Genetic determinants of inter-individual variations in DNA methylation

Dissertation zur Erlangung des Doktorgrades
der Naturwissenschaften (Dr. rer. nat.) der Fakultät für Biologie und
Vorklinische Medizin der Universität Regensburg

vorgelegt von
Julia Wimmer (geb. Wegner)

aus
Neubrandenburg

im Jahr
2014

The present work was carried out at the Clinic and Polyclininc of Internal Medicine III at the University Hospital Regensburg from March 2010 to August 2014.

Die vorliegende Arbeit entstand im Zeitraum von März 2010 bis August 2014 an der Klinik und Poliklinik für Innere Medizin III des Universitätsklinikums Regensburg.

Das Promotionsgesuch wurde eingereicht am:
01. September 2014

Die Arbeit wurde angeleitet von:
Prof. Dr. Michael Rehli

Prüfungsausschuss:

Vorsitzender:	Prof. Dr. Herbert Tschochner
Erstgutachter:	Prof. Dr. Michael Rehli
Zweitgutachter:	Prof. Dr. Axel Imhof
Drittprüfer:	Prof. Dr. Gernot Längst
Ersatzprüfer:	PD Dr. Joachim Griesenbeck

Unterschrift:

_____

Who seeks shall find.

(*Sophocles*)

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 The concept of epigenetics

The one thing all known organisms share is the matter of genetic information. However, pure genetic material is not enough to bring an organism to life. The genetic code needs to be interpreted and that, in case of multicellular organisms, in various ways. The initial step of interpretation is transcription of distinct parts of the present genetic makeup, which is thought to be initiated and controlled by three mechanisms. Firstly, intrinsic promoters and the transcription machinery itself seem to partly control the process. Secondly, specific transcription factors influence and target the transcription machinery[1,2]. Finally, the genetic material, deoxyribonucleic acid (DNA), is not found naked in the nucleus, but associated with proteins in a structure called chromatin. Post-replicational modifications of the DNA as well as of other chromatin components, especially histones, affect the accessibility of genetic material.

The later is nowadays addressed by the term “epigenetics”, which was coined initially by Conrad Waddington to describe the “causal interactions between genes and their products, which bring the phenotype into being”[3]. The human organism, for example, is composed of over 400 different, more or less stable cell types, which descended from one single cell, the zygote. Despite sharing an identical genotype (with a few exceptions), each emerging cell type is characterized by a specific gene expression profile and function. The different interpretations of the same genetic information are facilitated by the epigenome, which allows for stable and, ideally, inheritable changes of gene expression without altering the actual DNA sequence. Along with non-coding RNAs[4], DNA methylation and histone modifications encompass the major epigenetic mechanisms[5-7] and will be addressed in this work.

The functionality of epigenetic mechanisms as well as their own regulation is of high interest. Many common diseases such as cancer and degenerative diseases exhibit an altered epigenome[8,9]. Still, even in the healthy organism the basic mechanisms regulating epigenetic modifications are not fully understood and key regulators are rarely known. Against this background, the present work particularly focuses on identification of genetic factors, which cause variations in the epigenome among individuals, especially at the level of DNA methylation.

1.1.1 DNA methylation

The genetic code is written by the four bases adenine, guanine, thymine, and cytosine, but modified versions of the latter also contribute to the informational content. One of them is 5'-methylcytosine (5mC), which results from covalent modification of the carbon 5 position of cytosines mainly in the context of cytosine followed by guanine (CpG dinucleotides, CpGs), a reaction catalyzed by DNA methyltransferases (DNMTs)[10,11]. DNA methylation is in general associated with gene silencing[12-14] and is implicated in a multitude of biological processes such as mammalian X-inactivation[15-17], genomic imprinting[18-20], genomic stability[21,22], silencing of harmful DNA sequences[23-25], and embryonic development[26,27]. Findings linking aberrant DNA methylation patterns to abnormal developmental processes and diseases underpin its importance[28,29]. For example, acute myeloid leukemia shows well-documented evidence of abnormal global DNA methylation phenotypes[30,31]. In normal cells, CpG dinucleotides are generally methylated on a global scale. They exhibit an unequal distribution throughout the mammalian genome, but appear less frequently than statistically expected. The phenomenon could be explained by evolutionary depletion of CpGs due to spontaneous deamination of 5mC. The resulting product thymine, which is a naturally occurring genomic base, is not detected as misplaced leading to an accumulation of Cytosine to Thymine conversions over time[32]. Exceptions of the rule constitute CpG islands (CGIs), where CpG dinucleotides cluster with very high density. CGIs were initially defined as sequence stretches of at least 200 bp with a GC content above 50% and a ratio of observed versus expected CpG frequency of 0.6[33]. CpGs of CGIs are, in contrast to those in CpG-poor regions, usually unmethylated[34]. CGIs are frequently associated with promoters of ubiquitously expressed genes (*housekeeping genes*) and genes that are important for embryonic development. Promoters of highly tissue-specific genes exhibit lower CpG frequencies[35,36].

1.1.1.1 Setting and erasing methylation at CpG dinucleotides

So far, three families of DNA methyltransferases (DNMTs) have been identified in eukaryotes, namely DNMT1, DNMT2, and DNMT3[37-41]. They share ten characteristic sequence motifs, most of them being highly conserved (Figure 1-1). Among the conserved motifs is the catalytically active prolylcysteinyl dipeptide, which is required for methylation[42,43]. DNMT1, DNMT3A, and DNMT3B are capable of methylating cytosines at the carbon 5 position in the presence of the methyl group donor S-adenosyl methionine (SAM). The proposed mechanism[44] is depicted in Figure 1-2A.

DNMT3a and DNMT3b are abundantly expressed during germ cell development and early embryogenesis but can hardly be detected in somatic cells, with the exception of adult stem cells[45]. Both enzymes function as *de novo* methyltransferases and are essential for correct embryonic development[26,28,46]. DNMT1, also referred to as maintenance methyltransferase, is ubiquitously expressed and has a 5-10 fold higher preference for hemi-methylated DNA[10,47]. It is recruited inter alia via nuclear protein 95 (*NP95* or *Uhrf1*) to replication foci[48] and copies DNA methylation marks from the parental strand onto the newly synthesized daughter strand. The

clonal propagation of DNA methylation patterns facilitates inheritance during cell replication and thus maintenance of lineage commitment[49].

The remaining, structurally related members of the DNMT families do not methylate DNA. DNMT2 seems to be capable of methylating RNA molecules[50], whereas DNMT3L lacks the required amino acids for transmethylation completely and plays a role in stimulating *de novo* methylation, inter alia at imprinted loci[51-54].

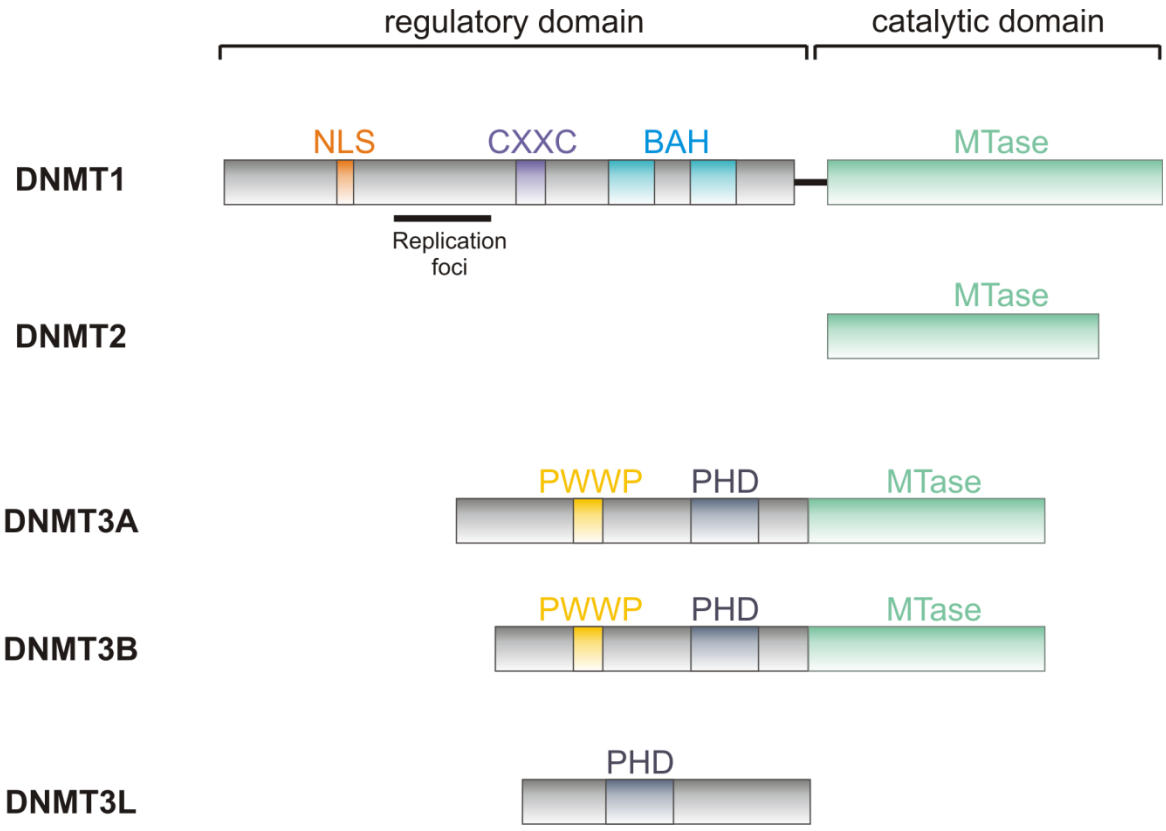


Figure 1-1 The DNA methyltransferase protein family
With the exception of DNMT3L, all known DNA methyltransferases (Dnmts) share the catalytic methyltransferase (MTase) domain, but only DNMT1 and DNMT3A/B accept DNA as substrate for methylation. The regulatory domains (cysteine-rich zinc finger (CXXC) domain, tandem bromo-adjacent homology (BAH) domains, replication foci targeting sequence) of DNMT1 seem to be involved in targeting the enzyme, but this function is still discussed controversially for CXXC and BAH. DNMT3 isoforms all contain a conserved proline-tryptophan-tryptophan-proline (PWWP) motif and/or a pleckstrin-homology domain (PHD) that are involved in chromatin targeting. Nuclear localization signal (NLS). (Adopted from 'Atlas of Genetics and Cytogenetics in Oncology and Haematology'[55])

Although DNA methylation is comparatively stable and maintained during cell proliferation, especially local tissue-specific methylation patterns are subject to dynamic changes[56-59]. Due to lacking evidence for a direct demethylation activity, passive loss of DNA methylation during replication (in the absence of maintenance activity) was initially assumed as the general mechanism for DNA demethylation[60,61]. However, cellular events like the genome-wide erasure of paternal methylation patterns during early embryogenesis could only be explained by active

removal of the methyl group[62,63]. In the past, evidence for a major active DNA demethylation process accumulated[27,64,65]. In 2009, Tahiliani *et al*. finally identified TET1, a member of the ten-eleven translocation (TET) family, to oxidize 5mC to 5'-hydroxymethyl cytosine (5hmC), which seems to be in most cases the initiation step in active DNA demethylation[66]. Since then, several mechanisms were proposed for the complete removal of the initial methyl mark including the base excision repair pathway[67,68] (Figure 1-2B). The Tet family comprises three members, Tet1, Tet2, and Tet3, which all share the catalytic domain required for oxidation of 5mC. Tet1 and Tet3 also share a CXXC domain, which function is largely unknown. The CXXC domain of Tet2 is encoded in a separate gene (*Cxxc4*)[69]. Although it was suggested that the CXXC domain might guide binding towards CpG-rich transcription start sites[70-75], the mechanisms targeting Tet enzymes are still under discussion[76]. All three members seem to be involved in the reprogramming events during early embryonic development. Tet1 and Tet2 were found highly expressed in primordial germ cells[77,78], whereas Tet3 is mainly expressed in the oocyte[79]. High levels of Tet1 and Tet2 were also detected in ES cells, the inner cell mass of blastocysts as well as the early mouse epiblast, but depletion of either Tet1 or Tet2, respectively, did not affect normal development[80-82]. Thus, Tet1 and Tet2 are assumed to have a compensatory role. In contrast, homozygous Tet3 mutations led to neonatal lethality and support its importance for normal development[79]. Tet proteins also seem to play a role in somatic tissues; Tet2, for example, is essential for the development of the hematopoietic lineage[83,84].

Finally, 5hmC does not only seem to be an intermediate of the active DNA demethylation process, but is assumed to also have a regulatory function by recruiting its own set of readers[85]. The consequences of the regulatory function are largely unknown as yet.

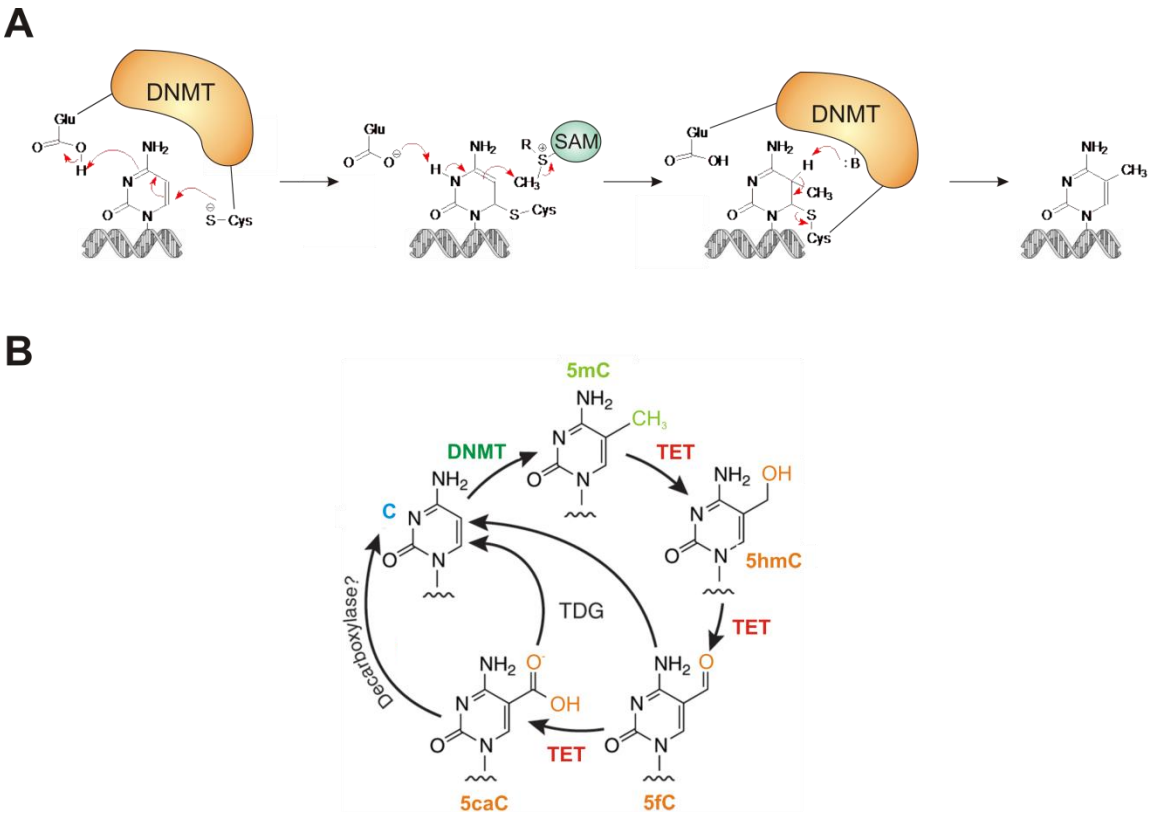


Figure 1-2 Cycle of DNA methylation and demethylation
(A) Proposed mechanism for CpG methylation: the target cytosine is everted from the DNA helix and inserted into the active site of the enzyme. Subsequently, a cysteine thiolate (prolylcysteinyl active site dipeptide) adds covalently to the 6th position and results in a reactive 4,5-enamine that attacks the methyl-group of SAM (S-adenosyl methionine, methyl-donor). Finally, β-elimination releases the enzyme.[44]
(B) Proposed pathways of active demethylation: the methyl group at cytosines (5mC) can be removed by repeated oxidation through TET oxygenase and include the intermediates 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxycytosine (5caC). Involvement of TDG, a member of the base excision repair pathway, and yet unknown decarboxylase were also proposed to return from oxidized intermediates to unmethylated cytosine (C). (Adopted from Huang *et al*.[86])

Although the factors directly involved in setting, maintaining, and erasing DNA methylation, namely the DNMTs and TET proteins, are identified and characterized to a certain extent, very little is known about how *de novo* methylation is targeted to specific sites in the genome. Three possible mechanisms were proposed. Firstly, target loci might be recognized by DNMT3 enzymes themselves by specific DNA or chromatin binding domains. The PWWP domain of DNMT3b, for example, shares structural similarities with the tudor domain of p53-binding protein 1 (53BP1). 53BP1 that is implicated in DNA repair is directed by its tudor domain to methylated H3K79 residues[87,88]. However, the PWWP domain does not show any sequence preference[89] and thus requires additional factors to direct sequence-specific binding. Secondly, the RNA-mediated interference (RNAi) system might facilitate *de novo* methylation of specific DNA sequences. In plants, RNAi-mediated transcription silencing often results in *de novo* methylation of the silenced gene[90]. Similar mechanisms were reported for mammalian cell-culture systems[91,92]. Thirdly, DNMT3a and DNMT3b might be recruited through protein-protein interactions. Interaction of DNMTs with known transcription factors could be demonstrated[93-95]. Lately, evidences for novel

epigenetic modifiers that target DNA methylation have accumulated. Ratnam *et al*. characterized a novel epigenetic modifier, which recognizes its target sequence, a transgene promoter, by a zinc finger motif and recruits chromatin modifier to promote silencing[96]. Another group applied a mutagenesis screening to identify *trans*-acting epigenetic modifiers that affect methylation of a GFP transgene as well as the agouti viable yellow (A^{vy}) allele[97,98]. Transgenes integrate in the genome generally in a multi-copy array. Methylation of the A^{vy} allele depends on silencing of an integrated retrotransposons. Based on these findings, it was assumed that the epigenetic modifiers influence silencing of repetitive sequences. Furthermore, an emerging role of long non-coding RNAs in guiding the silencing machinery to distal genomic regions was described[99,100]. Identifying additional *trans*-acting epigenetic modifier will help to increase the understanding of their regulatory mechanisms and thus the regulatory network that forms the basis of epigenetics.

1.1.1.2 Methyl-CpG binding proteins, the readers of DNA methylation

Silencing through DNA methylation is assumed to be mediated mainly by two major mechanisms. Firstly, the methyl group can mask the genome by simply preventing binding of transcription factors (TFs) required for expression[101-103]. The second mechanism requires "readers" of the methyl mark to transmit the epigenetic information to downstream regulatory proteins and integrate DNA methylation in the epigenetic network[104,105].

One group of proteins that bind to methylated CpGs (methyl-CpG binding proteins, MBPs) constitutes the methyl-binding domain (MBD) protein family[41] (Figure 1-3). The shared MBD domain was initially characterized in the chromosomal Methyl-CpG binding protein 2 (MeCP2). Via its MBD domain, MeCP2 binds to DNA strands with one or more methylated CpGs by recognizing their hydration shell[106,107]. The conserved sequence of the MBD domain was used to identify the other four members of the MBD family. MeCP1, which was actually the 1st identified methyl-CpG binding activity, is a complex containing MBD2[108,109]. MBD2 promotes binding to unrelated DNA sequences with methylated CpGs (mCpGs) in a salt concentration-dependent manner[110,111]. Together with MBD1, which in contrast seems to exhibit sequence preferences[112], these three members of the MBD family have a long-accepted implication in methylation-dependent repression of transcription[113].

The other two members of the MBD family were just recently associated with mCpG-mediated repression. MBD3, although bearing a MBD domain, shows no mCpG binding activity, but binds to unmethylated CpGs instead[114]. Nevertheless, MBD3 is associated with the Nucleosome Remodeling Deacetylase (NuRD) co-repressor complex, which seems to be targeted inter alia by biochemical interaction of MBD3 with MBD2. The importance of MBD3 is supported by the lethality of post-implantational embryos in *Mbd3*$^{-/-}$ mice[115]. The 5th member of the MBD family, MBD4, was initially thought to be involved DNA mismatch repair due to its carboxy-terminal DNA glycosylase domain and interaction with mutL homolog 1 (MHL1)[116]. However, the binding of MBD4 to a limited number of oncogene promoters suggests a combinatory mechanism of the MBD and the mismatch repair domains. The hypermethylated and thus repressed state of these cancer-promoting genes seems to be maintained by reversing 5mC to T conversions[117]. All

members of the MBD family, except for MBD3, seem to bind mCpGs depending on the methylation density. This is in line with their redundant function[114,118]. However, the methylation-independent binding patterns are specific and differ between the different MBD proteins[114].

Kaiso and the SET-Ring finger associated (SRA) domain-containing family comprise an exception among MBPs since they completely lack the MBD domain. On the one hand, Kaiso achieves recognition of methylated DNA by a carboxy-terminal Cys_2His_2 zinc finger domain[119]. In addition, it also recognizes a non-methylated consensus sequence making it a bifunctional chromosomal protein[120]. On the other hand, the SRA domain recognizes hemi-methylated DNA[121]. A prominent member of the SRA domain-containing protein family is *NP95*/*Uhrf1* that plays an important role of maintaining methylation patterns by recruiting Dnmt1 to replication foci[48,122] (Figure 1-3).

The binding sites of each of the MBPs seem to be specific and mutually exclusive, but the targeting mechanisms are not completely understood as yet[123-125].

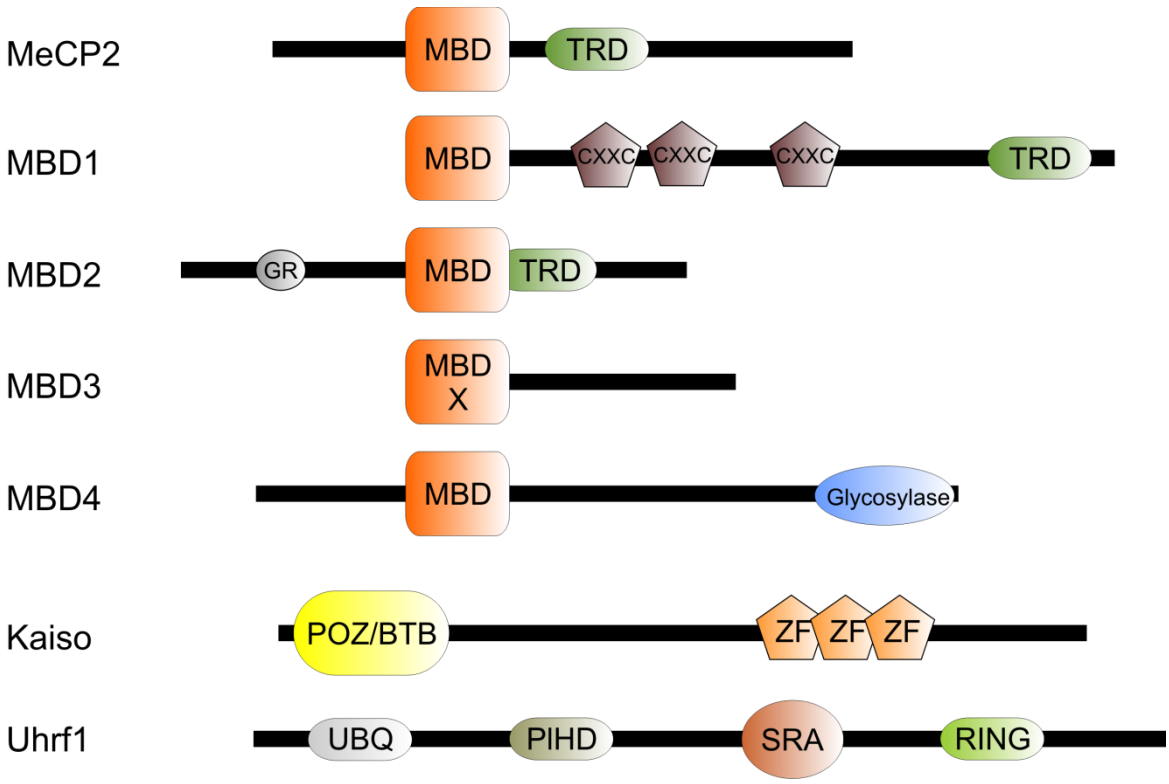


Figure 1-3 Methyl-CpG binding proteins

The family of MBPs comprises all members of the methyl-binding domain (MBD) protein family and Kaiso. MeCP2 is the founding member of the MBD family and harbors a conserved MBD domain to bind methylated DNA sequences as well as a transcriptional repressor domain (TRD). In addition to these domains, MBD1 also contains 3 cysteine rich zinc fingers (CXXC), which seems to provide an interaction platform with other proteins[126]. In MBD2, the MBD domain overlaps with the TRD domain and the protein also harbors a glycine-argenine (GR) repeat. The MBD domain of MBD3 is mutated and not able to bind to methylated DNA. MBD4 contains a MBD domain to bind to methylated DNA, but lacks a TRD domain; instead it contains a glycosylase activity which is important for its function in DNA repair. Kaiso depends on zinc finger (ZF) domains to recognize methylated DNA and promotes transcriptional repression by its Broad-Complex, Tramtrack and Bric a brac (POZ/BTB) domain. Uhrf1 is a member of the SET-Ring finger associated (SRA) domain family and binds to hemi-methylated DNA. It also contains a ubiquitin-like (UBQ) domain, a plant homeodomain (PHD) as well as a Ring finger (RING) domain. (Adapted from Klose *et al.*[127]).

Except for MBD4, all MBD family members interact with or are part of complexes containing histone deacetylases[108,115,124,128-131]. Histone deacetylases belong to histone-modifying activities that promote gene silencing by higher compaction of the chromatin and will be discussed together with other modifying enzymes in the following chapter.

There are additional ways how DNA methylation could mediate its repressiveness. It was shown that transcription factors are not always repelled by the methyl mark, but can also recognize and specifically bind to methylated CpGs[85,132]. DNA methylation was also found to influence histone 1 (H1) binding, which lashes the wrapping of DNA in nucleosomes[133] and seems to diminish elongation capacity of the RNA Polymerase II transcription machinery within the gene body[134]. Still, the underlying mechanisms are poorly understood and under discussion.

1.1.2 The histone code

In vivo, the DNA molecule is not found naked in the nucleus, but associated with proteins, a structure known as chromatin. Although the methyl group itself can directly prevent binding of transcription factors by masking their binding site[101], transcriptional silencing by DNA methylation seems to be more efficient in a chromatin context[135,136]. In line with these findings, DNMTs associate with chromatin co-repressor complexes containing histone methyltransferases and deacetylases[137-139], thereby contributing directly to transcriptional repression. The indirect silencing effect of DNMTs via CpG methylation is mediated by MBPs, which are also associated with histone-modifying complexes[108,115,124,128,130].

The most basic feature of DNA packaging is the nucleosome. Nucleosomes are composed of 146 Watson-Crick base pairs of DNA wrapped around a cylindrical histone complex. They form a structure similar to beads on a string separated by DNA stretches termed linker DNA. Histones constitute the core component of the nucleosome by forming an octamer composed of two molecules of each of the histones H2a, H2b, H3, and H4. A 5th member of the histone family, linker histone H1, completes and tightens the nucleosome assembly by localizing to the site of the core particle where DNA enters and exits[140-143]. Long non-structured amino(NH_2)-terminal tails protrude from the main globular domains, which reside within the nucleosome. Post-transcriptional modification of histones appears mainly at the NH_2-terminal tails with a variety of different chemical groups. The modifications regulate the transcriptional state of the associated DNA sequence[144,145]; a mechanism referred to as the "histone code" [146]. The accessibility and flexibility of the tails allow for interaction with other nucleosomes and the connecting linker DNA as well as regulatory factors[147,148]. In this context it was found that the regulatory outcome of individual modifications depends on adjacent modifications and that some of them are interdependent. Thus, a specific histone mark can have either repressive or activating consequences, depending on the modification environment it is found in.

"Writers" of the histone code add and remove at least eight distinct chemical residues including methyl and acetyl groups on over 60 distinct histone positions[149] (Figure 1-4). Site-specific

targeting of histone-modifying enzymes is achieved in several ways. The transcription as well as the replication machinery seems to recruit histone modifiers[150-152]. Furthermore, methylated DNA attracts histone-modifying enzymes via MBPs[108,115,124,128,130]. Recently, is was also proposed that non-coding RNAs direct histone-modifying activities[100,150,153]. This was shown for the Xist RNA in X chromosome inactivation[154].

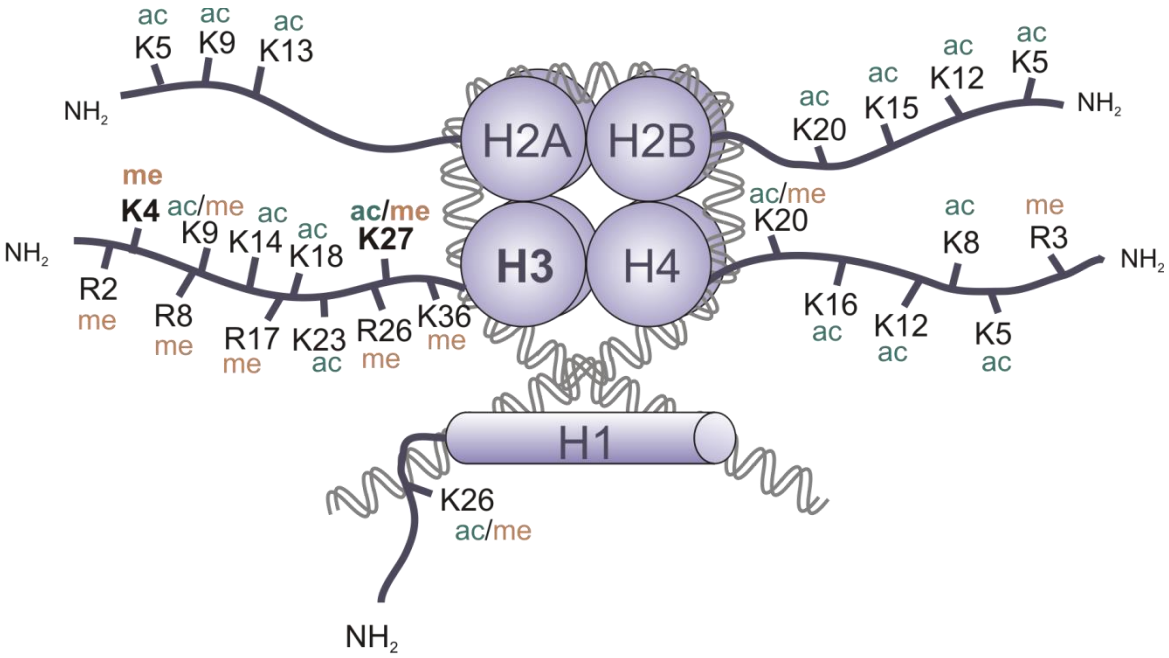


Figure 1-4 Post-translational modifications of histone tails
The NH_2-terminals of the four core histones (H2A/B, H3, and H4), which are part of the nucleosome, as well as of the linker histone H1 are subject to numerous post-translational modifications. In addition to acetylation and methylation of the indicated residues, many other amino acids are subject to post-translational modifications, which include phosphorylation, sumoylation, and ubiquitinylation. The well-studied residues (lysine 4 and 27 of histone H3) are highlighted by bold letters. (Adopted from "Pluripotent stem cells"[155])

Two well-studied histone modifications are lysine acetylation and methylation. Depending on the modified residue, these modifications represent key switches of repression as well as transcription. Due to their essential regulatory function, the following sections will focus on these two modifications.

1.1.2.1 Histone acetylation

The four core histones can be acetylated by histone acetyltransferases (HATs) at specific lysines[156-158]. This results in neutralizing their positive charge. A more open chromatin structure is the consequence of the altered electrostatic histone-DNA interaction by hyperacetylation[159,160]. Furthermore, acetylated residues can provide binding sites for bromodomain-containing factors involved in gene activation[158,161] and prime repressed genes for subsequent activation[162]. In this context, H3K27ac is of much interest since it marks active enhancers but also promoters and gene bodies[163,164].

Histone deacetylases (HDACs) can reverse active chromatin structure by removing the acetyl mark from lysine residues. In addition to inducing a closed chromatin structure, HDACs interact with other repressor proteins. They are recruited to sites of methylated CpGs by MBPs as well as sites marked by repressive histone marks and associate with transcriptional repressor complexes.

1.1.2.2 Histone methylation

Another well-studied histone modification is methylation, which can occur at specific lysines and argenines of histone H3 and H4 in the mono-, di-, or in case of lysines, trimethylated form[149]. Histone methyltransferases (HMTs) are generally more specific about their target residue than HATs and almost all HMTs contain a SET domain. Histone methylation can either be active or repressive, depending on the modified residue and neighboring modifications. H3K4 methylation, which is the prime example for actively marked regions and is associated with the presence of RNA Polymerase III, is catalyzed inter alia by trithorax-group (TrxG) proteins and their mammalian homologues[165]. H3K27 is methylated by polycomb-group (PcG) proteins and associated with repression[166,167]. Both protein families were initially identified as key regulators of developmental processes in *Drosophila*[168].

Promoters of actively transcribed genes generally show high levels of H3K4 trimethylation (H3K4me3)[167,169]. However, genes with bivalently marked promoters (H3K4me3/H3K27me3) appear inactive and could be poised for lineage-specific activation or repression[170,171]. H3K4me1 is also associated with activation but mainly found in intragenic regions of actively transcribed genes and at enhancers[172,173]. In combination with H3K27ac or H3K9ac, H3K4me1 marks active enhancers, whereas inactive and poised enhancers are marked by H3K4me1 and H3K27me3[163,164]. These findings suggest that PcG activity generally dominates over TrxG activity. Furthermore, there seem to be two distinct modes of regulation of promoter activity depending on the CpG content. In ES cells, CpG-rich promoters are targeted by mammalian homologues of TrxG proteins (mixed-lineage Leukemia, MLL) and drive transcription unless actively repressed by PcG activity[174]. Another factor that is involved in shaping the unique CpG islands chromatin architecture in mammals is the CXXC zinc finger protein 1 (CFP1). It recognizes non-methylated CpG-rich regions by its CXXC zinc finger domain and recruits the Set1 complex. The Set1 complex contains a histone methyltransferase resulting in increased methylation of H3K4 at CpG islands[175-177]. A subset of CpG-rich promoters seems to lose the recruiting capacity for TrxG proteins upon lineage commitment. Other epigenetic modifications such as DNA methylation

might account for this phenomenon. CpG-poor promoters behave in the exact opposite way. They seem to be inactive by default and do not need further repression by PcG proteins. Selective activation of the promoters could be propagated by cell type- or tissue-specific factors[167,178].
Other methylated histone residues such as H3K9me3 are associated to silencing of centromeres, transposons and tandem repeats[167,179].
In addition to the direct effect on chromatin structure (histone-histone and histone-DNA interaction, respectively), effector proteins "read" histone modifications and mediate their regulatory information[180].

1.1.2.3 Translation of the histone code

"Readers" of the histone code bind to specific modifications on distinct histone residues by chromatin-binding domains and transmit the detected epigenetic information. The readers often provide the effector function themselves and induce further modifications, ATP-dependent remodeling of the chromatin fiber, transcriptional elongation, and DNA repair mechanisms[149,181,182]. Two well-characterized binding domains are the bromodomain and the chromodomain, which bind to acetylated and methylated histone residues, respectively[149].
A well-studied example for gene silencing is the hierarchical recruitment of PcG activity. Polycomb repressor complex 2 (PRC2) binds to and promotes H3K27me3 by its enhancer of zeste homolog 2 (Ezh2) subunit (SET domain-containing methyltransferase), a histone mark associated with transcriptional repression. H3K27me3 is recognized by the chromodomain of Polycomb (Pc), a subunit of PRC1, resulting in a resistance to remodeling complexes and thus stabilization of the repressive chromatin structure[183,184]. A recently published work by Cao *et al*. also demonstrated that interaction of PRC2 and PRC1 is not only achieved by trimethylation of H3K27, but also by protein-protein interaction promoted by enhancer of embryonic development (Eed)[185]. Furthermore, Ezh2 was shown to interact with DNMTs, allowing for *de novo* methylation to finally lock the gene in a repressed state[186] (Figure 1-5A).
The TrxG activity is an example for gene activation. TrxG protein complexes and their mammalian homologues can recognize and promote H3K4 methylation residues by their SET domain. Furthermore, the associated complexes also contain HATs and remodeling factors, enzymatic activities that lead to a more accessible chromatin structure and thus facilitate transcription[187-189] (Figure 1-5B).

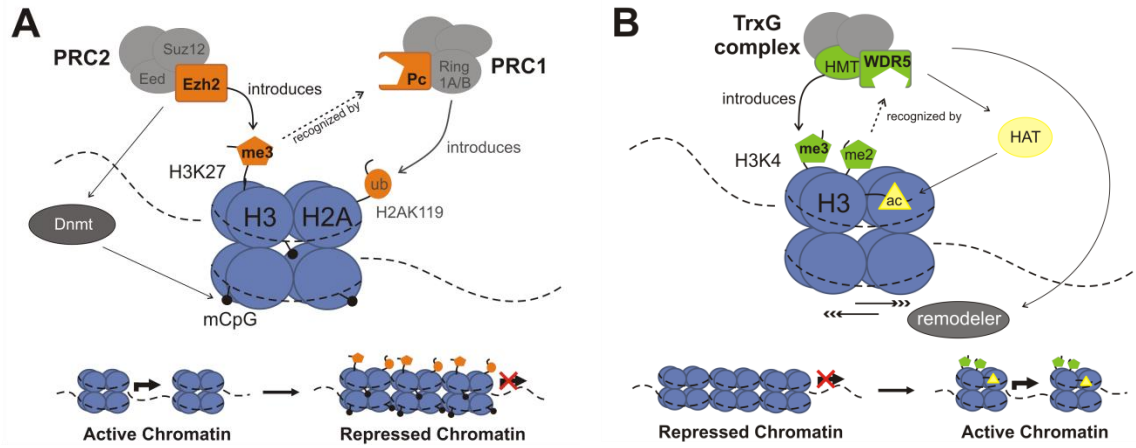


Figure 1-5 ***Repressive and active chromatin structures exemplified by (A) Polycomb- and (B) Trithorax-mediated regulation***

(A) Enhancer of zeste 2 (Ezh2) is a histone methyltransferase (HMT) part of the Polycomb repressive complex 2 (PRC2) and promotes H3K27 trimethylation (me3). H3K27me3 recruits PRC1 by Polycomb (Pc) and ubiquitinylates (ub) H2AK119. These histone modifications lead to increased compaction of the chromatin and can also attract Dnmts to lock the repressive state by methylated CpGs. **(B)** The Trithorax group (TrxG) complex recognizes H3K4 dimethylation (me2) by the Set domain of WD repeat domain 5 (WDR5) and promotes further methylation of H3K4 by HMTs that are part of the complex. The TrxG complex also interacts with histone acetylases (HATs) and chromatin remodelers to facilitate a more accessible chromatin structure. (Adopted from Veazey *et al.*[190])

1.1.3 Non-coding RNAs

In addition to DNA methylation and post-translational modification of histones, a third mechanism associated with epigenetic regulation of the genetic make-up is provided by non-coding RNAs (ncRNAs)[4]. In contrast to protein-coding RNAs, this class of transcripts is not translated into proteins and can be retained in the nucleus. They can interact with other RNAs, but also with transcription factors, chromosomal proteins and the DNA itself. In addition to transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), which are involved in the translational process required for protein synthesis[191], many other functionally important classes of ncRNAs have been identified. The regulatory potential of RNA molecules was first observed in plants[192-194] and later described as RNA interference (RNAi) in *C. elegans*[195]. The silencing effect of the involved small RNA molecules (micro RNAs (miRNAs) and small interfering RNAs (siRNAs)) is controlled by the RNA-induced silencing complex (RISC) and complementary interaction with mRNA[196,197]. In addition, small RNAs were also found to have an implication in gene activation when they are complementary to the promoter of target gene, but the mechanism is understood less well[198,199]. Furthermore, the class of small ncRNAs comprise other RNA molecules such as small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), exosomal RNAs (exRNAs), and Piwi-interacting RNAs (piRNAs).

In addition to the armada of small ncRNAs, long non-coding RNAs (lncRNAs) are larger than 200 nt[200] and seem to affect gene expression at the pre-transcriptional level[201]. The best-studied lncRNA is involved in mammalian X chromosome inactivation and encoded by the *Xist* gene[202,203]. In females, the XIST RNA is transcribed from the future inactivated X chromosome and acts in *cis*

on the chromosome of origin. XIST-dependent recruitment of PRC2 and PRC1 leads to the chromosome-wide accumulation of repressive histone modifications such as H3K27me3 and fortified also by DNA methylation. Another example for a long non-coding, but *trans*-acting RNA is HOTAIR, which is required for silencing of the HOXD locus[204]. HOTAIR also seems to involve the recruitment of the Polycomb silencing machinery. Further examples for regulatory lncRNA, which seemed to mainly act in *cis* or at least in the context of the chromosome of origin, exist and also include RNAs that promote gene activation[99,201,205-207]. Since most of the eukaryotic genome is transcribed, the regulatory mechanisms involving ncRNAs receive increasing attention.

The discussed epigenetic mechanisms are made responsible for one-way commitment during development. But in addition to the developmental program, environmental impacts manifest also in the epigenome. To restart the circle of life, all acquired memory from tissue-specificity as well as from external influences need to be reversed. This reprogramming takes places during early embryonic development and is the subject of the following chapter.

1.2 Epigenetic reprogramming during embryogenesis

Somatic methylation patterns are comparatively stable and facilitate lineage restriction during development. In contrast, germ cells and preimplantated embryos show genome-wide epigenetic reprogramming as depicted in Figure 1-6. The reprogramming events are essential to reset imprinting patterns and restore totipotency. They are also assumed to prevent transgenerational inheritance of epimutations[208].

For premordial germ cells (PGCs), the direct progenitors of sperm and oocytes, demethylation seems to restore their developmental potential[209,210]. PGCs arise around embryonic development day 7.5 (E7.5) from the epiblast[211,212]. Despite still harbouring a high level of pluripotency, the epiblast is already committed to somatic lineages and globally methylated[213]. Global DNA demethylation in PGCs appears to be active and involves oxidation as well as deamination of mCpGs and subsequent removal by base excision repair (BER)[212,214-216]. Since the actual window of DNA demethylation in PGCs is not fully determined as yet, a combination of active conversion of mCpG and passive loss of the intermediates by cell division is possible. Intracisternal A-particle (IAP) retrotransposons constitute an exception to the global wave of DNA demethylation and have the potential to facilitate transgenerational epigenetic inheritance[217,218]. PGC reprogramming goes beyond DNA demethylation and affects the general chromatin structure by exchange of histone variants and increasing levels of H3K27me3[213,219]. Regarding imprinted regions, the parental DNA methylation pattern is removed and re-established depending on the sex of the organism[19]. Mature gametes once again exhibit higher levels of DNA methylation achieved by *de novo* DNA methyltransferases DNMT3A and DNMT3B. Each of the enzymes is recruited by their non-

catalytic ortholog DNMT3L to distinct loci marked by histones lacking H3K4 methylation[220-222]. However, the underlying mechanisms are still poorly understood.

A second phase of global reprogramming is found in the zygote immediately after fertilization and is thought to erase epigenetic signatures in gametes (except at imprinted loci) as well as to regain nuclear totipotency essential for normal development[208]. Epigenetic reprogramming of the parental genomes takes place while both genomes are still separated in the two pronuclei. It is characterized by asymmetric, independent DNA demethylation of the two parental haploid genomes. In sperm, the paternal genome, tightly packed in protamines (instead of histones) and with a DNA methylation landscape similar to somatic cells[223,224], is maximally condensed and inactive to the greatest possible extend. Demethylation occurs rapidly and actively on a global scale shortly after fertilization[62,63] and includes pluripotency genes such as *Oct4* and *Nanog*[225]. Several mechanisms seem to contribute to the active removal of mCpGs at specific loci and form a complex demethylation network. 5mC can be erased by iterative TET3-catalysed oxidation via 5hmC[79,226]. In addition, the elongator complex was shown to be required for complete paternal demethylation[227]. Finally, components of the BER seem to specifically localize to the paternal nucleus and are necessary to reach low DNA methylation levels[77]. Paternal demethylation is completed passively by replication-dependent depletion with the lowest levels of DNA methylation measured in blastocysts[228,229]. Again, IAP sequences escape demethylation, as well as paternally imprinted loci and centromeric regions marked by heterochromatin[217,230]. The maternal genome exhibits generally lower levels of DNA methylation[223,224] and is protected from active demethylation by PcG protein PCG7. PCG7 further protects paternally imprinted loci[231,232]. Furthermore, the oocyte-derived maintenance activity (DNMT1o) is excluded from the nuclei during early cell divisions[233]. Loss of methylation in the maternal genome completely depends on passive, replication-dependent depletion of mCpGs[226]. Similar to the paternal genome, lowest levels of DNA methylation are measured in the blastocyst stage[208].

After specification of the inner cell mass (ICM) and implantation of the embryo, the process of gastrulation starts. It is accompanied by global remethylation of the genome, which contributes to lineage restriction and loss of pluripotency[213,234-236]. Loss of *Tet1* expression during implantation, which replaces Tet3 activity in later stages of preimplantated embryos[237], seems to be essential for re-methylation in the epiblast and subsequent stable silencing of genes in somatic tissues[238].

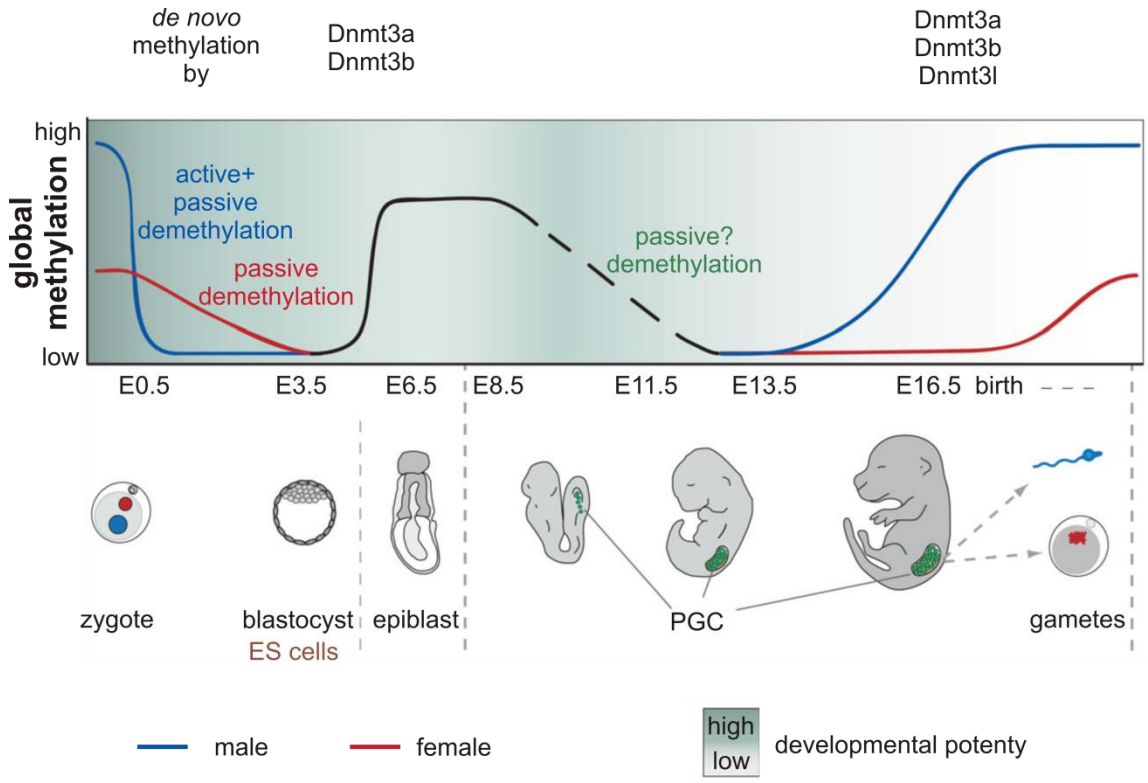


Figure 1-6 Reprogramming events during early embryonic development
In the first reprogramming event following fertilization, DNA methylation marks of the two parental gametes are erased in two independent processes. Firstly, the paternal pronucleus (in blue) undergoes rapid demethylation, potentially in a combination of active and passive mechanisms. Subsequently, the maternal pronucleus (in red) gradually loses DNA methylation in a passive process involving several cell divisions. Re-establishment of DNA methylation starts in the inner cell mass (ICM) of the embryonic blastocyst in the course of advancing differentiation and constitutes an epigenetic barrier (dashed line) in the developmentally more restricted epiblast. The epigenetic signature is inherited to the primordial germ cells (PGCs), but erased again on a global scale in a second reprogramming event to restore full developmental pluripotency. Differential DNA methylation patterns at imprinted regions, which are protected from first reprogramming event in the fertilized zygote, are also reset in PGCs. With progressing development into fully specialized gametes, DNA methylation is re-established leading to a restricted developmental potency. This epigenetic barrier (dashed line) will be removed once more in the zygote of the next generation. (Adopted from Seisenberg *et al.*[208])

1.3 Regulators of inter-individual epiphenotypes

Nowadays the term „epigenetics“ is used for a plethora of regulatory mechanisms that alter gene activity and expression without a change in the actual genomic sequence. The importance of this layer of regulation is accepted widely. In particular, the number of diseases associated with epigenetic aberrations is increasing rapidly and comprises, among many others, cardiovascular[239,240], pulmonary[241], psychiatric[242], and dementia disorders[243,244] as well as cancer[245-247]. In this context, it is still not fully understood if aberrant epigenetic patterns are the cause or consequence of these illnesses[9,248]. For some illnesses, distinct epigenetic patterns were linked to disease susceptibility[249]. In addition to the implicated effects on medical conditions, epigenetic variations affect the healthy organism by shaping behavioral, biochemical, and physiological phenotypes. Deeper understanding of factors that regulate the epigenome will be crucial to further unwind the complex network that determines cell identity and thus the fundaments of health and disease.

The individual convertibility of the epigenome can be addressed, for example, in studies with monozygotic twins. Despite sharing the identical genetic blueprint, epigenetic variations were described for monozygotic twins[250-253]. Although environmental effects such as nutrition are known to affect the epigenome[254-256], their influence does not seem to be sufficient to explain the entire spectrum of observed variations. Furthermore, the majority of studies addressed environmental effects *in utero* and should thus not affect twins. The idea of spontaneous epimutations and the formation of epialleles arose[257]. However, the changes might also be caused by spontaneous genetic aberrations introduced, for example, by transposable elements or point mutations[258].

One of the most-intensively studied epigenetic modifications is DNA methylation, which is involved in gene silencing[12,14,136]. Still, very little is known about how *de novo* methylation is targeted to specific sites in the genome. In case of the maintenance DNA methyltransferase *Dnmt1*, the local pattern of hemi-methylated CpG dinucleotides could be sufficient to target the enzymatic activity[10,47]. However, *de novo* DNA methyltransferases, which exhibit no sequence specificity[24,38,259], require other factors involved in a targeting mechanism. Several studies could demonstrate that the local genomic sequence affects site-specific epiphenotypes[260-262]. A possible explanation might be that each allele permits the binding of distinct regulators that recruit different sets of epigenetic modifiers. However, *cis*-acting genetic determinants of epigenetic variations reveal little about factors that direct site-specific epigenetic regulation. These factors, however, are important, as they provide fundamental insight into epigenetic regulation. Despite the increasing number of studies characterizing factors of the epigenetic targeting machinery[96,98,99], the catalog of described targeting factors is rather insufficient. Hence, further investigation on the influence of genetic determinants on the epigenetic variation is required.

2 Research Objectives

Epigenetic alterations contribute to phenotypic variations and are assumed to affect disease susceptibility. The effects of epigenetic mechanisms were extensively studied in the past decade. However, very little is still known about the key players directing epigenetic modifications to their target site. Recent studies revealed that although epigenetic mechanisms do not alter the genomic sequence of their target, their recruitment depends on genetic determinants that can either act in *cis*[260-262] or in *trans*[98,99,263]. *Trans*-acting epigenetic modifiers are of particular interest, since the knowledge of their function will improve the basic understanding of epigenetic regulation and malfunction.

The main focus of the present thesis was the identification of the epigenetic modifier responsible for strain-specific DNA methylation at the *Isoc2b* promoter region. Linkage analysis in hybrids of the differentially methylated mouse strains should determine the genomic localization of the responsible genetic trait. Since the differentially methylated region (DMR) seemed to be established during early embryogenesis, an *in vitro* embryonic stem cell (ESC) differentiation model should recapitulate the epiphenotypes observed *in vivo*. ESCs would represent a convenient model for further characterization of the *trans*-acting epigenetic modifier. Moreover, global profiling of active histone marks should help to identify novel DMRs and thus further potential targets of the epigenetic modifier. Additional target sites would facilitate the understanding of the involved regulatory network.

3 Materials & Equipment

3.1 Equipment

8-Channel PipettorImpact2 Equalizer 384	Thermo Fisher Scientific, Hudson, NH, USA
Autoclave	Technomara, Fernwald, Germany
Bioanalyzer 2100	Agilent Technologies, Böblingen, Germany
Biofuge fresco	Heraeus, Osterode, Germany
BioPhotometer	Eppendorf, Hamburg, Germany
Caliper LabChip	Perkin Elmer, Waltham, MA, USA
Casy Cell Counter	Innovatis/Roche, Basel, Schweiz
Covaris S2	Covaris, Woburn, USA
Cryo 1 °C Freezing Container	Thermo Fisher Scientific, Hudson, NH, USA
Electrophoresis equipment	Biometra, Göttingen, Germany
Electrophoresis equipment (PFGE)	Biostep, Jahnsdorf, Germany
Gel Doc XR+ System	Bio-Rad Laboratories, Munich, Germany
GenePulser II	Bio-Rad, Munich, Germany
Heat sealer	Eppendorf, Hamburg, Germany
Hemocytometer	Marienfeld, Lauda-Königshofen, Germany
Illumina Genome Analyzer	Illumina, San Diego, CA, USA
Illumina HiSeq1000	Illumina, San Diego, CA, USA
Incubators	Heraeus, Hanau, Germany
Intelli-Mixer RM-2L	Elmi-Tech, Riga, Latvia
LC480 Lightcycler	Roche, Basel, Switzerland
Magnetic Partical Concentrator	Life Technologies, Carlsbad, CA, USA
MassARRAY Compact System	Sequenom, San Diego, CA, USA
MassARRAY MATRIX Liquid Handler	Sequenom, San Diego, CA, USA
MassARRAY Phusio chip module	Sequenom, San Diego, CA, USA
Megafuge 3,0 R	Heraeus, Osterode, Germany
Microscopes	Zeiss, Jena, Germany
Multifuge 3S-R	Heraeus, Osterode, Germany
Multipipettor Multipette plus	Eppendorf, Hamburg, Germany
NanoDrop 1000	PeqLab, Erlangen, Germany
PCR Thermocycler PTC-200	MJ-Research/Biometra, Oldendorf, Germany
pH-Meter	Knick, Berlin, Germany
Picofuge	Heraeus, Osterode, Germany
Pipetboy	Integra Biosciences, Fernwald, Germany
Pipettes	Gilson, Milddleton, WI, USA
Pipettes	Eppendorf, Hamburg, Germany

Power supplies	Biometra, Göttingen, Germany
Pulsed Field Electrophoresis equipment	GE Healthcare, Chalfont St Giles, UK
Pulsed field gel electrophoresis equipment	Biostep, Jahnsdorf, Germany
QIAvac 96	Qiagen, Venlo, The Netherlands
Qubit	Life Technologies, Carlsbad, CA, USA
Realplex Mastercycler epGradient S	Eppendorf, Hamburg, Germany
Rocking plattform HS250	IKA Labortechnik, Staufen, Germany
Sonifier 250	Branson, Danbury, USA
Sonifier 250	Branson, Danbury, CT, USA
Sonorex Ultrasonic Bath	Bandelin, Berlin, Germany
Sorvall RC 6 plus	Thermo Fisher Scientific, Hudson, NH, USA
Speed Vac	Christ, Osterode, Germany
Thermomixer	Eppendorf, Hamburg, Germany
TissueLyser	Qiagen, Venlo, The Netherlands
Typhoon 9200	Molecular Dynamics, Krefeld, Germany
UV Stratalinker 4800	Stratagene, La Jolla, CA, USA
Veriti 384 well Thermal Cycler	Life Technologies, Carlsbad, CA, USA
Vortexer	Scientific Industries Ink., Bohemia, NY, USA
Water purification system	Millipore, Eschborn, Germany
Waterbath	Julabo, Seelstadt, Germany

3.2 Consumables

8-channel pipettor tips Impact 384	Thermo Fisher Scientific, Hudson, NH, USA
Adhesive PCR sealing film	Thermo Fisher Scientific, Hudson, NH, USA
Ampure XP Magnetic Beads	Beckman Coulter Genomics, Krefeld, Germany
Cell culture dishes	Nunc/Thermo Fisher Scientific, Hudson, NH, USA
Cell culture dishes	Greiner Bio-one, Frickenhausen, Germany
CLEAN resin	Sequenom, San Diego, CA, USA
Cryo tubes	Corning, Corning, NY, USA
Disposable scalpel (No. 11)	Feather, Osaka, Japan
Dynabeads Magnetic Beads	Life Technologies, Carlsbad, CA, USA
GenePulser Electroporation Cuvette (1mm)	Bio-Rad, Munich, Germany
Heat sealing film	Eppendorf, Hamburg, Germany
Hybridisation Mesh	Labnet, Woodbridge, NJ, USA
MATRIX Liquid Handler D.A.R.Ts tips	Thermo Fisher Scientific, Hudson, NH, USA
Micro test tubes (0.5 ml, 1.5 ml, 2 ml)	Eppendorf, Hamburg, Germany
Micro test tubes DNA LoBind (0.5, 1.5, 2 ml)	Eppendorf, Hamburg, Germany
Non-Adhesive sealing film (LightCycler)	Roche, Basel, Switzerland
nProteinA Sepharose 4 FastFlow	GE Healthcare, Munich, Germany
Nylon membrane	Roche, Basel, Switzerland
PCR plate 384 well (LightCycler)	Roche, Basel, Switzerland
PCR plate 384 well (MassARRAY)	Thermo Fisher Scientific, Hudson, NH, USA
PCR plate Twin.tec 96 well	Eppendorf, Hamburg, Germany
Pipettes	Costar, Cambridge, USA
Plug molds	Bio-Rad, Munich, Germany
Sepharose CL-4 Beads	Sigma-Aldrich, Munich, Germany
SpectroCHIP bead array	Sequenom, San Diego, CA, USA
Sterile combitips for Eppendorf multipette	Eppendorf, Hamburg, Germany
Sterile micropore filters	Millipore, Eschborn, Germany
Sterile plastic pipettes	Costar, Cambridge, USA
Syringes and needles	Becton Dickinson, Heidelberg, Germany
Tubes (5ml, 15 ml, 50 ml, 220 ml)	Falcon, Heidelberg, Germany

3.3 Chemicals

All chemicals were purchased from Sigma-Aldrich (Deisendorf, Germany) or Merk Millipore (Darmstadt, Germany) unless otherwise mentioned.

3.4 Enzymes, Kits, and products for molecular biology

Agarase	NEB, Frankfurt, Germany
Baysilone (Silicone for PFGE)	Bayer, Leverkusen, Germany
Bioanalyzer DNA HS Kit	Agilent Technologies, Santa Clara, CA, USA
Bioanalyzer mRNA Nano Kit	
Bioanalyzer mRNA Pico Kit	
Bioanalyzer RNA 6000 Kit	
Bovine Serum Albumin	Sigma-Aldrich, Munich, Germany
CasyTON	Roche, Basel, Switzerland
Cell Culture Medium Supplement	Life Technologies, Carlsbad, CA
DIG PCR labeling Kit	Roche, Basel, Switzerland
DNA Ladder 1 kb plus	Invitrogen/Life Technologies, Carlsbad, CA, USA
DNA ladder 50 bp	NEB, Frankfurt, Germany
DNA ladder Generuler 50 bp	Thermo Fisher Scientific, Waltham, MA, USA
DNA ladder λ HindIII Fragments	Invitrogen/Life Technologies, Carlsbad, CA, USA
DNA light loading dye (6x)	NEB, Frankfurt, Germany
DNeasy Blood & Tissue Kit	Qiagen, Venlo, The Netherlands
dNTPs	GE Healthcare, Buckinghamshire, UK
dUTP	GE Healthcare, Buckinghamshire, UK
Enzymatics enzymes for NGS library Prep	NEB, Frankfurt, Germany
Ethidium bromide	Sigma-Aldrich, Munich, Germany
Expand High Fidelity PCR System	Roche, Basel, Switzerland
EZ DNA methylation kit	Zymo Research, Orange, USA
Fermentas DNA loading dye (6x)	Thermo Fisher Scientific, Waltham, MA, USA
Fetal Bovine Serum	Gibco/Life Technologies, Carlsbad, CA, USA
Glycogen	Ambion/Life Technologies, Carlsbad, CA, USA
L(+)-Arabinose	Sigma-Aldrich, Munich, Germany
LabChip XT DNA Chips	Perkin Elmer, Waltham, MA, USA
Lipofectamin 2000	Invitrogen/Life Technologies, Carlsbad, CA, USA
Lysozyme	Sigma-Aldrich, Munich, Germany
MinElute PCR Purification Kit	Qiagen, Venlo, The Netherlands
NEBNext DNA Library Prep Reagent Set	NEB, Frankfurt, Germany
NEXTflex indexed adapters	Bioo Scientific, Austin, TX, USA
Nuclease-free water	Gibco/Life Technologies, Carlsbad, CA, USA

NucleoBond BAC 100 Kit	Macherey-Nagel, Düren, Germany
NucleoSpin Plasmid Quick Pure Kit	Macherey-Nagel, Düren, Germany
Oligo Only Kit	Illumina, San Diego, CA, USA
Phusion Hot Start High-Fidelity DNA Polymerase	Thermo Fisher Scientific, Waltham, MA, USA
Phusion Taq Polymerase	Thermo Fisher Scientific, Waltham, MA, USA
PI-SceI	NEB, Frankfurt, Germany
Proteinase K	Roche, Basel, Switzerland
QIAquick Gel Extraction Kit	Qiagen, Venlo, The Netherlands
QIAquick PCR Purification Kit	Qiagen, Venlo, The Netherlands
QuantiFast SYBR green	Qiagen, Venlo, The Netherlands
Qubit DNA HS Kit	Life Technologies, Carlsbad, CA
Random Decamers	Ambion/Life Technologies, Carlsbad, CA, USA
Restriction endonucleases	NEB, Frankfurt, Germany Roche, Basel, Switzerland
Reverse Transkriptase SuperSkript II Kit	Promega, Madison, WI, USA
RiboZero Kit (mouse)	Epicentre, Madison, Wi, USA
RNase	Roche, Basel, Switzerland
RNase A	Sequenom, San Diego, CA, USA
Rnase T1 (cloned)	Life Technologies, Carlsbad, CA
Rneasy Micro & Mini Kit	Qiagen, Venlo, The Netherlands
rRNA Removal Magnetic Kit (mouse)	Epicentre, Madison, Wi, USA
SAM (S-adenosyl-methionine)	NEB, Frankfurt, Germany
ScriptSeq Kits	Epicentre, Madison, WI, USA
Shrimp Alkaline Phosphatase (SAP)	Sequenom, San Diego, CA, USA
Spectinomycin	Sigma-Aldrich, Munich, Germany
SssI CpG methylase	NEB, Frankfurt, Germany
StrataClone™ PCR Cloning Kit	Agilent Technologies, Santa Clara, CA, USA
Taq DNA Polymerase	Roche, Basel, Switzerland
T-Cleavage MassCleave Reagent kit	Sequenom, San Diego, CA, USA

3.5 Antibiotics

Ampicillin	Ratiopharm, Ulm, Germany
Chloramphenicol	Sigma-Aldrich, Munich, Germany
Kanamycin	Roth, Karlsruhe, Germany
Spectinomycin	Sigma-Aldrich, Munich, Germany
Geneticin	Gibco/Life Technologies, Carlsbad, CA, USA

3.6 Molecular weight standards

Generuler 50 bp, 50bp, 1 kb Plus, and Lambda DNA-HindIII Fragments DNA Ladders were purchased from Thermo Fisher Scientific (Waltham, MA, USA), Invitrogen/Life Technologies (Darmstadt, Germany), and NEB (Frankfurt am Main, Germany), respectively.

3.7 Cell lines

JM8	murine embryonic stem cells (C57BL/6N background), provided by AG Gessner (University Hospital Regensburg, Regensburg, Germany)
BALB/c	murine embryonic stem cells (BALB/cJ background), provided by AG Gessner (University Hospital Regensburg, Regensburg, Germany)
MEF	Mouse embryonic fibroblast generated from mouse embryos E14.5 (129/SvJ x C57BL/6 background) in our laboratory

3.8 *E. coli* strains

StrataClone SoloPack competent cells express	Agilent, La Jolla, CA, USA
SW106	provided by Don Court Lab (GRCBL, Frederick National Laboratory for Cancer Research, Frederick, MD, USA)
DH10B	Invitrogen/Life Tech., Darmstadt, Germany

3.9 Deoxyribonucleic acids

3.9.1 Murine genomic DNA

Murine genomic deoxyribonucleic acid (DNA) was in general prepared according to section 4.3.1.2 from mice that were bred either at Jackson Laboratories or Charles River Laboratories or, as stated for some hybrids, in our own stable. The only exception is genomic DNA from BALB/cJ and BALB/cByJ spleen that was prepared and distributed by Jackson Laboratories.

3.9.2 Circular DNA

3.9.2.1 Plasmids

pSC-A-amp/kan	Agilent, La Jolla, CA, USA
pNelγ	provided by George Kotzamanis (Dept. Histology and Embryology Medical School, University of Athens, Greece)
pJM2545	provided by the Larin Monaco Group (Wellcome Trust Centre for Human Genetics, Oxford, UK)

3.9.2.2 Bacterial Artificial Chromosomes

Using DIG-labeled probes to screen for BACs representing parts of the 9 Mb candidate region on chromosome 12 (*mm9*_chr12:16902483-26024113), the BACs listed in the following table were selected from the CHORI-28 BAC library (BALB/cByJ background). The anchor BAC was selected from the RPCI-94 BAC library (*C. briggsea* background).

Table 3-1 List of Bacterial Artificial Chromosomes

RP94-1A1	*Anchor BAC*	CH28-313E22	CH28-334B18	CH28-360N6
CH28-277O6		CH28-314I19	CH28-335P5	CH28-363H2
CH28-290B2		CH28-315E5	CH28-335M3	CH28-363M8
CH28-292M2		CH28-315O11	CH28-335E2	CH28-363M15
CH28-292G19		CH28-316I23	CH28-339I15	CH28-363K13
CH28-295K18		CH28-316E21	CH28-339K16	CH28-364E16
CH28-295F17		CH28-317N16	CH28-339P13	CH28-365B22
CH28-295G20		CH28-317O18	CH28-340B11	CH28-365M12
CH28-296C19		CH28-317D4	CH28-343C13	CH28-365O7
CH28-297K5		CH28-318E2	CH28-343N22	CH28-365O12
CH28-300I24		CH28-320F16	CH28-343O11	CH28-367A20
CH28-301D22		CH28-320E21	CH28-344E3	CH28-367A18
CH28-301N22		CH28-320I8	CH28-345K5	CH28-368L22
CH28-302A6		CH28-322L14	CH28-345O3	CH28-369P21
CH28-301J16		CH28-322M24	CH28-347B19	CH28-369L12
CH28-302G5		CH28-324E19	CH28-348E10	CH28-371D6
CH28-302K17		CH28-325D12	CH28-352J5	CH28-372J22
CH28-303D21		CH28-325I8	CH28-354K16	CH28-372P3
CH28-304M14		CH28-327A7	CH28-355C15	CH28-373C16
CH28-306H10		CH28-327L17	CH28-355I8	CH28-375E19
CH28-307N18		CH28-328L21	CH28-356C17	CH28-375N9
CH28-310I2		CH28-330D9	CH28-357O17	CH28-375K3
CH28-311A16		CH28-331F12	CH28-358G15	
CH28-311E14		CH28-332I14	CH28-358K12	
CH28-311M1		CH28-332M24	CH28-359M10	

3.10 Oligonucleotides

The following oligonucleotide sequences are complementary to the indicated reference genome, exceptions are marked. The primers were purchased from Metabion or Sigma-Aldrich, respectively.

3.10.1 PCR primers

Table 3-2 PCR primer sequences

Gene symbol	Genome	Primer sequence (sense & antisense)
Nol10	mm9	5'-TGTCATGTAATCTGATGGGCT-3'
		5'-CTTCATTAGTCTCTGCTTGGT-3' (antisense specific for C57BL/6 strains)
		5'-CTTCACTAGTCTCTGCTTGGT-3' (antisense specific for BALB/c strains)

3.10.2 RT-qPCR primers

Table 3-3 RT-qPCR primer sequences

Gene symbol	Genome	Primer sequence (sense & antisense)
Isoc2a	mm9	5'-GATGCCTGTTCTTCTCGAAGCC-3'
		5'-GGAAAGCTCCACTCTGTCTCATCC-3'
Isoc2b	mm9	5'-TGGCGTCTTTCTGTCCACGAG-3'
		5'-TTGCCTTGGAAGAAGCCTAACAG-3'
Nol10_exon 12-14	mm9	5'-TTAATGATGTTTGCCTCTATCCCAGCTC-3'
		5'-CAAGTTGTCCAGGAAGGAACACCA-3'
Nol10_exon19-20	mm9	5'-AGGAGCAACTGAAGGAGGACCA-3'
		5'-GCTTCAAGTTTCAAACGATCTTCCAGAG-3'
Oct4	mm9	5'-GAAAGCGAACTAGCATTGAGAACCGT-3'
		5'-TTACAGAACCATACTCGAACCACATCC-3'
Rn18s (18S)	hg18	5'-ACCGATTGGATGGTTTAGTGAG-3'
		5'-CCTACGGAAACCTTGTTACGAC-3'

3.10.3 qPCR primers

Table 3-4 qPCR primer sequences

Gene symbol	Genome	#	Primer sequence (sense & antisense)
Gapdh	mm9		5'-CCCATCACGTCCTCCATCATCC-3'
			5'-ACGGGTCTAGGGATGCTGGT-3'
"empty"	mm5		5'-GAAACCCTCACCCAGGAGATACAC-3'
			5'-TGCAGTGGGACTTTATTCCATAGAAGAG-3'
"intergenic"	mm9		5'-GGGAGCGAAGGATACAGGGA -3'
			5'-CTACAGAGGACAGCAGAGAAGCA-3'
Isoc2b	mm9	1	5'-GCAAGGGAGGTCTCCAACAC-3'
			5'-CATTCAGGGAGCACAGATGGGA-3'
		2	5'-TTGTGAGCCATTCTGGAAGAGGAG-3'
			5'-GGCTATGCAGTCTCAGGTCAAACAC-3'
		3	5'-CTCAGTTCGCCAGATTTGATCGAG-3'
			5'-CAATTACGTGCTCCTCTGTATCGCT-3'
		4	5'-TCTGAATGACGGGATGGTGGG-3'
			5'-CGAACTCGAACCTGGGACCT-3'
		5	5'-CGTTCGCCTCACTCTGTTCTACTC-3'
			5'-TTTCATTGGTCCTACCCTCTGTCTC-3'
		6	5'-AATCTTGGCTAGTGTAGGGTTCATGG-3'
			5'-ACTCGTGCTTCAGCCTTCCT-3'
		7	5'-ATGGTGGCGCAAATGTTTAGTTCC-3'
			5'-TTTCGAGACAGGACTTACTTCCGT-3'
		8	5'-GTACTCAGAAGTGGAGACAGGAGG-3'
			5'-GGGTTTACAGTCCTCGTTAGCCA-3'
		9	5'-ATGGGATGGGACTGGGAAGG-3'
			5'-GCACTTGAGTTCACTGTTCTGGG-3'
Myod1	mm9		5'-GCATTCCAACCCACAGAACCT-3'
			5'-GAGTCGTCTTAACTTTCTGCCACTC-3'
Snrpn	mm8		5'-ACATTCCGGTCAGAGGGACAGAG-3'
			5'-CCGCAATGGCTCAGGTTTGTC-3'
"TARBAC2.1"	-		5'-CAATGCCTGCCGTATATCCTTACTG-3'
			5'-CGTATGCTGTATCTGTTCGTTGACC-3'
TRPE	*E. coli*		5'-TGGTTTCCGTGCCGCATAT-3'
			5'-AATTTCTCCAGCGCGAATCG-3'

3.10.4 Primers for the MassARRAY® system (Sequenom)

The primers were purchased from Sigma-Aldrich.

