2013-03-25

Age-dependent variation of genotypes in MHC II transactivator gene (CIITA) in controls and association to type 1 diabetes

Gyllenberg, Alexandra; Asad, Samina; Piehl, Fredrik; Swanberg, Maria; Padyukov, Leonid; van Yserloo, Brian; Graham, Jinko; Rutledge, Elizabeth; Orho-Melander, Marju; Lindholm, Eero; Graff, Caroline; Forsell, Charlotte; Åkesson, Karin; Landin-Olsson, Mona; Carlsson, Annelie; Forsander, Gun; Ivarsson, Sten; Larsson, Helena; Alfredsson, Lars; Åkesson, Kristina; Olsson, Tomas; Kockum, Ingrid; Lindblad, Bengt; Ludvigsson, Johnny; Marcus, Claude; Lernmark, Åke; McNeney, Brad; Swedish Childhood Diabetes Study Group; Diabetes Incidence in Sweden Study Group; Better Diabetes Diagnosis Study Group

Genes Immun. 2012 Dec;13(8):632-40.
Springer Nature
http://doi.org/10.1038/gene.2012.44
http://hdl.handle.net/10616/41464

Age Dependent Variation of Genotypes in MHCII Transactivator gene (CIITA) in Controls and Association to Type 1 Diabetes

Short title: Age dependent variation in CIITA

Alexandra Gyllenberg[1], Samina Asad[1], Fredrik Piehl[1], Maria Swanberg[2], Leonid Padyukov[3], Brian Van Yserloo[4], Elizabeth A. Rutledge[5], Brad McNeney[6], Jinko Graham[6], Marju Orho-Melander[7], Eero Lindholm[7], Caroline Graff[8,9], Charlotte Forsell[8,9], Karin Åkesson[10], Mona Landin-Olsson[11], Annelie Carlsson[12], Gun Forsander[13], Sten A. Ivarsson[14], Helena Larsson[14], Bengt Lindblad[15], Johnny Ludvigsson[16], Claude Marcus[17], Åke Lernmark[14,18], Lars Alfredsson[19], Kristina Åkesson[20,21], Tomas Olsson[1], Ingrid Kockum[1]

for the Swedish Childhood Diabetes Study Group[*], the Diabetes Incidence in Sweden Study Group[*], and the Better Diabetes Diagnosis Study group[*].

1. Neuroimmunology Unit, Dep. of Clinical Neuroscience, Karolinska Institutet, Stockholm Sweden.
2. Neurodegeneration and Inflammation Genetics Unit, Dep. of Experimental Medical Science, Lund University, Sweden

3. Rheumatology Unit, Dep. of Medicine, Karolinska Institutet, Stockholm Sweden

4. Molecular and Genetics Core, Diabetes Endocrinology Research Center, University of Washington, Seattle, WA, United States

5. Salish Kootenai College, Pablo, MT, United States

6. Dep. of Statistics and Actuarial Science, Simon Fraser University, Burnaby, Canada.

7. Diabetes and Cardiovascular Disease, Genetic Epidemiology, Dep. of Clinical Sciences, Skåne University Hospital, Lund, Sweden

8. Dep. of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

9. Dep. of Geriatric Medicine, Karolinska University Hospital Huddinge, Stockholm, Sweden

10. Dep. of Pediatrics, Ryhov County Hospital, Jönköping, Sweden

11. Dep. of Medicine, Skåne University Hospital, Lund, Sweden.

12. Dep. of Pediatrics, Skåne University Hospital, Lund, Sweden

13. Dep. of Pediatrics, The Queen Silvia Children's Hospital, Sahlgrenska/Ostra University Hospital, Göteborg, Sweden

14. Dep. of Clinical Sciences, Skåne University Hospital, Malmö, Sweden

15. Dep. of Pediatrics, Institute of Clinical Science, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

16. Division of Pediatrics and Diabetes Research Center, Dep. of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

17. Division of Pediatrics, Dep. of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden

18. Dep. of Medicine, University of Washington, Seattle, WA. United States

19. Dep. of Environmental Medicine, Karolinska Institutet, Stockholm Sweden.

20. Clinical and Molecular Osteoporosis Research Unit, Dep. of Clinical Sciences, Lund University, Sweden

21. Dep. of Orthopaedics, Skåne University Hospital, Malmö, Sweden

[*] Members of the Swedish Childhood Diabetes Study Group, the Diabetes Incidence in Sweden Study Group and the Better Diabetes Diagnosis Study group are listed in supplementary material.

Corresponding author:

Alexandra Gyllenberg

Dep. of Clinical Neurosciences

Karolinska Insitutet

Centre for Molecular Medicine, L8:05

Karolinska University Hospital

S171 76 Stockholm

Sweden

Tel: +46-8-51776258

Fax: +46-8-51770885

e-mail: alexandra.gyllenberg@ki.se

Abstract

The *Major Histocompatibility Complex Class II Transactivator (CIITA)* gene (16p13) has been reported to associate with susceptibility to multiple sclerosis, rheumatoid arthritis and myocardial infarction, recently also to celiac disease at genome wide level. However, attempts to replicate association have been inconclusive. Previously, we have observed linkage to the *CIITA* region in Scandinavian type 1 diabetes families.

Here we analyze five Swedish type 1 diabetes cohorts and a combined control material from previous studies of *CIITA*. We investigate how the genotype distribution within the *CIITA* gene varies depending on age, and the association to type 1 diabetes.

Unexpectedly we find a significant difference in the genotype distribution for markers in *CIITA* (rs11074932, $p=4x10^{-5}$ and rs3087456, p=0.05) with respect to age, in the collected control material. This observation is replicated in an independent cohort material of about 2000 individuals (p=0.006, p=0.007).

We also detect association to type 1 diabetes for both markers, rs11074932 (p=0.004) and rs3087456 (p=0.001) after adjusting for age at sampling. The association remains independent of the adjacent type 1 diabetes risk gene *CLEC16A*. Our results indicate an age dependent variation in *CIITA* allele frequencies, a finding of relevance for the contrasting outcomes of previously published association studies.

Keywords

Type 1 diabetes (T1D); CIITA; autoimmunity; association, age

Introduction

Type 1 diabetes (T1D) is a multifactorial disease where a number of genes are thought to be involved in regulating disease susceptibility. The major histocompatibility complex (MHC) is

known to be the major T1D susceptibility locus, accounting for approximately half of the genetic susceptibility to T1D [1]. The MHC region is thought to contain several susceptibility genes with two of the major ones being *DRB1* and *DQB1*[1]in the MHC class II (MHCII) region. In the Caucasian population, there are two major susceptibility haplotypes for T1D, *DRB1*03-DQA1*05:01-DQB1*02:01* and *DRB1*04-DQA1*03:01-DQB1*03:02*, and one protective haplotype, *DRB1*15-DQA1*01:02-DQB1*06:02* [2, 3].

The regulation of the MHCII genes is mainly at the transcriptional level, and one of the crucial factors is the class II transactivator, encoded by the *CIITA* gene (16p13). The CIITA protein is a non-DNA binding co-activator which acts as a platform for the assembly of transcription factors that bind to MHC II promoters and control transcription [4, 5]. *CIITA* is regarded as the master control factor for the expression of MHCII genes [6], and the lack of *CIITA* expression leads to an almost complete lack of MHCII expression [7]. Four independent and cell type specific *CIITA* promoters (PI-PIV) have been identified in humans [8]. Promoter I (PI) mainly controls *CIITA* expression in myeloid dendritic cells and macrophages, while PIII is active in B-cells, activated T cells and plasmacytoid dendritic cells and the PIV promoter regulates IFNγ-inducible *CIITA* expression in cells of non-hematopoetic origin and in thymic epithelium [9]. The function of the PII promoter in humans has not yet been fully characterised [8].

Since *CIITA* has a unique role in the control of MHC II expression and the MHC II locus is the major genetic determinant for susceptibility to autoimmune diseases, *CIITA* is an interesting candidate gene in the study of autoimmune diseases like T1D [10]. Previously, we found genome-wide significant linkage on chromosome 16 in the region of *CIITA* to T1D (LOD=3.7), among T1D patients who also carry the *DRB1*03* and *DRB1*04* alleles [11]. The HLA association to T1D is also know to vary with age at onset, such that the association is stronger in younger patients compare to older [3, 12, 13] .

Genome wide significant association to celiac disease has recently been reported for markers in the *CIITA* gene[14]. In addition, increased susceptibility to myocardial infarction (MI), rheumatoid arthritis (RA) and multiple sclerosis (MS) has been demonstrated for a polymorphism (rs3087456) in the 5´ region of type III *CIITA* [15]. However, this association has not always been replicated in later studies, and the outcome of the analysis is varying depending on which control group that has been used.

In this study, we test the hypothesis that *CIITA* is a T1D susceptibility gene. In addition, we investigate if the allele distribution in the gene varies depending on age, and how this may affect the evidence for association.

Results

Primary association to T1D in DISS2

The tag SNPs described under methods were first genotyped in the DISS2 material, a T1D cohort consisting of DNA samples from 586 incident T1D patients and 836 controls [16]. Association was found in the area between PI and PIII. Five markers were added upstream of the PI promoter to better define the associated region in *CIITA*. One extra marker was also added in the *CLEC16A* gene, which is a known T1D susceptibility gene which maps close to *CIITA* [17], in order to exclude that association to T1D in the *CIITA* gene was due to LD with this gene. The initial analyses indicated that several of the SNPs were associated to T1D with the most significant association being to rs11074932 *(S1)*. After correcting for multiple testing, only rs11074932 remained significant.

