

Analysis of SNPs Identifies Major Type 1A Diabetes Locus Telomeric of the MHC

Theresa A. Aly, MD, PhD^{1,2*}, Erin E. Baschal, BA^{1,2*}, Mohamed M. Jahromi, PhD¹, Maria S. Fernando, BS¹, Sunanda R. Babu, PhD¹, Tasha E. Fingerlin, PhD^{2,3}, Adam Kretowski, MD^{1,4}, Henry A Erlich, PhD⁵, Pamela R. Fain, PhD^{1,2}, Marian J. Rewers, MD, PhD^{1,3}, George S. Eisenbarth, MD, PhD^{1,2}

¹Barbara Davis Center for Childhood Diabetes and ²Human Medical Genetics Program, University of Colorado at Denver and Health Sciences Center, Aurora, Colorado 80045 USA;

³Department of Preventive Medicine and Biometrics, University of Colorado at Denver and Health Sciences Center, Aurora, CO 80045 USA; ⁴Medical Academy of Bialystok, Poland;

⁵Roche Molecular Systems, Alameda, CA 94710, USA

*equal contributors

Running title: Type 1A Diabetes MHC Locus Identified

Corresponding Author:

George S. Eisenbarth, MD, PhD
Barbara Davis Center for Childhood Diabetes
University of Colorado Health Sciences Center
Mail Stop B140
PO Box 6511
Aurora, CO 80045-6511 USA
George.Eisenbarth@UCHSC.edu

Received for publication 3 July 2007 and accepted in revised form 21 November 2007.

Additional information for this article can be found in an online appendix at <http://diabetes.diabetesjournals.org>.

ABSTRACT

Objective: *HLA-DRB1*03-DQB1*0201/DRB1*04-DQB1*0302 (DR3/4-DQ8)* siblings who share both MHC haplotypes identical-by-descent with their proband siblings have a higher risk for type 1A diabetes than *DR3/4-DQ8* siblings who do not share both MHC haplotypes identical-by-descent. Our goal was to search for non-*DR/DQ* MHC genetic determinants that cause the additional risk in the *DR3/4-DQ8* siblings who share both MHC haplotypes.

Research Design and Methods: We completed an extensive SNP analysis of the extended MHC in 237 families with type 1A diabetes from the United States and 1,240 families from the Type 1 Diabetes Genetics Consortium.

Results: We found evidence for an association with type 1A diabetes (rs1233478, $p=1.6 \times 10^{-23}$, allelic odds ratio=2.0) in the *UBD/MASIL* region, telomeric of the classic MHC. We also observed over 99% conservation for up to 9 million nucleotides between chromosomes containing a common haplotype with the *HLA-DRB1*03*, *HLA-B*08*, and *HLA-A*01* alleles, termed the “8.1 haplotype.” The diabetes association in the *UBD/MASIL* region remained significant both after chromosomes with the 8.1 haplotype were removed (rs1233478, $p=1.4 \times 10^{-12}$), and after adjustment for known HLA risk factors *HLA-DRB1*, *HLA-DQB1*, *HLA-B* and *HLA-A* ($p=0.01$).

Conclusions: Polymorphisms in the region of the *UBD/MASIL* genes are associated with type 1A diabetes independent of HLA class II and I alleles.

Genetic susceptibility to immune mediated diabetes is determined by polymorphisms/mutations in multiple genes in both man and animal models. This includes rare immunologic syndromes such as the Autoimmune Polyendocrine Syndrome Type 1 (APS-1) and the Immune dysfunction, Polyendocrinopathy, Enteropathy, X-linked (IPEX) syndrome, in which mutations of single genes greatly enhance the risk of developing autoimmune diabetes (1-3). The majority of individuals born with mutations of the *FOXP3* gene (IPEX syndrome) develop diabetes as neonates or young infants, indicating that, in the absence of *FOXP3* dependent function (most likely due to lack of a major class of regulatory T cells), there is high risk for immune destruction of islet beta cells (4). In contrast, for common forms of type 1A diabetes, single genes with major influence outside of the MHC region have apparently been ruled out, because whole genome association of large and independent populations have not identified single non-MHC genes with high odds ratios for diabetes risk (5). The MHC is the principal determinant for type 1A diabetes risk, with the largest contribution attributed to the *HLA-DR/DQ* alleles, including contributions from *HLA-DRB1*04* subtypes and *HLA-DP* alleles (6,7).

Several reports have suggested the existence of MHC and MHC-linked loci that contribute to diabetes risk in addition to the *HLA-DR/DQ* loci (8-12). We have recently reported that risk for anti-islet autoimmunity may be as high as 80% for *HLA-DRB1*0301-DQB1*0201/DRB1*04-DQB1*0302* (*DR3/4-DQ8*) siblings who share both of their MHC haplotypes with their proband identical-by-descent, but risk is “only” 20% for those *DR3/4-DQ8* siblings who share one or no haplotypes identical-by-descent with their sibling proband (13). This strongly implicates major diabetogenic MHC-linked alleles not

accounted for by molecular genotyping of known *HLA-DR/DQ* sequences.

There is a classic body of work documenting existence of what have been termed “extended” or “ancestral” MHC haplotypes on which a series of polymorphisms extending over approximately 1 million nucleotides are in almost total linkage disequilibrium (14-16). With the advent of high-density SNP analysis and resequencing of extended regions of the MHC, extended haplotypes, such as the *HLA-DRB1*03, HLA-B*08, HLA-A*01* (8.1) haplotype, have been shown to have over 99% conservation for 3.2 Mb of the MHC region, including alleles of the *HLA-DRB1, HLA-DQB1, HLA-B, HLA-C* and *HLA-A* genes (17,18). The 8.1 haplotype is a common haplotype that is present in approximately 9% of Caucasian MHC control haplotypes and 18% of MHC case haplotypes from families with type 1A diabetes (19). Even though there are other conserved extended haplotypes (e.g. *HLA-DRB1*03, HLA-B*18, HLA-A*30*) (20), none have even half the frequency of the 8.1 haplotype in the Caucasian population with type 1A diabetes (16). We report here strong evidence for an association between type 1A diabetes and a locus in the *UBD/MASIL* region that is apparently independent of HLA class II and I alleles, and also independent of linkage disequilibrium with the common MHC 8.1 haplotype.

