Charles University

Third Faculty of Medicine

Doctoral Thesis

Regulation of HLA class II genes expression

Regulace exprese genů HLA II. třídy

Supervisor: Prof. MUDr. Marie Černá, DrSc.

Declaration

Hereby I declare that this thesis was written by me and is based on experiments performed during my PhD studies. I declare that I correctly cited all resources and literature used and that this work was not used to obtain another or the same degree.

I agree with the permanent storage of the electronic version of my work in the database of the inter-university project Theses.cz for the purpose of continuous control of the similarities between the qualification theses.

Prague, 12.4.2018 Marta Zajacová

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem řádně uvedla a citovala všechny použité prameny a literaturu. Současně prohlašuji, že práce nebyla využita k získání jiného nebo stejného titulu.

Souhlasím s trvalým uložením elektronické verze mé práce v databázi systému meziuniverzitního projektu Theses.cz za účelem soustavné kontroly podobnosti kvalifikačních prací.

V Praze, 12.4.2018 Marta Zajacová

Identification record

ZAJACOVÁ, Marta. *Regulace exprese genů HLA II. třídy [Regulation of HLA class II genes expression]*. Praha, 2018. 155 stran, 5 příloh. Dizertační práce. Univerzita Karlova, Praha, 3. lékařská fakulta, Ústav lékařské genetiky. 3. LF UK. Vedoucí dizertační práce: Prof. MUDr. Marie Černá, DrSc.

Klíčová slova: HLA, polymorfismus, alela, exprese, epigenetika

Keywords: HLA, polymorphism, allele, expression, epigenetics

Acknowledgement

I would like to thank my supervisor, Prof. Marie Černá, MD, PhD for drawing my focus to the exciting topic of HLA class II gene expression and for providing the laboratory facilities and support for this work. I appreciate her patience and her giving me a free hand to follow the ideas I found interesting.

Sincere gratitude to Anna Kotrbova-Kozak, PhD for her broad support with the laboratory work, methodology and during writing this thesis. She generously shared her knowledge with me and her perseverance and positive approach to all challenges has been an inspiration to me.

Big thanks to Michaela Říhová, MD for her relentless help with collecting samples.

I want to thank Pavlína Čejková, PhD and Katarína Kološtová, PhD for always finding time to answer my questions and for their insightful notes regarding the RT-qPCR methodology.

Thanks to Anna Cibuľová, MSc, for sharing the excitement for research and for making the days in lab fun during beginning of our PhD studies.

I would like to thank all people who donated their blood for this study and thus made it possible.

I am grateful to my family for their support and understanding during writing this thesis.

Finally, I want to thank Rahul for his patience and project management skills, and for challenging me to always seek for the root cause and never stop asking why.

Contents

List of abbreviations

5meC	5-methylcytosine
A, C, T, G	Adenine, Cytosine, Thymine, Guanine,
APC	Antigen-presenting cell
bp	Base pair
CD	Cluster of differentiation; CD followed by a numeric code denotes a cell membrane molecules according to the international classification
CD4, CD14, CD16, CD19, CD20, CD25, CD122	Cluster of differentiation 4, 14, 16, 19, 20, 25, 122
cDNA	Complementary DNA
CIITA	Class II transactivator
CLIP	Class II-associated invariant chain
CREB	cAMP responsive element binding protein
CTCF	CCCTC-binding factor
DC	Dendritic cell
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DNMT1	DNA methyltransferase 1
DNMT3a, DNMT3b, DNMT3L	DNA methyltransferase 3a, 3b, 3L
E $_{\text{assay name}}$	Efficacy of the denoted qPCR assay
FoxP3	Forkhead box P3
H3,H4	Histone 3, Histone 4
H(3,4)K#ac	Histone H3 (H4) acetylated at lysine at the denoted position
H(3,4)K#me(1,2,3)	Histone H3 (H4) mono-/di-/trimethylated at lysine at the denoted position
H3K4me1, H3K4me3	Histone H3 mono/trimethylated at lysine 9
H3K6me3, H3K9me3	Histone H3 trimethylated at lysine 6 / lysine 9
H3K27ac, H3K27me3	Histone H3 acetylated/trimethylated at lysine 27
H4Ac	Acetylated histone H4
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HLA	Human leukocyte antigen
IFNγ	Interferon gamma
Ii	Invariant chain
IL-2, IL-4	Interleukin 2, Interleukin 4
MBD	Methyl-CpG-binding domain protein
MBD1, MB2, MBD3, MBD4	Methyl-CpG-binding domain protein 1, 2, 3, 4
MHC	Major histocompatibility complex
mRNA	Messenger RNA
NFY	Nuclear factor Y, trimer composed of NF-YA, NF-YB and NF-YC

PBMC	Peripheral blood mononuclear cells
p-MHC	Protein-MHC interaction
p-MHC-TCR	Protein-MHC-TCR interaction
PCR	Polymerase chain reaction
PPIA	Peptidylprolyl isomerase A (cyclophilin A), endogenous control for qPCR
QAP	*DQA1* gene promoter
RFX	Regulatory factor X, trimer composed of RFX5, RFXANK and RFXAP
RNA	Ribonucleic acid
RT-qPCR	Reverse transcription quantitative PCR
SNP	Single nuclear polymorphism
SXY	DNA region containing specific S, X and Y sequences for binding of MHC class II transcription factors
T1D	Type 1 diabetes
TCR	T cell receptor
Th1, Th2	Type 1 (Type 2) helper cells
Th17	T helper 17 cells
Treg	Regulatory T cells
TSS	Transcription start site
UTR	Untranslated region
WBC	Whole blood cells
XL1, XL4, XL8, XL9	X-like sequences 1, 4, 8, 9; sequences highly homologous to X and Y boxes of the SXY module

Summary

HLA class II genes are known to be highly polymorphic, even in the regulatory non-coding gene regions. Polymorphism in the promoter region potentially forms a strong basis for an uneven allele-specific expression. Even though it is known that the amount of HLA class II molecules on the cell surface has a significant role in shaping immune response, HLA class II expression polymorphism has not yet been thoroughly measured.

The thesis aims to shed light into allele-specific mRNA expression and promoter DNA methylation of HLA class II genes. Two studies, each addressing different aspects of the HLA class II allele expression regulation, were conducted.

Study A examines the DNA methylation of 10 *DQA1* promoter and its effect on the *DQA1* mRNA expression. DNA methylation in whole blood cells was determined with bisulfite sequencing and mRNA expression was measured using RT-qPCR. Even though inter-allelic differences in overall methylation were observed (the most methylated alleles were *DQA1*02:01* and **04:01*), the expected negative correlation between the *DQA1* promoter DNA methylation density and the allele expression was not observed. We suggest that the genetic polymorphism in the region (especially region upstream of position -400, which is almost completely methylated in all alleles) may lead to different interpretation of the same 5meCpG mark in different allelic contexts.

Study B analyses *DQA1* and *DQB1* mRNA expression in whole blood cells, B lymphocytes and monocytes using RT-qPCR. Class II transcription level is higher (with the exception of *DQB1*06* alleles) and shows lower interallelic variation in B cells compared to monocytes. The *DQB1* expression hierarchy can be generalized into *DQB1*06 > *03*,**05 > *02* pattern in monocytes and *DQB1*06 > *02*, **03*, **05* in B cells. The *DQA1* expression hierarchy is *DQA1*03 > *01 ≥ *02 > *05* in monocytes and *DQA1*03*, **05 > *01, *02* in B cells. Because of the low number of samples for certain allele-cell type combinations, only some of these results were statistically significant. CIITA isoforms are discussed as the factors that could drive the cell-type specific expression of class II alleles. Finally, we tried to relate DQ dimer expression level to the risk it carries for the development of the autoimmune diabetes.

In conclusion, this thesis confirms and describes mRNA expression polymorphism of the *HLA-DQA1* and *-DQB1* genes, and provides the expression hierarchy of these alleles in the B cells, monocytes and whole blood cells. It presents a DNA methylation profile of the *DQA1* gene promoter and shows that DNA methylation by itself is not able to explain the inter-allelic expression differences. In our simplified model of DQ dimer expression, the expression level by itself is not able to explain dimer association with T1D.

Shrnutí

Geny HLA II. třídy jsou vysoce polymorfní, a to také v regulačních nekódujících genových oblastech. Polymorfismus v promotorové oblasti vytváří silný potenciál pro rozdílnou expresi jednotlivých alel. I když je známo, že množství molekul HLA II. třídy na povrchu buněk má významnou úlohu při utváření imunitní odpovědi, polymorfismus v expresi genů HLA třídy II dosud nebyl podrobně analyzován.

Cílem této práce bylo zkoumat expresi mRNA a metylaci promotorové DNA u alel genů HLA II. třídy. Byly provedeny dvě studie, které se zabývaly různými aspekty regulace.

V **části A** byla zkoumána metylace DNA u 10 alel promotoru *DQA1* a její vliv na expresi mRNA tohoto genu. DNA metylace v plné krvi byla stanovena pomocí bisulfitového sekvenování a exprese mRNA byla měřena za použití RT-qPCR. I když byly pozorovány mezialelické rozdíly v celkové metylaci (nejvíce metylované alely byly *DQA1*02:01* a **04:01*), nebyla pozorována očekávaná negativní korelace mezi methylační hustotou DNA promotoru *DQA1* a expresí alely. Mezialelické rozdíly v metylaci jednotlivých CpG pozic prokázány nebyly, ale domníváme se, že genetický polymorfismus v oblasti (zejména oblast upstream od pozice -400, která je téměř kompletně metylovaná ve všech alelách) může vést k různým interpretacím 5meCpG na stejné pozici v různých alelických kontextech.

V **části B** byla analyzována úroveň exprese mRNA alel genů *HLA-DQA1* a *-DQB1* v plné krvi a B lymfocytech a monocytech pomocí RT-qPCR. Transkripce genů II. třídy byla vyšší (s výjimkou alel *DQB1*06*) a vykazovala nižší mezialelickou variaci v B lymfocytech ve srovnání s monocyty. Hierarchii exprese *DQB1* lze zobecnit vztahem *DQB1** 06 > **03*, **05* > **02* v monocytech a *DQB1*06* > **02*, **03*, **05* v B lymfocytech. Hierarchie exprese *DQA1* je *DQA1*03* > **01* ≥ **02* > **05* v monocytech a *DQA1 *03*, **05* > **01*, **02* v B lymfocytech. Vzhledem k nízkému počtu vzorků pro určité kombinace typu alel a buněk byly pouze některé z těchto výsledků statisticky významné. Isoformy CIITA jsou diskutovány jako faktory, které mohou být odpovědné za rozdílnou expresi alel HLA II. třídy v specifických buněčných typech. Nakonec jsme se pokusili objasnit vztah mezi expresí DQ dimerů a rizikem rozvoje autoimunitního diabetu.

Závěrem, tato práce potvrzuje a popisuje polymorfismus v expresi mRNA genů *HLA-DQA1* a *-DQB1* a stanovuje expresní hierarchii těchto alel v B lymfocytech, monocytech a plné krvi. Přináší profil metylace DNA promotoru genu *DQA1* a ukazuje, že metylace DNA není sama o sobě schopna vysvětlit mezialelické rozdíly v expresi. V našem zjednodušeném modelu není úroveň povrchové exprese DQ dimerů sama o sobě schopna vysvětlit jejich asociaci s T1D.

1 Introduction to Regulation of HLA Class II Genes and Their Role in Immunity

Variability in the protein structure and expression levels of HLA class II molecules is an important factor influencing the target and form of the immune response in different individuals within a human population. This thesis focuses on studying the differences in expression of alleles of *HLA-DQA1* and -*DQB1* genes and tries to identify the reasons and consequences of their expression variability.

The first part of the thesis introduction provides a high-level view on function of MHC (Major Histocompatibility Complex) class II molecules and on their properties important for immune response.

The second part of introduction reviews information on genetic and epigenetic regulation of expression of human MHC—i.e. HLA (Human Leukocyte Antigen)—class II genes; the effect of genetic and epigenetic differences on expression of individual allelic variants, and how these differences translate into functional differences between alleles.

1.1 HLA class II properties and function

1.1.1 HLA class II structure and isotypes

HLA class II molecules are dimers composed of non-covalently associated transmembrane chains α and β. Extracellular parts of protein chains together make up an antigen-binding site and short cytoplasmic tails serve for outside-in signalling (that can, for example, cause apoptosis in the antigen-presenting cell and thus limit the extent of an immune response) (1). Three isotypes of class II molecules can form on a cell surface, HLA-DR (formed by molecules encoded by the *DRA* and *DRB1, 3, 4* and *DRB5* genes), HLA-DQ (*DQA1, DQB1*) and *HLA-DP* (*DPA1, DQB1*) (2). HLA-DM and -DO non-classical isotypes are not expressed on the cell surface, and instead they assist with intracellular antigen loading onto classical class II molecules (3).

1.1.2 HLA class II polymorphism

In accordance with their role in the organism defense where an ability to react to a wide spectrum of pathogens plays an important role, HLA class II genes exhibit an exceptionally high level of polymorphism. Tens to thousands variants are known for each of classical class II genes, with the exception of *DRA*, which is basically non-polymorphic (7 alleles coding for 2 protein variants). Variation of genes encoding for DM and DO molecules is also very limited (**Table 1**) (4).

Table 1. Number of HLA class II genes variants known to the date (# of alleles/# of protein variants). Only expressed variants are listed. According to http://hla.alleles.org/nomenclature/stats.html. N/A: not available.

DRA	*DRB1*	*DQA1*	*DQB1*	DPA1	DPB1	DMA	DMB	DOA	DOB	DRB3	DRB4	DRB5	DQA2	DQB2
7/2	1883/ 1369	69/33	911/6 22	43/21	644/5 29	7/4	13/7	12/3	13/5	77/63	24/16	26/22	N/A	N/A

Genetic variance affects all parts of the gene: in the coding region it influences the spectrum of presented peptides, in the non-coding regions of the gene it has a potential to affect efficiency of an allele transcription. Together, variation on both levels creates unique combination of properties for each allele. Because of a strong linkage disequilibrium in the region, the "HLA functional unit" can be often extended to the whole haplotype, in a particular population.

1.1.3 HLA class II properties – pMHC-TCR interaction quality and quantity

Function of MHC class II molecules is to bind fragments of primarily exogenous (but also endogenous (5)) proteins and expose them on the cell surface where they can be recognized by CD4+ T cells. Specific recognition of MHC-peptide complexes by the antigen-specific T cell receptors (TCR) enables an adaptive immune response targeted against the eliciting antigen as well as establishing and maintaining the self-tolerance (6, 7).

The outcome of an encounter between an antigen presenting cell (APC) and T cell (whether the T cell will respond to the antigen and how the response will look like) is determined by several factors, like cytokines in the environment, presence of the costimulatory molecules on the APC and most importantly, by a signal delivered through TCR after interaction with a cognate pMHC complex. Both **quality (affinity) of TCR-pMHC interaction** and a **cumulative quantity of the TCR signal** are important. These properties can, for example, affect T cell gene expression upon TCR ligation. A subset of genes responds specifically to the pMHC-TCR affinity, some (small) subset to the pMHC dose, and finally, in a subset of genes, pMHC-TCR affinity and dose are interchangeable: in this case, high pMHC dose can compensate for the low pMHC-TCR affinity, for example, in the induction of proliferation *in vivo* or in the induction of negative regulators of the immune response, as is CTLA4 (8).

As the purpose of the chapter is rather to present principles than explicit details of the interaction and as the interesting concepts arose from human as well as mice studies on both MHC I and MHC II molecules (and the results rather complement than out-rule each other), the chapter includes results from all these types of experiments. For the same reason, I will use the term 'MHC' instead of 'HLA' here. Even though the cell/organism response to the signal quality and quantity may differ, in some experimental setups it is not possible to separate effects of these properties, in these place we discuss them together.

1.1.3.1 pMHC-TCR affinity

A major (but not the only) function of the pMHC-TCR interaction affinity is to promote IL-2 secretion from naive cells (7). In addition, cells carrying more affine pMHC ligands form longer lasting contacts with dendritic cells (DCs), express higher levels of the IL-2 receptor chains CD25 and CD122 and produce more IL-2 upon restimulation (8).

Affinity of the pMHC-TCR interaction is relatively low, and bonds with halftimes of several seconds are sufficient for generating an immune response (9). Optimal *in vivo* T cell effector responses (measured as production of IL-2 and IFNγ) are to ligands with intermediate TCR-pMHC half-lives (9).

The pMHC-TCR interaction is not only low affinity but also degenerate: that is, many TCRs recognize the same pMHC complex and many pMHC complexes are recognized by the same TCR (reviewed by (10)). This TCR promiscuity is enabled by flexibility of TCR conformation and by the fact that TCR often focuses on interaction with only few residues of the presented peptide. As a result, T cells are extremely cross-reactive: individual T cell clones are able to recognize more than a 10^6 different peptides in the context of a single MHC molecule (10). This brings a potential for autoimmunity: T cells sensitivity towards an antigen can increase up to 50-fold during their development ((10) and next chapter) and T cells activated by peptides from pathogens can cross-recognize and get activated by self-derived peptides with an affinity for TCR much lower that that of the original pathogen-derived peptide (10).

1.1.3.2 pMHC-TCR affinity and dose

A pMHC-TCR interaction of a certain minimal strength (determined by both, interaction quality and quantity) is needed for a T cell survival after an antigen encounter in periphery. T cells stimulated with a weaker signal are able to secrete IL-2 (and thus can provide support to the T cells with more specific TCRs), but fail to produce lineage-specific cytokines (IL-4 and IFNγ), commit to Th1 or Th2 lineage and do not persist in the organism (7).

A comprehensive study of a T cell differentiation into IFNγ-producing effector cells in response to a range of pMHC concentrations (10^6 range) and pMHC-TCR affinities (10^6 range) showed that for all affinities, the cytokine response to increasing pMHC concentration is bell-shaped with the reduced cytokine production at low and high pMHC concentrations (11). The pMHC-TCR affinity has no effect amplitude of the response, but is inversely correlated to the pMHC concentration needed to reach the maximal response (11). Similar observations were done in study on the peripheral induction of FoxP3 regulatory T cells: for each peptide, there is an optimal concentration (peak of the bell-shaped dose-response) for the FoxP3 induction, which is inversely correlated with the pMHC-TCR affinity (12). However, in this case, the affinity and quantity of the signal were not interchangeable: the most efficient stimulus for induction of persisting FoxP3+ cells is a low dose of a strong agonist—a stimulation with high amount of low-affinity

pMHC ligands induces comparable amount of the regulatory cells, but they do not survive (12). The presence of TGFβ extends the TCR signal strength range over which a FoxP3 induction occurs (12).

In thymus, an antigen presentation plays a crucial role in determining the T cell developmental fate and in shaping the T cell repertoire. T cells in thymus undergo two types of selection based on the signals received through TCR: positive selection (in the thymic cortex) ensures that only the cells that are able to recognize peptide-MHC complexes survive, negative selection (mostly in thymic medulla) ensures deletion of the T cells that recognize MHC complexes with self-peptides with high affinity (6). Negative selection executors in the medulla are thymic epithelial cells and bone marrow-derived APCs that uptake antigen in thymus (indirect presentation) and efficiently delete self-reactive cells (13). In addition, T cells that receive TCR signals that are strong, but not strong enough for T cell deletion, develop into natural FoxP3+ regulatory T cells (Tregs) (6). It is also possible that either a low T cell clone frequency or a low dose of pMHC, perhaps presented by a small number of APCs, is needed for an effective Treg cell selection (14).

1.1.3.3 pMHC-TCR dose

Several sets of experiments were performed that explored an effect of varying concentration of a particular pMHC complex (i.e., at the constant pMHC ligand affinity), often by simply altering the dose of the antigenic peptide. Variation of the antigen dose and the number of MHC II molecules on the APC surface leads to preferential development of cells with different effector phenotypes due to influence on the cytokines production and Th1–Th2 equilibrium of effector cells. The effect of a particular antigen dose is not consistent in all studies, however, generally it seems that both very low and very high level favors production of IL-4 and Th2 responses, and intermediate levels favor IFNγ and Th1 development (15–19). High antigen dose is also needed for development of inflammatory Th17 cells. If naïve T cells encounter a low-dose antigen in different cytokine environments, Th17 effectors do not differentiate at antigen doses that are sufficient to drive differentiation of Th1 and Th2 effectors (20). On the other hand, very high antigen doses can drive development of Th17 cells even in the absence of exogenous polarizing cytokines (19).

The amount of antigen present during the T cell development from naïve into effector cell also sets the **avidity setpoint**—the threshold of antigen concentration to which the cell is able to respond on a secondary encounter. The antigen dose to which the cells are exposed during primary encounter inversely correlates with the cells sensitivity (avidity) towards the antigen. The avidity setpoint is further affected by the cytokine environment, with Th1 differentiation conditions leading to effectors with the highest avidity, followed by Th17 and finally Th2 conditions (20).

Thus, differences in pMHC amount on the cell surface cause preferential development of different effector phenotypes. On the other hand, for each effector phenotype, there is an optimal **level** of TCR stimulation that directs cells towards differentiation into the phenotype.

Low MHCII expression was associated with autoimmunity through a less effective presentation of autoantigens to the protective regulatory T lymphocytes (21), or inefficient deletion of autoreactive cells in thymus through the similar mechanism as it happens with low-stability pMHC complexes (22).

Thus, pMHC amount on the cell surface is an important regulator of the immune response. Final pMHC amount is a result of multiple factors: intrinsic stability of the MHC αβ dimer, stability of the peptide-MHC complex, production rate of the of α and β chains, antigen dose, and finally pMHC localization within a cell.

1.1.3.3.1 Protein-MHC complex stability

The p-MHC complex stability and the resulting halftime of the complex on the cell surface correlates well with the **peptide immunogenicity** and is a better predictor than a closely related characteristics, affinity of p-MHC bond (23). The reason for the different effect of these two measures lies in the mechanism of the antigen presentation and immune response generation: once the MHC II loads a peptide in the dedicated cell compartment (MIIC—MHC class II compartment) and is transported to the cell surface, it has no chance to re-load the peptide. In addition, the p-MHC complex must last long enough to be, in a case of the pathogen encounter, transported to the lymph node and stimulate T cells. Therefore, p-MHC bond stability is of utmost importance (24). It seems that a minimal productive complex halftime is 1 hour, as under this threshold, the peptide does not elicit immune reaction under any tested circumstances (23–25). In addition, the complex must be stable within a pH ranging from the acidic pH=5.5 of the endosomal compartments to neutral pH of the cell surface (25).

Higher p-MHC complex stability leads to higher **T cell activation** (26, 27) and weak (unstable) bonds cause reduced p-MHC half-time on the cell surface and **T cell anergy** (28) or **inability of T cells to survive** (7). p-MHC stability also influences other aspects of the immune activation: it is inversely correlated with an optimal peptide concentration (i.e., the peptide concentration needed for the optimal T cell stimulation). Furthermore, it seems that high affinity p-MHC bonds lead to more complex T cell **repertoire** (29).

The p-MHC stability correlates with a number of complexes on the cell surface (30) and changes in the stability may be able to cause difference in a cell-surface numbers and thus affect the T cell activation.

1.1.3.3.2 MHCII αβ dimer stability

One of the factors affecting the p-MHC complex stability is an intrinsic stability of the MHCII dimer, which is to a high extent determined by an ability of the α and β chains to form hydrogen bonds with each other (31). Stability of MHCII dimers differs in between haplotypes and has consequences for the repertoire and amount of peptides presented by a particular dimer.

The intrinsically **unstable** αβ dimers are expected to present more narrow repertoire of peptides, as only class II complexes with high-affinity peptides will be stable enough to last on the cell surface (Note:

the authors talk about the high-affinity peptides, but in the context of previous chapter we should probably understand it as peptides that form stable complexes with MHCII). Despite unstability, these dimers can be present on the cell surface in high numbers, depending on the availability of high-affinity peptides and accessory molecules (32).

Stable αβ dimers, on the other hand, are expected to present the peptides with wider affinity ranges and their expression on the cell surface is more constant and not limited by a dose of high-affinity peptide (32).

Inability to form stable pMHC complex with certain self-epitopes in thymus may enable escape of self-reactive T cells to the periphery and the instability of αβ dimer was repeatedly associated with a susceptibility to autoimmune diseases (32, 33). For example, HLA-DQ dimers formed by the autoimmunity-prone DR3-*DQA1*05:01-DQB1*02:01* and DR4-*DQA1*03:01-DQB1*03:02* haplotypes (34) are amongs the least stable ones (32), and the T1D risk correlates with the DQ dimer instability for majority of DQ haplotypes, including the most predisposing and protective ones (32). Thus, MHCII dimer unstability may be a risk factor for a subset of autoimmune diseases.

In addition to *cis* dimers, HLA-heterozygotes can form ***trans*-dimers** by assembly of α and β chains from the two haplotypes (32, 35), with a unique antigen binding abilities (36). Some of these *trans*-dimers are stable enough to be expressed on the cell surface (32, 35) and as a result, certain heterozygotes may utilize up to 4 different MHCII dimers of a particular HLA isotype. To add to the complexity, the balance in the formation of *trans*-dimers may be affected by a widespread phenomenon of **random monoallelic expression**, where, at a given point of time, as a result of stochastic processes a cell expresses only a single allele out of a pair (37). This phenomenon was observed in HLA genes as well, where 77% of Raji B cell line cells expressed *HLA-DRB1* from both alleles, and only 23% expressed *DQA1* from both alleles (38). This means that even in heterozygotes that are able to form stable *trans*-dimers, a fraction of cells may express only 2, or even a single type of DQ molecules.

α and β chains of different HLA isotypes can assemble into functional heterodimers as well, e.g., DRα can pair with DQ5 and DQ6 (but not with DQ7,8,9) β chains (35). Formation of DRα-DQβ dimers was observed in B cell lines and authors suggests that mixed isotype pairing may further enhance ability to present various peptides and may be reason why certain diseases associate with haplotype rather than with genotype (39). Mixed isotype dimers are able to effectively present antigen, as was shown for the *DQA1*01:02-DRB1**15:01 dimer that was able to stimulate the development of the autoimmune disease in the murine model of multiple sclerosis (40).

The frequency of the mixed-isotype formation is not known, but it is plausible that this is not the preferred way of pairing and occurs only after favourable interactions within an isotype are all used (39).

1.1.3.3.3 Presence of HLA-DM, HLA-DO and CLIP – editing the peptide repertoire

During αβ dimer assembly in the ER, the peptide-binding site is protected from premature binding of proteins by an invariant chain (Ii) and after the Ii cleavage by its proteolytic fragment, class II-associated Ii peptide (CLIP)(41). The CLIP needs to be removed from the peptide-binding groove to free it for the interaction with a specific peptide derived from the antigenic protein. This process takes place in the late endosomal MIIC compartment (MHC class II compartment) where it is facilitated by an HLA-DM molecule, which is expressed by all class II expressing cells (3). In addition, HLA-DM stabilizes empty HLA class II molecules, and thus prevents them from aggregation, inactivation and proteolytic degradation (42). It seems that concentration of DM required for its chaperon activity is higher than concentration needed for the peptide exchange function (43).

DM complexes with empty HLA dimers are stable and the interaction can be broken only by binding of high-affinity peptides. By ensuring preferential presentation of highly stable p-MHCII complexes (44), DM significantly narrows the repertoire of presented peptides (30) and increases the frequency of high-stability p-MHC complexes on the cell surface.

HLA-DM activity is altered by the HLA-DO molecule that binds to HLA-DM and inhibits its function (44). In contrast to HLA-DM, HLA-DO is expressed only in subsets of class II expressing cells: in certain types of dendritics cells (45), B cells and thymic epithelial cells (3), but not in monocytes and macrophages (45). The effect of HLA-DM and -DO molecules in a particular cell type depends on the ratio of free (active) DM relative to the pool of inactive DM-DO complexes and corresponds with the activation state of the cell: in APCs expressing high levels of HLA-DO (immature DCs, resting B cells), inhibition of the -DM molecule enables presenting of more diverse spectrum of lower stability peptides. In this case, low stability and low frequency of individual p-MHC complexes decrease the chance of an immune response targeted against the presented peptide and promote the self-tolerance (44). The APCs expressing HLA-DM only (e.g., activated DCs, activated B cells) present the more narrow spectrum (implying also higher frequency) of the stably bound peptides, enabling efficient immune response against the peptides presented on the activated APCs (reviewed in (44)). DM presence also increases total abundance of certain subset of MHCII alleles (alleles with low CLIP affinity (see below) – DP4, DQ1, DR**04*04) on the cell surface (43).

As the majority of peptides in the non-activated cells are self-peptides, HLA-DO expression may be a mechanism preventing development of autoimmunity (44). In line with this hypothesis it was shown, that constitutive expression of DO in murine dendritic cells is able to prevent development of diabetes in the T1D prone NOD mice by altering the spectrum of presented peptides (46).

The mechanism behind unbalanced DM and DO expression leading to increase in HLA-DM:DO ratio during immune activation may lie in their different responsibity to the IFNγ: both HLA-DM chains are,

similar to classical class II genes, inducible by IFNγ, however, HLA-DOβ expression is independed of CIITA and in most cell types irresponsive to IFNγ (3).

According to model presented by Busch, the fate of the molecules that **bypass the HLA-DM editing** (even in the HLA-DM sufficient cells a certain fraction of class II molecules may bypass HLA-DM) depends on their CLIP-binding capability and their intrinsic stability in a following way: αβ dimers that strongly bind CLIP will retain this peptide; dimers with weak CLIP binding will loose this peptide and either aggregate (intrinsically unstable αβ dimers) or load other endosomal peptides (stable αβ dimers). In the latter case, the class II molecules loaded in the absence of HLA-DM carry also low-affinity peptides and this **altered peptide repertoire** may be prone to exchange on the cell surface, e.g., for extracelluler self-peptides during an inflammation (42). In this way, **low CLIP-αβ affinity may bring an autoimmune potential**, and indeed, the class II alleles conferring susceptibility to autoimmune diseases often have low affinity for CLIP (e.g., DR4 subtypes predisposing to rheumatoid arthritis)(42) and are resistant to DM editing, as is the case of T1D-predisposing DQ2 and DQ8 molecules (47).

1.1.3.3.4 Expression level

It is difficult to separate and study the effect of class II transcription rate on autoimmunity *per se*, however, the connection between them is supported by the fact that various CIITA polymorphism are associated with increased risk of autoimmune diseases (21, 48). High class II expression correlated with SLE risk carried by a haplotype in a study of Raj (49) and an increased transcription of transgenic HLA class II molecules in DC was associated with disesase severity in a murine model of an autoimmune disease (50).

1.1.3.3.5 Posttranslational mechanisms regulating pMHC availability on the cell surface

The availability of the pMHCII complexes on the cell surface depends not only on the rate of the class II dimer production and its stability, but also on its distribution within the cell and the rate of its degradation.

HLA class II molecules can take several routes to load the peptide and present it on the cell surface: In the **classical** class II presenting **pathway**, newly synthetized HLA class II dimer loads peptides derived from endocytosed proteins in the acidic antigen-processing compartment and exposes them on the cell surface.

Some HLA class II molecules take their path further and recycle from the cell surface (**recycling pathway**): HLA complexes with bound protein re-enter cell, exchange their bound peptide for a new one in the early endosomes and travel back to the plasma membrane through recycling endosomes (reviewed in (51)). Recycling pathway is crucial for presenting epitopes that are prone to destruction by proteases in the highly acidic environment of late endosomes (52), for example, certain viral peptides (53).

Furthermore, in immature DC, class II molecules are able to reach cell surface empty and capture antigen directly from the medium, serving as endocytic receptors or to broaden the peptide repertoire by presenting peptides that would be otherwise terminally degraded in endosomes (54).

The pMHC distribution within the cell is mostly function of the cell activation state: in resting APC, most pMHCII complexes either stay in antigen-processing compartments or are degraded in lysosomes. Upon activation, pMHC molecules are redistributed to the cell surface, and their recycling is preferred over degradation (55). For example in mice, half-time of a MHC dimer can increase from 10 h in the immature DCs to over 100 hours following DC maturation (56). Furthermore, each type of APC has its preferred antigen processing and loading pathway (22). These differences can lead to preferential presenting of peptides with different properties (basically, different peptide repertoires presented by different cell types) (42).

Cytoplasmic tails of HLA class II are not required for conventional presentation pathway, but their ubiquitinylation is important in recycling pathway and in degradation (41, 57). Removing last 10 and 12 aminoacids of α or β cytoplasmic tails, respectively, basically eliminates internalization of HLA-DR molecules and presentation of peptides depending on the recycling pathway (53). Ubiquitin-independent mechanisms may contribute to the process as well, as removing the part of the DRβ cytoplasmic domain not containing lysine (which is required for ubiquitinylation) is able to cause small, but significant inhibition of the DR dimer internalization (58). *DQB1* gene exon 5 does not contain any lysine residues, but it is possible that its presence (**05:03* and **06:01* alleles) or absence (others alleles) could influence the recycling pathway usage by *DQB1* alleles and consequently the properties of presented peptides in a similar way, however, this was not tested.

In vivo studies in mice revealed minor inter-allelic differences in the degradation of MHC class II proteins (difference in alleles turnover ≤ 20%). However, the major half-time determinant was a cell lineage: turnover of all 3 tested isotypes and genotypes was faster in DC ($t_{1/2}$ cca 5–6 hours) than in B cells ($t_{1/2}$ cca 10–12 h) (59).

Similar results were obtained in studies of HLA class I molecules, where the protein turnover (ranging from undetectable to several hours) in normal human cells depended mainly on the context (cell type and activation state) and on the class I isotype, but was allele-independent (60). However, it became allele-dependent in cells with defective peptide loading, probably reflecting different abilities of alleles to utilize alternative antigen processive pathways (60).

Therefore, from the known data it seems that under normal circumstances allele identity has no major influence on rate of protein degradation or intracellular trafficking.