Age stratification

Due to the inconclusive results when trying to replicate earlier association in the area, and different outcome depending on what control group that was used, we wanted to investigate if

the age among the controls could have an effect on the association. The same three markers (rs11074932, rs3087456 and rs4774) used in the combined T1D study below was investigated. Control cohorts were collected as described in the methods section, and a test for trend showed that there is a significant trend where the frequency of the genotypes changes over age for two of the markers (rs11074932, $p=4x10^{-5}$; rs3087456, p=0.05) among the controls *(fig1)*. For both markers the frequency of the major allele homozygote is increasing with age. The rs3087456 marker is significant only when the oldest age group (over 75 years of age) is included in the analysis. It should be noted that this group consist of several autopsy cases from a brain bank originally included in the Alzheimer's disease control group. No age dependent effect was evident for the rs4774 marker.

Replication of age variation in an independent population based cohort material

The same markers in the *CIITA* gene were investigated in two cohorts of 25- and 75 year old women (PEAK and OPRA), respectively. Here we could confirm the significant variation in genotype for marker rs11074932 (p=0.006) and rs3087456 (p=0.007) with the minor allele homozygote genotype frequency decreasing with age. Additionally, an earlier undiscovered variation could also be seen for the rs4774 marker (p=0.03), but here the minor allele is instead increasing with age *(fig2 and table 1).*

Age corrected association to type 1 diabetes in combined material

To increase power to detect association to T1D we combined cohorts from five T1D case-control studies and corrected for age at sampling using logistic regression. Due to the big discrepancies in age between the T1D cases and the other control cohorts, only T1D cohorts were included in this analysis.

Four SNPs (rs11074930, rs11074932, rs3087456 and rs4774) were chosen on the basis that they either had remained significant after correcting for age by matching each case to two controls in an initial analysis in a smaller (DISS2) cohort (rs11074930, rs11074932) *(S1)*, or been found associated to autoimmune disease in earlier publications (rs3087456, rs4774) [15, 18-20]. Rs11074930 was found to deviate from Hardy-Weinberg equilibrium in a section of the control population and was discarded from further studies. Two markers showed some association (rs11074932 p=0.004 and rs3087456 p=0.001) *(table 2*) after correcting for age in the combined cohorts of T1D.

An age-stratified meta-analysis was performed for the associated markers to get an overview of the influence of different age-groups and clarify the association. Only individuals under 40 years of age were included in this analysis since the majority of the patient falls into this group. We find that for both rs11074932 and rs3087456, heterozygotes are negatively associated, while major allele homozygotes are positively associated with T1D *(fig3).*The association of *CIITA* SNPs to T1D was independent of *CLEC16A*, a nearby gene which is an established T1D locus [21], as association to *CIITA* SNPs remained after adding *CLEC16A* marker rs12708716 to the logistic regression model in the DISS2 cohort. In addition, an LD plot of the area was made in the DISS2 cohort demonstrating that the rs12708716 SNP in *CLEC16A* is not in LD with any of our typed SNPs *(fig4).* We also tested whether gender would affect the association, and added gender as a covariate in the logistic regression model. This did not alter the association.

Interaction analysis with HLA

We also investigated if there was an interaction between T1D HLA risk alleles and *CIITA*. Since *CIITA* is a key protein in the control of expression of MHC class II alleles and we have previous findings of increased linkage in *DRB1*03/DRB1*04* positive patients in this region

of chromosome 16 [11], we wanted to further investigate the role of interaction between *CIITA* and HLA *DRB1*. Interaction analyses on both the additive and multiplicative scale was performed. These analyses were done for presence of minor allele compared to major allele homozygotes. For both associated markers (rs11074932 and rs3087456) the results showed that there is a significant additive interaction for major allele homozygotes and absence of *DRB1*15* in relation to T1D, but not in any of the other HLA types. There was no interaction on the multiplicative scale (data not shown). This means that individuals risk for T1D associated with the joint lack of the protective *DRB1*15* allele and presence of major allele homozygotes for any of these two SNPs are greater than the sum of the risk associated with *DRB1*15* and *CIITA* SNPs individually *(fig5)*. The analysis shows that the association is depending on the absence of *DRB1*15*, which is not unexpected since the majority of T1D cases are *DRB1*15* negative and *DRB1*15* is known as a strong protective factor for T1D. The proportion of T1D is higher among DR15 negative individuals for major allele positive compared to major allele negatives for both markers ($p<0.001$ for rs11074932 and $p<0.0003$ for rs3087456) while no such difference is seen among DR15 positive individuals. This supports our conclusion that the CIITA association seem to be among DR15 negative individuals.

Since the association to T1D remained for two SNP markers (rs11074932 and rs3087456) even after controlling for age, gender and HLA, we suggest that these polymorphisms or other polymorphisms in close proximity are related to T1D susceptibility.

Discussion

Given the pivotal role of *CIITA* for regulation of MHC class II gene transcription [6], it is an obvious candidate for affecting susceptibility to autoimmune diseases known to be associated

to MHC class II haplotypes[22]. Previous studies have reported positive [19, 20] but also negative [23, 24] results regarding association of *CIITA* to different autoimmune diseases.

In this study, we have performed LD mapping of the *CIITA* gene followed by an analysis of association to T1D. We identified a region extending between the PI and PIII promoters of the *CIITA* gene, which is modestly associated to T1D. This association remains significant after correcting for age, gender, HLA association and association to the neighboring *CLEC16A* gene, which is a nearby established T1D risk gene [21]. In a genome-wide association (GWA) scan performed by The Wellcome Trust Case Control Consortium (WTCCC) [17]the 16p13 area was found to be significant associated to T1D. The *CLEC16A* gene has been thoroughly evaluated and found to be the major associated T1D risk gene in the area, and independent of the *CIITA* gene [25]. However this doesn't rule out the possibility that *CIITA* can have a minor role as well at a more modest level. In the mentioned investigation of the area there were inconclusive results regarding the association of CIITA, when significant association to T1D was found in a family material but not in the case-control cohort [25]. Our findings regarding the association of the *CIITA* gene to T1D is not genome wide significant, but in the complex settings of the T1D many genes with moderate influence are believed to have an impact on the etiology of the disease. This association between CIITA and T1D does however need to be confirmed in other materials to remain of interest.

We have in our study observed a significant difference in the genotype distribution for *CIITA* SNPs in a control cohort across different age groups, which we also confirm in an independent material of approximately 2000 individuals. The change in genotype frequency with age among controls in this gene can be one reason for the conflicting results of previously published *CIITA* case-control association studies. Thus, if one genotype or allele is

less abundant among older control individuals, it could give a false association result for that genotype when compared with a younger patient group.

Possibly, this phenomenon may be relevant also for other genomic regions, and therefore suggest that proper matching of cases and controls with respect to age are important in genetic studies. This effect may not be evident when only analyzing an age restricted sample material. For example, no age effect was found among the T1D patients. However, with most of the T1D patients having an onset before the age of 30, they cover only part of the age span included in the investigation. The variation for rs4774 with age was not discovered in our initial combined control material, where only 18 individuals over the age of 75 were genotyped for this marker, but seen in the cohort study including about 1000 75 year old individuals. Possibly the effect for this marker is more profound for the oldest age groups, and the same tendency is seen for the rs3087456 marker, where the significance level improves when adding additional 75 year- or older individuals to the analysis. Although the discrepancies in results between association studies of the *CIITA* gene in part can be due to lack of statistical power to detect a small effect size, variability in outcome could also be affected by differences in age between patients and controls, a variable which often is not accounted for in genetic case-control studies.

There can be different underlying reason for an age effect on allele frequencies among the controls, which we will address in this discussion.

One reason could be population stratification, where the genotype frequencies vary among different population cohorts. Detailed data on ethnic origin was not available for all the cohorts included in this investigation, but when available, individuals with non-Scandinavian origin have been removed. The rs11074932 SNP is the marker varying most across the different age groups, with the *TT* genotype increasing with age at the expense of the other genotypes. The minor allele in the European and African populations (*C*) is the major allele

in Japanese and Chinese populations. The distribution is similar for marker rs3087456, where the minor allele (*G*) in Europeans is the major allele in Japanese and African populations, but not much information about population diversity regarding genotypes for these two markers is available. Theoretically infiltration of these populations could affect our findings. However, since the controls used in this study were all collected from Swedish residents and individuals with non-Scandinavian descent were removed, no or very few Japanese/Chinese or African individuals have to our knowledge been included. Also, one could argue that these are only minor immigrant groups in Sweden, and their participation in study cohorts like the one described here are likely to be small. Hence, the variation we see in our material is unlikely to depend solely on population heterogeneity.

