The impact of intragenic CpG content on epigenetic control of transgene expression in mammalian cells

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES
DER NATURWISSENSCHAFTEN (DR. RER. NAT.)
DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN
DER UNIVERSITÄT REGENSBURG

vorgelegt von

Simone Krinner
aus Straubing

im Jahr 2012

Das Promotionsgesuch wurde eingereicht am 6. November 2012
Die Arbeit wurde angeleitet von Prof. Dr. Ralf Wagner

_____

Simone Krinner

Prüfungsausschuss:

Vorsitz:	Prof. Dr. Stephan Schneuwly
Erstgutachter:	Prof. Dr. Ralf Wagner
Zweitgutachter:	Prof. Dr. Christopher Baum
Drittprüfer:	Prof. Dr. Gernot Längst

1 Zusammenfassung

Die effiziente Produktion rekombinanter Therapeutika in Säugerzellen und Verbesserung gentherapeutischer Verfahren sind bedeutende und expandierende Felder in der medizinischen und pharmazeutischen Forschung. Plasmid-DNA (pDNA)-basierte Vektorsysteme stellen aufgrund ihrer Stabilität, der kostengünstigen Produktion sowie ihres hervorragenden Sicherheitsprofils ein innovatives Gentransfer-System dar. Trotz dieser Vorteile ist der Einsatz von pDNA-Vektoren angesichts begrenzter Transgen-Expressionsraten gegenüber Virus-basierten Verfahren limitiert. Dies erfordert neue Strategien zur Optimierung von pDNA-basierten Genexpressionssystemen, wie beispielsweise durch die gezielte Nutzung transkriptionsregulierender Mechanismen der Zielzelle. CpG Dinukleotide in Transgenen haben sich diesbezüglich als entscheidende Expressions-modulierende Elemente erwiesen.

Anhand der Reportergene codierend für das murine Makrophagen inflammatorische Protein 1 alpha (MIP-1α) und das humanisierte grün fluoreszierende Protein (GFP) konnte bereits in früheren Studien ein proportionaler Zusammenhang zwischen CpG Dinukleotiden im offenen Leserahmen und einem erhöhten Genexpressionslevel gezeigt werden. Dazu wurden die Nukleinsäure-Sequenzen der *mip-1α* und *gfp* Gene unter Verwendung alternativer Codons modifiziert. Ausgehend vom *mip-1α* Wildtyp wurde ein Codon-optimiertes Gen, sowie eine CpG-freie und eine CpG-maximierte Genvariante hergestellt. Weiterhin dienten das für humane Zellen Codon-optimierte *gfp* Gen und darauf basierend ein CpG-freies *gfp* Gen als Ausgangskonstrukte für Genexpressionsanalysen. Es konnte gezeigt werden, dass intragenische CpG Dinukleotide einen positiven Einfluss auf die Genexpression in Säugerzellen ausüben, während eine CpG-Depletion zu starken Expressionsverlusten führt. Während keine Hinweise auf veränderte CpG-basierte posttranskriptionelle Regulations-mechanismen zu finden waren, konnte eine deutliche Korrelation zwischen intragenischen CpG Dinukleotiden und gesteigerter *de novo* synthetisierter mRNA hergestellt werden.

In der vorliegenden Arbeit sollten die durch differenziellen intragenischen CpG-Gehalt hervorgerufenen Regulationsmechanismen von *gfp* und *mip-1α* aufgeklärt werden. Das relative Expressionsprofil der CpG-modifizierten *gfp* Transgene in CHO Flp-In Zellen konnte über den Zeitraum von mindestens einem Jahr durch antibiotischen Selektionsdruck konstant gehalten werden. Die Abwesenheit selektiver Bedingungen resultierte dagegen in sukzessiven Expressionseinbußen, welche sowohl auf Transgenverluste als auch DNA-Methylierung zurückzuführen waren. Während eine hohe intragenische CpG-Frequenz zu gesteigerten Methylierungsraten des Transgen-kontrollierenden Promoters führte, hatte eine intragenische CpG-Depletion einen beschleunigten Transgenverlust zur Folge. Der Genexpressions-Rückgang nach Selektionsrestriktion korrelierte weiterhin bei allen *gfp* Varianten mit einer höheren Chromatin-Dichte. Interessanterweise ging auch die CpG-Depletion der in Flp-In CHO und HEK 293 stabil und unter Selektionsdruck integrierten *gfp* und *mip-1α*

Transgenvarianten mit einer Chromatin-Verdichtung einher. Darüber hinaus bewirkte der variable CpG-Gehalt in *gfp* eine veränderte *in vitro*-Positionierung von Nukleosomen. Die Detektion vermehrt aktiv transkribierender RNA Polymerasen II am Gen-Ende CpG-maximierter *mip-1α* Transgene in stabil transfizierten HEK 293 Flp-In Zellen ließ auf erhöhte Elongationsraten als Folge von CpG-Maximierung schließen. Expressionsanalysen von *gfp* Chimären konnten zeigen, dass sich nicht nur die CpG-Frequenz, sondern vielmehr die räumliche Nähe intragenischer CpG Dinukleotide zum Transkriptionsstart (TSS) positiv auf die Expressionseffizienz auswirken.

Um die Effekte intragenischer CpG Dinukleotide auf die Transgenexpression in einem Gentherapie-relevanten Zellsystem zu testen, wurden murine, embryonale pluripotente Stammzellen der Linie P19 mittels lentiviraler Vektoren stabil mit den *gfp* CpG-Varianten unter verschiedenen Promotoren transduziert. Der Promotor des Cytomegalovirus (CMV) wies in diesem Expressionssystem eine erhöhte Disposition bezüglich *gene silencing* auf. Im Vergleich zum CMV Promotor führte der Promotor des humanen Elongationsfaktors 1 alpha (EF-1α) zu verzögerten, dennoch deutlichen, Expressionsverlusten. Im Gegensatz dazu verhinderte der bidirektionale, divergent transkribierte Promoter A2UCOE aufgrund seiner ubiquitären Chromatin-öffnenden Eigenschaften eine Transgen-Stilllegung komplett. In Bezug auf den intragenischen CpG-Gehalt konnte auch dieses Expressionssystem trotz hohem *gene silencing*-Potentials unter bestimmten Bedingungen von der Anwesenheit intragenischer CpG Dinukleotide profitieren. So wies das CpG-angereicherte *gfp,* exprimiert durch den EF-1α Promotor, auch in P19 Zellen eine deutlich erhöhte Expressionseffizienz auf. Weiterhin konnte die Gen-Stilllegung des CMV Promotor-kontrollierten *gfp* durch intragenische CpG Dinukleotide leicht verzögert werden. Die durch den A2UCOE Promotor vermittelte Transkription hingegen wurde durch intragenische CpG Dinukleotide in *gfp* nicht beeinflusst. Es wird vermutet, dass die Chromatin-öffnende Funktion des A2UCOE Elements eine Chromatin-Kompaktierung als Folge der CpG-Depletion verhindern kann. Mit dieser Eigenschaft scheint A2UCOE die Nachteile der CpG-Depletierung durch Chromatin Verdichtung aufheben zu können.

Insgesamt konnten die anhand der Transgene *gfp* und *mip-1α* gewonnenen Daten zeigen, dass sich intragenische CpG Dinukleotide in TSS-Nähe positiv auf die Transkriptionseffizienz auswirken. Die durchgeführten Analysen deuten darauf hin dass dieser Effekt auf die Delokalisierung und Destabilisierung des +1 Nukleosoms durch TSS-proximale intragenische CpG Dinukleotide zurück geht, während eine intragenische CpG-Depletion eine Chromatin-Kondensation zur Folge hat. Diese Veränderungen der Chromatinstruktur werden als Ergebnis epigenetischer Regulationsmechanismen postuliert, die durch die An-, beziehungsweise Abwesenheit intragenischer CpG Dinukleotide hervorgerufen werden. Die genauen Mechanismen dieses Phänomens sind weiterhin nicht vollständig geklärt.

2 Abstract

The improvement of gene therapy applications and efficient production of recombinant therapeutics in mammalian cells is a growing field of interest in medical and pharmaceutical research. Plasmid-DNA (pDNA)-based vector systems offer an innovative gene transfer strategy due to their high stability, cost efficient production and their excellent safety profile. Despite these advantages, the application of pDNA-vectors is limited compared to viral-vector-based gene transfer regarding transgene expression rates. This requires new strategies to optimize pDNA-based gene expression systems. The directed utilization of transcription regulating mechanisms in the target cell is a major strategy towards this aim. In this regard, CpG dinucleotides in transgenes have proven to serve as crucial expression-modulating elements.

Previous studies have demonstrated a strong correlation between the presence of CpG dinucleotides in transgenes and the level of gene expression by means of the reporter genes coding for the murine macrophage inflammatory protein 1 alpha (MIP-1α) and humanized green fluorescent protein (GFP). The DNA sequence of *mip-1α* and *gfp* was modified by using alternative codons. Based on the *mip-1α* wild type sequence, a codon optimized, CpG-depleted and CpG-enriched *mip-1α* gene variant were generated. Additionally, the CpG-rich *gfp*, optimized for human codon usage, and the CpG-depleted *gfp,* provided the basis for gene expression analyses. Decreased gene expression was observed as a result of intragenic CpG depletion, whereas the enrichment of intragenic CpG dinucleotides led to a dramatic increase of gene expression. No evidence for CpG-based posttranscriptional regulation mechanisms could be found. Instead, intragenic CpG dinucleotides clearly correlated with enhanced *de novo* synthesized mRNA.

This study aimed to shed light on the CpG-induced mechanisms responsible for expression efficiency variations in *gfp* and *mip-1α*. The relative expression profile of CpG-modified *gfp* transgenes in CHO Flp-In cells could be maintained over at least a year under antibiotic selection pressure. Withdrawal of selective conditions resulted in gradual decrease in *gfp* expression which was shown to be a consequence of both transgene loss and DNA methylation. While a high intragenic CpG frequency promoted DNA methylation rates of the mediating promoter, intragenic CpG depletion led to accelerated transgene loss. Moreover, gene expression decline upon selection pressure withdrawal correlated with a higher chromatin density in both *gfp* variants. Notably, chromatin compaction also correlated with intragenic CpG depletion in *gfp* and *mip*-1α, stably expressed in Flp-In CHO and HEK 293 cells under selection pressure. CpG variations in *gfp* were furthermore shown to influence nucleosome positions *in vitro.* The detection of increased actively transcribing RNAPII at the gene end of CpG-maximized *mip-1α* transgenes in stably transfected HEK 293 Flp-In cells indicated enhanced elongation rates resulting from CpG enrichment. Expression analyses of *gfp* chimera revealed that not only the CpG frequency, but rather the proximity of intragenic

CpG dinucleotides to the transcription start site (TSS) is beneficial for transgene efficiency.

To test the effects of intragenic CpG dinucleotides on transgene expression efficiency in a gene therapy-relevant cell system, murine embryonic pluripotent stem cells of the line P19 were stably transduced with lentiviral vectors (LV) containing the respective *gfp* variants under different promoters. The promoter of the cytomegalovirus (CMV) revealed a high disposition for gene silencing in this expression system. Compared to the CMV promoter, *gfp* transcription by the elongation factor 1 alpha (EF-1α) promoter resulted in delayed, yet significant transgene silencing in P19 cells. In contrast, the bidirectional, dual divergently transcribed A2UCOE promoter prevented transgene silencing via its chromatin opening abilities completely. With regard to CpG frequency, the LV-P19 expression system could also benefit from the presence of intragenic CpG dinucleotides under certain conditions, in spite its high gene silencing potential. EF-1α-promoter-controlled expression of the CpG-maximized *gfp* variant was clearly increased over the CpG-depleted *gfp* in P19 cells. CMV promoter-mediated *gfp* expression revealed slightly delayed gene silencing in CpG-rich compared to CpG-depleted *gfp*. In contrast, A2UCOE-mediated transcription was not affected by intragenic CpG dinucleotides. It is assumed that A2UCOE can overcome chromatin compaction arising from intragenic CpG depletion due to its chromatin opening property.

The sum of data could show that TSS-adjacent intragenic CpG dinucleotides in *gfp* and *mip-1α* transgenes positively influence transcription efficiency. The results gained in this work imply that this effect results from delocalization and destabilization of the +1 nucleosome, whereas intragenic CpG depletion leads to a higher level of chromatin density. These chromatin changes are assumed to result from a complex epigenetic regulation network triggered by intragenic CpG changes. The exact mechanism of this phenomenon remains to be elucidated.

3 Introduction

3.1 Eukaryotic gene transcription

The regulation of gene transcription is fundamental for cellular differentiation, proliferation and the proper response to environmental changes. To achieve the high level of specialization of cells that have a common set of genetic information, gene transcription is subjected to multiple regulatory mechanisms. In prokaryotes, gene regulation allows a single cell to respond to environmental changes by switching genes on and off [1]. In multicellular eukaryotic systems, gene regulation not only serves to adjust to environmental changes. The biologically more important purpose of gene control is to provide the proliferation of many different cell types that compose a multicellular organism. Eukaryotic transcription is an immensely complicated process that is regulated by a large number of proteins (Figure 1) [2]. Sequence-specific binding factors/transcription factors interact with their DNA motifs in response to cellular signals [3]. They recruit transcriptional co-regulators to alter the local chromatin environment and facilitate assembly of the pre-initiation complex (PIC) [4], which is composed of the general transcription factors (GTFs) and Polymerase II (RNAPII) [5]. Among the three eukaryotic Polymerases, RNAPII, consisting of 12 subunits, is responsible for the transcription of protein coding genes [6]. GTFs, comprising TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, are essential for exact positioning of RNAPII at the promoter. Associated as the basal transcription machinery, RNAPII and GTFs form a preinitiation complex (PIC) at the core promoter, which is usually located upstream of the translated region [7]. Most core promoters contain a TATA box or equivalent motifs as an essential recognition feature for the basal transcription machinery [8]. TATA-boxes are present in the core promoter region and are typically 30–60 base pairs (bp) upstream of the transcription start site. In addition to these promoter motifs, the initiator (Inr) or downstream promoter element (DPE) interact with various components of the basal transcription machinery [9]. Another feature found at promoters of expressed genes in the yeast genome is the nucleosome-free region (NFR) [10]. What exactly creates an NFR is not fully understood, although some studies could correlate NFRs to poly-dA-dT tracts [11] or CpG islands [12]. Besides promoter regions, enhancers, also termed distal regulatory elements (DREs), contain binding sites for transcription factors. They can be located up to several thousand base pairs away from the actual initiation site [13]. Sequence-specific DNA binding transcription factors act as activators or repressors of transcription. They simultaneously recognize both promoter or enhancer sequences and other co-regulators through their DNA-binding domains and activation domains [4]. Whether a sequence-specific regulator activates or represses gene transcription depends on the genomic context and recruited co-regulators [2]. Co-regulators mainly comprise chromatin-modifying and/or chromatin-remodeling enzymes and the mediator complex [14]. The mediator complex facilitates the interaction between DNA-binding transcription factors, co-regulators and the basal transcription machinery [15].

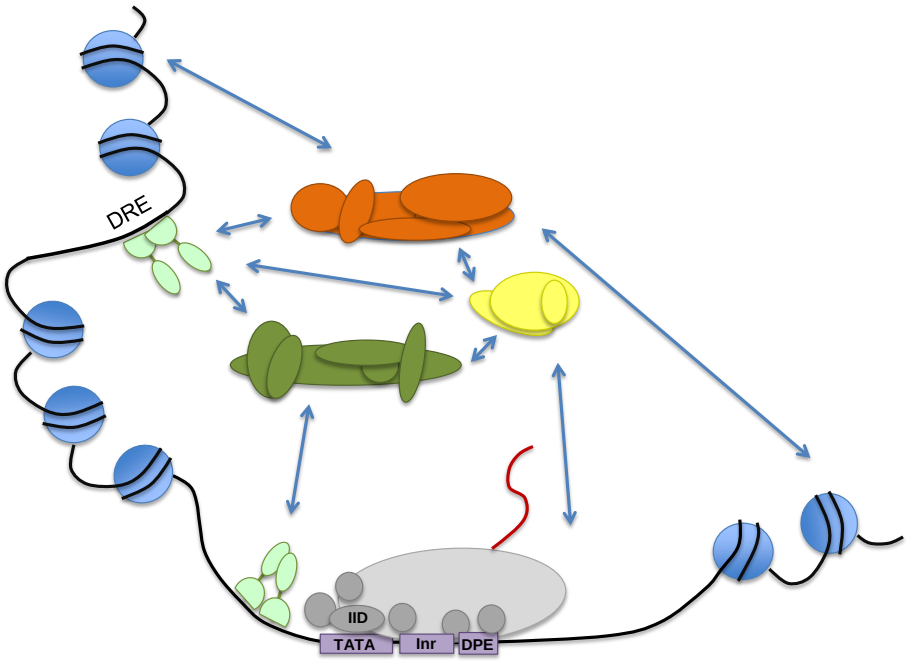

Figure 1 | Regulation of eukaryotic transcription (simplified). Assembly of the PIC, containing RNAPII (light grey) and GTFs (dark grey) is initiated by binding of TFIID to core promoter elements like TATA box, Initiator (Inr) or downstream promoter element (DPE) (purple). Transcriptional gene regulation involves: the binding of sequence specific binding factors (light green) to distal regulatory elements (DREs) and proximal promoter regions; interactions of DNA-binding factors with co-regulators like mediator (yellow), histone modifying complexes (green), chromatin remodelers (orange) and the basal transcription machinery (grey). The C-terminal domain (CTD) (red wavy line) is unphosphorylated in the PIC and becomes multiply phosphorylated upon initiation. As RNAPII traverses a transcription unit, the phosphorylation pattern changes resulting in the recruitment of different proteins. The concerted function of all these factors is to express a subset of genes as dictated by a complex interplay of environmental signals.

The C-terminal domain (CTD) of the largest subunit of the eukaryotic RNAPII contains several YSPTSPS heptad repeats (52 in mammals) that are unphosphorylated in the PIC of RNAPII and become multiply phosphorylated upon initiation [16]. As RNAPII traverses a transcription unit, the phosphorylation pattern changes resulting in the recruitment of different proteins to the CTD [17]. Phosphorylation has predominantly been found at serine 2 and serine 5 of the heptad repeats. Phosphorylation of the serine 5 residue occurs during transcription initiation and has been connected to multiple processes of transcription such as promoter clearance for transition from initiation to early elongation and 5′-end capping of pre-mRNA [18]. Modification of serine 2 is found

when the polymerase is associated with the coding region and has been implicated in productive elongation and the 3′-end processing of the transcript [19].

Several regulatory proteins specifically recognize the respective phosphorylation pattern of the CTD. Thereby, the CTD of RNAPII coordinates events during the transcription cycle by recruiting co-regulators involved in histone modifications and/or remodeling, transcription elongation, termination and mRNA processing [2].

3.2 Chromatin

Eukaryotic DNA is up to a thousand times longer than the cell's length [20]. Therefore, an organized packaging system is needed to fit the DNA into the nucleus. The nucleoprotein complex that meets this requirement is called chromatin. The term was first used by Walther Flemming, who discovered a visible cell substance with staining characteristics and therefore named it chromatin, which means "stainable material" [21].

Different states of chromatin, called euchromatin and heterochromatin, are found in the nucleus. They correlate with transcriptional active or repressed genes. Euchromatin undergoes a process of condensation and decondensation during cell cycle. It constitutes the majority of the chromosomal material and contains genes that are actively expressed. Heterochromatin remains highly condensed during the cell cycle. It is mostly found at the centromers and telomers of chromosomes as well as along the entire inactive X chromosome in female mammals [22].

3.2.1 The nucleosome

Nucleosomes are the primary structural units of chromatin, composed of DNA and histones. Histones are highly conserved, basic proteins of 11 to 21 kilo Dalton (kDa) (Table 1). In 1997, the structure of a nucleosome core particle could be resolved by X-ray diffraction at a resolution of 2.8 Å (Figure 2) [23]. It shows a nucleoprotein complex of approximately 147 bp of genomic DNA wrapped in a left handed superhelix 1.7 times around a histone octamere which has a diameter of 11 nm in length and 5.5 nm in height.

Table 1 | Molecular weight and size of histones. Values given are derived from bovine histones. Modified from [24].

Histone protein	Molecular weight [kDa]	Number of amino acids
H1	21,130	223
H2A	13, 960	129
H2B	13,774	125
H3	15,273	135
H4	11,236	102

A histone octamere contains two copies each of histones H2A, H2B, H3 and H4. All four histone proteins have a similar structural motif in common. The trihelical histone fold core mediates both binding between histones itself and between histones and DNA. Each histone has polypeptide extensions with NH_2- and/or COOH-terminal ends that stick out from the globular regions. These tails are targets for posttranslational modifications like acetylation and methylation [25]. Different from the rest of the histones, histone H1 is involved in the chromatin packing into a higher-order structure [20].

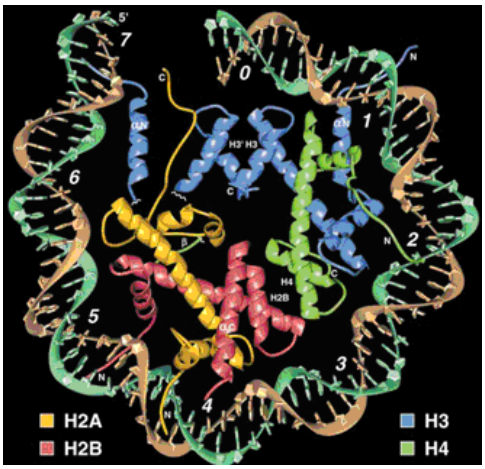

Figure 2 | Structure of the nucleosome core particle. The model shows the DNA double helix (brown and torquiouse) wound around the central histone octamere, consisting of two copies each of histones H2A, H2B, H3 and H4. Hydrogen bonds and electrostatic interactions between histones and DNA keep the nucleosome in place [23].

3.2.2 Chromatin organization

Nucleosomes are connected by nucleosome-free linker DNA to form a 10-nm fiber, also called the "beads-on-a-string array" [26][27]. The length of linker DNA varies among species, ranging from about 20 to 60bp. The linker region and parts of the nucleosomal DNA are associated with the linker histone H1, which binds to the nucleosome and causes the assembly of nucleosomes into a higher-order structure, the 30-nm filament [25][20]. While the X-ray crystal structure of the nucleosome core particle has early been resolved in atomic detail [23], the structure of the 30-nm chromatin fiber has been an issue of debate. In 1976, Finch and Klug postulated the "solenoidal model for superstructure in chromatin", which would direct the linker DNA between two nucleosomes into a strong bend [28]. For another model of organization, the so called

zig-zag structure, it was assumed that the linker DNA is straight and crosses the center of the 30-nm fiber [29]. X-ray analysis of a tetra nucleosome seems to support the zig-zag structure, which falls into the category of the 'two-start helix' type [30]. By contrast, electron microscope measurements provide evidence for the solenoid model characterized by interdigitated nucleosomes [31]. Both models agree on the function of the linker histone to determine the topology and degree of chromatin compaction [32]. Very recent analyses indicate that the 30-nm fiber involves both zigzag and bent linker motifs, depending on physiological conditions [33]. The 30-nm chromatin fiber results in an approximately 50-fold compaction of DNA. To obtain a higher level of organization, a hierarchical folding of chromatin structure, schematically illustrated in Figure 3, is needed [22]. A series of loops of 30-nm fibers are anchored at their base to the chromatin scaffold to form the 300-nm fiber [34]. The chromatin scaffold consists of non-histone proteins and has the shape of a metaphase chromosome. On average, each loop encompasses 20.000 to 500.000bp of DNA and is about 300nm in length. Tight helical coiling of the 300nm fiber produces the scaffold-associated chromatin structure. This helix is again packed and folded to generate an individual 700nm wide chromatid, two of which compose a metaphase chromosome [22] (Figure 3).

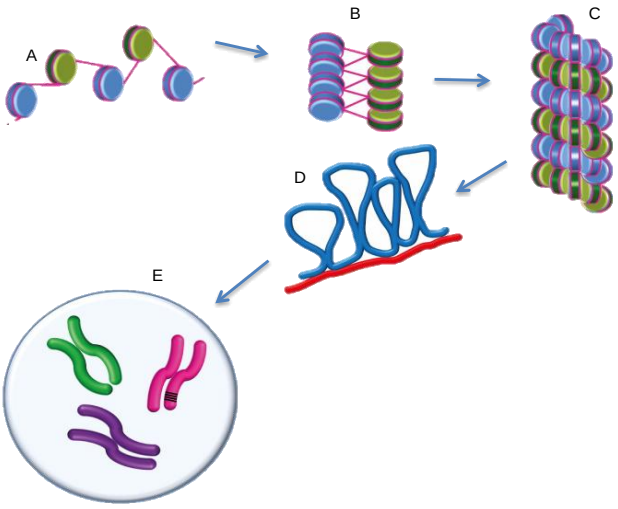

Figure 3 | Hierarchical folding of chromatin. (A) Beads-on-a-string array. Alternating nucleosomes are depicted with blue and green surfaces; (B) The 30-nm fiber twists further and forms a more compact fiber (C) that is arranged in loops (blue), with some portions attached to a protein scaffold (red) (D); (F) metaphase chromosome. Modified from [35].

The mechanism of higher order chromatin formation, ultimately resulting in metaphase chromosome formation, is still poorly understood. Multiple chromatin-associated proteins (CAPs) have been suggested to play an important role in the formation and dissociation of the chromatin structure beyond the 30-nm fiber. H1 is considered to be an important CAP in the organization of higher chromatin structure by stabilizing the folded state as was revealed by electron microscopy [36]. An important process for chromosome organization is the interaction of core histone domain tails which are also targets of multiple modifications in the course of gene transcription [37].

3.3 Transcriptional control by chromatin

Chromatin generally limits the accessibility of specific DNA sequences and inhibits the initiation and progression of the polymerase during transcription. There are basically three different ways by which the chromatin structure can be altered: i)By chromatin remodeling, ii)histone modification and iii) the replacement of core histones by histone variants. Together with DNA methylation and RNA binding, these regulation mechanisms are summarized as epigenetic control [38].

3.3.1 Histone modifications

To date, more than a hundred of histone modifications have been found. Several recent reviews cover this complex topic [39][40][41]. In the following sections, only a selection of modifications controlling gene activity is discussed. Among the many types of histone modifications that have been detected so far, acetylation, methylation and phosphorylation are the most frequently detected and best understood (Table 2). Over 60 different histone residues have been identified to be a target of modification, and in the case of methylation, multiple modifications (mono-, di- and trimethyl) can occur at one lysine or arginine [42][43].

Table 2 | Overview of the most important types of histone modifications in mammals. Modified amino acids include Lysine (K), Arginine (R), Serine (S), Threonine and Proline (P). Modified from [44].

Modifications	Residues Modified	Modification Position	Impact on Transcription
Acetylation	K-ac	H3 (9,14,18,56), H4 (5,8,13,16), H2A, H2B	Activation
Methylation (lys)	K-me1 K-me2 K-me3	H3 (4,36,79) H3 (9,27), H4 (20)	Activation Repression
Methylation (arg)	R-me1 R-me2a R-me2s	H3 (17,23), H4 (3)	Activation
Phosphorylation	S-ph T-ph	H3 (3,10,28), H2A, H2B	Activation
Ubiquitylation	K-ub	H2B (120) H2A (119)	Activation Repression
Sumoylation	K-su	H2B (6/7), H2A (126)	Repression
Isomerization	P-cis > P-trans	H3 (30-38)	Activation/ Repression

Depending on type and position of modification, opposed effects on transcription rate have been observed. The acetylation of histones generally activates a gene cumulatively, whereas methylation can have opposing effects (Table 2). Modifications that have been connected with transcription activation have been described as euchromatin modifications. Those that have been mapped to inactive genes are referred to as heterochromatin modification [45].

Genome-wide studies have revealed that individual histone modifications can be mapped to specific states of gene activity [46] (Figure 4). For example, the modifications H3K4me2/3 (histone H3 lysine4 di- and trimethylation) are mainly found in actively transcribing promoters, and H3K36me3 is frequently found in the body of actively transcribed genes, increasing towards the 3' end. By contrast, modifications like H3K27me3 and H4K20me3 are mostly mapped to regions where transcription is repressed [39]. Some modifications, such as H3K27me3 and H3K4me3 are however coincident with both activation and repression of gene transcription, respectively (Figure 4).

