central role in T-helper cell activation and influence the quality of the immune response. As outlined in Hong et al, and originally reported by Sturniolo et al (38), polymorphic amino acid residue 71 affects the binding specificity of pocket 4 of the peptide binding groove. It is therefore biologically plausible that this amino acid variant plays an important role in the development of antigen-specific immune responses and the development of allergy. Future studies employing targeted genotyping approaches are now needed to properly resolve the *DRB1* gene association and complex haplotype structure in order to clarify functional and clinical significance. Given that SNP arrays are limited by LD structure they indicate a region of interest, and targeted genotyping with functional studies would be required in future to assess causality.

The strengths of this study include the well-characterized cohort in which clinical phenotypes were determined by oral food challenge, and clinical assessment protocols were harmonised across the discovery and replication populations. Despite this we cannot exclude some heterogeneity between our discovery and replication populations as a potential caveat. The discovery population were young infants (age 12 months) whilst the replication studies (*de novo* and *in silico*) consisted of infants from a range of ages (mean age of peanut allergics in UFA was 3.5 years and 7 years in Chicago study). This was reflected to some extent as variations in peanut specific IgE levels, with the older children exhibiting higher antibody titres (HealthNuts: mean=13.5, SD=19.6 kU/L, UFA: mean=57.5, SD=95.9 kU/L and Chicago: mean 65.6, SD 87.4 kU/L), and mean wheal diameters (HealthNuts: mean=8.7 mm, de novo replication children ranged from 5.71 to 18.8mm across the studies **Supplementary Table 2**). Subjects from the *in silico* analysis were entirely Caucasian, whilst the discovery and de novo genotyping populations consisted of Caucasian and Asian ancestries. Other caveats include the limited sample size, which precluded the ability to detect all but substantial effect sizes. Moreover, our analysis was restricted to common variants, which is typical for most GWAS, and alternative study designs will be required to address the role of rare variants.

In summary, this study supports a role for *HLA-DRB1* alleles as genetic risk factors of IgE-mediated peanut allergy.

Acknowledgements: The genome wide 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), Career Development Fellowship ID 1053756 (S.L.) Australian Research Council Future Fellowship ID FT120100253 (JE), Early Career Fellowships 1072752, 1054592 (DM, JE); and by the Victorian Life Sciences Computation Initiative (VLSCI) grant number VR0240 on its Peak Computing Facility at the University of Melbourne, an initiative of the Victorian Government, Australia (S.L.). Research at the Murdoch Childrens Research Institute was supported by the Victorian Government's Operational Infrastructure Support Program. S.L. is a partner in Peptide Groove LLP. The UFA study was supported by a Clinical Exchange Program fellowship of the Experimental and Clinical Research Center of Max-Delbrück-Center and Charité Medical School. We thank Markus M Noethen (Institute of Human Genetics, University of Bonn, Germany) and Sven Cichon (Division of Medical Genetics, University Hospital Basel and Department of Biomedicine, University of Basel) for providing genotype data of the Heinz Nixdorf RECALL (HNR) study.

References

1. Sicherer SH. Clinical update on peanut allergy. ANAI. 2002 Apr;88(4):350–61–quiz361–2–394.

2. Osborne NJ, Koplin JJ, Martin PE, Gurrin LC, Lowe AJ, Matheson MC, et al. Prevalence of challenge-proven IgE-mediated food allergy using population-based sampling and predetermined challenge criteria in infants. J Allergy Clin Immunol. 2011 Mar;127(3):668–76.e1–2.

3. Finkelman FD. Peanut allergy and anaphylaxis. Curr Opin Immunol. 2010 Dec;22(6):783–8.

4. Sicherer SH, Furlong TJ, Maes HH, Desnick RJ, Sampson HA, Gelb BD. Genetics of peanut allergy: a twin study. Journal of allergy and clinical immunology. 2000 Jul;106(1 Pt 1):53–6.

5. Hourihane JO, Dean TP, Warner JO. Peanut allergy in relation to heredity, maternal diet, and other atopic diseases: results of a questionnaire survey, skin prick testing, and food challenges. BMJ. BMJ Publishing Group Ltd; 1996 Aug 31;313(7056):518–21.

6. Tsai HJ, Kumar R, Pongracic J, Liu X, Story R, Yu Y, et al. Familial aggregation of food allergy and sensitization to food allergens: a family-based study. Clin Exp Allergy. 2009 Jan;39(1):101–9.

7. Tan H-TT, Ellis JA, Koplin JJ, Matheson MC, Gurrin LC, Lowe AJ, et al. Filaggrin loss-of-function mutations do not predict food allergy over and above the risk of food sensitization among infants. J Allergy Clin Immunol. 2012 Nov;130(5):1211–3.

8. Brown SJ, Asai Y, Cordell HJ, Campbell LE, Zhao Y, Liao H, et al. Loss-of-function variants in the filaggrin gene are a significant risk factor for peanut allergy. Journal of allergy and clinical immunology. 2011;127(3):661–7.

9. Venkataraman D, Soto-Ramírez N, Kurukulaaratchy RJ, Holloway JW, Karmaus W, Ewart SL, et al. Filaggrin loss-of-function mutations are associated with food allergy in childhood and adolescence. J Allergy Clin Immunol. 2014 Oct;134(4):876–882.e4.

10. Thyssen JP, Tang L, Husemoen LLN, Stender S, Szecsi PB, Menné T, et al. Filaggrin gene mutations are not associated with food and aeroallergen sensitization without concomitant atopic dermatitis in adults. J Allergy Clin Immunol. 2015 Feb 10;135(5):1375–1378.e1.

11. Brough HA, Liu AH, Sicherer S, Makinson K, Douiri A, Brown SJ, et al. Atopic dermatitis increases the effect of exposure to peanut antigen in dust on peanut sensitization and likely peanut allergy. J Allergy Clin Immunol. 2014 Nov 18.

12. Brough HA, Simpson A, Makinson K, Hankinson J, Brown S, Douiri A, et al. Peanut allergy: effect of environmental peanut exposure in children with filaggrin loss-of-function mutations. J Allergy Clin Immunol. 2014 Oct;134(4):867–875.e1.

13. Kusunoki T, Okafuji I, Yoshioka T, Saito M, Nishikomori R, Heike T, et al. SPINK5 polymorphism is associated with disease severity and food allergy in children with atopic dermatitis. Journal of allergy and clinical immunology. 2005 Mar;115(3):636–8.

14. Campos Alberto EJ, Shimojo N, Suzuki Y, Mashimo Y, Arima T, Matsuura T, et al. IL-10 gene polymorphism, but not TGF-beta1 gene polymorphisms, is associated with food allergy in a Japanese population. Pediatr Allergy Immunol. 2008 Dec;19(8):716–21.