A maybe more reasonable and interesting theory concerns how the genotype affects health. It does not have to be a dramatic effect on longevity or survival, a small influence on the severity or recurrence of a common cold could be enough. If a certain genotype is associated with being healthier it might be more likely that these individuals are included as healthy controls for a medical study. Such an effect may be more pronounced for older individuals and will result in a skewing of genotypes in the study cohorts. Blood-donors are generally regarded as healthier than the general population. When we investigated the blood-donor control group from the MS cohort 1 (n=1217), we found that they have a lower minor allele frequency for rs3087456 (0.22) than the combined control material together (0.24). Similarly, the minor allele frequency for rs11074932 was 0.27 among blood donors compared to 0.28 in the whole material, regardless of age. This correlates with the findings that the older age groups in this study had a lower frequency of the minor allele homozygote genotype than did the younger groups.

The minor allele of rs3087456 has been associated with lower expression of MHC class II chains after stimulation of leukocytes with interferon-γ [15]. Possibly, a higher frequency of the

minor allele for these markers leads to less expression of the *CIITA* gene and subsequently less MHCII molecules, which in turn could lead to less effective antigen presentation and immune defense. If this also leads to more severe or more occurrence of infections, it will affect the possibility for those individuals to become a blood-donor or participate in a clinical study as a healthy control, maybe even more so if you are older. Theoretically, healthy older individuals in this study and blood donors are then expected to have increased expression of *CIITA* and therefore likely a more efficient clearance of infections.

A perhaps less likely explanation is that an infection or other environmental influence cause selection pressure for a certain genotype in the whole population, affecting the individual chance of survival. For this to occur, the infection in question must have conferred a high rate of mortality in young individuals, a scenario similar to what occurred due the Spanish flu epidemic.

Another reason for the variation of genotypes could be events that affect genotyping results. Since the genotyping of the different cohorts in this study has been performed with different methods we had in some cases overlapping results for the same individuals from different genotyping platforms (RA cohort, DASH and Taqman technology). We could not find any major discrepancy in genotyping results in this test sample (one mismatch out of 664 samples). Also, the RA cohort was re-genotyped with Taqman technology with a 100% concordance in results. It could be considered advantageous that different methods have been used, since it minimizes the effect of an introduced systematic error. All genotyping platforms used are well established and described.

Whatever the cause is, it is important to consider the effect it has on association studies and thoroughly choose well-matched controls in these cases. This variance could clearly affect the results of an association study that doesn't consider the age among the control group compared to the patient cohort.

The association found to T1D in this study is of moderate strength and not genome wide significant. However the importance of the *CIITA* gene's function in the immune system makes it plausible that it can have an effect on the pathogenesis of T1D. The two markers (rs3087456 and rs11074932) investigated here are in quite high linkage disequilibrium, 76% (r^2), with each other *(fig4)*, and for both markers it is the major allele homozygote that is associated to T1D. Since there are earlier findings of lower expression of MHC class II chains after stimulation of leukocytes with interferon-γ [15] for the minor allele of rs3087456, it is possible that a lower expression of *CIITA* and subsequently MHCII molecules could be protective to T1D in the way of less efficient self-antigen presentation, and higher or "normal" levels of *CIITA* correlating to the major allele genotype are involved in susceptibility. This correlates with our theory regarding the health aspect described above. Further functional studies are needed to investigate how different alleles may affect the expression of MHC class II in the setting of T1D.

In summary, we report an age dependent variation in SNPs located in the *CIITA* gene among control materials. In addition, we demonstrate a significant, albeit modest association to T1D for SNPs located in the genomic region of the PI and PIII promoters of the *CIITA* gene after controlling for age. Taken together these results suggest that replication of this association and other association studies in case-control materials should be performed with careful matching for age.

Materials and Methods

Ethics Statement

All included cohorts, patient material and analyses in this study were approved by the Regional Ethical Review Boards in the citites of Stockholm, Lund and Umeå in Sweden

(www.epn.se). Informed consent from all study participants or their parents was obtained. Investigations were carried out according to guidelines from the Declaration of Helsinki.

Subjects: Type 1 diabetes patients and controls

Diabetes Incidence Study in Sweden 1 (DISS1)

The DISS1 cohort consists of DNA samples from 839 T1D patients and 625 sex, age and residence matched controls. Blood was collected from incidence patients from the Diabetes Incidence Study in Sweden (DISS) registry, diagnosed between 1987 and 1989, at the age of 15-36 years [26]. 667 of the patients were classified with T1D by the treating physician. Sufficient DNA was available for 431 T1D cases and 348 controls in this cohort, these were the individuals included in the current investigation.

Diabetes Incidence Study in Sweden 2 (DISS2)

The DISS2 cohort consists of DNA samples from 778 incident diabetes patients aged 15-36 years and from the DISS registry during 1992 and 1993, and 836 sex and age- matched controls [16]. 586 of the patients were classified with T1D at follow-up and these subjects are included in this study.

Swedish Childhood Study (Sv2)

497 cases of children between 0-14 years with newly diagnosed T1D collected from the Swedish Childhood registry. Controls were geographically, gender and age matched to all cases above 7 years of age (n=423). For patients under the age of 7 years a control was selected among patients being treated at the hospital for reasons other than T1D (n=53) [27].

Better Diabetes Diagnosis Study (BDD)

2700 incident diabetes patients under the age of 18 years at diagnosis, collected between 2005 and 2009 from 40 pediatric clinics in Sweden for the Better Diabetes Diagnosis study[28].

Only patients diagnosed with T1D and with all parents and grandparents born in Scandinavia are included in this study.

Diabetes Registry in Southern Sweden (DR)

804 T1D patients, 436 men and 368 women, with onset age between 1 and 75 years of age, from the Diabetes Registry in Southern Sweden, all enlisted at the Dep. of Endocrinology at Malmö University Hospital, Sweden and collected between 1996 and 2005. Additionally, 2312 healthy controls, 1695 men and 617 women between 45 and 75 years of age [29].

Individuals of known non-Scandinavian origin were excluded (n=100).

Due to risk for overlap among patients in the DISS2 and DR cohorts, those individuals that could possibly occur in both cohorts were identified and removed from this study (n=73). To our knowledge there are no other cohorts in this study that could be overlapping.

Extra control cohorts:

Rheumatoid arthritis (RA)

1426 healthy controls matched to RA patients by age, sex and residential area. The recruitment of affected individuals and controls was described previously in connection with EIRA study [30]. 373 of these controls were used in the SNP tagging analysis.

Individuals deviating in a PCA analysis or of known non-Scandinavian origin were excluded from this study (n=93) [31].

Multiple sclerosis (MS)

From the Swedish MS 1 cohort, the control group consisted of 1215 healthy blood donors originating from Sweden or other Nordic countries [32].

From the Swedish MS 2 cohort, we used 663 controls matched for age, sex, and residential location to newly diagnosed MS cases resident throughout Sweden in the EIMS studies [33].

All individuals of known non-Scandinavian origin were excluded from the current study (n=190).

Myocardial infarction – SCARF (MI)

From the SCARF[15] study the control group consists of 387 sex- and age-matched healthy persons between 40-60 years of age, and recruited from the general population of the same county as cases with MI, of self-reported Caucasian origin.

Alzheimer's disease – SNAC (AD)

424 healthy controls of 60-73 years of age, from an earlier study of AD [34], randomly selected and sex-matched from 3500 individuals included in the longitudinal study: The Swedish National Study on Aging and Care in Kungsholmen (SNACK), in Stockholm, Sweden. Also, originally added to this cohort are 39 individuals which are autopsy cases from the Karolinska Brainbank who died from cardiovascular or malignant diseases, between 56-91 years of age and without a medical history of dementia [34].

Population based control cohorts from Osteoporosis study (PEAK-25 and OPRA)

The PEAK-25 cohort consists of 1005 healthy women of Swedish or Northeuropean ancestry. The women are all 25 years old and randomly selected from the Malmö city files between 1999 and 2003.[35]

The second cohort consists of 1010 healthy controls from the Malmö Osteoporosis Prospective Risk Assessment (OPRA) study, all aged 75 years and of Swedish or Northeuropean ancestry, randomly selected from the Malmö city files between 1995 and 1999[36].

Additional information regarding the cohorts is available as a supplementary table *(S3)*.

Genotyping methods

In the DISS2 cohort the allelic discrimination method for SNPs (single nucleotide polymorphisms) TaqMan ABI 7900 (Applied Biosystems, Inc ABI, Sweden) [37] was used for all markers except for rs4774, rs3087456 and rs8052709 for which the DASH[38] method was used.

SNP genotyping for DISS1 and SV-2 studies was performed using the MassArray chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (Sequenom Inc., San Diego, CA, USA) using the HME chemistry as described [39].

T1D samples and controls from Diabetes registry in Southern Sweden (DR) and the BDD cohort were genotyped with the TaqMan method.

In the gathered control material from RA, MS and MI, samples was genotyped as previously described [15]. Additional samples from these cohorts were genotyped with the TaqMan method. The controls from the AD cohort were genotyped with DASH.

In the PEAK-25 and OPRA cohorts, markers rs4774 and rs3087456 was genotyped with IPLEX (Sequenom Inc., San Diego, CA, USA) and rs11074932 was genotyped with Taqman.

HLA typing

HLA typing in the different cohorts were performed as follow;

DISS1 and SV2: restriction fragment-length polymorphism (RFLP) was used for DR typing, and genotyping for *DQB1*, *DQA1* and *DRB1* was performed with PCR amplification followed by dot blot hybridizations [13].