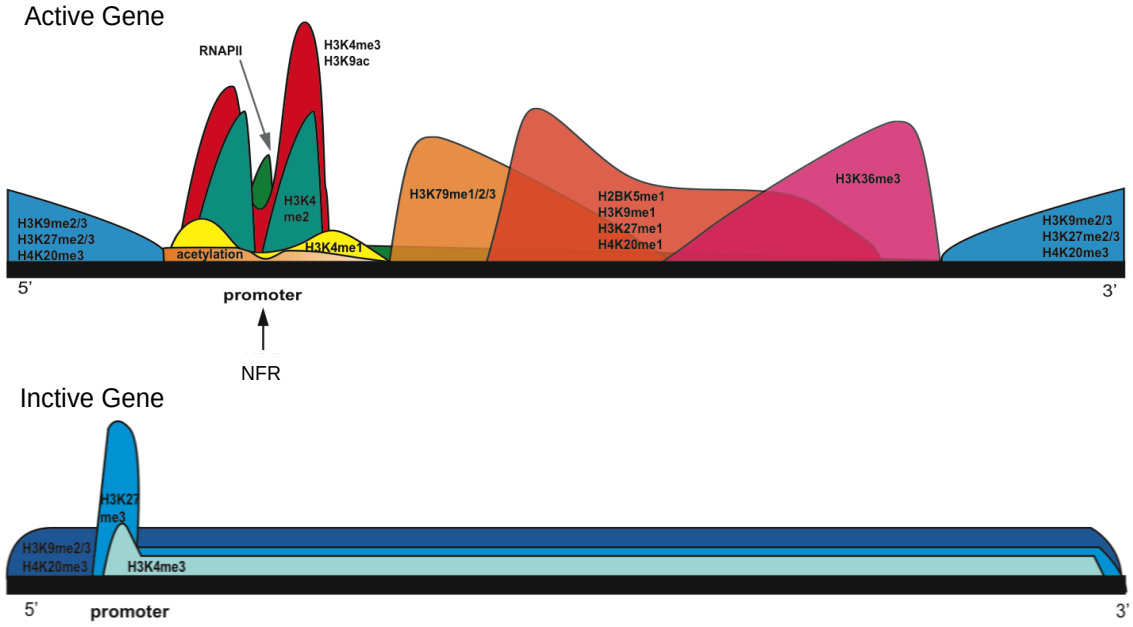

Figure 4 | Distribution of histone modifications on active and inactive genes. Modification patterns differ on actively transcribed and silenced genes, which is displayed as a schematic view of modification distribution over the gene. Promoters of actively transcribed genes carry high levels of active modifications such as acetylations and methylation of H3K4. At the transcriptional start site there is a nucleosome-free region (NFR) within the promoter. Inactive genes have a fairly even distribution of silencing modifications, such as H3K9 methylation and H4K20 methylation, whereas H3K27 methylation is enriched in the promoter. Modified from [39].

Strahl and Allis postulated the hypothesis of a histone code, proposing that the combination of histone modifications at a certain genomic locus determines the activity state of the underlying gene [47]. This hypothesis of a histone code is heavily discussed within epigenetic research, arguing that gene regulation by histone modifications might rather reflect a cumulative more than a combinatorial effect [48]. Nevertheless, the frequently made observation of distinct histone patterns demonstrates that histone modifications can indeed serve as indicator for gene activity or inactivity. In what respect these histone distributions are a matter of cause or consequence of gene activity is however not fully understood [39].

Histone modification is carried out by a variety of enzymes, categorized as acetyltransferases, methyltransferases, kinases etc. A detailed list of histone modifying enzymes is reviewed by Kouzarides [44]. The co-presence of both modifying and de-modifying enzymes indicates that histone modification is a highly dynamic process.

There are two major functions of histone modifications. First of all, histone modifications result in the weakening of inter- and intranucleosomal as well as histone-DNA interactions, thereby relaxing the chromatin structure. A simple consideration that led to this assumption is the fact that, apart from methylation, histone modifications all result in a net charge change of nucleosomes [39]. The disruption of chromatin contacts allows transcription factors to bind to their targets and is therefore fundamental for transcription. The second purpose of histone modification is the direct recruitment of regulatory proteins or DNA-methyltransferases (DNMTs) to their cognate binding sites [49]. An example for such co-regulators is the SET domain-containing histone methyltransferase enzyme SUV39H1, which is responsible for trimethylation of H3K9 and heterochromatinization of pericentromeric satellite repeats. These proteins are also required to recruit *de novo* methyltransferases to methylate CpG dinucleotides in the satellite sequence [50]. In addition to transcription factors and DNA modifying enzymes, histone modification patterns interact with remodeler complexes [43].

3.3.2 Chromatin remodeling

The dynamic property of DNA is maintained by chromatin remodeling complexes. These multi-protein complexes are essential for many chromatin functions such as the proper spacing of nucleosomes during nucleosome assembly, DNA repair or the binding of transcription factors to specific genes in the course of transcription regulation [51]. A broad range of remodeler complexes has been identified. All of them contain an ATPase domain which belongs to the superfamily II (SFII). On the basis of sequence similarities of the ATPases, remodeller complexes can be grouped into a number of subfamilies [52][53]. Most of these subfamilies have been designated to the archetypal member, such as *S.cerevisiae* Snf2p (Snf2 subfamily), *Drosophila melanogaster* Iswi (Iswi subfamily), or *Mus muculus* Chd1 (Chd subfamily). Several of them, e.g. members of the Iswi subfamily, have been reported to possess DNA-translocation activity [54]. Different remodelers affect the structure of the nucleosome array in a particular way and thereby influence a widespread number of nuclear processes, reviewed in [52]. For instance, the members of Iswi, namely the NURF (nucleosome remodeling factor), CHRAC (chromatin accessibility factor) and ACF (ATP-utilizing chromatin-assembly and remodeling factor) predominantly position nucleosomes in a manner to repress transcription [55]. By contrast, RSC (remodels the structure of chromatin), a member of the Swi/Snf family, mediates pathways that both activate and repress transcription [56]. Different than the variety of remodelers with regard to substrate specificity and chromatin product, the mechanism by which nucleosomes are rearranged has been suggested to be uniform. According to the 'loop recapture model', DNA translocation against a histone octamere is achieved by the successive detachment of DNA, starting from the edge of the nucleosome, its bending and recapturing by the octamere to form a loop that is carried along the DNA strand [57].

3.3.3 Sequence dependent nucleosome positioning

It is now well established that the DNA sequence itself determines the strength of DNA-histone interactions and the bending flexibility of the DNA helix around a histone octamere [58][59]. Poly (A) and poly (T) regions result in conformationally rigid molecules and therefore require high energy to incorporate into nucleosomes. By contrast, dinucleotides form nucleosomes of high stability: AA, TT and TA dinucleotides are favored approximately every 10bp where both DNA strands face towards the nucleosome core. GC dinucleotides are favored approximately every 10bp where both phosphodiester backbones face outward (Figure 5). A study of Gupta et al. has identified a 3bp periodicity of CG and GC dinucleotides to be a highly nucleosome favored sequence [60].

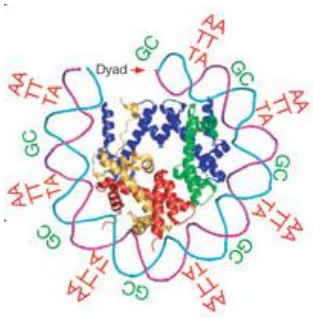

Figure 5 | Three dimensional structure of one-half of a symmetric nucleosome. Bends around the nucleosome core are favored by the dinucleotides AA/TT/TA that oscillate approximately 10bp periodically in phase with each other and out of phase with GC dinucleotides recurring every approximately 10bp as indicated [59].

The sequence preference calculation is based on a thermodynamic model that evaluates the free energy for any nucleosome constellation [59]. This includes the calculation of sterically allowed nucleosome organizations and competition between positions at each dinucleotide. A genome-wide analysis of nucleosome positioning demonstrated that approximately 50% of the *in vivo* nucleosome organization is solely determined by sequence preferences of nucleosome occupation [61]. By using high-density tiling arrays over the yeast genome, it was shown that a nucleosome-free region (NFR) was a common feature of promoters [10]. The so-called "–1" and "+1" nucleosomes are located in canonical regions upstream and immediately downstream of the NFR, respectively. These well-positioned nucleosomes encompassing the NFR at promoters have regulatory functions of transcriptional regulation (see chapter 3.1) [62].

3.4 Cytosine Guanine Dinucleotides

Nucleosome positioning is influenced by short periodic repeats of cytosines followed by a guanine [60]. These so called CpG dinucleotides are significantly underrepresented throughout the vertebrate genome than would be calculated from base composition. [63][64][65]. Since cytosines within CpGs are the exclusive targets for methylation in vertebrates, it was anticipated that this deficiency was related to DNA methylation [66]. The selective pressure resulting in this CpG loss was provided by the inherent mutability of methylated cytosine. The deamination of cytosine results in uracil, which is easily recognized and removed by uracil glycosylases. By contrast, the deamination of methyl cytosine gives rise to thymine, which is not recognized as foreign and therefore leads to a transition mutation in the subsequent replication. As a result, methylated CpG dinucleotides in the germ line tend to be lost over time [67]. Organisms with high levels of DNA methylation therefore tend to exhibit the most pronounced CpG deficiency [65].

3.4.1 CpG methylation

DNA methylation patterns among eukaryotes are not uniform. The most frequent pattern found in invertebrate animals is the so-called 'mosaic methylation'. It is characterized by moderate levels of methyl-CpG dinucleotides accumulated in domains of methylated DNA, interspersed with unmethylated domains. Vertebrates, on the other hand, exhibit high levels of methylated CpG dinucleotides distributed over the entire genome, except for small methylation free regions at transcriptionally active regions. This pattern is referred to as the 'global methylation' [68]. The transition from the ancestral mosaic methylation to the vertebrate global methylation is believed to have evolved in the evolution of CpG DNA immunity. The genomes of most bacteria and DNA viruses are rich in unmethylated CpG dinucleotides. These CpG motifs of several microbial parasites are detected by pattern recognition receptors, such as the Toll-like receptor 9 (TLR9), during the innate immune response in some vertebrates [69]. Since methylated CpGs have no potential to activate this defense, the genome of the host vertebrates prevents an auto immune response. The CpG-poor, globally methylated vertebrate genome is therefore believed to be a prerequisite of the CpG immunity [68]. The DNA methylation patterns in mammalian cells are usually well regulated and tissue-specific [70][71]. DNA methylation patterns of specific cell types are established during mammalian development and maintained in adult somatic cells [72]. In mammalian germ cells and early embryos, dramatic reprogramming with complete removal of methylation occurs, followed by renewed *de novo* methylation [73]. Not only global methylation changes, but also gene-specific *de novo* methylation and demethylation have been observed, for example during differentiation of hematopoietic progenitors [74]. DNA methylation in mammalian cells is mostly correlated with gene silencing, which is virtually always the case if this concerns promoter elements [75][76]. However, DNA methylation of gene bodies is also found to be positively correlated with transcription [77][78][79].

The majority of methylated DNA in differentiated cells is however harbored by non-coding transposable elements such as SINEs (short interspersed nuclear elements), LINEs (long interspersed nuclear elements) and endogenous retroviruses. These elements encompass approximately 42% of the human genome [80][81].

Methylation occurs at the 5-position of the cytosine residue within CpG dinucleotides, resulting in 5-methylcytosine (m^5C). The reaction is catalyzed by DNA methyltransferases (DNMTs), which catalyze the transfer of a methyl group from S-adenosyl-L-methionine to cytosine [82]. There are three enzymatically active DNMTs, which can be divided into *de novo* and maintenance methyltransferases. *De novo* methyltransferases act after the replication in unmethylated DNA. Maintenance methyltransferases catalyze the addition of methyl groups to hemi-methylated DNA during replication [82]. DNMT1 is the major maintenance methyltransferase [83]. DNMT3A and DNMT3B are *de novo* methyltransferases acting on unmethylated DNA. They are responsible for establishing methylation patterns during early development and each of them has distinct functions [84]. DNMT3L is a protein that is homologous to DNMT3A and DNMT3B but contains no catalytic activity. Instead, DNMT3L assists the methylation during gametogenesis by recruiting *de novo* methyltransferases [85]. DNA demethylation can be accomplished either passively, by leaving the new DNA strand unmethylated after replication, or actively. Some studies support the existence of active demethylation in zygotes [86] and in somatic cells [87]. So far, the exact mechanism is still not fully understood.

3.4.2 Unmethylated CpG dinucleotides

CpG dinucleotides are largely depleted throughout the mammalian genome as a consequence of their high susceptibility to mutation [66]. The result is that CpGs are relatively rare unless there is selective pressure to keep them or a region is not methylated due to active regulation of gene expression. Those genomic loci are mostly promoter regions of housekeeping genes that comprise at least half of the genes in the human genome [76].

It has been suggested that the unmethylated state of CpG dinucleotides is also dependent on germ line and early embryonic transcription. As a result of this lack of methylation, CpG dinucleotides in these regions are less suppressed and consequently appear relatively CpG-rich compared with the rest of the genome [88]. These stretches of mostly non-methylated CpGs are called CpG islands. CpG islands, defined by Bird in 1986, are on average 1000bp of length, have a C+G content of 0.5 or higher and an observed to expected CpG dinucleotide ratio of 0.6 or higher within a range of 200bp or greater [89][90]. CpG Islands are mostly found within the promoter and the first exon of several genes, particularly housekeeping genes [67][91]. In addition to housekeeping promoters, the average of protein coding genes in the human genome display a significant excess of CpG dinucleotides in exons, most pronounced in the first exon, compared to introns [67][92][93][94].

3.4.3 Gene control mechanisms directed by CpG dinucleotides

The high frequency of CpG dinucleotides in promoters and gene bodies of constitutively expressed genes versus the low frequency of CpG dinucleotides in mostly non-functional DNA already points to the outstanding role of this element as a transcriptional regulator. Despite more than 25 years of intensive study on CpG islands/regulatory CpG motifs since their discovery [89], the exact mechanisms by which CpG dinucleotides affect gene transcription are still poorly understood.

Trans acting proteins have been found that interact with unmethylated CpG dinucleotides leading to a unique chromatin architecture [95]. The transcription factor Sp1, for instance, has been demonstrated to bind to unmethylated CpG Islands to protect them from *de novo* methylation, which ensures active gene transcription [96]. In addition to Sp1, the CRE binding factor (CREB) [97] and CCCTC binding factor (CTCF) [98] contain CpG in their binding recognition site and DNA recognition is impaired upon CpG methylation.

More than 15 years ago, another important factor binding to unmethylated CpG dinucleotides was found in tobacco: the nuclear CpG-binding protein 1 (CGBP-1) binds with high affinity to unmethylated CpG dinucleotides [99]. A human CpG binding protein (hCGBP) was isolated a few years later, revealing specific binding for unmethylated CpG dinucleotides and thereby functioning as a transcriptional activator [100]. Subsequently, this protein was renamed as CXXC finger protein 1 (CFP1) [101]. CFP1 has frequently been localized in nuclear regions that are associated with euchromatin, which underlines its exclusive function as a transcriptional activator [102].

The key feature of CFP1 is a cysteinrich CXXC DNA-binding domain [100]. This zinc-finger like domain is highly conserved and frequently found in proteins involved in epigenetic regulation, such as the DNA methyltransferase 1 (Dnmt1) [103], methyl-CpG binding proteins MBD [104] and histone H3-Lys4 methyltransferase [105]. CFP1 was shown to associate with a histone H3K4 methyltransferase complex (SET1 complex) catalyzing the addition of the tri-methyl modification (H3K4me3) [106]. H3K4me3 coincides with promoters and 5' end of actively transcribed genes [107] (see also chapter 3.3.1). Histone lysine methylation marks are recognized by specific effector proteins containing plant homeodomain (PHD) finger domains or chromatin organization modifier (chromo) domains. PHD finger proteins can activate gene transcription, such as via TFIID [108] and the nucleosome remodeling factor (NURF) [109]. Another transcription factor binding to unmethylated CpG dinucleotides via the zink finger CXXC domain is the H3K36-specific lysine demethylase enzyme KDM2A. Binding of KDM2A to CpG results in removal of H3K36 methylation, thereby creating a "CpG island chromatin" that is depleted of this repressive modification [110].

The binding of unmethylated CpG dinucleotides by CpG-specific transcription factors, which are able to affect histone modifying activities, suggests that CpG dinucleotides may use chromatin associated processes to provide a transcriptionally active surface [95]. In addition to chromatin mediating abilities, early studies of CpG island chromatin revealed a distinct depletion of Histone H1 at CpG islands [111]. Histone H1 represses transcription [112] due to stabilization of chromatin structure [113].

Methylated CpG dinucleotides of regulatory elements have also been found to direct numerous gene control processes. For example, CpGs involved in tumori-genesis [114] or genomic imprinting [115] become methylated during cellular differentiation. DNA methylation has been shown to block the recruitment of zink finger CXXC proteins which then creates a repressive chromatin environment [107][110].

Additionally, methylated CpG dinucleotides provide binding sites for methyl CpG-binding domain proteins (MBDs) that interact with further co-regulators like histone deacetylase (HDAC) eventually leading to inhibition of gene expression [116]. A prominent mediator between DNA and histone modification is the DNMT3A/B homolog DNMT3L. DNMT3L binds to histone H3, and thereby recruits *de novo* methyltransferases to DNA. Once H3K4 becomes methylated, the interaction between DNMT3L and the nucleosome is inhibited [117]. Histone methyltransferases responsible for trimethylation of H3K9 are simultaneously required for the recruitment of DNMT3A and DNMT3B in order to methylate CpG dinucleotides, eventually leading to heterochromatinization at satellite sequences [50]. This process of heterochromatinization is initiated by a Dicer-mediated mechanism that recognizes RNA duplexes found at satellite sequences. The resulting RNA-induced silencing complex (RISC) is then specifically targeted back to pericentromeric regions where it probably recruits enzymes involved in this heterochromatin pathway [118][119][120].

Apparently, the interactions between histone and DNA modifying events can work in both directions: CpG methylation provides the template for some histone modifications, and histone modifications can recruit DNMTs. It seems that histone modifications provide more labile transcriptional repression, whereas DNA methylation is a rather stable epigenetic mark that is not easily reversed [49].

The mechanisms mentioned above are just a small insight into the many pathways that are directed by unmethylated or methylated CpG dinucleotides, respectively. Their extensive implications in epigenetic mechanisms underpin their role as a key player in transcriptional regulation. Despite recent advances in the understanding of regulatory CpG elements, there are still many gaps in the knowledge of this field that need to be filled to better understand cellular responses to the environment. Further to that, the understanding of CpG-mediated transcriptional control would be useful in the design of optimized transgene expression systems.

3.5 Transgene expression

The design of optimized transgene systems is crucial for gene therapy applications and the production of recombinant proteins. Prokaryotic and simple eukaryotic expression systems are inexpensive, fast growing and easy to handle. Nevertheless, these systems lack a suitable native glycosylation machinery and may not fold and secrete the recombinant proteins correctly [121][122]. Due to these limitations, mammalian cell culture has become the standard system for recombinant protein production. Accordingly, about 60–70% of all recombinant pharmaceuticals are produced in mammalian cells, particularly CHO and HEK 293 cells [123]. The growing demand for therapeutic proteins requires the establishment of highly effective and sustainable expression systems. Besides optimization of the translational or secretory capacity of host cells, the maximization of transgene expression levels is a major attempt to increase protein yields [124]. The first step of successful transgene expression in the target cell is the choice of the appropriate gene delivery system. There are currently two major delivery categories used for transgene expression: plasmid-based and viral vector-based [125] (Figure 6).

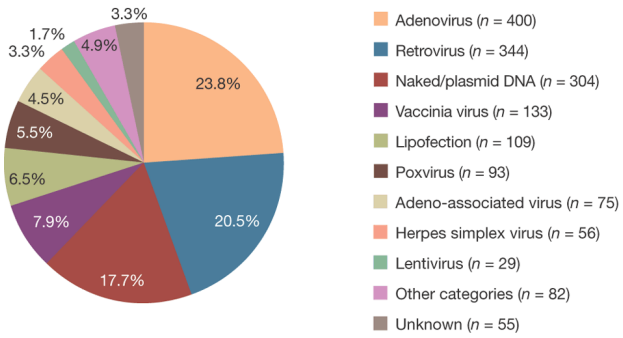

Figure 6 I Vectors used in gene therapy trials. Viral vectors, in particular retro- and adenoviruses, are the most frequently used vehicles for gene transfer to human cells. The development of efficient expression systems has made plasmid-based transgene delivery to the third most frequently used vector system in gene therapy trials [125].

3.5.1 Viral vector-based transgene expression

Viral vectors are mostly genetically modified, replication deficient viruses. They are able to transduce cells with high delivery efficiency and can be used in a variety of cells [125]. While DNA-based viral vectors, such as adeno and adeno-associated viruses (AAV), usually persist as episomal DNA in the host cell [126], retroviruses have the ability to confer long-term transgene expression through gene integration [127].

3.5.1.1 Retroviral vectors

Retroviral vectors are generated by exchanging replication elements by the gene of interest. Necessary *cis*-acting RNA regions, primarily the long terminal repeat (LTR), which is necessary for packaging, reverse transcription, integration and transcription regulation, are retained. All viral genes are usually deleted from the viral vector. The production of attenuated retroviral vectors takes place in packaging cells that provide all essential viral proteins in *trans*. Transgenes are delivered into the cell by receptor mediated fusion of viral and host cell lipid membrane. Upon entrance of the viral vector into the cell, reverse transcription is initiated. The viral genome is converted to a double-stranded DNA provirus, which is then inserted into the host genome [127].

One subclass of retroviruses often used in gene therapy trials comprises lentiviral vectors. In addition to the three essential gag, pol and env gene products, lentiviruses contain accessory viral proteins that regulate viral gene expression and infectivity [128]. These viral proteins interact with the nuclear import machinery to mediate the active transport of the viral preintegration complex through the nucleopore. This ability enables lentiviruses to transduce non-dividing cells [129].

Lentiviruses preferably integrate into or in the proximity of active transcription units [130]. Self-inactivating retroviral vectors (SIN LVs) have a deleted U3 region of the 3'LTR containing the viral enhancer sequence. This ability provides gene transfer with higher safety due to the reduced risk of enhancer-mediated mutagenesis [131]. Transgene expression in LVs has been shown to undergo epigenetic modifications, eventually leading to gene silencing [132][133][134].

3.5.1.2 Ubiquitously acting chromatin opening elements (UCOEs)

An attractive approach to overcome transgene silencing in LVs is the introduction of ubiquitously acting chromatin opening elements (UCOEs). UCOEs are regions containing CpG islands extending over dual divergently transcribed promoters derived from housekeeping gene loci [135][136]. UCOEs have been reported to provide stable transgene expression in cell culture systems even when integrated into heterochromatin regions [135]. This feature confers considerable utility for gene therapy and recombinant therapeutic applications.

3.5.2 Plasmid-based transgene expression

Alternative to virus-based delivery systems, which still bear several safety risks, plasmid-based gene delivery has become a common technique in gene therapy, DNA vaccination and the production of recombinant proteins in mammalian cells [137]. Plasmid DNA can be delivered to cells either physically or by synthetic particles. These particles typically consist of DNA complexed with cationic lipids, peptides or polymers capable of efficient gene transfer into the target cell. The easiest physical method of transgene delivery is by needle injection into the target tissue, i.e. muscle cells [138], skin [139], liver [140] or tumor [141]. Needle injection is the major application of DNA vaccination [142]. Other physical methods include electroporation [143], ballistic DNA administration [144] or sonoporation [145][146], just to name the most commonly used physical techniques. For review, see Kamimura et al [147].

Among the synthetic compounds, liposomes, particularly those composed of cationic lipids, have been reported to be most effective for gene delivery [148]. Liposomes are particles consisting of lipid bilayers encompassing an aqueous compartment. They are formed spontaneously when lipids are hydrated in an aqueous solution [147].

Alternative to liposomes, numerous polymer-based compounds such as polyethylenimine (PEI) [149], polyamidoamine [150], polyallylamine [151] and chitosan [152] are being widely employed today. These cationic polymers condense DNA into positively charged particles and prevent DNA from degradation. The cellular uptake of these complexes occurs via endocytosis [147].

Besides simplicity of delivery, the advantages of plasmid-based transgene expression are low toxicity and sustainability. The main disadvantage of plasmid-based techniques compared to viral-based methods is the low gene delivery efficiency. Large efforts have been made to modify the carrier or delivery vehicle to achieve higher transfection rates [137]. High transfection rates are however useless if transgene expression is ineffective. Once inside the cell, plasmid DNA is subjected to the cells regulation mechanisms that can directly be influenced by sequence elements of the plasmid DNA [137].

Plasmid-based vectors have a large capacity for transgene DNA. Rational plasmid design aims for the manipulation of a variety of regulatory factors that impact on gene transfer and gene expression. A plasmid accommodates the expression cassette (EC), which contains the gene(s) of interest and any regulatory sequences required for expression in mammalian cells, such as the promoter and the poly A site. The rest of the plasmid, the bacterial backbone (BB), usually contains an antibiotic resistance gene and an origin of replication required for the production of the plasmid DNA in bacteria [153]. Numerous efforts have been made to establish systems providing efficient plasmid-based transgene expression. One approach to improve transgene expression is to generate minicircles. In minicircles, the BB is removed by site-specific recognition sequences, which results in the generation of two smaller supercoiled minicircles. The minicircle harboring the EC is then separated from the other circle containing unwanted BB elements [154] such as antibiotic resistant genes or elements provoking DNA methylation and heterochromatin-associated histone modifications [137]. Another

strategy to avoid transgene silencing is the inclusion of a scaffold matrix attachment region (S/MAR). S/MARs are AT-rich sequences derived from eukaryotic DNA where the nuclear matrix attaches. They have been shown to contain DNA-unwinding elements and binding sites for transcription factors and topoisomerase II. Since S/MARs harbor mammalian origins of replication, they can promote sustainable episomal replication and maintenance in mammalian cells [155]. Another crucial factor for successful transgene expression is the careful choice of an appropriate promoter. Dependent on the type of application and target cell or tissue, different promoters should be selected. Endogenous housekeeping promoters express at low but constitutive rates. Due to this ability, they are recently preferred over viral promoters that provide high but often unstable transgene expression due to gene silencing [137]. Furthermore, a tissue-specific promoter has the potential of improved specificity and safety [156][157].

The adaptation of the codon usage has proven to be extremely effective in promoting transgene expression [158][159][160]. According to the codon bias of the host cell, the respective protein sequence is translated back into the DNA sequence, selecting only the most frequently used tRNAs of the respective organism. The use of plasmids free of CpG dinucleotides has been reported to minimize inflammation and provide prolonged transgene expression [161]. On the other hand, CpG dinucleotides in the EC have conversely been demonstrated to provide improved transgene expression in mouse tissue [162].

3.5.2.1 Applications of plasmid-based transgene technologies

Optimizing plasmid DNA not only promotes gene therapy applications. It also benefits plasmid DNA vaccination strategies [163] and transfection of mammalian cells providing for recombinant protein production [164]. Conventionally, transient expression or random integration techniques are used for recombinant protein expression. These approaches however usually result in random integration and irreproducible levels of gene expression. To overcome these problems, stable integration systems have been developed that generate stable mammalian cell lines with defined integration sites and reproducible level of protein expression [165]. The Flp-In recombinase system which is based on the site-specific recombinase (Flp) from *Saccharomyces cerevisiae* offers a single targeted integration site, has been used for applications like the production of antibodies [166][167] or vaccine immunogens [168]. Initially, this site specific integration system was developed for basic research to study and compare transcriptional reporter gene activities as it allows the expression of numerous reporter gene constructs at an identical genomic location [165]. It is therefore a useful tool to investigate the impact of regulatory plasmid vector elements on transgene expression in the host cell.

There are many different plasmid DNA modification approaches to enhance transgene expression. However, to systematically generate improved expression vectors, the complex regulation of transgenes within the host cell has to be unraveled. A substantiated knowledge of epigenetic control, chromatin dynamics, DNA binding effectors and the contribution of sequence elements is essential to gain a more comprehensive picture of transgene regulation in eukaryotic cells.

3.6 Overview of preceding CpG studies

Previous studies in our research group demonstrated a direct influence of intragenic CpG frequency on gene expression [169][170]. With the use of selected reporter genes, a recurring effect has been observed: The depletion of intragenic CpG content results in repressed gene expression, whereas the augmentation of intragenic CpG dinucleotides increases gene expression.

3.6.1 The model genes *hgfp* and *mmip-1α*

The green fluorescent protein (GFP), originating from the jellyfish *Aequorea Victoria,* has the ability to emit fluorescence. This feature makes GFP a popular marker for gene expression. A synthetic version of *gfp* has been adapted to human codon usage, denoted as *humanized hgp* (*hgfp*) [171]. The increase of the CAI in the transgene sequence positively influences the efficiency of protein translation [160]. The CAI is a measure of directional synonymous codon usage bias of a given protein coding gene sequence in a given host organism. *hgfp* was used as a reporter gene in previous studies [169] and in the present study.