15. Liu X, Beaty TH, Deindl P, Huang S-K, Lau S, Sommerfeld C, et al. Associations between specific serum IgE response and 6 variants within the genes IL4, IL13, and IL4RA in German children: the German Multicenter Atopy Study. Journal of allergy and clinical immunology. 2004 Mar;113(3):489–95.

16. Woo JG, Assa'ad A, Heizer AB, Bernstein JA, Hershey GKK. The -159 C-->T polymorphism of CD14 is associated with nonatopic asthma and food allergy. Journal of allergy and clinical immunology. 2003 Aug;112(2):438–44.

17. Torgerson TR, Linane A, Moes N, Anover S, Mateo V, Rieux-Laucat F, et al. Severe food allergy as a variant of IPEX syndrome caused by a deletion in a noncoding region of the FOXP3 gene. Gastroenterology. 2007 May;132(5):1705–17.

18. Amoli MM, Hand S, Hajeer AH, Jones KP, Rolf S, Sting C, et al. Polymorphism in the STAT6 gene encodes risk for nut allergy. Genes Immun. 2002 Jun;3(4):220–4.

19. Hong X, Hao K, Ladd-Acosta C, Hansen KD, Tsai H-J, Liu X, et al. Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children. Nat Commun. 2015;6:6304.

20. Howell WM, Turner SJ, Hourihane JO, Dean TP, Warner JO. HLA class II DRB1, DQB1 and DPB1 genotypic associations with peanut allergy: evidence from a family-based and case-control study. Clin Exp Allergy. 1998 Feb;28(2):156–62.

21. Hand S, Darke C, Thompson J, Stingl C, Rolf S, Jones KP, et al. Human leucocyte antigen polymorphisms in nut-allergic patients in South Wales. Clin Exp Allergy. Blackwell Science Ltd; 2004 May;34(5):720–4.

22. Madore A-M, Vaillancourt VT, Asai Y, Alizadehfar R, Ben-Shoshan M, Michel DL, et al. HLA-DQB1*02 and DQB1*06:03P are associated with peanut allergy. Eur J Hum Genet. 2013 Feb 26.

23. Shreffler WG, Wanich N, Moloney M, Nowak-Wegrzyn A, Sampson HA. Association of allergen-specific regulatory T cells with the onset of clinical tolerance to milk protein. Journal of allergy and clinical immunology. 2009 Jan;123(1):43–7.

24. Koplin JJ, Wake M, Dharmage SC, Matheson M, Tang ML, Gurrin LC, et al. Cohort Profile of the HealthNuts study: Population prevalence and environmental/genetic predictors of food allergy. Int J Epidemiol. 2015 Jan 21.

25. Koplin JJ, Tang MLK, Martin PE, Osborne NJ, Lowe AJ, Ponsonby A-L, et al. Predetermined challenge eligibility and cessation criteria for oral food challenges in the HealthNuts population-based study of infants. J Allergy Clin Immunol. 2012 Apr;129(4):1145–7.

26. Consortium T1GP. An integrated map of genetic variation from 1,092 human genomes. Nature. Nature Publishing Group; 2012 Nov 1;491(7422):56–65.

27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.

28. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. Schork NJ, editor. PLoS Genet. Public Library of Science; 2009 Jun;5(6):e1000529.

29. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. Nature Publishing Group; 2010 Jul 1;11(7):499–511.

30. Dilthey A, Leslie S, Moutsianas L, Shen J, Cox C, Nelson MR, et al. Multi-population classical HLA type imputation. Browning S, editor. PLoS Comput Biol. 2013;9(2):e1002877.

31. Jia X, Han B, Onengut-Gumuscu S, Chen W-M, Concannon PJ, Rich SS, et al. Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. Tang J, editor. PLoS ONE. Public Library of Science; 2013 Jun 6;8(6):e64683.

32. Vuillermin P, Saffery R, Allen KJ, Carlin JB, Tang ML, Ranganathan S, et al. Cohort Profile: The Barwon Infant Study. Int J Epidemiol. 2015 Aug;44(4):1148–60.

33. Zurzolo GA, Allen KJ, Taylor SL. Peanut Allergen Threshold Study (PATS): validation of eliciting doses using a novel single-dose challenge protocol. Allergy Asthma Clin …. 2013.

34. Tang MLK, Ponsonby A-L, Orsini F, Tey D, Robinson M, Su EL, et al. Administration of a probiotic with peanut oral immunotherapy: A randomized trial. J Allergy Clin Immunol. 2015 Mar;135(3):737–8.

35. Bousman CA, Yung AR, Pantelis C, Ellis JA. Effects of NRG1 and DAOA genetic variation on transition to psychosis in individuals at ultra-high risk for psychosis. Translational …. 2013.

36. Thornton T, McPeek MS. Case-control association testing with related individuals: a more powerful quasi-likelihood score test. The American Journal of Human Genetics. 2007.

37. Schmermund A, Möhlenkamp S, Stang A, Grönemeyer D, Seibel R, Hirche H, et al. Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. Am Heart J. 2002 Aug;144(2):212–8.

38. Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat Biotechnol. 1999 Jun;17(6):555–61.

Figures and Tables

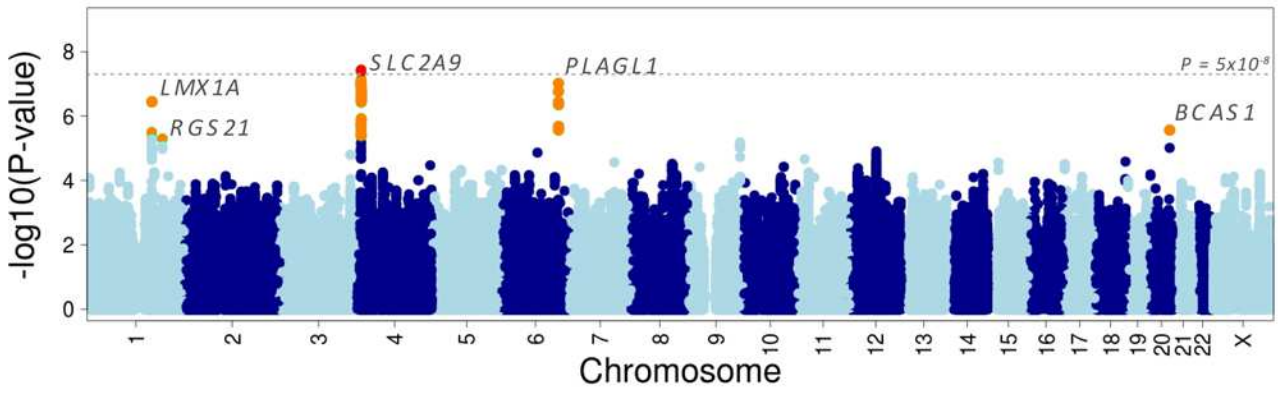


Figure 1. Summary of association results for 3,814,967 SNPs tested in 73 peanut food allergy cases and 148 controls. Suggestive associations ($P <= 10^{-5}$) are shown in orange and genome-wide significant associations ($P = 5x\ 10^{-8}$) in red.

