Genes and Immunity (2012), 1–7

www.nature.com/gene

ORIGINAL ARTICLE

A strategy for combining minor genetic susceptibility genes to improve prediction of disease in type 1 diabetes

C Winkler[1,2], J Krumsiek[3], J Lempainen[1,4], P Achenbach[1,2], H Grallert[5], E Giannopoulou[1,2], M Bunk[1,2], FJ Theis[3], E Bonifacio[2,6,7] and A-G Ziegler[1,2,7]

Genome-wide association studies have identified gene regions associated with type 1 diabetes. The aim of this study was to determine how the combined allele frequency of multiple susceptibility genes can stratify islet autoimmunity and/or type 1 diabetes risk. Children of parents with type 1 diabetes and prospectively followed from birth for the development of islet autoantibodies and diabetes were genotyped for single-nucleotide polymorphisms at 12 type 1 diabetes susceptibility genes (*ERBB3, PTPN2, IFIH1, PTPN22, KIAA0350, CD25, CTLA4, SH2B3, IL2, IL18RAP, IL10* and *COBL*). Non-human leukocyte antigen (HLA) risk score was defined by the total number of risk alleles at these genes. Receiver operator curve analysis showed that the non-HLA gene combinations were highly effective in discriminating diabetes and most effective in children with a high-risk HLA genotype. The greatest diabetes discrimination was obtained by the sum of risk alleles for eight genes (*IFIH1, CTLA4, PTPN22, IL18RAP, SH2B3, KIAA0350, COBL* and *ERBB3*) in the HLA-risk children. Non-HLA-risk allele scores stratified risk for developing islet autoantibodies and diabetes, and progression from islet autoimmunity to diabetes. Genotyping at multiple susceptibility loci in children from affected families can identify neonates with sufficient genetic risk of type 1 diabetes to be considered for early intervention.

Genes and Immunity advance online publication, 30 August 2012; doi:10.1038/gene.2012.36

Keywords: type 1 diabetes; type 1 diabetes susceptibility genes; islet autoimmunity

INTRODUCTION

Genome-wide association studies have identified numerous gene regions that confer susceptibility to type 1 diabetes.[1,2] With the exception of the human leukocyte antigen (HLA) and the *INS* gene regions, the contribution of any single locus to type 1 diabetes susceptibility is relatively small.[3–5] Therefore, although informative with respect to disease mechanisms and drug targets, identification of these susceptibility genes has provided relatively little improvement in disease prediction.[1–5] Moreover, it appears that there may be differences between cohorts in how much single gene regions contribute to risk and prediction.[6–9] It is conceivable that although contributions of single genes may be small and inconsistent, the sum of susceptible alleles from multiple gene regions could provide a more universal risk estimate and provide sufficient risk so as to be helpful for risk stratification.[10] Here we have examined how the combination of single-nucleotide polymorphisms (SNPs) within 12 type 1 diabetes-associated gene regions is associated with the initiation of autoimmunity and development of diabetes. We further applied a novel analysis of combinatorial gene risk assessment that aided in identifying gene combinations for disease risk stratification. We have examined prospectively followed first-degree relatives of patients with type 1 diabetes where we expect genetic contributions to disease development may be enhanced. Moreover, the prospectively followed cohort allowed us to examine effects on both the development of islet autoimmunity and progression to type 1 diabetes. The findings indicate that models in which the total genetic load is quantified will be valuable in stratifying type 1 diabetes risk in neonates.

RESULTS

The *CTLA4 SNP rs3087243* and the previously reported *IFIH1* SNP *rs2111485*(ref. 6) were associated with diabetes development in the BABYDIAB cohort (Supplementary Table 1). Although many of the SNPs in the other 10 tested gene regions had hazard ratios >1, none was significantly associated with diabetes development when examined alone in the cohort. We therefore asked whether combining information from multiple SNPs would aid risk stratification. For all 12 gene SNPs, a score of 2 was given if the child was homozygous for the susceptible allele, 1 if heterozygous and 0 if homozygous for the non-susceptible allele. The sum of the scores for the 12 genes was assigned as the combined risk score for each child. Scores ranged from 6 to 21 (median, 14; interquartile range, 12–15). The distribution of the combined risk sores was different in the 47 children who developed type 1 diabetes (mean, 14.7 ± 2.4) as compared with the 1243 who had not developed diabetes (mean, 13.6 ± 2.2; $P = 0.001$; Figure 1).

Receiver operator curve (ROC) analysis was performed on the combined risk scores in order to identify thresholds that may be

[1]Institute of Diabetes Research, Helmholtz Zentrum München and Forschergruppe Diabetes, Klinikum rechts der Isar, Technische Universität, München, Germany; [2]Forschergruppe Diabetes e.V. Neuherberg, Germany; [3]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Neuherberg, Germany; [4]Department of Pediatrics, Immunogenetics Laboratory, University of Turku, Turku, Finland; [5]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany and [6]Center for Regenerative Therapies-Dresden, Dresden University of Technology, Dresden, Germany. Correspondence: Professor Dr A-G Ziegler, Institute of Diabetes Research, Helmholtz Zentrum München and Forschergruppe Diabetes, Klinikum rechts der Isar, Technische Universität, Ingolstaedter Landstrasse 1, Neuherberg 85764, Germany or Professor Dr E Bonifacio, Center for Regenerative Therapies-Dresden, Dresden University of Technology, Fetscherstrasse 105, Dresden 01307, Germany.
E-mail: anette-g.ziegler@helmholtz-muenchen.de or ezio.bonifacio@crt-dresden.de
[7]Joint senior authorship.
Received 1 May 2012; revised 18 July 2012; accepted 23 July 2012