1.1.4 Conclusion

The p-MHC-TCR interaction is an important factor affecting the outcome of the antigen-specific immune response. The most crucial characteristics of the interaction are the p-MHC stability and the pMHC-TCR affinity and quantity, discussed above.

These characteristics are not unrelated: p-MHC stability affects the number of pMHC complexes on the cell surface and thus the pMHC-TCR signal quantity. The formation of the pMHC complex requires presence of both, peptide and the MHC molecule, so the amount of the complexes is in addition affected by the antigen dose and the availability of the MHC molecules.

The availability of the MHC molecules is a function of their production and degradation. Regulation of MHC production is the topic of this thesis and is described in more detail in the following chapter.

1.2 HLA class II expression – regulation

The following chapter describes factors that determine expression of HLA class II molecules. It presents general concepts that apply to all HLA (and MHC molecules), as well as factors that affect expression levels of certain alleles, with the special focus on the alleles of *DQA1* and *DQB1* genes.

1.2.1 HLA class II genomic region

Class II genes are located within MHC region on chromosome 6p21.3. Whole region comprises 3.6 Mb and can be divided into class I, class II and class III regions, each of them containing many genes functioning in innate and adaptive immunity. HLA class II region is located on the centromeric end and spans 0.5 Mb (61). In addition to loci coding for classical and non-classical α and β chains, most HLA haplotypes carry an additional DRB expressing locus, containing either *DRB3*, *DRB4*, or *DRB5* paralogous gene. Furthermore, there are several pseudogenes in the region: 0–3 out of 5 pseudogenes (*DRB2 and DRB6–9* (62) in DR region, the number depending on haplotype; *DQB3*; and *DPA2 and DPB2* (4). *DQA2* and *DQB2*, once thought to be pseudogenes, encode the DQX molecule expressed on specialized cell types (63).

Expression of HLA class II molecules is tightly regulated—under normal conditions, they are found only on immune cells: thymic epithelial cells, activated human T cells and antigen-presenting cells such as B lymphocytes, cells of monocyte-macrophage lineage and immature dendritic cells (constitutive expression). However, they can be induced on almost any cell by interferon-γ (inducible expression) (64) and their expression is further fine-tuned depending on several parameters, like developmental stage, activation status of the cell or extracellular stimuli (64).

1.2.2 Regulation of HLA class II expression—genetic regulation

The expression of class II molecules is regulated mainly on the transcriptional level. Several types of DNA sequences are involved in the regulation, and the most important and best defined of them is an SXY regulatory module in the proximal promoter (64). However, even though essential, the SXY module is not able to ensure normal pattern of class II expression *per se*, at least in mice (65). Several other regulatory sequences—often several kilobases distant from the transcription start site, providing long-range regulation of MHC class II expression—play important role as well (66). These regulators will be also described in the following sections.

1.2.2.1 SXY regulatory module of proximal promoter

Regulatory sequences in the proximal promoter of class II genes have been studied since the late 1980s and a mechanism of their functioning is well described.

Proximal promoter of all class II genes, including accessory genes (Ii, HLA-DM, HLA-DO (67)) contains conserved regulatory module located 150–300 base pairs (bp) upstream of transcription initiation site. Module consists of three sequences, so called S (also referred to as W region, or considered to be a part of the wider W region), X (further divided into X1 and X2) and Y boxes. Sequences of SXY module are cooperatively bound by ubiquitous, constitutively expressed protein factors, forming a multiprotein enhanceosome. The enhanceosome is not able to activate the transcription by itself, but it serves as a platform for protein-protein interaction mediated binding of major class II transcriptional regulator, class II transactivator (CIITA). CIITA recruits various transcriptional coactivators and chromatin-remodeling enzymes that trigger transcription, and thus its binding is the critical step that induces class II expression (64). In contrast to the proteins that bind the SXY module DNA, expression of CIITA is highly regulated.

An SXY module functions together as a single regulatory unit, and exact stereospecific alignment of its elements is very important for a proper arrangement of all proteins forming the enhanceosome. Therefore, not only the sequences of S, X and Y boxes, but also the distances between them are highly conserved, and polymorphisms affecting the pattern have an impact on the gene and allele expression. The important conserved characteristics of the SXY module motifs and the consequences of altering them are listed below (in the downstream direction). For a review of the regulatory sequences in the proximal promoter of class II genes, see (68)).

1.2.2.1.1 S motif

S motif is the most conserved part of the wider Z/W area (68). Specific factor recognizing this sequence was not identified yet (64), however, a CIITA recruitment requires both, an intact S motif and a correct spacing between the S and X boxes.

Functional effect of S box sequence variation is not known. There is no variation in the S box sequence within any of the class II isotypes, so if a polymorphism in the S box contributes to variant regulation of the individual alleles, the effect would present only on the inter-isotype level.

1.2.2.1.2 S-X spacer region

Absolute distance between S and X boxes has a strong influence on the CIITA binding to the promoter. Native S-X distance of the *DRA* gene is 16 nucleotides, and even 1–2 bp change in the spacer length is able to strongly reduce CIITA binding to the mutated promoter (69). Increasing the length to 18 and 26 nucleotides (nt), or decreasing it to 15 and 14 nt results in 40% and 30%, or 65% and 20% of CIITA binding compared to the original length. Only the effect of +1 nt increase is comparable to the non-modified sequence (95%)(69). However, it is interesting that a decrease in expression after inserting 5 nucleotides can be overcome by the CIITA overexpression (70).

The effect is a function of the distance, not of the orientation of the two elements: inserting either 5 or 10 nt (i.e., either half or a full helical turn) between *DRA* S and X box has the same impact on the SXY module—both mutations are able to completely abolish the module function (measured as expression of a reporter gene) (70).

TG polymorphism in S-X linking region

TG dinucleotide is present at the position -179–180 in the S-X linker region of several *DQB1* alleles. The TG insertion affect the promoter in the two opposing ways: it increases an absolute distance between S and X elements with a negative impact on the promoter activity; and it also serves as a binding site for a TG-specific nuclear protein, thus positively contributing to the allele expression (71). Nonetheless, the overall effect on the promoter strength is negative, and TG insertion is responsible for 2-fold reduced activity of *DQB1*03:02* compared to **03:01* allele, as was shown on the promoter fragment driving an expression of the reporter gene. Accordingly, the expression of **03:01* allele was higher compared to **03:02* in skin fibroblasts from heterozygous donors (71).

The TG insertion is present in several *DQB1* alleles and thus can affect their promoter activity.

1.2.2.1.3 X1 box

X1 box (also referred to as X box) lies 15–17 bp downstream of the S box (68). It is bound by a heterotrimeric protein RFX (72) that functions by promoting cooperative binding interactions with other proteins that bind MHC promoters (i.e., NFY and X2BP)(72). RFX constitutes of three subunits, RFXAP, RFXANK and RFX5. Crucial role of these three proteins was identified in studies of patients with a bare lymphocyte syndrome (an MHC class II deficiency), where mutations of either of the genes were found to be causal for the disease. In the absence of RFX, not only X, but also X2 and Y boxes are unoccupied (72) and the MHC class II expression is completely abrogated (72).

RFX shows *DRA* > *DPA* > *DQA* affinity gradient towards the promoter X boxes (73). Hasegawa *et al.* identified X1 box residues that are most important for the RFX binding, employing competition experiments where native and mutated *DRA* promoter variants competed for the RFX binding (74). They showed that positions -100, -101, and -106 (-151, -152 and -157 from start of translation) are the most important for RFX binding (mutating them reduces competitive ability of the fragment up to 90%), while positions -110, -108, -107, -105, -104, -103, and -102 have smaller, but important effect (20–90% reduction in competition). Thus, natural variation in these positions can contribute to the inter- and intra-allelic variation in class II promoter activity.

A single nucleotide G→A change at the position –159 at the beginning of the X1 region, and A→G change at position -146 in X1/X2 region is able to decrease both basal- and agonist-induced transcription of the *DQA1* promoter constructs containing X-Y region. This could partially explain the lower transcription rate of *DQA1*05:01* allele (-159 A and -146 G genotype) compared to *DQA1*03:01* (G-A genotype). However, the finding is not unequivocal as the effect of these polymorphic nucleotides also depends on the surrounding sequence: transferring the whole X region from the **03:01* allele into the **05:01* background does not increase the activity of the promoter construct (75).

1.2.2.1.4 X2 box

X2 box partially overlaps with the X1 box and its sequence is homologous to either CRE or TRE elements (12-O-Tetradecanoylphorbol-13-acetate (TPA) response element) (68). It is bound by the cAMP-response binding element, CREB (previously X2 binding protein, X2BP) (76).

CREB binds the X2 box as a homodimer (77) and the most important residues for interaction are -98, -96 and -90 on the coding strand and the -95 and -87 on the non-coding strand (77). CREB recognition sequence is CRE (TGACGTCA) (78), but CRE factors also recognize half-site motif CGTCA, which is less active than the full CRE for CREB binding and cAMP responsiveness (79). CREB is somewhat tolerant to the variation in its recognition sequence, and the importance of individual residues depends on the cell type examined (78). The X2 box sequence varies both between class II isotypes and between alleles within an isotype, but there is no study on effect of this variation on the class II promoter strength.

TRE element (TGA(C/G)TCA) is present in the *DRA* and *DPB* promoters and is recognized by the transcription factor AP-1(80). AP-1 is needed for maximal transcription of these genes (81).

1.2.2.1.5 Relative orientation of X and Y elements

X1 and Y boxes are separated by approximately 2 helical turns (20 nucleotides) of DNA and are aligned on the same side of the DNA helix (82). Requirements for exact spacing between X and Y elements are much looser compared to spacing of S and X elements, provided their relative orientation on the DNA

helix is preserved. Inserting up to four integral turns of DNA helix between X and Y boxes of *DRA* gene promoter does not abolish CAT reporter gene expression (it can even raise it), but adding half of integral turn strongly impairs the gene function (20–40% of the original level) (70). This effect is not caused by an inability of X and Y boxes to bind their respective proteins (82).

1.2.2.1.6 Y box

Y box lies 19–20 bp downstream of X box (68). It is an inverted CCAAT box recognized by a trimeric protein NFY, composed of NF-YA, -YB, -YC subunits. CIITA interacts with NF-YB and -YC but only minimally with NF-YA (70), which is the subunit responsible for the DNA binding (83) (Nardone, 2016).

Based on the A/G polymorphic site in the position -118 in the Y box, *DQA1* alleles can be divided into two groups: *DQA1*01*,**02* and **03* (G allele); and *DQA1*04*,**05* and **06* allele groups (A allele) (4, 75, 84). Y boxes of G-alleles are preferred targets of NF-Y binding, which is accompanied by higher mRNA expression of these alleles (85), both in basal and cytokine (TNFα, IFNγ) induced transcription (tested on the promoter construct in reporter gene vector (75). However, an A-allele Y box seems to be bringing an increased responsivity to the TNFα induction (75).

Y boxes of the *DQA1*01:01* and **05:01* alleles (but not *DRA* Y box) are able to bind DEK, a nuclear protein that changes the topology of chromatin (86). Importance of this observation is unknown (86) and DEK binding on the Y box and its function was not followed upon in any further study.

1.2.2.2 Other regulatory elements

In addition to the SXY module, which is common to all class II genes, there are also *cis*-acting gene-specific elements, such as:

- Octamer motif (*DRA* gene), recognized by the NF-Al and NF-A2 proteins was speculated to be responsible for a high activity of the *DRA* isotype (68), however, mutations of the octamer element have no effect on either *DRA* induction by CIITA or IFN-gamma, or on expression of reporter gene in transfection assays (70).
- An inverted X-box right downstream of a *DRA* gene S-box (68).
- VDRE (vitamin D response element) present downstream of SXY module in certain *DRB1* haplotypes (87) with unknown functional implications.
- A putative NF-κB binding site (*DQA1* gene).
- TNF response element (a sequence homologous to the TNF-response element in the promoter of murine Ia class II gene (88)), which is present in all *DQA1* alleles (68).

For an overview of SXY module position and sequence on the promoter of *DQA1* and *DQB1* alleles, see **Figure 1**. All *DQA1* gene sequence elements are summarized in article by (89).

In conclusion, proximal promoters of HLA class II genes are highly polymorphic, and natural or artificial variation in the sequence can have a significant effect on the binding of transcription factors (85) and promoter activity. However, the high cooperativity of class II transcription factors binding to the promoter can be an effective mechanism to overcome or decrease the deleterious effect of certain sequence variants, and explain the coordinated regulation of all genes despite sequence differences (77).

1.2.2.3 CIITA

A master regulator of HLA (and MHC) class II genes transcription is the CIITA protein. It controls transcription of all HLA class II genes (including *Ii*, *DM* and *DO* (72)) very strictly and in a coordinated fashion (64, 90). Although some exceptions do exist (91), generally we can say that without CIITA, there is no class II transcription; furthermore, there is a strong positive correlation between levels of the CIITA protein or mRNA in the cell and the level of any of the HLA class II transcripts (90, 92). In addition, basically all inflammatory stimuli that boost HLA class II expression act through CIITA (93).

CIITA as well contributes to the expression of MHC class I and several other genes (FasL, IL-4 - studies on mice), however its role is not so fundamental as for MHC class II genes (94).

1.2.2.3.1 Mechanism of class II genes activation

CIITA acts both as a transcriptional activator and as a general transcription factor (95). After binding to the MHC enhanceosome, CIITA activates transcription by several mechanisms: it recruits general transcription factors (TFIID and TFIIB), induces phosphorylation of RNA polymerase II, interacts with positive transcription elongation factor b (P-TEFb), and recruits co-activators that alter chromatin accessibility by inducing histone modifications (acetylation, methylation) and chromatin remodeling (64).

As observed on *DRA* promoter in melanoma cell line stimulated with IFNγ, induction of class II genes expression has two phases, both associated with chromatin remodeling in the region. In first phase, CIITA binding to enhanceosome is accompanied by a global increase in histone 4 acetylation (H4Ac) over a large region surrounding promoter (at least -5kb to +1kb from TSS) and by generation of intergenic transcripts from the whole region (with peak of transcripts production at distal SXY module, see below (96). Finally, four hours (and later) after an emergence of CIITA mRNA, nascent class II transripts appear and activation acetylation and methylation marks are deposed on histone 3 around TSS, a process dependent on ongoing transcription elongation (96).

1.2.2.3.2 CIITA protein structure and function

An N-terminal part of CIITA contains transcription-activation domains that are considered to mediate interactions with effector proteins that start transcription, including components of the general transcription machinery, factors involved in chromatin remodeling and other co-activators. The C-terminal two-thirds of

the protein play a role in self-association, localization to the nucleus and recruitment to the enhanceosome (64).

Activity of CIITA protein can be altered by a variety of posttranslational modifications on multiple residues: effect of serine phosphorylation varies depending on which residue was modified, resulting in either inhibition (more common) or stimulation of target gene transcription, oligomerization and nuclear export. Lysine acetylation supports nuclear localization of CIITA and deacetylation by HDAC2 and SIRT has lowering and supporting effect on CIITA stability, resulting in diminished interaction with RFX5 and increased class II activation, respectively. Mono- and polyubiquitinylation is required for enhanceosome formation and leads to increased class II activation (reviewed in (97)).

1.2.2.3.3 CIITA isoforms and their function

CIITA is transcribed from 4 different promoters, each with its own role, resulting in 3 protein isoforms that differ by their N-terminal parts: the sequence of the shortest isoform IV is shared by all variants, but the isoforms I and III have an additional, isoform-specific, N-terminal part. Each isoform functions in different cell types:

CIITA isoform I

Activation of CIITA promoter I leads to production of CIITA isoform I that regulates constitutive HLA class II expression in dendritic cells (DC) and cells of macrophage lineage (both with and without induction by IFNɣ) (94).

CIITA isoform III

Promoter III and CIITA isoform III regulate constitutive expression in B lymphocytes (98) and in plasmacytoid DC; and inducible expression in activated human T cells (99). Transcription of this form is also induced by IFNγ, although to a lower extent compared to isoform IV (98).

CIITA isoform III has the highest turnover rate (half-life of around 30–40 minutes) and highest transactivation potential (over 5-times higher than forms I and IV), both properties conditioned by an integrity of first 10 amino acids acting as a (portable) degron and transactivation sequence (100).

This isoform is also more efficiently recruited to the *HLA-DRA* promoter *in vivo* with and shows increased interaction with components of the transcription machinery compared to forms I and IV (100).

CIITA isoform IV

Isoform IV, transcribed from promoter IV, is responsible for IFNɣ inducible CIITA expression in cells of both non-hematopoietic and hematopoietic origin (94, 98). Inhibition of transcription from pIV seems to be a point on which several cytokines (IL-10, TGF-β) and pathogens (*Chlamydia,* Cytomegalovirs) inhibiting CIITA expression act (94, 101).

The usage of individual promoters is not completely exclusive, as significant fraction (around 1/3) of CIITA in DCs originates also from pIII (98), and after IFNγ induction, not only pIV, but pI is activated to small extent as well.

The differences between the protein isoforms are summarized in the **Figure 2**. Despite the role of N-terminal domain in the transcriptional activation, no information (except the study of Beaulieu) is available on the functional consequences of these differences.

We can only assume that function of all isoforms will be equally affected by posttranslational modifications, as they take place only in the non-variant region of the protein.

1.2.2.4 Inverted SXY module (XL sequences)

In addition to the proximal SXY module, several sequences highly homologous to X and Y boxes, but in reverse orientation, were identified within MHC class II and extended class II region.

First of these promoter distal regions containing X' (X like) and Y' (Y like) sequences was identified approximately 2.2 kb upstream of *HLA-DRA* gene (102). The sequence was bound by NF-Y *in vitro* (66), and by RFX and CIITA *in vivo* (103). CIITA and RFX binding to this region established a large domain of open chromatin spanning up to proximal promoter, and enriched in acetylation of histone 3 (H3) and histone 4 (H4). Distal S'X'Y' element also recruits polymerase II (Pol II) and low abundance bidirectional extragenic transcripts are generated from this region. (103). Binding of CIITA to the distal SXY leads to hyperacetylation and opening of chromatin up to 16 kb upstream (103).

Team of Krawczyk *et al.* searched for SXY homology sequences not associated with promoter and identified 4 new sequences within the class II area. The sequences bound RFX, CREB, and in MHC positive B cell lines also CIITA, to the levels comparable with the *DRA* promoter. In MHC II negative cell lines, binding of (RFX and) CIITA was inducible by IFNγ (104). CIITA binding to these regions induced histone acetylation over several kb long regions (upto 4 kb upstream; and upto 4 kb downstream, where measured), and the SXY sequences were able to function as enhancers—replacing *DRA* SXY sequence in proximal promoter (-150 - +10) with a sequence #4 from the DQ region or a sequence #8 from the DP region was able to drive an expression of a reporter gene (104). Interestingly, 2 active SXY sequences were identified also in the intron 1 of invariant chain gene on chromosome 5. These two sequences are also associated with CREB, RFX, and CIITA, and S"-Y" motif drove expression of reporter with efficacy of 50–80% compared to SXY of *DRA* gene.

Further bioinformatics search by Gomez and colleagues (105) revealed 32 X-like sites accompanied with Y-like sites 18–35 bp downstream in class II and extended class II region. The sequences were named XL (X-like), and were numbered in the order in which they are present on chromosome in a centromeric direction. At a closer examination of XL1–8 sites located in area spanning from upstream of *DRA* to DRB3

gene (**Figure 3**), XL4 (which is identical to the first X'Y' sequence characterized by Feriotto *et al.*, XL7 and XL8 were shown to bind RFX5 and CIITA in Raji B cell line. The amount of these factors bound to XL elements was lower compared to *DRA* proximal promoter, however, for the most bound XL7 its level was still within the binding range of biological significance. Similarly to proximal SXY module, CIITA binding was connected with chromatin remodeling activity, and high level of H3 and H4 acetylation and other histone modifications were observed at the XL elements. Many histone modifications of "active" XL4, 7, and 8 boxes were shared with proximal promoter SXY module of *DRA* gene. Furthermore, these elements were able to drive expression of reporter gene up to 65% of normal *DRA* X-Y module driven levels (105).

The human studies where the XL regions would be inactivated are not available to answer a question if they are necessary for the class II expression, however, a region homologous to XL4 is necessary for a normal expression pattern of I-Eα (*HLA-DRA* homologue) in murine B cells (102). It was pointed out that the high evolutionary conservation of these sequences (there are no differences between human XL4 box and its murine counterpart) also implicates their functional importance (102).

Several authors suggested that these distal sequences could act as distal enhancers (104) or locus-control-regions (LCR) (103). This is supported by observations of direct interaction between regulatory complexes binding distal modules and proximal promoters (105), induction of long range chromatin modifications and generation extragenic transcripts from these regions.

The regulatory potential is consistent with location of the XL sites at strategic positions separating the individual class II isotypes (104), **Figure 3**.

1.2.2.5 CTCF (insulator) binding sequences

The next category of sequences important for class II expression are the CTCF binding sequences. CTCF (CCCTC binding factor) is a zinc-finger DNA binding protein that acts in a complex with cohesin to form chromatin loops, or domains, by bringing together a pair of CTCF binding sites, which then form a base of the loop (106). Orientation of CTCF sites matters, as looping preferentially occurs between convergent CTCF motifs (107). In chromatin loops, contacts between regulatory elements from different loops are inhibited and interactions within a loop are favored, thus the distinct loops are (partially) functionally separated. This way, CTCF function has been implicated in regulating the chromatin architecture (by dividing the genome into functionally separate units), gene expression, and functioning as a gene insulator (reviewed in (106)). Disruption of the looping pattern is able to alter gene expression (107).

Ten functional binding sites for CTCF are located within HLA class II region, in the key positions between individual subregions (108, 109) (**Figure 3**). Each of these CTCF binding sites is able to interact with several of its neighbours (108), in both, class II expressing and non-expressing cells. In the presence of CIITA, 3D structure of region reorganizes and CTCF sites form interactions with CIITA, and through

them with class II promoters as well (each promoter preferably interacting with its closest CTCF sites) (38, 108). This organization seems to be crucial for an efficient expression of all class II genes associated with antigen presentation, because CTCF is required for reaching maximal expression levels (108). Abrogation of CTCF binding either by epigenetic modification or artificially by siRNA reduces class II production up to 20–40% of the original levels (38, 108, 110).

The three-dimensional arrangement of the class II region by CTCF proteins may also play role in a coordinated regulation of the whole locus by organizing all genes within the locus into the same area of the nucleus and forming a domain of high transcriptional activity (108). Located between individual isotypes (and even between the pseudogene pairs), CTCF sites can separate transcriptional units of the individual isotypes (109).

The first CTCF binding site in the class II locus was identified between *DQA1* and *DRB1*, in a proximity (within 500 bp downstream) to XL9 SXY homology region (see previous section), and thus was referred to as XL9 (109). Chromatin region around XL9 element was highly acetylated, but the element itself did not bind RFX nor CIITA and did not function as an enhancer nor repressor. XL9 element instead interacted with nuclear matrix and functioned as an enhancer blocking element (109).

Since then, further investigation of the XL9 region revealed that in addition to the CTCF binding site, XL9 region also contains 30 binding sites for other proteins regulating transcription and chromatin structure, all located within 1 kb distance (49). Several SNPs in these binding sites (e.g., an SNP in the IRF4 binding site) seem to have regulatory role, as they are associated with higher expression of the linked *DRB1*, *DQA1* and *DQB1* alleles: an mRNA expression level in the homozygotes for one allele could be up to 2-fold increased compared to homozygotes for the second allele, and an expression levels in heterozygotes is a mean of these 2 extremes. These genetic differences also translated into different amount of the class II chains on the cell (monocyte-derived DCs) surface (49).

In conclusion, there are distal regulatory modules within extended class II regions. They bind transcription factors, interact with proximal promoters, and induce long range chromatin modifications. Genetic and epigenetic variation in these regions has an ability to affect class II gene transcription rate.

1.2.2.6 Intragenic sequences

Several intragenic sequences can be involved in regulation of class II expression: two enhancer sequences, one in the intron 1 of the *DQA1* gene and one spanning from intron 2 to exon 4 of the *DQB1* gene, were shown to increase an expression of a reporter gene in DQ positive BJAB B cell line and DQ negative Jurkat T cell line transfected with the sequence-reporter construct (111). In the *DRA* gene, a fragment from the central part of intron 1 is able to function as an enhancer in lymphoid cells, but not in

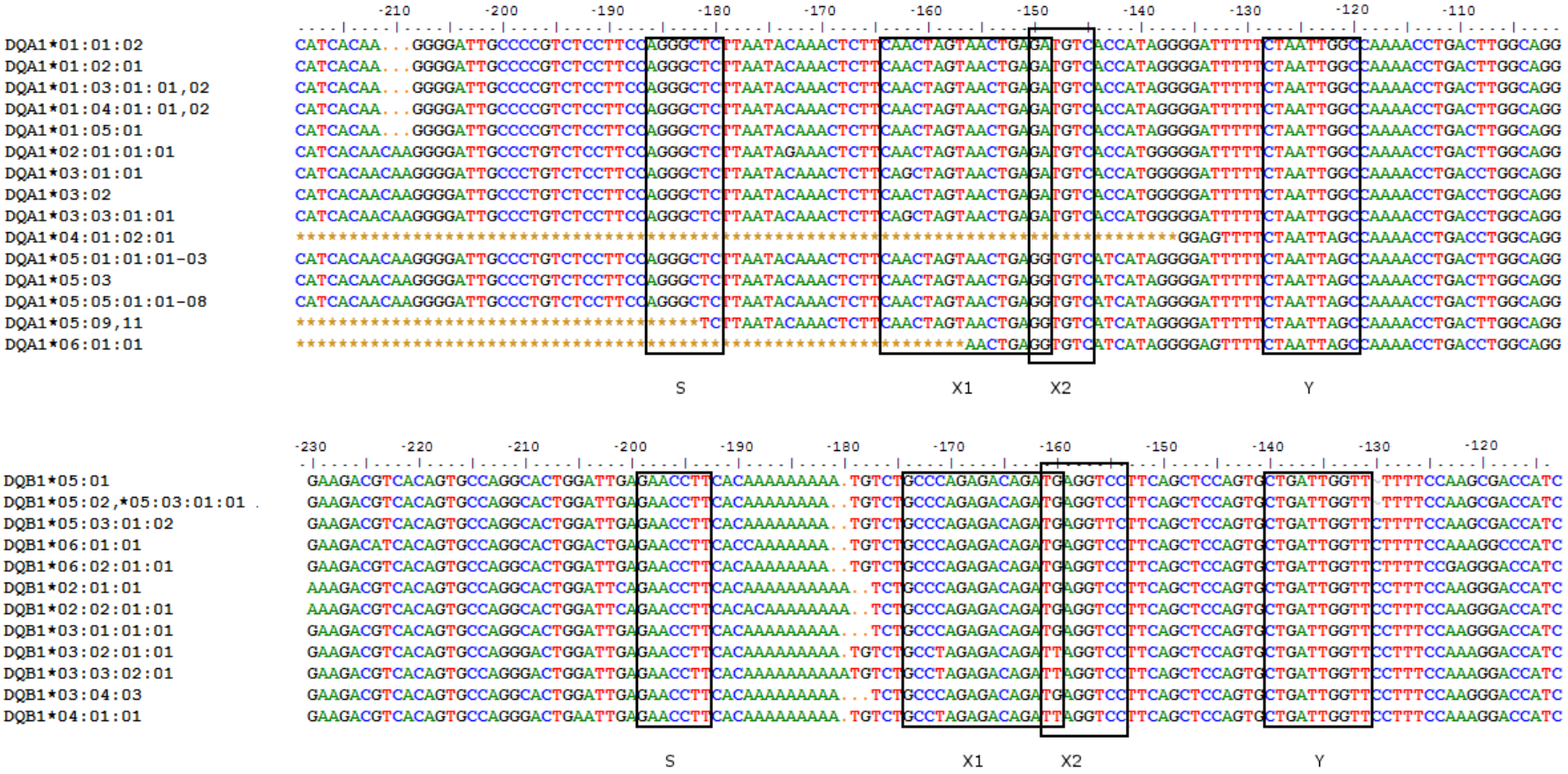

Figure 1. Sequences of *HLA-DQA1* (A) and *DQB1* (B) proximal promoter region. The boxes mark location of conserved elements (S, X1, X2 and Y boxes) of the regulatory SXY module. The allele sequences were copied from https://www.ebi.ac.uk/ipd/imgt/hla/align.html using the IMGT/HLA database Sequence Alignment Tool (112); and aligned and visualized with BioEdit software. The regulatory boxes are marked based on sequences probvided by Benoist *et al.* (113).

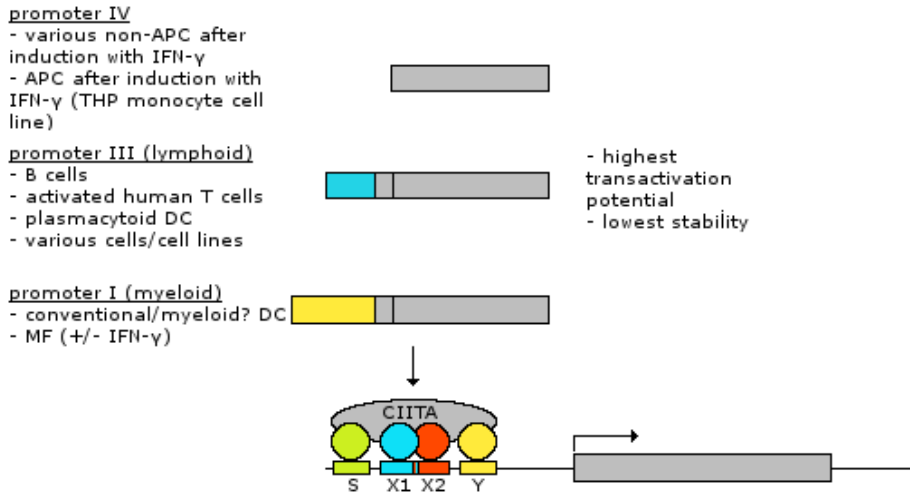

Figure 2. **Class II enhanceosome and CIITA**. (a) CIITA protein isoforms. (b) SXY module present in promoter of all class II genes binds ubiquitous transcription factors RFX, CREB and NFY that together form MHC class II enhanceosome. Expression of CIITA is highly regulated and binding of this factor triggers transcription from class II promoters. Figure based on (64, 94, 98).

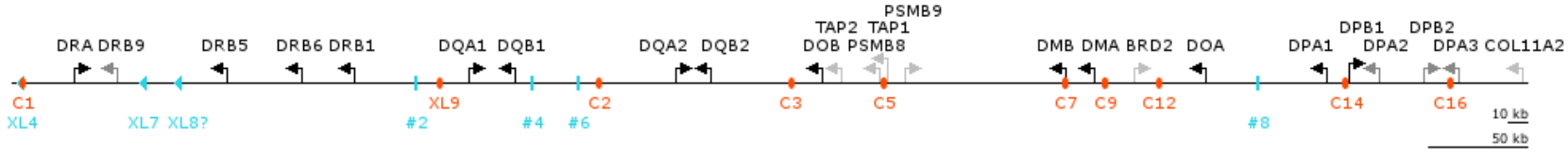

Figure 3. XL elements and CTCF binding sites in the HLA class II region. Non-promoter **SXY** elements are represented by blue triangles, whose orientation coresponds to the orientation of an SXY element (104, 105).
CTCF binding sites are marked by red ovals. The sites are located at the borders of the evolutionary subdomains and at the edges of the whole class II locus (108). Only "active" (i.e., binding transcription factors or CTCF) XL a CTCF sites are shown. The sites are marked by the names used by authors who described them. HLA class II genes are presented in black, pseudogenes in dark grey and non-HLA genes in light grey. Non-class II pseudogenes and non-translated mRNAs are not depicted.

fibroblasts (not even after induction with IFN-y) (114). However, there was no follow up on these studies to examine regulatory function of these sequences in a greater detail.

DQA1 gene displays a microsatellite polymorphism in the intron 3. A CTTT tetranucelotide repeat is present in one copy in the *DQA1*01:01–DQA1*03:03* alleles; in three copies in *DQA1*04:01–DQA1*05:03* and **06:01*; and finally *DQA1*05:05* allele contains 10–20 repetitions (4, 115). Moreover, the *DQA1*05:05* allele shows also inter-individual polymorphism. Authors mention the possible role of microsatellites in the regulation of expression, but the biological significance of the *DQA1* gene microsatellite was not explored in this (115) nor in further studies.

1.2.3 Regulation of HLA class II expression—posttranscriptional regulation

Even though the major control over class II expression is provided on the transcriptional level, the final number of class II dimers that appears on cell surface to present peptide fragments depends on other factors as well. The availability of α and β peptide chains is a function on their production rate (that can be affected by alternative mRNA splicing and mRNA stability), protein stability and degradation. Finally, amount of peptide-HLA complex on the cell surface depends on the peptide availability. The posttranslational mechanisms affecting class II genes are summarized below, with a focus on *HLA-DQA1* and *DQB1* genes and on allele-specific regulation.

1.2.3.1 Posttranscriptional mechanisms

1.2.3.1.1 Alternative splicing

Sequence polymorphism within class II genes can alter mRNA splice sites and other positions important for splicing. Effect of the alternative splicing depends on which part of gene it affects (UTR, cytoplasmic domain, etc.), and is discussed later in this chapter.

Alternative splicing of DQB1 gene alleles

DQB1 gene has 6 exons: exon 1 encodes the leader peptide, exons 2 and 3 encode two extracellular domains, exon 4 encodes a transmembrane region and exons 5 and 6 encoding cytoplasmic part of the protein (116, 117).

