**Online supplementary material**

# Genome-wide association study identifies variants in *HORMAD2* associated with tonsillectomy

Bjarke Feenstra, PhD[1], Peter Bager, PhD[1], Xueping Liu, PhD[1], Henrik Hjalgrim, MD, DMSc[1,2], Ellen A Nøhr, PhD[3], David M. Hougaard, MD, DMSc[4], Frank Geller, MSc[1], Mads Melbye, MD, DMSc[1,5,6].

**Affiliations:**

[1]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark

[2]Department of Haematology, Rigshospitalet, Copenhagen, Denmark

[3]Research Unit of Gynaecology & Obstetrics, Department of Clinical Research, University of Southern Denmark, Odense

[4]Danish Centre for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark

[5]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

[6]Department of Medicine, Stanford University School of Medicine, Stanford, California, USA

*Correspondence should be addressed to B.F. (fee@ssi.dk).

**Table of contents**

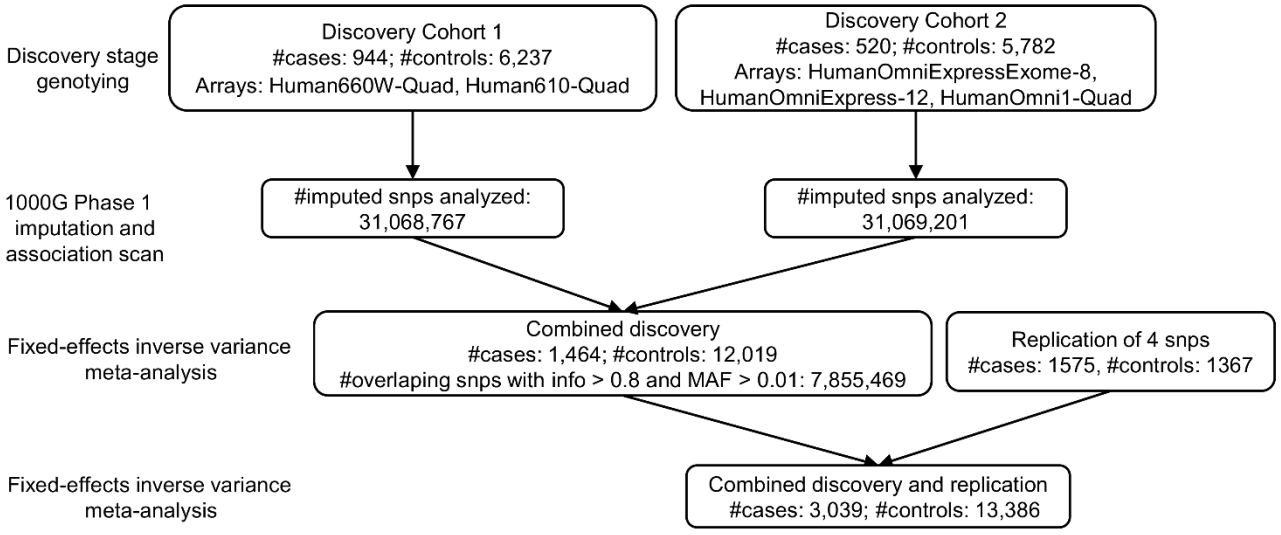


**Figure S1.**

Study design

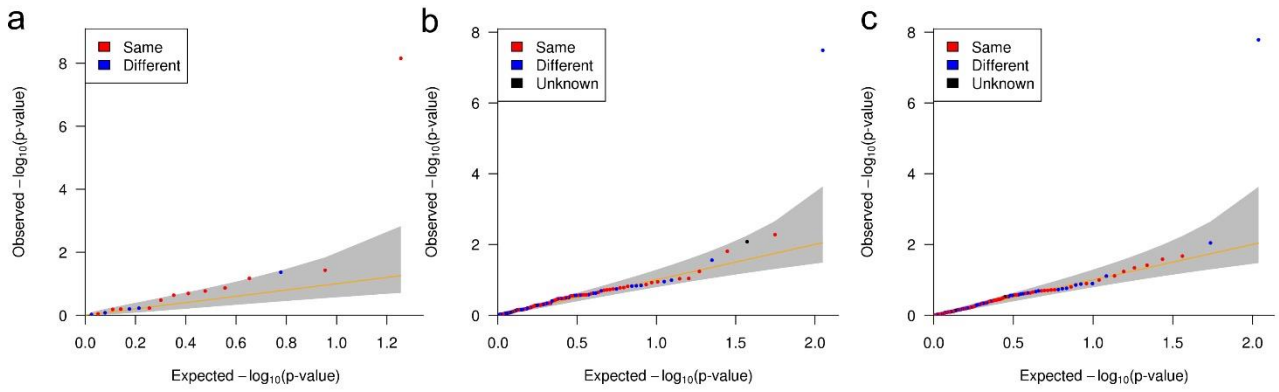


**Figure S2.**

Quantile-quantile plot of associations (from the meta-analysis of 13,483 discovery samples) between tonsillectomy and (a) 17 SNPs known to be associated with IgA nephropathy, (b) 111 SNPs known to be associated with Crohn's disease, and (c) 108 SNPs known to be associated with IBD. To be consider already known, the association had to be reported in the GWAS catalog with $p<5\times10^{-8}$. Observed versus expected $-\log_{10}$ p values are plotted for all SNPs and the orange line represents expected $-\log_{10}$ p values under the null distribution. The gray area defines the 95% concentration bands, which are an approximation to the 95% confidence intervals around the expected line. Directions of effects are indicated in colors; red dots are SNPs where the risk allele for tonsillectomy also corresponds to increased risk of the disease, blue dots are SNPs where the risk allele for tonsillectomy corresponds to decreased risk the disease, and black dots are SNPs where the risk allele of the disease was not reported in the GWAS catalog.

**Table S1.**

Sample characteristics.

| | Discovery1 | | Discovery2 | | Replication | |
|---|---|---|---|---|---|---|
| | cases | controls | cases | controls | cases | controls |
| Boys, No. (%) | 117 (12.4) | 1319 (21.1) | 224 (43.1) | 3072 (53.1%) | 21 (1.3) | 605 (44.3) |
| Year of birth, range | 1958-2008 | 1953-2008 | 1968-2008 | 1957-2009 | 1956-2006 | 1986-2010 |
| Year of birth, mean (SD) | 1977 (12) | 1979 (14) | 1992 (8) | 1996 (8) | 1971 (7) | 1998 (3) |
| Age of tonsillectomy, range, years | 1-44 | NA | 1-30 | NA | 1-48 | NA |
| Age of tonsillectomy, median [interquartile range], years | 16 (8-21) | NA | 6 (4-14) | NA | 17 (14-24) | NA |
| Age of tonsillectomy, mean (SD), years | 15.8 (9.2) | NA | 8.9 (6.6) | NA | 18.9 (9.0) | NA |