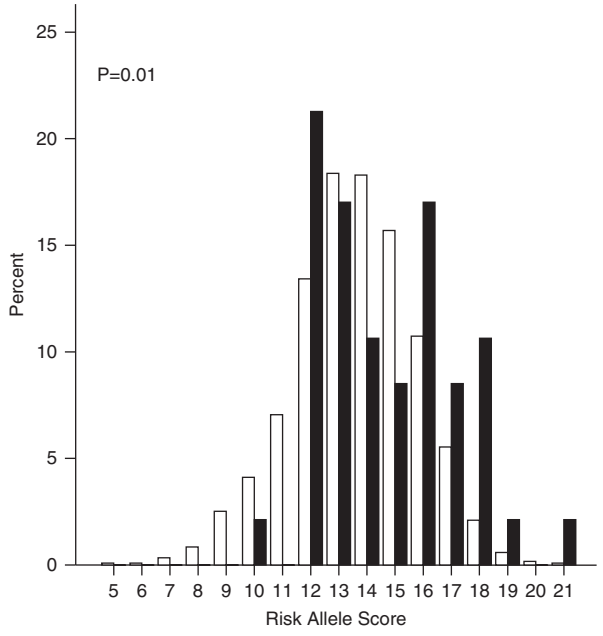

Figure 1. Non-HLA gene SNP-risk allele score distribution. Distribution of risk allele scores derived from 12 type 1 diabetes susceptibility genes (*ERBB3 rs705704, PTPN2 rs1893217, IFIH1 rs2111485, PTPN22 rs6679677, KIAA0350 rs12935413, CD25 rs11594656, CTLA4 rs3087243, SH2B3 rs3184504, IL2 rs4505848, IL18RAP rs917997, IL10 rs3024505* and *COBL rs4948088 SNPs*) in nondiabetic islet autoantibody-negative BABYDIAB children (unfilled bars) and in BABYDIAB children who progressed to type 1 diabetes (filled bars).

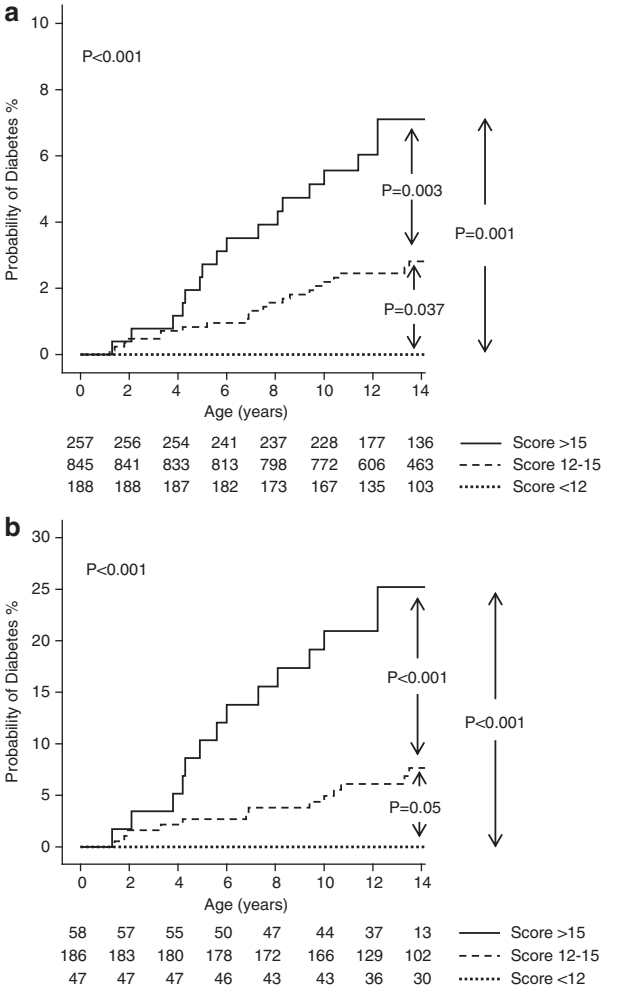

Figure 2. Cumulative risk for the development of type 1 diabetes by the combined risk allele score. Kaplan–Meier analysis of diabetes development in the BABYDIAB children categorized on the basis of their combined risk allele score derived from the 12 susceptibility genes analyzed. Categories are <12 risk alleles (dotted line), 12–15 risk alleles (dashed line) and >15 risk alleles (solid line). (**a**) Children from the whole-genotyped cohort. (**b**) Children who carry high type 1 diabetes risk HLA genotypes as defined by the TEDDY study.[19] *P*-values are calculated by the log-rank test. Numbers under the abscissa are the number of children still followed at each time point.

helpful in type 1 diabetes risk stratification (Supplementary Figure 1). Points of inflexion in the ROC curve could be observed at scores of <12 and >15, representing the lower 14.6% and upper 19.9% of scores, respectively. As predicted from the ROC analysis, these thresholds were able to stratify diabetes risk in the cohort from 0% by 14 years in children with scores <12, 2.8% (95% confidence interval (CI) 1.6–4.0; $P=0.037$ vs <12) by 14 years in children with scores 12–15 and 7.1% (95% CI 3.8–10.4; $P=0.001$ vs <12; $P=0.003$ vs 12–15) by 14 years in children with scores >15 (Figure 2a). Diabetes risk stratification was particularly strong in children who also had HLA-risk genotypes (Figure 2b, Table 1). Overall, 58 (4.5%) of the 1290 children with full genotypes had high-risk HLA genotypes plus a combined risk score of >15. The risk for type 1 diabetes was 25.2% (95% CI 13.5–36.9) by age 14 years in these 58 children.

