APPENDIX

SUPPLEMENTAL METHODS

Study population

NHS began in 1976, and is a prospective study of causes of cancer and other disease among US female nurses who were age 30 to 55 years at baseline. Blood was collected from 32,826 participants between 1989 and 1990. HPFS is the complementary all-male study among US health professionals. The study began in 1986 and men were age 40 to 75 years at baseline. Between 1993 and 1995, 18,255 men provided a blood sample. In both cohorts, DNA was extracted from white blood cells using the QIAmp (Qiagen, Inc, Chatsworth, CA) blood protocol and all blood samples were processed at the same laboratory. The participants included in the current project included men and women from European descent for whom prior GWAS data from nested case-control studies of type 2 diabetes, coronary heart disease, and breast cancer were available in NHS and HPFS using three different platforms: Affymetrix 6.0, Illumina HumanHap arrays (550 or 610), or Illumina OmniExpress arrays, as described in detail elsewhere.[1]

Asparagus anosmia definition

Information on asparagus anosmia was collected as part of a supplemental questionnaire sent in 2010 to living participants with existing GWAS data, of whom over 90% responded to the asparagus question. This was part of a broader questionnaire to evaluate genetics and smell. Participants were asked to

respond to the prompt: “People differ in their sense of smell. Please mark the response that best applies to you. After eating asparagus, you notice a strong characteristic odor in your urine.” For the primary analysis, participants who responded “Strongly agree” were categorized as able to smell asparagus metabolites and those who responded “Moderately agree”, “Slightly agree”, “Slightly disagree”, “Moderately disagree”, and “Strongly disagree” were categorized as having asparagus anosmia. Individuals who responded “I don’t eat asparagus” were excluded from the analysis. In addition, we performed a secondary analysis in which those who responded either “Strongly agree” or “Moderately agree” were considered to be able to smell asparagus odor.

Genome-wide association study

Genotyping and sample quality control measures (QC) for each study have been previously described.[1] The individual datasets genotyped on the same platform were combined, and SNPs with missing call rate >5% and duplicate IDs were excluded.[2] Genotypes were imputed using the 1000 Genomes Project Phase 1 v3 ALL panel to increase genome coverage. Imputation was performed using ChunkChromosome, MACH, and Minimac.[1] SNPs with imputation r^2<0.3 or minor allele frequency (MAF) <1% were excluded. There were about 9 million variants tested in this study.

Statistical Analysis

To further explore the association between genetic variation and asparagus anosmia, we conducted conditional genome-wide association analyses using GCTA-COJO.[3 4] This method approximates a stepwise conditional regression analysis using the marginal summary statistics from the GWAS meta-analysis and information on the linkage disequilibrium (LD) patterns among tested SNPs. We applied the method to all SNPs in a region on chromosome 1 defined by the furthest SNP upstream and the furthest SNP downstream in LD (r^2>0.3) with any anosmia-associated marker ($p < 1x10^{-8}$). The 1000 Genomes Phase1 v3 European sample genotypes were used to calculate linkage disequilibrium (LD). For each of the significant independent lead markers (conditional $p<5 \times 10^{-8}$) we identified sets of likely causal variants by retaining all SNPs in LD with the lead marker (r^2>0.8). We estimated the posterior odds against causality using a likelihood ratio comparing the likelihood for a SNP to the likelihood for the lead marker, and estimated posterior odds against causality smaller than 100:1.[5]

We identified coding variants in the anosmia-associated region on chromosome 1 using Variant Effect Predictor (VEP), and assessed whether these were in strong LD with any of the independent lead markers. We explored the possible impact of missense mutations on protein changes using the PolyPhen analysis tool (Version 2.2.2, http://genetics.bwh.harvard.edu/pph2/).

SUPPLEMENTAL RESULTS

Test statistic inflation at the meta-analysis level revealed no evidence of notable underlying population substructure ($\lambda = 0.998$, **Supplemental Figure 1**). Overall, 871 SNPs reached genome-wide significance ($p<5x10^{-8}$) for asparagus anosmia (**Supplemental Table 1)**. All significant SNPs were located in a 0.46 Mb region on chromosome 1 (248139851-248595299), which was split into two subregions by a recombination hotspot (**Supplemental Figure 2a and 2b**).

