RESEARCH PAPER

Allele-specific methylation of type 1 diabetes susceptibility genes

Alida SD Kindt[a], Rainer W Fuerst[b], Jan Knoop[b], Michael Laimighofer[a], Tanja Telieps[c], Markus Hippich[b], Maria A Woerheide[a], Simone Wahl[d,e,f], Rory Wilson[d,e], Eva-Maria Sedlmeier[b], Angela Hommel[g], John A Todd[h], Jan Krumsiek[a,f], Anette-G Ziegler[b,i*], Ezio Bonifacio[c,g,i]

[a] Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany; alida.kindt@helmholtz-muenchen.de; michael.laimighofer@helmholtz-muenchen.de; jan.krumsiek@helmholtz-muenchen.de

[b] Institute of Diabetes Research, Helmholtz Zentrum München, Neuherberg, Germany jan.knoop@helmholtz-muenchen.de; markus.hippich@helmholtz-muenchen.de; eva-maria.sedlmeier@helmholtz-muenchen.de; anette-g.ziegler@helmholtz-muenchen.de

[c] Institute of Diabetes and Obesity, Helmholtz Zentrum München, Neuherberg, Germany; tanja.telieps@helmholtz-muenchen.de

[d] Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany; rory.wilson@helmholtz-muenchen.de

[e] Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[f] German Center for Diabetes Research (DZD), Neuherberg, Germany

[g] Center for Regenerative Therapies - Dresden, Faculty of Medicine Carl Gustav Carus, Technische Universität, Dresden, Germany; ezio.bonifacio@tu-dresden.de

[h] JDRF/Wellcome Trust Diabetes and Inflammation Laboratory, Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK; jatodd@well.ox.ac.uk

[i] Forschergruppe Diabetes e.V., Neuherberg, Germany

[*] Corresponding author: Anette-G Ziegler (anette-g.ziegler@helmholtz-muenchen.de)

Highlights

- Marked and persistent differential DNA methylation associated with HLA class II
- Carriers of HLA DR3/DQ2 have reduced HLA class II RNA and protein expression
- *cis*-metQTL observed in all type 1 diabetes susceptibility genetic regions
- Novel association of differential methylation of *LDHC* gene with early autoimmunity

Abstract

The susceptibility to autoimmune disease is influenced by genes encoding major histocompatibility complex (MHC) proteins. By examining the epigenetic methylation maps of cord blood samples, we found marked differences in the methylation status of CpG sites within the MHC genes (cis-metQTLs) between carriers of the type 1 diabetes risk haplotypes HLA-DRB1*03-DQA1*0501-DQB1*0201 (DR3-DQ2) and HLA-DRB1*04-DQA1*0301-DQB1*0302 (DR4-DQ8). These differences were found in children and adults, and were accompanied by reduced HLA-DR protein expression in immune cells with the HLA-DR3-DQ2 haplotype. Extensive cis-metQTLs were identified in all 45 immune and non-immune type 1 diabetes susceptibility genes analyzed in this study. We observed and validated a novel association between the methylation status of CpG sites within the LDHC gene and the development of insulin autoantibodies in early childhood in children who are carriers of the highest type 1 diabetes risk genotype. Functionally relevant epigenetic changes in susceptibility genes may represent therapeutic targets for type 1 diabetes.

Keywords: Epigenetics, autoimmunity, HLA, insulin autoantibodies, insulin gene, Lactate Dehydrogenase C

1. Introduction

Genetic susceptibility for autoimmunity is conferred by multiple genes encoding histocompatibility leukocyte antigens (HLA), proteins involved in T and/or B cell responses, and proteins relevant to the target tissue to which tolerance has been lost[1]. Type 1 diabetes, an autoimmune disease targeting insulin-producing pancreatic β cells, quintessentially follows this paradigm with strong susceptibility provided by the HLA-DRB*03-DQB1*0201 (DR3-DQ2) and the HLA-DRB1*04-DQB1*0302 (DR4-DQ8) haplotypes[2]. Susceptibility is also conferred by genes encoding T cell receptor and interleukin-2 signaling proteins, for example, and by polymorphisms in INS that affect insulin expression[3, 4]. There is also substantial overlap with susceptibility for other autoimmune diseases. For example, the HLA-DR3-DQ2 haplotype is associated with increased susceptibility for type 1 diabetes[5], celiac disease[6], thyroid disease[7], rheumatoid arthritis[8], myasthenia gravis[9], systemic lupus erythematosus[10] and other autoimmune diseases[11]. The HLA-associated susceptibility to autoimmune disease varies according to the other haplotype carried by the person. For example, a homozygous DR3-DQ2 genotype confers greatest susceptibility to celiac disease[6, 12-15] whereas heterozygosity of DR3-DQ2 and DR4-DQ8 confers greatest susceptibility for type 1 diabetes[16-18]. Genetic polymorphisms or mutations can not only be associated with the loss or gain of function of a gene, but can also determine a gene's methylation state[19]. Therefore, variation in the methylation status may be an indicator of susceptibility, and provide information regarding how variation in the DNA sequence may influence the gene's function.

Here, we analyzed the methylation status of susceptibility genes for type 1 diabetes in relation to genetic polymorphisms (metQTLs). We investigated the methylomes of newborns, children, and adults with genetic susceptibility for type 1 diabetes. We found remarkable variation in the methylation status of HLA-DR3-DQ2 at hundreds of CpG sites within the MHC region (cis-metQTLs). We subsequently showed reductions in HLA-DR transcript expression and reduced HLA-DR protein expression in dendritic cells (DCs) from children with HLA-DR3-DQ2, suggesting that peptide presentation to lymphocytes may be compromised in carriers. In accordance with, and beyond the extent recognized by recently described cis-genetic effects on methylation[20, 21], we observed allele-associated differential DNA methylation at CpG sites in all 45 type 1 diabetes susceptibility genes analyzed in this study. Finally, we observed a novel association between the methylation status of CpG sites within LDHC and the development of insulin autoantibodies (IAAs) in the first years of life. This cis-metQTL atlas for HLA and type 1 diabetes susceptibility provides a novel resource for the investigation of disease susceptibility, and may yield potential targets for manipulating gene expression to prevent or treat type 1 diabetes.

2. Results

2.1 Methylation of type 1 diabetes-associated HLA genotypes in newborns.

We determined the methylation status of approximately 450,000 CpG sites across the entire genome in newborns with high genetic risk for type 1 diabetes. The samples and cohorts used in this study are described in Supplementary Fig. 1. The methylation status was compared between newborns with the HLA haplotypes HLA-DRB1*03 DQB1*0201 (DR3-DQ2) and HLA-DRB1*04 DQB1*0302 (DR4-DQ8). The newborns were also stratified based on their genotype: DR3-DQ2 homozygous (DQ2/2; n = 26), DR4-DQ8 homozygous (DQ8/8; n = 22), and DR3-DQ2/DR4-DQ8

heterozygous (DQ2/8; n = 52). In order to avoid single nucleotide polymorphsims (SNP)-related methodological bias, we removed CpG sites with known SNPs within 10 bp of the 3' end of probes, leaving 440,052 CpG sites for analysis. The metQTL analysis showed that 196 of these CpG sites were differentially methylated (genome-wide false discovery rate [FDR] < 0.05) among the three HLA-DR-DQ genotypes. Consistent with strong cis-metQTLs, 181 of the sites were found on chromosome 6, and 41/181 (23%) sites were located within the chromosomal region spanning HLA-DRA to HLA-DQB2 (Fig. 1; Supplementary Table 1). Analysis of the remaining 36,142 sites with known SNPS within 10 bp of the 3' end of probes also identified abundant cis-metQTLs in the HLA region (Supplementary Fig. 2). The differential methylation status was confirmed at all five of the CpG sites tested by pyrosequencing in DNA obtained from peripheral blood mononuclear cells and in both sites tested by pyrosequencing in DNA from HLA class II-expressing cells, B lymphocytes, monocytes, plasmacytoid DCs (pDCs), and myeloid DCs purified from peripheral blood (Supplementary Fig. 3).

The direction of the cis-metQTL status of the DR3-DQ2 haplotype relative to the DR4-DQ8 haplotype was assessed using Scheffe's post-hoc test. For 80/181 (44%) differentially methylated sites on chromosome 6, we found a direct relationship between the number of DR3-DQ2 haplotypes (0, 1, or 2) and increased (n = 36) or decreased (n = 44) methylation (Fig. 1d). In addition, 58 sites were hypomethylated in one of the homozygote genotypes relative to the heterozygous or other homozygous genotype (36 sites were hypomethylated in DR3-DQ2/DR3-DQ2; 22 sites were hypomethylated in DR4-DQ8/DR4-DQ8), and 43 sites were hypermethylated in one homozygote genotype (17 were hypermethylated in DR3-DQ2/DR3-DQ2; 26 were hypermethylated in DR4-DQ8/DR4-DQ8). The HLA-DR3-DQ2 haplotype was

hypomethylated relative to HLA-DR4-DQ8 in 106/181 (59%) differentially methylated sites on chromosome 6.

We also found differential methylation among the three HLA genotypes at 15 CpG sites on other chromosomes. These included 5 sites near FBRSL1 on chromosome 12 and one site near ~~FAM136A and~~ HOXD9 on chromosome 2 (Supplementary Table 1, Supplementary Fig. 4).

2.2 *HLA-DR3-DQ2 and DR4-DQ8 cis-metQTLs are present in children and adults.*

Type 1 diabetes autoimmunity rarely starts at birth[22-24]. Therefore, we investigated whether the differentially methylated HLA class II alleles detected on chromosome 6 in cord blood samples were also present in toddlers (n = 32; age range, 1.8–3.0 years) or adults (n = 86; age range, 39–78 years) with the DR3-DQ2/DR3-DQ2, DR3-DQ2/DR4-DQ8, or DR4-DQ8/DR4-DQ8 genotypes. Of the 181 CpG sites on chromosome 6 with metQTLs that were detected in cord blood samples, all 181 were also analyzed in adults and 169 in toddlers. The number discrepancy was due to different platforms used for the toddlers (see Methods). The effect direction of the cis-metQTLs for these CpG probes in the neonates was significantly correlated with that in toddlers (Spearman's r = 0.83; Fig. 2a) and adults (Spearman's r = 0.88; Fig. 2b). These findings indicate that the differential methylation signals at many sites on chromosome 6 are present in toddlers and adults, and may persist throughout life.

2.3 HLA-DR3-DQ2 and DR4-DQ8 variation in HLA-DR expression.

The direction of the cis-metQTLs for the DR3-DQ2 vs DR4-DQ8 haplotypes varied across sites, and did not allow us to conclude on any possible downstream consequences of the DR-DQ genotype cis-metQTLs. Because HLA class II gene expression is affected by numerous factors encountered during childhood, including

environmental factors[25], we first examined HLA-DR protein expression in cord blood monocytes (Fig. 3a). The median fluorescence intensity (MFI) of HLA-DR in cord blood monocytes was significantly lower in newborns carrying one DR3-DQ2 haplotype than in newborns carrying one DR4-DQ8 haplotype ($p = 0.02$).

We next determined HLA-DR protein expression in different cell types using PBMCs from adolescents (Fig. 3b). We found a clear relationship between HLA-DR3-DQ2 and decreased HLA-DR expression for pDCs ($p = 0.008$). Additionally, HLA-DR expression was lower in the heterozygote HLA-DR3-DQ2/DR4-DQ8 group for monocytes ($p = 0.02$ vs. homozygotes) and B lymphocytes ($p = 0.03$ vs. homozygotes). The inadequacy of available antibodies to equally detect HLA-DQ2 and HLA-DQ8 surface expression prevented us to assess whether HLA-DQ was also differentially expressed on these cells.

Finally, we examined microarray data previously obtained for PBMC samples taken from 87 children in the first year of life[26]. Transcript levels were available for the *HLA-DRB5*, *DQA1*, and *DQB1* genes. *HLA-DRB5* transcript expression was lower for HLA-DR3-DQ2 than for DR4-DQ8 (Fig. 3c), but the expression levels of *DQA1* and *DQB1* were unaffected. We next correlated transcript levels with methylation status at 91 sites in the HLA-DR-DQ region. *DRB5* transcript expression was significantly associated with methylation status, and this association was driven by the underlying genotype (Supplementary Fig. 4).

2.4 Cis-metQTLs of type 1 diabetes susceptibility genes.

We investigated whether other genes associated with type 1 diabetes had cis-metQTLs using cord blood DNA samples. For *INS*, we found that the type 1 diabetes *INS* 5' variable number tandem repeat (VNTR) risk genotype AA was consistently associated with hypermethylation at all of the CpG sites analyzed (Fig. 4;

Supplementary Table 2). Similarly, all 45 type 1 diabetes susceptibility genes examined in this study showed cis-metQTLs, reaching statistical significance for 59 sites after correcting for multiple tests (FDR < 0.05) and adjusting for the number of CpG sites within the genetic region (Supplementary Fig. 5; Supplementary Table 3). By comparison, in an analysis of 17 susceptibility loci for type 2 diabetes, we observed significant cis-metQTLs at 26 sites in 11 of 17 SNPs analyzed (Supplementary Fig. 6, Supplementary Table 4). This analysis was extended to >200,000 SNPs included on the Immunochip, a genotyping array specifically designed to include loci in autoimmune genes[27] and genes associated with other inflammatory diseases. The results are shown in Supplementary Table 5.

2.5 Novel metQTLs associated with early IAA status.

The cohorts included in this study comprised children who were prospectively followed for the development of autoantibodies to β-cell antigens. Therefore, we had the opportunity to identify methylation signatures associated with early type 1 diabetes autoimmunity. Early development of IAA strongly predicts the development of multiple autoantibodies and the later onset of type 1 diabetes[28]. In order to avoid confounding by cis-metQTLs at HLA, we restricted the analysis to cord blood samples from 54 children heterozygous for HLA-DR3-DQ2/DR4-DQ8. We first divided the children into two groups: children who developed IAAs before 3 years of age (n = 6) and children who remained autoantibody negative (n = 48). We identified 13 metQTLs that were hypomethylated in IAA-positive children of the discovery cohort (BABYDIET). These included 7 significant metQTLs located around LDHC on chromosome 11 (Fig. 5, Table 1), which were validated in an independent replication cohort of 12 neonates, with 5 IAA-positive children. Using data from C57BL/6J and NOD mice, we confirmed that LDHC was highly expressed in the

testis but was absent in lymphoid tissues, ~~pancreas,~~ and islets of Langerhans (Supplementary Fig. 7). LDHC expression was also absent in islets of Langerhans exposed to low or high glucose (Supplementary Fig. 7).

3 Discussion

Genotype-dependent methylation is thought to influence disease susceptibility[29]. Associations of genotype and methylation, and thus presumed sequence regulation by means of methylation, is usually observed in cis[30], and extensive associations between methylation and expression variation were identified[20]. Here, we showed strong and abundant cis-metQTLs for HLA class II genotypes and for other susceptibility genes for autoimmune diabetes. Intriguingly, we found differential methylation and expression of HLA between two susceptible HLA haplotypes, which prompted us to postulate that these haplotypes confer susceptibility to autoimmune diabetes via multiple mechanisms. Based on our data, we suggest that part of the generalized susceptibility for autoimmunity conferred by the HLA-DR3-DQ2 haplotype is due to decreased expression and consequently decreased avidity of self-peptide-MHC complexes for autoreactive T cells, and hence a lower threshold for retaining T lymphocytes during thymic selection. Specifically for type 1 diabetes, the combination of a lowered thymic selection threshold by HLA-DR3-DQ2 and a β-cell peptide bias facilitated by glutamate at position 57 of DQβ in DQ2 and DQ8 could explain the relative increase in susceptibility provided by heterozygosity of DQ2 and DQ8 relative to their homozygous states[31, 32].

The HLA class II metQTLs in this study are supported by methylation data and eQTL (RNA and/or protein) data for samples obtained at birth and in childhood. The results were obtained by our ability to simplify the relationship with HLA class II to two haplotypes. Previous reports of smaller numbers of older subjects in health vs. disease

cohorts indirectly showed that cis-metQTLs are generally observed for HLA[29, 33, 34], supporting our findings for HLA-DR3-DQ2 versus HLA-DR4-DQ8. Decreased HLA class II expression is associated with autoimmunity in experimental models[35]. Our data show that the variable expression is particularly evident in peripheral blood pDCs, and that the expression variation is not uniform between cell types bearing HLA-DR. This suggests that there may be functional consequences in antigen presentation. However, it is necessary to analyze antigen-presenting cells in the thymus and peripheral lymphoid organs to confirm this possibility.

