

Received Date: 19-Aug-2015

Accepted Date: 04-Nov-2015

Article Type: Original Article

Characterization of genetic variation in *TLR8* in relation to allergic rhinitis

V. Henmyr¹, C. Lind-Halldén¹, D. Carlberg¹, C. Halldén¹, E. Melén^{2,3}, M. Wickman^{2,3}, A. Bergström², T. Säll⁴, L.O. Cardell⁵

¹ Biomedicine, Kristianstad University, Kristianstad, Sweden

² Institute of Environmental Medicine Karolinska Institutet, Stockholm, Sweden

³ Sachs Children's Hospital, Stockholm, Sweden

⁴ Department of Biology, Lund University, Lund, Sweden

⁵ Division of ENT diseases, CLINTEC, Karolinska Institutet, Huddinge, Sweden

Corresponding author:

Viktor Henmyr, MSc, Biomedicine,
Kristianstad University, Kristianstad, SE-291 39 Kristianstad, Sweden, Phone:
+46732626874, Email: viktor.henmyr@hkr.se

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/all.12805

This article is protected by copyright. All rights reserved.

Abstract

Background: A previous investigation of all 10 TLR-genes for associations with allergic rhinitis (AR) detected a number of significant SNPs in the *TLR8* locus. The associations indicated that an accumulation of rare variants could explain the signal. The present study therefore searches for rare variants in the *TLR8* region and also investigates the reproducibility of previous SNP associations.

Methods: The *TLR8* gene was re-sequenced in 288 AR patients from Malmö and the data was compared with publically available data. Seven previously AR-associated SNPs from *TLR8* were analyzed for AR-associations in 422 AR patients and 859 controls from the BAMSE cohort. The associations detected in present and previous studies were compared.

Results: Sequencing detected 13 polymorphisms (3 promotor, 10 coding) among 288 AR patients. Four of the coding polymorphisms were rare (MAF <1%) and three of those were novel. Two coding polymorphisms were benign missense mutations and the rest were synonymous. Comparison with 1000Genomes and Exome Aggregation Consortium data revealed no accumulation of rare variants in the AR cases. The AR-association tests made using the BAMSE cohort yielded 5 *P*-values < 0.05. Tests of IgE-levels yielded 4 significant SNP associations to birch pollen. Comparing results between different populations revealed opposing risk alleles, different gender effects and response to different allergens in the different populations.

Conclusions: Rare variants in *TLR8* are not associated with AR. Comparison of present and previous association studies reveal contradictory results for common variants. Thus, no associations exist between genetic variation in *TLR8* and AR.

Short title: *TLR8* variants are not associated with AR

Key words: Allergic rhinitis; association; rare variant; resequencing; *TLR8*.

Allergic rhinitis (AR) is a global health problem affecting approximately 25% of the world population. AR is characterized by nasal symptoms after allergen exposure. It is associated with an IgE mediated inflammatory response against allergens and is also often associated by symptoms of the eyes and nose (1). Interactions between both environmental and genetic factors play a major role in the development of AR, and the heritability of AR has been estimated to 0.6-0.8 (2,3). Although a large number of candidate gene studies in AR have been published, few of the results have been convincingly replicated (4,5). Four genome-wide association studies (GWAS) have been published investigating different phenotypes of AR (6–9). Three of these were large meta-GWAS and together identified a total of 37 loci at genome-wide or suggestive significance levels. Few of the identified loci were shared between the studies. However, 4 loci were shared between all three studies and two of these were later replicated in the BAMSE cohort (10).

Like the majority of published GWAS, none of the GWAS described above included the X-chromosome in their analysis and therefore no association results from this chromosome were reported. A number of genome-wide linkage studies reporting linkage to the X-chromosome of different allergic disorders such as allergic asthma, atopic dermatitis (AD) and AR have been published (11-14). In the specific case of AR, two linkage studies showed evidence of linkage to AR for the Xp21region (13-14). However, three additional genome-wide linkage studies aiming for AR could not detect any signs of linkage of AR to the X-chromosome (15–17).

A number of candidate gene studies have reported associations between various diseases with variation in *TLR7* and *TLR8*, which are both located on chromosome Xp22 (18,19). In *TLR8*, the common Met1Val missense polymorphism has been implicated in a number of diseases such as tuberculosis (20) and progression of HIV infection (21). In the case of allergic diseases, a family-based association study investigated the *TLR7* and *TLR8* genes for association with asthma, AR and AD in a total of 235 families. Two SNPs, one in *TLR7* (rs179008) and one in *TLR8* (rs2407992) showed associations with all tested phenotypes. In rhinitis, the strongest association was detected for the *TLR7* SNP when analyzing only males ($P = 0.0039$). The *TLR8* SNP showed an association with rhinitis when analyzing both females and males combined ($P = 0.012$), and when analyzing only females ($P = 0.037$) (22). A case-control study investigating all 10 TLR-genes for associations with AR detected a number of SNPs in the *TLR8* gene region associated with AR in one Swedish and one Singapore Chinese population. The associated alleles were the major alleles and accounted

for the most common haplotypes in both populations. Two major discrepancies were however detected when comparing the two populations. First, the associations were detected in females in the Swedish population and in males in the Singapore Chinese population. Second, the allele frequencies of the two populations were completely reversed, resulting in different risk alleles for disease. This led to the hypothesis that accumulation of rare variants as a group could account for the detected associations (23).

In the present study, this hypothesis is evaluated by re-sequencing the complete coding sequence of *TLR8* and its putative promoter region in 288 AR cases from Sweden. In addition, the previously detected associations in Nilsson *et al* (2012) (23) are also evaluated in a replication study using the comprehensively characterized BAMSE cohort.

