Chapter 7

General Discussion

7.1 Introduction

In this chapter, data from chapters 3, 4, 5 and 6 are discussed in the context of recent studies concerning the epigenetics and regulation of MHC gene expression. A discussion on array-based assays for the identification of differentially methylated regions (DMRs) (chapter 3) is followed by discussions on the two DMR screens I performed (tDMR and pDMR screens) and are described in chapters 4, 5 and 6. Plans for future work following the work described in this thesis are also presented.

Finally, I introduce and discuss: (i). the phenomenon of Long Range Epigenetic Silencing (section 7.5), and (ii). the association between recombination hotspots and epigenetic events (section 7.6). These two concepts, although not discussed before in this thesis, are both relevant to the MHC region and should be considered for future MHC-related studies.

7.2 Array-based assay for DMR identification

I constructed a 2kb genomic tiling array of the entire MHC region. At the time of the array design, whole genome tiling arrays were constructed from PACs and BACs resulting in an approximate resolution of 100kb (Fiegler et al., 2006). Although commercial arrays are now available at much higher resolution (5 - 50mers), the MHC tiling array is still of great value today, as it can be used for multiple applications and is freely available from the Wellcome Trust Sanger Institute Microarray facility.

With respect to applications, the array is compatible with chromatin immunoprecipitation (ChIP), methylated DNA immunoprecipitation (MeDIP), array comparative genomic hybridization (aCGH) and expression analysis. It can be used to: (i). generate histone modification and DNA methylation maps, (ii). to study structural variation (CNVs) and (iii). to generate gene expression profiles and strand-specific transcript maps along the MHC. Hence, this platform in combination with the abundant data regarding single

nucleotide polymorphisms (SNPs) within the MHC, is a great resource for studying how genetic and epigenetic variation interact and how this interplay affects expression patterns which could eventually result in MHC-linked complex diseases.

This array has been used here for DNA methylation analysis and DMR identification and, by another group, for the identification of chromatin loops within the MHC (Ottaviani, 2008), underlining the multiple-purpose design of this array.

7.2.1 Future directions

Today, oligonucleotide tiling arrays with increasingly high probe densities and improved coverage, resolution and cost-effectiveness have enabled high-resolution studies of cytosine methylation when combined with MeDIP. Such arrays have already been used for the completion of the first high-resolution analysis of the *A. thaliana* methylome (Zhang et al., 2008). In addition, the development of a novel algorithm employing a Bayesian de-convolution strategy to normalize MeDIP array data can be expected to further increase the potential of high-resolution arrays for DNA methylation analysis (Down et al., 2008).

Hence, if I were to decide on my experimental approach today, I would have taken a different path. I would either develop a higher resolution array covering the MHC region or use deep sequencing of MeDIP- or bisulphite-treated DNA as it was described recently (Down et al., 2008; Meissner et al., 2008) and is discussed in the general introduction of this thesis (chapter 1).

In conclusion, the development of the MeDIP-MHC tiling array approach for DMR detection was highly innovative and demanding at the time it was established and remains to be a valuable resource for MHC-related studies.

7.3 tDMR screen

7.3.1 tDMRs within the MHC

Following the publication of the HEP pilot study (Rakyan et al., 2004) I performed the tDMR screen aiming to generate more comprehensive tissue-specific methylation data within the MHC. I identified 55 tDMRs, of which 54 were not identified by the HEP study, emphasizing the advantage of my unbiased assay covering the whole MHC region. However, I failed to identify 11 tDMRs reported by the HEP. The reason for this may be the limited resolution (2kb) of my MHC tiling array compared to bisulphite sequencing (1 bp resolution) used by the HEP. Today, this limitation could be overcome by higher resolution microarrays.

7.3.2 Genomic Features of tDMRs

Our understanding of the biological function of DNA methylation in mammals has been growing steadily over the last few years but is still far from complete. Identification of genomic regions with known and as yet unknown features that show differential methylation patterns is expected to give further functional insights into DNA methylation. To this end, a number of large-scale and genome-wide DNA methylation studies have been conducted aiming to identify DMRs in normal and disease-associated samples (Eckhardt et al., 2006; Keshet et al., 2006; Rakyan et al., 2008; Weber et al., 2005; Weber et al., 2007). One of the most striking findings of these studies is that DMRs are not always present in 5'UTRs or in close proximity to TSSs of genes, supporting the notion that DNA methylation has a functional role beyond the mere control of transcription through promoter methylation.