DISS2: HLA genotyping for *DQB1*, *DQA1* and *DRB1* was performed with PCR amplification followed by dot blot hybridizations and by RFLP as previously described [13], except that allele-specific PCR amplification (PCR-SSP) of *DRB1* alleles was also used [40]

BDD: A method based on an asymmetrical PCR and a subsequent hybridization of allele-specific probes was used, as described previously[41]. Established haplotypes in the European

population were used to determine DR genotypes in the BDD cohort where only *DQA1* and *DQB1* were genotyped.

MS and RA: Genotyping by allele specific amplification as described earlier [40].

The other cohorts had no data concerning HLA status.

Tag SNPs and haplotype blocks

41 SNP markers were selected from dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) to be evenly spread across 66428 bp of the *CIITA* gene (before PI to 3' UTR). Previously validated SNPs were chosen preferentially. The average marker spacing was 1468 bp (range 86-4030 bp). All SNPs were first genotyped in a selected cohort of 373 controls. 13 of the selected SNPs were non-polymorphic and one did not follow Hardy-Weinberg equilibrium and were discarded from further studies (*S1*).

Genotypes from the remaining 28 SNPs were used to analyze linkage disequilibrium (LD) block structure and to identify haplotype tagging SNPs (htSNPs). The results from this analysis revealed five LD blocks (*S2*), and common haplotypes in each block could be resolved by typing 1 to 3 htSNPs. LD blocks and htSNPs were accomplished using the HapBlock analysis program [42-44]. The block partitioning algorithm was set as the dynamic programming algorithm,[42] the common haplotype method was used for block partitioning and the method for identifying htSNPs was capable of identifying all common (> 5%) haplotypes.

Age stratification

Controls from studies of RA, MS, AD and MI together with T1D controls (DISS1, DISS2, SV2, DR) were analysed for three SNP markers, rs11074932 (*C/T*) 3747 controls, rs3087456 (*A/G*) 7331 controls and rs4774 (*C/G*) 3317 controls *(S3).*

The controls were divided into fifteen 5 years intervals with respect to age at sampling (0-4, 5-9 .... 65-69, and >70).

Statistical analysis

Pearson's Chi-squared test was used to detect overall association, logistic regression analysis using generalized linear modeling between the persons with diabetes and the general population, was used to correct for the effect of age and HLA on the CIITA association. When correcting for age in logistic regression, the age-group 7 (30-35 years) was used as a reference group. In the logistic regression model where HLA was included as a factor, only individuals with data regarding *DRB1*04* haplotype were included in these analyses to be able to compare results, since this was a limiting factor. HLA coding was otherwise defined as presence or absence of allele for *DRB1*15* and *DRB1*03*, for *DRB1*04*, only individuals with *DRB1*04*DQB1-03:02* were considered positive for *DRB1*04*.

Chi-squared Test for Trend in Proportions was used to detect the overall trend in variation of genotype over age. Pearson's Chi-squared test was also used to detect differences in genotype and allele frequencies between the two age groups in the population based cohorts used for replication.

A joint age-stratified ("meta-analysis over age-groups") test of association in CIITA with T1D in the combined cohort of five case–control studies was performed for the first 8 age-groups, (0-39 years) using fixed effect Mantel–Haenszel analysis and Woolf's test for heterogeneity in R using the meta.MH command in the rmeta package.

In rs3087456_*GG*, age group 1 (0-4 yrs) was removed from the analysis due to heterogeneity between groups (p=0.04)

To investigate interaction between the CIITA gene and HLA haplotypes, departure from additivity was estimated by calculating attributable proportion (AP) due to interaction. These

analyses were performed as described [45], using the generalized linear modeling (glm) in R and the vcov command to get the covariance matrix. As suggested by Knol et al, the group with the lowest OR when both factors are considered jointly have been used as reference group when calculating AP [46], however the OR plotted in fig.5 are from the analysis when DR15 and CIITA minor allele positives were used as a reference group to clarify the interaction effect. The modeling included the SNP marker of investigation, as well as HLA haplotypes and age groups described as above.

All statistical analyses were performed in the statistical computer program R version 2.6.2[47], except the initial association analysis which was performed in Unphased[48] using the cocaphase command. An LD plot over the markers in the DISS2 cohort was performed in Haploview 4.2 *(fig4)*[49].

Supplementary information is available at Genes&Immunity´s website.

Acknowledgments

We thank Jana L. Hoehna and Marta Janer for their contribution to this work.

We acknowledge Professor Laura Fratiglioni and co-workers in the SNACK-project for the SNACK controls and Dr Behnosh F Björk for the sample handling.

This work was supported by grants from the Juvenile Diabetes Research Foundation International (2-2000-570) and (1-2001-873), the Swedish Research Council, Swedish Diabetes Foundation (Svenska Diabetes Fonden), Swedish Child Diabetes Foundation (Barndiabetes Fonden), Novo Nordisk Foundation, Magnus Bergvalls Foundation and Neuropromise (LSHM-CT-2005-018637). The Molecular and Genetics Core of the Diabetes Endocrinology Research Center is supported by NIH grant DK-17047.

The project was also supported by the Swedish Brain Power initiative, Gun and Bertil Stohne´s Foundation, Foundation for Old Servants, Alzheimer Foundation and LIONS

Foundation for Research of Age Related Disorders, the AFA foundation, the Söderberg Foundation, Knut and Alice Wallenbergs Foundation.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

The authors declare no financial, personal or professional conflict of interest.

References

1. Todd JA. Genetic analysis of type 1 diabetes using whole genome approaches. *Proc Natl Acad Sci U S A* 1995; **92**(19)**:** 8560-5.

2. Kockum I, Sanjeevi CB, Eastman S, Landin-Olsson M, Dahlquist G, Lernmark A. Complex interaction between HLA DR and DQ in conferring risk for childhood type 1 diabetes. *Eur J Immunogenet* 1999; **26**(5)**:** 361-72.

3. Graham J, Kockum I, Sanjeevi CB, Landin-Olsson M, Nystrom L, Sundkvist G *et al.* Negative association between type 1 diabetes and HLA DQB1*0602-DQA1*0102 is attenuated with age at onset. Swedish Childhood Diabetes Study Group. *Eur J Immunogenet* 1999; **26**(2-3)**:** 117-27.

4. Masternak K, Muhlethaler-Mottet A, Villard J, Zufferey M, Steimle V, Reith W. CIITA is a transcriptional coactivator that is recruited to MHC class II promoters by multiple synergistic interactions with an enhanceosome complex. *Genes Dev* 2000; **14**(9)**:** 1156-66.

5. Zhou H, Glimcher LH. Human MHC class II gene transcription directed by the carboxyl terminus of CIITA, one of the defective genes in type II MHC combined immune deficiency. *Immunity* 1995; **2**(5)**:** 545-53.

6. Steimle V, Siegrist CA, Mottet A, Lisowska-Grospierre B, Mach B. Regulation of MHC class II expression by interferon-gamma mediated by the transactivator gene CIITA. *Science* 1994; **265**(5168)**:** 106-9.

7. Klein C, Lisowska-Grospierre B, LeDeist F, Fischer A, Griscelli C. Major histocompatibility complex class II deficiency: clinical manifestations, immunologic features, and outcome. *J Pediatr* 1993; **123**(6)**:** 921-8.

8. Muhlethaler-Mottet A, Otten LA, Steimle V, Mach B. Expression of MHC class II molecules in different cellular and functional compartments is controlled by differential usage of multiple promoters of the transactivator CIITA. *EMBO J* 1997; **16**(10)**:** 2851-60.

9. LeibundGut-Landmann S, Waldburger JM, Reis e Sousa C, Acha-Orbea H, Reith W. MHC class II expression is differentially regulated in plasmacytoid and conventional dendritic cells. *Nat Immunol* 2004; **5**(9)**:** 899-908.

10. Sartoris S, Brendolan A, Degola A, Testi MG, Chignola R, Scarpa A *et al.* Analysis of CIITA encoding AIR-1 gene promoters in insulin-dependent diabetes mellitus and rheumatoid arthritis patients from the northeast of Italy: absence of sequence variability. *Hum Immunol* 2000; **61**(6)**:** 599-604.

11. Nerup J, Pociot F. A genomewide scan for type 1-diabetes susceptibility in Scandinavian families: identification of new loci with evidence of interactions. *Am J Hum Genet* 2001; **69**(6)**:** 1301-13.

12. Caillat-Zucman S, Garchon HJ, Timsit J, Assan R, Boitard C, Djilali-Saiah I *et al.* Age-dependent HLA genetic heterogeneity of type 1 insulin-dependent diabetes mellitus. *J Clin Invest* 1992; **90**(6)**:** 2242-50.

13. Graham J, Hagopian WA, Kockum I, Li LS, Sanjeevi CB, Lowe RM *et al.* Genetic effects on age-dependent onset and islet cell autoantibody markers in type 1 diabetes. *Diabetes* 2002; **51**(5)**:** 1346-55.

14. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A *et al.* Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* 2011; **43**(12)**:** 1193-201.

15. Swanberg M, Lidman O, Padyukov L, Eriksson P, Akesson E, Jagodic M *et al.* MHC2TA is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction. *Nat Genet* 2005; **37**(5)**:** 486-94.