Our finding that HLA-DR expression is lower in HLA-DR3-DQ2 cells than in HLA-DR4-DQ8 cells is consistent with those of a recent study, which showed increased expression on peripheral blood monocytes in individuals homozygous at a vitiligo-associated 3-SNP haplotype between HLA-DRB1 and DQA1[35]. Indirectly corroborating our findings, the 3-SNP haplotype was strongly associated with HLA-DR4 but not with HLA-DR3. HLA-DRB1, DQA1 and DQB1 expression were reported to be increased in cells from individuals homozygous for systemic lupus erythematosus (SLE)-associated SNPs in the XL9 regulatory complex located between HLA-DR and DQ[36]. Susceptible alleles of the SLE-associated XL9 SNP are in linkage disequilibrium with and partially overlap with HLA-DR3-DQ2, suggesting that the findings in SLE are inconsistent with ours. Unlike our study and the study by Cavalli et al. [35], which examined expression in non-manipulated monocytes or DCs, protein expression in the SLE study was examined in macrophages and DCs obtained after in vitro conversion of monocytes, characterized by upregulated HLA-DR and HLA-DQ. It is also noteworthy that our findings were consistent in samples taken at birth, adolescence and in adults. Furthermore, HLA-DQ expression in the SLE study was examined using the antibody clone Tu169, but the

manufacturer reported that this clone binds to HLA-DQ8-associated DQ3 molecules weakly in comparison to HLA-DQ2 molecules. Clearly, the use of non-manipulated cells, and the correct choice of antibodies and methods to examine expression are crucial factors, and further cell-based studies are needed to determine immune responses or tolerance.

In addition to the cis-metQTLs, we observed a small number of trans-metQTLs for the two HLA class II haplotypes. Three regions harboring these trans-metQTL sites were of particular interest. Of the 16 sites on chromosomes other than chromosome 6, 5 were on chromosome 12 in or near FBRSL1. This gene was identified in a study investigating the epigenetic mediation of HLA-DR and DQ genes in peanut allergy[33]. Moreover, a SNP located 150 bp upstream of FBRSL1 was associated with fasting glucose levels[37]. The second region was within DEFB4A, which is locally regulated by inflammation, and encodes β-defensin-4, an antibacterial peptide with chemotactic activity on immature DCs and memory T cells[38]. This gene harbors the only CpG site exhibiting a heterozygote-specific effect; the heterozygote was hypermethylated compared with the homozygotes. Finally, one site was observed within GP2, which encodes a protein that binds pathogens, including enterobacteria, and is involved in innate and adaptive immune responses[39].

Additionally to the HLA class II associated cis-metQTLs, we provide a resource of cis-metQTLs associated with immune and autoimmune susceptibility genes. We observed extensive cis-metQTL in >7000 genes examined (Supplementary Table 5), and in all 45 non-HLA type 1 diabetes susceptibility genes examined. The cis-metQTL in INS is particularly relevant to type 1 diabetes. We observed hypermethylation of all significant sites in the type 1 diabetes susceptibility allele. This is consistent with data obtained in non-obese diabetic mice and in humans

showing lower insulin expression in the thymus and, therefore, a lower threshold for the selection of insulin autoreactive T cells associated with the susceptibility allele. Our data also help explain the previously reported association between type 1 diabetes and the methylation status of CpG sites in INS[40] and suggest that the disease association is secondary to the strong cis-metQTL. Indeed, the extensive cis-metQTLs observed in our study and in previous studies[20, 21, 41] should be considered in future analyses of disease-associated metQTLs. Researchers should investigate whether the methylation status beyond the cis-metQTLs at susceptibility regions is associated with disease. This requires knowledge of metQTLs for susceptibility regions, which we have provided for autoimmunity in this study, as well as a greater number of subjects for analysis.

We also identified eight differentially methylated CpG sites in genetically predisposed children who developed IAAs before 3 years of age. In IAA-positive children, all of these sites were hypomethylated within LDHC. The expression of LDHC is normally kept low in β-cells to ensure adequate function[2, 42]. LDHC is expressed in the testes and is associated with infertility in males with type 1 diabetes[43]. However, the link to IAA has not been shown before now.

In summary, we observed strong cis-metQTLs for HLA class II and for genes associated with type 1 diabetes. The downstream relationship to protein expression observed for HLA-DR suggests that these cis-metQTLs may be functionally relevant at the expression level. For HLA, we suggest that the cis-metQTLs provide a mechanism for the loss of tolerance to an autoantigen during early childhood, and may partly explain the increased risk for type 1 diabetes provided by the heterozygous DR3-DQ2/DR4-DQ8 genotype.

Table 1. Hypomethylation in LDHC on chromosome 11 in the discovery and replication cohorts

CpG	bp	Discovery cohort (n=54)				Replication cohort (n=12)		
		Estimate	SE	p-value	FDR p-value	Estimate	SE	p-value
cg07093428	18433500	−0.14	0.02	p < 0.001	p < 0.001	−0.08	0.04	0.08
cg19767548	18433554	−0.15	0.02	p < 0.001	p < 0.001	−0.09	0.04	0.06
cg14332815	18433564	−0.14	0.02	p < 0.001	p < 0.001	−0.08	0.05	0.10
cg11821245	18433683	−0.20	0.03	p < 0.001	p < 0.001	−0.11	0.07	0.12
cg14259717	18433732	−0.12	0.02	p < 0.001	p < 0.001	−0.06	0.04	0.19
cg08418111	18433745	−0.09	0.02	p < 0.001	0.03	−0.06	0.04	0.13
cg05740244*	18434015	−0.25	0.04	p < 0.001	p < 0.001	−0.16	0.08	0.05
cg07415359	18434354	−0.07	0.02	p < 0.001	0.005	−0.06	0.02	0.02

BP, base pair; SE, standard error; FDR, false discovery rate; * CpG with SNP within 10 base pairs of 3'UTR

Figure legends

Figure 1 │ Results of Illumina Infinium methylation array in neonates with HLA-DR-DQ genotypes associated with type 1 diabetes. (a) Manhattan plot of the differentially methylated CpG sites in neonates with HLA-DR3-DQ2/DR3-DQ2 (2/2; n = 26), DR3-DQ2/DR4-DQ8 (2/8; n = 52), and DR4-DQ8/DR4-DQ8 (8/8; n = 22) genotypes. This investigation was performed in the BABYDIET cohort using the Illumina Infinium methylation array. The maximum -log10 P value was on chromosome 6. Significant signals were also detected on other chromosomes. (b,c) Higher magnification views of the boxed region of chromosome 6 in panel a. Genes are shown in orange. (d) Illustration of the six representative patterns of methylation differences (left panels) and the frequencies of these patterns (right panel) among the significant sites on chromosome 6. The plots in the left panels show methylation in arbitrary units (A.U.) for each neonate categorized into one of the three HLA-DR-DQ genotypes and as box plots. Hyper, hypermethylated; hypo, hypomethylated; pos, positive; neg, negative. In boxplots: centers are median, error bars are 95% confidence intervals.

Figure 2 │ Stability of the epigenetic changes in the HLA region.

(a,b) Comparison of p-values for the differences in methylation at CpG sites within the HLA class II region between the DR3-DQ2/DR3-DQ2, DR3-DQ2/DR4-DQ8, and DR4-DQ8/DR4-DQ8 genotypes in neonates (BABYDIET cohort), toddlers (BABYDIAB cohort), and adults (KORA cohort). The colors correspond to the methylation status in the neonate cohort. Blue indicates hypermethylation in the DR4-DQ8/DR4-DQ8 genotype compared with the DR3-DQ2/DR3-DQ2 genotype and green indicates hypermethylation in the DR3-DQ2/DR3-DQ2 genotype compared

with the DR4-DQ8/DR4-DQ8 genotype. The directed p-value is shown as the $-\log_{10}$ p-value multiplied by the direction of the effect. Strong and significant correlations of directed p-values were observed across all sites for (a) neonates in the BABYDIET study versus toddlers in the BABYDIAB study (Spearman's $\rho = 0.83$) and (b) neonates in the BABYDIET study versus adults in the KORA study (Spearman's $\rho = 0.88$). (c) Methylation patterns in the genomic region around DR-DQ in neonates (upper panel), toddlers (middle panel), and adults (lower panel).

Figure 3 | HLA protein and mRNA expression in relation to the HLA genotype. (a) Median fluorescence intensity (MFI) of HLA-DR expression in cord blood-derived monocytes from neonates with the HLA-DR3-DQ2 haplotype (2/x; n = 10) or HLA-DR4-DQ8 haplotype (8/x; n = 16). Results are shown as the mean and p-values were calculated with unpaired t tests. Right panel: FACS histogram of monocyte HLA-DR staining in a child carrying HLA-DR3-DQ2 (light grey) or HLA-DR4-DQ8 (dark grey). The median value for each histogram was included as a single data-point in the graph in the left panel. (b) MFI of HLA-DR expression in FACS-purified plasmacytoid dendritic cells (pDCs), conventional dendritic cells (cDCs), classical monocytes, and B lymphocytes according to the HLA-DR-DQ genotype. Left panel: MFI for individual children (n = 5 per genotype). HLA-DR protein expression was lower in pDCs from individuals with the HLA-DR3-DQ2/DR3-DQ2 genotype. The p-values were calculated by analysis of variance. Right panels: representative FACS histograms for HLA-DR staining in the cells selected for sorting. (c) Expression of three HLA class II transcripts, HLA-DRB5, DQA1 and DQB1 according to the HLA-DR-DQ genotype determined in peripheral blood mononuclear cells collected in the

first year of life. The HLA-DRB5 transcript shows a significant positive dosage effect. Centers are median, error bars are 95% confidence intervals.

Figure 4 | Methylation of CpG sites within the type 1 diabetes susceptibility gene, INS, according to the INS variability in the number of tandem repeats genotype. The CpG sites in INS are ordered by their genetic position. All of the sites showed hypermethylation in carriers of the A/A genotype, which is homozygous for the risk-carrying variant in the 5' untranslated region of INS. Centers are median, error bars are 95% confidence intervals. One site with a SNP in its binding probe is indicated by an asterisk.

Figure 5 | Methylation of CpG sites across LDHC in neonates according to the development of islet autoantibodies. Left panels: neonates in the BABYDIET cohort with the HLA-DR3-DQ2/DR4-DQ8 genotype were followed for the development of islet autoantibodies. Neonates were categorized as negative if they did not develop islet autoantibodies (n = 48) or early IAA group if they developed insulin autoantibodies by 3 years of age (n = 6). Right panels: methylation status of the CpG sites in LDHC in a replication set of children. The analysis was performed across all valid CpG sites and multiple CpG sites with the LDHC gene were significant. The methylation of eight CpG sites within LDHC is shown and the one site with a SNP in its binding probe is indicated by an asterisk. All sites were hypomethylated in the early IAA group. The depicted sites are ordered according to their genetic position. Centers are median, error bars are 95% confidence intervals.

Methods

Study cohorts

BABYDIET. Between 2000 and 2006, the BABYDIET study enrolled infants with a first-degree relative with type 1 diabetes and one of the type 1 diabetes risk HLA genotypes: DRB1*03DQA1*05:01-DQB1*0201/DRB1*04-DQA1*03:01-DQB1*03:02 (DR3-DQ2/DR4-DQ8), DRB1*04-DQA1*03:01-DQB1*03:02/DRB1*04-DQA1*03:01-DQB1*03:02 (DR4-DQ8/DR4-DQ8) or DRB1*03-QA1*05:01DQB1*02:01/DRB1*03-DQA1*05:01DQB1*02:01 (DR3-DQ2/DR3-DQ2). The children were prospectively followed from birth, and blood samples were obtained every 3–6 months to measure β cell autoantibodies. The median follow-up from birth was 10.3 years (interquartile range, 9.5–11.9 years)[44]. Cord blood DNA was analyzed in 100 children participating in the BABYDIET study using Infinium 450k methylation assays (Illumina). In addition, 454 PBMC samples were collected from 109 children at multiple follow-up times, and were used for transcriptomics studies[26, 45]. These transcriptomic data are deposited with ArrayExpress (accession number: E-MTAB-1724). The BABYDIET study was approved by the ethics committee of Ludwig-Maximilians University (Protocol No. 329/00) and is registered at ClinicalTrials.gov NCT01115621. The parents or guardians of each child provided informed consent for participation in the BABYDIET study.

BABYDIAB. The BABYDIAB study enrolled 1,650 children born to a mother or father with type 1 diabetes between 1989 and 2000, and prospectively examined the natural history of islet autoimmunity and type 1 diabetes[46, 47]. Children in the BABYDIAB study were followed longitudinally with blood samples taken at 9 months of age, 2 years of age, and then every 3 years. For the current study, we used

venous blood DNA samples taken at 2 years of age from 32 children with the HLA-DR3-DQ2/DR4-DQ8, DR4-DQ8/DR4-DQ8, or DR3-DQ2/DR3/DQ2 genotypes. Samples were analyzed using InfiniumEPIC arrays (Illumina). All samples were collected after obtaining signed informed consent. The BABYDIAB study was approved by the Ethics Committee of Bavaria, Germany (Bayerische Landesärztekammernummer 95357).

ImmunDiabRisk. The ImmunDiabRisk Study enrolled neonates to investigate early priming of cellular immune responses at birth in relation to genetic and familial risk for type 1 diabetes. For this study, we analyzed HLA-DR surface expression in fresh cord blood mononuclear cells obtained from 27 neonates with a HLA-DR3-DQ2 or a HLA-DR4-DQ8 haplotype. All samples were collected after obtaining signed informed consent. The ImmunDiabRisk study was approved by the Ethics Committee of the Technical University Munich (No. 5293/12).

TeenDiab. The TeenDiab study enrolled children and adolescents with a first-degree relative with type 1 diabetes[48]. We analyzed HLA-DR surface expression on fresh peripheral blood samples from 15 children with the HLA-DR3-DQ2/DR4-DQ8, DR4-DQ8/DR4-DQ8, or DR3-DQ2/DR3/DQ2 genotypes. Samples were obtained at a median age of 13.06 years (interquartile range, 11.84–14.79 years). HLA-DR expression was analyzed after fluorescence-activated cell sorting (FACS) and isolation of PBMC subpopulations, and DNA methylation was analyzed by pyrosequencing. The number of 5 per group was chosen on the basis of the differences observed in pyrosequencing methylation data when using whole blood DNA. All samples and information were collected after obtaining signed informed consent. The TeenDiab study was approved by the ethics committees of the Technical

University Munich (No. 2149/08) and Medizinische Hochschule Hannover (No. 5644).

KORA. The Cooperative Health Research in the Region of Augsburg (KORA) F4 study was performed between 2006 and 2008 and included 3080 participants (age range, 39–78 years) from Augsburg in Germany. DNA methylation was assessed in all 86 adults with the HLA-DR3-DQ2/DR4-DQ8, DR4-DQ8/DR4-DQ8, or DR3-DQ2/DR3/DQ2 genotypes. Written informed consent was obtained from all participants. The KORA F4 study was approved by the Ethics Committee of the Bavarian Medical Association (Bayerische Landesärztekammer; No. 06068)[49].

Infinium 450k methylation and Illumina InfiniumEPIC arrays. Genomic DNA (750 ng) was subjected to bisulfite conversion using the EZ-96 DNA Methylation Kit (Zymo Research). Genome-wide DNA methylation was investigated using the Illumina HumanMethylation 450 BeadChip (BABYDIET study, KORA study) or the Infinium MethylationEPIC BeadChip (BABYDIAB study). Briefly, whole genome amplification of the bisulfite-converted sample (4 µl) was followed by enzymatic fragmentation and hybridization of the samples to the BeadChips. Arrays were then fluorescently stained and scanned with the Illumina HiScan SQ scanner. GenomeStudio Software (version 2011.1) with the Methylation Module version 1.9) was used to analyze the raw image data generated by the BeadArray Reader. The methylation data for both chips were pre-processed as previously described.[50]. Raw data files containing the intensity values for probes and individuals were imported into R (version 3.0.1, R Project for Statistical Computing; https://www.r-project.org) and background correction was performed using the R package minfi[51]. The signal intensities were extracted and filtered by (a) their detection p-value (cutoff: $p \geq 0.01$)

or (b) the nbeads count (cutoff: nbeads ≤ 3). All samples had sample-wise detection rates > 95%. Data were normalized using quantile normalization of the raw signal intensities within six probe categories[50]. These intensities were Infinium I signals from beads targeting methylated CpG sites obtained in the red and the green color channels, Infinium I signals from beads targeting unmethylated CpG sites obtained in the red and the green color channels, and Infinium II signals obtained in the red and the green color channels[52]. Quantile normalization was performed using the R package limma, version 3.16.5[53]. After quantile normalization the signal intensities were used to calculate β values. Briefly, a β value equals M/(M + U + 100) where M is the methylated signal, U is the unmethylated signal and 100 is the typical offset used. Before the white blood cell estimates were calculated using the Houseman method, a mean imputation was performed. To this end, white blood cell proportions were estimated using 473 CpG sites present on the array of a total of 500 cell-specific CpG sites according to Houseman et al.[54].

To remove batch or chip effects, we used principal components (PCs), which were calculated for the 235 sample-independent, built-in control probes, to evaluate the experimental components. These control probes were designed to measure the technical influence rather than detect meaningful biological differences among samples. The methylation levels were corrected for technical variance by calculating the residuals of a linear regression model, in which the beta values of each CpG site was included as the dependent variable and the first 20 PCs, explaining 80% of the technical variance, were included as independent variables. These residuals were used in the subsequent analyses.

Stringent quality control was performed for all sites in the BABYDIET cohort to remove any sample or probe outliers. Briefly, individual sites in samples were set to

"NA" (not available) if the absolute z-score was > 4. Three samples were removed due to a large number of missing sites (> 4000). Autosomal sites were filtered and removed if there were > 20% missing values (n = 1,326). Sites on the X and Y chromosomes were removed. Furthermore, probes were removed if a SNP was found to be within ten base pairs of the 3'UTR, This left a total of 440,052 sites in 100 samples in the BABYDIET data.