RESEARCH DESIGN AND METHODS

We analyzed three datasets, beginning with a survey study that we called “Illumina 1.” In Illumina 1, samples from 45 families (143 individuals) of children in the prospective cohort Diabetes Autoimmunity Study of the Young (DAISY) study were selected for analysis at 656 MHC SNPs. The DAISY study has HLA-typed more than 30,000 newborns in Denver Colorado at the

HLA-DR/DQ loci with informed parental consent and Institutional Review Board (IRB) oversight at the University of Colorado at Denver and Health Sciences Center (21,22). A subset of these children with high, intermediate, and low risk for type 1 diabetes by family history and HLA type were followed with measurements of autoantibody levels (insulin, GAD65, and IA-2) (22). Given preliminary results from the Illumina 1 analysis, a follow-up study, “Illumina 2”, was performed to refine the regions of association in Illumina 1 and extend the coverage of the telomeric end of the MHC. In the Illumina 2 study, DAISY families (N=165 families, 523 individuals, 138 families independent of the Illumina 1 study) and Human Biological Data Interchange (HBDI) families (N=72 families, 241 individuals) were analyzed at 364 SNPs. All of the 237 families in the Illumina 2 study were independent of the Illumina 1 study except for 27 DAISY families that were included in both studies [families with a prospectively followed child at extremely high risk for developing type 1A diabetes (high-risk *DR3/4-DQ8* genotype and a relative with type 1A diabetes)]. As a replication study of Illumina 2, we analyzed data provided by the Type 1 Diabetes Genetics Consortium (T1DGC) (“Illumina 3”). This study included 1,240 families (6,297 individuals, primarily affected sib pairs and their parents) from the British Diabetes Association (BDA), Danish (DAN), Joslin (JOS), Human Biological Data Interchange (HBDI), and United Kingdom (UK) populations.

We designed custom SNP panels for our Illumina 1 (656 SNPs) and Illumina 2 (364 SNPs) analyses. The Illumina 1 SNP panel covered 3 million base pairs of the classic MHC and included mapping and exon-centric SNPs with an average of 1 SNP every 6.3 kb and a maximum interval of 30 kb. The Illumina 2 custom SNP panel concentrated on nominally significant SNPs from the Illumina

1 study and SNPs telomeric of the classic MHC. We selected 53 SNPs in the classical MHC region [including nominally significant SNPs from the Illumina 1 analysis ($p < 0.01$ within the peak of association around the *HLA-DRB1* and *-DQB1* genes or $p < 0.05$ outside of the *HLA-DR/DQ* major peak of association), SNPs near genes of interest (*MICA*, *TNF*, and *HLA-DPBI*), and 12 “sentinel” SNPs with alleles defining the DR3, B8, A1 (8.1) haplotype] (17). The Illumina 2 panel included all 26 SNPs from Illumina 1 and 113 additional SNPs to cover the 1 Mb region telomeric of *HLA-F* (average of 1 SNP every 7.2 kb). We included 192 additional SNPs to extend the region of analysis an additional 6 million base pairs (with an average of 1 SNP every 18 kb between 27-29 Mb and 1 SNP every 50 kb between 23-27 Mb from the telomere, Supplementary Table 1). Illumina™ multiplex technology, which uses an extension/ligation reaction (details available at the Illumina™ Technology and Applications Web site) was used to genotype the Illumina 2 samples, yielding a 93% (772/828) sample success rate and a 95% SNP success rate.

The Illumina 3 (T1DGC) analysis utilized the dense standard Illumina™ mapping and exon-centric MHC panels (1,536 SNPs in each panel with 115 overlapping SNPs, 2,918 of 3,072 SNPs were successfully typed for a 95% SNP success rate). Illumina 2 and Illumina 3 had 83 SNPs in common for our combined study of the classical MHC. In addition, rs1233478, a highly significant SNP in the Illumina 3 analysis that was not included in the initial Illumina 2 panel, was genotyped with Taqman® probes in the Illumina 2 study samples (hybridization/extension reaction with a fluorescence detection system developed by Applied Biosystems Incorporated).

For all three Illumina™ SNP datasets, the PedCheck program (<http://watson.hgen.pitt.edu>) (23) was used to

ensure that the SNP genotype data in each family demonstrated a Mendelian inheritance pattern. Merlin software (www.sph.umich.edu/csg/abecassis/Merlin) (24) was used to phase the SNP genotype data from families into haplotypes. For those SNPs for which phase was ambiguous for an individual, we excluded that haplotype from analysis for that specific SNP. Using AFBAC (affected family based control) methodology, chromosomes were assigned case status if they were present in any family member who was anti-islet autoantibody positive (Illumina 2) or diagnosed with diabetes (Illumina 2 and Illumina 3) (25). Chromosomes were assigned control status if they were in unaffected families or were not carried by any affected family members (25).

SNP allele frequencies from case versus control parental chromosomes were analyzed with the Fisher's two-tailed exact test (25). To verify that statistically significant findings in Illumina 3 were not caused by population stratification, the Transmission Disequilibrium Test (TDT) was also computed using the mapping panel data from Illumina 3 with Haploview software (26) (<http://www.broad.mit.edu/mpg/haploview>). Prism software (www.graphpad.com/prism) was used to perform survival curve analysis, using the log-rank test with the alpha level for significance set at 0.05. SAS software (www.SAS.com) was used to perform logistic regression analysis.

RESULTS

The first two projects performed involved association analyses of MHC SNPs in families with type 1A diabetes from the Diabetes Autoimmunity Study of the Young (DAISY: a study of prospectively followed HLA-characterized children at risk for type 1 diabetes who were selected for prospective follow-up after HLA typing more than 30,000 newborns) and the Human Biological Data Interchange (HBDI) study. SNP association

analyses were performed by comparing allele frequencies in case versus control chromosomes using AFBAC methodology (27). In a survey study of 45 DAISY families, the "Illumina 1" study, we detected a cluster of 6 SNPs that were nominally associated with type 1 diabetes ($p < 0.05$) in a region telomeric of the *HLA-F* locus (29.2-29.5 Mb) at the telomeric end of the classic MHC. We were particularly interested in this region with significant SNPs given its long distance (over 3 million nucleotides) from the diabetes-associated *HLA-DR/DQ* alleles.