3.10.4.1 EpiTYPER primers

Table 3-5 EpiTYPER primer sequences

Gene symbol	Genome	ID		Primer sequence
Isoc2b	mm8	Epi19	10F	5'-aggaagagagGGGGATATAGATGGGAGAGGATTTA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAACTATTTACAAACCCACCATCCC-3'
		Epi21	10F	5'-aggaagagagGGTGGGTTTGTAAATAGTTAGTAAGTAA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAACTCTCAACCAAAACCCTCCCT-3'
		Epi22	10F	5'-aggaagagagTTGGTAAGTAGTGGGGGAAAATGTTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAATTCATAACACCATCCACTCCATC-3'
Nanog	mm9	JW4	10F	5'-aggaagagagTGGTAGAATTAAGAGGTTGTTAGGTTGT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTTAAACCCATAAATCAAAAAAATCAAAATA-3'
Pdgfrb	mm8	Epi26	10F	5'-aggaagagagGGTGATTGAAGGTTTTAGGTTGTTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTCCAAAAACCCTCCACTAAAAATAAAC-3'

Table 3-6 EpiTYPER primer sequences for novel candidate DMRs

Gene symbol	Genome	ID		Primer sequence
1700012B07Rik	mm9	SQ62_C57BL/6	10F	5'-aggaagagagGGGAAATTTTTGGGTTAGTAGGAAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAAACACAAACAAATCCTATCCCC-3'
		SQ62_BALB/c	10F	5'-aggaagagagGGGAAATTTTTGGTTTAGTAGGAAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAACCACAAACAAATCCTATCCCC-3'
		SQ63_C57BL/6	10F	5'-aggaagagagGTTGGTTTAGGGGTTTTTTATTTTTTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACCAACCCATCCCACTAATTTCATA-3'
		SQ63_BALB/c	10F	5'-aggaagagagGTTGGATTAGGGGTTTTTTATTTTTTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACCATCCCATCCCACTAATTTCATA-3'
1700030K09Rik	mm9	LaMa06	10F	5'-aggaagagagGATTTGAGTTGTTTATTGGTGGGTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAACCCTAAAAAATCATACCCCATC-3'
		LaMa07	10F	5'-aggaagagagTGTTTAAATTGGGAGTTATTGGTGGAT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAATAAACAACTCAAATCACAATCCC-3'
2410066E13Rik	mm9	LaMa42	10F	5'-aggaagagagTTTAGGTGGGGGAGGTGAGTTTAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACTACCAAAACATCTCCATAAAAACCC-3'
2610035D17Rik	mm9	SQ77_C57BL/6	10F	5'-aggaagagagTGGGTTAAAGGAAAGGTATTGGTTAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAAATAACACCCATATATCCCTATTC-3'
		SQ77_BALB/c	10F	5'-aggaagagagTGGGTTAAAGGAAAGGTATTGGTTAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAAATAACACTCATATATCCCTAATC-3'
4931429I11Rik	mm9	SQ29	10F	5'-aggaagagagGTATAGTTTGTGGTTTGGGAGGAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctATCCAAAATTACCCTTTTTAAAACAAATAA-3'
		SQ30	10F	5'-aggaagagagGTATAGTTTGTGGTTTGGGAGGAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAACAATAAATCCAAAATTACCCTTTTTAAA-3'
6330545A04Rik	mm9	SQ75	10F	5'-aggaagagagTTGTGATTGTGGAGAAGTTGGTTAGAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCCAAACTCAAAATACAACCAATTCAATC-3'
		SQ76	10F	5'-aggaagagagTTGGTTTTGTTTGTTTTGTTTGTAATTAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTCAAAAATAACACTTAATCTCCCCC-3'
a	mm9	LaMa11	10F	5'-aggaagagagAGGGTAGAATAGGTTTTGAGAGTGATAATG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTTTCAAAAATCCAACCAACAAATAAC-3'
		LaMa12	10F	5'-aggaagagagTAATTTTTTAAGAGTTGTAGAAGGGTTTAA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAAACTAATAAAACCTAAAACCAAAAC-3'
		LaMa13	10F	5'-aggaagagagGTTTTTTGGGAGGGTTTTTAGGAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAAACTATTCTTACCCCTCCCTTAACC-3'

Gene symbol	Genome	ID		Primer sequence
Acacb	mm9	SQ37	10F	5'-aggaagagagAGTAGGATGGAAGGTTTGGTTTATGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAAAACACAACAAAATCACCAACTC-3'
		SQ38_C57BL/6	10F	5'-aggaagagagAGGTGTTAAGGGGTAGATAGTTGATTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTAATTTAAATTTCCCAAAATCTAAAACAAT-3'
		SQ38_BALB/c	10F	5'-aggaagagagAGGTGTTAAGGGGTAGATAGTTGATTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAATTTAAATTTCCCAAAATCTAAAACAAT-3'
AI480653	mm9	LaMa52	10F	5'-aggaagagagGGATTGGTTATATGTGTTAGTTTTTTGAAT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTAAAATCTCCCCAACTCTCTCCTAC-3'
		LaMa53_C57BL/6	10F	5'-aggaagagagTAGAGGGTTGGGTAGATTTAGGAGAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCCCAAAACACTTACACAAAAATCAAC-3'
		LaMa53_BALB/c	10F	5'-aggaagagagTAGAGGGTTGGGTAGATTTAGAGAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCCCAAAACACTTACACAAAAATCAAC-3'
Aim1l	mm9	SQ61_C57BL/6	10F	5'-aggaagagagTTTGAGGTTGTTGGAGTTATTGAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACCCAACCCCCAAAAAATACTAACT-3'
		SQ61_BALB/c	10F	5'-aggaagagagTTTGAGGTTGTTGGAGTTATTGAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACCCAATCCCCAAAAAATACTAACT-3'
Ang	mm9	LaMa54	10F	5'-aggaagagagTTTAGGGGTTTATGTAGATTGAGTGAAT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAACCAAAAAACTTCCTTTCCTAC-3'
		LaMa55_C57BL/6	10F	5'-aggaagagagTTTGAAATTTAGATTTAAAGGATATTTTTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTAAAATCCTAAATACCTCCAACAAC-3'
		LaMa55_BALB/c	10F	5'-aggaagagagTTTGAAATTTAGATTTAAAGGATATTTTTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTAAAATCCTAAATACCTCCAAAAAC-3'
Apobec2	mm9	LaMa28	10F	5'-aggaagagagTTTTTAGGGGAGGAAAGGAAGTAGAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAAACAAAAACTAAAACCACACCCC-3'
		LaMa29	10F	5'-aggaagagagTGTTTATTTTGTGGGTTTGGGATTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTTTCTTCCTCTCCCATAAATTTAAAATTTC-3'
Apobec3	mm9	SQ35_C57BL/6	10F	5'-aggaagagagGTGGGTAGTTTAGTTGTAAGTGGTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAAACCTTAATCAAAAAAATAAAATC-3'
		SQ35_BALB/c	10F	5'-aggaagagagGTGGGTAGTTTAGTTGTAAGTGTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAAACCTTAATCAAAAAAATAAAATC-3'
		SQ36_C57BL/6	10F	5'-aggaagagagTTTGGTTTAATTGGTAGTTTTGGTTGTAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTACAAACCCCACCTTCTTCCTACAT-3'
		SQ36_BALB/c	10F	5'-aggaagagagTTTGGTTTAATTGGTAGTTTTGTTTGTAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTACAAACCCCACCTTCTTCCTACTT-3'
Atp9a	mm9	SQ17	10F	5'-aggaagagagGGTATAAAGAGTTTATAAGATGGTAGTTGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTAAAAAAAATACAAAATCCCAAACC-3'
		SQ18	10F	5'-aggaagagagGGGTAGGGAGGAGGGGATATTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAACCAAACACAAAAAACCCACAA-3'
B3galtl	mm9	SQ64	10F	5'-aggaagagagTGGGGTGGGTTTATTTGGTAGATAGT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACCAAAACTTCTATTCCCTCAACCA-3'
C530028O21Rik	mm9	SQ15	10F	5'-aggaagagagTTGGGTGTTTTTTTTAAATTTTTTGGT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACTCCAAAACATACCTACCCCTTC-3'
		SQ16	10F	5'-aggaagagagAGGGATTTGGAGGGAGTATTTAGAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAAAACCCAAAACCCCTACCC-3'
Ccnjl	mm9	LaMa04	10F	5'-aggaagagagGAGTTAGTGATTTAGTTAGTTTTGGGAGAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCCTACCCTCAATAACCCTAATCCTC-3'
Col6a4	mm9	SQ65_C57BL/6	10F	5'-aggaagagagTATTTTTAGAGGTAGGTTGGGTGTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAAAAATTTTCCTTCCAAAAAAAA-3'
		SQ65_BALB/c	10F	5'-aggaagagagTATTTTTAGAGGTAGGTTGGATGTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAAAAATTTTCCTTCCAAAAAAAA-3'

Gene symbol	Genome	ID		Primer sequence
Dcaf7	mm9	LaMa37	10F	5'-aggaagagagGGGGTGGGATTTTATATATAATATATTTTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCTCCTTATAAATCCAAAATCTAAACC-3'
		LaMa38	10F	5'-aggaagagagGGTTGTGGAGTTTTTGTTTTTAGTTGAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCCCAACCTCCCTAAAAATCCCTATA-3'
Dnahc2	mm9	SQ31	10F	5'-aggaagagagTATTTGAGATTATAGTTGAGGGGAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctATATTCAAAATCTTCAAAACCTCTC-3'
Dnajc4	mm9	SQ40	10F	5'-aggaagagagGTAGGGGTAGTTGGAGTAATAGGGA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTACCCCCAATAAAAAATCAAAAAC-3'
		SQ41	10F	5'-aggaagagagATTGAGATTTAATGGATTTTTGGGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAACAAACCAAAACTCAAATTCCC-3'
		SQ42	10F	5'-aggaagagagTTAAGGAGTTTGTGAGTGGGGTAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAACCCCCAAAAATCCATTAAATCTC-3'
Dst	mm9	SQ03	10F	5'-aggaagagagTTTAGGGTTTTTTAAAGTTATTTATTTTAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTACTTTCCCCAACCCATCTTAAAAC-3'
		SQ04	10F	5'-aggaagagagTAGTAATTTGGGTTATGGTGAGTGGA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTTCTATTTAAAACCCACAACAAATACC-3'
		SQ05	10F	5'-aggaagagagTATTTTTATAGGTTTGGAGGTATTTTTGTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAACTTCCAACCTCCTCAATAAACAC-3'
Fabp6	mm9	LaMa31	10F	5'-aggaagagagTGAGGAGGGTTTATAGAGAGAGAAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAAAAAACTCAAACTTTCCCAAAC-3'
		LaMa32	10F	5'-aggaagagagTTTTTGTTTGTAATTGTGTAGGTTAAAAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACAACTCATCCCCTCAAAATTCTATC-3'
Foxn3	mm9	SQ72	10F	5'-aggaagagagGGAAGGTGGTTTTTATTTGGTAAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACTTAATCTTACTCCTTTCTCCCCTATTTC-3'
		SQ73	10F	5'-aggaagagagTGTAAATTTTTTGAGAAAAATTATAGAGTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctATAATAACTTACCAAAAATAACATTCTCTC-3'
Fut10	mm9	SQ32	10F	5'-aggaagagagAAGTATTGATATGATGTTTTGAAATAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctATAATATAATCTCTTCAACAAACCC-3'
		SQ33	10F	5'-aggaagagagGGAATAAGGTTAGTTGATGTTTAGGAT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAAACCAAATCAAACCCTACCTA-3'
Galnt10	mm9	LaMa47	10F	5'-aggaagagagAGGGAAATTTGAAGGTTGTTAAAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAATATCACCCAACAAAAAAACAATTC-3'
		LaMa48_C57BL/6	10F	5'-aggaagagagGGAAGTTTTTGTGGTTTTTGTTTAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAAACAACCCAACCCTCTATAACC-3'
		LaMa48_BALB/c	10F	5'-aggaagagagGGAAGTTTTTGTGGTTTTTGTTTAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTAAAACAACCCAACCCTCTATAATC-3'
Gm8884	mm9	LaMa25	10F	5'-aggaagagagATGGGAGATGGTTAAGTTTTAGGGTAAAT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAATACCCTACCCCCACCAAAAAT-3'
H2-Ke6	mm9	LaMa17	10F	5'-aggaagagagGAGGGAGGGGGAGGTTAATATTAAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTCACAAATTCAAAAATTCCCTCTCAA-3'
		LaMa18	10F	5'-aggaagagagTTAGTTTAGAGGGAGATGGGTTTTGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTATCCTTCCCCTAACTCCTAATAAAAACAA-3'
H2-T10	mm9	LaMa05	10F	5'-aggaagagagGTTTGGGGAGAGGAGGGTTATAGAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAAAAATCCCTCCCAAACAATAACTC-3'
Hdhd3	mm9	SQ19	10F	5'-aggaagagagAGGTAGGTTGAGGTTTAGTTAGGAATTAGA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAAACAAAAAACCCAAAACTCAAC-3'
		SQ20	10F	5'-aggaagagagGGGGTTAAGTTTATTGTTGGTTGGTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAATCCCCTAAATATCTCCATCCCAC-3'
Hes1	mm9	LaMa16	10F	5'-aggaagagagAAGGGGAATTAAGGGGAATTATATTTGTTA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAAACCCCTACCTCCAACCAAC-3'

Gene symbol	Genome	ID		Primer sequence
Hist2h2bb	mm9	LaMa14	10F	5'-aggaagagagTTTTTGTTGAGTGAAGATTGTTTGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAACCAACCTAAAAAAAATTCCACTATTTTC-3'
		LaMa15	10F	5'-aggaagagagGTAGTTATTAGTGTAATTTGTGGTAGGAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACCACCATCATTTAATTATAAACTATAACC-3'
Hist4h4	mm9	LaMa01	10F	5'-aggaagagagTTGTTTGGGTGTATTTTTTTATTGGAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTAAAATTCCAAACCTCCACACCTAC-3'
Hopx	mm9	SQ06_C57BL/6	10F	5'-aggaagagagTTTAAAAAGAGGATGGTGTTGAGATAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTCCTCCTCTCCATCCTTAATCAAATTA-3'
		SQ06_BALB/c	10F	5'-aggaagagagTTTAGAAAGAGGATGGTGTTGAGATAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTCCTCCTCTCCATCCTTAATCAAATTA-3'
		SQ07	10F	5'-aggaagagagAGTTGTTTTAGGGATTGAGGATATT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctATACTCATAACTCAACCTTCTCCAC-3'
Hoxa1	mm9	SQ12	10F	5'-aggaagagagTTAAGGATGGGGTATTTTAGAAAGGAGTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCCCTTTAATCCCAATACTCCAAATC-3'
		SQ13	10F	5'-aggaagagagTTTGTTATTGAGTGATTGGATTTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAAAAAAACCAAAAAAACCTCATCC-3'
		SQ14	10F	5'-aggaagagagTTAGTTTAGTTTTTTGGGTTGGGTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAATTCATTCTTACATTATCCATCTATCA-3'
Krt10	mm9	SQ50	10F	5'-aggaagagagTTTTGTTATGTGTAGTGTTGAAGGAAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACAACTCTCCCAAATTCAAAACCAA-3'
		SQ51	10F	5'-aggaagagagTTTGGGTTTGGGTAAGTTTTGTTATGT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAATCTTCCTAAATCCCATATTTCACCTTAA-3'
Krt222	mm9	LaMa26	10F	5'-aggaagagagATTTTGGATTTGATTTTTTTGATGGTATTA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAAAAACCTTAAAAACTCCCTACAAACC-3'
		LaMa27	10F	5'-aggaagagagTTGATAATGTTGTTTATTGGTTTGTAGGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAACCAACATAACTTACCACAAAACCC-3'
Lnx2	mm9	LaMa49	10F	5'-aggaagagagTTTGTTTTTTTTGATGATAAGGATTTAGAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAAACCTCACATTCACTAACAAAATTACTT-3'
		LaMa50_C57BL/6	10F	5'-aggaagagagTTTGTGTGGGAATATTTTTAGGATAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCTCAACCAACTATTTCTAATAACCAAAATC-3'
		LaMa50_BALB/c	10F	5'-aggaagagagTTTGTGTGGGAATATTTTTAGGATAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCTCCACCATCTATTTCTAATAACCAAAATC-3'
LOC100233207	mm9	LaMa51_C57BL/6	10F	5'-aggaagagagGGAGGTTTTAATTAGGGGAAGTGTTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACAAAAAAACTAAAACACAAAACCAACC-3'
		LaMa51_BALB/c	10F	5'-aggaagagagGGAGGTTTTAATTAGGGGAAATGTTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACAAAAAAACTATAACACAAAACCAACC-3'
Lpar2	mm9	LaMa36	10F	5'-aggaagagagTAATTTTGGGTTAGGAGAAGGATTGTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACCCCCAAAAACCAAAAACTAAAAC-3'
Ly6d	mm9	SQ57	10F	5'-aggaagagagGGGATTTGGGTTGTTATTTATTTGAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCCCTCATCCTTCAACCTAAACACTT-3'
Lynx1	mm9	SQ34	10F	5'-aggaagagagGTTTTGTTTTAGGTTTGGTTTGAGGTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctATCCATTACCAAAAAAAATTTATCCC-3'
Mllt4	mm9	SQ68	10F	5'-aggaagagagGTTTTAGGAAAAAGGAAGTTTAGGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAACCCACCCCAAAATAAAACTAC-3'
		SQ69	10F	5'-aggaagagagATTGGGGTAGTTTGGGGAAGAATAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACTCCCAACTAAAAAAACAAAAACTCTAAC-3'
Morn3	mm9	LaMa34	10F	5'-aggaagagagGGTTAATAAGATTGGTGTTTATGGGATAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAAATAATAAAACACCTTCCTACTTTCAAA-3'
		LaMa35	10F	5'-aggaagagagTGAAAGTAGGAAGGTGTTTTATTATTTGTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctATAATCTTTAAATCCCCACTATTAATCTCC-3'

Gene symbol	Genome	ID		Primer sequence
Naip1	mm9	LaMa21	10F	5'-aggaagagagAAATTTTGAATATATTTTTGGGTGATAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACCTAACTCCATTTCCTTCAAAACC-3'
		LaMa22	10F	5'-aggaagagagTTTTTTATTTTAGGATAATAGGAGAGAA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAACCAATAAAAATTTCCAAAAACTC-3'
		LaMa23	10F	5'-aggaagagagGTGTGAAGTTTGTGGTAAGATATGATGTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACTCCCTCCAAAAAATTACCATTACTAAC-3'
		LaMa24	10F	5'-aggaagagagGAAGAGAATAAATAAGGGTTAGGATTGTAA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCCCACTCTTATACAAAAAAACTTCACAC-3'
Obfc1	mm9	SQ22	10F	5'-aggaagagagAGGGGATTATTATTAGTTGTGAAGAAATAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACAAAAACAACCAACCAAAATTACC-3'
		SQ23	10F	5'-aggaagagagTGTTAGGGGATTATTATTAGTTGTGAAGAA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAACAAAAACAACCAACCAAAATTACC-3'
Parp2	mm9	SQ58	10F	5'-aggaagagagTATAGTAGTGTAAAAAGGTGAGTTTGGAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAACCACACCAAAAAAAACTAAAAAAAACA-3'
Pbx4	mm9	LaMa02	10F	5'-aggaagagagGTTTTGGGAATAGGAGATATTGGAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctATACCCTCTATATCCACCCAACTCC-3'
		LaMa03	10F	5'-aggaagagagGAGGATTAGAAGAGAATTTTTGTTTTTTAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctATCCATCACTCCAAACTCAACTACAC-3'
Phf20	mm9	LaMa30	10F	5'-aggaagagagTGTGTTTATGAAGTGGTTTGGGTAGAT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAACCACAAACACAAAACTAAACCTCC-3'
Pisd-ps2	mm9	SQ08_C57BL/6	10F	5'-aggaagagagGATGTTGTTAGTTTTAGGTTTTTGTTTTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAACCATAATAAAATCTCACCTAAATAAAC-3'
		SQ08_BALB/c	10F	5'-aggaagagagGATGTTGTTAGTTTTATGTTTTTGTTTTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAACCATAATAAAATCTCACCTAAATAAAC-3'
		SQ09_C57BL/6	10F	5'-aggaagagagGGGGTTTTATTTGAGTGAGTATGAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCCCCAAATTCCTATCTCTACTTTTCC-3'
		SQ09_BALB/c	10F	5'-aggaagagagGGGGTTTTATTTGAGTGAGTATGAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCCCCATATTCCTATCTCTACTTTTCC-3'
Prelid2	mm9	SQ21	10F	5'-aggaagagagGGGGATGTGGGTGTGATTTATAGTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACCCACACCCTCTACCAAAAAAAC-3'
Ptpn14	mm9	SQ01	10F	5'-aggaagagagTTGGAAGTTAGTGGTGGGGATTTAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAACACAAAACAAACTTTAAACTCTAC-3'
		SQ02_C57BL/6	10F	5'-aggaagagagAGGTATAGGATAAGTTTTGGGTTTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAATTCTTAACCCTTTCTCCTCCCC-3'
		SQ02_BALB/c	10F	5'-aggaagagagAGGTATAGAATAAGTTTTGGGTTTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAATTCTTAACCCTTTCTCCTCCCC-3'
Pxmp4	mm9	SQ46	10F	5'-aggaagagagGGAGAGTTGTGTAGTTAGAGTTTAGGGTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTAAAAACAAAACAAAAAAAACCCCAATT-3'
		SQ47	10F	5'-aggaagagagTGTTAGTTTTTGGTTAGTTAGGATTTTTGA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAAAACAAAAAAAACCCCAATTTAAAAC-3'
Rpn1	mm9	SQ48_C57BL/6	10F	5'-aggaagagagAAAAATTTGGTATTGTAATTAAAAGATTAT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTAAACAAAAATACCTATCCACACAC-3'
		SQ48_BALB/c	10F	5'-aggaagagagAAAAATTTGGTATTGTAATTAAAAGATTAT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTAAACAAAAATACCTATCCACACAC-3'
		SQ49	10F	5'-aggaagagagTTTTGTTTGTTGATGTTAATGGTTTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCTAAAAAACCAAACTTCATTCCTTTAACC-3'
Scml4	mm9	SQ70	10F	5'-aggaagagagTAAATTGTGTTTTTATTTTAAATTTTTTTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCCTCAACAAACAACACCCTAATCAA-3'
		SQ71	10F	5'-aggaagagagATGGAAATTTATTGTTGTAGTTATGGTAAA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAAAACAAACATACAAACCCTCCC-3'

Gene symbol	Genome	ID		Primer sequence
Sertad2	mm9	SQ66	10F	5'-aggaagagagTTTTTGGGTAGTGTATGGAGTTGTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAACCACTTTATTCCATTTTACTATCTTCCT-3'
		SQ67	10F	5'-aggaagagagTTTTTGGGGGAGGTTTTATTTGTAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAACACCTAAACTAATTTATTTCCTTCAT-3'
Sfpq	mm9	LaMa39	10F	5'-aggaagagagTTTTAATGATAATGGAGTAATGGAGTGTGT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAAATCTCTCAACCTTCCTCAAACC-3'
Sh3pxd2a	mm9	SQ59	10F	5'-aggaagagagGGTTAATAGTTTATTTGATTTTTTTTGTTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTTAAAATACTCTCTCTCCTATCTCCCC-3'
		SQ60	10F	5'-aggaagagagATTTTTTTTAAGTTAGTTTTGAAAGTTAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAACACCTACTATATACCAAACACC-3'
Ska1	mm9	SQ39	10F	5'-aggaagagagGGTTGGTTGTGGTTATAGAGGATAATTAAT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACTTCCAAAACCCAAATTCAAACAA-3'
Slc25a28	mm9	SQ74	10F	5'-aggaagagagAAAGGTTTATGATTATTGGAAAATAAGAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAACATTTCACCAAATAACATTTCACTTAAA-3'
Slc38a1	mm9	SQ45	10F	5'-aggaagagagGTTAAGTTTGGTTTGGGAGTTGGAGT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctATCCCCTAACAACTAACCCCCTACA-3'
Spaca4	mm9	LaMa33	10F	5'-aggaagagagTGGTATTTAGTGTTGTGAAGGTTTGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACCCTTCCCCTCTTTCAACTAAAATT-3'
St6galnac5	mm9	LaMa19	10F	5'-aggaagagagGATTTTGTTTTAAGGTTTTAATTGGT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAATTTAACTAAATCCACAAATCTTTTT-3'
		LaMa20	10F	5'-aggaagagagAATTTTGTATAAGGGGTGGTATTAAAAGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAAAATTACTTCTACACTCCAAACCAAC-3'
Stt3b	mm9	LaMa45	10F	5'-aggaagagagTGGGGTTATGAGTTTAGAGTTTTTGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctATCCACATCCTATCCCCTAATCAAAC-3'
		LaMa46_C57BL/6	10F	5'-aggaagagagTTGTAGTTATATTGGGAAGATGGAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCTAAACCTATATCCAAACTCATCCTTATAC-3'
		LaMa46_BALB/c	10F	5'-aggaagagagTTGTAGTTATATTGGGAAGATGGAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCTAAACCTATATCCAAACTCATTCTTATAC-3'
Sv2c	mm9	SQ26	10F	5'-aggaagagagTAGTGGAAGTTAGTAGGATGGGGATG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTTCCAACACATTAAACCAAAAAAAA-3'
		SQ27	10F	5'-aggaagagagTTGTGTTTTTGTTTTTTGTAGAGTAA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAATTTATATTCTAATACTTAACCACCTTTC-3'
Tns3	mm9	SQ54	10F	5'-aggaagagagGAAGTGAGTAAATAGGTTAAGTTTTAAAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAATAAACAACCACAAAAAAATCTA-3'
		SQ55	10F	5'-aggaagagagGAGGTTGAAGGAGGGAATTAAAATTATGT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACCAAATCCAATACAAACCAAACCTT-3'
		SQ56	10F	5'-aggaagagagTTTTGGAAAAGGGGATTTTGAAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACCCCATCCCAATATACACTCCATA-3'
Tspan32	mm9	LaMa08	10F	5'-aggaagagagAGAGATATGTGTTTTTTGGGATTTTTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAATCTTATTCCTCTCCAACTCTTATCCC-3'
Tusc1	mm9	SQ10	10F	5'-aggaagagagGGTTGGAGTTTAGGAGTTTAGGATTTTAAA-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCTACAACTACACCTAAAACAACAAAAC-3'
		SQ11_C57BL/6	10F	5'-aggaagagagGTTTTTGTAGTTGTATTTAGAGTAGTAGGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTTCAAAAACTCAAAACTTCAAACTTCC-3'
		SQ11_BALB/c	10F	5'-aggaagagagGTTGTTGTAGTTGTATTTAGAGTAGTAGGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTTCAAAAACTCAAAACTTCAAACTTCC-3'
Usp51	mm9	SQ28	10F	5'-aggaagagagAGTTGGAGTTGGTGTGATAAGTAGT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTTCTTCCCAAACTTAAAAATAAACC-3'

Gene symbol	Genome	ID		Primer sequence
Utp23	mm9	SQ24	10F	5'-aggaagagagGGTGGTTTAAATAGGTATGAAGATTATTAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCACAACTAAATCTCTCCCATAAAATAAC-3'
		SQ25	10F	5'-aggaagagagTTATTTAGAAGGATTTGGTAGGGTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAAACTCCTCTACACAAAAAACTCC-3'
Wasf2	mm9	LaMa40	10F	5'-aggaagagagGAGGTTTAAATTGGGATGAATGGTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTAAAACTCTAAACCCCCTCCCAATC-3'
		LaMa41	10F	5'-aggaagagagTGGGTTTGAATGTTTTTTGTTTGTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctAACCAACCAATCCTTCTTCCTATACC-3'
Ykt6	mm9	SQ52	10F	5'-aggaagagagTTTTATTGAGGGTTTGGGGTTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCTCACCAACCACCCCTACAAACTA-3'
		SQ53	10F	5'-aggaagagagGAGAGAGGGAAATTGTTTGAGTTTTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACCCAAAATATTCCCTACTACCCAC-3'
Ypel3	mm9	LaMa43	10F	5'-aggaagagagTTATTAAATATGGGGAAAGAAAGAAGTTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCCACTACCCTCAACCTAAAAAATCC-3'
		LaMa44	10F	5'-aggaagagagATTGGGTTTTTTGTGGTAGATTTTTAAAAG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAACCAAAAACAAAAACCAAAATTATCA-3'
Zfp459	mm9	LaMa09	10F	5'-aggaagagagGTTTTTTATTTTGTTTTGTAATTTTATTTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACCTAAACTCCAAAAAATTCTTCTC-3'
		LaMa10	10F	5'-aggaagagagTTTTTTTGGAAATTTGTAGTAGAGTGTGG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctACACATAACCTCAAAAAACATTCTTCC-3'
Zfp760	mm9	SQ43	10F	5'-aggaagagagAGGATGGTAGTGTTTATGGTGAGGTT-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctCAACCCACTAATTCAAATCCCTTATC-3'
		SQ44	10F	5'-aggaagagagTGGTTAAGGATGGTAGTGTTTATGGTG-3'
			T7R	5'-cagtaatacgactcactatagggagaaggctTAACCCAACCCACTAATTCAAATCC-3'

3.10.4.2 **iPLEX primers**

The primer sequences for the genome-wide linkage analysis as well as for the fine-mapping approach will be available upon publication.

3.10.5 Primers for hybridization probes

The primers were purchased from Sigma-Aldrich.

Table 3-7 Primers for hybridization probes (BAC library screening)

ID	50 kb bin (*mm9*)	Primer sequence (sense & antisense)
bin1	chr12:16981570-17031570	5'-TACACTGCCTTCCACATCACAC-3'
		5'-ACATTCTGACTCCTCCCTCCCT-3'
bin2	chr12:17031571-17081571	5'-GGACTCTGTGTATCTCTTCTGGG-3'
		5'-AGTTGAAGTCTCAGGGTACTCCA-3'
bin3	chr12:17081572-17131572	5'-ACAACTCAAGATTCCAAGCCAG-3'
		5'-CCCTACCCACATATCCCTCAG-3'
bin4	chr12:17131573-17181573	5'-GTCTGTTCCATAGCCACTCCTG-3'
		5'-GGCAATAAACCCAACAAACTTCC-3'
bin5	chr12:17181574-17231574	5'-GAGGAAGGTGAGAAAGACAAGAAGG-3'
		5'-GCTGATTGAGACTGATTGAGTGAC-3'
bin6	chr12:17231575-17281575	5'-CCCATTGAACTAAACACAGCTCAG-3'
		5'-ATTTCCAGGTAGTCGGTAGCAG-3'
bin7	chr12:17281576-17331576	5'-AGTAAACTATCTGCCCTTCTCCTG-3'
		5'-GAACCTCTTCACTCGGACACTG-3'
bin8	chr12:17331577-17381577	5'-TCAATGTATTAGCCAGAGTCACC-3'
		5'-CAGCATGGGTAAGTAGTTTATCTTGTC-3'
bin9	chr12:17381578-17431578	5'-TGAGGCAACACCAACAAACTG-3'
		5'-AAACAAGAAATCTGAAGACCACTCC-3'
bin10	chr12:17431579-17481579	5'-GAAACAGCCAGACCTAACCAC-3'
		5'-AATGATCCTTTGACACGTAGCCA-3'
bin11	chr12:17481580-17531580	5'-CCCAGTTACATAACCAAGACAGCA-3'
		5'-CACAAGGACCCATCAGGAAGAG-3'
bin12	chr12:17531581-17581581	5'-CCTCCTTCCCTACCTTCTGTC-3'
		5'-GGCATTTCACTCTTCTCTTGGG-3'
bin13	chr12:17581582-17631582	5'-GTGGATGATCAAGGCTCAATGGG-3'
		5'-CTAGCAAAGCGACCAGAAACAG-3'
bin14	chr12:17631583-17681583	5'-ATAGCCGTGTGATAAGCCTCC-3'
		5'-TGTAATCTGTAACTAACCTCTGCCC-3'
bin15	chr12:17681584-17731584	5'-GATGACGCTCTCATTTCCTCCT-3'
		5'-AGGGTGTCTAACAGGATTTGGA-3'
bin15	chr12:17681584-17731584	5'-CCTACAACCCAGCGATAATCCC-3'
		5'-CCTCAGCCCATCCCAATATCTC-3'
bin16	chr12:17731585-17781585	5'-CAGCCCTAGACTCTCATAGTTCC-3'
		5'-TCCCTTAGCACCACCAAACATCC-3'
bin17	chr12:17781586-17831586	5'-CTGACCCACCCTGTAATGATCC-3'
		5'-GTTCTCCAGCTAGCCAAGATCC-3'
bin18	chr12:17831587-17881587	5'-AGCTTCATCTCAGGGTCTCCT-3'
		5'-ATAGCTCCGTGCACAAGTGACC-3'
bin19	chr12:17881588-17931588	5'-CTTCCGACAGCACATCAATCAG-3'
		5'-TTCCACCATTAACCTCAGAAGCC-3'
bin20	chr12:17931589-17981589	5'-AAGTCCTGAGTGTCCTGTCCT-3'
		5'-CCACGACTGAGCTGTATCTGAG-3'
bin21	chr12:17981590-18031590	5'-GCTTATAACGTCCTTCACATTCCC-3'
		5'-AGCCGAATGGAAATTCAAACCC-3'
bin22	chr12:18031591-18081591	5'-CTACCCTGCTCTATCCTAGTCCTG-3'
		5'-GATGCCCTCTACAATTCATTCCC-3'

ID	50 kb bin (*mm9*)	Primer sequence (sense & antisense)
bin23/20/44	chr12:18081592-18131592	5'-TGTGTTCTTATGTTGTGTCCCTG-3'
	chr12:17931589-17981589	5'-GCCCATTATCCTCAGTGTTTGTC-3'
	chr12:19131613-19181613	
bin24	chr12:18131593-18181593	5'-CAGCCTGTACCATAACACAATACC-3'
		5'-CTGCCAACAATACAGCCTAGAG-3'
bin28	chr12:18331597-18381597	5'-GTAGCCCGATAAGACTCCCAC-3'
		5'-TGAACAGAAGCCCACAATACCA-3'
bin29	chr12:18381598-18431598	5'-CCAAACCCTTCCCTACTTCTGTC-3'
		5'-TGTTCTGCCTCTAGTTATTCCCA-3'
bin30	chr12:18431599-18481599	5'-ACAGCAGCAAGAATATCCTAAACC-3'
		5'-ATTGTGCGAATATGACAGTCCCA-3'
bin31	chr12:18481600-18531600	5'-CTGACTGACCATCCAACTTCCA-3'
		5'-TCCTTCTGTGAGTGTTCTATATTAGCC-3'
bin32	chr12:18531601-18581601	5'-CATGTTGGGAAGAGACAGGGA-3'
		5'-CATTGCTCTATGGATTTAGGTGACTG-3'
bin33	chr12:18581602-18631602	5'-CTTTAGTAATCCCTAAGCCGCAC-3'
		5'-ACCTCTGTCTTCAGCCTTACC-3'
bin34	chr12:18631603-18681603	5'-GTTGGTCATCTTGGTTGTGCTC-3'
		5'-GGAATGTGAGGAATGTGGGTG-3'
bin35	chr12:18681604-18731604	5'-TCCATGCCCTAGCTGTATCCT-3'
		5'-TCTCTGCCATCAACAGTTTCAC-3'
bin36	chr12:18731605-18781605	5'-GTCCCTAACCCTCACAATCGT-3'
		5'-GCTCCATCTGAGTACAGCACC-3'
bin39	chr12:18881608-18931608	5'-CCCTAGAGACTTAGACCAGCAG-3'
		5'-CTTCCAGACATCCGTTTCCAG-3'
bin40	chr12:18931609-18981609	5'-AAGTCCTATAGAATGTCCTTAACACCC-3'
		5'-TTTAGGCTCCAACAACTCCCAG-3'
bin42	chr12:19031611-19081611	5'-ACACATTATGTCTGTCTTTCTTCCCTG-3'
		5'-CAGGTCCAGGCTCTTTAACCC-3'
bin43	chr12:19081612-19131612	5'-TCCTGTTCTCTGGTATCACCC-3'
		5'-CCCAGAACACTCTAGGAAAGACAG-3'
bin45/37	chr12:19181614-19231614	5'-AACCCTGAGCTACCTTCTCCA-3'
	chr12:18781606-18831606	5'-GTCCATCTTGCTTCCAAACTTCC-3'
bin46/59/100	chr12:19231615-19281615	5'-CCATGTAGTGACCTTCTTTGCC-3'
	chr12:19881628-19931628	5'-ATTCATAGTGTCTGATGCCCTGTG-3'
	chr12:21931669-21981669	
bin47	chr12:19281616-19331616	5'-GCATCACCACCATTGTTCAGAG-3'
		5'-GGACCCATAGACCCTCTTTCTG-3'
bin48	chr12:19331617-19381617	5'-GGATGGAAGAATCAATATGCTAGTGAG-3'
		5'-TAACCAGCTTAGAAATCACAACGG-3'
bin49	chr12:19381618-19431618	5'-CCAGGACTTCATAGCGATTTGG-3'
		5'-AGACAACAGCCTTGGAGATCAG-3'
bin50	chr12:19431619-19481619	5'-CCCAAAGTCCAGATCTATCATCCTG-3'
		5'-AAATGTCCTTAAACTCCAAACTCCC-3'
bin51	chr12:19481620-19531620	5'-GCTACATCCTTCCTTCCATCCC-3'
		5'-TACAGCCTTCCCTTCTACCCA-3'
bin53	chr12:19581622-19631622	5'-AAGTCCCAGTGACATATCCCTG-3'
		5'-GTCTTGCCTCACCTAAAGTCCC-3'
bin54	chr12:19631623-19681623	5'-CCAAGATAGCAACATGTCCTTATTTCC-3'
		5'-CTTTCCTGAATATTGTAACCGGCT-3'
bin55/44	chr12:19681624-19731624	5'-AAAGATTGAGTGGTCTACTGTCCT-3'
	chr12:19131613-19181613	5'-CTTATGTGCTGTGAGAAGTGGG-3'
bin56	chr12:19731625-19781625	5'-TCCTCACACTCCTATTAACAAGCAG-3'
		5'-GAGACAGAGAAACTGAACCACCA-3'

ID	50 kb bin (*mm9*)	Primer sequence (sense & antisense)
bin57	chr12:19781626-19831626	5'-TTTCTGAGAGTCACCTACAGCC-3'
		5'-TCACAGATACAGCTCAGTCATGG-3'
bin58	chr12:19831627-19881627	5'-GACAATCATTGCCCTCACTCAG-3'
		5'-AACAAGGGTAAACGCATAAAGCAG-3'
bin59	chr12:19881628-19931628	5'-CCTGCTACAGAAAGTGATTCACCT-3'
		5'-GATGTCACCTGTGTAATGCTCTG-3'
bin60	chr12:19931629-19981629	5'-CCACTATTTGCCACAGACAGAG-3'
		5'-CATCGGTTCACTACTGCTACAC-3'
bin61	chr12:19981630-20031630	5'-CAAATCTGCTTGTCCCATTACATTCTG-3'
		5'-CTCAGGTTGGCATATTGTCCC-3'
bin62	chr12:20031631-20081631	5'-TTCACAGTAGTAAGCAGCATCTTCC-3'
		5'-GTTTGTTCCAGTACTCCTCTCCTC-3'
bin63	chr12:20081632-20131632	5'-GGCTCCCGTGGAACTATAGAC-3'
		5'-TCTTGTCCTATTCTGTATCAGTGTCTC-3'
bin64	chr12:20131633-20181633	5'-AAATTCCCAGCTCCCACTATCC-3'
		5'-TCTTTGTTATCTTCTATGTCCCTCTCC-3'
bin65	chr12:20181634-20231634	5'-CTGCTTCTGCGTTTACTTCTTGG-3'
		5'-GCTTGGATCACTCCTTTGTTCTC-3'
bin66	chr12:20231635-20281635	5'-CTTCTTGGGTTGGTCATCTTCGT-3'
		5'-AGAAATGTGAGGGATGTGGGTG-3'
bin67	chr12:20281636-20331636	5'-TGGGATCATTAGAGGAAGCACAG-3'
		5'-TGTCTGGGATGGAGTAGAAGTG-3'
bin69	chr12:20381638-20431638	5'-GCTGCTTCTTCATTCACTTCTTGG-3'
		5'-GGATCACTCCTTTGCTCTCCTG-3'
bin70	chr12:20431639-20481639	5'-AGAACTGATAAATCTGCCTTGCTC-3'
		5'-CCTCTTATCTGATGCCTAATGCCTG-3'
bin72	chr12:20531641-20581641	5'-CTCAGCCCACAGGAAGAAGAC-3'
		5'-CATCAAAGATACTCATCACCAGACC-3'
bin73/80	chr12:20581642-20631642	5'-TCTCCTCGCTGTCAAAGAATGTG-3'
	chr12:20931649-20981649	5'-CATTGATTTATCTCCCTGGCTCC-3'
bin74	chr12:20631643-20681643	5'-GAGGAGGTGATGTGGTGAGAG-3'
		5'-GTGGACAAGCCTCAGTATGGT-3'
bin75	chr12:20681644-20731644	5'-CTCAGAAAGACACAGATGCTATTAGG-3'
		5'-TACTTCTTATACACCCTTGCCCAG-3'
bin76	chr12:20731645-20781645	5'-GTACCTGGTTCTGCTCTCCTG-3'
		5'-CACAGCCAACCTGCAAATTCTC-3'
bin78	chr12:20831647-20881647	5'-GCTCTTTCACTCTACCCTTTGAC-3'
		5'-TACTCCATCCCAGACACCCTC-3'
bin79	chr12:20881648-20931648	5'-GGCTCACACTTCATACCATACAC-3'
		5'-ATAATTTCCTCTTCCCTCCCTC-3'
bin80/125	chr12:20931649-20981649	5'-CCCTTGGCTCTTGTTACTCTCC-3'
	chr12:23181694-23231694	5'-GCTCTTGGAAATAGCCCTTCAG-3'
bin81	chr12:20981650-21031650	5'-TCTTCATAGCTTCATTCATTCCTACCC-3'
		5'-GAAACCAATCCCACCCTAACTG-3'
bin82	chr12:21031651-21081651	5'-GCCGAAGTAAAGAAATGACCCAC-3'
		5'-GGTTAGACTCTCCTAGATTCCAAACTG-3'
bin83	chr12:21081652-21131652	5'-AAAGCTGTCACTCAGAAAGCCT-3'
		5'-TGATCCCAAATAGGTCTCAGCA-3'
bin84	chr12:21131653-21181653	5'-CTCAGCCCATCTTTGTTCCTC-3'
		5'-GCCCAAACCACTTTCATTTCAC-3'
bin85	chr12:21181654-21231654	5'-GTGTTCTGCTAATAAATCCTCGTTCAG-3'
		5'-CAAATGCCACCAAGAAGCCTC-3'

ID	50 kb bin (*mm9*)	Primer sequence (sense & antisense)
bin86	chr12:21231655-21281655	5'-TTATCTCCCGCTCCCAAACAG-3'
		5'-GACTAATTCTGCCAGTAATCCCAC-3'
bin87	chr12:21281656-21331656	5'-TTTCCCTCCCACAGAACTCTC-3'
		5'-ATCTCTGGCTTGGTTGACTCAC-3'
bin88	chr12:21331657-21381657	5'-CAAATGAAACTACATCAGTCGCCAC-3'
		5'-GGGAGCAAACTGACTTATCTGAC-3'
bin89	chr12:21381658-21431658	5'-TCTGCTCAATAAACACCCTTCTG-3'
		5'-AGAGTAGAGTCACTTTCCAACCA-3'
bin90	chr12:21431659-21481659	5'-ACTCCTTCCACACCTAAATGCC-3'
		5'-GCTCTAATCCCACTTCCTGCTC-3'
bin91	chr12:21481660-21531660	5'-AGTAAGTCTCCCACCTTCCAG-3'
		5'-GTTCTGCCTTTGAGCATTTCCA-3'
bin92/165	chr12:21531661-21581661	5'-CTTTGATGTTTCTTGGTTTGGTCTG-3'
	chr12:25181734-25231734	5'-ACTCACTGGTCATTCTTCACCC-3'
bin93/25	chr12:21581662-21631662	5'-GCCTTCCAATCTGATGACTTCTG-3'
	chr12:18181594-18231594	5'-GCTACATCTCCCACCTCATCC-3'
bin94/77	chr12:21631663-21681663	5'-GTGTTAACTCTTGTCAGCTTCGG-3'
	chr12:20781646-20831646	5'-GAGGCTTGTGTTCCAACAGAC-3'
bin95/71/100	chr12:21681664-21731664	5'-GCAACGAATCCATGTGTCACTG-3'
	chr12:20481640-20531640	5'-AGAACTGAGTGAGGAAATACCTGAG-3'
	chr12:21931669-21981669	
bin97	chr12:21781666-21831666	5'-ACACATACATCCATCCTCTTATTCCC-3'
		5'-TGGACATCTGCCCTACATCCT-3'
bin98/69	chr12:21831667-21881667	5'-TATTCACAGTTGCCTGCCTGAG-3'
	chr12:20381638-20431638	5'-TCTTCCCAGACATTGCCTTCC-3'
bin99/68	chr12:21881668-21931668	5'-CACCAGAGAAGCCCATACTTGAG-3'
	chr12:20331637-20381637	5'-TAAGCCACTTTATTTGCCAGGAG-3'
bin101	chr12:21981670-22031670	5'-ATGGTCGGTGAAATCCTTGTG-3'
		5'-TTCCTGTTCTCTGGTATCATCCC-3'
bin102/124/159	chr12:22031671-22081671	5'-ATGACACATGCACAATCTCTGG-3'
	chr12:23131693-23181693	5'-TCAGATTTGTTTGGGTACTTGGG-3'
	chr12:24881728-24931728	
bin103/117	chr12:22081672-22131672	5'-CATCTGCCAGTATTACTGACACCT-3'
	chr12:22781686-22831686	5'-GTATCCGTTCTCCTTCACCCT-3'
bin104/116	chr12:22131673-22181673	5'-TATCTTCTCTATCCCTCTCCTACTCC-3'
	chr12:22731685-22781685	5'-ATTCCCAACTCCTGCTTTCCC-3'
bin105/115	chr12:22181674-22231674	5'-CAATGCCACCATTAGCCTCAG-3'
	chr12:22681684-22731684	5'-CCTCTGGTTCTCTTGGGTGTC-3'
bin106/111/114	chr12:22231675-22281675	5'-CTGCCCATTATCCTTAGTGTTTGTC-3'
	chr12:22481680-22531680	5'-TGTGTTCTTATGTTGTGTCCCTG-3'
	chr12:22631683-22681683	
bin107/143/136	chr12:22281676-22331676	5'-AAGGTAGGGACACAGATTTGGA-3'
	chr12:24081712-24131712	5'-TCTGGAACGTCTTGAATCACTTAGG-3'
	chr12:23731705-23781705	
bin108/137/143	chr12:22331677-22381677	5'-ACAGTAACTTTCCAGCCTCCTC-3'
	chr12:23781706-23831706	5'-TCTTTCCCTTCCTCTCATCCC-3'
	chr12:24081712-24131712	
bin110/104/115	chr12:22431679-22481679	5'-CTCTCACTTCAGATTCTAGCAGGG-3'
	chr12:22131673-22181673	5'-CCAGGACTTCATAGGGATTTGGT-3'
	chr12:22681684-22731684	
bin113/107/136/142	chr12:22581682-22631682	5'-AAGGCAGACATCAAGGTAGGG-3'
	chr12:22281676-22331676	5'-AGCAAAGCACTCGTATTCCTG-3'
	chr12:23731705-23781705	
	chr12:24031711-24081711	

ID	50 kb bin (*mm9*)	Primer sequence (sense & antisense)
bin118/101/149	chr12:22831687-22881687	5'-GACTCAAGCATCTATACTTCAAGAACC-3'
	chr12:21981670-22031670	5'-GACATTCCACTACACCCAGACAG-3'
	chr12:24381718-24431718	
bin119/67/70	chr12:22881688-22931688	5'-CCCATCTCACTGCTTCCTCTG-3'
	chr12:20281636-20331636	5'-CTTCTAAGTCCCTCCCTTGGT-3'
	chr12:20431639-20481639	
bin120/148	chr12:22931689-22981689	5'-TCCTTGCTGCCAAACAATGTG-3'
	chr12:24331717-24381717	5'-TTGATTCATCTCCCTGGCTCC-3'
bin121/44/123	chr12:22981690-23031690	5'-AAATCCTTCCCTGCCTCTGTC-3'
	chr12:19131613-19181613	5'-TGTTCTGCCTCTAGTTATTCCCA-3'
	chr12:23081692-23131692	
bin123/160	chr12:23081692-23131692	5'-AAGTCCCAGTGACATTTCCCTG-3'
	chr12:24931729-24981729	5'-GTCTTGCTTCACCTAAAGTCCC-3'
bin126	chr12:23231695-23281695	5'-CTGTAGAGGGATGTGGAGAATGG-3'
		5'-GAGCAAAGGAAATGTAATCATCGGT-3'
bin128/52/155	chr12:23331697-23381697	5'-AATTCCCAGCTCCCACTTTCC-3'
	chr12:19531621-19581621	5'-TCTTCTCTGTCCCTCTCCTACTC-3'
	chr12:24681724-24731724	
bin129/153	chr12:23381698-23431698	5'-GTGCTCTGCTCTGTTTCTTTCTG-3'
	chr12:24581722-24631722	5'-GACCACAGCATTCATTCACCT-3'
bin138/134/145	chr12:23831707-23881707	5'-GAAGCTGTTTCCCTTTCCTACC-3'
	chr12:23631703-23681703	5'-CATTGTGTCTGTCTTTCTTCCCTG-3'
	chr12:24181714-24231714	
bin139/133/146	chr12:23881708-23931708	5'-TCTTGTTCTTTGCCCTCACCTC-3'
	chr12:23581702-23631702	5'-GGCTGTTTGTTGAATTTGCTCC-3'
	chr12:24231715-24281715	
bin142/106/112	chr12:24031711-24081711	5'-GGAAGGAAGGACACTCAGGAC-3'
	chr12:22231675-22281675	5'-CATCTCTATGCCTCCTACTCGCC-3'
	chr12:22531681-22581681	
bin144/41/108/137	chr12:24131713-24181713	5'-AAAGGAAGGAGGAAAGGTTTGGG-3'
	chr12:18981610-19031610	5'-GGATACAGGGTTTGTCTTCGTAGG-3'
	chr12:22331677-22381677	
	chr12:23781706-23831706	
bin145	chr12:24181714-24231714	5'-CTGCTTCCTGGTCTTTCACTG-3'
		5'-GTGGTATGTTCTAAACATCTGCCTC-3'
bin146	chr12:24231715-24281715	5'-CTCTCCCTATGCTCTTTCCCA-3'
		5'-ATACCCAAGGATTGACCCTCTC-3'
bin147/38/132/140	chr12:24281716-24331716	5'-TCCTTCTCACTGTCCCTCTCC-3'
	chr12:18831607-18881607	5'-CCTGTCTGTTCCCATTTCCCTC-3'
	chr12:23531701-23581701	
	chr12:23931709-23981709	
bin148/131/151	chr12:24331717-24381717	5'-GTAAGAAATGCCTCCATACCTGATCTC-3'
	chr12:23481700-23531700	5'-TGTGTCCCTGTTCCAAGTGTG-3'
	chr12:24481720-24531720	
bin149/130/152	chr12:24381718-24431718	5'-AGGTAATAAGACTAAGCAGGAGAGG-3'
	chr12:23431699-23481699	5'-CAACACATCCCTAAGATTACAATCCC-3'
	chr12:24531722-24581722	
bin150	chr12:24431719-24481719	5'-GTAGATGCCCTCTAAATTCATTCCC-3'
		5'-AGATATTCTACCCTGCTCTATCCTAGTC-3'
bin151/131/148	chr12:24481720-24531720	5'-CACGCTAATAACTCCCACCTATCAC-3'
	chr12:23481700-23531700	5'-GCCACTTTGCCTTTCTGAACTG-3'
	chr12:24331717-24381717	
bin152	chr12:24531722-24581722	5'-TGTCTGTCAAATCTAATCCCTGCT-3'
		5'-ATCCACTCACCTCACTGAACAC-3'

ID	50 kb bin (*mm9*)	Primer sequence (sense & antisense)
bin153	chr12:24581722-24631722	5'-ACAAACTCTGAAACAATGGTGGTG-3'
		5'-ACACAAATGCACAAATGGTACACAG-3'
bin154	chr12:24631723-24681723	5'-CTGATCTCACTGGCTGTCTCTG-3'
		5'-CCTGGAACATGAACGATGGAG-3'
bin155	chr12:24681724-24731724	5'-CATCTTCATCACAGAGGACACCA-3'
		5'-CCTTTAGTTTGTTGAATCTGCCGA-3'
bin156/127	chr12:24731725-24781725	5'-AATCATTGTCCTCATTCGGTCTACAG-3'
		5'-AAGAGTGACAAGGCATGTCCC-3'
bin157	chr12:24781726-24831726	5'-AGACTATTCACAGATGCCTGCC-3'
		5'-TCTTCCCAGACATTGCCTTCC-3'
bin158/124	chr12:24831727-24881727	5'-CAAGAGTTACGGAACCAGAGGAG-3'
		5'-CCAAGAGAAAGCACATACACCC-3'
bin159	chr12:24881728-24931728	5'-CTTGGATCATGCCTTTGCTCTC-3'
		5'-GCTTCTGCGTTCACTTCTTGG-3'
bin160/122	chr12:24931729-24981729	5'-CCACACCTTGAACATGAACACTG-3'
		5'-ACTCCATCCATACTAGGCCCA-3'
bin161/122	chr12:24981730-25031730	5'-CACAAAGCATCCACTCACCTC-3'
		5'-CTATGTCAACGTTCTTCCTGCC-3'
bin162	chr12:25031731-25081731	5'-GGACTTCACCTTAGCATTTCTCCC-3'
		5'-CACAAGCTCTATTGCCCTCAG-3'
bin164	chr12:25131733-25181733	5'-GGCAGGTTCACAAGGTTACAG-3'
		5'-ACATCTACATTTCTCACATCCACGA-3'
bin165	chr12:25181734-25231734	5'-GAGTCCATGAGTATCGGTGAGAG-3'
		5'-TGTTTCCTTGCTGGTATCATCCC-3'
bin166	chr12:25231735-25281735	5'-GCCATCTACATCTTTAAACCCAGAG-3'
		5'-TCTAAGACACCTTCCCAAGCC-3'
bin167	chr12:25281736-25331736	5'-GCTCCTCACTCTGACCTATGTC-3'
		5'-AGAAATGAAGAACACCGTATACCCT-3'
bin168	chr12:25331737-25381737	5'-AGAACCTCCTCTCCACCTCTG-3'
		5'-TCAGCCTACAGTTATGTAACACCA-3'
bin169	chr12:25381738-25431738	5'-CCTGCCACCATCACATTTCAC-3'
		5'-CCATAAAGCTCATTTGTCCGCC-3'
bin170	chr12:25431739-25481739	5'-AGAATCACGCTGGTATGAGATGAC-3'
		5'-CATTATAAGTCCGTATGTTCTCCTCTCC-3'
bin171	chr12:25481740-25531740	5'-TAAGTGCATACAACCCTAGCCT-3'
		5'-ACCACATCTTTCTAAGCGCCT-3'
bin172	chr12:25531741-25581741	5'-CGAGCAGGTAAGAGTGATAGGT-3'
		5'-ATGGGTTCTGAAGATGTGAGGG-3'
bin173	chr12:25581742-25631742	5'-CCGTCTCCGTTCCAGATACTC-3'
		5'-TCGTTCTCTGCTCATCCACAG-3'
bin174	chr12:25631743-25681743	5'-CCCTTCAAAGATCCTTGTACAACCA-3'
		5'-CCCTGTCATTCCTACTGCCTC-3'
bin175	chr12:25681744-25731744	5'-ACACCACTAATAAAGGCCACCA-3'
		5'-CTGTCTTCGTCACTACCACCC-3'
bin176	chr12:25731745-25781745	5'-GATTTCAGACAACCCAGGACAG-3'
		5'-GTTCACCACAGCTAAGTCACAC-3'
bin177	chr12:25781746-25831746	5'-CCCTGCTATGTGTTCCTTCCTC-3'
		5'-CCTACCCTCCCTCTTATCTCCTC-3'
bin178	chr12:25831747-25881747	5'-TCTTTGACCAGATCTACTTTCTCACC-3'
		5'-TTGACCTGTGGCTGATAAGAACC-3'
bin179	chr12:25881748-25931748	5'-GCCCATAAAGCTTACTGTTCCC-3'
		5'-ATTCCTGATGCCTCTAACTCCA-3'
bin180	chr12:25931749-25981749	5'-TAGAGGTAATCAGGTAGAGGGAGG-3'
		5'-CAAGGTAGCAAGGGATAGAGATGTG-3'
bin181	chr12:25981750-26031750	5'-CAGTACCGAGACAATCTTAGAACCC-3'
		5'-GATACTCAACTCCTCCATCAGCC-3'

Table 3-8 Primers for anchor hybridization probe (BAC library screening)

ID	Genomic region (*C. briggsea*)	Primer sequence (sense & antisense)
anchor	chrIII:13541331-13541734	5'-CAGATAGAAGTAGATGAACACCGCC-3'
		5'-GTCATTCATCATGTGGCTTCGTC-3'

3.11 Antibodies for immunoprecipitation experiments

Table 3-9 Antibodies for immunoprecipitation experiments

Antigene	Assay	Origin	Company	Order#/Lot#
5hmC	hmeDIP	rabbit	Active Motif	ab39791/15413007
H3K4me1	ChIP	rabbit	Abcam	ab8895/937164 & 962543
H3K4me3	ChIP	rabbit	Millipore	04-745/DAM1494280
H3K27ac	ChIP	rabbit	Abcam	ab4729/961080 & GR45851-1
IgG	ChIP	rabbit	Millipore	12-370/DAM1713234 & NG1893918

3.12 Databases and software

The following databases and software tools were used to design experiments and process data, respectively. If no version is stated, the detailed information can be found in the corresponding methods sections.

AxioVision Rel. 4.8	Zeiss, Oberkochen, Germany
BioEdit version 7.0.9.0	http://www.mbio.ncsu.edu/BioEdit/bioedit.html
Bowtie	http://bowtie-bio.sourceforge.net/index.shtml
Bowtie2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
CorelDraw X4	Corel, Ottawa, Ontario, Canada
EpiTYPER 1.2	Sequenom, San Diego, CA, USA
Fastqc	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
GeneRunner version 3.05	http://www.generunner.com
GEO Database	http://www.ncbi.nlm.nih.gov/geo/
Homer	http://homer.salk.edu/homer/
ImageLab v4.0	Bio-Rad, Munich, Germany
MethPrimer	http://www.urogene.org/methprimer
Microsoft Excel 2003/2007	Microsoft Deutschland GmbH
Mouse Genome Informatics (MGI)	http://www.informatics.jax.org/
Mouse Genomes Project	http://www.sanger.ac.uk/resources/mouse/genomes/
ORF Finder	http://www.ncbi.nlm.nih.gov/gorf/gorf.html
PerlPrimer version 1.1.14	http://perlprimer.sourceforge.net/
PubMed	http://www.ncbi.nlm.nih.gov/entrez
R	http://www.r-project.org/

R Studio v0.97.551	http://www.rstudio.com/
RSEM	http://deweylab.biostat.wisc.edu/rsem/
Trinity	http://trinityrnaseq.sourceforge.net/
Typer 4.0	Sequenom, San Diego, CA, USA
USCS Genome Browser	http://www.genome.ucsc.edu

4 Methods

Unless otherwise mentioned, all methods were based on protocols described in “Current protocols of Molecular Biology”[264], and in the “Molecular Cloning: A Laboratory Manual”[265].

4.1 General cell and bacteria culture methods

4.1.1 Cell line cultures

4.1.1.1 Murine embryonic stem cells

Culture conditions and passaging

The murine embryonic stem (ES) cell line JM8 was cultured in ES cell (ESC) medium at 37°C with FCS (fetal calf serum, 10%), CO_2 (7%) and high humidity (95%) in general without mouse embryonic fibroblasts (MEFs). Instead, new ESC culture dishes are coated with 0.1% gelatine solution prior to use (min. 5 min at room temperature (RT)[266].

The cells were seeded with ~$4x10^4$ cells / cm^2 at a concentration of ~$3x10^5$ cells / ml, supplied with fresh media every day and passaged every second day (in a 1:3-1:4 ratio).

For passaging, the adherent cells were first washed twice with PBS (volume similar to that of the culture medium) and then trypsinized for 7 min at 37 °C, 7% CO_2. Trypsinization was stopped by adding 4 Vol of ESC medium. The cell suspension was either splitted on 3-4 dishes of the same size or transferred completely to a dish with the next bigger diameter (12-well < 6-well < 6 cm < 10 cm). The suspension volume was adjusted with fresh ESC medium. Approximate cell numbers for each size of culture dish as well as volumes of medium and trypsin are listed in Table 4-1.

Table 4-1 ESC culture: Volumes and cell numbers for various culture dishes

Culture dish	Area	Cell number	Volume	
			ESC medium (10% FBS)	Trypsin (1x)
12-well	3.5 cm^2	~0.45 Mio	600 µl	80 µl
6-well	9.6 cm^2	~1.2 Mio.	1.3 ml	150 µl
6 cm	28.3 cm^2	~3.5 Mio.	3.7 ml	400 µl
10 cm	78.5 cm^2	~10 Mio.	10 ml	1 ml

Undifferentiated murine ES cells form compact colonies with a refractive, defined edge and have a high nucleus/cytoplasm ratio. The colonies are bigger in size when cultured without MEFs and less convex.

Culture on MEFs and MEF removal

In case of BALB/c ESCs, cells start to differentiate when cultured without MEFs longer than 3 passages. Cytostatic MEFs were plated (~10^5 cells/cm^2) on the designated dish shortly before a frozen ESC aliquot is thawed or before ESCs are trypsinized for passaging. MEFs sediment faster than ESCs and will form a single-cell layer on, which the ESCs will form their colonies.

In order to differentiate the ESCs, the MEFs have to be removed from the culture. Due to their mitotic inactivation, MEFs are subsequently diluted in the co-culture. To accelerate the removal, the co-culture suspension was transferred to an uncoated dish of the same size after washing and trypsinizing. After the plate was incubated at normal culture conditions for up to two hours the cells still in suspension or loosely attached to the dish surface (mainly ESCs), respectively, were transferred to a gelatin-coated dish of the same side leaving the already adherent MEFs behind.

Unspecific differentiation of ESCs into embryoid bodies

Adherent, MEF-free cells were washed and trypsinized as for passaging. Undissociated cell aggregates were removed by sedimentation by gravity for 10-15 min at 4°C to obtain a single cell suspension. The supernatant was transferred to a new vial and cells were counted with CASY and/or Neubauer hemocytometer (see section 4.1.1.2). Cells were seeded for embryoid body (EB) formation at a concentration of ~2.5x10^5 cells / ml in EB medium according to the following table.

Transfection with Lipofectamine® 2000

One day prior transfection, adherent MEF-free ESCs are washed and trypsinized and seeded into gelatin-coated 12-well dishes (1x10^5 cells / well in 600 µl ESC medium). After o/n culture at normal conditions the medium is changed (710 µl) 1 h before transfection. In the meantime, 3.8 µl of Lipofectamine® 2000 are mixed with 95 µl transfection medium by pipetting and incubated for at least 5 min at RT. 500 ng of BAC DNA is also mixed in separate 95 µl transfection medium by pipetting. Both solutions are mixed by pipetting and incubated at RT for 20 min. The mix is added to the cells dropwise while the plate is shaked. The cells are incubated under normal culture conditions and supplemented with fresh ESC medium the next day (600 µl). Selection with Geneticin (G418, 1:250 diluted in ESC medium) was started at day 2 after transfection to select for transfected ESC clones. As negative control, one well was transfected with 1xTE without any BAC DNA. ESC medium containing G418 (1:250) was changed every 2^{nd} to 3^{rd} day. After 10 days of G418 selection (all cells in negative control were dead by then) the cells were washed and trypsinized (incubation at 37 °C for 9 min to achieve better dissociation of colonies) and plated into new, gelatin-coated 12-well dishes with ESC medium without G418. The transfected ESC clones are passaged three more times without G418 selection prior to EB formation (see above).

Freezing and thawing ESCs

In general, when cells were about to be splitted (1:3 to 1:4) the supernatant was removed and the adherent cells were washed twice with PBS (volume similar to that of the culture medium). Cells were trypsinized (Table 4-1) for 7 min at 37 °C, 7% CO_2. Trypsinization was stopped by adding 4 Vol of ESC medium. For each plate, the cells were split in two aliquots and centrifuged for 5 min, 200xg, RT. The supernatant was removed and the cells were resuspended in 200-1000 µl ice-cold ES Freeze medium (ESC medium containing 10% DMSO) depending on plate size. The cell suspension was transferred to a 2 ml cryo vial (Corning) and after o/n freezing in a Cryo 1 °C Freezing Container (Nalgene) at -80 °C stored at -196 °C.

For thawing, the cryo vials are removed from -196 °C and placed in an incubator at 37 °C. Immediately after complete thawing, ESC medium (10x the Volume of the freezing medium) is added to the cell suspension to dilute DMSO. If the Volume exceeds the required culture Volume, the cells are harvested by centrifugation (5 min, 200xg, RT) and resuspended in the appropriate Volume of ESC medium.

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

ESC medium	per 500 ml		DMEM Knock-Out (Gibco/Life Tech. #10829-018)
	0.1 mM	1 ml	β-Mercaptoethanol (Gibco/Life Tech. # 31350-010, 50 mM)
	2 mM	5 ml	L-Glutamine (Biochrom KG, #K0302, 200 mM)
	50 U/ml	2.5 ml	Penicillin-Streptomycin (Gibco/Life Tech. #15140-122, 10000 U/ml)
	1800 U/ml	90 µl	LIF (ESGRO®) (Chemicon/Millipore #ESG1107, 10^7 U/ml)
	10%		55 ml ESC culture-suitable fetal bovine serum (FBS) (Gibco/Life Tech., Batch 41F5283K)
		→ Store at 4 °C	
EB medium	similar to ESC medium, but without LIF supplementation		
Gelatin Solution	0.1%	1 g	Gelatin (Sigma-Aldrich, #G1890) → Add 1 l ddH_2O, autoclaved
Trypsin Solution	1x	4 ml	Trypsin/EDTA/NaOH (PAN, #P10-024100, 10x)
		→ Ad 40 ml PBS (1x), store at 4 °C	

PBS	1x	Phosphate-buffered saline (Sigma-Aldrich, #D8537)
Transfection reagent		Lipofectamine® 2000 (Invitrogen/Life Tech., #11668-019) → Store at 4 °C
Transfection medium	1x	OPTI-MEM® I (Gibco/Life Tech., # 31985-062) → Store at 4 °C
Selective Antibiotic (G418)	50 mg/ml	Geneticin® (G418 Sulfate) (Gibco/Life Tech., # 10131-019) → Store at 4 °C
Culture Dishes		
• ESC culture	12-well	Nunc, #150628
	6-well	Nunc, #140675
	6 cm	Nunc, #150288
	10 cm	Nunc, #172958
• EB Formation	6-well	Greiner Bio-one, #657970
	6 cm	Greiner Bio-one, #628102
	9.4 cm	Greiner Bio-one, #633102

4.1.1.2 Assessing cell number and vitality

Neubauer Hemocytometer

The total number of cells and their vitality can be determined microscopically using Trypan blue exclusion. The cell suspension was suitably diluted with Trypan blue solution and counted in a Neubauer hemocytometer. Dead cells appear dark blue and are clearly distinguishable from living cells in the microscope. The concentration of viable cells was then calculated using the following equation:

$$\text{Number of viable cells/ml} = N \times D \times 10^4$$

N: average of unstained cells per corner square (1 mM containing 16 sub-squares)
D: dilution factor

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

Trypan blue solution	0.2%	200 mg Trypan blue (Sigma-Aldrich, #T6146) → Ad 100 ml 0.9% NaCl solution → Store at 4 °C
NaCl solution	0.9%	900 mg Sodium chloride (AppliChem, #A3597) → Ad 100 ml ddH_2O, autoclave

CASY® Cell Counter (Innovatis)

Cell suspension was diluted 1:200 in CasyTON solution. Cells are soaked through a glass capillary three times and counted and sized. According to the size distribution settings listed below, cells per ml (total, viable) and their average cell diameter are determined:

Murine ESC: CL 7.65 µm NL 6.3 µm CR 20 µm NR 20 µm
(Dilution: $2x10^{-2}$; 50 µl cell suspension in 10 ml CasyTON)

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

CasyTON	CasyTON (Roche, #05651808001)

4.1.2 Bacterial culture

4.1.2.1 Bacterial growth medium

E .coli strains were streaked out on solid LB_{Amp} agar and grown for 12-36 h. DH10B, DH5α, PIR1, and TOP10 strains were grown at 37 °C, SW106 strains at 30 °C. Single colonies were picked and inoculated in LB medium containing the appropriate selection antibiotics and grown overnight at 37 °C or 30 °C, respectively, on a shaker at 200 rpm.

Table 4-2 Antibiotics for selective bacterial culture

Antibiotic	*Solvent*	*Stock concentration*	*Dilution in media*	*Final concentration*
Ampicillin (Amp)	ddH_2O	100 mg/ml	1:1000	100 µg/ml
Chloramphenicol (CM)	Ethanol (100%)	20 mg/ml	1:2000	12.5 µg/ml
Kanamycin (Kan)	ddH_2O	50 mg/ml	1:1000	50 µg/ml
Spectinomycin (Sp)	ddH2O	50 mg/ml	1:1000	50 µg/ml

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

Lysogeny broth (LB) medium	170 mM	10 g	Sodium chloride (AppliChem, #A3597)
	1%	10 g	Bacto™ Tryptone (BD, #211705)
	0.5%	5 g	Bacto™ Yeast extract (BD, #212720)
		→ Adjust with NaOH to pH 7.5	
		→ Ad 1 l ddH_2O, autoclave	
LB agar plates	1.5%	15 g	Bacto™ Agar (BD, #214010)
	170 mM	10 g	Sodium chloride (AppliChem, #A3597)
	1%	10 g	Bacto™ Tryptone (BD, #211705)
	0.5%	5 g	Bacto™ Yeast extract (BD, #212720)
		→ Adjust with 1 M NaOH to pH 7.5	
		→ Ad 1 l ddH_2O, autoclave, and cool to 50 °C	
		→ Add required antibiotics (see *Table 4-2*)	
		→ Pour into 10 cm Petri dishes, store inverted at 4 °C	

NaOH	1 M	Sodium hydroxide solution (Merck/Millipore, #109137)
Petri dishes	9.4 cm	Greiner Bio-one, #633102

4.1.2.2 Transformation of *E. coli*

Heat shock transformation

Subcloning of PCR products was performed by transformation of chemically competent *E. coli* cells part of either the StrataClone™ PCR Cloning Kit (Agilent Technologies) or the TOPO® TA Cloning® Kit (Life Technologies), respectively, according to the manufacturer's instructions (see section 4.3.3.1).

Electroporation

Transformation of larger DNA constructs and subsequent transformation was done by electroporation rather than heat shock.

To obtain electrocompetent cells, the *E. coli* strain of interest was inoculated from glycerol stock overnight on a shaker at 30°C/37°C (as required, see section 4.1.2.1) in 10 ml LB medium and, where necessary, with the appropriate antibiotics. The next day, the preculture was inoculated in fresh medium with a 1:100 dilution and incubated on a shaker (vigorously) at 30 °C/37 °C until OD_{600} reaches 0.6-0.9. Bacteria suspension was cooled on ice for 15 min while shaking and then harvested (30 min, 2900xg, 4 °C). The pellet was washed twice with 200 ml and 100 ml ice-cold sterile ddH_2O per 250 ml suspension. Each pellet from 250ml suspension was washed with 20 ml ice-cold sterile glycerol/ddH_2O (10%). Each pellet was finally resuspended in 500 µl ice-cold sterile glycerol/ddH_2O (10%), which rendered ~ 1 ml of electrocompetent bacteria. The bacteria were used directly for electroporation (40 µl per reaction) or stored at -80 °C in aliquots.

500 ng plasmid or 2 µl of BAC mini prep (see section 4.3.1.1) were placed on the wall in the gap of a 1 mM electroporation cuvette (Bio-Rad). 40 µl of electrocompetent *E. coli* were pipetted over the DNA into the gap. Electroporation was performed with the GenePulser II (Bio-Rad) at 200Ω, 1.8 kV, 25 µF. The time constant for successful electroporation lies within 4.5-5.5 msec. After 1 h recovery at 30/37 °C on a shaker (250 rpm) the cells are plated onto LB agar plates with suitable selective antibiotics.

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

LB (Lysogeny broth) medium	170 mM	10 g	Sodium chloride (AppliChem, #A3597)
	1%	10 g	Bacto™ Tryptone (BD, #211705)
	0.5%	5 g	Bacto™ Yeast extract (BD, #212720)
		→ Adjust with NaOH to pH 7.5	
		→ Ad 1 l ddH_2O, autoclave	
LB plates	1.5%	15 g	Bacto™ Agar (BD, #214010)
	170 mM	10 g	Sodium chloride (AppliChem, #A3597)
	1%	10 g	Bacto™ Tryptone (BD, #211705)
	0.5%	5 g	Bacto™ Yeast extract (BD, #212720)
		→ Adjust with NaOH to pH 7.5	
		→ Ad 1 l ddH_2O, autoclave, and cool to 50 °C	
		→ Add required antibiotics (see *Table 4-2*)	
		→ Pour into 10 cm Petri dishes, store inverted at 4 °C	
NaOH	1 M		Sodium hydroxide solution (Merck/Millipore, #109137)
Glycerol/ddH_2O	10%	10 ml	Glycerol (Sigma-Aldrich, #G5516)
		→ Ad 100 ml nuclease-free ddH_2O, autoclave	
Petri dishes	9.4 cm		Greiner Bio-one, #633102

4.1.2.3 Glycerol stock

For long-term storage, bacteria were stored in autoclaved LB-Glycerol medium (60%). 250 µl of bacterial suspension was added to 750 µl of LB-Glycerol media, mixed, and stored at -80 °C.

Required reagents and materials:
(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

LB (Lysogeny broth) medium	170 mM	10 g	Sodium chloride (AppliChem, #A3597)
	1%	10 g	Bacto™ Tryptone (BD, #211705)
	0.5%	5 g	Bacto™ Yeast extract (BD, #212720)
		→ Adjust with NaOH to pH 7.5	
		→ Ad 1 l ddH_2O, autoclave	
LB-Glycerol medium	60%	60 ml	Glycerol (Sigma-Aldrich, #G5516)
		40 ml	LB (Lysogeny broth) medium
		→ Autoclave	

4.2 Laboratory animals

All laboratory animals were purchased from Charles River Laboratories (Sulzfeld, Germany) as direct or indirect distributor if not stated otherwise.

4.3 General molecular biological methods

4.3.1 Preparation and analysis of DNA

4.3.1.1 Isolation of circular DNA from E. Coli

Plasmid isolation

Mini preparation

Plasmid preparation for up to 15 µg from single E. coli colonies was done with the NucleoSpin® Plasmid Quick Pure Kit (Macherey-Nagel) following the manufacturer's instructions. Insert-containing plasmid preparations were identified by restricted digest with EcoRI (recognition sites adjacent to cloning site, 2 h at 37 °C) (see Table 4-3).

Cell pellets can be stored at -20 °C until preparation. Plasmid preparations were stored at -20 °C.

Midi preparation

Plasmid preparation for up to 100 µg from single E. coli colonies was done with the Plasmid Midi Kit (Qiagen) following the manufacturer's instructions. Insert-containing plasmid preparations were identified by restricted digest with EcoRI (recognition sites adjacent to cloning site) (see Table 4-3). Cell pellets can be stored at -20 °C until preparation. Plasmid preparations were stored at -20 °C.

Control digest and Gel Electrophoresis

The obtained plasmid preparation was digested with EcoRI (NEB) for 2h at 37°C.

Table 4-3 Reaction composition for plasmid control digest

Component	*Volume*	*Final concentration*
Plasmid prep	3 µl	n.a.
Nuclease-free ddH_2O	5.5 µl	n.a.
EcoRI Buffer (10×)	1 µl	1×
EcoRI (20 U/µl)	0.5 µl	10 U/reaction
Final Volume	***10 µl***	

The resulting fragments (empty plasmid and the inserted amplicon) were separated by gel electrophoresis (1% agarose gel, 100 V, 40min) together with a suitable DNA ladder to test for insert presence and size.

BAC isolation

Mini preparation

The protocol was adopted from "BAC transgenesis protocol v2.9"[267].

Cells from 1.5ml of an overnight culture were harvested by centrifugation (> 8000xg, 20 sec, 4 °C) in a 2ml tube and resuspended in 250 µl Solution 1. Cells were lysed by addition of 250 µl Solution 2. Tubes were inverted two times to mix resuspended cells with Solution 2. The suspension was neutralized by addition of 350 µl ice-cold Solution 3. Tubes were inverted two times to mix lysed cells with Solution 3. The suspension was incubated on ice for 5 min prior to centrifugation (> 8000xg, 4 min, 4 °C). The BAC-containing supernatant was transferred to new 1.5 ml tube containing 750 ml Isopropanol (100%). The BAC DNA was precipitated (> 13000xg, 10 min, 4 °C), washed with 500 µl Ethanol (70%) (> 13000xg, 3 min, 4°C), and resuspended in 10 µl nuclease-free water (O/N, 37 °C).

If the BAC DNA was subject to restricted digest, 25 U of RNase T1 (cloned) were added to the pellet. Cell pellets can be stored at -20 °C until BAC preparation. BAC preparations were stored at 4 °C after resuspension.

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

Tris/HCl (pH 8)	1 M	121.1 g	Tris ultrapure (AppliChem, #A1086) → Adjust with HCl to pH 8 → Ad 1 l nuclease-free ddH_2O
HCl	37%		Hydrochloric acid (fuming) (Merck/Millipore, #100317)
EDTA/NaOH (pH 8)	0.5 M	186.1 g	Ethylenediaminotetraacetic acid disodium salt dehydrate (AppliChem, #A2937) → Dissolve in 500 ml nuclease-free ddH_2O → Adjust with NaOH to pH 8.0 → Ad 1 l nuclease-free ddH_2O
NaOH	1 M		Sodium hydroxide solution (Merck/Millipore, #109137)
SDS	10%	100 g	Sodium dodecyl sulfate (Sigma-Aldrich, #L4390) → Ad 1 l nuclease-free ddH_2O

Solution 1	50 mM	1.08 g	D-(+)-Glucose (Sigma-Aldrich, #G8270)
	25 mM	5 ml	Tris/HCl (1 M, pH 8.0)
	10 mM	4 ml	EDTA/NaOH (0.5 M, pH°8.0)
		→ Ad 200 ml nuclease-free ddH_2O, filtrate to sterilize	
Solution 2	1%	20 ml	SDS (10%)
	0.2 M	40 ml	NaOH (1 M)
		→ Ad 200 ml nuclease-free ddH_2O, filtrate to sterilize	
Solution 3	3 M	58.8 g	Potassium acetate (Merck/Millipore, #104820)
		→Dissolve in 100-160 ml nuclease-free ddH_2O	
	11.5%	23 ml	Acetic acid (glacial) (Merck/Millipore, #100063, 100%)
		→ Ad 200 ml nuclease-free ddH_2O, autoclave	

Maxi preparation

Plasmid preparation for up to 150 µg from single E. coli colonies was done with the NucleoBond® Xtra BAC Kit (Macherey-Nagel) following the manufacturer's instructions.

250 – 400 ml of $LB_{Antibiotics}$ were inoculated from a pre-culture (250 rpm, 8 h, 30 °C; 1:500 dilution) and grown overnight (200 rpm, max. 16 h, 30 °C). Cells were harvested by centrifugation (2900xg, 20 min, 4 °C).

Purified BAC DNA was eluted from the column with 6.5 ml ELU-BAC buffer and resuspended in 25-200 µl Tris/HCl (10 mM, pH 8.0) (o/n, RT) after isopropanol precipitation.

Cell pellets can be stored at -20 °C until preparation. BAC preparations were stored at 4 °C after resuspension.

Control digest and gel electrophoresis

The obtained BAC preparation was digested with HindIII restriction enzyme (NEB) and the fragment pattern assessed by gel electrophoresis (1%, 35-75 V, 5.5-15 h).

Table 4-4 Reaction composition for BAC control digest with HindIII

Component	*BAC mini Prep*	*BAC Maxi Prep*	*Final concentration*
BAC Preparation	10 µl	2 µl	n.a.
Nuclease-free ddH_2O	2 µl	10 µl	n.a.
NEB Buffer 2 (10x)	1.5 µl	1.5 µl	1x
HindIII (20 U/µl)	1.5 µl	1.5 µl	30 U/rxn
Final Volume	***15 µl***	***15 µl***	

4.3.1.2 Isolation of genomic DNA from mammalian cells

Genomic DNA was isolated using the Blood & Cell Culture DNA Midi Kit (Qiagen) or, for smaller cell numbers, the DNeasy Blood & Tissue Culture Kit (Qiagen) according to the manufacturer's instructions. DNA concentration was determined with the NanoDrop spectrophotometer and quality was assessed by agarose gel electrophoresis (see section 4.3.1.4).

Genomic DNA preparation from single cell suspensions

Genomic DNA from single cells suspensions were prepared according to the Animal Blood Spin-Column Protocol (DNeasy Blood & Tissue Kit) for up to $5x10^6$ cells or the 100/G Cell Culture Protocol (Blood & Cell Culture DNA Kit), respectively, for larger cell numbers.

Genomic DNA preparation from tissues and embryoid bodies

Genomic DNA from tissues and embryoid bodies were prepared according to the Animal Tissue Spin-Column Protocol (DNeasy Blood & Tissue Kit) for up to 25 mg. Embryoid bodies were lysed in proteinase K-containing ATL buffer for at least 30 min at 56 °C on a shaker (1100 rpm). Tissues were disrupted with the TissueLyser (Qiagen) (see Table 4-5) and lysed in proteinase K-containing ATL buffer over night at 56 °C on a shaker (1100 rpm). In case of spleen tissue, another 10 µl proteinase K were added the next day and incubation at 56 °C was continued for two hours.

Table 4-5 TissueLyser settings

Tissue	*TissueLyser Settings*
Testis	1x 20 sec, 15 Hz
Cerebellum	1x 20 sec, 15 Hz
Brain	1x 20 sec, 15 Hz
Spleen	2x 20 sec, 15 Hz
Liver	3x 20 sec, 15 Hz

4.3.1.3 Fragmentation of genomic DNA and chromatin

Sonication of genomic DNA

Genomic DNA can also be fragmented by sonication. The advantage compared to enzymatic digestion is a homogeneous fragmentation. Before sonication, gDNA was initially sheared using a 20 gauge needle (BD) attached to a 2 ml syringe (BD) to reduce viscosity before quantification with the NanoDrop ND 1000 spectrophotometer (Peqlab). Sonication to a mean fragment size of 300–400 bp was carried out with the Branson Sonifier 250 (Danbury) using the following settings for 5 µg gDNA in 500 µl TE (1x, pH 8) buffer:

duty cycle	80%
output	0.5
sonication time	2 x 30sec

Fragment range was controlled by agarose gel electrophoresis (1%, 110 V, 30 min; see section 4.3.1.4).