HLA and Immunochip genotyping. HLA genotyping was performed using PCR-amplified DNA and non-radioactive sequence specific oligonucleotide probes[4, 44, 55] for samples from the BABYDIET and BABYDIAB studies, and by next-generation sequencing for samples from the TeenDiab study[48]. For children in the ImmunDiabRisk study and adults in the KORA study, the genotypes were determined by SNP typing at rs3104413, rs2854275, and rs9273363[56].
Additional SNP genotyping was performed using the Immunochip (Illumina). The Immunochip is a custom Illumina Infinium HD array, which was designed using public databases to provide deep coverage of all known immune loci[27]. It provides genotypes for 196,524 polymorphisms (718 small insertion deletions, 195,806 SNPs), including 186 loci previously associated with 12 autoimmune diseases (e.g. autoimmune thyroid disease or type 1 diabetes)[27]. After quality control[27, 37], filtering of variants,[27, 57] and conversion to hg19, a total of 166,263 SNPs covered 8,266 genes.

HLA-DR surface protein expression. Cord blood was collected in CPD-blood collection bags (Macopharma) and mononuclear cells were freshly isolated by density centrifugation over Lymphoprep (Progen Biotechnik GmbH). $CD14^+$ monocytes were

enriched by positive selection via magnetic activated cell sorting (Miltenyi), and stained using the monoclonal antibodies anti-CD14 FITC (clone HCD14; BioLegend), and anti-HLA-DR APC (clone L243; BioLegend). Cells were stained for 20 min at 4°C in PBS containing 0.5% bovine serum albumin, washed in PBS and incubated for 15 min at 4°C with a fixable dead cell staining kit (Life Technologies) to determine cell viability. Cells were fixed for 30 min at 4°C in 1.5% formalin, acquired on a Becton Dickinson LSRFortessa flow cytometer, and analyzed with FlowJo software (Version 10; TreeStar Inc.). The MFI of HLA-DR was determined for the $CD14^+$ monocyte population.

The cell surface expression of HLA-DR on immune cell subsets was measured on PBMCs from 5 adolescents with DR3-DQ2/DR3-DQ2 (mean age, 14.3 years; range, 11.3–17.3 years; 1 female), 5 with DR3-DQ2/DR4-DQ8 (mean age, 13.6 years; range, 12.2–16.4 years; 2 females), and 5 with DR4-DQ8/DR4-DQ8 (mean age, 11.9 years; range, 11.1–13.1 years; 4 females). To analyze HLA-DR expression on B cells and monocytes, 1×10^6 PBMCs were stained with anti-HLA-DR APC (clone L243; BioLegend), anti-CD14 FITC (clone HCD14; BioLegend), anti-CD16 APC-Cy7 (clone 3G8; BioLegend), and anti-CD19 eFluor 450 (clone SJ25C1; eBioscience). To analyze surface HLA-DR expression on different DC subsets, 1×10^6 PBMCs were stained with anti-CD3 APC (clone HIT3a; BD Bioscience), anti-CD19 APC (clone HIB19; BD Bioscience), anti-CD14 PE-Cy7 (clone M5E2; BD Bioscience), anti-CD56 PE-Cy7 (clone B159; BD Bioscience), anti-CD16 BV605 (clone 3G8; BioLegend) as exclusion markers, and anti-CD1c Alexa Fluor 488 (clone L161; BioLegend), anti-CD11c Alexa Fluor 700 (clone 3.9; eBioscience), anti-CD123 BV510 (clone 6H6; both BioLegend), and anti-HLA-DR APC-H7 (clone G46-6; BD Bioscience). All samples were stained for 20 min at 4°C in PBS lacking Ca^{2+} and

Mg^{2+} (Gibco) supplemented with 0.5% BSA, and were subsequently incubated for 5 min with 7AAD (BD Bioscience) before acquisition on an ARIA III flow cytometer (Becton Dickinson). The MFI of HLA-DR surface proteins on immune cell subsets was analyzed with FlowJo software (Version 10; TreeStar Inc.). B cells ($CD19^+$), classical monocytes ($CD14^+/CD16^-$), conventional DCs ($CD11c^+$), and pDCs ($CD123^+$) were sorted in PBS and stored at −80°C for further methylation analysis by pyrosequencing.

Pyrosequencing. Genomic DNA was isolated from PBMCs and FACS-sorted cells using Quick-gDNA™ MiniPrep (Zymo Research). The entire genomic DNA was bisulfite converted using EpiTect Fast Bisulfite Conversion Kit (Qiagen). For bias-controlled quantitative methylation analysis, a combination of methylation-sensitive high resolution melting (MS-HRM) followed by pyrosequencing was performed as previously described[3]. We used PyroMark Assay Design Software 2.0 (Qiagen) to design the PCR primers (biotinylated forward: 5'-TGGAGTTAGGGTTGTTTTATAATTG-3'; reverse: 5'-ACYAAACATACAAAATTAAAACACYCTC-3'; annealing temperature: 59.1°C) and the sequencing primer (5'- AAYCAAAAATYAAAATATCTTATT-3') to cover cg21543081 and cg12192813 in the HLA-DQB1 gene. MS-HRM was performed in a final volume of 70µl containing 1× SensiFAST™ HRM Mix (Bioline GmbH), 250 nM of the forward primer, 500 nM of the reverse primer, and all of the bisulfite-converted genomic DNA. Samples were incubated at 95°C for 5 min with 45 cycles of 95°C for 10 s and 59.1°C for 15 s, followed by a melting curve (5 s, 0.1°C/cycle) with a final hold at 20°C. The products were subsequently purified with a Select-a-Size DNA Clean & Concentrator™ Kit (Zymo) using a double-size selection protocol

with 80 µl of 95% ethanol for depletion of the top fragments and 70 µl of 95% ethanol for depletion of the bottom fragments. Pyrosequencing was performed on the Pyromark Q48 Autoprep using a PyroMark Q48 Advanced CpG Reagents Kit (Qiagen) with 100 µM of the sequencing primer. Methylation was quantified using PyroMark Q48 Autoprep 2.4.2. software with CpG analysis for each of the two investigated CpG sites. Primers for cg 13047157 (forward: 5'-Biotin-TTAAGTGTAGAGGTTTATGTAAAATTAG-3'; reverse: 5-CAAAAAAAAAMAAAAAACCCTATCCCT-3'; annealing temperature 59°C; sequencing primer 5'-ACCTATAAACCRAAWCAAAAAACAAT-3') cg13778567 (forward: 5'-AGAGGTTTTTGAGGTTATAGTGT-3'; reverse: 5'biotin-ACCCATTACTCAACCAWATAATATTAAC-3'; annealing temp 61°C; sequening primer 5'-AGGTTATAGTGTTTTTTAAGTTT-3') and cg25597745 (forward 5'-GAGTATTTTRGTGGTTGAGGGTAGAGAT-3', reverse 5'-biotin-ACCCCAACCCAAAAAAAACCTATT-C-3'; annealing temperature: 59°C; sequencing primer 5'-AGTGTAGGGTAGTTG-3') were designed as described above. MS-HRM was performed in a volume of 10µl containing 1× SensiFAST™ HRM Mix (Bioline GmbH) 500 nM of the PCR primers (exept for the cg13047157 forward primer, which was used with 250m) and 4ng of the bisulfite-converted DNA. DNA. Samples were incubated at 95°C for 5 min with 40 cycles of 95°C for 10 s and 59°C or 61°C (depending on the annealing temperature) for 15s, followed by a melting curve (5 s, 0.1°C/cycle), 30 sec 20°C and with a final hold at 4°C. Pyrosequencing was performed and analyzed as described above with the respective sequencing primers (5µM).

LDHC expression. C57BL/6J male mice (13/14 weeks old) and NOD male mice (12/18 weeks old) were purchased from Janvier. Mice were maintained and sacrificed to collect the pancreas, liver, and testes and to isolate pancreatic islets. Mice were used in accordance with the institutional animal ethics guidelines (24-9168.24-1/2012-7). Islet samples were incubated in a buffer containing a high glucose concentration (25 mM) for 2h, as previously described[58]. As a control, freely suspended islets were rested in medium containing 2 mM glucose for 2 h before the assay. Cells were centrifuged at 300 × g for 5 min before extracting total RNA. Total RNA was extracted from the pancreas, isolated islets, liver, and testes using an RNeasy Mini Kit and DNase I digestion (Qiagen), and cDNA was synthesized with RNA to cDNA EcoDry Premix (Takara). The SYBR Premix exTaq kit (Takara Bio) and a Mastercycler ep realplex thermal cycler (Eppendorf) were used to analyze cDNA triplicates. The following primers were used: HPRT, 5'-GTCAACGGGGGACATAAAAG-3' and 5'-AGGGCATATCCAACAACAAAC-3'; β-ACTIN, 5'-TGGAATCCTGTGGCATCCATGAAA-3' and 5'-TAAAACGCAGCTCAGTAACAGTCC-3'; and LDHC, 5'-TTGCTGGCGTAACTCTGAAG-3' and 5'-GCGTGGTAACAGGATGCACT-3'.

Statistical analyses.

Numbers of samples selected used for measurements were chosen on the basis of all children available for the discovery set and subsequently on reaching 90% power in the validation set, where possible. Specifically, metQTL analysis of the BABYDIET study included all children who fulfilled HLA genotype criteria and who had stored cord blood DNA (n=100). HLA cis-metQTL strengths in this discovery set (Figure 1) were used to select numbers of samples measured from the BABYDIAB dataset and

in the pyrosequencing validation, whereas all information from the KORA dataset was used in the adult validation analysis. All available RNA expression data from the BABYDIET study was used for HLA class II gene expression analysis. All available samples from the ImmuneDiabRisk study were used for HLA protein expression discovery analysis, and the validation experiments in immune cell subsets were chosen for 90% power on the basis of the differences observed between genotypes in the pyrosequencing validation experiment (Supplementary Fig. 2). The replication cohort used to analyse LDHC methylation included all children with early insulin autoimmunity with samples for DNA methylation and who did not participate in the BABYDIET study.

All measurements were performed using coded material to ensure that investigators were blinded to the HLA genotype and autoantibody status of the samples. After depositing of results, data was unblinded and analysed. For the analysis of epigenome-wide associations in the neonate cohort, two linear regression models were calculated for each CpG site, where the methylation status of each CpG site was modelled as the outcome:

$$\text{Model 1:}\quad y_j = \beta_0 + \beta_{gt,j} * GT_j + \beta_{FDR,j} * FDR_j + \beta_{sex,j} * sex_j + \sum_{k>1}^{k} \beta_{CE,j,k} * CE_{j,k} + \varepsilon_j$$

$$\text{Model 2:}\quad y_j = \beta_0 + \beta_{FDR,j} * FDR_j + \beta_{sex,j} * sex_j + \sum_{k>1}^{k} \beta_{CE,j,k} * CE_{j,k} + \varepsilon_j$$

where y_j is the beta value adjusted for technical variation at the analyzed CpG site of child j, β_0 is the coefficient of the intercept, $\beta_{gt,j}$ is the coefficient of the genotype (GT_j) of child j (DR3-DQ2/DR3-DQ2, DR3-DQ2/DR4-DQ8, or DR4-DQ8/DR4-DQ8), $\beta_{FDR,j}$ is the coefficient of the first-degree relative status (FDR_j) of child j (0 if only one relative, 1 if more than 1 first-degree relative had type 1 diabetes), $\beta_{sex,j}$ is

the coefficient of sex_j, which is the sex of child j. $\sum_{k>1}^{k} \beta_{CE,j,k} * CE_{j,k}$ is the sum of the coefficients $\beta_{CE,j,k}$ of the cell estimates $CE_{j,k}$, as calculated by Houseman et al.[54], for child j, and ϵ_j is the independent and identically distributed normal noise. For analyses of the BABYDIAB and the KORA cohorts, age was included as a variable in both models. The significance of the genotype was assessed using the ANOVA function in R, which compares two nested linear regression models using the F-test. The genotype effect was significant if the p-value was < 0.05 after correcting for multiple testing using FDR for all analyzed methylation sites. The methylation pattern according to HLA genotype was identified using Scheffe's post-hoc test (package agricolae, version 1.2-4) as a single-step multiple comparison procedure.

For the analysis of HLA-DQA1, HLA-DQB1, and HLA-DRB5, the data were first corrected using the following generalized additive mixed model:

$$y_{i,j} = \beta_{0,i} + \sum_{k} b_{k,i}(t_{i,j})\beta_{k,i} + b_{ID,i}\, x_{ID,j} + \epsilon_{i,j}$$

where $y_{i,j}$ is the gene expression for gene i in child j, and $\beta_{0,i}$ is the intercept or average gene-wise expression. Adjusting for the different measurements over time, a spline function, $\sum_k b_{k,i}(t_{i,j})\beta_{k,i}$, was included per gene over all samples, where $t_{i,j}$ is the age at sampling, b_k are the basis functions of the spline with k being the number of knots, and $\beta_{k,i}$ is the corresponding spline coefficient. We also included a child specific random effect $x_{ID,j}$ with $b_{ID,i} \sim N(0,\sigma_i)$, where σ_i was estimated per gene, to correct for multiple measurements per child, and an error term $\epsilon_{i,j}$ as i.i.d. normally distributed noise. The residuals of this model were used to correlate the adjusted gene expression with the HLA genotypes. This model was estimated for all three HLA class II genes.

In order to assess how methylation status and gene expression correlate, a linear regression analysis was performed using the methylation status of cord blood samples and the gene expression in the first sample for which gene expression was available for each child (HLA-DR3-DQ2/DR3-DQ2, n = 12; HLA- DR3-DQ2/DR4-DQ8, n = 36; HLA-DR4-DQ8/DR4-DQ8, n = 15). The analysis was performed for 91 methylation sites between HLA-DRA and HLA-DQB2 that were differentially methylated among the HLA-DR-DQ genotypes and corrected for the Houseman cell estimates. To investigate whether the expression of the transcripts was driven by methylation or HLA status, we compared two linear models for each gene and each methylation site. The comparison was performed using the anova function in R, which assesses the difference of two nested models using the F-test.

Model 1: $y_j = \beta_0 + \beta_{GT,j} * GT_j + \beta_{CpG,j} * CpG_j + \beta_{age,j} * age_j + \varepsilon_j$

Model 2: $y_j = \beta_0 + \beta_{CpG,j} * CpG_j + \beta_{age,j} * age_j + \varepsilon_j$

where y_j is the gene expression status of the analyzed gene in child j, β_0 is the coefficient of the intercept, $\beta_{GT,j}$ is the coefficient of the genotype (GT_j) of child j (DR3-DQ2/DR3-DQ2, DR3-DQ2/DR4-DQ8, or DR4-DQ8/DR4-DQ8), $\beta_{CpG,j}$ is the methylation status at one of the CpG sites (CpG_j) of child j (corrected for sex of the child, first degree relative status and the cell estimates), $\beta_{age,j}$ is the coefficient of age_j, which is the age of child j at the time of sampling, and ϵ_j is the independent and identically distributed normal noise. The first model estimated the effect of methylation on the transcriptome using linear regression, while the second model used linear regression to explain the transcriptome by methylation site and HLA

status. Both models included the child's age at the time of sample collection as a confounding variable.

Analysis of variance (ANOVA) was used to evaluate the differences in pyrosequencing data and HLA-DR protein expression in relation to the HLA genotype. HLA-DR protein expression in cord blood monocytes was compared between neonates with the HLA-DR3-DQ2 and HLA-DR4-DQ8 haplotypes using unpaired two-sided Student's t-test.

Acknowledgements

The study was funded by grants from the German Federal Ministry of Education and Research (BMBF) to the German Center for Diabetes Research (DZD e.V.) and to E.B. (FZ 01KU0152A), Helmholtz Zentrum Muenchen, Juvenile Diabetes Research Foundation (17-2013-525), and iMed–the Helmholtz Initiative on Personalized Medicine. E.B. is supported by the DFG Research Center and Cluster of Excellence - Center for Regenerative Therapies Dresden (FZ 111).

Author contributions

Study concept and design: AZ, EB. Statistical analysis: AK, ML, MW, SW, RW. Acquisition of data: RF, JKn, TT, ES, MH, AH, JT, AZ. Analysis and interpretation of data: AK, RF, TT, ES, JKn, JKr, AH, AZ, EB. Drafting of the manuscript: AK, JKn, JKr, MH, AH, EB, AZ. Critical revision of the manuscript for important intellectual content: AK, JKr, EB, JT, AZ. All authors read and approved the final manuscript

Competing financial interests

The authors declare no competing financial interests.

Data availability

The data and computational code that support the findings of this study are available from the corresponding author upon request.

Corresponding author

Anette-G Ziegler: anette-g.ziegler@helmholtz-muenchen.de

References

[1] Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM et al. Defining the role of the MHC in autoimmunity: a review and pooled analysis. PLoS Genet, 2008;4:e1000024.

[2] Bonifacio E. Predicting type 1 diabetes using biomarkers. Diabetes Care, 2015;38:989-96.

[3] Serr I, Furst RW, Achenbach P, Scherm MG, Gokmen F, Haupt F et al. Type 1 diabetes vaccine candidates promote human Foxp3(+)Treg induction in humanized mice. Nat Commun, 2016;7:10991.

[4] 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;46:712-20.