Skipping of exon 5. Due to a polymorphism in the splice acceptor site, *DQB1*05:03* and **06:01* are the only *DQB1* alleles that utilize 24 nucleotide long exon 5 coding for a part of cytoplasmic domain (4, 117). Spliced versions of these alleles with exon 5 skipped are common in the B cell lines, forming up to 25–50% of *DQB1*05:03* and **06:01* transcripts, and were found in peripheral blood lymphocytes of healthy individuals as well (117). The importance of exon 5 is not clear: in study of Senju *et al.*, the antigen presenting ability of spliced variants was not compromised—they were able to associate with DQα chain, be expressed on cell surface, and stimulate proliferation of CD4+ T cells treated with mitomycin C (117).

The longer version seemed to be somewhat more efficient in an antigen presentation only at closer look, indicating a role of exon 5-coded cytoplasmic domain in the signal transduction (117). The signaling role of the cytoplasmic domain was confirmed later: the domain contains ubiquitination, cytoskeleton-interaction and cAMP signaling motifs. cAMP signaling is important for B cell activation and antibody production and is, in murine B cells, directly dependent on the cytoplasmic domain of class II β chain – more specifically, on its GP motif, which is present only in alleles containing exon 5. In all remaining alleles, GP is replaced by GL, which is unable to drive cAMP signaling (118). Based on this finding, the B cell function of individuals carrying only *DQB1* alleles that exclude exon 5 (majority) could be compromised. Based on the allele frequencies in Czech Republic from our previous study (119), this would mean that only around 4% of Czech population are responsive to cAMP stimulation of B lymphocytes; however, the cAMP signaling in B lymphocytes based on their class II haplotype was not measured (118). Furthermore, it is possible that the cAMP unresponsiveness caused by missing exon 5 of *DQB1* gene may be compensated for by a similar GH motif of *DRB1* genes, GH motif activity was not tested though (118).

Skipping of exon 4. Exclusion of exon 4 was observed in a proportion of *DQB1*03:01*, **03:02*, **03:03*:02, **04:01*, **04:02*, **05:01*, **05:02* and **06:02* transcripts (in HLA homozygous cell lines) (117, 120, 121); and was never detected in *DQB1*02:01* and **02:02* alleles (120, 121). As the exon 4 encodes a transmembrane domain of DQβ chain, the cell lines with alternatively spliced alleles produce secreted forms of HLA-DQβ protein (120). Spliced alleles can form up to 20% of primary transcripts of *DQB1*04:02* allele (122).

One consequence of skipping the transmembrane exon is a decrease in the number of class II molecules on the cell surface. Function of the secreted splice-forms themselves is unknown, however, secreted class I molecules were found to regulate immune responses by inducing apoptosis in activated CD8+ T cells (123).

Alternative splicing of DQA1 gene alleles

DQA1 gene has 4 exons: exon 1 encodes the leader peptide, exons 2 and 3 encode two extracellular domains, and exon 4 encodes a transmembrane region and a cytoplasmic part of the protein (116).

Alternative splicing of the 3' UTR. Alternative splicing of the last exon 4 (downstream of the stop codon) and usage of alternative polyadenylation signals causes length polymorphism of *DQA1* mRNA 3' UTR (124). As a result, *DQA1*01:01*, **01:02*, and **01:03* alleles form four mRNA isoforms (two spliced variant cleaved at two distinct polyadenylation sites), and *DQA1*02:01*,**03:01*, **04:01*, and **05:01* can generate 6 mRNA forms (three spliced variant cleaved at two polyadenylation-sequence signal), as observed in homozygous cell lines (124). Each allele preferentially (> 80%) generated one product, and the length of the most abundant isoforms followed **04:01*,**05:01 > *01:01, *01:02, *01:03 > *02:01*, **03:01* gradient (124).

3’ UTRs can contain sequences important for regulation of the mRNA stability or translation and variation in this region could be a mechanism of post-transcriptional regulation of an allele expression. However, in the study above, no considerable differences in half-lives between mRNA splice isoforms were observed (124).

1.2.3.1.2 mRNA stability

Variation in both inter-isotype and inter-allelic mRNA stability was detected for class II genes, possibly adding another layer of regulation to the class II expression. When comparing the class II genes, the *DRA* mRNA is more stable than *DRB1* mRNA ($t_{½}$ = 3.24 h and 2.6 h, respectively)(125), and much less stable than *DQB1* gene mRNA (*DQB1*02:01* and **04:02* alleles had $t_{½}$ of around 24 h)(122).

Interesting observation was done by Pisapia *et al.* who noticed an approximately 25% increase in the stability of both *DRA* and *DRB1* mRNA after transfection with a vector expressing mRNA coding for either of the two cognate chains. The effect was observed also for the DQ isotype. The authors suggested that this may be a cells mechanism to maintain proper stoichiometry between amounts of individual chains and balanced dimer formation on the cells surface. The co-regulation of the two chains was observed in both class II expressing and non-expressing cell types, included regulation of both, the transcription rate and export from nucleus to cytoplasm, and was mediated by binding of 5’ and 3’ UTRs to the same ribonucleoproteic complex (125).

The interallelic differences in mRNA stability of the *DQA1* and *DQB1* genes were described as well, with half-life of *DQA1*05* (3 h) shorter than that of *DQA1*01* (4 h) and *DQB1*02* half-life shorter than that of *DQB1*03* and *DQB1*05* (126). A different study observed no difference in the stability of *DQB1*04:02* and **02:01* alleles (with much longer half-lifes of around 24 h) (122).

1.2.4 Regulation of HLA class II expression—epigenetic regulation

Epigenetics studies heritable changes in phenotype that are not caused by an underlying change in genotype. Main epigenetic mechanisms are DNA methylation, covalent posttranslational modifications of histone proteins and RNA-mediated gene regulation. Different chromatin modifications are interconnected and influence each other.

1.2.4.1 DNA methylation

One of the epigenetic modifications in eukaryotes is DNA methylation, which involves covalent attachment of methyl group on the fifth carbon of cytosine in the CpG sequence, creating a 5-methylcytosine (5meC). This sequence is palindromic—i.e. CpG sequence is present on both strands and usually both complementary CpG dinucleotides are methylated (127).

1.2.4.1.1 Function of DNA methylation

DNA methylation together with histone modifications play an important role in modulation of chromatin structure, control of gene expression and other chromatin-dependent processes. It is responsible for genomic imprinting, suppression of transposable and retroviral elements in the mammalian genome, and plays role in the X-chromosome inactivation. It enables cellular differentiation by creating and maintaining diverse gene expression patterns (127–129).

Most CpG dinucleotides in mammalian genome are methylated. Effect of DNA methylation depends on its position in genome, but both, in gene regulatory regions as well as in gene body is important for regulation of its expression. Methylation in the gene body decreases ability of RNA polymerase to transcribe a gene on the level of RNA polymerase binding, transcription initiation and elongation (127).

For regulation of transcription initiation, the most important is methylation located in regulatory regions (in a promoter or a first intron) of the gene**.** Promoter methylation is generally associated with gene silencing and the mechanism of the association can be described by two non-exclusive models:

- Methyl group directly blocks an access of transcriptional coactivators to their cognate sequences
- 5meCpG is recognized by methyl-CpG-binding proteins (MBD) that induce repressive state of chromatin (127, 128). In this case, it is plausible that the strength of the effect could depend on the local concentration of methylated cytosines (130).

To study effect of promoter DNA methylation on gene expression, Weber *et al.* (130) divided gene promoters into three categories based on their C, G and CpG density. Then they explored promoter DNA methylation status and expression of the linked gene. Promoters with low CpG content were enriched among tissue-specific genes, were generally methylated, and their methylation status did not preclude the gene activity. Authors assume that at low methylation density, MBD binding is probably not sufficient for active repression of expression, however, this does not exclude the role of low-density methylation in reducing the transcriptional noise.

Promoters with high CpG content are more frequently (but not exclusively) associated with housekeeping genes and are usually non-methylated regardless of promoter activity. Promoters with intermediate CpG content were intermediary methylated and a higher correlation between promoter methylation and activity was observed. In promoters with high and intermediate CpG content, DNA methylation, if present, represses gene expression (130).

In contrary, another team observed a negative correlation between a level of promoter methylation and gene transcription across promoters with high to low CpG densities (131)**.**

According to criteria of the study above, all class *DQA1* and *DQB1* allele promoters belong among low-CpG promoters, which should be methylated and their methylation should not be an obstacle to the

allele expression. A CpG rich region is located in the intron 1 of *DQB1* gene. As DNA methylation of regulatory element located in the first intron of a gene is able to suppress transcription of interleukin 4 gene (132), it is possible that the same mechanism could regulate the *DQB1* gene as well.

1.2.4.1.2 Methyl-CpG binding proteins

Interpretation of the DNA methylation mark is mediated by methyl-CpG-binding proteins (MBP) that specifically recognize methylated CpG dinucleotides through their methyl-CpG binding domain (MBD). The only exception is Kaiso protein that recognizes methylated DNA through zinc-finger domains and proteins using a SRA (SET- and RING-associated) domain for recognition. All MBPs associate with corepressor complexes which modify chromatin and in this way suppress the gene expression. MBPs are to a high extent redundant, as loss of DNA methylation leads to the death in an embryonic period, but knock-out of MBP genes leads to much milder phenotype (127, 133).

Proteins with MBD domain are MBD1, MBD2, MBD3, MBD4 and MeCP2. MBD2 is, together with chromatin-remodeling complex NuRD/Mi2 (nucleosome remodelling and histone deacetylation; it contains, for example, histone-deacetylases 1 and 2) part of the MeCP1 complex (methyl-CpG binding protein 1) that efficiently binds densely methylated CpG regions and suppresses transcription depending on the DNA methylation density. It is involved, for example, in suppressing of the non-induced *IL-4* gene (133).

MeCP2 consists of MBD and a transcription-suppressing domain. Its activity depends on association with histone deacetylases. It recognizes methylated CpG sites flanked by a sequence of four A or T nucleotides. It makes a complex with DNMT1 and in this way participates in maintenance DNA methylation, and it may also function as a regulator of alternative pre-mRNA splicing. Mutations of this protein are cause of Rett syndrome (127).

MBD3 contains aminoacid changes that prevent it from binding to mCpG, however, despite this it is a part of NuRD corepressor complex. This complex can be recruited to DNA by various repressor proteins, and a role of MBD3 in the complex is not clear.

MBD4 is the only MBD-family protein that probably does not play a role in repression of transcription—instead, it mediates repair of errors in DNA sequence caused by spontaneous deamination of 5-methylcytosine to thymine. MBD4 binds to symmetrically methylated CpG sites, though it has a greater affinity for the mismatching 5mCpG/TpG base pair, from which it removes thymine using its glycosylase domain. Thus, T/G mismatches are preferentially corrected to C/G basepairs, but this process is imperfect, and C→T and G→A transitions make up more than 20% of all substitutions that are the cause of genetic diseases in human (133).

Kaiso preferentially binds the methylated CGCG sequence, but it also recognizes a specific sequence that does not contain CpG. It associates with the N-CoR corepressor complex that includes histone deacetylases (127).

UHRF1 binds to hemimethylated DNA via its SRA domain and brings a DNMT1 to it (134).

1.2.4.1.3 DNA methyltranserases

A transfer of the methyl group from the S-adenosyl methionine donor (SAM) to the DNA is carried out by DNA-methyltransferase enzymes (DNMT).

DNMT1 is a 'maintenance' methyltransferase. It adds a methyl to cytosine in hemimethylated CpG dinucleotides during replication, when it is associated with a replication fork; and during a DNA repair (129). DNMT3a and DNMT3b are *de novo* methyltransferases. They recognize and methylate sequences in which none of the pairing CpG dinucleotides is methylated. DNMT3L protein has structure similar to other methyltransferases, but lacks the conserved residues necessary for the catalytic activity as well as the DNA-binding domain. Instead, it acts as a regulatory protein—by its C-terminal domain it associates with the C-terminal (catalytic) domain of DNMT3a and DNMT3b and increases their methylation activity. In DNMT3a, this is achieved by stabilization of the conformation of the active site (129). However, a border between de novo and 'maintenance' methyltransferases is not entirely strict—DNMT1 preference of hemimethylated sites compared to unmethylated sites is not significant (therefore, it co-operates with the helper proteins that direct it to the target site), and DNMT3a and 3b are involved in the 'maintenance' DNA methylation as well (134).

DNMT3a and 3b are redundant to some extent, but not entirely, as each of them has its unique methylation targets. This may be caused by their DNA substrate preferences: DNMT3a prefers the RCGY sequence (where R = A or G and Y = C or T) and methylates the CpG sites in the DNA connecting the nucleosomes in the absence of histone H1, while DNMT3b favors the YCGR sequence and methylated DNA wound around histones (135). Furthermore, all DNMTs may or may not be DNA sequence-specific, depending on the proteins they form complexes with (136).

DNMT3a forms dimers with DNMT3L, that can further dimerize and form the DNMT3L-3a-3a-3L tetramers, in which the two active sites of DNMT3a reach a position in which they can bind to a major groove of DNA in a 8–10 bp distance. According to this model, such complexes would preferentially methylate CpG sites spaced 8–10 base pairs, and this periodicity in DNA methylation was indeed observed (134).

DNA methyltransferases themselves participate in gene silencing through interaction with histone-deacetylases and histone-methyltransferases (128).

1.2.4.1.4 Environmental impacts on DNA methylation

A DNA methylation pattern is established at an early stage of ontogenesis. In the dividing zygote, the methylation pattern is largely erased and then re-established after implantation. The extent of *de novo* methylation gradually decreases during differentiation and is rare, but not absent, in the postgastrulation period. The aging of the organism itself is associated with hypomethylation of repetitive sequences and hypermethylation of CpG islands associated with some genes (137).

DNA methylation is affected by exposure to various environmental or behavioral factors. Changes in DNA methylation are, for example, caused by exposure to heavy metals (cadmium, nickel, mercury, arsenic), which deplete SAM; increase, decrease or inhibit DNA-methyltransferases, depending on the chemical and exposure duration (138). Exposure to UVB-radiation decreases DNMT1 catalytic activity and causes DNA hypomethylation in patients with systemic lupus erythematosus (139). It was shown that sleep deprivation exerts its effect on metabolic changes and cognitive defects through DNA hydroxy/methylation (in addition to changes in histone modifications and expression of non-coding RNA)(140). It is possible that methylation of DNA is affected by the food components that may serve as methyl group donors, for example folate: in a progeny of mice fed during pregnancy with a methyl donor rich diet, a higher proportion of individualsis able to methylate an IAP retroviral element, compared to the progeny of females fed with a standard diet. Increased methylation of this element has even been observed in the grandchildren of the original females (141).

1.2.4.1.5 Allele-specific DNA methylation and its effect on gene expression

Differential expression of two alleles of one gene, ranging from mild overexpression to exclusive expression of one allele, has been described by several authors in non-imprinted genes (142–145). As the relative expression of two alleles was observed in the samples from one individual, the effect of *trans*-acting factors is minimized, and the observed differences are apparently due to factors acting in *cis*, such as polymorphism in the promoter region. Another plausible option is that the genetic polymorphisms translate into variant methylation of individual alleles. There was observed variant DNA methylation of two alleles of one gene, correlating with a sequence of a given allele, as well as association of this variant methylation with an allele expression (144–146). This phenomenon can affect up to 10% of human genes (145).

1.2.4.1.6 HLA class II DNA methylation

Direct blocking of transcription factors binding in the HLA region

The only factors involved in regulation of class II expression that contain a CpG dinucleotide in their recognition sequence and whose binding is sensitive to cytosine methylation are CREB and CTCF.

CpG in the CREB binding site (X2 box) is present only in the SXY module in the promoter of the *HLA-DRA* gene. The CpG cytosine is mutated in all other class II isotypes, and thus *DRA* is the only class II gene in which CREB binding can be directly regulated by DNA methylation.

There are several CTCF sites interspersed through class II region. It was shown that DNA methylation of CTCF sites surrounding DQ locus (**Figure 3**) is responsible for the lack of DQ expression in DQ negative B cell line. DNA methylation was associated with formation of heterochromatin and loss of CTCF binding to these sites; and with loss of RFX and CIITA binding to the *DQA1* and *DQB1* promoters (though methylation of promoter regions of these genes could also contribute to the effect) (110).

Effect on chromatin structure in the HLA region

Except the examples above, all CpGs in the class ll genes are located outside of known binding sites of regulatory proteins, therefore, if DNA methylation affects the gene expression, it has to be through modifying the chromatin accessibility.

Until now, most studies exploring HLA class II gene DNA methylation were performed on practically non-polymorphic *HLA-DRA* gene. These studies done in 1980s and early 1990s employed methylation sensitive restriction endonucleases (usually enzymes HhaI and/or HpaII).

Recognition sites of these enzymes present only minor fraction of all CpG sites in the whole *DRA* gene, and with the exception of 2 sites (Hh1 a Hh2, located in position -16 and 56 relative to start codon) are mostly situated far from promoter or other known regulatory areas. One (147) or both (148, 149) of these 2 sites were usually methylation free and their hypomethylation seemed to be important factor allowing for the gene expression (149). Role of DNA methylation in proximal promoter regulation is supported by the fact that artificial methylation of the region represses its function, and addition of DNA methyltransferase inhibitor azacytidine to the promoter construct increases transcription (150).

The association of non-promoter *DRA* gene methylation with expression was not so conclusive, some studies found that demethylation of a particular site/s is associated with genes expression (151, 152), some have found the contrary (153, 154). Again, the authors have found the common theme where most part of the gene was generally highly methylated and this extensive methylation did not preclude the expression (148, 149, 155). However, all these studies analyzed only few CpG sites recognized by the restriction enzyme, located over the thousans of kilobases.

The cells seem to have a mechanism in place to prevent class II genes silencing. Binding of RFX protects HLA class II promoters from DNA methylation and this function is independent of the CIITA presence. This effect was observed in both professional APC and non-APC, in B cell and fibroblast cell lines and in multiple class II promoter types (*DRA*, *DQB*, *DOA*), on multiple CpG sites in the proximal promoter (156). RFX binding is even able to revert promoter DNA methylation (156) and activate methylated *DRA* promoter (150), though the effect is isotype specific – for *HLA-DOA*, epigenetic

modifications induced by the RFX absence are not completely reversed after complementing with RFX. Authors speculate that the reason of promoters being protected from epigenetic silencing is to keep class II genes poised for potential transcription in class II negative cells – precursors of APC, and non-professional APCs (156).

Not much is known about class II allele-specific DNA methylation. We are aware of only two studies that studied methylation of individual alleles of HLA genes. In 1985, Uitterlinden *et al.* found variant DNA methylation between class I MHC haplotypes in rats (157). In 1992, using methylation sensitive restriction endonuclease, Toyoda *et al.* found differences in methylation of *HLA-DQB1* alleles (158). His team examined methylation of CCGG sites closest to the 5' UTR of the gene using sequence-specific probe and pair of restriction enzymes recognizing the same sequence: methylation sensitive HspII and non-sensitive MspI. Due to high polymorphism of *DQB1* gene, examined CCGG sites were not present at the same positions in different alleles (this is also reflected by different size of observed restriction fragments), but even with this limitation, it was obvious that methylation pattern of particular allele seems to be constant—an individual CpG site was either methylated in all examined sequences, or unmethylated in all sequences. The only allele with variant methylation was DR4-associated DQ7 (i.e., *DQB1*03:01* or **03:04*), where only a fraction (4/7) of sequences was methylated.

It is possible that allele-specific methylation of class II genes, even if present outside of proximal promoter or identified regulatory sequences, will translate into a distinct phenotype: for example, there are significant differences in methylation of CpG site located 1487 nucleotides upstream of translation start site of the *HLA-DQB1* gene between monozygotic twins discordant for type 1 diabetes (159), and interestingly, the T1D-specific methylation pattern precedes the start of disease (159).

1.2.4.2 Histone modifications and their connection with DNA methylation

Histones undergo posttranslational modifications on specific amino acid residues in the N-terminal part of the protein. The message carried by modifications depends on the type of modification (acetylation, methylation, phosphorylation, ubiquitination, and more), its position within a nucleosome (which residue on which histone is modified) and on the degree of modification (e.g., mono-, di-, and trimethylation). The marks deposed on histones are specifically recognized by chromatin-remodeling proteins that translate them into the change of chromatin state. The resulting effect on chromatin and on gene expression depends on combination of modifications; genomic context (gene promoter, enhancer, body of the gene, CpG island) and on the pattern of the modification distribution over the region (160).

Histone acetylation is a reversible lysine modification controlled by histone-acetyltransferases (HAT) and histone deacetylases (HDACs), which typically function as transcription co-activators (HAT)

and corepressors (HDAC), respectively. Deregulation of histone acetylation, like deregulation of DNA methylation, can lead to aberrant gene expression and tumorigenesis.

Acetylation of histones is found to a small extent in most part genome as a result of the balance between the activity of HAT and HDAC. Gene activity is controlled by acetylation of histones in its promoter. The active transcription of the gene is associated with increased acetylation of histones H3 and H4. Removal of the acetyl group causes chromatin compaction and repression of gene transcription. A low level of acetylation may induce further epigenetic modifications, including methylation of DNA. At the same time, histone acetylation protects DNA from methylation, and vice versa, DNA methylation prevents histone acetylation. Which of these modifications occurs first is unknown, and there is evidence supporting both claims (128).

Another type of modification is **histone methylation**. Lysines and arginines can be modified by the addition of one to three methyl groups. Histone methylation can be associated with gene activation or repression, depending on the exact modification and its genomic context: H3K4me1 is associated with active promoters and enhancers, H3K4me3 is present in active promoters only and CpG islands (128, 161). H3K27me3 on other side is associated with gene silencing and compacted chromatin (161). Triggering of gene expression is also associated with a decrease in H3K9me3 in the promoter region (128)

DNA methylation is strongly associated with the methylation status of lysine 4 and lysine 9 on the histone H3 (H3K4 and H3K9)—specifically, the presence of DNA methylation is associated with the absence of H3K4 methylation and the presence of H3K9 methylation (128).

The main linkage of histone methylation with DNA methylation are DNMT proteins. DNMT1 through its ADD domain specifically interacts with the N-terminus of H3, but only if it is unmethylated. Thus, it acts as a H3K4 methylation sensor that and in the absence of H3 methylation induces *de novo* methylation of DNA. The question, however, is to what extent this mechanism is functioning in somatic cells, because they express DNMT3L very weakly, if at all. In this case, this function could be performed by a DNMT3a that can recognize unmethylated H3K4 (again through the ADD domain) independently of DNMT1 (129).

The link of H3K9 methylation with DNA methylation is mediated by the above-mentioned UHRF protein which binds not only DNMT1 (and even DNMT3a and 3b) and hemimethylated sites, but also a methylated H3K9 (134).

Interestingly, the distribution of some modifications over the gene is not even, for example, the high levels of H3K36me3 associated with gene expression are found only in some of its exons (134, 162).

The role of link between histone modifications and HLA class II expression is played by a CIITA transcriptional coactivator. This protein not only associates with histone-acetyltransferases, but also acts as

one itself. The importance of this role is underlined by the fact that MHC II genes can be induced by HDAC inhibitor Trichostatin A (TSA) even in the absence of CIITA (64).

1.2.4.2.1 Allele-specific histone modifications and chromatin state

DNA sequence variation affects histone modifications in *cis* as well, as was repeatedly observed on association of polymorphic SNPs with H3K4me1, H3K27ac, H3K4me3 and H3K27me3 levels in the surrounding region: in three individuals examined (AA, AB, BB genotypes), the levels of histone modification in heterozygote lied between the homozygotes values (160). An allele-dependency of chromatin state was observed in proximity (within 50 kb) of the marker SNP, as well as in distal (1 Mbp) region, the interaction mediated through chromatin looping (chromatin in the linking region is not allele-dependent) (160, 163). Study on lymphoblastoid cell lines from 10 individuals identified hundreds of expression quantitative trait loci (SNPs that translate into variant expression of the respective alleles) whose chromatin state correlated with genotype, often on several layers: not only histone modifications, but also Pol II occupancy, DNAse I sensitivity sites and nucleosome positioning; and again in both, distal regulatory sites and gene promoters (164).

The connection between genotype and chromatin state is supported by the fact that the patterns of open chromatin, histone modifications and transcription factor binding are heritable (160, 163), as well as pattern of DNAse I hypersensitive sites and CTCF binding (165). Around 10% of active chromatin sites is allele-specific (165).

Even though the source of inheritance—whether it is genetic, epigenetic, or other—cannot be distinguished by some studies (165), at least some variance seems to originate in genotype, because variable regions of active chromatin marks are enriched for motif-disrupting SNPs (163).

1.2.4.2.2 Allele-specific chromatin modifications of HLA class II genes

As discussed above, allele-specific histone modifications are quite a common feature in the human genome (164), however, there is no information on allele specific histone modifications of class II genes.

The few works are available that provide information on of class II chromatin in healthy individuals and patients with autoimmune conditions: for example, Miao and colleagues explored histone modifications in autoimmune diabetes. They found altered histone methylation in lymphocytes from T1D patients that affected genes relevant to the pathology of type 1 diabetes, including loci upstream of *DQB1*, *DRB1* and *DPB2* gene (166, 167). Some of these changes may be a result, not a cause of the disease, as hyperglycemia in both T1D and T2D increases H3K9 acetylation in promoters of inflammatory genes, *in vitro* as well as *in vivo* (168).

To conclude, HLA alleles behavior is determined by their peptide sequence as well as by their production rate. High polymorphism on the genetic level may translate into variation in the epigenetic

marks (DNA methylation, histone modifications) and consequently into uneven expression of the individual alleles, with an impact on their function in immune system. The aim of our work was to add missing points to this picture and analyze HLA allele-specific DNA methylation and mRNA expression.

2 Aims and Hypotheses

The overall aim of this thesis was to study allele-specific mRNA expression and DNA methylation of the HLA class II genes. The work was divided into two studies, each addressing different aspects of the HLA class II allele expression regulation:

Study A: HLA class II DNA methylation and its effect on the mRNA expression

Hypothesis: DNA methylation can regulate expression of HLA class II genes.

Aim: Determine DNA methylation of individual alleles of the *HLA-DQA1* gene. Analyze inter-allelic differences in DNA methylation. Analyze the relationship between allele DNA methylation and mRNA expression.

Study B: HLA class II mRNA expression (allele- and cell-type-specific expression)

Hypothesis 1: There are differences in HLA class II expression, but are not well described.

Aim 1: Determine the mRNA expression hierarchy of HLA class II *DQA1* and *DQB1* alleles.

Hypothesis 2: Expression hierarchy of HLA class II alleles may differ in the different immune cell types.

Aim 2: Determine expression hierarchy of the *DQA1* and *DQB1* alleles in the whole blood cells, B cells and monocytes.

3 Material and Methods

The study was divided into 2 parts that partially differed in the methodology and approach and thus will be described separately, where applicable.

3.1 Subject of the study

The test group comprised of 158 individuals of European descent. From this group we selected a subset of people for each study based on their HLA genotype and age. The final number of samples analyzed in the studies (which is the number indicated here) is lower than the number of recruited individuals, as we had to exclude several samples because of an insufficient quality of obtained data. All subjects were healthy, i.e., without an autoimmune disease (self-reported).

A test group for the **study A** consisted of 59 healthy volunteers of European descent (18 male, 41 female), 19–39 years old (mean 25.5±3.9). The age restriction of individuals included into the group was chosen to limit the possible confounding effect of age on DNA methylation (137). A test group for the **study B** consisted of 42 healthy volunteers of European descent, 19–69 years old (mean 31.2±12.1).

The experiments were undertaken with the understanding and written consent of each subject. The study methodologies conformed to the standards set by the Declaration of Helsinki. This study was approved by The Ethical Committee of the Third Faculty of Medicine of Charles University in Prague.

3.2 Sample collection

Peripheral blood was collected into EDTA tubes. For RNA isolation, blood was stored at room temperature and processed within 4 hours after collection.

3.3 Material

3.3.1 Chemicals

Most of general chemicals of analytical reagent grade were bought from the following companies: Penta (Prague, Czech Republic), SERVA (Heidelberg, Germany), Sigma Aldrich (St. Louis, Missouri, United States).

The specific chemicals and ready-made kits were bought from the following companies: Applied Biosystems, Fermentas, Invitrogen—now all part of Thermo Fisher Scientific (Waltham, Massachusetts, USA), BD Difco (Franklin Lakes, New Jersey, USA), IDT (Coralville, Iowa, USA), Olerup SSP AB (Stockholm, Sweden), Promega (Fitchburg, Wisconsin, USA), Qiagen (Venlo, Netherlands), Sigma Aldrich

(St. Louis, Missouri, United States), Stratagene (La Jolla, California, USA). The chemicals are listed below, sorted by method.

B lymphocyte and monocyte isolation

Dynabeads® CD19 Pan B (Invitrogen)

Dynabeads® CD14 (Invitrogen)

PBS (Applied Biosystems)

DNA extraction

RCLB (Red Cell Lysis Buffer): 320 mM sucrose, 1% (v/v) Triton X-100, 12 mM Tris-HCl pH=7.5, 5 mM $MgCl_2$

WCLB (White Cell Lysis Buffer): 120 mM EDTA pH=8, 375 mM NaCl

Proteinase K from *Tritirachium album* 30 units/mg (Sigma Aldrich)

SDS; NaCl; 96% ethanol

RNA extraction

QIAamp RNA Blood Mini Kit (Qiagen)

GenElute™ Mammalian Total RNA Miniprep Kit (Sigma Aldrich)

RNA later stabilization solution (Applied Biosystems)

DNA conversion with bisulfite

Epitect® Bisulfite kit (Qiagen)

cDNA synthesis

High Capacity cDNA Reverse Transcription Kit (Applied Biosystems)

HLA Genotyping

Olerup SSP™ genotyping kits (Olerup SSP AB): *Olerup* SSP™ DR low resolution, DQ low resolution, DQA1, DQB1*02, DQB1*03, DQB1*04, DQB1*05, DQB1*06

Taq DNA polymerase (recombinant) 500U, 5U/µl (Fermentas)

PCR, qPCR primers and probes

Primers

Primers for PCR and qPCR were synthetized by IDT. Primers were desalted by the manufacturer. Upon receipt, lyophilized primers were resuspended in PCR-grade H_2O to 100 µM and stored at –20 °C.

Probes

qPCR probes (PrimeTime qPCR Probes) were synthetized, verified by a mass spectrometry and HPLC-purified by IDT. Probes were labelled with 6-carboxyfluorescein (6-FAM) fluorophore at the 5’ end and double-quenched with Iowa Black FQ (IAbFQ) quencher at the 3' end and

internal ZEN quencher at 9 bp from the fluorophore. Upon receipt, lyophilized probes were resuspended in TE buffer (10 mM Tris pH 8.0, 0.1 mM EDTA) to 50 µM and stored at –20 °C.

Ready-made assays

PPIA TaqMan Gene Expression Assay Hs04194521_s1 (Applied Biosystems)

PCR amplification

dNTPs: dATP, dCTP, dTTP, dGTP 100 mM (Fermentas), mixed into equimolar solution and diluted with PCR-grade H_2O to the work concentration of 10 mM

PCR reaction kit (Fermentas): Taq DNA polymerase (recombinant) 500U, 5U/µl 10X; Taq Buffer with $NH_4(SO4)_2$ 10X; 25 mM $MgCl_2$

Water, biotechnology performance certified (Sigma Aldrich)

qPCR amplification

Gene Expression Master Mix (Applied Biosystems)

Water, biotechnology performance certified (Sigma Aldrich)

PCR product purification

QIAquick Gel Extraction kit (Qiagen)

isopropanol

Agarose gel electrophoresis

GelRed® nucleic acid gel stain (Thermo Fisher Scientific)

Bromophenol Blue Loading Solution (Promega)

6x Loading Dye Solution (Fermentas)

Marker pUC19 DNA/MspI (Fermentas)

Electrophoresis for PCR product detection

TBE buffer: 890 mM Tris base, 890 mM boric acid, 20 mM EDTA pH=8

Agarose (Invitrogen)

Electrophoresis for PCR product purification

TAE buffer: 40 mM Tris-acetate, 1 mM EDTA

Low gelling temperature agarose (Sigma Aldrich)

Agar plates preparation

Bacto Agar (BD Difco)

LB Broth base (Invitrogen)

Ampicillin sodium salt (Sigma Aldrich)

X-Gal (Sigma Aldrich)

IPTG; DMSO; X-Gal

3.3.2 Bacterial strain and vectors

Bacterial strains

XL1-Blue *E.coli* (genotype: recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F′ proAB lacIqZΔM15 Tn10 (Tetr)]) (Stratagene)

Vector

pGEM-T Easy Vector (Genotype: Ap^R *oriV* ColE1 *lacZ*, 3015 bp) (Promega)

3.3.3 Software

7000 Sequence Detection Software 1.2.3 (Applied Biosystems)

BioEdit

GeneSnap

GraphPad Prism 3

Vector NTI Advance™ 11

3.3.4 Services

DNA sequencing

DNA sequencing was provided by Macrogen (Seoul, Korea). The service was arranged by Laboratory of DNA Sequencing at Faculty of Natural Sciences, Charles University in Prague

3.4 Methods

3.4.1 DNA extraction

Genomic DNA from peripheral leukocytes was extracted using the salt-extraction (ethanol-precipitation) method (modified (169)). Red blood cells from 0.5 ml of whole blood were lysed by adding 1 ml of RCLB and mixing for 30 s. Samples were spinned at 16 000 g for 3 min, supernatant was discarded, and samples were washed by two rounds of resuspending in 1 ml dH_2O followed by spinning and discarding the supernatant. To lyse white blood cells, sediment was resuspended in 235 µl dH_2O and 80 µl WCLB. Proteins were digested by adding 40 µl of 10% SDS and 15 µl of proteinase K followed by 30 min incubation at 55 °C on carousel. After cooling to the room temperature, proteins were precipitated by adding 100 µl of 6 M NaCl to the solution and shaking vigorously for 15 s. To remove the proteins, samples were spinned at 16 000 g for 6 min, supernatant was transferred to a fresh tube and again spinned at 16 000 g for 3 min. Supernatant was transferred to a fresh tube and after adding 1 ml of 96% ethanol pre-cooled to –20 °C, DNA was precipitated by inverting the tubes. Tubes were incubated in freezer at –20 °C, spinned for 3 min and

supernatant was discarded. DNA was hydrated by 3 min incubation on carousel with 1 ml of 70% ethanol and collected by spinning for 3 min. After discarding the supernatant, DNA was air-dried until translucent (around 30 min) and resuspended in 200 µl of PCR-grade H_2O. DNA was dissolved at 4 °C for 24 h, and after measuring concentration, diluted to 100 ng/µl.