In this context and by using the annotation provided by the Ensembl genome browser, I extracted the genomic features overlapping with the 55 MHC loci characterised as

tDMRs. More specifically, I reported the genomic features mapping within the genomic boundaries of the tDMRs (average size 2kb).

In agreement with what has been reported in other studies, only one of the tDMRs identified within this study overlaps with a TSS and only two with RNA polII binding sites. Based on my analysis, H3 lysine 4 tri-methylation (H3K4me3) is the most highly correlated genomic feature with differential DNA methylation. According to a recent publication, DNA methylation is in strong inverse correlation with H3K4me3 (Meissner et al., 2008) indicating that histone marks may drive the formation of DNA methylation patterns. Generation of maps for histone marks across the MHC using the MHC tiling array may therefore be informative in this context and may give further insights into the interplay between histone marks and DNA methylation. Histone marks and DNA methylation are the two major components defining the epigenome.

7.3.3 Copy number variation and DNA methylation

As part of the analysis conducted for the tDMR screen (chapter 4), I correlated the tDMRs overlapping with MHC transcripts with the corresponding expression data available from the GNF atlas of gene expression. This analysis revealed that tDMRs within the *C4A* and *C4B* loci show inverse correlation with *C4A* and *C4B* expression levels, implicating DNA methylation in the mechanism regulating their expression.

C4A and *C4B* genes are located in the MHC class III region, show more that 99% sequence similarity and are examples of copy number variants (CNVs) in the human genome. In the Caucasian population 55% of the MHC haplotypes have the 2-locus C4A-C4B configuration and 45% have an unequal number of *C4A* and *C4B* genes. This indicates that MHC haplotypes are subjected to duplications/deletions within the region encoding for *C4A* and *C4B* loci (Blanchong et al., 2001).

Gene duplication is commonly regarded as the main evolutionary mechanism towards the gain of a new gene function (Jiang et al., 2007). It has been suggested that epigenetic silencing protects newly born duplications from degradation to pseudogenes (Rodin and Riggs, 2003), leading to functional divergence between duplicated genes. This is further supported by the notion that the frequency of young gene duplicates is higher in organism that have cytosine methylation (*H. sapiens, M. musculus* and *A. thaliana*) than in organisms that do not have methylated genomes (*S. cerevisiae*, *D. melanogaster*, and *C. elegans*) (Lynch and Conery, 2000).

Based on the above and on my data, I have reasoned that duplicated genes with otherwise normal expression levels may be silenced by DNA methylation. This is supported by association studies reporting that gene duplications are not always in positive correlation with gene expression (Stranger et al., 2007). In this context and in collaboration with Vardhman Rakyan, I have already generated methylation data for a number of samples used for the CNV project (Redon et al., 2006). Analysis and correlation of these data with the available CNV, HapMap (SNPs) and expression data for these samples is expected to provide great insights into how DMRs, CNVs and SNPs interact to form complex phenotypes. In addition, this analysis may provide further insights into the evolutionary mechanism that lead to the generation of new genes by duplication.

7.3.4 Future directions

While acknowledging the progress that has been made in DNA methylation profiling technology, the tDMR screen (using the MeDIP-MHC tiling array approach) can be followed up by additional experiments as described below:

i. Recently the term ‘population epigenetics’ was introduced (Richards, 2008) underlining one of the greatest challenges in the field of epigenetics at moment: the determination of

the proportion of natural epigenetic variation in the human population. Understanding the significance of epigenetic polymorphism requires: (i). systematic approaches cataloguing epigenetic variation across the genome, including cytosine methylation and histone tail modifications and (ii). association of epigenetic variability with changes in local gene expression. Analyses of samples from different human populations, different tissues and cell types as well as from different phenotypes are necessary.

In this context, I would analyse additional tissue types and biological samples. This will allow the identification of additional tDMRs and the estimation of inter-individual variability in DNA methylation levels which has been reported for the MHC region (Rakyan et al., 2004) as well as in germ cells (Flanagan et al., 2006) and repetitive elements (Sandovici et al., 2005).