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

TE (1x, pH 8.0)	10 mM	10 ml	Tris/HCl (1 M, pH 8.0)
	1 mM	2 ml	EDTA/NaOH (0.5 M, pH 8.0)
		→ Ad 1 l	nuclease-free ddH_2O
Tris/HCl (pH 8)	1 M	121.1 g	Tris ultrapure (AppliChem, #A1086)
		→ Adjust with HCl to pH 8	
		→ Ad 1 l	nuclease-free ddH_2O
HCl	37%		Hydrochloric acid (fuming) (Merck/Millipore, #100317)
EDTA/NaOH (pH 8)	0.5 M	186.1 g	Ethylenediaminotetraacetic acid disodium salt dehydrate (AppliChem, #A2937)
		→ Dissolve in 500 ml	nuclease-free ddH_2O
		→ Adjust with NaOH to pH 8.0	
		→ Ad 1 l	nuclease-free ddH_2O
20 gauge needle			20 G Microlance 3 (BD, #301300)
2 ml syringe			BD Discardit® II (BD, #300928)

Sonication of chromatin

Crosslinked chromatin, e.g. for ChIP assays, can be fragmented by sonication. After crosslinking with formaldehyde and lysing the fixed cells according to section 4.3.1.17, sonication of chromatin was performed with the following setting:

output 3
sonication time 5 x 10sec
(after each sonication step, place sample on ice to cool again)

The fragmented lysate was centrifuged at 4 °C for 5 min at 13000g to pellet remaining cell debris. The supernatant was transferred to new 1.5 ml tube. To control and assess fragmentation range by agarose gel electrophoresis (see section 4.3.1.4), 30 µl of the sonicated and cleared lysate were prepared according to section 4.3.1.17.

Focused Ultrasonication of genomic DNA (Covaris)

In terms of reproducibility, fragmentation by focused ultrasonication with the Covaris S2 system (Covaris) is favored in general.

Ultrasonication was done in microTUBEs (Covaris). The best fragmentation result is obtained when diluting the DNA sample to a concentration of 25 ng/µl, but up to 5 µg (in 130 µl) can be processed per reaction.

Table 4-6 Parameters for Focused Ultrasonication with Covaris

Target base pair (Peak)	200 bp	300 bp	400 bp	650 bp
Intensity	5	4	4	3
Duty Cycle	10%	10%	10%	5%
Cycles per Burst	200	200	200	200
Treatment Time	180 sec	80 sec	55 sec	65 sec
Temperature	7 °C	7 °C	7 °C	7 °C
Water Level S2	12	12	12	12

Fragment range was controlled by agarose gel electrophoresis (1%, 110 V, 30 min; see section 4.3.1.4)

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

TE (1x, pH 8.0)	10 mM	10 ml	Tris/HCl (1 M, pH 8.0)
	1 mM	2 ml	EDTA/NaOH (0.5 M, pH 8.0)
		→ Ad 1 l	nuclease-free ddH_2O
Tris/HCl (pH 8)	1 M	121.1 g	Tris ultrapure (AppliChem, #A1086)
		→ Adjust with HCl to pH 8	
		→ Ad 1 l	nuclease-free ddH_2O
HCl	37%		Hydrochloric acid (fuming) (Merck/Millipore, #100317)
EDTA/NaOH (pH 8)	0.5 M	186.1 g	Ethylenediaminotetraacetic acid disodium salt dehydrate (AppliChem, #A2937)
		→ Dissolve in 500 ml	nuclease-free ddH_2O
		→ Adjust with NaOH to pH 8.0	
		→ Ad 1 l	nuclease-free ddH_2O
20 gauge needle			20 G Microlance 3 (BD, #301300)
2 ml syringe			BD Discardit® II (BD, #300928)
microTUBE			microTUBE (Covaris, #520045)

4.3.1.4 Agarose gel electrophoresis

The required amount of agarose (according to *Table 4-7*) was added to the corresponding amount of TAE (1x). The slurry was heated in a microwave oven until the agarose was completely dissolved. Ethidium bromide was added after cooling the solution to 50 °C. The gel was cast, mounted in the electrophoresis tank and covered with TAE (1x). DNA-containing samples were diluted 4:1 with DNA loading dye (5x), mixed, and loaded into the slots of the submerged gel. Depending on the size and the desired resolution, gels were run at 35-100 V for 30 min to 16 h.

Table 4-7 Agarose concentrations for different separation ranges

Agarose concentration in gel in%	*Efficient range of separation in kb*
0.5	genomic DNA
1.0	0.4-6
1.5	0.2-3
2.0	0.1-2

DNA fragments of specific sizes can be extracted after separation (see section 4.3.1.6).

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

SDS	10%	100 g	Sodium dodecyl sulfate (Sigma-Aldrich, #L4390)
		→ Ad 1 l	nuclease-free ddH_2O
NaOH	1 M		Sodium hydroxide solution (Merck/Millipore, #109137)
Glacial AcOH			Acetic acid (>99.7%) (Sigma-Aldrich, #320099)
NaOAc/AcOH (pH 7.8)	1 M	82 g	Sodium acetate anhydrous (Sigma-Aldrich, #2889)
		→ Dissolve in 500 ml	nuclease-free ddH_2O
		→ Adjust with	glacial AcOH to pH 7.8
		→ Ad 1 l	nuclease-free ddH_2O, autoclave
TAE (50x)	2 M	252.3 g	Tris ultrapure (AppliChem, #A1086)
	250 mM	250 ml	NaOAc/HOAc, pH 7.8
	50 mM	18.5 g	Ethylenediaminotetraacetic acid disodium salt dehydrate (AppliChem, #A2937)
		→ Ad 1 l	nuclease-free ddH_2O, autoclave

TE (1x, pH 8.0)	10 mM	10 ml	Tris/HCl (1 M, pH 8.0)
	1 mM	2 ml	EDTA/NaOH (0.5 M, pH 8.0)
		→ Ad 1 l nuclease-free ddH_2O	
Tris/HCl (pH 8)	1 M	121.1 g	Tris ultrapure (AppliChem, #A1086)
		→ Adjust with HCl to pH 8	
		→ Ad 1 l nuclease-free ddH_2O	
HCl	37%		Hydrochloric acid (fuming) (Merck/Millipore, #100317)
EDTA/NaOH (pH 8)	0.5 M	186.1 g	Ethylenediaminotetraacetic acid disodium salt dehydrate (AppliChem, #A2937)
		→ Dissolve in 500 ml nuclease-free ddH_2O	
		→ Adjust with NaOH to pH 8.0	
		→ Ad 1 l nuclease-free ddH_2O	
DNA loading dye (5x)	50 mM	500 µl	Tris/HCl, pH 8.0
	1%	500 µl	SDS (20%)
	50 mM	1 M	EDTA/NaOH (0.5 M, pH 8.0)
	40%	4 ml	Glycerol (anhydrous) (AppliChem, #A1123)
	1%	10 mg	Bromphenol blue (Sigma-Aldrich, #B0126)
		→ Ad 10 ml nuclease-free ddH_2O, store at 4°C	
X% Agarose	X%	X g	Biozym LE Agarose (Biozym, # 840004)
		→ Add 100 ml TAE (1x) per 1 g agarose and heat in a microwave until agarose is completely dissolved	
		→ Cool to 50 °C and add 5 µl Ethidium bromide (10 mg/ml) (Sigma)	
DNA ladder		50 bp	50 bp DNA ladder (NEB, #N3236)
		100 bp	1 kb Plus DNA ladder (Invitrogen/Life Tech., #10787018)
		>100 bp	λ DNA/Hind III Fragments (Invitrogen/Life Tech., 5612013)

4.3.1.5 **Pulsed-field gel electrophoresis (PFGE)**

BAC preparation with low contamination of genomic DNA from the host strain (*E. coli*) is needed for sequence analysis. PFGE enables separation of larger fragments with different sizes.

600 µl LB_{CM} are inoculated from a glycerol stock and incubated at 30°C on a shaker (150 rpm) for 6 h. 15 ml LB_{CM} are inoculated with 500 µl preculture in a 50 ml conical tube (Falcon) and incubated o/n at 30°C on a shaker (150 rpm) for max. 16 h. The cells are harvested by centrifugation (30 min, 1000xg, 4°C) and washed once with ice-cold 10 ml PET-IV. Cell pellets are carefully resuspended by vortexing (volume depends on bacterial growth and is roughly 140 µl). The resuspended pellets are subsequently warmed to 42 °C and mixed with the same Volume of prewarmed LowMelt agarose (2%, 42°C) by pipetting. The agarose-cell mix was immediately transferred to a disposable plug molder (Bio-Rad) and cooled for at least 15 min at 4 °C for hardening. The solid cell-containing agarose beads are transferred to a 96-deep well plate and incubated o/n at 37 °C with 250 µl lysozyme mix to lyse immobilized cells and to degrade RNAs. The beads were washed twice with TE (1x, pH 8) before addition of 250 µl PK mix and o/n incubation at 47 °C. Beads should appear pallid and clear and kept at 4 °C. The beads were washed twice with 250 µl TE (1x, pH 8) and store in fresh 250 µl TE (1x, pH 8) on at 4 °C. Remaining protein and cell debris were removed by gel electrophoresis in a 1% agarose gel (0.5x TAE, 90 min, 160 V, 4 °C). The immobilized BAC DNA was linearized with PI-SceI after o/n incubation at 4 °C with 250 µl PI-SceI digest mix for 3-3.5 h at 37 °C. The beads were washed twice with 250 µl TE (1x, pH 8) and store again in fresh 250 µl TE (1x, pH 8) on at 4 °C. The linearized BAC DNA was then separated by Pulsed Field Gel Electrophoresis in 1% LowMelt agarose (0.5x TAE, 20 h, 400 mA, 100 W) according to the following settings:

8 h: 3 sec→10 sec (in total 4430 pulses)

12 h: 10 sec→80 sec (in total 958 pulses)

The gel was stained for 60 min in 300 ml EtBr-containing TAE (0.5x) (20 µl aqueous EtBr solution (1%) per 300 ml 0.5x TAE). BAC-containing bands were cut out with disposable plastic cuvettes and transferred to 2 ml DNA loBind reaction tubes (Eppendorf). Agarose was melted at 72 °C in a water bath (15-30 min) and then cooled to 42 °C in a heating block. Melted agarose was digested by adding 5 U Agarase (NEB) and 25 µl TAE (12.5x) per tube and incubating o/n at 42 °C on a shaking (350 rpm) heating block.

BAC DNA was recovered from the remaining solution by Isopropanol precipitation. The solution was mixed with 70 µl NaOAc (3 M, pH 5.2) and incubated on ice for 30 min. Undigested remaining agarose was removed by centrifugation (15 min, >13000xg, 4 °C) and the supernatant was transferred to a new 2 ml tube containing 6 µl NaOAc (3 M, pH 5.2) and 2 µl Linear acrylamide (5 mg/ml) (Invitrogen/Life Tech.). After o/n incubation at -20 °C the precipitated DNA was pelleted by centrifugation (min. 45 min, >13000xg, 4 °C) and washed once with 70% EtOH (-20 °C). The DNA pellet was dried at 37 °C and resolved in 20 µl TE (1x, pH 8) o/n at 4 °C.

BAC-*E. coli* DNA ratio was determined by qPCR (see section 4.3.1.10) with BAC- and *E. coli*-specific primers (see Table 3-4).

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

LB_0 medium and antibiotics			see section 4.1.2.1
PET-IV buffer	10 mM	1,21 g	Tris ultrapure (AppliChem, #A1086)
		→ Dissolve in 500 ml nuclease-free ddH_2O	
		→ Adjust with HCl to pH 7.6	
	1 M	58.44g	Sodium chloride (AppliChem, #A3597)
		→ Ad 1 l nuclease-free ddH_2O, autoclave	
Glacial AcOH			Acetic acid (>99.7%) (Sigma-Aldrich, #320099)
NaOAc/AcOH (pH 5.2)	3 M	24.6 g	Sodium acetate anhydrous (Sigma-Aldrich, #2889)
		→ Dissolve in 50 ml nuclease-free ddH_2O	
		→ Adjust with glacial AcOH to pH 5.2	
		→ Ad 100 ml nuclease-free ddH_2O, autoclave	
NaOAc/AcOH (pH 7.8)	1 M	82 g	Sodium acetate anhydrous (Sigma-Aldrich, #2889)
		→ Dissolve in 500 ml nuclease-free ddH_2O	
		→ Adjust with glacial AcOH to pH 7.8	
		→ Ad 1 l nuclease-free ddH_2O, autoclave	
TAE (50x)	2 M	252.3 g	Tris ultrapure (AppliChem, #A1086)
	250 mM	250 ml	NaOAc/AcOH, pH 7.8
	50 mM	18.5 g	Ethylenediaminotetraacetic acid disodium salt dehydrate (AppliChem, #A2937)
		→ Ad 1 l nuclease-free ddH_2O, autoclave	
TE (1x, pH 8.0)	10 mM	10 ml	Tris/HCl (1 M, pH 8.0)
	1 mM	2 ml	EDTA/NaOH (0.5 M, pH 8.0)
		→ Ad 1 l nuclease-free ddH_2O	
Tris/HCl (pH 8)	1 M	121.1 g	Tris ultrapure (AppliChem, #A1086)
		→ Adjust with HCl to pH 8	
		→ Ad 1 l nuclease-free ddH_2O	

EDTA/NaOH (pH 8)	0.5 M		186.1 g Ethylenediaminotetraacetic acid disodium salt dehydrate (AppliChem, #A2937) → Dissolve in 500 ml nuclease-free ddH_2O → Adjust with NaOH to pH 8.0 → Ad 1 l nuclease-free ddH_2O
NaOH	1 M		Sodium hydroxide solution (Merck/Millipore, #109137)
HCl	37%		Hydrochloric acid (fuming) (Merck/Millipore, #100317)
LowMelt Agarose	2%	2g	Biozym Plaque Agarose (Biozym, #840100) → Ad 10 ml TAE (1x), heat in microwave to dissolve
Agarose			Biozym LE Agarose (Biozym, # 840004)
Lysozyme mix		22.5 ml	EC-Lyse solution
		2.5 ml	Lysozyme solution (10x)
	0.5 µg/ml	1.25 µl	RNase (10 mg/ml, Sequenom)
		→ Prepare freshly	
EC-Lyse buffer	6 mM	0,72 g	Tris ultrapure (AppliChem, #A1086)
	1 M	58.44 g	Sodium chloride (AppliChem, #A3597)
	100 mM	37.22 g	Ethylenediaminotetraacetic acid disodium salt dehydrate (AppliChem, #A2937)
	0.5%	0.5 g	Brij-58 (Sigma-Aldrich, # P5884)
	0.2%	0.2 g	Deoxycholate (Sigma-Aldrich, #D6850)
	0.5%	25 ml	Laurosyl-sarcosine sodium salt solution (20%) (Fluka, #61747)
		→ Dissolve in 500 ml nuclease-free ddH_2O	
		→ Adjust with NaOH to pH 9.0 to enable EDTA/NaOH solving	
		→ Adjust with HCl to pH 7.5	
		→ Ad 1 l nuclease-free ddH_2O, store at 4 °C	
Lysozyme solution	10x	7.2 mg	Lysozyme from chicken egg white (Sigma-Aldrich, #L6878)
		→ Dissolve in 1 ml Tris/HCl (10 mM, pH 8)	
		→ Sterile-filtrated, stable at 4 °C for up to 1 month	

NDS buffer	0.5 M	186.1 g	Ethylenediaminotetraacetic acid disodium salt dehydrate (AppliChem, #A2937)
		→ Ad 950 ml nuclease-free ddH_2O, autoclave	
	1%	50 ml	Laurosyl-sarcosine sodium salt solution (20%) (Fluka, #61747)
		→ Store at room temperature	
PK mix	50 µg/ml	25 µl	Proteinase K (20 mg/ml) (Ambion/Life Tech., # AM2548)
		→ Ad 10 ml NDS buffer	
PI-SceI digest mix	20 U/µl	100 µl	PI-SceI (NEB, #R0696L; 5 U/µl)
	1x	2.5 ml	PI-SceI buffer (10x) (NEB)
	1x	250 µl	BSA (100x) (NEB)
		→ Ad 25 ml nuclease-free ddH_2O	
Plastic cuvettes			PLASTIBRAND™ standard disposable Cuvettes (Sigma-Aldrich, #Z330396)
Plug molds			50-well disposable plug molds (Bio-Rad, #1703713)

4.3.1.6 Purification of DNA fragments by gel extraction

DNA fragments were size-separated by agarose gel electrophoresis (with DNA-intercalating dye, e.g. Ethidium bromide) (see section 4.3.1.4). The fragment of interest was excised under UV illumination. Fragments were purified by gel extraction using QIAquick Gel Extraction Kit (Qiagen) following the supplier's instructions.

4.3.1.7 Purification of DNA by phenol/chloroform extraction

Before starting, two MaXtract High Density tubes (Qiagen) per sample were centrifuged (13000xg for 5 min) to pellet the containing gel. 1 Vol Phenol:Chloroform:Isoamylalcohol (25:24:1; pH 8) was mixed with the DNA-containing solution directly in a prepared MaXtract tube. The tube was centrifuged at 13000xg for 5 min to separate the aqueous and the organic phase. The upper DNA-containing, aqueous phase was transferred to new MaXtract tubes and mixed with 1 Vol Chloroform:Isoamylalcohol (24:1). Both phases were separated again by centrifugation. The aqueous phase was then transferred to a new 1.5 ml tube and DNA was recovered by ethanol precipitation (see section 4.3.1.9).

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

Phenol:Chloroform:Isoamylalcohol	25:24:1		Phenol-chloroform-isoamyl alcohol mixture (Sigma-Aldrich, #77617)
		→ Store at 4 °C	
Chloroform:Isoamylalcohol	24:1		Chloroform:Isoamyl alcohol (Sigma-Aldrich, #C0549)
		→ Store at 4 °C	

4.3.1.8 Polyethylene glycol 8000 (PEG8000) precipitation of DNA

For precipitation of DNA from small Volumes, e.g. PCR reactions or endonuclease digestions, one Volume of PEG8000 Mix was added to the DNA-containing solution, vortexed and incubated for 10 min at RT. After centrifugation (10 min, > 13000xg, RT), the supernatant was discarded and the precipitated DNA was washed by carefully adding 200 µl EtOH (100%) to the (often invisible) pellet. Following centrifugation (10 min, > 13000xg, RT), the supernatant was carefully removed. The pellet was dried and resuspended in H_2O in 50-75% of the initial Volume.

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

PEG8000 Mix	26.2%	26.2 g	PEG 8000
	0.67 M	22 ml	NaOAc/AcOH (3 M, pH 5.2)
	0.67 M	660 µl	$MgCl_2$ (1 M)
		→ Ad 100 ml nuclease-free ddH_2O	
Glacial AcOH			Acetic acid (>99.7%) (Sigma-Aldrich, #320099)
NaOAc/AcOH (pH 5.2)	3 M	24.6 g	Sodium acetate
		→ Dissolve in 50 ml nuclease-free ddH_2O	
		→ Adjust with AcOH to pH 5.2	
		→ Ad 100 ml nuclease-free ddH_2O	
$MgCl_2$	1 M	9.5 g	Magnesium chloride (Sigma-Aldrich, #M8266)
		→ Ad 100 ml nuclease-free ddH_2O	

4.3.1.9 Alcohol precipitation of DNA

Ethanol precipitation

To precipitate DNA from a solution, 0.1 Vol NaOAc/AcOH (3 M, pH 5.2) and 2.5 Vol EtOH (100%) were added and the mixture was incubated for min. 1 hour at -20 °C. The precipitated DNA was pelleted by centrifugation (45 min, >13000xg, 4 °C). Precipitated DNA was washed once with 1 ml EtOH (80%) and dissolved in TE (1x, pH 8.0), Tris/HCl (10 mM, pH 8.0) or nuclease-free ddH_2O (pH 8.0), respectively. In case of low initial DNA concentration, carriers such as glycogen or linear acrylamide (1-3 µg/precipitation) can be added to the solution prior to precipitation.

Isopropanol precipitation

To precipitate DNA from a large Volume, precipitation of DNA with isopropanol is preferred. 0.1 Vol NaOAc (3 M, pH 5.2) and 0.4 Vol isopropanol (100%) were added to the solution, mixed, and immediately centrifuged (45 min, >13000xg, 4 °C) to avoid co-precipitation of salt. Precipitated DNA was washed once with 1 ml EtOH (70%) and dissolved in TE (1x, pH 8.0), Tris/HCl (10 mM, pH 8.0) or nuclease-free ddH_2O (pH 8.0), respectively. In case of very low initial DNA concentration, carriers such as Glycogen or Linear acrylamide (1-3 µg/precipitation) can be added to the solution prior to precipitation.

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

EtOH	70%	70 ml	Ethanol p.a. (Roth, #9065.3, ≥99.8%) → Ad 100 ml nuclease-free ddH_2O
	80%	80 ml	Ethanol p.a. (Roth, #9065.3, ≥99.8%) → Ad 100 ml nuclease-free ddH_2O
Isopropanol			2-Propanol p.a. (Merck/Millipore, #1096342511)
Glacial AcOH			Acetic acid (>99.7%) (Sigma-Aldrich, #320099)
NaOAc/AcOH (pH 5.2)	3 M	24.6 g	Sodium acetate → Dissolve in 50 ml nuclease-free ddH_2O → Adjust with AcOH to pH 5.2 → Ad 100 ml nuclease-free ddH_2O

NaOH	1 M		Sodium hydroxide solution (Merck/Millipore, #109137)
TE (1x, pH 8.0)	10 mM	10 ml	Tris/HCl (1 M, pH 8.0)
	1 mM	2 ml	EDTA/NaOH (0.5 M, pH 8.0)
		→ Ad 1 l	nuclease-free ddH_2O
Tris/HCl (pH 8)	1 M	121.1 g	Tris ultrapure (AppliChem, #A1086)
		→ Adjust with HCl to pH 8	
		→ Ad 1 l	nuclease-free ddH_2O
HCl	37%		Hydrochloric acid (fuming) (Merck/Millipore, #100317)
EDTA/NaOH (pH 8)	0.5 M	186.1 g	Ethylenediaminotetraacetic acid disodium salt dehydrate (AppliChem, #A2937)
		→ Dissolve in 500 ml	nuclease-free ddH_2O
		→ Adjust with NaOH to pH 8.0	
		→ Ad 1 l	nuclease-free ddH_2O
Glycogen	5 mg/ml		Glycogen (Ambion/Life Tech. #AM9510)
Linear acrylamide	5 mg/ml		Linear acrylamide (Ambion/Life Tech. #AM9520)

4.3.1.10 Polymerase chain reaction

Polymerase chain reaction (PCR) allows the amplification of specific DNA segments. A thermo stable DNA polymerase synthesizes the sister strand of a heat-denatured single-stranded DNA when deoxynucleotide triphosphates are added under appropriate conditions[268]. The polymerization reaction is “primed” with small oligonucleotides that anneal to the template DNA strand through base pairing giving the reaction its specificity by defining the borders of the amplified segment.

Standard PCR to amplify specific target regions was performed with the Expand High Fidelity PCR System (Roche) according to the following protocol.

Table 4-8 Reaction composition for standard PCR

Component	*Volume*	*Final concentration*
Mix 1		
genomic DNA template	x µl	10-250 ng
nuclease-free ddH_2O	21.6 – x µl	n.a.
dNTPs (25 mM each)	0.4 µl	0.2 mM each
Primer_forward (10 µM)	1.5 µl	0.3 µM
Primer_reverse (10 µM)	1.5 µl	0.3 µM
Final Volume	***25 µl***	
Mix 2		
nuclease-free ddH_2O	19.25 µl	n.a.
10x Buffer	5 µl	1x
Taq Polymerase (3.6 U/µl)	0.75 µl	2.6 U
Final Volume	***25 µl***	

Mix 1 and 2 were mixed on ice immediately before starting the cycling protocol.

Table 4-9 Cycling protocol for standard PCR

Cycle step	*Temperature*	*Time*	*Number of cycles*
Initial denaturation	94 °C	2 min	1×
Denaturation	94 °C	15 sec	32×
Annealing	45-65 °C	30 sec	
Extension	72 °C	45 sec – 2 min	
Final extension	72 °C	7 min	1x
Cooling	*4 °C*	*hold*	

Quantitative real-time PCR (qPCR)

QPCR was used to test for specific DNA enrichment by Methyl-CpG immunoprecipitation (MCIp, see section 4.3.1.15), Chromatin immunoprecipitation (ChIP, see section 4.3.1.17), Pulsed-Field Gel Electrophoresis (PFGE, see section 4.3.1.5), as well as for local expression analysis after reverse transcription PCR (see section 4.3.2.2). QPCR was performed with the QuantiFast SYBR Green Kit (Qiagen) in 96-well format adapted to the Eppendorf Realplex Mastercycler EpGradient S system (Eppendorf) and 384-well format adapted to the LightCycler480 system (Roche) (*Table* 4-10 & *Table 4-11*). The relative amount of amplified DNA was measured through the emission of light by the SYBR green dye after each extension step. Specificity of the amplification product was determined by a melting curve.

Table 4-10 Reaction composition for quantitative PCR

Component	*Volume*	*Final concentration*
SYBR Green mix (2×)	5 µl	1×
Nuclease-free ddH_2O	2 µl	n.a.
Primer_forward (10 µM)	0.5 µl	0.5 µM
Primer_reverse (10 µM)	0.5 µl	0.5 µM
Template DNA	2 µl	n. a.
Final Volume	***10 µl***	

Table 4-11 Cycling protocol for quantitative PCR

Cycle step	*Temperature*	*Time*	*Number of cycles*
Initial denaturation	95 °C	5 min	1×
Denaturation	95 °C	8 sec	45×
Annealing & Extension	60 °C	20 sec	
Final denaturation	95 °C	15 sec	1×
Final extension	60 °C	15 sec	
Melting Curve	60 → 95 °C	10 min	1×
Cooling	*4 °C*	*hold*	

To calculate amplification efficiency, a dilution series (e.g. 1:10; 1:50; 1:100, 1:1000) of a suitable DNA sample was additionally measured for each amplicon. The software automatically calculated relative DNA amounts based on the generated slope and intercept. Specific amplification was determined by melting curve analysis. Data were imported and processed in Microsoft Excel 2010. All samples were generally measured in duplicates.

Unless otherwise mentioned genomic sequences for primer design were extracted from the *mm9* reference sequence using the UCSC Genome Browser. In general, primers were designed with

PerlPrimer software and evaluated using *in silico* PCR and BLAT functions of the UCSC Genome Browser and GeneRunner software. The following settings were used to design specific primers:

Primer T_m:	65-68 °C (max 2 °C difference)
Primer length:	18-28 bp
Amplicon size:	70-150 bp

Required reagents and materials:
(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

QuantiFast SYBR Green PCR Kit	Qiagen (#204056)

4.3.1.11 *In vitro* methylation and demethylation of genomic DNA

In vitro *methylation*

In order to generate fully methylated DNA as a control for methylation analysis, gDNA was methylated using SssI methyltransferase (New England Biolabs). S-adenosyl methionine (SAM) was used as a methyl donor. In general the following reaction conditions were used:

Table 4-12 Reaction composition for* in vitro *methylation

Component	*Volume*	*Final concentration*
Nuclease-free ddH_2O	84.5 – x µl	n. a.
genomic DNA	x µl	20 µg
Buffer NEB2 (10×)	10 µl	1×
S-adenosyl methionine (32 mM)	0.5 µl	160 µM
SssI (4 U/µl)	5 µl	20 U
Final Volume	***100 µl***	

After 2 hours incubation at 37°C the reaction was supplied with additional 0.75 µl SAM and 1 µl SssI followed by incubation of another hour.
To inactivate the SssI methyltransferase the sample was incubated at 65°C for 20min.
If necessary, *in vitro* methylated genomic DNA was cleaned up with the QIAquick PCR Purification according to the manufacturer's instructions.

Required reagents and materials:
(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

CpG methyltransferase (SssI)	M. SssI NEB (#M0226S)

In vitro *demethylation*

Unmodified cytosines as well as modified (e.g. methylated cytosine) pair with guanine during replication of DNA. Thus, genomic DNA was *in vitro* demethylated by whole genome amplification with either the REPLI-g Mini or Midi Kit (Qiagen) according to the manufacturer's instruction. If necessary, *in vitro* demethylated genomic DNA was cleaned up with the QIAquick PCR Purification according to the manufacturer's instructions.

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

REPLI-g Kit — Qiagen (#150023, 150043)

4.3.1.12 Bisulfite treatment of genomic DNA

Cytosine derivates undergo reversible reactions with sodium bisulfite, yielding a 5,6-Dihydro-6-sulfonate, which deaminates under alkaline conditions leaving uracil whereas 5-methylcytosine is not affected by this reaction[269]. After PCR amplification, unmethylated cytosine appears as thymine in contrast to 5-methylcytosine, which remains as cytosine.

Bisulfite conversion was performed with the EZ DNA Methylation Kit (Zymo Research) according to the manufacturer's instructions. Bisulfite conversion was optimised with the following incubation protocol:

Table 4-13 Cycling protocol for bisulfite treatment

Cycle step	*Temperature*	*Time*	*Number of cycles*
Denaturation	95 °C	30 sec	20×
Sulfonation	50 °C	15 min	
Cooling	*4 °C*	*hold*	

The bisulfite-treated DNA was recovered by adding nuclease-free water (10 µl/10 ng input DNA) directly to the column centre and centrifugation for 30 seconds at 9500xg. The procedure yields bisulfite-converted DNA with a concentration of 7-8 ng/µl.

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

EZ DNA Methylation Kit (single columns) — Zymo Research (#D5002)

EZ-96 DNA Methylation Kit (96-well format) — Zymo Research (#D5004)

4.3.1.13 Sanger sequencing of DNA and sequence analysis

To amplify specific genomic regions, primers were designed and evaluated as described in section 4.3.1.10. Thermocycling parameters are also enlisted in section 4.3.1.10. Positive and specific amplification was evaluated by agarose gel electrophoresis (see section 4.3.1.4). PCR reactions were either gel-purified using agarose gel electrophoresis followed by purification with QIAquick Gel Extraction Kit (Qiagen) (see section 4.3.1.4 & 4.3.1.6) or directly used for PEG8000 precipitation (see section 4.3.1.8). Sanger sequencing was done by Geneart/Life Technologies (Regensburg, Germany). Purified PCR products can directly be sequenced using one of the PCR primers. To obtain sequence information of single DNA strands, the PCR product was subcloned with the StrataClone PCR Cloning Kit (Agilent Technologies) or the TOPO TA Cloning Kit (Invitrogen/Life Technologies), respectively (see section 4.3.3.1). Plasmids from individual clone were isolated with NucleoSpin® Plasmid Quick Pure Kit (Macherey-Nagel) and sequenced with one of the backbone primers. The obtained sequence was analyzed with the UCSC Browser and GeneRunner as well as the BioEdit Software.

Bisulfite sequencing

A common method for analyzing cytosine methylation is bisulfite sequencing[269]. Bisulfite treatment of DNA generates methylation-dependent sequence variations: Cytosine (C) is deaminated to uracil, which generates thymine (T) after PCR amplification whereas 5-methylcytosine (5mC) remains unaffected (see section 4.3.1.12). This methylation-dependent difference can be detected by sequencing. All reagents were purchased from Sequenom unless otherwise mentioned.

Table 4-14 Reaction composition for PCR amplification prior to bisulfite sequencing

Component	*Volume*	*Final concentration*
Nuclease-free ddH_2O	6.44 µl	n. a.
PCR buffer (10×)	1 µl	1×
dNTPs (25 mM)	0.08 µl	200 µM each
Primer_10F (10 µM)	0.2 µl	0.2 µM
Primer_T7R (10 µM)	0.2 µl	0.2 µM
Bisulfite-converted DNA	2 µl	~ 20 ng
DNA Polymerase (5 U/µl)	0.08 µl	0.4 U/reaction
Final Volume	*10 µl*	

Table 4-15 Cycling protocol for PCR amplification prior to bisulfite sequencing

Cycle step	*Temperature*	*Time*	*Number of cycles*
Initial denaturation	95 °C	1 min	1×
Denaturation	95 °C	15 sec	40×
Annealing	56-62 °C	15 sec	
Extension	72 °C	30 sec	
Final extension	72 °C	7 min	1×
Cooling	*4 °C*	*hold*	

To determine the methylation status of each CpG dinucleotide in dependence of the genetic context, the PCR product was in general subcloned prior to sequencing (see section 4.3.3.1).
PCR primers were especially designed to amplify bisulfite-treated DNA (MethPrimer, www.urogene.org/methprimer/). This software excludes regions with CpG dinucleotides as potential primer binding sites in order to obtain methylation status-independent primers.
The following settings were used for the MethPrimer software:

	Min.	*Opt.*	*Max.*
Product size (bp)	250	350	450
Primer Tm (°C)	55	59	65
Primer size (bp)	20	25	30

Product CpGs: 4	Primer non-CpG ‘C’s: 4
Primer Poly X: 5	Primer Poly T: 5

Specific 5’ tags were added to the primer sequences to allow for mass spectrometric analyses of bisulfite-treated DNA with the MassARRAY compact system (see section 4.3.1.14).

Required reagents and materials:
(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer’s instructions or at RT if not stated otherwise)

Complete PCR Reagent Set	Sequenom

4.3.1.14 MassARRAY® system (Sequenom)

The MassARRAY system is a matrix-assisted laser desorption/ionization time of flight mass spectrometer (MALDI-TOF MS) to quantitatively and qualitatively analyze nucleic acids. In our laboratory, we use this system for quantitative assessment of DNA methylation (EpiTYPER) and genotyping (iPLEX). All reagents were purchased from Sequenom unless otherwise mentioned. All centrifugation steps were carried out at 3000rpm unless otherwise mentioned.

EpiTYPER: Quantitative DNA methylation analysis

DNA samples for quantitative methylation analysis are initially bisulfite-treated, resulting in the conversion of unmethylated cytosines to uracil, whereas methylated cytosines remain unchanged. This conversion reaction allows for accurate discrimination between methylated and unmethylated cytosines at CpG dinucleotides. Following bisulfite treatment, genomic DNA consists of two non-complementary single-stranded DNA populations. Subsequently, PCR primer pairs for a region of interest are designed to amplify either the forward and reverse strand of bisulfite-treated genomic DNA. A T7 polymerase promoter tag is added to the 5' end of the reverse primer to facilitate *in vitro* transcription. A 10-mer tag is added to the 5' end of the forward PCR primer to minimize melting temperature differences between both primers during PCR cycling.

Following PCR, unincorporated dNTPs are dephosphorylated by treatment with Shrimp alkaline phosphatase (SAP). Reverse transcription is performed using a chemically modified T7 RNA polymerase, which utilizes a mixture of ribonucleotides and deoxyribonucleotides when synthesizing the RNA strand. In parallel with the reverse transcription the cleave reaction is achieved using the pyrimidine-specific Ribonuclease A (RNase A) enzyme, which cleaves at pyrimidines (C and T) only on the newly synthesized transcript. By incorporating a non-cleavable dCTP (deoxyribonucleotide) into the transcript, RNase A is unable to cleave at C and can only cleave at T (T-specific cleavage) yielding a population of single-stranded cleavage fragments (Figure 4-1). A methylated cytosine is represented by a G nucleotide in the cleavage fragment, whereas an unmethylated cytosine is represented by an A nucleotide. The mass difference of 16 Dalton between G (329 Da) and A (313 Da) is easily detected by MALDI-TOF MS. Depending on the number of methylated CpG sites within a cleavage fragment, the difference in mass will increase in 16 Da units. As already mentioned, in the following procedure this methylation-specific difference is not addressed by sequencing but for generating methylation-depending mass differences to be analyzed by mass spectrometry. A detailed description of the method is given by Ehrich *et al*.[270] and in the EpiTYPER User Guide (www.sequenom.com).

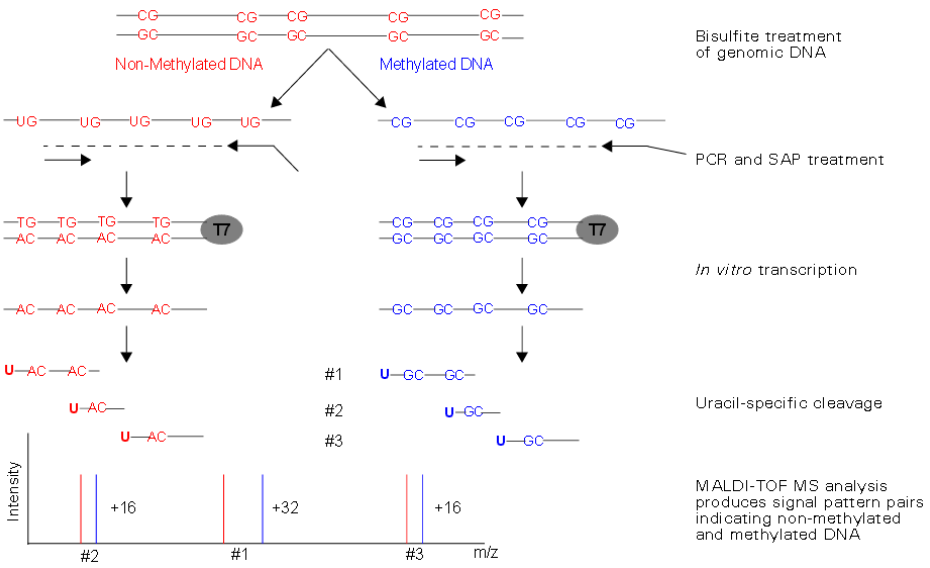


Figure 4-1 Schematic outline of the EpiTYPER technology
Genomic DNA is isolated and bisulphite treated to generate methylation specific mass differences. Regions of interest are amplified by PCR following *in vitro* transcription and base specific cleavage (in the figure U-specific, in the assay used for this work T-specific). Mass differences are then analyzed with MALDI-TOF MS (Figure taken from www.sequenom.com).

Primer design

Genomic DNA sequences were downloaded from the University of California, Santa Cruz genome browser (http://www.genome.ucsc.edu). In order to maximize coverage, both the forward and reverse strand of a target region were included for amplicon design. The selected genomic sequence was subsequently exported to the primer design software MethPrimer (http://www.urogene.org/methprimer). Once genomic DNA is uploaded into this application, an *in silico* bisulfite conversion is performed to facilitate the primer design. PCR primer design criteria consist of the following: An optimal primer melting temperature of 62 °C (range: T_m=56-64 °C); Primer length ranges from 20-30 nucleotides, excluding tag addition. Amplicons vary from 100-500 bp in length, with a desired length of 400 bp. All primers designed for methylation analysis using the MassCLEAVE assay are listed in section 3.10.4.1. Primers were ordered in 96-well format at 100 µM concentration (Sigma-Aldrich). Prior to PCR set up, a 96-well primer mix plate (Sarstedt V-bottom, Newton, USA) was assembled, with each well containing 1 µl of both the forward and reverse primers of a primer pair and 198 µl ddH2O to give a final concentration of 0.5 µM each.

To facilitate reverse transcription of the amplified target region by T7 RNA Polymerase (see "*Reverse transcription and RNA base-specific cleavage"*), all reverse primers were tagged at the 5' end with the T7 promoter sequence; to balance for the elongated length of the reverse primer, all forward primer carry a 10mer tag at the 5' end:

Forward tag	(10F)	5'-AGGAAGAGAG
Reverse tag (T7R)		5'-CAGTAATACGACTCACTATAGGGAGAAGGCT

Bisulfite treatment

Bisulfite conversion of genomic DNA was performed with the EZ DNA Methylation Kit™ (Zymo Research) according to section 4.3.1.12.

PCR amplification

PCR master mixes were prepared in 384-well plates (Thermo Scientific) and made as follows per reaction according to the manufacturer's instructions:

Table 4-16 Reaction composition for EpiTYPER PCR

Component	*Volume per single reaction*	*Final concentration*
Nuclease-free ddH_2O	1.42 µl	n. a.
PCR Buffer (10×)	0.5 µl	1×
dNTPs (25 mM each)	0.04 µl	200 µM
Taq Polymerase (5 U/µl)	0.04	0.2 U
DNA template	1 µl	5 – 10 ng
Final Volume	*3 µl*	

To each reaction 2 µl primer mix was added giving a final reaction Volume of 5 µl, with the final concentration of 500 pM of the forward and reverse primer. The plate was sealed with adhesive PCR sealing foil (Thermo Scientific), centrifuged, and incubated in a Veriti 384-well thermal cycler (Applied Biosystems) with the following program:

Table 4-17 Cycling protocol for EpiTYPER PCR

Cycle step	*Temperature*	*Time*	*Number of cycles*
Initial denaturation	94 °C	4 min	1×
Denaturation	94 °C	20 sec	45×
Annealing	59 °C	30 sec	
Extension	72 °C	1 min	
Final extension	72 °C	3 min	1×
Cooling	*4 °C*	*hold*	

Shrimp alkaline phosphatase (SAP) treatment

Unincorporated nucleotides would disturb downstream applications and are enzymatically inactivated. Under alkaline conditions SAP removes phosphate groups from several substrates including deoxynucleotide triphosphates, rendering it unavailable for further polymerase-catalyzed

reactions. The SAP solution was prepared according to the manufacturer's instructions (Table 4-18).

Table 4-18 Reaction composition for EpiTYPER SAP treatment

Component	*Volume per single reaction*
Nuclease-free ddH_2O	1.7 µl
SAP mix	0.3 µl
Final Volume	*2 µl*

2 µl of the SAP mixture was added to each PCR reaction with the 96-channel pipetting robot MassARRAY Liquid Handler. The plate was sealed with adhesive PCR sealing foil (Thermo Scientific), centrifuged, and incubated as follows on a Veriti 384-well thermal cycler (Applied Biosystems):

Table 4-19 Cycling protocol for EpiTYPER SAP treatment

Cycle step	*Temperature*	*Time*
Dephosphorylation	37 °C	20 min
SAP heat inactivation	85 °C	5 min
Cooling	*4 °C*	*hold*

Reverse transcription and RNA base-specific cleavage

The PCR reaction was *in vitro* transcribed into RNA with the T7 RNA polymerase, which was guided to the amplified PCR-products by the introduced T7 promoter tag in the reverse primer. Transcription/Cleavage reaction was performed using a single mix containing according to the manufacturer's instructions.

Table 4-20 Reaction composition for EpiTYPER Transcription/Cleavage reaction

Component	*Volume per single reaction*
Nuclease-free ddH_2O	3.21 µl
T7 polymerase buffer (5×)	0.89 µl
Cleavage mix (**T mix**)	0.22 µl
DTT (100 mM)	0.22 µl
T7 R&DNA Polymerase (50 U/µl)	0.4 µl
RNaseA	0.06 µl
Final Volume	*5 µl*

5 µl of the transcription/cleavage mix and 2 µl of the SAP-treated PCR reaction were transferred into a new 386-well plate with the 96-channel pipetting robot MassARRAY Liquid Handler, sealed

with adhesive PCR sealing foil (Thermo Scientific),centrifuged, and incubated on a Veriti 384-well thermocycler C (Applied Biosystems) for three hours at 37°C.

Desalting the cleavage reaction

Because salt ions are co-vaporized when acquired during MALDI-TOF analysis they would be visible in the mass spectra. Since this would irritate the analysis of the mass spectra the reactions need to be desalted. For desalting of the transcription/cleavage mix 15 µl water was added to each reaction with the MassARRAY Liquid Handler (Matrix) followed by the addition of 6 mg CLEAN resin per reaction. This mix was rotated for slowly for 30 minutes and centrifuged for 10 min at 3000 rpm to collect the resin at the bottom of the wells.

Transfer on SpectroCHIP and acquisition

The SpectroCHIP holds the matrix-consisting of a crystallized acidic compound on, which the sample probes are spotted on. The analyte's solvent dissolves the matrix, which recrystallizes when the solvent evaporates. By that the analyte molecules are enclosed in the crystals.
The DNA samples are transferred on a SpectroCHIP with the Fusio™ Chip Module and analyzed with the MassARRAY Compact System MALDI-TOF MS (all Sequenom). The co-crystallized analyte is acquired with a laser while the matrix is predominantly ionized, protecting the DNA from the disruptive laser beam. Eventually the charge is transferred to the sample and charged ions are created, which are accelerated in a vacuum towards a detector that measures the particles' time of flight.

Data processing

Acquired data were processed with the EpiTYPER Analyzer software (v1.2, Sequenom). The MS is calibrated with a four point calibrant (Sequenom) containing 1479, 3004, 5044.4, and 8486.6 Da particles. Relative to this calibration, the accelerated analytes generate signal intensity (y-axis) versus mass (kDa, x-axis) plots. With the sequence of every amplicon known, the software can virtually process the sequence and predict the fragments from the *in vitro* transcription/RNase digestion and relocate CpG units. Each CpG-containing fragment is called a CpG unit, which can contain more than one CpG site. If more than one CpG sites are contained within one fragment, they get a sum methylation value since the software averages the methylation of the individual CpG sites. If expected and incoming information match, the signal intensities of the methylated and unmethylated DNA templates are compared and quantified. A normal calibrated system is able to measure fragments between a range of 1500 and 7000 Da. Fragments outside of this range and fragments whose mass peaks are overlapping with multiple other fragments cannot be analyzed.

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

EZ DNA Methylation Kit (single columns)	Zymo Research (#D5002)
EZ-96 DNA Methylation Kit (96-well format)	Zymo Research (#D5004)
Complete PCR Reagent Set MassCLEAVE T7 Kit SpectroCHIP® Arrays (incl. Resin)	Sequenom
384-well plate	Thermo Scientific (#TF0384)
Adhesive PCR Foil Seal	Thermo Scientific (#AB0558)

iPLEX: Genotyping polymorphisms

Similar to the mass difference between (hydroxy)methylated and unmethylated cytosines after bisulfite conversion, the iPLEX assay uses mass differences to distinguish alleles at single nucleotide polymorphisms (SNP) and short Insertions/Deletion. It uses single base extension (SBE) to incorporate mass-modified di-dNTP terminators, thereby creating allele-specific products with significant masses differences detectable by MALDI-TOF MS[271].

Starting with genomic DNA, the region of interest is amplified by PCR and subsequent SAP treatment neutralizes unincorporated dNTPs. Post-PCR SBE with mass-modified di-dNTP terminators is performed with a single primer binding in direct proximity to the allelic base. Depending on the zygosity of the analyzed region, either a unique SBE product (homozygous) or two distinct SBE products (heterozygous) are detected. In case of two SBE products, the mass difference is unique depending on the incorporated nucleosides and makes them distinguishable in MALDI-TOF MS spectra. Using the Assay Designer Software (Sequenom), multiplex reactions combining up to 36 individual SBE reactions can be designed.

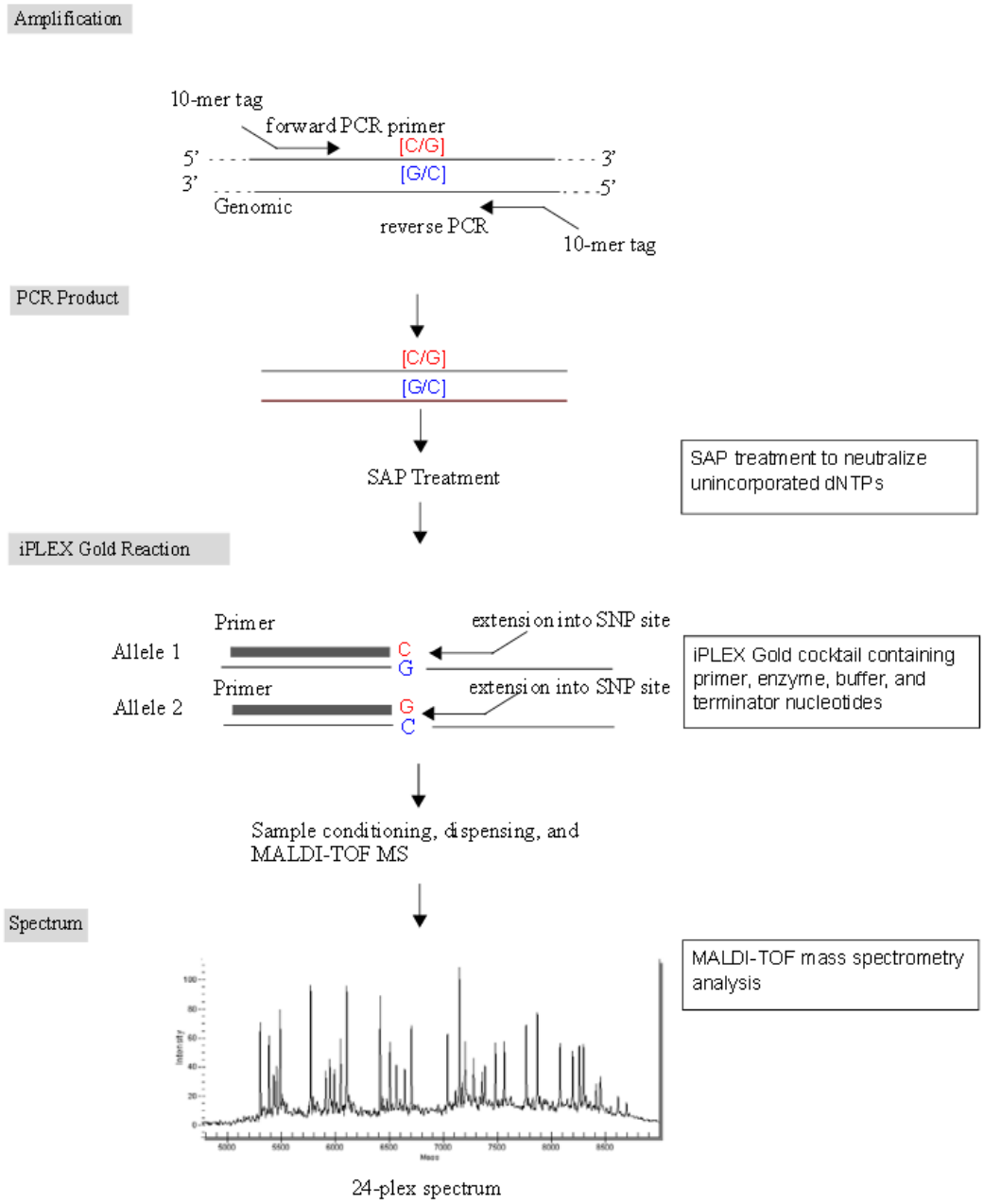


Figure 4-2 Schematic outline of the iPLEX technology

The region containing the SNP of interest is amplified by PCR rendering a fragment of ~100bp in size. Excess dNTPs are inactivated by shrimp Alkaline Phosphatase (SAP) treatment. The iPLEX Gold reaction elongates an oligonucleotide that binds with its 3' end to the SNP position by a single base into the SNP site. After a clean-up step the sample is spotted onto a 384 MassARRAY chip and subject to MALDI-TOF MS analysis. Up to 36 individual SNP positions can be assessed in one reaction by designing compatible capture and extension primer with the Sequenom Assay Design software. (Adopted from iPLEX Gold Application Guide)

PCR amplification

PCR Mastermixes were prepared according to the manufacturer's instructions and dispensed in 384-well plates (Thermo Scientific). The Mastermix (Table 4-21) contains multiplexed Capture Primer mixes according to assay design.

Table 4-21 Reaction composition for iPLEX PCR

Component	*Volume per single reaction*	*Final concentration*
Nuclease-free ddH_2O	1.850 µl	n. a.
PCR Buffer (10×)	0.625 µl	1.25x
$MgCl_2$ (25 mM)	0.325 µl	1.625 mM
dNTPs (25 mM each)	0.100 µl	500 µM
Primer mix (500 nM each)	1.000 µl	100 nM
Taq Polymerase (5 U/µl)	0.100 µl	0.5 U
Final Volume	*5 µl*	

10ng of the corresponding genomic DNA was added to each well.
The plate was sealed with adhesive PCR sealing foil (Thermo Scientific), centrifuged, and incubated in a Veriti 384-well thermal cycler (Applied Biosystems) (see Table 4-22).

Table 4-22 Cycling protocol for iPLEX PCR

Cycle step	*Temperature*	*Time*	*Number of cycles*
Initial denaturation	94 °C	15 min	1×
Denaturation	94 °C	20 sec	45×
Annealing	56 °C	30 sec	
Extension	72 °C	1 min	
Final extension	72 °C	3 min	1×
Cooling	*4 °C*	*hold*	

Shrimp alkaline phosphatase (SAP) treatment

Unincorporated nucleotides would disturb downstream applications and are enzymatically inactivated. Under alkaline conditions SAP removes phosphate groups from several substrates including deoxynucleotide triphosphates, rendering it unavailable for further polymerase catalyzed reactions. The SAP solution was prepared according to the manufacturer's instructions (see Table 4-23).

Table 4-23 Reaction composition for iPLEX SAP treatment

Component	*Volume per single reaction*
Nuclease-free ddH_2O	1.530 µl
SAP buffer (10x)	0.170 µl
SAP enzyme (1 U/µl)	0.300 µl
Final Volume	*2 µl*

2 µl of the SAP mixture was added to each PCR reaction with the 96-channel pipetting robot MassARRAY Liquid Handler. The plate was sealed with adhesive PCR sealing foil (Thermo Scientific), centrifuged, and incubated on a Veriti 384-well thermal cycler (Applied Biosystems) (see Table 4-24).

Table 4-24 Cycling protocol for iPLEX SAP treatment

Cycle step	*Temperature*	*Time*
Dephosphorylation	37 °C	40 min
SAP heat inactivation	85 °C	5 min
Cooling	*4 °C*	*hold*

iPLEX SBE reaction

For each assay, a specific extension primer binds next to the SNP site with its 3' end in direct proximity to the SNP position. The iPLEX Gold reaction cocktail contains modified nucleotides that prevented extension by more than one base. Depending on the plex level, the cocktail composition varies

Table 4-25 Reaction composition for iPLEX assay (2-18 plex)

Component	*Volume per single reaction*	*Final concentration (in 9 µl)*
Nuclease-free ddH_2O	0.7395 µl	n. a.
iPLEX Buffer Plus (10×)	0.2 µl	0.222x
iPLEX Termination Mix	0.1 µl	0.5 x
Primer mix (7 µM : 14 µM)	0.94 µl	0.73 µM : 1.46 µM
iPLEX enzyme	0.0205 µl	0.5 x
Final Volume	*2 µl*	

Table 4-26 Reaction composition for iPLEX assay (19-36 plex)

Component	*Volume per single reaction*	*Final concentration (in 9 µl)*
Nuclease-free ddH_2O	0.619 µl	n. a.
iPLEX Buffer Plus (10×)	0.2 µl	0.222x
iPLEX Termination Mix	0.2 µl	1 x
Primer mix (8 µM : 10 µM : 15 µM)	0.94 µl	0.52 µM : 1.04 µM : 1.57 µM
iPLEX enzyme	0.041 µl	1 x
Final Volume	*2 µl*	

Table 4-27 Cycling protocol for iPLEX SBE reaction

Cycle step	*Temperature*	*Time*	*Number of cycles*	
Initial denaturation	94 °C	30 sec	1×	
Denaturation	94 °C	5 sec	5×	40×
Annealing	52 °C	5 sec		
Extension	80 °C	5 sec		
Final extension	72 °C	3 min	1×	
Cooling	*4 °C*	*hold*		

Desalting the cleavage reaction

Because salt ions are co-vaporized when acquired during MALDI-TOF analysis they would be visible in the mass spectra. Since this would irritate the analysis of the mass spectra the reactions need to be desalted. For desalting of the transcription/cleavage mix 16 µl water was added to each reaction with the MassARRAY Liquid Handler (Matrix) followed by the addition of 6 mg CLEAN resin per reaction. This mix was rotated for slowly for 30 minutes and centrifuged for 10 min at 3000 rpm to collect the resin at the bottom of the wells.

Transfer on SpectroCHIP and acquisition

The SpectroCHIP holds the matrix-consisting of a crystallized acidic compound on, which the sample probes are spotted on. The analyte solvent dissolves the matrix, which recrystallizes when the solvent evaporates and encloses the analyte molecules.
The DNA samples are transferred on a SpectroCHIP with the Fusio™ Chip Module and analyzed with the MassARRAY Compact System MALDI-TOF MS (all Sequenom). The co-crystallized analyte is acquired with a laser while the matrix is predominantly ionized, protecting the DNA from the disruptive laser beam. Eventually the charge is transferred to the sample and charged ions are created, which are accelerated in a vacuum towards a detector that measures the particles' time of flight.

Data processing

Acquired data were processed with the TYPER software (v4.0, Sequenom). The MS is calibrated with a three point calibrant (Sequenom) containing 5044.4, 8486.6, and 9977.6 Da particles. Relative to this calibration the accelerated analytes generate signal intensity (y-axis) versus mass (kDa, x-axis) plots. With the sequence of every potential SBE product known, the software can virtually predict the fragments mass. The extension primers are designed to fall within a range of 4500 and 8500 Da after SBE. Due to the mass modification of the incorporated nucleotides during SBE, differentially extended extension primers theoretically separate well in the mass spectra allowing for SNP genotyping.

Selection criteria for SNPs used in linkage analysis

(i) SNPs needed to differ between C57BL/6 and BALB/c and be detected as homozygous in both strains (*to reduce the probability of being false-positive*)

(ii) SNPs enlisted in the dbSNP database and already evaluated for genotyping microarrays were preferable (*to reduce the probability of being false-positive*)

(iii) Only for genome-wide SNP panel: C57BL/6 SNP variants needed to be identical in less than six of the other 14 inbred strains (*to select for C57BL/6-specific SNPs*)

(iv) A minimum distance to adjacent SNPs or neighboring repeats exceeding 80 bp was preferable (*to provide for an unique sequence context*)

(v) On average, one informative SNP per 10 Mb (genome-wide analysis) or 0.5 Mb (fine mapping analysis), respectively, was selected (*to achieve an even coverage of the whole mouse genome*).

Default amplicon design settings (100 bp on average) for SNP capture PCR and the “high multiplex (36)” design option to facilitate maximum multiplex level. Assay design was generally performed with “constrict” settings to exclude repetitive sequences (*mm9* annotation available at www.repeatmasker.org) from assay design. Furthermore, in the case of closely located SNPs, the binding sites for the corresponding primers were confirmed to not intersect with SNPs annotated by the Mouse Genomes Project (release REL-1005) to ensure unbiased binding of both capture and SBE oligos.

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer’s instructions or at RT if not stated otherwise)

Complete PCR Reagent Set iPLEX Kit SpectroCHIP® Arrays (incl. Resin)	Sequenom
384-well plate	Thermo Scientific (#TF0384)
Adhesive PCR Foil Seal	Thermo Scientific (#AB0558)

4.3.1.15 **Methyl-CpG immunoprecipitation (MCIp)**

MCIp is based on an immobilizable MBD2-Fc fusion protein and enables rapid enrichment of CpG-methylated (mCpG) DNA as well as fractionation of DNA fragments according to their methylation level. DNA affinity of the antibody-like protein is dependent on the density of methylated CpGs and salt concentrations in the buffer. MCIp was performed with the EpiMark® Methylated DNA Enrichment Kit (NEB), which is based on a method established in our laboratory. The method allows for the fractionation of DNA fragments according to their mCpG level as well as the enrichment of methylated DNA fragments. Fractionation is achieved by washing the immobilized MBD2-Fc-bound DNA fragments with increasing NaCl concentration in the wash buffer. Murine genomic DNA fragments that exhibit low to intermediate methylation levels reside in fractions obtained from washing with buffer of 150-500 mM NaCl concentration. Remaining bound DNA fragments, which exhibit high methylation levels, can be totally recovered by washing with buffer of ≥ 1000 mM NaCl concentration. To enrich in general for methylated DNA fragments, all remaining bound DNA fragments after washing with buffer of 150 mM NaCl concentration can be recovered by incubating the immobilized DNA in nuclease-free ddH_2O at 65 °C for 15 min. Both protocols can be combined.

MCIp was performed according to the manufacturer's instructions. Fragments with low to intermediate methylation level were eluted by washing with 150 mM and 500 mM NaCl wash buffer. Highly methylated fragments were recovered by incubation with 52 µl nuclease-free ddH_2O at 65 °C for 15 min. Fractions obtained by increasing NaCl concentration in the wash buffer where desalted using the PCR Purification Kit (Qiagen) prior to subsequent assays.

Successful fractionation was confirmed by quantitative PCR (qPCR, at the imprinted *Snrpn* locus and other loci with known methylation levels. In somatic cells, imprinted regions exhibit allele-specific DNA methylation depending on the parental origin. For *Snrpn*, the paternal allele is, in contrast to the completely methylated maternal allele, free of DNA methylation. The paternal allele resides in the lower salt fractions (50% recovery) whereas the maternal allele was enriched in the fraction with the highest salt concentration (1000 mM) (50% recovery). In contrast, wash fractions with intermediate salt concentration are depleted of fragments from the *Snrpn* locus. Additional loci with known methylation pattern can be tested to validate successful enrichment of highly methylated DNA fragments (see Table 4-28).

Table 4-28 Control loci for MCIp enrichment by qPCR

Gene symbol	*Expected outcome in H_2O fraction (spleen, cerebellum)*
Gapdh (promoter)	Ubiquitously expressed → depletion of mCpGs
Myod1 (promoter)	Muscle-specifically expressed → enrichment of mCpGs
Snrpn (promoter)	Imprinted region → 50% enrichment of mCpGs
'empty'	CpG-free region → depletion of mCpGs
Isoc2b (promoter)	Only expressed in C57BL/6 → depletion of mCpGs in C57BL/6, enrichment in BALB/c

4.3.1.16 Hydroxymethylcytosine methylated DNA immunoprecipitation (hMeDIP)

Hydroxymethylation of CpGs (5hmC) occurs during the process of active DNA demethylation[66]. DNA fragments carrying this modification can be enriched using a specific antibody against 5hmC. The Covaris Ultra-Sonifier was used to guarantee reproducible statistically fragmentation results. 3.25 µg genomic DNA were diluted with TE (1x, pH 8.0) to a final concentration of 25 ng/µl and sonified according to the "200 bp" setting in Table 4-6. Successful sonication was assessed by gel electrophoresis (2% agarose, see section 4.3.1.4). Per immunoprecipitation, 1 µg of fragmented DNA was adjusted to 270 µl with TE (1x, pH 8.0). Diluted DNA was denatured at 95 °C for 10 min immediately followed by cooling on ice for 10 min to obtain single-stranded DNA fragments. After addition of 30 µl IP buffer (10x) and 1 µg of the 5hmC-specifc antibody the DNA antibody mix is incubated at 4 °C for 2-3 h on a rotator (~4-6 rpm). Per precipitation 30 µl of Dynabeads are batch-washed twice for 5 min at RT in BSA/PBS solution overhead on a shaker and recovered using a magnetic rack. The washed beads were resuspended in 30 µl IP buffer (1x) per 30 µl Dynabeads and added to the samples (30 µl each). After incubation at 4 °C for maximal 2 h on a rotator (~4-6 rpm) the beads are recovered using a magnetic rack and washed three times with 700 µl IP buffer (1x). Elution of the trapped DNA fragments was achieved by resuspension and incubation for 3 h at 50 °C in 250 µl Proteinase K digestion buffer and 3.5 µl Proteinase K. Trapping the beads on a magnetic rack allows the recovery of the DNA fragment-containing supernatant, which is then purified as well as 10% of the initial input DNA using QIAquick PCR Purification Kit (Qiagen) with 30 µl prewarmed EB buffer (55 °C, applied twice) for elution.Successful enrichment was confirmed by qPCR at other loci with known methylation levels. The choice depended on the cell type source of the processed genomic and is stated in detail in the results' sections.

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

Sodium phosphate (monobasic)	2 M	276 g	Anhydrous Sodium phosphate monobasic (Sigma-Aldrich, #S8282)
		→ Ad 1 l	nuclease-free ddH_2O
Sodium phosphate (dibasic)	2 M	284 g	Anhydrous di-Sodium hydrogen phosphate (Merck Millipore, #1065590500)
		→ Ad 1 l	nuclease-free ddH_2O
Sodium phosphate buffer (pH 7.0)	1 M	39 ml	Sodium phosphate (monobasic)
		61 ml	Sodium phosphate (dibasic)
		100 ml	nuclease-free ddH_2O

IP buffer (10x)	100 mM	100 ml	Sodium phosphate buffer (1 M, pH7.0)
	1.4 M	81.8 g	Sodium chloride (AppliChem, #A3597)
	0.5%	5 ml	Triton™ X-100 (Sigma-Aldrich, #T8787)
		→ Ad 1 l nuclease-free ddH_2O	
IP buffer (1x)	1x	1 ml	IP buffer (10x)
		9 ml	TE (1x, pH 8.0)
TE (1x, pH 8.0)	10 mM	10 ml	Tris/HCl (1 M, pH 8.0)
	1 mM	2 ml	EDTA/NaOH (0.5 M, pH 8.0)
		→ Ad 1 l nuclease-free ddH_2O	
Tris/HCl (pH 8)	1 M	121.1 g	Tris ultrapure (AppliChem, #A1086)
		→ Adjust with HCl to pH 8	
		→ Ad 1 l nuclease-free ddH_2O	
HCl	37%		Hydrochloric acid (fuming) (Merck/Millipore, #100317)
EDTA/NaOH (pH 8)	0.5 M	186.1 g	Ethylenediaminotetraacetic acid disodium salt dehydrate (AppliChem, #A2937)
		→ Dissolve in 500 ml nuclease-free ddH_2O	
		→ Adjust with NaOH to pH 8.0	
		→ Ad 1 l nuclease-free ddH_2O	
NaOH	1 M		Sodium hydroxide solution (Merck/Millipore, #109137)
BSA	1%	100 mg	Bovine serum albumin (Sigma-Aldrich, #A4503)
		→ Ad 10 ml nuclease-free ddH_2O	
BSA/PBS solution	0.1%	1 ml	BSA (10 mg/ml)
		9 ml	Phosphate-buffered saline (Sigma-Aldrich, #P4417)
Proteinase K digestion buffer	50 mM	2 ml	Tris/HCl (1 M, pH 8.0)
	10 mM	2 ml	EDTA/NaOH (0.5 M, pH 8.0)
	0.5%	5 ml	SDS (10%)
		→ Ad 100 ml nuclease-free ddH_2O	

SDS	10%	100 g	Sodium dodecyl sulfate (Sigma-Aldrich, #L4390)
		→ Ad 1 l	nuclease-free ddH_2O
Proteinase K (-20°C)	20 mg/ml	25 mg	Proteinase K (Roche, #03115836001)
		1.25 ml	nuclease-free ddH_2O
Dynabeads (4 °C)			Dynabeads® Protein A (Life Tech., #10008D)

4.3.1.17 Chromatin immunoprecipitation (ChIP)

Tissue preparation and mincing

ChIP samples were prepared from fresh tissue. DNA and its associated proteins (chromatin) are crosslinked by formaldehyde fixation. To obtain homogenous fixation results, tissue samples need to be dissociated to obtain single cell suspensions or suspensions close to single cells, respectively. Tissue samples were cut into small pieces in sterile culture dishes filled with cold PBS/PMSF on ice. For spleen, samples were minced by pressing the tissue through cell strainer in presence of cold PBS/PMSF. For cerebellum, the pieces were transferred together with 200 µl PBS/PMSF into a fresh 1.5 ml tube and homogenized with a mini pestle (pointed shape fits into bottom of 1.5 ml tube). The cell suspensions was centrifuged at 4 °C for 10 min at 365 xg and the pellet resuspended in 25 ml cold PBS/PMSF. Cells were washed twice with 15 ml PBS/PMSF and collected each time by centrifugation at 4 °C for 8 min AT 365xg. Afterwards, the cells were aliquoted. For spleen, cells were counted in a hemocytometer and the concentration was adjusted to 10 Mio cells per 10 ml PBS/PMSF. For cerebellum, one half of a cerebellum roughly comprises 10 Mio cells (according to tissue weight); two minced quarter portions of a cerebellum were combined and resuspended in 10 ml cold PBS/PMSF.

Crosslinking with formaldehyde

Formaldehyde causes crosslinking of molecules within the range of 2 Angstrom. Batch crosslinking of multiples of 10 Mio cells required adjustment of final PBS/PMSF volume. Per 10 ml cell suspension (10 Mio cells), 270 µl formaldehyde (37%) was added, mixed by inversion, and incubated at RT for 10 min. During the incubation time, the tube was inverted several times to prevent settling of cells. Crosslink reaction was stopped by addition of 1/20 Vol Glycine (2.625 M) and 5 min incubation at RT. Again, the tubes were inverted several times during incubation. The fixed cells were harvested by centrifugation at 4 °C for 10 min at 365 xg and the supernatant discarded. The pellet was washed twice with cold PBS/PMSF as before. Pellets can be stored at -80 °C up to months after snap freezing in N_2 (l).

Cell and nuclei lysis

Crosslinked chromatin was isolated from fixed cells after lysis. Per 10 Mio cells, the pellet was resuspended in 250 µl of Suspension Buffer and lysed on ice for 10 min by addition of 250 µl of Cell Lysis Buffer and mixing. The lysate was transferred to a new 1.5 ml tube and centrifuged at 4 °C for 5 min at 700 xg to pellet the nuclei. The nuclei were then lysed by resuspension in 450 µl Nuclear Lysis Buffer. The lysed nuclei can be stored in Nuclear Lysis Buffer at -80 °C up to months after snap freezing in N_2 (l).

Fragmentation of chromatin

The chromatin was sonicated with a Branson Sonifier 250 according to section 4.3.1.3. After sonication, the lysate was cleared from remaining cell debris by centrifugation at 4 °C for 5 min at 13000g. To evaluate the fragmentation size, 30 µl of the sonicated and cleared lysate was mixed with 14 µl NaCl (5 M) and 26 µl L2 (without PMSF and inhibitors) and incubated o/n at 65 °C to reverse the crosslinking. After digestion of RNAs with 2.3 µl RNase A for 1 h at 37 °C, the sample was purified with the QIAquick PCR Purification Kit according the manufactor's instructions. The sample was eluted in 30 µl EB buffer and size-separated by agarose gel electrophoresis (2% agarose gel, 50 bp ladder) according to section 4.3.1.4. The remaining lysate can be stored at -80 °C up to months.

Immunoprecipitation

For each immunoprecipitation (IP), 80 µl of cleared lysate (chromatin of ~2 Mio cells) were used. For normalization purposes, per sample 4 µl of lysate (5% of IP) were retained as input. Before the actual IP, all IP aliquots from one chromatin batch were subject to pre-clearing. Per IP, 50 µl pre-clearing beads were washed twice by repeated centrifugation at RT for 1 min at 600 xg and resuspension in TE (1x) buffer. The washed beads were diluted to the initial volume with Dilution buffer (DB, w/o inhibitors). After addition of 25 µl BSA (20%) and 4 µl Glycogen per ml pre-clearing beads, the beads were incubated on a rotor for at least 2 h at RT. For pre-clearing, the chromatin batch was diluted with 1.5 Volume DB (with inhibitors, results in 200 µl diluted lysate per IP) and incubated with 50 µl pre-clearing beads per IP on a rotor at 4 °C for 2 h. To recover the pre-cleared lysate, the mix was centrifuged at 4 °C for 5 min at 13000 xg and the supernatant transferred to a new 1.5 ml tube.

Per IP, 200 µl of pre-cleared lysate were transferred to 0.5 ml tubes and after addition of 2.4 µg of the desired antibody the mix was incubated o/n on a rotor at 4 °C. In parallel to the o/n incubation of the chromatin antibody mixture, 40 µl Protein A-coated beads per individual IP for the subsequent pull down of the antibody-bound chromatin fragments were prepared. The beads were washed twice by repeated centrifugation at RT for 1 min at 600 xg and resuspension in TE (1x) buffer. The washed beads were diluted to the initial volume with Dilution buffer (DB, w/o inhibitors). After addition of 25 µl BSA (20%) and 4 µl Glycogen per ml Protein A-coated beads, the beads were incubated o/n on a rotor at 4 °C. To pull down the chromatin-antibody complexes, 40 µl of blocked beads were added to each IP sample and incubated on a rotor for 4 h at 4 °C. The beads with the bound chromatin-antibody complex were centrifuged at 4 °C for 5 min at 1100xg. Unspecifically bound chromatin was washed away by repeated washing steps with

increasing stringency. The beads were resuspended twice in 400 µl of each of the three washing buffers (WBI-III) and incubated on a rotor for 5 min at RT after each resuspension. Final washing step was performed by three times resuspension in 400 µl TE (1x) buffer. To elute the bound chromatin-antibody complexes, the beads were incubated with 110 µl Elution Buffer (EB) for 20 min at RT (shaking every 5 min). After centrifugation at 4 °C for 1 min at 600xg, 100 µl supernatant were transferred to a new 1.5 ml tube. Elution was repeated with additional 110 µl Elution Buffer and incubation for 10 min and after centrifugation, another 100 µl supernatant was combined with the first 100 µl eluate. Crosslinking of the eluted chromatin fragments was reversed by incubation with 5 µl Proteinase K (20 µg/µl) o/n at 65 °C. In parallel, 4 µl input were diluted to 200 µl with Elution Buffer and also subject to Proteinase K treatment.

Clean up

The DNA fragments released from the recovered chromatin as well as from the input sample by Proteinase K treatment were purified by RNase A digestion (0.33 µg/µl, 2 h at 37 °C) and subsequent clean up with the QIAquick PCR purification kit according to the manufacturer's instructions. DNA is loaded onto the column by centrifugation at 6700xg and eluted after 2 min incubation at RT with 52.5 µl prewarmed Elution Buffer (55 °C, part of the kit).

Validation by qPCR

To validate specific enrichment by immunoprecipitation, 10 µl of the eluate were diluted with Elution Buffer for subsequent qPCR according to section 4.3.1.10. The following loci were tested for specific enrichment.

Table 4-29 Control loci for ChIP validation by qPCR

Gene symbol	*Expected outcome (spleen, cerebellum)*
Gapdh (promoter)	Ubiquitously expressed → enrichment of active histone marks
Myod1 (promoter)	Muscle-specifically expressed → depletion of active histone marks
'Intergenic'	Depletion of active histone marks
Isoc2b (promoter)	Only expressed in C57BL/6 → enrichment of active histone marks in C57BL/6, depletion in BALB/c

40 µl were processed for Next Generation Sequencing-compatible library preparation (see section 4.3.1.18).

Required reagents and materials:

(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

Formaldehyde	37%	Formaldehyde (Merck Millipore, #8187081000)

Glycine	2.625 M	9.85 g	Glycine (Roth, #3908.1)
		→ Ad 50 ml nuclease-free ddH_2O	
BSA	20%	2 g	Bovine serum albumin (Sigma-Aldrich, #A4503)
		→ Ad 10 ml nuclease-free ddH_2O	
Glycogen	5 mg/ml		Glycogen (Life Techn., #AM9510)
HEPES/KOH (pH 7.9)	1 M	238.3 g	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (Sigma-Aldrich, #H3357)
		→ Adjust with KOH to pH 7.9	
		→ Ad 1 l nuclease-free ddH_2O	
KCl	2 M	149.1 g	Potassium chloride (Merck Millipore, #1049360250)
		→ Ad 1 l nuclease-free ddH_2O	
EDTA/NaOH (pH 8)	0.5 M	186.1 g	Ethylenediaminotetraacetic acid disodium salt dehydrate (AppliChem, #A2937)
		→ Dissolve in 500 ml nuclease-free ddH_2O	
		→ Adjust with NaOH to pH 8.0	
		→ Ad 1 l nuclease-free ddH_2O	
PMSF	100 mM	1.74 g	Phenylmethanesulfonyl fluoride (Sigma-Aldrich, #P7626)
		→ Ad 100 ml Isopropanol (p.a.) (Merck Millipore, #1096342511)	
NaCl	5 M	29.2 g	Sodium chloride (AppliChem, #A3597)
		→ Ad 100 ml nuclease-free ddH_2O	
Tris/HCl (pH 7.4)	1 M	121.1 g	Tris ultrapure (AppliChem, #A1086)
		→ Adjust with HCl to pH 7.4	
		→ Ad 1 l nuclease-free ddH_2O	
TE (1x, pH 8.0)	10 mM	10 ml	Tris/HCl (1 M, pH 8.0)
	1 mM	2 ml	EDTA/NaOH (0.5 M, pH 8.0)
		→ Ad 1 l nuclease-free ddH_2O	

SDS	20%	20 g	Sodium dodecyl sulfate (Sigma-Aldrich, #L4390)
		→ Ad 100 ml nuclease-free ddH_2O	
Empigen BB	30%	3 ml	Empigen BB detergent (100%) (Sigma-Aldrich, #45165)
		7 ml	nuclease-free ddH_2O
Triton X-100	10%	1 ml	Triton X-100 (100%) (Sigma-Aldrich, #T8787)
		9 ml	nuclease-free ddH_2O
LiCl	2.5 M	10.6 g	Lithium chloride (Merck Millipore, # 438002)
		→ Ad 100 ml nuclease-free ddH_2O	
Deoxycholate	10%	1g	Sodium deoxycholate (Sigma-Aldrich, #D6750)
		→ Ad 10 ml nuclease-free ddH_2O	
$NaHCO_3$	1 M	84 g	Sodium hydrogen carbonate (Merck Millipore, #1063290500)
		→ Ad 1 l nuclease-free ddH_2O	
Cell Buffer Mix (CBM)	10 mM	1 ml	HEPES/KOH (1 M, pH 7.9)
	85 mM	4.25 ml	KCl (2 M)
	1 mM	200 µl	EDTA/NaOH (500 mM, pH 8.0)
		91.55 ml	nuclease-free ddH_2O
immediately before use, add the following inhibitors per ml CBM:			
	1 mM	10 µl	PMSF (100 mM)
		20 µl	Inhibitor Mix (50x) (Roche, #04693132001)
Suspension Buffer		900 µl	CBM
		100 µl	nuclease-free ddH_2O
Cell Lysis Buffer		900 µl	CBM
	1%	100 µl	NP-40 (10%) (Roche, #11332473001)

Nuclear Lysis Buffer	50 mM	5 ml	Tris/HCl (1 M, pH 7.4)
	1%	5 ml	SDS (20%)
	0.5%	1.667 ml	Empigen BB (30%)
	10 mM	2 ml	EDTA/NaOH (500 mM, pH 8.0)
		83.33 ml	nuclease-free ddH_2O

immediately before use, add the following inhibitors per ml Nuclear Lysis Buffer:

1 mM	10 µl	PMSF (100 mM)
	20 µl	Inhibitor Mix (50x) (Roche, #04693132001)

Dilution Buffer (DB)	20 mM	2 ml	Tris/HCl (1 M, pH 7.4)
	100 mM	2 ml	NaCl (5 M)
	2 mM	400 µl	EDTA/NaOH (500 mM, pH 8.0)
	0.5%	5 ml	Triton X-100 (10%)
		87.6 ml	nuclease-free ddH_2O

immediately before use, add the following inhibitors per ml Dilution Buffer:

1 mM	10 µl	PMSF (100 mM)
	20 µl	Inhibitor Mix (50x) (Roche, #04693132001)

Wash Buffer

WBI	20 mM	2 ml	Tris/HCl (1 M, pH 7.4)
	150 mM	3 ml	NaCl (5 M)
	0.1%	500 µl	SDS (20%)
	1%	10 ml	Triton X-100 (10%)
	2 mM	400 µl	EDTA/NaOH (500 mM, pH 8.0)
		84.1 ml	nuclease-free ddH_2O
WBII	20 mM	2 ml	Tris/HCl (1 M, pH 7.4)
	500 mM	10 ml	NaCl (5 M)
	1%	10 ml	Triton X-100 (10%)
	2 mM	400 µl	EDTA/NaOH (500 mM, pH 8.0)
		84.1 ml	nuclease-free ddH_2O
WBIII	10 mM	1 ml	Tris/HCl (1 M, pH 7.4)
(protect from light!)	250 mM	10 ml	LiCl (2.5 M)
	1%	10 ml	NP-40 (10%)
	1%	10 ml	Deoxycholat (10%)
	1 mM	200 µl	EDTA/NaOH (500 mM, pH 8.0)
		68.8 ml	nuclease-free ddH_2O

Elution Buffer	0.1 M	500 µl	$NaHCO_3$ (1 M)
	1%	250 µl	SDS (20%)
		4.25 ml	nuclease-free ddH_2O

4.3.1.18 Library preparation for next generation sequencing

Next generation sequencing (NGS) libraries were prepared to be sequenced on Illumina platforms (GenomeAnalyzer, HiSeq1000). In general, DNA loBind tubes (Eppendorf) were used. Library preparation was subsequently enhanced during the time of this thesis, resulting in three different preparation protocols (NGS library protocol I, II & III). The protocol used for the individual library preparations will be indicated in the results' section. Library preparation of RNA samples for NGS is described separately in section 4.3.2.3.