Genomic DNA from the B lymphocytes and monocytes was isolated using the same protocol without the first step (lysis of red blood cells).

3.4.2 HLA genotyping

DNA was diluted to the concentration of 30 ng/µl. *HLA-DRB1*, *HLA-DQA1* and *HLA-DQB1* were typed by polymerase chain reaction with sequence-specific primers (SSP-PCR) using Olerup SSP™ HLA typing kits according to the manufacturer‘s instructions. PCR products were separated by 2% agarose gel electrophoresis in TBE buffer for 50 mins at 5V/cm and visualised by UV transillumination. Genotype was determined based on the presence/absence of specific products according to the interpretation table provided by the manufacturer.

3.4.3 Analysis of DNA methylation

HLA-DQA1 gene promoter methylation was analyzed using bisulfite sequencing. Sodium bisulfite ($NaHSO_3$) deamidates unmethylated C in DNA to U and leaves 5meCpG unchanged. This enables detection of the methylation status of the studied region – after PCR amplification (which replaces U with T) and DNA sequencing, the initially methylated CpG sites remain unchanged and unmethylated sites are shown as CpA or TpG dinucleotides. Since DNA was isolated from whole blood and methylation may vary in individual cell types, the products were transformed into bacteria and individual clones were sequenced.

3.4.3.1 Bisulfite conversion

Genomic DNA was converted by sodium bisulfite using an EpiTect Bisulfite Kit according to the manufacturer’s protocol. Briefly, DNA was converted by incubation with sodium bisulfite at high temperature in the presence of DNA protecting buffer and washed. According to the protocol, DNA concentration in the conversion reaction should be 2 µg or less and we used input concentration of 1000 ng or less, depending on the DNA concentration in the sample. DNA could be eluted into a buffer provided by manufacturer or into water, we chose elution into PCR-grade H_2O.

3.4.3.2 PCR amplification of target region in the bisulfite treated DNA

Genes within the HLA region are highly homologous and bisulfite conversion further decreases complexity of the DNA sequence. To increase amplification specificity, the target segment was amplified by nested PCR.

Primer design

Sequences of *HLA-DQA1* (NCBI reference sequence: NC_000006.11) and *HLA-DQB1* (NC_000006.12) genes were obtained from www.pubmed.org. The sequences of the *HLA-DQA1* alleles were obtained from https://www.ebi.ac.uk/ipd/imgt/hla/align.html using the IMGT/HLA database Sequence Alignment Tool (112). Properties of individual primers and primer pairs were calculated with Vector NTI software.

HLA-DQA1 gene. Because of polymorphism in the *DQA1* promoter region, a separate set of primers had to be designed to amplify the *DQA1*04* alelle. The promoter region amplified by the DQA1*04:01 assay is shorter compared to the original assay, but the omitted area is free of CpG dinucleotides (see also **Figure 4**), therefore it is possible to compare overall methylation of this allele with overall methylation of other promoter alleles. The properties of primers are listed in the **Table 2A.**

HLA-DQB1 gene. Target section of the *HLA-DQB1* gene was a CpG-dense region spanning from intron 1 to intron 2. The region is 1500 bp long and as the resolution limit of the capillary electrophoresis used for reading of the DNA sequence is around 700 bp, the sequence was divided into two parts (marked A and B), each amplified by a specific set of inside primers. Because of polymorphism in the region, a separate set of outside primers had to be designed to amplify the *DQB1*02* and **03* alelles. These primers did not change the position of the analysed region.

Note:primers for amplifying HLA-DQB1 gene were designed here but were used for other projects of our laboratory (170).

PCR amplification

To increase the probability of capturing both alleles and multiple cell types in the sample, three independent reactions were employed to amplify the target segment in the *HLA-DQA1* gene promoter region from bisulfite-treated genomic DNA. Reaction mix for nested PCR reaction1/reaction2 consisted of 1xTaq buffer with $NH_4(SO_4)_2$, 6 mM/4 mM MgCl2, 0.5 mM dNTP mix, 0.8 µM of each primer, 1 U/25 µl of Taq and 1 µl/1.5 µl of DNA solution; and cycling conditions were 95 °C for 5 min, followed by 40/50 cycles of 95 °C for 1 min, 56 °C/65 °C (55 °C/57 °C for the *DQA1*04:01* allele) for 1.5 min, 72 °C for 1 min, and finally 72 °C for 10 min. Primer sequences and properties are listed in **Table 2A**. Presence of PCR products was verified by 2% agarose gel electrophoresis in TBE buffer for 50 mins at 5V/cm and visualisation by UV transillumination.

PCR product purification

PCR products form 2–3 independent PCR reactions for each sample were mixed, separated by 1% low-melting agarose gel electrophoresis in TAE buffer and visualized under UV transilluminator. The products were then purified using QIAquick Gel Extraction Kit according to the manufacturer's instructions.

Table 2. Sequences and properties of bisulfite sequencing primers. (A) Primers used to amplify DQA1 promoter area **(B)** Primers used to amplify DQB1 region spanning from intron 1 to intron 2. Primers amplifying the A and B parts of the gene are marked A and B.

(A)

Assay	Primer name	Sequence (5′→3′)	Length	Tm (°C)	Amplicon	
					Size (bp)	Location
DQA1_all	metF1-DQA	GGTTGTAAGTTAGAATATTTTGAAGGATG	29	63	643	-729 – -87
	metR1-DQA	CAAACCAAACCCTACCAAATCA	22	58		
	metF2-DQA	AGGTTGTTTAGAAATGTTTATTTTTGG	27	59	548	-673 – -126
	metR2-DQA	AAAATCCCCTATAATAACATCTCAATTAC	29	62		
DQA1_04	metF_04-DQA [a]	TTATTTATTTACGAGGTTGTTTAGAAATG	29	53	572	-686 – -115
	metR1_04-DQA	AACTAATTAAAAAACTCCCCTATAATAACAC	31	53		
	metR2_04-DQA	TTTAACAAAAAAATCCCCTAATTATAAC	28	52	501	-686– -186[a]

[a] metF_04-DQA was used as a forward primer in both, first and second, PCR reaction to amplify *DQA1*04* allele of *DQA1* gene.

(B)

Assay	Primer name	Sequence (5′→3′)	Length	Tm (°C)	Amplicon	
					Size (bp)	Location
DQB1_all	metF1-DQB	TAAATTGGTGATTGTTATAGTTTAATTGGAATTTAGT	37	59	1515	909–943
	metR1-DQB	CTCAAAAATCTCCGCCATTAATAATAACCATT	32	61.6		2392–2423
DQB1_02,03	metF1_02,03-DQB	TAATGGGAATTTAGTTTATTAAAGTTAAAAGTTTG	35	56.8	1538	931–965
	metR1_02,03-DQB	ATATATCAAATACCAACCATACAATAATTAAAACAC	36	56.9		2430–2468
DQB1_inA	metF2A-DQB	AGGGTAAATTTAGGTATGGGAAGGTAGGTAT	31	58.6	704	1057–1087
	metR2A-DQB	CTCCAAAACTTCCTTCTAACTATTCCAATACT	32	57.9		1730–1761
DQB1_inB	metF2B-DQB	AGTATTGGAATAGTTAGAAGGAAGTTTTGGAG	32	57.9	613	1730–1761
	metR2B-DQB	AATATCTTATTTCGCAACTATAATTACTAAATACCCTA	38	57.7		2306–2343

Briefly, products were cut out of the gel and agarose was dissolved with QG buffer. DNA was precipitated with isopropanol, bound to the silica gel columns, washed and eluted with PCR-grade water.

3.4.3.3 Cloning

To obtain methylation profile of individual DNA sequences, PCR products were cloned into bacteria and sequencing of individual clones was performed. Briefly, purified PCR product was ligated with the pGEM®-T Easy vector which was subsequently cloned into *E. coli* strain XL-1 Blue. The vector contains gene for an α-peptide of β-galactosidase interrupted with a cloning site and the E.coli strain is ampicillin-resistant, so their combination is suitable for blue-white selection on a medium supplemented with IPTG (isopropyl β-D-1-thiogalactopyranoside, a β-galactosidase inducer), X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, a β-galactosidase substrate that gives a blue product after digestion by the enzyme) and ampicillin: non-transformed bacteria do not form any colonies, bacteria transformed with empty vector / vector with a product form blue/white colonies, respectively.

Bacteria cultivation

E. coli were grown in a liquid or solid LB medium at 37 °C. Culture growth in the liquid media under aerobic conditions in an orbital shaker at 220–280 r.p.m. was monitored by measuring the culture optical density at 590 nm (OD_{590}).

Agar plates preparation

Complex Luria-Bertrani (LB) medium (pH = 7.0–7.2) was routinely used to cultivate *E. coli*. Cultivation plates were prepared according to a modified protocol Sambrook *et al.* (171): 20 g of LB medium and 15–18 g of agar were dissolved in 1 L of water and autoclaved at 121 °C for 15 min. After medium cools down to 65 °C (just before it starts solidifying), 1 ml of ampicillin solution (0.1 g/ml), 1 ml of 3% X-Gal solution (in DMSO) and 1 ml of 100 mM IPTG solution were added. The solution was poured into Petri dishes (cca 25 ml of medium for dish with 9 cm diameter). Once solid, the agar plates were stored upside-down at 4 °C.

Competent bacteria preparation

Competent bacteria were prepared according to a modified $CaCl_2$ method of Sambrook *et al.* (171): Bacteria (an amount covering approximately half of the inoculation loop) were transferred from an LB medium agar plate to a tube with 3 ml LB medium and incubated at orbital shaker at room temperature for 18–20 h or at 37 °C for 16–18 h. 100 ml of LB medium heated to 37 °C was poured into a sterile 500 ml flask, 1 ml of the bacterial culture was added and incubated on a shaker at 37 °C until the OD measured at 590 or 600 nm reached 0.5–0.6. The flask was placed on ice to stop the bacterial growth. The culture was aseptically transferred to sterile, ice-cold 50 ml polypropylene tubes, cells were harvested by spinning at 3 500 g for 5 min and the medium was discarded. Bacteria were washed by resuspending in 25 ml of ice-cold sterile 0.1 M $CaCl_2$, spinned at 3 500 g for 5 min. After discarding fluid, bacteria were resuspended in 25 ml of ice-cold sterile 0.1 M $CaCl_2$ and incubated on ice for 30 min. Cells were harvested by spinning at 3 500 g for 5 min and resuspended in 3 ml of ice-cold sterile 0.1 M $CaCl_2$ with 20% glycerol. The sterile suspension was transferred in 100 µl aliquots into 1.5 ml ice-cold sterile tubes. The cells were snap-frozen by inserting the tubes into ethanol pre-cooled to –70 °C and stored at this temperature.

Product ligation with vector and transformation of competent cells

PCR product was ligated into the vector using pGEM®-T Easy Vector system (Promega) according to the manufacturer's instructions with one modification—the reactions were performed with half volume of all reactants. Briefly, PCR products were ligated into pGEM-T Easy vector (insert to vector ratio was 3:1, i.e., each ligation reaction contained 14 ng of PCR product) and competent *E. coli* XL-1 Blue were heat-shock-transformed with the constructs. Successful transformants were selected based on their color (white, not blue) after cultivation on LB agar plates supplemented with IPTG, X-Gal and ampicillin.

Selection of successful transformants

Presence of the correct insert in the white colonies was verified by a colony PCR. In colony PCR, bacterial cells were used (instead of genomic DNA) as a template in the second reaction of the nested PCR that was used to amplify the original genomic template.

3.4.3.4 Sequencing

Clones whose transformation was verified were prepared for sequencing: a solution of 1% LB agar with ampicillin (0.1 mg/ml) was dispensed in 100 µl aliquots into wells of the 96-well PCR plate. Bacteria were transferred to the plate using the pipette tip, cultivated overnigt at room temperature and then stored in refrigerator until dispatch. The samples were sequenced by Macrogen using universal SP6 and T7 primers whose sequences flank the pGEM®-T Easy vector insertion site. Sequences obtained from the individual clones were processed using BioEdit software.

3.4.3.5 Sequence quality check

Damaged sequences, sequences that recombined during PCR amplification, and sequences where bisulfite conversion efficacy was smaller than 95% were discarded. The methylation status of the remaining 213 *DQA1* promoter sequences from 35 individuals was analyzed. Out of the 213 sequences, 182 covered the whole promoter region studied and 31 lacked methylation data for one or more methylation sites within the region. To analyze a site-specific DNA methylation, data from 182 complete sequences and data from the informative CpG sites of 31 incomplete sequences were used. To analyze an overall methylation, data from 182 full-length sequences were used.

3.4.3.6 Statistical analysis

Overall methylation, i.e., a number of methylated CpG-cytosines per sequence, of individual *DQA1* promoter alleles was compared by two-tailed Mann-Whitney non-parametric test with 95% confidence interval (CI). To compensate for multiple comparisons, the Bonferroni correction was used (10 alleles tested, $p_{corrected(c)} = p_{uncorrected(un)} \times 10$). The interallelic differences in methylation of individual CpG sites were analyzed by a two-tailed Fisher's exact test with a 95% CI and a significance level of $\alpha = 0.05$. With regard to the non-identical number and position of CpG sites in individual alleles, for each site, only the data of alleles containing a particular CpG site (**Figure 4B**) were compared. Obtained p-values were subjected to the Bonferroni correction (11 alleles tested, $p_c = p_{un} \times 11$).

Correlation between allele methylation and expression was analyzed with Spearman's rank correlation analysis.

3.4.4 Analysis of mRNA expression

3.4.4.1 Isolation of B lymphocytes and monocytes

For study B, B lymphocytes and monocytes were isolated from peripheral blood using Dynabeads® CD19 Pan B and Dynabeads® CD14. First, anti-CD14 coated magnetic beads were used to capture monocytes, then, monocyte-depleted blood was incubated with anti-CD19 magnetic beads to capture B

lymphocytes. The rest of the procedure was performed according to the manufacturer's protocol. The cells were either directly used for the RNA extraction or were resuspended in RNA later stabilization solution and stored at –80 °C.

3.4.4.2 RNA extraction

Wherever possible, RNA was isolated from the same blood draw as DNA isolated for bisulfite analysis.

Total RNA from whole blood cells (WBCs) was extracted using QIAamp RNA Blood Mini Kit according to the manufacturer's instructions. Kit uses silica gel coated spin columns to capture DNA and the isolation is performed in a presence of β-mercaptoethanol to inhibit RNAses. Briefly, after disintegration of red and white blood cells (by lysis and shredding, respectively), DNA was precipitated with ethanol, washed, and eluted into 30 µl of RNA-se free water. RNA was either immediately used for reverse transcription or stored at –80 °C.

RNA from B lymphocytes and monocytes was isolated by GenElute™ Mammalian Total RNA Miniprep Kit according to the manufacturer's instructions. The principle was the same as for RNA extraction from the whole blood.

3.4.4.3 Reverse transcription

Total RNA was reverse transcribed with random hexamer primers using High Capacity cDNA Reverse Transcription Kit according to the manufacturer's instructions. The input amount of RNA was 300 ng per RT reaction.

3.4.4.4 Quantification of mRNA expression

3.4.4.4.1 Primers and probes

Due to a high polymorphism of the target region, the assays were designed to quantify whole allele groups instead of individual alleles: DQA1*01, *02, *03, *04, *05 and DQB1*02, *03:01,*03x (amplifies alleles *03:02–03:05 and *04), DQB1*05, *06 assays. "DQA1 total" and "DQB1 total" assays were used to quantify total *DQA1* and *DQB1* mRNA, respectively, regardless of alleles present in the sample; "DRA" assay targeting the *HLA-DRA* gene was used as an endogenous control; and assays amplifying an intronic segment of the *DQA1* gene "DQA1 intron" and *DQB1* gene "DQB1 intron" were used to determine a genomic DNA contamination in the samples. The design of all PCR primers and fluorogenic probes was based on primers and probes used by Fernandez *et al.* (172) and modified according to published sequences at http://www.ebi.ac.uk/ipd/imgt/hla/align.html (4, 84).

After completing the measurements for the study A, a new version of reference genome assembly was released and we found out that the DQA1 assay targets a DQA2 gene as well. As the DQA2 gene was

reported to be expressed in some cell types (63, 173), an additional assay "DQA2" was designed for the study B to quantify mRNA expression of this gene. *DQA2* gene levels were found to be less than 0.01% of *DQA1* gene levels (i.e., at least 14-cycles difference) in all samples. Sequences and properties of all primers and probes are listed in **Table 3.**

3.4.4.4.2 Selecting a classical endogenous control

In addition to the *DRA* gene, a classical endogenous control was used to normalize the expression data as well. The control assay was selected out of 4 genes (*RPS13*, *PPIA*, *ACTB* and *TBP*) based low variation in its expression in 8 samples transcribed from whole blood RNA and on the proximity of its the ΔCt values to the analyzed *DQA1* gene.

3.4.4.4.3 qPCR amplification

All qPCR amplifications were performed in triplicate and contained 200 nM of probe, 300 nM of each primer and 1x Gene Expression Master Mix. Cycling conditions were 50 °C for 2 min and 95 °C for 10 min, followed by 50 cycles of 95 °C for 15 s and 60 °C for 1 min. For quantification, an Applied Biosystems 7500 Fast Real-Time PCR System was used. Amplification curves were visually inspected and outliers (reactions whose slopes strongly differed from the slopes of the same-assay reactions on the same plate) were omitted from further analysis. Values obtained for allele-specific assays were corrected according to the assays efficacy and subsequently according to the genomic DNA content of the sample to reflect proportions in the RNA component of the sample only.

3.4.4.5 Analysis of allele-specific mRNA expression

3.4.4.5.1 Measurement quality check

Relative expression of *DQA1* alleles was determined with reference to the amount of *HLA-DRA* mRNA. To ensure reliability of results, only samples where the ratio (allele 1 + allele 2 mRNA)/(total *DQA1* mRNA) belonged into the <0.85; 1.15> interval were included into the subsequent analyses.

3.4.4.5.2 Statistical analysis

Relative expression between all possible pairs of alleles was compared by a two-tailed Mann-Whitney test with 95% CI. All p-values were subjected to the Bonferroni correction for multiple comparisons (9 alleles tested, $p_c = p_{un}*9$). Out of 43 subjects included into the expression study, 4 were heterozygous for the *DQA1* alleles recognized by the same allele-specific assay, and therefore expression of individual alleles in these subjects could not be analyzed. In 4 homozygous individuals included to our study, both *DQA1* gene copies were considered to express mRNA equally, therefore the detected allele mRNA level in homozygotes represented double the amount of mRNA from a single gene copy and was divided by 2. In

Table 3. Sequences and properties of RT-qPCR primers and probes. The fluorogenic probes are FAM-labelled at the 5' end and double-quenched by Iowa Black at the 3' end and by ZEN internal quencher at 9 nucleotide distance from the dye. Efficacy of the DQA2 assay was not determined (n.d.).

Assay	Primer name	Sequence (5'→3')	Amplicon length (bp)	location	assay efficacy (study A)	assay efficacy (study B)
DQA1*01	DQA01F	GAAGGAGACTGCCTGGCG	106	exon 2	1.02	1.013
	DQA01R	CATGATGTTCAAGTTGTGTTTTGC				
	DQA01FAIB	CCTGCGGGTCAAAACCTCCAAATTTG				
DQA1*02	DQA02F	TTACGGTCCCTCTTGCCAGTT	124	exon 2	0.98	0.980
	DQA02R	TTGCGGGTCAAATCTAAGTCTGT [a]				
	DQA0203FAIB	CCACATAGAACTCCTCGTCTCCATCAAATTCAT				
DQA1*03	DQA03F	GGTCCCTCTGGGCAGTACAG	127	exon 2	0.97	0.989
	DQA03R	CAAATTGCGGGTCAAATCTTCT [a]				
	DQA0203FAIB	CCACATAGAACTCCTCGTCTCCATCAAATTCAT				
DQA1*04	DQA04F	GTACACCCATGAATTTGATGGAGAC	154	exon 1	0.98	0.956
	DQA04R	CAGGATGTTCAAGTTGTGTTTTGTC				
	DQA0405FAIB	ACTGTCTGGTGTTTGCCTGTTCTCAGACAA [a]				
DQA1*05	DQA05F	GATGAGCAGTTCTACGTGGACCT	152	exon 2	0.95	0.956
	DQA05R	GTAGAGTTGGAGCGTTTAATCAGAC				
	DQA0405FAIB	ACTGTCTGGTGTTTGCCTGTTCTCAGACAA [a]				
DQA1total	DQAtotalF	TACAGCTCAGAACAGCAACTGC	126	exon 1	1	0.998
	DQAtotalR	CCCACAATGTCTTCACCTCCA				
	DQAtotalFAIB	CTTTGTTTAGGATCATCCTCTTCCCAAGGC				
DQA1intron	DQAintrF	GTTGCCCGTTTCTTTCTCTCA [a]	80	intron 1	1	0.999
	DQAintrR	TGGACTCCTTTACCCACTCCC [a]				
	DQAintrFAIB	ACCTGTGCCAGTTCCCATGTGGAAAT				
DQA2	DQA2F	CAGCTCAGAGCAGCAACTGC	164	5'UTR–exon2	–	n.d.
	DQA2R	AGACTGGTAGAAGTTCACACCATAG				
	DQAtotalFAIB	CTTTGTTTAGGATCATCCTCTTCCCAAGGC				
DQB1*02	DQB02F	GCAACTGTGACCTTGATGCTGT	163	exon 1–2	–	1.008
	DQB02R	CGATCTCTTCTCGGTTATAGATGCT				
	DQB0203FAIB	ATCCTCGGGAGAGTCTCTGCCCTCA				
DQB1*0301	DQB0301F	GATGCTGTCGATGCTGAGCA	140	exon 1–2	–	1.011
	DQB0301R	CTCGGTTATAGATGTATCTGGTCACATA				
	DQB0203FAIB	ATCCTCGGGAGAGTCTCTGCCCTCA				
DQB1*03x	DQB03xF	GCCTTCGGGTAGCAACTGTG	175	exon 1–2	–	0.976
	DQB03xR	GCGTACTCCTCTCGGTTATAGATGTA				
	DQB0203FAIB	ATCCTCGGGAGAGTCTCTGCCCTCA				
DQB1*05	DQB05F	CGTCTCAATTATGTCTTGGAAGAAGT	110	5'UTR–exon1	–	1.052
	DQB05R	CTCTGCCCTCAGCCAGTGA				
	DQB0506FAIB	CTTCGGGTAGCAACTGTCACCTTGATGCT				
DQB1*06	DQB06F	CTTCGTCTCAGTTATGTCTTGGAAG	116	5'UTR–exon1	–	1.026
	DQB06R	GAGTCTCTGCCCTCAGCCTGTA				
	DQB0506FAIB	CTTCGGGTAGCAACTGTCACCTTGATGCT				
DQB1total	DQBtotalF	CCTTCGTCTCA**G**TTATGTCTTGGA	172	exon 1–2	–	1
	DQBtotalR	GTCCCGTTGGTGAAGTAGCAC				
	DQBtotalFAIB	CTGAGGGCAGAGACTCTCCCGAGGAT				
DQB1intron	DQBintrF	GGATGATGCCCACTTTGTGC	172	intron 4	–	1.010
	DQBintrR	ACAGAACTTCAGCTTGATGCAGAT				
	DQintrFAIB	CCACTAGCAGCCTCTTTCAGTCACTGGAA				
DRA	DRAF	GGACAAAGCCAACCTGGAAA [a]	120	exon 2–3	1	1
	DRAR	AGGACGTTGGGCTCTCTCAG [a]				
	DRA_FAIB	CAACTATACTCCGATCACCAATGTACCTCCAGAG				

[a] Primers adopted from the sequences published by Fernandez *et al.*, 2003.

the remaining 35 heterozygous individuals, altogether 22 allelic combinations were found. Two *DQA1*03* group alleles were found in our set—*DQA1*03:01* and **03:03* (as a part of two different *DRB1*04-DQA1*03-DQB1*03* haplotypes), however, since both these alleles are linked to the same promoter allele, *QAP 3.1*, their expression was analyzed together as a *DQA1*03* allele group.

3.4.4.5.3 Detection of genomic DNA contamination

The content of genomic DNA in the samples was determined by an assay amplifying a nonpolymorphic intronic segment (172) of the *DQA1* gene and *DQB1* gene. Genomic DNA contamination was found to be 0–14% (mean 5.2 ± 3.6%) in the whole blood samples in study A; and 0–19% (mean 4.7 ± 3.8%) for *DQA1* gene and 0–32% (mean 9.8 ± 7.3%) for *DQB1* gene in the study B. We were unable to identify a reason for discrepancy between assessment of DNA content by *DQA1* and *DQB1* intronic assay.

3.4.4.5.4 Verifying assay specificity

Specificity of each allele-specific assay was verified by using cDNA of individuals carrying off-target allelic groups as a template in amplification reaction. The proportion of the off-target amplicons was found to be less than 0.02% of target alleles amplification (=at least 12 cycle difference) for all assays used.

3.4.4.5.5 Assessment of relative amplification efficiencies

To determine the efficacy of allele-specific assays in **study A**, DQA1 total assay has been assigned efficacy $E_{total} = 1$, and mean Ct of DQA1 total assay was compared with mean Cts of DQA1*01, *02, *03, *04, *05 assays in all available individuals homozygous for a given DQA1 allele. There were no DQA1*04 homozygotes in our experimental set, therefore the DQA1*04 assay efficacy was calculated from the measurements in individuals heterozygous for this allele and an allele with already assessed efficacy. Only triplicates with overall SD < 0.1 were included into efficacy calculations. DRA assay efficacy (E_{DRA}) was determined by using 10^0 to 10^{-4} serial dilutions of input cDNA as a template for amplification in both, DRA and DQA1 total assays. Standard curves were constructed, E_{total} was assigned a value of 1, and E_{DRA} relative to E_{total} was assessed by comparing the slopes. E_{DRA} was found to be the practically same as E_{total} and therefore was also assigned a value of 1.

To determine the efficacy of allele-specific assays in **study B**, DQB1 total assay was assigned efficacy $E_{totalB} = 1$, and mean Ct of DQB1 total assay was compared with mean Cts of DQB1*02, *03:01, *03x, *05, and *06 in all available homozygotes for a given *DQB1* allele. Then, for each assay, the assays efficacy was calculated from the data of individuals heterozygous for this allele and also for the allele with already assessed efficacy. The assays efficacy was determined as an average of values obtained from "homozygote" and "heterozygote" calculations. If no homozygote for given allele was present in a sample, we used only "heterozygote" data as efficacy of the assay. Relative efficacies of "DQB1 total"(E_{totalB}), "DQB1 intron", "DQA1 total" and "DQA1 intron" were determined by using 2^0 to 2^{-7} serial dilutions of input cDNA as a template for amplification in all 4 assays. Standard curves were constructed, E_{totalB} was assigned a value of 1, and efficacy of the three assays relative to E_{totalB} was assessed by comparing the slopes. Then, efficacy of

DQA1 allele-specific assays was determined by comparing mean Ct of DQA1 total assay with mean Cts of allele-specific assays in a same way as it was done for the *DQB1* gene. E_{DRA} was previously in study A found to be practically the same as E_{totalA}; therefore, it was also assigned a value of 1. PPIA assay efficacy was determined by using 10^0 to 10^{-4} serial dilutions of input cDNA as a template for amplification in both, PPIA and DQA1 total assays and comparing the standard curves slopes for both assays. Efficacy of PPIA assay was 1.001. For details of all efficiency calculations, see the **Appendix**.

4 Results

The parts of this chapter are based on published articles (174, 175). A subset of data on *DQA1* promoter methylation was previously reported in my diploma thesis (176). DNA methylation and mRNA expression measurements from the study A were used to provide data for the group of healthy controls in the study of DNA methylation and mRNA expression of *HLA-DQA1* alleles in type 1 diabetes mellitus (174).

4.1 HLA genotyping

Results of genotyping of 158 individuals included into this study are summarized in **Table 4**. HLA genotyping results of the Czech subjects from this study combined with genotyping results of subjects from other studies conducted at the Department of Medical Genetics were published (119) and included into the public repository of HLA allele and haplotype frequencies www.allelefrequencies.net (177). Hardy-Weinberg equilibrium (HWE) was analyzed for haplotypes and for individual genes with χ^2 test. Alleles/haplotypes whose total number in the study group was less than 5 included into a common category „rare“. All characteristics were in HWE.

Table 4. **DRB1-DQA1-DQB1 haplotypes and their frequencies in the study population.**
n - number of haplotype carriers; f - haplotype frequency

DRB1*	DQA1*	DQB1*	n	f (%)	DRB1*	DQA1*	DQB1*	n	f (%)
11	05:05	03:01	50	15.9	04	03:03	03:02	4	1.3
15	01:02	06:02	39	12.4	13	01:02	06:09	3	1.0
01	01:01	05:01	38	12.1	10	01:05	05:01	2	0.6
07	02:01	02:02	31	9.8	03	05:05	03:01	1	0.3
03	05:01	02:01	29	9.2	04	03:01	03:01	1	0.3
13	01:03	06:03	23	7.3	04	03:01	03:05	1	0.3
04	03:01	03:02	20	6.3	04	03:03	02:02	1	0.3
16	01:02	05:02	13	4.1	04	03:03	03:04	1	0.3
12	05:05	03:01	9	2.9	09	03:01	03:03	1	0.3
13	05:05	03:01	9	2.9	10	01:05	06:02	1	0.3
08	04:01	04:02	8	2.5	11	01:03	06:03	1	0.3
13	01:02	06:04	8	2.5	13	01:03	03:01	1	0.3
14	01:04	05:03	7	2.2	13	01:02	05:01	1	0.3
07	02:01	03:03	6	1.9	15	01:03	06:01	1	0.3
04	03:03	03:01	5	1.6					

4.2 Determining *DQA1* promoter (*QAP*) alleles

The *DQA1* promoter (*QAP*) alleles identity was determined based on their sequence (study A) and haplotype association (study B). The nomenclature was taken from Brunnler *et al.* (178) and our previous work (176). Each *DQA1* allele is associated with one *QAP* allele, with the exception of *DQA1*01:02* which is associated with 3 different promoters (*QAP1.2K, 1.2L, 1.4*), depending on haplotype. *DQA1* alleles and their linked promoter alleles are listed in **Table 5** as a part of the respective HLA class II haplotype (178).

4.3 DNA methylation of *HLA-DQA1* promoter region (Study A)

Methylation status of 213 *HLA-DQA1* promoter (*QAP*) sequences from 35 individuals was analyzed. Out of these 35 subjects, mRNA expression data were available for 18 individuals. Out of them, 14 had both expression and DNA methylation analyses done on the RNA and DNA isolated from the same blood draw.

4.3.1 Methylation of individual CpG sites does not differ between alleles

The only difference in DNA methylation of individual CpG sites found to be significant after correction was between the *DQA1*01:01* and *DQA1*01:02* (DR15-linked) promoters at site −508 (p_{un} = 0.0043/p_c = 0.0473, RR CI = 0.9436 to 10.57). However, RR CI contained value 1 and therefore we could not reject the possibility that the association observed is only due to chance. No other differences in DNA methylation of any CpG site between any 2 alleles were observed (data not shown). The average methylation of individual CpG sites is shown in **Figure 4A**.

4.3.2 Overall methylation—the most methylated alleles are *DQA1*02:01* and **04:01*

Because of the low number (3) of *DQA1*01:04* promoter allele sequences, we did not include it into methylation analysis. Promoters of *DQA1*02:01* and *DQA1*04:01* were significantly more methylated than most of the other alleles (namely, *DQA1*01:01, *01:02* (DR16-linked), **01:03, *05:01, *05:05*); promoter of *DQA1*04:01* was also methylated more than promoter of *DQA1*03*, and methylation of *DQA1*01:02* (DR13-linked) promoter was higher than that of *DQA1*01:01*. Overall methylation ranged from 7.6 methylcytosines on average per promoter for *DQA1*01:03* allele to 10.5 methylcytosines per promoter in *DQA1*02:01* and 10.8 methylcytosines in *DQA1*04:01*. Overall methylation of individual alleles is depicted in **Figure 4A,** results of statistical analysis are shown in **Table 6.**

.

Table 6. Differences in overall methylation of *DQA1* promoter alleles. For each sequence, the total number of methylated CpG sites in the region examined (i.e. overall methylation) was counted. According to the polymorphisms present on the sequencing read, sequences were classified into appropriate allelic groups and differences in overall methylation between allelic groups were tested. Uncorrected p-values of all comparisons performed are shown. p-values that remained statistically significant after correction are highlighted in bold.

promoter allele of DQA1

	***01:01** (n=45)	***01:02**[a] (n=8)	***01:02**[b] (n=16)	***01:02**[c] (n=14)	***01:03** (n=16)	***02:01** (n=17)	***03** (n=5)	***04:01** (n=30)	***05:01** (n=10)	***05:05** (n=18)
***01:01**	-	NS	0.0412 [d]	**0.0041** [e]	NS	**<0.0001** [f]	NS	**<0.0001** [f]	NS	NS
***01:02**[a]	-	-	NS	NS	NS	**0.0027** [d]	NS	0.0011 [d]	NS	NS
***01:02**[b]	-	-	-	NS	0.0364 [d]	NS	NS	NS	NS	0.0217 [d]
***01:02**[c]	-	-	-	-	0.0077 [d]	NS	NS	NS	0.0092 [d]	0.0069 [d]
***01:03**	-	-	-	-	-	**0.0001** [f]	NS	**<0.0001** [f]	NS	NS
***02:01**	-	-	-	-	-	-	0.0086 [d]	NS	**0.0002** [f]	**<0.0001** [f]
***03**	-	-	-	-	-	-	-	**0.0047** [e]	NS	NS
***04:01**	-	-	-	-	-	-	-	-	**<0.0001** [f]	**<0.0001** [f]
***05:01**	-	-	-	-	-	-	-	-	-	NS

[a] Linked with DR16 (*QAP 1.2K*), [b] DR15 (*QAP 1.2L*), [c] DR13 (*QAP 1.4*).
[d] p_c = NS, [e] $p_c < 0.05$, [f] $p_c < 0.005$ by two-tailed Mann-Whitney test with 95% CI and after Bonferroni correction.