The second model gene used in this study codes for the murine macrophage inflammatory protein (mMIP-1α), which belongs to the large family of cytokines. Cytokines are small, multifunctional proteins that play critical roles in the regulation of the body's responses to diseases and infection. Among the clinical applications for cytokines are cancer immunotherapy [172], wound healing [173], allergy relief [174], animal health, [175], treatment of autoimmune disorders [176], and disease diagnosis [177]. The growing demand for human recombinant therapeutics is constantly promoting the development of enhanced expression systems. The generation of efficient expression vector systems is a major strategy towards this aim. Thus, *mmip-1α* has been chosen to serve as a model gene for previous studies [170] and the study at hand.

mmip-1α and *hgfp* have been subjected to multiple modifications with respect to codon adaptation to human cells and their intragenic CpG content. *hgfp* contains 60 CpG dinucleotides [171] and is referred to as hGFP-60 in the work at hand. On the basis of this sequence, hGFP-0, lacking intragenic CpGs, was generated (Table 3A).

Table 3A | *hgfp* variants and sequence characteristics. CpG - Amount of CpG dinucleotides. GC% - Percentage of guanine-cytosine content. TpA - Amount of TpA dinucleotides. CAI – Codon adaptation index, which indicates the deviation of a given gene sequence with respect to a reference set of genes for predicted gene expression levels, regarding codon usage [160]. Modified from [172].

Gene variant	modification	CpG	GC %	TpA	CAI
hGFP-0	depletion	0	55	15	0.93
hGFP-60	optimization	60	61	15	0.96

Table 3B | *mmip-1α* variants and sequence characteristics. CpG – Amount of CpG dinucleotides. GC% - Percentage of guanine-cytosine content. TpA - Amount of TpA dinucleotides. CAI – Codon adaptation index. Modified from [170].

Gene variant	modification	CpG	GC %	TpA	CAI
mMIP-wt	none	7	51	7	0.77
mMIP-13	optimization	13	58	7	0.96
mMIP-0	depletion	0	53	8	0.92
mMIP-42	maximization	42	63	5	0.73

The wild type sequence of *mmip-1α* was initially adapted to maximal codon quality thereby obtaining 13 CpGs, denoted as mMIP-13. On the basis of mMIP-13, the nucleotide sequence was further adapted to quantitatively deplete CpGs (mMIP-0) or maximize (mMIP-42) the intragenic CpG content within the ORF (Table 3B).

For all *hgfp* and *mmip-1α* gene variants, alternative codons were used to maintain the amino acid sequence. Throughout this optimization process cryptic splice sites, TATA-boxes and internal polyadenylation signals were avoided and neither codon distribution nor overall GC content or TpA amount were changed significantly.

Figure 7 shows a schematic depiction of the used intron-free gene variants and the CpG dinucleotide distribution within the ORF. Depending on the type of experiment and host cell, gene variants were inserted into different eukaryotic and viral expression vectors, respectively, and were controlled by various promoters, as described in the result sections below.

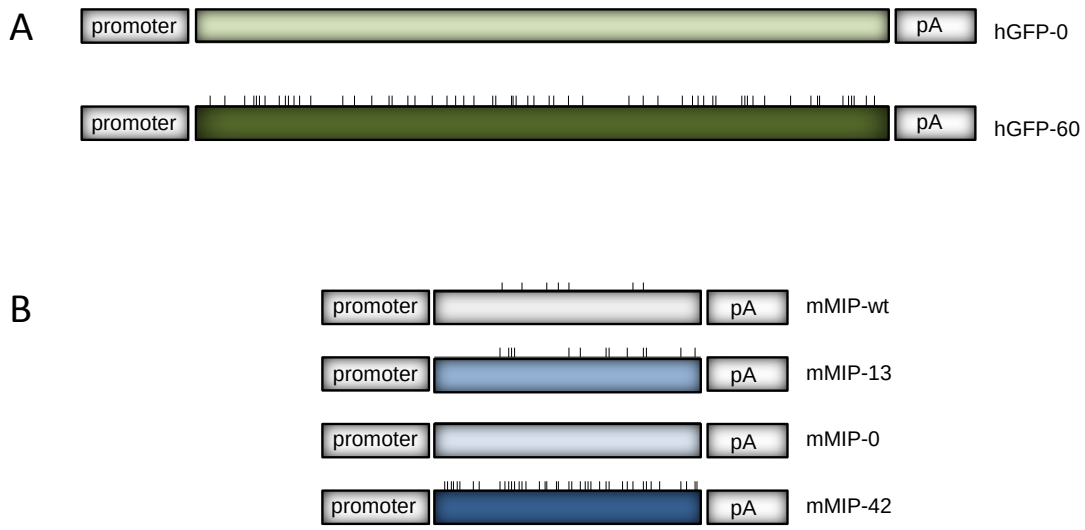

Figure 7 I Schematic depiction of the *hgfp* and *mmip-1α* gene variants as inserted into the respective expression vector. Number and distribution of CpG dinucleotides within the ORF of *hgfp* (720bp) and *mmip-1α* (279bp) of the sense strand is shown. CpG dinucleotides are indicated as vertical lines true to scale. None of the genes contains introns. The gene variants are driven by different promoters and pA signals, as described in the respective section

3.6.2 Impact of intragenic CpG content of *hgfp* and *mmip-1α* on gene expression

Based on *mmip-1α* and *hgfp* gene variants, processes underlying differential gene expression levels have been investigated. Enhanced gene expression of CpG-rich genes was shown to be irrespective of mRNA export from the nucleus, splicing activities, altered RNA stability or translational modifications. By using different promoters (CMV, EF-1α) and cell lines (H1299, HEK 293, CHO) the observed effect was proven to be not cell type- or promoter-specific. The sum of the results indicated that the mechanisms responsible for changed gene expression occur at the level of gene transcription and are triggered by unmethylated CpG dinucleotides within the ORF [170][172]. Nuclear run on experiments confirmed that CpG depletion led to decreased *de novo* synthesized mRNA levels, whereas CpG maximization clearly enhanced *de novo* mRNA rates (Figure 8).

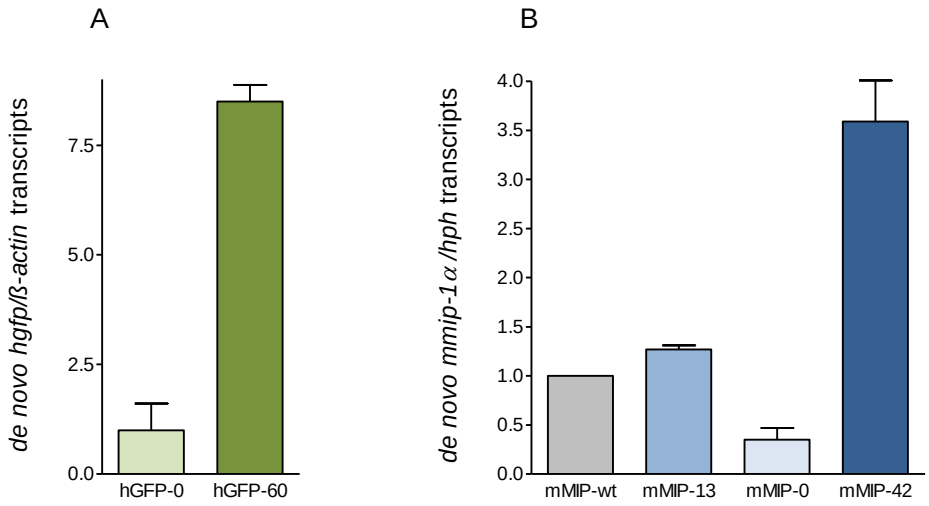

Figure 8 I Influence of CpG content on the *de novo* synthesis of *hgfp* (A) and *mmip-1α* (B) transcripts. The nuclear run-on assay was performed with stably transfected CHO Flp-In cells by supplying nuclei with biotin-16-UTP. Labelled transcripts were bound to streptavidin-coated magnetic beads and cDNA was synthesized by means of oligo-d$(T)_{15}$-primed reverse transcription of captured molecules. Absolute cDNA copy numbers obtained from newly synthesized mRNA transcripts were quantified via LightCycler and normalized to *ß-actin* (hGFP) and *hph* (mMIP-1α) transcripts, respectively. Values were normalized to *hgfp-o* and *mmip-wt*, respectively, which were set to 1. Results show the mean of 4 independent experiments each. Modified from [172] and [170] .

3.7 Aim of the study

The present study was based on the positive correlation of transcription efficiency and CpG content previously observed with CpG-modified transgenes *hgfp* and *mmip-1α* [169][170][172](see chapter 3.6). The aim of this study was to shed light on the impact of intragenic CpG content on epigenetic control of *hgfp* and *mmip-1α* variants in mammalian cells. Stable transgene integration into CHO and HEK 293 cells using the flippase recombinase (Flp-In) technique provided the basis for detailed molecular analyses. Long term hGFP expression capacities in stable CHO Flp-In cells were to be investigated with respect to intragenic CpG content, promoters and under variable growth conditions. Transgene maintenance, DNA methylation and chromatin structure were to be compared between CpG variants depending on selection pressure. Nucleosome positioning abilities among transgene variants *in vitro* should give additional insights into CpG-induced effects on chromatin dynamics. Total and actively transcribing RNA Polymerase II occupancy between CpG variants were to be correlated to transcription rates in stably transfected HEK 293 Flp-In cells. Expression analysis of transgene CpG-chimera in CHO Flp-In cells containing CpG clusters in distinct intragenic regions should reveal positional relevance of CpG dinucleotides within the ORF.

To evaluate transgene expression depending on intragenic CpG frequency in a gene therapy application suitable cell system, embryonic pluripotent stem cells of the line P19 were to be transduced with lentiviral vectors containing the respective *hgfp* transgenes with differing CpG content. Expression analyses were to be conducted in P19 cells to reveal the impact of intragenic CpG dinucleotides in this system, which displays a high potential of epigenetic activity. In addition to varying CpG content, the cytomegalovirus (CMV) immediate early promoter, the human promoter for the elongation factor 1α (EF-1α) and the ubiquitously acting chromatin opening element (UCOE) from the human *HNRPA2B1-CBX3* locus (A2UCOE) were to be compared regarding their capacity to mediate high and stable transgene expression.

4 Results

4.1 CpG-dependent differential transgene expression using mammalian Flp-In cells

To assess the impact of differential intragenic CpG content on long-term expression and regulation mechanisms, *hgfp* and *mmip-1α* gene variants were stably transfected into HEK (human embryonic kidney) 293 and chinese hamster ovary (CHO) cell lines using the flippase induced (Flp-In) recombination system [178]. CHO and HEK 293 cell lines were chosen as they are widely used for recombinant protein production [179]. The Flp-In system allows site-specific integration of a single copy transgene [165]. This system makes the established cell lines suitable for comparisons between transgene variants and enables their analysis within the same genomic environment. Homologous recombination is mediated by the flippase recombinase, which is encoded by the plasmid pOG44. The flp recombination target (FRT) is located at a defined region of the cell genome and determines the integration location.

hGFP and mMIP-1α expression was either driven by the cytomegalovirus (CMV) major immediate-early promoter or the human elongation factor 1α promoter (EF-1α), respectively. All gene variants have a Kozak sequence upstream of the start codon and are followed by the polyadenylation site of the bovine growth hormone (BGH pA), which is essential for the nuclear export, translation and stability of mRNA [180]. All expression cassettes have been inserted into the pcDNA5/FRT expression vector followed by stable Flp-In integration into HEK 293 and CHO cells, respectively.

The plasmid pcDNA5/FRT contains a hygromycin resistance gene (*hph*) lacking the ATG start codon. Therefore, hygromycin expression is not initiated until *hph* is brought in frame with the ATG codon located in the Flp-In host genome. This system allows the selection of stable transfectants in CHO and HEK 293 cells when exposed to culture medium supplemented with hygromycin B. The site-specific integration of *mmip-1α* or *hgfp* in cells was confirmed by PCR with primers flanking the respective ORF and by X-gal staining (not shown). If the gene of interest is correctly integrated, the genomic *lacZ* gene loses its functionality. Approximately four weeks after successful transfection, ß-galactosidase activity could no longer be detected in any of the transfectants. Upon stable integration, hGFP expression was analyzed by flow cytometry, and mMIP-1α production was assayed by enzyme linked immunosorbent assay (ELISA) (Figure 9).

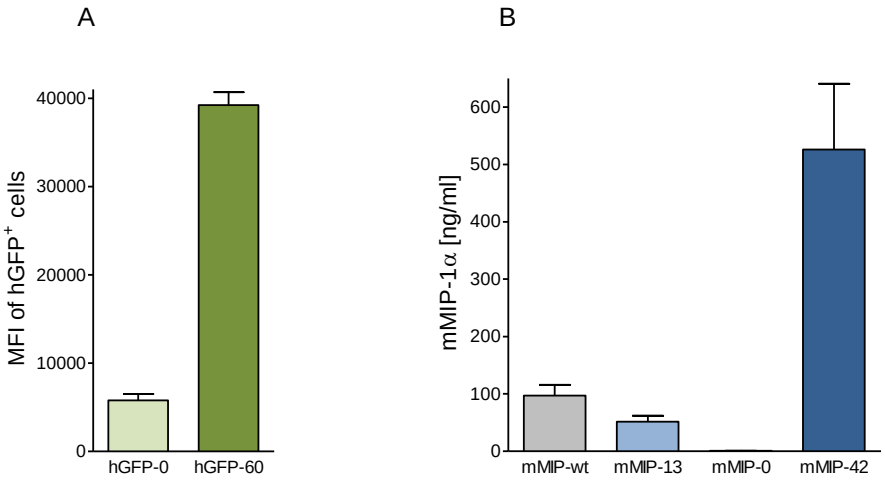

Figure 9 I Expression analyses of stably transfected CpG variants *hgfp* and *mmip-1α*. (A) hGFP expression of CHO Flp-In cells stably expressing the respective gene variant driven by the CMV promoter was assayed by flow cytometry. The mean fluorescent intensity (MFI) of hGFP positive cells (hGFP$^+$) is shown. (B) mMIP-1α concentration in the medium supernatants of HEK 293 Flp-In cells stably expressing the respective gene variant driven by the CMV promoter was measured. The charts show the mean of three measurements each; standard deviations are indicated as error bars.

In accordance with results obtained in previous studies [169][170][172], transgene expression was decreased upon CpG depletion in *hgfp* and *mmip-1α*, whereas intragenic CpG accumulation in *mmip-1α* led to a significant increase in protein levels. CpG depletion in *hgfp* resulted in a 6-fold decreased gene expression compared to the respective CpG-rich gene variant, exemplarily shown for CHO Flp-In cells and mediated by the CMV promoter in Figure 9A. For mMIP-1α, CpG depletion led to an almost complete loss of gene expression in stably transfected HEK 293 Flp-In cells, when driven by the CMV promoter (Figure 9B). By contrast, CpG maximization in mMIP-42 could achieve a more than 5-fold increase of wild type protein amount.

4.1.1 Long-term hGFP expression in the presence or absence of selection pressure

The Flp-In system used for expression studies of CpG variants ensures stable insertion of a single copy of the transgene at a specific genomic location within a chromatinized setting that resembles that of a transcriptionally active environment [165]. This allows the interaction of epigenetic mechanisms surrounding the Flp-In target region with the integrated transgene. Sustainability of transgene expression in Flp-In cell lines is usually maintained by the application of selective antibiotic pressure. To resist the antibiotic pressure, the hygromycin resistance gene (*hph*) is expressed at high rates. Since *hph* is located 2.7kb upstream of the transgene-driving promoter, the chromatin structure at the promoter and ORF of CpG variants might remain permissively open and the DNA unmethylated due to the constant and high *hph* transcription upstream. It was hypothesized that intragenic CpG dinucleotides might negatively affect expression levels upon selection pressure removal due to intragenic transgene methylation and chromatin compaction. To address this issue, CHO Flp-In cells stably transfected with *hgfp* variants were maintained either with (+ hygromycin) or without (- hygromycin) selection pressure over the course of one year (Figure 10). After one year of regular measurements, these two cell groups, each stably expressing either hGFP-0 or hGFP-60, were compared with regard to expression efficiency and in correlation to DNA methylation and chromatinized state. To determine a possible impact of promoter origin (cellular versus viral), expression capacities of CMV and EF-1α promoter-driven gene transcription were examined in parallel.

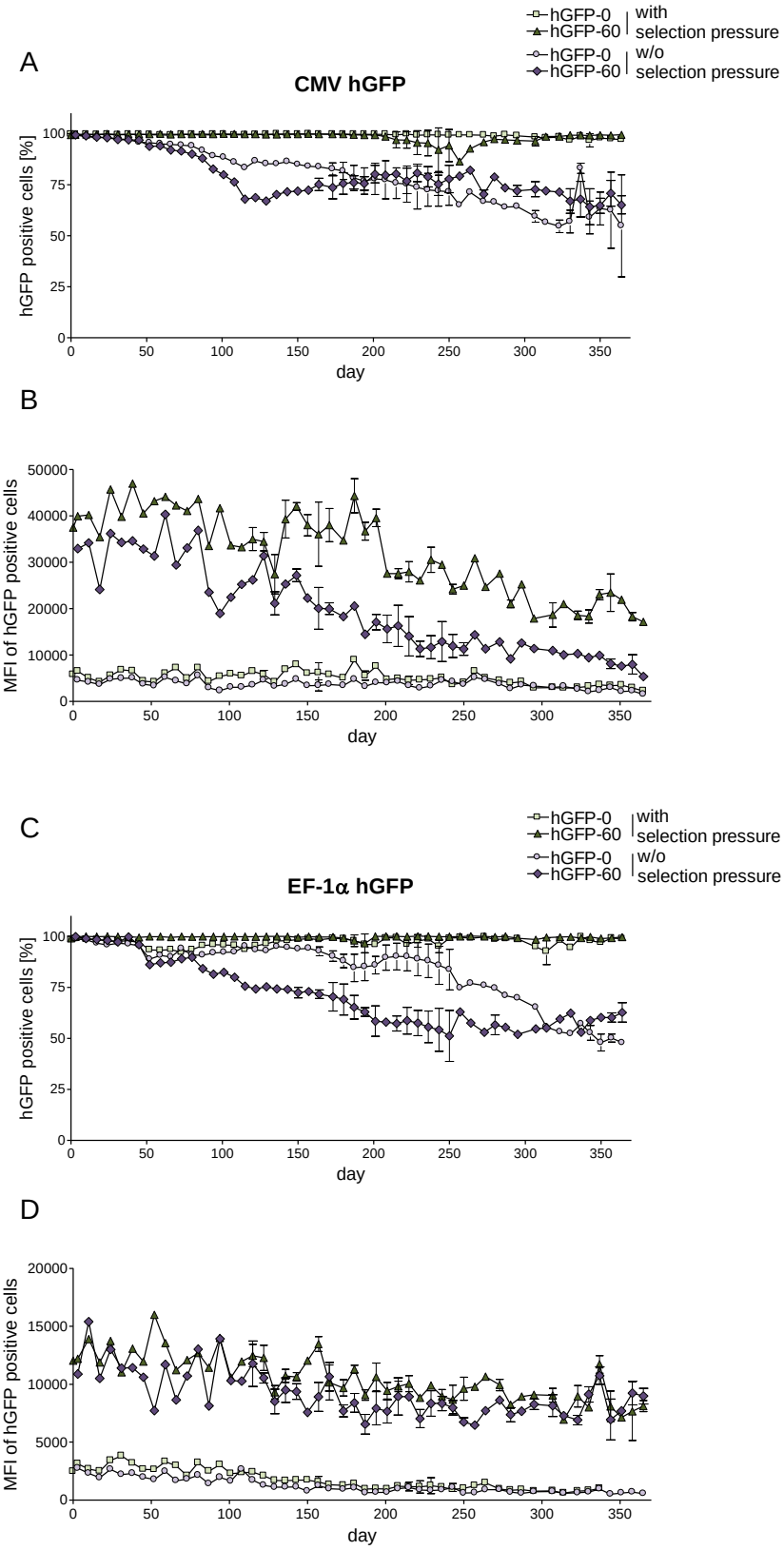

Figure 10 I hGFP long-term expression of stably transfected CHO Flp-In cells cultured with or without selection pressure as analysed by flow cytometry. The expression level of polyclonal CHO Flp-In cells stably transfected with *hgfp* variants driven by the CMV and the EF-1α promoter, respectively, with or without selection pressure by hygromycin was measured over the course of one year. The percentage of hGFP positive cells (A and C) and the MFI of hGFP positive cells (B and D) were measured weekly. The mean of two in parallel cultivated cell lines each is shown; standard deviations are indicated as error bars.

Upon abolishment of selection pressure, a slow but gradual decrease in expression efficiency could be observed in cell lines of both transgenes (Table 4). The amount of hGFP expressing CHO Flp-In cells was almost constantly 100% when cultivated under selection pressure and could even be maintained in a high percentage of CHO Flp-In cells until one year after selection pressure abolishment: at this time point, hGFP expression by hGFP-0 was still observed in 54% (CMV-promoter controlled) and 50% (EF-1α-promoter controlled) of the respective cell line, whereas expression by hGFP-60 was still observed in 67% (CMV-promoter controlled) and 64% (EF-1α-promoter controlled) of CHO Flp-In cells. The mean fluorescence intensity (MFI) of the remaining hGFP expressing cells was measured accordingly. In the presence of selection pressure, the MFI of hGFP-0 was decreased to 41% (CMV-promoter driven) and 32% (EF-1α-promoter-driven), whereas hGFP-60 expression was diminished to 46% (CMV-promoter driven) and 69% (EF-1α-promoter-driven) of the respective gene expression level at the start of the experiment. Without selection pressure, hGFP-0 expression decreased to 34% (CMV-promoter driven) and 22% (EF-1α-promoter-driven) of the initial MFI; hGFP-60 was diminished to 16% (CMV-promoter driven) and 69% (EF-1α-promoter-driven) of the initial MFI. The ratios of remaining hGFP expressing cells and MFIs are summarized in Table 4.

Table 4 I Percentage of hGFP$^+$ cells and MFI of cells after one year of cell cultivation (with or w/o selection pressure) compared to hGFP$^+$ cells and MFI at the start of the experiment.

Promoter	Selection pressure	Gene variant	% hGFP$^+$ cells	% of initial MFI
CMV	yes	hGFP-0	100%	41%
		hGFP-60	100%	46%
	no	hGFP-0	54%	34%
		hGFP-60	67%	16%
EF-1α	yes	hGFP-0	100%	32%
		hGFP-60	100%	69%
	no	hGFP-0	50%	22%
		hGFP-60	64%	69%

Surprisingly, intragenic CpG dinucleotides did not lead to accelerated gene silencing compared to CpG-lacking gene variants. Instead, the reduction of hGFP positive cells occurred even faster in hGFP-0, as reflected by the amount of remaining hGFP expressing cells. hGFP expression efficiency of hGFP positive cells, quantified as the MFI, however decreased faster in cell lines expressing hGFP-60 compared to hGFP-0 when controlled by the CMV promoter. Contrary and most notably, *hgfp* transcription driven by the EF-1α could resist gene silencing more effectively with an increased intragenic CpG content.

4.1.2 Sorting of CHO Flp-In cells according to hGFP expression levels

To shed light on the mechanism responsible for decreased gene expression over the course of time, *hgfp* variants were analyzed with regard to transgene control mechanisms. Since the loss of function was most pronounced in transgenes mediated by the CMV promoter, cells harboring transgenes driven by the EF-1α promoter were excluded from these analyses. CHO Flp-In cells expressing hGFP-0 and hGFP-60 cultivated without selection pressure were subjected to fluorescence activated cell sorting (FACS) one year after withdrawal of selection pressure. Cells of each cell line (hGFP-0 and hGFP-60) were sorted into the subpopulations "no", moderate ("mod") and maximum ("max") gene expression (Figure 11).

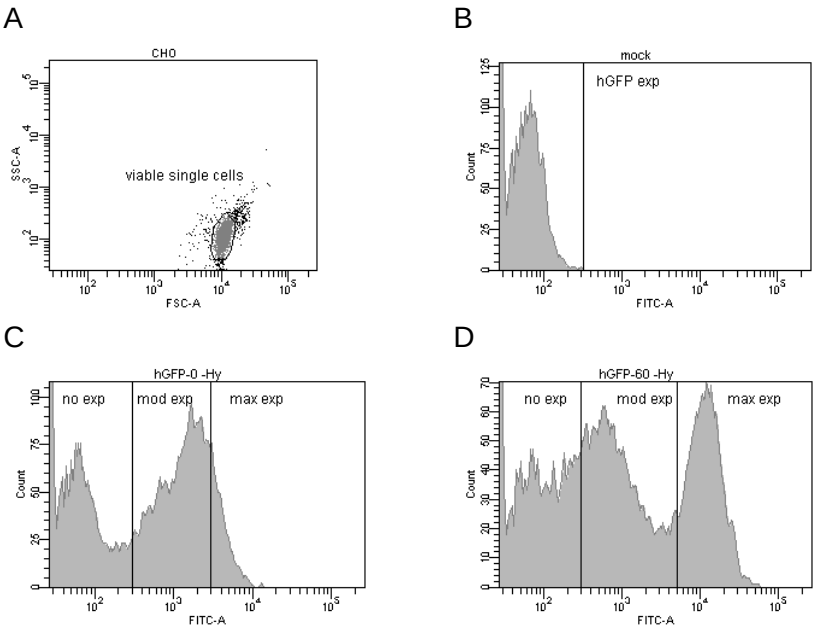

Figure 11 I Sorting of hGFP expressing CHO Flp-In cells into subpopulations according to their expression levels by FACS after one year of cell cultivation. Only viable single cells (A) were analyzed. Threshold for hGFP expressing cells was set according to the mock control (non-transfected CHO cells) (B). Cells below this threshold were regarded as hGFP negative (no expression). For hGFP-0 (C) and hGFP-60 (D), thresholds dividing moderately (mod) from maximal (max) expressing cells were set arbitrarily.

After sorting, the respective cell populations were cultivated without selection pressure for another two days to obtain enough cell material for subsequent analyses. Thereafter, genomic DNA of each of the cell-fractions was isolated for determination of *hgfp* copy numbers and investigation of the methylation status at the CMV promoter and ORF of *hgfp*.

4.1.3 Relative copy number and methylation status of *hgfp* in correlation to expression levels

Two reasons were conceivable to be responsible for the decrease in gene expression over the course of time. i) The missing selection pressure by the *hph* transcription stop led to transgene loss, either by ejection of the transfected plasmid at the Flp-In sites or by the outgrowth of cells that do not contain the transgene. ii) Alternatively, the transgenes were subjected to gene silencing via epigenetic regulation.

To address this issue, relative *hgfp* copy numbers of the respective cell populations, cultivated in the presence or absence of selection pressure, and sorted according to their expression levels, were compared. *hgfp* copy numbers of all cell populations were quantified relative to endogenous *ß-actin* by real-time PCR and normalized to hGFP-0 cultivated under selection pressure (Figure 12).

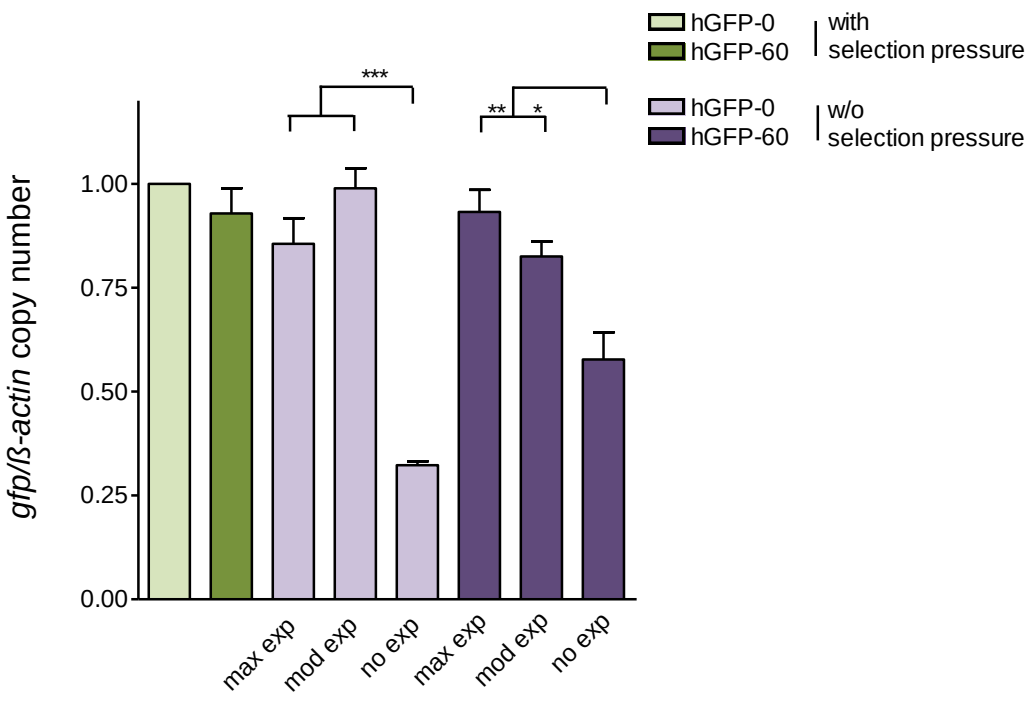

Figure 12 I Determination of *hgfp* copy numbers relative to *ß-actin* of CHO Flp-In cells stably expressing hGFP variants with or w/o selection pressure. Genomic DNA of cells sorted into the subpopulations no, moderately (mod) and maximal (max) gene expression was isolated and subjected to quantitative PCR. Primers encompassing the TSS of *hgfp* were used to determine the copy numbers of *hgfp* transgenes. All Ct values were normalized to the corresponding Ct values of *ß-actin*. hGFP-0 expressed under selection pressure was set to the value 1; the remaining gene variants were scaled accordingly. The mean and standard deviations of two DNA preparations of triplicates each are shown. Significance was calculated using ANOVA/Tukey's Multiple Comparison Test (*p<0.05; **p<0.01; ***p<0.001).