[5] Arnheim N, Strange C, Erlich H. Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: studies of the HLA class II loci. Proc Natl Acad Sci U S A, 1985;82:6970-4.

[6] Congia M, Cucca F, Frau F, Lampis R, Melis L, Clemente MG et al. A gene dosage effect of the DQA1*0501/DQB1*0201 allelic combination influences the clinical heterogeneity of celiac disease. Hum Immunol, 1994;40:138-42.

[7] Shi Y, Zou M, Robb D, Farid NR. Typing for major histocompatibility complex class II antigens in thyroid tissue blocks: association of Hashimoto's thyroiditis with HLA-DQA0301 and DQB0201 alleles. J Clin Endocrinol Metab, 1992;75:943-6.

[8] Morling N, Andersen V, Fugger L, Georgsen J, Halberg P, Oxholm P et al. Immunogenetics of rheumatoid arthritis and primary Sjogren's syndrome: DNA polymorphism of HLA class II genes. Dis Markers, 1991;9:289-96.

[9] Vieira ML, Caillat-Zucman S, Gajdos P, Cohen-Kaminsky S, Casteur A, Bach JF. Identification by genomic typing of non-DR3 HLA class II genes associated with myasthenia gravis. J Neuroimmunol, 1993;47:115-22.

[10] Morris DL, Taylor KE, Fernando MM, Nititham J, Alarcon-Riquelme ME, Barcellos LF et al. Unraveling multiple MHC gene associations with systemic lupus erythematosus: model choice indicates a role for HLA alleles and non-HLA genes in Europeans. Am J Hum Genet, 2012;91:778-93.

[11] Mangalam AK, Taneja V, David CS. HLA class II molecules influence susceptibility versus protection in inflammatory diseases by determining the cytokine profile. J Immunol, 2013;190:513-8.

[12] Lundin KE, Sollid LM, Qvigstad E, Markussen G, Gjertsen HA, Ek J et al. T lymphocyte recognition of a celiac disease-associated cis- or trans-encoded HLA-DQ alpha/beta-heterodimer. J Immunol, 1990;145:136-9.

[13] Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E. Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. J Exp Med, 1989;169:345-50.

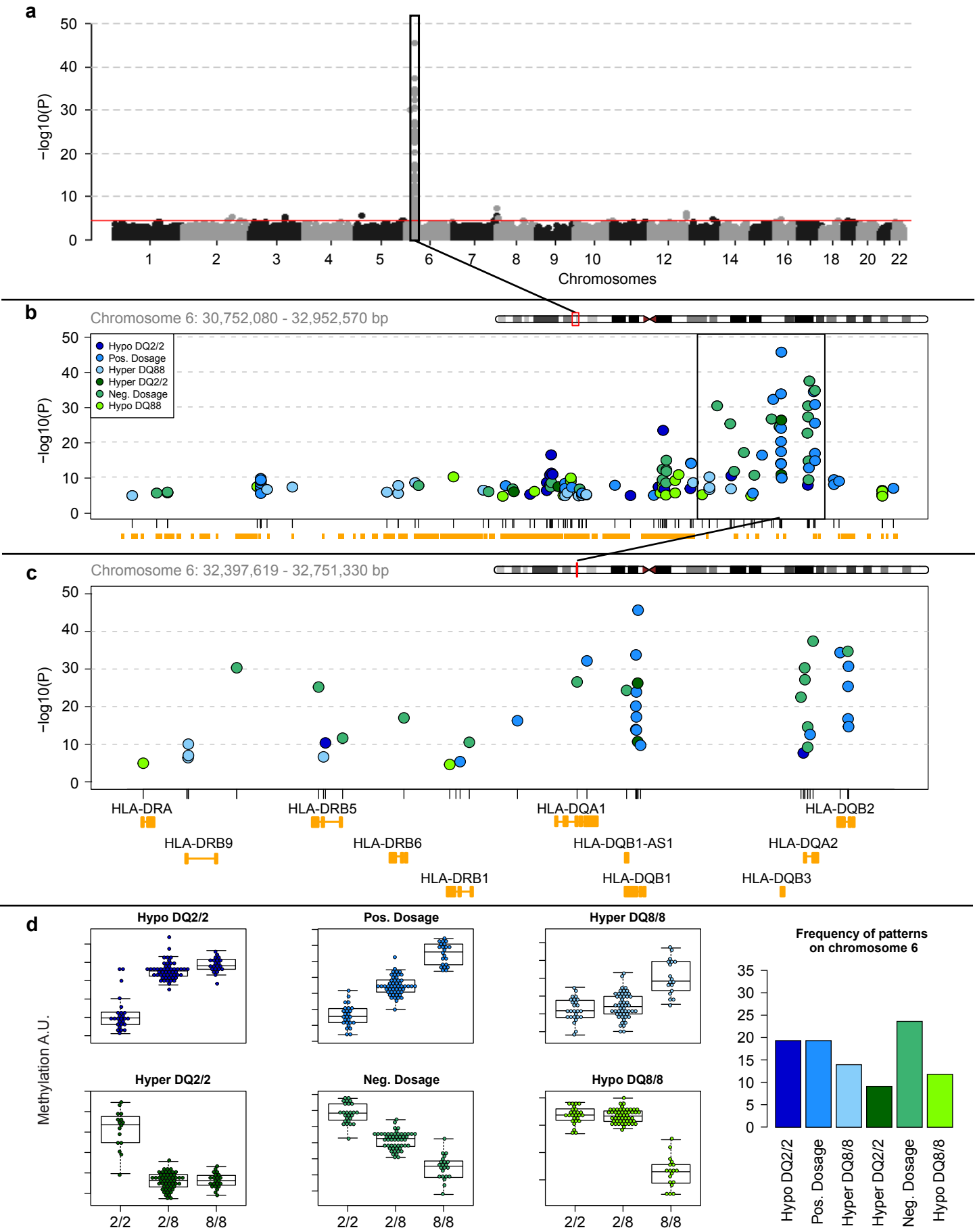
[14] Spurkland A, Sollid LM, Ronningen KS, Bosnes V, Ek J, Vartdal F et al. Susceptibility to develop celiac disease is primarily associated with HLA-DQ alleles. Hum Immunol, 1990;29:157-65.

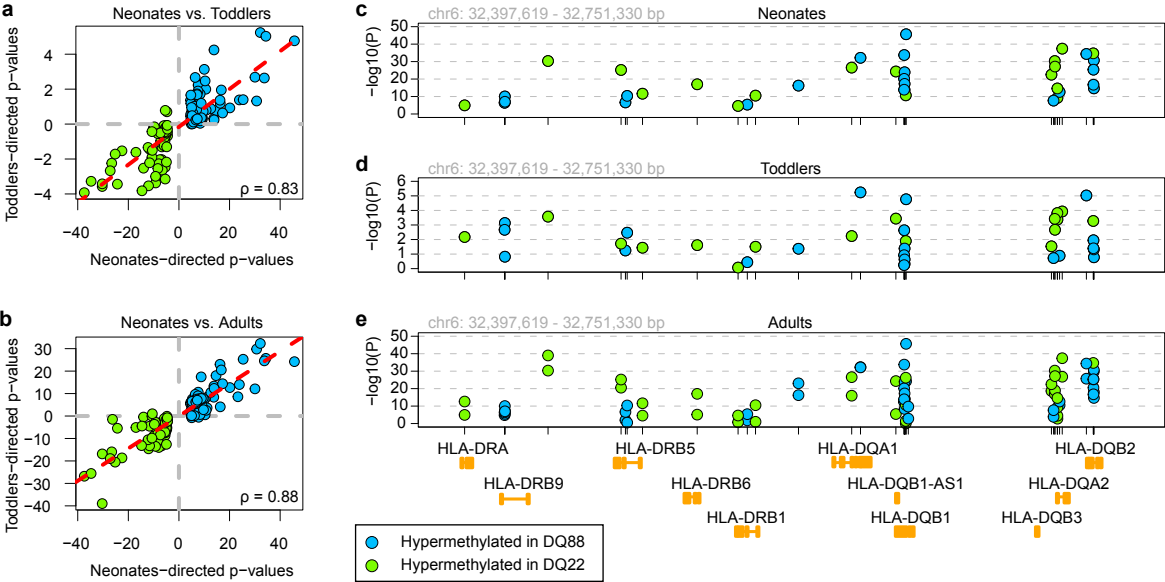
[15] Liu E, Lee HS, Aronsson CA, Hagopian WA, Koletzko S, Rewers MJ et al. Risk of pediatric celiac disease according to HLA haplotype and country. N Engl J Med, 2014;371:42-9.

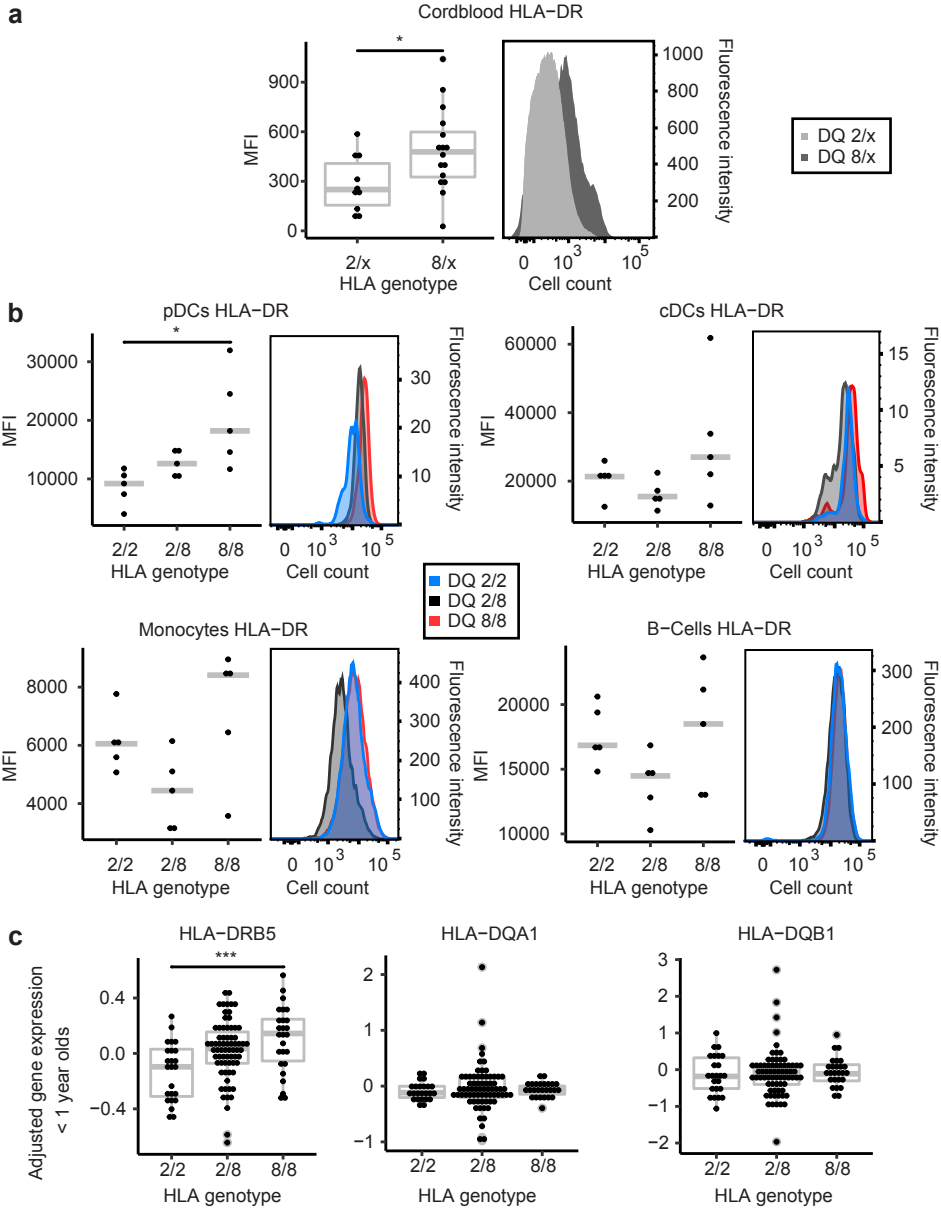
[16] Kockum I, Wassmuth R, Holmberg E, Michelsen B, Lernmark A. HLA-DQ primarily confers protection and HLA-DR susceptibility in type I (insulin-dependent) diabetes studied in population-based affected families and controls. Am J Hum Genet, 1993;53:150-67.
[17] Sheehy MJ, Scharf SJ, Rowe JR, Neme de Gimenez MH, Meske LM, Erlich HA et al. A diabetes-susceptible HLA haplotype is best defined by a combination of HLA-DR and -DQ alleles. J Clin Invest, 1989;83:830-5.
[18] Khalil I, Deschamps I, Lepage V, al-Daccak R, Degos L, Hors J. Dose effect of cis- and trans-encoded HLA-DQ alpha beta heterodimers in IDDM susceptibility. Diabetes, 1992;41:378-84.
[19] Jones MJ, Fejes AP, Kobor MS. DNA methylation, genotype and gene expression: who is driving and who is along for the ride? Genome Biol, 2013;14:126.
[20] Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martin D et al. Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. Cell, 2016;167:1398-414 e24.
[21] Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. Cell, 2016;167:1415-29 e19.
[22] Ziegler AG, Bonifacio E, Group B-BS. Age-related islet autoantibody incidence in offspring of patients with type 1 diabetes. Diabetologia, 2012;55:1937-43.
[23] Parikka V, Nanto-Salonen K, Saarinen M, Simell T, Ilonen J, Hyoty H et al. Early seroconversion and rapidly increasing autoantibody concentrations predict prepubertal manifestation of type 1 diabetes in children at genetic risk. Diabetologia, 2012;55:1926-36.
[24] Krischer JP, Lynch KF, Schatz DA, Ilonen J, Lernmark A, Hagopian WA et al. The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. Diabetologia, 2015;58:980-7.
[25] Handunnetthi L, Ramagopalan SV, Ebers GC, Knight JC. Regulation of major histocompatibility complex class II gene expression, genetic variation and disease. Genes Immun, 2010;11:99-112.
[26] Ferreira RC, Guo H, Coulson RM, Smyth DJ, Pekalski ML, Burren OS et al. A type I interferon transcriptional signature precedes autoimmunity in children genetically at risk for type 1 diabetes. Diabetes, 2014;63:2538-50.
[27] Onengut-Gumuscu S, Chen WM, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat Genet, 2015;47:381-6.
[28] Hummel M, Bonifacio E, Schmid S, Walter M, Knopff A, Ziegler AG. Brief communication: early appearance of islet autoantibodies predicts childhood type 1 diabetes in offspring of diabetic parents. Ann Intern Med, 2004;140:882-6.
[29] Liu Y, Ding J, Reynolds LM, Lohman K, Register TC, De La Fuente A et al. Methylomics of gene expression in human monocytes. Hum Mol Genet, 2013;22:5065-74.
[30] Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E et al. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. Nat Genet, 2008;40:904-8.