Methods

The Malmö study population was recruited at Malmö University hospital (Malmö, Sweden) between the years 2003 and 2009 and consists of unrelated subjects from the general population. It is comprised of 360 AR cases (169 females, 191 males, mean age 33 years) and 720 controls (294 females, 426 males, mean age 43 years) with no atopy and allergic symptomology. All cases were patients at the allergy clinic and were diagnosed with symptomatic birch and/or timothy grass pollen induced intermittent AR. Both cases and controls were of Caucasian origin, with both parents born in Sweden. Diagnostic procedures included personal interview of medical history and skin prick test (SPT) or Phadiatop test and were performed using standard panels of common airborne allergens. SPT were performed on the volar side of the forearm with saline buffer as negative and histamine chloride (10 mg/ml) as positive controls. A wheal reaction diameter of ≥ 3 mm was considered a positive SPT response. SPT was only performed if the AR cases had not taken any anti-allergic drugs for at least 3 days prior to the test. Atopy is defined as a positive SPT reaction to either one of allergens tested. AR is thus diagnosed based on the presence of atopic status and typical AR symptoms as defined by the Allergic Rhinitis Impact on Asthma (ARIA) guidelines (1). Conversely, the non-allergic controls are defined by having no atopy and no typical AR symptoms. For the sequencing experiment 288 patients, 148 males and 140 females, representing 428 chromosomes were selected. For the CNV experiment (see below) another 288 individuals, 163 patients and 125 controls were selected.

The BAMSE study is an unselected population-based birth cohort and consists of 4089 children recruited between the years 1994 and 1996. All participants were living in the

northern and central parts of Stockholm, Sweden, and were of mixed social economic status and from both urban and inner city areas. The participants and parents and/or legal guardians were given questionnaires concerning allergies, social economic status and housing. Follow-up studies were performed at 1, 2, 4, 8, 12 and 16 years of age. Additional details about the BAMSE study have been published elsewhere (24). The present study uses data from the 16 year follow-up and includes a total of 2153 children from the BAMSE cohort. At each follow-up, sera were screened with Phadiatop (mixture with 8 different inhalant allergens, Phadia AB, Uppsala, Sweden). Sera with IgE-value for Phadiatop ≥ 0.35 kU_A/l were then further analyzed for reactivity to single allergens. AR cases were defined according to the following definition: symptoms of sneezing, a runny or blocked nose, or itchy, red watery eyes after exposure to furred pets or pollen (last 12 months) and/or Doctor's diagnosis of allergic rhinitis from the age of 12 to the date of the 16 year follow-up, and IgE value for Phadiatop ≥ 0.35 kUA/l (25). Non-allergic controls were defined by having no symptoms of AR and an IgE value for phadiatop < 0.35 kUA/l

Genomic DNA was extracted from blood collected in EDTA using QIAamp DNA Blood Maxi or Mini kits (Qiagen, Hilden, Germany) and DNA concentrations determined by fluorometry using Pico Green (Molecular Probes, Invitrogen, Eugene, OR, USA). Primers were designed using NCBI Primer-BLAST and purchased from DNA Technology A/S (Risskov, Denmark). The primer systems were designed to amplify all coding sequence and 1000 bp upstream of the start of exon 1 (Table S1). Primary PCR was performed using KAPA Taq Extra HS PCR Kit (KAPA Biosystems, Cape Town, South Africa) and the following protocol: 1 cycle of 94°C for 3 min, 35 cycles of 94°C for 15 s, incubation at the annealing temperature (Table S1) for 15 s and 72°C for 1 min, and then 1 cycle of extension at 72°C for 10 min with a final soak at 4°C. Big Dye Terminator Sanger sequencing was performed in both directions using a 3130XL Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Primary PCR products were treated with ExoSAP-IT® (Applied Biosystems) and DNA sequencing was subsequently performed in a total volume of 5 µl containing 0.5X Big Dye sequencing ready reaction premix (Big Dye Terminator v 2.0, Applied Biosystems), 0.5X Big Dye Sequencing buffer and 3.2 pmol of the sequencing primer. The following PCR conditions were used: 1 cycle of 96°C for 1 min, 25 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 4 min. The sequencing reactions were purified using Xterminator (Applied Biosystems) according to the manufacturer's instructions. Sequences were interpreted and all polymorphisms were identified using SeqScape ver. 2.5 and

confirmed by manual inspection.

Long-range (LR) PCR systems were designed using NCBI/Primer-BLAST (Table S2). PCR was performed in a total reaction volume of 10 μ l, containing 20 ng of template DNA, 0.4 U LongRange PCR enzyme mix (Qiagen, Hilden, Germany), 1X LongRange PCR buffer, 0.2 mM dNTPs and 0.4 μ M of each primer. The PCR products were visualized on 1% agarose gels and analyzed with the Image Lab Software version 3.0 (Bio-Rad Laboratories Inc., Hercules, CA, USA).

Genotypes were determined using the Sequenom MassARRAY MALDI-TOF system (Sequenom Inc, San Diego, CA, USA). The system analyzes allele-specific primer extension products using mass-spectrometry. Assay design was made using the MassARRAY Assay Design v2.0 software. A total of 7 AR-associated SNPs were selected based on the results of Nilsson *et al* (2012) (23). All genotyping was performed at Mutation Analysis Core Facility, Karolinska Institutet, Huddinge, Sweden.

Statistical analyses were made using R statistical software (26) and PLINK v1.07 (27). For the association analysis of the BAMSE population, genotype frequencies were calculated and tested for Hardy-Weinberg equilibrium according to Zheng *et al* (2007) in both cases and controls (28). Allele and genotype frequencies were then investigated for association with AR using a χ^2 -homogeneity test. The allele tests were calculated for males and females separately as well as together, whereas the genotype tests were calculated for females only. The Kruskal-Wallis rank sum test was used for the analysis of genotype effects on the level of IgE-response.

To test for accumulation of rare variants in the coding region of *TLR8* (Malmö population), a simulation test was made using data from the Exome Aggregation Consortium (ExAC) (Cambridge, MA (URL: <http://exac.broadinstitute.org>) [date (03, 2015) accessed]). ExAC is a resource containing allele frequencies from 60,706 unrelated individuals from various disease-specific and population genetic studies. The non-Finnish European population of ExAC consists of > 30,000 individuals. A total of 118 polymorphisms (excluding indels) present in the coding region of *TLR8* from the non-Finnish European population were extracted from ExAC. For each of the 100 000 iterations, the number of expected variants were calculated using the minor allele frequency from ExAC data as the probability of finding a variant for 428 (148 males and 140 females individuals) chromosomes, corresponding to the number of chromosomes in the Malmö case population. The number of

variants from the simulation test were then compared with the number of detected variants in the Malmö population for two different categories; 1) total number of alternative allele counts of variants with MAF <1% and 2) total number of distinct variants with MAF <1%. A score was tallied each time the simulated dataset from ExAC had an excess of variants compared to the Malmö population. The test quantity was calculated as the total score over the number of iterations. In order to evaluate the power of such a simulation test, the simulation was repeated and each time the number of observed rare alleles was increased by 5%. It was then noted when fewer than 5% and 1% of the expected rare alleles were greater or equal to the observed.