Table 1 Top 5 HLA associations stratified by variant class ranked according to P-value

Amino acids and imputed SNP	CHR	BP	OR	SE	L95	U95	STAT	P
AA_DRB1_37_32660037_FY	6	32552059	0.3	0.31	0.16	0.55	-3.9	9.80E-05
AA_DRB1_71_32659935_R	6	32551957	0.34	0.29	0.19	0.59	-3.79	1.50E-04
SNP_DRB1_32659935_C	6	32551957	0.34	0.29	0.19	0.59	-3.79	1.50E-04
AA_DRB1_37_32660037_NS	6	32552059	0.34	0.29	0.19	0.6	-3.7	2.16E-04
AA_DRB1_47_32660007	6	32552029	0.35	0.29	0.2	0.62	-3.64	2.76E-04
Genotyped SNPs								
rs154975	6	32900601	2.6	0.25	1.59	4.24	3.83	1.30E-04
rs2858305	6	32670464	2.22	0.22	1.44	3.4	3.64	2.72E-04
rs2856717	6	32670308	2.19	0.22	1.43	3.36	3.59	3.28E-04
rs2858332	6	32681161	2.28	0.23	1.45	3.58	3.58	3.48E-04
rs2858309	6	32668713	2.17	0.22	1.42	3.33	3.57	3.61E-04
2-digit HLA alleles								
HLA_DQB1_06	6	32631061	2.26	0.25	1.38	3.69	3.24	1.20E-03
HLA_DPB1_03	6	33049368	3.23	0.41	1.45	7.21	2.86	4.25E-03
HLA_DQA1_02	6	32608306	0.36	0.42	0.16	0.82	-2.44	1.48E-02
HLA_DRB1_15	6	32552064	2.15	0.32	1.15	3.99	2.41	1.60E-02
HLA_DRB1_07	6	32552064	0.34	0.46	0.14	0.82	-2.39	1.68E-02
4-digit HLA alleles								
HLA_DPB1_0301	6	33049368	3.23	0.41	1.45	7.21	2.86	4.25E-03

HLA_DQA1_0201	6	32608306	0.36	0.42	0.16	0.82	-2.44	1.48E-02
HLA_DRB1_0701	6	32552064	0.34	0.46	0.14	0.82	-2.39	1.68E-02
HLA_DQB1_0602	6	32631061	2.11	0.31	1.14	3.91	2.38	1.72E-02
HLA_DQA1_0102	6	32608306	1.81	0.25	1.11	2.95	2.38	1.75E-02

CHR = Chromosome, BP = base position, OR= odds ratio, SE = standard error, L95= lower 95% confidence interval, U95 = upper 95% confidence interval, STAT = value of the test statistic, P= P-value.

Clinical and Experimental Allergy

November 7, 2016

Genome-wide association study of peanut allergy reproduces association with amino acid polymorphisms in *HLA-DRB1*

David J Martino [1 *], Sarah Ashley [1,4, *], Jennifer Koplin [1,2], Justine Ellis [1], Richard Saffery [1], Shyamali C Dharmage [2], Lyle Gurrin [2], Melanie C Matheson [2], Birgit Kalb [5,6,7], Ingo Marenholz [6,7], Kirsten Beyer [5], Young-Ae Lee [6,7], Xiumei Hong [8], Xiaobin Wang [8], Damjan Vukcevic [1,3], Allan Motyer[1], Stephen Leslie [1,3], Katrina J. Allen [1] and Manuel A.R. Ferreira [9].

SUPPLEMENTARY FIGURES AND TABLES

Supplementary Table 1. Clinical characteristics of 221 infants included in the GWAS.

	Non-atopic controls (NA) N=148	Peanut allergic (PA) N=73
Infant demographics		
Age in months at recruitment (mean, SD)	12.6 (0.67)	12.8 (0.80)
Gender (% male)	43.2%	69.4%
Genetically inferred ancestry		
Asian	6.8%	19.2%
European	80.4%	69.9%
Mixed European/Asian	12.8%	11.0%
Infant clinical characteristics		
Peanut SPT wheal size mm (mean, SD)	0.037 (0.23)	8.7 (3.7)
Peanut positive SPT	0%	
Peanut specific IgE kU/L (mean, SD)	0.07 (0.29)	13.5 (19.6)
History of eczema*	43.7%	88.4%
Other food allergies		
Egg allergy	0%	73.9%
Sesame allergy	0%	12.7%
Family characteristics**		
Any siblings	50.70%	44.40%
Asthma		
Maternal asthma	16.20%	23.60%
Paternal asthma	16.90%	13.90%
Sibling asthma	8.80%	9.70%
Hay fever		
Maternal hay fever	37.80%	43.10%
Paternal hay fever	33.10%	36.10%
Sibling hay fever	5.40%	2.80%
Eczema		
Maternal eczema	17.60%	25.00%
Paternal eczema	10.80%	11.10%
Sibling eczema	14.20%	18.10%
Food allergy		
Maternal food allergy	7.40%	0%
Paternal food allergy	10.80%	2.80%
Sibling food allergy	6.80%	2.80%

*History of eczema diagnosis or itchy rash treated with topical steroids or nurse-observed eczema at recruitment

** Parent reported

Supplementary Table 2. Clinical characteristics of 497 infants included in the *de novo* genotyping replication stage.

	HealthNuts (HN)		PPOIT	PAT	BIS	
	Non-atopic controls (NA) N=63	Peanut allergic (PA) N=26	Peanut allergic (PA) N=26	Peanut allergic (PA) N=59	Non-atopic controls (NA) N=317	Peanut allergic (PA) N=6
Infant demographics						
Gender (% male)	34.69%	57.69%	67%	51%	50.16%	100%
Genetically inferred ancestry						
European	80.95%	61.54%	66.67%	77.97%	91.17%	66.67
Mixed European/Asian	19.05%	38.46%	33.33%	22.03%	8.83%	33.33%
Infant clinical characteristics						
Peanut SPT wheal size (mean, SD)	0.25 (1.03)	7.87 (4.87)	18.8 (5.3)	15.2 (6.4)	0.62 (1.4)	5.71 (2.73)
Other food allergies						
Egg allergy	0.00%	68.00%	42.86%	41.86%	0.00%	61.54%
Sesame allergy	0.00%	8.70%	-/-	11.72%	0.00%	0.00%

-/- Data was not collected upon recruitment

PPOIT: Peanut oral immunotherapy trial, PAT: Peanut allergen threshold study, BIS: Barwon infant study

Supplementary Table 3. Association between peanut allergy SNPs ($P < 5 \times 10^{-5}$) and variation in gene expression levels in published eQTL studies.