We performed a follow-up project in a larger panel of DAISY and HBDI families, the "Illumina 2" study, genotyping a greater density of SNPs telomeric of the *HLA-F* gene, and with extension of the SNP analysis 6 Mb farther telomeric of the MHC. In the Illumina 2 study, we also genotyped 12 sentinel SNPs reported to define the 8.1 haplotype (boxed in Figure 1) and SNPs in the MHC region that were significantly associated with diabetes and autoimmunity in the Illumina 1 study (17). We again found a cluster of nominally significant SNPs telomeric of the *HLA-F* gene, some with far more significant p-values (Figure 1, seven SNPs with $p < 1.0 \times 10^{-5}$). The most significant SNP in this cluster, rs389419, is located 2 kb telomeric of the ubiquitin D (*UBD*) gene. As expected, many of the SNP alleles of the 8.1 consensus haplotype were significantly associated with case chromosomes given almost complete linkage disequilibrium with *HLA-DRB1*0301*, an allele associated with high risk for type 1A diabetes.

For more convincing replication, we compared MHC SNP allele frequencies in case and control chromosomes from family data provided by the T1DGC from three countries (United Kingdom, United States and Denmark), "Illumina 3" study. In this analysis of the T1DGC data, rs389419 was again significantly associated with diabetes ($p = 2.3 \times 10^{-12}$, Figure 2). The most significant

SNP telomeric of the *HLA-F* gene was rs1233478, located 48 kb telomeric of the *UBD* gene and 23 kb centromeric of the *MASIL* gene [$p=2.7 \times 10^{-23}$, odds ratio=2.2 (95% CI, 1.9, 2.6), Figure 2]. Allele “A” of rs1233478 is present in 27.5% (817/2967) of case chromosomes and 14.7% (228/1550) of control chromosomes. Analysis of the T1DGC data for distorted transmission of alleles [transmission disequilibrium test (TDT)] showed that the “A” allele for rs1233478 was significantly over-transmitted to children with diabetes ($p=2.9 \times 10^{-27}$; 964 transmitted:544 not transmitted). A haplotype with the “A” allele for rs1233478 and alleles from 4 adjacent SNPs was also significantly over-transmitted ($p=4.1 \times 10^{-22}$). The odds ratio for rs1233478 with combined analysis of DAISY, HBDI, and T1DGC data was 2.0 (1.7-2.3, $p=1.6 \times 10^{-23}$). Since rs1233478 was not typed as part of Illumina 2, we subsequently typed it in our Illumina 2 families and found that it was significantly associated with diabetes and autoimmunity [$p=1.1 \times 10^{-6}$, odds ratio=2.2 (1.6, 2.9)].

There were highly significant SNPs close to the *HLA-DRA* gene (e.g. rs3129871, $p=1.9 \times 10^{-15}$ in the Illumina 2 data, $p=1.1 \times 10^{-125}$ in the Illumina 3 data) and also near the *HLA-B* and *HLA-C* genes (e.g. rs2596560, $p=8.8 \times 10^{-51}$ in Illumina 3 data) (Online Appendix Tables 1 and 2 [available at <http://diabetes.diabetesjournals.org>]).

However, we lacked sufficient information to distinguish whether these class II and class I MHC SNPs remained significant independent of linkage disequilibrium with *HLA-DR/DQ* alleles. There was also a significant SNP in the Illumina 2 data near the *PRSS16* gene (rs996247, $p=8.5 \times 10^{-5}$ in the Illumina 2 data), telomeric of the classic MHC (Online Appendix Tables 1 and 2), but this region was not included in the Illumina 3 T1DGC data.

Others have previously identified MHC regions of SNP identity between unrelated individuals using DNA sequence or

microsatellite data (14-16), and we have described a 3 million nucleotide conserved MHC region on the 8.1 haplotype using SNP analyses (18). This means that almost all 8.1 haplotypes from different individuals have close to 100% identity for the 3 million nucleotides of the MHC between the *HLA-DQB1* and *HLA-A* loci. Analysis of the Illumina 3 study data demonstrated that 91% of the 8.1 (HLA-DR3, B8, A1) haplotypes (N=467) had essentially identical SNP alleles for the 3 Mb between *HLA-DQB1* and *HLA-F*, confirming the remarkable conservation of the 8.1 haplotype. The Illumina 2 study extends the SNP analysis over 6 Mb farther telomeric than the Illumina 3 study (Figure 3), and shows that the region of identity between unrelated chromosomes with 8.1 haplotypes gradually breaks up telomeric of *HLA-A*. In a subset of 8.1 chromosomes, the region of identity extends as far as 9 million nucleotides telomeric of the *HLA-DR/DQ* loci. Also of note, the conservation of the case 8.1 haplotypes is similar to the conservation of the control 8.1 haplotypes.

In addition, we found that two sentinel SNPs can be used to identify 8.1 haplotypes (rs17196849 and rs3117573). For 99% of chromosomes with the 8.1 haplotype, rs17196849 and rs3117573 both have “G” alleles, whereas less than 5% of non-8.1 haplotypes have “G” alleles for the two SNPs.

Considering that the 8.1 haplotype has an unusually high frequency for an extended haplotype (present in 18% of case haplotypes) and its consensus sequence is associated with the high-risk *HLA-DRB1*03* allele, we wanted to correct the association analysis of rs1233478 for a possible confounding effect due to linkage disequilibrium with the 8.1 haplotype. The linkage disequilibrium between the “A” allele of rs1233478 and the 8.1 haplotype is moderate, but incomplete (Illumina 3 study: $D'=0.81$, $r^2=0.28$). However, even with exclusion of all of the 8.1 chromosomes (N=467) from the Illumina 3

dataset, the association of the “A” allele of rs1233478 with case chromosomes remained significant (rs1233478, $p=1.4 \times 10^{-12}$).

To evaluate whether rs1233478 accounted for the association in the region or whether other SNPs were independently associated, we performed stepwise logistic regression using the Illumina 3 data. The fifteen most significant *UBD/MASIL* SNPs were included as explanatory variables and a 0.05 level of significance was required for each addition to and retention within the model (28). The SNP with the lowest p-value in the final model with stepwise selection was rs1233478 [$p=6.6 \times 10^{-8}$, odds ratio=1.8 (1.4, 2.2)]. Two other SNPs were included in the model, but were not significant after correction for HLA alleles [rs3131020 $p=0.02$, odds ratio=1.1 (1.0, 1.2) and rs1611527 $p=0.0001$, odds ratio=1.3 (1.2, 1.6)].