A permutation test was used to investigate the probability of obtaining equal or larger proportions of significant results due to chance for the test of IgE-response variables. In each of the 10 000 iterations, the IgE-response variable were randomly distributed among the cases and a *P*-value was calculated for each SNP using the Kruskal-Wallis test. Each time the total number of significant *P*-values for all tested SNPs were equal to or exceeded the observed number a score was tallied. The test quantity was calculated as the total score over the number of iterations. This test was used to investigate the IgE and SPT variables with observed significances corresponding to previous results for timothy grass in the Malmö population (23), and to previous results for *Blomia tropicalis* in the Singapore population (23) and to birch in the BAMSE population of the present study.

To investigate whether the association signals in the present study (BAMSE population) and a previous study (Malmö and Singapore populations) (23), could be due to chance simulation tests were performed. For the Malmö (23) and BAMSE populations, the results from the allele tests of males and females combined, allele tests of females only and the genotype tests were simulated. The three tests were also combined into a total test. In the total test, all three categories each had to show equal or greater number of significant *P*-values compared to the observed number. Similarly, in the test for the Singapore population (23) the allele tests of males and the allele tests of males and females combined were simulated. Also in this case the two categories were combined in a total test. Each population was analyzed in the following way: 1) Haplotypes were phased using SHAPEIT v.2.5 (29). These haplotypes were then used as a training set together with recombination frequencies for the region to simulate a population of 20 000 haplotypes with no disease model utilizing HAPGEN2 (30) 2) New case and control populations were randomly picked from the simulated dataset, using the same population sizes and female to male ratios as in the original tests, 3) All SNPs were

then tested for the relevant association test using PLINK v.1.07 (27) and each time a test was significant ($P < 0.05$) a score, m , was given, 4) This score was then compared to the observed number of significant tests from the respective original test. If the score was greater or equal to the observed amount a score, n , was given, 5) Step 2-4 was iterated 10 000 times resulting in a test quantity = $n/10\ 000$.

Publically available polymorphisms in the *TLR8* gene region were extracted from the 1000Genomes (<http://www.1000genomes.org/>) database. The variants were obtained from the Integrated Variant Set of the 1000Genomes Project release April 2012 (<http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/>). Variants of the promoter and coding regions were extracted for 1092 individuals using tabix (31). The promoter region was defined as the sequence 1000 bp upstream the start of exon 1. The average coverage of the sequence data was $> 50X$ in the coding regions and 2-6X in the promoter regions (32). Using VCFtools (33) a dataset containing sequence data from 379 European individuals (EUR) was extracted. The EUR population consists of five different subpopulations: 89 individuals from England and Scotland (GBR), 85 from Utah with northern and western European ancestry (CEU), 98 from Italy (TSI), 93 from Finland (FIN) and 14 from Spain (IBS). Allele frequencies were calculated for each variation using the same allele as referent for all populations. Missense mutations identified in the study population and in the 1000Genomes population were investigated using SIFT (34) and Polyphen-2 (35).

A short outline of the present study can be found in Text S1 which describes which sources of data were used in the different analyses.

Results

Detection of rare variants in *TLR8*

Sanger sequencing was used to search for rare variants in the *TLR8* gene in a total of 288 AR cases. The putative promoter region, defined as 1000 bp upstream the start of exon 1, and the complete coding sequence of *TLR8* were re-sequenced. The sequence data was generally of high quality with $> 95\%$ of bases having a Phred score of 30 or higher in $> 95\%$ of individuals. The screening detected a total of 13 variants, 3 in the promoter and 10 in the coding region, among the 288 AR cases. Three coding variants were novel and detected in single chromosomes (Table 1). Variants with MAF < 0.01 represent only 5 chromosomes and cannot explain the association results from the previous study (23). A total of 19

chromosomes are represented when investigating variants with $MAF < 0.02$. When the presences of these variants are compared to the genotypes of rs3761624, which was the strongest association signal in the previous study (23), it is seen that the five genotypes (A/0, G/0, A/A, A/G and G/G) are evenly distributed among the 19 chromosomes. Thus, there is no general accumulation of low frequency variants on the major haplotype giving rise to the previous association signal.

In order to investigate a general accumulation of rare variants among patients, data from the European population (EUR) of the 1000Genomes was retrieved. The same number of variants (13, with 4 in the promoter and 9 in the coding region) was present in the EUR population. In the promoter region a total of three common SNPs ($MAF > 1\%$) were detected in both the AR cases and the EUR population. No rare variant ($MAF < 1\%$) was detected among the AR cases whereas one was found in the EUR population. One promoter SNP (rs3764879) showed a frequency difference of 9 % between the two populations. In the coding region, a total of 10 polymorphisms were detected of which three were novel, eight were synonymous and two were missense mutations, M1V and M391T (Table 1). The common M1V variant showed complete linkage disequilibrium with the rs3764879 SNP located in the promotor region and therefore also showed a frequency difference of 9 % between the AR cases and the EUR population. The M391T variant detected in the AR cases was novel and detected in a single chromosome. However, both missense mutations were classified as benign by PolyPhen2. In addition to the M391T missense mutation two rare synonymous variants were AR-specific whereas two were EUR-specific. Thus, there was no obvious excess of rare variants in the AR cases compared to the EUR population. This was confirmed using a simulation test for accumulation of rare variants ($MAF < 0.01$) using ExAC data. In fact, the results for the total number of rare alleles showed that in 95% of the iterations the simulated number of rare alleles was larger than the observed number for the AR cases. In contrast, the total number of distinct variants in the AR cases was as expected when compared to ExAC ($P = 0.25$). In order to assess the power conditions for the investigation the expected P -value was calculated for different levels of increased frequencies of rare variants relative to the ExAC dataset. The mean number of rare alleles for 428 chromosomes picked randomly from the ExAC dataset was found to be 6.4. To reach a significance level < 0.05 it would require 190% (12.0 alleles) increase in the number of observed rare alleles in the Malmö population and 220% (14.1 alleles) increase to reach $P <$

0.01. However, since the observed relative frequency of rare variants was lower in the Malmö population relative the ExAC dataset this information is of limited interest.