Table 5. Analyzed *DQA1* promoter (*QAP*) alleles as a part of *DRB1-QAP-DQA1-DQB1* haplotype.

DRB1	QAP [a]	DQA1	DQB1
01	**1.1**	**01:01**	05:01
16	**1.2K**	**01:02**[b]	05:02
15	**1.2L**	**01:02**[b]	06:02
13	**1.3a**	**01:03**	06:03
14	**1.3b**	**01:04**	05:03
13	**1.4**	**01:02**[b]	06:04, 06:09
07	**2.1**	**02:01**	02:02, 03:03
04	**3.1**	**03**	03, 02:02
11, 12	**4.1A**	**05:05**	03:01
03	**4.1B**	**05:01**	02:01
08	**4.2**	**04:01**	04:02

[a] The available nomenclature of *QAP* alleles (178) did not distinguish all alleles present in our samples, so we split some of the existing alleles into 2 groups and used our own notation to name them. Based on the sequence differences, allele *QAP 1.2* was split into groups *QAP 1.2K* and *QAP 1.2L*, *QAP 1.3* into groups *QAP 1.3a* and *1.3b*, and *QAP 4.1* into *QAP 4.1A* and *4.1B*. These "new" alleles differing in their sequence were also parts of different HLA-haplotypes.

[b] Allele *DQA1*01:02* is associated with 3 different promoters (in 3 different haplotypes).

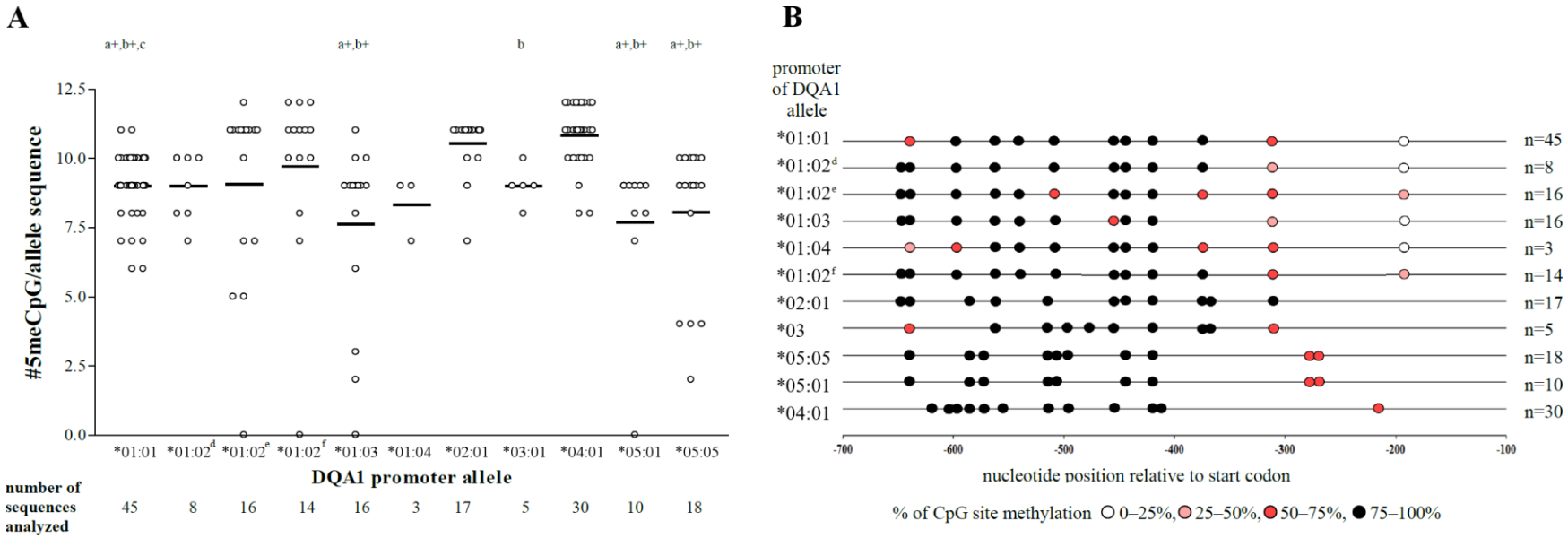

Figure 4. *DQA1* promoter DNA methylation. According to the polymorphisms present on the sequencing read, sequences were classified into appropriate allelic groups. **(A) Overall methylation of *DQA1* promoter alleles.** For each sequence, the total number of methylated CpG sites in the region examined (i.e. overall methylation) was counted. Each point in the graph represents the number of methylated CpGs per 1 sequence. The horizontal line indicates mean methylation level of each allele. **(B) CpG methylation status of individual CpG sites of *DQA1* promoter alleles**.The matrix represents an amplicon that contains 9–12 CpG sites (number depends on promoter allele identity) obtained for 500 bp region overlapping promoter of the *DQA1* gene. Each colour-coded circle within the matrix represents one CpG site and its colour indicates the percentage of sequences that have the cytosine at the given site methylated. Unmethylated–low methylated sites are indicated by open circles. No interallelic difference in methylation of individual CpG sites was found to be significant after Bonferroni correction.
n – number of sequences analyzed
[a] [(a+)] Allele promoter methylation significantly, $p < 0.05$ (highly significantly, $p < 0.005$) lower than that of *DQA1*02:01* allele.
[b] [(b+)] Allele promoter methylation significantly, $p < 0.05$ (highly significantly, $p < 0.005$) lower than that of *DQA1*04:01* allele.
[c] Allele methylation significantly lower than that of *DQA1*01:02* (DR13-linked) allele.
[d] Linked with DR16 (*QAP 1.2K*), [e] DR15 (*QAP 1.2L*),[f] DR13 (*QAP 1.4*)

4.4 mRNA expression of HLA class II genes *DQA1* and *DQB1*

4.4.1 mRNA expression of *DQA1* alleles in whole blood cells (Study A)

We analyzed mRNA expression data from 43 individuals. Data included 10 *HLA-DQA1* alleles linked with 12 different promoter alleles. Since allele *DQA1*01:02* is linked to 3 different *DQA1* promoters in 3 different haplotypes, we treated it as 3 different alleles (DR13-, DR15-, and DR16-linked *DQA1*01:02* allele). *DQA1* alleles and their linked promoter alleles are listed in **Table 5** as a part of the respective HLA class II haplotype (178).

Because of a low number of sequences obtained (1, 2 and 1, respectively), expression data of *DQA1*01:02* (DR16-linked), *DQA1*01:04*, and *DQA1*01:05* alleles were not included into the analysis. Average mRNA expression levels of individual alleles normalized to *DRA* varied over 3-fold (0.07 for the least expressed *DQA1*05:05* and 0.23 for the most expressed *DQA1*03*). Statistical analysis showed that *DQA1*03* alleles were significantly overexpressed compared to most other alleles (*DQA1*01:01*, **01:02* (DR13-linked), **01:03*, **02:01*, **05:01*, **05:05*), and *DQA1*05:05* allele was expressed less than *DQA1*01:03*, **02:01*, **04:01* alleles; the *DQA1*01:03* allele was expressed more than *DQA1*01:01* allele (**Figure 5**).

To analyze the differences in more detail, we calculated the ratio between mean expression levels of all possible allele pairs. For alleles whose mean expressions proved to differ significantly in the previous analysis, the ratio ranged from 1.4 (*DQA1*02:01*/*DQA1*05:05*) to 3.6 (*DQA1*03*/*DQA1*05:05*) (**Table 7**). The allele expression hierarchy was preserved in the heterozygous individuals (**Table 8**). In heterozygotes, *DQA1*03* allele was always expressed more than the other allele, on the other side *DQA1*05:05* allele was always the less expressed one (with a single exception where *DQA1*01:05*/**05:05* expression ratio was 1.09). In single heterozygous individuals we observed distinct *DQA1*01:03* > *DQA1*02:01*,**05:01*, **05:01*, and *DQA1*01:02* (DR13-linked) > *DQA1*05:01* relationship (allele ratio 1.3–1.7); and less pronounced (allele ratio close to one) *DQA1*02:01* > *DQA1*01:02* (DR15-linked) and *DQA1*02:01* > *DQA1*05:01* relationship. As these observations were done in single individuals only, no further conclusions regarding the hierarchy of allele expression could be made.

We also wondered whether and how the presence of specific *DQA1* alleles in an individual's genotype affects total *DQA1* mRNA expression. An inter-individual variation in the allele expression of a particular allele was low and it seemed that an allele mRNA expression level tends to stay the same irrespective of the identity of the other allele present in a heterozygous combination. If this is the case, one should be able to calculate the total relative *DQA1* mRNA level in the subject only by adding up the known values of a relative expression of the alleles carried by the subject. Indeed, we observed that the amount of total *DQA1* gene

mRNA seen in individual samples (**Table 9A**) followed the "theoretical" *DQA1* total mRNA level, which was calculated as a sum of mean expression of alleles present in a sample (**Table 9B**).

4.4.2 No correlation between allele promoter DNA methylation and expression

Next, we aimed to see if there is a correlation between allele promoter DNA methylation and its expression. For each allele, mean allele mRNA expression was plotted against its mean methylation (mean number of methylated cytosines per allele). For the analysis, a *DQA1* promoter was divided into multiple subregions, each ranging from position 0 to a promoter CpG site (0 to -193, 0 to -215, ...). For each subregion, relationship was analyzed with Spearman's rank correlation test. There was no correlation between a number of methylated cytosines and allele expression in none of the regions (data not shown).

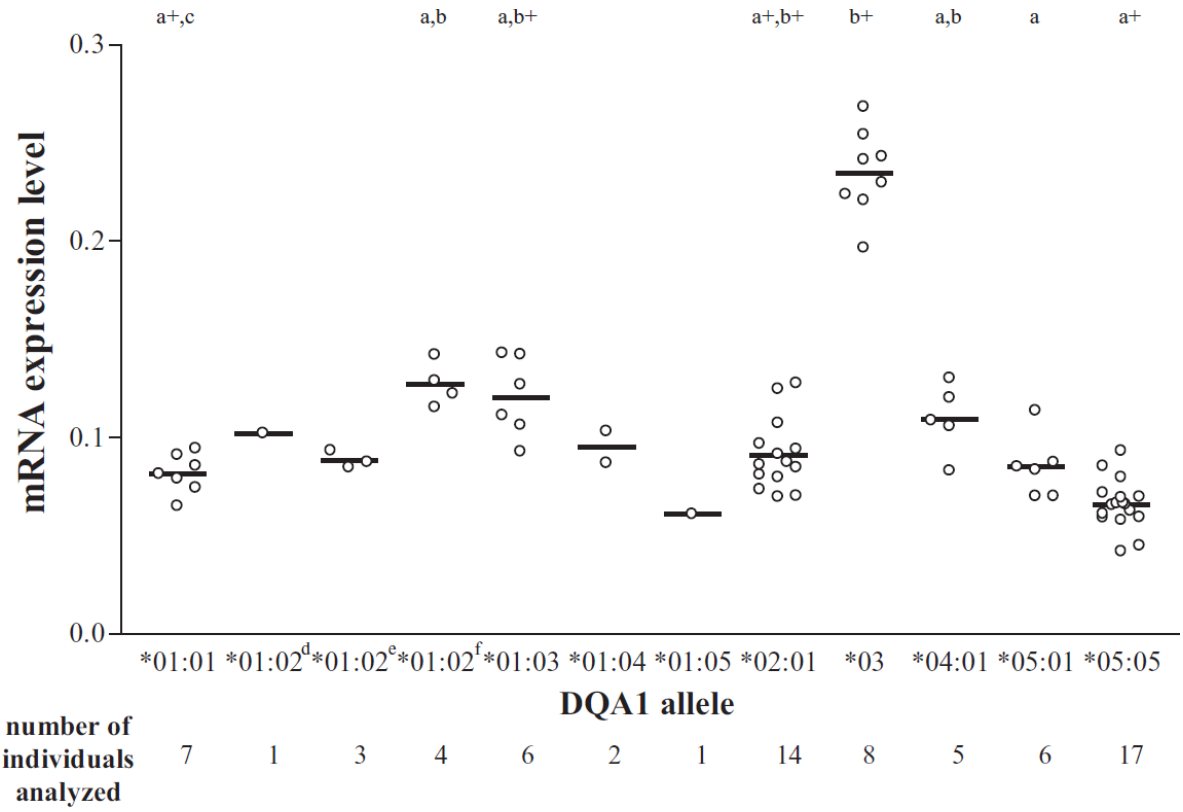

Figure 5. mRNA expression levels of *DQA1* alleles. Expression of *DQA1* alleles was determined by RT-qPCR with allele-specific primers and the mean of 3 independent measurements was used for the analysis. Each point in the graph represents normalized value for 1 individual and 1 allele. The horizontal line indicates mean relative expression of the allele.

[a(a+)] Allele expression significantly, $p < 0.05$ (highly significantly, $p < 0.005$) lower than that of *DQA1*03* allele.
[b(b+)] Allele expression significantly, $p < 0.05$ (highly significantly, $p < 0.005$) higher than that of *DQA1*05:05* allele.
[c] Allele expression significantly lower than that of *DQA1*01:03* allele.
[d] Linked with DR16 (*QAP 1.2K*), [e] DR15 (*QAP 1.2L*), [f] DR13 (*QAP 1.4*).

Table 7. Mean expression ratio of *DQA1* alleles. For each allele, the mean relative expression level was calculated (mean allele expression). To obtain the expression ratio of two alleles, the mean expression of one allele (in this table in the title row) was divided by the mean expression of the second allele (title column). Data for allele pairs whose expression proved to differ significantly (**Figure 5**) are highlighted in bold.

		DQA1 allele								
		***01:01**	***01:02[a]**	***01:02[b]**	***01:03**	***02:01**	***03**	***04:01**	***05:01**	***05:05**
	mean allele expression	0.081	0.088	0.127	0.120	0.091	0.235	0.109	0.085	0.066
	(SEM)	(0.004)	(0.003)	(0.006)	(0.008)	(0.005)	(0.008)	(0.008)	(0.006)	(0.003)
DQA1 allele	***01:01**	1	1.08	1.57	**1.48**	1.12	**2.88**	1.34	1.04	0.81
	***01:02[a]**	0.92	1	1.44	1.36	1.03	2.67	1.24	0.97	0.75
	***01:02[b]**	0.64	0.69	1	0.94	0.72	**1.85**	0.86	0.67	**0.52**
	***01:03**	**0.68**	0.73	1.06	1	0.76	**1.96**	0.91	0.71	**0.55**
	***02:01**	0.89	0.97	1.40	1.32	1	**2.58**	1.20	0.93	**0.73**
	***03**	**0.34**	0.37	**0.54**	**0.51**	**0.39**	1	**0.46**	**0.36**	**0.28**
	***04:01**	0.74	0.81	1.17	1.10	0.83	**2.16**	1	0.78	**0.61**
	***05:01**	0.95	1.04	1.49	1.41	1.07	**2.76**	1.28	1	0.78
	***05:05**	1.23	1.33	**1.92**	**1.82**	**1.38**	**3.56**	**1.66**	1.29	1

[a] Linked with DR15 (*QAP 1.2L*), [b] DR13 (*QAP 1.4*).

Table 8. Mean expression ratio of *DQA1* alleles in heterozygotes. For each heterozygous individual in our sample, the expression ratio of the two alleles was calculated by dividing expression of one allele (in this table listed as first) by expression of the second allele (listed as second). When data from multiple individuals with the same allelic combinations were available, the mean value for each combination was calculated, and used.

DQA1 allele combination		expression ratio (allele1/allele2)			DQA1 allele combination		expression ratio (allele1/allele2)
(allele1, allele2)	n	mean	range	SD	(allele1, allele2)	n	
*01:01, *05:05	2	1.32	1.10–1.54	0.22	*01:02[a], *05:05	1	1.54
*01:02[c], *05:05	2	1.82	1.80–1.83	0.02	*01:02[b], *05:05	1	1.34
*01:03, *05:05	2	1.73	1.67–1.80	0.06	*01:02[c], *05:01	1	1.65
*01:04, *05:01	2	1.16	1.12–1.21	0.05	*01:03, *02:01	1	1.59
*02:01, *01:01	3	1.10	1.02–1.23	0.09	*01:03, *05:01	1	1.27
*02:01, *05:05	3	1.40	1.23–1.61	0.16	*01:05, *05:05	1	0.92
*03, *01:03	2	2.33	1.76–2.90	0.57	*02:01, *01:02[b]	1	1.20
*03, *05:05	2	4.84	4.40–5.28	0.44	*02:01, *05:01	1	1.08
*04:01, *05:05	3	1.85	1.73–1.92	0.08	*03, *01:01	1	2.44
*04:01,*02:01	2	1.12	1.04–1.19	0.07	*03, *01:02[b]	1	2.88
					*03, *01:02[c]	1	1.70
					*03, *02:01	1	3.01

[a] Linked with DR16 (*QAP 1.2K*), [b] DR15 (*QAP 1.2L*), [c] DR13 (*QAP 1.4*).

Table 9. Total relative *DQA1* expression. Expression data of *DQA1* alleles were normalized to the *DRA* expression level. **(A) Total relative *DQA1* expression – measured**. The graph shows the sum of expression levels of both alleles (*DQA1* allele 1 and *DQA1* allele 2) present in an individual. In heterozygotes whose alleles' expression is indiscernible by our approach (i.e. *DQA1*01:0x/*01:0y* and *DQA1*05:01/*05:05* heterozygotes), normalized *DQA1* total expression value was used. If data from multiple individuals with given allelic combinations were available, the mean value for each combination is shown. Dash (–) marks allele combinations that are not present in our sample. **(B) Total relative *DQA1* expression – predicted.** For each allele, the mean relative expression level was calculated (allele mean). The mean expression levels of two alleles were added up to calculate predicted total *DQA1* mRNA level. Data for allelic combinations not present in our study sample are denoted in smaller font. The background shade marks difference between the predicted and measured values: □ < 10%, □ < 20%, □ < 30%.

A

DQA1 allele 2 \ DQA1 allele 1	***01:01**	***01:02[a]**	***01:02[b]**	***01:02[c]**	***01:03**	***01:04**	***01:05**	***02:01**	***03**	***04:01**	***05:01**	***05:05**
***01:01**	0.18	–	–	–	0.28	0.19	–	0.17	0.32	–	–	0.14
***01:02[a]**		–	–	–	–	–	–	–	–	–	–	0.17
***01:02[b]**			–	–	–	–	–	0.18	0.33	–	–	0.16
***01:02[c]**				–	–	–	–	–	0.38	–	0.18	0.19
***01:03**					–	–	–	0.18	0.36	–	0.19	0.22
***01:04**						–	–	–	–	–	0.17	–
***01:05**							–	–	–	–	–	0.13
***02:01**								0.16	0.34	0.20	0.18	0.18
***03**									–	–	–	0.25
***04:01**										–	–	0.17
***05:01**											0.17	0.14
***05:05**												–

B

DQA1 allele 2	**DQA1 allele 1**	***01:01**	***01:02[a]**	***01:02[b]**	***01:02[c]**	***01:03**	***01:04**	***01:05**	***02:01**	***03**	***04:01**	***05:01**	***05:05**
	allele mean	0.081	0.102	0.088	0.127	0.120	0.091	0.061	0.095	0.235	0.109	0.085	0.066
***01:01**	0.081	0.16	0.18	0.17	0.21	0.20	0.18	0.14	0.17	0.32	0.19	0.17	0.15
***01:02[a]**	0.102		0.20	0.19	0.23	0.22	0.20	0.16	0.19	0.34	0.21	0.19	0.17
***01:02[b]**	0.088			0.18	0.22	0.21	0.18	0.15	0.18	0.32	0.20	0.17	0.15
***01:02[c]**	0.127				0.25	0.25	0.22	0.19	0.22	0.36	0.24	0.21	0.19
***01:03**	0.120					0.24	0.22	0.18	0.21	0.36	0.23	0.21	0.19
***01:04**	0.095						0.19	0.16	0.19	0.33	0.20	0.18	0.16
***01:05**	0.061							0.12	0.15	0.30	0.17	0.15	0.13
***02:01**	0.091								0.18	0.33	0.20	0.18	0.16
***03**	0.235									0.47	0.34	0.32	0.30
***04:01**	0.109										0.22	0.19	0.18
***05:01**	0.085											0.17	0.15
***05:05**	0.066												0.13

[a] Linked with DR16 (*QAP 1.2K*), [b] DR15 (*QAP 1.2L*),[c] DR13 (*QAP 1.4*).

4.4.3 mRNA expression of HLA class II genes *DQA1* and *DQB1* in B lymphocytes and monocytes (study B)

After the initial study A, we decided to explore the expression differences between the alleles more thoroughly: we included a *DQB1* gene to the study and compared the expression between the different cell types.In total, the transcription rates in 42 individuals were analyzed. Data included 10 *HLA-DQA1* alleles and 13 *DQB1* alleles.

4.4.3.1 Cell types studied in this work

As class II expression is both cell-type specific and developmentally regulated, here we provide the details about the cell types that we analyzed (i.e., the cell types of peripheral blood) and their class II expressing abilities.

Isolated cell types

B cells. B cells were isolated based on the CD19 marker, which is expressed by all developmental stages of B cells (179). B cells in the bloodstream are mostly mature naïve B cells (60–70%), followed by memory B cells (20–30%), short-lived transitional (immature) B cells (10%), plasmablasts (< 2%) (179) (http://celentyx.com/b-cells.html) and regulatory B cells (180), all of them expressing class II molecules (179).

Monocytes. Monocytes were isolated from the whole blood based on the CD14 marker. In the peripheral blood, there are 3 monocyte subpopulations that differ in their function, marker expression and class II expression (181):

- Migratory monocytes that are able to migrate into tissues in response to appropriate signals (182):
    - CD14++ CD16- classical monocytes (85% of total monocytes)
    - CD14++ CD16+ intermediate monocytes (5% of total)
- Patrolling monocytes that remain in circulation and patrol the blood vessel endothelium
    - CD14+, CD16++ non-classical monocytes (10% of total).

All these populations express HLA class II molecules and the expression level is higher in the CD16+ subtypes (99, 183).

Whole blood cells. The proportion of individual cell types in the whole blood shows high interindividual variability, and was not measured in our study. A rough idea about composition of the analyzed samples can be given by the reference ranges for healthy individuals (**Table 10**).

Table 10. Representation of cell types in whole blood. Based on https://www.dartmouth.edu/~dartlab/uploads/CellTypes%20StemCell.pdf and (184).

Class II expressing cell type: yes	% of peripheral blood leukocytes
B cells	1–7%
monocytes	2–12%
DCs	0.3–0.9%
T cells: effector	0.7–3.7% (1/3 of CD8+ T cells)
Class II expressing cell type: no	
NK cells	1–6%
T cells: naive	1.4–12.6%
granulocytes	35–80% (mostly neutrophils)
Class II expressing cell type: not known	
T cells: memory	4–18%

Analyzed cell types

DNA methylation study. DNA for the methylation study was extracted from the unsorted whole blood, whose nucleated cells (leukocytes) are mostly neutrophils; followed by T cells; then B cells, monocytes and NK cells; and finally dendritic cells.

mRNA expression study. Class II mRNA expression analysis could, by definition, detected class II mRNA only in the cell types that express this molecule, which in the whole blood are B cells, monocytes, DCs and activated T cells. Thus, the class II expression we measured in whole blood is a weighted average of the B cell and monocyte expression (measured separately for both subsets) and DC and T cell expression (unmeasured by separate assays). Based on average proportions of these cell types in blood (**Table 10**), we can say that we measured mostly expression in B cells and monocytes (in 1:2 ratio).

4.4.3.2 Relative expression of HLA class II alleles in different cell types

DQA1 and DQB1 gene expression in cell types

First, we analyzed differences in *DQA1* and *DQB1* gene alleles expression between cell types irrespective of an allele identity. The highest ***DQA1*** level was observed in B lymphocytes. Whole blood cells (WBC) expressed around 1/2 and monocytes around 1/3 of the B cell levels, respectively (**Figure 6A**). The differences were significant for all cell-type pairs tested (B lymphocytes—monocytes; B lymphocytes—WBC; monocytes —WBC, each $p_c < 0.0003$).

For ***DQB1*** gene alleles, the only significant difference was observed between B lymphocytes and monocytes ($p_c = 0.0213$), where *DQB1* allele expression was on average 1.25 times higher in B cells (**Figure 6B**).

DQA1 and DQB1 allele expression in cell types

Then we examined differences in expression between cell types on the level of individual alleles (**Figure 7**). For the ***DQA1*** gene alleles, mean allele mRNA level was always highest in B lymphocytes and lowest in monocytes. Difference in allele mRNA levels between these two cell populations varied from 1.3-fold for the *DQA1*03* group of alleles to more than 13-fold for the *DQA1*05* alleles (**Table 11A**). Expression levels in WBC lied between the values for B cells and monocytes. Due to a low number of samples in some allelic groups, we were able to statistically compare only expression of *DQA1*01:01*, **02:01*, **03:01* and **05:05* alleles. For *DQA1*02:01*,**03:01* and **05:05* alleles, differences in expression between B cells and monocytes proved to be significant, as well as differences between B cells and WBC for alleles *DQA1*02:01* and **05:05*.

The situation was not so unambiguous for the ***DQB1*** gene. The mean allele mRNA level was higher in B lymphocytes compared to monocytes. This was the same pattern as in the *DQA1* gene, but the differences were less pronounced (1.8 to 4-fold for all alleles in this group). The situation was reversed for the *DQB1*06* allele group where B cells expressed 0.6 to 1.0-fold of the monocyte mRNA levels. The expression of *DQB1*03:03* and **05* seemed to be similar in both cell types (**Table 11B**). We were able to statistically test differences between cell types for alleles *DQB1*02:02*, **03:01* and **05:01*. The expression of *DQB1*02:02* and **03:01* alleles was significantly higher in B lymphocytes compared to monocytes and compared to WBC.

As in the study A, we calculated expression ratios in the heterozygous individuals. The allele expression hierarchy was generally preserved in the heterozygotes (**Table 12A, C**). An extremly high ratio (up to 20) was observed in three allelic combinations (*DQA1*0x:0x/*05*) in three individuals. These individuals differed in age and sex, thus these are not likely to be the factors that caused the anomaly. However, in all three measurements, the amount of the less expressed *DQA1*05* allele was very low (relative mRNA expression < 0.01), therefore we cannot exclude the possibility that a small error in measurement (in the $\pm 10^{-3}$ order), which would stay hidden in more expressed alleles, translated into a big change in the resulting ratios.

Theoretical DQ dimer expression

The DQ molecule is a dimer and its amount on the cell surface depends on expression of both subunits. Thus we calculated theoretical expression of DQ dimers based on mRNA expression of the *DQA1* and *DQB1* alleles in B cells and monocytes (**Table 13**). The most and least „theoretically expressed“ DQ molecules in both cell types would be DQ6 and molecules with *DQB1*02* and **03*-coded β chains, respectively. The differences in DQ expression between cell types would be highest for the DQ7.5 (4.9-fold) and lowest for the DQ9.2 (same expression on B cells and monocytes).

4.4.3.3 Inter-allelic differences in expression of *DQA1* and *DQB1* genes

Next, we inspected the differences in relative expression between individual alleles. In ***DQB1*** gene, alleles of **02*, **03*, **05* groups were less expressed than **06* group alleles, and this relationship was observed in all cell types examined. The inter-allelic differences were more pronounced in monocytes, where mean mRNA level of the most (*DQB1*06:03*) and least (**02:02*) expressed alleles differed up to 15-fold. In B cells, the expression of individual alleles was more balanced; the most (*DQB1*06:04*) and the least (**02:01*) expressed alleles differed 3.5-fold (**Figure 7**, **Table 14**).

Expression hierarchy of the ***DQA1*** gene followed the same pattern in monocytes and WBC, with lowest mRNA levels in *DQA1*05:05* and **05:01* alleles, gradually increasing through **02:01* and **01:01* alleles, **01:02* and **01:03* alleles, and peaking in *DQA1*03* alleles (3-fold higher than in the **05* allele group). In B lymphocytes, similarly to monocytes, *DQA1*01:01*, **01:02* (DR15) and **02:01* were on the low end and **03* group alleles on the high end of the expression spectrum. However, it was surprising to see that *DQA1*05* alleles, whose expression level in monocytes was the lowest out of all alleles, were among the most expressed ones in B lymphocytes, with levels similar to that of **03* allele group. As for *DQB1* gene, expression of individual alleles was more balanced in B cells; the most (*DQA1*05:01*) and the least (*DQA1*01:01*) expressed alleles differed 1.6-fold, compared to 10-fold difference between *DQA1*03:03* and **05:05* in monocytes.

4.4.3.4 Correlation of *PPIA* and *DRA* gene expression

To analyze the strength of HLA class II promoting stimuli in each cell type, we analyzed an expression of *DRA* gene normalized against a classic endogenous control, *PPIA*. *DRA* expression was significantly higher in B lymphocytes (mean relative *DRA* expression 4.2) and monocytes (4.9) compared to whole blood (1.9; both $p_c < 0.0003$) (**Figure 9A**). There was significant monotonic and linear correlation between Ct of *DRA* and Ct of *PPIA* in each cell type (**Figure 9B**). A value of Spearman's correlation coefficient was 0.88 in whole blood, 0.94 in B lymphocytes and 0.89 in monocytes, and Pearson's correlation coefficient was 0.99 in whole blood, 0.93 in B lymphocytes and 0.90 in monocytes (all six p values < 0.0001). Linear correlation between Cts means that ratio of *PPIA* and *DRA* in individual cell types tends to be constant among individuals. However, the slopes and constants of the equations describing this linear relationship were different for each cell type, indicating a cell type-specific relationship between the amount of *DRA* and *PPIA* transcripts.

4.4.4 mRNA expression of HLA class II genes *DQA1* and *DQB1* in B lymphocytes and monocytes in patients with T1D (pilot study)

mRNA expression of *HLA-DQA1* alleles in WBC of patients with T1D has been analyzed in our laboratory previously (185). In the study, higher expression of *DQA1*02:01* in healthy controls compared to patients was was found. As a next step, a study in was conducted to compare the expression of *DQA1* and *DQB1* alleles in cell populations in patients with T1D (Kotrbova-Kozak A, Cepek P, Zajacova M, Cerna M, 2016, unpublished). The results are shown in **Figure 8**.

Interallelic differences in expression of DQA1 and DQB1 genes

The expression pattern seemed to be the same in controls and patients for both genes in all cell types analyzed. Mean mRNA expression for each allele-cell type combination was calculated and compared between controls and patients with two-tailed Mann-Whitney test. There were enough samples for comparison for the following combinations: *DQA1*01:01, *03:01* and *DQB1*02:01, *02:02, *03:02, *05:01* in B cells, *DQA1*01:01, *02:01, *03:01* and *DQB1*02:01, *02:02, *03:02* in monocytes and *DQA1*03:01* and *DQB1*02:01, *03:02* in WBC. After Bonferroni correction ($p_c=p_{un}*4$), a single difference between expression of *DQB1*02:02* alleles in B cells was significant ($p_c=0.038$), with the expression higher in T1D patients.

Similarly to our previous studies, an allele expression ratio in heterozygotes was calculated (**Table 12B, D**). Again, the allele expression hierarchy was generally preserved in the heterozygous individuals.

Correlation of PPIA and DRA gene expression

Comparison of *DRA* expression normalized to *PPIA* between controls and patients (in WBC, control group from study A, group of patients from study (185)) did not reveal any differences between the two groups. The only notable difference were few outlying values in the patient group (**Figure 10**). We aimed to compare *DRA/PPIA* ratio between controls and patients as well, however, due to damaged batch of the *PPIA* assay which was recognized too late (and there was no opportunity to repeat the measurements), it was not possible.

Table 11. **Comparison of mRNA expression of *DQA1* (A) and *DQB1* (B) alleles in different cell types.** Expression data for each allele were normalized relative to *DRA* gene and mean relative expression was calculated for each combination of allele and cell type. Then ratio of expression in all combinations of cell types was calculated. For each allele, the number of B lymphocyte (B), monocyte (M), and whole blood (wb) samples is indicated.