Because the total set of genes examined proved effective in type 1 diabetes risk stratification, we asked whether there were particular combinations that were important for diabetes risk prediction in the cohort. The total number of combinations with the 12 SNPs is 4095. To determine which combinations and which genes were likely to be most useful in stratifying risk, ROC plots were performed and the area under the curve (AUC) calculated for all 4095 combinations in the total cohort and in the 291 children with HLA-risk genotypes (Figure 3, Supplementary Figure 2). ROC analysis was performed for both diabetes and islet autoantibodies (positive for insulin autoantibody (IAA), glutamic acid decarboxylase (GAD), insulinoma antigen 2 (IA-2) or zinc transporter 8 (ZnT8) autoantibodies in at least two consecutive samples) as outcomes. First, we reasoned that if the AUC of the 4095 combinations were collectively significantly >0.5, the gene set was likely to be helpful in discriminating type 1 diabetes or islet autoantibody risk. Accordingly, the median AUC for ROCs using diabetes outcome was 0.588 ($P=0.001$) in the whole cohort and 0.656 ($P<0.001$) in the HLA-risk children, and the median AUC using islet autoantibody outcome was 0.521 ($P=0.945$) in the whole cohort and 0.573 ($P=0.02$) in the HLA-risk children (Supplementary Figure 2). AUCs were greater when diabetes was used as the outcome as compared with islet autoantibodies ($P<0.0001$), suggesting that the genes were more relevant to the development of diabetes than to the development of islet autoimmunity. AUCs were also significantly higher in children with high-risk HLA than in the total cohort ($P<0.0001$), suggesting that some gene effects were dependent upon the presence of HLA-susceptible genotypes and/or that their effect on susceptibility is more pronounced in children with high *a priori* risk.

Next, we anticipated that some genes would be more useful in combinations than others and examined the prevalence of individual gene SNPs in the combinations that yielded highly significant AUCs. Because the highest risks were obtained for diabetes outcome in HLA-risk children, we used this as the outcome and data set to establish the model. Representation of the genes included in combinations as a function of AUC *P*-values

Table 1. Performance of 12 gene-combined risk score categories in identifying children who developed diabetes among 291 HLA risk in the BABYDIAB cohort

Risk category	*Sensitivity[a] cases, (%)*	*Specificity[a] % (95% CI)*	*Positive PV[b] % (95% CI)*	*Negative PV[b] % (95% CI)*
Low (score < 12), $n = 47$	0	82% (77–86)	0%	100%
Moderate (12–15), $n = 186$	16 (52%)	35% (29–41)	7.6 (3.5–11.7)	92.4% (88.3–96.5)
High (> 15), $n = 58$	15 (48%)	83% (78–87)	25.2% (13.5–36.9)	74.8% (63.1–86.5)

Abbreviations: CI, confidence interval, HLA, human leukocyte antigen; PV, predictive value. [a]Sensitivity and specificity are calculated from the number of children who did or did not develop diabetes without considering time of follow-up. Because a number of children were not followed for 14 years, the values may change with more follow-up. [b]Positive and negative PVs are derived from the actual time to event diabetes risks until 14 years of age.

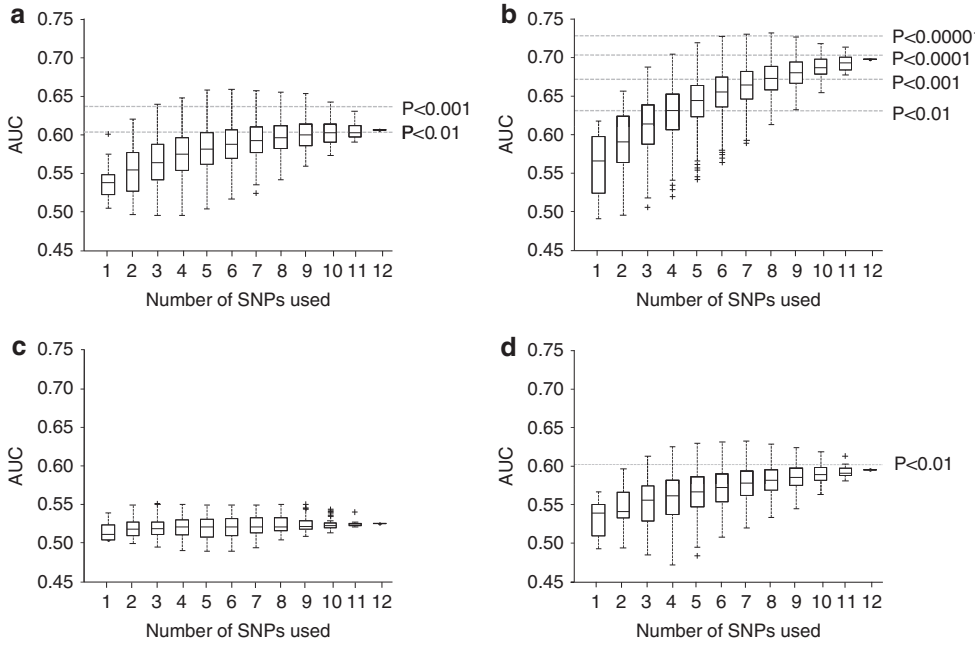

Figure 3. AUC for all 4095 possible combinations of the 12 SNPs. AUC calculated from the ROC analysis for all 4095 possible combinations of the 12 SNPs using diabetes as outcome in all children (**a**), diabetes as outcome in children with HLA-risk genotypes (**b**), islet autoimmunity as outcome in all children (**c**) and islet autoimmunity as outcome in children with the HLA-risk genotypes (**d**). Values are represented as box plots showing median and 25th to 75th percentiles in boxed regions, and the extended bars indicate the highest and lowest points in the data that are not considered outliers (that is, within 1.5 times the interquartile range above or below the box, respectively). The horizontal gray dashed lines represent the AUC for the listed *P*-values.