16. Jensen RA, Gilliam LK, Torn C, Landin-Olsson M, Karlsson FA, Palmer JP *et al.* Multiple factors affect the loss of measurable C-peptide over 6 years in newly diagnosed 15- to 35-year-old diabetic subjects. *J Diabetes Complications* 2007; **21**(4)**:** 205-13.

17. Consortium WTCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447**(7145)**:** 661-78.

18. Bronson PG, Caillier S, Ramsay PP, McCauley JL, Zuvich RL, De Jager PL *et al.* CIITA variation in the presence of HLA-DRB1*1501 increases risk for multiple sclerosis. *Hum Mol Genet* 2010; **19**(11)**:** 2331-40.

19. Iikuni N, Ikari K, Momohara S, Tomatsu T, Hara M, Yamanaka H *et al.* MHC2TA is associated with rheumatoid arthritis in Japanese patients. *Ann Rheum Dis* 2007; **66**(2)**:** 274-5.

20. Martinez A, Sanchez-Lopez M, Varade J, Mas A, Martin MC, de Las Heras V *et al.* Role of the MHC2TA gene in autoimmune diseases. *Ann Rheum Dis* 2007; **66**(3)**:** 325-9.

21. Cooper JD, Walker NM, Smyth DJ, Downes K, Healy BC, Todd JA. Follow-up of 1715 SNPs from the Wellcome Trust Case Control Consortium genome-wide association study in type I diabetes families. *Genes Immun* 2009; **10 Suppl 1:** S85-94.

22. Larsen CE, Alper CA. The genetics of HLA-associated disease. *Curr Opin Immunol* 2004; **16**(5)**:** 660-7.

23. Eyre S, Bowes J, Spreckley K, Potter C, Ring S, Strachan D *et al.* Investigation of the MHC2TA gene, associated with rheumatoid arthritis in a Swedish population, in a UK rheumatoid arthritis cohort. *Arthritis Rheum* 2006; **54**(11)**:** 3417-22.

24. Harrison P, Pointon JJ, Farrar C, Harin A, Wordsworth BP. MHC2TA promoter polymorphism (-168*G/A, rs3087456) is not associated with susceptibility to rheumatoid arthritis in British Caucasian rheumatoid arthritis patients. *Rheumatology (Oxford)* 2007; **46**(3)**:** 409-11.

25. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 2007; **39**(7)**:** 857-64.

26. Landin-Olsson M, Karlsson FA, Lernmark A, Sundkvist G. Islet cell and thyrogastric antibodies in 633 consecutive 15- to 34-yr-old patients in the diabetes incidence study in Sweden. *Diabetes* 1992; **41**(8)**:** 1022-7.

27. Dahlquist G, Blom L, Tuvemo T, Nystrom L, Sandstrom A, Wall S. The Swedish childhood diabetes study--results from a nine year case register and a one year case-referent study indicating that type 1 (insulin-dependent) diabetes mellitus is associated with both type 2 (non-insulin-dependent) diabetes mellitus and autoimmune disorders. *Diabetologia* 1989; **32**(1)**:** 2-6.

28. Carlsson A, Kockum I, Lindblad B, Engleson L, Nilsson A, Forsander G *et al.* Low risk HLA-DQ and increased body mass index in newly diagnosed type 1 diabetes children in the Better Diabetes Diagnosis study in Sweden. In: *Int J Obes (Lond).* 2011/06/30 ed, 2011.

29. Lindholm E, Agardh E, Tuomi T, Groop L, Agardh CD. Classifying diabetes according to the new WHO clinical stages. *Eur J Epidemiol* 2001; **17**(11)**:** 983-9.

30. Stolt P, Bengtsson C, Nordmark B, Lindblad S, Lundberg I, Klareskog L *et al.* Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. *Ann Rheum Dis* 2003; **62**(9)**:** 835-41.

31. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B *et al.* TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. *N Engl J Med* 2007; **357**(12)**:** 1199-209.

32. Roos IM, Kockum I, Hillert J. The interleukin 23 receptor gene in multiple sclerosis: a case-control study. *J Neuroimmunol* 2008; **194**(1-2)**:** 173-80.

33. Hedstrom AK, Baarnhielm M, Olsson T, Alfredsson L. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. *Neurology* 2009; **73**(9)**:** 696-701.

34. Bjork BF, Katzov H, Kehoe P, Fratiglioni L, Winblad B, Prince JA *et al.* Positive association between risk for late-onset Alzheimer disease and genetic variation in IDE. *Neurobiol Aging* 2007; **28**(9)**:** 1374-80.

35. McGuigan FE, Larzenius E, Callreus M, Gerdhem P, Luthman H, Akesson K. Variation in the BMP2 gene: bone mineral density and ultrasound in young adult and elderly women. *Calcif Tissue Int* 2007; **81**(4)**:** 254-62.

36. Gerdhem P, Isaksson A, Akesson K, Obrant KJ. Increased bone density and decreased bone turnover, but no evident alteration of fracture susceptibility in elderly women with diabetes mellitus. *Osteoporos Int* 2005; **16**(12)**:** 1506-12.

37. Ekelund E, Saaf A, Tengvall-Linder M, Melen E, Link J, Barker J *et al.* Elevated expression and genetic association links the SOCS3 gene to atopic dermatitis. *Am J Hum Genet* 2006; **78**(6)**:** 1060-5.

38. Prince JA, Feuk L, Howell WM, Jobs M, Emahazion T, Blennow K *et al.* Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. *Genome Res* 2001; **11**(1)**:** 152-62.

39. Roach JC, Deutsch K, Li S, Siegel AF, Bekris LM, Einhaus DC *et al.* Genetic mapping at 3-kilobase resolution reveals inositol 1,4,5-triphosphate receptor 3 as a risk factor for type 1 diabetes in sweden. *Am J Hum Genet* 2006; **79**(4)**:** 614-27.

40. Olerup O, Zetterquist H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. *Tissue Antigens* 1992; **39**(5)**:** 225-35.

41. Kiviniemi M, Hermann R, Nurmi J, Ziegler AG, Knip M, Simell O *et al.* A high-throughput population screening system for the estimation of genetic risk for type 1 diabetes: an application for the TEDDY (the Environmental Determinants of Diabetes in the Young) study. *Diabetes Technol Ther* 2007; **9**(5)**:** 460-72.

42. Zhang K, Deng M, Chen T, Waterman MS, Sun F. A dynamic programming algorithm for haplotype block partitioning. *Proc Natl Acad Sci U S A* 2002; **99**(11)**:** 7335-9.

43. Zhang K, Jin L. HaploBlockFinder: haplotype block analyses. *Bioinformatics* 2003; **19**(10)**:** 1300-1.

44. Zhang K, Qin ZS, Liu JS, Chen T, Waterman MS, Sun F. Haplotype block partitioning and tag SNP selection using genotype data and their applications to association studies. *Genome Res* 2004; **14**(5)**:** 908-16.

45. Andersson T, Alfredsson L, Kallberg H, Zdravkovic S, Ahlbom A. Calculating measures of biological interaction. *Eur J Epidemiol* 2005; **20**(7)**:** 575-9.

46. Knol MJ, VanderWeele TJ. Recoding preventive exposures to get valid measures of interaction on an additive scale. *Eur J Epidemiol* 2011; **26**(10)**:** 825-6; author reply 826.

47. R Development Core Team. R: A language and environment for statistical computing. In. Vienna, Austria: R Foundation for Statistical Computing,, 2008.

48. Dudbridge F. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum Hered* 2008; **66**(2)**:** 87-98.

49. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**(2)**:** 263-5.

Figure Legends

Figure 1. Genotype distribution with age in *CIITA* among controls.

Genotype frequencies in percent for A) rs11074932 B) rs3087456 and C) rs4774 related to age-group (number of individuals in parenthesis). P-value for respectively genotype in test for trend-over-age. For B) the p-value for the AA genotype is 0.09 if the oldest group (over75 yrs of age) is removed.

Figure 2. Genotype frequencies in cohort material.

Genotype frequencies in marker rs11074932, rs3087456 and rs4774 in the two groups of the cohort material (25 years old, n=1005 and 75 years old, n=1010 individuals).
P-values for genotypes in 2x3 chi-sqr test.

Figure 3. Age-stratified analysis of association in *CIITA* with type 1 diabetes.

Association in the combined cohort of five Scandinavian case–control studies for individuals under 40 years of age, of rs11074932 and rs3087456 in *CIITA* with Type 1 diabetes.
Meta analysis of age-groups with 5 years intervals and using the frequency of the genotypes for each marker.

Figure 4. Linkage disequlibrium plot of *CIITA* and *CLEC16A* in the DISS2 cohort.

Linkage disequilibrium of the *CIITA* to *CLEC16A* gene region in the DISS2 cohort; darker gray indicates higher r^2 between markers. (Haploview 4.2)

Figure 5. Interaction between *DRB1*15* and markers in *CIITA*.

Lack of *DRB1*15* together with the major allele homozygote increases the OR for type 1 diabetes. Error bars are 95% CI of OR estimates.

Attributable proportion (AP) is the proportion of the incidence among individuals exposed to both associated factors compared to the factors individually. The AP value is significant if separate from zero.

HLA is coded as absence of HLA allele for DRB1*15. The CIITA SNPS were coded as absence of minor allele for both markers.