To investigate for the presence of copy number variants (CNVs) a total of six overlapping long range PCR systems were used to analyse 288 individuals (163 AR cases and 125 controls). The systems covered the complete *TLR8* region including 8kbp upstream exon 1. All six systems produced clear bands of expected sizes after optimization, but no variants were detected among the 288 individuals.

Replication of previous *TLR8* associations

A replication study of the associations previously detected in the Malmö population (23) was performed using the BAMSE cohort. This replication used 7 SNPs in association tests of 422 AR cases and 859 controls, as well as testing for genotype differences in IgE-response among the 422 AR cases. Investigating the allele and genotype frequencies in males and females separately as well as combined in addition to investigating IgE responses to timothy grass and birch among AR cases resulted in a total of 42 tests (28 for AR and 14 for IgE; Table 2, Table S3). The 28 association tests yielded 5 *P*-values < 0.05, 4 of which were detected for females only. The 14 tests of IgE-response yielded 4 significant results, all for reaction to birch pollen. The permutation tests of the number of significant comparisons were borderline significant for both association and IgE response tests (Table 3). Thus, to a large extent the association results from the BAMSE cohort yields an internally consistent picture compatible with an effect of *TLR8* in women. The lack of any apparent effects in males is however striking.

Evaluation of present and previous studies of *TLR8*

When the result of the BAMSE cohort is compared to the results of the original study based on cases and controls from Malmö (23), a number of inconsistencies appear (Table 2 and Figure 1). Comparing the association tests in women, 6 out of 7 were discordant with respect to risk alleles. Moreover, comparing the results of the skin-prick test with the IgE test the signals were detected for timothy grass in the original study whereas in the present study the effects were observed for birch. The only real consistency between the studies is that the association effects were observed for females in both. In the Singapore population on the other hand (23), the risk alleles are to a large extent the same as in the BAMSE cohort but the effects are observed only in males. When the simulation tests of the number of significant

results are applied to the Malmö and the Singapore data, the results are similar to the present study, showing a marginally significant overrepresentation of low P -values (Table 3). Thus, each study by itself implicates *TLR8* as a gene hosting risk alleles. When the results from the different populations are compared, however, the picture becomes highly inconsistent.

Discussion

The present study was initially motivated by the results in (22, 23) which in different ways indicated that genetic variation in *TLR8* and flanking sequences is associated with risk for developing AR. Nilsson *et al.* 2012 showed that it was the major alleles for a numbers of SNPs that acted as risk alleles in both a Malmö population and a Singapore Chinese population. However, the major alleles in the Malmö population were the minor alleles in the Singapore population and vice versa and thus the risk alleles were reversed in the two populations. This generated a hypothesis that the risk was really due to an accumulation of rare variants on common haplotypes. One primary aim of the present study was therefore to test this hypothesis by sequencing AR patients from Malmö for rare variants. However, no accumulation of rare variants was observed, not in general and not specifically on the common haplotype that acted as the risk haplotype in the Malmö AR population. If anything the comparison to the ExAC population indicated a deficit of rare variants. One interpretation of this would be the existence of rare protective alleles in the healthy population. Although such variants have been reported (36, 37), the present data does not allow such a conclusion. First, the result is not significant using a two sided test. Second, since the initial hypothesis were one-sided, an independent replication would be absolutely necessary.

To further investigate a possible AR-association for genetic variants in *TLR8*, a replication study was made using an extensively characterized longitudinal cohort, the BAMSE cohort (24). One advantage of this material is the matching between patients and controls inherent from the study design. The results of this replication showed a set of risk alleles acting primarily in females, exactly as found also in the Malmö population. However, the risk alleles were reversed in 6 out of 7 cases, including the rs3761624 SNP that is in complete linkage disequilibrium with the well-studied M1V missense polymorphism that have been previously indicated in a number of diseases such as tuberculosis (20) and HIV infection (21). The BAMSE and Malmö populations also differed with respect to the agents that gave rise to the significant phenotypic responses; birch in BAMSE and timothy grass in Malmö. Overall, it seems very unlikely that these discrepancies could be due to alleles and genotypes

reacting differently in the study populations from Malmö and Stockholm. In addition, if the data from the single concordant SNP (rs1548731) for the Malmö and BAMSE populations are combined a P -value of 0.00018 is obtained. However, there are at the same time 6 other SNPs that are discordant with respect to their risk alleles. In the comparison between the BAMSE and Singapore populations the risk alleles are to a large extent concordant, but the sexes are reversed. Since the Malmö and Singapore populations were discordant not only with respect to what alleles that conferred risk, but also with respect to ethnicity and what sex that was affected, the contradictory results could possibly have an origin in different environment by genotype interactions being present in the two ethnically different populations.

The present association study based on the BAMSE cohort, and the previously published Malmö and Singapore studies, all show an aggregation of P -values < 0.05 and each of them also show a considerable degree of internal consistency, e.g. the same risk alleles appear in both the allele, genotype and phenotypic tests. Yet, there is a high level of inconsistency between the studies. The BAMSE study was motivated by an initial signal in the *TLR8* locus in the Malmö and Singapore materials and by the absence of an accumulation of rare variants in *TLR8*. The most likely outcomes of the present study would either be a confirmation of the original effects, at least with respect to risk alleles, or a complete absence of signals in the BAMSE population. Instead a very complex picture emerges where opposing risk alleles, opposing effects in the sexes and varying effects to agents in different populations are observed. For this to be true, very dramatic genotype by environment interactions are necessary. As pointed out above, each investigation on its own represents a consistent result each indicating *TLR8*, but when the investigations are compared multiple inconsistencies become obvious. The pattern of low P -values indicates an additional variance component which acts systematically within each population, but randomly between them. The nature of such a hypothetical effect is beyond the imagination of the authors of this study. LD naturally influences the overall association pattern, but even when LD is taken into account there is still a surplus of significant results (Table 3). A question that is often raised when a replication of an association fails is that of sufficient power. The fact that there is an accumulation of significant results in each of the three studies, but in different directions shows that it is not simply a matter of reduced power. Thus, the enigma remains – are there genetic variants in *TLR8* that contribute to the AR phenotype?