Primers encompassing the transcription start site (TSS) were used instead of the ORF to avoid different primer efficiencies due to different template sequences between *hgfp* variants. The quantification of relative *hgfp* copy numbers revealed that the removal of selection pressure did not lead to changed transgene copy numbers in moderately and maximally hGFP expressing cell populations, compared to CHO Flp-In cells maintained under selection pressure. This was shown for both hGFP-0 and hGFP-60. This result was expected as the threshold for both moderate and maximal hGFP expression was set above the fluorescence level of the mock control. By contrast, the cell fractions sorted into the category 'no expression' exhibited a decreased *hgfp* copy number in comparison to cells exhibiting higher levels of hGFP expression. While *hgfp* transgenes were retained in 58% of cells transfected with hGFP-60, only 32% of hGFP-0 transfected cells contained the *hgfp* transgene after one year of selection pressure abolishment. The fact that a subset of cells exhibited

complete deficiency of hGFP expression and yet still contained the transgene implies that the loss of function must additionally be due to epigenetic repression.

Isolated genomic DNA of cell populations used for copy number determination was simultaneously used for evaluation of the methylation state of *hgfp*. To this end, genomic DNA was subjected to bisulphite genomic sequencing following a published protocol [181]. Sodium bisulphite selectively deaminates unmethylated cytosines to uraciles, whereas methylated cytosines stay unchanged. In the subsequent PCR reaction, uraciles are replaced by thymines resulting in a C to T conversion. To validate the method, the pcDNA5 plasmid containing *hgfp-60* (phGFP-60) was subjected to quantitative *in vitro* methylation prior to bisulphite sequencing. Chromatograms were evaluated using the software Chromas. The methylation levels of CpG dinucleotides were determined by measuring the ratio of each of the cytosine peak heights to the sum of respective cytosine and thymine peak heights in automated DNA sequencing traces, according to a technique published by Jiang et al [182].

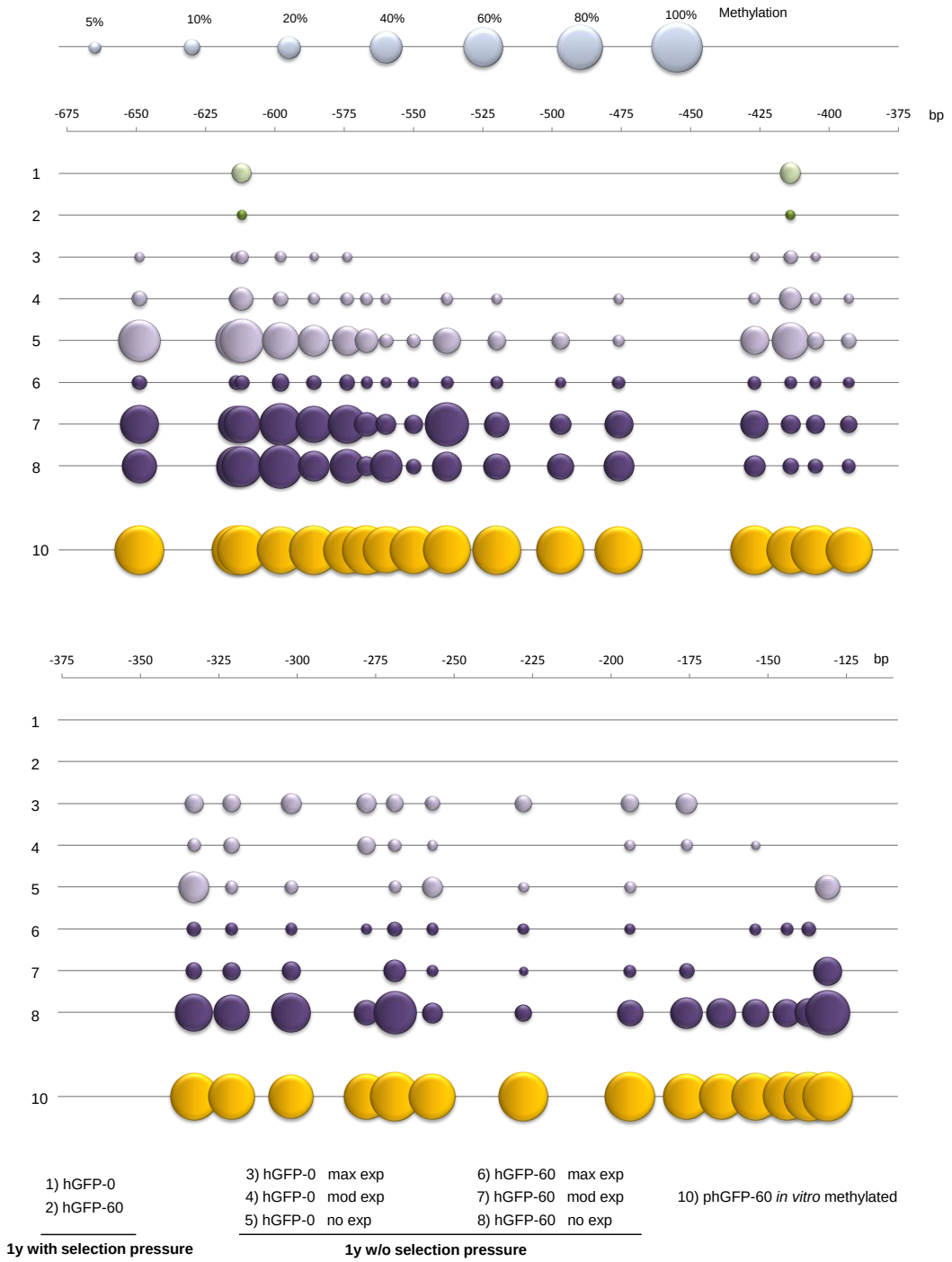

Figure 13A I Methylation levels of the CMV promoter. Genomic DNA of CHO Flp-In cells expressing hGFP variants with selection pressure and CHO-hGFP cells w/o selection pressure sorted into the fractions no, moderate (mod) and maximum (max) gene expression was isolated and subjected to bisulfite sequencing. *In vitro* methylated phGFP-6o served as a positive control. The methylation level is reflected by the size of the bubbles, as shown in the scale above the diagram. Numbers above the charts represent the distance from the *hgfp* start codon. Examined cell lines are characterized below the diagrams. The methylation level of CpGs was determined by measuring the ratio of the cytosine peak height to the sum of cytosine and thymine peak heights in automated DNA sequencing traces [182].

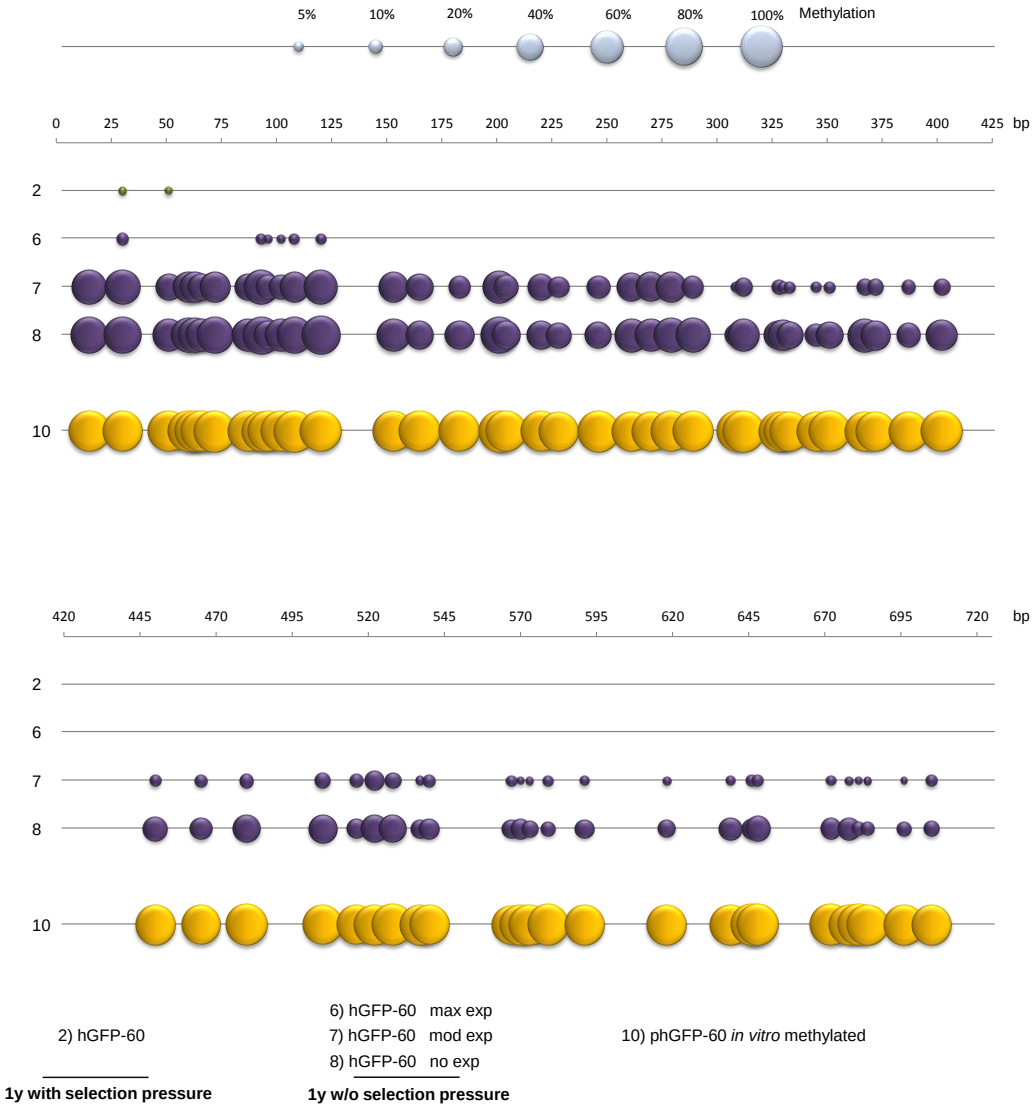

Figure 13B I Methylation levels of the *hgfp* ORF. Genomic DNA of CHO Flp-In cells expressing hGFP variants with selection pressure and CHO-hGFP cells w/o selection pressure sorted into the fractions no, moderate (mod) and maximum (max) gene expression was isolated and subjected to bisulfite sequencing. *In vitro* methylated phGFP-6o served as a positive control. The methylation level is reflected by the size of the bubbles, as shown in the scale above the diagram. Numbers above the charts represent the distance from the *hgfp* start codon. Examined cell lines are characterized below the diagrams. The methylation level of CpGs was determined by measuring the ratio of the cytosine peak height to the sum of cytosine and thymine peak heights in automated DNA sequencing traces [182].

The graphic charts of chromatogram evaluations revealed substantial differences in methylation levels between different cell populations according to their expression levels (Figure 13 A, B). While cells expressing hGFP-0 and hGFP-60 cultivated in the presence of selection pressure exhibited virtually no methylation neither in the promoter nor in the ORF (in case of hGFP-60), cells cultivated in the absence of selection pressure showed gradually increasing levels of DNA methylation both in the promoter and in the ORF. The overall methylation levels (Table 5) for the CMV promoter mediating hGFP-0 expression were 4.8% (maximum expression), 5.1% (moderate expression) and 20.2% (no expression). For the CMV promoter controlled hGFP-60 expression, the evaluation yielded 5.4% (maximum expression), 23% (moderate expression) and 35.9% (no expression) of methylated CpGs. The ORF of *hgfp* exhibited an overall methylation of 0.5% (maximum expression), 24.3% (moderate expression) and 44.7% (no expression). It has to be noted that the DNA isolation for bisulfite treatment was conducted two days after the cell sorting. Thus, the expression profile might have undergone slight changes compared to the day of DNA sorting. Cytosines of *in vitro* methylated phGFP-60 were virtually completely resistant to bisulphite treatment, resulting in 90% (CMV promoter) and 95% (ORF) of overall cytosine methylation, respectively. While the methylation levels in the promoter is highest at the borders and lower in the center, CpG methylation in the ORF of *hgfp* is highest in the 5' end and gradually decreases towards the 3' end. In *in vitro* methylated phGFP-60, CpG dinucleotides are methylated evenly.

Table 5 I Summary of methylation levels of the CMV promoter and *hgfp* ORF, displayed as bubble chart in Figure 13, and relative *hgfp* copy no as quantified by real-time PCR, with hGFP-0 cultivated under selection pressure set to the value 1 (see Figure 12).

Description of cell line			Methylation level [%]		
Selection pressure	**Gene variant**	**Relative expression level**	**CMV promoter**	**ORF of *hgfp***	**relative copy no (qPCR)**
yes	hGFP-0	low	< 1	-	1
	hGFP-60	high	< 1	< 1	0.9
no	hGFP-0	maximal (max)	4.8	-	0.9
		moderate (mod)	5.1	-	1
		no expression (no)	20.2	-	0.3
	hGFP-60	maximal (max)	5.4	0.5	0.9
		moderate (mod)	23.1	24.3	0.8
		no expression (no)	35.9	44.7	0.6
-	phGFP-60	positive control	90.2	94.8	

The decline of hGFP expression in the absence of selection pressure therefore seems to be a combination of both transgene loss and DNA methylation of the expression cassette, as can be seen from Table 5. Methylation occurred in cytosines of the promoter and the ORF. While hGFP-0 provides no methylation targets in the ORF, hGFP-60 can be methylated at both the promoter and the ORF. The ratio of methylated cytosines in the promoter was higher in hGFP-60 compared to hGFP-0. In particular the moderately hGFP expressing cell fraction displayed a more than 4-fold higher methylation rate in the CMV promoter driving hGFP-60 compared to the CMV promoter controlling hGFP-0. Hence, loss of function ("no hGFP expression") in hGFP-0 resulted only 20% from DNA methylation and approximately 70% (according to the relative copy no of 0.3; see Table 5) from transgene loss. Contrarily, loss of function by hGFP-60 was achieved by 36% methylation in the promoter and an additional 45% in the ORF. According to the relative copy no of 0.6 (Table 5), only 40% of cells have lost the transgene hGFP-60. Due to extensive silencing, cells containing hGFP-60 less readily lost the transgene compared to hGFP-0 harbouring cells.

4.1.4 Impact of intragenic CpG dinucleotides on chromatin structure

Since DNA methylation is frequently accompanied by chromatin changes, it was investigated whether the observed differences in expression efficiency are reflected by changes in chromatin density.

4.1.4.1 Chromatin structure of *hgfp* variants *in vivo*

At the time point of copy number determination and methylation analysis, stably hGFP expressing CHO Flp-In cells cultivated in the presence, respectively absence of selection pressure were subjected to FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). Out of the cell population cultivated without selection pressure, the fraction of repressed hGFP expression (denoted as "no expression") was examined. By FAIRE, chromatin is cross-linked with formaldehyde, sheared to fragments of 200-500bp by sonication, and phenol-chloroform extracted. This procedure results in preferential enrichment of nucleosome-depleted genomic regions that can be quantified by real-time PCR. The assay was performed at the transcription start site (TSS) and the ORF (position +32 to +152 relative to the start codon) of the respective gene variant. A region between the 4^{th} and 5^{th} exon of the housekeeping gene *ß-actin* served as a comparison control. All values were normalized to genes coding for rRNA (*rdna*). The fraction of FAIRE-extracted DNA was found to be generally higher at both the TSS and ORF of *hgfp* compared to the corresponding *ß-actin* control (Figure 14). This reflects the particularly open chromatin environment at the Flp-In recombination target site containing the respective CpG gene variant.

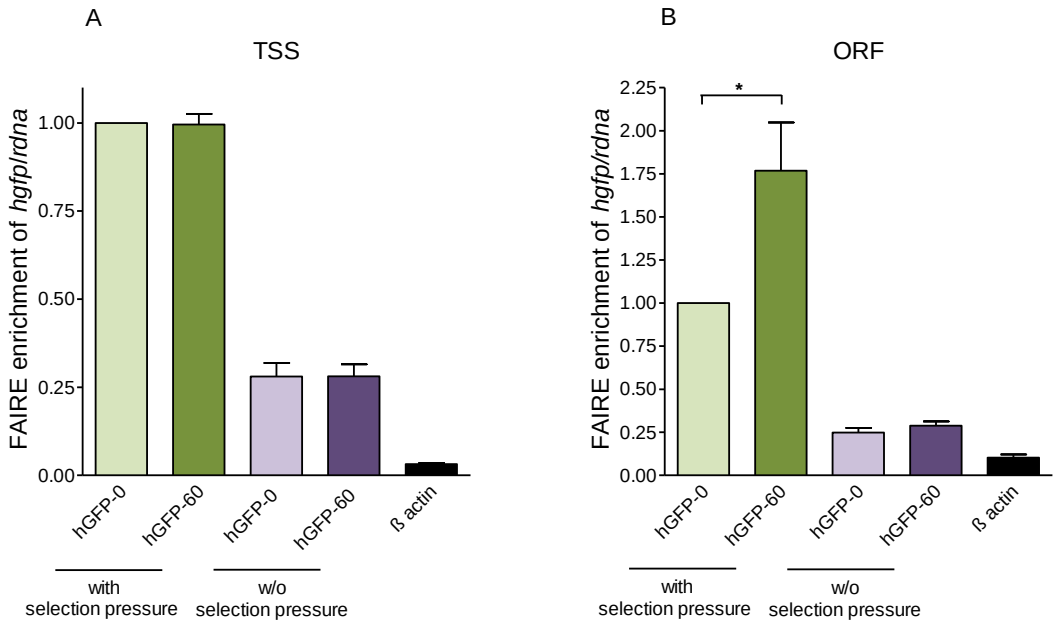

Figure 14 I Inverse chromatin densities of CHO Flp-In cells stably expressing hGFP variants *in vivo* as analyzed by FAIRE. Enrichment for nucleosome-depleted chromatin by FAIRE extraction was performed, and DNA from the aqueous phase was quantified by real-time PCR using primer pairs specific for (A) the TSS (-87 to -17 relative to the start codon) and (B) a region of the *hgfp* ORF (+32 to +152 relative to the start codon). A region between the 4^{th} and 5^{th} exon of *ß-actin* served as a control. The mean and standard deviation of *ß-actin* values of all four cell lines is shown. The values are presented as the ratio of DNA recovered from cross-linked cells divided by the amounts of the same DNA in the corresponding non-cross-linked samples. All results were normalized to *rdna* and referred to hGFP-0 cultured under selection pressure, which was set to the value 1. The data reflect the degrees of nucleosome depletion in the respective genomic regions. The mean and standard deviations of two FAIRE preparations with a duplicate each are shown. Significance was calculated using ANOVA/Tukey's Multiple Comparison Test (* $p<0.05$).The colors of the bars reflect the respective cell populations in Figure 10.

The amount of extracted nucleosome-free DNA detected at the TSS and ORF of *hgfp* clearly correlated with the presence of selection pressure. Thus, abolishment of selection pressure not only induced DNA methylation but also a significantly increased chromatin density at the TSS of hGFP-0 and hGFP-60. Quantification of isolated nucleosome-free DNA in the ORF not only revealed an association with selection pressure. It furthermore showed a significantly higher degree of chromatin density at the ORF of hGFP-0 relative to hGFP-60 (ANOVA; $p<0.05$). This relative difference was clearly visible in CHO Flp-In cells cultivated under selection pressure, but was not observed in CHO Flp-In cells in which hGFP expression was repressed due to the absence of antibiotic selection. Thus, intragenic CpG depletion led to a higher chromatin density at *hgfp* in CHO Flp-In cells growing under selective conditions, thereby supposedly impeding transcription efficiency, whereas a high intragenic CpG content maintained an open chromatin structure. Upon selection pressure withdrawal, the chromatin opening abilities of intragenic CpG dinucleotides seem to get lost.

4.1.4.2 Chromatin structure of *mmip-1α* variants *in vivo*

In order to verify the association of transcription efficiency and chromatin structure among genes differing in intragenic CpG content, FAIRE was analogously performed in HEK 293 Flp-In cells stably expressing the mMIP-1α variants mMIP-wt, mMIP-13, mMIP-0 and mMIP-42 (Figure 15). Due to the highly divergent DNA sequences among the variants, primers were used binding to regions in direct proximity to the ORF, to avoid a bias by different primer efficiencies. Hence, the transcription start site (TSS) and the 3'untranslated region (3'UTR) immediately downstream of the respective *mmip-1α* variant were examined. The second exon-intron junction of the housekeeping gene *ß2-microglobulin (ß2-m)* was used as an endogenous control and all values were normalized to *rdna*. In all of the cell lines, a large amount of FAIRE-extracted DNA was observed from both the TSS and the 3'UTR of *mmip-1α* compared to the constitutively expressed *ß2-m*. This indicates a very open chromatin structure at the *mmip-1α* locus of all gene variants, as was already shown by CHO Flp-In *hgfp* variants. While mMIP-wt, mMIP-13 and mMIP-42 exhibited very similar levels of nucleosome density, mMIP-0 revealed a significantly denser nucleosome occupancy at the TSS and 3' UTR compared to the rest of mMIP-1α variants (ANOVA; $p<0.05$). In accordance with *hgfp* FAIRE analyses, increased chromatinization correlated with transcription loss resulting from CpG depletion. However, no correlation could be detected between transcription efficiency and chromatin density between the wild type and CpG-maximized mMIP-42. It is assumed that the additional accumulation of CpGs within the ORF of mMIP-42 did not lead to lower chromatin density due to saturation effects of the already very open chromatin structure at the Flp-In recombination locus.

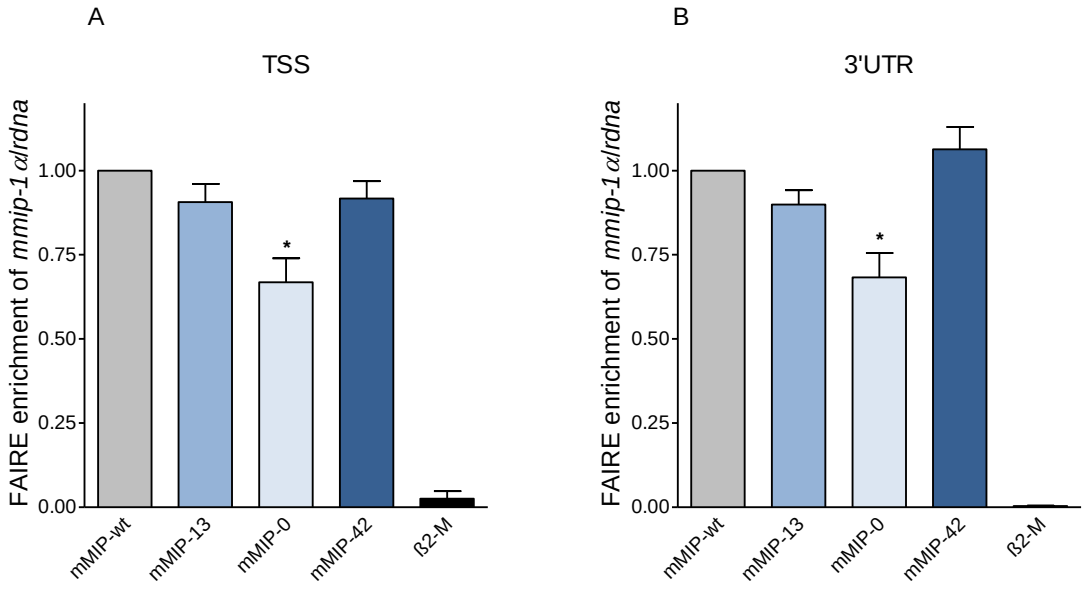

Figure 15 I Inverse chromatin densities of HEK 293 Flp-In cells stably expressing mMIP-1α variants *in vivo* as analyzed by FAIRE. Enrichment for nucleosome-depleted chromatin by FAIRE extraction was performed, and DNA from the aqueous phase was quantified by real-time PCR using primer pairs specific for (A) the TSS (-87 to -17 relative to the start codon) and (B) a region 10bp to 97bp downstream of the ORF stop codon (3' UTR). The 2nd exon-intron junction of *ß2-microglobulin (ß2-m)* served as a control. The mean and standard deviation of *ß2-m* values of all four cell lines is shown. The values are presented as the ratio of DNA recovered from cross-linked cells divided by the amounts of the same DNA in the corresponding non-cross-linked samples. All results were normalized to *rdna* and referred to mMIP-wt, which was set to the value 1. The data reflect the degrees of nucleosome depletion in the respective genomic regions. The mean and standard deviations of two FAIRE preparations with a duplicate each are shown. Significance was calculated using ANOVA/ Tukey's Multiple Comparison Test (* p<0.05).

4.1.4.3 H3K4me3 occupation of *mmip-1α* variants

The correlation between intragenic CpG depletion and increased chromatin density raised the question of which histones or histone modifications might be involved in the observed chromatin changes. The histone modification of H3 tri-methylated at lysine 4 (H3K4me3) is strongly and preferentially associated with transcribed regions of active genes [183]. It was further shown to co-localize with CpG islands and regulators of active gene transcription, such as the CpG-binding protein CFP1 [107]. Due to these features, Chromatin Immunoprecipitation (ChIP) was used to examine whether H3K4 was selectively trimethylated at CpG enriched regions. Occupation of histone H3, representing one of the five basal histones of the nucleosome served as a control. Chromatin was cross-linked with formaldehyde, sheared to fragments of 300-800bp by sonication and the DNA fragments were precipitated with the

according antibodies (see Material and Methods). Quantitative PCR following ChIP was conducted at the CMV promoter and 3'UTR of HEK 293 Flp-In cells stably expressing the respective mMIP-1α variants. The first exon-intron-junction of *gapdh* was used as a control. The results are presented as relative output-to-input, *mmip-1α* to *gapdh* (H3) ratios, or H3K4me3/H3, respectively.

Figure 16 shows that all gene variants were less occupied by H3 and H3K4me3 at the 3'UTR and CMV promoter region than the control region of the housekeeping gene *gapdh*. This observation affirms the high level of nucleosome depletion of all gene variants already detected by FAIRE. In contrast to FAIRE analyses, the differences of chromatin density as reflected by H3 occupancy were not significantly different between the *mmip-1α* variants. The same situation applied for H3K4me3 precipitates. One reason for these discrepancies between chromatinization levels detected by FAIRE and ChIP might be due to the higher sensitivity of the FAIRE assay. During ChIP performance optimization, a fragment size not smaller than 300-800bp of sonicated genomic DNA was determined to be required for successful DNA precipitation by ChIP. In contrast, fragment lengths of only 200-500bp was sufficient to obtain enough template for FAIRE analyses. Due to the short length of the CMV promoter (588bp) and *mmip-1α* (279bp), ChIP fragments might have been too large to reflect the actual H3-binding at the promoter and 3'UTR, but rather the H3-binding of proximal regions, which are most likely more intensively occupied by histones.