**Table S2.** Discovery, replication and combined results for one additional SNP at each of the 22q12.2 and 7p21.3 loci.

| Chromosome Position (bp) SNP (effect/ alternate allele) | Sample sets | Effect allele frequency | | Number | | OR (95% CI) | p | $I^2$ | $p_{het}$ |
|---|---|---|---|---|---|---|---|---|---|
| | | cases | controls | cases | controls | | | | |
| 22 30540590 rs2412975 (T/C) | Discovery1 | 0.606 | 0.550 | 944 | 6237 | 1.26 (1.14-1.40) | 3.81E-6 | | |
| | Discovery2 | 0.616 | 0.559 | 520 | 5782 | 1.28 (1.12-1.46) | 0.00025 | | |
| | Combined discovery | | | 1464 | 12019 | 1.27 (1.17-1.38) | 6.25E-9 | 0 | 0.90 |
| | Replication | 0.602 | 0.574 | 1567 | 1364 | 1.13 (1.01-1.25) | 0.027 | | |
| | All combined | | | 3031 | 13383 | 1.21 (1.14-1.21) | 2.50E-9 | 67.6 | 0.08 |
| 7 7554505 rs17168369 (G/A) | Discovery1 | 0.923 | 0.897 | 944 | 6237 | 1.38 (1.15-1.65) | 0.00028 | | |
| | Discovery2 | 0.935 | 0.896 | 520 | 5782 | 1.66 (1.29-2.14) | 2.64E-5 | | |
| | Combined discovery | | | 1464 | 12019 | 1.47 (1.27-1.70) | 3.18E-7 | 27.8 | 0.24 |
| | Replication | 0.891 | 0.895 | 1575 | 1362 | 0.97 (0.82-1.14) | 0.69 | | |
| | All combined | | | 3039 | 13381 | 1.22 (1.09-1.36) | 0.00038 | 92.7 | 0.00022 |

OR, odds ratio; CI, confidence interval; $I^2$, heterogeneity estimate; $p_{het}$, p *value* from the *Cochran Q* test of heterogeneity.

**Table S3.** Discovery, replication and combined results for the top SNPs at the 22q12.2 and 7p21.3 loci after requiring all pairs of individuals in the study to have an estimated identity by descent proportion less than 0.1875.

| Chromosome Position (bp) SNP (effect/ alternate allele) | Sample sets | Effect allele frequency | | Number | | OR (95% CI) | *P* | $I^2$ | $P_{het}$ |
|---|---|---|---|---|---|---|---|---|---|
| | | cases | controls | cases | controls | | | | |
| 22 30494371 rs2412971 (G/A) | Discovery1 | 0.607 | 0.550 | 941 | 6189 | 1.27 (1.15-1.40) | 3.18E-6 | | |
| | Discovery2 | 0.616 | 0.561 | 517 | 5567 | 1.26 (1.11-1.44) | 0.00051 | | |
| | Combined discovery | | | 1458 | 11756 | 1.27 (1.17-1.37) | 9.96E-9 | 0 | 0.98 |
| | Replication | 0.603 | 0.573 | 1567 | 1367 | 1.14 (1.02-1.26) | 0.017 | | |
| | All combined | | | 3025 | 13123 | 1.22 (1.14-1.30) | 2.01E-9 | 59.7 | 0.12 |
| 22 30540590 rs2412975 (T/C) | Discovery1 | 0.606 | 0.550 | 941 | 6189 | 1.27 (1.15-1.40) | 3.46E-6 | | |
| | Discovery2 | 0.617 | 0.561 | 517 | 5567 | 1.27 (1.11-1.45) | 0.00041 | | |
| | Combined discovery | | | 1458 | 11756 | 1.27 (1.17-1.37) | 8.89E-9 | 0 | 0.99 |
| | Replication | 0.602 | 0.574 | 1567 | 1364 | 1.13 (1.01-1.25) | 0.027 | | |
| | All combined | | | 3025 | 13120 | 1.21 (1.14-1.29) | 3.40E-9 | 66.5 | 0.08 |
| 7 7488601 rs10255201 (T/C) | Discovery1 | 0.915 | 0.886 | 941 | 6189 | 1.39 (1.17-1.65) | 0.00013 | | |
| | Discovery2 | 0.920 | 0.884 | 517 | 5567 | 1.53 (1.21-1.93) | 0.00019 | | |
| | Combined discovery | | | 1458 | 11756 | 1.43 (1.25-1.65) | 4.39E-7 | 0 | 0.52 |
| | Replication | 0.880 | 0.884 | 1571 | 1360 | 0.97 (0.82-1.13) | 0.68 | | |
| | All combined | | | 3029 | 13116 | 1.21 (1.09-1.34) | 0.00043 | 92.5 | 0.00027 |
| 7 7554505 rs17168369 (G/A) | Discovery1 | 0.922 | 0.897 | 941 | 6189 | 1.37 (1.14-1.64) | 0.00040 | | |
| | Discovery2 | 0.935 | 0.896 | 517 | 5567 | 1.67 (1.30-2.16) | 2.52E-5 | | |
| | Combined discovery | | | 1458 | 11756 | 1.46 (1.26-1.70) | 4.58E-7 | 36.4 | 0.21 |
| | Replication | 0.891 | 0.895 | 1575 | 1362 | 0.97 (0.82-1.14) | 0.69 | | |
| | All combined | | | 3033 | 13118 | 1.22 (1.09-1.36) | 0.00047 | 92.5 | 0.00026 |