is shown in Figure 4a. SNPs from four genes (*CTLA4, PTPN22, IL18RAP* and *IFIH1*) were almost always present in combinations that gave highly significant AUCs (Figure 4a). In contrast, the *IL2* gene SNP was infrequent in such combinations. Overall, there were 161 (3.9%) gene combinations that resulted in AUC with *P*-values < 0.0001 for diabetes as an outcome in the HLA-risk children. These 161 combinations contained between 4 and 11 of the 12 gene SNPs. Heat maps show which genes were represented in these 161 gene combinations (Figure 4b). SNPs in the *CTLA4* gene (97% of the combinations), *PTPN22* (92%), *IL18RAP* (92%), and *IFIH1* (84%) were highly represented. *INS* gene polymorphisms are also associated with type 1 diabetes risk and have hazard ratios that exceed those of the 12 genes examined in the current study.[2,4,11] We therefore added the *INS* genotype to the 12 SNPs and performed ROC analysis on the 8191 possible combinations. Remarkably, 1117 combinations had an AUC with $P < 0.0001$ and 117 combinations had an AUC with $P < 10^{-5}$ (Supplementary Figure 3). Again, the *CTLA4* gene (99% of the combinations), *PTPN22* (97%), *IL18RAP* (91%) and *IFIH1* (89%) were highly represented in these 117 combinations, together with the *INS* gene (97%).

Having demonstrated that the ROC analysis was able to identify gene combinations that were strongly associated with type 1 diabetes, we examined risk stratification for the gene set yielding the highest AUC. The maximum AUC achieved for any of the 12 gene SNP combinations was 0.732 ($P = 7 \times 10^{-6}$), and included eight genes (*IFIH1, CTLA4, PTPN22, IL18RAP, SH2B3, KIAA0350, COBL, ERBB3*) for diabetes outcome in the children with HLA-risk genotypes. Discrimination between children who did and did not develop diabetes was significantly better with this set of eight genes than with the combination of all 12 genes ($P = 0.046$). The ROC curve for the highest diabetes-risk gene combination was analyzed and categories for high intermediate and low risk were assigned based upon sensitivity/specificity trade-off (Supplementary Figure 4). Risk stratification with these categories was effective for diabetes (Figure 5a) with risk ranging from 0% by age 14 years for children in the low-risk score category to 26.9% (95% CI, 15.2–38.6%) for children in the high-risk score category ($P < 0.0001$ high-risk vs moderate-risk scores; $P = 0.004$ high-risk vs low-risk scores). The high-risk score category identified 65 (22%) of the genotyped HLA-risk children, including 16 (55%) of the 29 HLA-risk children who developed diabetes by 14 years of age. The

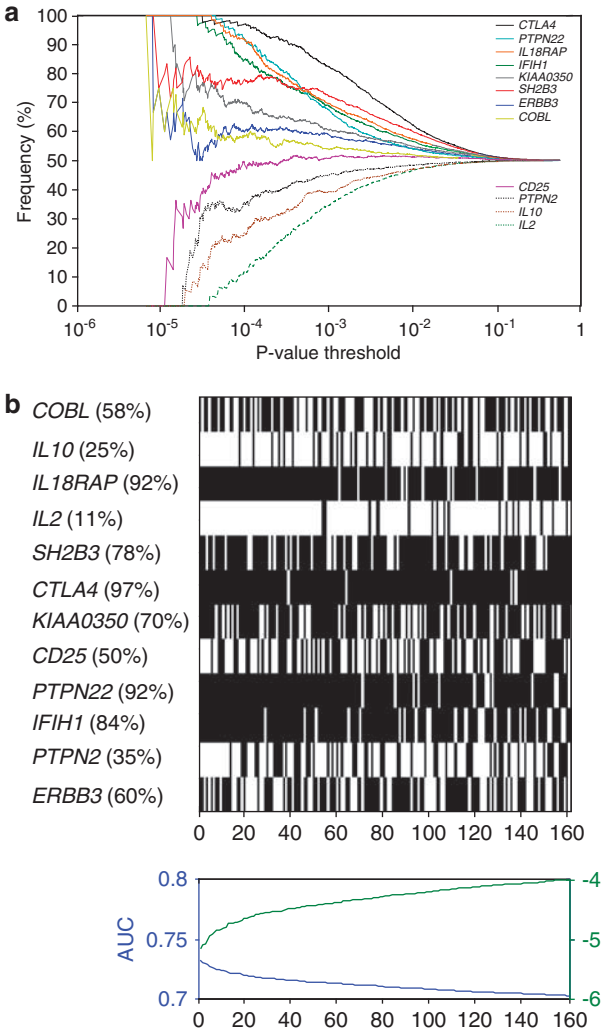

Figure 4. Frequency of each SNP in combination ranked upon *P*-values. (**a**) The frequency of each gene (Y-axis) found in combinations that yielded ROC AUC *P*-values below the value indicated on the X-axis. For example, the *CTLA4* gene (highest line, black) was found in 100% of combinations with AUC *P*-values $<5\times10^{-4}$, whereas the *IL2* gene (lowest line, dashed light green) was found in 0% of combinations with AUC *P*-values $<5\times10^{-4}$. The analysis is done for diabetes outcome in children who carry high type 1 diabetes risk HLA genotypes as defined by the TEDDY study.[19] The lines representing each of the genes are indicated on the figure. (**b**) The combinations of the 12 genes that resulted in ROC AUCs with *P*-values <0.0001 when diabetes was used as outcome in children with HLA-risk genotypes ($n=161$ combinations) are shown as heat maps. The heat maps represented the 161 combinations horizontally for each of the 12 genes listed vertically. Combinations are arranged from the highest (left) to the lowest (right) AUC, and the corresponding AUC and $-\log_{10}$ *P*-values are provided in the lower panel. The presence of the gene in a combination is indicated as a black bar. The frequency of the gene in the 161 combinations is indicated.