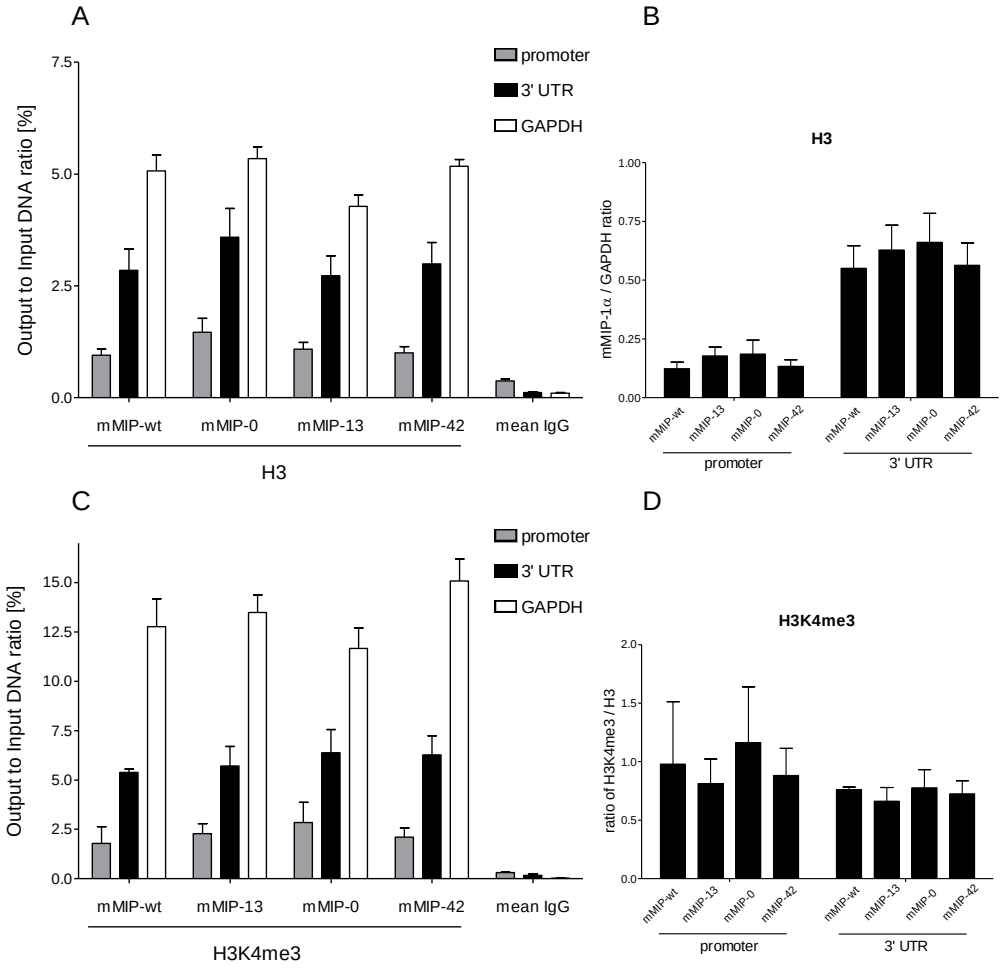

Figure 16 I ChIP analysis of H3 (A,B) and H3K4me3 (C,D) at the promoter and 3'UTR of mMIP-wt, mMIP-13, mMIP-0 and mMIP-42. ChIP was performed by cross-linking of HEK 293 Flp-In cells stably expressing mMIP-1α variants, sonication, incubation with the appropriate antibody and isolation of bound DNA by sepharose A beads. After precipitation DNA was quantified by real-time PCR using primer pairs specific for the CMV promoter and the 3'UTR immediately downstream of the stop codon. The first exon-intron-junction of *gapdh* was used as internal control. Results for the promoter and 3'UTR were either expressed as input to output ratio (A;C) or normalized to *gapdh* (C). H3K4me3 values were normalized to H3 (D). Normal polyclonal rabbit IgG served as a negative control. The mean-IgG levels of all cell lines were either expressed as input to output ratio of the respective gene locus (A;C) or subtracted from the ChIP results of the corresponding precipitated protein (B;D). The mean and standard deviation of three independent experiments is shown.

4.1.4.4 Influence of CpG dinucleotides in *hgfp* on nucleosome positioning *in vitro*

The observation of changed chromatin density between CpG variants detected by FAIRE led to the question whether the positioning of nucleosomes was affected by CpG dinucleotides as well. Sequence patterns can directly affect nucleosome positioning by determining biophysical properties of DNA like the bending flexibility around a histone octamere [184]. It was therefore hypothesized that sequence modifications in the ORF act via an altered nucleosome binding to change transcription performance. To analyze impaired DNA-histone interactions among CpG variants, *in vitro* nucleosome reconstitution assays were performed.

Mononucleosomes were formed mixing histone octamers and PCR fragments spanning distinct regions of the ORF of hGFP-0 and hGFP-60, respectively, by salt dialysis [185]. Nucleosome positions were resolved by native polyacrylamid gel electrophoresis (PAGE) and detected by ethidium bromide staining followed by ultraviolet exposure. The ORF of *hgfp* comprises 720bp. Due to its length, a fragment comprising the entire gene would form polynucleosomes when assembled with histone octamers that cannot be resolved by native PAGE. Hence, *hgfp* was partitioned into the three fractions I, II and III with approximately the same CpG content (22, 21 and 24 CpGs), which were generated by PCR. Primers were designed to amplify overlapping fragments of 280 to 300bp within the ORF of *hgfp* (Figure 17).

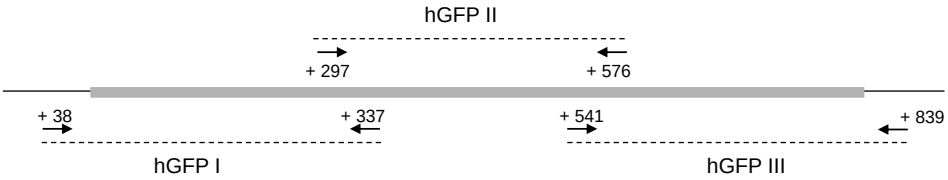

Figure 17 I Amplification of hGFP fragments I, II and III. Three fragments of similar length, hGFPI (300bp), hGFPII (280bp) and hGFPIII (299bp), indicated as dashed lines, were amplified. The ORF of *hgfp* is represented by a grey bar. Primers are marked as arrows. Sequence positions are indicated relative to the TSS [186].

The position of a histone octamere within the DNA fragment affects its electrophoretic mobility: centrally located nucleosomes migrate slower than nucleosomes located at the end of a DNA fragment. Once the optimal histone: DNA ratio was established by a test assembly (not shown) nucleosome reconstitutions were performed with each of the DNA fragments. Comparative analysis of mononucleosome band patterns revealed different positioning preferences among the gene variants (Figure 18).

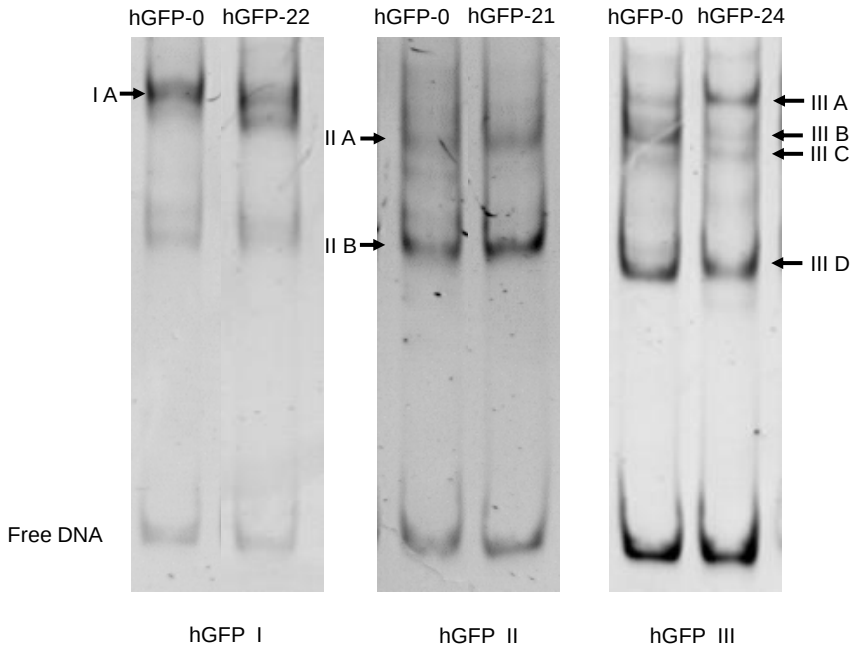

Figure 18 I Nucleosome positions among variant hGFP fragments. hGFP fragments I, II and III (see Figure 17) were reconstituted into nucleosomes, followed by native PAGE, ethidium bromide staining and UV detection. The CpG frequency of each fragment is denoted in the fragments annotation above each PAGE illustration. Bands representing nucleosome positions that are specified in the text below are indicated as black arrows. One representative set out of two reconstitutions is shown [186].

Both hGFP I variants preferably bound histones in a central region (IA). The distinctiveness of nucleosome binding is more defined in hGFP-0 than in hGFP-22. hGFP II variants were mostly occupied by histones at the 5' or 3' end of the fragments (IIB), whereas central regions were bound very unspecifically (IIA). Despite favoring histone binding within the same regions (IIIA-D), the preference for certain histone locations seems to vary between the two fragments hGFP-0 III and hGFP-24 as judged by the respective band intensities. The results clearly demonstrate that CpG-variations in hGFP directly affect nucleosome-positioning abilities *in vitro*.

4.1.5 Influence of intragenic CpG dinucleotides on RNAPII occupation

The observation that intragenic CpG dinucleotides alter chromatin structure in a manner that apparently correlates with gene transcription led to the question whether these changes coincide with a changed RNA Polymerase II (RNAPII) occupancy and transcription rate. Due to the broader spectrum of CpG frequencies among *mmip-1α* compared to *hgfp* variants, *mmip-1α* transgenes were used as model for this experiment. The C-terminal domain (CTD) of RNAPII becomes multiply

phosphorylated upon initiation. Phosphorylation of the serin-2 residue occurs when RNAPII is associated with the coding region and has been implicated in productive elongation and the 3′-end processing of the transcript [19]. Commercially available antibodies directed against different CTD phosphorylation states therefore allow distinguishing between certain stages of the transcription cycle.

To address whether transcriptional changes observed in CpG variants correlate with altered RNAPII-binding, HEK 293 Flp-In cells harboring the respective *mmip-1α* variants were subjected to ChIP. Chromatin was cross-linked with formaldehyde, sheared to fragments of 300-800bp by sonication and the DNA fragments were precipitated with the according antibodies (see Material and Methods). Quantitative PCR following ChIP was conducted at the CMV promoter and 3'UTR. The first exon-intron-junction of *gapdh* was used as a control. The binding of total and transcriptionally active RNAPII was examined using antibodies raised against the N-terminus of RNAPII and the CTD of RNAPII, phosphorylated at serine-2 (Ser2P Pol II), respectively. The results are presented as relative output-to-input ratios and *mmip-1α* to *gapdh* ratios, respectively.

The evaluation of absolute RNAPII-bound values revealed an increased amount of RNAPII precipitated by *mmip-1α* transgenes compared to RNAPII bound at *gapdh* (Figure 19 A, C). This discrepancy is even more pronounced in Ser2 RNAPII-precipitates. This result repeatedly confirms that the transgenes are situated in the transcriptionally active region of the recombination site.

ChIP results normalized to endogenous *gapdh* demonstrate that total RNAPII-binding at the promoter is not significantly changed between mMIP-wt, mMIP-13, mMIP-0 and mMIP-42 (Figure 19 B). At the 3'UTR, a trend of decreased RNAPII at mMIP-0 and increased RNAPII at mMIP-42 was observed compared to mMIP-wt and mMIP-13. To evaluate the fraction of bound Polymerase II that is actively engaged in elongation, relative quantification of Ser2P RNAPII at the promoter and 3'UTR was examined. In accordance with total RNAPII, similar amounts of Ser2P Pol II were detected at the promoter between mMIP-wt, mMIP-13 mMIP-0 and mMIP-42 and at the 3'UTR between mMIP-wt, mMIP-13 and mMIP-0 when normalized to *gapdh* (Figure 19D). In mMIP-42 however, a trend of increased Ser2P RNAPII occupancy could be observed. The correlation of increased Ser2P RNAPII occupancy at the 3'UTR with increased mRNA transcripts in mMIP-42 implies that mMIP-42 exhibits a higher elongation rate than the CpG-reduced/-lacking gene variants.

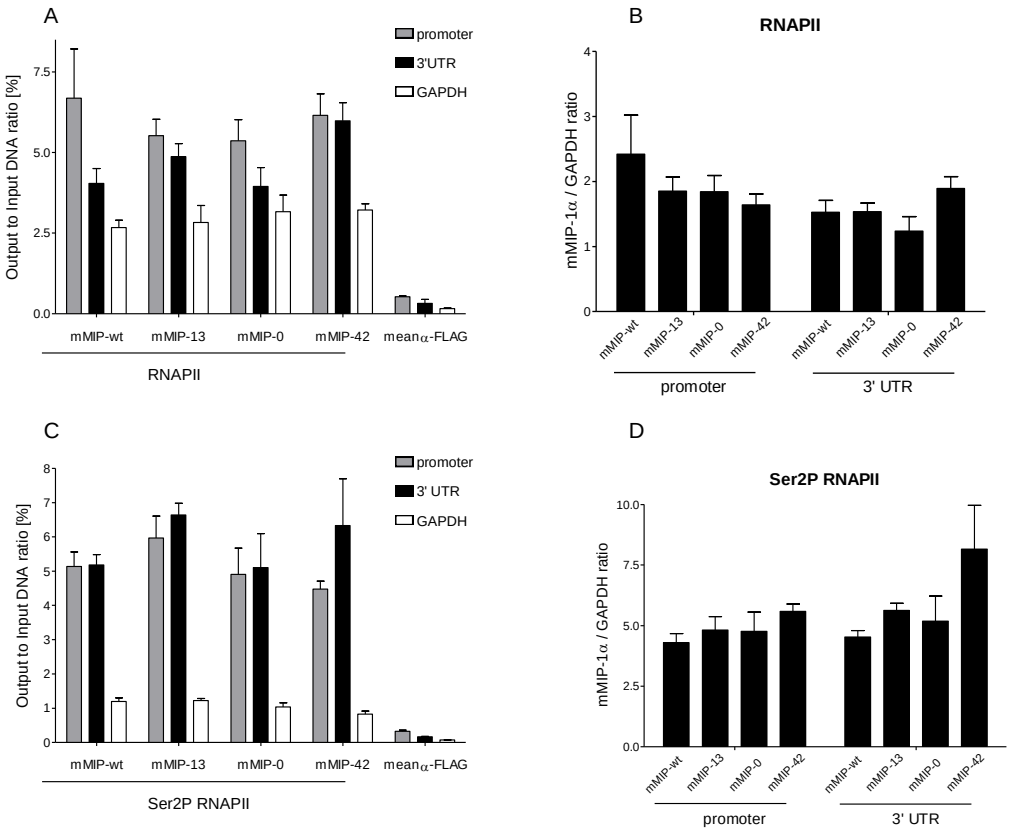

Figure 19 I ChIP analysis of Pol II (A,B) and Ser2P RNAPII (C,D) at the promoter and 3'UTR of mMIP-1α variants. ChIP was performed by cross-linking of HEK 293 Flp-In cells stably expressing mMIP-1α variants, sonication, incubation with the appropriate antibody and isolation of DNA loaded antibodies by sepharose A beads. After precipitation, DNA was quantified by real-time PCR using primer pairs specific for the CMV promoter and 3'UTR. The first exon-intron-junction of *gapdh* was used as internal control. Results for the promoter and 3'UTR were expressed as input to output ratio (A,C) or normalized to *gapdh* (B,D). Polyclonal rabbit anti-FLAG antibody served as a negative control. The mean-IgG levels of all cell lines were either expressed as input to output ratio of the respective gene locus (A;C) or subtracted from the ChIP results of the corresponding precipitated protein (B;D). The mean and standard deviation of three independent experiments and duplicates each is shown.

4.1.6 Impact of intragenic CpG distribution on gene expression in *hgfp*

The ability of CpG dinucleotides to enhance elongation rates of *mmip-1α* led to the question whether a specific motif or region within the ORF of CpG variants was responsible for this effect. To shed light on the positional relevance of CpGs within the transgene ORF, chimeras were generated by fusion PCR to create genes with CpG clusters in distinct intragenic 5' and 3' regions of the ORF, respectively. *hgfp* was preferred over *mmip-1α* fragments for this experiment, since their ORF is 2.5 fold longer than *mmip-1α*. Thus, the position effect was assumed to be more pronounced.

Based on hGFP-0 and hGFP-60, chimera illustrated in Figure 20, were created. Expression cassettes containing the gene-chimera were stably transfected into CHO Flp-In cells and expression levels were measured by flow cytometry exactly as was done in previous expression analyses. The protein levels observed generally correlated with the amount of intragenic CpG content (Figure 20, right panel). An exception to this trend was the chimera containing only 13 CpGs in the 5' region. This variant conferred a higher gene expression than would be expected by its CpG frequency.

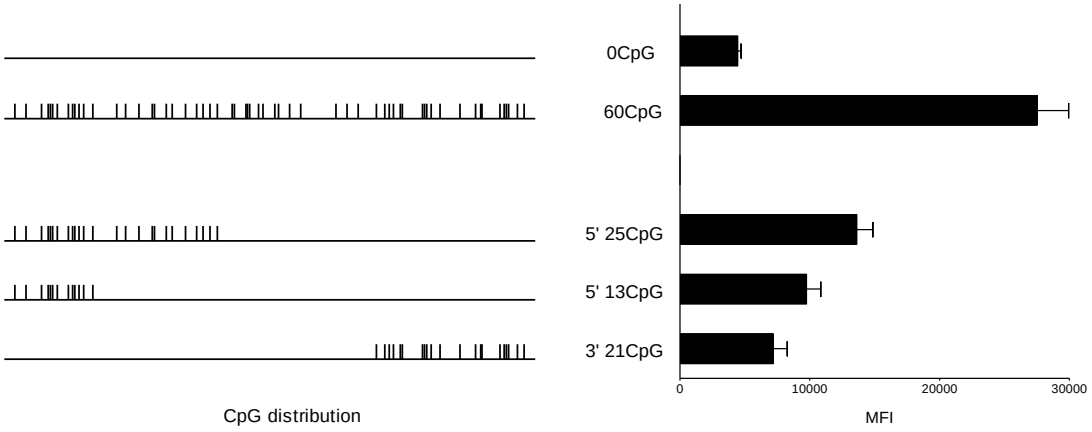

Figure 20 I Expression analysis of *hgfp* CpG-chimera. On the basis of hGFP-0 and hGFP-60, gene chimera with different CpG distribution were generated by fusion PCR, followed by stable transfection into CHO Flp-In cells. hGFP expression of the respective gene variants was analyzed by flow cytometry. The mean and standard deviations of three measurements is shown.

Correlating the expression efficiency (indicated by the MFI) to the CpG frequency, it becomes evident that the expression levels decrease with increasing distance of CpG dinucleotides from the start codon (Table 6). While hGFP-60 and hGFP containing 25 CpGs in their 5'gene end achieved similar MFI/CpG frequency ratios, hGFP containing only 13 CpGs in the very 5'gene border revealed a 2-fold increased

MFI/CpG frequency over hGFP with 21 CpGs in the 3'gene end. Thus, not only the mere amount but also the proximity of CpG dinucleotides to the transcription start site significantly accounts for increased transcription rates.

Table 6 I Ratio of MFI/CpG frequency of hGFP-60 and gene chimera as depicted in Figure 20.

Gene variant	Ratio of MFI/CpG frequency
hGFP-0	-
hGFP-60	458
5' 25 CpG	545
5' 13 CpG	750
3' 21 CpG	343

4.2 CpG-dependent differential transgene expression in murine embryonic carcinoma cells P19

The results presented in section 4.1 provided convincing evidence that an augmented CpG content within the ORF of *hgfp* and *mmip-1α* significantly increased gene transcription in mammalian Flp-In cells. Long-term hGFP expression analysis of stable CHO Flp-In cells demonstrated gradual but surprisingly slow decrease in transgene expression over the period of one year. The gene expression decline correlated with transgene loss, DNA methylation and a higher degree of chromatinization *in vivo*. The Flp-In cell system used in this study provided site-specific integration of the respective transgenes within a transcriptionally active and epigenetically constant genomic environment under standardized conditions. Due to these abilities, the Flp-In system was the tool of choice to compare regulation mechanisms responsible for differential transcription efficiencies between CpG-variants. On the other hand, the consistent transgene integration into one specific genomic locus might restrict CpG-mediated mechanisms to a limited spectrum of epigenetic regulation. Moreover, the Flp-In system does not represent a relevant technique for gene therapy applications, the optimization of which is one of the long-term goals of this project. Instead, retro- and lentiviral vectors are frequently applied in gene therapy trials [187][188][189] and stem cells are the major source for regenerative medicine [190][191][192]. However, embryonic stem (ES) cells have a much higher potential of epigenetic activity than differentiated somatic cells [193]. Thus, expression sustainability by different regulatory transgene elements in ES cells has to be elaborately tested prior to their application.

P19 embryonic carcinoma cells are pluripotent stem cells with a high potential for gene silencing [194]. P19 cells were therefore chosen to examine the sustainability of hGFP expression depending on intragenic CpG frequency and regulating elements. Self-inactivating lentiviral vectors (SIN-LVs) incorporating respective expression cassettes were used to introduce *hgfp* variants of different intragenic CpG content into P19 cells. Three different promoters were compared for their ability to confer expression of hGFP variants within this system: The CMV promoter, EF-1α promoter and ubiquitously acting chromatin opening element (UCOE) from the human *HNRPA2B1-CBX3* locus (A2UCOE). A2UCOE was reported to sustain stable transgene expression in cell culture systems even in the absence of selection pressure, or when integrated into heterochromatin region [135].

4.2.1 Generation of SIN-LVs incorporating *hgfp* variants

SIN-LVs were produced by transient transfection of HEK 293 cells with the envelope plasmid pcDNA3.1-VSV-G, the packaging plasmid psPAX2 and a LV plasmid containing either hGFP-0 or hGFP-60 mediated by the CMV, EF-1α or the A2UCOE promoter (Figure 21).

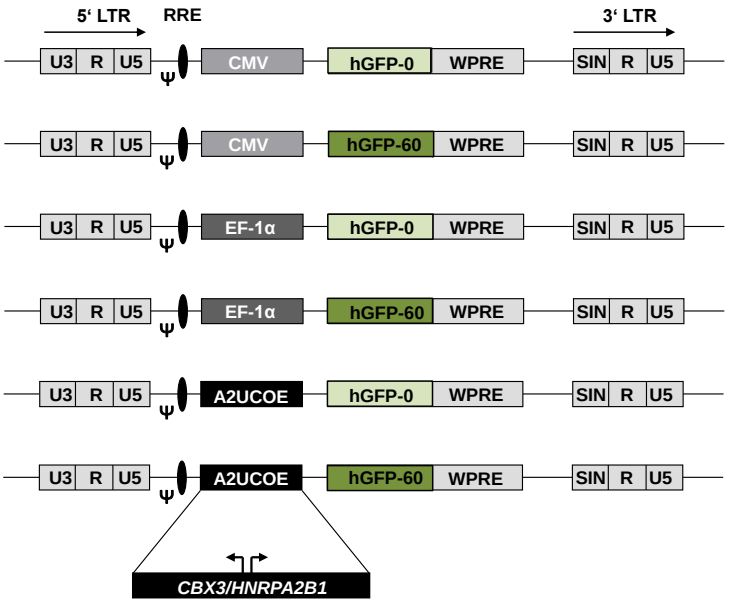

Figure 21 I Schematic of the lentiviral transfer vectors used in this study. LTR - long-terminal repeat, RRE - rev-response element, Ψ - packaging signal, WPRE - Woodchuck hepatitis posttranscriptional regulatory element. The divergent transcription directions at the *CBX3/HNRPA2B1* locus are indicated as arrows. The region covering the minimal 2.2-kb A2UCOE element [135] and the promoters CMV and EF-1α control the transcription of hGFP-0 and hGFP-60, respectively.

The titer of virus preparations was determined by transducing HEK 293 cells with serial dilutions of virus batches followed by quantification of hGFP expressing cells assayed by flow cytometry. The transduction of P19 cells by SIN-LVs incorporating the various constructs by an equal MOI as calculated from hGFP positive HEK 293 cells resulted in different proportions of hGFP expressing P19 cells among the variant hGFP constructs. Titrations of different initial virus concentrations for the transduction of P19 cells were therefore necessary to reach a similar percentage of hGFP positive cells at the start of the experiment. An equal proportion of hGFP positive cells was desired to obtain a comparable baseline for the comparison of expressional changes over time. A low MOI was used to avoid the integration of multiple copies per cell. An exception to this had to be made for CMV-promoter containing vectors. In this case, a high MOI was required to reach the hGFP

detection limit and to obtain comparable proportions of hGFP positive cells (Table 7). hGFP expression was assayed every three days up to 20 days.

Table 7 I Summary of used MOI to yield similar proportions of hGFP positive cells, vector copy numbers (VCN) per cell as calculated by copy numbers of WPRE relative to the 2-copy gene *mouse telomerase reverse transcriptase tert* (*mtert*) at day 5 in P19 cells, proportion of hGFP positive cells ($hGFP^+$) as assayed by flow cytometry at day 5 in P19 cells and ratio of hGFP expressing cells to VCN at day 5.

promoter	MOI	VCN/cells at day5	$hGFP^+$ cells at day 5 [%]	$hGFP^+$ cells/VCN at day 5
CMV	11	34.5	2.00	1/1725.3
	5	3.2	3.60	1/90.1
EF-1α	0.3	0.4	7.80	1/4.9
	0.2	0.2	14.20	1/1.1
A2UCOE	0.8	0.3	13.10	1/2.1
	0.3	0.3	15.00	1/1.9

4.2.2 Long-term expression of hGFP variants in P19 cells using different promoters

Flow cytometry analysis of hGFP expression two days after transduction of P19 cells with SIN-LVs revealed a similar proportion of hGFP positive cells among all cell lines (Figure 22).

Despite similar proportions of hGFP expressing cells at the start of the experiment, the stability of hGFP expression exhibited high variations between different promoters and CpG variants over the course of 20 days (Figure 23). hGFP expression by the CMV promoter declined rapidly within 5 days (hGFP-0: 12.5 – 2%; hGFP-60: 27.4 – 3.6% in 5 days) and remained at this low level thereafter. EF-1α-driven hGFP expression declined at a slower but constant rate to reach a similar low proportion of hGFP positive cells after 20 days (hGFP-0: 18.6 – 1.7%; hGFP-60: 18.7 – 5.6% in 20 days). In marked contrast, hGFP expression mediated by the A2UCOE element clearly increased within 11 days (hGFP-0: 14.3 – 23.5%; hGFP-60: 17 – 27.9% in 11 days) and fell then progressively back to the initial level of hGFP positive cells.

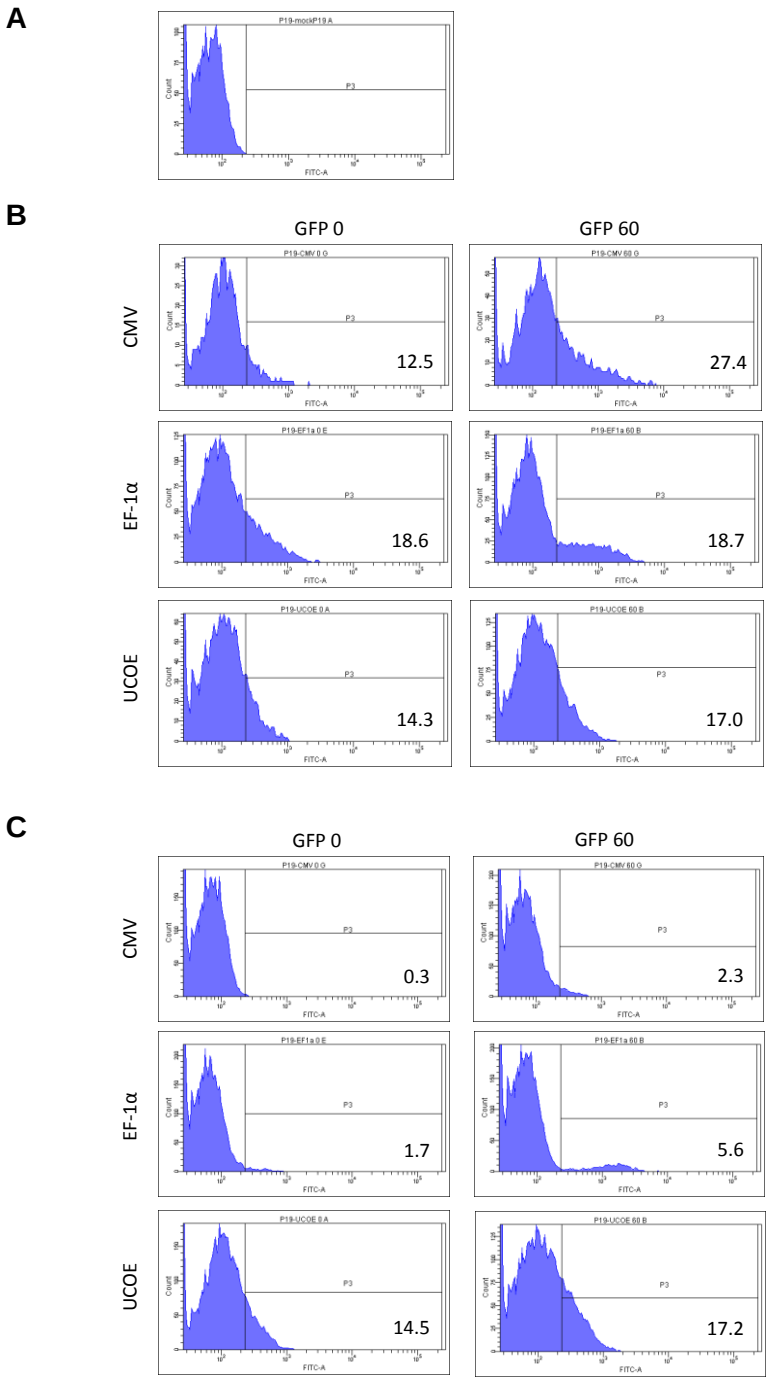

Figure 22 I Flow cytometry histograms of hGFP expression of non-transduced P19 cells (A) and P19 cells transduced with LVs at day 2 (B) and day 20 (C) after transduction. Percentage of hGFP positive cells [%] is shown in the right corner of each histogram.