Supplementary material

S1. SNP positions and association analysis for T1D in the DISS2 cohort.

S2. LD block analysis in 373 Caucasian controls

S3. Cohorts – numerical summary

S4. Members of the Swedish Childhood Diabetes Study Group, the Diabetes Incidence in Sweden Study Group and the Better Diabetes Diagnosis Study group.

Tables

Table 1 Allele frequency and p-values for OPRA and PEAK cohorts

Minor allele frequency	PEAK- 25 yrs old	OPRA- 75 yrs old	p-value allele[1]	p-value genotypes[2]	p-value minor allele homozygote genotype[3]
rs11074932-C	0.303	0.271	0.03	0.02	0.006
rs3087456-G	0.270	0.235	0.01	0.01	0.007
rs4774-C	0.301	0.311	0.52	0.64	0.03

[1]2x2 Chi Sqr test for variation between groups in allele distribution
[2]2x3 Chi Sqr test for variation between groups in genotype frequency
[3]2x2 Chi Sqr test for variation between groups in minor allele homozygotes vs. other genotypes frequency

Table 2 Association of *CIITA* to Type 1 Diabetes in the combined T1D cases and controls.

Marker	cases % (n)	controls% (n)	chi2 [1] (p-value)	log reg [2] (p-value)	logreg corrected for age[3] (P-value)	logreg minor allele corrected for age[4] (P-value)
rs11074932						
TT	53% (1374)	48% (595)				
CT	39% (1012)	43% (533)		0.007	0.006	
CC	8% (195)	9% (115)		0.15	0.15	
Total n	2581	1243	0.005			0.004
rs3087456						
AA	59% (1945)	55% (1981)				
AG	35% (1139)	38% (1369)		0.001	0.003	
GG	6% (200)	7% (251)		0.040	0.09	
Total n	3284	3601	0.002			0.001
rs4774						
GG	48% (605)	47% (616)				
GC	42% (528)	43% (563)		0.58	0.58	
CC	10% (126)	10% (131)		0.88	0.96	
Total n	1259	1310	0.86			0.62

[1] 2x3 Chi Sqr test for association of genotypes to T1D without correction for age at sampling
[2] Logistic regression analysis testing association of heterozygotes and minor allele homozygotes with major allele homozygotes as reference group
[3] Logistic regression analysis testing association of heterozygotes and minor allele homozygotes with major allele homozygotes as reference group, age at sampling was included as covariates.
[4] Logistic regression analysis testing association of presence of minor allele with major allele homozygotes as reference group, age at sampling was included as covariates.

Fig 1

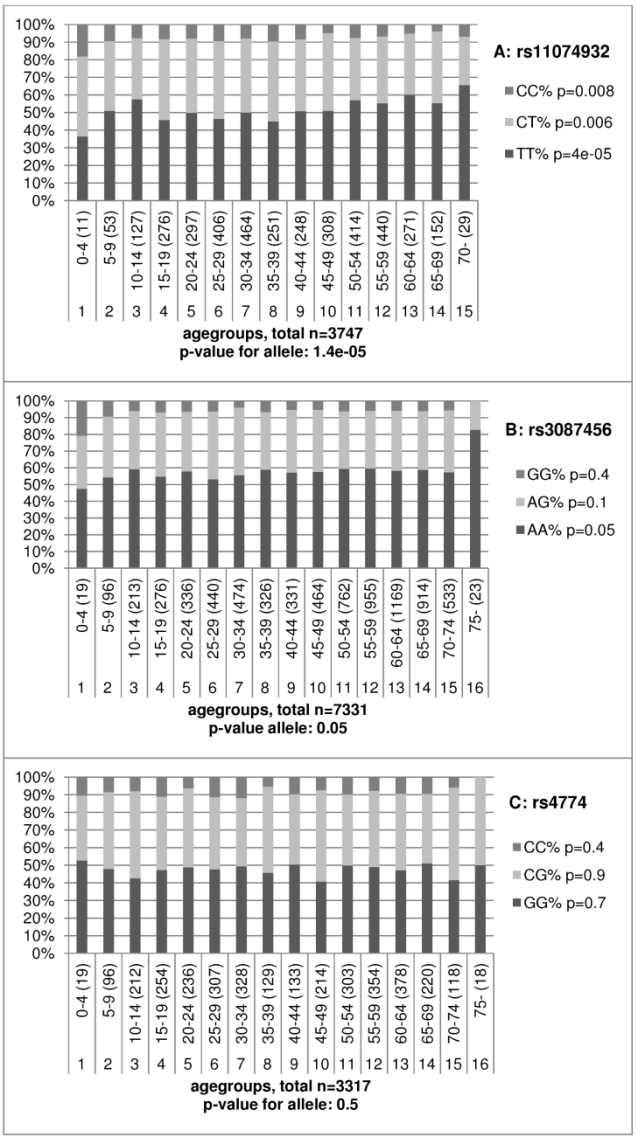

A: rs11074932
CC% p=0.008
CT% p=0.006
TT% p=4e-05
0-4 (11)
5-9 (53)
10-14 (127)
15-19 (276)
20-24 (297)
25-29 (406)
30-34 (464)
35-39 (251)
40-44 (248)
45-49 (308)
50-54 (414)
55-59 (440)
60-64 (271)
65-69 (152)
70- (29)
agegroups, total n=3747
p-value for allele: 1.4e-05
B: rs3087456
GG% p=0.4
AG% p=0.1
AA% p=0.05
0-4 (19)
5-9 (96)
10-14 (213)
15-19 (276)
20-24 (336)
25-29 (440)
30-34 (474)
35-39 (326)
40-44 (331)
45-49 (464)
50-54 (762)
55-59 (955)
60-64 (1169)
65-69 (914)
70-74 (533)
75- (23)
agegroups, total n=7331
p-value allele: 0.05
C: rs4774
CC% p=0.4
CG% p=0.9
GG% p=0.7
0-4 (19)
5-9 (96)
10-14 (212)
15-19 (254)
20-24 (236)
25-29 (307)
30-34 (328)
35-39 (129)
40-44 (133)
45-49 (214)
50-54 (303)
55-59 (354)
60-64 (378)
65-69 (220)
70-74 (118)
75- (18)
agegroups, total n=3317
p-value for allele: 0.5

Fig 2

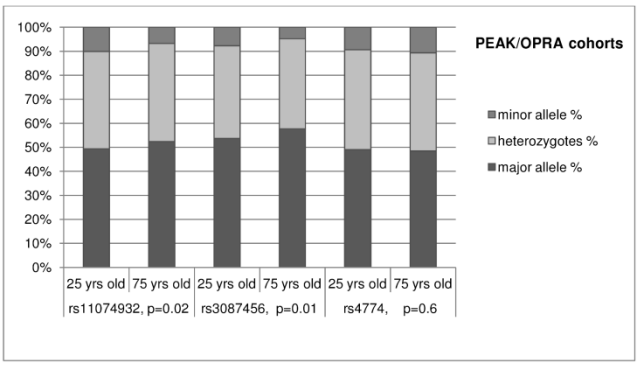

PEAK/OPRA cohorts
minor allele %
heterozygotes %
major allele %
25 yrs old
75 yrs old
rs11074932, p=0.02
rs3087456, p=0.01
rs4774, p=0.6