A

	relative expression of allele in cell type			expression ratio in different cell types			number of samples
	B	**M**	**wb**	**B/M**	**B/wb**	**wb/M**	**B; M; wb**
DQA1*01:01	0.13	0.06	0.08	2.4	1.7	1.4	4; 5; 6
DQA1*01:02[c]	0.16	0.07	0.10	2.4	1.6	1.5	3; 2; 4
DQA1*01:02[d]	0.14	0.06	0.09	2.4	1.7	1.4	4; 3; 3
DQA1*01:02[e]	0.15	0.08	0.10	1.9	1.5	1.3	1; 1; 1
DQA1*01:03	0.18	0.08	0.12	2.6	1.5	1.7	4; 4; 5
DQA1*02:01	0.17	0.06	0.08	3.0[a]	2.2	1.4[a]	5; 7; 7
DQA1*03:01	0.20	0.14	0.17	1.4[b]	1.2	1.2	8; 7; 5
DQA1*03:03	0.21	0.16	0.17	1.3	1.2	1.1	1; 3; 2
DQA1*05:01	0.22	0.02	0.07	14.8	2.7	5.5	5; 4; 5
DQA1*05:05	0.18	0.02	0.06	13.1[a]	3.3	4.0[b]	12; 11; 13
max/min	1.6	10.4	3.0				

B

	relative expression of allele in cell type			expression ratio in different cell types			number of samples
	B	**M**	**wb**	**B/M**	**B/wb**	**wb/M**	**B; M; wb**
DQB1*02:01	0.04	0.02	0.03	2.0	1.6	1.2	4; 4; 6
DQB1*02:02	0.07	0.02	0.03	4.1[a]	2.7[b]	1.5	6; 6; 7
DQB1*03:01	0.07	0.03	0.04	2.8[b]	1.7[a]	1.7	10; 8; 12
DQB1*03:02	0.10	0.05	0.07	1.8	1.4	1.3	5; 4; 5
DQB1*03:03	0.06	0.06	0.05	1.0	1.0	1.0	2; 2; 2
DQB1*03:05	0.08	0.05	0.01	1.5	5.4	0.3	1; 1; 1
DQB1*05:01	0.08	0.05	0.07	1.5	1.1	1.3	5; 6; 5
DQB1*05:02	0.11	0.11	0.12	1.0	0.9	1.0	1; 1; 1
DQB1*05:05	-	0.05	0.06	-	-	1.1	0; 1; 1
DQB1*06:02	0.13	0.13	0.16	1.0	0.9	1.2	5; 4; 4
DQB1*06:03	0.15	0.25	0.22	0.6	0.7	0.9	2; 3; 2
DQB1*06:04	0.16	0.20	0.19	0.8	0.8	0.9	3; 2; 3
DQB1*06:09	0.14	-	0.18	-	0.8	-	1; 0; 1
max/min	3.5	15.4	16				

[a]p-corrected < 0.05, [b]p-corrected < 0.005

[c]Linked with DR13 (*QAP 1.4*), [d]DR15 (*QAP 1.2L*), [e]DR16 (*QAP 1.2K*).

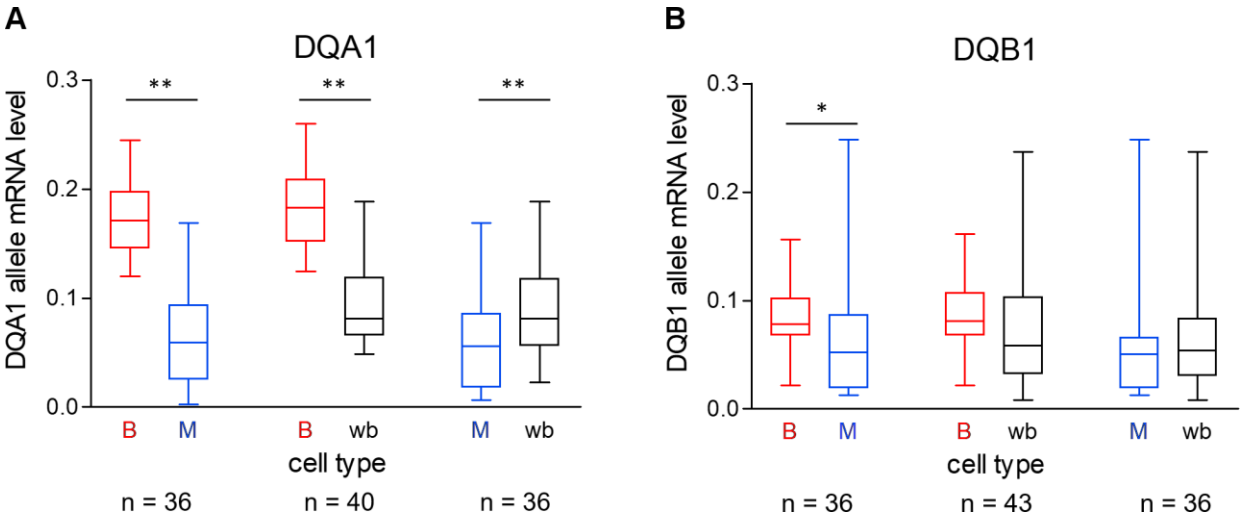

Figure 6. Relative mRNA expression of *DQA1* (A) and *DQB1* (B) alleles in different cell types. Normalized expression data of all *DQA1* (A) and *DQB1* (B) alleles in the indicated cell type were grouped together irrespective of allele identity, and differences in allele expression between cell types were analyzed by two-tailed Wilcoxon matched pairs test. Only individuals for whom measurements in both compared cell types were available (paired measurements) were included into analysis. Number of samples analyzed (n) and cell type (wb – whole blood cells, B – B lymphocytes, M – monocytes) is indicated. *$p_c < 0.05$, ** $p_c < 0.005$.

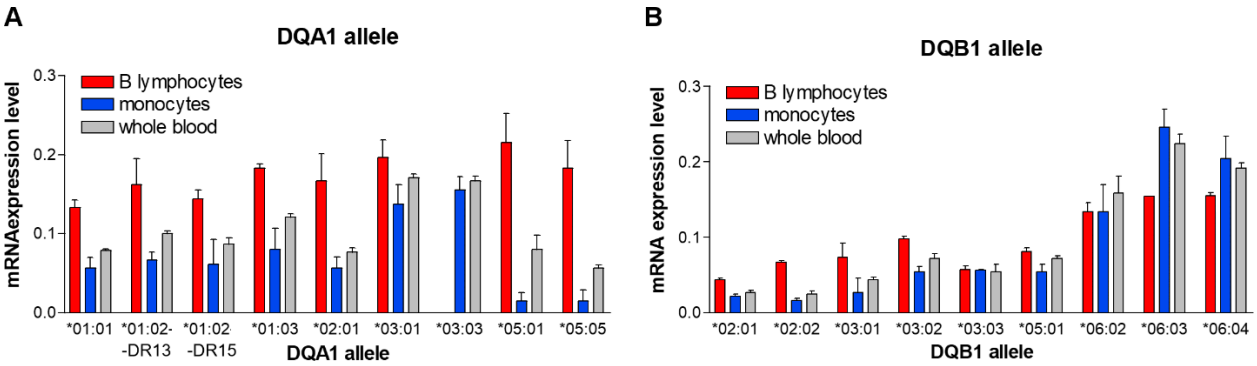

Figure 7. mRNA expression levels of *DQA1* (A) and *DQB1* (B) alleles in different cell types. Expression of *DQA1* (A) and *DQB1* (B) alleles was determined by RT-qPCR with allele-specific primers and the mean of 3 independent measurements was used for the analysis. Each column represents mean of data of at least 2 individuals, SD is indicated.

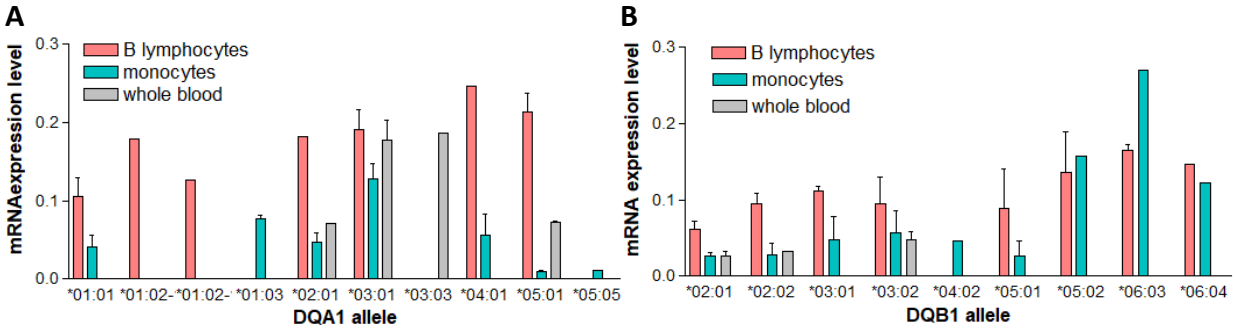

Figure 8. mRNA expression levels of *DQA1* (A) and *DQB1* (B) alleles in different cell types from T1D patients. Expression of *DQA1* (A) and *DQB1* (B) alleles was determined by RT-qPCR with allele-specific primers and the mean of 3 independent measurements was used for the analysis. SD is indicated.

Table 12. Mean expression ratio of DQA1 and DQB1 alleles in healthy heterozygotes and patients with T1D. For each heterozygous individual in our sample, the expression ratio of the two alleles was calculated by dividing expression of one allele (in this table listed as first) by expression of the second allele (listed as second). When data from multiple individuals with the same allelic combinations were available, the mean value for each combination was calculated, and used. Parts (A) and (C) list DQA1 and DQB1 expression ratios in healthy controls (HC), parts (B) and (D) list DQA1 and DQB1 expression ratios in patients with T1D (T1D). For each allele, the number of B lymphocyte (B), monocyte (M), and whole blood (wb) samples is indicated.

(A)

DQA1 allele combination	expression ratio (allele1/allele2) in HC						
(allele1/allele2)	B		M		wbc		
	mean	SD	mean	SD	mean	SD	n
*01:01, *02:01	0.90	-	1.00	-	0.98	0.06	1;1;2
*01:01, *03:01	0.65	-	0.44	-	0.52	-	1;1;1
*01:01, *05:01	0.66	-	4.97	-	1.46	-	1;1;1
*01:01, *05:05	-	-	4.84	-	1.62	-	-;1;1
*01:02[b], *03:01	0.75	-	0.33	-	0.50	-	1;1;1
*01:02[a], *03:03	0.64	-	0.51	-	0.57	-	1;1;1
*01:02[a], *05:01	-	-	5.69	-	1.56	-	1;1;1
*01:02[a], *05:05	0.91	0.02	-	-	1.59	0.16	2;-;2
*01:02[b], *05:05	0.95	0.07	7.75[d]	5.88	1.52	0.29	3;2;2
*01:02[c], *05:05	0.85	-	5.17	-	1.78	-	1;1;1
*01:03, *02:01	-	-	2.59	-	2.64	-	-;1;1
*01:03, *03:01	0.97	-	0.58	-	0.76	-	1;1;1
*01:03, *03:03	-	-	0.61	-	-	-	-;1;-
*01:03, *05:01	0.85	-	-	-	2.06	-	1;-;1
*01:03, *05:05	0.87	0.06	4.92	-	1.55	0.02	2;1;2
*01:04, *05:05	-	-	6.72	-	6.72	-	-;1;1
*02:01, *03:01	0.69	-	0.46	-	-	-	1;1;-
*02:01, *05:01	0.85	-	2.65	-	1.18	-	1;1;1
*02:01, *05:05	0.69	-	3.76	-	1.40	0.16	1;1;2
*03:01, *05:05	1.14	0.14	15.43[d]	9.45	3.60	0.27	3;2;2
*03:03, *05:05	-	-	22.04[d]	-	6.98	-	1;1;1

(B)

DQA1 allele combination	expression ratio (allele1/allele2) in T1D						
(allele1/allele2)	B		M		wbc		
	mean	SD	mean	SD	mean	SD	n
*01:01, *02:01	0.61	-	1.17	-	-	-	1;1;-
*01:01, *03:01	0.61	0.03	0.28	0.05	-	-	2;2;-
*01:01, *05:01	0.50	-	4.59	-	0.50	-	1;1;1
*01:02[a], *05:01	0.94	-	-	-	-	-	1;-;-
*01:02[c], *05:01	0.52	-	-	-	-	-	1;-;-
*01:03, *03:01	1.07	-	0.66	0.19	-	-	1;3;-
*02:01, *03:01	-	-	0.35	-	0.51	-	-;1;1
*02:01, *04:01	-	-	0.64	-	-	-	-;1;-
*02:01, *05:01	0.70	-	3.17	0.24	1.55	-	1;3;1
*02:01, *05:05	-	-	5.52	-	-	-	-;1;-
*03:01, *04:01	0.66	-	3.02	-	-	-	1;1;-
*03:01, *05:01	1.03	-	20.39[d]	-	-	-	1;1;-
*03:03, *05:01	-	-	-	-	2.64	-	-;-;1

(C)

DQB1 allele combination	expression ratio (allele1/allele2) in HC						
(allele1/allele2)	B		M		wbc		
	mean	SD	mean	SD	mean	SD	n
*02:01, *03:01	0.72	0.18	0.97	0.13	0.78	0.06	2;2;4
*02:01, *03:03	0.77	-	0.41	-	0.48	-	1;1;1
*02:01, *05:03	-	-	0.21	-	-	-	-;1;-
*02:01, *05:05	-	-	0.38	-	0.36	-	-;1;1
*02:01, *06:09	0.32	-	-	-	0.14	-	1;-;1
*02:02, *03:01	0.90	-	0.84	-	0.80	0.05	1;1;2
*02:02, *03:05	0.92	-	0.44	-	1.12	-	1;1;1
*02:02, *05:01	0.94	-	0.27	-	0.43	-	1;1;1
*02:02, *06:02	0.44	-	0.09	-	0.19	-	1;1;1
*03:01, *03:02	0.83	-	-	-	0.51	-	1;-;1
*03:01, *05:01	-	-	0.36	-	-	-	-;1;-
*03:01, *05:02	0.68	-	0.18	-	0.32	-	1;1;1
*03:01, *06:02	0.62	0.12	0.36	0.24	0.44	0.11	3;2;2
*03:01, *06:04	0.55	0.00	0.07	-	0.22	0.01	2;-;2
*03:02, *05:01	1.11	-	0.88	-	0.81	-	1;1;1
*03:02, *06:02	0.63	-	0.41	-	0.49	-	1;1;1
*03:02, *06:03	0.65	-	0.19	-	0.35	-	1;1;1
*03:02, *06:04	0.72	-	0.35	-	0.39	-	1;1;1
*03:03, *05:01	0.55	-	0.94	-	0.72	-	1;1;1
*05:01, *06:03	0.53	-	0.25	-	0.35	-	1;1;1

(D)

DQB1 allele combination	expression ratio (allele1/allele2) in T1D						
(allele1/allele2)	B		M		wbc		
	mean	SD	mean	SD	mean	SD	n
*02:01, *03:02	0.71	0.02	0.64	0.01	0.60	0.07	0.07
*02:01, *05:02	0.75	-	-	-	-	-	-
*02:01, *06:04	0.45	-	0.16	-	-	-	-
*02:02, *03:01	-	-	1.26	0.22	-	-	-
*02:02, *03:02	1.09	0.16	0.78	0.50	0.62	-	-
*02:02, *04:02	-	-	0.58	-	-	-	-
*02:02, *05:01	1.13	-	0.35	-	-	-	-
*03:01, *05:02	0.67	-	0.44	-	-	-	-
*03:01, *06:03	0.63	-	-	-	-	-	-
*03:02, *05:01	0.92	-	-	-	-	-	-
*03:02, *06:03	0.77	-	0.21	-	-	-	-

[a] Linked with DR13, [b] DR15, [c] DR16,

[d] The allele ratio in some *DQA1*0x:0x/*05* heterozygotes was extremly high. The individuals differed in age and sex, thus these are not likely to be the factors that caused the anomaly. As in all measurements the amount of the less expressed *DQA1*05* allele was very low (relative mRNA expression < 0.01), we cannot exclude the possibility that a small error in measurement (in the ±10-3 order), which would stay hidden in more expressed alleles, translated into a big change in the resulting ratios.

Table 13. Theoretical expression of DQ dimers. Expression of HLA-DQ dimers on the cell surface based on mRNA expression of *DQA1* and *DQB1* genes. DQ expression is determined as equal to the mRNA expression of the less expressed gene in the haplotype. Fold DQ expression for a particular DQ molecule is determined by dividing its DQ expression by 0.015 (DQ expression level of the least expressed DQ molecule). This is a simplified model that does not take into account post-translational mechanisms affecting the DQ expression on the cell surface.

Haplotype		DQ dimer	DQ expression		Fold DQ expression		
DRB1	DQA1*-DQB1*		B cells	mono	B cells	mono	B/mono
01	01:01-05:01	DQ5.1	0.081	0.054	5.4	3.6	1.5
03	05:01-02:01	DQ2.5	0.044	0.015	2.9	1.0	2.9
04	03:01-03:02	DQ8.1	0.098	0.054	6.5	3.6	1.8
07	02:01-02:02	DQ2.2	0.067	0.016	4.5	1.1	4.2
07	02:01-03:03	DQ9.2	0.057	0.056	3.8	3.7	1.0
11	05:05-03:01	DQ7.5	0.073	0.015	4.9	1.0	4.9
13	01:02-06:04	DQ6.4	0.155	0.080	10.3	5.3	1.9
13	01:03-06:03	DQ6.3	0.154	0.066	10.3	4.4	2.3
15	01:02-06:02	DQ6.2	0.134	0.061	8.9	4.1	2.2

Table 14. Expression hierarchy of individual alleles of *DQA1* (A) and *DQB1* (B) gene. Expression data were normalized relative to *DRA* gene and mean relative expression was calculated for each combination of allele and cell type. Only alleles with measurements from at least 2 individuals are listed. Then alleles were sorted according to expression means, and mRNA ratio of the most and the least expressed allele in each cell type was calculated. B lymphocytes (B), monocytes (M), whole blood (wb).

A

	mean *DQA1* allele expression	statistically significant results	most/least expressed
B	01:01 < 01:02 (DR15) < 01:02 (DR13) < 02:01 < 05:05 = 01:03 < 03:01 < 05:01	01:01, 01:02 (DR15) < 03:01	1.62
M	05:05 = 05:01 < 02:01 < 01:01 < 01:02 (DR15) < 01:02 (DR13) < 01:03 < 03:01 < 03:03	05:05 < 01:01, 01:03, 02:01, 03:01; 01:01, 02:01 < 03:01	10.4
wb	05:05 < 05:01 < 02:01 = 01:01 < 01:02 (DR15) < 01:02 (DR13) < 01:03 < 03:03 < 03:01	05:05 < 01:01, 01:02, (DR13), 01:03, 03:01; 01:01, 02:01 < 01:03, 03:01	3.04

B

	mean *DQB1* allele expression	statistically significant results	most/least expressed
B	02:01 < 03:03 < 02:02 < 03:01 < 05:01 < 03:02 < 06:02 < 06:03 < 06:04	-	3.52
M	02:02 < 02:01 < 03:01 < 03:02 = 05:01 < 03:03 < 05:02 < 06:02 < 06:04 < 06:03	02:02, 03:01 < 06:02; 02:02 < 03:02, 05:01	15.38
wb	02:02 < 02:01 < 03:01 < 03:03 < 03:02 = 05:01 < 05:02 < 06:02 < 06:04 < 06:03	02:01, 02:02 < 03:01, 03:02, 05:01; 03:01 < 03:02, 05:01, 06:02	8.96

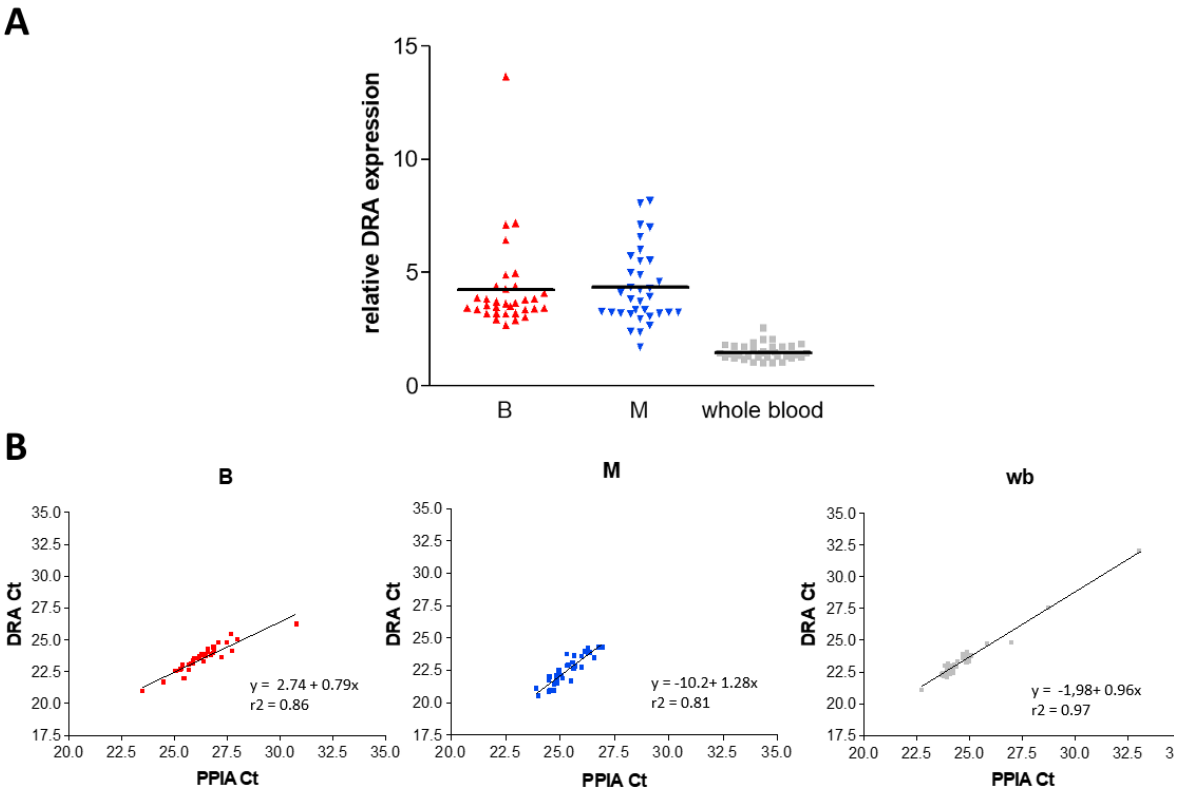

Figure 9. Expression of *DRA* mRNA in different cell types. Expression of *DRA* and *PPIA* gene was measured by RT-qPCR and the mean of 3 independent measurements was used for the analysis. (A) *DRA* mRNA levels normalized against *PPIA* in B lymphocytes (B), monocytes (M) and whole blood (wb). (B) Correlation between *PPIA* and *DRA* mRNA expression in different cell types. Ct values of *DRA* gene assay are plotted against Ct values of *PPIA* assay. Linear regression curves and r2 values are shown for each curve.

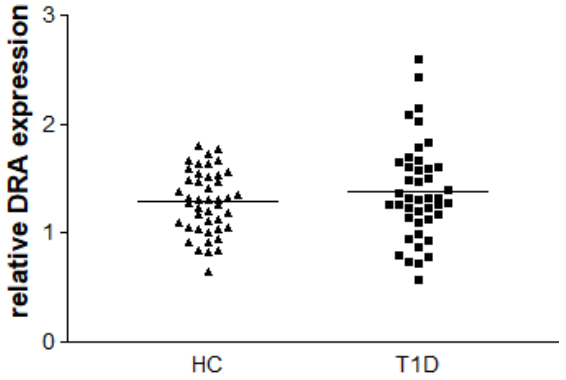

Figure 10. Relative *DRA* mRNA exoression in healthy controls (HC) and patients with T1D (T1D) in whole blood cells. Expression of *DRA* and *PPIA* gene was measured by RT-qPCR and the mean of 3 independent measurements was used for the analysis.

5 Discussion

The chapters 5.1 and 5.2 of the Discussion are based one the published articles (174, 175).

5.1 DNA methylation of the polymorphic *HLA-DQA1* promoters and its effect on the mRNA expression

We aimed to find out whether the observed differences in the mRNA expression of *DQA1* alleles could be caused by differences in their DNA methylation. As there are no CpG dinucleotides present in the *DQA1* gene SXY box (**Figure 1**) whose methylation could directly block the access of transcriptional coactivators to their cognate sequences, we focused on the differences that could affect the recruitment of chromatin remodeling proteins.

We would expect the strength of allele repression to correlate with the **local concentration of methylated cytosines** (130) (i.e., with the overall methylation of the promoter region). However, this was not the case for the *DQA1* alleles: in the region examined (-641 to -193 relative to first ATG), promoters of *DQA1*02:01* and *DQA1*04:01* were methylated significantly more than most of the other alleles, and *DQA1*05:01* was the least methylated allele. The high methylation density of the *DQA1*02:01* and **04:01* alleles did not match their average expression levels, and the low methylation density of the *DQA1*05:01* allele was not in accordance with its low expression.

Alternatively, the expression variability could be caused by **variable ability of the *DQA1* alleles to match the MBP requirements for the sequence surrounding the methylated cytosine**. This hypothesis is based on the observation that, at least *in vitro*, proteins recognizing 5meCpG show different preferences to this mark depending on the DNA sequence that surrounds it (186–188). As a consequence, their affinity towards alleles that have CpGs placed in different sequence contexts can differ, even when the alleles are methylated to the same extent. This could be of importance in *DQA1* and other highly polymorphic class II promoters, whose CpG dinucleotides vary not only in a number, but also in the distribution pattern (most sites are present only in a subset of alleles) and sequence context. For example, methyl-CpG binding protein 2 (MeCP2) prefers CpG sites adjacent to $(A/T)_{n\geq4}$ stretch (186), methyl-CpG-binding domain protein 1 (MBD1) recognizes best 5meCpG within a repeat of several methylated CpGs, and has somewhat higher affinity to CpG in T(G)CCGCA context too (187). Another methyl-CpG binding protein, Kaiso, had the highest affinity towards the sequences that contain two adjacent CpG dinucleotides (188). There are no MBD1 binding sites in any of the *DQA1* promoter alleles, but 1 Kaiso binding site in promoter of *DQA1*02:01* and *DQA1*01* group of alleles. It is interesting though that the least expressed alleles, *DQA1*05:01* and *DQA1*05:05*, have most theoretical MeCP2 binding sites (5 compared to 4 in *DQA1*03*

and 3 in most *DQA1*01* alleles). As most of the CpG sites in all alleles are methylated (**Figure 4B**), these context differences have a chance to manifest. This could be one of the factors that add up and lead to low expression of the *DQA1*05* alleles. Thus, we suggest that if DNA methylation plays a role in the *DQA1* alleles expression, it will not exert its effect through a local concentration of methylated cytosines, but rather through differences in their distribution and sequence context. This idea could be confirmed by studying allele-specific histone modifications and a presence/absence of the MBPs in the allele promoter (e.g., by ChIP experiments). A pilot projects from our laboratory on histone modifications of the class II alleles hint low H3 acetylation of the *DQA1*05* group promoters (significantly (189) and insignificantly (190) lower than of the **01* group), which is in line with the low *DQA1*05* mRNA expression level observed here. Based on this, we could speculate that the combination of MBP binding sites in the **05* allele promoter make it especially prone to the silencing by DNA methylation, but further experiments (as mentioned above) would be needed to confirm this. These pilot projects did not reveal correlation betweeen allele histone modification and mRNA expression – however, repeating these pilot experiments (to get measurements with lower SD), possibly in an extended genomic region (as in proximal region, the specific signal presents a very low percentage of input, which likely signifies an absence of histones in this region) may provide more informative results.

In all *DQA1* alleles we could observe a common pattern where most of the promoter CpG sites were methylated and as we approached the transcription start site, the methylation decreased to almost zero. This is in line with results of several studies which suggest that for class II gene expression, the most TSS-proximal area should be methylation free, as was shown for the *DRA* (149, 150), *DQA1* and *DQB1* genes. Majumder *et al.* showed that DNA hypermethylation of the promoter region of the *DQA1* and *DQB1* genes in cancer cells is able to suppress the transcription of the gene even in the presence of CIITA and all other factors necessary for class II genes expression (38, 110). For *DQA1* gene, the region in question largely overlapped with the region that was unmethylated in our samples. This seems in contract to results of Weber *et al.* (130): according to their criteria, all *DQA1* allele promoters belong among low-CpG promoters, a class of promoters with low content of CpG that tend to be highly methylated whether they are expressed or not, and whose high methylation should not be an obstacle to alleles expression. However, Weber's analysis simplified the gene expression level to yes/no sense, and their findings do not exclude fine regulation of gene expression by DNA methylation even in this promoter class.

To conclude, the overall level of *DQA1* promoter methylation is not able to explain expression differences between alleles. We suggest that the genetic polymorphism in the region (especially region upstream of position -400, which is almost completely methylated in all alleles) may lead to different interpretation of the same 5meCpG mark in different allelic contex. In addition, lack of DNA methylation

in the region very proximal to the transcription start site may be required for the gene expression and we observed that this area tends to be methylation free.

5.2 mRNA expression

5.2.1 mRNA expression of the *HLA-DQA1* alleles in the whole blood cells

According to our findings, the hierarchy of mRNA expression levels of *DQA1* alleles is *DQA1*03 > *01:03 > *01:01,*05:05,* and *DQA1*03 > *02:01 > *05:05*, with the average mRNA expression level of *DQA1*03* alleles being approximately 2–4 times higher than the average mRNA expression of any other allele. In the heterozygous individuals present in our study sample, *DQA1*03* allele was always overexpressed as compared to any other allele. In contrast, *DQA1*05:05* allele was always the less expressed one (with an exception of single individual, in whom **05:05/*01:05* ratio was found to be 1.09).

Our results are in agreement with the expression hierarchy observed in peripheral blood cells by Donner *et al.* (*DQA1*03 > *01 > *02:01 > *05*) (191) and Maffei *et al.* (*DQA1*03:0 > *05:01*) (192), and in peripheral blood mononuclear cells by Britten *et al.* (*DQA1*03,*01:03 > *05:01*) (143). The exception to this trend was reported by Fernandez *et al.*, who observed the highest expression of *DQA1*04*, and no interallelic differences between other *DQA1* alleles in peripheral blood mononuclear cells (PBMC) (172). This discrepancy is most probably caused by their team not considering differences in efficacy of individual assays. In conclusion, there seems to be an agreement on highest expression of *DQA1*03* and lowest expression of *DQA1*05* group of alleles.

Our results suggest that relative mRNA level of particular *DQA1* allele does not show much interindividual variability, tends to remain constant for a particular allele, and is not dependent on the second allele present in heterozygous individual. If this is true, then the total relative amount of *DQA1* mRNA (which is the sum of relative *DQA1* mRNA levels of individual alleles) should differ between individuals in accordance with the alleles an individual carries. It was indeed observed that the sum of relative *DQA1* allele expression in heterozygotes (**Table 9A**) paralleled theoretical values obtained by adding up mean expression levels of the two alleles in the given heterozygous combination (**Table 9B**). An interesting consequence of this finding is that not only can an amount of *DQA1* mRNA differ between alleles, but also **the total (i.e., sum of both alleles) amount of *DQA1* mRNA can differ between individuals based on the particular alleles they carry**. Indeed, we observed higher than twofold differences in total *DQA1* mRNA expression in individuals in our cohort (from *DQA1*/*DRA* mRNA ratio 0.15 in *DQA1*01:01/*05:05* heterozygotes, to more than 0.30 in most *DQA1*03*/other heterozygotes).

These differences could be even more pronounced in individuals homozygous for the most and the least expressed alleles, but these were not present in the study group (however, we observed up to 4-times

higher expression of total relative *DQA1* mRNA in *DQA1*03/*03* homozygotes compared to *DQA1*05:01/*05:01* homozygotes in group of patients with autoimmune disease (185)).

Low interindividual variability in the allele expression is in agreement with results of Fernandez *et al.*(172), however, they did not observe effect of gene dosage in the expression of *DQA1*01* and *DQA1*05* alleles in homozygotes and heterozygotes as we did. Similarly, Pisapia *et al.*(126) observed similar amount of DQ2.5 dimers on a surface of homozygous and heterozygous APCs. The reason for this observation is not known, and it cannot be explained by *trans*-dimer formation, as haplotype combinations employed in their study do not form stable DQ dimers (32).

5.2.2 mRNA expression of the *HLA-DQA1* alleles in B lymphocytes and monocytes

In the next phase, we analyzed the expression of *DQA1* and *DQB1* gene alleles in purified cell populations of CD14+ monocytes and CD19+ B lymphocytes. To provide a reference frame and compare data with the previous study, we again included whole blood cells into the analysis.

The ***DQA1*** expression hierarchy in the **whole blood** was **01:03, *03:01 > *01:01 > *05:05* (statistically significant results) and *DQA1*03:01*, **03:03 > *01:03 > *01:02* (DR13, DR15-linked) > **01:01 >*02:01 > *05:01*, **05:05* (descriptive), which is in accordance with our pilot study on *DQA1* gene allele expression in whole blood and with other studies mentioned above.

Essentially the same expression hierarchy as for whole blood was observed in **monocytes** (**Table 14**, **Figure 7A**). However, the situation was different in **B lymphocytes**: although *DQA1*03:01* was still highly expressed, it was surprising to see **05* group alleles among the most expressed alleles as well, as they were the least expressed ones in monocytes in our study, and were also consistently reported to be the lowest expressed allele group in various whole blood cell subsets by others (143, 191, 193). This discrepancy is likely to be caused by the actual difference in the *DQA1* allele expression regulation in B cells compared to mixed cell populations used in these studies, as the only other work that observed higher expression of *DQA1*05* compared to 01 allele was also done on B lymphocytes (126). The authors reported *DQA1*05:01* allele mRNA levels 1.9–4.6 higher compared to the **01:01* allele in heterozygous **01:01/*05:01* B lymphocyte-derived lymphoblastoid cell lines from both celiac disease patients and the healthy controls. The same team observed identical *DQA1*01:01 > *05:01* relationship in monocyte-derived DCs (126), which is inconsistent with low expression of *DQA1*05* alleles observed in this study; however, we cannot exclude the possibility that during the maturation of monocyte derived DC, different alleles are induced to different levels (122) and expression ratio changes.

When different cell populations were compared, the *DQA1* allele expression was higher in B lymphocytes than in monocytes for each allele. Also, the expression of alleles in B lymphocytes was more balanced (maximum ratio between mean allele expression was 1.6), while in monocytes, the mean levels of

the transcripts of most expressed allele *DQA1*03:03* were 10 times higher than those of the least expressed alleles of **05* group.