At this point I would like to mention that currently there are a lot of large collaborative projects both in the USA and in Europe that aim to determine and elucidate the significance of epigenetic variation in the human population (Jones and Martienssen, 2005; Qiu, 2006).

ii. Although most of the genes in the MHC class I and III regions are expressed in all somatic cell types, MHC class II gene expression is largely restricted to antigen presenting cells. Cytokines such as IFN-γ can induce expression of classical MHC class II genes and up-regulate genes in the MHC class I and III regions (Boehm et al., 1997; Rohn et al., 1996). It is also known that epigenetic events, including histone marks and non-coding RNAs (Wright and Ting, 2006), can control MHC class II gene expression. These epigenetic events were shown to be induced by IFN-γ (Morris et al., 2002; Pattenden et al., 2002). It would be interesting to investigate further the role of DNA methylation in the selective expression of MHC class II molecules and how DNA methylation patterns change upon treatment with cytokines. To this end, it would be informative to apply the MeDIP-MHC tiling array approach to cell lines either expressing

or not expressing the classical MHC class II genes (*HLA-DP*, *-DQ*, and *-DR*), aiming to identify DMRs associated with MHC class II expression.

7.4 pDMR screen

7.4.1 pDMRs within the MHC

I identified two pDMRs that could be associated with the MHC class I^- phenotype. Of those only one was found to be overlapping with two of the genes involved in the MHC class I antigen and presentation pathway. This pDMR maps to the bidirectional promoter of the *TAP1/PSMB9* genes. Interestingly, this pDMR could not be associated with *PSMB9* down-regulation as it is also present in cell lines expressing this gene. Therefore, this pDMR is likely to be associated with the *TAP1* gene only. In addition, it was found to be associated with the down-regulation of the *HLA-A, HLA-B* and *PSMB8* gene expression levels. A second pDMR (within the *NMR* locus) was also associated with the down-regulation of these four genes: *HLA-A, HLA-B, TAP1* and *PSMB8*.

Although the association is high, proving the functional connection between the two pDMRs and the expression of the four genes is complicated due to our limited knowledge regarding the functional role of DNA methylation. It is possible that hypermethylation blocks a distant control element for *HLA-A, HLA-B, TAP1* and *PSMB8* genes (figure 4.1). This is possible to occur in a genomic region like the MHC where chromatin loops are known to be associated with transcriptional regulation (Ottaviani, 2008). Deletion of the two regions containing the two pDMRs and subsequent expression analysis would be an experimental approach to investigate this possibility. In addition chromatin conformation capture (3C) assay (Dekker et al., 2002) can be employed to test the interaction of distant regions within the MHC. This approach has been used previously to show that DMRs within the imprinted genes Igf2 and H19 interact (Murrell et al., 2004). The regulatory role of these pDMRs could be further

verified by experiments looking for factors binding to these regions. A DNase footprinting assay would detect any DNA-protein interactions within the corresponding regions. Subsequent mass-spectrometric analysis could be use to reveal the identity of these proteins. Finally, additional bisulphite sequencing analysis may also be necessary to identify the exact CpG sites within these pDMRs that undergo differential methylation.

Interestingly, no pDMRs were identified within the coding regions of *HLA-A, -B* and *–C* genes. According to a previous publication, the promoters of these three genes are hypermethylated in human oesophageal squamous cell carcinomas (Nie et al., 2001) that display the HLA class I^- phenotype. The authors of this paper claimed that hypermethylation of the promoter regions of the *HLA-A, HLA-B* and *HLA-C* genes is a major mechanism of transcriptional inactivation. This deviation can be explained by: (i). the fact that DNA hypermethylation of MHC class I genes is a specific characteristic of oesophageal squamous cells (not tested here); (ii). the low MHC tiling array resolution (2kb); and (iii). the high sequence similarity (>80%) between MHC class I genes; it is possible that co-hybridization of highly similar DNA molecules is masking the effect of differential methylation. I have attempted to perform methylation analysis of the promoters of *HLA-A, -B*, and *–C* genes but was not successful in designing bisulphite primers that were locus-specific; MHC class I loci are highly polymorphic.