We wanted to further account for linkage disequilibrium with other HLA alleles to confirm that the association with rs1233478 was independent of known HLA effects. Stepwise logistic regression analysis was performed including *HLA-DRB1*, *HLA-DQB1*, *HLA-B*, *HLA-A*, and rs1233478 as explanatory variables. The final model included *HLA-DRB1*, *HLA-DQB1*, *HLA-B* and rs1233478, but not *HLA-A* [rs1233478 $p=0.01$, OR=1.3 (1.1, 1.7), Table 2A]. Even though the p-value for rs1233478 is not as striking in this analysis as in our previous analyses and the estimate of effect is attenuated, proper adjustment for all of the HLA alleles required the estimation of many additional parameters, causing the power for the specific test of rs1233478 to be reduced. Details of *HLA-DRB1*, *HLA-DQB1*, *HLA-B*, and *HLA-A* types included as input for the logistic regression analysis can be found in Online Appendix Tables 3-6.

HLA-A24 and *HLA-B39* are both class I MHC alleles that have been reported to contribute high-risk for type 1A diabetes (29,30). Within our the Illumina 3

chromosomes, *HLA-A24* is present in 10% of cases and 7% of controls, *HLA-B39* in 4% of cases and 1% of controls, and the rs1233478 “A” allele in 28% of cases and 15% of controls. We completed a stepwise logistic regression analysis by analyzing *HLA-DRB1*, *HLA-DQB1*, *HLA-B39*, *HLA-A24* and rs1233478 as explanatory variables. The final model includes all of the variables [rs1233478 $p=0.01$, odds ratio=1.3 (1.1, 1.6), Table 2B].

Finally, D6S2223 has been reported to influence type 1A diabetes risk (8,31,32), and therefore logistic regression analysis was performed including D6S2223 as an explanatory variable, in addition to those listed above (with *HLA-B* and *HLA-A* as class variables). All variables were included in the best fitting final model except for *HLA-A* [D6S2223 $p=0.001$, odds ratio=1.4 (1.1, 1.7), rs1233478 $p=0.01$, odds ratio=1.4 (1.1, 1.7)], as in the previous analysis.

The DAISY study is prospectively following a cohort of children with an immediate family history of type 1A diabetes and children from the general population (with no immediate family history of type 1 diabetes) for anti-islet autoimmunity and diabetes. As illustrated in Figure 4, we stratified children with the high-risk *HLA-DR3/4-DQ8* genotype by their genotypes for rs1233478 (A/A, A/C, and C/C), and analyzed their risk for progression to type 1A diabetes. Children with the A/A genotype had a higher risk than children with either the A/C or C/C genotypes (43.7% (± 16.8 SEM, A/A) vs. 13.8% (± 6.1 , A/C) vs. 11.8% (± 3.8 , C/C) by age 12, $p=0.05$, log rank test).

DISCUSSION

The MHC region has a major influence on risk for type 1A diabetes that has been predominantly ascribed to HLA alleles, and, in particular, to *HLA-DR/DQ* alleles. Nevertheless, it is estimated that individuals from the general population with the highest risk

*HLA-DRB1*0301-*

*DQB1*0201/DRB1*04-DQB1*0302 (DR3/4-DQ8)* genotype have only a 5% risk of developing childhood onset type 1A diabetes. This is significantly greater than the approximate general population risk of 0.3%, but significantly less than our recently reported risk of 80% for persistent anti-islet autoantibody expression among the *DR3/4-DQ8* siblings who share both MHC regions identical-by-descent with their proband siblings (13). The *DR3/4-DQ8* siblings who do not share both MHC haplotypes with their proband siblings have a risk of 20%. The high risk in the *DR3/4-DQ8* siblings who share both haplotypes identical-by-descent strongly suggests that additional loci within or linked to the MHC contribute to risk. Our analysis of data from multiple independent populations confirms a major locus marked by SNPs in the *UBD/MASIL* region.

The *UBD* protein (also known as ubiquitin D, diubiquitin, and FAT10) was so-named because it has two tandem head-to-tail ubiquitin-like domains. Ubiquitin proteins are generally known for their ability to bind to proteins, targeting them for degradation by the proteasome. They are also involved in regulation of members of the immunoreceptor family, including B-cell antigen receptors (33). The *UBD* gene is predominantly expressed in B cells and dendritic cells, which are antigen presenting cells that initiate the T cell response that triggers diabetes and these cells also participate in T cell tolerance (34). Ocklenburg et al. recently reported that the *UBD* protein regulates CD4+ T-cell anergy as a downstream element of the *FOXP3* protein, possibly by upregulating *LGALS3* protein expression (35). Expression studies of the *UBD* gene show that the highest mRNA levels are in the two most relevant organs for diabetes, the pancreas and the thymus (<http://t1dbase.org>, tissue expression site). The two other genes in the *UBD/MASIL* region, *MASIL* (Mas-related G-protein coupled receptor MRG (MAS-R)), and

LOC729653 (a gene with a hypothetical protein), have less well characterized expression patterns and functions.

Our extended analysis of SNP alleles of the 8.1 haplotype demonstrates that the remarkable conservation for this common haplotype can extend as far as 9 million base pairs telomeric of the *HLA-DR/DQ* locus. Analysis of 467 8.1 chromosomes showed that more than 91% of 8.1 haplotypes are essentially completely conserved between *HLA-DQB1* and *HLA-F*. We analyzed a subset of 12 SNPs reported by Smith, W.P. et al. to define the 8.1 haplotype (17), and identified two sentinel SNPs (rs17196849 “G” allele and rs3117573 “G” allele) that can be used to identify 8.1 haplotypes with high sensitivity (99%) and specificity (95%).

When we excluded chromosomes with the 8.1 haplotype from the analysis, the association of the *UBD/MASIL* region with diabetes remained significant. This diabetes association of the *UBD/MASIL* region with removal of the 8.1 haplotypes indicates that the differences in allele frequencies between the case and control chromosomes in the *UBD/MASIL* region are independent of linkage disequilibrium with the 8.1 haplotype.

An important aspect of finding a major diabetes-associated region is that it has the potential to improve genetic prediction of type 1A diabetes. Survival curve analysis of children with the high-risk *DR3/4-DQ8* genotype indicates that children with the A/A genotype for rs1233478 in the *UBD/MASIL* region have a higher risk for type 1A diabetes by the age of 12 than children with an A/C or C/C genotype [43.7% (A/A) vs. 13.8% (A/C) vs. 11.8% (C/C)].