The MFI of hGFP expressing cells reached the highest levels when transcription was controlled by CMV at the start of the experiment. After day 8, expression efficiency by CMV decreased to MFI levels lower than transgene expression controlled by the EF-1α. The MFI of transgenes driven by A2UCOE revealed slightly lower but very constant gene expression levels. For transgene expression controlled by the EF-1α promoter, a 2-fold increased MFI was observed for hGFP-60 compared to hGFP-0, whereas expression by CMV and A2UCOE was not affected by intragenic CpG dinucleotides. To verify the reproducibility of the obtained results, analogous assays were conducted with P19 cells transduced at higher MOIs, resulting in an overall higher percentage of hGFP positive cells at the start of the experiment but a comparable hGFP expression profile for all vectors used in this study (data not shown). Thus, the observed effects are irrespective of LV load.

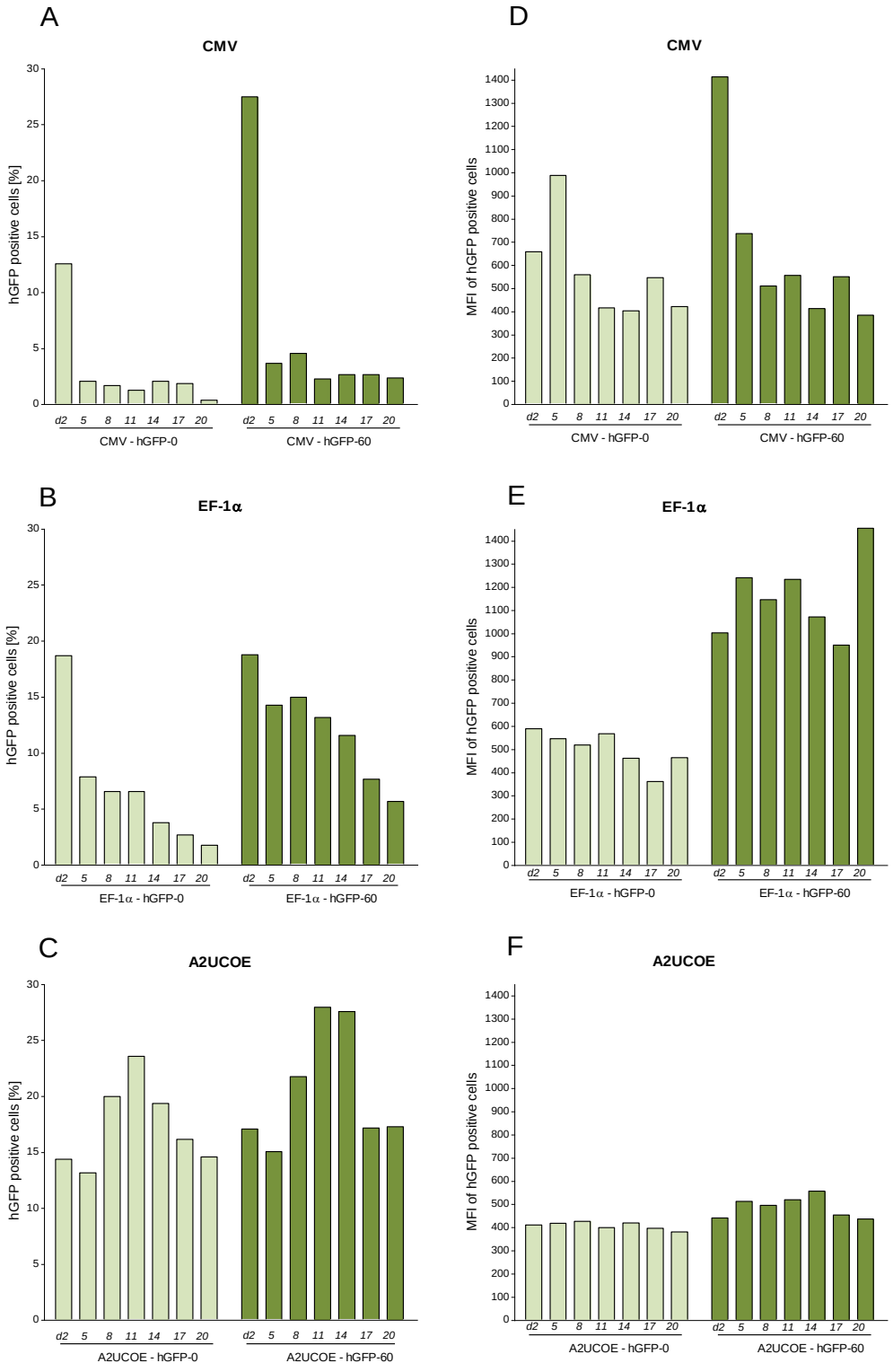

Figure 23 I Percentage of hGFP$^+$ cells (left panel) and MFI of hGFP$^+$ cells (right panel). P19 cells were transduced with vectors shown in Figure 21 at different MOIs, shown in Table 7 to yield a similar amount of hGFP positive cells measured by flow cytometry. hGFP expression was measured 2 days after transduction and subsequently every 3 days up to 20 days, as indicated.

Quantitative PCR was performed on day 5 after transduction of P19 cells to determine the average vector copy number (VCN) for each of the vector variants. *wpre* copies were compared to the endogenous 2-copy gene *mouse telomerase reverse transcriptase* (*mtert*). VCNs summarized in Table 7 revealed a similar copy number compared to the used MOI (as determined by HEK 293 transduction) for EF-1α and UCOE vectors indicating predominance of single copy integrations. In contrast, CMV controlled constructs revealed a much higher copy number than the MOI calculated from hGFP positive cells by flow cytometry. This result was already anticipated considering the high virus concentration applied to reach the hGFP detection limit of this gene variant. This is due to the tendency of the CMV promoter to get silenced very rapidly immediately after integration, resulting in a lower proportion of hGFP expressing cells than hGFP containing cells. Additionally, hGFP-0 tends to confer very weak gene expression, which might have led to the sorting of false negative hGFP expressing cells. The high ratio of hGFP expressing cells to VCP of EF-1α- and UCOE-driven transgenes at day 5 after infection indicated a relatively stable hGFP expression compared to CMV-mediated hGFP expression (Table 7). This silencing effect is even more pronounced in the CMV hGFP-0 transgene.

4.2.3 Partial prevention of *hgfp* silencing in P19 cells by DNMT inhibition

Such severe silencing effects are virtually always connected with DNA methylation, which was therefore assumed to be a major contributor of the loss of function in P19 cells. To test this hypothesis, P19 cells were supplemented with the DNA methyltransferase (DNMT) inhibitor 5-Aza-2'-deoxycytidine (5'aza) at day 20 after transduction, followed by 2 days of incubation and subsequent flow cytometry analysis (Figure 24). An increased proportion of hGFP positive cells was detected when grown in the presence of 5'aza, which was associated with an augmented MFI. Conversely, the respective cell population containing CMV- and EF-1α constructs cultured without 5'aza supplementation led a further repressed state at day 22 after infection. UCOE-driven transgene expression stayed continually stable without 5'aza treatment and could even be increased up to an average of 20.4% (hGFP-0) and 27.65% (hGFP-60) upon 5'aza supplementation. The ratio of hGFP expressing cells in the presence of 5'aza to their respective control cell group correlated with the extent of gene silencing (Table 8). The highest reestablishment of transgene expression was observed in CMV hGFP-0, which concordantly also showed the most severe gene silencing effects throughout the experiment. The inability of 5'aza treatment to re-establish initial high levels of hGFP expressing cells is assumed to be due to further epigenetic silencing effects, such as histone modifications, which are not affected by 5'aza. This hypothesis is supported by similar observations in previous transgene expression analysis conducted in P19 cells [195].

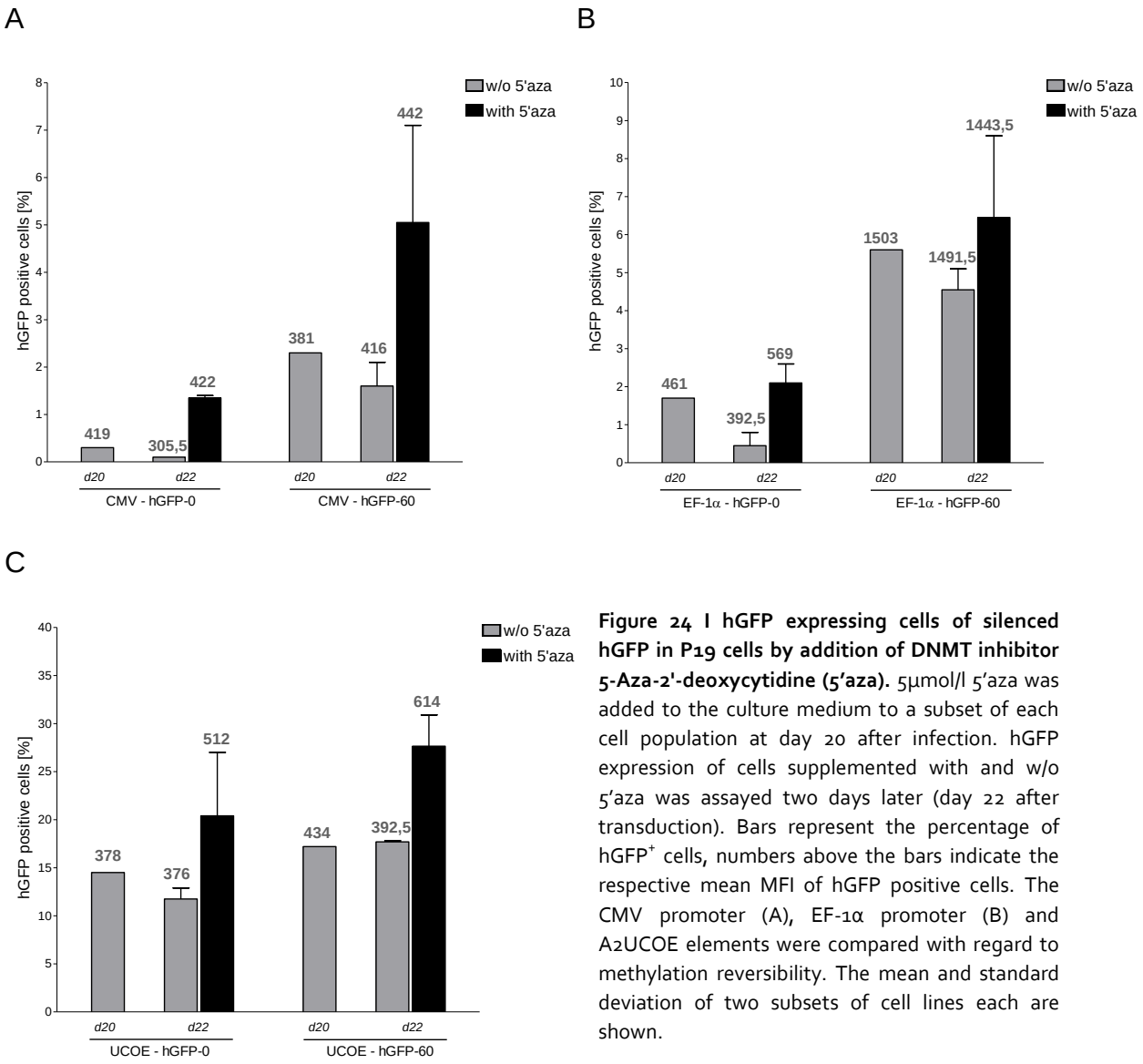

Figure 24 I hGFP expressing cells of silenced hGFP in P19 cells by addition of DNMT inhibitor 5-Aza-2'-deoxycytidine (5'aza). 5µmol/l 5'aza was added to the culture medium to a subset of each cell population at day 20 after infection. hGFP expression of cells supplemented with and w/o 5'aza was assayed two days later (day 22 after transduction). Bars represent the percentage of hGFP$^+$ cells, numbers above the bars indicate the respective mean MFI of hGFP positive cells. The CMV promoter (A), EF-1α promoter (B) and A2UCOE elements were compared with regard to methylation reversibility. The mean and standard deviation of two subsets of cell lines each are shown.

Table 8 I Ratio of hGFP$^+$ cells supplemented with 5'aza to hGFP$^+$ cells without 5'aza at day 22. Values were calculated from the ratios of of hGFP positive cells in Figure 24. 5'aza treatment was carried out for two days.

promoter	Gene variant	Ratio of hGFP$^+$ cells at d22 with/ w/o 5'aza
CMV	hGFP-0	13.5
	hGFP-60	3.2
EF-1α	hGFP-0	4.7
	hGFP-60	1.4
A2UCOE	hGFP-0	1.7
	hGFP-60	1.6

5 Discussion

5.1 Evolution of CpG frequency in the mammalian genome

Due to the degeneration of the genetic code, some nucleotides in the open reading frames can be exchanged without changing the resulting amino acid sequence [196]. In the wobble position of codons, all thymines can be replaced by cytosines, and almost all adenines can be replaced by guanine. Consequently, every gene can be encoded by a large number of different sequences. Since the first and second position of the codons of almost all amino acids cannot be substituted, the exchange of entire dinucleotides would however result in a nonsense mutation. Exceptions to this are the dinucleotides ApG, TpC, CpT and CpG. Due to the absent pressure to remain in the open reading frames one would assume that the frequency of these dinucleotides might decrease compared to other dinucleotides in gene sequences. No evidence can however be found in the literature that the ApG, TpC or CpT content is significantly different from other dinucleotides within genes of the mammalian genome. In marked contrast, genome-wide studies have shown that particularly first exons and the 5'region of exons were shown to be rich in CpGs [67][92][93][94]. This is particularly striking since CpGs are actually significantly underrepresented throughout the mammalian genome as a consequence of their high susceptibility to mutation [66]. Despite this negative selection and the absent intragenic pressure, evolutionary processes seem to have maintained a high CpG frequency within these distinct regions of human genes. It seems obvious that the evolution of this intragenic CpG overrepresentation must confer gene expression and its regulation any selective advantage over other nucleotide combinations.

A vast number of studies has been published addressing the role of CpG islands and their implication on transcriptional regulation [12][64][89][90][91][95][96][111][197][198]. These stretches of mostly unmethylated CpGs are predominantly found within the promoter and the first exon of several genes, particularly housekeeping genes [67][89]. CpG islands have long been known as transcriptional promoting elements [91]. Apart from CpG islands, which compose only 1-2% of the vertebrate genome [199], most of the CpG dinucleotides in the mammalian genome are methylated. The majority of methylated DNA in normal adult tissue is harbored by non-coding transposable elements such as SINEs, LINEs and endogenous retroviruses [80][81]. DNA methylation within these non-functional stretches of the genome function to maintain the repressed chromatin state and therefore stably silence promoter activity [200].

5.2 CpG dinucleotide usage is pivotal for transgene expression

Depending on the origin, their genomic surrounding, developmental stage and type of the cell, CpG dinucleotides can have opposing effects on gene expression efficiency (see chapter 3.4). Thus, CpG dinucleotides should be used with caution in gene design for applications such as gene therapy or the production of recombinant proteins in mammalian cells.

Several studies have reported that CpG depletion from transgenes improves the persistence of expression in host cells [201][202][203]. *In vivo* studies in transgenic mice using a CpG containing reporter gene have shown that methylation of the upstream EF-1α promoter is induced resulting in transcriptional gene silencing, whereas the CpG-depleted reporter had no influence on promoter methylation and led to an extended reporter activity [204]. CpG-containing transgenes were even reported to diminish the expression of an adjacent reporter gene located on the same plasmid, while the CpG-depleted equivalent transgene had no negative effect on reporter expression [205]. These observations led to the accepted theory that CpG dinucleotides should be avoided when designing transgene expression vectors. In contrast, CpGs within genes have also been found to be beneficial for expression levels, such as by improving RNA stability [206]. Previous studies in our laboratory have shown that the depletion of CpGs from different transgenes such as hGFP [172], the capsid protein of HIV and murine erythropoietin (mEPO) [162] resulted in a drastic loss of reporter activity in mammalian cell lines and mice, respectively. Conversely, intragenic CpG enrichment clearly enhanced gene expression of the reporter mMIP-1α in H1299, CHO and HEK 293 cells [170]. It was demonstrated that intragenic CpG dinucleotides positively correlate with altered *de novo* mRNA synthesis [162][170][172].

5.3 Intragenic CpG abundance determines expression levels of hGFP and mMIP-1α

The reporter genes *hgfp* and *mmip-1α* used in previous CpG studies served as basis for experimental analyses of CpG-mediated epigenetic regulation mechanisms in the thesis at hand. *mmip-1α* and *hgfp* genes were previously modified with regard to optimal codon usage according to the CAI [160] and their intragenic CpG content. Humanized GFP (hGFP-60) [171], harboring 60CpG in its ORF, was used as the basis for the generation of the CpG-lacking hGFP-0 [169]. The wild type *mmip-1α* gene sequence was subjected to computer-assisted optimization strategies with initial focus on codon usage (mMIP-13). Based on mMIP-13, intragenic CpG dinucleotides were depleted (mMIP-0) and maximized (mMIP-42), respectively [170]. In accordance with previous data obtained for hGFP and mMIP-1α, expression analysis of both transiently transfected H1299/HEK 293 cells [170] as well as of stably

transfected CHO Flp-In (hGFP)/ HEK 293 (mMIP-1α) cells could verify the positive correlation between intragenic CpG content and transcription efficiency. Notably, changes in protein levels between CpG variants were even more pronounced in stably transfected cells compared to transient transfected cells (data not shown). This already implies that chromatin dynamics or epigenetic marks in the proximity of the transgene affect its transcription regulation in a CpG-dependent manner.

The CpG depletion in *hgfp* resulted in a CAI of 0.93, which is insignificantly lower than that of hGFP-60 (0.96). These minor differences in CAI are unlikely to result in the 6-fold increase in protein levels of stably expressed hGFP-60 compared to hGFP-0 (CHO Flp-In cells; Figure 9). The applied sequence modifications in *mmip-1α* resulted in a high CAI for the codon optimized (mMIP-13) and the CpG-depleted (mMIP-0) cytokine variants, but a CAI below wild type (mMIP-wt) for the CpG-maximized (mMIP-42) cytokine variant. Despite low CAI, mMIP-42 showed the highest expression levels (5-fold in stable HEK 293 Flp-In cells compared to mMIP-wt), while mMIP-0 gave protein yields barely above the detection limit despite a higher CAI than mMIP-wt and mMIP-42 (Figure 9). Adapting the codon usage usually results in increased expression yields [158], which could not be observed for mMIP-13 in the given experimental setting. Since gene-optimization was not the object of the study at hand, this correlation was not further scrutinized. In order clarify the effects of codon optimization on gene expression, these analyses will be repeated. Nevertheless, according to the 5-fold increased protein levels upon CpG maximization, gene optimization of *hgfp* and *mmip-1α* with regard to codon usage can be excluded to be the crucial modification promoting gene expression in a CpG-dependent manner.

It was shown that short half-lives observed for transiently expressed genes, such as lymphokines, cytokines and transcription factors, correlate with the presence of AU-rich elements in their 3'UTR [207]. UpA-dinucleotides are preferred targets of endoribonuclease cleavage which results in decreased mRNA stability [206]. It is therefore tempting to argue that increasing the CpG content simultaneously decreased the TpA and AU content, respectively, thereby increasing mRNA stability. During the modification process of *hgfp* and *mmip-1α* however, TpA amounts were not changed significantly (Table 3).

Another study found a correlation of high mRNA levels and increased GC frequency [208]. This raises the question whether an increased GC rate might actually be responsible for the effects ascribed to CpG dinucleotides in this study. This possibility can also be excluded, since the overall GC content among transgene variants was barely changed, and the observed changes in protein yields do not generally correlate with GC content (Table 3; Figure 9). In addition to similar GC and TpA frequencies, care was taken during the modification process to not create any TATA-boxes, cryptic splice sites, internal polyadenylation sites or other regulatory elements. In search of the responsible mechanism for the observed phenotypes, altered gene expression was proven to be irrespective of biased transfection rates, translation efficiencies, mRNA export and splicing activities. Instead, it was shown that intragenic CpG dinucleotides lead to increased gene transcription.

CpG-dependent effects were verified with different mammalian cell lines and various promoters [162][170][172].

5.4 Intragenic CpG dinucleotides confer no disadvantage for long-term expression in mammalian Flp-In cells

Effective and sustainable transgene expression in mammalian cells is a major aspect in the production of recombinant protein [179]. Once a transgene is stably integrated into the host cell, it is affected by the genomic DNA surrounding the integration locus. Especially epigenetic marks spreading their repressive modifications to the proximal transgene are a crucial issue [124]. Another problem leading to reduced expression sustainability is the loss of the transgene over time due to the outgrowth of a less productive, metabolically favored sub-population, or the ejection of the complete transgene via its recombination site [209]. Hence, the production of recombinant proteins usually requires some kind of selection pressure in order to maintain sustainability of transgene expression.

Transgenes stably integrated into mammalian cells via the Flp-In recombination system served for comparative analysis of CpG-associated epigenetic mechanisms. The Flp-In system was chosen as it ensures stable insertion of a single copy of the transgene at a specific location within an active chromatinized setting [165]. All plasmids used for transgene delivery into mammalian Flp-In cells contained a hygromycin resistance gene (*hph*) which was stably integrated into the cells together with the transgene. To maintain selective conditions, hygromycin was added to the culture medium. *hph*, which was located in the proximity of the transgene-driving promoter, therefore had to be constantly expressed. The transgenes integration into a transcriptionally active genomic region, and also the constant and high transcription of the upstream *hph* gene have important functions; first of all, these features provide a selective advantage for cells carrying the transgenes, and secondly, they keep the chromatin structure at the promoter and ORF permissively open and the DNA unmethylated. There are, however, utilities for recombinant protein production that need cells to grow under antibiotic-free conditions, such as to minimize cellular stress [210] or to avoid the contamination of cells with antibiotics in industrial fermentation processes [211].

It was hypothesized that intragenic CpG dinucleotides might negatively affect expression levels upon selection pressure withdrawal due to methylation of intragenic CpG dinucleotides and chromatin compaction. To address this issue, CHO Flp-In cells stably transfected with *hgfp* variants were cultivated either with or without selection pressure over the course of one year. Both the CMV and the EF-1α promoter were examined for their ability to confer stable transgene expression.

While selective conditions preserved 100% hGFP expressing cells, the withdrawal of selection pressure resulted in gradually decreasing ratios of hGFP positive cells (Figure 10, Table 4). Surprisingly, intragenic CpG dinucleotides did, however, not lead to accelerated gene expression loss compared to CpG-lacking gene variants. Instead,

the reduction of hGFP-0 expressing cells occurred even slightly faster compared to hGFP-60, irrespective of the used promoter. The efficiency of hGFP expression (quantified as MFI) on the other hand decreased both in the absence and in the presence of selection pressure. In CMV driven constructs, the decrease occurred faster in cell lines expressing hGFP-60 compared to cells harbouring hGFP-0. Contrary and most notably, *hgfp* transcription driven by the EF-1α promoter could resist gene silencing more effectively with an increased intragenic CpG content. The results imply that selection pressure at least fulfilled one purpose – the prevention of transgene ejection or the outgrowth of transgene lacking cells. This was achieved for both CpG variants and for both promoters analyzed. Upon selection pressure withdrawal however, the ratio of hGFP positive cells of all gene variants and promoters decreased gradually. To distinguish whether this effect was a consequence of transgene loss or epigenetic downregulation, cells maintained without selection pressure were sorted into subpopulations according to their expression level, followed by detailed transgene analysis.

5.5 Intragenic CpG dinucleotides cause increased DNA methylation rates, whereas low CpG content promotes transgene loss

CHO Flp-In cells stably expressing hGFP variants and maintained without selection pressure were sorted into the subpopulations maximal (max), moderate (mod) and no expression. Quantitative PCR of isolated DNA of the respective cell populations could show that absent (no) hGFP expression coincided with a decreased relative copy number compared to moderately and highly expressing cells. The transgene loss was almost 2-fold higher in hGFP-0 compared to hGFP-60 (Figure 12). To elucidate to what extent the decreasing hGFP expression efficiency arouse from DNA methylation, bisulfite sequencing of genomic DNA isolated from respective subpopulations was carried out. Evaluation of the according chromatograms revealed that selection pressure could not only eliminate the risk of transgene loss, but also prevent from DNA methylation. Withdrawal of selective conditions, on the other hand, resulted in increased levels of methylation (Figure 13). The methylation level of transgenes clearly correlated with expression efficiency in both hGFP-0 and hGFP-60. hGFP-0 lacks methylation targets in the ORF, whereas hGFP-60 provides many potential sites for *de novo* methylation. *De novo* DNA methyltransferases (DNMTs) have the tendency to spread bidirectional in the genome [212]. Hence, it is assumed that the excessive amount of intragenic CpG dinucleotides in hGFP-60 attracts many *de novo* DNMTs, leading to ORF methylation, which then spreads to the promoter and reinforces the silencing effect. This hypothesis is further supported by the finding of an increased methylation level of the ORF in transcription start site (TSS) proximity, which gradually decreases towards the 3'end. This methylation gradient

also implies a greater regulatory function of CpG dinucleotides in the 5'intragenic region compared to downstream loci.

Considering the inverse rates of DNA methylation and transgene ejection between *hgfp* variants (Table 5), it seems obvious that the reduced methylation targets of hGFP-0 compared to hGFP-60 were compensated by an increased frequency of complete transgene loss.

Cell lines cultivated under selective conditions neither showed any signs of transgene loss, nor DNA methylation. Yet, the transcription efficiency slowly decreased over the period of one year (CMV-hGFP-0 41%; CMV hGFP-60 46%; Table 4). This decline of expression efficiency was already observed in previous experiments in our laboratory with the reporter mMIP-1α. By the supplementation of sodium butyrate (NaB), which is an inhibitor of histone deacetylases (HDACs), this expression decline could partly be reversed [186]. HDACs remove acetyl groups from histones and thereby contribute to gene repression [44].

Altogether, these results indicate that transcription of CpG variants is regulated by several epigenetic processes working in concert to control transgene expression.

5.6 Intragenic CpG dinucleotides alter chromatin structure

Transcriptional activity is affected by chromatin structure and *vice versa*. Histone modifications, histone variants, chromatin remodeling and the DNA sequence itself impact on transcriptional events (see chapter 3.3). It was thus interesting to evaluate in what respect differential transcription rates observed upon intragenic CpG variations were reflected by chromatin changes.

5.6.1 Chromatin density of *hgfp* transgenes is affected by intragenic CpG dinucleotides and growth conditions *in vivo*

To examine the chromatin density of CpG variants *in vivo*, genomic DNA of CHO Flp-In cells expressing hGFP variants was subjected to FAIRE. Cell populations cultivated either with or without selection pressure were analyzed. Out of the cell population cultivated without selection pressure, only the fraction in which hGFP expression was no longer detectable (denoted as "no expression") was examined.

The amount of extracted nucleosome-free DNA isolated from the TSS and ORF of *hgfp* clearly correlated with the presence of selection pressure. Thus, withdrawal of selective conditions not only induced DNA methylation but also a clearly increased chromatin density at the TSS of hGFP-0 and hGFP-60.