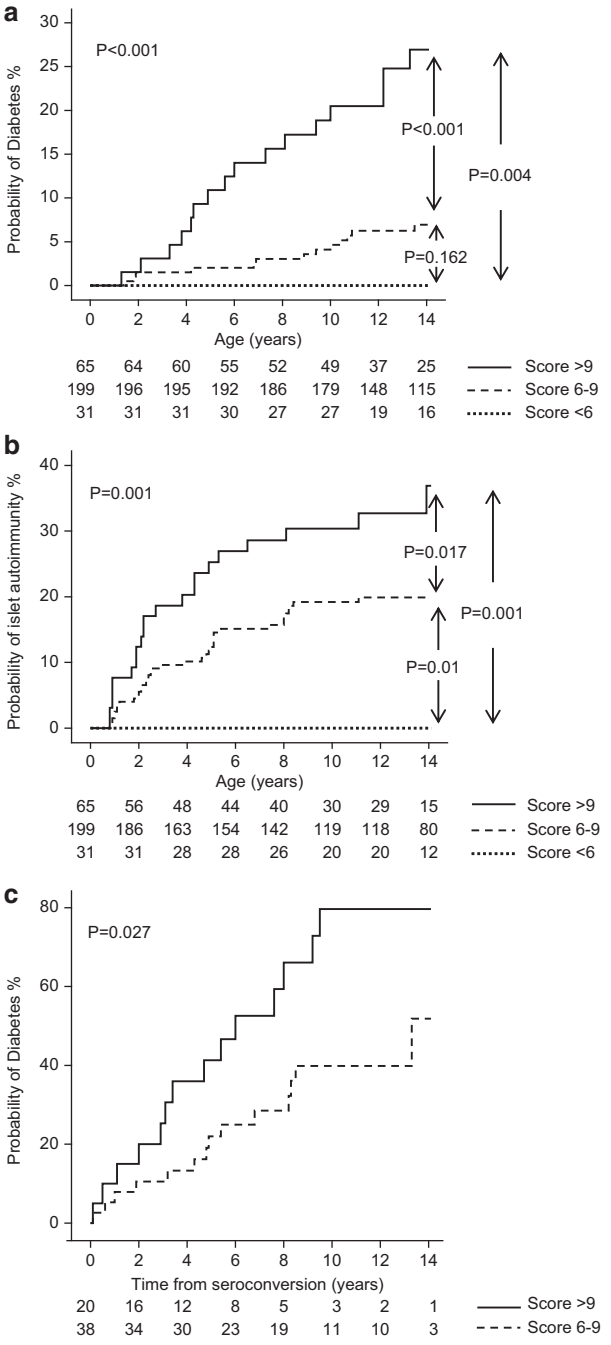

Figure 5. Cumulative risk for the development of type 1 diabetes and islet autoantibodies by the eight gene-combined risk allele score. Kaplan–Meier analysis of diabetes development (**a**), islet autoantibody development (**b**) and progression from the age of islet autoantibody development to diabetes in islet autoantibody-positive children (**c**). The analysis was performed in the BABYDIAB children who carry high type 1 diabetes risk HLA genotypes as defined by the TEDDY study.[19] The eight genes used to calculate scores were those that provided the highest ROC AUC for diabetes outcome in these children (*IFIH1, CTLA4, PTPN22, IL18RAP, SH2B3, KIAA0350, COBL* and *ERBB3*). Categories are <6 risk alleles (dotted line), 6–9 risk alleles (dashed line) and >9 risk alleles (solid line). *P*-values are calculated by the log-rank test. Numbers under the abscissa are the number of children still followed at each time point.

gene-combination scores also stratified type 1 diabetes risk in the children with the highest risk HLA DR3/DR4-DQ8 genotype ($P=0.01$, data not shown).

Finally, the children were prospectively followed for the development of islet autoantibodies and diabetes, providing the opportunity to assess where the combined set of genes affected the preclinical pathogenesis. The three categories significantly modified risk for islet autoantibodies (low scores, 0% by 14 years of age; moderate scores, 20%; and high scores, 37%; $P=0.001$; Figure 5b), and for progression to diabetes from the first appearance of islet autoantibodies (moderate scores, 40% by

10 years; high scores, 80% by 10 years; $P = 0.03$; Figure 5c), suggesting that this combination of eight genes affects both the onset of islet autoimmunity and progression to diabetes.

DISCUSSION

Allelic variations at numerous loci affect type 1 diabetes susceptibility.[1,2,4] Apart from HLA class II genes and the *INS* gene, the remainder are associated with relatively small odds ratios.[1–5] We considered that the risk conferred by these loci may be cumulative and show here that the sum of risk alleles for 12 type 1 diabetes susceptibility SNPs could stratify risk for disease in first-degree relatives, and in particular when combined with HLA class II genotype. We further provide a model for identifying combinations of genes to obtain maximal disease risk stratification. Using this model, we show that scores derived from combinations of genes provided significant increased discrimination over that which could be achieved by any single gene, and that the highest discrimination was obtained when combining HLA susceptibility with SNPs from 8 of the 12 susceptibility genes.

The study was intended to provide a proof of principle. It is not exhaustive in the type 1 diabetes susceptibility genes examined and it was performed in prospectively followed first-degree relatives of patients. Although it cannot be assumed that findings in relatives will also apply to children from unaffected families, relatives are currently the main contenders for recruitment into prevention studies and therefore of practical relevance.[12] Furthermore, because the majority of these type 1 diabetes susceptibility genes were identified in unselected case–control cohorts,[1,2] we suggest that inclusion of combinations of these and other SNPs will also improve the identification of at-risk neonates with no *a priori* family history of type 1 diabetes. A limitation of our study is the number of children who developed diabetes during follow-up and the absence of a confirmatory cohort. Therefore, it will be important to test the strategy in other cohorts. Although larger cohorts may not identify the same gene combinations and risk scores as those described for the BABYDIAB cohort, we expect the strategy to be robust and that the general findings will be confirmed. Another limitation is that only 12 non-HLA type 1 diabetes susceptibility genes were typed and analyzed. The findings may have differed with respect to the number and identity of genes that stratify type 1 diabetes risk and the degree of stratification that can be achieved if additional type 1 diabetes susceptibility genes were included.