NGS library protocol I

Using the NEBNext DNA Library Prep Reagent Set for Illumina (NEB) and the Oligo Only Kit (Illumina) immunoprecipitation (IP) samples were first subject to end repair to convert overhangs of double stranded DNA fragments into phosphorylated blunt ends (see Table 4-30).

Table 4-30 Reaction composition for end repair (NGS library protocol I)

Component	*Volume*	*Final concentration*
Nuclease-free ddH_2O	0.8 µl	n.a.
T4 DNA Ligase buffer with 10 mM ATP (10x)	5 µl	1x (1 mM ATP)
dNTP mix (10 mM each)	2 µl	0.4 mM each
T4 DNA polymerase	1 µl	n.a.
Klenow DNA Polymerase	0.2 µl	n.a.
T4 Polynucleotide Kinase	1 µl	n.a.
IP DNA sample*	40 µl	n.a.
Final Volume	***50 µl***	

* For input: 10 ng, add 40 µl nuclease-free ddH_2O

The reaction mix was incubated at 20 °C for 30 min in a thermal cycler and subsequently purified using the QIAquick PCR Purification Kit (Qiagen) with an advanced protocol: After binding the sample to the column, the flow-through was transferred back onto the column and centrifuged again. After washing with 750 µl PE buffer, the column was transferred to a new collection tube and the empty column centrifuged at RT for 2 min at 13000xg. The column was transferred to a new 1.5 ml tube and the residual PE buffer was aspirated from the purple ring inside the column. After air-drying the column for 1 min, 36 µl of prewarmed EB buffer (50 °C) was applied to the column center and incubated at RT for 5 min. To recover the eluate, the column was centrifuged at RT for 1 min at 9500xg. Residual EB buffer on the purple ring as well as the flow-through was transferred back onto the column that was centrifuged again. The final volume of the eluate was 34 µl.

The next step introduced an A-overhang to the 3' end of the phosphorylated blunt ends of the DNA fragments and was required for subsequent ligation of adapters (see Table 4-31).

Table 4-31 Reaction composition for A-overhang introduction (NGS library protocol I)

Component	*Volume*	*Final concentration*
Klenow Buffer (10x)	5 µl	1x
dATP (1 mM)	10 µl	200 µM
Klenow fragment (3' to 5' exo minus)	1 µl	n.a.
End-repaired IP sample	34 µl	n.a.
Final Volume	***50 µl***	

The reaction mix was incubated at 37 °C for 30 min in a thermal cycler and subsequently purified using the QIAquick MinElute PCR Purification Kit (Qiagen) according to the advanced protocol described above. Elution was performed with 12 µl prewarmed EB buffer (50 °C) to recover 10 µl of the eluate.

The next step ligated adapters to the end of the DNA fragments to allow later hybridization to an Illumina Next Generation Sequencing flow cell.

Table 4-32 Reaction composition for adapter ligation (NGS library protocol I)

Component	*Volume*	*Final concentration*
Quick DNA ligase buffer (2x)	15 µl	1x
Adapter (1:10 dilution, 1.5 pmol/µl)	1 µl	50 µM
Quick DNA ligase	1 µl	n.a.
Nuclease-free ddH_2O	3 µl	n.a.
IP sample (with A-overhang)	10µl	n.a.
Final Volume	***30 µl***	

The reaction mix was incubated at 20 °C for 15 min in a thermal cycler and subsequently purified using the QIAquick MinElute PCR Purification Kit (Qiagen) according to the advanced protocol described above. Elution was performed with 12 µl prewarmed EB buffer (50 °C) to recover 10 µl of the eluate.

The next step removed excess adaptors and also includes a size selection. For each sample, a separate 2% agarose gel (50 ml) with small, thin pockets was prepared with TAE (1x) buffer. To each of the samples, 3 µl DNA light loading dye (6x) was added. In parallel, 0.5 µg Generuler 50 bp ladder was run on both sides of the sample with one empty lane in between (60 V, 90 min. With a clean razor, the region of 175-225 bp as indicated by the ladder was excised from the gel. The DNA fragments trapped in the excised gel were extracted using the QIAquick Gel Extraction

Kit (Qiagen) with an advanced protocol. The gel pieces was weighted and transferred to 6 Vol QG buffer and incubated at RT on a rotor until gel is dissolved completely (~ 10 min). The dissolved gel sample was mixed with 2 Vol Isopropanol by pipetting. The mix was loaded onto a column by centrifugation at RT for 30 sec at 13000xg. The flow-through was loaded again onto the column and the column centrifuged until the entire volume has passed through the column. The flow-through was discarded and the column washed with 500 µl QC buffer. The column was washed twice with 750 µl PE buffer and centrifuged at RT for 45 sec at 13000xg. The empty column was centrifuged again at RT for 1 min at 13000xg and transferred to a new 1.5 ml tube. After residual PE buffer was aspirated from the purple ring, the column was air-dried for 1 min. To elute the bound DNA fragments, the column was incubated with 38 µl prewarmed EB buffer (50 °C) for 3 min at RT and subsequently centrifuged at RT for 1 min at 13000xg. The column was rotated by 180 ° within the 1.5 ml tube and centrifuged again. 36 µl elute were recovered.
Following excess adapter removal and size selection, the adapter-modified DNA fragments were enriched by PCR according to Table 4-33 & Table 4-34.

Table 4-33 Reaction composition for PCR enrichment (NGS library protocol I)

Component	*Volume*	*Final concentration*
Phusion HF buffer (5x)	10	1x
dNTP mix (10 mM each)	1.5 µl	300 µM each
PCR primer 1.1 (25 µM)	1 µl	0.5 µM
PCR primer 2.1 (25 µM)	1 µl	0.5 µM
Phusion polymerase	0.5	n.a.
Size-selected DNA fragments	36 µl	n.a.
Final Volume	***50 µl***	

Table 4-34 Cycling protocol for PCR enrichment (NGS library protocol I)

Cycle step	*Temperature*	*Time*	*Number of cycles*
Initial denaturation	98 °C	30 sec	1×
Denaturation	98 °C	10 sec	
Annealing	65 °C	30 sec	18×
Extension	72 °C	30 sec	
Final extension	72 °C	5 min	1×
Cooling	*4 °C*	*hold*	

The PCR-enriched DNA fragments were purified with the QIAquick MinElute PCR Purification Kit (Qiagen) according to the advanced protocol described above. To increase purity and yield, the column was washed twice with PE buffer and elution was performed in two steps with 10 µl prewarmed EB buffer (50 °C) each and 3 min incubation at RT. The final volume of the eluate was 16 µl. To validate the successful enrichment of the size-selected DNA fragments, 1 µl of the library sample was analyzed on a Bioanalyzer DNA HS assay.

The library samples were further processed for NGS data acquisition by the KFB (Biopark, Regensburg).

NGS library protocol II

Using Enzymatics enzymes (NEB) and NEXTflex indexed adapters (Bioo Scientific) the IP sample was first subject to end repair to convert overhangs of double stranded DNA fragments into phosphorylated blunt ends (see Table 4-35).

Table 4-35 Reaction composition for end repair (NGS library protocol II)

Component	*Volume*	*Final concentration*
Nuclease-free ddH_2O	3.34 µl	n.a.
T4 DNA ligase buffer (10x)	5 µl	1x
dNTP mix (10 mM each)	1 µl	0.2 mM each
T4 DNA polymerase	0.3 µl	n.a.
Klenow Fragment	0.06 µl	n.a.
T4 Polynucleotide Kinase	0.3 µl	n.a.
IP DNA sample*	40 µl	n.a.
Final Volume	***50 µl***	

* For input: 10 ng, add 40 µl nuclease-free ddH_2O

The reaction mix was incubated at 20 °C for 30 min in a thermal cycler and subsequently purified using the QIAquick MinElute PCR Purification Kit (Qiagen) with an advanced protocol: After binding the sample to the column by centrifugation at RT for 30 sec at 9500xg, the flow-through was transferred back onto the column and centrifuged again. After washing twice with 750 µl PE buffer (centrifugation at RT for 45 sec at 13000 g), the column was transferred to a new collection tube and the empty column centrifuged at RT for 2 min at 13000xg. The column was transferred to a new 1.5 ml tube and the residual PE buffer was aspirated from the purple ring inside the column. After air-drying the column for 1 min, 10.25 µl of prewarmed EB buffer (50 °C) was applied twice to the column center and incubated at RT for 5 min. To recover the eluate, the column was centrifuged at RT for 1 min at 9500xg. Residual EB buffer on the purple ring as well as the flow-through was transferred back onto the column that was centrifuged again. The final volume of the eluate was 18 µl.

The next step introduced an A-overhang to the 3'-end of the phosphorylated blunt ends of the DNA fragments and was required for subsequent ligation of adapters (see Table 4-36).

Table 4-36 Reaction composition for A-overhang introduction (NGS library protocol II)

Component	*Volume*	*Final concentration*
Blue Buffer (10x)	2 µl	1x
dATP (10 mM)	0.4 µl	195 µM
Klenow fragment (3' to 5' exo minus)	0.2 µl	n.a.
End-repaired IP sample	18 µl	n.a.
Final Volume	***20.6 µl***	

The reaction mix was incubated at 37 °C for 30 min in a thermal cycler.
The next step ligated adapters to the end of the DNA fragments to allow later hybridization to an Illumina Next Generation Sequencing flow cell. Beforehand, 0.5 µl of indexed NEXTflex adapters (Bioo Scientific) were diluted 1:20 with 1x T4 DNA ligase buffer (10x T4 DNA ligase buffer diluted 1:10 with nuclease-free ddH_2O). Then, 20.6 µl of the prepared IP sample were mixed on ice with 22.6 µl Rapid Ligation Buffer (2x). Subsequently, 1 µl of the diluted adapter mix were added on ice, as well as 1 µl of T4 DNA Ligase. The reaction mix was mixed on ice and incubated at 30 °C for 10 min. Finally, 4.8 µl nuclease-free ddH_2O were added to adjust the volume to 50 µl.

Excess adapters were removed by clean up with AMPure XP magnetic beads (equilibrated for 30 min at RT on a rotor) at a ratio 1:1.1. 55 µl beads were added to the ligation mix (50 µl), vortexed (speed 6) two times, inverted and vortexed again and then incubated at RT for 5 min. The tube was placed on a magnetic particle concentrator (MPC) for 2 min to remove the supernatant. The trapped beads were washed twice with 500 µl Ethanol (70%, prepared freshly). To remove residual Ethanol, the beads were air-dried at 37 °C for 2-5 min. The bound DNA fragments were eluted by incubation with 50 µl nuclease-free ddH_2O at RT for 5 min. After centrifugation at RT for 1 sec at 13000xg, the tubes were placed on a MPC for 2 min and the supernatant was transferred to a new 1.5 ml tube for a second round of clean up. Elution volume of the second clean up depended on the IP procedure; ChIP samples were eluted with 37.5 µl nuclease-free ddH_2O and MCIp samples with 25 µl, respectively.

Following excess adapter removal, the adapter-modified DNA fragments were enriched by PCR according to the following tables.

Table 4-37 Reaction composition for PCR enrichment of ChIP samples (NGS library protocol II)

Component	*Volume*	*Final concentration*
Phusion HF buffer (5x)	10	1x
dNTP mix (10 mM each)	1.5 µl	300 µM each
BiooPrimer1 (100 µM)	0.25 µl	0.5 µM
BiooPrimer2 (100 µM)	0.25µl	0.5 µM
Phusion polymerase	0.5	n.a.
Adapter-ligated ChIP sample	37.5 µl	n.a.
Final Volume	***50 µl***	

Table 4-38 Reaction composition for PCR enrichment of MCIp samples (NGS library protocol II)

Component	*Volume*	*Final concentration*
Phusion HF buffer (5x)	10	1x
Betaine (5 M)*	13	1.3 M
dNTP mix (10 mM each)	1.0 µl	200 µM each
BiooPrimer1 (100 µM)	0.25 µl	0.5 µM
BiooPrimer2 (100 µM)	0.25µl	0.5 µM
Phusion polymerase	0.5	n.a.
Adapter-ligated ChIP sample	25 µl	n.a.
Final Volume	***50 µl***	

* to dissolve GC-rich sequences

Table 4-39 Cycling protocol for PCR enrichment of ChIP samples (NGS library protocol II)

Cycle step	*Temperature*	*Time*	*Number of cycles*
Initial denaturation	98 °C	30 sec	1×
Denaturation	98 °C	10 sec	18×
Annealing	65 °C	30 sec	
Extension	72 °C	30 sec	
Final extension	72 °C	5 min	1×
Cooling	*4 °C*	*hold*	

Table 4-40 Cycling protocol for PCR enrichment of MCIp samples (NGS library protocol II)

Cycle step	*Temperature*	*Time*	*Number of cycles*
Initial denaturation	98 °C	2 min*	1×
Denaturation	98 °C	30 sec*	18×
Annealing	65 °C	30 sec	
Extension	72 °C	45 sec*	
Final extension	72 °C	5 min	1×
Cooling	*4 °C*	*hold*	

* to dissolve GC-rich sequences

Excess primers were removed by two times clean up with AMPure XP magnetic beads (equilibrated for 30 min at RT on a rotor) at a ratio 1:1.1. 55 µl beads were added to the PCR reaction mix (50 µl) and processed as before. After the first clean up, the bound DNA fragments were eluted with 50 µl nuclease-free ddH_2O, after the second clean up with 12 µl.

Size selection of the amplified DNA fragments was performed gel electrophoresis and excision of the desired fragment size. The eluate (12 ml) was mixed with 2 µl loading dye (6x) and loaded together with 1 µg 50 bp DNA ladder (NEB) (separated by one lane) on a 2% low-melting agarose gel (50 ml, 1x TAE). The gel was run for 60 min at 100 V with pre-cooled running buffer. Using a clean razor, fragments in the range of 200-350 bp were excised and subject to DNA extraction with the QIAquick Gel Extraction Kit. The gel piece was weighted and dissolved in 6 Volumes OC buffer by incubation at RT (approx. 5 min). The dissolved gel sample was mixed with 2 Volumes Isopropanol and loaded twice onto the column by centrifugation at RT for 30 sec at 9500xg. The following clean up procedure was similar to NGS library protocol I. The bound DNA fragments were eluted with 13 µl prewarmed EB buffer (50 °C). The final volume of the eluate was 11 µl. To validate the successful enrichment of the size-selected DNA fragments, 1 µl of the library sample was analyzed on a Bioanalyzer DNA HS assay.
The library samples were further processed for NGS data acquisition by the KFB (Biopark, Regensburg).

NGS library protocol III

This protocol is similar to NGS protocol II except for the PCR enrichment and size selection procedure. The protocol stated was established for ChIP and MCIp samples. For this work, NGS library protocol III was only performed on a BAC pool sample, which was treated similar as ChIP samples. Relevant alterations to the established protocol are stated accordingly.
The pool of BACs was fragmented by Focused Ultrasonication (Covaris, see section 4.3.1.3) using the 200 bp setting for genomic DNA in 130 µl TE (0.09x). The volume was reduced by lyophylization to ~30 µl and the concentration determined with the Qubit HS DNA assay. In parallel, 130 µl TE (0.09x) were also lyophilized to the same degree for later size selection with the Caliper LapChip approach.

Following excess adapter removal, the adapter-modified DNA fragments were enriched by PCR in two rounds; round one was performed according to the following tables. In between the two PCR reactions, size selection using the Caliper LabChip approach (PerkinElmer) was performed.

Table 4-41 Reaction composition for PCR enrichment of ChIP samples (NGS library protocol III)

Component	*Volume*	*Final concentration*
Phusion HF buffer (5x)	10	1x
dNTP mix (10 mM each)	1.5 µl	300 µM each
BiooPrimer1 (100 µM)	0.25 µl	0.5 µM
BiooPrimer2 (100 µM)	0.25µl	0.5 µM
Phusion polymerase	0.5	n.a.
Adapter-ligated ChIP sample	37.5 µl	n.a.
Final Volume	***50 µl***	

Table 4-42 Reaction composition for PCR enrichment of MCIp samples (NGS library protocol III)

Component	*Volume*	*Final concentration*
Phusion HF buffer (5x)	10	1x
Betaine (5 M)*	13	1.3 M
dNTP mix (10 mM each)	1.0 µl	200 µM each
BiooPrimer1 (100 µM)	0.25 µl	0.5 µM
BiooPrimer2 (100 µM)	0.25µl	0.5 µM
Phusion polymerase	0.5	n.a.
Adapter-ligated ChIP sample	25 µl	n.a.
Final Volume	***50 µl***	

* to dissolve GC-rich sequences

Table 4-43 Cycling protocol for PCR enrichment (part 1) of ChIP samples (NGS library protocol II)

Cycle step	*Temperature*	*Time*	*Number of cycles*
Initial denaturation	98 °C	30 sec	1×
Denaturation	98 °C	10 sec	
Annealing	65 °C	30 sec	4×
Extension	72 °C	30 sec	
Final extension	72 °C	5 min	1×
Cooling	*4 °C*	*hold*	

Table 4-44 Cycling protocol for PCR enrichment (part 1) of MCIp samples (NGS library protocol II)

Cycle step	*Temperature*	*Time*	*Number of cycles*
Initial denaturation	98 °C	2 min*	1×
Denaturation	98 °C	30 sec*	4×
Annealing	65 °C	30 sec	
Extension	72 °C	45 sec*	
Final extension	72 °C	5 min	1×
Cooling	*4 °C*	*hold*	

* to dissolve GC-rich sequences

Excess primers were removed by clean up with AMPure XP magnetic beads (equilibrated for 30 min at RT on a rotor) at a ratio 1:1.8. 90 µl beads were added to the PCR reaction mix (50 µl) and processed as before. The bound DNA fragments were eluted with 10 µl nuclease-free ddH_2O and transferred to a new tube for the Caliper approach. Normally, 2 µl dye was added to the eluted fragments and fragments of 275 bp ± 15% (234-316 bp) using the LabChip XT DNA 300 Chip were recovered. Input DNA was recovered at a size range of 250 bp ± 15% (213-287 bp). The recovery volume was 20 µl. In case of the pooled BACs, the ladder was diluted in the lyophilized TE sample

After adjusting the sample volume with nuclease-free ddH2O to 37.5 µl for ChIP samples or 25 µl for MCIp samples, respectively, the second round of PCR enrichment of the size selected DNA fragments was performed. For round two, reaction composition is the same as for round one (see Table 4-41 and Table 4-42), but the cycle number was increased (see Table 4-45 & Table 4-46)

Table 4-45 Cycling protocol for PCR enrichment (part 2) of ChIP samples (NGS library protocol II)

Cycle step	*Temperature*	*Time*	*Number of cycles*
Initial denaturation	98 °C	30 sec	1×
Denaturation	98 °C	10 sec	12×
Annealing	65 °C	30 sec	
Extension	72 °C	30 sec	
Final extension	72 °C	5 min	1×
Cooling	*4 °C*	*hold*	

Table 4-46 Cycling protocol for PCR enrichment (part 2) of MCIp samples (NGS library protocol II)

Cycle step	*Temperature*	*Time*	*Number of cycles*
Initial denaturation	98 °C	2 min*	1×
Denaturation	98 °C	30 sec*	12×
Annealing	65 °C	30 sec	
Extension	72 °C	45 sec*	
Final extension	72 °C	5 min	1×
Cooling	*4 °C*	*hold*	

* to dissolve GC-rich sequences

The PCR product was again depleted of excessive primers with AMPure XP beads (twice) as described above (1:1.1 ratio). Elution volume of the first round was 50 µl nuclease-free ddH_2O, of the second 11 µl. 1 µl of the final eluate was run on a Bioanalyzer DNA HS Chip to validate successful enrichment of size-selected DNA fragments.

The library samples were further processed for NGS data acquisition by the KFB (Biopark, Regensburg).

4.3.1.19 Next generation sequencing on the Illumina platform

High-throughput DNA Sequencing with the Solexa Sequencing-by-Synthesis Technology (Illumina) (next generation sequencing, NGS) is based on iterative sequencing using reversible terminator chemistry. Labeling the modified nucleotides with different fluorescent dyes permits parallel readout of all immobilized library molecules. The sequencing-by-synthesis workflow is illustrated in Figure 4-3. The advantage to Sanger sequencing is the massive parallel sequencing of millions of DNA fragments. Due to the high sequencing depth, up to 4 billion reads per lane, multiplexing of samples by using barcoded libraries dramatically reduces costs. Samples analyzed by NGS in this work included ChIP, MCIp, RNA, and pooled BAC samples.

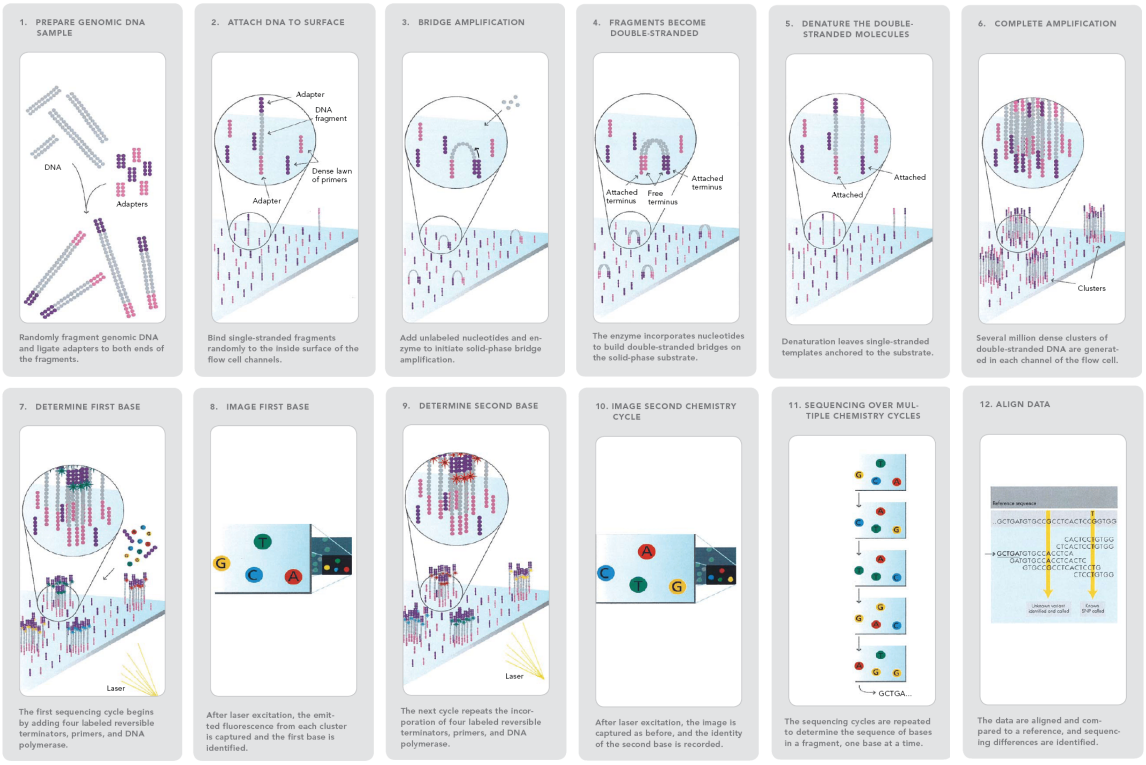


Figure 4-3 Workflow of Illumina's NGS technology
1, 2 – Single strands of the NGS library generated according to the NGS library protocols listed in section 4.3.1.18 bind to complementary adapters. 3-6 – Clusters with more than 100 copies of the starting library molecule are generated by bridge amplification. 7 – Sequencing primer anneal to the adapter sequence of the cluster molecule and 3'-terminated, fluorescent-labeled nucleotides are incorporated at the 3' end of the sequencing primer by a modified polymerase (due to the 3' terminating modification, only one nucleotide per cycle is incorporated). 8 – Using different lasers for excitation and filters, an image is generated of the attached fluorophores, which reflects the incorporation positions of the individual nucleotides. 9-11 – Following each imaging cycle, the 3' terminators as well as the fluorescent labels are chemically removed permitting for a new cycle of incorporation and imaging. The read length for all clusters is determined by the numbers of repeated incorporation and imaging cycles. Barcodes of pooled libraries are encoded in the adapter sequence and assessed accordingly. Different sequences of left and right adapter allows for paired-end sequencing. 12 – Following sequencing, the Illumina base caller analyzes the images. The collected sequence data are aligned and compared with a reference sequence. (Taken from Illumina, www.illumina.com)

4.3.1.20 **BAC library screening with DIG-labeled probes**

BACPAC Resources harbours BAC libraries from a multitude of organisms; however, most of BAC clones have not been mapped yet, inter alia BALB/cByJ (CHORI-28). To identify BACs of interests, BACPAC Resources provides filter membranes with spotted BAC clones representing the whole library. Each membrane carries over 18000 distinct BAC clones in duplicates and statistically represents the genome once. To statistically obtain 3x coverage of the region of interest three membranes were screened.

Probe design

Hybridization probes were generated by PCR. PCR primers were designed with the PerlPrimer software according to the following criteria.

(i) one amplicon per 50 kb window within region *mm9*_chr12:16902483-26024113 (*to enhance the detection signal for each BAC*), (ii) The melting temperature T_m was restricted to 63-65 °C (Δ_{max}=2 °C) (*to yield an annealing temperature of ~60 °C*), (iii) The amplicon length was set to 400-500 bp (to *provide for a reasonable number of DIG-labeled bases per amplicon*), (iv) the length of the oligos was set to 21-28 bp (to facilitate BLATing in UCSC genome browser as well as appropriate T_ms), (v) sequences with high GC content were excluded (*for similar PCR conditions*), (vi) GC clamp at the 3' end of oligos was preferred (*to ensure primer specificity*), (vii) at least one oligonucleotide per amplicon needed to map less than three times by BLAT (UCSC genome browser) (*to generate distinct amplicons*), (viii) the amplicons returned by In-Silico PCR (UCSC genome browser) must to fall within the 9 Mb candidate region and if criterion (vii) could not be fulfilled, were limited to three distinct amplicons per oligonucleotide pair.

- Primer binding sites must not to contain known SNPs
- Primer binding sites should not overlap with repetitive sequences

If criterion (vii) was omitted and up to three different 50 kb windows within the candidate region were covered with one probe, additional probe design for these regions was refrained from.

Furthermore, the theoretical hybridization temperature for the amplicons was calculated according to the following formula:

$$T_{hyb}=T_m - (20 \text{ to } 25\ ^\circ C)$$

$$T_{m'=}49.82 + 0.41(\%GC) - 600/length(bp)$$

(for 80-100% sequence homology)

Since probe design for the repetitive regions was difficult in general, only probes of the conserved regions were limited to an intermediate hybridization temperature of T_{Hyb}=42-50°C (formula see below).

PCR labeling

PCR was initially performed with normal dNTPs to optimize the amplification conditions as well as to control for specificity. PCR was performed as described in Table 4-47 & Table 4-48, but scaled up to 50 µl total reaction volume and with dNTP mix (2mM) instead of PCR DIG Probe Synthesis Mix (10x). 5 µl of the PCR product were analyzed by gel electrophoresis on a 1.5% agarose gel (120 V, 120 min). In case of insufficient PCR product, the setting were adjusted individually (setting ranges are listed below Table 4-48). The residual 45 µl of probes with multiple amplicons were separated on a second 1.5% agarose gel (120 V, 120 min) and the amplicons with the expected size were cut out and amplified DNA purified with the QIAquick GelExtraction Kit. The concentration of the purified amplicons was assessed with the NanoDrop apparatus and adjusted to ~300 pg/µl with nuclease-free ddH_2O. The probes were labeled with Digoxigenin during PCR amplification of genomic DNA from BALB/cByJ or amplicon DNA, respectively, with the PCR DIG Probe Synthesis Kit (Roche).

PCR was performed in 96-well PCR plates. Mixes for multiple reactions were prepared according to Table 4-47 and dispensed to the wells in the following order. 6 µl of nuclease-free ddH_2O was dispensed to each well as well as 2 µl Primer mix (containing sense and antisense Primers at a concentration of 6.25 µM each). Subsequently, 7 µl of Mix1 was added to each well, as well as 10 µl of Mix2.

Table 4-47 Reaction composition for DIG PCR labeling

Component	*Volume*	*Final concentration*
Mix1 (for genomic DNA)		
Nuclease-free ddH_2O	4 µl	n.a.
PCR DIG Probe Synthesis Mix (10x)	2.5 µl	1x (200 µM of each nucleotide)
genomic DNA	0.5 µl	10 ng
Final Volume	*7 µl*	
Mix1 (for amplicon DNA)		
Nuclease-free ddH_2O	3.5 µl	n.a.
PCR DIG Probe Synthesis Mix (10x)	2.5 µl	1x (200 µM of each nucleotide)
amplicon DNA (1:60 dilution)	1 µl	5 pg
Final Volume	*7 µl*	
Mix2		
Nuclease-free ddH_2O	7.125 µl	n.a.
PCR Buffer (10x)	2.5 µl	1x
Enzyme Mix (3.5 U/ml)	0.375 µl	1.3 U
Final Volume	*10 µl*	

The plate content was mixed on a vortexer and incubated on a thermal cycler according to Table 4-48.

Table 4-48 General cycling protocol for DIG PCR labeling

Cycle step	*Temperature*	*Time*	*Number of cycles*
Initial denaturation	95 °C	4 min	1×
Denaturation	95 °C	30 sec	
Annealing	62 °C*	30 sec	30×***
Extension	72 °C	25 sec**	
Final extension	72 °C	7 min	1×
Cooling	*4 °C*	*hold*	

* 56-62 °C, ** 25-30 sec, ***30 or 32 cycles

For hybridization, 2-4 µl of the individual probe would be mixed with 1 ml DIG Easy Hyb buffer. The individual volume of each probe depended on the labeling efficiency that was assessed by Dot plot analysis (see below). Since the probes were pooled for parallel hybridization, less amount of the individual probe was required per ml DIG Easy Hyb buffer. Depending on the number of combined probes and the labeling efficiency of the individual probes, 0.029 – 0.087 µl of the individual probe was contained per ml hybridization solution.

Dot plot

Prior to hybridization the labeling efficiency for each probe was tested in a dot plot experiment. A dilution series of each probe was spotted on a positively charged nylon membrane, UV-crosslinked (120 mJ, 3-5min), and subject to DIG-antibody-based chemiluminescent detection (see following section).

Probe hybridisation and detection

Hybridizing colony filters for the 1st time, cell debris from the spotting procedure have to be removed. The membranes were incubated twice for 10 min at RT in 150 ml SSC (2x) on a shaker followed by two times incubation for 15 min at RT in 150 ml SSC (2x, 1.5% SDS) on a shaker. The membrane was then placed on a soaked 3MM Whatman paper (2x SSC), DNA side facing up, and covered firmly with a wet sheet of 3MM Whatman paper (sterile ddH_2O). A ruler was used to apply constant pressure. Afterwards the membrane was washed twice for 3 min at RT in 150 ml sterile ddH_2O on a shaker.

The precleared membrane was then equilibrated for min. 15 min at RT in 150 ml SSC (2x) and placed on a hybridization mesh also soaked in SSC (2x). The same applies for stripped membranes. Both mesh and membranes were rolled up, placed in a hybridization bottle containing 40 ml prewarmed DIG Easy Hyb buffer (Roche) and incubated in a rotating hybridization oven for 1.5-3 h at the theoretically determined hybridization temperature (T_{hyb}) determined above.

Hybridization solution was prepared by adding 2 µl denatured probe mix (98 C, 10 min → on slushy ice, min. 5 min) per ml prewarmed (T_{hyb}) DIG Easy Hyb buffer (Roche). Used hybridization solutions can be stored at -20 °C and reused 3-4 times. Prior to reusing stored hybridization solutions it was incubated at 68 °C for 10 min and cooled immediately on slushy ice for 10 min.

After pre-hybridization, DIG Easy Hyb Buffer was replaced by prewarmed hybridization solution and the membrane was incubated with the hybridization solution for 24 h at T_{hyb} in a rotating hybridization oven.

The membrane was washed 3 times for 3 min in 250 ml Low Stringency Buffer in the hybridization bottle, which was shaked by hand. The membrane was then transferred to a new, prewarmed (68 °C) hybridization bottle, filled with 250 ml prewarmed (68 °C) High Stringency Buffer and placed in the hybridization oven. The High Stringency Buffer was replaced 3 times during 30 min incubation at 68 °C. Hybridized membranes can be stored or used directly for chemiluminescent detection of hybridized probes.

For storage, membranes were air-dried and kept between two sheets of 3MM Whatman paper in a sealed bag at 4 °C. Stored membranes cannot be stripped and re-probed after detection.

For chemiluminescent detection of hybridized probes the membrane was washed twice in 250 ml Washing Buffer (1x, Roche) for 2 min and blocked for at least 30 min (up to 3 h) in 400 ml Blocking Solution (1x, Roche). The blocked membrane was incubated in 75 ml Antibody Solution (Roche) for 30 min and washed again twice in 300 ml Washing Buffer (1x, Roche) for 15 min. The antibody binds to the DIG epitope of hybridized probes, whereas unspecifically bound antibodies are washed away. The membrane was then equilibrated in 60 ml Detection Buffer (1x, Roche) for 2-5 min. All previous steps were carried out at RT on a shaker. For detection the equilibrated membrane was placed onto plastic foil with the DNA side facing up, sprinkled with 3 ml/100 cm^2 of CDP-Star substrate (Roche) and immediately covered with another sheet of plastic foil in a way that allows for even spreading of the CDP-Star substrate. The bound antibody is linked to an enzyme emitting detectable light upon the addition of substrate. After 5 min incubation at RT, excessive substrate was removed and the two sheets of plastic foil were weld together. The chemiluminescent signal was detected on GelDoku (Bio-Rad) for 60-90 min with pictures taken every 90 sec. The signal is detectable up to 2 days after substrate addition.

Membranes that were not air-dried after probe hybridization were stripped after chemiluminescent detection of the hybridized probes. The membrane was rinsed in 300 ml sterile ddH_2O for 1 min and incubated twice in 250 ml Stripping Buffer for 15 min at 37 °C on a shaker. The membrane was rinsed in SSC (2x) for 5 min at RT and stored wet between two sheets of 3MM Whatman paper in SSC (2x) at 4 °C.

Required reagents and materials:
(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

SSC	2x	50 ml	UltraPure SSC Buffer (20x) (Gibco/Life Tech. # 15557-044)
		450 ml	sterile ddH_2O
SDS	10%	100 g	Sodium dodecyl sulfate (Sigma-Aldrich, #L4390)
		→ Ad 1 l	nuclease-free ddH_2O
SSC (2x, 1.5% SDS)		50 ml	UltraPure SSC Buffer (20x) (Gibco/Life Tech. # 15557-044)
		75 ml	SDS (10%)
DIG Easy Hyb Buffer			DIG Easy Hyb (Roche, #11603558001)
Low Stringency Buffer		75 ml	UltraPure SSC Buffer (20x) (Gibco/Life Tech. # 15557-044)
		7.5 ml	SDS (10%)
		667.5 ml	sterile ddH_2O
High Stringency Buffer		5 ml	UltraPure SSC Buffer (20x) (Gibco/Life Tech. # 15557-044)
		10 ml	SDS (10%)
		985 ml	sterile ddH_2O
DIG Wash and Block Buffer Set			Roche (#11585762001)
• Maleic Acid Buffer (1x)		10 ml	Maleic Acid Buffer (1x, RT)
		90 ml	sterile ddH_2O
		→ Stable at RT;	
• Blocking Solution (1x)		10 ml	Blocking Solution (10x, RT)
		90 ml	Maleic Acid Buffer (1x, RT)
		→ Prepare fresh and place at 4 °C;	
• Washing Buffer (1x)		10 ml	Washing Buffer (10x, RT) (shake well!)
		90 ml	sterile ddH_2O
		→ Stable at RT;	
• Detection Solution (1x)		10 ml	Detection Buffer (10x)
		90 ml	sterile ddH_2O
		→ Stable at RT;	

Antibody Solution		Anti-Digoxigenin-AP Fab fragments (Roche, #11093274910, 4 °C)
	→ ↺ 5min, 10’000 rpm, 4 °C	
	1 µl	‘cleared’ Antibody
	10 ml	Blocking Solution (1x)
	→ Prepare fresh, stable for 12 h at 4 °C;	
CDP-Star substrate		CDP-Star ready-to-use (Roche, #12041677001, 4 °C)
Stripping Buffer	60 ml	Sodium hydroxide solution (1M) (Merck/Millipore, #109137)
	3 ml	SDS (10%)
	236 ml	sterile ddH_2O
Hybridization bottle		Large hybridization bottle (Biometra #846052001)
Hybridization oven		Hybridization Oven 6/12 (Unitherm)
Hybridization mesh		Mesh (23 cm x 23 cm) (Labnet, #H9088)
Plastic tray		Nunc™ Square BioAssay Dishes (Thermo Scientific, #240835)
Plastic foil		Weichfolien-Prospekthülle (DIN-A4) (Milan #803)
Welding apparatus		HS 250 (Janke & Kunkel)

4.3.2 Preparation and analysis of ribonucleic acids (RNA)

4.3.2.1 Isolation of total RNA (larger than 200 nt)

Total RNA (>200 nt) was isolated using the RNeasy Midi, mini or Micro Kit (Qiagen) depending on the available number of cells according to the manufacturer's instructions. In case of fresh or RNAlater-stabilized tissues, the tissue was homogenized with the TissueLyser (3 min, 20 Hz) and filtering through a QIAshredder column by centrifugation at RT for 3 min at 13000xg. To remove potential DNA contaminations, on-column DNA digestion with the RNase-free DNase Set (Qiagen) was integrated according to the protocol. RNA concentration was determined with the NanoDrop 1000 spectrophotometer. RNA integrity and quality was assessed with the Agilent Bioanalyzer RNA 6000 Chip assay (Agilent Technologies) according to the manufacturer's instructions.

4.3.2.2 Reverse transcription quantitative real-time PCR (RT-qPCR)

To analyze transcription levels, total RNA was transcribed into complementary DNA (cDNA) with the M-MLV RT (H- point mutant) enzyme (Promega). Random decamers (Promega) were used to prime cDNA synthesis. Individual expression levels were subsequently assessed by quantitative PCR (see section 4.3.1.10) on cDNA level.

The volume of 1 µg of total RNA was adjusted to 13 µl with nuclease-free ddH_2O and mixed with 1 µl Random Decamers (Promega) and 1 µl dNTPs (10 mM) on ice. Secondary structures are dissolved by 5 min incubation at 65 °C followed by immediate incubation ice for 1 min. After mixing with 4 µl M-MLV Buffer (5x) the sample was incubated at 42 °C for 2 min. Reverse transcription started upon addition of 1 µl RT enzyme (50 min, 42 °C) and was stopped by heat inactivation of the enzyme (15 min, 70 °C). CDNA samples are stable at -20 °C.

Required reagents and materials:
(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

M-MLV RT (H- point mutant)			Promega (#M368B), -20 °C
M-MLV Buffer (5x)			Promega (#M531A), -20 °C
Random Decamers (Retroscript)			Ambion/Life Tech. (#5722G), -20 °C
dNTPs	10 mM	100 µl	dATP/dCTP/dGTP/dTTP each (100mM) (Roche, # 11969064001) → Ad 1 ml nuclease-free ddH_2O, store at -20 °C

4.3.2.3 Whole genome expression analysis by RNA-seq

The Next generation sequencing technology can also be applied to analyze whole genome gene expression based on cDNA derive from RNA samples (RNA-seq). RNA-seq libraries from differentiating ESCs were prepared from RNA isolated according to section 4.3.2.1. 4.5 µg of RNA from BALB/c day0, day1, day2, and day4 as well as JM8 day0 and day4 was subject to ribosomal RNA (rRNA) depletion using the RiboZero (Mouse) Kit (Epicentre) according to the manufacturer's instruction. The following variations of the protocol were applied: After washing the bead-bound RNA strands, the mixture was centrifuged for 5 min at 12000 g; to recover the eluted RNA strands, centrifugation was prolonged to 1:30 min; Ethanol precipitation to recover the cleaned rRNA-depleted RNA samples was performed o/n. The precipitated RNA was resuspended in 5.5 µl nuclease-free ddH_2O. Successful rRNA depletion was validated with the Bioanalyzer mRNA Nano assay (Agilent Technologies). According to the Bioanalyzer assay, recovered RNA amounts varied between 50 and 280 ng. Up to 210 ng rRNA-depleted RNA was further processed ScriptSeq (v1) RNA-seq Library Kit according to the manufacturer's instructions. CDNA purification was performed with the QIAquick MinElute PCR Purification Kit according to the NGS library protocol I (see 4.3.1.18) and elution with 20 µl EB buffer (37 °C). To amplify the cDNA fragments, an alternate PCR cycling program to account for the use of Phusion Taq polymerase (which requires 72 °C for Elongation and is twice as fast).

Table 4-49 Cycling protocol for PCR enrichment (RNA-seq library protocol v1)

Cycle step	*Temperature*	*Time*	*Number of cycles*
Initial denaturation	95 °C	30 sec	1×
Denaturation	95 °C	30 sec	
Annealing	55 °C	30 sec	11*×
Extension	72 °C	1:45 min	
Final extension	72 °C	7 min	1×
Cooling	*4 °C*	*hold*	

* in case of BALB/c d0 (50 ng input) 12 cycles were used.

The library fragments were purified using AMPure XP beads according to the manufacturer's instructions and eluted in 25 µl nuclease-free ddH_2O. Libraries were further processed for NGS at the KFB (Biopark, Regensburg).

The RNA-seq libraries for JM8 day1 and day2 performed at a later stage and with newer versions of the above used Kits. RRNA removal was performed on 4.5 µg RNA with the rRNA Removal Magnetic Kit (Epicentre) according to the manufacturer's instructions. After Ethanol precipitation o/n, the precipitated RNA was resuspended in 10 µl nuclease-free ddH_2O. To validate rRNA-removal, 1 µl of the rRNA-depleted RNA was diluted with 9 µl nuclease-free ddH_2O and analyzed on a Bioanalyzer mRNA Pico assay (Agilent Technologies). As recommended by the manufacturer's instructions, 15 ng of rRNA-depleted RNA were further processed with the

ScriptSeq v2 RNA-Seq Library Kit. PCR enrichment was validated on a Bioanalyzer DNA HS Chip assay (Agilent Technologies) and the libraries further processed for NGS at the KFB (Biopark, Regensburg).

4.3.3 Molecular cloning

4.3.3.1 Cloning of vectors for Sanger sequencing

Subcloning of PCR products into a sequencing vector was either done with the StrataClone™ PCR Cloning Kit (Agilent Technologies) or the TOPO® TA Cloning® Kit (Life Technologies), respectively.

Chemically competent E. coli were thawed on ice. Transformation with the ligation mix was performed according to the manufacturer's instructions except that in general only half of the recommended Volumes were used. 250 µl SOC medium was added to the bacteria and the suspension was incubated at 37°C for one hour on a shaker in order to express the resistance. For blue/white screening of insert-containing clones, 40 µl of X-gal was dispersed on a prewarmed LB_{amp} plate prior to use and incubated at 37°C for an additional 30 min. Afterwards 50-150 µl of the transformation reaction were plated and incubated at 37°C over night on LB_{amp} plates.

Required reagents and material:
(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

SOC medium	2%	20 g	Bacto™ Tryptone (BD, #211705)
	0.5%	5 g	Bacto™ Yeast extract (BD, #212720)
	10 mM	0.6 g	Sodium chloride (AppliChem, #A3597)
	3 mM	0.2 g	Potassium chloride (Merck Millipore, #1049360250)
		→ Ad 1 l nuclease-free ddH_2O, autoclave	
		→ Add to cooled solution:	
	10 mM	10 ml	$MgCl_2$ (1 M), sterile-filtered
	10 mM	10 ml	$MgSO_4$ (1 M), sterile-filtered
	20 mM	10 ml	Glucose (2 M), sterile-filtered

LB (Lysogeny broth) medium	170 mM	10 g	Sodium chloride (AppliChem, #A3597)
	1%	10 g	Bacto™ Tryptone (BD, #211705)
	0.5%	5 g	Bacto™ Yeast extract (BD, #212720)
		→ Adjust with NaOH to pH 7.5	
		→ Ad 1 l ddH_2O, autoclave	
X-gal	40 mg/ml	40mg	X-gal (5-bromo-4-chloro-3-indolyl-ß-D-galactoside)
		→ Dissolve in 1 ml DMF	
		→ Store at -20°C protected from light	

4.3.3.2 BAC recombination with the Cre/LoxP system

Cloning additional sequences like resistency-conferring genes can be achieved by recombination of LoxP sequence elements in the presence of active Cre recombinase. In a topoisomerase I like mechanism the enzyme performs site specific recombination event between two DNA segments containing the 34 bp long LoxP recognition site (ATAACTTCGTATANNNTANNNTATACGAAGTTAT)[272].

BACs selected from the CHORI-28 Mouse BALB/cByJ (*Mus musculus*) BAC libraries according to section 4.3.1.20 only contain a Chloramphenicol resistency not suitable for selection of transfected eukaryotic cells. Using the LoxP/Cre recombination system, the pNelγ plasmid, which carries similar to the BAC clones a LoxP recognition site, was integrated into the BAC backbone. PNelγ promotes prokaryotic resistency to Kanamycin and Spectinomycin, but also eukaryotic resistency to Geneticin (G418). PNelγ also encodes for an Enhanced Green Fluorescent Protein (*eGFP*) gene under the control of the eukaryotic CAG promoter and thus enables visualization of transfection efficiency. Without integration into the BAC backbone, pNelγ cannot be maintained in neither DH10B nor SW106 cells since its R6K ori requires the expression of the *pir* gene for replication.

Retrofitting in the Cre-expressing* E. coli *strain SW106

The SW106 cells express Cre recombinase after induction with L-(+)-Arabinose. The transformation of the BAC was done with 2 µl BAC preparations according to section 4.3.1.1 with the exception that the precipitated BAC DNA was resuspended in 6 µl nuclease-free ddH_2O. SW106 cells were in general cultured at 30 °C. BAC DNA was electroporated into electrocompetent SW106 cells according to section 4.1.2.2. Selection for BAC-containing clones (sBACs) was done on LB_{CM} agar plates incubated o/n at 30 °C (see Table 4-2). In general, one Glycerol stock per BAC was prepared from a single colony. For electroporation of sBACs with pNelγ 20 ml LB_{CM} was inoculated with 1 ml of 2 ml LB_{CM} o/n culture and shaked vigorously at 30 °C for 1 h. Cre expression was induced by adding 0.1% L-(+)-Arabinose and further incubation at 30 °C for 1h. The cells were cooled on ice on a shaker for 15 min and harvested by

centrifugation (15 min, 2900xg, 4 °C). The pellet was washed twice with 2 ml and 1 ml ice-cold sterile ddH_2O and once with 1 ml ice-cold sterile Glycerol/ddH_2O (10%). The pellet was finally resuspended in 80 µl ice-cold sterile Glycerol/ddH_2O (10%), which rendered ~90 µl of electrocompetent, Cre-expressing bacteria. 40 µl of the electrocompetent sBACs were used directly for electroporation, the remaining cells were stored at -80 °C as back up.

The cells were transformed with 500 ng of pNelγ according to section 4.1.2.2 and plated on $LB_{CM+Kan+Sp}$ agar plates (see Table 4-2). After 24-36 h incubation at 30 °C several clones were picked and the retrofitted BACs (nBACs, in SW106 strain) were prepared according to section 4.3.1.1. The HindIII digestion pattern of the individual nBACs was compared to the original BAC (oBACs, in DH10B strain) and a glycerol stock was prepared for a positive nBAC.

Retrofitting with the Cre-encoding pJM2545 plasmid

The pJM2545 plasmid constantly expresses Cre recombinase, but has a temperature-sensitive replication ori. BAC-containing cells co-transformed with both pNelγ and pJM2545 will experience LoxP/Cre recombination until they are incubated at high temperature, which leads to the loss of pJM2545 during replication. Individual oBAC clones were prepared for electroporation as described above for sBACs with the exception that the cells were incubated at 37 °C and no supplement was added to the medium except selecting antibiotic. Furthermore, the washed cell pellet was resuspended 30 µl ice-cold sterile Glycerol/ddH_2O (10%) and used completely for electroporation with 200 ng pJM2545 and 500 ng pNelγ (settings according to section 4.1.2.2). After 1 h recovery in 250 µl SOC medium, the cells were plated on $2xYT_{CM+Sp}$ agar plates and incubated for 1 h at 30 °C followed by o/n incubation at 42 °C. Several clones were picked and the retrofitted BACs (njBACs, in DH10B strain) were prepared according to section 4.3.1.1. The HindIII digestion pattern of the individual njBACs was compared to the oBACs and a Glycerol stock was prepared for a positive njBAC.

Required reagents and materials:
(Stated volume-/weight-% or molarities refer to final concentrations, storage according to manufacturer's instructions or at RT if not stated otherwise)

SOC medium	2%	20 g	Bacto™ Tryptone (BD, #211705)
	0.5%	5 g	Bacto™ Yeast extract (BD, #212720)
	10 mM	0.6 g	Sodium chloride (AppliChem, #A3597)
	3 mM	0.2 g	Potassium chloride (Merck Millipore, #1049360250)
		→ Ad 1 l nuclease-free ddH_2O, autoclave	
		→ Add to cooled solution:	
	10 mM	10 ml	$MgCl_2$ (1 M), sterile-filtered
	10 mM	10 ml	$MgSO_4$ (1 M), sterile-filtered
	20 mM	10 ml	Glucose (2 M), sterile-filtered

LB (Lysogeny broth) medium	170 mM	10 g	Sodium chloride (AppliChem, #A3597)
	1%	10 g	Bacto™ Tryptone (BD, #211705)
	0.5%	5 g	Bacto™ Yeast extract (BD, #212720)
		→ Adjust with NaOH to pH 7.5	
		→ Ad 1 l ddH_2O, autoclave	
LB agar plates	1.5%	15 g	Bacto™ Agar (BD, #214010)
	170 mM	10 g	Sodium chloride (AppliChem, #A3597)
	1%	10 g	Bacto™ Tryptone (BD, #211705)
	0.5%	5 g	Bacto™ Yeast extract (BD, #212720)
		→ Adjust with 1 M NaOH to pH 7.5	
		→ Ad 1 l ddH_2O, autoclave, and cool to 50 °C	
		→ Add required antibiotics (see *Table 4-2*)	
		→ Pour into 10 cm Petri dishes, store inverted at 4 °C	
2xYT agar plates	1.5%	15 g	Bacto™ Agar (BD, #214010)
	85 mM	5 g	Sodium chloride (AppliChem, #A3597)
	1.6%	16 g	Bacto™ Tryptone (BD, #211705)
	1%	10 g	Bacto™ Yeast extract (BD, #212720)
		→ Adjust with 1 M NaOH to pH 7.0	
		→ Ad 1 l ddH_2O, autoclave, and cool to 50 °C	
		→ Add required antibiotics (see *Table 4-2*)	
		→ Pour into 10 cm Petri dishes, store inverted at 4 °C	
NaOH	1 M		Sodium hydroxide solution (Merck/Millipore, #109137)
Petri dishes	9.4 cm		Greiner Bio-one, #633102

4.4 Analysis of NGS data sets

Bioinformatic analyses of next generation sequencing (NGS) data were performed on a Linux Debian workstation with the indicated programs and versions. Raw read files were delivered in fastq format from the sequencing facilities.

4.4.1 ChIP-seq and MCIp-seq data

ChIP-seq data sets from spleen chromatin were generated by non-indexed single 36 bp sequencing on a GAII (Illumina) using sequencing chemistry SBSv1 (Illumina). ChIP-seq data from cerebellum chromatin as well as MCIp-seq data from both tissues were generated by indexed (6 samples/lane) single 50 bp sequencing on a HiSeq1000 (Illumina) using sequencing chemistry SBSv3 (Illumina). The general NGS workflow is stated in section 4.3.1.19.

4.4.1.1 Identification of novel candidate DMRs using ChIP-seq

ChIP-seq data from spleen chromatin (both C57BL/6 and BALB/c; H3K4me1, H3K4me3, H3K27ac, and input each) were initially analyzed according to the following script to identify regions marked by active histone in a strain-specific manner. Regions for MALDI-TOF MS-based DNA methylation analysis listed in Appendix IV (Table 10-4) were selected from strain-specific ChIP-seq peaks.

The following script for C57BL/6 spleen H3K4me3 is representative for all analyzed samples. Strain- and mark-specific alterations are indicated. If not stated otherwise in the command line, default settings were applied.

```
###Software versions
#Fastqc        0.10.1
#bowtie        0.12.7
#homer         3.0


###Quality assessment with Fastqc.


###Alignment of raw reads to mm9 reference genome with bowtie.
% bowtie –q --best –k 1 –m 1 –p <#CPU> mm9 /path-to-input-file/C57_spl_K4me3.fastq > \
/path-to-output-file/C57_spl_K4me3.txt;
```

```
###Creating Tag Directories with the homer tool pipeline.
% makeTagDirectory /path-to-output-folder/C57_spl_K4me3 \
/path-to-alignment-file/C57_spl_K4me3.txt;
###Creating UCSC custom tracks to visualize aligned sequence tags for each sample (homer
###v3.0).
% makeUCSCfile /path-to-tag-directory/C57_spl_K4me3 -o auto;
##The UCSC browser track is accessible at http://genome.ucsc.edu/cgi-
bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=juwelll&hgS_otherUserSessionName=C57
vsBc_spleen_%20%09ChIP%2Dseq_H3K4me1_H3K4me3_H3K27ac_mm9


###Identifying peaks (ChIP-enriched regions) with the homer tool pipeline.
% findPeaks /path-to-tag-directory/C57_spl_K4me3 -i /path-to-tag-directory/C57_spl_input \
-size 1000 -L 0 -min-Dist 2500 -o /path-to-output-file/C57_spl_K4me3_peaks.txt;

###Merging peak files from both strains for individual histone marks with the homer tool
###pipeline.
% mergePeaks -d 500 /path-to-peak-file1/C57_spl_K4me3_peaks.txt \
/path-to-peak-file2/Bc_spl_K4me3_peaks.txt > \
/path-to-output-file/merged_C57+Bc_spl_K4me3_peaks.txt;

###Annotate peaks (tag information and mm9 genomic information) in merged peak file with
###the homer tool pipeline.
% annotatePeaks.pl /path-to-merged-peak file/merged_C57+Bc_K4me3_peaks.txt mm9 > \
/path-to-output-file/merged_C57+Bc_K4me3_peaks.ann.txt;
#The annotated merged peak files for H3K4me1, H3K4me3, and H3K27ac
#(merged_C57+Bc_*_peaks.ann.txt) were loaded and combined in an Excel (2007) file and
#further analyzed to identify regions (peaks) strain-specifically enriched for active
#histone marks as stated in section 5.2.1.
```

4.4.1.2 Comparison of ChIP-seq and MCIp-seq data sets

ChIP-seq data from spleen were complemented with data from cerebella of the same individuals for all three histone marks (both C57BL/6 and BALB/c; H3K4me1, H3K4me3, H3K27ac, and input each). Furthermore, genome-wide DNA methylation data from both tissues in both strains were accessible by MCIp-seq data. The following scripts were used to compare ChIP-seq data from different tissues as well as ChIP-seq and MCIp-seq data within the same strain/tissue and between different strains/tissues.

The following script for C57BL/6 spleen H3K4me3 and/or MCIp samples is representative for all analyzed samples. Strain- and mark-specific alterations are indicated. If not stated otherwise in the command line, default settings were applied.

```
###Software versions
#Fastqc        0.10.1
#bowtie2       2.1.1
#homer         4.6
#R             3.1.0
###Quality assessment with Fastqc.
```

```
###Alignment of raw reads to mm9 reference genome with bowtie2.
% bowtie2 -x /path-to-bowtie2-programs/indexes/bt2-mm9 /path-to-file/C57_spl_K4me3.fastq \
-S /path-to-file/C57_spl_K4me3.sam;


###Creating Tag Directories with the homer tool pipeline.
##For MCIp: add '-keepone' option
% makeTagDirectory /path-to-output-folder/C57_spl_K4me3 /path-to-file/C57_spl_K4me3.sam \
-genome mm9 -checkGC (-keepone);


###Creating MultiWigHub for layered visualization in UCSC genome browser with the homer
###tool pipeline.
##For ChIP-seq, one MultiWigHub was created for each strain and each tissue containing
##input, H3K4me1, H3K4me3, and H3K27ac.
% makeMultiWigHub.pl C57_spl_ChIP mm9 -fsize 1e8 -d \
/path-to-tag-directory/C57_spl_input /path-to-tag-directory/C57_spl_K4me1 \
/path-to-tag-directory/C57_spl_K4me3 /path-to-tag-directory/C57_spl_K27ac -webdir \
/path-to-web-directory/ -url http:/www.ag-rehli.de/ChIP+MCIp;

## For MCIp-seq, one MultiWigHub was created for each strain containing aligned MCIp-seq
##reads for both tissues.
% makeMultiWigHub.pl C57_MCIp mm9 -fsize 1e8 -d /path-to-tag-directory/C57_spl_MCIp \
/path-to-tag-directory/C57_cer_MCIp -webdir /path-to-web-directory/ \
-url http:/www.ag-rehli.de/ChIP+MCIp;
#The contents of the track information file (trackdb.txt) of all tracks within one hub
#were combined in one trackdb.txt file to facilitate layered visualization.
#The multi-layered track hubs are accessible at http://genome-euro.ucsc.edu/cgi-
bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=juwelll&hgS_otherUserSessionName=mm9
_DRRs%26DMRs.


###Identifying peaks (enriched regions) for each sample with the homer tool pipeline.
##H3K4me3, H3K27ac (to account for mapping bias, adjust enrichment over input: C57 default
##(-F 4), BALB/c -F 3).
% findPeaks /path-to-tag-directory/C57_spl_K4me3 -region -size 250 -L 0 -minDist 350 \
-ntagThreshold 10 (-F 3) -i /path-to-input-tag-directory/ \
-o /path-to-output-file/C57_spl_K4me3_regions.txt;

##H3K4me1 data are processed differently to account for broader and lower H3K4me1 peaks
##(to account for mapping bias, adjust enrichment over input: C57 -F 3, BALB/c -F 2).
% findPeaks /path-to-tag-directory/C57_spl_K4me1 -region -size 400 -L 0 -minDist 600 \
-ntagThreshold 10 -F <2 or 3> -o /path-to-output-file/C57_spl_K4me1_regions.txt;

##MCIp data are processed differently, since no Input data was available.
%findPeaks /path-to-tag-directory/C57_spl_MCIp -region -size 250 -L 0 -minDist 350 \
ntagThreshold 10 -o /path-to-output-file/C57_spl_MCIp_regions.txt;
```

```
###Identifying strain-specifically enriched regions for each mark with the homer tool
###pipeline.
##Merging corresponding region files for subsequent identification of differential regions
##(=differential peaks) with homer v4.6.
% mergePeaks /path-to-region-file1/C57_spl_K4me3_regions.txt \
/path-to-region-file2/Bc_spl_K4me3_regions.txt > \
/path-to-output-folder/merged_C57+Bc_spl_K4me3_regions.txt;

##Identifying differentially enriched regions within merged region files for each
##strain/tissue combination and mark with the homer tool pipeline.
% getDifferentialPeaks /path-to-merged-region-file/ merged_C57+Bc_spl_K4me3_regions.txt \
/path-to-target-tag-directory/C57_spl_K4me3 \
/path-to-background-tag-directory/Bc_spl_K4me3 –F 2 (-rev) > \
/path-to-output-file/C57_over_Bc_F2_spl_K4me3.txt;
#'-rev' option looks for specified fold-enrichment of background tag directory over
#target directory and is applied separately for all strain/tissue combination and marks.

##Identifying tissue-independently enriched regions for H3K4me3 and H3K27ac in each
#strain.
#Merging regions with peak center distance <300bp (use –prefix option to separate common
#and unique regions). The command was used in both directions for both histone marks.
% cd /path-to-output-folder/;
% mergePeaks -d 300 -prefix tissue300 \
/path-to-input-file1/Bc_over_C57_F2_cer_ac.txt \
/path-to-input-file2/Bc_over_C57_F2_spl_ac.txt\;
#The two output files containing tissue-independent H3K4me3 (or H3K27ac) regions for each
#direction (enriched in C57BL/6 or BALB/c, respectively) were combined (>F2_H3K4me3.txt,
#F2_H3K27ac.txt).

#Merging the combined files for tissue-independent H3K4me3 and H3K27ac regions.
% mergePeaks -d 300 /path-to-input-file1/F2_H3K27ac.txt \
/path-to-input-file2/F2_H3K4me3.txt \
> /path-to-output-folder/F2_H3K4me3andorH3K27ac.txt;
#Annotating the resulting list of tissue-independently strain-specific H3K4me3 and/or
#H3K27ac regions with genomic context (mm9) and tag information.
% annotatePeaks.pl /path-to-region-file/ F2_H3K4me3andorH3K27ac.txt mm9 -size given \
-noann –d <space-separated list of tag directories> \
> /path-to-output-folder/F2_H3K4me3andorH3K27ac_ann.txt
#Regions associated with the same gene were grouped.

###Comparing general enrichment of active histone marks and MCIp-enriched regions within
###same strain and tissue by generating scatterplots.
##Merging corresponding ChIP-seq region file and MCIp-seq region file.
% mergePeaks /path-to-region-file1/C57_spl_K4me3_regions.txt \
/path-to-region-file2/C57_spl_MCIp_regions.txt > \
/path-to-output-file/C57_spl_K4me3+MCIp_regions.txt;
```

```
##Figure 5-17A: Annotating merged region files with corresponding tag information (log-
##transformed).
% annotatePeaks.pl /path-to-merged-region-file/C57_spl_K4me3+MCIp_regions.txt mm9 \
-size given –noann –log –d /path-to-tag-directory/C57_spl_input \
/path-to-tag-directory/C57_spl_K4me3 /path-to-tag-directory/C57_spl_MCIp > \
/path-to-output-file/C57_spl_K4me3+MCIp_regions.ann.txt;
#Log-transformed tag information for all (common and unique) H3K4me3 and MCIp peaks
#within one strain and tissue were plotted against each other in R on a Windows7 PC to
#generate scatterplots.


###Figure 5-17B: Creating histograms of strain-specifically enriched histone marks and
###MCIp-enriched regions by annotating strain-specific region files with tag information
###of all other marks with the homer tool pipeline.
% annotatePeaks.pl /path-to-region-file/C57_spl_K4me3 _regions.txt mm9 -size 6000 \
–hist 25 –d <all desired tag directories in a space-separted list> \
/path-to-output-file/C57_spl_K4me3.hist.txt;
#Respective tag counts were plotted in Excel 2010 on a Windows7 PC against genomic
#position around the underlying regions file to create histograms with genomic
#distributions.


###Figure 5-17C: Calculating statistics of pairwise comparisons for Venn diagrams
###homer v4.6.
% mergePeaks -d 300 /path-to-region-file1/C57_spl_K4me3_regions.txt \
/path-to-region-file2/C57_cer_K4me3_regions.txt –venn venn_C57_K4me3.txt;
#Output files contain numbers of unique and common peaks and were visualized by Venn
#diagrams.
```

4.4.1.3 **Repeat analysis with MCIp-seq data**

When NGS reads are aligned to a reference genome, in general only uniquely mapable reads are kept for subsequent analysis. However, to analyze enrichment of repetitive regions, the alignment can be adjusted to retain the best alignment of repetitive sequences.

The HOMER pipeline allowed for analysis of repetitive regions with the analyzeRepeats.pl tool, which used a repeat-annotated genome to determine read abundancy at repetitive regions. Subsequent statistical analysis with EdgeR (implemented in tool pipeline) was performed to identify differentially enriched repeat classes between C57BL/6 and BALB/c samples.

The following script for C57BL/6 spleen MCIp sample is representative for all analyzed samples. If not stated otherwise in the command line, default settings were applied.

```
###Software versions
#bowtie2       2.1.1
#homer         4.6
#R             3.1.0
#EdgeR         3.4.2


###Alignment of MCIp reads to the mm9 reference genome.
##Mapping MCIp reads to the mm9 reference genome with bowtie 2 (same as in 'Comparison of
##ChIP-seq and MCIp-seq data').
% bowtie2 -x /path-to-bowtie2-programs/indexes/bt2-mm9 /path-to-file/C57_spl_MCIp.fastq \
-S /path-to-file/C57_spl_MCIp.sam;

##Creating Tag Directories for the aligned reads with the '-keepone' option to include
##reads mapped to repetitive regions.
% makeTagDirectory /path-to-output-folder/C57_spl_K4me3 /path-to-file/C57_spl_K4me3.sam \
-genome mm9 –checkGC (-keepone);


###Table 5-5: Determing read abundancy at (a subset of) annotated repeats.
##Options:
#repeats       triggers loading of repeat definition from UCSC for the indicated reference
#genome
#-strand both  for non-strand-specific reads
#-noadj        reports the exact number of counted reads
#-L2 SINE      restrict analysis to Alu repeats)
% analyzeRepeats.pl repeats mm9 -strand both -noadj (-L3 Alu) \
 -d <space separated listing of /paths-to-tag-directories/> \
> /path-to-output-folder/MCIp_allRepeats.txt;


##Estimating differential enrichment.
#Statistical analysis using the implemented EdgeR script (with pairing of cerebellum and
#spleen samples from the same strand as de facto replicates).
% getDiffExpression.pl /path-to-input-file/MCIp_allRepeats.txt Bc Bc C57 C57 \
-repeats >/path-to-output-folder/DiffExpr_MCIp_allRepeats_replicates.txt;
#Results were further analyzed in Excel 2010.
```

4.4.2 RNA-seq data from differentiating ESCs

RNA-seq data from differentiating ESCs were generated by non-indexed paired-end 200 bp sequencing on a HiSeq1000 (Illumina) using sequencing chemistry SBSv3 (Illumina). The general NGS workflow is stated in section 4.3.1.19. Raw data was delivered as compressed fastq files
Since important RNA-seq generated from BALB/c ESCs might be omitted by mapping reads to the C57BL/6-derived *mm9* reference genome, an unbiased *de novo* assembly of the reads into transcripts was favored. The Trinity software package includes this assembly step and functions independent of any reference genome.

The following script was applied to identify differentially expressed transcripts between differentiating C57BL/c and BALB/c ESCs (see section 5.1.7).
The following script for JM8 (C57BL/6) RNA-seq data is representative for all analyzed samples.
De novo transcript assemblies were performed on the athene1.uni-regensburg.de cluster due to memory limitation on the available Linux workstation. All subsequent analyses were performed on the in house Linux workstation.

```
###Software versions
#Trinity       trinityrnaseq_r20140413p1
#bowtie        0.12.7
#RSEM          1.2.11
#EdgeR         3.4.2
#R             3.0.2


###To reduce verbosity of the command line for the de novo assemblies, all read 1 and
###read2 fastq files, respectively, from each sample were concatenated.
% cat /path-to-read1-folder_C57_d0/*_1.fastq > /path-to-output-file/C57_d0_read1.fq;
% cat /path-to-read1-folder_C57_d0/*_2.fastq > /path-to-output-file/C57_d0_read2.fq;


###De novo assembly of transcripts with the Trinity tool pipeline.
##*_read1.fastq and *_read2.fastq files were transferred to the athene cluster and
##analyzed with trinity using job scripts (example see below). Several options were used
to ##reduce the memory requirements to enable de novo assembly at the local cluster.
#Options:
#--normalize_reads
#--normalize_max_read_cov 30
#--normalize_by_read_set
#--min_kmer_cov 2     keeps only those kmers in RAM that appeared at least 2x (also
#reduces RAM requirement, but only minor impact on sensitivity)

#Exemplary job script for the de novo transcript assembly from all available JM8 samples.
>

#PBS -q serial
#PBS -l nodes=1:ppn=16
#PBS -l mem=128000m
#PBS -l walltime=48:00:00     (was manually extended by tech support if required)
#PBS -m be
#PBS -M julia2.wimmer@ukr.de

$ path-to-Trinity/Trinity --normalize_reads --normalize_max_read_cov 30 -- \
normalize_by_read_set --seqType fq --JM 120G --min_kmer_cov 2 \
--left /path-to-C57_d0_read1.fq/C57_d0_read1.fq, \
/path-to-C57_d1_read1.fq/C57_d1_read1.fq, \
/path-to-C57_d2_read1.fq/C57_d2_read1.fq, \
/path-to-C57_d4_read1.fq/C57_d4_read1.fq, \
--right /path-to-C57_d0_read2.fq/C57_d0_read2.fq, \
/path-to-C57_d1_read2.fq/C57_d1_read2.fq, \
```

```
/path-to-C57_d2_read2.fq/C57_d2_read2.fq, \
/path-to-C57_d4_read2.fq/C57_d4_read2.fq, \
--SS_lib_type FR --CPU 16 --output /path-to-output-folder/;

<
#The generated files were transferred back to the Linux work station for further
#analysis. The generated Trinity.fasta files contain all de novo assembled transcripts
#generated from the respective samples and are used in subsequent analyses (e.g. by RSEM
#for abundancy estimations or for similarity analyses).


###Quality assessment of the de novo assembled transcriptomes.
##Obtain the total number of transcripts for each assembly.
% grep -c '>' /media/disk4/JW_TRINITY/trinity_C57_norm30_kmercov2_128G/Trinity_C57.fasta

##Examine the average length coverage of annotated transcripts.
#Create a reference transcriptome database based on the mRNA sequence of annotated
#RefGenes (mm10). The required information in fasta format (refGene_mm10_mRNA.fasta) was
#obtained on the XXX from the UCSC table browser.
% makeblastdb -in /path-to-fasta-file/refGene_mm10_mRNA.fasta -dbtype nucl;

#Query the de novo assembly fasta file against the reference transcriptome database with
#megablast and high stringency (evalue <1e-20) and analyze the top database hits for the
#length coverage.
% blastn -query /path-to-C57-assemmbly-file/Trinity.fasta \
-db /path-to-reference-database/refGene_mm10_mRNA.fasta \
-out /path-to-output-folder/TrinityC57_vs_refGene_mm10_mRNA.blastn \
-evalue 1e-20 -dust no -task megablast -num_threads 6 -max_target_seqs 1 -outfmt 6;
%analyze_blastPlus_topHit_coverage.pl \
/path-to-blastn-result/TrinityC57_vs_refGene_mm10_mRNA.blastn \
/path-to-C57-assemmbly-file/Trinity.fasta \
/path-to-reference-database/refGene_mm10_mRNA.fasta \
> /path-to-output-folder/;
#The generated *.blastn.hist file contains the number of top database hits (= reference
#transcripts with at least 20% sequence coverage by de novo assembled transcripts) as well
#as the number of reference transcripts with higher coverage (in 10% bins). Here
#exemplarily shown for de novo assembled transcripts from all JM8 samples.
"#hit_pct_cov_bin     count_in_bin   >bin_below
100     4142    4142
90      1304    5446
80      1476    6922
70      1811    8733
60      2173    10906
50      2329    13235
40      2282    15517
30      2479    17996
20      2992    20988
10      0       20988
0       0       20988"
```

```
##Estimating transcript abundancy with RSEM to assess expression levels of control genes
##as well as for a general comparison of the individual RNA-seq data sets.
#Prepare the assembly files for alignment and abundancy estimation.
% align_and_estimate_abundance.pl \
--transcripts /path-to- C57-assemmbly-file/Trinity.fasta --est_method RSEM \
--aln_method bowtie --trinity_mode --prep_reference;

#Individual read abundancy estimation for each sample with RSEM.
% align_and_estimate_abundance.pl \
--transcripts /path-to- C57-assemmbly-file/Trinity.fasta \
--seqType fq --left /path-to-C57_d0_read1.fq/C57_d0_read1.fq \
--right /path-to-C57_d0_read2.fq/C57_d0_read2.fq --est_method RSEM --aln_method bowtie \
--SS_lib_type FR --trinity_mode --output_dir /path-to-output-folder/ \
--output_prefix Bc_d0 --thread_count 6;
#The generated *.genes.results and *.isoforms.results files contain inter alia the raw
#read counts for each analyzed sample (on gene and isoform level, respectively).

#Comparing estimated transcript abundancies of individual samples of each strain with
#EdgeR.
% abundance_estimates_to_matrix.pl --est_method RSEM --out_prefix <gene or isoform> \
<space-separated list of required *.genes.results/*.isoforms.results files> \
> /path-to-output-folder/;
% run_DE_analysis.pl --matrix /path-to-afore-generated-read-count-matrix/*.counts.matrix -
-method edgeR;
#For each pairwise comparison of samples represented in the expression matrix a MA plot is
#generated. The MA plots for pairwise comparison of BALB/c samples or JM8 samples based on
#the strain-specific de novo transcript assembly are depicted in Figure 5-13.

##Estimating transcript abundancy of selected control genes.
#Screening the de novo assembled transcriptomes for transcripts similar to selected
#control genes: Query the assembly fasta files against a database of annotated
#Mus musculus proteins (here: annotated RefSeq proteins[273]).
#Create a database for blastx.
% segmasker -in /path-to-protein-fasta-file/mouse.protein.faa -infmt fasta \
-parse_seqids -outfmt maskinfo_asn1_bin \
-out /pat-to-output-folder/refseq_mouse.protein_seg.asnb;
% makeblastdb -in / path-to-protein-fasta-file/mouse.protein.faa -dbtype prot \
-parse_seqids -mask_data /path-to-masking-file/refseq_mouse.protein_seg.asnb \
-out /path-to-output-folder/refseq_mouse.protein -title "RefSeq Mouse Protein Database";

#Check the created database to retrieve masking algorithm ID.
% blastdbcmd -db /path-to-database/refseq_mouse.protein -info;
"Database: RefSeq Mouse Protein Database
       77,941 sequences; 51,201,547 total residues

Date: Jul 28, 2014  1:32 PM   Longest sequence: 35,463 residues
Available filtering algorithms applied to database sequences:
Algorithm ID  Algorithm name      Algorithm options
21            seg                 window=12; locut=2.2; hicut=2.5

Volumes:
       /path-to-database/refseq_mouse.protein"
```

```
#Create a list (one entry per line) of GI number of the proteins of interest (retrieved
#from original mouse.protein.faa fasta file of annotated Mus musculus RefSeq proteins).

#Query the assembly fasta files against the created database for the selected genes
#represented by the GI list with high stringency (evalue <1e-20).
% blastx -query /path-to-de-novo-assembly-fasta-file/Trinity_C57.fasta \
-gilist /path-to-gilist/ControlGenes_GIlist -db /path-to-database/refseq_mouse.protein \
-outfmt 7 -out /path-to-output-folder/C57_blastx_controlgenes.out -db_soft_mask 21 \
-evalue 1e-20 -num_threads 6;
% grep -e "^c" /path-to-similar-transcripts-file/C57_blastx_controlgenes.out \
> /path-to-output-folder/all_blastx_controlgenes_hits.out;
% sed -e 's/^\([^\t]*\).*/\1/' /path-to-hits-file/C57_blastx_controlgenes_hits.out \
> /path-to-output-folder/C57_blastx_controlgenes_hits.short.out;
#The generated *_hits.short.out file contains the list of de novo assembled transcripts
#with high similarity to proteins listed in the GI list. These identifiers were then used
#to retrieve the expression values from the respective read count matrix generated before.

##Assess representation (coverage) of control genes by de novo assembled transcripts by
##retrieving sequences of similar transcripts from de novo assembly fasta file.
% sed -e 's/^\([^\t]*\).*/\1/' /path-to-de-novo-assembly-fasta-file/Trinity.fasta \
> /path-to-output-folder/Trinity.short.fasta;
% xargs samtools faidx /path-to-shortened-fasta-file/Trinity.short.fasta \
< /path-to-list-of-selected-transcripts/C57_blastx_controlgenes_hits.short.out \
> /path-to-output-folder/Trinity_C57_controlgenes.short.fasta;
#The obtained transcript sequences were then subject to UCSC's blat algorithm
#(www.ucsc.com).

###Analyzing transcripts similar to the candidate epigenetic modifier Nol10.
##The same protein database was queried with the same settings as for control genes. The
##GI list was adjusted to query only for transcript similar to Nol10 (GI identifier
##retrieved from original mouse.protein.faa fasta file of annotated Mus musculus RefSeq
##proteins).
% blastx -query /path-to-de-novo-assembly-fasta-file/Trinity_C57.fasta \
-gilist /path-to-gilist/Nol10_GIlist -db /path-to-database/refseq_mouse.protein \
-outfmt 7 -out /path-to-output-folder/C57_blastx_Nol10.out -db_soft_mask 21 \
-evalue 1e-20 -num_threads 6;
% grep -e "^c" /path-to-similar-transcripts-file/C57_blastx_Nol10.out \
> /path-to-output-folder/all_blastx_Nol10_hits.out;
% sed -e 's/^\([^\t]*\).*/\1/' /path-to-hits-file/C57_blastx_Nool10_hits.out \
> /path-to-output-folder/C57_blastx_Nol10_hits.short.out;
#The generated *_hits.short.out file contains the list of de novo assembled transcripts
#with high similarity to the Nol10 protein. These identifiers were then used to retrieve
#the expression values from the respective read count matrix generated before. Retrieving
#the corresponding transcript sequences from the original Trinity.fasta file was performed
#as for the control genes.
```

4.4.3 General representation of the *MIso*-encoding candidate region by BAC clones

Sequencing data of a pool of BACs were generated by indexed (6 samples/lane) single 50 bp sequencing on a HiSeq1000 (Illumina) using sequencing chemistry SBSv3 (Illumina). The general NGS workflow is stated in section 4.3.1.19.
The following script was applied to evaluate the coverage of region *mm9*_chr12:16902483-26024113 selected by BAC library screening (see section 4.3.1.20 & 5.1.6.2).

```
###Software versions
#bowtie2       2.1.0
#samtools      0.1.18
#bedtools      v2.16.2
#R             3.0.2


###Indexing reference genomes for bowtie2.
##For chromosome 12 (mm9)
##(http://hgdownload.cse.ucsc.edu/goldenPath/mm9/chromosomes/chr12.fa.gz)
% bowtie2-build /path-to-fasta-file/mm9_chr12.fa \
/path-to-bowtie2-programs/indexes/mm9_chr12only;

##For E. coli genome (E. coli K12, NC_000913.3)
##(http://www.ncbi.nlm.nih.gov/nuccore/556503834?report=fasta)
% bowtie2-build /path-to-fasta-file/EcoliK12.fa \
/path-to-bowtie2-programs/indexes/EcoliK12;

###Removing reads generated by E. coli gDNA contamination with bowtie2.
% bowtie2 -p7 -x EcoliK12 -U /path-to-fastq-file/BAC_Mix_159.fastq \
S /path-to-alignment-output-file/alignmentEcoliK12.sam \
--un /path-to-file-for-unaligned-reads/BACmix159_woEcolireads.fastq;
#The BACmix159_woEcolireads.fastq contained raw reads removed from E. coli contaminations
#and was further processed.

###Mapping E. coli-depleted raw reads to chromosome 12 with bowtie2.
% bowtie2 -p7 -x mm8_chr12only -U /path-to-fastq-file/BACmix159_woEcolireads.fastq \
-S /path-to-alignment-output-file/BACmix159_woEcoli_chr12only.sam;

###Extracting reads, which mapped to candidate region (mm9_chr12:16902483 26024113).
##Converting SAM file to BAM file with samtools.
% samtools view -bSh /path-to-SAM-file/BACmix159_woEcoli_chr12only.sam > \
path-to-output-file/BACmix159_woEcoli_chr12only.bam;
##Converting BAM file to BED file.
% bamToBed -i /path-to-BAM-file/BAC159aligned_woEcoli_chr12only.bam > \
/path-to-output-file/BAC159aligned_woEcoli_chr12only.bed;
```

```
##Removing all reads not mapped to candidate region.
#BAC159aligned_woEcoli_chr12only.bed was opened in text editor and positions
#< chr12:16902483 were manually removed and saved as
#BAC159aligned_woEcoli_chr12only_16902483-end.bed. Positions greater chr12:26024113 were
#removed using the sed command.
% sed -n '1,/HWI-1KL143:88:D2DVPACXX:7:1101:9955:67001/ p' \
/path-to-BED-file/BAC159aligned_woEcoli_chr12only_16902483-end.bed > \
/path-to-output-file/BAC159aligned_woEcoli_chr12only_16902483-26024113.bed;
#'HWI-1KL143:88:D2DVPACXX:7:1101:9955:67001' was read name of first read, which mapped to
#position > chr12: 26024113 (looked up manually).


###Removing reads mapped to known repeats (mm9) using bedtools ((rmsk_mm9.bed created with
UCSC table browser).
% intersectBED -v \
-a /path-to-BED file1/BAC159aligned_woEcoli_chr12only_16902483-26024113.bed \
-b /path-to-BED-file2/rmsk_mm9.bed > \
/path-to-output-file/BAC159aligned_woEcoli_chr12only_16902483-26024113_worepeats.bed;


###Adjusting read length (50 bp) to average fragment length (170 bp, identified by size
###distribution on a Bioanalyzer DNA HighSensitivity Chip) with bedtools.
% bedtools slop \
-i /path-to-bed-file/BAC159aligned_woEcoli_chr12only_16902483-26024113_worepeats.bed \
-g /path-to-bedtools-programs/genomes/mouse.mm9.genome -s -r 120 -l 0 > \
path-to-output-file/BAC159aligned_woEcoli_chr12only_16902483-
26024113_worepeats_120bp_added.bed;


###The following steps were done for both BED files
###(BAC159aligned_woEcoli_chr12only_16902483-26024113_worepeats.bed,
###BAC159aligned_woEcoli_chr12only_16902483-26024113_worepeats_120bp_added.bed).

##Creating a read count table (one entry per base) for candidate region
##(mm9_chr12:16902483 26024113) excluding known repeats and assembly gaps of mm9
##(rmsk_mm9.bed & gaps_mm9.bed created with UCSC table browser).
% bash -c 'for i in {16902483..26024113}; do echo \
-e "chr12\tjust\tgenomicposition\t$i\t$i\t1000\t.\t.\t$i"; done' > \
/path-to-output-file/9MB.gff;
#(subsequently add first line [seqname generated by feature start stop score
#strand frame group] manually to 9MB.gff).
% intersectBed -v -a /path-to-file1/9MB.gff -b /path-to-file2/gaps_mm9.bed > \
/path-to-output-file/9MB_wo_gaps.gff;
% intersectBed -v -a /path-to-file1/9MB_wo_gaps.gff -b /path-to-file2/rmsk_mm9.bed > \
/path-to-output-file/9MB_wo_gaps_wo_repeats.gff;
```

```
##Counting reads from BED files into read count table with bedtools.
% bedtools coverage \
-a /path-to-BED-file/BAC159aligned_woEcoli_chr12only_16902483-26024113_worepeats.bed \
-b /path-to-gff-file/9MB_wo_gaps_wo_repeats.gff > \
/path-to-output-file/countsperbase_wor_nonadj.gff
% bedtools coverage \
-a /path-to-BED-
file/BAC159aligned_woEcoli_chr12only_16902483-26024113_worepeats_120bp_added.bed \
-b /path-to-gff-file/9MB_wo_gaps_wo_repeats.gff > \
/path-to-output-file/countsperbase_wor_120.gff

##Counting bases with zero reads and non-zero reads with R to estimate BAC coverage of
##candidate region. The following script is representative for nonadjusted and fragment
##size-adjusted reads.
#Loading read count table with read counts into R.
% countsperbase_wor_nonadj<-
read.table(file="/media/disk3/JuliaW_Analysis/BAC159mix_50bp_se/coverage/countsperbase_wor
_nonadj.gff", nrows=4487730,
col.names=c("seqname","source","feature","start","end","score","strand","frame","group","r
eads","non-zero bases","featurelength","%non-zero bases"));
#Retrieving numerical vector of read counts.
% reads <-countsperbase_wor_nonadj$reads;
#Creating a logical vector (TRUE/FALSE) for reads_wor_nonadj regarding entries with value
#'0'.
% zero.reads<- reads ==0;
#Returning total numbers of TRUE (= zero BAC reads) and FALSE (= non-zero BAC reads)
#entries in vector 'zero.reads'.
% table(zero.reads);

##Creating coverage plot with R.
#Loading read count table with read counts into R.
% countsperbase_wor_120<-
read.table(file="/media/disk3/JuliaW_Analysis/BAC159mix_50bp_se/coverage/countsperbase_wor
_130.gff", nrows=4487730,
col.names=c("seqname","source","feature","start","end","score","strand","frame","group","r
eads","non-zero bases","featurelength","%non-zero bases"));
#Retrieving numerical vectors of read counts and base position.
% base<-countsperbase_wor_120.gff$start
% reads<-countsperbase_wor_120.gff$reads
% plot(base,reads, xlab="chr12:16902483-26024113 (mm9)", ylab="read per base", type="l")

#Calculating average read coverage of covered (non-zero) bases.
% nonzero_countsperbase_wor_120<-countsperbase_wor_120[!countsperbase_wor_120$reads==0, ]
% mean(nonzero_countsperbase_wor_120)
```

5 Results

5.1 Identification of *MIso*: a strain-specific epigenetic modifier, which regulates DNA methylation at the *Isoc2b* promoter

The effects of DNA methylation have been subject of extensive studies, but still very little is known about the extent, establishment, and inheritance of individual patterns of this mark. Inbred mice represent a well-suited model organism to address these questions. They have a distinct genetic background and reproducible breeding conditions minimize environmental effects on epigenetic modifications.