We have concentrated on the *UBD* region signal, because it was confirmed in multiple populations, and is not caused by linkage disequilibrium with class I and class II HLA alleles. However, it is likely that other MHC region alleles/SNPs influence risk for type 1A diabetes. Specific *HLA-B* alleles, *HLA-A*

alleles (e.g. *HLA-A24*) and a microsatellite (D6S2223) have been associated with diabetes in previous reports. In our data, there were clusters of diabetes-associated SNPs near, but not contiguous with, *HLA-B* and *HLA-A*, and a very strong signal at *HLA-DRA* (Online Appendix Tables 1 and 2). Stepwise logistic regression analysis with *HLA-DR*, *HLA-DQ*, *HLA-A24*, *HLA-B39*, D6S2223 and the *UBD* SNP rs1233478 showed that rs1233478 has an independent effect on diabetes risk. Of note, a SNP marking the *ITPR3* gene, rs2229634, was not significantly associated with type 1A diabetes ($p=0.4$). Extension of these studies to larger populations as well as finer mapping of SNPs in the *UBD/MASIL* region will be important to assess the full impact of this region on the prediction and pathogenesis of type 1A diabetes.

ACKNOWLEDGEMENTS

This research utilizes resources provided by the Type 1 Diabetes Genetics Consortium,

a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. We thank Elise Eller and Jean Jasinski for their bioinformatics assistance. This work was supported by the National Institutes of Health (DK32083, DK057538), Diabetes Autoimmunity Study in the Young (DAISY, DK32493), Autoimmunity Prevention Center (AI050864), Diabetes Endocrine Research Center (P30 DK57516), Clinical Research Centers (MO1 RR00069, MO1 RR00051), the Immune Tolerance Network (AI15416), the American Diabetes Association, the Juvenile Diabetes Research Foundation, and the Children's Diabetes Foundation.

REFERENCES

1. Wildin,RS, Ramsdell,F, Peake,J, Faravelli,F, Casanova,JL, Buist,N, Levy-Lahad,E, Mazzella,M, Goulet,O, Perroni,L, Bricarelli,FD, Byrne,G, McEuen,M, Proll,S, Appleby,M, Brunkow,ME: X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. *Nat Genet* 27:18-20, 2001
2. Aaltonen,J, Björnses,P, Perheentupa,J, Horelli-Kuitunen,N, Palotie,A, Peltonen,L, Lee,YS, Francis,F, Hennig,S, Thiel,C, Lehrach,H, Yaspo,M-L: An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat Genet* 17:399-403, 1997
3. Halonen,M, Eskelin,P, Myhre,AG, Perheentupa,J, Husebye,ES, Kampe,O, Rorsman,F, Peltonen,L, Ulmanen,I, Partanen,J: AIRE Mutations and Human Leukocyte Antigen Genotypes as Determinants of the Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy Phenotype. *J Clin Endocrinol Metab* 87:2568-2574, 2002
4. Wildin,RS, Smyk-Pearson,S, Filipovich,AH: Clinical and molecular features of the immunodysregulation, polyendocrinopathy, enteropathy, X linked (IPEX) syndrome. *J Med.Genet.* 39:537-545, 2002
5. Concannon,P, Erlich,HA, Julier,C, Morahan,G, Nerup,J, Pociot,F, Todd,JA, Rich,SS: Type 1 diabetes: evidence for susceptibility loci from four genome-wide linkage scans in 1,435 multiplex families. *Diabetes* 54:2995-3001, 2005
6. Baschal,EE, Aly,TA, Babu,SR, Fernando,MS, Yu,L, Miao,D, Barriga,KJ, Norris,JM, Noble,JA, Erlich,HA, Rewers,MJ, Eisenbarth,GS: HLA-DPB1*0402 protects against type 1A diabetes autoimmunity in the highest risk DR3-DQB1*0201/DR4-DQB1*0302 DAISY population. *Diabetes* 56:2405-2409, 2007
7. Sheehy,MJ, Scharf,SJ, Rowe,JR, Neme de Gimenez,MH, Meske,LM, Erlich,HA, Nepom,BS: A diabetes-susceptible HLA haplotype is best defined by a combination of HLA-DR and -DQ alleles. *J Clin Invest* 83:830-835, 1989
8. Lie,BA, Sollid,LM, Ascher,H, Ek,J, Akselsen,HE, Ronningen,KS, Thorsby,E, Undlien,DE: A gene telomeric of the HLA class I region is involved in predisposition to both type 1 diabetes and coeliac disease. *Tissue Antigens* 54:162-168, 1999
9. Lie,BA, Ronningen,KS, Akselsen,HE, Thorsby,E, Undlien,DE: Application and interpretation of transmission/disequilibrium tests: transmission of HLA-DQ haplotypes to unaffected siblings in 526 families with type 1 diabetes. *Am J Hum Genet* 66:740-743, 2000
10. Lie,BA, Thorsby,E: Several genes in the extended human MHC contribute to predisposition to autoimmune diseases. *Curr.Opin.Immunol.* 17:526-531, 2005
11. Nejentsev,S, Gombos,Z, Laine,AP, Veijola,R, Knip,M, Simell,O, Vaarala,O, Akerblom,HK, Ilonen,J: Non-class II HLA gene associated with type 1 diabetes maps to the 240-kb region near HLA-B. *Diabetes* 49:2217-2221, 2000
12. Zavattari,P, Lampis,R, Motzo,C, Loddo,M, Mulargia,A, Whalen,M, Maioli,M, Angius,E, Todd,JA, Cucca,F: Conditional linkage disequilibrium analysis of a complex disease superlocus, IDDM1 in the HLA region, reveals the presence of independent modifying gene effects influencing the type 1 diabetes risk encoded by the major HLA-DQB1, -DRB1 disease loci. *Hum.Mol.Genet.* 10:881-889, 2001
13. Aly,TA, Ide,A, Jahromi,MM, Barker,JM, Fernando,MS, Babu,SR, Yu,L, Miao,D, Erlich,HA, Fain,PR, Barriga,KJ, Norris,JM, Rewers,MJ, Eisenbarth,GS: Extreme Genetic Risk for Type 1A Diabetes. *Proc Natl Acad Sci USA* 103:14074-14079, 2006