Author contribution: All authors contributed to the design of the experiments. DN, VH, CH, TS performed statistical analysis. CLH performed Sanger sequencing, MW, EM, LOC acquired phenotypic data and all authors contributed to the writing of the manuscript and approved the final version of the manuscript.

Conflicts of interest: The authors declare no conflicts of interest.

Supporting information

Additional supporting information may be found in the online version of this article:

Text S1. Description of study outline

Table S1. Sanger sequencing primers for *TLR8* promotor and exons

Table S2. Long-range PCR primers for *TLR8* region

Table S3. Minor allele frequencies of genotyped SNPs in BAMSE and Malmö populations

References

1. Bousquet J, Khaltaev N, Cruz a a, Denburg J, Fokkens WJ, Togias a et al. Review article Allergic Rhinitis and its Impact on Asthma (ARIA) 2008 * Review Group : *Allergy* 2008;**63**:8–160.
2. Willemsen G, van Beijsterveldt TCEM, van Baal CGCM, Postma D, Boomsma DI. Heritability of self-reported asthma and allergy: a study in adult Dutch twins, siblings and parents. *Twin Res Hum Genet* 2008;**11**:132–142.
3. Fagnani C, Annesi-Maesano I, Brescianini S, D’Ippolito C, Medda E, Nisticò L et al. Heritability and shared genetic effects of asthma and hay fever: an Italian study of young twins. *Twin Res Hum Genet* 2008;**11**:121–131.
4. Vercelli D. Discovering susceptibility genes for asthma and allergy. *Nat Rev Immunol* 2008;**8**:169–182.
5. Nilsson D, Andiappan AK, Halldén C, Tim CF, Säll T, Wang DY et al. Poor Reproducibility of Allergic Rhinitis SNP Associations. *PLoS One* 2013;**8**. doi:10.1371/journal.pone.0053975
6. Andiappan AK, Wang DY, Anantharaman R, Parate PN, Suri BK, Low HQ et al. Genome-wide association study for Atopy and Allergic Rhinitis in a Singapore Chinese population. *PLoS One* 2011;**6**:1–9.
7. Ramasamy A, Curjuric I, Coin LJ, Kumar A, McArdle WL, Imboden M et al. A genome-wide meta-analysis of genetic variants associated with allergic rhinitis and

grass sensitization and their interaction with birth order. *J Allergy Clin Immunol* 2011;**128**:996–1005.

8. Hinds D a, McMahon G, Kiefer AK, Do CB, Eriksson N, Evans DM et al. A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat Genet* 2013;**45**:907–911.
9. Bønnelykke K, Matheson MC, Pers TH, Granell R, Strachan DP, Alves AC et al. Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. *Nat Genet* 2013;**45**:902–906.
10. Nilsson D, Henmyr V, Halldén C, Säll T, Kull I, Wickman M et al. Replication of genome-wide associations with allergic sensitization and allergic rhinitis. *Allergy* 2014;**69**:1506–1514.
11. Haagerup a, Bjerke T, Schiøtz P. Asthma and atopy—a total genome scan for susceptibility genes. *Allergy* 2002;**57**:680–686.
12. Haagerup A, Bjerke T, Schiøtz PO, Dahl R, Binderup HG, Tan Q et al. Atopic dermatitis - A total genome-scan for susceptibility genes. *Acta Derm Venereol* 2004;**84**:346–352.
13. Haagerup a., Børghlum a. D, Binderup HG, Kruse T a. Fine-scale mapping of type I allergy candidate loci suggests central susceptibility genes on chromosomes 3q, 4q and Xp. *Allergy Eur J Allergy Clin Immunol* 2004;**59**:88–94.
14. Haagerup a, Bjerke T, Schiøtz PO, Binderup HG, Dahl R, Kruse T a. Allergic rhinitis--a total genome-scan for susceptibility genes suggests a locus on chromosome 4q24-q27. *Eur J Hum Genet* 2001;**9**:945–952.
15. Bu LM, Bradley M, Söderhäll C, Wahlgren CF, Kockum I, Nordenskjöld M. Genome-wide linkage analysis of allergic rhinoconjunctivitis in a Swedish population. *Clin Exp Allergy* 2006;**36**:204–210.
16. Dizier M-H, Bouzigon E, Guilleud-Bataille M, Bétard C, Bousquet J, Charpin D et al. Genome screen in the French EGEA study: detection of linked regions shared or not shared by allergic rhinitis and asthma. *Genes Immun* 2005;**6**:95–102.
17. Yokouchi Y, Shibasaki M, Noguchi E, Nakayama J, Ohtsuki T, Kamioka M et al. A genome-wide linkage analysis of orchard grass-sensitive childhood seasonal allergic rhinitis in Japanese families. *Genes Immun* 2002;**3**:9–13.
18. Shen N, Fu Q, Deng Y, Qian X, Zhao J, Kaufman KM et al. Sex-specific association of X-linked Toll-like receptor 7 (TLR7) with male systemic lupus erythematosus. *Proc Natl Acad Sci U S A* 2010;**107**:15838–15843.
19. Saruta M, Targan SR, Mei L, Ippoliti AF, Taylor D, Rotter JI. High-frequency haplotypes in the X chromosome locus TLR8 are associated with both CD and UC in females. *Inflamm Bowel Dis* 2009;**15**:321–327.