Preliminary work in our laboratory identified the nuclear *Isoc2b* locus as one of numerous mouse strain-specifically methylated regions. Depending on the methylation status of its promoter, the *Isoc2b* gene, which encodes for the isochorismatase domain containing 2b protein, is differentially expressed in inbred mouse strains[261]. Analyses regarding the inheritance of the methylation status suggested a regulation in *trans* by a yet unknown epigenetic modifier.

Since only few site-specific epigenetic regulators are known so far, the identification of further factors, their targets as well as their mode of action are of great interest. The present work aimed at the identification of the strain-specific epigenetic modifier responsible for the differential DNA methylation pattern at the *Isoc2b* promoter.

5.1.1 Preliminary work

The experimental work, which identified the differentially methylated region at the *Isoc2b* promoter, was published in the journal Genome Research[261]. In summary, microarray-based expression analysis of bone marrow-derived macrophages (BMM) from inbred C57BL/6J and BALB/cAnNCrl mouse identified *Isoc2b* to be C57BL/6-specifically expressed, which was confirmed by RT-qPCR. The *Isoc2b* locus was analyzed regarding strain-specific DNA methylation (together with other differentially expressed genes) using a comparative Methyl-CpG immunoprecipitation(MCIp)-on-ChIP approach established in our laboratory. A representative set of 18 differentially methylated regions (DMRs) was chosen for validation and further quantitative methylation analysis at single CpG dinucleotides (CpG) resolution by Matrix-Assisted Laser Desorption/Ionisation combined with mass spectrometry analysis by Time-of-flight (MALDI-TOF

MS) (EpiTYPER technology, Sequenom). This approach facilitates simultaneous quantification of the methylation level at multiple CpGs across various samples (see method section 4.3.1.14 for more details).

The selected DMRs comprised promoter-associated, intergenic as well as intragenic regions. Average DNA methylation levels at the selected DMRs correlated well in BMM and spleen of male and female mice of both strains excluding cell culture artefacts as well as gender-dependent methylation patterns[261].

Since the inter-generational inheritance of epigenetic patterns is poorly understood, DNA methylation levels of somatic tissues and germ line cells (testis) were compared between wild type mice and F1 hybrids (reciprocal crosses). The majority of the validated DMRs seemed to be re-established within each generation. The differential DNA methylation patterns detected in somatic cells were scarcely found in germ line cells. In addition, DNA methylation levels in somatic tissues from F1 hybrids showed an intermediate degree of methylation compared to parental wild type strains. In most cases, MALDI-TOF MS-derived average DNA methylation level of individual CpGs in F1 hybrids were similar to the mean parental methylation levels indicating allele-specific inheritance of DNA methylation patterns. Only three of the analyzed DMRs displayed a loss or gain of methylation in F1 hybrids (Figure 5-1A).

To address the notion of allele-specific inheritance, bisulfite sequencing (see section 4.3.1.13), which is able to determine the methylation status of individual CpGs within their genomic context, is required. The EpiTYPER technology only allows for average quantification of the methylation level of individual CpGs. However, allele-specific bisulfite sequencing depends on allelic sequence variations still detectable after bisulfite treatment. Not all DMRs were suitable for allele-specific bisulfite sequencing, due to the lack of informative sequence variations. In total, six regions were analyzed including the *Isoc2b* promoter. Except for *Isoc2b*, all analyzed regions showed allele-specific DNA methylation patterns, i.e. their methylation status seemed to be determined by local sequence variation (*cis* regulation). The *Isoc2b* promoter exhibited reduced, but unbiased methylation of both alleles in F1 hybrids. These findings argued for an epigenetic modifier regulating the methylation status in *trans* (Figure 5-1B).

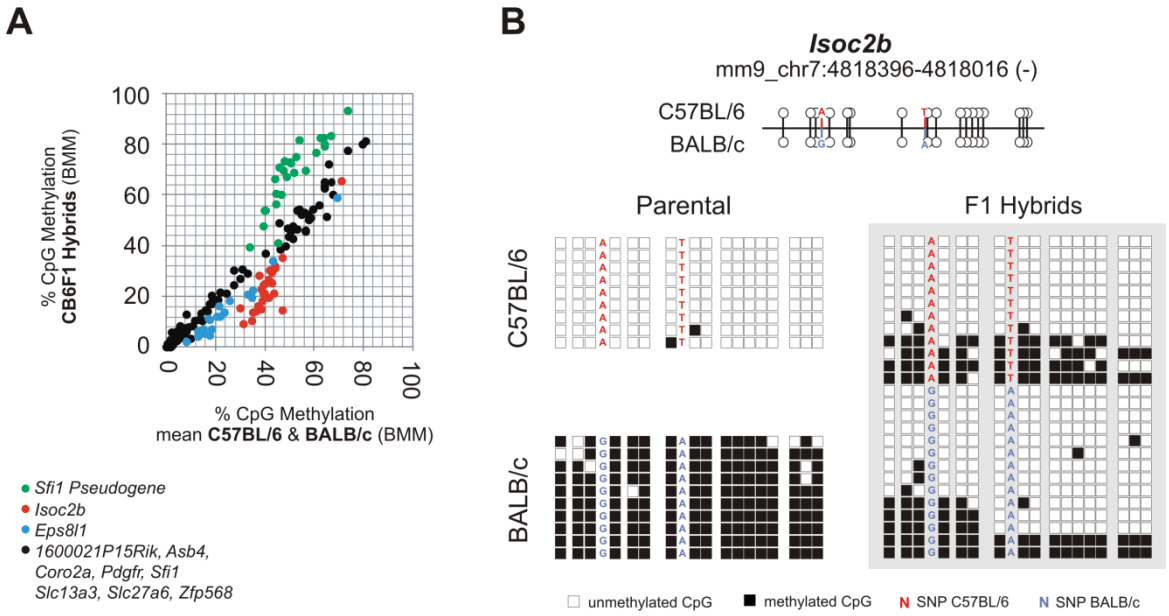


Figure 5-1 ***Preliminary work – With only few exceptions, strain-specifically methylated regions are regulated in* cis**

(A) Methylation levels of individual CpGs were assessed by MALDI-TOF MS in BMM of wild type mice (n=3 for each strain) and F1 hybrids (n=5). Mean average CpG methylation levels of parental BMM are plotted against average CpG methylation levels of F1 hybrids (exemplarily shown for CB6F1 hybrids) for individual CpGs. Regions that either gained or lost methylation in F1 hybrids compared to the mean parental methylation level are highlighted. **(B)** Allele-specific bisulfite sequencing of *Isoc2b* promoter in BMM from C57BL/6, BALB/c, and F1 hybrids (reciprocal crosses). The genomic position of CpGs within the amplicons is shown at the top. The sequence variations used to distinguish the different parental alleles are marked in red (C57BL/6) and blue (BALB/c). Individual CpGs are represented by either white (unmethylated) or black (methylated) squares. Lines of squares represent independently sequenced clones derived from two independent sample preparations. (Adopted from Schilling *et al.* [261])

The general notion drawn from these findings was that DMRs were mainly regulated in *cis* and re-established after fertilization. Interestingly, bisulfite sequencing of the DMR at the *Isoc2b* promoter indicated regulation in *trans* by an epigenetic modifier rather than in *cis* by the local sequence. This was of particular interest, since epigenetic modifiers have been scarcely characterized as yet. With the present work, I aimed at the identification and characterization of the responsible epigenetic modifier.

5.1.2 Epigenetic profile of the *Isoc2b* promoter region in somatic tissues

To recapitulate the preliminary MCIp-on-Chip data for the *Isoc2b* promoter region, I performed methylated CpG immunoprecipitation (MCIp) experiments with spleen- and cerebellum-derived genomic DNA and assessed the enriched DNA fragments by next generation sequencing (MCIp-seq). To address additional epigenetic modifications, next generation sequencing data of chromatin immunoprecipitation (ChIP-seq) of active histone marks (H3K4me1/3, H3K27ac) for both tissues from C57BL/6 and BALB/c mice were generated (both data sets are further discussed in section 5.2.4). In line with the published data of Schilling *et al.* [261], only C57BL/6-derived tissues were marked with active histone marks, whereas enrichment of highly methylated fragments by MCIp was only detected in BALB/c-derived tissues (Figure 5-2A).

Interestingly, the *Isoc2b* homolog, *Isoc2a*[274], which is located downstream of *Isoc2b* in a head-to-head orientation, did not display strain-specific epiphenotypes (Figure 5-2A) and was expressed in both mouse strains (Figure 5-2B). Moreover, BALB/c seemed to express *Isoc2a* at a level twice as high as C57BL/6, thereby potentially compensating the lack of *Isoc2b* expression. The comparison of the nucleotide sequences (1000 bp) upstream of the coding sequences of both homologs by Blastn (NCBI) revealed no significant sequence similarities.

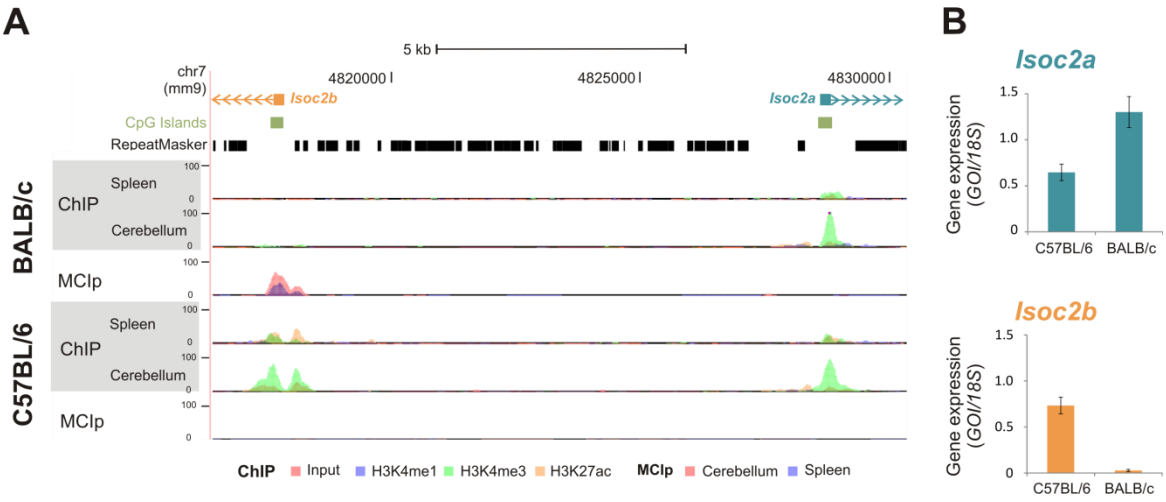


Figure 5-2 ***Epigenetic profiles of the* Isoc2b *and* Isoc2a *promoter regions, respectively, are in line with the detected expression statuses***

(A) The represented 15 kb window displays the promoter regions of the duplicated genes *Isoc2a* (right) and *Isoc2b* (left). Enrichment of regions by MCIp and ChIP (H3K4me1/3, H3K27ac) experiments with spleen and cerebellum are indicated by color-coded peaks. **(B)** Expression levels of each of the two duplicated genes of interest (GOI) were assessed by RT-qPCR with total RNA from spleen tissue (n=2) and normalized to 18S ribosomal RNA (error bars represent standard deviation).

5.1.3 The *Isoc2b* promoter region displays two distinct DNA methylation phenotypes in homozygous inbred strains

To further test the notion of regulation in *trans* by an epigenetic modifier, I analyzed the *Isoc2b* promoter methylation phenotype in spleen DNA from six additional inbred mouse strains (129S1/SvImJ, A/J, C3H/HeJ, CBA/J, DBA/2J, NOD/ShiLtJ) by MALDI-TOF MS. Five of the six strains displayed a hypomethylated phenotype at the *Isoc2b* promoter region similar to C57BL/6. Only A/J exhibited a hypermethylated phenotype similar to BALB/c (Figure 5-3A). According to the Mouse Genomes Project[275] (REL-1003), all six strains display BALB/c SNP and indel variants at those positions that distinguish BALB/c and C57BL/6 mice in the analyzed region (*mm9*_chr7:4817304-4818396). These findings further argue against a regulation in *cis* by the local sequence.

DNA methylation analysis in substrains of BALB/c and C57BL/6 revealed consistent strain-specific methylation phenotypes for the *Isoc2b* promoter region (Figure 5-3B).

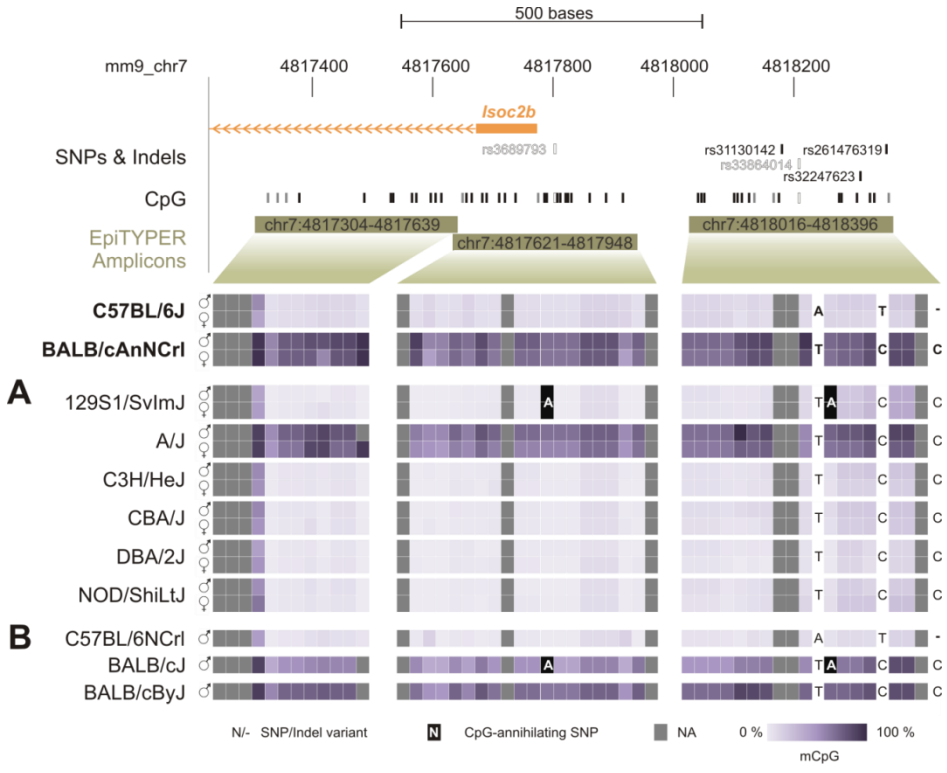


Figure 5-3 ***DNA methylation analysis for the* Isoc2b *promoter region in common inbred mouse strains reveals two distinct phenotypes***

Methylation levels of individual CpGs were obtained by MALDI-TOF MS analysis of bisulfite-treated DNA in spleen of indicated mouse strains and from pale purple (0% methylation) to dark purple (100% methylation). The relative position of sequence variations, CpGs, and EpiTYPER amplicons within the *Isoc2b* promoter region as well as the detected methylation levels for the generally used BALB/c and C57BL/6 strain are indicated at the top. **(A)** *Isoc2b* promoter methylation in spleen from six common inbred strains (n=2 per gender). **(B)** *Isoc2b* promoter methylation in spleen of the BALB/cJ and BALB/cByJ substrains (2 technical replicates each) as well as the C57BL/6N substrain (n=2).
Strain-specific variants of indicated SNPs and Indels are positioned relative to analyzed CpGs.

5.1.4 Linkage analysis maps *MIso* to chromosome 12

Preliminary data revealed distinct differential DNA methylation levels at the *Isoc2b* promoter region in BMM from C57BL/6, BALB/c, and F1 hybrids (Figure 5-1B). In F1 hybrids, both parental alleles display similar DNA methylation patterns excluding *cis* regulation by local sequence differences. Furthermore, the low methylation level observed in F1 hybrids could be explained by a gene dose effect. These findings support the notion of *trans* regulation by an epigenetic modifier. To identify the *modifier of* Isoc2b (*MIso)*, I mapped its approximate localization within the mouse genome by linkage analysis (LA).

5.1.4.1 Principles of linkage analysis

Genetic markers, for example SNPs, are polymorphic features of the genomic sequence that can be followed over generations and enable to distinguish individuals by their alleles. The characteristics of these markers are accessible by various genotyping approaches, such as pyrosequencing[276] and mass spectrometric assays (i.e. iPLEX technology, see section 4.3.1.14).

Two genetic markers located on different chromosomes will be inherited independently and separated with 50% probability in the next generation according to the 2nd Mendelian Law (*Law of Independent Assortment*)[277]. Genetic markers on the same chromosome are not inherited independently[278,279], but can be separated in the next generation due to meiotic recombination events during gametogenesis[191]. The probability of linked inheritance increases with declining distance between two genetic markers on the same chromosome due to a decreased frequency of recombination events between the two markers[279].

The same accounts for any genetic trait. Located in close vicinity, a phenotype-causing genetic trait and a genetic marker will be inherited together more frequently than expected by random chance (linkage disequilibrium)[280]. This phenomenon provides the basis for LA, an approach to investigate the shared inheritance of selected genetic markers among 1st degree relatives, which are similar regarding the phenotypic characteristic of interest[281]. If a marker is not located close to the phenotype-causing genetic trait, its alleles will be randomly distributed among the individuals. With decreasing distance to the genetic trait, the allele frequency observed in 1st degree relatives grouped according to the relevant phenotype will diverge from the expected frequency.

Murine backcrosses, i.e. offspring of a heterozygous individual mated with a homozygous individual, are particularly suitable for LA. The haploid gametes arising from heterozygous individuals display a mosaique array of the heterozygous parental diploid genome due to meiotic recombination between homologous chromosomes. Recombination also takes place in the homozygous individuals, but is without consequence resulting in a haploid genome similar to the homozygous parental genome. If the heterozygous parent represents a F1 hybrid (generated by mating of two homozygous inbred strains), the expected allele frequency is 50%. This assumption also accounts for the generally expected genotype frequency of individual genetic markers (heterozygous versus homozygous) in backcrosses (1st Mendelian Law – *Law of Segregation*) grouped according to the phenotype of interest. However, regions linked to the phenotype-causing

genetic trait will display an increased number matching (*concordant*) phenotypes-genotype combination (*linkage disequilibrium*). Based on the observed concordant phenotypes-genotype combinations, the logarithm of the odds (LOD) score is used to estimate the likelihood that two loci (e.g. genetic marker and genetic trait) are subject to linked inheritance during meiosis and thus located in close vicinity[282].

$$LOD = log_{10} \frac{1 - (\binom{n}{k} p^k (1-p)^{n-k})}{\binom{n}{k} p^k (1-p)^{n-k}}$$

n	total number of analyzed individuals
k	number of individuals with concordant phenotype/genotype combinations
p	expected allele frequency for independent inheritance

Genetic markers not linked to the phenotype-causing genetic trait will return similar numbers of concordant and discordant phenotype-genotype combination resulting in low LOD scores. In contrast, linked genetic markers will display a high level of concordance resulting in a high LOD score. LOD scores higher than three indicate linkage. It implies that the observed ratio of concordant and discordant combinations is with a 10^3-fold higher probability caused by linked rather than by independent inheritance.

Evidences support the existence of recombination hotspots as well as linkage blocks of different size that are always inherited together[280]. However, when calculating the requirements for LA, a genome-wide random and unbiased appearance of recombination events is assumed. In mouse, the total length of the genetic map was determined as 1361 cM[283]. Taking the size of the mouse genome (2.58 Gb[284]) into account, one cM represents on average 2 Mb. The theoretical number of backcrosses required for a specific linkage map resolution can be calculated according to the following formula:

$$N = \frac{100}{x\ Mb / 2\ Mb}$$

N	number of required backcrosses
x	requested resolution

5.1.4.2 Genome-wide linkage analysis in C;B6 hybrids

To map the assumed *trans*-acting epigenetic modifier responsible for the strains-specific *Isoc2b* promoter methylation, I performed linkage analysis in backcrossed C;B6 hybrids ((BALB/cxC57BL/6)F1 x BALB/c). F1 hybrids were generated from the C57BL/6NCrl substrain due to availability at the breeding center (Charles River Laboratories). Although the DMR was initially identified in C57BL/6J mice, DNA methylation analysis in C57BL/6NJ (Figure 5-3B) as well as derived F1 hybrids (Figure 5-4) revealed similar DNA methylation levels in the two substrains and their offspring. Recently, C57BL/6NJ, BALB/cJ as well as 15 other commonly used mouse

inbred strains were subject to resequencing by the Mouse Genomes Project of the Wellcome Sanger Institute[275]. The provided database was used to select informative genetic markers for LA.
As described above, backcrossed C;B6 hybrids (CB6F1 x BALB/c) displays a unique alteration of homozygous and heterozygous regions due to meiotic recombination in the F1 parent and represent 1st-degree relatives suitable for LA. Since the DMR at the *Isoc2b* promoter was also detected in spleen, which is enriched for macrophages (Figure 5-3), LA was performed with spleen-derived genomic DNA. The breeding strategy to generate C;B6 hybrids was restricted to BALB/c sires, since allelic methylation levels in F1 hybrids were not influenced by the parental origin of the alleles (non-imprinted locus) (Figure 5-1B).

Backcrosses display two distinct Isoc2b *promoter methylation phenotypes*

Taking advantage of high-throughput DNA methylation analysis with the EpiTYPER technology, the methylation frequency at the *Isoc2b* promoter was easily accessible. Assays applied in other studies to examine differential DNA methylation patterns included bisulfite sequencing and a combination of mCpG-sensitive digestion, southern blotting, and (radioactive) site-specific probing. These approaches mainly relied on the methylation status of only few CpGs[285-287]. In contrast, this non-radioactive method yields reproducible quantitative average levels of DNA methylation of multiple CpGs of numerous samples within a short time and requires only small amounts of input DNA (see section 4.3.1.14).
Assuming regulation in *trans* by a monogenetic trait, C;B6 hybrids displaying high DNA methylation levels at the *Isoc2b* promoter were expected to be homozygous for the genetic trait (similar to BALB/c mice). Vice versa, individuals with intermediate methylation levels were expected to be heterozygous for the genetic trait (similar to the F1 hybrids) (compare with Figure 5-1B). These two genotype-phenotype combinations represent concordant combinations.
Average DNA methylation levels at the *Isoc2b* promoter region (*mm9*_chr7:4818016-4818396) of spleen genomic DNA from BALB/c mice (n=6), F1 hybrids (both reciprocal crosses; CB6F1: n=5 / B6CF1: n=3), and an initial set of 45 C;B6 hybrids (10-12 weeks old) were acquired by MALDI-TOF MS. Based on the average DNA methylation frequency, I calculated an initial phenotype discriminator of 58.5% average DNA methylation by clustering the individuals with the *Partitioning Around Medoids* (PAM) Algorithm (number of groups *k*=2) (“intermediate”/“high”, Figure 5-4). The discriminator was used to group the C;B6 hybrids for genome-wide linkage analysis (see section 5.1.4.2).

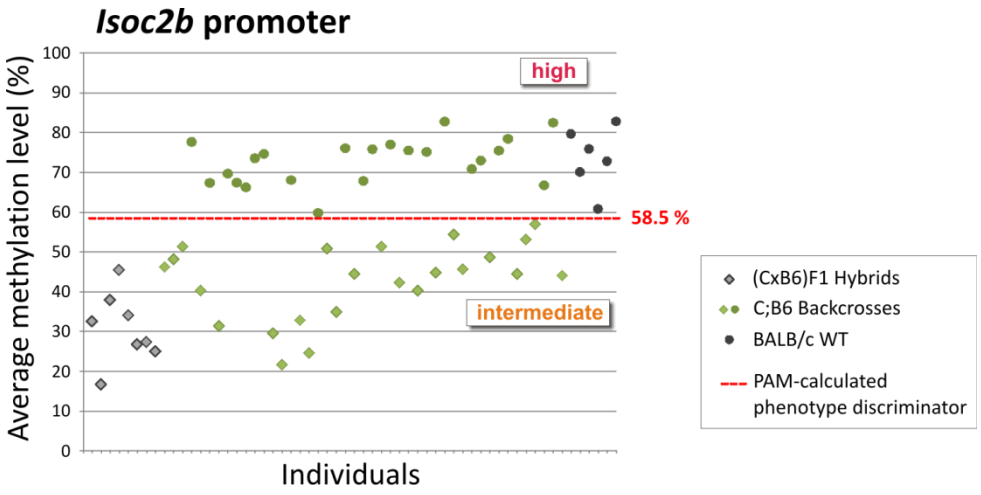


Figure 5-4 C;B6 backcrosses display two distinct* Isoc2b *promoter methylation phenotypes
Mean DNA methylation of the *Isoc2b* promoter region was determined by MALDI-TOF MS in spleen DNA from an initial set of 45 C;B6 backcrosses (CB6F1 x BALB/c) and compared to BALB/c wild type (WT) mice ("high" methylation phenotype, n=6) and F1 hybrids ("intermediate" methylation phenotype, CB6F1: n=5 / B6CF1: n=3). Values represent averages of all measured CpGs within chr7:4818016-4818396 (*mm9*). A methylation level of 58.5% was calculated by PAM clustering (*k*=2) to separate the analyzed C;B6 hybrids in two distinct groups with either an "intermediate" or a "high" methylation phenotype.

SNP genotyping deciphers genetic mosaicism in C;B6 hybrids

Using the resequencing data available from the Mouse Genomes Project[275] (REL-1003), I chose C57BL/6- and BALB/c-discriminating SNPs for an initial genome-wide linkage analysis. The selected SNPs needed to fulfill the criteria listed in section 4.3.1.14 to reduce the probability of being uninformative. In total, 265 SNPs located across the entire mouse genome were selected and genotyped with the iPLEX approach (Sequenom, see section 4.3.1.14). The major advantage of this genotyping approach is its multiplexing capacity; up to 36 distinct genetic positions in one reaction can be genotyped, while only requiring minimal amounts of input material (10 ng DNA per reaction). The use of mass-modified terminator-tagged nucleotides in a single base extension (SBE) MassEXTEND assay enables precise detection of SNP variants and facilitates reliable discrimination of homozygous and heterozygous genotypes (AA versus aa versus Aa). The genomic context of each SNP was obtained from the *mm9* reference genome. Assays for the selected SNPs were designed and multiplexed with Typer4.0 software (Sequenom).

The initial panel of 265 SNPs was evaluated by genotyping five individuals of each wild type strain (C57BL/6 and BALB/c) as well as four F1 hybrids. All assays were further tested for perturbing primer dimerization in a template-free reaction. Of the initial SNP panel, 238 SPNs proved suitable for linkage analysis. All 45 C;B6 hybrids were genotyped for the evaluated reduced SNP panel. The individuals were grouped according to the determined phenotype discriminator of 58.5% (Figure 5-4). The LOD score for each genetic marker was calculated based on the observed numbers of concordant and discordant phenotype-genotype combinations for the 45 C;B6 hybrids. As depicted in Figure 5-5A, a subset of adjacent SNPs showed increased LOD scores. The highest LOD score (13.5) was determined for SNP rs29160274. For this marker, the PAM-

calculated phenotype discriminator of 58.5% average DNA methylation yielded exclusively concordant phenotype-genotype combination indicating a more than 10^{13}-fold higher probability of linked compared to independent assortment. Although another region rendered increased LOD scores of up to 3.8, the strong indication for linkage to SNP rs29160274 supported the idea of a monogenetic trait responsible for the observed strain-specific DNA methylation phenotypes. These findings suggested that the genetic trait encoding for *MIso* is most probably located in a 19 Mb region on chromosome 12, which is confined by the SNPs rs29180275 and rs4229294 (*mm9*_chr12:7990826-26992720). The genotyping and phenotyping results as well as the corresponding LOD scores are listed in Table 10-1 (see Appendix I).

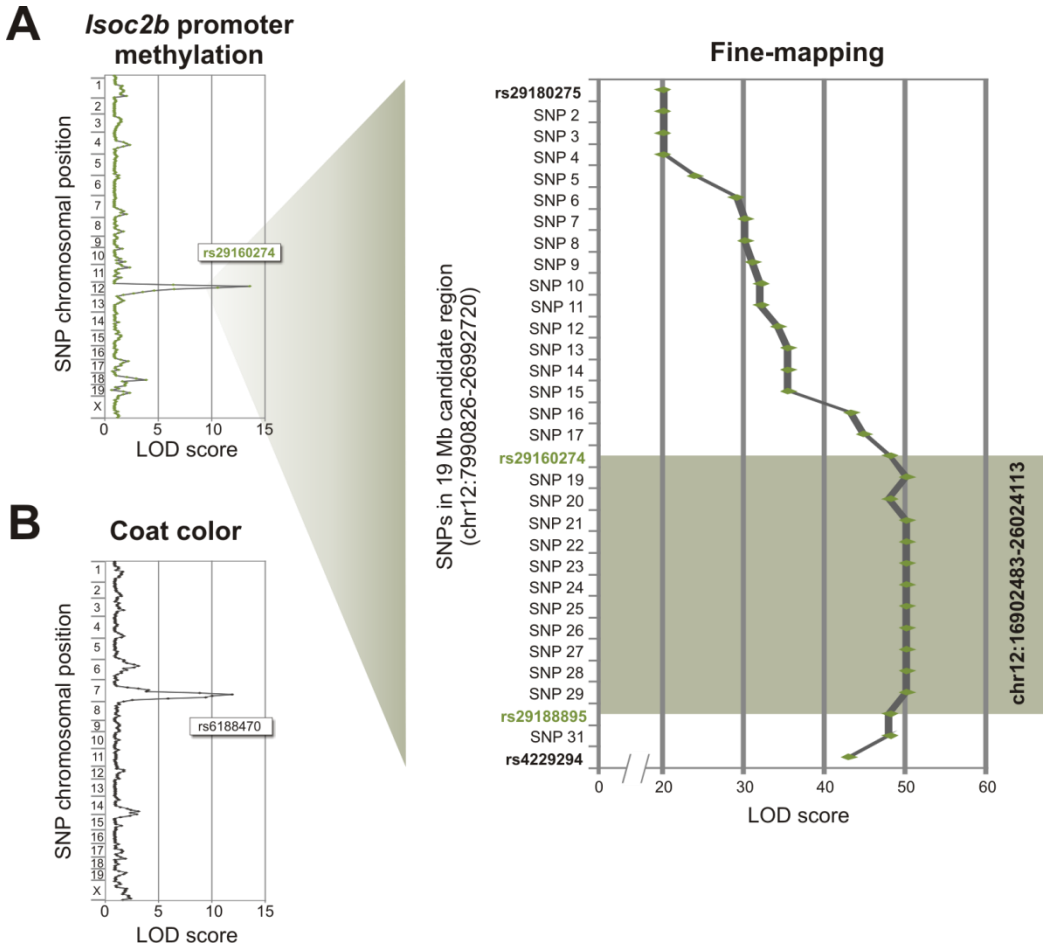


Figure 5-5 ***Linkage analyses in C;B6 hybrids reliably maps* MIso *to a 9 Mb region on chromosome 12***

(A) LOD scores indicate linkage probabilities between genetic markers and *Isoc2b* promoter methylation (*mm9*_chr7:4818016-4818396). For genome-wide mapping (left), 238 genome-wide SNPs were genotyped in spleen-derived genomic DNA from 45 C;B6 hybrids. Absolute concordance (45 of 45 individuals) of zygosity and *Isoc2b* methylation phenotype was detected for SNP rs29160274. Fine mapping the initial 19 Mb candidate region surrounding SNP rs29160274 (right) was performed with 184 additional C;B6 hybrids at 29 additional SNPs. **(B)** Linkage analysis in the initial set of 45 C;B6 hybrids positively maps the genetic trait for the differential coat color of C57BL/6 (black) and BALB/c (white) (*Tyr*, *mm9*_chr7:94575915-94641921) to a 17 Mb region on chromosome 7 (*mm9*_chr7:82114282-98849561).

To evaluate the linkage analysis approach, I used the coat color of the same C;B6 hybrids to map the genetic trait responsible for the albinism found in BALB/c mice. A loss-of-function point mutation in the *tyrosinase* gene (*Tyr*, *mm9*_chr7:94575915-94641921) in BALB/c mice results in the complete lack of melanin pigments in the skin and the retina[288] (Figure 5-5B). The C;B6 hybrids exhibited either a white (homozygous background of the genetic trait) or a brownish grey coat color (heterozygous background of the genetic trait). Using the genotyping data from the genome-wide SNP panel, the genetic trait for the coat color could be positively mapped to the *Tyr*-encoding region on chromosome 7 (*mm9*_chr7:82114282-98849561).

5.1.4.3 Fine-mapping the genetic trait for *MIso* on chromosome 12

I repeated LA within the candidate region on chromosome 12 with additional C;B6 hybrids and a higher marker density to further narrow down the *MIso*-encoding region. SNPs were selected according to the criteria listed in 0 with an average distance of 0.5 Mb between genetic markers (39 SNPs within *mm9*_chr12: 7990826-26992720). The fine-mapping SNP panel was designed and evaluated as described above resulting in panel of 32 SNPs (with, on average, one genetic marker per 0.6 Mb). According to the formula in section 5.1.4.1, a minimum of 334 individuals would be required to achieve this resolution. In addition to eight individuals from the genome-wide LA, which exhibited recombination in the 19 Mb candidate region, 184 new C;B6 backcrosses were analyzed regarding the DNA methylation phenotype at the *Isoc2b* promoter region and genotyped with the fine-mapping SNP panel. Twenty-nine of the additional 184 C;B6 hybrids were informative and displayed recombination within the 19 Mb region. Due to an overlap of the "intermediate" and "high" methylation phenotype for 154 non-informative heterozygous and homozygous C;B6 hybrids, the phenotype discriminator was re-defined to a level of 61.5±3.5% average DNA methylation (Figure 5-6A). In total, 20 C;B6 hybrids, including 4 informative individuals, exhibited DNA methylation levels within this range and were excluded from further analysis. Together with the eight informative individuals from the genome-wide LA, the remaining 164 individuals were used for fine mapping LA. Applying the re-defined methylation discriminator, 99.4% concordance with only one non-concordant individual was observed for a subset of fine-mapping SNPs. The resulting 9 Mb candidate region (*mm9*_chr12:16902483-26024113) was confined by the SNPs rs29160274 and rs29188895 (Figure 5-5A). The genotyping results for the fine-mapping SNP panel for all informative C;B6 hybrids are listed in Table 10-2 (see Appendix II). Since the recombination frequency within the 19 Mb candidate region was unexpectedly low, the phenotyping and genotyping of additional C;B6 hybrids was not continued. High sequence variability among C57BL/6 and BALB/c mice in the candidate region could interfere with homologous recombination.

To improve the recombination frequency, further backcrosses were bred from BALB/c and NOD/ShiLtJ mice. Together with five other inbred mouse strains included in the Mouse Genomes Project[275], the NOD/ShiLtJ strain exhibited an *Isoc2b* methylation phenotype similar to C57BL/6 (see Figure 5-3), but seemed genetically more similar to BALB/c with regards to the total number of distinguishing SNPs and indels in the candidate region (REL-1003). In total, 293 C;NOD

hybrids (2 weeks old, bred and raised in the in-house animal facilities) were phenotyped and genotyped. This time backcrosses derived from reciprocal crosses with CNODF1 as well as BALB/c sires. The fine-mapping SNP panel was adjusted to contain 17 genetic markers to distinguish BALB/c from NOD alleles in the 9 Mb candidate region (*mm9*_chr12:16902483-26024113) in agreement with the criteria listed in 5.1.4.2 at an average distance of 0.5 Mb and evaluated with wild type DNA samples.

However, LA with C;NOD hybrids failed to further reduce the size of the candidate region. One individual showed recombination in the candidate region, but only at the outer border of the previously defined 9 Mb candidate region between SPNs rs29181570 and rs29201729. Furthermore, C;NOD hybrids could not be separated well according to the two expected methylation phenotypes (“intermediate”/”high”) (Figure 5-6B, top panel). The blurred phenotype pattern could not be explained by mixing of both reciprocal crosses since separate analysis did not alleviate the phenotype grouping (Figure 5-6B, bottom panel). Also an effect of characterizing younger individuals could be excluded, since the methylation pattern is established during embryogenesis (see section 5.1.5). The simplest explanation might be differential environmental effects on the backcrosses and their parents, since factors such as nutrition seem to have a general effect on epigenetic patterns[251,289] CNODF1 as well as C;NOD hybrids were produced in the in-house animal facilities, whereas the breeding involving C57BL/6 was carried out at Charles River Laboratories.

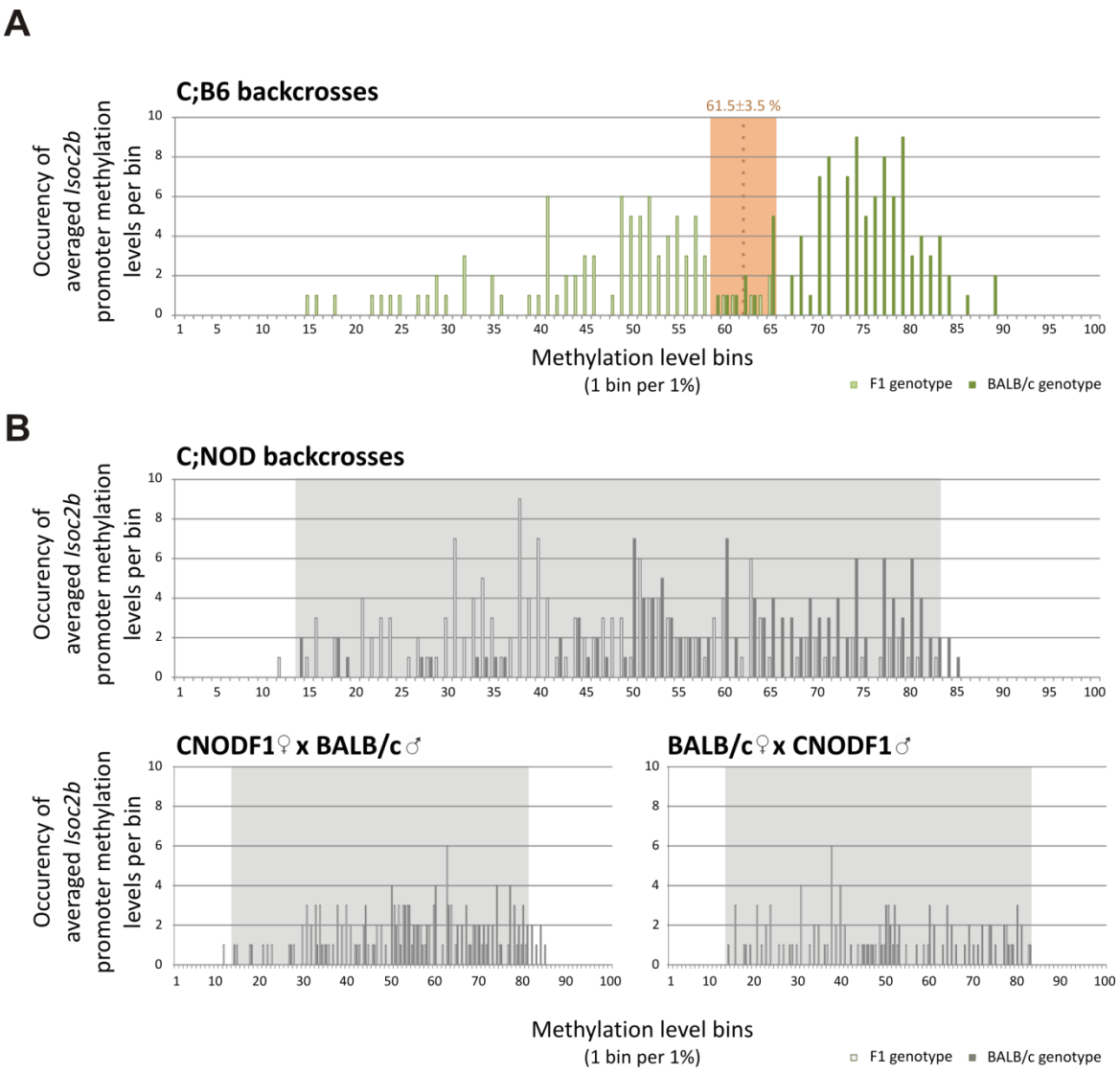


Figure 5-6 ***Discrimination of* Isoc2b *promoter methylation phenotypes in C;B6 and C;NOD backcrosses***

Non-informative individuals without any recombination in the initial 19 Mb candidate region were identified by MALDI-TOF MS-based genotyping and grouped according to their zygosity (transparent bars: heterozygous (F1) genotype, solid bars: homozygous (BALB/c) genotype). For each zygosity group, non-informative **(A)** C;B6 hybrids and **(B)** C;NOD hybrids, respectively, were clustered in 1% bins based on the average methylation DNA levels of the *Isoc2b* promoter region (*mm9*_chr7:4818016-4818396) determined by MALDI_TOF MS. The height of each bin represents the number of comprised individuals. Overlapping DNA methylation levels of the two genotype groups are highlighted by **(A)** orange and **(B)** grey background. The dashed line in **(A)** represents the redefined phenotype discriminator of 61.5 +/- 3.5%.

In summary, the initial genome-wide LA identified a monogenetic trait for *MIso* encoded within a 19 Mb candidate region on chromosome 12. This region was further reduced by fine-mapping with additional animals and SNPs to 9 Mb (*mm9*_chr12:16902483-26024113). The 9 Mb region contains about 50 annotated genes as well as excessive repetitive regions. Initially, none of the included genes was regarded as a suitable candidate gene to encode *MIso*. Therefore, region-wide approaches were applied to further characterize *MIso*.

5.1.5 *In vitro* ESC differentiation models display DNA methylation phenotypes observed *in vivo*

Preliminary DNA methylation data from somatic tissues (BMM, spleen) as well as germ line cells indicated *de novo* methylation at the *Isoc2b* promoter region in each generation of BALB/c strain and its offspring.

Analyzing DNA methylation levels at the *Isoc2b* promoter in additional somatic tissues (brain, liver) of both wild type strains, I could demonstrate that the differential methylation pattern is established not only in mesodermal tissues, but also in somatic tissues arising from the embryonic ectoderm and endoderm, respectively (Figure 5-7). The lack of tissue-specificity indicated that *de novo* methylation at the *Isoc2b* promoter takes place during early embryonic development before the embryo gives rise to the three germ layers. This process, called gastrulation, starts 6.5 days *post coitum* (dpc) (embryonic day (E) 6.5)[290]. At this point, the mouse embryo forms the primitive streak and the precursors of the different germ layers are allocated. The primordial germ cells (PGCs), which later give rise to the mature germ line cell lineage, start to specify around E7.5.

The *Isoc2b* promoter is most probably affected by the genome-wide reprogramming in PGCs that erases DNA methylation patterns acquired in the epiblast and leads to the generally unmethylated state found in germ line cells (testis) of both strains.

Since early embryonic stages render very little material to address DNA methylation and gene transcription, I applied an *in vitro* embryonic stem cells (ESC) differentiation model to overcome these limitations. Similar to germ line cells, pluripotent ESCs of both strains (generated from male E3.5 blastocysts) displayed a hypomethylated phenotype at the *Isoc2b* promoter (Figure 5-7) as assessed by MALDI-TOF MS. Furthermore, ESCs can be driven into unspecific as well as cell type-specific differentiation[291-293]. These findings emphasize the potential of the ESC differentiation model in further investigations regarding *MIso*.

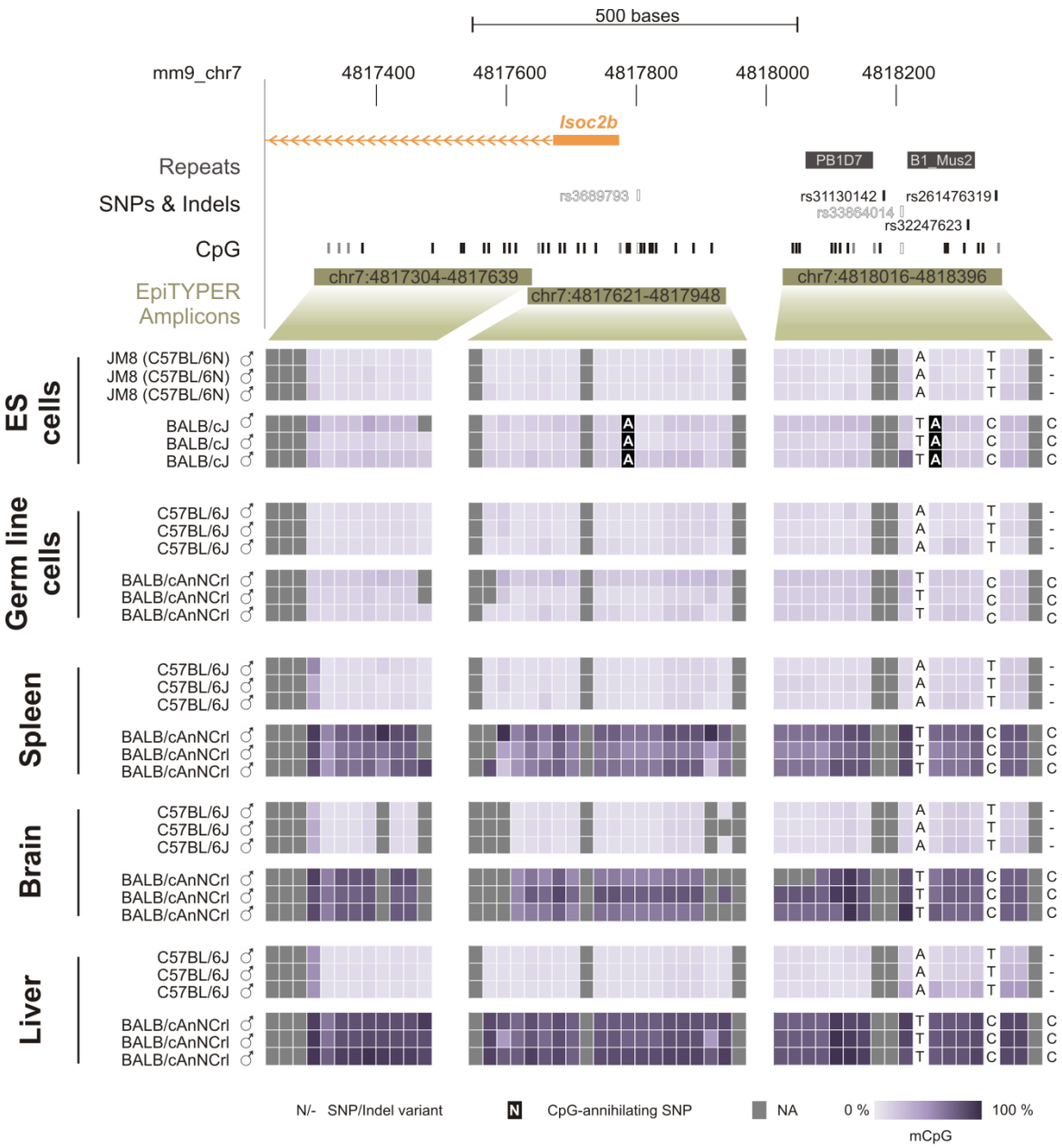


Figure 5-7 ***The strain-specific* Isoc2b *methylation phenotype is present in somatic tissues, but not in germ line cells and ES cells***

The relative position of repeats, sequence variations, CpGs, and EpiTYPER amplicons within the *Isoc2b* promoter region are indicated at the top. Methylation levels of individual CpGs obtained by MALDI-TOF MS analysis of bisulfite-treated DNA in indicated cells and tissues of C57BL/6 and BALB/c are shown color-coded. The scale ranges from pale purple (0% methylation) to dark purple (100% methylation). For ESCs, three independent batches of the same cell lines were analyze. For germ line cells and somatic tissues, three biological replicates are analyzed. Strain-specific variants of indicated SNPs and Indels are positioned relative to analyzed CpGs.

5.1.5.1 *De novo* methylation of the *Isoc2b* promoter in differentiating ESCs

Weng *et al*. have already shown in 1995 that strain-specific developmental gene silencing can be recapitulated in differentiating ESCs[294]. I adapted a procedure described by Bibel *et al*.[266] to drive pluripotent ESCs of both strains into unspecific differentiation by embryoid body (EB) formation (see section 4.1.1.1).

Although the DMR at the *Isoc2b* promoter was initially identified in C57BL/6J and BALB/cAnNCrl mice, I could show that the substrains of C57BL/6 and BALB/c, respectively, revealed similar methylation phenotypes at the *Isoc2b* promoter region (Figure 5-3B). Thus, working with the available ESCs from C57BL/6NJ (JM8) as well as BALB/cJ was expected to have no effect on the experimental outcome.

In agreement with the assumed early onset of *de novo* methylation in BALB/c wild type mice, I detected gradually increasing levels of DNA methylation at the *Isoc2b* promoter shortly after EB formation in the BALB/c ES cell line, but not in the JM8 cells (Figure 5-8A). The dynamics of *de novo* methylation observed in repeated differentiation kinetics of BALB/c ESCs varied, but generally reached a methylation level similar to that in spleen of the corresponding the BALB/cJ substrain (Figure 5-8A). The variable dynamics between different kinetics and even between EBs within the same culture was previously described[295,296].

Differences in the cellular potential of the BALB/c ESCs used for EB formation might also constitute a possible explanation as it is visible in the altered methylation of undifferentiated ESCs (d0) depicted in Figure 5-8A. BALB/c ESC cells could only be maintained in a pluripotent state when cultured on mouse embryonic fibroblasts (MEF). However, removal of MEFs is required for EB formation, since MEFs represent a differentiated cell type contaminating the differentiating ESCs. Counterproductively, BALB/c ESCs seemed to lose their pluripotency rapidly when cultured without MEFs. In contrast, JM8 ESCs were maintained without MEFs on gelatine-coated culture dishes without any effects on their pluripotency. The pluripotent state of ESCs and EBs was assessed by expression analyses of pluripotency genes, which are repressed upon differentiation[296]. RT-qPCR data of total RNA from differentiating ESCs revealed an inverse correlation of *de novo* methylation at the *Isoc2b* promoter with expression levels of the pluripotency markers *Oct4*[225] (Figure 5-8D). An additional marker of the pluripotency potential of ESCs seemed to be the methylation level at an intragenic region of the pluripotency marker *Nanog* (*mm9*_chr6:122658540-122658757). Assessed by MALDI-TOF MS, the detected average DNA methylation levels at this locus correlated well with the *Isoc2b* promoter methylation in differentiating BALB/c ESCs (Figure 5-8B). Increasing levels of DNA methylation at this control locus was be used as an indicator for the progression of ESC differentiation.

Comparing DNA methylation levels of the *Nanog*-associated region between parallel time courses of differentiating JM8 and BALB/c ESCs did not correlate as well and indicated a deferred differentiation of JM8 ESCs in comparison to BALB/c ESCs (Figure 5-8C). This was also detectable at the expression level of the pluripotency marker *Oct4* (Figure 5-8D). However, after

ten days of differentiation not only BALB/c ESCs seemed to generally reach the maximum of *Isoc2b* promoter methylation, but also JM8 cells achieved a differentiation level similar to BALB/c ESCs (according to *Nanog* methylation).

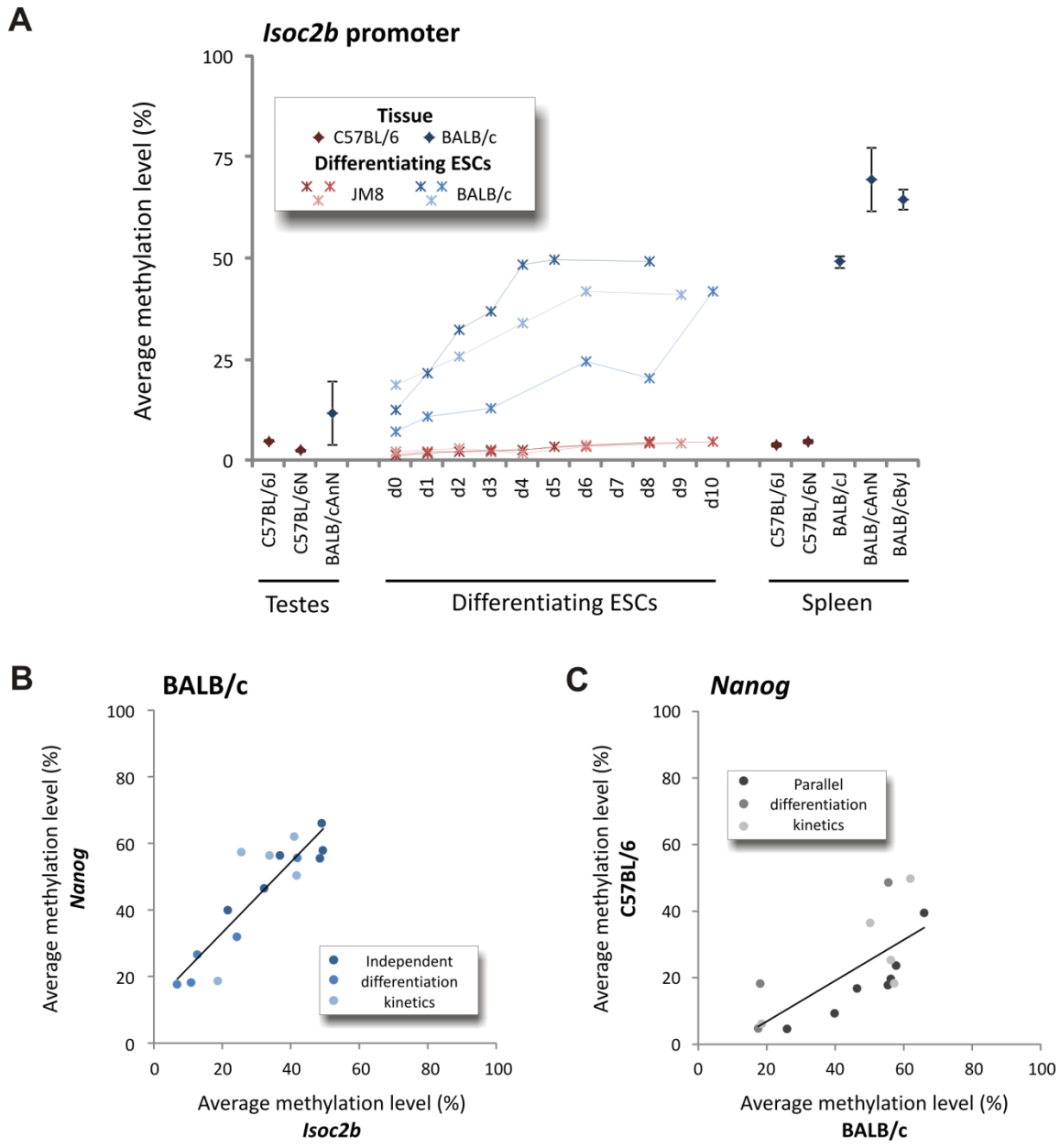


Figure 5-8 DNA methylation and gene expression analysis in differentiating ESCs
(A) – (C) Average DNA methylation levels of the stated genomic regions were assessed by MALDI-TOF MS. **(A)** *Isoc2b* promoter methylation (*mm9*_chr7: 4817621-4817948) in germ line cells (left, n=2-3), differentiating ESCs (middle) and spleen (right, n=2-6) of both wild type strains (C57BL/6: red, BALB/c: blue). For differentiating ESCs, three independent time courses are depicted (matching color intensities indicate parallel generation of differentiation kinetics for JM8 and BALB/c ESCs). The light red and light blue kinetics were generated over the course of a master thesis[297]. **(B)** Correlation of DNA methylation levels in three independent differentiation kinetics of BALB/c ESCs at the *Isoc2b* promoter (*mm9*_chr7: 4817621-4817948) and an intragenic region of *Nanog* (*mm9*_chr6:122658540-122658757) (Pearson correlation, r^2=0.8104). **(C)** Correlation of DNA methylation levels in three parallel differentiation kinetics of JM8 and BALB/c ESCs at an intragenic region of *Nanog* (*mm9*_chr6:122658540-122658757) (Pearson correlation, r^2=0.5045). **(D)** Expression analysis of the pluripotency marker *Oct4* in three independent kinetics of differentiating ESCs of both wild type strains. Expression levels were determined by RT-qPCR and normalized to 18S.

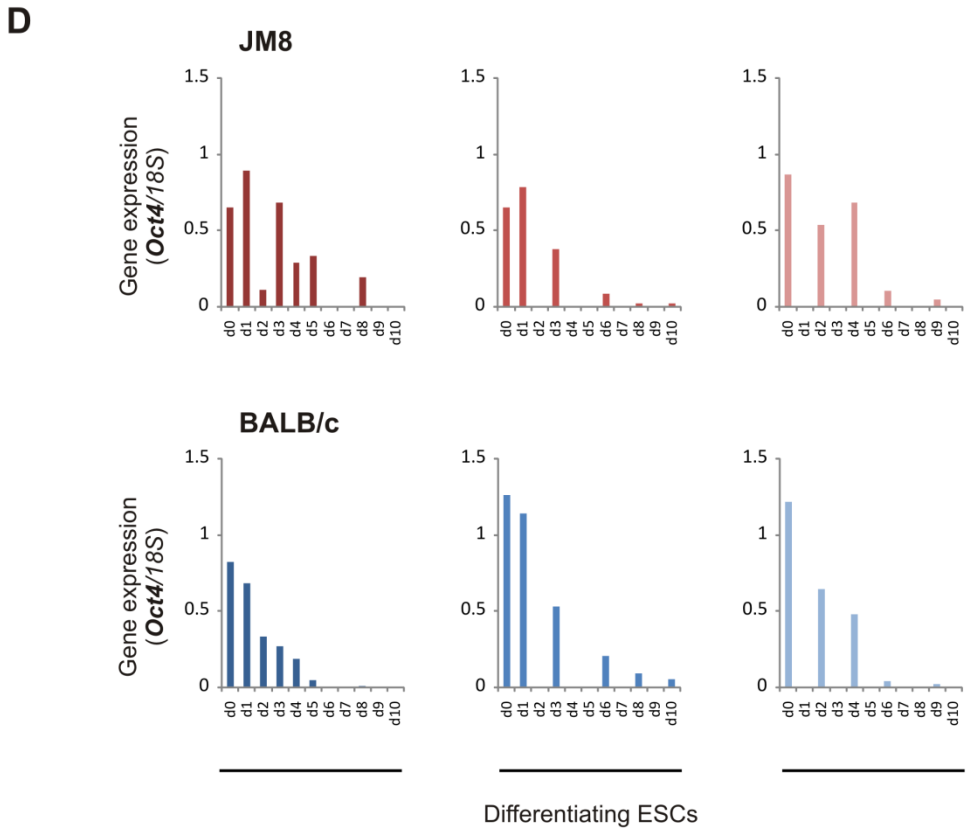


Figure 5 -8
Continued from previous page.

These findings further supported the notion of a strain-specific epigenetic modifier and demonstrated the applicability of the *in vitro ESC* differentiation model for further experiments such as expression analysis in the candidate region during *de novo* methylation.
The results are presented in the following section.

5.1.5.2 Hydroxymethylcytosine accounts for higher methylation level in BALB/c ESCs

Although germ line cells as well as embryonic stem cells exhibited a hypomethylated phenotype at the *Isoc2b* promoter in both strains when compared to somatic tissues, BALB/c ESCs always displayed a slightly increased methylation level. Another modification of the DNA, hydroxymethylation of cytosines (5hmC, an intermediate of the active DNA demethylation pathway), was previously detected at its highest level in mammals in neuronal cells, but also in ESCs[82,298,299]. However, bisulfite treatment-based methylation analysis do not allow for the discrimination of contributions of methylated cytosines (5mC) and 5hmC, respectively[300]. To investigate the presence and dynamics of 5hmC during ESC differentiation, I performed hydroxymethylcytosine methylated DNA immunoprecipitation (hMeDIP), which involves a 5hmC-specific antibody. Using qPCR, local enrichment of 5hmC was assessed.

I performed the hMeDIP assay with one set of parallel differentiation kinetics depicted in Figure 5-8A (JM8: dark red, BALB/c: dark blue) and subsequently analyzed 5hmC enrichment in a 3 kb region surrounding the *Isoc2b* promoter by qPCR. The relative locations of the qPCR amplicons and the EpiTYPER amplicons used for MADI-TOF MS-based DNA methylation analysis are depicted in Figure 5-9A. Specific enrichment of hydroxymethylated DNA fragments was confirmed by parallel analysis of a control region devoid of CpGs ("*empty*", see section 3.10.3),.

Interestingly, undifferentiated BALB/c ESCs exhibited a higher level of 5hmC in the core region (grey background) compared to JM8 ESCs (Figure 5-9B), which might account for the slightly increased DNA methylation s observed in BALB/c ESCs (Figure 5-8A). Furthermore, 5hmC levels at the *Isoc2b* promoter as well as adjacent regions declined rapidly upon differentiation of BALB/c ESC, whereas JM8 ESCs seemed to steadily gain 5hmC during differentiation at the border of the DMR.

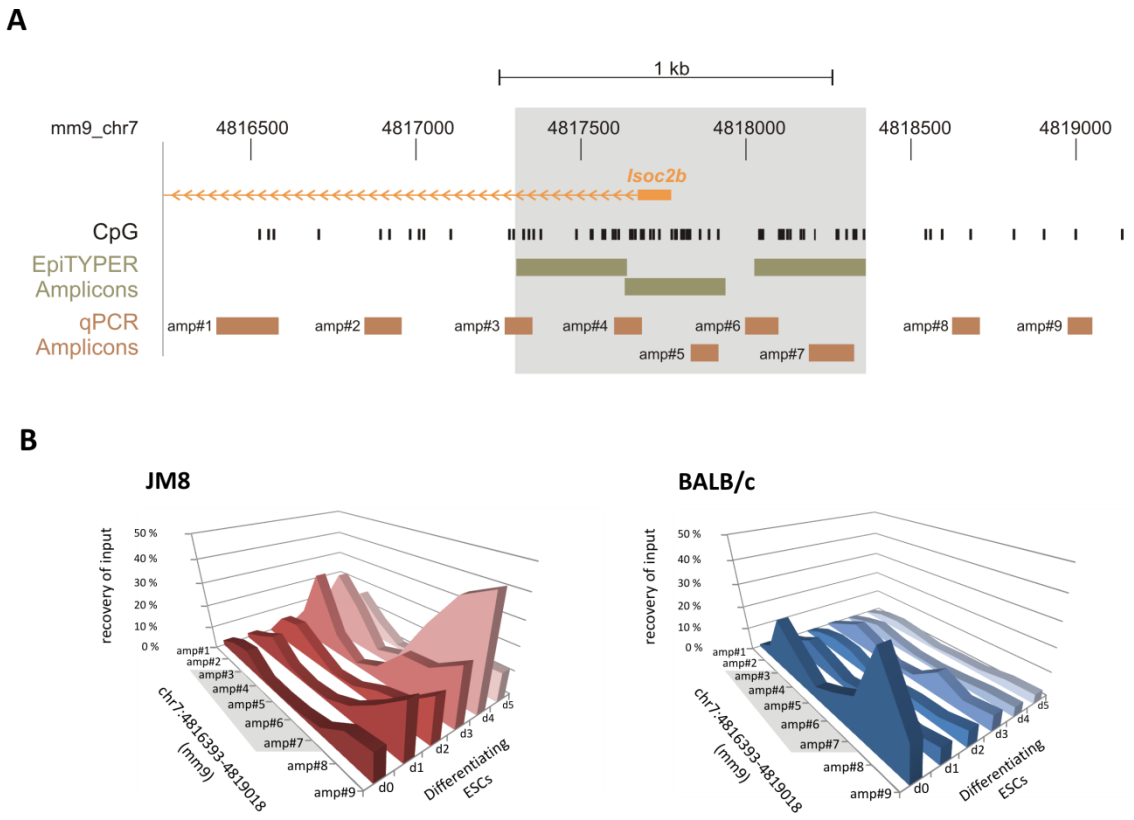


Figure 5-9 ***JM8 and BALB/c ESCs exhibit dynamic, but individual patterns of 5hmC during differentiation***

(A) The represented 3 kb window displays the relative position of qPCR (red) and EpiTYPER amplicons (green) within the promoter regions of the *Isoc2b* gene. **(B)** 5hmC levels were determined in differentiating JM8 (left) and BALB/c ESCs (right) by hMeDIP and qPCR for the core region (grey background) as well as adjacent regions (as indicated in (A)).

5.1.6 Complementation assay in ESCs

In addition to limited availability of mouse embryos, manipulation of the latter is a difficult task, especially during early embryonic stages. However, stain-specific *de novo* methylation of the *Isoc2b* promoter takes place during these early stages of embryogenesis. In contrast, ESCs are comparably easy to manipulate and represent a convenient model to examine the effects of exogenous factors on DNA methylation. Since *de novo* methylation occurs to the same extent at both alleles in F1 hybrids, a transgene encoding for *MIso* was assumed to be able to complement the genetic trait and enforce the associated phenotype to ESCs.

The existing data did not reveal whether *MIso* prevents or promotes *de novo* methylation at the *Isoc2b* promoter. Known epigenetic modifiers, which target DNA methylation during early embryogenesis, often involve a DNA methylation-promoting mechanism (Table 6-1). Based on these findings, I performed complementation assays in normally non-methylating JM8 ESCs with sequences derived from the methylating strain (BALB/c) to subsequently analyze the *Isoc2b* promoter methylation after differentiation in EBs. For that purpose, the CHORI-28 BAC library (BALB/cByJ, obtained from BACPAC Resources) was screened for Bacterial Artificial Chromosomes (BACs) representing the 9 Mb candidate region (see section 4.3.1.20). Since BALB/cByJ exhibited DNA methylation levels at the *Isoc2b* promoter (Figure 5-3B) similar to BALB/cAnNCrl, sequences from both substrains were assumed to have the same effect in complementation assays.

5.1.6.1 Identification of candidate BACs

For library screening purposes, filter membranes are available from BACPAC resources. Each membrane contains over 18000 individual BAC clones of specific BAC libraries in duplicates. The CHORI-28 BAC library is represented by eleven membranes (2 sets of five and six membranes are available), each spotted with BAC clones carrying genomic inserts at an average of 170 kb. Each membrane theoretically provides a one-fold coverage of the BALB/cByJ genome.

Probe design was impeded by the repetitiveness of the candidate region. To provide for specific detection of candidate BACs, I designed oligos to generate labeled probes with PerlPrimer software[301] according to the criteria listed in section 4.3.1.20.

Probes were generated and labeled by PCR with the designed primers on spleen DNA from BALB/cByJ (purchased from Jackson Laboratories) (see section 3.10.5). In total, 146 probes were hybridized to each of three different membranes from the second segment of the CHORI-28 BAC library (7H, 8H, 9H) yielding a theoretical three-fold coverage of the candidate region. The readout of hybridized membranes based on the relative arrangement of the duplicate clones and their position on the membrane was performed according to BACPAC's instruction.

In total, 160 candidate BACs were identified. Based on the size of the candidate region (9 Mb) and the average genomic insert per BAC (170 kb), the total number of identified BAC clones was in line with an estimated three-fold coverage and suggesting a low false-positive rate despite the

repetitiveness of the candidate region. Candidate BACs selected from the CHORI-28 BAC library are listed in Table 3-1.

5.1.6.2 Representation of the candidate region by selected BACs

As an initial attempt to estimate the coverage of the candidate region, I sequenced a pool of 159 BACs prepared by pulsed-field gel electrophoresis (see section 4.3.1.5 and 4.3.1.18 - protocol III) on an Illumina HiSeq platform (indexed 50 bp single-end reads). BAC32 was omitted due to preparation difficulties. The relative concentration as well as residual contamination by *E. coli* gDNA was determined by qPCR with primers specific for the pTARBAC2.1 backbone of the BACs and the *E. coli TRPE* gene locus, respectively. With few exceptions, *E. coli* contaminations were below 10%. To account for the different content of BAC DNA in the individual preparations, the contributing volume to the BAC pool was adjusted according to the individual BAC concentration determined by qPCR. The NGS library of pooled BACs was generated according to NGS library protocol III (see section 4.3.1.18). The complete command line for processing the NGS data of the pooled BACs is stated in section 4.4.3.

Eighty-nine percent of the candidate region was found to be covered by at least one 50 bp read. After adjusting the read length to the actual average fragment length of 170 bp (see section 4.3.1.18) 94% of the bases in the candidate region were covered. In addition to the six annotated 50 kb gaps of the *mm9* reference genome, only 10 regions of the *mm9* reference genome showed major alignment gaps larger than 10 kb in size (Figure 5-10), but were mainly comprised of repetitive sequences (Table 5-1). The largest gap with ~ 235 kb was located in a conserved and well annotated part of the candidate region. The only annotated gene residing in this region is the 5' end of *Mboat2*, a lysophospholipid acyltransferase-encoding gene. Since this gene does not appear to be a suitable factor to confer the observed methylation phenotype, this gap was not assumed to have a negative effect on the complementation assay. Nevertheless, the Mouse Genomes Project[275] (REL-1211) lists many strain-specific SNPs and Indels for C57BL/6 and BALB/c for this region as well as eleven larger structural variances. Hence, closing the gap with additional BAC clones would be advisable.

In summary, the candidate BACs provided for an acceptable region-wide coverage. However, complete sequencing of the individual BAC will be necessary to obtain a more reliable picture of the actual representation of the candidate region.