14. Alper,CA, Awdeh,Z, Yunis,EJ: Conserved, extended MHC haplotypes. *Exp.Clin.Immunogenet.* 9:58-71, 1992
15. Alper,CA, Awdeh,ZL, Raum,DD, Yunis,EJ: Extended major histocompatibility complex haplotypes in man: role of alleles analogous to murine t mutants. *Clin.Immunol.Immunopathol.* 24:276-285, 1982
16. Alper,CA, Larsen,CE, Dubey,DP, Awdeh,ZL, Fici,DA, Yunis,EJ: The haplotype structure of the human major histocompatibility complex. *Hum Immunol* 67:73-84, 2006
17. Smith,WP, Vu,Q, Li,SS, Hansen,JA, Zhao,LP, Geraghty,DE: Toward understanding MHC disease associations: Partial resequencing of 46 distinct HLA haplotypes. *Genomics* 87:561-571, 2006
18. Aly,TA, Eller,E, Ide,A, Gowan,K, Babu,SR, Erlich,HA, Rewers,MJ, Eisenbarth,GS, Fain,PR: Multi-SNP Analysis of MHC Region: Remarkable Conservation of HLA-A1-B8-DR3 Haplotype. *Diabetes.* 55:1265-1269, 2006
19. Awdeh,ZL, Raum,D, Yunis,EJ, Alper,CA: Extended HLA/complement allele haplotypes: evidence for T/t-like complex in man. *Proc.Natl.Acad.Sci.U.S.A* 80:259-263, 1983
20. Bilbao,JR, Calvo,B, Aransay,AM, Martin-Pagola,A, Perez,dN, Aly,TA, Rica,I, Vitoria,JC, Gaztambide,S, Noble,J, Fain,PR, Awdeh,ZL, Alper,CA, Castano,L: Conserved extended haplotypes discriminate HLA-DR3-homozygous Basque patients with type 1 diabetes mellitus and celiac disease. *Genes Immun* 7:550-554, 2006
21. Rewers,M, Bugawan,TL, Norris,JM, Blair,A, Beaty,B, Hoffman,M, McDuffie,RSJr, Hamman,RF, Klingensmith,G, Eisenbarth,GS, Erlich,HA: Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY). *Diabetologia* 39:807-812, 1996
22. Barker,JM, Barriga,K, Yu,L, Miao,D, Erlich,H, Norris,JN, Eisenbarth,GS, Rewers,M: Prediction of autoantibody positivity and progression to type 1 diabetes: Diabetes Autoimmunity Study in the Young (DAISY). *J Clin Endocrinol Metab* 89:3896-3902, 2004
23. O'Connell,JR, Weeks,DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63:259-266, 1998
24. Abecasis,GR, Cherny,SS, Cookson,WO, Cardon,LR: Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat.Genet.* 30:97-101, 2002
25. Thomson,G: Mapping disease genes: family-based association studies. *Am.J.Hum.Genet.* 57:487-498, 1995
26. Barrett,JC, Fry,B, Maller,J, Daly,MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 21:263-265, 2005
27. Thomson,G: Mapping disease genes: family-based association studies. *Am.J.Hum.Genet.* 57:487-498, 1995
28. Cordell,HJ, Clayton,DG: A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am.J.Hum.Genet.* 70:124-141, 2002
29. Noble,JA, Valdes,AM, Bugawan,TL, Apple,RJ, Thomson,G, Erlich,HA: The HLA class I A locus affects susceptibility to type 1 diabetes. *Hum.Immunol* 63:657-664, 2002
30. Valdes,AM, Erlich,HA, Noble,JA: Human leukocyte antigen class I B and C loci contribute to Type 1 Diabetes (T1D) susceptibility and age at T1D onset. *Hum.Immunol.* 66:301-313, 2005
31. Lie,BA, Todd,JA, Pociot,F, Nerup,J, Akselsen,HE, Joner,G, Dahl-Jorgensen,K, Ronningen,KS, Thorsby,E, Undlien,DE: The predisposition to type 1 diabetes linked to the

human leukocyte antigen complex includes at least one non-class II gene. *Am.J Hum.Genet.* 64:793-800, 1999

32. Johansson,S, Lie,BA, Todd,JA, Pociot,F, Nerup,J, Cambon-Thomsen,A, Kockum,I, Akselsen,HE, Thorsby,E, Undlien,DE: Evidence of at least two type 1 diabetes susceptibility genes in the HLA complex distinct from HLA-DQB1, -DQA1 and -DRB1. *Genes Immun* 4:46-53, 2003

33. Mukhopadhyay,D, Riezman,H: Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science* 315:201-205, 2007

34. Bates,EE, Ravel,O, Dieu,MC, Ho,S, Guret,C, Bridon,JM, Ait-Yahia,S, Briere,F, Caux,C, Banchereau,J, Lebecque,S: Identification and analysis of a novel member of the ubiquitin family expressed in dendritic cells and mature B cells. *Eur.J.Immunol.* 27:2471-2477, 1997

35. Ocklenburg,F, Moharreg-Khiabani,D, Geffers,R, Janke,V, Pfoertner,S, Garritsen,H, Groebe,L, Klempnauer,J, Dittmar,KE, Weiss,S, Buer,J, Probst-Kepper,M: UBD, a downstream element of FOXP3, allows the identification of LGALS3, a new marker of human regulatory T cells. *Lab Invest* 86:724-737, 2006

TABLE 1. Summary of Study Design

Study	Subject Source	Number of Families	Number of Individuals	Illumina SNPs
Illumina 1	DAISY	n=45	n=143	Custom 656
Illumina 2	DAISY HBDI	n=165 n=72	n=772 n=523	Custom 364
Illumina 3	T1DGC BDA, Danish, Joslin, HBDI, UK	n=1,240	n=6,297	Standard 2,918 SNPs

TABLE 2. Logistic Regression Analysis

Part A: Type 3 Analysis of Effects (variables within the model),
HLA-A and *HLA-B* are class variables

Variable	DF	Wald Chi-square	Pr>ChiSq
<i>HLA-DRB1</i>	12	180.6766	3.10E-32
rs1233478	1	6.6617	9.85E-03
<i>HLA-B</i>	28	56.6897	1.06E-03
<i>HLA-DQB1</i>	10	163.4422	6.30E-30

Part B: Type 3 Analysis of Effects (variables within the model), A24 and B39

Variable	DF	Wald Chi-square	Pr>ChiSq
<i>HLA-DRB1</i>	12	215.9578	1.63E-39
rs1233478	1	6.2467	1.24E-02
A24	1	9.0515	2.62E-03
B39	1	12.786	3.49E-04
<i>HLA-DQB1</i>	10	194.6782	2.08E-36

FIGURE LEGENDS:

Figure 1. AFBAC results for Illumina 2 (DAISY and HBDI). The x-axis represents the distance of each SNP from the telomere, and the y-axis represents the $-\log(p)$ value for each SNP. The circled SNPs represent the $-\log(p)$ values for the most significant SNPs telomeric of *HLA-F*. The most significant SNP, rs389419, is 2 kb telomeric of the *UBD* gene. The SNPs surrounded by boxes (N=12) are sentinel SNPs that identify and are specific for the 8.1 haplotype.