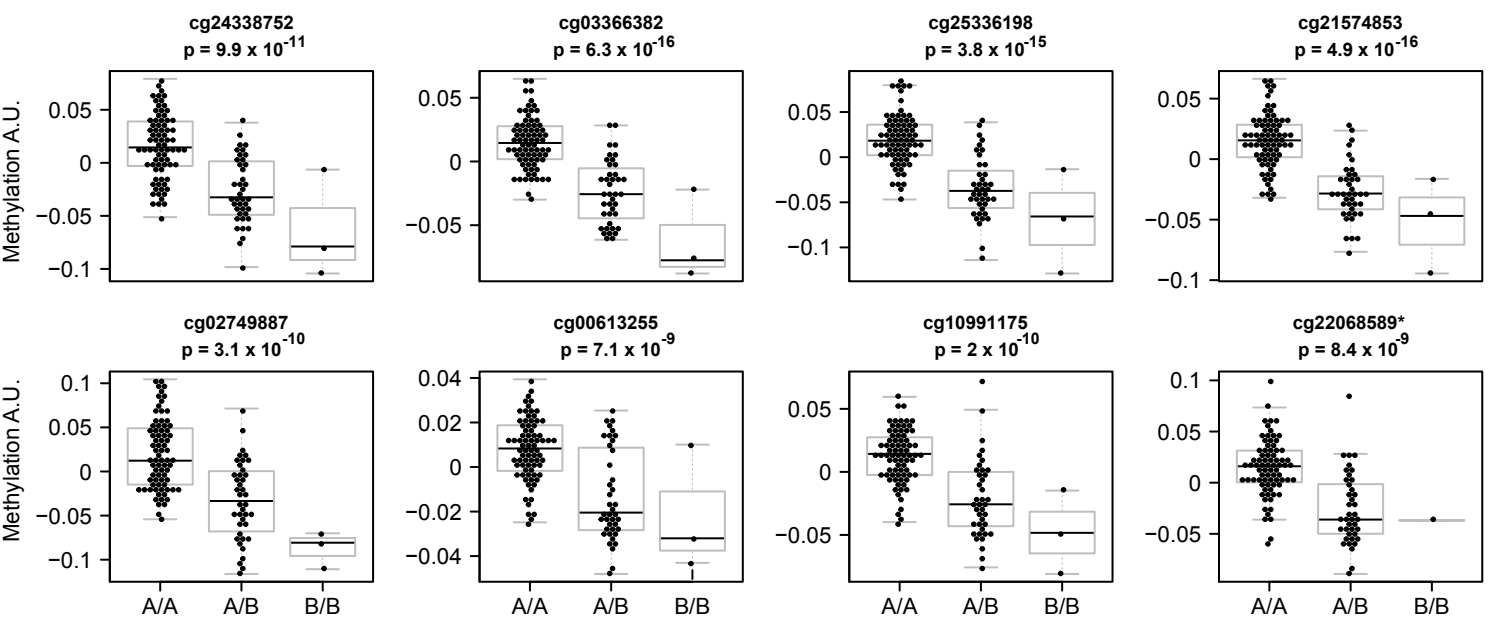
[31] Hu X, Deutsch AJ, Lenz TL, Onengut-Gumuscu S, Han B, Chen WM et al. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. Nat Genet, 2015;47:898-905.
[32] Noble JA, Erlich HA. Genetics of type 1 diabetes. Cold Spring Harb Perspect Med, 2012;2:a007732.
[33] Hong X, Hao K, Ladd-Acosta C, Hansen KD, Tsai HJ, Liu X et al. Genome-wide association study identifies peanut allergy-specific loci and evidence of epigenetic mediation in US children. Nat Commun, 2015;6:6304.
[34] Cepek P, Zajacova M, Kotrbova-Kozak A, Silhova E, Cerna M. DNA methylation and mRNA expression of HLA-DQA1 alleles in type 1 diabetes mellitus. Immunology, 2016;148:150-9.
[35] Cavalli G, Hayashi M, Jin Y, Yorgov D, Santorico SA, Holcomb C et al. MHC class II super-enhancer increases surface expression of HLA-DR and HLA-DQ and affects cytokine production in autoimmune vitiligo. Proc Natl Acad Sci U S A, 2016;113:1363-8.
[36] Raj P, Rai E, Song R, Khan S, Wakeland BE, Viswanathan K et al. Regulatory polymorphisms modulate the expression of HLA class II molecules and promote autoimmunity. Elife, 2016;5.
[37] Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet, 2012;44:991-1005.
[38] Yang D, Chertov O, Bykovskaia SN, Chen Q, Buffo MJ, Shogan J et al. Beta-defensins: linking innate and adaptive immunity through dendritic and T cell CCR6. Science, 1999;286:525-8.
[39] Werner L, Paclik D, Fritz C, Reinhold D, Roggenbuck D, Sturm A. Identification of pancreatic glycoprotein 2 as an endogenous immunomodulator of innate and adaptive immune responses. J Immunol, 2012;189:2774-83.
[40] Fradin D, Le Fur S, Mille C, Naoui N, Groves C, Zelenika D et al. Association of the CpG methylation pattern of the proximal insulin gene promoter with type 1 diabetes. PLoS One, 2012;7:e36278.
[41] Stunnenberg HG, International Human Epigenome C, Hirst M. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. Cell, 2016;167:1145-9.
[42] Achenbach P, Koczwara K, Knopff A, Naserke H, Ziegler AG, Bonifacio E. Mature high-affinity immune responses to (pro)insulin anticipate the autoimmune cascade that leads to type 1 diabetes. J Clin Invest, 2004;114:589-97.
[43] Alves MG, Martins AD, Rato L, Moreira PI, Socorro S, Oliveira PF. Molecular mechanisms beyond glucose transport in diabetes-related male infertility. Biochim Biophys Acta, 2013;1832:626-35.
[44] Hummel S, Pfluger M, Hummel M, Bonifacio E, Ziegler AG. Primary dietary intervention study to reduce the risk of islet autoimmunity in children at increased risk for type 1 diabetes: the BABYDIET study. Diabetes Care, 2011;34:1301-5.
[45] Dopico XC, Evangelou M, Ferreira RC, Guo H, Pekalski ML, Smyth DJ et al. Widespread seasonal gene expression reveals annual differences in human immunity and physiology. Nat Commun, 2015;6:7000.

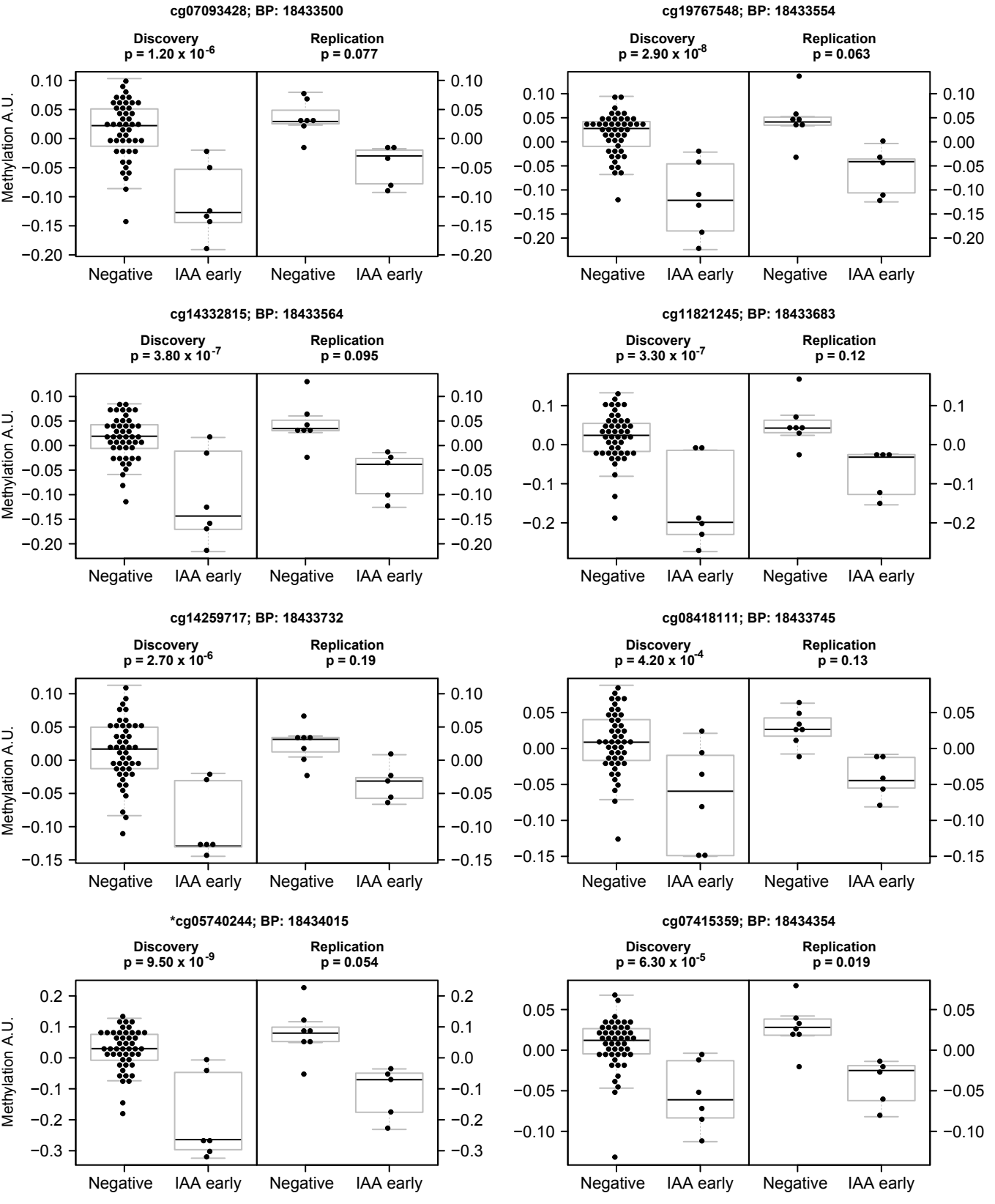
[46] Ziegler AG, Hillebrand B, Rabl W, Mayrhofer M, Hummel M, Mollenhauer U et al. On the appearance of islet associated autoimmunity in offspring of diabetic mothers: a prospective study from birth. Diabetologia, 1993;36:402-8.
[47] Ziegler AG, Hummel M, Schenker M, Bonifacio E. Autoantibody appearance 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;48:460-8.
[48] Ziegler AG, Meier-Stiegen F, Winkler C, Bonifacio E, Group TS. Prospective evaluation of risk factors for the development of islet autoimmunity and type 1 diabetes during puberty--TEENDIAB: study design. Pediatr Diabetes, 2012;13:419-24.
[49] Rathmann W, Kowall B, Tamayo T, Giani G, Holle R, Thorand B et al. Hemoglobin A1c and glucose criteria identify different subjects as having type 2 diabetes in middle-aged and older populations: the KORA S4/F4 Study. Ann Med, 2012;44:170-7.
[50] Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome Biol, 2015;16:37.
[51] Fortin JP, Hansen KD. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. Genome Biol, 2015;16:180.
[52] Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM et al. High density DNA methylation array with single CpG site resolution. Genomics, 2011;98:288-95.
[53] Smyth GK, Michaud J, Scott HS. Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics, 2005;21:2067-75.
[54] Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics, 2012;13:86.
[55] 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;42:671-7.
[56] Nguyen C, Varney MD, Harrison LC, Morahan G. Definition of high-risk type 1 diabetes HLA-DR and HLA-DQ types using only three single nucleotide polymorphisms. Diabetes, 2013;62:2135-40.
[57] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet, 2006;38:904-9.
[58] Knoch KP, Meisterfeld R, Kersting S, Bergert H, Altkruger A, Wegbrod C et al. cAMP-dependent phosphorylation of PTB1 promotes the expression of insulin secretory granule proteins in beta cells. Cell Metab, 2006;3:123-34.


a
−log10(P)
Chromosomes
b
Chromosome 6: 30,752,080 - 32,952,570 bp
Hypo DQ2/2
Pos. Dosage
Hyper DQ88
Hyper DQ2/2
Neg. Dosage
Hypo DQ88
c
Chromosome 6: 32,397,619 - 32,751,330 bp
HLA-DRA
HLA-DRB5
HLA-DQA1
HLA-DQB2
HLA-DRB9
HLA-DRB6
HLA-DQB1-AS1
HLA-DQA2
HLA-DRB1
HLA-DQB1
HLA-DQB3
d
Methylation A.U.
Hypo DQ2/2
Pos. Dosage
Hyper DQ8/8
Hyper DQ2/2
Neg. Dosage
Hypo DQ8/8
2/2
2/8
8/8
Frequency of patterns on chromosome 6


a
Neonates vs. Toddlers
Toddlers−directed p−values
Neonates−directed p−values
ρ = 0.83
b
Neonates vs. Adults
Toddlers−directed p−values
Neonates−directed p−values
ρ = 0.88
c
chr6: 32,397,619 - 32,751,330 bp
Neonates
−log10(P)
d
chr6: 32,397,619 - 32,751,330 bp
Toddlers
−log10(P)
e
chr6: 32,397,619 - 32,751,330 bp
Adults
−log10(P)
HLA-DRA
HLA-DRB5
HLA-DQA1
HLA-DQB2
HLA-DRB9
HLA-DRB6
HLA-DQB1-AS1
HLA-DQA2
HLA-DRB1
HLA-DQB1
HLA-DQB3
Hypermethylated in DQ88
Hypermethylated in DQ22


a
Cordblood HLA−DR
MFI
Fluorescence intensity
HLA genotype
Cell count
2/x
8/x
DQ 2/x
DQ 8/x
b
pDCs HLA−DR
cDCs HLA−DR
Monocytes HLA−DR
B−Cells HLA−DR
2/2
2/8
8/8
DQ 2/2
DQ 2/8
DQ 8/8
c
HLA−DRB5
HLA−DQA1
HLA−DQB1
Adjusted gene expression
< 1 year olds


cg24338752
p = 9.9 x 10-11
cg03366382
p = 6.3 x 10-16
cg25336198
p = 3.8 x 10-15
cg21574853
p = 4.9 x 10-16
cg02749887
p = 3.1 x 10-10
cg00613255
p = 7.1 x 10-9
cg10991175
p = 2 x 10-10
cg22068589*
p = 8.4 x 10-9
Methylation A.U.
A/A
A/B
B/B


cg07093428; BP: 18433500
Discovery
p = 1.20 x 10-6
Replication
p = 0.077
cg19767548; BP: 18433554
Discovery
p = 2.90 x 10-8
Replication
p = 0.063
cg14332815; BP: 18433564
Discovery
p = 3.80 x 10-7
Replication
p = 0.095
cg11821245; BP: 18433683
Discovery
p = 3.30 x 10-7
Replication
p = 0.12
cg14259717; BP: 18433732
Discovery
p = 2.70 x 10-6
Replication
p = 0.19
cg08418111; BP: 18433745
Discovery
p = 4.20 x 10-4
Replication
p = 0.13
*cg05740244; BP: 18434015
Discovery
p = 9.50 x 10-9
Replication
p = 0.054
cg07415359; BP: 18434354
Discovery
p = 6.30 x 10-5
Replication
p = 0.019
Methylation A.U.
Negative
IAA early

Supplementary Information

Supplementary Tables 1-5 and Supplementary Figures 1-7

Allele-specific methylation of type 1 diabetes susceptibility genes

Alida SD Kindt[a], Rainer W Fuerst[b], Jan Knoop[b], Michael Laimighofer[a], Tanja Telieps[c], Markus Hippich[b], Maria A Woerheide[a], Simone Wahl[d,e,f], Rory Wilson[d,e], Eva-Maria Sedlmeier[b], Angela Hommel[g], John A Todd[h], Jan Krumsiek[a,f], Anette-G Ziegler[b,i*], Ezio Bonifacio[c,g,i]

[a] Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Germany; alida.kindt@helmholtz-muenchen.de; michael.laimighofer@helmholtz-muenchen.de; jan.krumsiek@helmholtz-muenchen.de

[b] Institute of Diabetes Research, Helmholtz Zentrum München, Neuherberg, Germany jan.knoop@helmholtz-muenchen.de; eva-maria.sedlmeier@helmholtz-muenchen.de; anette-g.ziegler@helmholtz-muenchen.de

[c] Institute of Diabetes and Obesity, Helmholtz Zentrum München, Neuherberg, Germany; tanja.telieps@helmholtz-muenchen.de

[d] Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany; rory.wilson@helmholtz-muenchen.de

[e] Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[f] German Center for Diabetes Research (DZD), Neuherberg, Germany

[g] Center for Regenerative Therapies - Dresden, Faculty of Medicine Carl Gustav Carus, Technische Universität, Dresden, Germany; ezio.bonifacio@tu-dresden.de

[h] JDRF/Wellcome Trust Diabetes and Inflammation Laboratory, Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK; jatodd@well.ox.ac.uk

[i] Forschergruppe Diabetes e.V., Neuherberg, Germany

[*] Corresponding author: Anette-G Ziegler (anette-g.ziegler@helmholtz-muenchen.de)

Table of Contents

Supplementary Files

Supplementary Tables

Supplementary Table 1 │ Differentially methylated CpG sites (n = 181)

bp, base pair; TSS, transcription start site; FDR, false-discovery rate; bonf, Bonferroni p value

Excel File: YJAUT 2017 488R1 Kindt_et_al_Supplementary_Table 1.xls

Supplementary Table 2 │ Differentially methylated CpG sites in *INS* (n = 8)

bp, base pair; * marks the one CpG site where a SNP lies within 10 base pairs of 3'UTR

CpG site	p-value	Benjamini-Hochberg p-value	Bonferroni p-value	Chromosome	bp	Gene
cg24338752	< 0.001	< 0.001	< 0.001	11	2182672	*INS*
cg03366382	< 0.001	< 0.001	< 0.001	11	2182644	*INS*
cg25336198	< 0.001	< 0.001	< 0.001	11	2182618	*INS*
cg21574853	< 0.001	< 0.001	< 0.001	11	2182573	*INS*
cg02749887	< 0.001	< 0.001	< 0.001	11	2182540	*INS*
cg00613255	< 0.001	< 0.001	< 0.001	11	2182376	*INS*
cg10991175	< 0.001	< 0.001	< 0.001	11	2182300	*INS*
cg22068589*	< 0.001	< 0.001	< 0.001	11	2181072	*INS*

Supplementary Table 3 │ Differentially methylated CpG sites (n = 61) in type 1 diabetes susceptibility loci