The study has a number of important strengths. First, it provides a novel approach to maximize the effectiveness of combinations of risk factors for disease prediction. Here we apply it to gene combinations. Importantly, over 100 different combinations from 12 gene SNPs plus HLA could discriminate future patients from controls with $P < 0.0001$. Some genes were frequent in these high-discrimination combinations (*CTLA4*, *PTPN22*, *IL18RAP* and *IFIH1*), and indeed these four genes alone could discriminate cases from controls at relatively high *P*-value (0.00009). The addition of *INS* gene polymorphisms provided even greater discrimination. It is notable, however, that a number of combinations that included variable gene sets with respect to the number and the identify of genes achieved high ROC AUCs. It is also notable that the maximum discrimination was achieved by HLA plus eight of the non-HLA genes, which is less than the full complement of genes analyzed. These observations are relevant to the application of risk gene combination to other populations. We predict that different combinations of genes may achieve high discrimination in ethnically or environmentally diverse populations. A limitation of the ROC analysis on gene combinations we described is that in its current form, it does not consider time to event and therefore loses some of the benefit of prospective follow-up data, its assessment of performance and identification of thresholds. Second, scores did not consider dominant and recessive effects of susceptible alleles, nor the weight of genes in conferring susceptibility. Incorporating modifications to consider these aspects may enhance the performance of the approach.

A second strength of our study is that it applies risk stratification to a prospectively followed cohort, in which timing of autoimmunity and diabetes onset is monitored. This allowed assessment of effects on both development of autoimmunity and progression to diabetes. We have previously seen effects of genes, such as *IFIH1*, on progression rather than the development of autoimmunity.[6] Here we see that the combination of genes affects both the development of autoimmunity and progression. We therefore expect that some of the eight genes included in the final stratification influence events that lead to autoimmunity, and others influence events that occur after the initiation of autoimmunity. Also, informative is the observation that the non-HLA gene combinations were more discriminatory for type 1 diabetes in children who already present with HLA-risk genotypes than in the unselected cohort. Much of this will be explained by the Bayesian principles due to the increase in *a priori* risk, but we cannot exclude that some of the genes exert susceptibility in the presence of HLA-risk genotypes.

The prospective nature of the study allowed us to examine risk stratification using time to event analyses. Diabetes risk was around 25% in the highest risk categories. Remarkable was that this risk was achieved without substantial loss in sensitivity. Screening by TEDDY HLA genotypes plus non-HLA gene SNP combination scores would have preselected $<5\%$ of all the first-degree relatives for follow-up. These relatively few relatives would have included over one-third of all diabetes cases that occurred by 14 years of age in the whole BABYDIAB cohort. Such numbers are approaching as to what may be practical for multicenter early primary prevention studies using type 1 diabetes rather than islet autoimmunity as their primary outcome measure.

This proof of principle study provides an approach to the analysis and use of combinations of weak odds ratio SNPs to help stratification of type 1 diabetes risk. It is versatile and can include additional genes and other risk factors in developing risk algorithms. It is likely that it can be further improved by weighting of individual genes and/or genotypes within genes. We suggest that once this is applied and tested in additional cohorts, it may serve as a model for identifying risk combinations that can effectively stratify risk for type 1 diabetes well beyond what is currently achieved by HLA and family history alone.

SUBJECTS AND METHODS

Study cohort, participants and samples

The study was performed in children from the BABYDIAB study, a longitudinal study examining the natural history of islet autoimmunity and type 1 diabetes in 1650 children born to a mother or father with type 1 diabetes.[13] Recruitment began in 1989 and ended in 2000. All children were recruited from Germany. The cohort is not population based and 97% of families are German Caucasian. Venous blood samples were obtained from children at study visits scheduled at 9 months, and at 2, 5, 8, 11, 14, 17 and 20 years of age. Autoantibodies against IAA, GAD autoantibody, IA-2 autoantibody and ZnT8 autoantibody were measured in samples taken at all scheduled visits, and every 6 months in children with islet autoantibodies. The median follow-up time from birth to last sample was 12.5 years (interquartile range, 6.5–15.5 years) and from birth to last contact was 14.1 years (interquartile range, 11.9–15.6 years). The BABYDIAB study was approved by the ethical committee of Bavaria, Germany (Bayerische Landesärztekammer no. 95357). All families gave written informed consent to participate in the study. Investigations were carried out in accordance with the principles of the Declaration of Helsinki, as revised in 2000.

Follow-up for diabetes

Families were asked to report the occurrence of symptoms of diabetes. In children with islet autoantibodies, a yearly oral glucose tolerance test was

performed. Diabetes onset was defined according to the American Diabetes Association criteria, which include unequivocal hyperglycemia with acute metabolic decompensation, or the observation on at least two occasions of a 2-h plasma glucose $>200\,mg\,dl^{-1}$ after an oral glucose challenge, or a random blood glucose $>200\,mg\,dl^{-1}$ if accompanied by unequivocal symptoms. Since 1997, fasting blood glucose $>126\,mg\,dl^{-1}$ on two occasions was added to the diabetes diagnosis criteria.[14]

Islet autoantibody measurements

IAA, GAD autoantibody, IA-2 autoantibody and ZnT8 autoantibody were determined centrally by the Institute of Diabetes Research Munich using radiobinding assays as previously described.[13,15] Briefly, IAAs were measured by protein A/G radiobinding assays using [^{125}I]-recombinant human insulin labeled at tyrosine amino acid 14. GAD autoantibody, IA-2 autoantibody and ZnT8 autoantibody were measured separately by protein A radiobinding assays using [^{35}S]-methionine-labeled *in vitro* transcribed/translated recombinant human GAD65, IA-2ic and the carboxy-terminal portion of ZnT8 for the two major variants at amino acid 325, respectively. The upper limit of normal for each assay was determined using Q–Q plots and corresponded to the 99th percentile of control children. Offsprings were considered islet autoantibody-positive when two consecutive samples collected after birth were positive. Islet autoantibody assays were evaluated by the Diabetes Autoantibody Standardization Program (laboratory 121).[16,17]