20. Davila S, Hibberd ML, Dass RH, Wong HEE, Sahiratmadja E, Bonnard C et al. Genetic association and expression studies indicate a role of Toll-like receptor 8 in pulmonary tuberculosis. *PLoS Genet* 2008;**4**. doi:10.1371/journal.pgen.1000218
21. Oh D-Y, Taube S, Hamouda O, Kücherer C, Poggensee G, Jessen H et al. A functional toll-like receptor 8 variant is associated with HIV disease restriction. *J Infect Dis* 2008;**198**:701–709.
22. Møller-Larsen S, Nyegaard M, Haagerup a, Vestbo J, Kruse T a, Børghlum a D. Association analysis identifies TLR7 and TLR8 as novel risk genes in asthma and related disorders. *Thorax* 2008;**63**:1064–1069.
23. Nilsson D, Andiappan A, Halldén C, De Yun W, Säll T, Tim C et al. Toll-like receptor gene polymorphisms are associated with allergic rhinitis: a case control study. *BMC Med Genet* 2012;**13**:66.
24. Wickman M, Kull I, Pershagen G, Nordvall SL. The BAMSE project: presentation of a prospective longitudinal birth cohort study. *Pediatr Allergy Immunol* 2002;**13 Suppl 1**:11–13.
25. Fuertes E, Söderhäll C, Acevedo N, Becker A, Brauer M, Chan-Yeung M et al. Associations between the 17q21 region and allergic rhinitis in 5 birth cohorts. *J Allergy Clin Immunol* 2015;**135**:573–576.
26. R Development Core Team R. R: A Language and Environment for Statistical Computing. R Found. Stat. Comput. 2011;**1**:409.
27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M a R, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559–575.
28. Gang Z, Jungnam J, Chun Z, Geller NL. Testing Association for Markers on the X Chromosome. *Genet Epidemiol* 2007;**31**:834–843.
29. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 2013;**10**:5–6.
30. Su Z, Marchini J, Donnelly P. HAPGEN2: Simulation of multiple disease SNPs. *Bioinformatics* 2011;**27**:2304–2305.
31. Li H. Tabix: Fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* 2011;**27**:718–719.
32. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE et al. An integrated map of genetic variation from 1 , 092 human genomes. *Nature* 2012;**491**:56–65.
33. Danecek P, Auton A, Abecasis G, Albers C a., Banks E, DePristo M a. et al. The variant call format and VCFtools. *Bioinformatics* 2011;**27**:2156–2158.

34. Sim N-L, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 2012;**40**:W452–W457.
35. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;**7**:248–249.
36. Maloney JA, Bainbridge T, Gustavsson A, Zhang S, Kyauk R, Steiner P et al. Molecular mechanism of Alzheimer disease protection by the A673T allele of amyloid precursor protein. *J. Biol. Chem.* 2014;**45**:30990-31000.
37. Xu X, Hawi Z, Brooks KJ, Anney R, Bellgrove M, Franke B et al. Replication of a rare protective allele in the noradrenaline transporter gene and ADHD. *Am J Med Genet B Neuropsychiatr Genet.* 2008;**8**:1564-1567.

Table 1 *TLR8* single nucleotide variants detected in 288 AR cases and compared with 379 European individuals from the 1000Genomes project

SNP ID	Position	Location	Alleles	AR		EUR		Amino acid change	SIFT/PolyPhen2
				HWE	MAF	HWE	MAF		
rs148539975	12923891	Promoter	T/C	1.0	0.017	1.0	0.026	-	-
rs189513707	12924182	Promoter	C/T	-	0.000	1.0	0.002	-	-
rs5741883	12924221	Promoter	C/T	0.24	0.274	0.0035	0.264	-	-
rs3764879	12924697	Promoter	C/G	0.61	0.177	1.0	0.272	-	-
rs3764880	12924826	Coding	A/G	0.82	0.175	1.0	0.272	M1V	Failed Sift/Benign
rs56194919	12937234	Coding	C/T	1.0	0.002	1.0	0.017	C25C	-
rs2159377	12937513	Coding	C/T	0.17	0.149	0.82	0.171	D118D	-
rs5744080	12937804	Coding	C/T	0.77	0.364	0.90	0.383	H215H	-
Novel	12938331	Coding	T/C	1.0	0.002	-	0.000	M391T	Tolerated/Benign
rs181415824	12938905	Coding	T/C	-	0.000	1.0	0.003	N582N	-
rs2407992	12939112	Coding	G/C	0.56	0.369	0.90	0.390	L651L	-
rs5744083	12939406	Coding	A/G	1.0	0.012	1.0	0.021	K749K	-
rs3747414	12939412	Coding	C/A	0.48	0.322	1.0	0.321	I751I	-
rs143150953	12939415	Coding	C/T	-	0.000	1.0	0.003	N752N	-
Novel	12939469	Coding	C/T	1.0	0.002	-	0.000	H770H	-
Novel*	12939928	Coding	C/T	1.0	0.005	-	0.000	D923D	-

Positions are according to NCBI build 37.3

Alleles are given in the order major/minor

HWE, Hardy-Weinberg equilibrium calculated according to Gang Z, Jungnam J, Chun Z, Geller NL. Testing Association for Markers on the X Chromosome. *Genet Epidemiol* 2007;**31**:834-843.

MAF, minor allele frequency

*, Novel compared to the EUR population of the 1000Genomes project (dbSNP: rs2109135).

Table 2 AR-association analysis of SNVs in *TLR8* in the BAMSE population and comparison of significances in present and previous studies

SNP ID	Position	Location	Alleles	Present study							Previous study*						
				P _(allele)				P _(genotype)	Specific IgE		P _(allele)				P _(genotype)	Skin prick test	
				RA	Combined	Males	Females	Females	P _(birch)	P _(timothy grass)	RA	Combined	Males	Females	Females	P _(birch)	P _(timothy grass)
rs2269809	12910619	Promoter	G/T	-	-	-	-	-	-	-	G	0.58	0.91	0.40	0.028	0.81	0.054
rs5935438	12913022	Promoter	G/C	G	0.19	0.31	0.021	0.054	0.12	0.56	C	0.28	0.99	0.22	0.0059	0.81	0.019
rs178998	12917787	Promoter	C/T	T	0.046	0.80	0.027	0.090	0.029	0.32	C	0.19	0.86	0.080	0.15	0.10	0.23
rs3788935	12922659	Promoter	A/G	G	0.54	0.93	0.40	0.49	0.0058	0.21	A	0.0036	0.36	0.0037	0.012	0.44	0.23
rs3761624	12923681	Promoter	A/G	G	0.53	0.95	0.32	0.37	0.0034	0.17	A	0.0080	0.49	0.0056	0.016	0.40	0.16
rs5741883	12924221	Promoter	C/T	-	-	-	-	-	-	-	T	0.28	0.96	0.17	0.068	0.67	0.038
rs17256081	12926045	Intron	T/C	-	-	-	-	-	-	-	C	0.80	0.64	0.98	0.021	0.76	0.0066
rs4830805	12927759	Intron	G/A	A	0.44	0.96	0.27	0.43	0.023	0.32	G	0.0069	0.25	0.013	0.042	0.34	0.080
rs1548731	12927947	Intron	C/T	T	0.11	0.54	0.025	0.0017	0.35	0.14	T	0.50	0.89	0.48	0.046	0.94	0.49
rs2407992	12939112	Intron	G/C	C	0.090	0.70	0.086	0.18	0.28	0.090	G	0.63	0.75	0.40	0.67	0.42	0.040