SNP, single nucleotide polymorphism; bp, base pair

SNP	Chromosome	bp	Gene	CpG site	p-value	Benjamini-Hochberg p-value
scobl rs4948088	1	118472455	*WDR3*	cg15080701	<0.001	<0.001
tagap rs1738074	1	153934248	*SLC39A1*	cg09079613	<0.001	<0.001
rgs1 rs2816316	1	192542576	*RGS1*	cg24886076	0.001	0.037
rgs1 rs2816316	1	192544902	*RGS1*	cg02586212	<0.001	<0.001
prkcq rs7090696	2	20688014	*RHOB*	cg18301717	<0.001	<0.001
rasgrp1 rs17574546	2	42160150	*C2orf91*	cg15258521	<0.001	<0.001
tnfaip3 rs6920220	2	74648708	*WDR54*	cg10386659	<0.001	<0.001
ccr3 rs11711054	3	46394550	*CCR2*	cg07743747	<0.001	<0.001
rs10517086	3	62358980	*FEZF2*	cg19629292	<0.001	<0.001
gpr rs9585056	3	115417878	*GAP43*	cg01710607	<0.001	<0.001
stat4 rs7574865	4	3519599	*LRPAP1*	cg10359931	<0.001	<0.001
prkd2 rs425105	4	40058985	*LOC344967*	cg04811621	<0.001	<0.001
ifih1 rs2111485	4	83821584	*THAP9*	cg27182555	<0.001	<0.001
cd226 rs763361	5	137801033	*EGR1*	cg08611430	<0.001	<0.001
gab3 rs2664170	6	30524112	*GNL1*	cg26791826	<0.001	<0.001
c6orf173 rs9388489	6	31276187	*HLA-C*	cg15225267	<0.001	0.018

glis3 rs7020673	**6**	**153476415**	***RGS17***	**cg22867315**	**<0.001**	**<0.001**
ormdl3 rs2290400	**6**	**168502888**	***FRMD1***	**cg06444195**	**<0.001**	**<0.001**
sirpg rs2281808	**7**	**923365**	***GET4***	**cg18612044**	**<0.001**	**<0.001**
ptpn22 rs2476601	**7**	**5265722**	***WIPI2***	**cg20239604**	**<0.001**	**<0.001**
sh2b3 rs3184504	**7**	**41982794**	***GLI3***	**cg14258366**	**<0.001**	**<0.001**
ikzf1 rs10272724	**7**	**75677469**	***MDH2***	**cg22813794**	**<0.001**	**<0.001**
bach2 rs3757247	**7**	**100881842**	***CLDN15***	**cg08636573**	**<0.001**	**0.005**
il2r rs12722495	**10**	**6178319**	***PFKFB3***	**cg17191567**	**<0.001**	**<0.001**
il2r rs12722495	**10**	**6186542**	***PFKFB3***	**cg19801868**	**<0.001**	**0.023**
il2 rs2069763	**10**	**118884059**	***KIAA1598***	**cg16175006**	**<0.001**	**<0.001**
ptpn22 rs6679677	**11**	**126703882**	***KIRREL3-AS2***	**cg08336938**	**<0.001**	**<0.001**
ifih1 rs1990760	**12**	**56652374**	***ANKRD52***	**cg11875676**	**<0.001**	**<0.001**
cyp27b1 rs10877012	**12**	**58146742**	***CDK4***	**cg00599273**	**<0.001**	**<0.001**
cyp27b1 rs10877012	**12**	**58146821**	***CDK4***	**cg03829839**	**<0.001**	**0.001**
erbb3 rs2292239	**12**	**129299332**	***SLC15A4***	**cg05253716**	**<0.001**	**<0.001**
il10 rs3024505	**13**	**61913482**	***PCDH20***	**cg08438164**	**<0.001**	**<0.001**
dlk1 rs941576	**13**	**96742905**	***HS6ST3***	**cg24974423**	**<0.001**	**<0.001**
zfp36l1 rs1465788	**14**	**69261079**	***ZFP36L1***	**cg16751098**	**0.002**	**0.049**
tlr8 rs5979785	**15**	**35837840**	***DPH6-AS1***	**cg26481784**	**<0.001**	**<0.001**
rnls rs7068821	**15**	**73075587**	***ADPGK-AS1***	**cg01171597**	**<0.001**	**<0.001**
ctsh rs3825932	**15**	**79237157**	***CTSH***	**cg07448499**	**<0.001**	**0.025**
ctsh rs3825932	**15**	**79243526**	***CTSH***	**cg17506061**	**<0.001**	**<0.001**
kiaa350 rs12708716	**16**	**11127037**	***CLEC16A***	**cg00121339**	**<0.001**	**0.018**
kiaa350 rs12708716	**16**	**11256561**	***CLEC16A***	**cg05080940**	**<0.001**	**<0.001**
rs12927773	**16**	**11350112**	***SOCS1***	**cg27540841**	**<0.001**	**0.018**
rs12927773	**16**	**11439710**	***RMI2***	**cg00044050**	**<0.001**	**<0.001**
il27 rs151181	**16**	**28608288**	***SULT1A2***	**cg00931491**	**<0.001**	**<0.001**
il27 rs151181	**16**	**28608894**	***SULT1A2***	**cg04270652**	**<0.001**	**<0.001**
il27 rs151181	**16**	**28609220**	***SULT1A2***	**cg04643524**	**<0.001**	**0.028**
rs7202877	**16**	**75279566**	***BCAR1***	**cg16519668**	**<0.001**	**0.008**
tyk2 rs2304256	**17**	**5993686**	***WSCD1***	**cg23098912**	**<0.001**	**<0.001**
rs7221109	**17**	**38805042**	***SMARCE1***	**cg07438660**	**<0.001**	**<0.001**
ptpn2 rs45450798	**18**	**12775841**	***PSMG2***	**cg20581490**	**<0.001**	**<0.001**
ptpn2 rs45450798	**18**	**12777645**	***PSMG2***	**cg23598886**	**<0.001**	**<0.001**
ptpn2 rs45450798	**18**	**12777786**	***PSMG2***	**cg23544223**	**<0.001**	**<0.001**
ptpn2 rs45450798	**18**	**12777974**	***PSMG2***	**cg09945482**	**<0.001**	**<0.001**
ptpn2 rs45450798	**18**	**12778029**	***PSMG2***	**cg24737193**	**<0.001**	**<0.001**
rs4900384	**19**	**36120171**	***RBM42***	**cg24429831**	**<0.001**	**<0.001**
scap2 rs7804356	**20**	**1373655**	***FKBP1A-SDCBP2***	**cg21332426**	**<0.001**	**<0.001**
ubash3a rs3788013	**21**	**43823749**	***UBASH3A***	**cg27111890**	**<0.001**	**0.013**
ubash3a rs3788013	**21**	**43824071**	***UBASH3A***	**cg00134539**	**<0.001**	**0.007**
ubash3a rs3788013	**21**	**43835592**	***UBASH3A***	**cg14174221**	**<0.001**	**<0.001**
ctla4 rs3087243	**22**	**29781841**	***AP1B1***	**cg11145632**	**<0.001**	**<0.001**
rs5753037	**22**	**30112403**	***CABP7***	**cg27665648**	**<0.001**	**<0.001**
il2rb rs229541	**22**	**37585549**	***C1QTNF6***	**cg05209515**	**<0.001**	**<0.001**

Supplementary Table 4 | Differentially methylated CpG sites (n = 26) in type 2 diabetes susceptibility loci

SNP, single nucleotide polymorphism; bp, base pair

SNP	Chromosome	bp	Gene	CpG site	p-value	Benjamini-Hochberg p-value
gckr_rs780094	2	27665017	*KRTCAP3*	cg18428193	< 0.001	< 0.001
gckr_rs780094	2	27665079	*KRTCAP3*	cg04845466	< 0.001	< 0.001
gckr_rs780094	2	27665128	*KRTCAP3*	cg24768116	< 0.001	< 0.001
gckr_rs780094	2	27665139	*KRTCAP3*	cg12000995	< 0.001	< 0.001
gckr_rs780094	2	27665141	*KRTCAP3*	cg12648201	< 0.001	< 0.001
gckr_rs780094	2	27665150	*KRTCAP3*	cg21248554	< 0.001	< 0.001
gckr_rs780094	2	27665306	*KRTCAP3*	cg17158414	< 0.001	< 0.001
gckr_rs780094	2	27665507	*KRTCAP3*	cg02592271	< 0.001	< 0.001
gckr_rs780094	2	27665543	*KRTCAP3*	cg11618577	< 0.001	< 0.001
gckr_rs780094	2	27665638	*KRTCAP3*	cg20102877	< 0.001	< 0.001
gckr_rs780094	2	27665711	*KRTCAP3*	cg26034919	< 0.001	< 0.001
fto_rs8050136	3	149470375	*COMMD2*	cg18759346	< 0.001	< 0.001
fto_rs8050136	3	149471298	*COMMD2*	cg01143510	0.001	0.025
kcnq1_rs163184	3	183167772	*LINC00888*	cg19415736	< 0.001	< 0.001
slc30a8_proxy_rs3802177	3	69160779	*LMOD3*	cg05798911	< 0.001	< 0.001
igf2bp2_rs4402960	8	41639488	*NKX6-3*	cg19442889	< 0.001	< 0.001
pparg_rs1801282	10	23384686	*MSRB2*	cg12046168	< 0.001	< 0.001
kcnj11_proxy_rs2074314	11	17408437	*KCNJ11*	cg03864215	< 0.001	< 0.001
kcnj11_proxy_rs2074314	11	17409602	*KCNJ11*	cg15432903	< 0.001	< 0.001
kcnj11_proxy_rs2074314	11	17415270	*KCNJ11*	cg09575421	< 0.001	< 0.001
kcnj11_rs5215	11	17408437	*KCNJ11*	cg038642151	< 0.001	< 0.001
kcnj11_rs5215	11	17409602	*KCNJ11*	cg154329031	< 0.001	< 0.001
kcnj11_rs5215	11	17415270	*KCNJ11*	cg095754211	< 0.001	0.008
kcnj11_rs5219	11	17409602	*KCNJ11*	cg154329032	< 0.001	< 0.001
cdkal1_proxy_rs4712526	17	25621092	*WSB1*	cg19383689	< 0.001	< 0.001
fto_proxy_rs12149832	17	40947843	*COA3*	cg08380385	< 0.001	< 0.001

Supplementary Table 5 | CpG relationships to single nucleotide polymorphisms in 8179 genes identified using the Immunochip

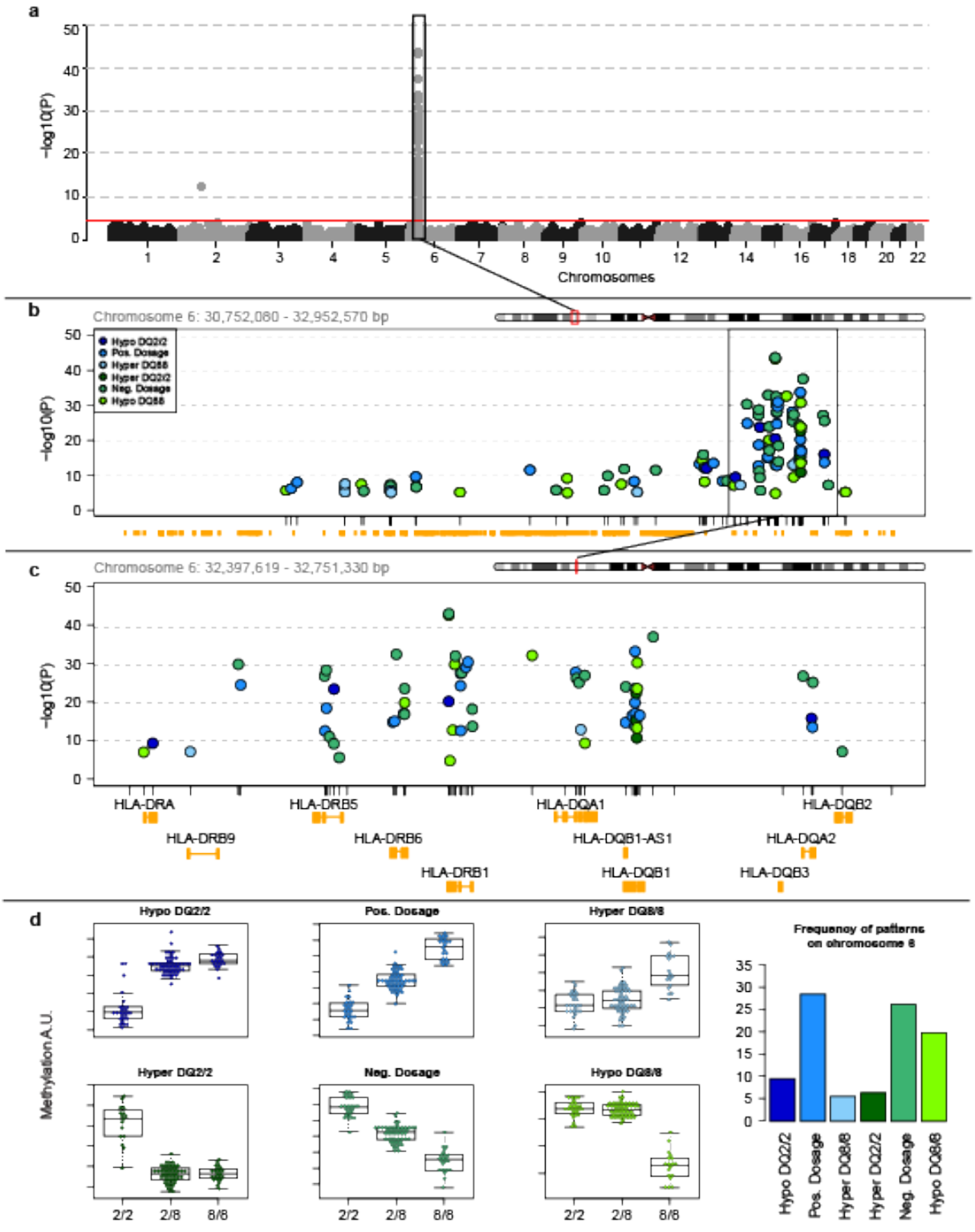
The association p-value of SNPs within a gene is given for each CpG site within the gene. The p-value originates from an ANOVA model and is not corrected for multiple testing.

Excel file: YJAUT 2017 488R1 Kindt_et_al_Supplementary_Table 5.xlsx

Supplementary Figures

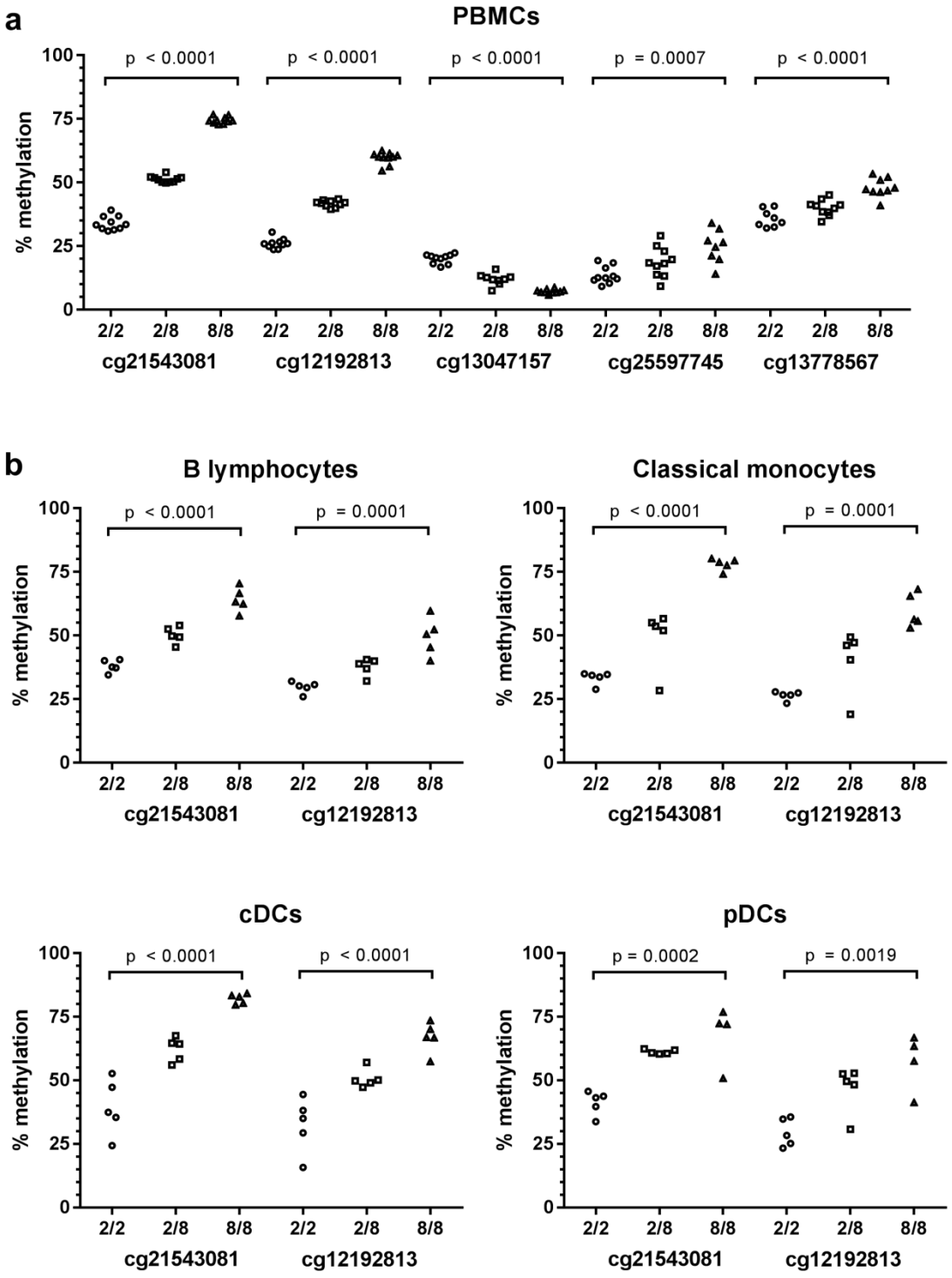
	Neonates 0 years	Toddlers < 1 year	Adolescents 6-18 years	Adults 39-78 years
Genomics	*BABYDIET* n = 109			*KORA* n = 86
Methylation Array	**Cord Blood** *BABYDIET* n = 100	**Venous Blood** *BABYDIAB* n = 36		**Venous Blood** *KORA* n = 86
Pyrosequencing		**Venous Blood** *BABYDIAB* n = 36	**pDCs, cDCs, Monocytes, B cells** *TeenDiab* n = 15	
Transcripts		**PBMCs** *BABYDIET* n = 87		
Proteins	**Monocytes** *ImmunDiabRisk* n = 26		**pDCs, cDCs, Monocytes, B cells** *TeenDiab* n = 15	

Supplementary Figure 1 | Sample and data sources. cDC, conventional dendritic cells; PBMCs, peripheral blood mononuclear cells; pDCs, plasmacytoid dendritic cells.