It is well established that processes involving DNA methylation and (modified) histones are tightly connected [49]. An example of proteins connecting histone and DNA modification is DNMT3L, which assists the binding of *de novo* methyltransferases to DNA (see chapter 3.3.1). Once H3K4 loses its

methylation, the DNMT-DNMT3L complex can bind to DNA and methylates CpG dinucleotides [117][49]. Inversely, histone methyltransferases have been found to recruit *de novo* DNA-methyltransferases, thereby silencing gene expression [50]. The crosstalk between DNA methylation and histone modification or chromatin compaction can go in both directions. DNA methylation and chromatin condensation in hGFP expressing cells are supposed to have affected each other. The direction of causality and the identification of histone modifications involved in the silencing process could however not be determined by this experiment.

Quantitative comparison of isolated nucleosome-free DNA of hGFP-0 and hGFP-60 stably expressed by cells under selective conditions revealed another striking aspect. hGFP-0 lacking intragenic CpGs showed a significantly higher chromatin density at the ORF compared to hGFP-60 (Figure 14). It is assumed that the increased chromatin density resulting from intragenic CpG depletion is a major contributor of impeded transcription efficiency. Changed chromatin density of *hgfp* variants was clearly visible in CHO Flp-In cells cultivated under selection pressure, but was not observed in CHO Flp-In cells cultured without antibiotic selection. It seems that the effect of nucleosome reorganization resulting from CpG variations is lost upon selection pressure removal in favor of a more intensive chromatin condensation. Altogether, these results imply that chromatin structure plays a crucial role in the CpG-mediated transcription regulation.

5.6.2 Chromatin density of *mmip-1α* transgenes is increased upon CpG depletion *in vivo*

To verify that CpG depletion leads to a chromatin condensation also for a different reporter gene and cell line, FAIRE was analogously carried out to analyze the chromatin density of HEK 293 Flp-In cells stably expressing the mMIP-1α variants at the TSS and the 3'UTR under selection pressure. The results revealed similar levels of nucleosome density between mMIP-wt, mMIP-13 and mMIP-42. In contrast, mMIP-0 exhibited a significantly higher degree of chromatinization, both at the TSS and the 3'UTR (Figure 15). These results are in concordance with hGFP FAIRE analysis and confirm that CpG depletion leads to local condensation of chromatin. This conclusion is further supported by a study of nucleosome remodeling in mammalian primary response genes by Toll-like receptors that has revealed a tendency of CpG islands to exhibit a reduced nucleosome occupancy as a direct result of their nucleotide content [213]. CpG maximization, however, did not lead inversely to further chromatin de-compaction. The missing correlation of chromatin density and expression efficiency for mMIP-42 can be explained by the features of the Flp-In cell system. This system mediates transgene integration within a transcriptionally accessible chromatin conformation [165]. Taking this into account, the CpG enrichment in mMIP-42 presumably did not lead to further disaggregation of

chromatin *in vivo* due to saturation effects of the already very open chromatin structure at the Flp-In recombination locus.

The relative differences of chromatin densities among CpG variants could not be observed in H3 and H3K4me3 ChIP analyses (Figure 16). It is assumed that DNA fragments generated by sonication in the ChIP procedure were too large to clarify the local differences of histone occupation on the small *mmip-1α* gene and the promoter. Furthermore, H3K4me3 is a histone modification usually associated with the promoters of active genes (see chapter 3.3.1) [39]. The presence of H3K4me3 in all gene variants irrespective of intragenic CpG content reflects that the genes are generally located in a transcriptionally active locus. It is assumed that this active surrounding and the promoter itself are responsible for the recruitment of H3K4me3. This histone modification is therefore unlikely to explain the severe changes in transcription efficiency. Further gene activating histone modifications, particularly those coinciding with gene bodies such as H3K9me1, H3K27me1 or H3K36me (see chapter 3.3.1) [39] should be of great interest in future experiments to identify the responsible CpG-triggered control mechanisms.

5.6.3 CpG dinucleotides in *hgfp* affect nucleosome positioning abilities *in vitro*

The differences of chromatin compaction in correlation to intragenic CpG changes led to the assumption that also the nucleosome positioning abilities might be affected by CpG changes. *In vitro* nucleosome reconstitutions with PCR fragments of *hgfp* variants were assembled by salt dialysis. The comparison of the nucleosome position pattern indeed revealed individual positioning capabilities among *hgfp* variants *in vitro* (Figure 18). Altered nucleosome positions as a result of CpG variations could previously be shown for the reporter gene *mmip-1α* [186].

Nucleosome preferences *in vitro* originate from the sequence-dependent mechanics of the wrapped DNA itself [214]. Early studies postulated that intrinsic DNA sequence preferences also affect nucleosome positioning *in vivo* [58][215]. Subsequent analyses demonstrated that *in vitro* nucleosome preferences indeed often reflect *in vivo* locations [216][217] and that nucleosome positioning *in vivo* can be predicted based on the genomic DNA sequence alone [218]. According to a genome-wide analysis of nucleosome positioning, approximately 50% of the *in vivo* nucleosome organization is solely determined by sequence preferences of nucleosome occupation [59]. The *in vivo* nucleosome occupancy map of *human mip-1α* (*hmip-1α*) reveals a strong positioning preference of the +1 nucleosome at the 5' end of the coding sequence of *hmip-1α* that is shifted 40 nucleotides upstream when activated upon inflammation (unpublished data; group of Prof. Längst, Regensburg). Assuming that the *mmip-1α* transgene model used in this study is comparable to the endogenous situation, it is concluded that the positioning abilities of the +1 nucleosome plays an important role in the observed transcriptional changes of transgenes.

Considering the sum of chromatin analyses, it is hypothesized that differential transcription efficiencies of CpG variants result from the creation of a transcriptionally more, respectively less, favorable nucleosome arrangement facilitated by intragenic CpG dinucleotides. More precisely, altered gene transcription efficiency is supposed to result from a combinational effect of changed nucleosome positioning and chromatin density.

5.7 Intragenic CpG dinucleotides increase transcription elongation of *mmip-1α*

FAIRE analyses in this work could elaborately show that the Flp-In system targets the transgenes into an active genomic environment. This transcriptionally active setting of the transgene Flp-In locus was furthermore reflected by the abundance of total and engaged RNAPII at the promoter of all *mmip- 1α* variants compared to *gapdh*, as quantified by ChIP analysis. Similar RNAPII binding to the promoter between the gene variants indicated that RNAPII recruitment is not modulated by CpG dinucleotides downstream (Figure 19). In fact, genome-wide studies of transcription regulation in human cells have demonstrated that approximately 20% of unexpressed genes are constantly occupied by preloaded RNAPII prior to transcription initiation [219]. Phosphorylation of serine-2 (Ser2P) at the carboxy-terminal domain (CTD) of RNAPII is a modification that occurs later in the transcription cycle during the elongation process [19]. Among all *mmip-1α* variants, an equal amount of Ser2P RNAPII at their promoters was found. Upstream Ser2P RNAPII density therefore also excludes abortive transcription initiation to be responsible for CpG-divergent transcription rates. However, a trend of an increased amount of actively transcribing RNAPII at the 3'UTR of mMIP-42 compared to mMIP-wt, mMIP-13 and mMIP-42 could be detected. The synthetic *mmip-1α* used in this study contains no introns and hence is a small gene of 279bp, which is too short to be occupied by more than one Polymerase simultaneously. Determining the efficiency of transcriptional elongation as *de novo* mRNA transcripts per unit density of elongating Pol II, it is concluded that RNAPII molecules traverse the ORF of mMIP-42 at a higher elongation rate than of mMIP-wt, mMIP-13 and mMIP-0.

5.8 Gene expression benefits from TSS-proximity of intragenic CpG dinucleotides

The ability of CpG dinucleotides to enhance elongation rates led to the question whether a specific region within the ORF of CpG variants was responsible for this effect. Thus, gene chimeras with CpG clusters in distinct intragenic 5' and 3' regions of the ORF were generated. Analyses of CHO Flp-In cells stably expressing hGFP

chimera, containing CpG clusters in either 5' or 3' regions of the ORF, showed that not only the amount of CpG dinucleotides, but particularly the localization of CpG dinucleotides in proximity to the start codon, respectively TSS, is pivotal for efficient gene transcription (Figure 20).

Interestingly, genome-wide studies of regulatory regions in the human genome have shown that the average of protein coding genes in the human genome display a significant excess of CpG dinucleotides in the 5' ends of exons, most pronounced in the first exon [67][92][93][94]. In a genome-wide transcriptome analysis of different human cell lines, a positive correlation of expression efficiency and CpG frequency at the TSS +500bp downstream was found (unpublished data; cooperation with group of Prof. Längst, Regensburg). CpG frequency in the gene body was low in both gene sets. CpG dinucleotides are depleted throughout the mammalian genome as a consequence of their high susceptibility to mutation [66]. Despite this negative selection, evolutionary processes apparently seem to have maintained a high CpG frequency at the TSS-proximity in a group of genes in correlation to their expression performance. A genome-wide study by Choi et al. investigated the nucleosome deposition and DNA methylation at regulatory regions in human cells. The authors have likewise revealed a group of genes with exceptionally high frequency of mainly unmethylated CpGs at the 5' gene end [220]. These genes were furthermore found to exhibit high expression rates compared to average expression levels, and even higher than in genes controlled by a promoter CpG island. The authors ascribe the enhanced expression levels to effects of elongation control. These data obtained by previous investigations and the findings described in the study at hand point to the very important function of intragenic CpGs proximal the intragenic 5' region. In the case of *hgfp* and *mmip-1α* transgenes, 5'-adjacent CpGs seem to play a major role in the configuration of chromatin architecture, allowing efficient transcription elongation.

5.9 CpG frequency and type of promoter determines transgene stability in pluripotent stem cells P19

CHO Flp-In cells demonstrated gradually decreased transgene expression over the period of one year without selection pressure (4.1). The gene expression decline correlated with transgene loss, DNA methylation and a higher degree of chromatinization *in vivo*. Despite these repressive events, transgene expression was still clearly detectable in a surprisingly high percentage of cells at the end point of the experiment. The Flp-In cell system used in this study provided site-specific single copy integration of the respective transgenes within a transcriptionally active and epigenetically constant genomic environment under standardized conditions. Due to these abilities, the Flp-In system was the tool of choice to compare regulation mechanisms responsible for differential transcription efficiencies between CpG-variants. On the other hand, the consistent transgene integration into one specific genomic locus might restrict CpG-mediated mechanisms to a limited spectrum of epigenetic regulation. Moreover, the Flp-In system does not represent a relevant

technique for gene therapy applications, the optimization of which is one of the long-term goals of this project. Instead, retro- and lentiviral vectors are frequently applied in gene therapy trials [187][188][189]. The mammalian Flp-In cells used for comparative expression analyses were exclusively differentiated cells. However, not differentiated somatic cells, but rather stem cells are the major source for regenerative medicine [190][191][192]. Embryonic stem (ES) cells have a much higher potential of *de novo* methylation [221] and epigenetic control in general [193] than differentiated somatic cells. Thus, expression sustainability by different regulatory transgene elements in ES cells has to be elaborately tested prior to their application.

LVs have a preference for integration into the proximity of active transcription units [130]. However, the integration location is not directed to one identical region in all cells, as in the Flp-In system. Particularly the development of self-inactivating retroviral vectors (SIN LVs) with a deleted U3 region of the 3'LTR containing the viral enhancer sequence provides gene transfer with higher safety due to the reduced risk of enhancer-mediated mutagenesis [131][222]. SIN-LVs incorporating respective expression cassettes were used to introduce *hgfp* variants of different intragenic CpG content into P19 cells. P19 embryonic carcinoma cells are pluripotent stem cells with the ability to differentiate into various cell types such as neuronal, glial, cardiac and skeletal muscle [223][224]. The ability to change their phenotype completely arises from their immense potential of epigenetic regulation [225]. This feature was utilized to challenge transgene stability. P19 cells were transduced with SIN-LVs incorporating *hgfp* transgenes containing a different CpG content to investigate the susceptibility of this reporter gene to be silenced dependent on intragenic CpG frequency. Further to that, the stability of transgene expression from different promoters was tested. The CMV promoter, the human EF-1α and the A2UCOE element from the human *HNRPA2B1-CBX3* locus were compared regarding their capacity to mediate stable transgene expression.

5.9.1 CMV- and EF-1α-promoter-mediated hGFP expression is gradually silenced in P19 cells

Despite similar infection rates at the start of the experiment, the stability of hGFP expression revealed substantial differences depending on the driving promoter (Figure 23).

The CMV promoter is widely used due to its strong gene expression potential in several tissues [226][227]. This feature was elaborately verified in this study by all experiments conducted in Flp-In cells. The CMV promoter is however also known to confer very variable expression depending on the cell type [228], which seems most critical in ES cells [229][230]. Indeed, the disposition to undergo extensive epigenetic repression in ES cells became already apparent at the start of the experiment, reflected by the requirement of a very high MOI for the transduction of P19 cells to reach the hGFP detection limit (Table 7). Progressing silencing forces became increasingly obvious through the rapid decline in hGFP expression immediately after

gene transduction albeit multiple transgene copies. These characteristics of a high silencing activity of CMV controlled transgenes were even more pronounced in hGFP-0. Thus, a decreased intragenic CpG content was not able to prevent gene silencing, which is a widely used strategy in transgene expression applications [161][201][202]. Instead, the generally low potential of CpG-lacking transgene expression seems to even promote the rapid silencing of CMV- controlled hGFP expression. Despite the ability of hGFP-60 to induce a higher expression rate at the start of the experiment (as quantified by the MFI), gene silencing led to a decreased MFI similar to hGFP-0 after just five days.

The human EF-1α promoter is a widely used element to regulate retroviral transgene expression claiming robust and constitutive gene expression[228][231]. According to a low required MOI to reach detectable hGFP signals and the observation of a low vector copy number (VCN) when transduced at a low MOI, the EF-1α promoter indeed appeared to be a more suitable regulating element compared to the CMV promoter. Nevertheless, also EF-1α-mediated transgene expression declined gradually over the period of 20 days. Interestingly, the EF-1α-controlled expression level of hGFP (as quantified as MFI) in P19 was increased by 2-fold in hGFP-60 compared to hGFP-0. Consistent with CMV-promoter vectors, EF-1α promoter-mediated hGFP expression therefore seems to benefit from intragenic CpG content, leading to a delayed gene repression. Despite the delayed silencing effects of the EF-1α compared to the CMV promoter, the amount of hGFP expression was still almost completely diminished after 20 days. This also makes the EF-1α promoter appear to confer insufficiently stable gene expression for ES cell applications.

5.9.2 A2UCOE confers stable hGFP expression in P19 cells and prevents hGFP repression upon intragenic CpG depletion

Several *cis*-elements have been proposed to avoid transgene silencing, such as locus control regions (LCRs), chromatin insulators or scaffold/matrix attachment regions (S/MARs) [232]. Further to these elements, ubiquitous chromatin opening elements (UCOEs) consisting of divergently transcribed promoters of housekeeping genes, surrounded by a methylation free CpG island with a chromatin opening ability, were demonstrated to induce stable transgene expression. UCOE derived from the human *HNRPA2B1-CBX3* locus (A2UCOE) was shown to confer stable levels of transgene expression in a variety of different cell lines, including P19 cells [195][233]. Due to these observations, A2UCOE was included in the expression analyses of this study to test the impact of these chromatin modifying features on hGFP expression with respect to varying intragenic CpG content. Indeed, the number of hGFP expressing cells as well as the MFI of hGFP-0 and hGFP-60 remained constant during the period of the experiment. Moreover, a high ratio of hGFP positive cells to VCN was observed. This indicates a high proportion of actively hGFP expressing cells and a low silencing tendency. Remarkably, gene expression was equally efficient between hGFP-0 and hGFP-60, as reflected by an equal and constant MFI.

Flp-In expression analyses demonstrated that the lack of intragenic CpGs in *hgfp* led to decreased transcription efficiency and that this effect can at least in part be ascribed to a transcriptionally more unfavorable chromatin structure and nucleosome position. A2UCOE can provide a transcriptionally active environment through its chromatin opening features. It is assumed that the methylation-free CpG islands of A2UCOE interact with active histone modifications and that the bidirectional transcription by the UCOE closely spaced dual divergent promoters is associated with an inherent chromatin opening function [195]. Based on these observations and in concordance with the findings of this study, the chromatin opening features of A2UCOE seem to overcome the establishment of a more repressive chromatin state induced by the lack of CpG dinucleotides.

5.9.3 DNMT inhibition partly prevents *hgfp* silencing in P19 cells depending on promoter usage

Previous studies elaborately demonstrated a clear correlation of DNA methylation and declined transgene expression in P19 embryonic carcinoma cells [195]. The spleen focus-forming virus (SFFV) LTR exhibited an almost complete state of CpG methylation, EF-1α was methylated to a significantly lower degree compared to SFFV, whereas the UCOE element showed only very weak levels of methylation. In the light of these findings and in concordance with the observations described above, DNA methylation was assumed to be a major contributor to the loss of function in P19 cells. Detailed methylation levels of *hgfp*-transduced P19 cells were not elaborately examined within the scope of this work. However, a contribution of DNA methylation events on *hgfp* repression was investigated by application of the DNA methyltransferase (DNMT) inhibitor 5-Aza-2'-deoxycytidine (5'aza) to P19 cells stably integrated with the respective gene variants. Indeed, the ratio of hGFP positive cells grown in the presence of the 5'aza correlated with the extent of gene silencing. The highest re-establishment of silenced transgenes was observed in CMV hGFP-0, whereas A2UCOE-driven transgenes were almost not affected by the de-methylating agent (Figure24; Table8). The inability of 5'aza treatment to re-establish initial high levels of hGFP expressing cells is assumed to be due to further epigenetic silencing effects, such as histone modifications, which are not affected by 5'aza. This hypothesis is supported by similar observations in previous transgene expression analyses conducted in P19 cells [195].

The data clearly demonstrated that the stability of transgene expression in SIN LV-transduced P19 carcinoma stem cells depend on the choice of promoter and transgene sequence. The strong viral CMV is unsuitable due to its high potential to become silenced. The EF-1α promoter is silenced as well, albeit at a lower rate. Furthermore, transgene expression from EF-1α benefits from an augmented intragenic CpG frequency, reflected by delayed silencing effects and a two-fold increase in expression efficiency up to at least 20 days after transduction. In contrast,

CpG accumulation seemed to have no positive effect on A2UCOE-driven transgenes, as this element itself induces a permissive chromatin state.

5.10 Proposed CpG-mediated transcriptional control mechanism and outlook

Considering the sum of observations obtained from hGFP and mMIP-1α analyses in the study at hand, the following scenario is proposed: CpG differences within the ORF of *hgfp* and *mmip-1α* cause an altered arrangement of nucleosomes. Intragenic CpG depletion results in a more compact chromatin structure, thereby impeding effective transcription elongation. By contrast, the accumulation of CpG dinucleotides induces nucleosome rearrangement and instability. It is suggested that particularly the removal of the +1 nucleosome from the start codon facilitates effective transgene elongation. Similar conclusions were drawn by independent investigations. Choi et al have suggested that the deposition of nucleosomes downstream of the TSS by CpGs and their modification plays a pivotal role in epigenetic regulation [220]. In a Review, Harinder Singh described CpG islands as the "transcriptional tee off areas of the mammalian genome that provide a nucleosome-depleted surface" for efficient transcription elongation [12]. Support of this hypothesis is further provided by a study of mammalian primary response genes by Toll-like receptors, in which the authors claim CpG-island promoters to facilitate "promiscuous induction from constitutively active chromatin without a requirement for [...] nucleosome remodeling complexes" [213]. It is assumed that intragenic CpGs in mMIP-42 and hGFP-60 mimic CpG islands and thereby exhibit CpG-island typical features.

One consequence of the weakened chromatin structure facilitated by CpGs is the greater accessibility of the underlying DNA to transcriptional regulators *in vivo*. Several transcription factors have been found to affect transcriptional elongation. The DRB sensitivity-inducing factor (DSIF) and the negative elongation factor (NELF) are known to induce transcriptional pausing. By contrast, the transcription factor IIF (TFIIF), protein kinase P-TEFb and the eleven-nineteen lysine-rich in leukemia (ELL) activate efficient elongation. A detailed overview of Pol II elongation factors is reviewed in [234]. Trans-activating factors directly associated with unmethylated CpG dinucleotides might also contribute to the observed phenotype. The ubiquitously expressed CXXC finger protein 1 (CFP1) specifically and exclusively binds unmethylated CpG dinucleotides, thereby trans-activating transcription [235]. CFP1 is discussed to play an important role as an epigenetic regulator in modulating gene expression via CpG dinucleotides [236]. Both the overexpression and downregulation of CFP-1 in 293 Flp-In cells stably expressing mMIP-1α did however not show significant changes in mMIP-1α expression levels (data not shown). KDM2A is another factor binding to CpG dinucleotides. It removes H3K36 methylation, thereby creating a "CpG island chromatin" that is depleted of this repressive modification [110]. It is very likely that there exist several more transcription factors

not identified so far, representing probable candidates for CpG-based effectors of transcription enhancement. Complex epigenetic processes can be triggered by CpG dinucleotides (see chapter 3.4.3). Looking at the ever increasing number of discovered epigenetic regulation mechanisms, a substantial impact of CpG dinucleotides on transcription regulation becomes apparent. Occupation of the histone modification H3K4me3, which is often found in CpG islands [107][237] was not significantly changed between the CpG variants of mMIP-1α. However, looking at the wide range of histone modifications associated with gene regulation, it is very likely to discover histone modifications, remodeler or histone variants that act in concert to generate the observed CpG-mediated differences in expression efficiencies in the near future. The identification of such regulators and responsible DNA sequence elements will provide new perspectives regarding CpG-rich transgenes designed for efficient expression in mammalian cells.

6 Materials

6.1 Cell lines

Cell type	Origin	Comments	ATTC no.
HEK 293T/17	*Homo sapiens*	Ad5 transformed embryonic kidney cells	CRL-11268
HEK 293 Flp-In	*Homo sapiens*	HEK 293 cells with a single stably integrated FRT site (Invitrogen R750-07)	CRL-1573
CHO Flp-In	*Cricetulus griseus*	Based on Chinese hamster ovary cells CHO-K1 (Invitrogen R758-07)	CCL-61
P19	*Mus musculus*	Derived from an embryonal carcinoma induced in a C3H/He mouse	CRL-1825

6.2 Bacterial strains

Bacterial strain	Description
DH5α	f- *sup*E44 Δ*lac*U169 (ϕ80 *lac*ZΔM15) *hsd*R1 *rec*A1 *end*A1 *gyr*A96 *thi*1 *rel*A1

6.3 Media and supplements

Medium	Composition	Supplements	Organism
LB_0 (Luria Bertani)	1% Bacto-tryptone 0.5% yeast extract 1% NaCl NaOH adjusted to pH 7.0 autoclaved		DH5α
LB_{Amp}	sterile LB_0	100µg/ml ampicillin	DH5α
Agar plates	LB_{Amp}	1.5% agar	DH5α
DMEM (Dullbecco's modified eagle medium	See manufacturer's product information (Gibco)	10% FBS (fetal bovine serum; PAN) 100µg/ml penicillin (PAN) 100µg/ml streptomycin (PAN)	293
Ham's 12 (Invitrogen)	See manufacturer's product information (Gibco)	10% FBS (fetal bovine serum; PAN) 2mM L-Glutamine (PAN) 100µg/ml penicillin (PAN) 100µg/ml streptomycin (PAN) 100µg/ml zeocin or 500µg/ml hygromycin	CHO

MEM-α	See manufacturer's product information (Gibco)	10% FBS (fetal bovine serum; PAN) 100µg/ml penicillin (PAN) 100µg/ml streptomycin (PAN) 2mM L-Glutamine (PAN) 1% NEAA (non essential amino acids)	P19

6.4 Kits

Name	Application	Supplier
CCL_3/MIP-1 alpha DuoSet Mouse	ELISA	R&D
EpiTect Bisulfite-Kit	Bisulfite Converision	Qiagen
QIAamp DNA Mini Kit	gDNA Isolation	Qiagen
QIAGEN Plasmid Maxi Kit	pDNA Isolation	Qiagen
QIAGEN Plasmid Midi Kit	pDNA Isolation	Qiagen
QIAquick Gel Extraction Kit	Gel extraction of DNA fragments	Qiagen
QIAquick PCR Purification Kit	Purification of DNA fragments	Qiagen

6.5 Buffers and reagents

Buffer/reagent	Composition	Application
APS	10% Ammoniumperoxidsulfat in H_2O	PAGE
ChIP buffer IA	10mM HEPES/KOH pH 7.9 85mM KCl 1mM EDTA 1x protease inhibitor cocktail (Roche) Phosphatase inhibitor cocktail SetV (Calbiochem, 524629)	ChIP
ChIP buffer IB	10mM HEPES/KOH pH 7.9 85mM KCl 1mM EDTA 10% Nonidet P-40 1x protease inhibitor cocktail (Roche) Phosphatase inhibitor cocktail SetV (Calbiochem, 524629)	ChIP
ChIP buffer II	50mM Tris/HCl pH 7.4 1% SDS, 0.5% Empigen BB 10mM EDTA pH 8.0 1x protease inhibitor cocktail (Roche)	ChIP

ChIP dilution buffer	150 mM NaCl 20mM Tris–HCl pH 8.1 1.2mM EDTA 1% Triton X-100 0.01% SDS Protease inhibitor cocktail (Roche) Phosphatase inhibitor cocktail SetV (Calbiochem, 524629)	ChIP
ChIP low salt buffer	20mM Tris [pH 8.1] 150 mM NaCl 2 mM EDTA 1% Triton X-100 0.1% SDS	ChIP
ChIP high salt buffer	20mM Tris [pH 8.1] 0.5M NaCl 2mM EDTA 1% Triton X-100 0.1% SDS	ChIP
DNA sample buffer (6x)	0.001% Bromphenol blue (w/v) 0.001% Xylene cyanol (w/v) 50mM EDTA pH 8.0 30% Glycerin (w/v)	Agarose Gel electrophoresis
FAIRE buffer IA	10mM HEPES/KOH pH 7.9 85mM KCl 1mM EDTA 1x protease inhibitor cocktail (Roche)	FAIRE
FAIRE buffer IB	10mM HEPES/KOH pH 7.9 85mM KCl 1mM EDTA 10% Nonidet P-40 1x protease inhibitor cocktail (Roche)	FAIRE
FAIRE buffer II	50mM Tris/HCl pH 7.4 1% SDS 0.5% Empigen BB 10mM EDTA pH 8.0 1x protease inhibitor cocktail (Roche)	FAIRE
FACS buffer	1mg/ml NaN_3 1% FCS in PBS	Flow cytometry
Fixation solution	2% formaldehyde (v/v) 0.2% glutaraldehyde (v/v) In PBS	X-Gal staining
High Salt buffer	10mM Tris/HCl ph 7.6 2M NaCl 1mM EDTA 0.05% NP40 1mM ß-mercaptoethonol	Nucleosome reconstitution
Laemmli sample buffer (6x)	300mM Tris/HCl pH 6.8 12% SDS (w/v) 60% Glycerin (w/v) 10% Mercapto ethanol (v/v) 0.025% Bromphenol blue	PAGE

LiCl buffer	10mM Tris [pH 8.1] 0.25M LiCl 1mM EDTA 1% Igepal-CA630 1% deoxycholic acid)	ChIP
Low salt buffer	10mM Tris/HCl ph 7.6 50mM NaCl 1mM EDTA 0.5% NP40 1mM ß-mercaptoethonol	Nucleosome reconstitution
native PAA gel (5%)	8.3ml Acrylamide (Rotiphorese; 30%) 41.6ml 0.4x TBE 300µl APS (10%) 30µl TEMED	Native PAGE
PBS	7mM Na2HPO4 3mM NaH2PO4 130mM NaCl	Diverse applications
PBS-T	PBS with 0.5% Triton X-100	ELISA
Ponceau solution	2% Ponceau red 3% trichloroacetic acid (TCA)	Western blot
RIPA buffer	50mM Tris/HCl pH 8.0 150mM NaCl 0.1% SDS (w/v) 1% Nonidet P-40 (w/v) 0.5% Natriumdesoxycholat (w/v) 2 tablets protease inhibitor (complete mini; Roche)	Cell lysis
TE buffer	10mM Tris [pH 8.0] 1mM EDTA	Diverse applications
TBE Buffer (10x)	1M Tris 1M Boric acid 20mM EDTA pH 8.0	Agarose gel electrophporesis Native PAGE
TBS buffer	150mM NaCl 50mM Tris/HCl pH 7.5	Western blot
Transfer buffer pH 8.3	0.3% Tris 1.45% Glycin 20% Methanol 0.2% SDS	Western blot
Trypan blue reagent	0.5% Trypan blue	Cell staining
TTBS	0.05% Tween-20 in TBS	Western blot
X-gal staining reagent	4mM Ferricyanid 4mM Ferrocyanid 2mM $MgCl_2$ 5% X-Gal (20mg/ml in Dimethylsulfonamid (DMSO)	X-gal staining