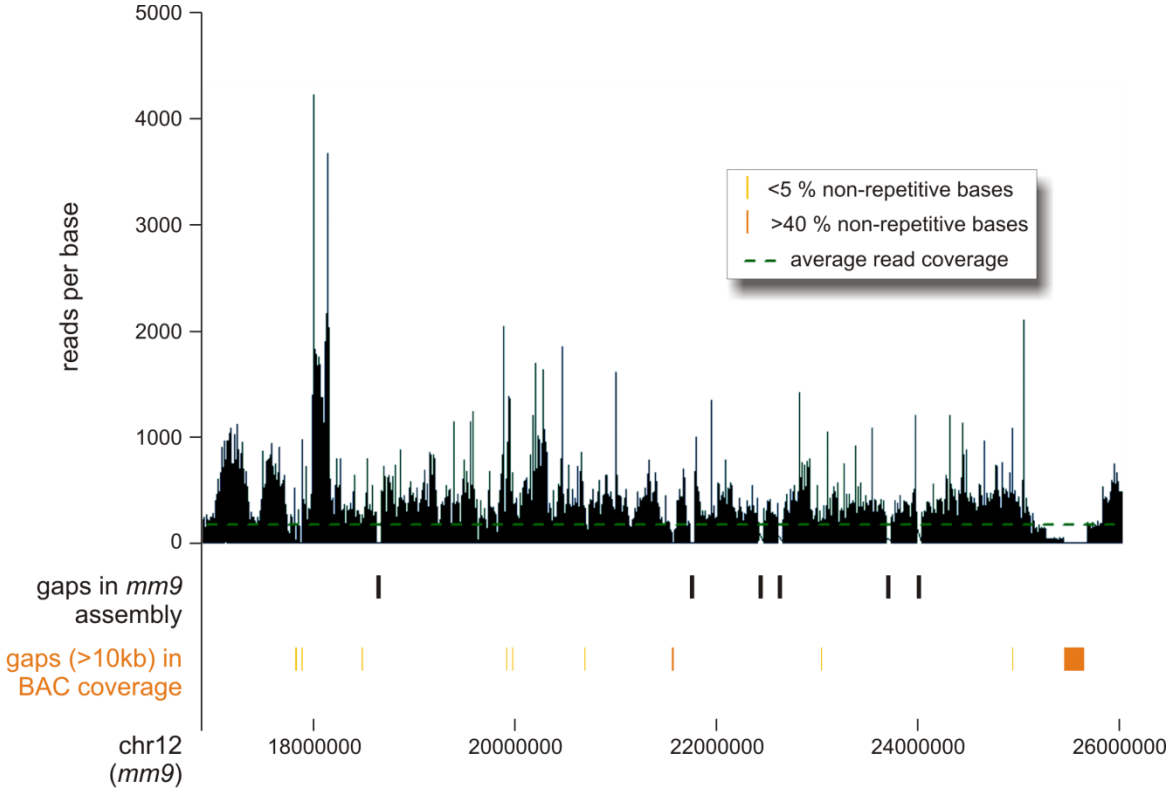


Figure 5-10 Coverage of the candidate region by BAC clones

A pool of 159 candidate BACs was sequenced and mapped to chromosome 12 of the *mm9* reference genome. At the top, the number of reads per non-repetitive base within the candidate region (*mm9*_chr12:16902483-26024113) as well as the average read depth (170 reads per base) at covered bases (green dashed line) is depicted. At the bottom, the relative position of the six known *mm9* assembly gaps (black) within the candidate region as well as coverage gaps > 10 kb (orange). The color intensity of the coverage gaps indicates proportion of non-repetitive bases within the gap (light: ≤ 5%, dark: >5%, see also Table 5-1).

Table 5-1 Major gaps (>10 kb) in BAC coverage

Genomic location (*mm9*)	Length	Fraction of repetitive sequences
chr12:17865353-17885915	20562 bp	> 99%
chr12:17963033-17976874	13841 bp	> 99%
chr12:18441887-18455390	13503 bp	> 99%
chr12:19713003-19724841	11838 bp	96%
chr12:19817985-19828645	10660 bp	> 99%
chr12:20713551-20723740	10189 bp	99%
chr12:21560633-21582762	**22129 bp**	**59%**
chr12:23176181-23186853	10672 bp	> 99%
chr12:24844859-24855529	10670 bp	> 99%
chr12:25432995-25667219	**234224 bp**	**44%**

5.1.6.3 Retrofitting BACs with an eukaryotic selection marker

In order to perform complementation experiments, a eukaryotic selection marker to select for stable transfected ESC clones was a prerequisite. However, the original CHORI-28 BACs only carried a Chloramphenicol resistency gene in the pTARBAC2.1 vector backbone to allow for maintenance in bacteria. Since a LoxP (ATAACTTCGTATANNNTANNNTATACGAAGTTAT) site was encoded in the BAC backbone, I modified (*retrofitted*) the latter using a LoxP/Cre recombinase system (Figure 5-11A) with a Neomycin/Geneticin resistency gene. This gene encodes for an aminoglycoside phosphotransferase that inactivates the otherwise cytotoxic compound Geneticin (G418) by phosphorylation[302]. The eukaryotic selection marker was introduced to the BAC backbone by recombination with the pNelγ plasmid (see section 4.3.3.2). In addition to the LoxP recombination site and the resistency gene, this 9 kb plasmid harbours an *eGFP* gene. Since the resistency gene as well as the *eGFP* gene is under the control of the eukaryotic CAG promoter, ESCs transfected stably with pNelγ-retrofitted BACs could be selected by G418-supplemented medium and visualized by their intrinsic green fluorescence (Figure 5-11B). Of the initial 160 candidate BACs, 148 BACs could be retrofitted with the eukaryotic selection marker and were subject to stable transfection in JM8 ESCs.

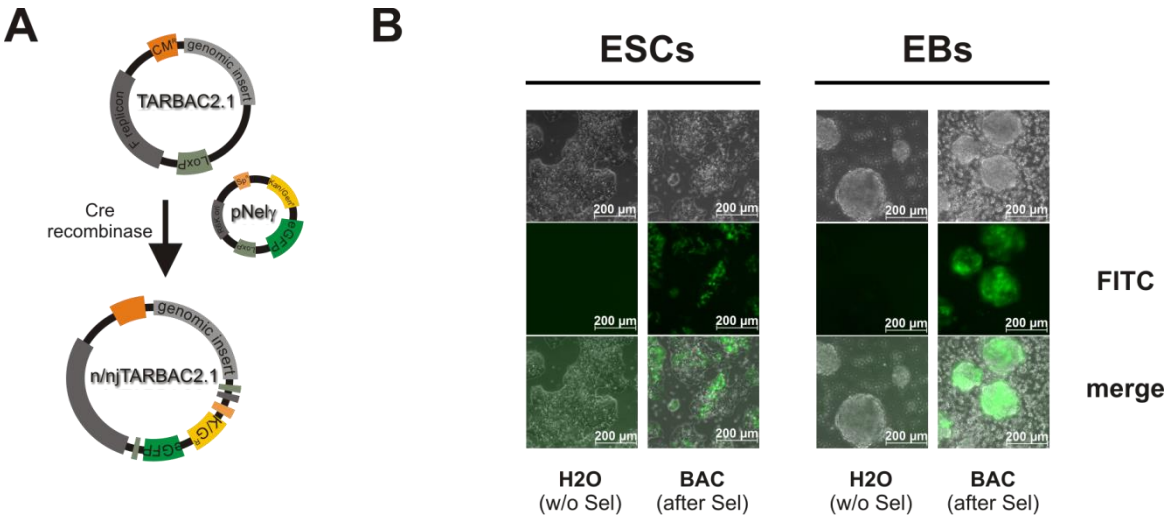


Figure 5-11 ESCs can be stably transfected with retrofitted BACs
(A) Retrofitting system: A eukaryotic selection marker (Gen^R) was introduced to the BAC backbone (TARBAC2.1) together with an eGFP expression cassette provided by pNelγ via the LoxP/Cre recombination system. Two different retrofitting strategies were applied ("n"/"nj", see section 4.3.3.2). **(B)** Fluorescent microscopy of embryonic stem cells (ESCs) and embryoid bodies (EBs) after transfection with H_2O (without G418 selection) or a representative BAC (after selection with G418 and expansion). EGFP was excited at 488 nm for 500 ms.

5.1.6.4 DNA methylation analysis of transfected embryoid bodies

MIso was also assumed to favor DNA methylation and thus to be encoded and functional in BALB/c strains. Consequently, complementation assays in JM8 ESCs with BALB/c-derived *MIso*-encoding sequences might facilitate *de novo* methylation at the *Isoc2b* promoter during differentiation. Therefore, I transfected pluripotent JM8 ESCs individually with the 148 retrofitted candidate BACs selected from the CHORI-28 BAC library (BALB/cByJ). After 10 days of selection with G418 and subsequent three passages without selection pressure, stably transfected pluripotent cells were subject to unspecific differentiation according to section 4.1.1.1. After a differentiation period of 10 days, formed EBs were collected and DNA methylation levels at the *Isoc2b* promoter were determined by the MALDI-TOF MS (see section 4.3.1.14) after bisulfite treatment (see section 4.3.1.12). Stable transfection and hence the presence of BAC-derived exogenous DNA sequences was confirmed for all differentiated ESC clones by PCR with primers specific for the BAC backbone (data not shown).

Confirmed for both cell lines, the transfection and selection procedure itself did not alter the outcome of *Isoc2b* promoter methylation phenotype after differentiation (Figure 5-12A). As described before (see section 5.1.5.1), variation in the differentiation kinetics seems to be a common phenomenon. Since the EBs for each depicted time point were collected from separate culture dishes, it was not expected to see an exclusive incline in methylation levels at the assessed loci. Nevertheless, the combined analyses of methylation levels at the *Isoc2b* promoter as well as control regions (*Pdgfrb*, *Nanog)*, which also gained methylation during differentiation, allowed for reliably estimating the general differentiation state of the respective EB sample. Transfected BALB/c ESC clones exhibited substantial levels of DNA methylation at the *Isoc2b* promoter after 10 days of differentiation. Hence, this differentiation period was assumed to be sufficient to allow for *de novo* methylation at the *Isoc2b* promoter by an exogenous copy of *MIso* in complementation assays with JM8 ESCs. However, *de novo* methylation at the *Isoc2b* promoter region could not be detected in any of the transfected JM8 ESC clones despite their advanced differentiation state (as assessed by methylation levels at the control regions, Figure 5-12B).

Possible explanations for the failure of the complementation assay will be discussed in chapter 6.

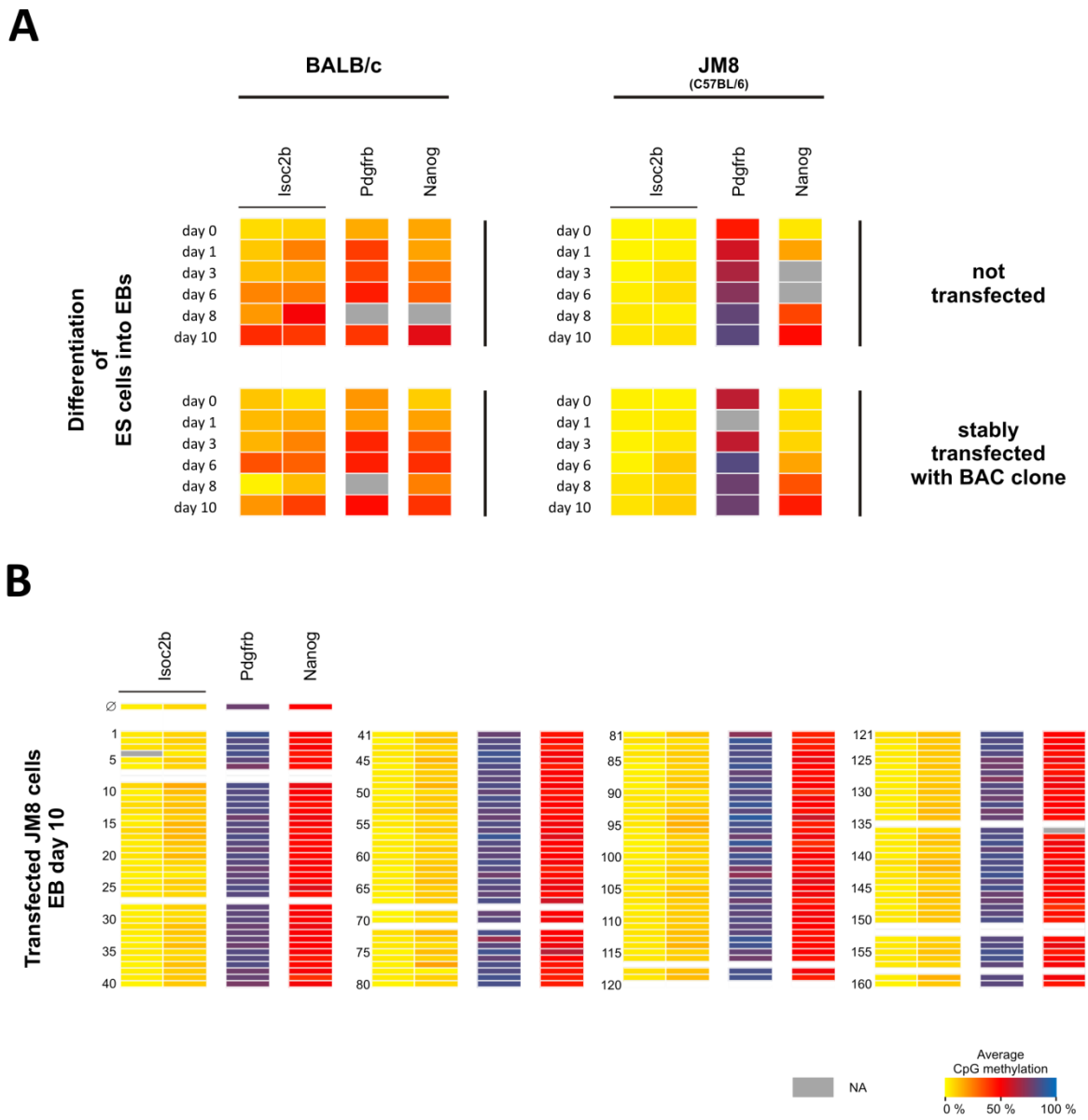


Figure 5-12 DNA methylation analysis of stable transfected differentiating ESCs
The degree of average DNA methylation detected by MALDI-TOF MS at the stated genomic regions is shown color-coded. **(A)** Comparison of methylation dynamics in untransfected (top) as well as BAC-transfected (bottom) BALB/c (left) and JM8 ESCs (right) during differentiation. Depicted are two regions within the DMR at the *Isoc2b* promoter region (left: *mm9*_chr7:4817623-4817948; right: mm9_chr7:4828016-4818396, only the first 7 CpGs) as well as two control regions (*Pdgfrb*: *mm9*_chr18:61225274-61225631; *Nanog*: *mm9*_chr6:122658540-122658757). **(B)** Methylation analysis in JM8 EBs (stably transfected with BALB/c-derived BAC clones) after ten days of differentiation at the *Isoc2b* DMR as well as the control regions (genomic locations of the analyzed regions are stated in (A)).

5.1.7 Comparative expression analysis in differentiating ESCs reveals a candidate gene for *MIso*

Since the complementation assays with BALB/c BAC clones in JM8 ESCs failed to further reduce the candidate region, I used comparative expression analysis of differentiating ESCs as an alternative approach for the identification of *MIso*. Upon differentiation, ESCs established the strain-specific *Isoc2b* methylation phenotype observed in somatic tissues. The *Isoc2b* promoter region was *de novo* methylated in differentiating BALB/c ESCs, whereas it remained unmethylated in differentiating JM8 ESCs. Assuming that the phenotype is regulated by a *trans*-acting epigenetic modifier, its presence and activity must coincide with the process of active *de novo* methylation. Based on these findings, region-wide expression levels in both cell lines were investigated.

After identifying the window of active *de novo* DNA methylation in differentiating BALB/c ESCs (see Figure 5-8A), strain-specific transcripts arising from the 9 Mb candidate region were assessed by RNA-seq experiments. Several time points during differentiation of both JM8 and BALB/c ES cell lines were selected for comparative expression analysis. I purified total RNA from pluripotent ESCs as well as from EBs of both cell lines after one, two, and four days of unspecific differentiation and processed the samples to generate strand-specific RNA-seq libraries. The libraries contained cDNA fragments generated from rRNA-depleted RNA samples with molecules longer than 200 nt, thus allowing for sequencing of both protein-coding and long non-coding RNAs. Corresponding DNA methylation levels of the *Isoc2b* promoter at the selected time points are depicted in Figure 5-8A (BALB/c: dark red time course; JM8: dark blue time course). Ribosomal RNA-depleted RNA-seq libraries were prepared according to section 4.4.2 and sequenced on the Illumina HiSeq1000 platform (100 bp paired-end reads, see section 4.4.4) with one sample per lane. RNA-seq data were analyzed with the Trinity software package as described in section 4.4.2. As a *de novo* assembling software, the Trinity tool pipeline circumvents the difficulties of a not yet available BALB/c reference genome. The complete command line for RNA-seq data analysis is listed in section 4.4.2.

Three different *de novo* assemblies were generated; one for each strain from all reads from all available time points and one for all reads from both strains from all available time points. The *de novo* assemblies counted in total 615402 (BALB/c), 1287454 (JM8), and 1490539 (combined) *'trinity transcripts'* (reconstructed sequence contigs from RNA-seq data) (transcript length > 200 bp). The large number of transcripts assembled from JM8-derived reads and hence combined reads were caused by the increased number of sequencing reads derived from JM8 day 1 and 2 samples (Table 5-2). This inconsistency is most probably due to the applied advanced library preparation and sequencing protocol used for these two samples (see section 4.4.2); the libraries were generated and processed at a later time point when only follow-up versions of the originally applied kits were available. These technical effects on the reproducibility of RNA-seq data can be visualized by pairwise comparison of samples from the same differentiation kinetic. Although no

biological replicates were available, samples from consecutive time points of individual kinetics can be interpreted as such. To test for the similarity of samples derived from the same differentiation kinetic, abundancy levels of the *de novo* assembled transcriptome generated from all samples were estimated and compared. As expected for quasi biological replicates, differentially expressed genes were rare between consecutive BALB/c samples, but increase with proceeding distance of the samples in the differentiation kinetic (Figure 5-13A). In contrast to the BALB/c RNA-seq libraries that were generated with the same kits in one experiment, the comparison of JM8 samples revealed less reproducibility most probably due to the different processing of day 0 and 4 and day 1 and 2 samples (Figure 5-13B).

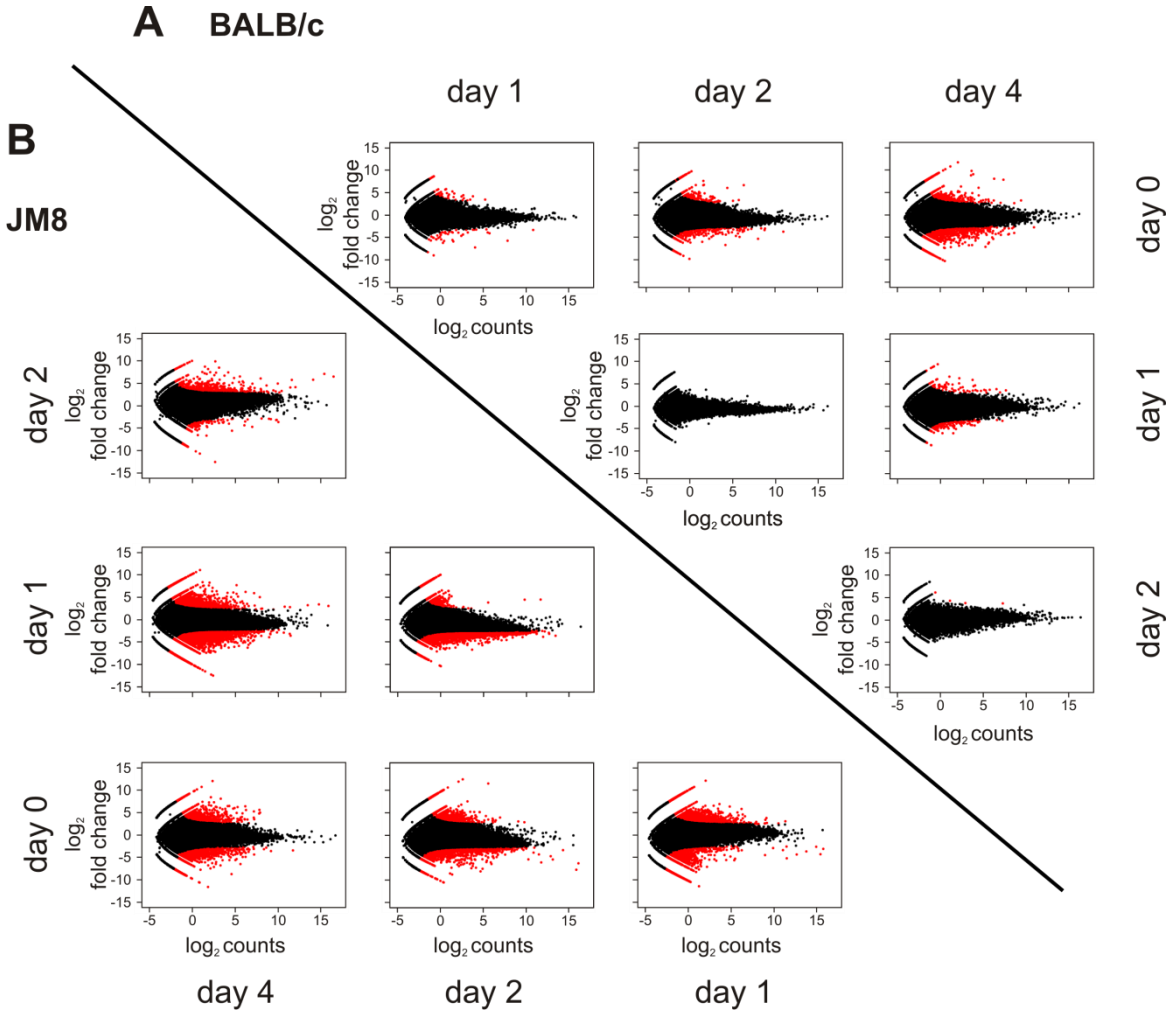


Figure 5-13 Reproducibility of RNA-seq data
MA plots of pairwise comparisons (vertical over horizontal) of BALB/c (**A**) and C57BL/6 (**B**) samples, respectively, are depicted. Transcript abundancies were calculated for transcripts *de novo* assembled from the combined analysis of RNA-seq reads of all samples. Differential expression (DE) and false discovery rate (FDR) were estimated using the EdgeR script stated in 4.4.2. Red dots represent differentially expressed transcripts ($p<0.05$, FDR <0.05) in each individual comparison.

Nevertheless, quality assessment of the generated *de novo* assemblies implicated successful and reliable assemblies. The transcripts from each *de novo* assembly were compared to the mRNA nucleotide sequences of the RefSeq Genes transcriptome (*mm10*, obtained from UCSC table browser) by performing megablast (NCBI). Transcripts of the BALB/c, JM8, and combined *de novo* assemblies that covered reference transcripts by more than 10% (alignment expectation value $<1e^{-20}$) represented 56% (18566), 64% (20988), and 67% (22248), respectively, of the 33004 reference transcripts. Of these represented reference transcripts, 58% (10798/18566), 52% (10906/20988), and 53% (11745/22248), respectively, were covered by more than 50%. The partial coverage of the reference transcriptome can be explained by the complexity of the gene expression programs found in a highly developed multicellular organism such as *Mus musculus*. On the one hand, only parts of the available genes are expressed in each cell type. On the other hand, the deep sequencing approach will also detect minimal amounts of RNA fragments derived from transcriptional background noise resulting in incomplete transcripts with very low abundancy. This background noise is accumulating with increasing read depth, which accounts for the samples JM8 day 1 and 2. To test this hypothesis, a JM8 *de novo* assembly was also generated only from JM8 day 0 and 4 samples. This assembly rendered similar results when compared to the reference transcriptome: The 565736 *de novo* assembled transcripts covered 54% (17946) of the reference transcripts (33004) by more than 10% (expectation value $<1e^{-20}$) and of these reference transcripts 51% (9215/17946) were covered by more than 50%. Nevertheless, the informational content of the *de novo* assembled transcriptomes from all four JM8 samples and all eight combined samples, respectively, is expected to be higher, despite increased background noise. Therefore, further analyses were performed with the transcriptomes assembled from all available samples.

The high number of *de novo* assembled transcripts, especially those generated from JM8 and the combination of all samples, is also associated with an increased number of incomplete isoforms, especially for highly expressed genes as shown for several control genes (Table 5-2). These control genes represent housekeeping genes (*Actb*, *Gapdh*), pluripotency-associated genes (*Nanog*, *Oct4*, *Sox2*) as well as genes known to be expressed in differentiated cells (*Isoc2b*, *Myod1*) and were all (except for *Myod1*) represented in full length by at least one of the assembled transcripts. Normalized expression values for corresponding ‘trinity genes’ (sum of related isoforms) are listed in Table 5-3. Housekeeping genes were detected at high abundancy at all time points, whereas pluripotency-associated genes were gradually repressed with progressing differentiation. In contrast, genes that are known to be transcribed only differentiated cells were detected only at background levels. Again, although accompanied by an increased background noise (no-full-length isoforms), using sequencing data from all available JM8 samples rendered a more complete picture of the transcriptome as assessed by detection of full-length isoforms.

Table 5-2 Overview of detected isoforms of selected control genes

Assembly	Samples	Read counts (2x 100bp)	*De novo* assembled transcripts	Detected isoforms / full-length isoforms (identity, mismatches, gaps)						
				Housekeeping genes		Pluripotency-associated genes			Differentiation-specific genes	
				Actb	*Gapdh*	*Nanog*	*Pou5f1*	*Sox2*	*Isoc2b*	*Myod1*
combined assembly	JM8 day 0-4 BALB/c day 0-4	-	1490539	105 / 2 (100%, 0, 0) (93%, 0, 1)	58 / 1 (100%, 0, 0)	3 / 1 (93%, 0, 1)	16 / 1 (100%, 0, 0)	12 / 2 (100%, 0, 0) (100%, 0, 0)	8 / 0	1 / 0
JM8-specific assembly	JM8 day 0 JM8 day 1 JM8 day 2 JM8 day 4	371417662 653891042 424313042 341866426	1287454	104 / 4 (99%, 4, 0) (99%, 4, 0) (99%, 4, 0) (98%, 6, 0)	46 / 2 (100%, 0, 0) (100%, 0, 0)	7 / 4 (93%, 0, 1) (93%, 0, 1) (93%, 0, 1) (93%, 0, 1)	28 / 1 (100%, 0, 0)	11 / 1 (100%, 0, 0)	8 / 0	1 / 0
BALB/c-specific assembly	BALB/c day 0 BALB/c day 1 BALB/c day 2 BALB/c day 4	282936352 367029208 335016080 374718578	615402	60 / 1 (100%, 0, 0)	39 / 1 (100%, 0, 0)	2 / 0	16 / 1 (100%, 0, 0)	15 / 1 (100%, 0, 0)	4 / 0	-
JM8-specific assembly	JM8 day 0, day 4	-	565736	54 / 0	19 / 2 (100%, 0, 0) (100%, 0, 0)	6 / 3 (93%, 0, 1) (93%, 0, 1) (93%, 0, 1)	18 / 2 (100%, 0, 0) (100%, 0, 0)	10 / 0	4 / 0	-

Table 5-3 Normalized expression values of selected control genes

Gene name	Trinity gene ID	Normalized expression values (fpkm*) BALB/c day 0	BALB/c day 1	BALB/c day 2	BALB/c day 4	JM8 (C57BL/6) day 0	JM8 (C57BL/6) day 1	JM8 (C57BL/6) day 2	JM8 (C57BL/6) day 4
House keeping genes									
Actb	c1247068_g1	2535.370	1918.918	1115.922	1690.708	1027.806	1329.987	1069.797	1382.951
Gapdh	c1257627_g3	280.405	197.886	145.817	135.255	218.785	717.390	488.449	237.839
Pluripotency-associated genes									
Nanog	c1240255_g4	25.061	8.900	9.017	11.042	44.783	26.929	48.881	19.824
Oct4	c1256103_g1	355.599	210.487	168.430	54.556	516.166	351.421	161.258	102.258
Sox2	c1254521_g4	65.289	27.480	22.497	4.090	119.149	103.397	67.193	29.369
Differentiation-associated genes									
Isoc2b	c1234267_g4	0.000	0.000	0.484	0.000	0.000	0.000	1.142	0.443
Myod1	c1289921_g1	0.000	0.000	0.000	0.266	0.000	0.000	2.049	0.000

*fpkm: fragments per kilo base per mapped million reads

General and local quality assessment supported the reliability of the available RNA-seq data sets as well as successful *de novo* assembly of the transcriptome present in differentiating ESCs. Further analyses of the *de novo* assemblies generated from all BALB/c- and JM8-derived samples ('BALB/c-specific', 'JM8-specific', 'combined') were aimed at the identification of strain-specific transcripts potentially encoding *MIso*.

In addition to the search for novel, unannotated strain-specific transcript within the candidate region, the list of annotated genes located within *mm9*_chr12:16902843-26024113 (*mm10*_chr12:16895677-25339248) was revised. Initially, none of the annotated genes was associated with the observed strain-specific *Isoc2b* methylation phenotypes. However, recently published IP experiments by Cao *et al.* implicated an interaction of nucleolar protein 10 (Nol10) with Eed[185], a subunit of the Polycomb repressor complex 2 that is involved in gene silencing during embryonic development[93]. Nol10, encoded at *mm10*_chr12:17348493-17430095 and composed of 687 amino acids (aa), belongs to a yet uncharacterized family of nucleolar proteins that share the NUC153 domain[303]. Adjacent to this domain of unknown function, Nol10 exhibits a WD40-repeat-containing domain with five adjoining WD40 repeats. WD40 repeats are found in all eukaryotes and seem to act as sites for protein-protein interaction, but can also serve as platforms

for the assembly of protein complexes. The specificity of proteins harbouring WD40-repeat-containing domains seems to be determined by the sequence outside the repeat-containing domain[304,305]. Examining the available genotyping data generated by the Mouse Genomes Project[275] (Rel-1303) for C57BL/6 and BALB/c as well as for the other six inbred mouse strains (129S1/SvImJ, A/J, C3H/HeJ, CBA/J, DBA/2J, NOD/ShiLtJ) that were also analyzed for their methylation phenotype at the *Isoc2b* promoter (see Figure 5-3A) further supported a possible involvement of Nol10 in the establishment of the observed methylation phenotype. Only the hypermethylated strains (BALB/c, A/J) have a consequential SNP variant of rs29160165 (*mm10*_chr12:17416714). The G to A mutation in BALB/c and A/J has two consequences according to the Mouse Genome Project annotation. On the one hand, the SNP variant induces a non-synonymous variant resulting in an alternate protein sequence. On the other hand, the SNP position overlaps with a splice junction and might result in an alternate splice variant.

Visual inspection of *de novo* assembled transcripts highly similar to the Nol10 protein (expectation value $<1e^{-20}$) revealed that, although none of the transcripts represented a full-length isoform of *Nol10*, the combination of related transcripts rendered nearly 100% *Nol10* transcript coverage. Only transcripts assembled from all JM8 (day 0 to day 4) displayed larger gaps, as the first exon as well as a connection of exon 14 and 15 was missing. These gaps, however, were covered by transcripts assembled from JM8 day 0 and day 4 (Figure 5-14A). The estimated transcript abundancy on gene level further revealed, that *Nol10* is expressed to a similar extend in both ES cell lines at all analyzed time points (Table 5-4). These findings were validated by RT-qPCR (Figure 5-14B).

Since the non-synonymous SNP had an implication in the post-transcriptional processing of the *Nol10* transcript, the corresponding *de novo* assembled transcripts were analyzed in more detail. Although rs29160165 resides in a splice region of intron 18, no effect on the splicing of the *Nol10* transcript was detected in RNA-seq-derived transcripts nor on cDNA level. Nevertheless, the SNP variant (Guanine to Adenine) seems to have an effect on the amino acid sequence of the Nol10 protein in BALB/c ESCs (see Appendix III, Table 10-3). The annotated SNP variant was verified on transcript and genomic sequence level and causes an amino acid exchange (glutamic acid to lysine) at position 532. The WD40 repeats and the NUC153 domain are not affected by the non-synonymous SNP variant (Figure 5-14C).

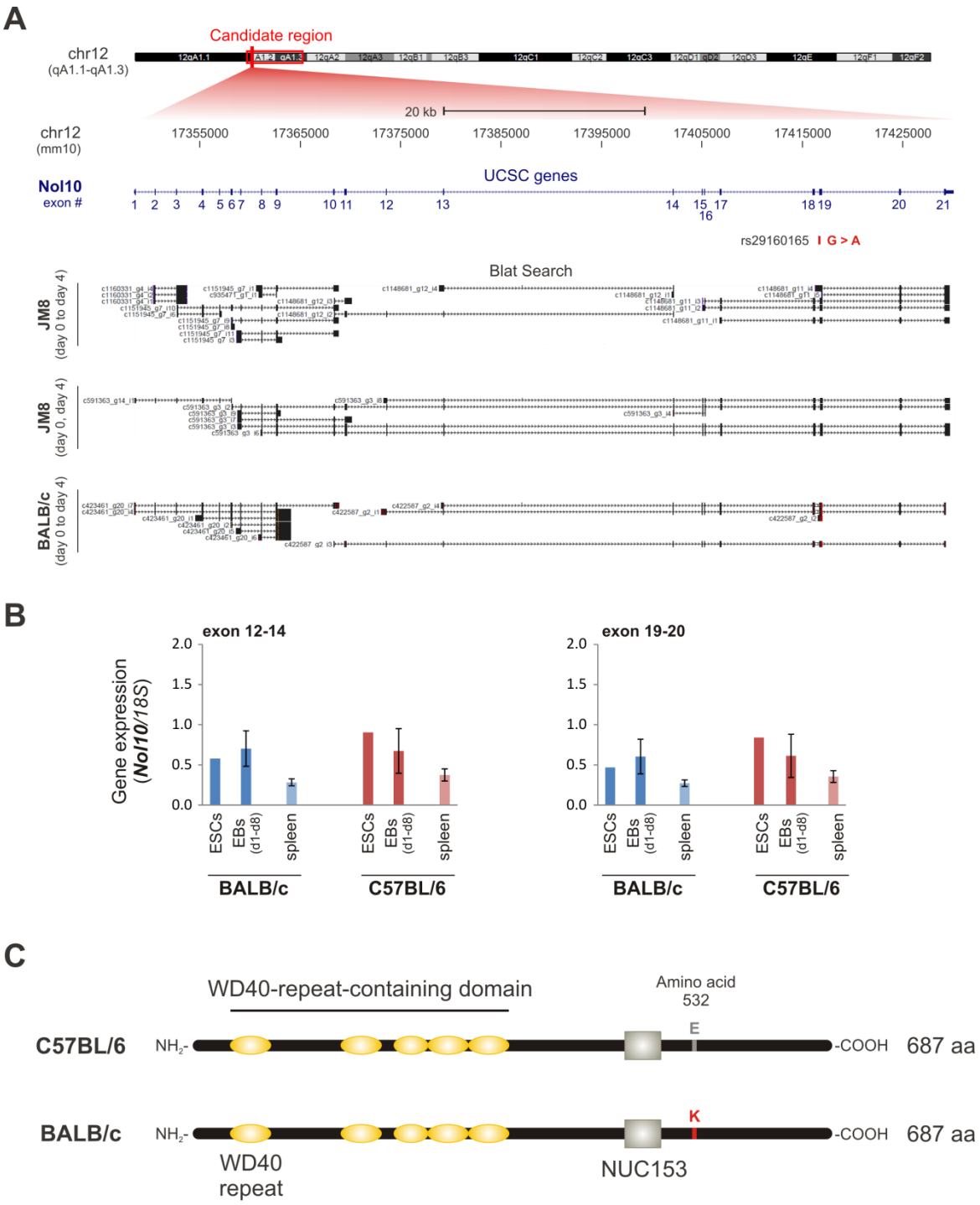


Figure 5-14 Differentiating ESCs express strain-specific variants of Nol10

(A) Depicted at the top is the UCSC genome browser track of the *Nol10*-encoding genomic region (*mm10*_chr12:17348493-17430095). Below BLAT results for *de novo* assembled transcripts similar to Nol10 are shown for the individual Trinity assemblies. The non-synonymous SNP variant (G > A) of rs29160165 resides in a splice region, but does not affect splicing. Other SNP variants between C57BL/6 and BALB/c mainly affect intronic sequences or are synonymous. **(B)** *Nol10* expression in ESCs, EBs (mean of differentiated EBs day 1-5, day 8) and spleen (n=3-4) assessed by RT-qPCR (normalized to 18s, error bars represent standard deviation). **(C)** The non-synonymous SNP variant (G > A) of rs29160165 (see (A)) at the 5'-end of exon 19 results in an amino acid exchange (glutamic acid > lysine) at position 532. The beta-transducin (WD40) repeats as well as the nucleolar protein domain 153 (NUC153) of Nol10 are not affected.

***Table 5-4 Normalized expression values of* de novo *assembled transcripts similar to* Nol10**

Transcript assembly	Trinity gene ID	Normalized expression values (fpkm*) day 0	day 1	day 2	day 4
JM8 (day 0 to day 4)	c935471_g1	0.259	0.050	0.552	0.000
	c1148681_g11	18.197	42.345	22.404	11.559
	c1148681_g12	58.296	62.887	32.784	36.583
	c1151945_g7	6.263	16.918	14.160	5.448
	c1160331_g4	40.713	57.860	33.324	42.928
JM8 (day 0, day 4)	c591363_g3	31.631	-	-	18.447
	c591363_g14	24.380	-	-	17.541
BALB/c (day 0 to day 4)	c422587_g2	18.689	19.490	21.047	16.912
	c423461_g20	16.160	13.558	14.739	10.166

*fpkm: fragments per kilo base per mapped million reads

Although further investigations on the functional effect of the strain-specific isoforms of *Nol10* are required, these findings implicate a potential role of *Nol10* in the establishment of the strain-specific methylation phenotype at the *Isoc2b* promoter and thus a potential candidate locus encoding for *MIso*.

Nol10 is a promising candidate gene, but its actual involvement still has to be confirmed. Therefore, the RNA-seq data and the generated *de novo* assemblies will be subject to further analysis regarding strain-specific novel transcripts within the candidate region.

5.1.8 DNA methylation frequency at repetitive sequences

The known epigenetic modifiers were often associated with the silencing of repetitive sequences[96,98,306,307]. The strain-specific methylation phenotype observed at the *Isoc2b* promoter might also be caused due to differential methylation of repetitive sequences among the two mouse strains, since two repeats of the Alu family (PB1D7, B1_Mus2) reside within the DMR (see Figure 5-7). Using MCIp-seq data from spleen and cerebellum of both strains (further described in section 5.2.4), differential DNA methylation at known repeats was assessed with the Homer tool pipeline. The complete command line for processing the MCIp-seq data is stated in section 4.4.1.2. Although the methylation profile of several annotated repeats seemed to generally differ significantly between both strains ($p<0.05$, FDR<0.05), no difference was found for the repeats located within the *Isoc2b* DMR (Table 5-5, PB1D7 and B1_Mus2 are highlighted).

Table 5-5 Strain-specific MCIp enrichment of selected repeats

RepeatID (Name|Class|Family)	Length (bp)	Copies (*mm9*)	MCIp tags (normalized to 10 Mio. reads)				C57BL/6 versus BALB/c		
			Spleen		Cerebellum				
			C57BL/6	BALB/c	C57BL/6	BALB/c	logFC	p value	FDR
significantly differentially methylated repeats									
(TCCTG)n|Simple_repeat|Simple_repeat	80.4	149	866	147.6	1081.8	95.5	3.05	1.66E-27	2.02E-24
(CGGAG)n|Simple_repeat|Simple_repeat	83.1	17	1937.21	442.8	1667.12	419.66	2.05	2.23E-23	1.36E-20
(TATTG)n|Simple_repeat|Simple_repeat	65.4	147	221.33	29.35	192.27	19.28	3.09	6.79E-21	2.76E-18
(CATAT)n|Simple_repeat|Simple_repeat	72.5	263	19.34	132.06	13.39	170.8	-3.25	5.86E-15	1.79E-12
SUBTEL_sa|Satellite|Satellite	162.4	31	2159.62	701.75	1837.48	618.01	1.59	1.87E-14	4.57E-12
5S|rRNA|rRNA	80.5	1035	970.22	365.12	963.77	305.79	1.54	2.88E-12	5.85E-10
(CATAG)n|Simple_repeat|Simple_repeat	89.1	52	134.3	504.09	131.42	333.34	-1.63	1.63E-10	2.85E-08
MER70A|LTR|ERVL	280.5	181	377.13	164	324.91	101.01	1.43	7.48E-09	1.14E-06
(TTCCG)n|Simple_repeat|Simple_repeat	76.6	5	399.69	182.99	382.1	137.74	1.30	9.87E-08	1.34E-05
(CTAGT)n|Simple_repeat|Simple_repeat	58.6	24	22.56	3.45	24.34	2.75	2.88	1.79E-05	2.19E-03
MER96|DNA|hAT	116.9	105	44.05	115.66	41.37	126.72	-1.52	1.98E-05	2.20E-03
(CTTCG)n|Simple_repeat|Simple_repeat	40.3	3	106.37	33.66	98.57	35.81	1.55	2.19E-05	2.23E-03
(TTTCC)n|Simple_repeat|Simple_repeat	107.7	1279	645.74	358.21	446.59	203.86	0.97	2.46E-05	2.31E-03
(GCCTG)n|Simple_repeat|Simple_repeat	65.2	78	79.51	23.31	102.22	28.47	1.82	2.99E-05	2.61E-03
(CAGAC)n|Simple_repeat|Simple_repeat	94.1	55	69.84	33.66	90.05	14.69	1.78	5.70E-05	4.61E-03
(CATAA)n|Simple_repeat|Simple_repeat	68.1	94	60.17	15.54	34.07	10.1	1.86	6.04E-05	4.61E-03
tRNA-Ile-ATA|tRNA|tRNA	47.2	26	20.41	1.73	6.08	0	3.72	1.17E-04	8.40E-03
(TGGG)n|Simple_repeat|Simple_repeat	65.1	1064	507.14	308.15	444.16	251.61	0.77	3.66E-04	2.49E-02
G-rich|Low_complexity|Low_complexity	86.7	6134	11920.9	5147.04	12831.96	5821.06	1.17	5.17E-04	3.32E-02
LTR87|LTR|ERVL?	190.4	211	154.72	236.51	114.39	256.2	-0.90	7.95E-04	4.85E-02
Alu repeats									
FAM|SINE|Alu	82.6	215	56.01	54.61	40.03	45.76	-0.05	8.92E-01	1.00E+00
B1_Mur1|SINE|Alu	122.8	37980	12867.93	13030.13	11517.92	10523.13	0.06	4.18E-01	1.00E+00
FLAM_A|SINE|Alu	69.5	73	12.22	7.91	8.01	12.93	0.00	1.00E+00	1.00E+00
B1_Mur4|SINE|Alu	122.2	40539	13592.04	14013.08	12474.74	11557.74	0.04	8.01E-01	1.00E+00
PB1D9|SINE|Alu	96.7	29287	8373.58	8304.49	7973.53	7923.68	0.02	8.15E-01	1.00E+00
Alu|SINE|Alu	48.9	31	0	7.12	0	3.98	-5.53	2.49E-03	4.97E-02
B1F2|SINE|Alu	110.2	16782	5548.44	5392.81	4842.84	4503.52	0.08	1.74E-01	1.00E+00
B1_Mm|SINE|Alu	131.2	42397	22718.24	22108.64	27817.29	28952.04	-0.01	9.64E-01	1.00E+00
B1_Mur2|SINE|Alu	122	31802	10506.18	10882.18	9424.12	8818.02	0.02	7.34E-01	1.00E+00
B1_Mur3|SINE|Alu	124	24658	8138.32	7893.74	7417.05	6747.82	0.09	1.17E-01	1.00E+00
PB1D10|SINE|Alu	94.4	72983	22350.58	22539.97	20097.31	20304.13	-0.01	9.60E-01	1.00E+00
AluF_3|SINE|Alu	67.9	166	73.33	87.06	29.36	44.77	-0.35	3.94E-01	1.00E+00
B1F|SINE|Alu	109.3	45775	13893.49	14426.21	12322.61	12478.93	-0.03	8.25E-01	1.00E+00
Alu_5|SINE|Alu	44.7	19	0	0.79	2.67	0	1.18	1.00E+00	1.00E+00
B1_Mus2|SINE|Alu	**128.1**	**70597**	**36720.7**	**35968.18**	**42476.59**	**43056.5**	**0.01**	**9.74E-01**	**1.00E+00**
B1_Mus1|SINE|Alu	133.4	93733	36266.48	36305.33	36650.23	37949.13	-0.02	9.17E-01	1.00E+00
PB1|SINE|Alu	95.7	13667	4247.89	4043.42	3775.25	3690.76	0.06	2.98E-01	1.00E+00
PB1D7|SINE|Alu	**96.1**	**24985**	**7933.61**	**7768.69**	**7132.81**	**7421.3**	**-0.01**	**8.91E-01**	**1.00E+00**
B1F1|SINE|Alu	107.8	20755	6876.48	7332.61	6174.65	6161.87	-0.04	4.28E-01	1.00E+00
AluG_3|SINE|Alu	80	282	68.24	76.77	66.72	47.75	0.11	6.44E-01	1.00E+00

5.2 *Trans*-regulated inter-individual variations in DNA methylation are the exception of the rule

5.2.1 Identification of novel differentially methylated regions in inbred mouse strains

As discussed in section 5.1, inbred mouse strains represent an ideal model organism to study epigenetic variations. The MCIp-on-ChIP approach established in our lab[261] proved suitable to detect genome-wide differentially methylated regions (DMRs). However, the method was limited to selected differentially expressed regions (DERs) identified by microarray-based expression analysis. With the microarray technology available at that time (Whole Mouse Genome Expression Array (4x44K), Agilent) only parts of the annotated mouse genome could be screened for DERs. Thus, a new approach to discover DMRs on a more global scale was tested. To provide for a functional relationship between DNA methylation and gene expression, chromatin immunoprecipitation combined with next generation sequencing (ChIP-seq) analysis for active histone marks (methylated H3K4, methylated H3K27) were combined with DNA methylation analysis. H3 histones methylated at lysine 4 and acetylated at lysine 27 are associated with active enhancers and promoters. Since DNA methylation is in general associated with repression, an inverse correlation of active histone marks and DNA methylation at regulatory genomic elements was expected between the two mouse strains. To test this hypothesis as well as to identify additional DMRs, ChIP-Seq and MCIp-seq experiments for cerebellum and spleen tissue of BALB/c and C57BL/6 mice were performed and analyzed.

Initially, only ChIP-seq data for spleen tissue of both wild type strains (n=1 each, male) were available (library preparation according to protocol I in section 4.3.1.18). In two turns, I selected candidate regions for further DNA methylation analysis by MALDI-TOF MS. The premises for Set 1 were four-fold enrichment of C57BL/6-specific tags for either H3K4me3 or H3K27ac within defined peaks and a minimal coverage of 80 tags per peak (normalized to 10^7 total reads). To compensate mapping disadvantages of BALB/c sequences to the C57BL/6-derived *mm9* reference genome, the requirements to detect BALB/c-specifically enriched regions were less strict (three-fold enrichment, 60 tags). For Set 2, stringency was further reduced to a general fold change of two and tag coverage of 60. The complete command line for processing the ChIP-seq data is stated in section 4.4.1.1. The complete list of the 70 novel candidate regions is stated in Appendix IV (Table 10-4). Set 1 contained many CpG-rich (15 of 41 regions overlapped CGIs) and promoter-associated (13 of 41, seven were CGI-associated) sites similar to the differentially methylated region at the *Isoc2b* promoter. However, most of the analyzed CpG-rich regions were detected as demethylated, which is consistent with the generally hypomethylated state of CpG islands[34]. Hence, regions selected for Set 2 contained less CpG-rich (three of 29 regions

overlapped CGIs) and promoter-associated (five of 29, two were CGI-associated) regions. MALDI-TOF MS-based DNA methylation analysis in spleen from both wild type strains (n=2-3 for each strain) identified 18 differentially methylated regions (DMRs) (Table 5-6) among the 70 selected candidate regions.

Table 5-6 Annotated list of novel strain-specific DMRs

Differential histone mark	Peak localization (*mm9*)	*Nearest gene* (Rel. position)	Enriched in (log2 FC)	Epityper amplicon localization (*mm9*)	CpGs	Average CpG methylation in Spleen*	
						C57BL/6	BALB/c
Set 1							
H3K4me3	chr17:21844143-21845143	***Zfp760*** (promoter)	BALB/c (2.1x)	chr17:21844358-21844570	8	17%	1%
H3K27ac	chr4:140766411-140767411	***6330545A04Rik*** (distal)	C57BL/6 (4.7x)	chr4:140766797-140767168	8	72%	92%
H3K27ac	chr4:133613961-133614961	***Aim1l*** (distal)	C57BL/6 (2.4x)	chr4:133614214-133614597	13	30%	54%
H3K27ac	chr15:79722157-79723157	***Apobec3*** (promoter)	C57BL/6 (2.2x)	chr15:79724111-79724508	7	16%	68%
				chr15:79722557-79722852	12	7%	4%
H3K27ac	chr9:105973860-105974860	***Col6a4*** (intragenic)	BALB/c (7.7x)	chr9:105974425-105974596	6	66%	44%
H3K27ac	chr14:51443888-51444888	***Parp2*** (intragenic)	C57BL/6 (2.7x)	chr14:51444053-51444423	8	25%	59%
H3K27ac	chr19:47515636-47516636	***Sh3pxd2a*** (intragenic)	C57BL/6 (3.3x)	chr19:47515951-47516448	10	25%	51%
H3K27ac	chr13:96730805-96731805	***Sv2c*** (intragenic)	C57BL/6 (2.1x)	chr13:96731300-96731697	19	21%	62%
				chr13:96731126-96731439	16	17%	64%
Set 2							
H3K4me3	chr3:96077378-96078378	***Hist2h2bb*** (distal)	BALB/c (4.1x)	chr3:96077670-96077953	7	25%	1%
				chr3:96077988-96078283	7	76%	9%
H3K4me3	chr7:150191605-150192605	***Tspan32*** (intragenic)	C57BL/6 (2.1x)	chr7:150192677-150192988	6	50%	86%
H3K27ac	chr17:48565966-48566966	***Apobec2*** (intragenic)	C57BL/6 (1.8x)	chr17:48566705-48566973	4	21%	52%
				chr17:48566071-48566366	6	31%	79%
H3K27ac	chr11:105894819-105895819	***Dcaf7*** (promoter)	C57BL/6 (5.6x)	chr11:105895046-105895376	7	56%	80%
H3K27ac	chr16:30155069-30156069	***Hes1*** (distal)	C57BL/6 (5.5x)	chr16:30155216-30155568	6	15%	35%
H3K27ac	chr5:147874677-147875677	***Lnx2*** (intragenic)	BALB/c (1.8x)	chr5:147874763-147875096	7	46%	10%
				chr5:147875405-147875618	4	80%	25%
H3K27ac	chr5:123488913-123489913	***Morn3*** (intragenic)	C57BL/6 (2.1x)	chr5:123489136-123489456	17	3%	26%
				chr5:123489428-123489728	7	6%	68%
H3K27ac	chr13:101179103-101180103	***Naip1*** (intragenic)	BALB/c (7.8x)	chr13:101179250-101179593	5	84%	7%
H3K27ac	chr7:52996858-52997858	***Spaca4*** (intragenic)	C57BL/6 (3.5x)	chr7:52997182-52997586	9	10%	57%
H3K27ac	chr4:132650946-132651946	***Wasf2*** (distal)	C57BL/6 (2.5x)	chr4:132651381-132651641	5	19%	53%
				chr4:132651025-132651364	6	10%	47%

**mean of 2-3 individuals (8-12 weeks old) for each strain*

5.2.2 Tissue specificity of novel DMRs

In terms of identifying additional targets of MIso, novel DMRs exhibiting a tissue-independent strain-specific methylation pattern were of particular interest. Tissue specificity of the newly identified DMRs was assessed by MALDI-TOF MS-based DNA methylation analysis in various somatic tissues from both wild type strains (Table 5-7). According to the detected different tissue specificities, the DMRs were grouped in three categories: *Category 1* comprises all regions showing strain-specific methylation patterns in all tested somatic tissues as well as germ line cells (testis), which applied to two of 18 regions; *Category 2* comprises all DMRs with differential DNA methylation in all tested somatic tissues, but similar methylation levels in germ line cells, which applied four of 18 regions; *Category 3* comprises the remaining DMRs with tissue-specific differential methylation patterns, which applied to the majority (12) of all analyzed regions.

The only target known so far for *MIso* is strain-specifically *de novo* methylated during early embryonic development and its resulting differential methylation pattern is present in tissues from all three germ layers. Hence, DMRs from *Category 2* – especially those close to *Aim1l* and *Morn3,* as they were detected as BALB/c-specifically methylated – seemed promising.

To further assess the mode of inheritance and regulation of the newly identified DMRs and their potential role as novel targets of *MIso*, DNA methylation analysis was extended to F1 hybrids.

Table 5-7 Average CpG Methylation of novel DMRs in various somatic tissues and germ line cells

Epityper amplicon localization (*mm9*)	*Nearest gene* (Rel. position)	Average CpG methylation* in									
		Spleen		BMM		Cerebellum		Liver		Testis**	
		C57BL/6	BALB/c	C57BL/6	BALB/c	C57BL/6	BALB/c	C57BL/6	BALB/c	C57BL/6	BALB/c
***Category 1*: Tissue-independent strain-specific DMRs (somatic & germ line)**											
chr3:96077670-96077953	***Hist2h2bb*** (distal)	25%	1%	25%	2%	85%	1%	72%	3%	92%	4%
chr3:96077988-96078283		76%	9%	88%	8%	95%	16%	96%	33%	96%	60%
chr7:52997182-52997586	***Spaca4*** (intragenic)	10%	57%	13%	53%	38%	62%	39%	67%	63%	46%
***Category 2*: Tissue-independent strain-specific DMRs (only somatic)**											
chr4:133614214-133614597	***Aim1l*** (distal)	30%	54%	43%	59%	33%	51%	28%	45%	9%	15%
chr11:105895046-105895376	***Dcaf7*** (promoter)	56%	80%	73%	90%	45%	86%	61%	87%	84%	87%
chr5:123489136-123489456	***Morn3*** (intragenic)	3%	26%	5%	29%	6%	19%	3%	7%	2%	7%
chr5:123489428-123489728		6%	68%	11%	75%	22%	65%	5%	25%	4%	8%
chr4:132651381-132651641	***Wasf2*** (distal)	19%	53%	23%	61%	4%	46%	5%	15%	60%	62%
***Category 3*: Tissue-specific DMRs**											
chr4:140766797-140767168	***6330545A04Rik*** (distal)	72%	92%	77%	94%	72%	89%	88%	92%	94%	96%
chr17:48566705-48566973	***Apobec2*** (intragenic)	21%	52%	67%	82%	62%	77%	75%	81%	79%	80%
chr17:48566071-48566366		31%	79%	72%	64%	68%	55%	83%	67%	78%	54%
chr15:79724111-79724508	***Apobec3*** (promoter)	16%	68%	90%	67%	88%	92%	77%	76%	87%	89%
chr9:105974425-105974596	***Col6a4*** (intragenic)	66%	44%	89%	66%	94%	91%	84%	85%	96%	98%
chr16:30155216-30155568	***Hes1*** (distal)	15%	35%	6%	20%	22%	30%	61%	68%	14%	9%
chr5:147874763-147875096	***Lnx2*** (intragenic)	46%	10%	70%	18%	75%	74%	55%	55%	76%	83%
chr5:147875405-147875618		80%	25%	87%	52%	71%	54%	39%	50%	82%	62%
chr13:101179250-101179593	***Naip1*** (intragenic)	84%	7%	78%	7%	67%	50%	69%	48%	30%	50%
chr14:51444053-51444423	***Parp2*** (intragenic)	25%	59%	10%	66%	80%	61%	77%	59%	85%	66%
chr19:47515951-47516448	***Sh3pxd2a*** (intragenic)	25%	51%	18%	53%	92%	93%	84%	87%	89%	92%
chr13:96731300-96731697	***Sv2c*** (intragenic)	21%	62%	29%	79%	10%	11%	22%	58%	81%	77%
chr13:96731126-96731439		17%	64%	23%	74%	12%	14%	17%	60%	52%	79%
chr7:150192677-150192988	***Tspan32*** (intragenic)	50%	86%	50%	92%	94%	94%	85%	93%	100%	n.a.
chr17:21844358-21844570	***Zfp760*** (promoter)	17%	1%	26%	2%	4%	2%	8%	2%	3%	3%

**mean of 2-4 individuals (8-12 weeks old) for each strain. **only 1 individual (8-12 weeks old) tested.*

5.2.3 Inheritance of strain- and tissue-specific DNA methylation patterns in F1 hybrids

In concordance to previously published data[261], the newly identified DMRs also often lacked differential methylation patterns in germ line cells (Table 5-7). At most regions described by Schilling *et al*. sequence-dependent (allele-specific) *de novo* methylation or demethylation seemed to render the distinct methylation patterns found in somatic tissues. Except for the *Isoc2b* promoter region that was the only confirmed region where the local sequence did not affect *de novo* methylation; in contrast, an epigenetic modifier seemed to regulate the somatic phenotype in *trans*. To identify novel *trans*-regulated DMRs and thus potential targets of *MIso*, DNA methylation analysis of regions identified in section 5.2.1 was extended to F1 hybrids.

Although allele-specific DNA methylation analysis is not possible by MALDI-TOF MS, I compared averaged CpG methylation ratios of both wild type strains with the average CpG methylation levels in F1 hybrids to obtain an initial indication regarding the mode of regulation. As demonstrated by Schilling *et al*., DMRs that were mainly regulated in a sequence-dependent, allele-specific manner rendered almost identical average methylation level in F1 hybrids when compared to the mean methylation level in parental strains[261].

In line with these previous findings, the majority (14) of the 18 analyzed regions exhibited similar averaged CpG methylation levels in spleen of F1 hybrids and parental mice indicating allele-specific differential DNA methylation patterns. Only four regions either gained (*6330545A04Rik*, *Svc2*) or lost (*Naip1*, *Tspan32*) methylation in F1 hybrids compared to the parental strains suggesting the involvement of a potential *trans*-regulating factor in the establishment of the strain-specific methylation pattern (Figure 5-15).

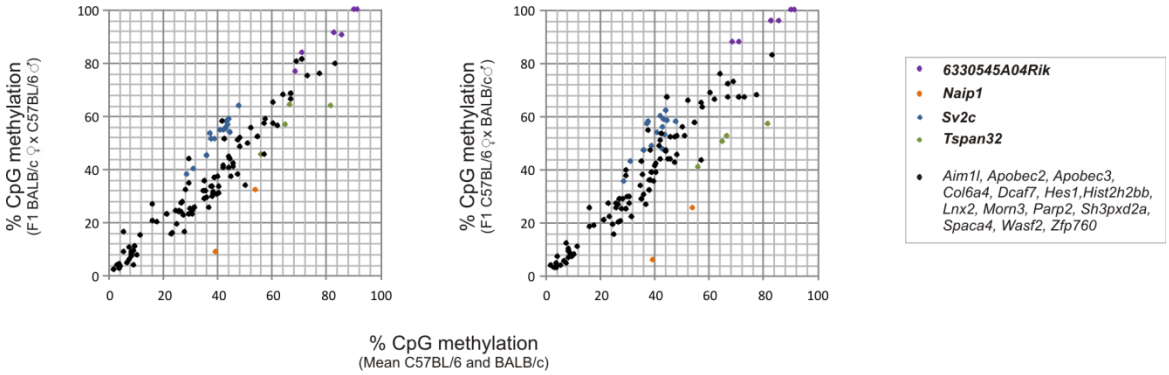


Figure 5-15 Comparison of mean parental methylation ratios with average methylation levels of F1 hybrids at distinct CpGs provides initial indication for mode of inheritance

Methylation levels of individual CpGs were assessed by MALDI-TOF MS in spleen of wild type mice (n=3 for each strain) and F1 hybrids (n=5). Mean average CpG methylation levels of parental spleen are plotted against average CpG methylation levels of F1 hybrids derived from BALB/s (left, n=3) or C57BL/6 (right, n=1-3) dams, respectively. Regions that either gained or lost methylation in F1 hybrids compared to the mean parental methylation level are highlighted.

Nevertheless, reliable conclusions regarding the mode of regulation can only be assessed by bisulfite sequencing (see section 4.3.1.13) that does not only assess the methylation status of individual CpGs, but also their sequence context. In the presence of strain-specific sequence variations this method allows for allele-specific methylation analysis.

Unfortunately, only the DMR within the *Svc2* gene was suitable for allele-specific bisulfite sequencing. The DMRs associated with *Tspan32* and *6330545A04Rik* did not comprise any sequence variation or indels to distinguish the parental alleles in F1 hybrids and were not further analyzed. In case of *Naip1*, PCR of bisulfite-treated DNA did not render a specific amplicon resulting in mixed and thus inconclusive sequencing results (data not shown) and might also distort the relationship of F1 and parental methylation depicted in Figure 5-15. Despite the implied *trans* regulation, I found the DMR associated with *Sv2c* to be also differentially methylated in dependence of the local sequence (Figure 5-16A).

As a control, one of the potential *cis*-regulated regions (*Sh3pxd2a*) was also analyzed by bisulfite sequencing confirming its allele-specific methylation pattern (Figure 5-16B).

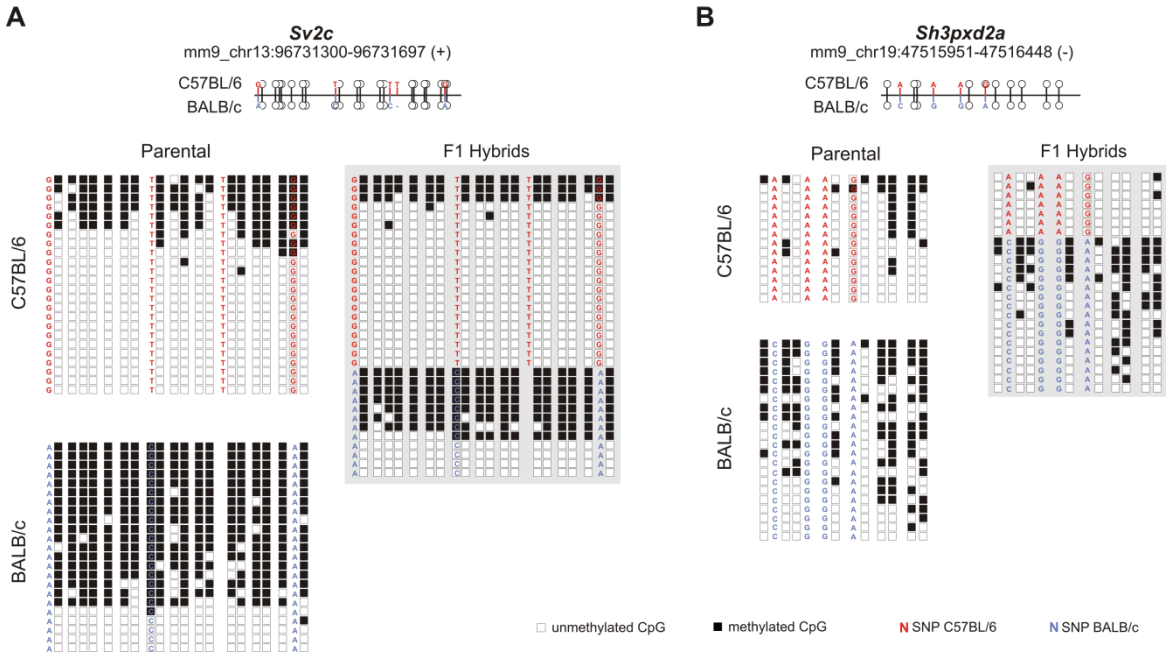


Figure 5-16 Allele-specific bisulfite sequencing of selected DMRs determines mode of inheritance for strain-specific methylation phenotypes

The genomic position of CpGs within the amplicons of *Sv2c* **(A)** and *Sh3pxd2a* **(B)** are indicated at the top. Sequence variations that distinguish the parental alleles are marked in red (C57BL/6) and blue (BALB/c). The methylation status of the individual CpGs is depicted by white (unmethylated) and black (methylated) squares. Each line of squares represents an independently sequenced clone derived from two independent sample preparation (one sample preparation for each reciprocal cross for F1 hybrids).

5.2.4 Comparison of ChIP-seq and MCIp-seq data sets

Since the search for novel *trans*-regulated DMRs and potential targets of *MIso* was not successful as yet, I complemented the ChIP-seq data set from spleen with ChIP-seq data from cerebellum of the same individuals (library preparation according to protocol II in section 4.3.1.18). The combination of ChIP-seq data from two tissues should alleviate the identification of tissue-independent differentially regulated regions. Furthermore, the newly established approach of high-throughput sequencing of MCIp-enriched DNA fragments (MCIp-seq) was also applied to genomic DNA from spleen and cerebellum of both wild type strains (C57BL/6J and BALB/cAnNCrl, n=1 each, female) in order to supplement the ChIP-seq data sets with genome-wide information on highly methylated CpG-rich regions (library preparation according to protocol II in section 4.3.1.18). The complete command line for processing the ChIP-seq and MCIp-seq data is stated in section 4.4.1.2.

As it is widely accepted, the vast majority of regions marked by active histone marks (H3K4me3, H3K27ac) and highly methylated regions (enriched by MCIp) exclude each other within the same specimen (visualized for spleen in Figure 5-17A). However, the assumption to identify novel DMRs based on an inverse correlation of active histone marks and highly methylated DNA between both wild type strains was not supported by the data. For C57BL/6- and BALB/c-enriched peaks, a general enrichment of highly methylated DNA fragments by MCIp in the opposite strain could not be detected (Figure 5-17B). Since DMRs at additional target sites of *MIso* are most probably established during early embryonic development, regions with similar epiphenotypes in both cerebellum and spleen were of particular interest. However, the level of joint presence of strain-specifc active histone marks in both tissues is very low (Figure 5-17C). Visual inspection of the 82 regions that showed enrichment (fold change >2) of at least one of the strong active histone mark (H3K4me3 or H3K27ac) in both analyzed tissues identified only nine regions with inverse correlation of active histone marks and accumulated mCpGs (including the *Isoc2b* promoter). Instead, MCIp-seq data generally displayed strain- and tissue-independent high or low enrichment of mCpGs at these sites. Due to the specificity of the MCIp approach for CpG-rich hypermethylated regions, it cannot be excluded that DMRs at CpG-poor regions were missed. Taken together, active and repressive marks generally exclude each other within a strain, but inverse correlation between strains cannot be generally assumed.

Since the total number of detected tissue-independent strain-specifically regulated regions based on active histone modifications is very low, future experiments will focus on the validation of tissue-independent strain-specifically methylated region derived from the available MCIp-seq data sets (fold change >2 rendered approximately 2200 regions).

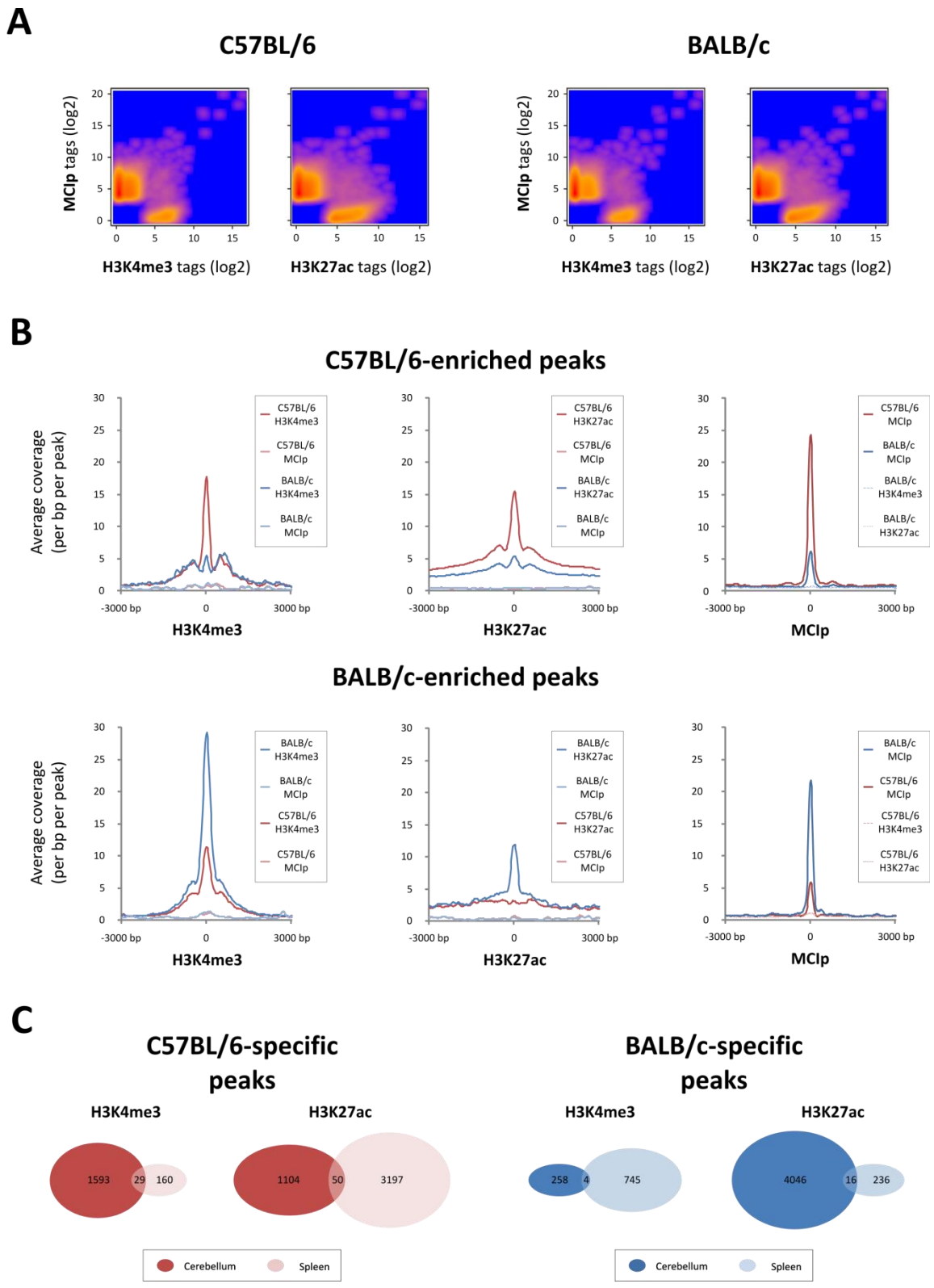


Figure 5-17 Comparison of ChIP-seq and MCIp-seq data sets

(A) Correlation of fragment enrichment by MCIp and ChIP experiments for the indicated histone marks in spleen tissue from C57BL/6 (left) and BALB/c (right) mice. **(B)** The genomic location of averaged C57BL/6(top)- and BALB/c(bottom)-specific H3K4me3 (left), H3K27ac (middle), and MCIp peaks (right) is centered in the middle of each plot (fold change >2). Average genomic distribution of the corresponding signal in the opposite strain as well as other histone marks and MCIp signals are plotted relative to the center of the strain-specifically enriched peaks. **(C)** Overlap (peak center distance <300 bp) of strain-specifically enriched histone mark peaks (fold change >2) in cerebellum and spleen. The size of the Venn circles does not represent relative proportions.

6 Discussion & Perspectives

Although epigenetic mechanisms are widely accepted to affect gene expression without changing the actual genomic sequence, little is known about how epigenetic mechanisms are regulated themselves. The fact that alterations in the epigenome have implications in disease susceptibility[8,9] demonstrates the importance of this knowledge. In this regard, the extent of inter-individual variations and how differential epiphenotypes are established and inherited is of particular interest. Several publications demonstrated that genetic determinants acting in *cis* and in *trans* affect individual epigenomes[96,98,260-262].

With this thesis, I aimed at the identification of the *trans*-acting epigenetic modifier responsible for the strain-specific epiphenotype at the *Isoc2b* promoter in C57BL/6 and BALB/c mice, termed *modifier of* Isoc2b (*MIso*). Epigenetic variations at this particular locus comprise DNA methylation as well as histone modifications. It could be demonstrated that *MIso* is a monogenetic trait encoded in a 9 Mb region on the short arm of chromosome 12 and acts during early embryonic development. *Nucleolar protein 10* (*Nol10*) was identified as a potential candidate gene encoding *MIso*. I further tried to define additional targets of *MIso*. However, none of the newly identified and evaluated differentially methylated regions seemed to be regulated in *trans*.

6.1 A novel pipeline to identify mouse strain-specific epigenetic modifiers

The strategies to identify epigenetic modifiers in mice are still limited. A commonly used strategy is to employ mutagenesis screenings. Thereby, the expression and epiphenotype of a reporter gene is analyzed in the offspring of mutagen-treated mice. Mutations affecting general epigenetic modifiers will lead to an up- or downregulation of the reporter gene and can be subsequently mapped by linkage analysis. The approach was initially developed in the *Drosophila* model, in order to identify factors involved in position effect variegation during development[308,309]. For example, with the help of a N-ethyl-N-nitrosourea (ENU) mutagenesis screening, the group around Emma Whitelaw were able to detect known and novel epigenetic modifiers[97,98]. However, this approach is not designed to identify site-specific epigenetic modifier as it relies on the effect on a single reporter gene. Epigenetic modifiers not affecting the reporter gene, but with targets elsewhere in the genome, are missed. Presented in this work is an alternative strategy, which is in theory less restrictive. By comparing the individual epigenomes of inbred mouse strains,

differences in the general portfolio of epigenetic modifiers, including those with sequence specificity, can be examined. This second strategy further allows for the investigation of non-artificial epiphenotypes with functional effects in healthy organisms. Subsequent localization of the strain-specific epigenetic modifiers is based on linkage analysis similar to the first strategy.

In the present study, I used two commonly inbred mouse strains (C57BL/6 and BALB/c) as a model system to assess genetic determinants of inter-individual variations in DNA methylation, the most stable epigenetic modification. Inbred mouse strains have the advantage of consistent breeding conditions, hence minimizing environmental effects on the individual epigenome. In addition, they provide a stable genetic background and an unlimited cohort.

A few years back, reliable and efficient detection of methylated cytosines on a global scale was challenging. Since DNA methylation is generally associated with repression[14,136], strain-specific gene expression differences were used to pre-define candidate DMRs in an earlier study in our laboratory[261]. In the present work, I defined regions displaying active strain-specific histone mark patterns as candidate regions for differential DNA methylation. Thereby, candidate differentially methylated regions (DMRs) were no longer restricted to promoter regions of differentially expressed regions, but also included potential distal regulatory elements, which can interact with the transcription start site of target genes by DNA looping[310-312]. Limiting candidate DMRs to the vicinity of differentially expressed genes tends to ignore regulatory elements at distal sites. The switch to global chromatin immunoprecipitation(ChIP)-derived histone mark patterns in order to define candidate DMRs partly overcomes the restriction. Active histone marks at transcription start sites not only correlate with transcription (H3K4me3, H3K27ac), but also mark intergenic enhancers (H3K4me1, H3K27ac)[164,313,314].

There are limitations to the approach chosen. Advancing differential DNA methylation from strain-specific patterns of active histone marks will miss DMRs without inverse correlation to the analyzed histone modification. Moreover, analysis of next generation sequencing ChIP (ChIP-seq) data from two different tissues (spleen and cerebellum) demonstrated that only a minor fraction of the defined histone mark peaks were strain-specifically enriched in both tissues in the individual mouse strains. I also generated next generation sequencing methyl-CpG immunoprecipitation (MCIp-seq) data, which captured highly methylated regions on a global scale. In comparison to the ChIP-seq results, analysis of the MCIp-seq data identified a much larger number of regions with strain-specific enrichment in both tissues. Still, enrichment of methylated CpG dinucleotides (mCpGs) by MCIp is restricted to highly methylated CpG-rich region. Hence, high-throughput sequencing of bisulfite-treated DNA would be the method of choice. However, the method is still very expensive when performed on a global scale. It is further difficult to analyze due to PCR biases generated from the unbalanced GC content of methylated and unmethylated fragments[315,316]. Since the differential methylation pattern at the *Isoc2b* promoter is captured well by MCIp-seq, new candidate DMRs and thus potential novel targets of *MIso* will be selected based on the MCIp-seq data sets in future studies.

The method of choice to evaluate DNA methylation levels at selected candidate regions was the EpiTYPER technology (MassARRAY, Sequenom), which is based on Matrix-assisted laser desorption/ionisation combined with mass spectrometry analysis by time-of-flight

(MALDI-TOF MS). The method allows for accurate high-throughput quantitative DNA methylation analysis up to single CpG resolution at distinct loci[270,317] and was used to validate strain-specific DNA methylation at selected candidate DMRs.
Previous data generated in our laboratory and also by other groups suggested that the methylation phenotype is mainly regulated in *cis*[260-262]. Only few DMRs appear to be regulated in *trans* by a strain-specific epigenetic modifier. To address the mode of regulation, MALDI-TOF MS-based methylation analysis is only applicable for an initial estimation. Mean parental methylation levels at individual CpGs are compared with the average methylation of the corresponding CpGs in F1 hybrids. DMRs with an allele-specific methylation phenotype (i.e. that are regulated in *cis*) generally exhibit similar methylation levels in this comparison, whereas *trans*-regulated regions display diverging methylation levels. A reliable validation of the mode of regulation (*cis* or *trans*) can only be achieved by bisulfite sequencing (BS). BS detects the methylation state of individual CpGs within their genomic context, but depends on a high number of individually sequenced DNA strands to permit for quantitative methylation analysis. Since the major purpose of identifying additional DMRs was to find other targets of *MIso*, I used the initial implication on the mode of regulation obtained by MALDI-TOF MS-based methylation analysis to restrict BS to potential *trans*-regulated regions.

After identifying a *trans*-regulated DMR, such as the *Isoc2b* promoter, the next step in the described pipeline towards characterizing responsible epigenetic modifiers is mapping their genomic localization. Genetic linkage analysis (LA) is the method of choice. In contrast to earlier studies[285-287,318], single nucleotide polymorphisms (SNPs) instead of microsatellites were used as genetic markers. Several reasons favor SNPs: firstly, SNPs seem to be less error prone[319] ; secondly, strain-specific SNPs were available at a very high density allowing for high resolution owing to the Mouse Genomes Project of the Sanger Wellcome Trust Institute[275]; finally, the iPLEX technology (MassARRAY, Sequenom) a powerful MALDI-TOF MS-based method to reliably genotype a high number of SNPs simultaneously was at hand[271].
LA categorically relies on homologous recombination. Although its resolution is theoretically determined by the density of genetic markers and the number of genotyped individuals, the lack of recombination events at highly divergent regions can foreclose a desired resolution in practice. This was the case for fine mapping *MIso* within the initial candidate region. Although the experimental setup allowed for a minimal resolution of 1 Mb, no recombination events were observed within the current 9 Mb candidate region.