Chr	BP	Food Allergy SNP [A]	Food Allergy Risk allele	eQTL SNP	r2 with [A]	eQTL effect allele	haplotype †	eQTL effect (Beta)*	Concordance**	eQTL Pvalue	eQTL target Gene	eQTL Tissue	eQTL Reference
4	10004805	rs10018666	T	rs13129697	0.57	G	CG/TT	-6.4	Y	1.50E-10	SLC2A9	Blood	Westra
				rs10939650	0.74	NA	CC/TT	NA	NA	7.70E-04	SLC2A9	Monocytes	Fairfax
6	144315219	rs6928827	G	rs9484836	0.6	A	AA/GG	3.5	N	5.00E-04	STX11	Blood	Westra
9	132008809	rs4240433	T	rs7871818	0.99	NA	CT/TC	NA	NA	3.20E-10	IER5L	LCLs	Lappalainen
				rs927927	0.99	T	CT/TC	-5.4	Y	6.30E-08	CRAT	Blood	Westra
				rs2104727	1	G	CG/TA	-2.9	Y	3.30E-03	PPP2R4	Blood	Westra
12	68603179	rs7300806	G	rs1861492	0.54	G	AG/GA	-4.6	Y	3.60E-06	MDM1	Blood	Westra
				rs7132188	0.66	NA	AC/GA	NA	NA	1.00E-05	MDM1	Monocytes	Fairfax
				rs10748099	0.56	T	AT/GC	3.4	N	6.40E-04	IFNG	Blood	Westra
7	115842729	rs73220497	T	rs10250473	0.66	C	TC/GT	-5.7	N	1.00E-08	TES	Blood	Westra
				rs1881288	0.64	C	TC/GA	-4	N	7.60E-05	CAV2	Blood	Westra
8	104883146	rs16870788	G	rs10955332	0.93	NA	GG/AA	NA	NA	5.40E-04	UBR5	Monocytes	Fairfax
12	21594028	rs7131777	T	rs2192176	0.99	NA	CC/TA	NA	NA	8.20E-17	PYROXD1	LCLs	Lappalainen
				rs10770810	0.73	G	CT/TG	-0.877	N	1.20E-06	PYROXD1	Normal skin	Ding
				rs12423381	0.65	A	CA/TT	0.82	N	1.40E-06	IAPP	Uninvolved skin	Ding
				rs10841832	0.72	A	CG/TA	-0.82	N	2.30E-06	IAPP	Normal skin	Ding
				rs3782660	0.98	C	CT/TC	3.2	Y	1.50E-03	RECQL	Blood	Westra

				rs3782660	0.98	C	CT/TC	3	Y	2.40E-03	GOLT1B	Blood	Westra
5	179411289	rs864481	A	rs864481	1	A	Same SNP	5.8	Y	6.80E-09	RNF130	Blood	Westra

† *shows phasing of FA allele and eQTL*

* *shows the direction of effect for the association between the eQTL allele and gene expression levels*
** *indicates whether the FA risk allele is in phase with the eQTL allele that increases gene expression (NA indicates missing data)*

Westra H-J, Peters MJ, Esko T, et al (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 45:1238–1243. doi: 10.1038/ng.2756

Fairfax BP, Makino S, Radhakrishnan J, et al (2012) Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet 44:502–510. doi: 10.1038/ng.2205

Lappalainen T, Sammeth M, Friedländer MR, et al (2013) Transcriptome and genome sequencing uncovers functional variation in humans. Nature 501:506–511. doi: 10.1038/nature12531

Ding J, Gudjonsson JE, Liang L, et al (2010) Gene expression in skin and lymphoblastoid cells: Refined statistical method reveals extensive overlap in cis-eQTL signals. Am J Hum Genet 87:779–789. doi: 10.1016/j.ajhg.201

Supplementary Table 4. Top 21 independent SNP associations with challenge-proven peanut food allergy ($P<5x10^{-5}$).

Chr	Bp position	SNP	A1	Freq	OR	SE	*P*-value	Nearest gene, kb distance	Candidate target gene(s) based on published eQTL studies
4	10004805	rs10018666	T	0.76	5.86	0.38	$4x10^{-8}$	*SLC2A9*,*	*SLC2A9*
6	144315219	rs6928827	G	0.88	13.98	0.66	$1x10^{-7}$	*PLAGL1*,14*	*STX11*
1	165082110	rs6686894	G	0.11	0.06	0.81	$4x10^{-7}$	*LMX1A*,89	-
20	52489424	rs11700330 #	A	0.22	0.23	0.35	$3x10^{-6}$	*BCAS1*,71	-
1	192351266	rs12142904	G	0.24	3.51	0.29	$5x10^{-6}$	*RGS21*,15	-
9	132008809	rs4240433#	T	0.77	3.61	0.32	$7x10^{-6}$	*IER5L*,68	*IER5L, CRAT, PRR2B4*
12	68603179	rs7300806	G	0.80	0.28	0.31	$1x10^{-5}$	*IL26*,8*	*MDM1, IFNG*
6	90476452	rs9362681#	G	0.31	2.83	0.25	$1x10^{-5}$	*MDN1*,53*	-
3	180686365	rs6763069	A	0.63	0.38	0.24	$2x10^{-5}$	*FXR1*,9*+	-
11	20123190	rs2439871#	C	0.59	0.38	0.24	$1x10^{-5}$	*NAV2*,20*	-
18	76652861	rs73971133	A	0.06	0.07	0.85	$3x10^{-5}$	*SALL3*,87	-
7	115842729	rs73220497	T	0.08	0.06	0.98	$3x10^{-5}$	*TES*,8	*TES, CAV2*
15	25840403	rs17555239	T	0.43	2.58	0.24	$3x10^{-5}$	*ATP10A*,83	-
8	104883146	rs16870788	G	0.16	3.58	0.32	$3x10^{-5}$	*RIMS2*,9	*UBR5*
17	2545473	rs8077351	G	0.07	0.05	1.00	$3x10^{-5}$	*PAFAH1B1*,43*	-
4	186704292	rs57144668#	C	0.65	0.37	0.25	$3x10^{-5}$	*SORBS2*,174*	-
10	102476167	rs6584390#	T	0.76	3.56	0.33	$4x10^{-5}$	*PAX2*,29	-
9	28757900	rs10812871#	C	0.53	0.38	0.25	$4x10^{-5}$	*LINGO2*,39	-
12	21594028	rs7131777	T	0.50	2.55	0.24	$4x10^{-5}$	*PYROXD1*,3*+	*PYROXD1,IAPP,RECQL,COLT1B*
5	179411289	rs864481#	A	0.23	2.91	0.27	$5x10^{-5}$	*RNF130*,29*	*RNF130*
5	75659270	rs10474468#	T	0.41	0.37	0.26	$5x10^{-5}$	*SV2C*,38	-