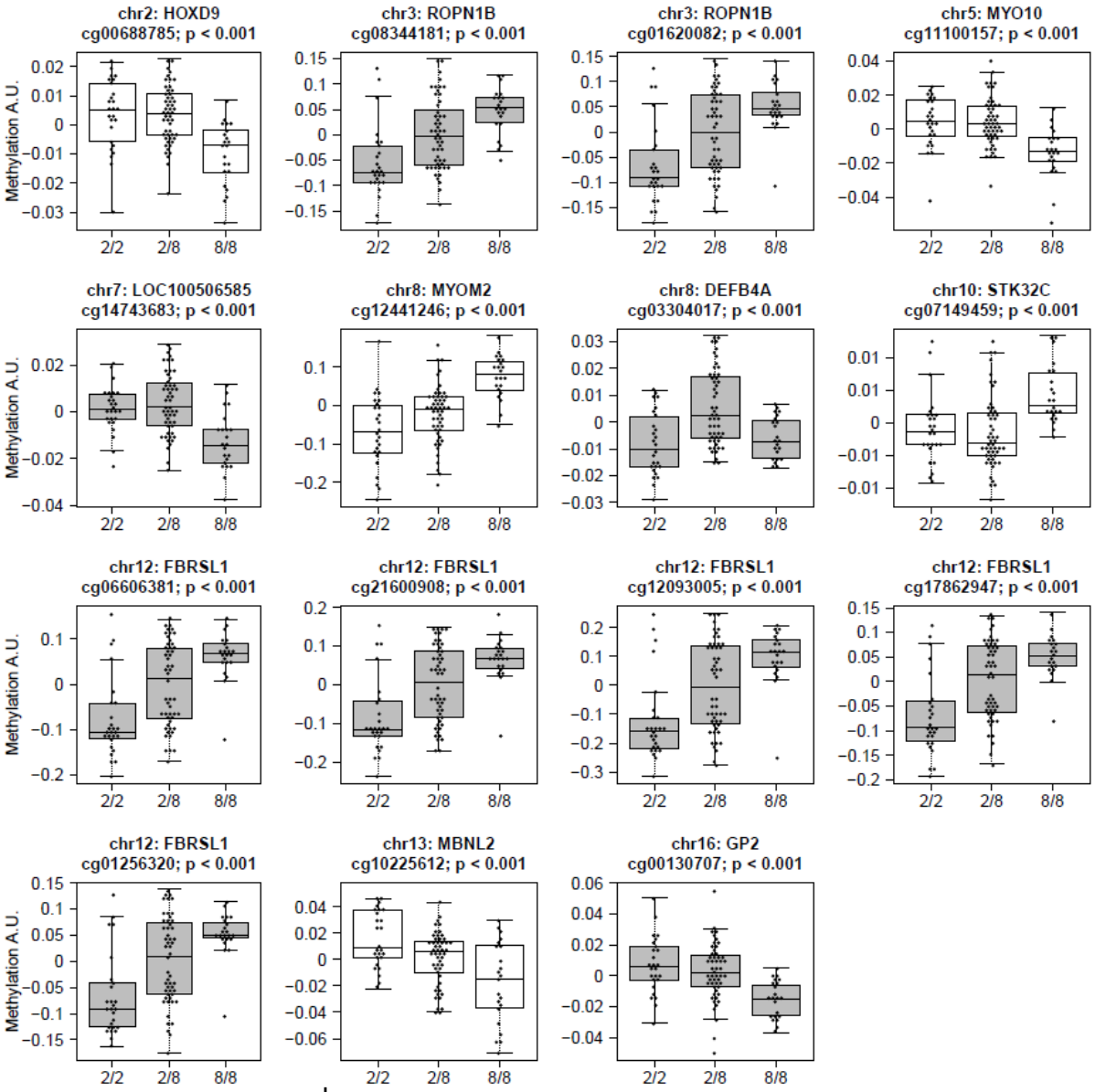
Supplementary Figure 2 | Results of the analysis of probes with a SNP within 10 base pairs in the 3'UTR. (a) Manhattan plot of differential methylation signal of CpG sites with a SNP in the probe. The maximum -log10 P value was on chromosome 6. Significant signals were also detected on other chromosomes. (b,c) Higher magnification views of the boxed

region of chromosome 6 in panel a. Genes are shown in orange. (d) Illustration of the six representative patterns of methylation differences (left panels) and the frequencies of these patterns (right panel) among the significant sites on chromosome 6. The plots in the left panels show methylation in arbitrary units (A.U.) for each neonate categorized into one of the three HLA-DR-DQ genotypes and as box plots. Hyper, hypermethylated; hypo, hypomethylated; pos, positive; neg, negative. In boxplots: centers are median, error bars are 95% confidence intervals.



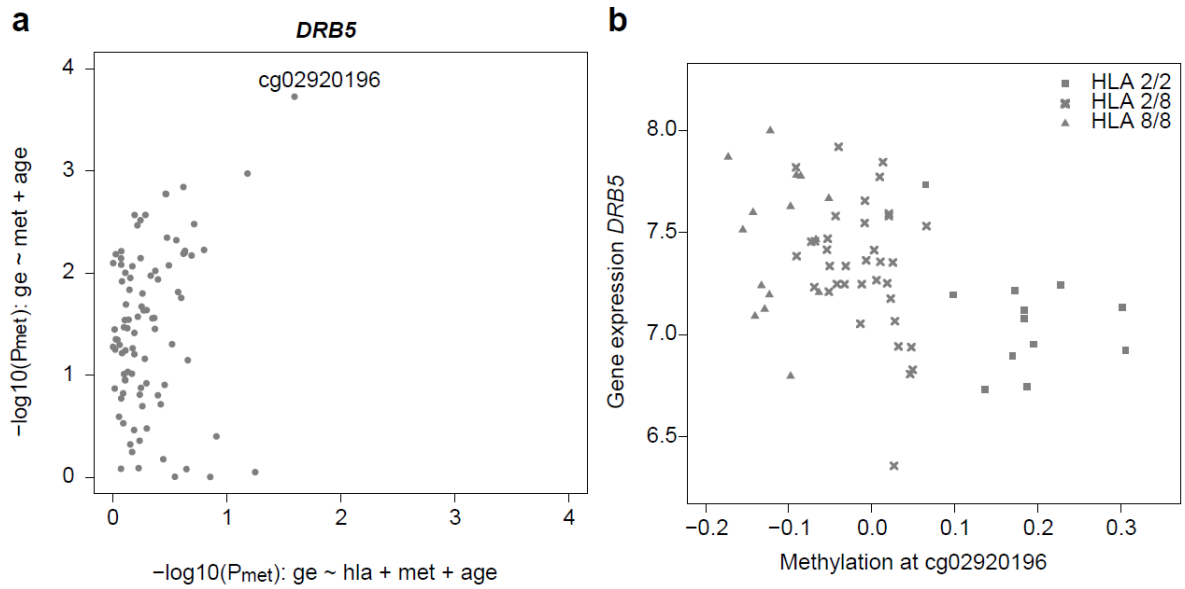
Supplementary Figure 3 | Confirmation of the HLA-DR-DQ genotype-associated methylation differences at the CpG sites cg21543081, cg12192813, cg13047457, cg25597745, and cg13778567 using pyrosequencing. (**a**) Cord blood-derived peripheral blood mononuclear cells (PBMCs) (10 neonates per genotype, 5 probes) and (**b**) B

lymphocytes, classical monocytes, conventional dendritic cells (cDCs), and plasmacytoid dendritic cells (pDCs) (5 individuals per genotype, 2 probes). In each case, the direction of the methylation change corresponded to that observed in the array generated data. For each cell type, the methylation was greater in cells from individuals with the HLA-DR4-DQ8/DR4-DQ8 (8/8) genotype than in individuals with the HLA-DR3-DQ2/DR3-DQ2 (2/2) genotype.

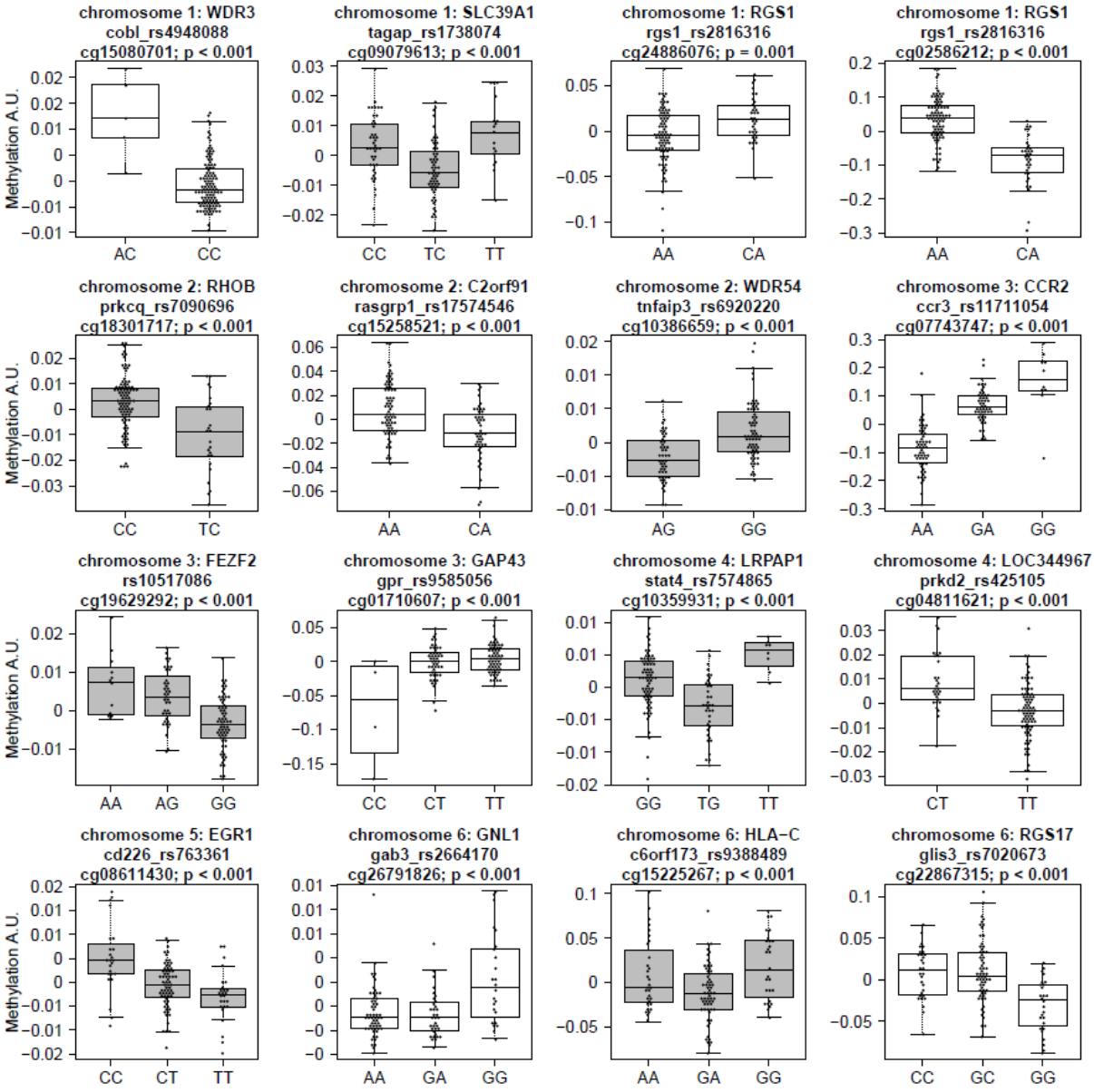


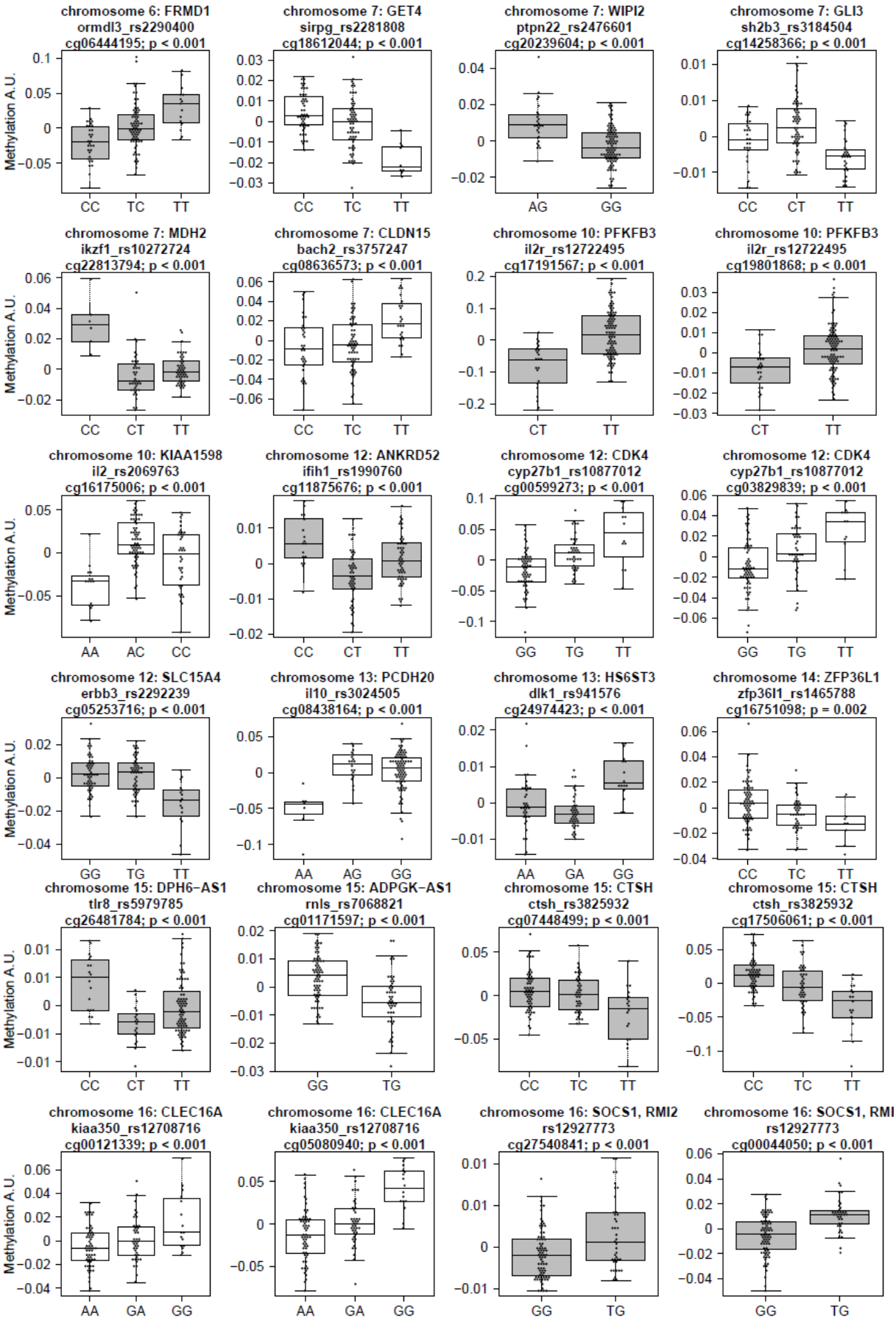
Supplementary Figure 4 | Differentially methylated CpG sites outside chromosome 6 in relation to the HLA-DR-DQ genotype. The plots show methylation in arbitrary units (A.U.) for each neonate with the HLA-DR3-DQ2/DR3-DQ2 (2/2; n = 26), HLA-DR3-DQ2/DR4-