7.4.2 DMRs within the TNF cluster

I have identified three DMRs within the TNF cluster that can be associated with the expression of *LTA, LTB* and *TNF-α* genes. This agrees with previous data showing that *TNF-α* expression is controlled epigenetically (Sullivan et al., 2007). I showed that in addition to the *TNF-α* promoter, the gene bodies of *TNF-α*, *LTB* and *LTA* were hypermethylated in the majority of the cell lines tested. Hypermethylation of multiple loci within the TNF cluster can happen either simultaneously or it can follow a spreading

model for DNA methylation (Clark and Melki, 2002; Turker, 2002). Based on this model, hypermethylation of the TNF cluster can be a two-step process. Initially, CpG sites within the 5'UTR of *TNF-α* may be hypermethylated by *de novo* methylation (5m-CpG seeds). Subsequently, these 5m-CpG seeds may act as foci for methylation spreading to distal 5' and 3' CpG sites, resulting in the observed hypermethylation of the TNF-α, LTA and LTB gene bodies. Additional functional studies are required to verify this model.

Interestingly, the DMRs within the TNF-α loci are also associated with the HLA class I^- phenotype. TNF-α, together with IFN-γ, is a cytokine known to be an immune modifier acting on the MHC class I processing and presentation pathway by inducing expression of the *PSMB8, PSMB9, TAP1, TAP2* and MHC class I genes. A kB-like element within the promoter of these genes is responsible for the response upon TNF-α stimulation. *TAPBP* and *B2M* are known not to respond to TNF-α (Dovhey et al., 2000; Johnson, 2003; Johnson and Pober, 1994).

It is possible that the up-regulation I observed in MHC class I gene expression levels after 5-aza-CdR treatment is the result of demethylation of the TNF cluster and subsequent up-regulation of *TNF-α*. This speculation is supported by the fact that *B2M* and *TAPBP* do not respond significantly to 5-aza-CdR treatment. However as *HLA-C, PSMB9* and *TAP2* show normal expression levels in some cell lines with reduced *TNF-α* expression, further experiments are required before I can draw a conclusion. Also, there is one cell line (CCRF-CEM) that shows up-regulation of the *PSMB9* gene; interestingly CCRF-CEM displays higher levels of *TNF-α* gene expression compared to the other cell lines tested here.

It has been reported that TNF-α acts in synergy with interferons for the transcriptional activation of the MHC class I heavy and light chain genes (Johnson and Pober, 1994). Hence, it should be expected that in the absence of *TNF-α* (as it is the case for the cell

lines tested here) other cytokines would be sufficient to stimulate MHC class I expression and presentation. One such cytokine is IFN-γ which is the most prominent inducer of MHC class I expression.

It would have been interesting to analyse the expression levels of *IFN-γ* in the cell lines tested here. Recent evidence implicates epigenetics in the regulation of *IFN-γ* expression as well (Schoenborn et al., 2007; Spilianakis and Flavell, 2007) indicating that the two cytokines, TNF-α and IFN-γ, may be down-regulated simultaneously by DNA hypermethylation. Methylation analysis of the *IFN-γ* gene in the cell lines tested here should clarify this matter.

Finally, previous studies using MCF7, T47D and MDA-MB-231 cells (cell lines tested here) have shown that stimulation of MHC class I molecules was induced by IFN-γ or TNF-α (Dejardin et al., 1998) treatment. Hence, it is possible that low expression of the two cytokines (possibly due to promoter hypermethylation) in combination with pDMRs or other epigenetic modifications in the MHC region result in the MHC class I^- phenotype. It may be informative to treat the cancer cell lines tested here with TNF-α and IFN-γ . If my speculation is correct, this treatment should have similar effects as 5-aza-CdR on the expression levels of my candidate genes. Combined treatment with 5-aza-CdR and TNF-α/IFN-γ should result in an additive effect on expression levels of genes involved in the MHC class I pathway.

7.4.3 Transcriptional silencing and DNA hypermethylation

It has been proposed that gene silencing is the critical precursor of DNA methylation, as it may change the dynamic interplay between *de novo* methylation and demethylation of

CpG islands and tilts the balance in favour of DNA hypermethylation (Clark and Melki, 2002; Turker, 2002). This model can be used to explain hypermethylation in the promoter regions of the *PSMB8* and *B2M* genes in the cell lines tested here that show the lowest expression levels for the corresponding genes. However, this is only a speculation made based on the presence of hypermethylated DMRs in cell lines with the lowest expression levels.