Sequential conditional analysis revealed three loci independently associated with asparagus anosmia in this region (rs13373863, rs71538191, rs6689553; conditional $p<5x10^{-8}$). These three SNPs were all imputed (rs13373863 Rsq =0.81 for Affymetrix and Illumina; rs71538191 Rsq =0.50 for Affymetrix and Rsq =0.39 for Illumina; rs6689553 Rsq =0.81 for Affymetrix and Rsq =0.73 for Illumina), and are not in strong LD with each other (r^2 rs13373863-rs71538191 = 0.002; r^2 rs13373863-rs6689553 = 0.01; r^2 rs71538191-rs6689553 = 0.43). Two of the SNPs (rs71538191 and rs6689553) are located 3' of a recombination hotspot near *OR2M2* (**Supplemental Figure 2a**); the third SNP (rs13373863) is located 5' of the hotspot (**Supplemental Figure 2b**). The SNPs rs13373863 and rs6689553 tag sets of likely causal variants with $r^2 > 0.8$ of size 63 and 9 SNPs, respectively. There were no SNPs in strong LD ($r^2 > 0.8$) with rs71538191 (**Supplemental Table 2**). Cochran's Q chi-square and p-value from the combined (males and females) analysis show no significant variation in study outcomes (**Supplemental Table 2**). In addition, we present the percentages of

those who can and cannot smell the odor across genotypes (**Supplemental Table 3**)

We also explored whether the three SNPs in Table 2 that tag the three independent signals at 1q44 were in strong linkage disequilibrium with SNPs in regulatory regions or known eQTLs using HaploReg and the GTEx portal. None of these SNPs had an r2>0.8 with any known eQTL or a SNP in promoter histone marker, enhancer histone markers or DNAse hypersensitive regions.[6]

Two of these missense SNPs (rs7555310 and rs7555424 in *OR2M7*) were in LD with one of the SNPs identified in the conditional analysis, rs6689553 (r^2 = 0.80) (**Supplemental Table 4**). These were the only genome-wide significant missense variants in one of the three sets of likely causal variants (**Supplemental Table 5**).

We undertook a secondary analysis and classified the ability to smell urinary metabolites of asparagus to include both “Strongly agree” and “Moderately agree” and asparagus anosmic for the others. Using this definition, 47% of participants were classified as anosmic with a similar proportion of men (46%) and of women (47%). The genetic loci identified with the genome-wide significant SNPs were similar to those associated with the SNPs identified in the primary analysis (data not shown).

References for Appendix

1. Sara Lindstrom SL, Constance Chen, Hongan Huang, Jinyan Huang, Hugues Aschard, Andrew T Chan, Hyon Choi, Marilyn Cornelis, Gary Curhan, Immaculata De Vivo, A. Heather Eliassen, Charles Fuchs, Michael Gaziano, Susan E. Hankinson, Frank Hu, Majken Jensen, Jae H. Kang, Christopher Kabrhel, Liming Liang, Louis R. Pasquale, Eric Rimm, Meir J. Stampfer, Rulla M. Tamimi, Shelly S. Tworoger, Janey L. Wiggs, David J. Hunter, Peter Kraft. A comprehensive survey of genetic variation in 20,691 subjects from four large cohorts. Manuscript submitted, 2015.
2. Yu K, Wang Z, Li Q, et al. Population substructure and control selection in genome-wide association studies. PLoS One 2008;**3**(7):e2551.
3. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet 2012;**44**(4):369-75, S1-3.
4. Yang J, Lee SH, Goddard ME, et al. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011;**88**(1):76-82.
5. Udler MS, Tyrer J, Easton DF. Evaluating the power to discriminate between highly correlated SNPs in genetic association studies. Genetic epidemiology 2010;**34**(5):463-8.
6. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012;**40**(Database issue):D930-4.