Since no other *trans*-regulated region was validated as yet, I have executed the complete pipeline only for the DMR at the *Isoc2b* promoter so far. The newly identified DMRs were exclusively regulated in *cis*. This observation may pose another indication for the developmental importance of epigenetic modifiers. It was reported that the crossing of heterozygous mutants leads to reduced litter sizes due to embryonic lethality of homozygous mutant offspring[98,287]. The successful development of a multicellular organism from the fertilized zygote depends on a well-orchestrated network of epigenetic mechanisms and seems to have little tolerance for

dysfunctional components. Based on these findings, only few strain-specific epigenetic modifiers can be expected in the healthy system.
An alternative explanation is that the applied methods to measure DNA methylation are biased towards detecting *cis*-regulated regions. Allele-specific BS depends on strain-specific sequence variations to distinguish parental alleles in F1 hybrids, therefore not all potential *trans*-regulated DMRs could be evaluated. In addition, the necessity for sequence variations might favor the selection of *cis*-regulated DMRs. Paradoxically, the sequence variations tend to impede comparative MALDI-TOF MS-based methylation analysis due to variation-caused differences in the fragmentation pattern; hence, regions with expected disparate fragmentation pattern were excluded *a priori*.

Nevertheless, the pipeline for genome-wide LA in 45 C;B6 hybrids is available and experiments to identify additional *trans*-regulated regions are on-going. As soon as a novel *trans*-regulated candidate region is determined, the corresponding methylation data of the 45 C;B6 hybrids can be processed in the described pipeline to localize the responsible epigenetic modifier. Although being limited to BS-resistant strain-specific SNPs to validate potential *trans*-regulated DMRs by allele-specific BS of F1 hybrids, the power of the pipeline could be demonstrated in practice. Genome-wide LA was also performed for the potential *trans*-regulated DMR at the *Sfi1* pseudogene identified in the initial study[261]. Due to the lack of BS-resistant SNPs in the DMR, allele-specific DNA methylation in F1 hybrids could not be completely excluded at this locus. C;B6 hybrids further displayed a more complex methylation pattern at the *Sfi1* DMR then at the *Isoc2b* DMR indicating the involvement of more than one factor. Nevertheless, LA with the initial 45 C;B6 hybrids mapped one potential epigenetic modifier to a region on chromosome 4, which overlapped with the encoding region for a recently identified novel epigenetic modifier[96].

In addition to using the iPLEX technology for LA, the assay is of great interest for other applications, for example for studying the genetic background of designed mouse models. Recent insights in the phenotypic diversity of related mouse substrains[320] emphasize the importance of a pure genetic background in genetically engineered mouse models. With the help of the resequencing data provided by the Mouse Genomes Project[275], SNP panels for specific strains can be designed to verify the genetic background of a given mouse strain. Determing the degree of cross contamination in individuals could also be used to speed up backcrossing.

6.2 *MIso*, a strain-specific epigenetic modifier, is active during early embryonic development

MALDI-TOF MS-based methylation analysis and bisulfite sequencing (BS) in parental C57BL/6 and BALB/c mice as well as F1 hybrids implied that the methylation phenotype at the *Isoc2b* promoter was not inherited in an allele-specific manner. The results suggested regulation with semi-dominant penetrance by an epigenetic modifier encoded elsewhere, since none of the parental phenotypes was fully reestablished in F1 hybrids. This is in line with earlier studies demonstrating that the establishment of epigenetic patterns is dose-dependent[287,309]. I performed genome-wide LA in C;B6 hybrids to localize the genetic trait of *MIso*. The experiment not only identified a region on chromosome 12 with a very high LOD score (>13), but also indicated the lack of additional regions with comparably increased LOD scores. Based on these finding, *MIso* was assumed to be a monogenetic trait. This is in line with the observed phenotypes and fine mapping results in C;B6 hybrids. Since the recombination frequency within the initial candidate region was unexpectedly low in C;B6 hybrids, additional backcrosses for further fine-mapping were generated with another strain than C57BL/6. Several other inbred strains that were part of the Mouse Genomes Project[275] were tested for their DNA methylation frequency at the *Isoc2b* promoter. With the exception of A/J, which displayed a hypermethylated phenotype at the *Isoc2b* promoter, all other strains tested (129S1/SvImJ, C3H/HeJ, CBA/J, DBA2/J, and NOD/ShiLtJ) exhibited a hypomethylated phenotype similar to C57BL/6. However, methylation analysis in C;NOD hybrids, derived from BALB/c and NOD/ShiLtJ mice, lacked the clear separation of two distinct methylation phenotypes. The hybrids displayed a continuous methylation pattern rather than two distinct methylation phenotypes suggesting a polygenic control in C;NOD hybrids. Nevertheless, for C57BL/6 and BALB/c mice the difference in DNA methylation at the *Isoc2b* promoter regions seems to depend on a monogenetic trait located in a 9 Mb region on the short arm of chromosome 12, according to LA.

So far, the available results do not permit for the classification of *MIso* as either a methylation-promoting modifier active in the C57BL/6 strain or a methylation-preventing modifier active in the BALB/c strain. Interestingly, BALB/c seemed to compensate the hypermethylated phenotype and the resulting lack of Isoc2b protein with an increased expression of the neighboring homolog *Isoc2a*. Although initially identified in bone marrow-derived macrophages from C57BL/6 and BALB/c mice, I could verify the strain-specific methylation pattern at the *Isoc2b* promoter also in other somatic tissues. Not only spleen, a tissue enriched for macrophages, but also brain and liver exhibited the strain-specific epiphenotype. Germ line cells obtained from testis of both strains were not differentially methylated but displayed a hypomethylated phenotype in both mouse strains. Consequently, the differential epiphenotype seems to be re-established in each generation. Since the analyzed somatic tissues originate from the three different germ layers during embryonic development, strain-specific *de novo* methylation at the *Isoc2b* promoter was

expected to take place at pregastrulational stages during early embryogenesis. This assumption is consistent with reports connecting the activity of epigenetic modifiers to early developmental stages[294,321]. Employing an *in vitro* embryonic stem cell (ESC) differentiation model with ES cell lines derived from C57BL/6NJ and BALB/cJ mice, I could confirm the activity of *MIso* during early embryonic development. Similar to the findings in germ line cells and somatic tissues, both ES cell lines displayed a hypomethylated phenotype in the pluripotent state. Upon differentiation, BALB/c ESCs actively gained cytosine methylation, whereas C57BL/6 ESCs retained the hypomethylated phenotype. Nevertheless, pluripotent ESCs as well as germ line cells from BALB/c always exhibited slight increases in the average DNA methylation level when compared to C57BL/6. A possible explanation could be that levels of 5-hydroxymethylcytosine (5hmC) differed between C57BL/6 and BALB/c. 5hmC is an intermediate of the active demethylation process and found at its highest levels in neuronal cells[322], but also in substantial amounts in early embryos and ESCs[66,82]. Methylation analyses based on bisulfite treatment do not distinguish between 5-methylcytosine (5mC) and 5hmC[300]. In order to assess the presence and dynamics of 5hmC at the *Isoc2b* promoter, I performed hydroxymethylcytosine methylated DNA immunoprecipitation (hMeDIP) on genomic DNA from differentiating C57BL/6 and BALB/c ESCs. In fact, undifferentiated BALB/c ESCs exhibited increased 5hmC levels throughout the whole DMR at the *Isoc2b* promoter in comparison to C57BL/6 ESCs. Upon differentiation, BALB/c ESCs rapidly lost the 5hmC modification. Active demethylation of 5mC seemed to retain the region in a general hypomethylated state in BALB/c ESCs. With the onset of differentiation, 5hmC levels decreased rapidly leading to the accumulation of 5mC in differentiating BALB/c ESCs. In contrast, the core of the DMR completely lacked 5hmC in C57BL/6 ESCs, but its borders slowly gained 5hmC upon differentiation. The presence of marginal 5hmC levels in C57BL/6 ESCs could block spreading of 5mC from the neighboring regions into the DMR. Still, *MIso* might actively contribute to both described scenarios. HMeDIP experiments were repeated with independent differentiation kinetics as well as with genomic DNA derived from testes over the course of a master thesis[297] to validate these findings. In C57BL/6 ESCs, 5hmC showed a similar dynamic pattern during differentiation with increasing amounts of 5hmC at the border of the DMR, which is in accordance to other studies[323,324]. As expected from the low methylation levels detected in testis, C57BL/6 germ line cells exhibit no 5hmC modification at the DMR. The hMeDIP experiment for BALB/c ESCs and germ line cells is still pending.

The 9 Mb candidate region on chromosome 12 consists of conserved as well as highly repetitive regions. The conserved region, which represents roughly one third of the candidate region, is found in three blocks and comprises numerous annotated genes. Initially, based on literature-based annotation, I considered none of the genes as a suitable candidate to promote the strain-specific DNA methylation phenotype. The bulk of the candidate region appears to be highly repetitive with 6 major assembly gaps in the *mm9* reference genome. To date, about 32000 SNPs, 5000 indels, as well as 264 structural variances are annotated for the candidate regions according to the Mouse Genomes Project[275] (REL-1211). The figures are comparable to other genomic regions of the same size. However, the low rate of recombination events within the

candidate region implies a more complex genomic architecture. Interestingly, the number of pseudogenes that seem to reside within the candidate region as annotated by various approaches in the UCSC genome browser (gene and gene prediction tracks ‘Yale Pseudo60’ (*mm9*), ‘Retroposed genes V2, incl. pseudogenes’ (*mm10*)) is more than 5-fold higher than in the neighboring regions. Thus it was speculated that the repetitive part of the candidate region, especially the annotated pseudogenes have the potential to encode an epigenetic modifier.

Although still a poorly understood area in the field of epigenetics, more and more factors targeting epigenetic modifications to specific sites in the genome are identified. Table 6-1 gives an overview of selected epigenetic modifiers and their suggested regulatory function.

Table 6-1 Epigenetic modifiers and their regulatory mechanisms

Class of epigenetic modifier	Examples	Mode of action	Target site(s)	Regulatory consequence	Proposed mechanism and co-factors
Krüppel-associated box (KRAB) repressor domain-containing zinc finger proteins (ZFP)	**Ssm1b**[96,325]	*trans*	Transgene	transcriptional repression	- (sex- and tissue-specific) targeted DNA binding by the zinc finger domain - interaction with KAP1/TRIM28 that recruits NuRD complex → deacetylation of histone residues SETDB1 → long-range spreading of H3K9me3 HP1 (binds to H3K9me3) → stabilization of repressive chromatin Dnmts → CpG methylation
	Rsl1[326]	*trans*	*Slp*	transcriptional repression	
	Znf382[327]	*trans*	oncogenes (e.g. *MYC*, *CDK6*), NF-κB downstream factors (e.g. *STAT3*, *ID1*)	transcriptional repression	
	KS1[328,329]	*trans*	oncogenes (e.g. *Ha-ras*, *Galpha12*)	transcriptional repression	
	ZFP57[330]	*trans*	group of maternally imprinted genes	transcriptional repression	
	Artifical KRAB-ZFP[331]	*trans*	designed to target *Oct4*	transcriptional activation	- the zinc finger domain targets the epigenetic modifier - repressor domains mediate context-dependent protein-protein interactions - both activation and repression were detected: → repressive mechanism as above → activating mechanism remains elusive
	ZFP263[332]	*trans*	distinct set of >5000 binding sites at promoters and in introns	transcriptional repession/activation	
SCAN repressor domain-containing ZFP	**ZFP206**[333,334]	*trans*	*Meis2*, *mir-124-a*, *mir-124-a2* (repression) *Oct4*, *Klf5*, *Jardic1c* (activation)	transcriptional repession/activation	
CXXC domain-containing proteins	**KDM2B**[335]	*trans*	subset of unmethylated CpG islands	transcriptional repression	- KDM2B forms a complex with RING2 and other PcG proteins - the complex is targeted to unmethylated CpG island bound by PRC1 and PRC2 via the CXXC domain of KDM2B - the complex ubiquitinylates H2AK119
	KDM2A[306]	*trans*	centromeric satellite repeats	transcriptional repression	KDM2A maintains heterochromatin at centromeric satellite repeats by interaction with HP1
	[336]		rDNA promoter	transcriptional repression	KDM2A affects levels of H3K36me0/1/2 at rDNA promoters, which might have functional effect on transcription level of rRNA
	[337,338]		unmethylated CpG islands	unique CpG island chromatin structure	KDM2A-mediated removal of H3K36me2 creates a functional unique open chromatin structure at CpG islands (independent of transcription)
	CFP1[176,177]	*trans*	unmethylated CpG islands	unique CpG island chromatin structure	CFP1 specifically recognizes unmethylated CpG-dense regions and interacts with the Set1 complex → methylation of H3K4
	Tet1/3, **IDAX**[66,76] (associated with Tet2)	*trans*	Marginal areas of CpG-rich promoters, CpG-poor regulatory sites, gene bodies	active DNA demethylation	- Tet1/3 and IDAX exhibit a distinct subclass of CXXC domains with varying binding affinity for cytosines and its modified derivates - mechanisms how Tet enzymes are recruited are still under discussion; one possbility is a general association with the transcription machinery

Class of epigenetic modifier	Examples	Mode of action	Target site(s)	Regulatory consequence	Proposed mechanism and co-factors
non-coding RNAs	**XIST**[203] (lncRNA)	*cis*	X chromosome	initiation of X chromosome inactivation	- Xist RNA coates future inactivated X chromosomes (regulation in *cis*) - mechanism only effective in females at early developmental stages - Xist-dependent, yet unknown recruitment of PRC2 and PRC1 → chromosome-wide accumulation of H3K27me3/ H2AK119ub, predominantly at promoters of X-linked genes - cascade of chromatin changes (also DNA methylation)
	Kcnq1ot1[207] (lncRNA)	*cis*	imprinted genes in *Kcnq1* cluster	transcriptional repression	Kcnq1ot1 recruits PRC2 via interaction with Ezh2 to imprinting *Kcnq1* cluster → deposition of H3K27me3
	Air[339,340] (lncRNA)	*cis*	imprinted genes in *Igf2r* cluster	transcriptional repression	- Air is required for silencing of the associated gene cluster - Air interacts with G9a HMTase at the *Slc22a3* promoter → deposition of H3K9me3
	HOTAIR[204,341] (lncRNA)	*trans*	*HOXD*	transcriptional repression	- HOTAIR recruits PRC2 in *trans* to HOXD locus → deposition of H3K27me3 - recruitment of Lysine-specific demethylase 1 (LSD1) → demethylation of H3K4
	PTENpg1 asRNA α[263] (lncRNA)	*trans*	*PTEN*	transcriptional repression	- PTENpg1 asRNA α directs EZH2 (PRC2) to PTEN promoter → deposition of H3K27me3 - also interacts with DNMT3A → *de novo* DNA methylation?
	Oct4pg5 asRNA[342] (lncRNA)	*trans*	*Oct4*	transcriptional repression	Oct4pg5 asRNA potentially interacts with EZH2 to maintain a repressive chromatin state (H3K27me3) in differentiated cells
	ncRNA[343]	*cis/ trans*	*DHFR*	transcriptional repression	- ncRNA forms stable triplex structure with major promoter → sequence-specific targeting - ncRNA interacts with TFIIB → disruption of the preinitiation complex required for transcription
	MILI, MIWI2 (piRNAs)[307,344]	*trans*	retrotransposons, imprinted *Rasgrf1* locus	transcriptional repression	- piRNAs guide *de novo* methyation of LINE and IAP elements in male germ line cells in a yet unknown mechanisms - a retrotransposon upstream of *Rasgrf1* is targeted by a piRNA in male germ lines resulting in *de novo* methylation and imprinting
	promoter-associated RNA (pRNA)[345]	*cis*	rDNA loci	transcriptional repression	pRNA forms RNA:DNA triple helix at the promoter, which is specifically recognized by DNMT3b → *de novo* DNA methylation

Class of epigenetic modifier	Examples	Mode of action	Target site(s)	Regulatory consequence	Proposed mechanism and co-factors
non-coding RNAs	**TUG1**[201] (lncRNA)	*trans*	E2F1-regulated growth genes	transcriptional repression	- TUG1 interacts with methylated Pc2/CBX4 (K191me2) under serum starvation conditions - Pc2 associated with TUG1 preferentially binds to H3K27me2 → localization to PcG bodies (repressive nulear environment)
	NEAT2/ MALAT1[201] (lncRNA)	*trans*		transcritional activation	- Upon serum stimulation Pc2/CBX4 is demethylated and interacts with NEAT2 - Pc2 associated with NEAT2 preferentially binds to acetylated H2A peptides → relocalization to interchromatin granule (active nuclear nvironment)
	Jpx[206,346] (lncRNA)	*cis/ trans*	*Xist*	transcriptional activation	- Jpx interacts with CTCF and competes CTCF binding to *Xist* promoter → CTCF removal is essential for *Xist* transcription, which in turn is required for X inactivation
	ecCEBPA[205] (lncRNA)	*cis*	*CEBPA*	transcriptional activation	- ecCEBPA is bound by DNMT1 → CEBPA promoter is shielded from DNMT1 binding and thus DNA methylation - loci with DNMT1-interacting RNAs are globally represented
	CHD RNA[347] (ssRNA)	*cis*	*E-cadherin*	transcriptional activation	Antisense ncRNA of the E-cadherin promoter binds to catalytic domain of DNMT3A → inhibition of DNMT3A
Other *trans* -acting factors	**Smchd1** [97,348,349]	*trans*	transgene, distinct set of CpG islands of inactivated X-linked genes, imprinted genes	transcriptional repression	- Smchd1 site-specific DNA binding is achieved by a hinge domain - targeting of slowly methylated inactivated X-linked loci - a homologuous protein in *Arabidopsis* was found to be involved in RNA-directed DNA methylation - mechanism in mammals remains elusive
	Rlf[98]	*trans*	transgene	transcriptional activation	- mutation analyses of Rlf indicate active role in preventing site-specific DNA methylation - mechanism remains elusive
	Dnmt3l [51,54,350,351]	*trans*	imprinted genes, gene bodies of housekepping genes	transcriptional repression	-expression correlated with establishment of genomic imprinting - targeting could be achieved by its PHD domain, but also by mediating interaction of Dnmt3a/b with specific transcription factors - direct interaction with Dnmt3a/b might stimulate *de novo* methylation activity
			promoters of bivalent developmental genes	maintenance of DNA hypomethylation	competitve interaction of PRC2 with Dnmt3a/b in ESCs seems to prevent *de novo* methylation

Of particular interest were Krüppel-associated box (KRAB) repressor domain-associated zinc fingers (KZF), since some of the annotated pseudogenes in the candidate region seem to contain KRAB or zinc finger domains. This protein family was associated with site-specific silencing in the past and is hypothesized to be involved in processes such as DNA damage response, cell cycle control, proliferation and tumorigenesis[352-355]. The established facultative heterochromatin was also shown to result in methylation of cytosines in the target region when propagated during early embryonic development[356] or prevented the conversion of 5mC to 5hmC in early embryos[232]. Very recently, a novel type of KZF (*Ssmb1*) was identified that methylated a transgene during mouse embryogenesis in a strain-specific manner[96]. Interestingly, the region on chromosome 4 encoding for *Ssm1b* matched the region potentially containing one of the factors regulating the strain-specific methylation phenotype of the *Sfi1* pseudogene (as mentioned above). In addition, other publications implicated a role of KZFs in the activation of specific genes[331,332]. Nevertheless, the involved mechanisms are not known as yet.

Although there are sequence variations between C57BL/6 and BALB/c affecting the KRAB-containing pseudogenes located within the candidate region for *MIso*, none of the variations detected to date seems to have a substantial effect on the putative coding sequences. As described above, the repetitiveness of the candidate regions hinders reliable sequence calling. Since the resequencing approach of the Mouse Genomes Project is still in progress, interesting pseudogenes should be under constant surveillance.

Non-coding RNAs (ncRNAs) constitute another major class of site-specific targeting factors. Chromatin modifiers often lack DNA-binding domains, but exhibit motifs for RNA binding[357]. NcRNAs that escape nuclear export are able to interact with the modifiers and guide them to specific target sites. The best-studied example is the long non-coding RNA (lncRNA) XIST, which is required for Polycomb-mediated inactivation of the X chromosome[358-360] (Table 6-1). Compared to this *cis*-acting lncRNA, HOTAIR was identified as a *trans*-acting regulator of human HOX loci[204]. HOTAIR seems to interact with the Polycomb repressor complex 2 (PRC2) to mediate gene silencing. Aberrant HOTAIR expression was detected in several human cancers underlining its regulatory importance in particular and that of lncRNA as epigenetic regulators in general[361]. Genome-wide studies on novel lncRNAs associated with PRC2 point towards additional *trans*-regulating candidates[100,153]. Moreover, it was shown that lncRNAs act during embryonic development[362]. The transcriptional regulation mediated by ncRNAs is not restricted to lncRNAs or repression, but also includes short ncRNAs and was shown to contribute to gene activation. Table 6-1 provides further examples for ncRNAs involved in epigenetic mechanisms. In summary, many of the known regulatory lncRNAs act as repressors by recruiting Polycomb repressor complexes. Sometimes they act in combination with DNA methyltransferases (DNMTs) and facilitate the establishment of repressive chromatin structures. But although sequence similarities might direct ncRNAs to their target sites, the involvement of other factors is also possible. LncRNAs often act in *cis* or seem to affect sites on the chromosome of origin, still inter-chromosomal targeting cannot be excluded.

Another class of epigenetic modifiers that seems to affect DNA methylation is the cysteine-rich CXXC domain-containing family of zinc finger proteins (Table 6-1). Members of the CXXC1

subfamily are known to bind to unmethylated CpGs located in CpG islands and account for their open chromatin architecture[175,337,363]. The open chromatin structure must not necessarily be associated with transcription. CXXC-containing zinc fingers were also shown to cause transcriptional repression, for example KDM2B was linked to Polycomb-mediated repression[335]. Aside from the targeting of hypomethylated, CpG-dense regions, other sequence-specific properties were not found for this class of epigenetic modifiers. Tet enzymes, which are to date the only known enzymatic activities to promote active DNA demethylation by converting 5mC into 5-hydroxymethylcytosine (5hmC) also exhibit a CXXC domain, but belong to another subfamily. The targeting mechanisms for Tet enzymes are not well understood as yet[76]. It was proposed that the CXXC domain might guide binding towards CpG-rich transcription start sites[70-74], but due to their broad spectrum of targets, Tet enzymes may simply be associated with the general transcription machinery.

The ENU mutagenesis screening performed by the group around Emma Whitelaw identified several known, but also novel modifiers of epigenetic reprogramming[98]. Among them the *rearranged L-myc fusion* (*Rlf*) gene, which, when mutated, caused increased expression of the transgenic GFP reporter plus a hypermethylated phenotype at the reporter locus. Genome-wide methylation analysis in *Rfl* null embryos implied an active role of Rlf in preventing DNA methylation at specific sites, but the underlying mechanism is still unclear[364]. In general, two mechanisms for maintaining a hypomethylated phenotype during embryonic development are conceivable. On the one hand, target sites might be shielded to prevent *de novo* methylation *a priori*. On the other hand, a process involving constant active demethylation by Tet enzymes might foreclose the accumulation of 5-methylcytosine (5mC).

In summary, the majority of described epigenetic modifiers is associated with repression and seems to propagate gene silencing by recruiting repressive chromatin modifiers, which introduce histone modifications such as H3K27me3 or H3K9me3. Epigenetic modifiers associated with activation are less frequent and their mode of action is largely unknown. Protein epigenetic modifiers target the epigenetic machinery by zinc finger domains in a *trans*-acting manner and mediate protein-protein interactions by their repressor domains. NcRNAs appear to mainly act in *cis* or at least on the chromosome of origin and often seem to provide a scaffold for the assembly of chromatin-modifying complexes. They are often involved in the establishment of imprinting patterns. Activation by ncRNAs seems to be achieved by site-specific competitive interaction with repressive factors (Table 6-1).

Since the strain-specific *de novo* methylation of the *Isoc2b* promoter can be recapitulated in differentiating ESCs, comparative expression analysis seemed promising to identify the responsible genetic trait. I performed next generation sequencing of RNA samples (RNA-seq) derived from of undifferentiated as well as differentiating ESCs of both strains. The selected time points of the differentiation kinetics corresponded with *de novo* methylation in BALB/c ESCs at the *Isoc2b* promoter. I chose the Trinity pipeline, a *de novo* assembly approach, to analyze the RNA-seq data to circumvent alignment biases due to mapping to the available reference genome and transcriptome (both derived from the C57BL/6 strain). Quality assessment of the *de novo*

assembled transcripts supported their validity. To date no strain-specific transcript related to the above-mentioned classes of described epigenetic modifiers has been identified. Although the reference genome is excluded during *de novo* assembly, subsequent characterization of transcripts involves comparison to an annotated genome and transcriptome impairing the identification of strain-specific transcripts. Analyses regarding strain-specific transcripts generated from the candidate region will be continued. To obtain a comprehensive analysis of the RNA-seq data, the knowledge of the actual genomic sequence of the candidate region in BALB/c mice might be essential.

In addition to RNA-seq experiments, I used complementation assays with BACs in differentiating ESCs to further narrow down the *MIso*-encoding region. Since the majority of the above-described epigenetic modifiers acted in a DNA methylation-promoting manner, ESCs from the non-methylating C57BL/6 strain were complemented with BALB/c-derived genomic sequences of the 9 Mb candidate region by stable transfection.
The BALB/c sequences were delivered as Bacterial Artificial Chromosomes (BACs) obtained from the CHORI-28 BAC library, which is based on BALB/cByJ genomic DNA. Since BALB/cByJ exhibited similar levels of *Isoc2b* promoter methylation in comparison to BALB/cAnNCrl and BALB/cJ, no negative effect on the outcome of the complementation assays was assumed. Initial next generation sequencing experiments with the selected BACs implied convenient coverage of the candidate region, but the repetitiveness of the candidate region made proper alignment difficult. Therefore, to exactly determine the coverage of the candidate region further sequencing of individual BACs is required.
Unfortunately, none of the transfected BACs induced the BALB/c-specific methylation phenotype in differentiated C57BL/6 ESCs. Several reasons could explain this outcome. Firstly, the *MIso*-encoding sequence might not be covered by the selected BACs. Screening of the BALB/c BAC library was performed with probes designed for the candidate region according the C57BL/6-derived *mm9* reference genome hence BALB/c-specific sequences might be missed. Moreover, not all selected BACs were part of the complementation assays due to unsuccessful retrofitting of the original BAC with a eukaryotic selection marker. Secondly, although the *MIso*-encoding region might be completely covered by the transfected BACs, *MIso* might not be fully represented by a single BAC. Thirdly, the *Isoc2b* promoter might be primed in BALB/c ESCs thus foreclosing *de novo* methylation in C57BL/6 ESCs. However, bisulfite sequencing results from F1 hybrids argue against this assumption, since both alleles were methylated independent of the parental origin. Finally, the design of the complementation assay assumed a methylation-promoting activity of *MIso*. Although the majority of epigenetic modifiers seem to repress their targets by promoting DNA methylation, methylation-preventing exceptions were also reported (Table 6-1).

Since none of the two candidate region-wide approaches (comparative expression analysis, complementation assays) identified a candidate gene, I decided to re-investigate the annotated genes encoded in the candidate region. Interestingly, recently published data indicated an interaction of *nucleolar protein 10* (*Nol10*) with Eed, a subunit of the PRC2 repressor complex[365],

in ESCs[185]. Nol10 is a member of a yet uncharacterized family of proteins that share the NUC153 domain of unknown function[303]. Nol10 also exhibits a WD40-repeat-containing domain, which is found in all eukaryotes and could serve as protein-protein interaction site[304,305]. Interestingly, the two hypermethylated mouse strains (BALB/c, A/J) share a non-synonymous SNP variant at rs29160165. All hypomethylated mouse inbred strains (C57BL/6 strains as well as 129S1/SvImJ, C3H/HeJ, CBA/J, DBA2/J, and NOD/ShiLtJ) share the reference variant of the SNP. The strain-specific SNP variants were confirmed on gDNA and RNA level. According to the Mouse Genomes Project[275] (REL-1303), the SNP variant resides in the splice region between exon 18 and 19 of *Nol10*, but splicing did not seem to be affected in transcripts detected in differentiating C57BL/6 and BALB/c ESCs. Nevertheless, the non-synonymous SNP variant of rs29160165 present in BALB/c and A/J causes an amino acid exchange in the COOH-terminal tail of the Nol10 protein. Since the specificity of WD40 repeat-mediate protein-protein interaction seems to be determined by sequences outside the repeat domain[304], the non-synonymous SNP variant found in BALB/c and A/J could have an influence on interaction partners. To understand how this notion can be brought in line with a potential interaction with the PRC2 complex will require further characterization of Nol10. Although Nol10 is not related to any of the site-specific epigenetic modifiers identified so far, its cellular localization as well as its domain repertoire makes it a promising candidate. In addition to BALB/c and A/J, WSB/EiJ is the third strain of the Mouse Genomes Project, which exhibits the respective SNP variant at rs29160165. The DNA methylation phenotype at the *Isoc2b* promoter in this strain might further support the involvement of Nol10. To date *Nol10* is the only promising candidate gene to encode *MIso*. Future studies will focus on the functionality of Nol10 in the context of the DMR at the *Isoc2b* promoter but also at other yet unknown target sites. Of note, it cannot be excluded that the correlation of the SNP variant at rs29160165 with the observed methylation phenotype might have appeared by chance. Therefore, the search for other candidate genes will be continued until the involvement of Nol10 is clarified.

The present work focused mainly on the analysis of the *MIso*-encoding candidate region on chromosome 12, but further studies concerning the DMR at the *Isoc2b* promoter also provided important information. Some of the epigenetic modifiers listed in Table 6-1 were found to affect the silencing of repetitive sequences[96,98,306,307]. Two retrotransposons of the SINE (short interspersed elements) class are located within the DMR at the *Isoc2b* promoter. I analyzed the generated MCIp-seq data regarding potential genome-wide differences in the methylation of repetitive sequences. Although the two inbred strains showed significant differences in a number of simple repeats, none of the retrotransposons present at the DMR associated with the *Isoc2b* promoter seemed enriched differentially by the MCIp-seq procedure. However, the read length (50 bp, single end) of the MCIp-seq data set is too short to estimate the abundancy at repetitive elements longer than 50 bp reliably.

In addition, ChIP-seq and expression analyses in differentiating ESCs are available, which were generated over the course of a master thesis[297] in our laboratory. The findings of this thesis indicated an involvement of Polycomb-mediated repression of the *Isoc2b* gene in BALB/c. Polycomb-mediated gene silencing is in general associated with the reversible repression of

differentiation genes during early embryonic development[366]. Mohn *et al*. could also demonstrate that additional CpG-rich Polycomb targets arise during ESC differentiation and that some of these targets were *de novo* methylated during lineage commitment[178]. *De novo* methylation could be achieved by recruitment of DNMTs by EZH2, a H3K27-specific histone methyltransferase part of PRC2[186], and might lock genes in a silent state. In line with these findings, pluripotent BALB/c ESCs displayed bivalent histone modifications: the active mark H3K4me3 as well as the repressive mark H3K27me3 was detected at the *Isoc2b* promoter region. Upon differentiation, the level of H3K4me3 declines, whereas the repressive mark H3K27me3 was retained. In contrast, C57BL/6 ESCs completely lacked the repressive H3K27me3 mark at the DMR at all analyzed time points, but retained high levels of H3K4me3 throughout differentiation. In accordance with these findings, expression of *Isoc2b* was detected only in differentiating C57BL/6 ESCs and started between day 9 and day 13 after removal of Leukemia Inhibiting Factor (LIF). In summary, the strain-specific epigenetic characteristics and expression profile found in differentiated ESCs are in agreement with phenotypes observed in somatic tissues. Due to limitations of chromatin specimens, other repressive histone marks have not been analyzed as yet. H3K9me3 will be of particular interest, since gene silencing by the recently identified strain-specific epigenetic modifier *Ssm1b* was associated with increasing levels of H3K9me3 and DNA methylation[96,325].

The *Isoc2b* locus seems to be poised to be activated in both strains as indicated by the hypomethylated state in both germ line cells and ESCs and the presence of H3K4me3 at the promoter in ESCs. The active state was only maintained in differentiating C57BL/6 ESCs and resulted in C57BL/6-specific expression of *Isoc2b*. In contrast, differentiating BALB/c ESCs established a repressive epiphenotype marked by the loss of H3K4me3 and the presence of H3K27me3 and DNA methylation impeding the expression of *Isoc2b*. Unexpectedly, BALB/c ESCs seemed to be already marked by H3K27me3. The expression level of the pluripotency marker *Oct4* was similar to other preparations of pluripotent ESCs. But DNA methylation levels of the undifferentiated BALB/c ESCs at the *Isoc2b* promoter and the *Nanog*-associated control region were slightly increased in comparison to other preparations. These findings imply that the BALB/c ESCs analyzed in the ChIP experiment might have reached a more committed state, which would explain the presence of H3K27me3 at the *Isoc2b* promoter. To evaluate these findings, ChIP experiments in combination with methylation and expression analyses in undifferentiated BALB/c ESCs will be repeated.

How the machinery that establishes the repressive phenotype in BALB/c is directed to the *Isoc2b* promoter remains unclear as yet. With respect to the hypothesized loss-of-function mutation of Nol10 in BALB/c, a potential mechanism might involve the spreading of heterochromatin from an upstream located H3K27me3 island (Geo database GSE49431[367]). In this scenario, Nol10 might take part in effectively blocking the spreading in C57BL/6, but its mutated version fails to do so in BALB/c. Further investigations regarding the functionality of Nol10 as well as the dynamic chromatin environment at the *Isoc2b* promoter during the process of *de novo* DNA methylation will help to untangle the regulatory network.

Taken together, the data provide strong experimental evidence that the strain-specific epiphenotype at the *Isoc2b* promoter is regulated in *trans* by a monogenetic trait during early embryonic development. With *Nol10* a potent candidate gene for *MIso* is on hand. The fact that so far no additional *trans*-regulated DMRs were identified emphasizes the importance of these regulators for accurate development and functionality of a multicellular organism. If their functionality was less import, their regulatory effect would be less essential and a higher number of *trans*-regulated variations among individuals were to be expected. Therefore, identification and characterization of *MIso* will provide insights to the regulatory mechanisms that shape the epigenome during early developmental stages.

6.3 Perspectives

Linkage analyses located the monogenetic trait encoding *MIso*, the *trans*-acting epigenetic modifier responsible for the strain-specific DNA methylation phenotype at the *Isoc2b* promoter, in a 9 Mb region on chromosome 12. With *Nol10*, which is encoded within the candidate region, a potential candidate gene is on hand. A non-synonymous SNP variant correlates with the observed methylation phenotypes not only in C57BL/6 and BALB/c, but also 129S1/SvImJ, A/J, C3H/HeJ, CBA/J, DBA2/J, and NOD/ShiLtJ. Still, its functional participation in the establishment of the differentially methylated region (DMR) needs to be validated. The next step will be to analyze DNA methylation levels in the third strain (WSB/EiJ) of the Mouse Genomes Project[275] that carries the same mutation as BALB/c and A/J. Assuming a positive correlation of the SNP variant and the methylation phenotype, bidirectional complementation assays will be performed with BACs representing the candidate gene. Not only C57BL/6 ESCs will be transfected with BALB/c-derived BACs, but also BALB/c ESCs with C57BL/6-derived BACs. The assay can further be performed with cDNA clones allowing for the expression of a tagged exogenous version of the protein. With the help of the tag, immunoprecipitation experiments can be applied to validate the presence of Nol10 at the *Isoc2b* promoter as well as to identify other Nol10 target sites throughout the genome. In addition, cDNA clones can be subject to targeted mutation of the coding sequence for functional analyses. In this context, ES cell lines with mutations at the endogenous locus will also be of interest. To date, 21 different ES cell clones with various mutations at the *Nol10* locus are listed in the MGI database. In the long term, generation of a transgenic mouse model would be desirable.

Since the involvement of Nol10 is not validated as yet, the available RNA-seq data sets will be subject to further analysis. Determing the actual BALB/c sequence of the *MIso*-encoding candidate region on chromosome 12 will be essential to identify strain-specific transcripts or transcript variants. Since to date only draft versions of the BALB/c genome are available, individual high-throughput sequencing of the screened BALB/c BAC clones should be performed to *de novo* assemble the BALB/c sequence for the *MIso*-encoding candidate region.

Furthermore, epigenetic profiling of the DMR at *Isoc2b* promoter in ESCs should be extended. Not only the presence of other histone modifications (such as H3K9me3), but also partial deletion of the DMR might provide information on the regulatory mechanism. In this context, it would be interesting to assess the *Isoc2b* promoter methylation in mouse clones of the first and second generation generated from ESCs with truncated endogenous *Isoc2b* promoter sequences. In addition to manipulating the endogenous *Isoc2b* promoter sequence, exogenous fragments of the promoter region could be introduced to ESCs to screen for the sequence part that is sufficient to recruit the strain-specific DNA methylation machinery.

Regarding repetitive elements as a potential target of the silencing machinery, it might be reasonable to reproduce the high-throughput sequencing of the MCIp samples with longer read length and paired ends. In this way, the representation of longer repetitive elements such as those present within the DMR at the *Isoc2b* promoter could be analyzed in a reliable manner.

Finally, the identification of additional *trans*-regulated DMRs will be of high priority. Not only DMRs targeted by *MIso,* but other factors, are of interest to augment the portfolio of strain-specific *trans*-acting epigenetic modifiers. Novel candidate DMRs will be defined based on the available MCIp-seq data sets of both spleen and cerebellum from C57BL/6 and BALB/c.

7 Summary

DNA methylation and histone modifications are epigenetic mechanisms that affect the chromatin structure and thus the readout of the underlying DNA sequence. Hence, variations in the DNA methylation profiles of individuals may contribute to phenotypic differences and disease susceptibility. However, the extent of such epigenetic variations is unclear. With regard to DNA methylation, little is known on how these differences are established or inherited.

This thesis was concerned with elucidating the role that genetic determinants play in maintaining specific DNA methylation patterns. Two common inbred mouse strains (C57BL/6 and BALB/c) were used to address inter-individual variations in DNA methylation. Little is known about *trans*-acting epigenetic modifiers that control the epiphenotype at distinct target loci. Therefore the presented work focussed primarily on the identification of the *trans*-acting epigenetic modifier regulating the strain-specific DNA methylation pattern at the *Isoc2b* promoter. It could be demonstrated that the DMR is established during early embryonic development in BALB/c, but not in C57BL/6. The DMR could further be found in other inbred mouse strains. Using embryonic stem (ES) cell differentiation models, the strain-specific DNA methylation patterns were also detectable *in vitro*. Linkage analysis in C;B6 backcrosses narrowed down the genomic localization of the modifier to a highly repetitive 9 MB region on chromosome 12. To further characterize the epigenetic modifier gene, complementation assays with BAC clones covering the candidate regions as well as expression analyses (RNA-seq) in the *in vitro* ES cell differentiation models were performed. The latter identified *Nol10* as a candidate gene to propagate the observed strain-specific epiphenotypes. Additionally, novel candidate DMRs based on strain-specific patterns of active histone modifications were selected and validated to identify additional *trans*-regulated DMRs. In future, the investigation of DMRs will be extended, since to date only novel *cis*-regulated DMRs have been discovered.

Further characterization of the potential epigenetic modifier Nol10 might provide fundamental insights into mechanisms that establish DNA methylation patterns during embryonic development and how inter-individual differences can be introduced.

Zusammenfassung

Epigenetische Mechanismen, wie DNA-Methylierung und die post-translationale Modifikation von Histonen, beeinflussen die Chromatinstruktur und damit die Zugänglichkeit und Expression der assoziierten DNA-Sequenz. Individuelle Variationen im DNA-Methylierungsprofil haben damit das Potenzial zu unterschiedlichen Phänotypen und der Anfälligkeit für bestimmte Krankheiten beizutragen. Dennoch ist aktuell noch wenig über das Ausmaß solch individueller epigenetischer Unterschiede bekannt, ebenso wie diese grundsätzlich etabliert und vererbt werden.

Diese Arbeit setzt sich mit dem Einfluss von genetischen Faktoren auf spezifische DNA-Methylierungsmuster auseinander. Zwei häufig verwendete Inzuchtmausstämme, C57BL/6 und BALB/c, wurden als Modellsystem gewählt, um der Frage nach inter-individuellen Unterschieden im DNA-Methylierungsmuster nachzugehen. Zum heutigen Zeitpunkt ist noch wenig über *trans*-agierende Faktoren bekannt, die gezielt den Epiphänotyp beeinflussen. Die vorliegende Arbeit zielte deshalb primär darauf ab, den verantwortlichen Faktor für das stammspezifische Methylierungsmuster einer bereits bekannten differentiell methylierten Region, dem *Isoc2b*-Promoter, zu identifizieren. Es konnte gezeigt werden, dass die Methylierung in BALB/c-Mäusen im Zuge der frühen Embryonalentwicklung in jeder Generation neu etabliert wird, während C57BL/6-Mäuse generell einen hypomethylierten Zustand beibehalten. Dieses differentielle Methylierungsmuster gilt auch in anderen Inzuchtmausstämmen und konnte außerdem in einem Differenzierungsmodell mit embryonalen Stammzellen (ES-Zellen) nachgewiesen werden. Mit Hilfe von Kopplungsanalysen in C;B6-Hybriden wurde die für den Faktor kodierende Region auf einen 9 Mb-Bereich auf Chromosom 12 eingeschränkt. Diese Region ist allerdings hoch repetitiv. Um das für den Faktor kodierende Gen weiter zu charakterisieren, wurden sowohl Komplementierungsassays mit BAC-Klonen, die die Kandidatenregion abdecken, als auch Expressionsanalysen mittels RNA-seq im ES-Zell-Differenzierungsmodell durchgeführt. Letztere identifizierten *Nol10* als mögliches Kandidatengen, um den beobachteten stammspezifischen Epiphänotyp hervorzurufen. Desweiteren wurden neue potentielle DMRs auf Basis von stammspezifischen Histonmodifikationsmustern ausgewählt und validiert. Da bisher aber nur weitere in *cis* regulierte DMRs gefunden wurden, wird die Suche nach *trans*-regulierten DMRs weiter fortgesetzt.

Die eingehende Charakterisierung des potentiellen epigenetischen Modifiers Nol10 gestattet möglicherweise fundamentale Einsichten darüber, wie spezifische Methylierungsmuster während der Embryonalentwicklung gesetzt und damit inter-individuelle Unterschiede eingeführt werden können.

8 References

1 Malik, S. & Roeder, R. G. Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends in biochemical sciences* **30**, 256-263, doi:10.1016/j.tibs.2005.03.009 (2005).

2 Carrera, I. & Treisman, J. E. Message in a nucleus: signaling to the transcriptional machinery. *Current opinion in genetics & development* **18**, 397-403, doi:10.1016/j.gde.2008.07.007 (2008).

3 Waddington, C. H. The epigenotype. 1942. *International journal of epidemiology* **41**, 10-13, doi:10.1093/ije/dyr184 (2012).

4 Mercer, T. R. & Mattick, J. S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nature structural & molecular biology* **20**, 300-307, doi:10.1038/nsmb.2480 (2013).

5 Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: a landscape takes shape. *Cell* **128**, 635-638, doi:10.1016/j.cell.2007.02.006 (2007).

6 Lennartsson, A. & Ekwall, K. Histone modification patterns and epigenetic codes. *Biochimica et biophysica acta* **1790**, 863-868, doi:10.1016/j.bbagen.2008.12.006 (2009).

7 Inbar-Feigenberg, M., Choufani, S., Butcher, D. T., Roifman, M. & Weksberg, R. Basic concepts of epigenetics. *Fertility and sterility* **99**, 607-615, doi:10.1016/j.fertnstert.2013.01.117 (2013).

8 Lechner, M., Boshoff, C. & Beck, S. Cancer epigenome. *Advances in genetics* **70**, 247-276, doi:10.1016/B978-0-12-380866-0.60009-5 (2010).

9 Van den Hove, D. L. *et al.* Epigenetically regulated microRNAs in Alzheimer's disease. *Neurobiology of aging* **35**, 731-745, doi:10.1016/j.neurobiolaging.2013.10.082 (2014).

10 Bestor, T. H. & Ingram, V. M. Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 5559-5563 (1983).

11 Singal, R. & Ginder, G. D. DNA methylation. *Blood* **93**, 4059-4070 (1999).

12 McGhee, J. D. & Ginder, G. D. Specific DNA methylation sites in the vicinity of the chicken beta-globin genes. *Nature* **280**, 419-420 (1979).

13 Kass, S. U., Landsberger, N. & Wolffe, A. P. DNA methylation directs a time-dependent repression of transcription initiation. *Current biology : CB* **7**, 157-165 (1997).

14 Siegfried, Z. *et al.* DNA methylation represses transcription in vivo. *Nature genetics* **22**, 203-206, doi:10.1038/9727 (1999).

15 Mohandas, T., Sparkes, R. S. & Shapiro, L. J. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* **211**, 393-396 (1981).

16 Goto, T. & Monk, M. Regulation of X-chromosome inactivation in development in mice and humans. *Microbiology and molecular biology reviews : MMBR* **62**, 362-378 (1998).

17 Chaligne, R. & Heard, E. X-chromosome inactivation in development and cancer. *FEBS letters* **588**, 2514-2522, doi:10.1016/j.febslet.2014.06.023 (2014).

18 Sapienza, C., Peterson, A. C., Rossant, J. & Balling, R. Degree of methylation of transgenes is dependent on gamete of origin. *Nature* **328**, 251-254, doi:10.1038/328251a0 (1987).

19 Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362-365, doi:10.1038/366362a0 (1993).

20 Li, X. Genomic imprinting is a parental effect established in mammalian germ cells. *Current topics in developmental biology* **102**, 35-59, doi:10.1016/B978-0-12-416024-8.00002-7 (2013).

21 Lehnertz, B. *et al.* Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Current biology : CB* **13**, 1192-1200 (2003).

22 Wong, J. Y. *et al.* The association between global DNA methylation and telomere length in a longitudinal study of boilermakers. *Genetic epidemiology* **38**, 254-264, doi:10.1002/gepi.21796 (2014).

23 Feenstra, A., Fewell, J., Lueders, K. & Kuff, E. In vitro methylation inhibits the promotor activity of a cloned intracisternal A-particle LTR. *Nucleic acids research* **14**, 4343-4352 (1986).

24 Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends in genetics : TIG* **13**, 335-340 (1997).

25 Crichton, J. H., Dunican, D. S., Maclennan, M., Meehan, R. R. & Adams, I. R. Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. *Cellular and molecular life sciences : CMLS* **71**, 1581-1605, doi:10.1007/s00018-013-1468-0 (2014).

26 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).

27 Messerschmidt, D. M., Knowles, B. B. & Solter, D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes & development* **28**, 812-828, doi:10.1101/gad.234294.113 (2014).

28 Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926 (1992).

29 Yin, L. J. *et al.* Insufficient maintenance DNA methylation is associated with abnormal embryonic development. *BMC medicine* **10**, 26, doi:10.1186/1741-7015-10-26 (2012).

30 Ehrlich, M. The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. *Clinical immunology* **109**, 17-28 (2003).

31 Cancer Genome Atlas Research, N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England journal of medicine* **368**, 2059-2074, doi:10.1056/NEJMoa1301689 (2013).

32 Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic acids research* **8**, 1499-1504 (1980).

33 Gardiner-Garden, M. & Frommer, M. CpG islands in vertebrate genomes. *Journal of molecular biology* **196**, 261-282 (1987).

34 Bird, A. P. CpG islands as gene markers in the vertebrate nucleus. *Trends in genetics : TIG* **3**, 342-347 (1987).

35 Antequera, F. & Bird, A. Number of CpG islands and genes in human and mouse. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 11995-11999 (1993).

36 Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 1412-1417, doi:10.1073/pnas.0510310103 (2006).

37 Bestor, T., Laudano, A., Mattaliano, R. & Ingram, V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal of molecular biology* **203**, 971-983 (1988).

38 Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature genetics* **19**, 219-220, doi:10.1038/890 (1998).

39 Yoder, J. A. & Bestor, T. H. A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Human molecular genetics* **7**, 279-284 (1998).

40 Aapola, U. *et al.* Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics* **65**, 293-298, doi:10.1006/geno.2000.6168 (2000).

41 Lan, J., Hua, S., He, X. & Zhang, Y. DNA methyltransferases and methyl-binding proteins of mammals. *Acta biochimica et biophysica Sinica* **42**, 243-252 (2010).

42 Lauster, R., Trautner, T. A. & Noyer-Weidner, M. Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. *Journal of molecular biology* **206**, 305-312 (1989).

43 Posfai, J., Bhagwat, A. S., Posfai, G. & Roberts, R. J. Predictive motifs derived from cytosine methyltransferases. *Nucleic acids research* **17**, 2421-2435 (1989).

44 Santi, D. V., Garrett, C. E. & Barr, P. J. On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. *Cell* **33**, 9-10 (1983).

45 Challen, G. A. *et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nature genetics* **44**, 23-31, doi:10.1038/ng.1009 (2012).

46 Tatton-Brown, K. *et al.* Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nature genetics* **46**, 385-388, doi:10.1038/ng.2917 (2014).

47 Gruenbaum, Y., Cedar, H. & Razin, A. Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature* **295**, 620-622 (1982).

48 Sharif, J. *et al.* The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**, 908-912, doi:10.1038/nature06397 (2007).

49 Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science* **293**, 1089-1093, doi:10.1126/science.1063443 (2001).

50 Goll, M. G. *et al.* Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science* **311**, 395-398, doi:10.1126/science.1120976 (2006).

51 Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536-2539, doi:10.1126/science.1065848 (2001).

52 Chedin, F., Lieber, M. R. & Hsieh, C. L. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16916-16921, doi:10.1073/pnas.262443999 (2002).

53 Hata, K., Okano, M., Lei, H. & Li, E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129**, 1983-1993 (2002).

54 Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H. & Tajima, S. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *The Journal of biological chemistry* **279**, 27816-27823, doi:10.1074/jbc.M400181200 (2004).

55 Huret JL, D. P., Bernheim A. *An internet database on genetics in oncology*, <http://AtlasGeneticsOncology.org> (2003).

56 Kim, M. S. *et al.* DNA demethylation in hormone-induced transcriptional derepression. *Nature* **461**, 1007-1012, doi:10.1038/nature08456 (2009).

57 Schmidl, C. *et al.* Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome research* **19**, 1165-1174, doi:10.1101/gr.091470.109 (2009).

58 Klug, M., Schmidhofer, S., Gebhard, C., Andreesen, R. & Rehli, M. 5-Hydroxymethylcytosine is an essential intermediate of active DNA demethylation processes in primary human monocytes. *Genome biology* **14**, R46, doi:10.1186/gb-2013-14-5-r46 (2013).

59 Varley, K. E. *et al.* Dynamic DNA methylation across diverse human cell lines and tissues. *Genome research* **23**, 555-567, doi:10.1101/gr.147942.112 (2013).

60 Matsuo, K. *et al.* An embryonic demethylation mechanism involving binding of transcription factors to replicating DNA. *The EMBO journal* **17**, 1446-1453, doi:10.1093/emboj/17.5.1446 (1998).

61 Hsieh, C. L. Evidence that protein binding specifies sites of DNA demethylation. *Molecular and cellular biology* **19**, 46-56 (1999).

62 Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403**, 501-502, doi:10.1038/35000654 (2000).

63 Oswald, J. *et al.* Active demethylation of the paternal genome in the mouse zygote. *Current biology : CB* **10**, 475-478 (2000).

64 Cervoni, N., Bhattacharya, S. & Szyf, M. DNA demethylase is a processive enzyme. *The Journal of biological chemistry* **274**, 8363-8366 (1999).

65 Ramchandani, S., Bhattacharya, S. K., Cervoni, N. & Szyf, M. DNA methylation is a reversible biological signal. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 6107-6112 (1999).

66 Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935, doi:10.1126/science.1170116 (2009).

67 Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472-479, doi:10.1038/nature12750 (2013).

68 Li, C. & Li, C. DNA Demethylation Pathways: Recent Insights. *Genetics & Epigenetics*, 43, doi:10.4137/geg.s12143 (2013).

69 Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell cycle* **8**, 1698-1710 (2009).

70 Zhang, H. *et al.* TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell research* **20**, 1390-1393, doi:10.1038/cr.2010.156 (2010).

71 Frauer, C. *et al.* Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PloS one* **6**, e16627, doi:10.1371/journal.pone.0016627 (2011).

72 Xu, Y. *et al.* Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Molecular cell* **42**, 451-464, doi:10.1016/j.molcel.2011.04.005 (2011).

73 Tan, L. & Shi, Y. G. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* **139**, 1895-1902, doi:10.1242/dev.070771 (2012).

74 Xu, Y. *et al.* Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. *Cell* **151**, 1200-1213, doi:10.1016/j.cell.2012.11.014 (2012).

75 Ko, M. *et al.* Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* **497**, 122-126, doi:10.1038/nature12052 (2013).

76 Pastor, W. A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature reviews. Molecular cell biology* **14**, 341-356, doi:10.1038/nrm3589 (2013).

77 Hajkova, P. *et al.* Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329**, 78-82, doi:10.1126/science.1187945 (2010).

78 Surani, M. A. & Hajkova, P. Epigenetic reprogramming of mouse germ cells toward totipotency. *Cold Spring Harbor symposia on quantitative biology* **75**, 211-218, doi:10.1101/sqb.2010.75.010 (2010).

79 Gu, T. P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606-610, doi:10.1038/nature10443 (2011).

80 Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129-1133, doi:10.1038/nature09303 (2010).

81 Koh, K. P. *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell stem cell* **8**, 200-213, doi:10.1016/j.stem.2011.01.008 (2011).

82 Ruzov, A. *et al.* Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. *Cell research* **21**, 1332-1342, doi:10.1038/cr.2011.113 (2011).

83 Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer cell* **18**, 553-567, doi:10.1016/j.ccr.2010.11.015 (2010).

84 Pronier, E. *et al.* Inhibition of TET2-mediated conversion of 5-methylcytosine to 5-hydroxymethylcytosine disturbs erythroid and granulomonocytic differentiation of human hematopoietic progenitors. *Blood* **118**, 2551-2555, doi:10.1182/blood-2010-12-324707 (2011).

85 Spruijt, C. G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146-1159, doi:10.1016/j.cell.2013.02.004 (2013).

86 Huang, Y. & Rao, A. New functions for DNA modifications by TET-JBP. *Nature structural & molecular biology* **19**, 1061-1064, doi:10.1038/nsmb.2437 (2012).

87 Maurer-Stroh, S. *et al.* The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends in biochemical sciences* **28**, 69-74, doi:10.1016/S0968-0004(03)00004-5 (2003).

88 Huyen, Y. *et al.* Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* **432**, 406-411, doi:10.1038/nature03114 (2004).

89 Qiu, C., Sawada, K., Zhang, X. & Cheng, X. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nature structural biology* **9**, 217-224, doi:10.1038/nsb759 (2002).

90 Matzke, M. A. & Birchler, J. A. RNAi-mediated pathways in the nucleus. *Nature reviews. Genetics* **6**, 24-35, doi:10.1038/nrg1500 (2005).

91 Kawasaki, H. & Taira, K. Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature* **431**, 211-217, doi:10.1038/nature02889 (2004).

92 Morris, K. V., Chan, S. W., Jacobsen, S. E. & Looney, D. J. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289-1292, doi:10.1126/science.1101372 (2004).

93 Di Croce, L. *et al.* Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor. *Science* **295**, 1079-1082, doi:10.1126/science.1065173 (2002).

94 Brenner, C. *et al.* Myc represses transcription through recruitment of DNA methyltransferase corepressor. *The EMBO journal* **24**, 336-346, doi:10.1038/sj.emboj.7600509 (2005).

95 Hormaeche, I. & Licht, J. D. Chromatin modulation by oncogenic transcription factors: new complexity, new therapeutic targets. *Cancer cell* **11**, 475-478, doi:10.1016/j.ccr.2007.05.005 (2007).

96 Ratnam, S. *et al.* Identification of Ssm1b, a novel modifier of DNA methylation, and its expression during mouse embryogenesis. *Development* **141**, 2024-2034, doi:10.1242/dev.105726 (2014).

97 Blewitt, M. E. *et al.* SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nature genetics* **40**, 663-669, doi:10.1038/ng.142 (2008).

98 Daxinger, L. *et al.* An ENU mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse. *Genome biology* **14**, R96, doi:10.1186/gb-2013-14-9-r96 (2013).

99 Johnsson, P., Morris, K. V. & Grander, D. Pseudogenes: a novel source of trans-acting antisense RNAs. *Methods in molecular biology* **1167**, 213-226, doi:10.1007/978-1-4939-0835-6_14 (2014).

100 Nazer, E. & Lei, E. P. Modulation of chromatin modifying complexes by noncoding RNAs in trans. *Current opinion in genetics & development* **25**, 68-73, doi:10.1016/j.gde.2013.11.019 (2014).

101 Watt, F. & Molloy, P. L. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes & development* **2**, 1136-1143 (1988).

102 Iguchi-Ariga, S. M. & Schaffner, W. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes & development* **3**, 612-619 (1989).

103 Reddington, J. P. *et al.* Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome biology* **14**, R25, doi:10.1186/gb-2013-14-3-r25 (2013).

104 Boyes, J. & Bird, A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* **64**, 1123-1134 (1991).

105 Bird, A. & Macleod, D. Reading the DNA methylation signal. *Cold Spring Harbor symposia on quantitative biology* **69**, 113-118, doi:10.1101/sqb.2004.69.113 (2004).

106 Meehan, R. R., Lewis, J. D. & Bird, A. P. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic acids research* **20**, 5085-5092 (1992).

107 Nan, X., Meehan, R. R. & Bird, A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic acids research* **21**, 4886-4892 (1993).

108 Ng, H. H. *et al.* MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nature genetics* **23**, 58-61, doi:10.1038/12659 (1999).

109 Feng, Q. & Zhang, Y. The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes & development* **15**, 827-832, doi:10.1101/gad.876201 (2001).

110 Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. & Bird, A. P. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58**, 499-507 (1989).

111 Menafra, R. *et al.* Genome-wide binding of MBD2 reveals strong preference for highly methylated loci. *PloS one* **9**, e99603, doi:10.1371/journal.pone.0099603 (2014).

112 Clouaire, T., de Las Heras, J. I., Merusi, C. & Stancheva, I. Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA. *Nucleic acids research* **38**, 4620-4634, doi:10.1093/nar/gkq228 (2010).

113 Bird, A. P. & Wolffe, A. P. Methylation-induced repression--belts, braces, and chromatin. *Cell* **99**, 451-454 (1999).

114 Baubec, T., Ivanek, R., Lienert, F. & Schubeler, D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153**, 480-492, doi:10.1016/j.cell.2013.03.011 (2013).

115 Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A. & Bird, A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes & development* **15**, 710-723, doi:10.1101/gad.194101 (2001).

116 Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J. & Bird, A. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* **401**, 301-304, doi:10.1038/45843 (1999).

117 Kondo, E., Gu, Z., Horii, A. & Fukushige, S. The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. *Molecular and cellular biology* **25**, 4388-4396, doi:10.1128/MCB.25.11.4388-4396.2005 (2005).

118 Caballero, M. The Methyl-CpG Binding Proteins Mecp2, Mbd2 and Kaiso Are Dispensable for Mouse Embryogenesis, but Play a Redundant Function in Neural Differentiation. *PloS one* **4**, doi:10.1371/journal.pone.0004315.t001 (2009).

119 Prokhortchouk, A. *et al.* The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes & development* **15**, 1613-1618, doi:10.1101/gad.198501 (2001).

120 Daniel, J. M., Spring, C. M., Crawford, H. C., Reynolds, A. B. & Baig, A. The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. *Nucleic acids research* **30**, 2911-2919 (2002).

121 Johnson, L. M. *et al.* SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* **507**, 124-128, doi:10.1038/nature12931 (2014).

122 Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y. & Shirakawa, M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* **455**, 818-821, doi:10.1038/nature07249 (2008).

123 Hutchins, A. S. *et al.* Gene silencing quantitatively controls the function of a developmental trans-activator. *Molecular cell* **10**, 81-91 (2002).

124 Sarraf, S. A. & Stancheva, I. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Molecular cell* **15**, 595-605, doi:10.1016/j.molcel.2004.06.043 (2004).

125 Klose, R. J. *et al.* DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Molecular cell* **19**, 667-678, doi:10.1016/j.molcel.2005.07.021 (2005).

126 Sakamoto, Y. *et al.* Overlapping roles of the methylated DNA-binding protein MBD1 and polycomb group proteins in transcriptional repression of HOXA genes and heterochromatin foci formation. *The Journal of biological chemistry* **282**, 16391-16400, doi:10.1074/jbc.M700011200 (2007).

127 Klose, R. J. & Bird, A. P. Genomic DNA methylation: the mark and its mediators. *Trends in biochemical sciences* **31**, 89-97, doi:10.1016/j.tibs.2005.12.008 (2006).

128 Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature genetics* **19**, 187-191, doi:10.1038/561 (1998).

129 Wade, P. A. *et al.* Histone deacetylase directs the dominant silencing of transcription in chromatin: association with MeCP2 and the Mi-2 chromodomain SWI/SNF ATPase. *Cold Spring Harbor symposia on quantitative biology* **63**, 435-445 (1998).

130 Yoon, H. G., Chan, D. W., Reynolds, A. B., Qin, J. & Wong, J. N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. *Molecular cell* **12**, 723-734 (2003).

131 Morey, L. *et al.* MBD3, a component of the NuRD complex, facilitates chromatin alteration and deposition of epigenetic marks. *Molecular and cellular biology* **28**, 5912-5923, doi:10.1128/MCB.00467-08 (2008).

132 Joel, P., Shao, W. & Pratt, K. A nuclear protein with enhanced binding to methylated Sp1 sites in the AIDS virus promoter. *Nucleic acids research* **21**, 5786-5793 (1993).

133 Gilbert, N. *et al.* DNA methylation affects nuclear organization, histone modifications, and linker histone binding but not chromatin compaction. *The Journal of cell biology* **177**, 401-411, doi:10.1083/jcb.200607133 (2007).

134 Lorincz, M. C., Dickerson, D. R., Schmitt, M. & Groudine, M. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nature structural & molecular biology* **11**, 1068-1075, doi:10.1038/nsmb840 (2004).

135 Buschhausen, G., Wittig, B., Graessmann, M. & Graessmann, A. Chromatin structure is required to block transcription of the methylated herpes simplex virus thymidine kinase gene. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 1177-1181 (1987).

136 Kass, S. U., Pruss, D. & Wolffe, A. P. How does DNA methylation repress transcription? *Trends in genetics : TIG* **13**, 444-449 (1997).

137 Fuks, F., Burgers, W. A., Brehm, A., Hughes-Davies, L. & Kouzarides, T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nature genetics* **24**, 88-91, doi:10.1038/71750 (2000).

138 Fuks, F., Burgers, W. A., Godin, N., Kasai, M. & Kouzarides, T. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *The EMBO journal* **20**, 2536-2544, doi:10.1093/emboj/20.10.2536 (2001).

139 Geiman, T. M. *et al.* DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. *Biochemical and biophysical research communications* **318**, 544-555, doi:10.1016/j.bbrc.2004.04.058 (2004).

140 Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868-871 (1974).

141 Levitt, M. How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 640-644 (1978).

142 McGhee, J. D. & Felsenfeld, G. Nucleosome structure. *Annual review of biochemistry* **49**, 1115-1156, doi:10.1146/annurev.bi.49.070180.005343 (1980).

143 Over, R. S. & Michaels, S. D. Open and closed: the roles of linker histones in plants and animals. *Molecular plant* **7**, 481-491, doi:10.1093/mp/sst164 (2014).

144 Allfrey, V. G., Faulkner, R. & Mirsky, A. E. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **51**, 786-794 (1964).

145 Nikolov, M. & Fischle, W. Systematic analysis of histone modification readout. *Molecular bioSystems* **9**, 182-194, doi:10.1039/c2mb25328c (2013).

146 de la Cruz, X., Lois, S., Sanchez-Molina, S. & Martinez-Balbas, M. A. Do protein motifs read the histone code? *BioEssays : news and reviews in molecular, cellular and developmental biology* **27**, 164-175, doi:10.1002/bies.20176 (2005).

147 Stefanovsky, V., Dimitrov, S. I., Russanova, V. R., Angelov, D. & Pashev, I. G. Laser-induced crosslinking of histones to DNA in chromatin and core particles: implications in studying histone-DNA interactions. *Nucleic acids research* **17**, 10069-10081 (1989).

148 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260, doi:10.1038/38444 (1997).

149 Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705, doi:10.1016/j.cell.2007.02.005 (2007).

150 Imhof, A. Epigenetic regulators and histone modification. *Briefings in functional genomics & proteomics* **5**, 222-227, doi:10.1093/bfgp/ell030 (2006).

151 Buratowski, S. & Kim, T. The role of cotranscriptional histone methylations. *Cold Spring Harbor symposia on quantitative biology* **75**, 95-102, doi:10.1101/sqb.2010.75.036 (2010).

152 Aranda, S., Rutishauser, D. & Ernfors, P. Identification of a large protein network involved in epigenetic transmission in replicating DNA of embryonic stem cells. *Nucleic acids research* **42**, 6972-6986, doi:10.1093/nar/gku374 (2014).

153 Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 11667-11672, doi:10.1073/pnas.0904715106 (2009).

154 Hall, L. L. & Lawrence, J. B. The cell biology of a novel chromosomal RNA: chromosome painting by XIST/Xist RNA initiates a remodeling cascade. *Seminars in cell & developmental biology* **14**, 369-378 (2003).

155 Bhartiya D., L. N. *Pluripotent Stem Cells*. (2013).

156 Allard, S. *et al.* NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. *The EMBO journal* **18**, 5108-5119, doi:10.1093/emboj/18.18.5108 (1999).

157 Kuo, M. H., vom Baur, E., Struhl, K. & Allis, C. D. Gcn4 activator targets Gcn5 histone acetyltransferase to specific promoters independently of transcription. *Molecular cell* **6**, 1309-1320 (2000).

158 Marmorstein, R. & Zhou, M. M. Writers and Readers of Histone Acetylation: Structure, Mechanism, and Inhibition. *Cold Spring Harbor perspectives in biology* **6**, doi:10.1101/cshperspect.a018762 (2014).

159 Mutskov, V. *et al.* Persistent interactions of core histone tails with nucleosomal DNA following acetylation and transcription factor binding. *Molecular and cellular biology* **18**, 6293-6304 (1998).

160 Shahbazian, M. D. & Grunstein, M. Functions of site-specific histone acetylation and deacetylation. *Annual review of biochemistry* **76**, 75-100, doi:10.1146/annurev.biochem.76.052705.162114 (2007).

161 Agalioti, T., Chen, G. & Thanos, D. Deciphering the transcriptional histone acetylation code for a human gene. *Cell* **111**, 381-392 (2002).

162 Roh, T. Y., Ngau, W. C., Cui, K., Landsman, D. & Zhao, K. High-resolution genome-wide mapping of histone modifications. *Nature biotechnology* **22**, 1013-1016, doi:10.1038/nbt990 (2004).

163 Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112, doi:10.1038/nature07829 (2009).

164 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

165 Ringrose, L. & Paro, R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annual review of genetics* **38**, 413-443, doi:10.1146/annurev.genet.38.072902.091907 (2004).

166 Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128**, 735-745, doi:10.1016/j.cell.2007.02.009 (2007).

167 Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560, doi:10.1038/nature06008 (2007).

168 Simon, J. Locking in stable states of gene expression: transcriptional control during Drosophila development. *Current opinion in cell biology* **7**, 376-385 (1995).

169 Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876-880, doi:10.1038/nature03877 (2005).

170 Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nature cell biology* **8**, 532-538, doi:10.1038/ncb1403 (2006).

171 Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326, doi:10.1016/j.cell.2006.02.041 (2006).

172 Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837, doi:10.1016/j.cell.2007.05.009 (2007).

173 Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics* **39**, 311-318, doi:10.1038/ng1966 (2007).

174 Jiang, H. *et al.* Regulation of transcription by the MLL2 complex and MLL complex-associated AKAP95. *Nature structural & molecular biology* **20**, 1156-1163, doi:10.1038/nsmb.2656 (2013).

175 Lee, J. H. & Skalnik, D. G. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *The Journal of biological chemistry* **280**, 41725-41731, doi:10.1074/jbc.M508312200 (2005).

176 Thomson, J. P. *et al.* CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082-1086, doi:10.1038/nature08924 (2010).

177 Clouaire, T. *et al.* Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes & development* **26**, 1714-1728, doi:10.1101/gad.194209.112 (2012).

178 Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular cell* **30**, 755-766, doi:10.1016/j.molcel.2008.05.007 (2008).

179 Martens, J. H. *et al.* The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *The EMBO journal* **24**, 800-812, doi:10.1038/sj.emboj.7600545 (2005).

180 Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447**, 407-412, doi:10.1038/nature05915 (2007).

181 Daniel, J. A., Pray-Grant, M. G. & Grant, P. A. Effector proteins for methylated histones: an expanding family. *Cell cycle* **4**, 919-926 (2005).

182 LeRoy, G., Rickards, B. & Flint, S. J. The double bromodomain proteins Brd2 and Brd3 couple histone acetylation to transcription. *Molecular cell* **30**, 51-60, doi:10.1016/j.molcel.2008.01.018 (2008).

183 Shao, Z. *et al.* Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**, 37-46, doi:10.1016/S0092-8674(00)80604-2 (1999).

184 Cao, R. & Zhang, Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Molecular cell* **15**, 57-67, doi:10.1016/j.molcel.2004.06.020 (2004).

185 Cao, Q. *et al.* The central role of EED in the orchestration of polycomb group complexes. *Nature communications* **5**, 3127, doi:10.1038/ncomms4127 (2014).

186 Vire, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871-874, doi:10.1038/nature04431 (2006).

187 Collins, R. T., Furukawa, T., Tanese, N. & Treisman, J. E. Osa associates with the Brahma chromatin remodeling complex and promotes the activation of some target genes. *The EMBO journal* **18**, 7029-7040, doi:10.1093/emboj/18.24.7029 (1999).

188 Petruk, S. *et al.* Trithorax and dCBP acting in a complex to maintain expression of a homeotic gene. *Science* **294**, 1331-1334, doi:10.1126/science.1065683 (2001).

189 Wan, M. *et al.* The trithorax group protein Ash2l is essential for pluripotency and maintaining open chromatin in embryonic stem cells. *The Journal of biological chemistry* **288**, 5039-5048, doi:10.1074/jbc.M112.424515 (2013).

190 Veazey, K. J., Muller, D. & Golding, M. C. Prenatal alcohol exposure and cellular differentiation: a role for Polycomb and Trithorax group proteins in FAS phenotypes? *Alcohol research : current reviews* **35**, 77-85 (2013).

191 Bruce Alberts, A. J., Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell*. 5th edn, (Garland Science, 2007).

192 Ecker, J. R. & Davis, R. W. Inhibition of gene expression in plant cells by expression of antisense RNA. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 5372-5376 (1986).

193 Napoli, C., Lemieux, C. & Jorgensen, R. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *The Plant cell* **2**, 279-289, doi:10.1105/tpc.2.4.279 (1990).

194 Ratcliff, F., Harrison, B. D. & Baulcombe, D. C. A similarity between viral defense and gene silencing in plants. *Science* **276**, 1558-1560 (1997).

195 Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811, doi:10.1038/35888 (1998).

196 Bagasra, O. & Prilliman, K. R. RNA interference: the molecular immune system. *Journal of molecular histology* **35**, 545-553, doi:10.1007/s10735-004-2192-8 (2004).

197 Shabalina, S. A. & Koonin, E. V. Origins and evolution of eukaryotic RNA interference. *Trends in ecology & evolution* **23**, 578-587, doi:10.1016/j.tree.2008.06.005 (2008).

198 Check, E. RNA interference: hitting the on switch. *Nature* **448**, 855-858, doi:10.1038/448855a (2007).

199 Vasudevan, S., Tong, Y. & Steitz, J. A. Switching from repression to activation: microRNAs can up-regulate translation. *Science* **318**, 1931-1934, doi:10.1126/science.1149460 (2007).

200 Perkel, J. M. Visiting "noncodarnia". *BioTechniques* **54**, 301, 303-304, doi:10.2144/000114037 (2013).

201 Yang, L. *et al.* ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* **147**, 773-788, doi:10.1016/j.cell.2011.08.054 (2011).

202 Froberg, J. E., Yang, L. & Lee, J. T. Guided by RNAs: X-inactivation as a model for lncRNA function. *Journal of molecular biology* **425**, 3698-3706, doi:10.1016/j.jmb.2013.06.031 (2013).

203 Gendrel, A. V. & Heard, E. Noncoding RNAs and Epigenetic Mechanisms During X-Chromosome Inactivation. *Annual review of cell and developmental biology*, doi:10.1146/annurev-cellbio-101512-122415 (2014).

204 Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323, doi:10.1016/j.cell.2007.05.022 (2007).

205 Di Ruscio, A. *et al.* DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**, 371-376, doi:10.1038/nature12598 (2013).

206 Sun, S. *et al.* Jpx RNA activates Xist by evicting CTCF. *Cell* **153**, 1537-1551, doi:10.1016/j.cell.2013.05.028 (2013).

207 Zhang, H. *et al.* Long noncoding RNA-mediated intrachromosomal interactions promote imprinting at the Kcnq1 locus. *The Journal of cell biology* **204**, 61-75, doi:10.1083/jcb.201304152 (2014).

208 Seisenberger, S. *et al.* Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **368**, 20110330, doi:10.1098/rstb.2011.0330 (2013).

209 Rosner, M. H. *et al.* A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* **345**, 686-692, doi:10.1038/345686a0 (1990).

210 Yamaguchi, S., Kimura, H., Tada, M., Nakatsuji, N. & Tada, T. Nanog expression in mouse germ cell development. *Gene expression patterns : GEP* **5**, 639-646, doi:10.1016/j.modgep.2005.03.001 (2005).

211 Ginsburg, M., Snow, M. H. & McLaren, A. Primordial germ cells in the mouse embryo during gastrulation. *Development* **110**, 521-528 (1990).

212 Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mechanisms of development* **117**, 15-23 (2002).

213 Seki, Y. *et al.* Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Developmental biology* **278**, 440-458, doi:10.1016/j.ydbio.2004.11.025 (2005).

214 Morgan, H. D., Dean, W., Coker, H. A., Reik, W. & Petersen-Mahrt, S. K. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *The Journal of biological chemistry* **279**, 52353-52360, doi:10.1074/jbc.M407695200 (2004).

215 Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042-1047, doi:10.1038/nature08752 (2010).

216 Yamaguchi, S. *et al.* Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell research* **23**, 329-339, doi:10.1038/cr.2013.22 (2013).

217 Lane, N. *et al.* Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* **35**, 88-93, doi:10.1002/gene.10168 (2003).

218 Guibert, S., Forne, T. & Weber, M. Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome research* **22**, 633-641, doi:10.1101/gr.130997.111 (2012).

219 Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877-881, doi:10.1038/nature06714 (2008).

220 Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96-99, doi:10.1038/nature02886 (2004).

221 Kaneda, M. *et al.* Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900-903, doi:10.1038/nature02633 (2004).

222 Ooi, S. K. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714-717, doi:10.1038/nature05987 (2007).

223 Smallwood, S. A. *et al.* Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature genetics* **43**, 811-814, doi:10.1038/ng.864 (2011).

224 Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339-344, doi:10.1038/nature10960 (2012).

225 Farthing, C. R. *et al.* Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS genetics* **4**, e1000116, doi:10.1371/journal.pgen.1000116 (2008).

226 Wossidlo, M. *et al.* 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature communications* **2**, 241, doi:10.1038/ncomms1240 (2011).

227 Okada, Y., Yamagata, K., Hong, K., Wakayama, T. & Zhang, Y. A role for the elongator complex in zygotic paternal genome demethylation. *Nature* **463**, 554-558, doi:10.1038/nature08732 (2010).

228 Inoue, A. & Zhang, Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334**, 194, doi:10.1126/science.1212483 (2011).

229 Inoue, A., Shen, L., Dai, Q., He, C. & Zhang, Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell research* **21**, 1670-1676, doi:10.1038/cr.2011.189 (2011).

230 Howlett, S. K. & Reik, W. Methylation levels of maternal and paternal genomes during preimplantation development. *Development* **113**, 119-127 (1991).

231 Nakamura, T. *et al.* PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nature cell biology* **9**, 64-71, doi:10.1038/ncb1519 (2007).

232 Nakamura, T. *et al.* PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* **486**, 415-419, doi:10.1038/nature11093 (2012).

233 Howell, C. Y. *et al.* Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell* **104**, 829-838 (2001).

234 Kafri, T. *et al.* Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes & development* **6**, 705-714 (1992).

235 Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196-199, doi:10.1038/nature05972 (2007).

236 Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nature genetics* **42**, 1093-1100, doi:10.1038/ng.708 (2010).

237 Lee, K. *et al.* Dynamics of TET family expression in porcine preimplantation embryos is related to zygotic genome activation and required for the maintenance of NANOG. *Developmental biology* **386**, 86-95, doi:10.1016/j.ydbio.2013.11.024 (2014).

238 Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398-402, doi:10.1038/nature10008 (2011).

239 Udali, S., Guarini, P., Moruzzi, S., Choi, S. W. & Friso, S. Cardiovascular epigenetics: from DNA methylation to microRNAs. *Molecular aspects of medicine* **34**, 883-901, doi:10.1016/j.mam.2012.08.001 (2013).

240 Aslibekyan, S., Claas, S. A. & Arnett, D. K. Clinical applications of epigenetics in cardiovascular disease: the long road ahead. *Translational research : the journal of laboratory and clinical medicine*, doi:10.1016/j.trsl.2014.04.004 (2014).

241 Barreiro, E. & Gea, J. Epigenetics and muscle dysfunction in chronic obstructive pulmonary disease. *Translational research : the journal of laboratory and clinical medicine*, doi:10.1016/j.trsl.2014.04.006 (2014).

242 Pena, C. J., Bagot, R. C., Labonte, B. & Nestler, E. J. Epigenetic Signaling in Psychiatric Disorders. *Journal of molecular biology*, doi:10.1016/j.jmb.2014.03.016 (2014).

243 Lu, H., Liu, X., Deng, Y. & Qing, H. DNA methylation, a hand behind neurodegenerative diseases. *Frontiers in aging neuroscience* **5**, 85, doi:10.3389/fnagi.2013.00085 (2013).

244 Cacabelos, R. & Torrellas, C. Epigenetic drug discovery for Alzheimer's disease. *Expert opinion on drug discovery*, 1-28, doi:10.1517/17460441.2014.930124 (2014).