**Table S4.** Association results for 145 SNPs with $p<10^{-4}$ across the 22q12.2 locus associated with tonsillectomy. The table is sorted by base pair position (NCBI build 37) and shows p value (combined discovery stage, after genomic control); effect and alternative allele; direction of effect in the two discovery cohorts; heterogeneity test statistic ($I^2$) and p value; odds ratio estimate with 95% confidence interval; squared Pearson correlation coefficient ($r^2$) of imputed SNP allele dosage to allele dosage for the top SNP at the locus (rs2412971); characteristics of the two discovery cohorts, including: indicator of whether the SNP was genotyped, info value from SNPTEST indicating imputation quality, effect allele frequency (overall and in cases and controls); eQTL results for the SNP from the GTEx and GEUVADIS consortia; variant class and nearest gene; reported associations for the SNP in the GWAS catalog. The table is provided as an additional Excel spreadsheet (online).

**Table S5.** Discovery, replication and combined results for rs2412971 for outcomes related to tonsillectomy.

| Outcome | Sample set | Effect allele (G) frequency | | Number | | OR (95% CI) | p | $I^2$ | $p_{het}$ |
|---|---|---|---|---|---|---|---|---|---|
| | | cases | controls | cases | controls | | | | |
| Acute tonsillitis | Discovery1 | 0.583 | 0.55 | 290 | 6237 | 1.14 (0.97-1.36) | 0.11918 | | |
| | Discovery2 | 0.607 | 0.559 | 219 | 5782 | 1.22 (1.00-1.49) | 0.043774 | | |
| | Combined discovery | | | 509 | 12019 | 1.18 (1.04-1.34) | 0.01293 | 0 | 0.62 |
| | Replication | 0.621 | 0.573 | 343 | 1350 | 1.22 (1.03-1.46) | 0.02215 | | |
| | All combined | | | 852 | 13369 | 1.19 (1.08-1.32) | 0.000782 | 0 | 0.73 |
| Chronic disease of the tonsils | Discovery1 | 0.601 | 0.55 | 704 | 6237 | 1.23 (1.10-1.38) | 0.000248 | | |
| | Discovery2 | 0.612 | 0.559 | 408 | 5782 | 1.25 (1.08-1.45) | 0.0026 | | |
| | Combined discovery | | | 1112 | 12019 | 1.24 (1.13-1.36) | 2.45E-06 | 0 | 0.88 |
| | Replication | 0.595 | 0.572 | 1019 | 1360 | 1.10 (0.98-1.24) | 0.1062 | | |
| | All combined | | | 2131 | 13379 | 1.19 (1.11-1.28) | 2.32E-06 | 59.7 | 0.12 |
| Appendectomy | Discovery1 | 0.59 | 0.548 | 487 | 5934 | 1.19 (1.04-1.36) | 0.010 | | |
| | Discovery2 | 0.616 | 0.558 | 215 | 5607 | 1.28 (1.05-1.56) | 0.015 | | |
| | Combined discovery | | | 702 | 11541 | 1.22 (1.09-1.36) | 0.00051 | 0 | 0.56 |
| | Replication | 0.583 | 0.574 | 168 | 1342 | 1.04 (0.82-1.3) | 0.76 | | |
| | All combined | | | 870 | 12883 | 1.18 (1.07-1.31) | 0.00113 | 34.8 | 0.22 |

OR, odds ratio; CI, confidence interval; $I^2$, heterogeneity estimate; $p_{het}$, p value from the *Cochran Q* test of heterogeneity.