Positions are according to NCBI build 37.3

Alleles are given in the order major/minor

RA, risk allele

Significant associations are given in bold

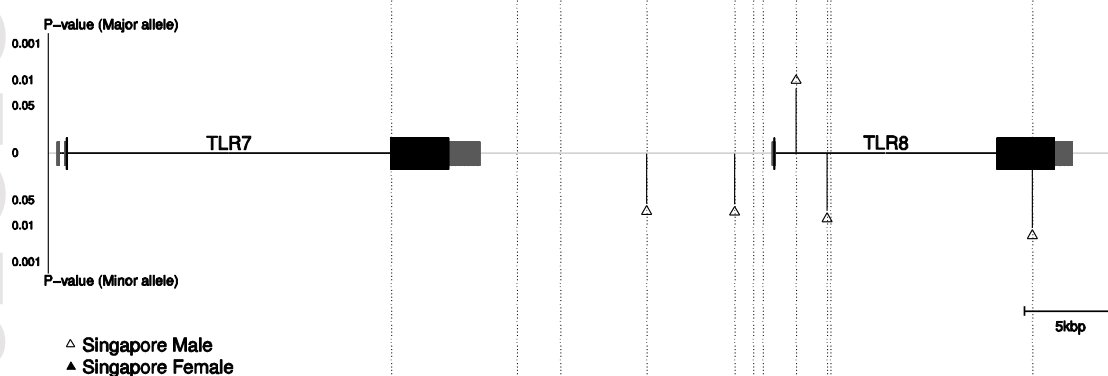
* Nilsson et al. 2012

Table 3 Permutation tests of observed number of significant p-values for the Malmö, Bamse and Singapore populations

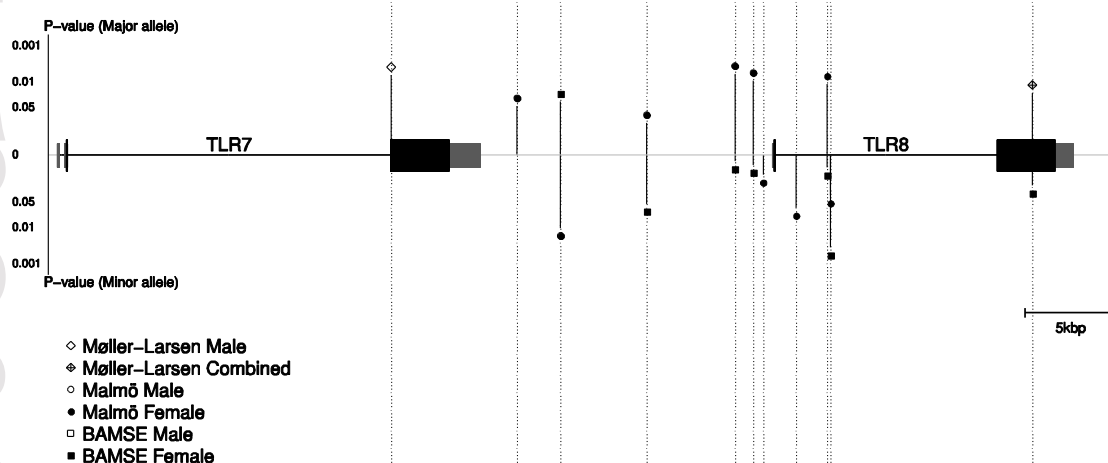
Test	Malmö		Bamse		Singapore	
	No. of P-values < 0.05	P(sim)	No. of P-values < 0.05	P(sim)	No. of P-values < 0.05	P(sim)
Allele all	3	0.19	1	0.20	1	0.47
Allele females	3	0.18	3	0.044	0	-
Allele males	0	-	0	-	5	0.051
Genotype	4	0.080	1	0.21	0	-
All tests*	3,3,4	0.034	1,3,1	0.024	1,5	0.027

*, require that all of the tests above show an equal or greater number of significant p-values for each category