Genotyping

HLA class II alleles HLA-DRB1, HLA-DQA1 and HLA-DQB1 were determined using PCR-amplified DNA and nonradioactive sequence-specific oligonucleotide probes as described previously.[18] Children were defined as having a high-risk HLA genotype on the basis of whether they had one of the TEDDY study inclusion genotypes for first-degree relatives: DR4-DQA1*030X-DQB1*0302$^{@}$/DR3-DQA1*0501-DQB1*0201; DR4-DQA1*030X-DQB1*0302$^{@}$/DR4-DQA1*030X-DQB1*0302$^{@}$; DR4-DQA1*030X-DQB1*0302$^{@}$/DR8-DQA1*0401-DQB1*0402; DR3-DQA1*0501-DQB1*0201/DR3-DQA1*0501-DQB1*0201; DR4-DQA1*030X-DQB1*0302$^{@}$/DR4-DQA1*030X-DQB1*020X; DR4-DQA1*030X-DQB1*0302$^{@}$/DR1-DQA1*0101-DQB1*0501; DR4-DQA1*030X-DQB1*0302$^{@}$/DR13-DQA1*0102-DQB1*0604; DR4-DQA1*030X-DQB1*0302/DR4-DQA1*030X-DQB1*0304; DR4-DQA1*030X-DQB1*0302$^{@}$/DR9-DQA1*030X-DQB1*0303; and DR3-DQA1*0501-DQB1*0201/DR9-DQA1*030X-DQB1*0303; where $^{@}$ includes DQB1*0302 and *0304.[19]

Additional type 1 diabetes susceptibility gene SNPs were considered and selected for typing on the basis of the genes that were reported to be susceptible in 2008/2009, the strength of their odds ratio and the SNP-typings that were successful when establishing the typing in Munich. This resulted in 12 SNPs from 12 genes: *ERBB3 rs705704, PTPN2 rs1893217, IFIH1 rs2111485, PTPN22 rs6679677, KIAA0350 rs12935413, CD25 rs11594656, CTLA4 rs3087243, SH2B3 rs3184504, IL2 rs4505848, IL18RAP rs917997, IL10 rs3024505* and *COBL rs4948088*. Genotyping of these SNPs was performed with the MassARRAY system using the iPLEX chemistry (Sequenom, San Diego, CA, USA) as previously described.[6] To control for reproducibility, 16.3% of samples were genotyped in duplicate with discordance rate $<0.5\%$. All SNPs were tested for deviation from Hardy–Weinberg equilibrium by means of Chi-square or Fisher's exact test. DNA samples for HLA genotyping were available from 1488 children and for genotyping of the additional type 1 diabetes risk genes from 1380 children. The number of children who were successfully typed are listed in Supplementary Table 1. Complete successful typing for HLA and all 12 additional type 1 diabetes genes were obtained for 1290 children. *INS* gene polymorphisms were obtained on 1382 children as previously described,[11] and including 1288 of those with HLA and all 12 gene SNP data.

Risk scores

A total risk score was created for each child by counting the total number of risk alleles across all SNP genotypes similar to what was previously described for obesity[20] and coronary artery disease.[10] Only one variant at each locus was chosen and only children with complete genotype data at all 12 variants were included in score analyses ($n=1290$ children). Risk scores were generated for all possible permutations of the 12 gene SNPs, considering any number of genes in the permutation, that is, 1–12 gene SNPs with a total of $n=2^{12}-1=4095$ permutations. This increased to $n=2^{13}-1=8191$ permutations when the *INS* genotype was also included. Risk scores for these permutations were generated using MATLAB software (R2011a; The MathWorks Inc., Nattick, MA, USA). Each permutation was treated as a separate risk score model.

Statistical analysis

Outcomes used for analyses were either type 1 diabetes or islet autoimmunity (defined as positive for antibodies to any of IAA, GAD, IA-2 or ZnT8 in at least two consecutive samples). Analyses were also performed in the total cohort and selectively in children who had high-risk HLA genotype. Total risk scores for the 12 gene SNP model were compared between children with and without diabetes using Student's *t*-test. ROC analysis[21,22] was used to assess the performance of risk score models in classification of outcome. The AUC and the significance (*P*-value) of the AUC as compared with an AUC with no case–control discrimination (0.5) were used as assessment of risk score models. The ROC analysis was repeated for all possible permutations. The significance of outcome classification by the risk score models was assessed using a permutation-based empirical *P*-value in which the median AUC of all possible combinations (risk score models) was compared with the median AUC of 1000 data sets with randomly shuffled class assignments.

Cumulative risk of type 1 diabetes development and development of islet autoimmunity was estimated by life table analysis. Follow-up was calculated from birth to the age of diagnosis of diabetes and to the age when islet autoimmunity first developed, or to the last contact/sample. Within islet autoantibody-positive children, life table analysis was used to calculate the risk of progression to type 1 diabetes. The length of follow-up was calculated from the age when islet autoimmunity first developed to the age of diagnosis of type 1 diabetes, or to the last contact. Risk score categories used in life table analyses were assigned as low, moderate and high by examining ROC curves for inflexion points (Supplementary Figure 1). Comparisons between categories were made using the log-rank test. Hazard ratios were determined using Cox proportional hazards model (with and without adjustment for HLA genotype).