Figure 2. AFBAC results for Illumina 3 (T1DGC). Comparison of SNP allele frequencies in 3,138 case and 1,681 AFBAC control chromosomes from T1DGC data indicates that a SNP between the *UBD* and *MAS1L* genes is significantly associated with type 1A diabetes (rs1233478, $p=2.7 \times 10^{-23}$).

Figure 3. *HLA-DRB1*03*, *HLA-B*08*, *HLA-A*01* (8.1) haplotypes are highly conserved in the MHC region. Data from chromosomes with the 8.1 haplotype from the Illumina 2 study were placed in columns of Microsoft Excel™ (N=237) with a row for each SNP analyzed (N=364). The allele for each SNP of each chromosome was compared with alleles of the consensus 8.1 SNPs. Yellow highlights SNPs identical to alleles of the consensus 8.1 SNPs, dark grey highlights variant SNPs different from the consensus 8.1 SNPs, and white highlights unphased or untyped SNPs. As depicted by these 237 8.1 haplotypes, 8.1 haplotypes are highly conserved for over 3 million nucleotides of the MHC region and the conservation extends up to 6 million nucleotides farther in the telomeric direction.

Figure 4. Survival analysis stratified by rs1233478 genotype. These survival curves illustrate the risk over time for type 1A diabetes associated with rs1233478 genotypes in children with the *DR3/4-DQ8* genotype. At the first time point, age 0 years, 100% of the children are “diabetes-free”. The percentages of children in each rs1233478 genotype group that remain “diabetes-free” changes as the children are followed over time. The *DR3/4-DQ8* children with the A/A genotype for rs1233478 have a diabetes risk of 43.7% (± 16.8 SEM, A/A) by age 12, versus a 13.8% (± 6.1) risk in *DR3/4-DQ8* children with the A/C genotype, and an 11.8% (± 3.8) risk with the C/C genotype ($p=0.051$, log rank test). Filled squares = AA genotype for rs1233478; filled triangles = AC genotype for rs1233478; open circles = CC genotype for rs1233478; Error bars= standard error of the mean. The numbers below the ages on the x-axis indicate the number of individuals being followed from each genotype cohort at different ages.