5.2.3 mRNA expression of the *HLA-DQB1* alleles in the whole blood cells, B lymphocytes and monocytes

We analyzed expression of 10 *DQB1* alleles belonging to *DQB1*02*, **03*, **05* and **06* allele groups. In all three cell populations examined, *DQB1*02* alleles were the least expressed ones, and *DQB1*06* alleles the most expressed (**Table 14**, **Figure 7B**). This is in agreement with previous works: low mRNA expression of *DQB1*02:01* in the heterozygous B lymphoblastoid cell lines (half of the *DQB1*04:02*) (122), the lowest expression of *DQB1*02* and highest of **06* allele group was observed also by Britten *et al.* in PBMC (143), who also confirmed the different strength of the respective promoters by expression in transfected B cell lines. Similar to our samples, *DQB1*06:02 > *05:01* (5-fold difference) and **05:01 > *03:01* (2-fold difference) hierarchy was observed also in monocytes (122).

Beaty *et al.* (71) observed approximately 2.5 higher transcription of the heterologous reporter gene under the control of *DQB1*03:01* promoter compared to **03:02* promoter in B cell lines. However, in our samples we observed *DQB1*03:02* allele expression slightly higher than that of *DQB1*03:01* allele (and the result reached significance in whole blood cells). Finding of 2.6-9 times higher mean expression of *DQB1*02:01* allele compared to the expression of *DQB1*03*, **05* and **06* mRNA in monocyte-derived DCs and B-LCL (126) is discrepant with our measurement that *DQB1*02:01* allele was the least expressed one in all cell types examined. Other studies examining unbalanced expression of *DQB1* alleles used heterozygotes for *DQB1*04* allele (122, 194) which were not present in our study sample, hence a comparison with our results was not possible.

Relationship between the expression of *DQB1* alleles in different cell types was not so clear as was the case with *DQA1* gene. For the *DQB1*02* group, **03:01*, **03:02* and **05:01*, the allele expression was higher in B lymphocytes compared to monocytes, for *DQB1*03:03* and **06:02* expression was at the same level, and finally for *DQB1*06:03* and **06:04* alleles it was lower in B cells. Similar to *DQA1* allele, the inter-allelic expression was more uniform in B cells (most/least expressed allele means ratio was 3.5) compared to monocytes (ratio of 15.4).

Differences in expression of *DQA1* and *DQB1* alleles in B cells and monocytes can possibly translate into DQ dimer expression hierarchies that differ in these cell types (see **Table 13**). Subsequently, some DQ molecules may have carry risk to autoimmunity depending on whether the disease has more important B cells or monocyte contribution.

5.2.4 Cell-type specific regulation of HLA class II genes

The major observations on the cell-type specific expression are:

- For most *DQA1* and *DQB1* gene alleles, an expression in B lymphocytes is higher
- Inter-allelic expression differences are much less pronounced in B lymphocytes than in monocytes.

The first observation could be explained by a **regulatory protein present in B cells, but not in monocytes.** This function could be exerted by CIITA itself. **CIITA isoform III** which regulates constitutive class II expression in B lymphocytes has the highest transactivation potential out of all isoforms, due to more efficient interaction with transcription machinery and promoter binding protein RFX (100), and could be the factor responsible for a higher expression observed in B lymphocytes.

The second observation, the one of distinct allele hierarchies in the B cells and monocytes is very intriguing**.** It implies that there must be **a transcription factor or coactivator that is both cell-type- and allele-specific** (and also isotype-specific, as it does not affect *DRA*). As discussed in the introduction, there are genetic differences between the promoter alleles that can affect the class II enhanceosome formation – possibly the **CIITA isoform III** could not only bind better to RFX, but also be more tolerant to variation in the enhanceosome structure, mitigating the effect of inter-allelic promoter variation on the transcription. However, this is just a hypothesis not explored in any study, and we did not follow up on this idea either. An alternative explanation involves a **transcription factor that would be more abundant in B cells than in monocytes**: saturation of all class II promoters in the presence of high protein concentration would lead to more allele-balanced expression in B cells; while a lower amount of the factor would lead to competition between promoters in monocytes. If this is the case, then **RFX5** protein is a potential candidate for such regulation, as its (mRNA) expression is approximately 3 to 4-fold higher in B cells compared to the monocytes (http://biogps.org/#goto=genereport&id=5993). All other basic subunits forming the enhanceosome complex (RFXAP, RFXANK, NF-YA, NF-YB, NF-YC) are expressed to similar extent in these cell types (http://biogps.org).

Of note, a switch in *DQB1*03:01* and **03:02* allele hierarchy between primary skin fibroblast (**03:01 > *03:02*) and adherent PBMC (i.e., monocytes and macrophages; **03:02 > *03:01*) was reported before and the authors considered different CIITA levels in these cell types as a possible reason (195).

Therefore, it is possible that promoter polymorphisms, cell type-specific CIITA isoforms and maybe some other, yet unidentified, cell type-specific factors interact and generate observed complexity of HLA class II alleles expression. In conclusion, we described differences in expression of *DQA1* and *DQB1* gene alleles in whole blood cells, monocytes and B lymphocytes. Our findings support the idea that expression level of HLA class II alleles is the result of both, promoter-, and cell type-specific factors.

5.2.5 mRNA expression of the *DRA* gene

Congruent with their role of the antigen-presenting cells, we observed almost three times higher expression of *DRA* gene (a measure of HLA class II promoting stimuli in a sample) in B lymphocytes and in monocytes compared to whole blood.

High linear correlation of Ct values of *DRA* and *PPIA* gene assays observed in B lymphocytes and monocytes indicated that individuals within our study population were more or less homogenous regarding HLA class II expression in these cell types. We observed a correlation between *DRA* and *PPIA* also in samples from unsorted whole blood cells. The donors were healthy individuals, not undergoing acute infection or having an autoimmune disease. It is thus reasonable to assume that the absence of inflammatory cytokines in their blood and amount of blood cell populations within normal, reference range (not tested) could lead to observed homogeneity. The formula describing the linear relationship was somewhat different in each cell type, suggesting differential regulation of HLA class II expression in these cell types.

5.2.6 What are the advantages of using the *HLA-DRA* gene as an endogenous reference?

The proportion of HLA class II expressing cells in the whole blood can vary between different individuals and depends on the proportion of APCs among leukocytes and signals that stimulate the HLA class II expression. Furthermore, expression of HLA class II molecules within the same cell type can differ on a cell-to-cell basis. Therefore, referring to classical endogenous control could lead to the false-positive finding of different *DQA1* (*DQB1*) allele expression between two samples only due to differences in the amount of HLA class II expressing cells or strength of inflammatory stimuli in the sample. However, as all HLA class II molecules are regulated coordinately, if we use another HLA class II molecule as an endogenous control, we eliminate all these influences. *HLA-DRA* gene is unique among HLA class II genes because it almost lacks their typical polymorphism. Thus, interindividual variation in its expression should not be affected by genetic variation. As a result, the *DRA*-normalized expression of *DQA1* (*DQB1*) allele is proportionate purely to the strength of allele promoter (caused by *cis*-acting genetic and epigenetic differences).

Using *DRA* as an endogenous control gives us one more advantage – we can perform additional expression studies on other HLA class II genes, normalize values to the *DRA* expression, and compare them with each other.

Further, even though a linear relationship between the amount of *DQA1* (*DQB1*) and *DRA* gene mRNA was one of the premises of our experimental setup, the approximately constant ratio of *DQA1* (*DQB1*) mRNA/*DRA* mRNA for each allele and also for *DQA1* (*DQB1*) total indeed confirms its validity.

5.3 HLA class II autoimmunity relevant properties - example of T1D

Finally, let's explore the effect of the HLA class II allele properties on the example of a common autoimmune disease, type 1 diabetes, in which immune system attacks insulin-producing β cells of pancreas (196). T1D initiation requires insulin-reactive CD4+ T cells. The pathogenic effectors are Th17 cells, and the protection is provided by Tregs. The Role of Th1 and Th2 subsets is not so unambiguous, though Th1 cells seem to promote and Th2 cells prevent T1D development. CD8+ cells function as the actual β-cell killers. B lymphocytes are important in promoting autoimmunity and maintaining memory T cells and DCs can either promote or prevent the disease by inducing either Tregs or pro-inflammatory effectors (196).

Evidence suggests that in T1D, T cell self-tolerance may be disrupted either in thymus, when high-avidity insulin-reactive thymocytes evade central tolerance in some patients, and in periphery by skewing the balance between the T cell subsets (196). Thus, factors affecting antigen presentation have a crucial role in T1D, which is supported by strong association of this disease with the class II genes.

5.3.1 Genetic risk – T1D

Approximately half of the T1D risk can be attributed to the genetics factors, and HLA genes are responsible for a half of the genetic risk (197). The T1D predisposition can be carried independently by individual alleles as well as by whole haplotypes. Multiple risk alleles can together form high-risk haplotypes as is *DRB1*04-DQA1*03:01-DQB1*03:02* (DR4-DQ8.1) and **03-*05:01-*02:01* (DR3-DQ2.5), and the highest risk is carried by the DR3/DR4 heterozygotes. The *DRB1*15-DQA1*01:02-DQB1*06:02* (DR15-DQ6.2) haplotype confers dominant protection (198).

5.3.2 Allele expression level and T1D risk

What are the properties of these haplotypes, and particularly DQ molecules, that bring the susceptibility? T1D risk has been previously explained by the ability of risk HLA molecules to present insulin (199), however, both protective and risk HLA alleles are able to present insulin-derived peptides, despite differences in the peptide binding groove (200). T1D seems to be strongly associated with low stability of DQ dimer and its short half-life on the cell surface (28, 32). Risk DQ2 and DQ8 dimer also display increased resistance to DM-mediated editing (47). In the context of these studies, we tried to explore how can allele expression level relate to the T1D risk.

It seems that protective DQ dimers are intrinsically stable molecules with high expression level in B cells and monocytes and risk DQ dimers are unstable molecules with low expression (**Table 15B**). This is in line with theory that increased antigen presentation in thymus supports generation of protective Tregs (199). The expression hierarchy of the T1D risk and protective haplotypes in this study is in agreement with observation of Raj (49). However, comparing OR (taken from (201)) of DQ dimers with their 'theoretical'

expression rate did not reveal significant correlation of these parameters in B cells nor in monocytes (Spearman's correlation coeficient, data not shown). Evidence supports strong relationship between HLA class II mRNA and protein (49, 126), yet it would be interesting to actually measure surface expression of individual DQ dimers and then relate it to the risk for T1D.

It has been suggested that HLA-DQ6.2 can protect from diabetes by competing for the same antigen with the risk DQ8.1 and DQ2.5 molecules (202, 203). High expression of DQ6.2 observed in this study is in agreement with this theory: expression of DQ2.5 and DQ8.1 is around 0.3 and 0.7 of DQ6.2 expression in B cells, and around 0.4 and 0.9 of DQ6.2 expression in monocytes. These differences may become more pronounced when we consider that DQ6.2 is approximately 10-fold more stable than these molecules (32).

We could also speculate that unstable molecules with high expression level could be the best candidates for autoimmunity-risk alleles: in thymus, with limited antigen availability, they can be expressed in low levels and escape central tolerance mechanisms. At the same time, they can be expressed in high levels if the antigen is abundant in periphery and in this way activate the effector T cells. In the most common haplotypes in our study population, there are no haplotypes that have at the same time low stability and high expression rate in neither B cells or monocytes (**Table 15**). This could potentially represent a result of selection against molecules with these properties.

Comparison of the *DQA1* and *DQB1* allele mRNA expression in healthy controls and patients with T1D enables us to assess non-genetic factors influencing T1D development. Previously (185) we observed an increased expression of the *DQA1*02:01* allele in WBC from healthy *DQA1*02:01/*03* heterozygotes. *DQA1*02:01* is a part of neutral *DR7-DQA1*02:01-DQB1*02:02* haplotype, however, it is possible, that its increased expression in the carriers of risk haplotypes protected them from developing the disease (for example, by forming stable *trans*-dimers (32) with the DQα chain trascribed from high-risk *DQA1*03:01* and thus basically preventing function of the DQ8.1 dimer). However, in this study we observed an increased expression of the *DQB1*02:02* allele from the same haplotype in B cells from T1D patients. This is in contrast with the previous theory and we have no explanation for it.

5.3.3 Epigenetics of T1D

T1D disesase-association studies employ monozygotic twins discordant for the disease to study effect of non-genetic factors on the disease development. Such studies revealed mutiple differences in **DNA methylation**, including hypermethylation of *HLA-DOB* and *HLA-DQA2* (204) and hypomethylation of the *HLA-DQB1* gene in patients (205), which could arise early in the proces, even before the disease manifestation (205). The association between DNA methylation and T1D is supported by observation that methylation status of CD14+ cells of a prediabetic quadruplet was intermediate between its affected and healthy siblings, suggesting a relationship between disease severity and DNA methylation (206). However,

Table 15. (A) Properties of DQ molecules in relation to the risk of type 1 diabetes: T1D risk carried by the DQ dimer: neutral (N), risk (R), protection (P), not available (-) (32, 33, 207); OR (for Caucasians in (201)) intrinsic stability: DQ dimer stability compared to the DQ6.2 (32); CLIP affinity towards the dimer: low (l), medium (2), high (3), not available (-) (43);dimer expression: as measured in this study and determined in **Table 13**. **(B) Relationship between HLA-DQ dimer stability and expression and T1D risk in B cells and monocytes**.

(A)

DQ dimer		dimer properties					
DQA1*-DQB1*	DQ dimer	T1D risk	OR (201)	intrinsic stability	CLIP affinity	expression in B cells	expression in monocytes
01:03-06:01	DQ0601	P	0.25	0.265	-	-	-
04:01-04:02	DQ4.2	N/R	-	0.087	-	-	-
03:01-04:01,2	DQ4.3	R	1.48	0.085	-	-	-
01:04-05:03	DQ5.3	P	0.38[a]	0.011	1	-	-
03:01-02:01	DQ2 *trans*	?	-	0.452	-	0.044	0.022
05:01-02:01	DQ2.5	R	3.64	0.100	3	0.044	0.015
02:01-03:03	DQ9.2	P	0.02	0.815	3	0.057	0.056
03:01-03:03	DQ9.3	R	0.53[b]	0.173	2	0.057	0.056
02:01-02:02	DQ2.2	N	0.32	0.565	3	0.067	0.016
03:01-03:01	DQ7.3	P	0.35	0.153	-	0.073	0.026
05:05-03:01	DQ7.5	P	0.36	0.202	-	0.073	0.015
01:01-05:01	DQ5.1	N	0.71	0.012	1	0.081	0.054
03:01-03:02	DQ8.1	R	3.47[a]	0.089	2	0.098	0.054
05:01-03:02	DQ8 *trans*	?	-	0.047	2	0.098	0.015
01:02-06:02	DQ6.2	P	0.03	1.000	1	0.134	0.061
01:03-06:03	DQ6.3	P	0.13	0.349	1	0.154	0.066
01:02-06:04	DQ6.4	N/P	0.87	0.378	1	0.155	0.080

[a] there are multiple DQ dimers whose risk depends on the linked *DRB1* allele, OR is calculated as a mean of ORs for all these DQ dimers; [b] DQ9.3 dimer is associated with risk in Asians and protection (non-significant) in Europeans (201).

(B)

B cells stability\expression	<0.015–0.043)	<0.015–0.043)	<0.071–0.099)	<0.099–0.127)	<0.127–0.155>
< 0.2		DQ2.5, DQ9.3	DQ5.1,DQ7.3, DQ8.1		
<0.2–0.4)	DQ0601		DQ7.5		DQ6.3,DQ6.4
<0.4–0.6)		DQ2.2, DQ2.31			
<0.6–0.8)					
≥ 0.8		DQ9.2			DQ6.2

monocytes stability\expression	<0.015–0.043)	<0.015–0.043)	<0.071–0.099)	<0.099–0.127)	<0.127–0.155>
< 0.2	DQ2.5,DQ7.3	DQ5.1,DQ8.1, DQ9.3			
<0.2–0.4)	DQ7.5	DQ6.3	DQ6.4		
<0.4–0.6)	DQ2.2,DQ2.31				
<0.6–0.8)					
≥ 0.8		DQ6.2,DQ9.2			

comparison of DNA methylation of the *HLA-DQA1* gene between HLA-matched T1D patients and healthy controls in our laboratory revealed no difference in DNA methylation of the proximal promoter of this gene (185).

Histone modifications in T1D were explored by a set of works by Miao. The authors observed differences in various histone modifications (H3K9Ac, H4K16Ac, H3K4me3, H3K9me2,3, H3K27me3) of genes involved in diabetes pathways between T1D and healthy controls (166, 167). The loci of variant chromatin modifications were also located close to *DQB1, DRB1* and *DPB2* gene: monocytes from T1D patients had lower levels of H3K9Ac 4 kb upstream of *HLA-DRB1* and higher levels of H3K9Ac 4 kb upstream of *HLA-DQB1* (167)—in the location of the #4 XL sequence identified before (see the Introduction). The increased acetylation at these sites correlates with increased transcription in the monocyte cell line (167). However, it is not possible to decide whether the differences are a cause of the disease or result of the disease-associated hyperglycaemia (168, 208). In addition, authors do not state whether the patients and controls were HLA-matched or not, so we cannot exclude the option that the observations are result of interallelic variation rather than a disease.

5.3.4 Study limitations

Whole blood cell subpopulations targeted in DNA metylation and mRNA expression analyses did not correspond. DNA methylation was studied in unsorted whole blood leukocytes, which in healthy individuals comprise mostly of class II negative cells (**Table 10**). At the same time, our mRNA expression methodology let us study only the class II expressing cell types. Even though it seems that under normal circumstances chromatin of all class II genes in all cell types is in the state poised for transcription (156) and differences in DNA methylation between cell types were shown to affect mostly regions distal from promoter (209), we cannot exclude that our approach confounded a potential relationship between allelic DNA methylation and mRNA expression (i.e., our methodology increased a risk of false negative findings). This limitation could be overcome analysing DNA methylation in purified cell types (as in pilot study of our laboratory (185)) or by determining the cellular content of the samples (210) and solving this issue may be needed in case futher studies in the wider class II region will be conducted.

Not analysing 5-hydroxymethylcytosine. The bisulfite sequencing that we used to detect 5meC is not able to discriminate this mark from another cytosine modification, 5-hydroxymethylation (211). Although the precise role of this mark in transcriptional regulation is not known yet, its distribution pattern at intergenic *cis*-regulatory elements and promoters with intermediate CpG density (212) or its effect on CTCF binding (213) make its regulatory role plausible.

Low-throughput DNA sequencing method. Combination of cloning and DNA sequencing by the Sanger's chain-termination method enabled us to distinguish DNA methylation pattern with a single-DNA

molecule resolution. However, with this approach we were able to study only a very limited genomic region in a limited number of samples. Therefore, for the future studies, I would recommend to use contemporary high-throughput methods that enable sequencing of, e.g., whole HLA class II region with the comparable budget and within a fraction of time, to fully exploit the research potential of this topic.

Correlation of class II mRNA and protein expression. We analyzed expression of *DQA1* and *DQB1* genes on the mRNA level, however, the amount of DQ dimers on the cell surface is also modified by post-transcriptional mechanisms, ability of protein chains to form stable dimers, ability to form *trans*-dimers and finally the availability and properties of the peptide ligand (discussed in the introduction). We should keep this limitation in mind when thinking about biological significance of our findings.

Non-HLA genes polymorphism. We did not analyze polymorphism of non-HLA genes that could possibly affect the class II expression.

6 Conclusions

This thesis focused on allele-specific mRNA expression and promoter DNA methylation of HLA class II genes. The work was divided into two studies, each addressing different aspects of the HLA class II allele expression regulation:

Study A: HLA class II DNA methylation and its effect on the mRNA expression

Aim: Determine DNA methylation of individual alleles of the *HLA-DQA1* gene. Analyze inter-allelic differences in DNA methylation. Analyze the relationship between allele DNA methylation and mRNA expression.

Results: In the **study A**, we determined expression hierarchy and DNA methylation profile of proximal promoter of the *HLA-DQA1* alleles. We observed inter-allelic differences in the **overall methylation**, which was highest for the *DQA1*02:01* and **04:01* alleles and did not match their average expression level, and the low expression of the *DQA1*05:01* allele was not in accordance with its low methylation density. There was no correlation between number of methylated cytosines per allele and its expression. Frequency of methylation of the **individual CpG sites** did not differ between alleles, however, due to the high polymorphism of class II genes, the CpG sites were often located in different positions and sequence contexts which could translate into epigenetic differences between alleles and subsequently into differences in mRNA expression levels.

To conclude, based on our results, DNA methylation density of the *DQA1* promoter alleles does not regulate their expression. It is possible that allele expression is regulated by methylation of individual CpG sites in different allelic contexts, but this hypothesis would require furtner testing.

Study B: HLA class II mRNA expression (allele- and cell-type-specific expression)

Aim 1: Determine the mRNA expression hierarchy of HLA class II *DQA1* and *DQB1* alleles.

Aim 2: Determine expression hierarchy of the *DQA1* and *DQB1* alleles in the whole blood cells, B cells and monocytes.

In the **study B,** we determined the mRNA expression hierarchy of the individual alleles and described its variation between the three cell populations. Class II transcription level is higher (with the exception of *DQB1*06* alleles) and shows lower interallelic variation in B cells compared to monocytes. The *DQB1* expression hierarchy can be generalized into *DQB1*06* > **03*,**05* > **02* pattern in monocytes and *DQB1*06* > **02*, **03*, **05* in B cells. The *DQA1* expression hierarchy is *DQA1*03* > *01 ≥ **02* > **05* in monocytes and *DQA1*03*, **05* > *01, **02* in B cells. Because of the low number of samples for certain allele-cell type combinations, only some of these results were confirmed statistically. CIITA isoforms are discussed as the factors that could drive the cell-type specific expression of class II alleles.

Finally, a relevance of these findings to the type 1 diabetes as an example of autoimmune disease is discussed. If we simplify ‘theoretical‘ amount of DQ dimers on the cell surface as an mRNA amount of the less expressed of its two chains, then the expression level by itself is not able to explain dimer association with T1D.

As a followup of this study, we could look deeper into several topics related to HLA class II regulation, primarily into the causes and consequences of variant expression of HLA class II alleles:

Transcriptional regulation of HLA class II expression: How does genetic polymorphism of class II genes affect their transcription rate and which sequences are responsible for differences between expression rates of individual alleles? This could be answered by a computational analysis of the sequence data (available online) and mRNA expression data (measured by this or similar studies).

Epigenetic regulation of HLA class II expression: Do differences in positioning of methylated CpG sites in the *DQA1* promoter lead to differences in MBPs binding pattern? As a consequence, are there differences in histone modifications (chromatin structure) between the HLA alleles? First step into this topic was taken by the pilot studies done in our laboratory (189, 190).

It could be interesting to assess DNA methylation and histone modifications of a more extensive part (or even whole) HLA class II region, for example, by the above mentioned high-throughput DNA sequencing methods.

Post-transcriptional regulation of HLA class II expression: How does mRNA expression translate into the dimer amount on the cell surface? Measurement of number of class II complexed on the cell surface would be inherently affected by the preference of class II alleles for different peptides.To cancel this effect, we could measure relationship between mRNA and DQ dimer expression in, e.g., B cells and monocytes from the same person (as they have different class II mRNA levels, but same-peptide preference).

Autoimmunity: To explain mechanism of positive or negative association of certain HLA class II alleles with particular autoimmune disease, a comprehensive approach may be needed. This would involve studying not only association between disease and genotype, but also functional properties of resulting DQ dimers such as stability, expression rate, and ability to bind particular autoantigen(s).

7 References

1. Al-Daccak R, Mooney N, Charron D (2004): MHC class II signaling in antigen-presenting cells. *Curr Opin Immunol*. 16: 108–113
2. Cerna M (2008): Genetics of autoimmune diabetes mellitus. *Wiener Medizinische Wochenschrift*. 158: 2–12
3. Alfonso C, Karlsson L (2000): Nonclassical MHC Class II Molecules. *Annu Rev Immunol*. 18: 113–142
4. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SGE (2015): The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res*. 43: D423–D431
5. Schmid D, Pypaert M, Münz C (2007): Antigen-loading compartments for major histocompatibility complex class II molecules continuously receive input from autophagosomes. *Immunity*. 26: 79–92
6. Perry JSA, Hsieh C-S (2016): Development of T-cell tolerance utilizes both cell-autonomous and cooperative presentation of self-antigen. *Immunol Rev*. 271: 141–55
7. Rogers PR, Croft M (1999): Peptide Dose, Affinity, and Time of Differentiation Can Contribute to the Th1/Th2 Cytokine Balance. *J Immunol*. 163: 1205–1213
8. Gottschalk RA, Hathorn MM, Beuneu H, Corse E, Dustin ML, Altan-Bonnet G, Allison JP (2012): Distinct influences of peptide-MHC quality and quantity on in vivo T-cell responses. *Proc Natl Acad Sci*. 109: 881–886
9. Corse E, Gottschalk RA, Krogsgaard M, Allison JP (2010): Attenuated T cell responses to a high-potency ligand in vivo. *PLoS Biol*. 8: e1000481
10. Sewell AK (2012): Why must T cells be cross-reactive? *Nat Rev Immunol*. 12: 669–677
11. Lever M, Lim H-S, Kruger P, Nguyen J, Trendel N, Abu-Shah E, *et al.* (2016): Architecture of a minimal signaling pathway explains the T-cell response to a 1 million-fold variation in antigen affinity and dose. *Proc Natl Acad Sci U S A*. 113: E6630–E6638
12. Gottschalk RA, Corse E, Allison JP (2010): TCR ligand density and affinity determine peripheral induction of Foxp3 in vivo. *J Exp Med*. 207: 1701–11
13. Gallegos AM, Bevan MJ (2004): Central tolerance to tissue-specific antigens mediated by direct and indirect antigen presentation. *J Exp Med*. 200: 1039–49
14. Bautista JL, Lio C-WJ, Lathrop SK, Forbush K, Liang Y, Luo J, *et al.* (2009): Intraclonal competition limits the fate determination of regulatory T cells in the thymus. *Nat Immunol*. 10: 610–7
15. Hosken NA, Shibuya K, Heath AW, Murphy KM, Garra AO (1995): The effect of antigen dose on CD4+ T helper cell phenotype development in a T cell receptor-αβ-transgenic model. *J Exp Med*. 182: 20–22
16. Carballido JM, Carballido-Perrig M, Terres G, Heusser CH, Blaser K (1992): Bee venom phospholipase A; specific T cell clones from human allergic and non-allergic individuals: Cytokine patters change in response to the antigen concentration. *Eur J Immunol*. 22: 1357
17. Constant BS, Pfeiffer C, Woodard A, Pasqualini T, Bottomly K (1995): Extent of T cell receptor ligation can determine the functional differentiation of naive CD4+ T cells. *J Exp Med*. 182: 0–5
18. DiMolfetto L, Neal HA, Wu A, Reilly C, Lo D (1998): The density of the class II MHC T cell receptor ligand influences IFN-gamma/IL-4 ratios in immune responses in vivo. *Cell Immunol*. 183: 70–9
19. Iezzi G, Sonderegger I, Ampenberger F, Schmitz N, Marsland BJ, Kopf M (2009): CD40-CD40L cross-talk integrates strong antigenic signals and microbial stimuli to induce development of IL-17-producing CD4+ T cells. *Proc Natl Acad Sci*. 106: 876–881
20. Shiner EK, Holbrook BC, Alexander-Miller MA (2014): CD4+ T cell subset differentiation and avidity setpoint are dictated by the interplay of cytokine and antigen mediated signals. *PLoS One*. 9: e100175
21. Swanberg M, Lidman O, Padyukov L, Eriksson P, Åkesson E, Jagodic M, *et al.* (2005): MHC2TA is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction. *Nat Genet*. 37: 486–494
22. Alvaro-Benito M, Morrison E, Wieczorek M, Sticht J, Freund C (2016): Human leukocyte Antigen-DM polymorphisms in autoimmune diseases. *Open Biol*. 6: 160165
23. Harndahl M, Rasmussen M, Roder G, Dalgaard Pedersen I, Sørensen M, Nielsen M, Buus S (2012): Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. *Eur J Immunol*. 42: 1405–1416
24. Lazarski CA, Chaves FA, Jenks SA, Wu S, Richards KA, Weaver JM, Sant AJ (2005): The kinetic stability of MHC class II:peptide complexes is a key parameter that dictates immunodominance. *Immunity*. 23: 29–40
25. Hall FC, Rabinowitz JD, Busch R, Visconti KC, Belmares M, Patil NS, *et al.* (2002): Relationship between

kinetic stability and immunogenicity of HLA-DR4/peptide complexes. *Eur J Immunol*. 32: 662–70
26. Nelson CA, Petzold SJ, Unanue ER (1994): Peptides determine the lifespan of MHC class II molecules in the antigen-presenting cell. *Nature*. 371: 250–252
27. Hwang I, Kim K, Choi S, Lomunova M (2017): Potentiation of T Cell Stimulatory Activity by Chemical Fixation of a Weak Peptide-MHC Complex. *Mol Cells*. 40: 24–36
28. Edwards LJ, Evavold BD (2013): Destabilization of peptide:MHC interaction induces IL-2 resistant anergy in diabetogenic T cells. *J Autoimmun*. 44: 82–90
29. Busch DH, Pamer EG (1998): MHC class I/peptide stability: implications for immunodominance, in vitro proliferation, and diversity of responding CTL. *J Immunol*. 160: 4441–8
30. Sant AJ, Chaves FA, Jenks SA, Richards KA, Menges P, Weaver JM, Lazarski CA (2005): The relationship between immunodominance, DM editing, and the kinetic stability of MHC class II:peptide complexes. *Immunol Rev*. 207: 261–278
31. Bondinas GP, Moustakas AK, Papadopoulos GK (2007): The spectrum of HLA-DQ and HLA-DR alleles, 2006: A listing correlating sequence and structure with function. *Immunogenetics*. 59: 539–553
32. Miyadera H, Ohashi J, Lernmark Å, Kitamura T, Tokunaga K, Lernmark A, *et al.* (2015): Cell-surface MHC density profiling reveals instability of autoimmunity-associated HLA. *J Clin Invest*. 125: 275–291
33. Ettinger RA, Liu AW, Nepom GT, Kwok WW (1998): Exceptional Stability of the HLA-DQA1*0102/DQB1*0602 αβ Protein Dimer, the Class II MHC Molecule Associated with Protection from Insulin-Dependent Diabetes Mellitus. *J Immunol*. 161: 6439–6445
34. Mangalam AK, Taneja V, David CS (2013): HLA class II molecules influence susceptibility versus protection in inflammatory diseases by determining the cytokine profile. *J Immunol*. 190: 513–8
35. Kwok WW, Kovats S, Thurtle P, Nepom GT (1993): HLA-DQ allelic polymorphisms constrain patterns of class II heterodimer formation. *J Immunol*. 150: 2263–72
36. van Lummel M, van Veelen PA, Zaldumbide A, de Ru A, Janssen GMC, Moustakas AK, *et al.* (2012): Type 1 diabetes-associated HLA-DQ8 transdimer accommodates a unique peptide repertoire. *J Biol Chem*. 287: 9514–24
37. Reinius B, Sandberg R (2015): Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. *Nat Rev Genet*. 16: 653–64
38. Majumder P, Gomez JA, Chadwick BP, Boss JM (2008): The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J Exp Med*. 205: 785–98
39. Lotteau V, Teyton L, Burroughs D, Charron D (1987): A novel HLA class II molecule (DRα-DQβ) created by mismatched isotype pairing. *Nature*. 329: 339–341
40. Kaushansky N, Eisenstein M, Boura-Halfon S, Hansen BE, Nielsen CH, Milo R, *et al.* (2015): Role of a novel human leukocyte antigen-DQA1∗ 01:02;DRB1∗ 15:01 mixed isotype heterodimer in the pathogenesis of "humanized" multiple sclerosis-like disease. *J Biol Chem*. 290: 15260–15278
41. Neefjes J, Jongsma ML, Paul P, Bakke O (2011): Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol*. 11: 823–36
42. Busch R, Rinderknecht CH, Roh S, Lee AW, Harding JJ, Burster T, *et al.* (2005): Achieving stability through editing and chaperoning: regulation of MHC class II peptide binding and expression. *Immunol Rev*. 207: 242–60
43. Rinderknecht CH, Roh S, Pashine A, Belmares MP, Patil NS, Lu N, *et al.* (2010): DM influences the abundance of major histocompatibility complex class II alleles with low affinity for class II-associated invariant chain peptides via multiple mechanisms. *Immunology*. 131: 18–32
44. Denzin LK (2013): Inhibition of HLA-DM Mediated MHC Class II Peptide Loading by HLA-DO Promotes Self Tolerance. *Front Immunol*. 4: 465
45. Hornell TMC, Burster T, Jahnsen FL, Pashine A, Ochoa MT, Harding JJ, *et al.* (2006): Human dendritic cell expression of HLA-DO is subset specific and regulated by maturation. *J Immunol*. 176: 3536–47
46. Yi W, Seth NP, Martillotti T, Wucherpfennig KW, Sant'Angelo DB, Denzin LK (2010): Targeted regulation of self-peptide presentation prevents type I diabetes in mice without disrupting general immunocompetence. *J Clin Invest*. 120: 1324–36
47. Zhou Z, Reyes-Vargas E, Escobar H, Rudd B, Rockwood AL, Delgado JC, *et al.* (2016): Type 1 diabetes associated HLA-DQ2 and DQ8 molecules are relatively resistant to HLA-DM mediated release of invariant chain-derived CLIP peptides. *Eur J Immunol*. 46: 834–45
48. Bronson PG, Caillier S, Ramsay PP, McCauley JL, Zuvich RL, De Jager PL, *et al.* (2010): CIITA variation in