** SNP is located within the gene boundaries; distance is reported to start or end of gene, whichever is nearest. A1 = minor allele, Freq = allele frequency, OR = odds ratio, SE = standard error.*

+ Was not successfully incorporated into a multiplexed assay for the replication stage

 # Genotyped SNP

Supplementary Table 5. Frequency of the risk allele[+] in cases and controls by ancestry strata for the top 21 variants associated with peanut food allergy SNPs ($P < 5 \times 10^{-5}$).

Locus	SNP	Risk allele	Frequency in HEALTHNUTS samples (cases, controls)[a]				Frequency in 1000G samples	
			Overall	Europeans	Asians	Mixed	Europeans	Asians
1	rs10018666	T	0.88,0.70	0.94,0.73	0.68,0.40	0.88,0.63	0.82	0.51
2	rs6928827	G	0.98,0.83	0.98,0.85	1.00,0.90	0.93,0.71	0.87	0.87
3	rs6686894	G	0.01,0.16	0.02,0.17	0.00,0.06	0.00,0.10	0.13	0.08
4	rs11700330	A	0.08,0.29	0.11,0.32	0.00,0.20	0.06,0.21	0.28	0.04
5	rs12142904	G	0.36,0.18	0.38,0.19	0.29,0.15	0.35,0.14	0.26	0.27
6	rs4240433	T	0.90,0.71	0.91,0.71	0.86,0.80	0.88,0.68	0.79	0.78
7	rs7300806	G	0.70,0.85	0.60,0.82	0.94,0.95	0.95,0.97	0.78	0.97
8	rs9362681	G	0.47,0.23	0.37,0.20	0.82,0.45	0.44,0.26	0.25	0.67
9	rs6763069	A	0.50,0.70	0.45,0.68	0.64,0.80	0.56,0.80	0.62	0.72
10	rs2439871	C	0.42,0.67	0.53,0.74	0.07,0.20	0.38,0.47	0.66	0.08
11	rs73971133	A	0.01,0.09	0.00,0.06	0.04,0.30	0.05,0.13	0.03	0.17
12	rs73220497	T	0.01,0.11	0.00,0.13	0.00,0.00	0.06,0.03	0.08	0.00
13	rs17555239	T	0.55,0.37	0.57,0.38	0.39,0.30	0.75,0.34	0.40	0.31
14	rs16870788	G	0.25,0.11	0.25,0.12	0.12,0.10	0.51,0.08	0.17	0.06
15	rs8077351	G	0.01,0.09	0.00,0.08	0.05,0.10	0.00,0.16	0.05	0.09
16	rs57144668	C	0.50,0.72	0.61,0.76	0.21,0.50	0.31,0.58	0.73	0.36
17	rs6584390	T	0.88,0.71	0.83,0.67	1.00,1.00	0.94,0.79	0.73	0.99
18	rs10812871	C	0.42,0.58	0.39,0.54	0.46,0.80	0.50,0.74	0.54	0.68
19	rs7131777	T	0.62,0.45	0.65,0.45	0.53,0.36	0.62,0.46	0.44	0.51
20	rs864481	A	0.35,0.18	0.33,0.18	0.43,0.20	0.31,0.16	0.17	0.28
21	rs10474468	T	0.27,0.48	0.33,0.53	0.07,0.10	0.19,0.37	0.46	0.03

[a] Sample size for each strata is as follows: Overall, 73 cases and 148 controls; Europeans, 51 cases and 119 controls; Asians, 14 cases and 10 controls; Mixed European Asians, 8 cases and 19 controls.

[+]Where risk allele refers to the reference allele conferring a risk effect (OR>1)

Supplementary Table 6. Association results in the Melbourne replication cohorts for 19 SNPs that were selected for validation and successfully genotyped

SNP	Effect allele	Ancestry strata[a]					
		Overall		Europeans		Mixed	
		OR	P	OR	P	OR	P
rs11700330	A	0.61	0.009*	0.57	0.007	0.86	0.742
rs57144668	T	1.50	0.014*	1.31	0.144	2.46	0.017
rs10812871	T	0.68	0.014*	0.71	0.064	0.57	0.088
rs9362681	G	1.43	0.037*	1.39	0.088	1.54	0.222
rs17555239	T	0.79	0.131	0.74	0.089	1.01	0.983
rs10474468	C	0.84	0.261	1.11	0.534	0.55	0.154
rs6686894	G	1.29	0.280	1.30	0.298	1.23	0.744
rs6928827	A	0.77	0.292	0.87	0.607	0.55	0.232
rs4240433	C	0.83	0.316	0.74	0.177	1.11	0.763
rs7300806	A	0.82	0.319	0.85	0.464	0.70	0.447
rs10018666	C	1.18	0.360	1.47	0.061	0.62	0.187
rs864481	A	1.09	0.681	1.00	0.996	1.51	0.357
rs2439871	G	0.94	0.723	0.99	0.976	0.74	0.448
rs73971133	A	0.87	0.723	0.84	0.690	1.00	1.000
rs16870788	G	0.93	0.734	0.99	0.973	0.67	0.448
rs8077351	G	1.07	0.820	1.06	0.858	1.11	0.884
rs6584390	C	1.03	0.864	0.99	0.976	1.27	0.604
rs73220497	T	1.04	0.891	1.08	0.797	0.91	0.859
rs12142904	G	1.02	0.905	1.11	0.607	0.75	0.464

Supplementary Table 7. Association results in the German UFA study replication study for the 21 SNPs selected for validation.