A



B



C

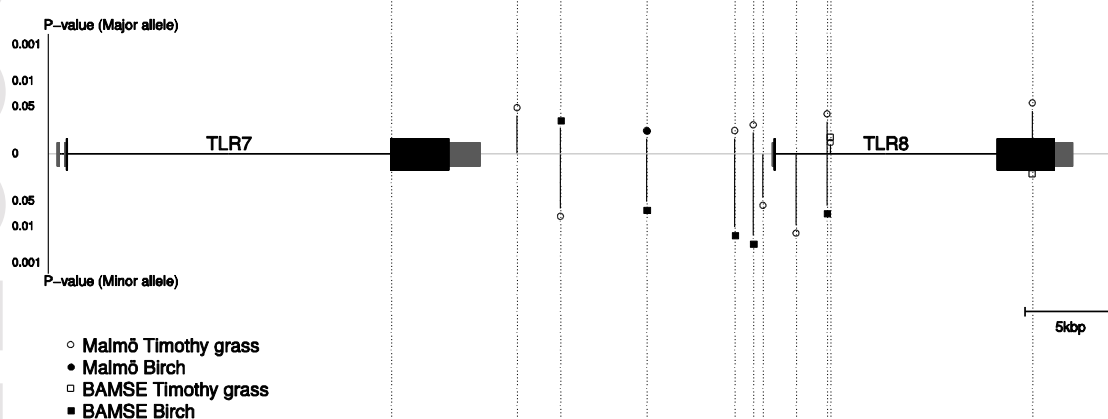


Table S1. Sanger sequencing primers for TLR8 promotor and exons

Primer ID	Position*	Sequence	Amplicon size (bp)	Annealing temp (°C)
1b_F1020	12924258-12924277	5'-AAGATGCTCGCTTTGCGGCC-3'	739	66
1b_R1758	12924975-12924996	5'-AGCAAGCCGCTTTACCTGCATT-3'		
2b_F382	12923620-12923639	5'-GCCACCCCTGCTGCCAATTT-3'	888	66
2b_R1269	12924483-12924507	5'-TGTGTGTACAGGGGAGAAAAAGGGA-3'		
7_F192	12928216-12928240	5'-GTGACAAATCCTACCTGCATAACCA-3'	631	57
7_R822	12928827-12928846	5'-TGCTGATGGCCACCAACAC-3'		
9_F109	12937071-12937090	5'-ACAAGTATGGGGCAGCGCTG-3'	530	63
9_R638	12937579-12937600	5'-AGACTCTGGCAAACCAGAGGGT-3'		
12_F549	12937511-12937530	5'-GACGGGGCATTCTCAACCT-3'	663	60
12_R1211	12938154-12938173	5'-GCGGGGCAGCATCGTTAAAA-3'		
14_F1127	12938089-12938109	5'-GCCTCATCTGAAGGTGCTGGA-3'	519	63
14_R1645	12938587-12938607	5'-CCATAAGCAGCACATTGTGGC-3'		
15_F1549	12938511-12938530	5'-TCCGGAAACGACGCTCAACA-3'	720	63
15_R2268	12939210-12939230	5'-CGAGACGAGGAACTGCTGGA-3'		
17_F2042	12939004-12939023	5'-CAGTGGCAATCGCCTTGACA-3'	581	63
17_R2622	12939565-12939584	5'-GATCCCAGGACTGGCACAA-3'		
20_F2369	12939331-12939350	5'-TTCCACCTACCCTCTGGCT-3'	544	60
20_R2912	12939855-12939874	5'-AAGGTGGTAGCGCAGCTCAT-3'		
22_F2720	12939682-12939701	5'-GGTTATGTTGGCTGCCCTGG-3'	501	60
22_R3220	12940164-12940182	5'-GGGTTGTCAGGCCACTGGA-3'		
24_F3175	12940137-12940156	5'-GGCAGCGGATCTGTAAGAGC-3'	385	60
24_R3559	12940500-12940521	5'-CCAGTGATGCCTATGTCTCTGG-3'		

* According to GRCh37

Table S2. Long-range PCR primers

Primer ID	Position*	Sequence	Amplicon size (bp)	Annealing temp (°C)
2cF	12916521-12916540	5'-ATGGACACCTTGGGCGGGGT-3'	4041	66
2cR	12920540-12920561	5'-TGAGGAGGAAAAGCCGCAGGA-3'		
7F	12919002-12919024	5'-AGCACTTTGGTGCCTATAGGGCA-3'	9844	63
7R	12928825-12928845	5'-GCTGATGGCCACCAACACGT-3'		
3bF	12919756-12919775	5'-TCAGGAAGCCCGAGATATTG-3'	6863	68
3bR	12926599-12926618	5'-GCGAAGGCTAACAACCTCTG-3'		
4bF	12925446-12925465	5'-CATAGCAGGTGGTCCGTTTT-3'	6862	68
4bR	12932287-12932307	5'-GGAGACCCATTTACCCTAACG-3'		
5bF	12930575-12930594	5'-CATGGTCACACAGCACTCCT-3'	6946	68
5bR	12937501-12937520	5'-ATGCCCCGTCTGTGATATTC-3'		
6bF	12934456-12934475	5'-GGCCACCCACTAGCTAATCA-3'	6623	68
6bR	12941059-12941078	5'-ATAAAACCGAACGAACCAC-3'		

* According to GRCh37

Table S3. Minor allele frequencies of genotyped SNPs in BAMSE and Malmö populations

SNP ID	Chromosome	Position	Location	Major/minor	Cases Malmö			Controls Malmö			Cases BAMSE			Controls BAMSE		
					Females	Males	Total	Females	Males	Total	Females	Males	Total	Females	Males	Total
rs2269809	X	12910619	Intergenic_7/8	G/T	0.45	0.54	0.48	0.48	0.54	0.51						
rs5935438	X	12913022	Intergenic_7/8	G/C	0.51	0.43	0.48	0.45	0.43	0.44	0.37	0.43	0.39	0.44	0.39	0.43
rs178998	X	12917787	Intergenic_7/8	C/T	0.30	0.35	0.32	0.36	0.34	0.35	0.44	0.42	0.44	0.38	0.41	0.39
rs3788935	X	12922659	Intergenic_7/8	A/G	0.16	0.22	0.18	0.24	0.26	0.25	0.28	0.24	0.26	0.25	0.23	0.25
rs3761624	X	12923681	Intergenic_7/8	A/G	0.16	0.22	0.18	0.24	0.25	0.24	0.28	0.23	0.26	0.25	0.23	0.24
rs5741883	X	12924221	Intergenic_7/8	C/T	0.27	0.24	0.26	0.23	0.25	0.24						
rs17256081	X	12926045	TLR8_Intron1	T/C	0.53	0.49	0.52	0.47	0.47	0.47						
rs4830805	X	12927759	TLR8_Intron1	G/A	0.15	0.19	0.17	0.22	0.23	0.23	0.26	0.21	0.24	0.23	0.20	0.22
rs1548731	X	12927947	TLR8_Intron1	C/T	0.29	0.27	0.28	0.27	0.26	0.27	0.32	0.31	0.32	0.26	0.33	0.28
rs2407992	X	12939112	TLR8_Exon2	G/C	0.36	0.40	0.38	0.39	0.38	0.39	0.46	0.45	0.45	0.40	0.43	0.41

Text S1. Description of study outline.

- 1) Sanger resequencing for single nucleotide polymorphisms (SNPs) in 288 patients from the Malmö population.
- 2) Screening for copy number variations (CNVs) using long-range PCR in 288 patients from the Malmö population.
- 3) Comparison of resequencing results to those of 379 European subjects from the EUR-population of 1000Genomes.
- 4) Test of an accumulation of rare variants in the Malmö population by comparison with the non-Finnish European ExAC population.
- 5) Association analysis of 7 SNPs in the BAMSE cohort with 422 cases and 859 controls.
- 6) Test for accumulation of significances in previous association results from the Malmö and Singapore populations in Nilsson *et al* (2012), and in present association results from the BAMSE population.