FIGURE 1

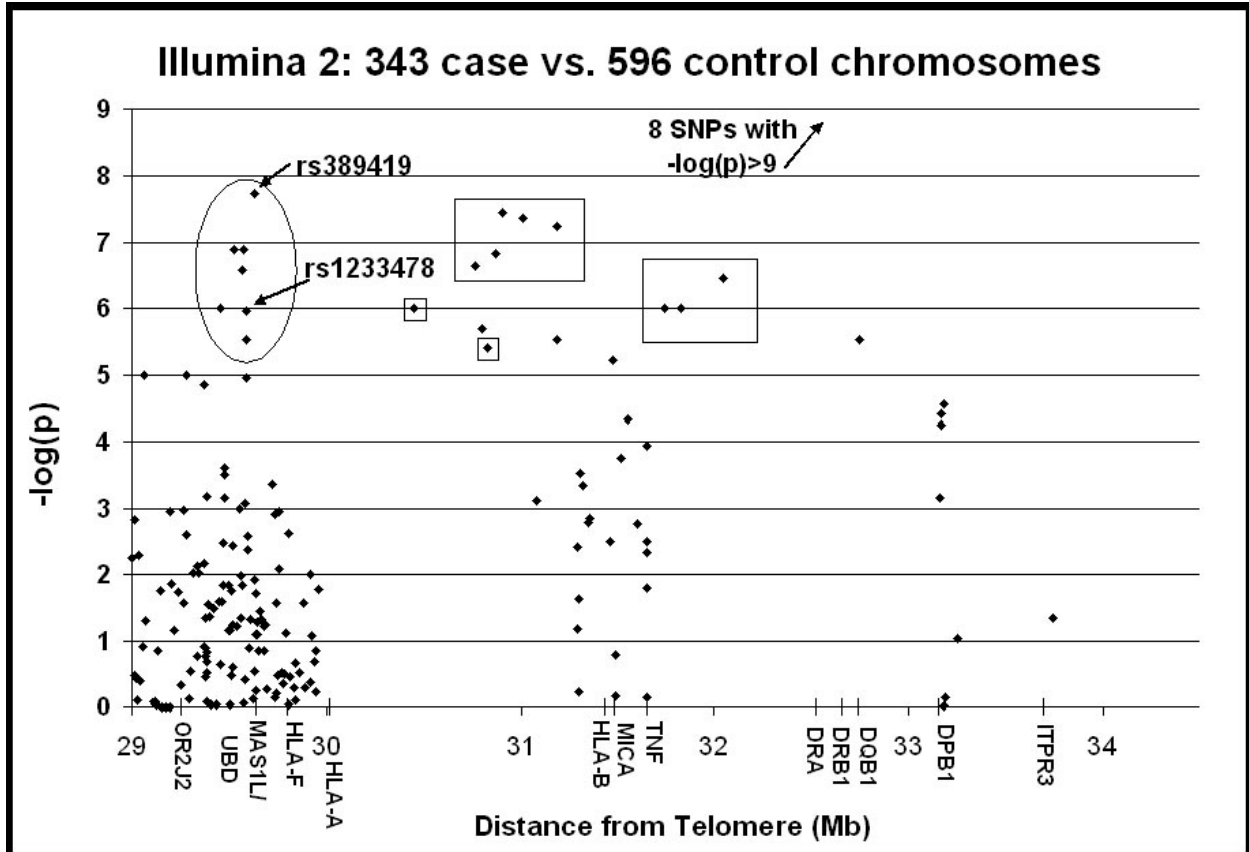


FIGURE 2

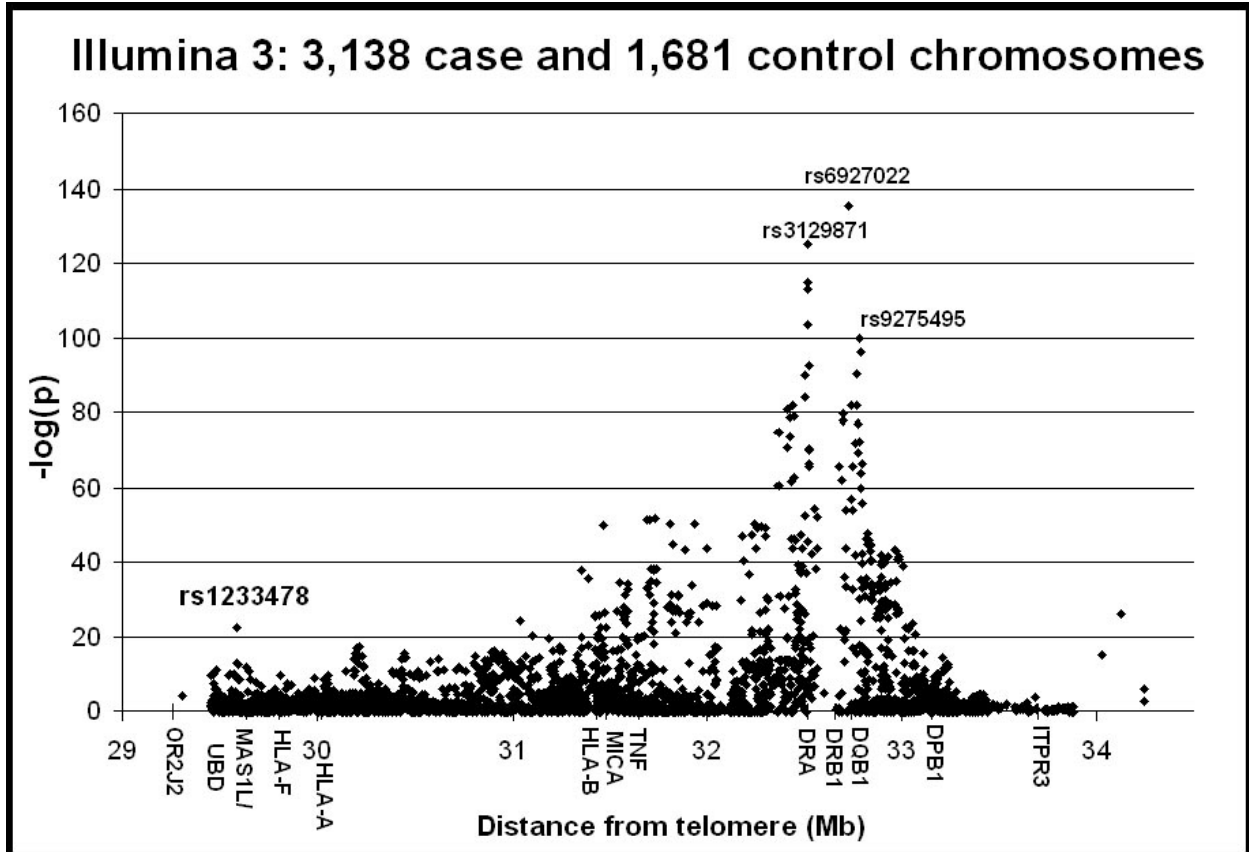


FIGURE 3

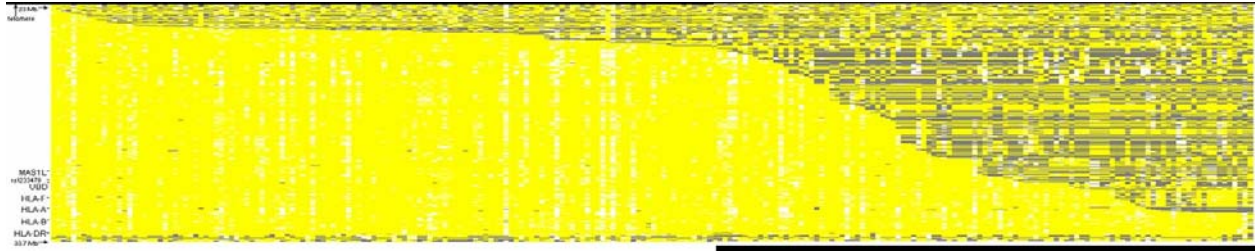
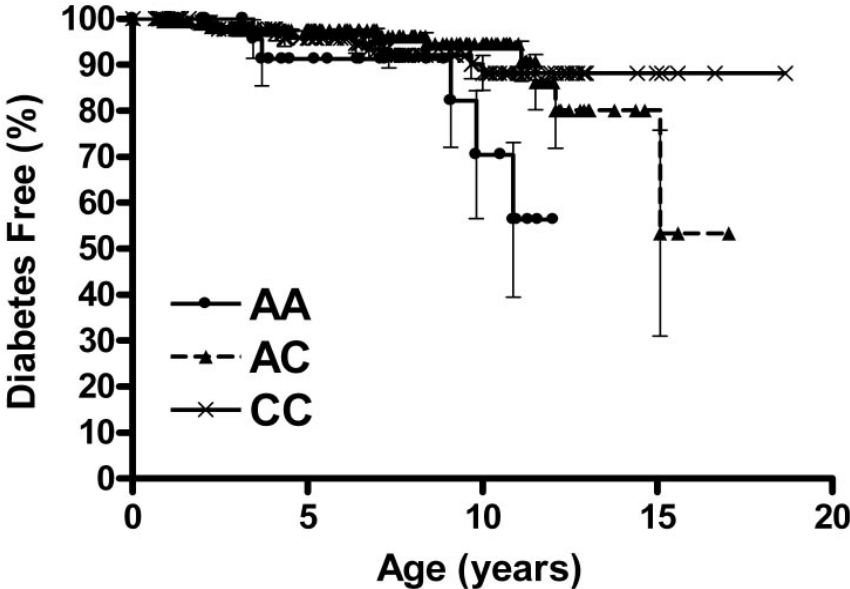


FIGURE 4



AA	32	18	6	0	0
AC	185	108	44	10	0
CC	158	98	45	13	0

p=0.051
 AA: 5 cases, AC: 10 cases, CC: 10 cases