SNP	Allele1	Allele2	OR	SE	*P*-value
rs10018666	T	C	0.87	0.13	0.30
rs6928827	G	A	0.87	0.16	0.39
rs6686894	A	G	0.94	0.19	0.75
rs11700330	G	A	0.86	0.14	0.28
rs12142904	A	G	1.10	0.13	0.47
rs4240433	T	C	0.87	0.12	0.26
rs7300806	G	A	0.96	0.12	0.75
rs9362681	A	G	1.05	0.12	0.70
rs6763069	A	T	0.85	0.12	0.15
rs2439871	G	C	1.09	0.12	0.49
rs73971133	G	A	1.28	0.27	0.36
rs73220497	G	T	1.57	0.22	0.03
rs17555239	C	T	0.98	0.11	0.88
rs16870788	A	G	0.79	0.14	0.09
rs8077351	A	G	0.84	0.21	0.42
rs57144668	C	T	1.16	0.13	0.23
rs6584390	T	C	1.04	0.12	0.75
rs10812871	C	T	0.99	0.11	0.94
rs7131777	C	T	0.97	0.11	0.81
rs864481	G	A	0.98	0.14	0.88
rs10474468	C	T	1.09	0.12	0.48

Supplementary Table 8. Association results for the 21 SNPs selected for validation from the Chicago Food Allergy Study .

SNP	Allele1	Allele2	Freq cases	Freq controls	*P-value*
rs10018666	C	T	0.22	0.19	0.65
rs2439871	G	C	0.28	0.27	0.21
rs10812871	T	C	0.45	0.50	0.18
rs864481	A	G	0.18	0.23	0.07
rs11700330	A	G	0.22	0.21	0.90
rs6763069	T	A	0.37	0.33	0.42
rs16870788	G	A	0.14	0.16	0.70
rs7131777	A	C	0.50	0.52	0.85
rs9362681	G	A	0.27	0.27	0.57
rs73220497	T	G	0.11	0.14	0.58
rs7300806	A	G	0.21	0.22	0.56
rs6686894	G	A	0.14	0.13	0.25
rs10474468	T	C	0.49	0.48	0.86
rs4240433	C	T	0.23	0.27	0.64
rs8077351	G	A	0.06	0.05	0.95
rs6928827	A	G	0.11	0.12	0.43
rs73971133	A	G	0.06	0.06	0.42
rs17555239	T	C	0.40	0.39	0.56
rs12142904	C	T	0.27	0.28	0.84
rs6584390	C	T	0.30	0.32	0.66
rs57144668	T	C	0.30	0.25	0.009

70

71

72 **Supplementary Table 9.** Genotyped panel of Ancestry Informative Markers to genetically infer ancestry of replication

73 population.

Locus	SNP	Frequency in 1000G samples							
		CEU				CHB			
		A	C	G	T	A	C	G	T
16	rs1002587		0.32	0.68			0.91	0.09	
12	rs10879311	0.2		0.8		0.83		0.17	
8	rs1227647	0.72		0.28		0.27		0.73	
8	rs12678324		0.92		0.08		0.55		0.45
8	rs1402851	0.95			0.05	0.51			0.49
4	rs16877243		0.81		0.19		0.38		0.62
3	rs1698042			0.06	0.94			0.47	0.53
10	rs1986420	0.96		0.04		0.21		0.79	
1	rs2759281		0.13		0.87		0.8		0.2
15	rs2934193		0.88		0.12		0.22		0.78
5	rs326626		0.94	0.06			0.94	0.06	
8	rs3912537		0.9	0.1			0.14	0.86	
8	rs4484738	0.75		0.25		0.24		0.76	
1	rs4653130		0.05		0.95		0.81		0.19
22	rs4824001		0.58	0.42			0.51	0.49	
17	rs4968382	0.23		0.77		0.55		0.45	
20	rs6141319	0.71		0.29		0.02		0.98	
19	rs6510332	0.24		0.76		0.64		0.36	
4	rs6552216		0.81		0.19		0.07		0.93
18	rs679832		0.39		0.61		0.97		0.03
7	rs10488619	0.42		0.58		0.9		0.1	
4	rs11098964	0.51		0.49		0.06		0.94	
1	rs11184898	0.54			0.46	0.05			0.95
10	rs11203006	0.83		0.17		0.1		0.9	
15	rs12595448		0.25	0.75			0.98	0.02	
4	rs12644851	0.43		0.57		0.86		0.14	
8	rs1347201		0.57		0.43		0.06		0.94
4	rs1488299		0.53		0.47		0.95		0.05
3	rs1519260		0.42		0.58		0.87		0.13
6	rs1538956			0.55	0.45			0.02	0.98
14	rs2193595	0.48		0.52		0.93		0.07	
5	rs2416504		0.42		0.58		0.96		0.04
9	rs2486448			0.4	0.6			0.89	0.11
8	rs2927385		0.38	0.62			0.84	0.16	
7	rs315280	0.75		0.25		0.12		0.88	
3	rs36110		0.38		0.62		0.91		0.09
16	rs4240793	0.56	0.44			0.01	0.99		

1	rs4265409	0.6		0.4	0.04		0.96
8	rs4841401	0.31	0.69		0.92	0.08	
22	rs5753625	0.39		0.61	0.86		0.14
5	rs6595142	0.36		0.64	0.89		0.11