Throughout the manuscript, CIs represent s.d. or 95% CIs. Two-tailed *P*-values are used without correction. The statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS 19.0; SPSS Inc., Chicago, IL, USA) and MATLAB (R2011a; The MathWorks Inc.).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Annette Knopff, Stephanie Krause, Claudia Matzke, Marlon Scholz, Daniela Höfelmann, Jesse Hansen, Joanna Stock, Claudia Ramminger, Julia Schenkel, Nadine Bruckmeier, Nadja Kocher, Melanie Herbst and Ina Geisler for data collection and expert technical assistance, and Ramona Puff for laboratory management. We also thank all pediatricians and family doctors in Germany for participating in the BABYDIAB study. The work was supported by grants from the Kompetenznetz Diabetes mellitus (Competence Network for Diabetes mellitus) funded by the Federal Ministry of Education and Research (FKZ 01GI0805-07), the European Union (EP7-HEALTH-2007, DIAPREPP N202013) and the Juvenile Diabetes Research Fund (No 17-2012-16). EB is supported by the DFG Research Center and Cluster of Excellence—Center for Regenerative Therapies Dresden (FZ 111). JL is supported by the Finnish Foundation for Pediatric Research and Turku University Hospital Foundation. This study was supported in part by a grant from the German Federal Ministry of Education and Research (BMBF) to the German Center for Diabetes Research (DZD e.V.).

REFERENCES

1 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**: 857–864.
2 Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 2009; **41**: 703–707.
3 Noble JA, Valdes AM, Cook M, Klitz W, Thomson G, Erlich HA. The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *Am J Hum Genet* 1996; **59**: 1134–1148.
4 Todd JA. Etiology of type 1 diabetes. *Immunity* 2010; **32**: 457–467.

5 Tait BD, Colman PG, Morahan G, Marchinovska L, Dore E, Gellert S *et al.* HLA genes associated with autoimmunity and progression to disease in type 1 diabetes. *Tissue Antigen* 2003; **61**: 146–153.
6 Winkler C, Lauber C, Adler K, Grallert H, Illig T, Ziegler AG *et al.* An interferon-induced helicase (IFIH1) gene polymorphism associates with different rates of progression from autoimmunity to type 1 diabetes. *Diabetes* 2011; **60**: 685–690.
7 Douroudis K, Laine AP, Heinonen M, Hermann R, Lipponen K, Veijola R *et al.* Association of CTLA4 but not ICOS polymorphisms with type 1 diabetes in two populations with different disease rates. *Hum Immunol* 2009; **70**: 536–539.
8 Lempainen J, Hermann R, Veijola R, Simell O, Knip M, Ilonen J. Effect of the PTPN22 and INS risk genotypes on the progression to Clinical Type 1 Diabetes After the Initiation of β-cell autoimmunity. *Diabetes* 2012; **61**: 963–966.
9 Steck AK, Wong R, Wangner B, Johanson K, Liu E, Romanos J *et al.* Effects of non-HLA gene polymorphisms on development of islet autoimmunity and type 1 diabetes in a population with high-risk HLA-DR,DQ genotypes. *Diabetes* 2012; **61**: 753–758.
10 Lluís-Ganella C, Lucas G, Subirana I, Sentí M, Jimenez-Conde J, Marrugat J *et al.* Additive effect of multiple genetic variants on the risk of coronary artery disease. *Rev Esp Cardiol* 2010; **63**: 925–933.
11 Walter M, Albert E, Conrad M, Keller E, Hummel M, Ferber K *et al.* IDDM2/insulin VNTR modifies risk conferred by IDDM1/HLA for development of Type 1 diabetes and associated autoimmunity. *Diabetologia* 2003; **46**: 712–720.
12 Skyler JS. Type 1 Diabetes TrialNet Study Group; update on worldwide efforts to prevent type 1 diabetes. *Ann NY Acad Sci* 2008; **1150**: 190–196.
13 Ziegler AG, Hummel M, Schenker M, Bonifacio E. Autoantibody appearance and risk for development of childhood diabetes in offspring of parents with type 1 diabetes: the 2-year analysis of the German BABYDIAB Study. *Diabetes* 1999; **48**: 460–468.
14 Puavilai G, Chanprasertyotin S, Sriphrapradaeng A. Diagnostic criteria for diabetes mellitus and other categories of glucose intolerance: 1997 criteria by the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (ADA), 1998 WHO consultation criteria, and 1985 WHO criteria. World Health Organization. *Diabetes Res Clin Pract* 1999; **44**: 21–26.
15 Achenbach P, Lampasona V, Landherr U, Koczwara K, Krause S, Grallert H *et al.* Autoantibodies to zinc transporter 8 and SLC30A8 genotype stratify type 1 diabetes risk. *Diabetologia* 2009; **52**: 1881–1888.
16 Törn C, Mueller PW, Schlosser M, Bonifacio E, Bingley PJ. Participating laboratories. diabetes antibody standardization program: evaluation of assays for autoantibodies to glutamic acid decarboxylase and islet antigen-2. *Diabetologia* 2008; **51**: 846–852.
17 Schlosser M, Mueller PW, Törn C, Bonifacio E, Bingley PJ. Participating laboratories. diabetes antibody standardization program: evaluation of assays for insulin autoantibodies. *Diabetologia* 2010; **53**: 2611–2620.
18 Schenker M, Hummel M, Ferber K, Walter M, Keller E, Albert ED *et al.* Early expression and high prevalence of islet autoantibodies for dr3/4 heterozygous and dr4/4 homozygous offspring of parents with type i diabetes: the german babydiab study. *Diabetologia* 1999; **42**: 671–677.
19 TEDDY Study Group 2007. The Environmental Determinants of Diabetes in the Young (TEDDY) study: study design. *Pediatr Diabetes* 2007; **8**: 286–298.
20 Elks CE, Loos RJ, Sharp SJ, Langenberg C, Ring SM, Timpson NJ *et al.* Genetic markers of adult obesity risk are associated with greater early infancy weight gain and growth. *PLoS Med* 2010; **7**: e1000284.
21 Fawcett T. An introduction to ROC analysis. *Pattern Recognition Lett* 2006; **27**: 861–874.
22 Clayton DG. Prediction and interaction in complex disease genetics: experience in type 1 diabetes. *PLoS Genet* 2009; **5**: e1000540.

Supplementary Information accompanies the paper on Genes and Immunity website (http://www.nature.com/gene)