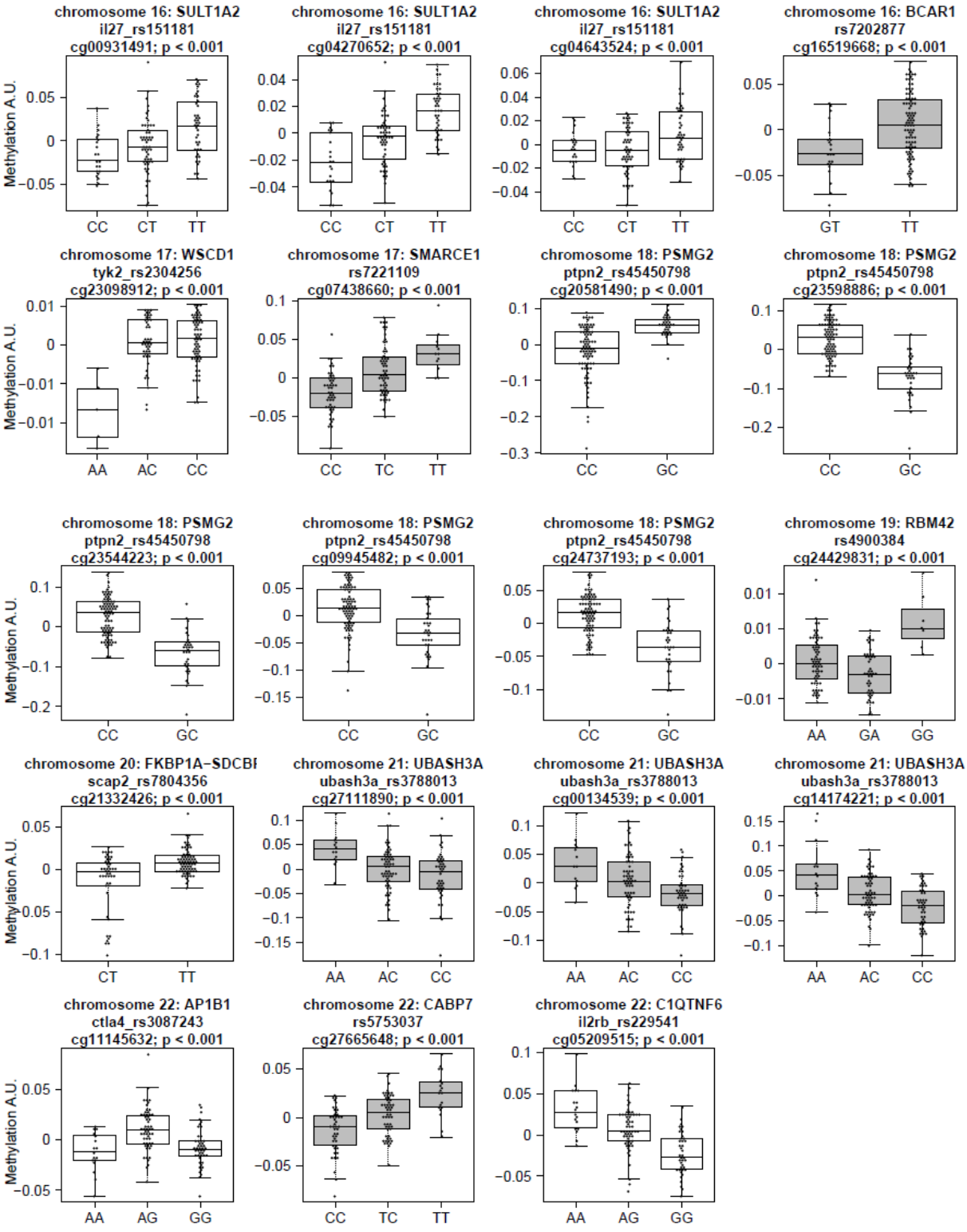
DQ8 (2/8; n = 52), or HLA-DR4-DQ8/DR4-DQ8 (8/8; n=22) genotype with box plots. The open and shaded boxes represent different genes. A.U.=Arbitrary Units.



Supplementary Figure 5 | Associations between genotype, gene expression, and methylation levels. (**a**) P-values for methylation in two models of *DRB5* transcript expression. The two models assessed gene expression levels with the inclusion (x-axis) or exclusion (y-axis) of the HLA-DR-DQ genotype. The p-values are less significant on the x-axis, which suggests that the genotype is a greater driver than the methylation state on gene expression. (**b**) Correlation between gene expression of *DRB5* and methylation at the most significant site (cg02920196) for HLA-DR3-DQ2/DR3-DQ2 (2/2; closed squares), DR3-DQ2/DR4-DQ8 (2/8; crosses), and DR4-DQ8/DR4-DQ8 (8/8; closed triangles). The interaction between methylation level and gene expression is driven by the genotype because children with HLA-DR3-DQ2/DR3-DQ2 have greater methylation levels and lower HLA DRB gene expression than children with HLA-DR4-DQ8/DR4-DQ8.

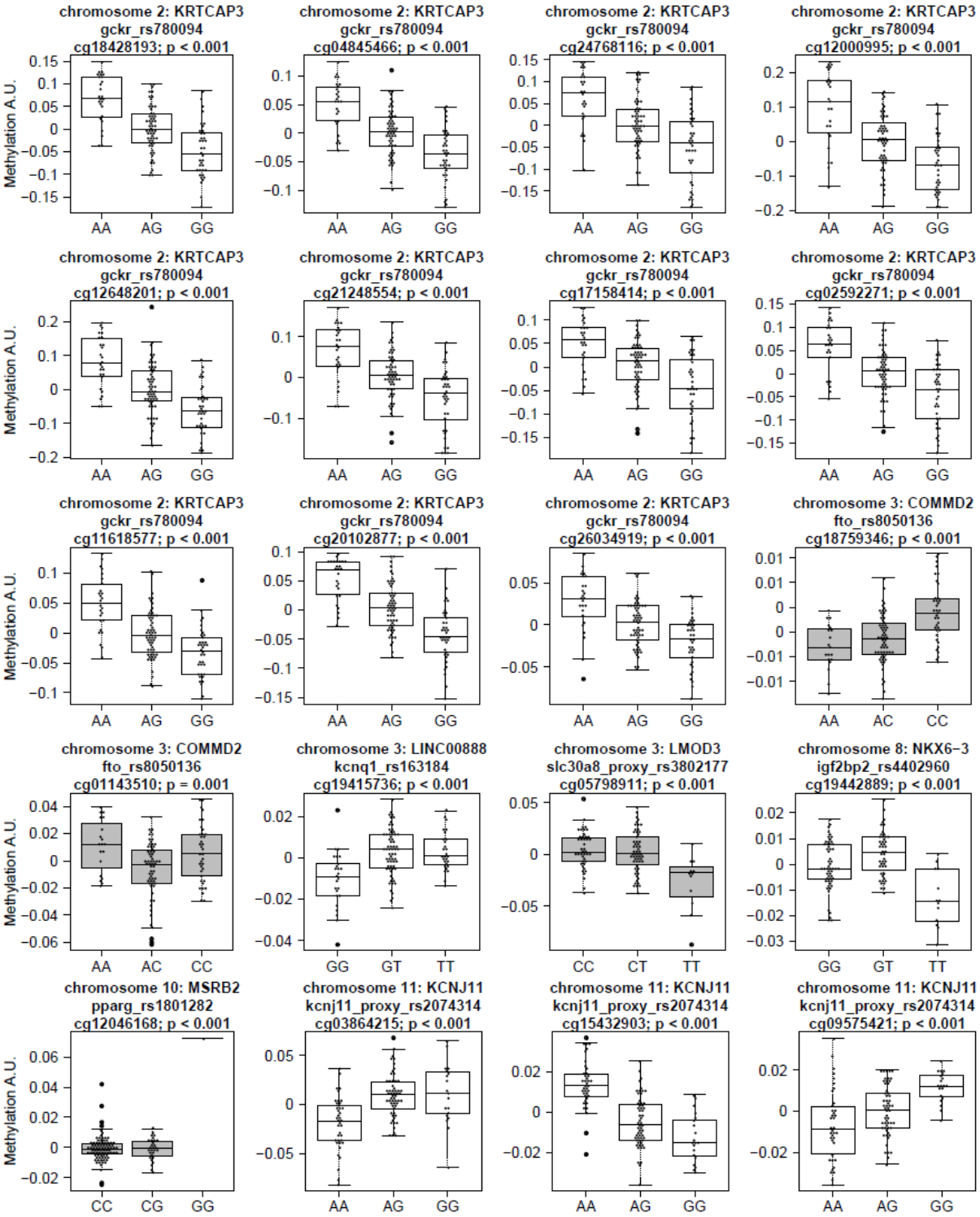

chromosome 1: WDR3
cobl_rs4948088
cg15080701; p < 0.001
chromosome 1: SLC39A1
tagap_rs1738074
cg09079613; p < 0.001
chromosome 1: RGS1
rgs1_rs2816316
cg24886076; p = 0.001
chromosome 1: RGS1
rgs1_rs2816316
cg02586212; p < 0.001
chromosome 2: RHOB
prkcq_rs7090696
cg18301717; p < 0.001
chromosome 2: C2orf91
rasgrp1_rs17574546
cg15258521; p < 0.001
chromosome 2: WDR54
tnfaip3_rs6920220
cg10386659; p < 0.001
chromosome 3: CCR2
ccr3_rs11711054
cg07743747; p < 0.001
chromosome 3: FEZF2
rs10517086
cg19629292; p < 0.001
chromosome 3: GAP43
gpr_rs9585056
cg01710607; p < 0.001
chromosome 4: LRPAP1
stat4_rs7574865
cg10359931; p < 0.001
chromosome 4: LOC344967
prkd2_rs425105
cg04811621; p < 0.001
chromosome 5: EGR1
cd226_rs763361
cg08611430; p < 0.001
chromosome 6: GNL1
gab3_rs2664170
cg26791826; p < 0.001
chromosome 6: HLA−C
c6orf173_rs9388489
cg15225267; p < 0.001
chromosome 6: RGS17
glis3_rs7020673
cg22867315; p < 0.001
Methylation A.U.

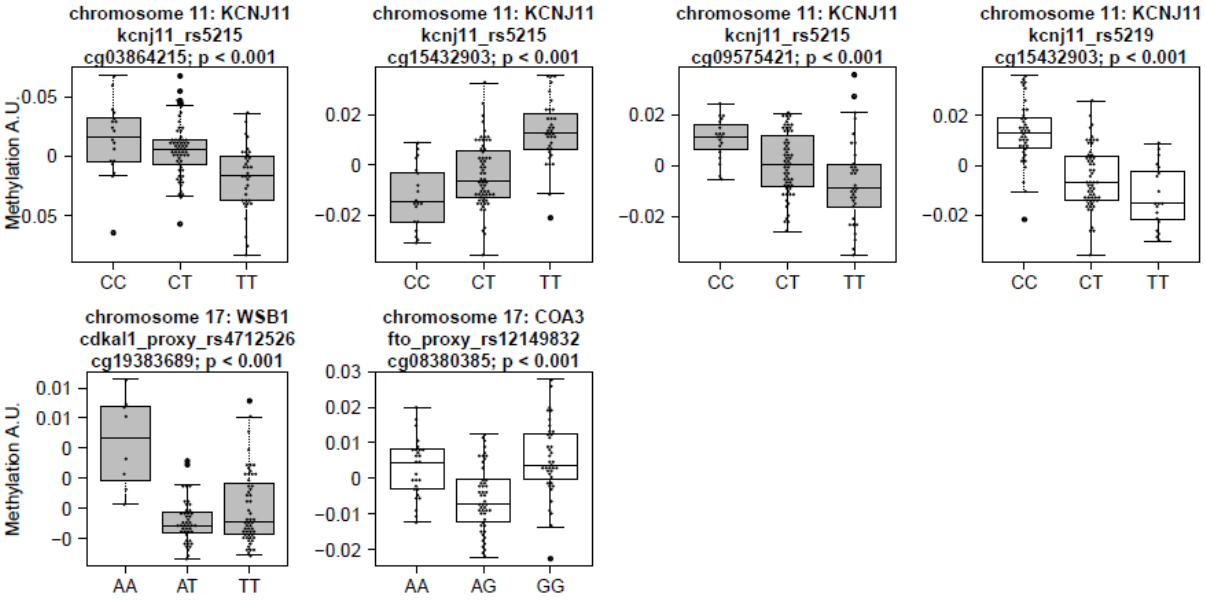

chromosome 6: FRMD1
ormdl3_rs2290400
cg06444195; p < 0.001
chromosome 7: GET4
sirpg_rs2281808
cg18612044; p < 0.001
chromosome 7: WIPI2
ptpn22_rs2476601
cg20239604; p < 0.001
chromosome 7: GLI3
sh2b3_rs3184504
cg14258366; p < 0.001
chromosome 7: MDH2
ikzf1_rs10272724
cg22813794; p < 0.001
chromosome 7: CLDN15
bach2_rs3757247
cg08636573; p < 0.001
chromosome 10: PFKFB3
il2r_rs12722495
cg17191567; p < 0.001
chromosome 10: PFKFB3
il2r_rs12722495
cg19801868; p < 0.001
chromosome 10: KIAA1598
il2_rs2069763
cg16175006; p < 0.001
chromosome 12: ANKRD52
ifih1_rs1990760
cg11875676; p < 0.001
chromosome 12: CDK4
cyp27b1_rs10877012
cg00599273; p < 0.001
chromosome 12: CDK4
cyp27b1_rs10877012
cg03829839; p < 0.001
chromosome 12: SLC15A4
erbb3_rs2292239
cg05253716; p < 0.001
chromosome 13: PCDH20
il10_rs3024505
cg08438164; p < 0.001
chromosome 13: HS6ST3
dlk1_rs941576
cg24974423; p < 0.001
chromosome 14: ZFP36L1
zfp36l1_rs1465788
cg16751098; p = 0.002
chromosome 15: DPH6-AS1
tlr8_rs5979785
cg26481784; p < 0.001
chromosome 15: ADPGK-AS1
rnls_rs7068821
cg01171597; p < 0.001
chromosome 15: CTSH
ctsh_rs3825932
cg07448499; p < 0.001
chromosome 15: CTSH
ctsh_rs3825932
cg17506061; p < 0.001
chromosome 16: CLEC16A
kiaa350_rs12708716
cg00121339; p < 0.001
chromosome 16: CLEC16A
kiaa350_rs12708716
cg05080940; p < 0.001
chromosome 16: SOCS1, RMI2
rs12927773
cg27540841; p < 0.001
chromosome 16: SOCS1, RMI
rs12927773
cg00044050; p < 0.001
Methylation A.U.



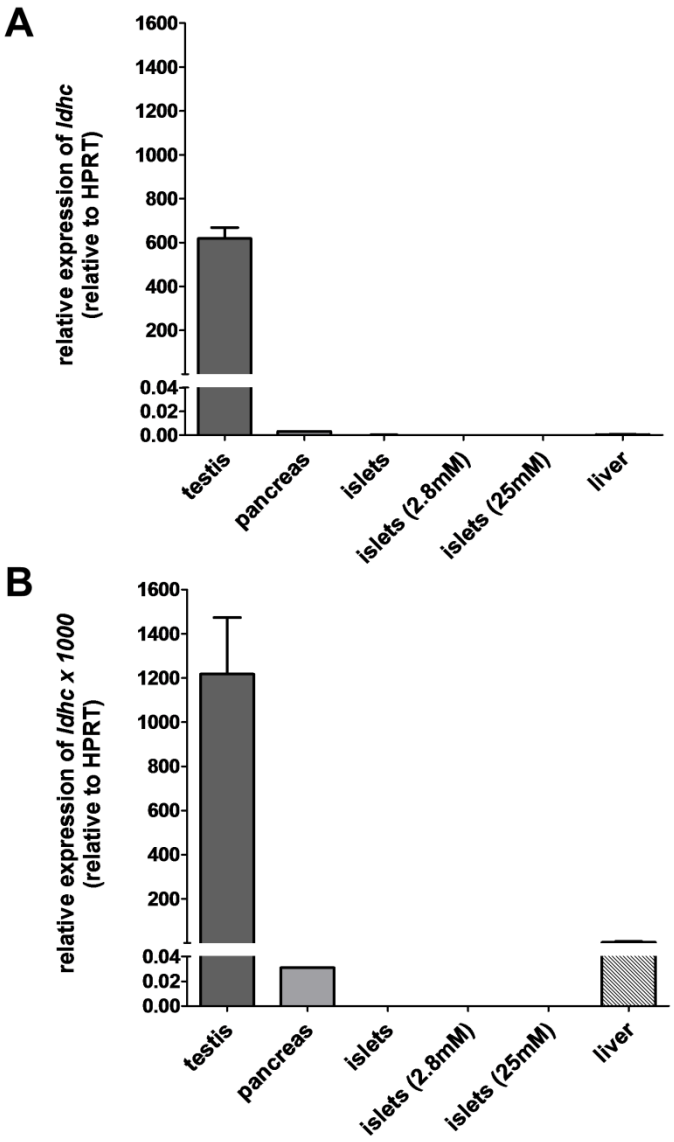
Supplementary Figure 6 | Methylation levels of 59 CpG sites associated with 45 single nucleotide polymorphisms in susceptibility genes for type 1 diabetes. The plots show methylation in arbitrary units (A.U.) for each neonate categorized according to the genotype

with box plots. The open and filled boxes represent different genes, utilized to improve readability.





Supplementary Figure 7 | Differential methylation at 26 CpG sites associated with 10 single nucleotide polymorphisms in nine type 2 diabetes susceptibility genes. The plots show methylation in arbitrary units (A.U.) for each neonate categorized according to the genotype with box plots. The open and filled boxes represent different genes, utilized to improve readability.



Supplementary Figure 8 | LDHC-expression in mouse tissues. mRNA expression of LDHC was determined by quantitative RT-PCR employing whole tissues or islets harvested from 13/14-week-old C57BL/6J mice (A, n = 2) or 12/18 week old male NOD mice (B, n = 4). HPRT was used for normalization. Data from C57BL/6J and NOD mice were generated in separate experiments.