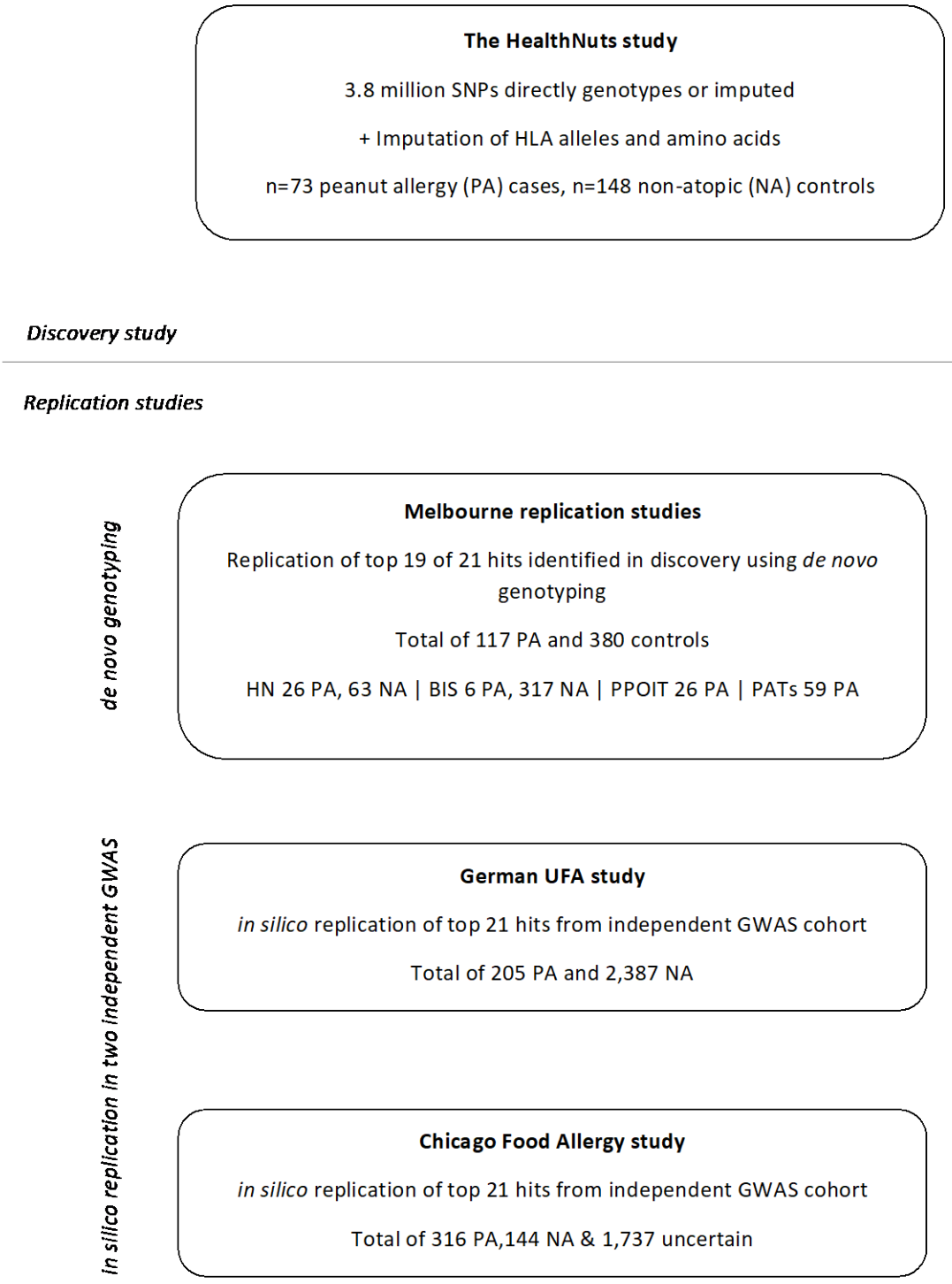
74

Supplementary Table 10. Allele frequencies shown for HealthNuts discovery and de novo genotyping r eplication studies

SNP	Risk allele	Frequency in REPLICATION samples (cases, controls)[a]			Frequency in 1000G samples	
		Overall	Europeans	Mixed	Europeans	Asians
rs6686894	G	0.11,0.10	0.14,0.10	0.08,0.08	0.13	0.08
rs12142904	G	0.26,0.28	0.30,0.28	0.18,0.24	0.26	0.27
rs10018666	T	0.71,0.78	0.75,0.81	0.74,0.63	0.82	0.51
rs57144668	C	0.61,0.72	0.68,0.73	0.53,0.72	0.73	0.36
rs10474468	T	0.66,0.51	0.48,0.51	0.79,0.69	0.46	0.03
rs864481	A	0.2,0.18	0.17,0.18	0.24,0.19	0.17	0.28
rs9362681	G	0.37,0.24	0.28,0.22	0.40,0.31	0.25	0.67
rs6928827	G	0.89,0.88	0.89,0.88	0.9,0.83	0.87	0.87
rs73220497	T	0.07,0.09	0.09,0.08	0.08,0.10	0.08	0
rs16870788	G	0.15,0.16	0.16,0.16	0.11,0.15	0.17	0.06
rs10812871	C	0.60,0.49	0.57,0.49	0.62,0.47	0.54	0.68
rs4240433	T	0.78,0.78	0.82,0.78	0.76,0.76	0.79	0.78
rs6584390	T	0.77,0.73	0.70,0.71	0.82,0.86	0.73	0.99
rs2439871	C	0.61,0.69	0.72,0.72	0.64,0.6	0.66	0.08
rs7300806	G	0.85,0.79	0.81,0.79	0.87,0.82	0.78	0.97
rs17555239	T	0.39,0.43	0.36,0.44	0.46,0.45	0.4	0.31
rs8077351	G	0.07,0.06	0.07,0.07	0.07,0.06	0.05	0.09
rs73971133	A	0.07,0.05	0.04,0.05	0.04,0.04	0.03	0.17
rs11700330	A	0.16,0.29	0.20,0.31	0.17,0.18	0.28	0.04

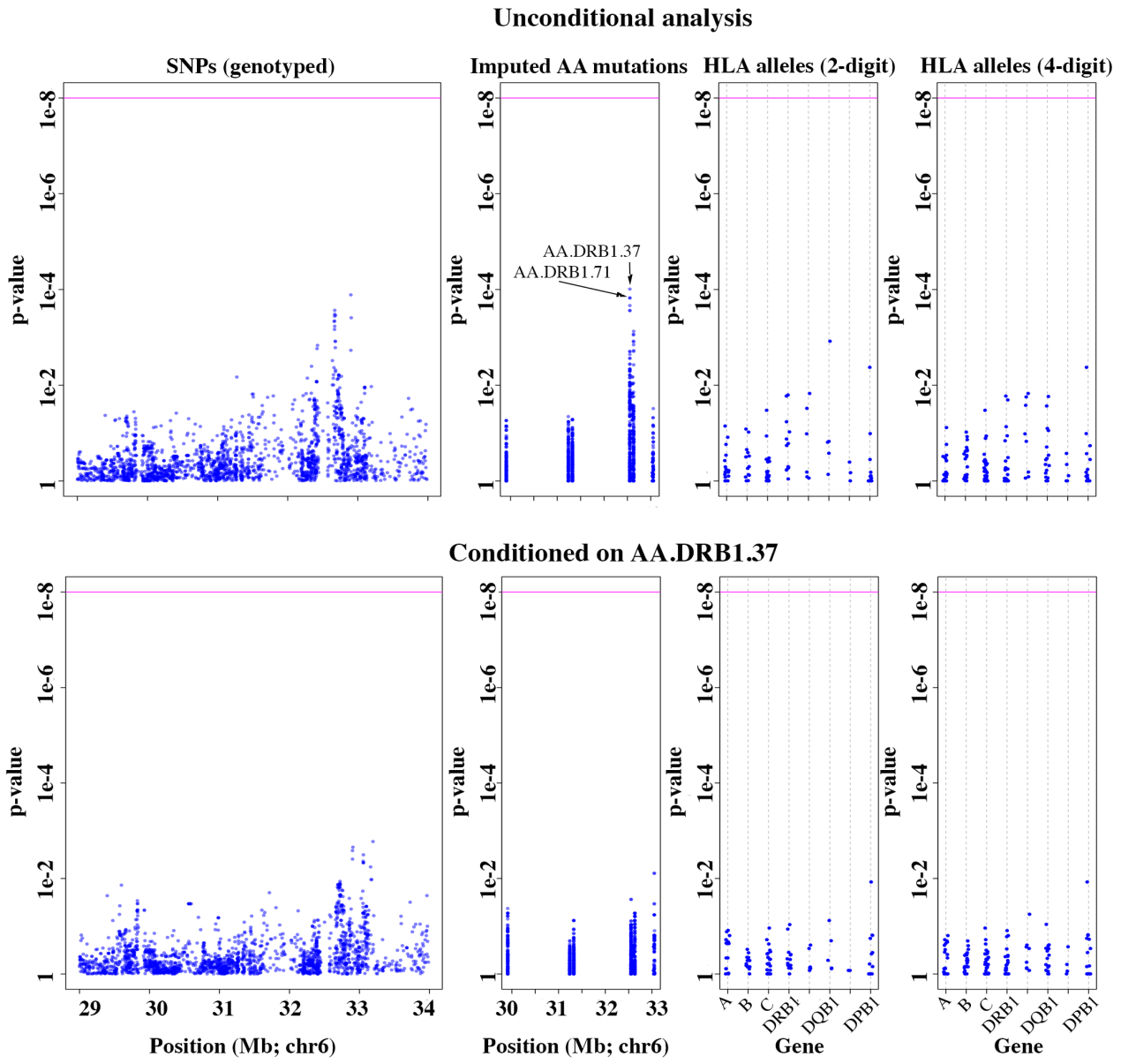
[+]Where risk allele refers to the reference allele conferring a risk effect (OR>1)