Fig 3

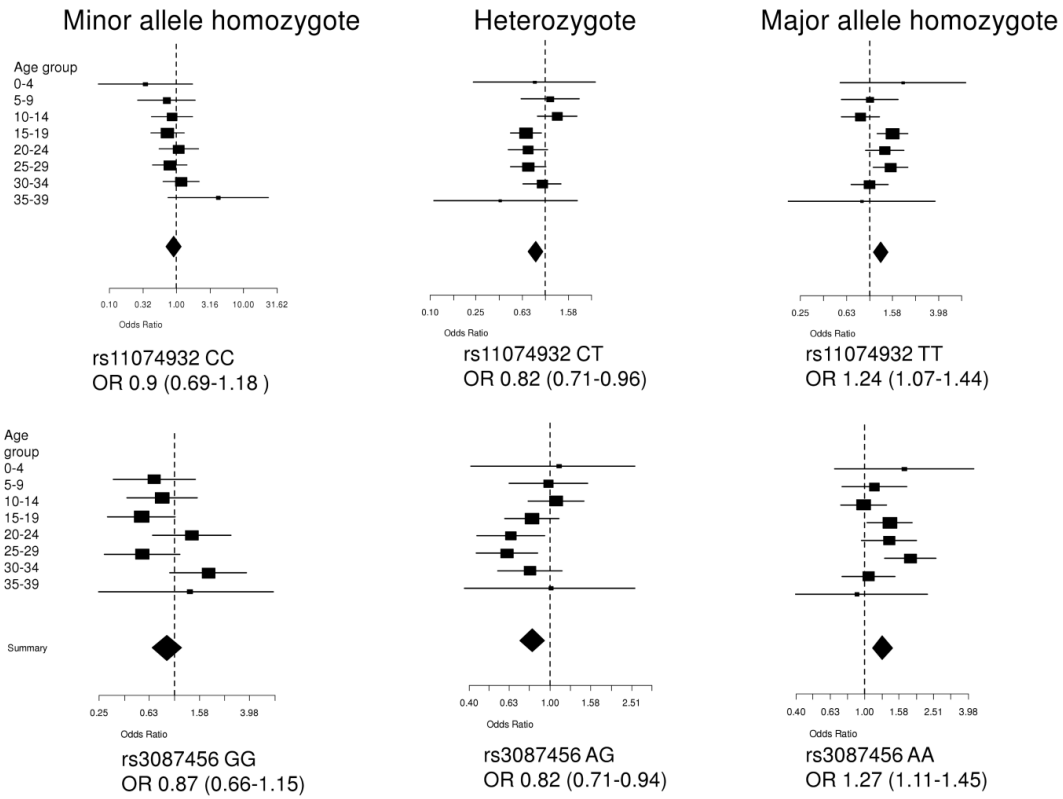

Minor allele homozygote
Heterozygote
Major allele homozygote
Age group
0-4
5-9
10-14
15-19
20-24
25-29
30-34
35-39
0.10 0.32 1.00 3.16 10.00 31.62
Odds Ratio
0.10 0.25 0.63 1.58
Odds Ratio
0.25 0.63 1.58 3.98
Odds Ratio
rs11074932 CC
OR 0.9 (0.69-1.18)
rs11074932 CT
OR 0.82 (0.71-0.96)
rs11074932 TT
OR 1.24 (1.07-1.44)
Age group
0-4
5-9
10-14
15-19
20-24
25-29
30-34
35-39
Summary
0.25 0.63 1.58 3.98
Odds Ratio
0.40 0.63 1.00 1.58 2.51
Odds Ratio
0.40 0.63 1.00 1.58 2.51 3.98
Odds Ratio
rs3087456 GG
OR 0.87 (0.66-1.15)
rs3087456 AG
OR 0.82 (0.71-0.94)
rs3087456 AA
OR 1.27 (1.11-1.45)

Fig 4

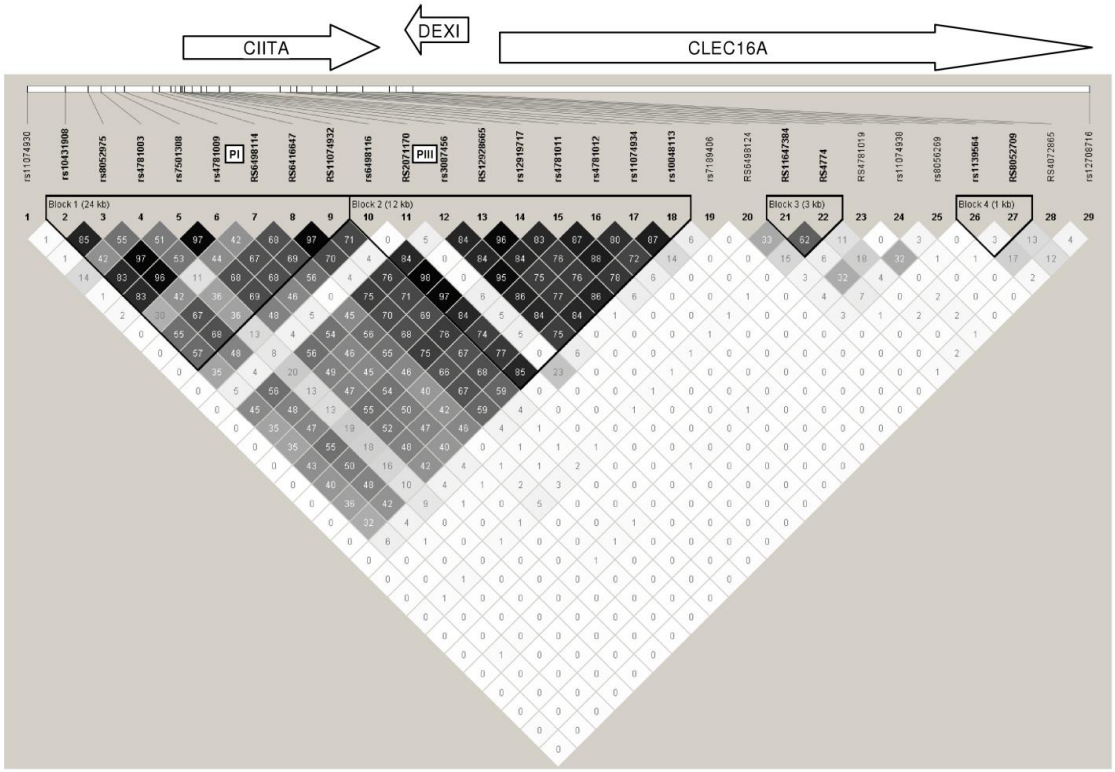
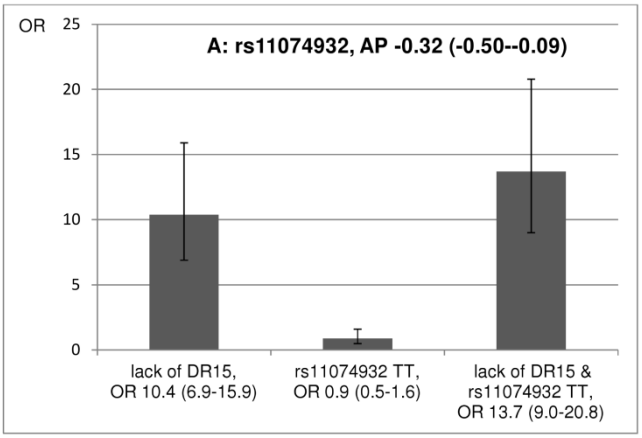
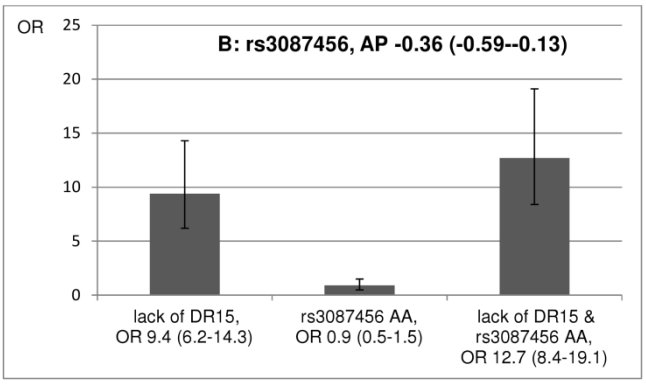

CIITA
DEXI
CLEC16A
PI
PIII
Block 1 (24 kb)
Block 2 (12 kb)
Block 3 (3 kb)
Block 4 (1 kb)

Fig 5


OR
25
20
15
10
5
0
A: rs11074932, AP -0.32 (-0.50--0.09)
lack of DR15,
OR 10.4 (6.9-15.9)
rs11074932 TT,
OR 0.9 (0.5-1.6)
lack of DR15 &
rs11074932 TT,
OR 13.7 (9.0-20.8)


OR
25
20
15
10
5
0
B: rs3087456, AP -0.36 (-0.59--0.13)
lack of DR15,
OR 9.4 (6.2-14.3)
rs3087456 AA,
OR 0.9 (0.5-1.5)
lack of DR15 &
rs3087456 AA,
OR 12.7 (8.4-19.1)

S1. SNP positions and association analysis for T1D in the DISS2 cohort.

SNP name	htSNP	Location/position[1]	H-W[2]	Minor allele frequency in DISS2		Association[3]	Association after matching cases and controls for age[4]
				patients	controls	P<	
rs11074930**	na	Before PI/ 10842650	ns	0.45	0.52	0.005	0.04
rs10431908**	na	Before PI/ 10851548	ns	0.23	0.28	0.01	0.39
rs8052975**	na	Before PI/ 10856764	ns	0.26	0.30	0.02	0.44
rs4781003**	na	Before PI/ 10859668	ns	0.16	0.19	0.1	0.60
rs7501308**	na	Before PI/ 10862957	ns	0.26	0.31	0.006	0.08
rs4781009*	No	Before PI/ 10865178	ns	0.26	0.30	0.04	0.32
rs7500908	No	PI/ 10867491	np	-	-	-	-
rs8059450	No	PI/ 10867746	np	-	-	-	-
rs6498114*	No	between PI and PIII/ 10871619	ns	0.22	0.26	0.03	0.16
rs6416647*	yes	between PI and PIII/ 10873098	ns	0.27	0.32	0.005	0.09
rs11074932*	No	between PI and PIII/ 10875837	ns	0.27	0.32	0.002	0.05
rs6498116*	No	between PI and PIII/ 10876783	ns	0.22	0.25	0.04	0.27
rs7404116	No	between PI and PIII/ 10877836	np	-	-	-	-
rs2071170*	No	between PI and PIII/ 10878128	ns				
rs3087456*	Yes	In PIII/ 10878403	ns	0.24	0.28	0.03	0.16
rs12928665*	No	between PIII and PIV/ 10878975	ns	0.22	0.25	0.06	0.28
rs12919717*	No	intron/ 10880686	ns	0.22	0.25	0.04	0.29
rs4781011*	Yes	intron/ 10882812	ns	0.24	0.27	0.06	0.22
rs11074933	No	intron/ 10884067	np	-	-	-	-
rs11074934*	No	intron/ 10886941	ns	0.26	0.28	0.17	0.16
rs10048113*	No	intron/ 10889417	ns	0.24	0.26	0.15	0.19
rs8043545	No	intron/ 10889846	np	-	-	-	-
rs8062705	No	intron/ 10892631	np	-	-	-	-
rs7195305*	No	intron/ 10894700	ns				
rs11074937	No	intron/ 10896875	np	-	-	-	-
rs4781015	No	intron/ 10899453	np	-	-	-	-
rs7189406*	No	intron/ 10900989	ns				
rs6498124*	Yes	intron/ 10903351	ns	0.44	0.44	0.92	0.1
rs11647384*	Yes	intron/ 10904790	ns	0.40	0.40	0.75	0.58
rs7404615*	No	intron/ 10907174	ns				
rs4774*	Yes	non-synonymous/ 10908349	ns	0.32	0.30	0.35	0.23
rs2229319	No	non-synonymous/ 10908823	np	-	-	-	-
rs7196089	No	intron/ 10910602	np	-	-	-	-
rs4781019*	Yes	intron/ 10911651	ns	0.46	0.47	0.52	0.54
rs11074938*	Yes	intron/ 10914044	ns	0.36	0.36	0.99	0.44
rs11647308	No	intron/ 10914145	np	-	-	-	-
rs4781021*	No	intron / 10916768	ns				
rs6498131*	No	intron / 10918127	ns				
rs8056269*	Yes	intron / 10920068	<0.0009	0.44	0.42	na	na
rs7203275	No	intron / 10924098	np	-	-	-	-
rs4781024*	No	intron / 10924559	ns				
rs1139564*	Yes	untranslated / 10926123	ns	0.17	0.17	0.67	0.81
rs8052709*	Yes	untranslated / 10927756	ns	0.27	0.30	0.11	0.17
rs11643328	No	untranslated / 10928828	np	-	-	-	-
rs3087519*	No	untranslated / 10930709	ns				
rs4072865*	Yes	untranslated / 10931606	ns	0.48	0.49	0.43	0.47
rs12708716**	na	CLEC16A gene / 11087374	ns	0.28	0.32	0.07	0.07