**Table S6.** Discovery, replication and combined results for rs2412971 for tonsillectomy association analyses stratified by sex and age, respectively.

| Stratification factor | Sample set | Effect allele (G) frequency | | Number | | OR (95% CI) | p | $I^2$ | $p_{het}$ |
|---|---|---|---|---|---|---|---|---|---|
| | | cases | controls | cases | controls | | | | |
| Sex (**M**ale/**F**emale) | Discovery1 (M) | 0.611 | 0.544 | 117 | 1319 | 1.32 (1.00-1.74) | 0.046 | | |
| | Discovery2 (M) | 0.617 | 0.567 | 224 | 3072 | 1.24 (1.01-1.52) | 0.035 | | |
| | Combined discovery (M) | | | 341 | 4391 | 1.27 (1.08-1.49) | 0.004216 | 0 | 0.72 |
| | Replication (M) | 0.476 | 0.56 | 21 | 605 | 0.72 (0.39-1.33) | 0.30 | | |
| | M Combined | | | 362 | 4996 | 1.22 (1.04-1.43) | 0.013 | 67.4 | 0.080 |
| | Discovery1 (F) | 0.606 | 0.552 | 827 | 4918 | 1.25 (1.13-1.39) | 3.49E-5 | | |
| | Discovery2 (F) | 0.615 | 0.55 | 296 | 2710 | 1.31 (1.10-1.56) | 0.0021539 | | |
| | Combined discovery (F) | | | 1123 | 7628 | 1.27 (1.16-1.39) | 4.24E-7 | 0 | 0.66 |
| | Replication (F) | 0.605 | 0.583 | 1546 | 762 | 1.10 (0.97-1.25) | 0.1529 | | |
| | F combined | | | 2669 | 8390 | 1.21 (1.12-1.3) | 7.90E-7 | 69.3 | 0.07107 |
| | All combined (M+F) | | | 3031 | 13386 | 1.21 (1.13-1.29) | 3.25E-8 | 0 | 0.90 |
| Age (**<10** / **>=10** years) | Discovery1 (<10) | 0.624 | 0.542 | 271 | 1790 | 1.42 (1.18-1.72) | 0.00024 | | |
| | Discovery2 (<10) | 0.613 | 0.556 | 334 | 3714 | 1.27 (1.08-1.50) | 0.0038712 | | |
| | Combined discovery (<10) | | | 605 | 5504 | 1.34 (1.18-1.51) | 5.44E-6 | 0 | 0.39 |
| | Replication (<10) | 0.633 | 0.591 | 225 | 197 | 1.20 (0.90-1.58) | 0.21 | | |
| | Combined (<10) | | | 830 | 5701 | 1.31 (1.17-1.47) | 3.06E-6 | 0 | 0.48 |
| | Discovery1 (>=10) | 0.6 | 0.554 | 673 | 4447 | 1.21 (1.07-1.36) | 0.0014899 | | |
| | Discovery2 (>=10) | 0.619 | 0.564 | 185 | 2068 | 1.26 (1.01-1.57) | 0.037 | | |
| | Combined discovery (>=10) | | | 858 | 6515 | 1.22 (1.10-1.35) | 0.00019 | 0 | 0.74 |
| | Replication (>=10) | 0.598 | 0.57 | 1342 | 1170 | 1.13 (1.01-1.26) | 0.039 | | |
| | Combined (>=10) | | | 2200 | 7685 | 1.18 (1.09-1.27) | 3.40E-5 | 0 | 0.32 |
| | All combined (all ages) | | | 3030 | 13386 | 1.22 (1.14-1.30) | 1.43E-9 | 57.8 | 0.12 |

OR, odds ratio; CI, confidence interval; $I^2$, heterogeneity estimate; $p_{het}$, p value from the *Cochran Q* test of heterogeneity.