Supplementary figure 1. Overview of experimental strategy



HN = HealthNuts, BIS = Barwon Infant Study, PPOIT = Peanut and Probiotic Oral Immunotherapy Trial, PAT= Peanut Allergy Threshold Study,

Supplementary Figure 2. Conditional analysis of HLA genotyped and imputed mutations and variants.



Clinical and Experimental Allergy

November 7, 2016

Genome-wide association study of peanut allergy reproduces association with amino acid polymorphisms in *HLA-DRB1*

David J Martino [1 *], Sarah Ashley [1,4, *], Jennifer Koplin [1,2], Justine Ellis [1], Richard Saffery [1], Shyamali C Dharmage [2], Lyle Gurrin [2], Melanie C Matheson [2], Birgit Kalb [5,6,7], Ingo Marenholz [6,7], Kirsten Beyer [5], Young-Ae Lee [6,7], Xiumei Hong [8], Xiaobin Wang [8], Damjan Vukevic [1,3], Allan Motyer[1], Stephen Leslie [1,3], Katrina J. Allen [1] and Manuel A.R. Ferreira [9].

SUPPLEMENTARY METHODS

Supplementary Methods

Genotyping and quality control

DNA samples were extracted from peripheral blood from 768 participants whom attended HealthNuts clinics, and stored at -80 degrees. 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 (1), 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.

Replication study – de novo genotyping

We directly genotyped the 19 SNPs and 41 ancestry markers in an additional 147 peanut allergy cases and 387 non-allergic non-sensitized controls from Melbourne (HealthNuts, Barwon, Peanut

Allergen Threshold, PPOIT studies) who were unrelated to the participants included in the discovery GWAS.

All individuals were genotyped for 19 of 21 selected SNPs associated with peanut allergy in the discovery GWAS, as 2 SNPs were not able to be incorporated into the iPLEX assay (designated in **Supplementary Table 4**), as well as for a panel of ancestry informative markers (2) (**Supplementary Table 9).** As in the discovery GWAS, the genotyping dataset was merged with data from the 1000 Genomes Project and MDS analysis of IBS was performed in PLINK (3) to infer genetic ancestry for all genotyped individuals. There were no significant differences in allele frequency for all SNPs between the replication dataset and 1000 Genomes data (all with $P>0.05$, based on European controls only; **Supplementary table 10**). Only participants from European (88 cases, 340 controls) or mixed European-Asian (36 cases and 40 controls) ancestry were included in the analysis, given the small sample size for the Asian (23 cases and 5 controls) and African (2 controls) ancestry groups. All analyses were adjusted for ethnic group to control for population structure and the genomic inflation factor was used to assess this. Association testing for the 19 candidates was performed using logistic regression under an additive model of SNP allelic dosage with sex and ancestry included as covariates

In silico replication – American and German studies

The Chicago Food Allergy study is a family based study recruited from the Chicago area (2005 – 2011). Genotyping was carried out on genomic DNA from peripheral blood white cells on the Illumina HumanOmni1-Quad BeadChip. The replication analysis was carried out by testing for association with the HealthNuts 21 candidates or proxy markers for these ($r^2 > 0.8$) by applying the modified quasi-likelihood score (MQLS) test, which leverages the family-based data and allows for both unaffected controls and controls of uncertain phenotype (4). A priori significant associations were defined as those exhibiting an adjusted P value < 0.05 and consistent direction of effect to the HealthNuts discovery

data.

In the Understanding Food Allergy (UFA) study, unrelated cases (n=205) were recruited at the Department of Pediatrics, Division of Pneumology and Immunology, at Charité University Medicine Berlin, Germany. Controls were 2,387 unrelated individuals from the German population based Heinz Nixdorf Recall Study (5). All cases and controls were of Central European origin. Genotyping was performed using the HumanOmniExpressExome-8 v1.2 array or HumanOmniExpress-12 v1.1 plus HumanExome-12 v1 or HumanOmni1M-4 v1 plus HumanExome-12 v1. Replication analysis was carried out by testing for association with the HealthNuts 21 candidates or proxy markers for these ($r^2 >$ 0.8) by logistic regression analysis. Results from the UFA study were meta-analysed with those from the HealthNuts study using a fixed-effects model, as implemented in METAL (6). Results from the Chicago Food Allergy study could not be included in the meta-analysis as the analysis was based on the MQLS approach, which does not produce an odds ratio and standard error directly comparable to those obtained in HealthNuts. For both the Chicago Food Allergy and UFA studies, use of genotyping data described here was approved by local ethics committees and all participants or their legal guardians gave informed consent.

1. Consortium T1GP. An integrated map of genetic variation from 1,092 human genomes. Nature. Nature Publishing Group; 2012 Nov 1;491(7422):56–65.

2. Bousman CA, Yung AR, Pantelis C, Ellis JA. Effects of NRG1 and DAOA genetic variation on transition to psychosis in individuals at ultra-high risk for psychosis. Translational …. 2013.

3. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.

4. Hong X, Hao K, Ladd-Acosta C, Hansen KD, Tsai H-J, Liu X, et al. Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children. Nat Commun. 2015;6:6304.

5. Schmermund A, Möhlenkamp S, Stang A, Grönemeyer D, Seibel R, Hirche H, et al. Assessment

of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. Am Heart J. 2002 Aug;144(2):212–8.

6. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010 Sep 1;26(17):2190–1.