All SNPs were first genotyped in a selected cohort of 373 controls. 13 of the selected SNPs were non-polymorphic and 1 did not follow Hardy-Weinberg equilibrium and were discarded from further studies

* SNPs run in the HapBlock program. All markers within the haplotype blocks where not typed in the DISS cohort.

**SNPs added in second round of genotyping.

[1] Chromosome position, genome build 36.3, contig NT 010393.15 (Reference sequence)

[2] Test of Hardy-Weinberg Equilibrium in 373 Caucasian controls, ns=p>0.05, np= non-polymorphic

[3] Association analyses performed in Unphased using cocaphase command in the DISS2 cohort. After correcting for multiple testing, significant association remained only for rs11074932.

[4] Association analyses performed in Unphased using cocaphase command in the DISS2 cohort after manually matching each case to one or if possible two controls by exact age.

S2. LD block analysis in 373 Caucasian controls

Haplotype Block No.	positions	SNP boundaries	Block size (bp)	Number of typed SNPs	htSNPs	% of uniquely distinguished haplotypes
Block 1	upstream of PI to between PI and PIII	rs4781009-rs6416647	7920 bp	3	rs6416647	89
Block 2	upstream of PIII to downstream of exon 8	rs11074932 – rs6498124	27514 bp	12	rs3087456 / rs4781011 / rs6498124	81
Block 3	upstream of exon 11 to exon 13	rs11647384 – rs4774	3559 bp	4	rs11647384 / rs4774	88
Block 4	the end of exon 13 to exon 20	rs4781019 - rs8056269	8417 bp	5	rs4781019 / rs11074938 / rs8056269	87
Block 5	spans the 3´UTR.	rs1139564 - rs4072865	5483 bp	4	rs1139564 / rs8052709 / rs4072865	94

LD block definition and htSNP identification was carried out in HapBlock program (24-26). The block partitioning algorithm used was the dynamic programming algorithm by Zhang et al (24) this minimizes the total number of tag SNPs in a region of interest. The method for block partitioning was the “common haplotype” method where a set of SNPs form a LD block if the common haplotypes in this block account for at leas 80% of the observed haplotypes. The method for identifying htSNPs was the “all common haplotypes” method. This method identifies the minimum set of SNPs that can distinguish all common haplotypes within an LD block. Common haplotypes were defined to have a frequency of 5% or more.

S3. Cohorts

a) Overview of T1D cohorts – n (%)

controls	**BDD**	**DISS1**	**DISS2**	**DR**	**SV2**	**Total sum:**
total	0	348	797	2312	342	3799
men	0	196 (56%)	448 (56%)	1695 (73%)	179 (52%)	2518
women	0	152 (44%)	349 (44%)	617 (27%)	163 (48%)	1281
mean age	0	25	25	63	10	
median age	0	24	26	63	11	
range yrs	0	15-34	15-36	45-74	1-15	
cases	**BDD**	**DISS1**	**DISS2**	**DR**	**SV2**	
total	1475	431	544	631	404	3485
men	844 (57%)	277 (64%)	349 (64%)	337 (53%)	213 (53%)	2020
women	631 (43%)	154 (36%)	195 (36%)	294 (47%)	191 (47%)	1465
mean age	10	25	24	22	8	
median age	10	25	25	21	9	
range yrs	0-18	15-36	15-35	0-70	1-14	

b) Summary of cohorts used for age-stratification analyses; controls only, age 0-91 yrs (n).

marker	**DISS1**	**DISS2**	**SV2**	**DR**	**MS**	**RA**	**MI**	**AD**	**Total sum:**
rs11074932	295	757	191	na	1177	942	385	na	3747
rs3087456	207	751	331	2312	1602	1324	387	417	7331
rs4774	204	777	329	na	477	738	386	406	3317

c) Summary of population based control cohorts for replication of age variation (n).

marker	**OPRA 75 yrs old**	**PEAK 25 yrs old**	**Total sum:**
rs11074932	1002	1002	2004
rs3087456	994	993	1987
rs4774	990	976	1966

S4. Members of the Swedish Childhood Diabetes Study Group:

All from Departments of Pediatrics in Sweden: M. Aili, Halmstad; L.E. Bååth, Östersund; E. Carlsson, Kalmar; H. Edenwall, Karlskrona; G. Forsander, Falun; B.W. Granström, Gällivare; I. Gustavsson, Skellefteå; R. Hanås, Uddevalla; L. Hellenberg, Nyköping; H. Hellgren, Lidköping; E. Holmberg, Umeå; H. Hörnell, Hudiksvall; Sten-A. Ivarsson, Malmö; C. Johansson, Jönköping; G. Jonsell, Karlstad; K. Kockum, Ystad, B. Lindblad, Mölndal; A. Lindh, Borås; J. Ludvigsson, Linköping; U. Myrdal, Vä sterås; J. Neiderud, Helsingborg; K. Segnestam, Eskilstuna; S. Sjöblad, Lund;L. Skogsberg, Boden; L. Strömberg, Norrköping; U. Ståhle, Ängelholm; B. Thalme, Huddinge; K. Tullus, Danderyd; T. Tuvemo, Uppsala; M. Wallensteen, Stockholm; O. Westphal, Göteborg; and J. Åman, Örebro.

Members of the Diabetes Incidence in Sweden Study (DISS) Group:

Hans Arnqvist, Department of Internal Medicine, University of Linköping, Linköping; Elisabeth Björck, Department of Medicine, University Hospital, Uppsala; Jan Eriksson, Department of Medicine, University of Umeå, Umeå; Lennarth Nyström, Department of Epidemiology and Public Health, University of Umeå, Umeå; Lars Olof Ohlson, Sahlgrenska Hospital, University of Göteborg, Göteborg; Bengt Scherstén, Department of Community Health Sciences, Dahlby, University of Lund, Lund; Jan Östman, Center for Metabolism and Endocrinology, Huddinge University Hospital, Stockholm.

Members of the BDD Study Group:

Anita Nilsson (Malmö), Helena Desaix (Borås), Kalle Snellman (Eskilstuna), Anna Olivecrona (Falun), Åke Stenberg (Gällivare), Lars Skogsberg (Gävle), Nils Östen Nilsson (Halmstad), Jan Neiderud (Helsingborg), Åke Lagerwall (Hudiksvall), Kristina Hemmingsson (Härnösand), Karin Åkesson (Jönköping), Göran Lundström (Kalmar),

Magnus Ljungcrantz (Karlskrona), Eva Albinsson (Karlstad), Karin Larsson (Kristianstad), Christer Gundewall (Kungsbacka), Rebecka Enander (Lidköping), Agneta Brännström (Luleå), Maria Nordwall (Norrköping), Lennart Hellenberg (Nyköping), Elena Lundberg (Skellefteå), Henrik Tollig (Skövde), Britta Björsell (Sollefteå), Björn Rathsman (Stockholm/Sacchska), Torun Torbjörnsdotter (Stockholm/Huddinge), Björn Stjernstedt (Sundsvall), Nils Wramner (Trollhättan), Ragnar Hanås (Uddevalla), Ingemar Swenne (Uppsala), Anna Levin (Visby), Anders Thåström (Västervik), Carl-Göran Arvidsson (Västerås), Stig Edvardsson (Växjö), Björn Jönsson (Ystad), Torsten Gadd