It would be interesting to follow up the impact of gene silencing to methylation patterns. This has already been done for a number of genes, including the *GSTP1* and *RASSF1A* (Song et al., 2002; Strunnikova et al., 2005) but further more systematic approaches are required to confirm the ability and the requirements for gene silencing to drive *de novo* methylation. This would give further mechanistic insights in *de novo* methylation that is observed in many diseases including cancer.

7.4.4 Future directions

The findings of the pDMR screen are consistent with the notion that DNA methylation is involved in the development of the MHC class I^- phenotype. In order to further support my findings, the following experiments could be performed:

(i). treatment of additional cell lines with and without the MHC class I^- phenotype with methylation inhibitors.

(ii). I would take advantage of recent developments in microarray technology and perform similar analysis using high-resolution (e.g. 50bp resolution) arrays, as it was discussed above. Using these array-platforms it may be possible to identify additional pDMRs and ease the effort to identify the exact CpG sites that undergo aberrant methylation in samples with the MHC class I^- phenotype.

(iii). study genetic variation (SNPs and CNVs) within the MHC region for the same samples tested under (ii). Meta-analysis of such genetic data with methylation data will

allow the identification of 'hepitypes' linked with MHC phenotypes. Hepitypes were introduced recently and refer to genetic haplotypes which when combined with specific methylation patterns (epitypes) may contribute to the development of a phenotype (Murrell et al., 2005). Sequence-dependent allele specific methylation patters (hepitypes) were recently identified in normal individuals (Kerkel et al., 2008) as well as in individuals with chronic lymphocytic leukaemia (CLL) (Raval et al., 2007). This analysis can also be implemented by expression analysis. Such meta-analysis can be expected to have great medical relevance for the diagnosis and treatment of MHC-linked diseases.

(iv) While the system and analysis described above is suitable to identify pDMRs and study their underlying mechanisms in cell lines, primary tissue samples will need to be analysed to confirm the involvement of such pDMRs in clinical samples displaying the same or similar phenotype.

7.5 Long Range Epigenetic Silencing

A recent study suggested that epigenetic changes in cancer are not always local but can be global encompassing large-scale chromosomal regions, resulting in concordant repression of large regions of DNA (Long Range Epigenetic Silencing – LRES) (Frigola et al., 2006). LRES was observant in a 4Mb band on chromosome 2q14.2. In a similar manner, LRES could be involved in the concordant silencing of multiple MHC (a 4Mb region on chromosome 6) loci. More comprehensive DNA methylation analysis, in combination with histone marks and expression profiling would be informative with respect to LRES within the MHC region.

7.6 Recombination hotspots and epigenetic events

Although not experimentally tested within this thesis, it has been shown that epigenetic events can be implicated in controlling events of recombination hotspots during meiosis. Meiotic recombination between highly similar duplicated sequences (non-allelic

homologous recombination, NAHR) generates deletions, duplications, inversions and translocations that frequently result in genomic disorders (Turner et al., 2008). It has been shown that in males, the presence of meiotic recombination hotspots is not influenced by genomic sequence but rather by distal regulatory elements or epigenetic events (Neumann and Jeffreys, 2006). The latter may control accessibility of these hotspots. The MHC class II region represents a prominent region where such hotspots have been detected (Kauppi et al., 2005).

Studying how epigenetic events within the MHC influence this phenomenon will be the basis for future studies regarding genomic disorders that are the result of genomic rearrangements as well as for studies aiming to elucidate the evolution of the MHC region.

7.7 Conclusion

This thesis describes the most comprehensive DNA methylation analysis of the human MHC region to date. I developed and used an unbiased array-based assay for the detection of differentially methylated regions (DMRs) that can be associated with particular tissues (tDMRs) and particular phenotypes (pDMRs). The study presented here, underlines the important role of epigenetic variation in phenotypic plasticity.

Current advances in epigenome mapping technologies and the various epigenome projects that have been established recently (Jones and Martienssen, 2005; Qiu, 2006) are expected to give critical insights into the interplay between the genotype, the epigenotype and the environment and serve as catalyst for future studies on human complex diseases.