the presence of HLA-DRB1*1501 increases risk for multiple sclerosis. *Hum Mol Genet*. 19: 2331–40
49. Raj P, Rai E, Song R, Khan S, Wakeland BE, Viswanathan K, *et al.* (2016): Regulatory polymorphisms modulate the expression of HLA class II molecules and promote autoimmunity. *Elife*. 5: e12089
50. Vander Lugt B, Khan A a, Hackney J a, Agrawal S, Lesch J, Zhou M, *et al.* (2014): Transcriptional programming of dendritic cells for enhanced MHC class II antigen presentation. *Nat Immunol*. 15: 161–7
51. Oh J, Shin JS (2015): Molecular mechanism and cellular function of MHCII ubiquitination. *Immunol Rev*. 266: 134–144
52. Sinnathamby G, Eisenlohr LC (2003): Presentation by Recycling MHC Class II Molecules of an Influenza Hemagglutinin-Derived Epitope That Is Revealed in the Early Endosome by Acidification. *J Immunol*. 170: 3504–13
53. Pinet V, Vergelli M, Martin R, Bakke O, Long EO (1995): Antigen presentation mediated by recycling of surface HLA-DR molecules. *Nature*. 375: 603–606
54. Santambrogio L, Sato AK, Fischer FR, Dorf ME, Stern LJ (1999): Abundant empty class II MHC molecules on the surface of immature dendritic cells. *Proc Natl Acad Sci U S A*. 96: 15050–5.
55. Roche PA, Furuta K (2015): The ins and outs of MHC class II-mediated antigen processing and presentation. *Nat Rev Immunol*. 15: 203–216
56. Cella M, Engering A, Pinet V (1997): Inflammatory stimuli induce accumulation of MHC class II complexes on dendritic cells. 388: 782–787
57. Shin J-S, Ebersold M, Pypaert M, Delamarre L, Hartley A, Mellman I (2006): Surface expression of MHC class II in dendritic cells is controlled by regulated ubiquitination. *Nature*. 444: 115–118
58. Jahnke M, Trowsdale J, Kelly AP (2012): Structural requirements for recognition of major histocompatibility complex class II by membrane-associated RING-CH (MARCH) protein E3 ligases. *J Biol Chem*. 287: 28779–28789
59. De Riva A, Varley MC, Bluck LJ, Cooke A, Deery MJ, Busch R (2013): Accelerated Turnover of MHC Class II Molecules in Nonobese Diabetic Mice Is Developmentally and Environmentally Regulated In Vivo and Dispensable for Autoimmunity. *J Immunol*. 190: 5961–5971
60. Prevosto C, Usmani MF, McDonald S, Gumienny AM, Key T, Goodman RS, *et al.* (2016): Allele-independent turnover of human leukocyte antigen (HLA) class Ia molecules. *PLoS One*. 11: 1–25
61. The MHC sequencing consortium (1999): Complete sequence and gene map of a human major histocompatibility complex. *Nature*. 401: 921–3
62. Andersson G (1998): Evolution of the human HLA-DR region. *Front Biosci*. 3: d739-45
63. Lenormand C, Bausinger H, Gross F, Signorino-Gelo F, Koch S, Peressin M, *et al.* (2012): HLA-DQA2 and HLA-DQB2 genes are specifically expressed in human Langerhans cells and encode a new HLA class II molecule. *J Immunol*. 188: 3903–11
64. Reith W, LeibundGut-Landmann S, Waldburger J-M (2005): Regulation of MHC class II gene expression by the class II transactivator. *Nat Rev Immunol*. 5: 793–806
65. Handunnetthi L, Ramagopalan S V, Ebers GC, Knight JC (2010): Regulation of major histocompatibility complex class II gene expression, genetic variation and disease. *Genes Immun*. 11: 99–112
66. Mischiati C, Feriotto G, Bianchi N, Rutigliano C, Giacomini P, Gambari R (1999): Analysis of the human HLA-DRA gene upstream region: Evidence for a stem-loop array directed by nuclear factors. *Biochimie*. 81: 219–228
67. van den Elsen PJ (2011): Expression regulation of major histocompatibility complex class I and class II encoding genes. *Front Immunol*. 2: 1–9
68. Benoist C, Mathis D (1990): Regulation of major histocompatibility complex class-II genes: X, Y and other letters of the alphabet. *Annu Rev Immunol*. 8: 681–715
69. Muhlethaler-Mottet A, Krawczyk M, Masternak K, Spilianakis C, Kretsovali A, Papamatheakis J, Reith W (2004): The S box of major histocompatibility complex class II promoters is a key determinant for recruitment of the transcriptional co-activator CIITA. *J Biol Chem*. 279: 40529–40535
70. Zhu XS, Linhoff MW, Li G, Chin KC, Maity SN, Ting JP (2000): Transcriptional scaffold: CIITA interacts with NF-Y, RFX, and CREB to cause stereospecific regulation of the class II major histocompatibility complex promoter. *Mol Cell Biol*. 20: 6051–61
71. Beaty JS, West KA, Nepom GT (1995): Functional effects of a natural polymorphism in the transcriptional regulatory sequence of HLA-DQB1. *Mol Cell Biol*. 15: 4771–82
72. Reith W, Mach B (2001): The Bare Lymphoycyte Syndrome and the Regulation of MHC Expression. *Annu Rev Immunol*. 19: 331–73

73. Reith W, Barras E, Satola S, Kobr M (1989): Cloning of the major histocompatibility complex class II promoter binding protein affected in a hereditary defect in class II gene regulation. 86: 4200–4204
74. Hasegawa SL, Sloan JH, Reith W, Mach B, Boss JM (1991): Regulatory factor-X binding to mutant HLA-DRA promoter sequences. *Nucleic Acids Res*. 19: 1243–9
75. Morzycka-Wroblewska E, Munshi A, Ostermayer M, Harwood JI, Kagnoff MF (1997): Differential expression of HLA-DQA1 alleles associated with promoter polymorphism. *Immunogenetics*. 45: 163–70
76. Moreno CS, Beresford GW, Louis-plence P, Morris AC, Boss JM (1999): CREB Regulates MHC Class II Expression in a CIITA-Dependent Manner. 10: 143–151
77. Westerheide SD, Boss JM (1999): Orientation and positional mapping of the subunits of the multicomponent transcription factors RFX and X2BP to the major histocompatibility complex class II transcriptional enhancer. 27: 1635–1641
78. Wang J, Zhuang J, Iyer S, Lin X-Y, Greven MC, Kim B-H, *et al.* (2013): Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Res*. 41: D171–D176
79. Mayr B, Montminy M (2001): Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nat Rev Mol Cell Biol*. 2: 599–609
80. Angel P, Imagawa M, Chiu R, Stein B, Imbra RJ, Rahmsdorf HJ, *et al.* (1987): Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. *Cell*. 49: 729–39
81. Ono SJ, Bazil V, Levi BZ, Ozato K, Strominger JL (1991): Transcription of a subset of human class II major histocompatibility complex genes is regulated by a nucleoprotein complex that contains c-fos or an antigenically related protein. *Proc Natl Acad Sci U S A*. 88: 4304–8
82. Vilen BJ, Cogswell JP, Ting JP (1991): Stereospecific alignment of the X and Y elements is required for major histocompatibility complex class II DRA promoter function. *Mol Cell Biol*. 11: 2406–15
83. Nardone V, Chaves-Sanjuan A, Nardini M (2017): Structural determinants for NF-Y/DNA interaction at the CCAAT box. *Biochim Biophys Acta*. 1860: 571–580
84. Robinson J, Malik A, Parham P, Bodmer J, Marsh S (2000): IMGT/HLA - a sequence database for the human major histocompatibility complex. *Tissue Antigens*. 55: 280–287
85. Indovina P, Megiorni F, Ferrante P, Apollonio I, Petronzelli F, Mazzilli MC (1998): Different binding of NF-Y transcriptional factor to DQA1 promoter variants. *Hum Immunol*. 59: 758–767
86. Adams BS, Cha HC, Cleary J, Haiying T, Wang H, Sitwala K, Markovitz DM (2003): DEK binding to class II MHC Y-box sequences is gene- and allele-specific. *Arthritis Res Ther*. 5: 226–233
87. Nolan D, Castley A, Tschochner M, James I, Qiu W, Sayer D, *et al.* (2012): Contributions of vitamin D response elements and HLA promoters to multiple sclerosis risk. *Neurology*. 79: 538–546
88. Freund YR, Dedrick RL, Jones PP (1990): cis-acting sequences required for class II gene regulation by interferon gamma and tumor necrosis factor alpha in a murine macrophage cell line. *J Exp Med*. 171: 1283–1299
89. Morzycka-Wroblewska E, Harwood JI, Smith JR, Kagnoff MF (1993): Structure and evolution of the promoter regions of the DQA genes. *Immunogenetics*. 37: 364–72
90. Otten LA, Steimle V (1998): Quantitative control of MHC class II expression by the transactivator CIITA. *Eur J Immunol*. 28: 473–478
91. Douhan J, Lieberson R, Knoll JH, Zhou H, Glimcher LH (1997): An isotype-specific activator of major histocompatibility complex (MHC) class II genes that is independent of class II transactivator. *J Exp Med*. 185: 1885–95
92. Cazalis M, Friggeri A, Cavé L, Demaret J, Barbalat V, Cerrato E, *et al.* (2013): Decreased HLA-DR antigen-associated invariant chain (CD74) mRNA expression predicts mortality after septic shock. *Crit Care*. 17: R287
93. Ting JP, Trowsdale J, Hill C, Carolina N, Allcock F (2002): Genetic Control of MHC Class II Expression. *Cell*. 109: S21–S33
94. LeibundGut-Landmann S, Waldburger J-M, Krawczyk M, Otten LA, Suter T, Fontana A, *et al.* (2004): Mini-review: Specificity and expression of CIITA, the master regulator of MHC class II genes. *Eur J Immunol*. 34: 1513–25
95. Devaiah BN, Singer DS (2013): CIITA and Its Dual Roles in MHC Gene Transcription. *Front Immunol*. 4: 476
96. Rybtsova N, Leimgruber E, Seguin-Estévez Q, Dunand-Sauthier I, Krawczyk M, Reith W (2007): Transcription-coupled deposition of histone modifications during MHC class II gene activation. *Nucleic Acids*

Res. 35: 3431–41
97. Wu X, Kong X, Luchsinger L, Smith BD, Xu Y (2009): Regulating the activity of class II transactivator by posttranslational modifications: exploring the possibilities. *Mol Cell Biol*. 29: 5639–5644
98. Muhlethaler-Mottet A, Otten LA, Steimle V, Mach B (1997): Expression of MHC class II molecules in different cellular and functional compartments is controlled by differential usage of multiple promoters of the transactivator CIITA. *EMBO J*. 16: 2851–60
99. Wong AW, Ghosh N, McKinnon KP, Reed W, Piskurich JF, Wright KL, Ting JP-Y (2002): Regulation and Specificity of MHC2TA Promoter Usage in Human Primary T Lymphocytes and Cell Line. *J Immunol*. 169: 3112–3119
100. Beaulieu YB, Leon Machado JA, Ethier S, Gaudreau L, Steimle V (2016): Degradation, Promoter Recruitment and Transactivation Mediated by the Extreme N-Terminus of MHC Class II Transactivator CIITA Isoform III. *PLoS One*. 11: e0148753
101. Reith W, Mach B (2001): The Bare Lymphocyte Syndrome and the Regulation of MHC Expression. *Annu Rev Immunol*. 19: 331–73
102. Feriotto G, Mischiati C, Bianchi N, Rutigliano C, Giacomini P, Gambari R (1995): Sequencing of an upstream region of the human HLA-DRA gene containing X' and Y' boxes. *Nucleic Acids Res*. 23: 1671–8
103. Masternak K, Peyraud N, Krawczyk M, Barras E, Reith W (2003): Chromatin remodeling and extragenic transcription at the MHC class II locus control region. *Nat Immunol*. 4: 132–7
104. Krawczyk M, Peyraud N, Rybtsova N, Masternak K (2004): Long Distance Control of MHC Class II Expression by Multiple Distal Enhancers Regulated by Regulatory Factor X. *J Immunol*. 173: 6200–6210
105. Gomez JA, Majumder P, Nagarajan UM, Boss JM (2005): X Box-Like Sequences in the MHC Class II Region Maintain Regulatory Function. *J Immunol*. 175: 1030–1040
106. Ghirlando R, Felsenfeld G (2016): CTCF: Making the right connections. *Genes Dev*. 30: 881–891
107. de Wit E, Vos ESM, Holwerda SJB, Valdes-Quezada C, Verstegen MJAM, Teunissen H, *et al.* (2015): CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell*. 60: 676–684
108. Majumder P, Boss JM (2010): CTCF Controls Expression and Chromatin Architecture of the Human Major Histocompatibility Complex Class II Locus. *Mol Cell Biol*. 30: 4211–4223
109. Majumder P, Gomez JA, Boss JM (2006): The Human Major Histocompatibility Complex Class II HLA-DRB1 and HLA-DQA1 Genes Are Separated by a CTCF-binding Enhancer-blocking Element. 281: 18435–18443
110. Majumder P, Boss JM (2011): DNA methylation dysregulates and silences the HLA-DQ locus by altering chromatin architecture. *Genes Immun*. 12: 291–9
111. Sullivan KE, Peterlin BM (1987): Transcriptional enhancers in the HLA-DQ subregion. *Mol Cell Biol*. 7: 3315–3319
112. Robinson J, Halliwell J, Hayhurst J, Flicek P, Parham P, Marsh S (2015): The IPD and IPD-IMGT/HLA Database: allele variant databases. *Nucleic Acids Res*. 43: D423-431
113. Benoist C, Mathis D (1990): Regulation of major histocompatibility complex class-II genes: X, Y and other letters of the alphabet. *Annu Rev Immunol*. 8: 681–715
114. Wang BY, Larsen AS, Peterlin BM (1987): A tissue-specific transcriptional enhancer is found in the body of the HLA-DR alpha gene. *J Exp Med*. 166: 625–36
115. Balas A, Aviles MJ, Alonso-Nieto M, Zarapuz L, Blanco L, García-Sánchez F, Vicario JL (2005): HLA-DQA1 introns 2 and 3 sequencing: DQA1 sequencing-based typing and characterization of a highly polymorphic microsatellite at intron 3 of DQA1*0505. *Hum Immunol*. 66: 903–911
116. Caillat-Zucman S (2017): New insights into the understanding of MHC associations with immune-mediated disorders. *HLA*. 89: 3–13
117. Senju S, Kimura A, Yasunami M, Kamikawaji N, Yoshizumi H, Sasazuki T (1992): Allele-specific expression of the cytoplasmic exon of HLA-DQB1 gene. *Immunogenetics*. 36: 319–325
118. Harton J, Jin L, Hahn A, Drake J (2016): Immunological Functions of the Membrane Proximal Region of MHC Class II Molecules. *F1000Research*. 5: 368
119. Zajacova M, Kotrbova-Kozak A, Cerna M (2016): HLA-DRB1, -DQA1 and -DQB1 genotyping of 180 Czech individuals from the Czech Republic pop 3. *Hum Immunol*. 77: 365–366
120. Briata P, Radka SF, Sartoris S, Lee JS (1989): Alternative splicing of HLA-DQB transcripts and secretion of HLA-DQ beta-chain proteins: allelic polymorphism in splicing and polyadenylylation sites. *Proc Natl Acad Sci U S A*. 86: 1003–1007
121. Seidl C, Lee JS (1992): Expression of alternatively spliced HLA class II transcripts in lymphoid and

nonlymphoid tissues. *Immunogenetics*. 35: 385–90
122. Cesari M, Caillens H, Cadet F, Pabion M, Cesari M, Cadet F, Pabion M (1999): In vivo analysis of HLA-DQ gene expression in heterozygous cell lines. *Immunogenetics*. 50: 309–318
123. Puppo F, Contini P, Ghio M, Brenci S, Scudeletti M, Filaci G, *et al.* (2000): Soluble human MHC class I molecules induce soluble Fas ligand secretion and trigger apoptosis in activated CD8+ Fas (CD95)+ T lymphocytes. *Int Immunol*. 12: 195–203
124. Hoarau J-J, Cesari M, Caillens H, Cadet F, Pabion M (2004): HLA DQA1 genes generate multiple transcripts by alternative splicing and polyadenylation of the 3' untranslated region. *Tissue Antigens*. 63: 58–71
125. Pisapia L, Cicatiello V, Barba P, Malanga D, Maffei A, Hamilton RS, Del Pozzo G (2013): Co-regulated expression of alpha and beta mRNAs encoding HLA-DR surface heterodimers is mediated by the MHCII RNA operon. *Nucleic Acids Res*. 41: 3772–3786
126. Pisapia L, Camarca A, Picascia S, Bassi V, Barba P, Del Pozzo G, Gianfrani C (2016): HLA-DQ2.5 genes associated with celiac disease risk are preferentially expressed with respect to non-predisposing HLA genes: Implication for anti-gluten T cell response. *J Autoimmun*. 70: 63–72
127. Klose RJ, Bird AP (2006): Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci*. 31: 89–97
128. Vaissiere T, Sawan C, Herceg Z (2008): Epigenetic interplay between histone modifications and DNA methylation in gene silencing. *Mutat Res*. 659: 40–48
129. Cheng X, Blumenthal RM (2008): Mammalian DNA methyltransferases: a structural perspective. *Structure*. 16: 341–350
130. Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, Schübeler D (2007): Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet*. 39: 457–66
131. Shen H, Qiu C, Li J, Tian Q, Deng H-W (2013): Characterization of the DNA methylome and its interindividual variation in human peripheral blood monocytes. *Epigenomics*. 5: 255–69
132. Hural JA, Kwan M, Henkel G, Benjamin M, Brown MA, Hock MB, Brown MA (2000): An Intron Transcriptional Enhancer Element Regulates IL-4 Gene Locus Accessibility in Mast Cells. *J Immunol*. 165: 3239–3249
133. Hendrich B, Tweedie S (2003): The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet*. 19: 269–277
134. Cheng X, Blumenthal RM (2010): Coordinated chromatin control: structural and functional linkage of DNA and histone methylation. *Biochemistry*. 49: 2999–3008
135. Lan J, Hua S, He X, Zhang Y (2010): DNA methyltransferases and methyl-binding proteins of mammals. *Acta Biochim Biophys Sin (Shanghai)*. 42: 243–52
136. Hervouet E, Peixoto P, Delage-Mourroux R, Boyer-Guittaut M, Cartron P-F (2018): Specific or not specific recruitment of DNMTs for DNA methylation, an epigenetic dilemma. *Clin Epigenetics*. 10: 17
137. Jaenisch R, Bird A (2003): Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*. 33 Suppl: 245–54
138. Ruiz-Hernandez A, Kuo C-C, Rentero-Garrido P, Tang W-Y, Redon J, Ordovas JM, *et al.* (2015): Environmental chemicals and DNA methylation in adults: a systematic review of the epidemiologic evidence. *Clin Epigenetics*. 7: 55
139. Wu Z, Li X, Qin H, Zhu X, Xu J, Shi W (2013): Ultraviolet B enhances DNA hypomethylation of CD4+ T cells in systemic lupus erythematosus via inhibiting DNMT1 catalytic activity. *J Dermatol Sci*. 71: 167–173
140. Gaine ME, Chatterjee S, Abel T (2018): Sleep Deprivation and the Epigenome. *Front Neural Circuits*. 12: 14
141. Wolff GL, Kodell RL, Moore SR, Cooney CA (1998): Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *FASEB J*. 12: 949–57
142. Lo HS, Wang Z, Hu Y, Yang HH, Gere S, Buetow KH, Lee MP (2003): Allelic Variation in Gene Expression Is Common in the Human Genome. *Genome Res*. 13: 1855–1862
143. Britten AC, Mijovic CH, Barnett AH, Kelly MA (2009): Differential expression of HLA-DQ alleles in peripheral blood mononuclear cells: alleles associated with susceptibility to and protection from autoimmune type 1 diabetes. *Int J Immunogenet*. 36: 47–57
144. Ghotbi R, Gomez A, Milani L, Tybring G, Syvänen A-C, Bertilsson L, *et al.* (2009): Allele-specific expression and gene methylation in the control of CYP1A2 mRNA level in human livers. *Pharmacogenomics J*. 9: 208–17
145. Zhang Y, Rohde C, Reinhardt R, Voelcker-Rehage C, Jeltsch A (2009): Non-imprinted allele-specific DNA

methylation on human autosomes. *Genome Biol*. 10: R138
146. Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, *et al.* (2008): Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet*. 40: 904–8
147. Feriotto G, Pozzi L, Ciucci A, Camarda G, Mischiati C, D'Agnano I, *et al.* (1993): Methylation state of the human HLA-DRA gene in T-lymphocytes and B-lymphocytes of transgenic mice. Lack of methylation at one 5'-GCGC site is not required for gene expression. *Eur J Biochem*. 218: 485–92
148. Barbieri R, Nastruzzi C, Volinia S, Villa M, Piva R, Giacomini P, *et al.* (1990): Methylation pattern of the HLA-DR alpha gene in human tissues. *J Immunogenet*. 17: 51–66
149. Gambari R, Barbieri R, Piva R, Tecce R, Fisher PB, Giacomini P, Natali PG (1987): Regulation of the expression of class II genes of the human major histocompatibility complex in tumor cells. *Ann N Y Acad Sci*. 511: 292–307
150. Niesen MI, Osborne AR, Yang H, Rastogi S, Chellappan S, Cheng JQ, *et al.* (2005): Activation of a methylated promoter mediated by a sequence-specific DNA-binding protein, RFX. *J Biol Chem*. 280: 38914–22
151. Carrington MN, Salter RD, Cresswell P, Ting JP (1985): Evidence for methylation as a regulatory mechanism in HLA-DR alpha gene expression. *Immunogenetics*. 22: 219–29
152. Wang Y, Peterlin BM (1986): Methylation patterns of HLA-DR alpha genes in six mononuclear cell lines. *Immunogenetics*. 24: 298–303
153. Lee JS, O'Neill L (1987): Methylation of the HLA-DR alpha gene is positively correlated with expression. *Immunogenetics*. 26: 92–8
154. Levine F, Pious D (1985): Different roles for cytosine methylation in HLA class II gene expression. *Immunogenetics*. 22: 427–40
155. Reitz MS, Mann DL, Eiden M, Michael F (1984): DNA Methylation and Expression of HLA-DR alpha. *Mol Cell Biol*. 4: 890–897
156. Seguín-Estévez Q, De Palma R, Krawczyk M, Leimgruber E, Villard J, Picard C, *et al.* (2009): The transcription factor RFX protects MHC class II genes against epigenetic silencing by DNA methylation. *J Immunol*. 183: 2545–53
157. Uitterlinden AG, Vijg J, Giphart MJ, Knook DL (1985): Variation in restriction fragment length and methylation pattern of rat MHC class I genes. *Exp Clin Immunogenet*. 2: 215–22
158. Toyoda H, Redford A, Magalong D (1992): Allele-specific methylation in the 5'-regulatory region of class II DQ beta genes in the human major histocompatibility complex (MHC): relationship to autoimmune disease susceptibility. *Dis Markers*. 10: 7–18
159. Rakyan VK, Beyan H, Down TA, Hawa MI, Maslau S, Aden D, *et al.* (2011): Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet*. 7: e1002300
160. Taudt A, Colomé-tatché M, Johannes F (2016): Genetic sources of population epigenomic variation. *Nat Rev Genet*. 17: 319–332
161. Rose NR, Klose RJ (2014): Understanding the relationship between DNA methylation and histone lysine methylation. *Biochim Biophys Acta*. 1839: 1362–72
162. Chou S-D, Tomasi TB (2008): Spatial distribution of histone methylation during MHC class II expression. *Mol Immunol*. 45: 971–80
163. Kasowski M, Kyriazopoulou-panagiotopoulou S, Grubert F, Judith B, Kundaje A, Liu Y, *et al.* (2013): Extensive Variation in Chromatin States Across Humans. *Science*. 342: 750–752
164. McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A, *et al.* (2013): Identification of genetic variants that affect histone modifications in human cells. *Science*. 342: 747–9
165. McDaniell R, Lee B, Song L, Liu Z, Boyle AP, Michael R, *et al.* (2010): Heritable Individual-Specific and Allele-Specific Chromatin Signatures in Humans. *Science*. 328: 235–239
166. Miao F, Smith DD, Zhang L, Min A, Feng W, Natarajan R (2008): Lymphocytes From Patients With Type 1 Diabetes Display a Distinct Profile of Chromatin Histone H3 Lysine 9 Dimethylation. *Diabetes*. 57: 3190–3198
167. Miao F, Chen Z, Zhang L, Liu Z, Wu X, Yuan YC, Natarajan R (2012): Profiles of epigenetic histone post-translational modifications at type 1 diabetes susceptible genes. *J Biol Chem*. 287: 16335–16345
168. Miao F, Gonzalo IG, Lanting L, Natarajan R (2004): In Vivo Chromatin Remodeling Events Leading to Inflammatory Gene Transcription under Diabetic Conditions. *J Biol Chem*. 279: 18091–18097
169. Miller SA, Dykes DD, Polesky HF (1988): A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 16: 1215
170. Lamborová V (2013): *Epigenetic regulation of HLA class II genes and its modification during lifetime*.

Diploma Thesis. Charles University in Prague, Faculty of Science
171. Sambrook J, Fritsch E, Maniatis T (1989): *Molecular Cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press
172. Fernandez S, Wassmuth R, Knerr I, Frank C, Haas JP (2003): Relative quantification of HLA-DRA1 and -DQA1 expression by real-time reverse transcriptase-polymerase chain reaction (RT-PCR). *Eur J Immunogenet*. 30: 141–8
173. Yu L, Sheehy M (1991): The cryptic HLA-DQA2 ("DX alpha") gene is expressed in human B cell lines. *J Immunol*. 15: 4393–7
174. Zajacova M, Kotrbova-Kozak A, Cepek P, Cerna M (2015): Differences in promoter DNA methylation and mRNA expression of individual alleles of the HLA class II DQA1 gene. *Immunol Lett*. 167: 147–54
175. Zajacova M, Kotrbova-Kozak A, Cerna M (2018): Expression of HLA-DQA1 and HLA-DQB1 genes in B lymphocytes, monocytes and whole blood. *Int J Immunogenet*.In Press. doi: 10.1111/iji.12367
176. Zajacová M (2010): *Assessment of expression markers of HLA class II genes*. Diploma Thesis. Charles University in Prague, Faculty of Science
177. Gonzalez-Galarza FF, Takeshita LYC, Santos EJM, Kempson F, Maia MHT, Silva ALS d., *et al.* (2015): Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res*. 43: D784–D788
178. Brünnler G, Haas J, Fan L, Petzl-Erler M, Volgger A, Yao Z, *et al.* (1997): DQA1 promoter polymorphism: 12th International Histocompatibility Workshop study. *Proc Twelfth Int Histocompat Work Conf Vol I Work*. (Vol. 1), pp 171–175
179. Murphy K, Travers P, Walport M, Janeway C (2012): *Immunobiology*, 8th ed. Garland Science
180. Rosser EC, Mauri C (2015): Regulatory B cells: origin, phenotype, and function. *Immunity*. 42: 607–12
181. Wong KL, Yeap WH, Tai JJY, Ong SM, Dang TM, Wong SC (2012): The three human monocyte subsets: Implications for health and disease. *Immunol Res*. 53: 41–57
182. Jakubzick C V., Randolph GJ, Henson PM (2017): Monocyte differentiation and antigen-presenting functions. *Nat Rev Immunol*. 17: 349–362
183. Ziegler-Heitbrock L (2006): The CD14+ CD16+ blood monocytes: their role in infection and inflammation. *J Leukoc Biol*. 81: 584–592
184. Sathaliyawala T, Kubota M, Yudanin N, Turner D, Camp P, Thome JJC, *et al.* (2013): Distribution and compartmentalization of human circulating and tissue-resident memory T cell subsets. *Immunity*. 38: 187–97
185. Cepek P, Zajacova M, Kotrbova-Kozak A, Silhova E, Cerna M (2016): DNA methylation and mRNA expression of HLA-DQA1 alleles in type 1 diabetes mellitus. *Immunology*. 148: 150–159
186. Klose RJ, Sarraf SA, Schmiedeberg L, McDermott SM, Stancheva I, Bird AP (2005): DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol Cell*. 19: 667–78
187. Clouaire T, de Las Heras JI, Merusi C, Stancheva I (2010): Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA. *Nucleic Acids Res*. 38: 4620–34
188. Prokhortchouk A, Hendrich B, Jørgensen H, Ruzov A, Wilm M, Georgiev G, *et al.* (2001): The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev*. 15: 1613–8
189. Chmel M (2015): *The study of epigenetic regulation of gene HLA II. Class within family relationships*. Diploma Thesis. Charles University in Prague, Faculty of Science
190. Cernoch M (2014): *The study of epigenetic regulation of HLA class II genes at the level of histone modifications*. Diploma Thesis. Charles University in Prague, Faculty of Science
191. Donner H, Seidl C, Rau H, Herwig J, Seifried E, Usadel KH, Badenhoop K (2002): Unbalanced amounts of HLA-DQA1 allele mRNA: DQA1*03 shows high and DQA1*0501 low amounts of mRNA in heterozygous individuals. *Eur J Immunogenet*. 29: 321–30
192. Maffei A, Pozzo GD, Prisco A, Ciullo M, Harris PE, Reed EF, Guardiola J (1997): Polymorphism in the 5' terminal region of the mRNA of HLA-DQA1 gene: identification of four groups of transcripts and their association with polymorphism in the alpha 1 domain. *Hum Immunol*. 53: 167–73
193. Maffei A, Harris PE, Reed EF, Del Pozzo G, Ciullo M, Suciu-Foca N, Guardioloa J (1997): Differential expression of insulin-dependent diabetes mellitus-associated HLA-DQA1 alelles in vivo. *Eur J Immunol*. 27: 1549–1556
194. Nepom GT, Chung J, West KA (1995): Differential expression of HLA-DQB1 alleles in a heterozygous cell line. *Immunogenetics*. 42: 143–8
195. Beaty JS, Sukiennicki TL, Nepom GT (1999): Allelic variation in transcription modulates MHC class II expression and function. *Microbes Infect*. 1: 919–927

196. Li M, Song L-J, Qin X-Y (2014): Advances in the cellular immunological pathogenesis of type 1 diabetes. *J Cell Mol Med*. 18: 749–58
197. Polychronakos C, Li Q (2011): Understanding type 1 diabetes through genetics: Advances and prospects. *Nat Rev Genet*. 12: 781–792
198. Jones EY, Fugger L, Strominger JL, Siebold C (2006): MHC class II proteins and disease: a structural perspective. *Nat Rev Immunol*. 6: 271–282
199. Unanue ER (2014): *Antigen Presentation in the Autoimmune Diabetes of the NOD Mouse*. *Annu Rev Immunol*. (Vol. 32). doi: 10.1146/annurev-immunol-032712-095941.
200. Astill TP, Ellis RJ, Arif S, Tree TIM, Peakman M (2003): Promiscuous binding of proinsulin peptides to Type 1 diabetes-permissive and -protective HLA class II molecules. *Diabetologia*. 46: 496–503
201. Erlich H, Valdes AM, Noble J, Carlson JA, Varney M, Concannon P, *et al.* (2008): HLA DR-DQ Haplotypes and Genotypes and Type 1 Diabetes Risk: Analysis of the Type 1 Diabetes Genetics Consortium Families. 57: 1084–1092
202. Nepom GT (1990): A unified hypothesis for the complex genetics of HLA associations with IDDM. *Diabetes*. 39: 1153–1157
203. Parry CS, Brooks BR (2008): A new model defines the minimal set of polymorphism in HLA-DQ and -DR that determines susceptibility and resistance to autoimmune diabetes. *Biol Direct*. 3: 42
204. Stefan M, Zhang W, Concepcion E, Yi Z, Tomer Y (2014): DNA methylation profiles in type 1 diabetes twins point to strong epigenetic effects on etiology. *J Autoimmun*. 50: 33–37
205. Rakyan VK, Hildmann T, Novik KL, Lewin J, Tost J, Cox A V, *et al.* (2004): DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol*. 2: e405
206. Disanto G, Vcelakova J, Pakpoor J, Elangovan RI, Sumnik Z, Ulmannova T, *et al.* (2013): DNA methylation in monozygotic quadruplets affected by type 1 diabetes. *Diabetologia*. 56: 2093–2095
207. Erlich H, Valdes AM, Noble J, Carlson JA, Varney M, Concannon P, *et al.* (2008): HLA DR-DQ Haplotypes and Genotypes and Type 1 Diabetes Risk: Analysis of the Type 1 Diabetes Genetics Consortium Families. *Diabetes*. 57: 1084–1092
208. Miao F, Wu X, Zhang L, Yuan YC, Riggs AD, Natarajan R (2007): Genome-wide analysis of histone lysine methylation variations caused by diabetic conditions in human monocytes. *J Biol Chem*. 282: 13854–13863
209. Schmidl C, Klug M, Boeld TJ, Andreesen R, Hoffmann P, Edinger M, Rehli M (2009): Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res*. 19: 1165–74
210. Waite LL, Weaver B, Day K, Li X, Roberts K, Gibson AW, *et al.* (2016): Estimation of Cell-Type Composition Including T and B Cell Subtypes for Whole Blood Methylation Microarray Data. *Front Genet*. 7: 23
211. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A (2010): The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One*. 5: e8888
212. Shen L, Zhang Y (2013): 5-Hydroxymethylcytosine: generation, fate, and genomic distribution. *Curr Opin Cell Biol*. 25: 289–96
213. Teif VB, Beshnova DA, Vainshtein Y, Marth C, Mallm J-P, Höfer T, Rippe K (2014): Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. *Genome Res*. 24: 1285–95