245 Conway O'Brien, E., Prideaux, S. & Chevassut, T. The epigenetic landscape of acute myeloid leukemia. *Advances in hematology* **2014**, 103175, doi:10.1155/2014/103175 (2014).

246 Easwaran, H., Tsai, H. C. & Baylin, S. B. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Molecular cell* **54**, 716-727, doi:10.1016/j.molcel.2014.05.015 (2014).

247 Kang, C., Song, J. J., Lee, J. & Kim, M. Y. Epigenetics: an emerging player in gastric cancer. *World journal of gastroenterology : WJG* **20**, 6433-6447, doi:10.3748/wjg.v20.i21.6433 (2014).

248 Chik, F., Szyf, M. & Rabbani, S. A. Role of epigenetics in cancer initiation and progression. *Advances in experimental medicine and biology* **720**, 91-104, doi:10.1007/978-1-4614-0254-1_8 (2011).

249 Dong, X., Liu, R. Y. & Chen, W. D. Correlation of Promoter Methylation in the MGMT Gene with Glioma Risk and Prognosis: a Meta-Analysis. *Molecular neurobiology*, doi:10.1007/s12035-014-8760-3 (2014).

250 Petronis, A. *et al.* Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance? *Schizophrenia bulletin* **29**, 169-178 (2003).

251 Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10604-10609, doi:10.1073/pnas.0500398102 (2005).

252 Wong, A. H., Gottesman, II & Petronis, A. Phenotypic differences in genetically identical organisms: the epigenetic perspective. *Human molecular genetics* **14 Spec No 1**, R11-18, doi:10.1093/hmg/ddi116 (2005).

253 Poulsen, P., Esteller, M., Vaag, A. & Fraga, M. F. The epigenetic basis of twin discordance in age-related diseases. *Pediatric research* **61**, 38R-42R, doi:10.1203/pdr.0b013e31803c7b98 (2007).

254 Painter, R. C. *et al.* Maternal nutrition during gestation and carotid arterial compliance in the adult offspring: the Dutch famine birth cohort. *Journal of hypertension* **25**, 533-540, doi:10.1097/HJH.0b013e328012135b (2007).

255 Feil, R. & Fraga, M. F. Epigenetics and the environment: emerging patterns and implications. *Nature reviews. Genetics* **13**, 97-109, doi:10.1038/nrg3142 (2011).

256 Dominguez-Salas, P. *et al.* Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. *Nature communications* **5**, 3746, doi:10.1038/ncomms4746 (2014).

257 Rakyan, V. K., Blewitt, M. E., Druker, R., Preis, J. I. & Whitelaw, E. Metastable epialleles in mammals. *Trends in genetics : TIG* **18**, 348-351 (2002).

258 Roy-Engel, A. M. & Belancio, V. P. in *eLS* (John Wiley & Sons, Ltd, 2001).

259 Dodge, J. E., Ramsahoye, B. H., Wo, Z. G., Okano, M. & Li, E. De novo methylation of MMLV provirus in embryonic stem cells: CpG versus non-CpG methylation. *Gene* **289**, 41-48 (2002).

260 Kerkel, K. *et al.* Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nature genetics* **40**, 904-908, doi:10.1038/ng.174 (2008).

261 Schilling, E., El Chartouni, C. & Rehli, M. Allele-specific DNA methylation in mouse strains is mainly determined by cis-acting sequences. *Genome research* **19**, 2028-2035, doi:10.1101/gr.095562.109 (2009).

262 Lienert, F. *et al.* Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. *PLoS genetics* **7**, e1002090, doi:10.1371/journal.pgen.1002090 (2011).

263 Johnsson, P. *et al.* A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nature structural & molecular biology* **20**, 440-446, doi:10.1038/nsmb.2516 (2013).

264 Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J.A., Struhl, K. *Current Protocols in Molecular Biology*. (John Wiley & Sons, 1987).

265 Green, M. R., Sambrook, J. *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, 2012).

266 Bibel, M. *et al.* Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nature neuroscience* **7**, 1003-1009, doi:10.1038/nn1301 (2004).

267 BAC transgenesis protocol v2.9. (SKI Stem Cell Research Facility).

268 Mullis, K. *et al.* Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor symposia on quantitative biology* **51 Pt 1**, 263-273 (1986).

269 Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 1827-1831 (1992).

270 Ehrich, M. *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15785-15790, doi:10.1073/pnas.0507816102 (2005).

271 Masouleh, A. K., Waters, D. L., Reinke, R. F. & Henry, R. J. A high-throughput assay for rapid and simultaneous analysis of perfect markers for important quality and agronomic traits in rice using multiplexed MALDI-TOF mass spectrometry. *Plant biotechnology journal* **7**, 355-363, doi:10.1111/j.1467-7652.2009.00411.x (2009).

272 Sauer, B. & Henderson, N. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 5166-5170 (1988).

273 *RefSeq Protein Database*, <ftp://ftp.ncbi.nlm.nih.gov/refseq/M_musculus/mRNA_Prot/>

274 Ouedraogo, M. *et al.* The duplicated genes database: identification and functional annotation of co-localised duplicated genes across genomes. *PloS one* **7**, e50653, doi:10.1371/journal.pone.0050653 (2012).

275 Keane, T. M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289-294, doi:10.1038/nature10413 (2011).

276 Harrington, C. T., Lin, E. I., Olson, M. T. & Eshleman, J. R. Fundamentals of pyrosequencing. *Archives of pathology & laboratory medicine* **137**, 1296-1303, doi:10.5858/arpa.2012-0463-RA (2013).

277 Monaghan, F. & Corcos, A. On the origins of the Mendelian laws. *The Journal of heredity* **75**, 67-69 (1984).

278 Bateson, W., Saunders, E. R., Punnett, R. C. Experimental studies in the physiology of heredity. *Reports to the Evolution Committee of the Royal Society* **2**, 1-55, 80-99 (1905).

279 Morgan, T. H., Cattell, E. Data for the study of sex-linked inheritance in Drosophila. *J. Exp. Zool.* **13**, 79-101 (1912).

280 Reich, D. E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411**, 199-204, doi:10.1038/35075590 (2001).

281 Sturtevant, A. H. The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. *J. Exper. Zool.* **14**, 43-59 (1913).

282 Xia, Y. *et al.* Bulk segregation mapping of mutations in closely related strains of mice. *Genetics* **186**, 1139-1146, doi:10.1534/genetics.110.121160 (2010).

283 Dietrich, W. F. *et al.* A comprehensive genetic map of the mouse genome. *Nature* **380**, 149-152, doi:10.1038/380149a0 (1996).

284 Jensen-Seaman, M. I. *et al.* Comparative recombination rates in the rat, mouse, and human genomes. *Genome research* **14**, 528-538, doi:10.1101/gr.1970304 (2004).

285 Engler, P. *et al.* A strain-specific modifier on mouse chromosome 4 controls the methylation of independent transgene loci. *Cell* **65**, 939-947 (1991).

286 Valenza-Schaerly, P. *et al.* A dominant modifier of transgene methylation is mapped by QTL analysis to mouse chromosome 13. *Genome research* **11**, 382-388, doi:10.1101/gr.163801 (2001).

287 Blewitt, M. E. *et al.* An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 7629-7634, doi:10.1073/pnas.0409375102 (2005).

288 Shibahara, S. *et al.* A point mutation in the tyrosinase gene of BALB/c albino mouse causing the cysteine----serine substitution at position 85. *European journal of biochemistry / FEBS* **189**, 455-461 (1990).

289 Wolff, G. L., Kodell, R. L., Moore, S. R. & Cooney, C. A. Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **12**, 949-957 (1998).

290 Nagy A., G. M., Vintersten K., Behringer R. *Manipulating the Mouse Embryo, A Laboratory Manual*. 3rd edn, (Cold Spring Harbour Laboratory Press, 2003).

291 Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156 (1981).

292 Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 7634-7638 (1981).

293 Brook, F. A. & Gardner, R. L. The origin and efficient derivation of embryonic stem cells in the mouse. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 5709-5712 (1997).

294 Weng, A., Magnuson, T. & Storb, U. Strain-specific transgene methylation occurs early in mouse development and can be recapitulated in embryonic stem cells. *Development* **121**, 2853-2859 (1995).

295 Evsikov, A. V. & Solter, D. Comment on " 'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". *Science* **302**, 393; author reply 393, doi:10.1126/science.1082380 (2003).

296 Mansergh, F. C. *et al.* Gene expression profiles during early differentiation of mouse embryonic stem cells. *BMC developmental biology* **9**, 5, doi:10.1186/1471-213X-9-5 (2009).

297 Maccarrone, L. *Identification and characterization for trans-regulated strain-specific methylated regions in Mus Musculus* Master of Science thesis, University Regensburg, (2014).

298 Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PloS one* **5**, e15367, doi:10.1371/journal.pone.0015367 (2010).

299 Szwagierczak, A., Bultmann, S., Schmidt, C. S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic acids research* **38**, e181, doi:10.1093/nar/gkq684 (2010).

300 Nestor, C., Ruzov, A., Meehan, R. & Dunican, D. Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. *BioTechniques* **48**, 317-319, doi:10.2144/000113403 (2010).

301 Marshall, O. J. PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics* **20**, 2471-2472, doi:10.1093/bioinformatics/bth254 (2004).

302 Davies, J. & Jimenez, A. A new selective agent for eukaryotic cloning vectors. *The American journal of tropical medicine and hygiene* **29**, 1089-1092 (1980).

303 Staub, E., Fiziev, P., Rosenthal, A. & Hinzmann, B. Insights into the evolution of the nucleolus by an analysis of its protein domain repertoire. *BioEssays : news and reviews in molecular, cellular and developmental biology* **26**, 567-581, doi:10.1002/bies.20032 (2004).

304 Li, D., Roberts, R. WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cellular and Molecular Life Sciences* **58**, 2085-2097 (2001).

305 Smith, T. F., Gaitatzes, C., Saxena, K. & Neer, E. J. The WD repeat: a common architecture for diverse functions. *Trends in biochemical sciences* **24**, 181-185 (1999).

306 Frescas, D. *et al.* KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. *Cell cycle* **7**, 3539-3547 (2008).

307 Kuramochi-Miyagawa, S. *et al.* DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes & development* **22**, 908-917, doi:10.1101/gad.1640708 (2008).

308 Henikoff, S. Position-effect variegation after 60 years. *Trends in genetics : TIG* **6**, 422-426 (1990).

309 Schotta, G., Ebert, A., Dorn, R. & Reuter, G. Position-effect variegation and the genetic dissection of chromatin regulation in Drosophila. *Seminars in cell & developmental biology* **14**, 67-75 (2003).

310 Mueller, P. R. & Wold, B. In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* **246**, 780-786 (1989).

311 Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-1311, doi:10.1126/science.1067799 (2002).

312 Schuettengruber, B. & Cavalli, G. Polycomb domain formation depends on short and long distance regulatory cues. *PloS one* **8**, e56531, doi:10.1371/journal.pone.0056531 (2013).

313 Pokholok, D. K. *et al.* Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517-527, doi:10.1016/j.cell.2005.06.026 (2005).

314 Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707-719, doi:10.1016/j.cell.2007.01.015 (2007).

315 Laird, P. W. Principles and challenges of genomewide DNA methylation analysis. *Nature reviews. Genetics* **11**, 191-203, doi:10.1038/nrg2732 (2010).

316 Krueger, F., Kreck, B., Franke, A. & Andrews, S. R. DNA methylome analysis using short bisulfite sequencing data. *Nature methods* **9**, 145-151, doi:10.1038/nmeth.1828 (2012).

317 Coolen, M. W., Statham, A. L., Gardiner-Garden, M. & Clark, S. J. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. *Nucleic acids research* **35**, e119, doi:10.1093/nar/gkm662 (2007).

318 Taylor, D. B. Genetics of screwworm: new genetic markers and preliminary linkage map. *The Journal of heredity* **80**, 425-432 (1989).

319 Ball, A. D. *et al.* A comparison of SNPs and microsatellites as linkage mapping markers: lessons from the zebra finch (Taeniopygia guttata). *BMC genomics* **11**, 218, doi:10.1186/1471-2164-11-218 (2010).

320 Simon, M. M. *et al.* A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. *Genome biology* **14**, R82, doi:10.1186/gb-2013-14-7-r82 (2013).

321 Weichman, K. & Chaillet, J. R. Phenotypic variation in a genetically identical population of mice. *Molecular and cellular biology* **17**, 5269-5274 (1997).

322 Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929-930, doi:10.1126/science.1169786 (2009).

323 Szulwach, K. E. *et al.* Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS genetics* **7**, e1002154, doi:10.1371/journal.pgen.1002154 (2011).

324 Wu, H. *et al.* Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes & development* **25**, 679-684, doi:10.1101/gad.2036011 (2011).

325 Padjen, K., Ratnam, S. & Storb, U. DNA methylation precedes chromatin modifications under the influence of the strain-specific modifier Ssm1. *Molecular and cellular biology* **25**, 4782-4791, doi:10.1128/MCB.25.11.4782-4791.2005 (2005).

326 Krebs, C. J., Schultz, D. C. & Robins, D. M. The KRAB zinc finger protein RSL1 regulates sex- and tissue-specific promoter methylation and dynamic hormone-responsive chromatin configuration. *Molecular and cellular biology* **32**, 3732-3742, doi:10.1128/MCB.00615-12 (2012).

327 Cheng, Y. *et al.* KRAB zinc finger protein ZNF382 is a proapoptotic tumor suppressor that represses multiple oncogenes and is commonly silenced in multiple carcinomas. *Cancer research* **70**, 6516-6526, doi:10.1158/0008-5472.CAN-09-4566 (2010).

328 Gebelein, B., Fernandez-Zapico, M., Imoto, M. & Urrutia, R. KRAB-independent suppression of neoplastic cell growth by the novel zinc finger transcription factor KS1. *The Journal of clinical investigation* **102**, 1911-1919, doi:10.1172/JCI1919 (1998).

329 Gebelein, B. & Urrutia, R. Sequence-specific transcriptional repression by KS1, a multiple-zinc-finger-Kruppel-associated box protein. *Molecular and cellular biology* **21**, 928-939, doi:10.1128/MCB.21.3.928-939.2001 (2001).

330 Zuo, X. *et al.* Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. *The Journal of biological chemistry* **287**, 2107-2118, doi:10.1074/jbc.M111.322644 (2012).

331 Juarez-Moreno, K. *et al.* Breaking through an epigenetic wall: re-activation of Oct4 by KRAB-containing designer zinc finger transcription factors. *Epigenetics : official journal of the DNA Methylation Society* **8**, 164-176, doi:10.4161/epi.23503 (2013).

332 Frietze, S., Lan, X., Jin, V. X. & Farnham, P. J. Genomic targets of the KRAB and SCAN domain-containing zinc finger protein 263. *The Journal of biological chemistry* **285**, 1393-1403, doi:10.1074/jbc.M109.063032 (2010).

333 Williams, A. J., Blacklow, S. C. & Collins, T. The zinc finger-associated SCAN box is a conserved oligomerization domain. *Molecular and cellular biology* **19**, 8526-8535 (1999).

334 Yu, H. B., Kunarso, G., Hong, F. H. & Stanton, L. W. Zfp206, Oct4, and Sox2 are integrated components of a transcriptional regulatory network in embryonic stem cells. *The Journal of biological chemistry* **284**, 31327-31335, doi:10.1074/jbc.M109.016162 (2009).

335 Farcas, A. M. *et al.* KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife* **1**, e00205, doi:10.7554/eLife.00205 (2012).

336 Tanaka, Y. *et al.* JmjC enzyme KDM2A is a regulator of rRNA transcription in response to starvation. *The EMBO journal* **29**, 1510-1522, doi:10.1038/emboj.2010.56 (2010).

337 Blackledge, N. P., Thomson, J. P. & Skene, P. J. CpG island chromatin is shaped by recruitment of ZF-CxxC proteins. *Cold Spring Harbor perspectives in biology* **5**, a018648, doi:10.1101/cshperspect.a018648 (2013).

338 Zhou, J. C., Blackledge, N. P., Farcas, A. M. & Klose, R. J. Recognition of CpG island chromatin by KDM2A requires direct and specific interaction with linker DNA. *Molecular and cellular biology* **32**, 479-489, doi:10.1128/MCB.06332-11 (2012).

339 Sleutels, F., Zwart, R. & Barlow, D. P. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810-813, doi:10.1038/415810a (2002).

340 Nagano, T. *et al.* The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717-1720, doi:10.1126/science.1163802 (2008).

341 Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689-693, doi:10.1126/science.1192002 (2010).

342 Hawkins, P. G. & Morris, K. V. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription* **1**, 165-175, doi:10.4161/trns.1.3.13332 (2010).

343 Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchev, A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**, 666-670, doi:10.1038/nature05519 (2007).

344 Watanabe, T. *et al.* Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. *Science* **332**, 848-852, doi:10.1126/science.1203919 (2011).

345 Schmitz, K. M., Mayer, C., Postepska, A. & Grummt, I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes & development* **24**, 2264-2269, doi:10.1101/gad.590910 (2010).

346 Tian, D., Sun, S. & Lee, J. T. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* **143**, 390-403, doi:10.1016/j.cell.2010.09.049 (2010).

347 Holz-Schietinger, C. & Reich, N. O. RNA modulation of the human DNA methyltransferase 3A. *Nucleic acids research* **40**, 8550-8557, doi:10.1093/nar/gks537 (2012).

348 Kanno, T. *et al.* A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nature genetics* **40**, 670-675, doi:10.1038/ng.119 (2008).

349 Gendrel, A. V. *et al.* Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Developmental cell* **23**, 265-279, doi:10.1016/j.devcel.2012.06.011 (2012).

350 Neri, F. *et al.* Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. *Cell* **155**, 121-134, doi:10.1016/j.cell.2013.08.056 (2013).

351 Pacaud, R. *et al.* DNMT3L interacts with transcription factors to target DNMT3L/DNMT3B to specific DNA sequences: Role of the DNMT3L/DNMT3B/p65-NFkappaB complex in the (de-)methylation of TRAF1. *Biochimie* **104**, 36-49, doi:10.1016/j.biochi.2014.05.005 (2014).

352 Schultz, D. C., Friedman, J. R. & Rauscher, F. J., 3rd. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes & development* **15**, 428-443, doi:10.1101/gad.869501 (2001).

353 Urrutia, R. KRAB-containing zinc-finger repressor proteins. *Genome biology* **4**, 231, doi:10.1186/gb-2003-4-10-231 (2003).

354 Sripathy, S. P., Stevens, J. & Schultz, D. C. The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Molecular and cellular biology* **26**, 8623-8638, doi:10.1128/MCB.00487-06 (2006).

355 Groner, A. C. *et al.* KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS genetics* **6**, e1000869, doi:10.1371/journal.pgen.1000869 (2010).

356 Quenneville, S. *et al.* The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. *Cell reports* **2**, 766-773, doi:10.1016/j.celrep.2012.08.043 (2012).

357 Bernstein, E. & Allis, C. D. RNA meets chromatin. *Genes & development* **19**, 1635-1655, doi:10.1101/gad.1324305 (2005).

358 Brockdorff, N. *et al.* The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515-526 (1992).

359 Herzing, L. B., Romer, J. T., Horn, J. M. & Ashworth, A. Xist has properties of the X-chromosome inactivation centre. *Nature* **386**, 272-275, doi:10.1038/386272a0 (1997).

360 Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Molecular cell* **5**, 695-705 (2000).

361 Cai, B., Song, X. Q., Cai, J. P. & Zhang, S. HOTAIR: acancer-related long non-coding RNA. *Neoplasma* **61**, 379-391 (2014).

362 Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295-300, doi:10.1038/nature10398 (2011).

363 Blackledge, N. P. *et al.* CpG islands recruit a histone H3 lysine 36 demethylase. *Molecular cell* **38**, 179-190, doi:10.1016/j.molcel.2010.04.009 (2010).

364 Oey H., H. S., Bourke L. Whitelaw E. (35th Lorne Genome Conference, 2014).

365 Di Croce, L. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nature structural & molecular biology* **20**, 1147-1155, doi:10.1038/nsmb.2669 (2013).

366 Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447**, 425-432, doi:10.1038/nature05918 (2007).

367 Kaneko, S., Son, J., Shen, S. S., Reinberg, D. & Bonasio, R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nature structural & molecular biology* **20**, 1258-1264, doi:10.1038/nsmb.2700 (2013).

9 Abbreviations

%	Percent
°C	Degree Celcius
µg	Microgram
µl	Microliter
µM	Micromolar
3‘	3-prime
5‘	5-prime
53BP1	P53-binding protein 1
5caC	5-carboxycytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
A	Adenine
aa	Amino acids
ac	Acetylation
AcOH	Acetic Acid
Amp	Ampicillin
ATP	Adenosine triphosphate
A^{vy}	Agouti viable yellow allele
B6	C57BL/6
B6CF1	C57BL/6 (f) x BALB/c (m)
BAC	Bacterial Artificial Chromosome
BAH	Bromo-adjacent homology domain
BER	Base excision repair
BMM	Bone marrow-derived macrophages
bp	Base pair
BS	Bisulfite sequencing
BSA	Bovine Serum Albumin
C	Cytosine
CB6F1	BALB/c (f) x C57BL/6 (m)
CBM	Cell Buffer Mix
C. briggsea	Caenorhabditis briggsea
cDNA	Complementary DNA
CFP1	CXXC zinc finger protein 1
CGI	CpG island
ChIP	Chromatin immunoprecipitation
Chr	Chromosome
cm	Centimeter
cM	Centimorgan
CM	Chloramphenicol

cm^2	Square centimeter
CNODF1	BALB/c (f) x NOD (m)
CO_2	Carbondioxide
CpG	CpG dinucleotide
CXXC	Cysteine-rich zinc finger domain
Da	Dalton
dATP	Desoxyadenosin triphosphate
DB	Dilution Buffer
ddH_2O	Double-destilled H2O
ddNTP	Dideoxyribonucleotide triphosphate
DER	Differentially expressed regions
DIG	Digoxigenin
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	Deoxyribonucleotide triphosphate
dpc	day *post coitus*
DTT	Dithiothreitol
dUTP	Deoxyuridine triphosphate
E	Embryonic day
E. coli	Escherichia coli
EB	Embryoid body
EDTA/Na_2EDTA	Ethylenediaminotetraacetic acid disodium salt dehydrate
Eed	Enhancer of embryonic development
eGFP	Enhanced green fluorescent protein
NEU	N-ethyl-N-nitrosourea
ESC / ES cell	Embryonic stem cell
EtBr	Ethidium bromide
EtOH	Ethanol
exRNA	Exosomal RNA
Ezh2	Enhancer of zeste homolog 2
FBS	Fetal bovine serum
G	Guanine
g	Gram
G418	Geneticin
Gapdh	Glyceraldehyde-3-phosphate dehydrogenase
GC	Guanin/Cytosine
gDNA	Genomic DNA
GR	Glycine-argenine repeat
H	Histone
h	Hour
HAT	Histone acetyltransferase
HCl	Hydrochloric acid
HCP	High-CpG promoter
HDAC	Histone deacetylase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hg18	Human genome assembly 18

hMeDIP	Hydroxymethylcytosine methylated DNA immunoprecipitation
HMT	Histone methyltransferase
Hz	Hertz
IAP	Intracisternal A-particle
ICM	Inner cell mass
IgG	Immunoglobulin G
IP	Immunoprecipitation
Isoc2a/b	Isochorismatase domain containing 2 a/b
K	Lysine
Kan	Kanamycin
kb	Kilo base
KCl	Potassium chloride
kDa	Kilodalton
KOH	Potassium hydroxide
KRAB	Krüppel-associated box domain
KZF	KRAB-containing zinc finger
l	Liter
LA	Linkage analysis
LB	Lysogeny broth
LIF	Leukemia inhibiting Factor
lncRNA	Long non-coding RNA
LOD	Logarithm 209oniza odds
LSD1	Lysine-specific demethylase 1
M	Molar
mA	Milliampere
MALDI-TOF MS	Matrix-assisted laser desorption/ionization coupled with mass spectrometry analysis by Time-of-fligth
Mb	Mega base
MBD	Methyl-binding domain
MBP	Methyl-CpG binding protein family
MCIp	Methyl-CpG immunoprecipitation
mCpG	Methylated CpG
me1	Monomethylation
me3	Trimethylation
MeCP	Methyl-CpG binding protein
MEF	Mouse embryonic fibroblast
mg	Milligram
MgCl2	Magnesium chloride
MHL	MutL homolog 1
min	Minute
Mio	Million
miRNA	Micro RNA
MIso	Modifier of *Isoc2b*
ml	Milliliter
mM	Millimolar
mm5/9/10	*Mus musculus* assembly 5/9/10
MPC	Magnetic particle concentrator

mRNA	Messenger RNA
msec	Millisecond
Myod1	Myogenic differention 1
NaCl	Sodium chloride
NaHCO3	Sodium hydrogen carbonate
NaOAc	Sodium acetate
nBAC	Retrofitted BAC in SW106
ncRNA	Non-coding RNA
ng	Nanogram
NGS	Next generation sequencing
NH_2	Amino
njBAC	BAC retrofitted in presence of pJM2545 plasmid in DH10B strain
NLS	Nuclear localization signal
nm	Nanometer
NOD	NOD/ShiLtJ
Nol10	Nucleolar protein 10
NP95/Uhrf1	Nuclear protein 95
nt	Nucleotide
NuRD	Nucleosome remodeling eacetylase
o/n	Overnight
oBAC	Original BAC in DH10B strain
ORF	Open reading frame
PAM	Partitioning Around Medoids
PBS	Phosphate-buffered saline
Pc	Polycomb
PcG	Polycomb-Group
PCR	Polymerase chain reaction
Pdgfrb	Platelet-derived growth factor receptor, beta polypeptide
PEG8000	Polyethylenglycol 8000
PFGE	Pulsed-field gel electrophoresis
PGC	Primordial germ cell
PHD	Pleckstrin-homology domain
piRNA	Piwi-interacting RNA
PMSF	Phenylmethanesulfonyl fluoride
Oct4	POU domain, class 5, transcription factor 1
POZ/BTB	Broad-Complex, Tramtrack and Bric a brac
PRC(1/2)	Polycomb repressor complex (1/2)
PWWP	Proline-tryptophan-tryptophan-proline motif
qPCR	Quantitative Real-time PCR
RT-qPCR	Reverse-transcription quantitative real-time PCR
RISC	RNA-induced silencing complex
Rn18s	18S ribosomal RNA
RNA	Ribonucleic acid
RNAi	RNA-mediated interferance
Rnase	Ribonuclase
rpm	Rounds per minute
rRNA	Ribosomal RNA

RT	Room temperature
SAM	S-adenosylmethionin
SAP	Shrimp alkaline phosphatase
sBAC	Original BAC in SW106 strain
SBE	Single base extension
SDS	Sodium dodecyl sulfate
sec	Second
SET	Su(var)3-9 and 'Enhancer of zeste' domain
SINE	Short interspresed elements
siRNA	Small interfering RNA
snoRNA	Small nucleolar RNA
SNP	Single nucleotide polymorphism
snRNA	Small nuclear RNA
Snrpn	small nuclear ribonucleoprotein N
T	Thymine
T_a	Annealing temperature
TAE	Tris/Acetate/EDTA
TE	Tris/EDTA
TET	Ten-eleven translocation
TF	Transcription factor
T_m	Melting temperature
TRD	Transcriptional repressor domain
Tris	Tris(hydroxymethyl)-aminomethan
tRNA	Transfer RNA
TRPE	component I of anthranilate synthase
TrxG	Trithorax-group
Tyr	Tyrosinase
U	Unit
ub	Ubiquitinylation
UCSC	University of California, Santa Cruz
UV	Ultra violet
V	Volt
Vol	Volume
Ω	Ohm
W	Watt
WBI-III	Wash buffer I-III
w/o	Without
WDR5	WD repeat domain 5
WT	Wild type
xg	G-force acceleration
X-gal	5-Brom-4-chlor-3-indoxyl-β-D-galactopyranosid
ZF	Zinc finger

10 Appendix

<u>Appendix I</u> – Genome-wide linkage analysis in C;B6 hybrids (section 10.1; page 214)

Table 10-1 lists the genotyping results for the genome-wide SNP panel for 45 C;B6 hybrids as well as their average methylation at the *Isoc2b* promoter (*mm9*_chr7:4818016-4818396). The homozygous genotype is indicated by black squares, the heterozygous genotype by white squares. The classification of the methylation phenotype as high or intermediate according to the separating level of 58.5% is indicated by color (black or white, respectively, to match the color of the concordant genotype; grey if genotype could not be called reliably). For each SNP, the number of concordant individuals (homozygous – high methylation, heterozygous – low methylation) in comparison to the total number of genotyped individuals determined the LOD score (see section 5.1.4.1). The candidate region potentially encoding the strain-specific epigenetic modifier *MIso* is highlighted by a green box.

The identity of the individual SNPs will be available upon publication.

<u>Appendix II</u> – Fine-mapping the candidate region on chromosome 12 (section 10.2; page 218)

Table 10-2 lists the genotyping results for the fine-mapping SNP panel for 33 informative (with recombination in candidate region) of 172 C;B6 hybrids as well as their average methylation at the *Isoc2b* promoter (*mm9*_chr7:4818016-4818396). The homozygous genotype is indicated by black squares, the heterozygous genotype by white squares. The classification of the methylation phenotype as high or intermediate according to the re-defined separating level of 61.5±3.5% is indicated by color (black or white, respectively, to match the color of the concordant genotype). For each SNP, the number of concordant individuals (homozygous – high methylation, heterozygous – low methylation) in comparison to the total number of genotyped individuals determines the LOD score (see section 5.1.4.1). The number of concordant, but non-informative individuals is also stated. The candidate region potentially encoding the strain-specific epigenetic modifier *MIso* is highlighted by a green box.

The identity of the individual SNPs will be available upon publication.

<u>Appendix III</u> – Strain-specific *Nol10* ORFs derived from Trinity–assemled transcripts (section 10.3; page 219)

Strain-specific open reading frames (ORFs) of the Nol10 transcript were analzyed with the ORF Finder tool using the combined sequences of the Trinity-assembled transcripts c591363_g14_i1/c591363_g3_i2 (JM8, day 0 and 4) and c423461_g20_i7/c422587_g2_i3 (BALB/c, day 0 to day 4), which rendered nearly full-length coverage of the *Nol10* coding sequence. The Trinity-derived transcripts were identified by Blastx (evalue <$1e^{-20}$). The annotated

Nol10 ORF is covered by both strains, but exhibits a non-synonymous SNP variant (rs29160165) in BALB/c causing an amino acid exchange (glutamic acid > lysine, highlighted) at position 532.

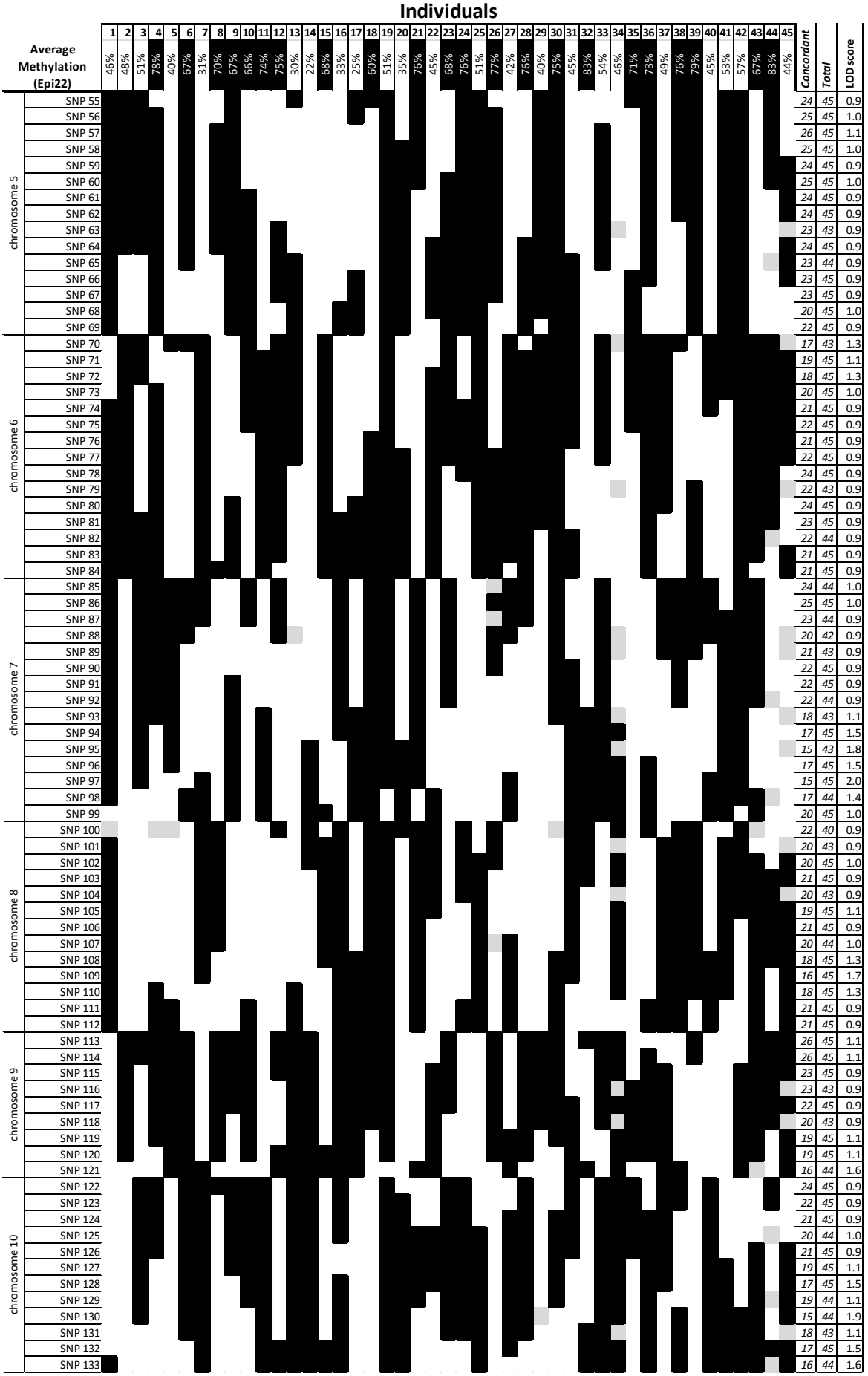
Appendix IV – DNA methylation anaylsis of novel candidate differentially methylated regions (DMRs) (section 10.4; page 221)

The novel candidate DMRs selected from regions strain-specifically marked by active histone marks (as determined by ChIP-seq experiments in spleen) are listed in Table 10-4.

Corresponding UCSC genome browser tracks are available at http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=juwelll&hgS_otherUserSessionName=C57vsBc_spleen_%20%09ChIP%2Dseq_H3K4me1_H3K4me3_H3K27ac_mm9.

10.1 Appendix I – Genome-wide linkage analysis in C;B6 hybrids

Table 10-1 ***Genotyping results of genome-wide linkage analysis with corresponding LOD scores***


Individuals
Average Methylation (Epi22)
1 46%
2 48%
3 51%
4 78%
5 40%
6 67%
7 31%
8 70%
9 67%
10 66%
11 74%
12 75%
13 30%
14 22%
15 68%
16 33%
17 25%
18 60%
19 51%
20 35%
21 76%
22 45%
23 68%
24 76%
25 51%
26 77%
27 42%
28 76%
29 40%
30 75%
31 45%
32 83%
33 54%
34 46%
35 71%
36 73%
37 49%
38 76%
39 79%
40 45%
41 53%
42 57%
43 67%
44 83%
45 44%
Concordant
Total
LOD score
chromosome 5
SNP 55 24 45 0.9
SNP 56 25 45 1.0
SNP 57 26 45 1.1
SNP 58 25 45 1.0
SNP 59 24 45 0.9
SNP 60 25 45 1.0
SNP 61 24 45 0.9
SNP 62 24 45 0.9
SNP 63 23 43 0.9
SNP 64 24 45 0.9
SNP 65 23 44 0.9
SNP 66 23 45 0.9
SNP 67 23 45 0.9
SNP 68 20 45 1.0
SNP 69 22 45 0.9
chromosome 6
SNP 70 17 43 1.3
SNP 71 19 45 1.1
SNP 72 18 45 1.3
SNP 73 20 45 1.0
SNP 74 21 45 0.9
SNP 75 22 45 0.9
SNP 76 21 45 0.9
SNP 77 22 45 0.9
SNP 78 24 45 0.9
SNP 79 22 43 0.9
SNP 80 24 45 0.9
SNP 81 23 45 0.9
SNP 82 22 44 0.9
SNP 83 21 45 0.9
SNP 84 21 45 0.9
chromosome 7
SNP 85 24 44 1.0
SNP 86 25 45 1.0
SNP 87 23 44 0.9
SNP 88 20 42 0.9
SNP 89 21 43 0.9
SNP 90 22 45 0.9
SNP 91 22 45 0.9
SNP 92 22 44 0.9
SNP 93 18 43 1.1
SNP 94 17 45 1.5
SNP 95 15 43 1.8
SNP 96 17 45 1.5
SNP 97 15 45 2.0
SNP 98 17 44 1.4
SNP 99 20 45 1.0
chromosome 8
SNP 100 22 40 0.9
SNP 101 20 43 0.9
SNP 102 20 45 1.0
SNP 103 21 45 0.9
SNP 104 20 43 0.9
SNP 105 19 45 1.1
SNP 106 21 45 0.9
SNP 107 20 44 1.0
SNP 108 18 45 1.3
SNP 109 16 45 1.7
SNP 110 18 45 1.3
SNP 111 21 45 0.9
SNP 112 21 45 0.9
chromosome 9
SNP 113 26 45 1.1
SNP 114 26 45 1.1
SNP 115 23 45 0.9
SNP 116 23 43 0.9
SNP 117 22 45 0.9
SNP 118 20 43 0.9
SNP 119 19 45 1.1
SNP 120 19 45 1.1
SNP 121 16 44 1.6
chromosome 10
SNP 122 24 45 0.9
SNP 123 22 45 0.9
SNP 124 21 45 0.9
SNP 125 20 44 1.0
SNP 126 21 45 0.9
SNP 127 19 45 1.1
SNP 128 17 45 1.5
SNP 129 19 44 1.1
SNP 130 15 44 1.9
SNP 131 18 43 1.1
SNP 132 17 45 1.5
SNP 133 16 44 1.6


Individuals
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
Average Methylation (Epi22)
46% 48% 51% 78% 40% 67% 31% 70% 67% 66% 74% 75% 30% 22% 68% 33% 25% 60% 51% 35% 76% 45% 68% 76% 51% 77% 42% 76% 40% 75% 45% 83% 54% 46% 71% 73% 49% 76% 79% 45% 53% 57% 67% 83% 44%
Concordant
Total
LOD score
chromosome 11
SNP 134 31 45 2.3
SNP 135 27 45 1.3
SNP 136 25 45 1.0
SNP 137 25 45 1.0
SNP 138 27 45 1.3
SNP 139 25 45 1.0
SNP 140 22 45 0.9
SNP 141 16 43 1.5
SNP 142 19 45 1.1
SNP 143 21 45 0.9
SNP 144 21 44 0.9
SNP 145 22 45 0.9
chromosome 12
rs29180275 38 44 6.4
rs29160274 45 45 13.5
rs4229294 43 45 10.6
SNP 149 36 41 6.5
SNP 150 36 45 4.6
SNP 151 34 45 3.5
SNP 152 32 45 2.7
SNP 153 29 45 1.7
chromosome 13
SNP 154 25 43 1.1
SNP 155 26 43 1.3
SNP 156 28 45 1.5
SNP 157 28 43 1.8
SNP 158 28 45 1.5
SNP 159 26 44 1.2
SNP 160 24 45 0.9
SNP 161 24 45 0.9
SNP 162 24 45 0.9
SNP 163 24 45 0.9
SNP 164 24 45 0.9
SNP 165 25 45 1.0
SNP 166 25 43 1.1
chromosome 14
SNP 167 23 45 0.9
SNP 168 24 45 0.9
SNP 169 24 44 1.0
SNP 170 24 45 0.9
SNP 171 24 45 0.9
SNP 172 27 45 1.3
SNP 173 23 44 0.9
SNP 174 24 43 1.0
SNP 175 23 43 0.9
SNP 176 25 45 1.0
SNP 177 23 43 0.9
SNP 178 26 44 1.2
SNP 179 27 45 1.3
chromosome 15
SNP 180 25 44 1.1
SNP 181 28 45 1.5
SNP 182 28 45 1.5
SNP 183 27 45 1.3
SNP 184 27 45 1.3
SNP 185 24 43 1.0
SNP 186 23 43 0.9
SNP 187 22 45 0.9
SNP 188 24 44 1.0
SNP 189 24 45 0.9
SNP 190 22 44 0.9
chromosome 16
SNP 191 19 45 1.1
SNP 192 20 45 1.0
SNP 193 20 45 1.0
SNP 194 22 45 0.9
SNP 195 22 44 0.9
SNP 196 23 44 0.9
SNP 197 24 43 1.0
SNP 198 28 45 1.5
SNP 199 30 44 2.2
chromosome 17
SNP 200 16 45 1.7
SNP 201 17 45 1.5
SNP 202 16 43 1.5
SNP 203 20 45 1.0
SNP 204 20 44 1.0
SNP 205 18 45 1.3
SNP 206 11 23 0.7
SNP 207 24 45 0.9
SNP 208 22 40 0.9
SNP 209 25 44 1.1


Individuals
Average Methylation (Epi22)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
46% 48% 51% 78% 40% 67% 31% 70% 67% 66% 74% 75% 30% 22% 68% 33% 25% 60% 51% 35% 76% 45% 68% 76% 51% 77% 42% 76% 40% 75% 45% 83% 54% 46% 71% 73% 49% 76% 79% 45% 53% 57% 67% 83% 44%
Concordant
Total
LOD score
chromosome 18
SNP 210 29 43 2.0
SNP 211 32 44 2.9
SNP 212 34 44 3.9
SNP 213 27 42 1.6
SNP 214 30 45 2.0
SNP 215 29 44 1.9
SNP 216 29 44 1.9
SNP 217 25 45 1.0
chromosome 19
SNP 218 18 45 1.3
SNP 219 9 18 0.6
SNP 220 16 45 1.7
SNP 221 14 45 2.3
SNP 222 16 45 1.7
SNP 223 18 45 1.3
SNP 224 18 44 1.2
chromosome X
SNP 225 18 44 1.2
SNP 226 21 45 0.9
SNP 227 19 44 1.1
SNP 228 21 45 0.9
SNP 229 20 44 1.0
SNP 230 22 45 0.9
SNP 231 21 44 0.9
SNP 232 21 44 0.9
SNP 233 21 45 0.9
SNP 234 18 44 1.2
SNP 235 20 44 1.0
SNP 236 18 44 1.2
SNP 237 17 43 1.3
SNP 238 18 44 1.2

10.2 Appendix II – Fine-mapping the candidate region on chromosome 12

Table 10-2 Genotyping results of the fine-mapping SNP panel for informative C;B6 hybrids

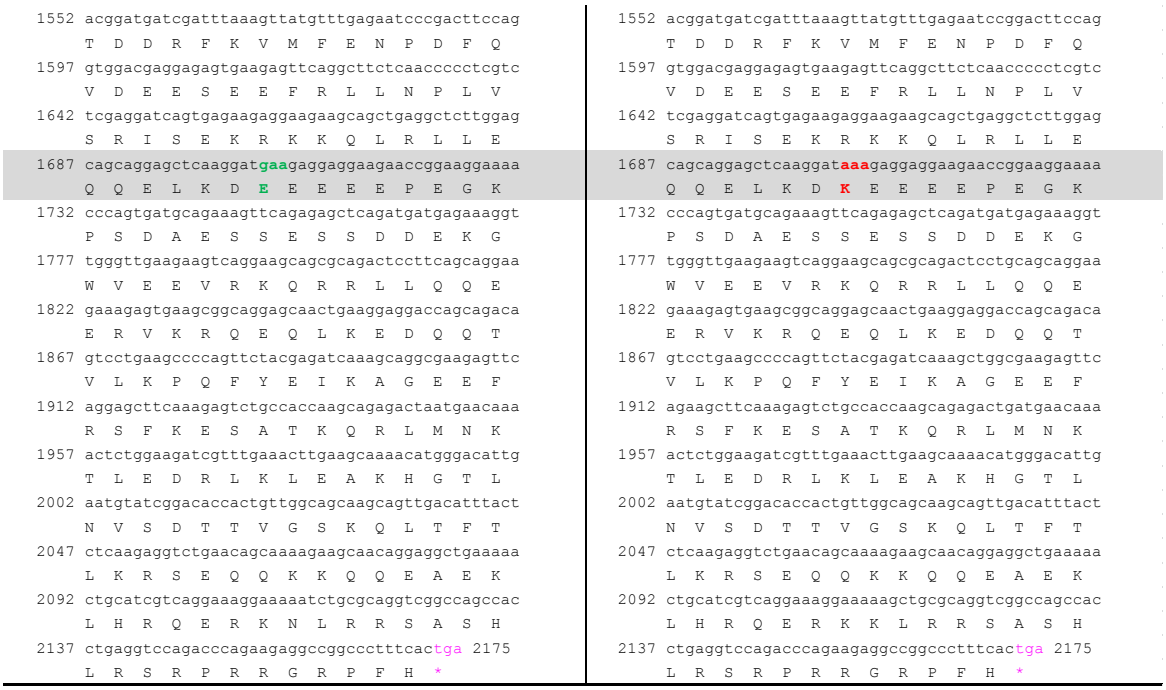
Individuals

		1	15	16	22	23	33	39	40	7	8	12	22	26	30	31	46	47	67	73	92	97	109	114	138	157	160	161	167	177	195	197	199	35	*Concordant, but w/o CO*	*Concordant*	*Total*	**LOD score**
	Average Methylation (Epi22)	46%	68%	33%	45%	68%	54%	79%	45%	55%	69%	48%	50%	55%	49%	84%	22%	81%	27%	52%	48%	68%	66%	74%	86%	49%	73%	73%	56%	41%	71%	49%	70%	56%				
chromosome 12	rs29180275																																		*137*	*144*	*172*	19.7
	SNP2																																		*137*	*144*	*172*	19.7
	SNP3																																		*137*	*144*	*172*	19.7
	SNP4																																		*137*	*144*	*172*	19.7
	SNP5																																		*137*	*149*	*172*	23.4
	SNP6																																		*137*	*155*	*172*	28.7
	SNP7																																		*137*	*156*	*172*	29.6
	SNP8																																		*137*	*156*	*172*	29.6
	SNP9																																		*137*	*157*	*172*	30.6
	SNP10																																		*137*	*158*	*172*	31.7
	SNP11																																		*137*	*158*	*172*	31.7
	SNP12																																		*137*	*160*	*172*	33.8
	SNP13																																		*137*	*161*	*172*	34.9
	SNP14																																		*137*	*161*	*172*	34.9
	SNP15																																		*137*	*161*	*172*	34.9
	SNP16																																		*137*	*167*	*172*	42.7
	SNP17																																		*137*	*168*	*172*	44.2
	rs29160274																																		*137*	*170*	*172*	**47.6**
	SNP19																																		*137*	*171*	*172*	**49.5**
	SNP20																																		*136*	*170*	*172*	**47.6**
	SNP21																																		*137*	*171*	*172*	**49.5**
	SNP22																																		*137*	*171*	*172*	**49.5**
	SNP23																																		*137*	*171*	*172*	**49.5**
	SNP24																																		*137*	*171*	*172*	**49.5**
	SNP25																																		*137*	*171*	*172*	**49.5**
	SNP26																																		*137*	*171*	*172*	**49.5**
	SNP27																																		*137*	*171*	*172*	**49.5**
	SNP28																																		*137*	*171*	*172*	**49.5**
	SNP29																																		*137*	*171*	*172*	**49.5**
	rs29188895																																		*137*	*170*	*172*	**47.6**
	SNP31																																		*137*	*170*	*172*	47.6
	rs4229294																																		*136*	*166*	*171*	42.4

10.3 Appendix III – Strain-specific *Nol10* ORFs derived from Trinity–assemled transcripts

Table 10-3 ***Strain-specific* Nol10 *ORFs***

Nol10 **ORF of JM8 (C57BL/6J) (687aa), derived from Trinity-assembled transcripts c591363_g14_i1 and c591363_g3_i2**	***Nol10*** **ORF of BALB/c (687aa), derived from Trinity-assembled transcripts c423461_g20_i7 and c422587_g2_i3**
112 atgcaggtctccagcctcaacgaagtgaagatttacagcctcagc	112 atgcaggtctccagcctcaacgaagtgaagatttacagcctcagc
M Q V S S L N E V K I Y S L S	M Q V S S L N E V K I Y S L S
157 tgcggcaagtctcttcccgagtggctttctgacaggaagaagaga	157 tgcggcaagtctcttcccgagtggctttctgacaggaagaagaga
C G K S L P E W L S D R K K R	C G K S L P E W L S D R K K R
202 gcactgcagaagaagaatgtcgatgttcgtaggagaattgaactt	202 gcactgcagaagaagaatgtcgatgttcgtaggagaattgaactt
A L Q K K N V D V R R R I E L	A L Q K K N V D V R R R I E L
247 attcaggacttcgaaatgcctactgtgtgtactactatcaaggtg	247 attcaggacttcgaaatgcctactgtgtgtactactatcaaggtg
I Q D F E M P T V C T T I K V	I Q D F E M P T V C T T I K V
292 tcaaaagatggacagtatattttagcaactggaacatataagcct	292 tcaaaagatggacagtatattttagcaactggaacatataagcct
S K D G Q Y I L A T G T Y K P	S K D G Q Y I L A T G T Y K P
337 cgggtccggtgttatgacacatatcagttatccctgaagtttgaa	337 cgggtccggtgttatgacacatatcagttatccctgaagtttgaa
R V R C Y D T Y Q L S L K F E	R V R C Y D T Y Q L S L K F E
382 aggtgcttggattctgaagttgtcacctttgagattttatcagat	382 aggtgcttggattctgaagttgtcacctttgagattttatcagat
R C L D S E V V T F E I L S D	R C L D S E V V T F E I L S D
427 gactattcaaagattgtcttcttacataatgacagatatattgaa	427 gactattcaaagattgtcttcttacataatgacagatatattgaa
D Y S K I V F L H N D R Y I E	D Y S K I V F L H N D R Y I E
472 tttcactcacaatcaggcttttactacaaaaccagaattcccaag	472 tttcactcacaatcaggcttttactacaaaaccagaattcccaag
F H S Q S G F Y Y K T R I P K	F H S Q S G F Y Y K T R I P K
517 ttcggccgagacttctcttaccactatccatcctgtgacttgtac	517 ttcggccgagacttctcttaccactatccatcctgtgacttgtac
F G R D F S Y H Y P S C D L Y	F G R D F S Y H Y P S C D L Y
562 tttgtaggtgcaagctctgaagtttatagattaaacttagaacaa	562 tttgtaggtgcaagctctgaagtttatagattaaacttagaacaa
F V G A S S E V Y R L N L E Q	F V G A S S E V Y R L N L E Q
607 gggcggtacctgaatcctctccagacggacgccgcggagaacaat	607 gggcggtacctgaatcctctccagacggacgccgcggagaacaat
G R Y L N P L Q T D A A E N N	G R Y L N P L Q T D A A E N N
652 gtttgtgacataaatgcagtgcatggcttgttcgccacaggaaca	652 gtttgtgacataaatgcagtgcatggcttgttcgccacaggaaca
V C D I N A V H G L F A T G T	V C D I N A V H G L F A T G T
697 atagagggtcgagtggaatgctgggatccaagagtaagaaaaagg	697 atagagggtcgagtggaatgctgggatccaagagtaagaaaaagg
I E G R V E C W D P R V R K R	I E G R V E C W D P R V R K R
742 gttggcgtgttagactgtgcattaaatagtgtcacagcagattca	742 gttggcgtgttagactgtgcattaaatagtgtcacagcagattca
V G V L D C A L N S V T A D S	V G V L D C A L N S V T A D S
787 gagataaacagcttacccaccatctctgcattgaagtttaatgga	787 gagataaacagcttacccaccatctctgcattgaagtttaatgga
E I N S L P T I S A L K F N G	E I N S L P T I S A L K F N G
832 gccttgagcatggcagttggaacatccacagggcaggtgttgtta	832 gccttgagcatggcagttggaacatccacagggcaggtgttgtta
A L S M A V G T S T G Q V L L	A L S M A V G T S T G Q V L L
877 tatgatcttcgctcggataagccgttgctagttaaggatcaccag	877 tatgatcttcgctcggataagccgttgctagttaaggatcaccag
Y D L R S D K P L L V K D H Q	Y D L R S D K P L L V K D H Q
922 tatgggctgccgattaagtcagttcatttccaggattccctagat	922 tatgggctgccaattaagtcagttcatttccaggattccctagat
Y G L P I K S V H F Q D S L D	Y G L P I K S V H F Q D S L D
967 ttggttttgtctgcagactctagaattgtcaaaatgtggaacaag	967 ttggttttgtctgcagactctagaattgtcaaaatgtggaacaag
L V L S A D S R I V K M W N K	L V L S A D S R I V K M W N K
1012 gactccgggaaaatatttacttcgttggagccagaacatgatctt	1012 gactccgggaaaatatttacttcgttggagccagaacatgatctt
D S G K I F T S L E P E H D L	D S G K I F T S L E P E H D L
1057 aatgatgtttgcctctatcccagctcaggaatgcttcttactgcc	1057 aatgatgtttgcctctatcccagctcaggaatgcttcttactgcc
N D V C L Y P S S G M L L T A	N D V C L Y P S S G M L L T A
1102 aatgaatcccccaagatgggcatttactacattccggttttgggt	1102 aatgaatcccccaagatgggcatttactacattccggttttgggt
N E S P K M G I Y Y I P V L G	N E S P K M G I Y Y I P V L G
1147 cctgctcccaggtggtgttccttcctggacaacttgacagaagag	1147 cctgctcccaggtggtgttccttcctggacaacttgacagaagag
P A P R W C S F L D N L T E E	P A P R W C S F L D N L T E E
1192 ttagaagagaatccagagagcacagtgtatgatgattacaaattt	1192 ttagaagagaatccagagagcacagtgtatgatgattacaaattt
L E E N P E S T V Y D D Y K F	L E E N P E S T V Y D D Y K F
1237 gtcaccaagaaagaccttgaaaatttagggctcacacatctcatt	1237 gtcaccaagaaagaccttgaaaatttagggctcacacatctcatt
V T K K D L E N L G L T H L I	V T K K D L E N L G L T H L I
1282 ggatcacctttcctccgagcatatatgcatgggtttttcatggat	1282 ggatcacctttcctccgagcatatatgcatgggtttttcatggat
G S P F L R A Y M H G F F M D	G S P F L R A Y M H G F F M D
1327 atcagactctaccacaaggtgaaattgatggtcaatccatttgct	1327 atcagactctaccacaaggtgaaattgatggtcaatccatttgct
I R L Y H K V K L M V N P F A	I R L Y H K V K L M V N P F A
1372 tatgaagaatataggaaagataaaattcgacagaagatagaagaa	1372 tatgaagaatataggaaagataaaattcgacagaagatagaagaa
Y E E Y R K D K I R Q K I E E	Y E E Y R K D K I R Q K I E E
1417 acacgtgcacagagagtccaattaaagaaactgccaaaagttaac	1417 acacgtgcacagagagtccaattaaagaaactgccaaaagttaac
T R A Q R V Q L K K L P K V N	T R A Q R V Q L K K L P K V N
1462 aaagaactggcacttaaattaattgaagaggaagaggagaaacaa	1462 aaagaactggcacttaaattaattgaagaggaagaggagaaacaa
K E L A L K L I E E E E E K Q	K E L A L K L I E E E E E E K Q
1507 aagtctacgctgaaaaagaaagttaagagcctccctaatatcctc	1507 aagtctacgctgaaaaagaaagttaagagcctccctaatatcctc
K S T L K K K V K S L P N I L	K S T L K K K V K S L P N I L


1552 acggatgatcgatttaaagttatgtttgagaatcccgacttccag
T D D R F K V M F E N P D F Q
1597 gtggacgaggagagtgaagagttcaggcttctcaacccccctcgtc
V D E E S E E F R L L N P L V
1642 tcgaggatcagtgagaagaggaagaagcagctgaggctcttggag
S R I S E K R K K Q L R L L E
1687 cagcaggagctcaaggatgaaggaggaggaagaaccggaaggaaaa
Q Q E L K D E E E E E P E G K
1732 cccagtgatgcagaaagttcagagagctcagatgatgagaaaggt
P S D A E S S E S S D D E K G
1777 tgggttgaagaagtcaggaagcagcgcagactccttcagcaggaa
W V E E V R K Q R R L L Q Q E
1822 gaaagagtgaagcggcaggagcaactgaaggaggaccagcagaca
E R V K R Q E Q L K E D Q Q T
1867 gtcctgaagccccagttctacgagatcaaagcaggcgaagagttc
V L K P Q F Y E I K A G E E F
1912 aggagcttcaaagagtctgccaccaagcagagactaatgaacaaa
R S F K E S A T K Q R L M N K
1957 actctggaagatcgtttgaaacttgaagcaaaacatgggacattg
T L E D R L K L E A K H G T L
2002 aatgtatcggacaccactgttggcagcaagcagttgacatttact
N V S D T T V G S K Q L T F T
2047 ctcaagaggtctgaacagcaaaagaagcaacaggaggctgaaaaa
L K R S E Q Q K K Q Q E A E K
2092 ctgcatcgtcaggaaaggaaaaatctgcgcaggtcggccagccac
L H R Q E R K N L R R S A S H
2137 ctgaggtccagacccagaagaggccggccctttcactga 2175
L R S R P R R G R P F H *
1552 acggatgatcgatttaaagttatgtttgagaatccggacttccag
T D D R F K V M F E N P D F Q
1597 gtggacgaggagagtgaagagttcaggcttctcaacccccctcgtc
V D E E S E E F R L L N P L V
1642 tcgaggatcagtgagaagaggaagaagcagctgaggctcttggag
S R I S E K R K K Q L R L L E
1687 cagcaggagctcaaggataaagaggaggaagaaccggaaggaaaa
Q Q E L K D K E E E E P E G K
1732 cccagtgatgcagaaagttcagagagctcagatgatgagaaaggt
P S D A E S S E S S D D E K G
1777 tgggttgaagaagtcaggaagcagcgcagactcctgcagcaggaa
W V E E V R K Q R R L L Q Q E
1822 gaaagagtgaagcggcaggagcaactgaaggaggaccagcagaca
E R V K R Q E Q L K E D Q Q T
1867 gtcctgaagccccagttctacgagatcaaagctggcgaagagttc
V L K P Q F Y E I K A G E E F
1912 agaagcttcaaagagtctgccaccaagcagagactgatgaacaaa
R S F K E S A T K Q R L M N K
1957 actctggaagatcgtttgaaacttgaagcaaaacatgggacattg
T L E D R L K L E A K H G T L
2002 aatgtatcggacaccactgttggcagcaagcagttgacatttact
N V S D T T V G S K Q L T F T
2047 ctcaagaggtctgaacagcaaaagaagcaacaggaggctgaaaaa
L K R S E Q Q K K Q Q E A E K
2092 ctgcatcgtcaggaaaggaaaaagctgcgcaggtcggccagccac
L H R Q E R K K L R R S A S H
2137 ctgaggtccagacccagaagaggccggccctttcactga 2175
L R S R P R R G R P F H *

10.4 Appendix IV – DNA methylation analysis of novel candidate differentially methylated regions (DMRs)

Table 10-4 Annotated list of novel candidate mouse strain-specific DMRs

Differential histone mark	Peak localization (*mm9*)	*Nearest gene* (Rel. position)	Enriched in (log2 FC)	Epityper amplicon localization (*mm9*)	CpGs	Average CpG Methylation in Spleen*		Enriched in MClp-seq
						C57BL/6	BALB/c	
			Set 1					
H3K4me3	**chr17:21844143-21845143**	***Zfp760*** **(promoter)**	**BALB/c (2.1x)**	**chr17:21844358-21844570**	**8**	**17%**	**1%**	**no**
H3K4me3	chr5:114625485-114626485	*Acacb* (intragenic)	BALB/c (2.7x)	chr5:114625776-114626064	16	3%	2%	no
				chr5:114625909-114626286	19	3%	3%	no
H3K4me3	chr2:168566763-168567763	*Atp9a* (promoter)	BALB/c (3.0x)	chr2:168566798-168567164	27	4%	5%	no
H3K4me3	chr11:69233014-69234014	*Dnahc2* (promoter)	BALB/c (3.0x)	chr11:69233742-69233972	17	2%	3%	no
H3K4me3	chr19:7066422-7067422	*Dnajc4* (promoter)	BALB/c (1.8x)	chr19:7066713-7067105	17	1%	2%	no
				chr19:7066441-7066802	14	15%	13%	no
H3K4me3	chr1:34067926-34068926	*Dst* (promoter)	BALB/c (1.8x)	chr1:34067828-34068212	15	4%	4%	no
				chr1:34068252-34068742	55	3%	3%	no
H3K4me3	chr6:52108100-52109100	*Hoxa1* (promoter)	BALB/c (1.6x)	chr6:52108174-52108564	27	4%	3%	no
H3K4me3	chr11:99247023-99248023	*Krt10* (intragenic)	BALB/c (2.3x)	chr11:99247785-99248012	10	12%	14%	no
H3K4me3	chr17:3083852-3084852	*Pisd-ps2* (promoter)	BALB/c (2.0x)	chr17:3084411-3084659	14	2%	2%	yes
H3K4me3	chr18:42110177-42111177	*Prelid2* (intragenic)	BALB/c (1.9x)	chr18:42110499-42110973	29	3%	4%	no
H3K4me3	chr1:191551704-191552704	*Ptpn14* (promoter)	BALB/c (2.2x)	chr1:191552381-191552797	37	5%	5%	no
H3K4me3	chr2:154428841-154429841	*Pxmp4* (promoter)	BALB/c (1.8x)	chr2:154429048-154429440	22	3%	2%	no
H3K4me3	chr10:42580350-42581350s	*Scml4* (intragenic)	BALB/c (1.8x)	chr10:42580614-42580915	6	10%	14%	no
				chr10:42580408-42580781	11	8%	11%	no
H3K4me3	chr18:74366816-74367816	*Ska1* (promoter)	BALB/c (1.7x)	chr18:74367379-74367621	18	1%	1%	no
H3K4me3	chr6:88024650-88025650	*Rpn1* (distal)	BALB/c (2.2x)	chr6:88025082-88025555	24	2%	2%	no
				chr6:88024951-88025305	22	2%	2%	no
H3K4me3	chrX:149441203-149442203	*Usp51* (intragenic)	BALB/c (1.7x)	chrX:149441822-149442102	16	31%	31%	yes (both)
H3K4me3	chr15:51708697-51709697	*Utp23* (intragenic)	BALB/c (1.9x)	chr15:51709053-51709247	12	2%	2%	no
				chr15:51708860-51709163	20	1%	3%	no
H3K4me3/ H3K27ac	chr6:124946693-124947693	*C530028O21Rik* (intragenic)	BALB/c (3.1x/2.1x)	chr6:124947234-124947572	17	6%	3%	no
				chr6:124946966-124947256	15	6%	7%	no
H3K4me3/ H3K27ac	chr4:62162837-62163837	*Hdhd3* (promoter)	BALB/c (1.6x/1.6x)	chr4:62162950-62163369	29	10%	10%	no
				chr4:62163090-62163566	29	5%	5%	no
H3K27ac	**chr4:140766411-140767411**	***6330545A04Rik*** **(distal)**	**C57BL/6 (4.7x)**	**chr4:140766797-140767168**	**8**	**72%**	**92%**	**no**
H3K27ac	**chr4:133613961-133614961**	***Aim1l*** **(distal)**	**C57BL/6 (2.4x)**	**chr4:133614214-133614597**	**13**	**30%**	**54%**	**yes (BALB/c)**
H3K27ac	**chr15:79722157-79723157**	***Apobec3*** **(promoter)**	**C57BL/6 (2.2x)**	**chr15:79724111-79724508**	**7**	**16%**	**68%**	**no**
				chr15:79722557-79722852	12	7%	4%	**no**
H3K27ac	**chr9:105973860-105974860**	***Col6a4*** **(intragenic)**	**BALB/c (7.7x)**	**chr9:105974425-105974596**	**6**	**66%**	**44%**	**no**
H3K27ac	**chr14:51443888-51444888**	***Parp2*** **(intragenic)**	**C57BL/6 (2.7x)**	**chr14:51444053-51444423**	**8**	**25%**	**59%**	**no**
H3K27ac	**chr19:47515636-47516636**	***Sh3pxd2a*** **(intragenic)**	**C57BL/6 (3.3x)**	**chr19:47515951-47516448**	**10**	**25%**	**51%**	**no**
H3K27ac	**chr13:96730805-96731805**	***Sv2c*** **(intragenic)**	**C57BL/6 (2.1x)**	**chr13:96731300-96731697**	**19**	**21%**	**62%**	**yes (BALB/c)**
				chr13:96731126-96731439	**16**	**17%**	**64%**	**yes (BALB/c)**

Differential histone mark	Peak localization (*mm9*)	*Nearest gene* (Rel. position)	Enriched in (log2 FC)	Epityper amplicon localization (*mm9*)	CpGs	Average CpG Methylation in Spleen*		Enriched in MCIp-seq
						C57BL/6	BALB/c	
H3K27ac	chr11:109721065-109722065	*1700012B07Rik* (distal)	C57BL/6 (2.3x)	chr11:109721598-109721882	7	2%	2%	no
				chr11:109721864-109722352	8	1%	2%	no
H3K27ac	chr11:113033555-113034555	*2610035D17Rik* (intragenic)	C57BL/6 (5.6x)	chr11:113033795-113034048	8	5%	6%	no
H3K27ac	chr9:40771667-40772667	*4931429I11Rik* (intragenic)	BALB/c (6.7x)	chr9:40771817-40772049	8	19%	17%	no
H3K27ac	chr5:150491416-150492416	*B3galtl* (intragenic)	BALB/c (4.8x)	chr5:150491557-150491924	8	88%	90%	no
H3K27ac	chr5:77523455-77524455	*Hopx* (intragenic)	C57BL/6 (2.6x)	chr5:77523733-77524198	38	2%	2%	no
				chr5:77524277-77524728	24	6%	7%	no
H3K27ac	chr15:74592965-74593965	*Ly6d* (intragenic)	C57BL/6 (2.4x)	chr15:74593160-74593293	4	23%	29%	no
H3K27ac	chr15:74582537-74583537	*Lynx1* (intragenic)	C57BL/6 (2.8x)	chr15:74583148-74583564	25	4%	4%	no
H3K27ac	chr17:13819968-13820968	*Mllt4* (distal)	C57BL/6 (2.3x)	chr17:13820648-13820981	8	7%	8%	no
				chr17:13820861-13821052	7	4%	8%	no
H3K27ac	chr19:47611036-47612036	*Obfc1* (promoter)	C57BL/6 (2.1x)	chr19:47611478-47611945	31	3%	4%	no
H3K27ac	chr11:20364229-20365229	*Sertad2* (distal)	C57BL/6 (2.1x)	chr11:20365098-20365450	8	76%	80%	no
				chr11:20364992-20365316	8	74%	77%	no
H3K27ac	chr19:43721462-43722462	*Slc25a28* (distal)	C57BL/6 (5.5x)	chr19:43721671-43721988	6	20%	24%	no
H3K27ac	chr15:96413816-96414816	*Slc38a1* (intragenic)	C57BL/6 (3.3x)	chr15:96414973-96415323	6	92%	94%	no
H3K27ac	chr11:8588764-8589764	*Tns3* (distal)	C57BL/6 (2.0x)	chr11:8588926-8589291	7	26%	39%	no
				chr11:8588507-8588756	5	86%	86%	no
H3K27ac	chr4:93001506-93002506	*Tusc1* (intragenic)	BALB/c (1.8x)	chr4:93001381-93001616	23	2%	1%	no
H3K27ac	chr11:5899700-5900700	*Ykt6* (intragenic)	C57BL/6 (2.1x)	chr11:5902166-5902660	7	77%	83%	no
				chr11:5901432-5901918	6	79%	83%	no
			Set 2					
H3K4me3	**chr3:96077378-96078378**	***Hist2h2bb* (distal)**	**BALB/c (4.1x)**	**chr3:96077670-96077953**	**7**	**25%**	**1%**	**yes**
				chr3:96077988-96078283	**7**	**76%**	**9%**	**yes (C57BL/6)**
H3K4me3	**chr7:150191605-150192605**	***Tspan32* (intragenic)**	**C57BL/6 (2.1x)**	**chr7:150192677-150192988**	**6**	**50%**	**86%**	**no**
H3K4me3	chr11:43341835-43342835	*Ccnjl* (promoter)	BALB/c (1.6x)	chr11:43342549-43342812	11	2%	2%	no
H3K4me3	chr6:136752255-136753255	*Hist4h4* (intragenic)	BALB/c (1.7x)	chr6:136753203-136753439	6	3%	2%	no
H3K4me3	chr13:67521833-67522833	*Zfp459* (promoter)	BALB/c (1.2x)	chr13:67522131-67522439	15	4%	11%	no
H3K27ac	**chr17:48565966-48566966**	***Apobec2* (intragenic)**	**C57BL/6 (1.8x)**	**chr17:48566705-48566973**	**4**	**21%**	**52%**	**no**
				chr17:48566071-48566366	**6**	**31%**	**79%**	**no**
H3K27ac	**chr11:105894819-105895819**	***Dcaf7* (promoter)**	**C57BL/6 (5.6x)**	**chr11:105895046-105895376**	**7**	**56%**	**80%**	**no**
H3K27ac	**chr16:30155069-30156069**	***Hes1* (distal)**	**C57BL/6 (5.5x)**	**chr16:30155216-30155568**	**6**	**15%**	**35%**	**no**
H3K27ac	**chr5:147874677-147875677**	***Lnx2* (intragenic)**	**BALB/c (1.8x)**	**chr5:147874763-147875096**	**7**	**46%**	**10%**	**no**
				chr5:147875405-147875618	**4**	**80%**	**25%**	**no**
H3K27ac	**chr5:123488913-123489913**	***Morn3* (intragenic)**	**C57BL/6 (2.1x)**	**chr5:123489136-123489456**	**17**	**3%**	**26%**	**yes (BALB/c)**
				chr5:123489428-123489728	**7**	**6%**	**68%**	**no**
H3K27ac	**chr13:101179103-101180103**	***Naip1* (intragenic)**	**BALB/c (7.8x)**	**chr13:101179250-101179593**	**5**	**84%**	**7%**	**no**
H3K27ac	**chr7:52996858-52997858**	***Spaca4* (intragenic)**	**C57BL/6 (3.5x)**	**chr7:52997182-52997586**	**9**	**10%**	**57%**	**no**
H3K27ac	**chr4:132650946-132651946**	***Wasf2* (distal)**	**C57BL/6 (2.5x)**	**chr4:132651381-132651641**	**5**	**19%**	**53%**	**no**
				chr4:132651025-132651364	**6**	**10%**	**47%**	**no**
H3K27ac	chr8:74968538-74969538	*1700030K09Rik* (intragenic)	C57BL/6 (1.7x)	chr8:74967739-74968002	11	2%	2%	no
				chr8:74967345-74967757	24	10%	13%	no
H3K27ac	chr2:154844650-154845650	*a* (intragenic)	C57BL/6 (5.2x)	chr2:154845337-154845688	14	6%	13%	no
				chr2:154844660-154844973	13	6%	7%	no
H3K27ac	chr16:31066530-31067530	*AI480653* (intragenic)	C57BL/6 (1.8x)	chr16:31066486-31066836	9	18%	32%	no
				chr16:31068482-31068909	11	76%	81%	no

Differential histone mark	Peak localization (*mm9*)	*Nearest gene* (Rel. position)	Enriched in (log2 FC)	Epityper amplicon localization (*mm9*)	CpGs	Average CpG Methylation in Spleen*		Enriched in MCIp-seq
						C57BL/6	BALB/c	
H3K27ac	chr14:51702668-51703668	*Ang* (distal)	BALB/c (4.1x)	chr14:51702905-51703187	12	3%	2%	no
				chr14:51702653-51702968	17	9%	9%	no
H3K27ac	chr11:57499273-57500273	*Galnt10* (intragenic)	C57BL/6 (1.8x)	chr11:57499808-57500157	5	66%	68%	no
				chr11:57500262-57500560	5	89%	82%	no
H3K27ac	chr9:48226482-48227482	*Gm8884* (distal)	C57BL/6 (1.8x)	chr9:48226834-48227177	8	72%	83%	no
H3K27ac	chr11:99099440-99100440	*Krt222* (intragenic)	C57BL/6 (1.9x)	chr11:99099757-99100069	11	7%	9%	no
				chr11:99100022-99100522	7	52%	58%	no
H3K27ac	chr11:57764784-57765784	*Larp1* (distal)	C57BL/6 (1.7x)	chr11:57765183-57765429	5	11%	11%	no
				chr11:57765947-57766302	4	29%	39%	no
H3K27ac	chr3:79056735-79057735	*LOC100233207* (distal)	C57BL/6 (2.1x)	chr3:79057262-79057600	7	11%	44%	no
H3K27ac	chr8:72345971-72346971	*Lpar2* (promoter)	C57BL/6 (1.8x)	chr8:72346241-72346530	17	3%	2%	no
H3K27ac	chr8:72356124-72357124	*Pbx4* (promoter)	C57BL/6 (1.9x)	chr8:72356942-72357146	6	1%	5%	no
				chr8:72356368-72356832	36	2%	5%	no
H3K27ac	chr2:156016009-156017009	*Phf20* (distal)	C57BL/6 (1.8x)	chr2:156016184-156016459	5	9%	22%	no
H3K27ac	chr4:126692793-126693793	*Sfpq* (distal)	C57BL/6 (3.3x)	chr4:126692940-126693230	8	6%	9%	no
H3K27ac	chr3:152591862-152592862	*St6galnac5* (intragenic)	C57BL/6 (3.8x)	chr3:152591738-152592016	6	68%	77%	no
				chr3:152590995-152591313	9	92%	93%	yes (both)
H3K27ac	chr9:115402330-115403330	*Stt3b* (distal)	C57BL/6 (2.0x)	chr9:115401618-115401996	5	12%	14%	no
				chr9:115402569-115403035	4	94%	87%	no
H3K27ac	chr7:133916447-133917447	*Ypel3* (intragenic)	BALB/c (4.1x)	chr7:133916194-133916470	6	91%	84%	no
				chr7:133916247-133916556	7	85%	83%	no

Acknowledgment

Als Erstes möchte ich meinem Betreuer und Doktorvater Prof. Dr. Michael Rehli danken. Sein Zuspruch und seine Zuversicht haben mich durch so manches finstere Promotions-Tal geleitet. Sein immer offenes Ohr und sein Vertrauen haben diese Arbeit erst ermöglicht.
Ebenso gilt mein Dank meinen beiden Mentoren Prof. Dr. Gernot Längst und Prof. Dr. Axel Imhof, die mich während der gesamten Arbeit begleitet und mir mit weisem Rat zur Seite gestanden haben.

Ein großes Dankschön möchte ich auch an die Abteilungschefs, Herrn Prof. Dr. Reinhard Andreesen und seinem Nachfolger Herrn Prof. Dr. Wolfgang Herr, richten. Ihrem Interesse an allen Mitarbeitern und ihrem Sinn für Zusammenhalt war das durchweg positive Arbeitsklima in der Abteilung Innere Medizin III zu verdanken.

Egal welche Experimente für dieses ehrgeizige Projekt anstanden, die Dimensionen waren meist gigantisch. Der schiere Umfang hätte einen schon in die Flucht schlagen können. Wärd da nicht ihr gewesen, Ireen, Laura und Johanna! Euch und ganz besonders dir, Johanna, gilt mein Dank für die Unterstützung im Labor. Zusammen haben wir den Humor nicht verloren, hoffe ich zumindest!
Apropos Humor, Almut, Sandra und Chris, euch als Kollegen und vor allem als Freunde zu haben, war ein dicker Lichtstreifen am Laborhorizont. Mit euch konnte ich die kleinen Freuden und großen Leiden im Laboralltag teilen, ertragen und genießen.
Auch dem Rest der (ehemaligen) Mannschaft – Lucia, Daggi, Claudia, Eddy, Julia, SandraP & Co - sowie den JCC-Mädels, der gesamten AG Kreutz und AG Edinger/Hoffmann ein ganz großes Danke! Obwohl definitiv sanierungsbedürftig, habt ihr den H1 zu einem wunderbaren Arbeitsplatz gemacht!
An dieser Stelle möchte ich auch Dr. Florian Meier, Dr. Michelle Boiani und Zbigniew Czyz nennen. Sie haben ihre kostbare Zeit dafür geopfert, mir mit ihrer Expertise das nötige Know-how für einige wichtige Experimente zu verschaffen.
Im Bezug auf die Weiterbildung meiner Person möchte ich auch RIGeL und der GRA RNA Biology danken. Nicht nur die Promotion selbst, sondern auch das große Spektrum an Weiterbildungsmaßnahmen und die Möglichkeit, mich persönlich einzubringen, haben mich in den letzten 4 Jahren geformt. Danke ganz besonders an Susanne und Phillip für die gemeinsame Zeit.

Babsi, Kathrin, Matze, UliO, Christian, UliS, Lissi, Ca, Lisa, Maria, Verena, Sandra, Inge, Manu, Mira und alle, die sonst noch auf meinem Leben wirken: Ihr habt mich zum Teil schon in der Schule, im Studium, und vor allem während diesem nervenaufreibenden, wahnsinnigen Lebensabschnitt namens Promotion begleitet, mir zugehört, mir gut zu gesprochen, mich aufgebaut. Dafür möchte ich euch danken. Ich hoffe ich kann auch in Zukunft auf euch zählen!

Um die Aufzählung komplett zu machen, möchte ich von ganzem Herzen meiner Familie, natürlich inklusive dem angeheirateten Teil, danken. Dabei ganz besonders Lisa, Mama, Papa und Peter. Ihr habt immer an mich geglaubt, mich unterstützt und vor allem im letzten Jahr zurückgesteckt. Ihr habt mir die nötige Erdung und Rückhalt gegeben.

Der Beste zum Schluss. Martin, du bist alles für mich. Geliebter Ehemann, bester Freund, Roomie. Du ganz besonders weißt, was ich durchgemacht habe, hast meine Launen ertragen, dich mit mir gefreut und gegrämt und mich immer wieder angespornt – danke, danke, danke!