6.6 Plasmids

Plasmid	Supplier
pcDNA5/FRT	V6010-20 (Invitrogen)
pOG44	V6005-20 (Invitrogen)
pcDNA5-CMV-mMIP-wt/13/0/42	Kindly provided by Dr. Bauer [170]
pcDNA5-CMV-hGFP-0/60	Kindly provided by Dr. Leikam [169]
pcDNA5-EF-1α-hGFP-0/60	Adopted from previous studies [186]
pcDNA3.1-VSV-G	Kindly provided by Dr. Alexander Kliche
psPAX2	Kindly provided by Dr. Alexander Kliche
pHR'SINcPPT-EF1α -eGFP-WPRE	Kindly provided by Dr. Zhang, Institute of Child Health, UCL, London, UK
pHR'SINcPPT-UCOE-EGFP-WPRE	Kindly provided by Dr. Zhang, Institute of Child Health, UCL, London, UK

6.7 Oligonucleotides

Name	Sequence (5'-3')	Application
Bis-CMV5' fwd	TTGTATGAAGAATTTGTTTAGGG	Bisulfit sequencing
Bis-CMV5' rev	TAATACCAAAACAAACTCCCAT	Bisulfit sequencing
Bis-CMV3' fwd	GGATTTTTTTATTTGGTAGTATATTTA	Bisulfit sequencing
Bis-CMV3' rev	CTCTAATTAACCAAAAAACTCTACTTATAT	Bisulfit sequencing
Bis-hGFP60-5' (915) fwd	TTGTTATTATGGTGAGTAAGGG	Bisulfit sequencing
Bis-hGFP60-5' (1359) rev	TAATTATACTCCAACTTATACCCCA	Bisulfit sequencing
Bis-huGFP60-3' (1206)-fwd	AGGAGTGTATTATTTTTTTTAAGGA	Bisulfit sequencing
Bis-huGFP60-3' (1685)-rev	TAAATATCTACAAAATTCCACCACA	Bisulfit sequencing
*EcoR*I-pc5/6351-2043-fwd	ATCGAATTCAGGCGTTTTGCG	Cloning of pHR vectors
GFP0_I_rev	TGCTGCTTCATGTGGTCTGG	Nucleosome reconstitution
GFP0_II_fwd	GTGCAGTGCTTCAGCAGATACC	Nucleosome reconstitution
GFP0_II_rev	TGCCATTCTTCTGCTTGTCTG	Nucleosome reconstitution
GFP0_III_ fwd	ATGTGTACATCATGGCAGACAAG	Nucleosome reconstitution
GFP60_I_rev	TGCTGCTTCATGTGGTCGGG	Nucleosome reconstitution
GFP60_II_fwd	GTGCAGTGCTTCAGCCGC	Nucleosome reconstitution
GFP60_II_rev	TGCCGTTCTTCTGCTTGTCG	Nucleosome reconstitution

GFP60_III_fwd	ACGTCTATATCATGGCCGACA	Nucleosome reconstitution
GFP ORF 954 fwd	GGGTGGTGCCCATCCTGGT	FAIRE
GFP ORF 1074 rev	GTGGTGCAGATGAACTTCAGGGT	FAIRE
*Nde*I(2)-pHR/9666-2974 rev	TATCATATGCGATGCGGGGAGG	Cloning of pHR vectors
rDNA fwd	GGCGGACTGTCCCCAGTG	FAIRE
rDNA rev	GTGGCCCCGAGAGAACCTC	FAIRE
*Sal*I-pHR/9691-650 fdw	TACGTCGACTGGCTAGCGTTTAAAC	Cloning of pHR vectors
*Sbf*I-pc5/6351-2043-rev	ATCCTGCAGGCCACACTGGACTA	Cloning of pHR vectors
TSS fwd	AGAGAACCCACTGCTT ACT GG CTTA	FAIRE
TSS rev	GCTAGCCAGCTTGGGTCT CCC TA	FAIRE
ß-Actin-2781-fwd	ACCACCATGTACCCAGGCATTG	FAIRE
ß-Actin-3020-rev	GAGCCACCGATCCACACAGAGT	FAIRE
ß2-M fwd	CGAGACATGTAAGCAGCATC	FAIRE
ß2-M rev	GCAGGTTGCTCCACAGGTA	FAIRE
3'UTR fwd	CTCGAGTCTAGAGGGCCCGTTT	FAIRE/ChIP
3'UTR rev	GAGGGGCAAACAACAGATGG	FAIRE/ChIP

6.8 Chemicals, enzymes and materials

All chemicals were supplied by SIGMA-ALDRICH, FLUKA CHEMIE, Roth or MERCK. Enzymes were supplied by NEB.

7 Methods

7.1 Cultivation of eukaryotic cells

7.1.1 Maintenance of cell lines

Cell culture media and supplements were supplied by PAN Biotech or Invitrogen. Culture vessels were purchased from Greiner or BD Bioscience. All eukaryotic cell lines were maintained in an atmosphere consisting of 37°C and 5% CO_2. HEK 293 cells were grown in DMEM medium, CHO cells were cultivated in Ham's F12 medium and P19 cells were maintained in MEM-α medium, containing the respective supplements (see 6.3).

7.1.2 Transient transfections

Transient transfections were performed using polyethylenimine (PEI; PeqLab) [149]. Cells were seeded ($3x10^5$ cells/6-well plate/3ml media) 24h before transfection. The medium was replaced with antibiotic- and FBS-free medium and cells were transfected with the respective expression vector(s). A total of 2µg DNA (6-well) and 8µl of a 1mg/ml PEI in H_2O solution, each ad 100µl NaCl, was mixed on a vortex device, left standing for 10min at room temperature and added drop-wise to the cell suspension. Transfection medium was replaced 6h later with 3ml of supplemented growth medium. For transfection in a smaller or larger vessel, the volume of DNA and all transfection reagents was scaled up or down accordingly.

7.1.3 Establishment of plasmid-based stable cell lines

The establishment of stable cell lines was achieved using the Flp-In system (Invitrogen). A *lacZ-zeocin* gene, controlled by the SV40 promoter, is stably integrated within a defined region of the CHO Flp-In genome. A FRT-region is located downstream of the promoter and the ATG start codon and provides a site for homologous recombination mediated by the Flp recombinase. The expression vector pcDNA5/FRT contains an identical FRT site. CHO Flp-In cells were transfected with the Flp-recombinase expression plasmid pOG44 together with pcDNA5/FRT including the transgene. As a result, the transgene is introduced into the host cell within a defined genomic region. The selection of positive transfectants is obtained by a hygromycin resistance gene, located within pcDNA5/FRT.

CHO cells were stably transfected employing PEI as previously described (7.1.2). A total of 2µg plasmid DNA (6 well) was transfected at a pOG44:pcDNA5 rate of 9:1.

Procedure of transfection, number of seeded cells and amount of reagents was analogous to transient transfections. The medium was supplemented with gradually increased concentrations of Hygromycin B (50µg/ml–500µg/ml within approximately 3 weeks) which led to the selection of positive cell clones.

7.1.4 Lentiviral vector (LV) preparation and transduction of cell lines

LVs were produced by transient co-transfection of HEK 293 cells with the envelope plasmid pcDNA3.1-VSV-G, the packaging plasmid psPAX2, and the respective lentiviral vector (see 7.5.5) at a molar ratio of 1:3:4, employing PEI as previously described (7.1.2). Cells were harvested 48h post transfection and cleared by centrifugation at 3000*g* for 10min. The supernatants containing the viruses were loaded onto a 30% sucrose cushion in PBS and ultra-centrifuged at 130000*g* for 2h. The pellets were resuspended in cold PBS and left on ice for 1h. After thorough resuspension of the viruses, the liquid was transferred to an eppendorf tube and centrifuged at 14000*g* at 4°C for 5min to remove any remaining debris. The virus stock was aliquoted and stored at -80°C. The titer of LVs was determined by transducing HEK 293 cells with virus serial dilutions and monitoring expression after three days by flow cytometry. Lentiviruses containing hGFP-0 and hGFP-60, respectively, under control of the CMV, EF-1α or A2UCOE promoters were used to transduce P19 cells at different MOIs to reach a similar amount of positive cells.

7.2 Cultivation of prokaryotic cells

For the cloning of DNA fragments and amplification of plasmids, *E.coli* strain DH5α (f- *sup*E44 Δ*lac*U169 (ϕ80 *lac*ZΔM15) *hsd*R1 *rec*A1 *end*A1 *gyr*A96 *thi*1 *rel*A1) was used. Bacteria were grown in LB medium or on agar plates, supplemented with ampicillin (100µg/ml) if required, at 37°C.

7.3 DNA methods

7.3.1 Isolation of genomic DNA

Genomic DNA was isolated using the QlAmp DNA Mini Kit (Quiagen) according to the manufacturers' instructions. The method is based on cell lysis by detergent and the degradation of proteins by Proteinase K. Additionally, RNAseA was used to prevent RNA contamination. DNA was eluted in 200µl sterile H_2O and stored at -20°.

7.3.2 DNA quantification

The DNA amount and purity was determined using the Nanodrop (Peqlab). 1.5µl of DNA sample was pipettet onto the pedestral. The absorbance of the DNA sample was measured within a range from 220nm–300nm. All absorbance values were normalized to the DNA solvent reagent.

7.3.3 Agarose gel electrophoresis

DNA was separated using 1 – 1.5% (W/V) agarose gels, depending on the fragment size. Gels were prepared with 1xTBE buffer, mixed with 0.5mg/l ethidium bromide. DNA samples were mixed with 6x running dye. Molecular weight standard (NEB) consisted of 100bp, 1kb or 2-log ladder. Gels were electrophoresed in the corresponding buffer at 80 – 200V, depending on the size of the gel. DNA was detected by ultraviolet exposure.

7.3.4 DNA purification from agarose gels

DNA was purified from gels using the QlAquick Gel Extraction Kit (Qiagen) according to the manufacturers' instructions. DNA was usually eluted in H_2O.

7.3.5 *In vitro* methylation

Methylation of phGFP-60 was carried out using the the M.SssI methylase (NEB) according to the manufacturer's protocol. Quantitative methylation was verified by digestion of 1 µg methylated DNA with the CG methylation insensitive enzyme *Sac*I and the CG methylation sensitive restriction enzyme *Apa*I (both from NEB) for 1h at 37 °C. Plasmids were subsequently purified using the PCR purification Kit (Qiagen).

7.3.6 Bisulfite conversion and sequence analysis

Genomic DNA was isolated from cells using the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's instructions. Sodium bisulfite treatment of genomic DNA was performed to convert unmethylated cytosine to thymine residues using the EpiTect Bisulfite-Kit (Qiagen) according to the manufacturer's instructions. Primers used for bisulfite treated DNA amplification were designed based on converted sequences (6.7). Primer binding sites were devoid of CpGs to allow equal amplification of methylated and unmethylated DNA. PCR products of bisulfite converted DNA were separated by gel electrophoresis and bands of the appropriate size were cut out, purified by the QIAquick Gel Extraction Kit (Qiagen) and sent to Geneart/Life techonologies for sequencing. Sequence alignment was conducted by

the software Lasergene DNAstar (Seqman; v5.0.3) and chromatograms were analyzed by the software Chromas (Version 2.32, Technelysium). The methylation levels of CpG dinucleotides were determined by measuring the ratio of each of the cytosine peak heights to the sum of respective cytosine and thymine peak heights in automated DNA sequencing traces, according to a technique published by Jiang et al [182]. The evaluation and presentation of methylation levels was done by the software Excel 2010.

7.4 Polymerase chain reaction (PCR)

PCR was performed using the myCycler (Biorad) or the PCR Thermal Cycler Gene Amp 2400 (PerkinElmar). Real-time PCR was performed using the StepOnePlus Real-Time PCR system (Applied Biosystems). Primers were usually designed using the online software Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi).

7.4.1 Quantitative PCR/real-time PCR

Quantitative PCR (qPCR) was performed to evaluate copy count numbers or to quantify nucleosome depleted (FAIRE) and immunoprecipitated (ChIP) DNA, respectively. The DyNAmo™ Flash SYBR® Green qPCR Kit (Finnzymes) and the TaqMan® Genotyping Master Mix (Life technologies), respectively, were used for qPCR applications according to the manufacturer´s protocol. Quantitative amplification was carried out in a StepOnePlus real-time PCR system (Applied Biosystems). Product specificity was assessed based on melting curves. Fluorescence was measured and expressed as crossing point (Cp) when exceeding background fluorescence of the PCR master mix by the StepOne Software v2.2.2 (Applied Biosystems).

For relative quantification analyses of *hgfp* transgene copy number in CHO Flp-In cells, the DyNAmo™ Flash SYBR® Green qPCR Kit (Finnzymes) was used. For *hgfp* copy number determination in P19 cells, a predesigned Custom TaqMan® Copy Number Assay (Life technologies) targeting the Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element (WPRE) element was applied. For relative quantification, this assay was combined with the TaqMan® Copy Number Reference Assay (Life technologies) specifically binding to the *endogenous mouse telomerase reverse transcriptase* (*mtert*).

PCR efficiencies (E) were determined by evaluation of serial dilutions of the respective templates. E can be calculated from the slope of the standard curve: $E = 10^{-1/slope}$. Primers were designed such that the E was approximately 2. Data were analyzed using the $2^{\Delta\Delta CT}$ method.

The procedures for qPCR following ChIP and FAIRE analyses are specified in the respective sections below.

7.4.2 DNA sequencing

Sequence analyses of several genes were performed by the company Geneart AG/life technology according to the Sanger method. Evaluation was performed with the software Lasergene DNAstar (Seqman; v5.0.3) or the online software Chromas (Version 2.32, Technelysium).

7.5 Plasmid construction

7.5.1 Ligation

Ligations were set up in a total volume of 10µl, consisting of 1:3 concentrations of prepared vector DNA and respective insert, 1µl of 10x ligase buffer, 1µl T4 ligase (NEB) and an appropriate volume of H_2O. Ligations were performed at RT for 1h.

7.5.2 Transformation of *E.coli*

100µl of frozen competent DH5α cells were thawed on ice following addition of the complete ligation volume and incubation on ice for 20 min. The heat shock was performed at 42°C for 45sec. 900µl of LB_0 medium was added and cells were incubated at 37°C for 1h. After centrifugation and resuspension in 200µl LB_0 cells were plated on LB_{Amp} agar and incubated at 37°C o/n. Positive cell clones were identified by means of colony PCR and restriction analyses after plasmid isolation.

7.5.3 Preparation of plasmid DNA

Plasmid DNA used for analytic purposes such as DNA sequencing was extracted according to the alkaline lysis. 2ml of an o/n bacteria culture were centrifuged (13000 rpm; 1 min; RT) and resuspended in 200µl buffer P1 (Quiagen). Subsequently, 200µl lysis buffer P2 was added and the mixture was incubated at RT for 5 min. Lysis was stopped by adding 200µl buffer P3 (Quiagen). The samples were incubated at ice for 5 min followed by centrifuging twice (13000 rpm; 1 min; RT). After each centrifugation step, the DNA present in the supernatant was transferred into a fresh eppendorf tube and subsequently precipitated by addition of 0.8 volumes isopropanol, followed by centrifugation (13000 rpm; 30 min; 4°C). The DNA precipitate was washed with 70% ethanol and eluted in 50µl sterile H2O. For the extraction of larger DNA amounts with high purity, plasmid DNA was isolated by using Midi- or Midi Kits (Quiagen) according to the manufacturers' instructions.

7.5.4 Cloning of *hgfp* chimera

All *hgfp* chimera were created by fusion PCR using the plasmid pcDNA5-CMV-hGFP-o and pcDNA5-CMV-hGFP-o as template and corresponding primers listed in 6.7. Amplified inserts were subsequently cloned into pcDNA5 containing the CMV promoter via *Hin*dIII and *Bam*HI, thereby replacing *hgfp*, to give the gene chimera pcDNA5-hGFP-5'25CpG, pcDNA5-hGFP-5'13CpG and pcDNA5-hGFP-3'21CpG.

7.5.5 Cloning of lentiviral transgene vectors

The lentiviral (LV) vectors pHR'SINcPPT-EF1α-eGFP-WPRE and pHR'SINcPPT-UCOE-EGFP-WPRE (kindly provided by Dr. Zhang, Institute of Child Health, UCL, London, UK) served as basis for LV construction. The UCOE element was created as previously described [233]. The element EF1α-eGFP was released from pHR'SINcPPT-EF1α-eGFP-WPRE via *EcoR*I and *Sbf*I. The elements CMV-hGFP-o, CMV-hGFP-60, EF1α-hGFP-o and EF1α-hGFP-60 were obtained by amplification from the plasmids pcDNA5-CMV/EF-1α-hGFPo/60, thereby obtaining the restriction sites *EcoR*I and *Sbf*I. CMV-hGFP-o, CMV-hGFP-60, EF1α-hGFP-o and EF1α-hGFP-60 were subcloned into pHR'SINcPPT-WPRE via *EcoR*I and *Sbf*I to obtain pHR'SINcPPT-CMV-hGFP-o-WPRE, pHR'SINcPPT-CMV-hGFP-60-WPRE, pHR'SINcPPT-EF1α-hGFP-o-WPRE and pHR'SINcPPT-EF1α-hGFP-60-WPRE. Using pHR'SINcPPT-CMV-hGFP-o-WPRE and pHR'SINcPPT-CMV-hGFP-60-WPRE as a template, hGFP-o and hGFP-60 were amplified and cloned into pHR'SINcPPT-UCOE-EGFP-WPRE via *Sal*I and *Nde*I, thereby replacing eGFP and creating pHR'SINcPPT-UCOE-hGFP-o-WPRE and pHR'SINcPPT-UCOE-hGFP-60-WPRE.

7.6 Protein methods

7.6.1 Determination of protein amount according to Bradford

The total protein amount was analyzed spectrophotometrically according to the method of Bradford by using the "Biorad-protein-assay" reagent (Biorad)[238]. The protein amount was determined by a BSA standard curve.

7.6.2 Enzyme linked Immunosorbent Assay (ELISA)

Quantification of mMIP-1α in cell culture supernatants was performed in 96-well MaxiSorb-plates (Nunc) using a commercial ELISA kit (CCL3/MIP-1 alpha DuoSet Mouse; R&D) according to the manufacturers' instructions. The washing was

performed using the application platform hydro flex (Tecan). Antigen-antibody complexes were detected with a TMB substrate solution (BD Bioscience) according to manufacturers' instructions and read out at a wavelength of 450nm using a spectrophotometer (Biorad; 680 Microplate Reader). MIP-1α concentration was determined by a standard curve in which a known concentration of mMIP-1α was plotted against the intensity of the emitted signal.

7.6.3 Flow cytometry

hGFP expression was analyzed by flow cytometry. At least $3x\ 10^5$ cells were washed in 500µl PBS scraped off and centrifuged (300*g*; 5min; RT). The washing was repeated before cells were resuspended in 200µl FACS buffer.

hGFP fluorescence of a cell population of at least 10000 cells was detected using the flow cytometer FACS Canto II device (BD, FACS Diva v6.1.3 software). Results were evaluated by the FACS Diva v6.1.3 software and statistic calculations were performed using the Excel software 2010.

7.7 Formaldehyde-assisted isolation of regulatory elements (FAIRE)

FAIRE was essentially performed according to a published protocol [239]. Approximately $3x10^7$ exponentially growing CHO and HEK 293 Flp-In cells, respectively, stably expressing hGFP, respectively, mMIP-1α variants, were cross-linked for 7 min at RT with 1% formaldehyde added directly to the culture medium. The reaction was quenched by the addition of glycine to a final concentration of 125mM. Cells were scraped off, washed twice with ice-cold phosphate-buffered saline and collected by centrifugation (700 × g; 5 min; 4°C). The cell pellet was snap-frozen at -80°C for storage or directly resuspended in FAIRE buffer IA and lysed on ice for 10 min in FAIRE buffer IB (For buffer composition, see chapter 6.5). Cell lysate was centrifuged at 700*g* for 5min and cell nuclei lysed in FAIRE buffer II. Samples were sonicated using a Bioruptor sonicator (Diagenode) to yield approximately 200-500bp DNA fragments. Cell debris was spun at 16100*g* for 5min and the clarified supernatant was treated with RNAse A at a final concentration of 0.33µg/µl for 1-2h at 37°C. 25% of the sheared chromatin was isolated, treated with proteinase K (0.5µg/µl) at 56°C for 1h and reverse cross-linked o/n at 65°C. Released DNA was isolated by adding an equal volume of phenol-chloroform/isoamyl alcohol (25:24:1) in Phase Lock Gel Light Tubes (5Prime). The remaining 75% of sheared chromatin was directly extracted by phenol chloroform in the same way without prior proteinase K treatment and reverse cross-link. DNA from the aqueous phase of both chromatin fractions (with/without reverse cross-link) was subsequently precipitated by the addition of ammonium acetate (pH 7.5) to a final concentration of 2.5M and an equal volume of isopropanol followed by an overnight incubation at –20°C. The precipitate was collected the next

day by centrifugation for (30min; 16000*g*; 4°C), washed with 70% ethanol, air dried, and resuspended in 200µl double-distilled water.

Quantification of purified DNA was carried out by real-time PCR using the DyNAmo Flash SYBR® Green qPCR Kit from Finnzymes according to the manufacturer's instructions. Primers were designed to cover the TSS, ORF (for hGFP variants), 3'UTR (for mMIP-1α variants), the second exon-intron junction of *ß-2-microglobulin* (*ß2-m*) and *rdna* (see chapter 6.7). Data were analyzed using the $2^{\Delta\Delta CT}$ method. All results were normalized to *rdna* and referred to hGFP-0 and mMIP-wt, respectively. They are presented as the ratio of DNA recovered from crosslinked cells divided by the amounts of the same DNA in the corresponding non-crosslinked samples. Data evaluation and statistical calculations were conducted using the Excel software 2010 and the software GraphPad Prism v.4. For all FAIRE evaluations, the ANOVA/ Tukey's multiple comparison test was used.

7.8 Chromatin Immunoprecipitation (ChIP)

Approximately $3x10^7$ recombinant CHO and HEK 293 Flp-In cells stably expressing mMIP-1α variants were cross-linked by 1% formaldehyde for 12min. The reaction was quenched by 0.125M glycine. Cells were washed three times in 1× PBS, collected into ChIP buffer IA and lysed in ChIP buffer IB on ice for 10min. Cell lysate was centrifuged (700x*g*; 5min) and pelleted cell nuclei were lysed in ChIP buffer II. Samples were sonicated using a Bioruptor sonicator (Diagenode) to yield approximately 300-800bp DNA fragments. Cell debris was spun at 16100*g* for 5min and the cleared chromatin stored at -80°C. Approximately $2x10^6$ cells were used for one Immunoprecipitation (IP) reaction. Additionally, 20% of each sonicated sample was removed as Input fraction to calculate the Output/Input ratio in subsequent IPs. Sheared chromatin samples were diluted 1:10 in dilution buffer and precleared with rotation for 2h with 60µl protein A agarose/salmon sperm DNA slurry (Millipore) at 4°C. Precleared chromatin was incubated with the appropriate antibody (see table below) at 4°C overnight with gentle rotation. The retained Input fraction was likewise diluted 1:10 in dilution buffer and isopropanol precipitated by 2.5M ammonium acetate overnight. Antibody-chromatin complexes were precipitated the next day by incubation with 60µl protein A agarose/salmon sperm DNA slurry for 4h at 4°C with gentle rotation, followed by centrifugation (5min; 500*g*; 4°C). The supernatant was aspirated, and the pellet was washed consecutively with 1ml each of ChIP low-salt buffer ChIP high-salt buffer, lithium chloride (LiCl) buffer, and twice with TE buffer. After final aspiration of the washed beads, a total of 100 µl of 10% Chelex (Biorad) (10*g*; 100ml H_2O) was added to the Output samples and the precipitated 20% Input fraction. After 15min of boiling, Proteinase K (100µg/ml) was added to the Chelex/protein A bead suspension and incubated for 1.5 h at 56 °C while shaking, followed by another 15min of boiling. The suspension was then applied onto Micro Bio-Spin Columns (Bio-Rad) and centrifuged at 500*g* for 5min for purification of nucleic acids. The eluate was used directly as a template in quantitative PCR. Primer

were designed to cover a 83bp 3'region of the CMV promoter, a 83bp region immediately downstream the open reading frame (3'UTR) of mMIP-1α and a 69bp sequence of the first exon-intron junction of *gapdh* (see chapter 6.7). Data were analyzed using the $2^{\Delta\Delta CT}$ method and reported as Output to Input fraction or as the relative enrichment of immunoprecipitated DNA by the specific antibody at the respective gene of interest (GOI) normalized to the endogenous control gene (*gapdh*) and relative to the Input, minus the respective IgG signal, according to the formula:

$$relative\ enrichment = \frac{2^{\Delta\ Ct\ (Input-Output_{specific\ AB}} - 2^{\Delta\ Ct\ (Input-Output_{IgG}}{2^{\Delta\ Ct\ (Input-Output_{specific\ AB}} - 2^{\Delta\ Ct\ (Input-Output_{IgG}}\ \begin{matrix} GOI \\ GAPDH \end{matrix}$$

Data evaluation and statistical calculations were conducted using the Excel software 2010 and the software GraphPad Prism v.4.

Antibody	Description	Conc. per IP	Manufacturer/ catalogue no
α-H3	Rabbit polyclonal against Histone H3	5µg	Abcam, 1791
α- H3K4me3	Rabbit polyclonal against Histone H3K4me3	5µg	Abcam, ab8580
α-RNAPII	Rabbit polyclonal antibody against a peptide mapping at the N-terminus of Pol II of mouse origin (N-20)	7,5µg	Santa Cruz, sc-899 X
α-Ser2P RNAPII	Rabbit polyclonal against RNAPII CTD repeat YSPTSPS (phospho S2)	7.5µg	Abcam, ab5095
α-FLAG	Polyclonal rabbit against the FLAG epitope; peptide sequence DYKDDDDK	7.5µg	Sigma Aldrich, F7425
α-IgG	IgG from rabbit serum	5µg	Sigma Aldrich, I5006-10MG

7.9 Analysis of reconstituted mononucleosomes *in vitro*

7.9.1 Amplification of CpG fragments for nucleosome reconstitutions

hGFP fragments used in reconstitution assays were generated by PCR using the according primers (see 6.7) and purified using the QIAquick PCR Purification Kit (Qiagen) according to the manufacturers' instructions. Before use in reconstitution assays, fragments were diluted in H_2O to a concentration of 0.3-0.5µg/µl.

7.9.2 Nucleosome assembly by salt dialysis

hGFP fragments generated by PCR were incorporated into mononucleosomes using the salt dialysis technique [185]. Core histones, extracted from *Drosophila* embryos,

were kindly provided by Prof. Dr. Längst (Regensburg). An arbitrary plasmid (pUC 19) was used as competitive DNA. BSA was added to the reaction and low binding reaction tubes (Biozym) were used to avoid adherence of nucleosomes to the tube walls. In order to determine the optimal histone: DNA ratio for a given combination of gene variants, a test assembly was performed in which increasing amounts of histone octamers were added to a constant amount of DNA. A typical test assembly contained histone: DNA ratios of 0:1, 0.8:1, 1:1, 1.2:1, 1.4:1, 1.6:1 and 1.8:1. All additional reaction compounds are listed in Table 1. The ideal ratio was selected for nucleosome positioning. For subsequent comparative analyses, the histone:DNA ratio was used in which approximately 70% of the DNA was incorporated into nucleosomes. Large histone-DNA complexes resulting from high histone concentrations were avoided. A typical assembly reaction was prepared in low binding reaction tubes as follows:

Table 1 | Typical assembly set-up. * The optimal amount of histones was detected in a test assembly, described above

volume	component
1µl	BSA (10mg/µl)
4µl	gene variant (0.3µg/µl)
4µl	competitive plasmid (0.5µg/µl)
xµl (*)	purified octamers (1µg/µl)
ad 50µl	high salt buffer

The salt concentration of the high salt buffer was gradually decreased by the addition of low salt buffer via pipes using a peristaltic pump (150ml/h). Analysis of mononucleosomes was performed by native PAGE, loading approximately 700ng nucleosomes in each lane.

7.9.3 Analysis of mononucleosomes by Native PAGE

Native PAGE was used to characterize reconstituted mononucleosomes. Free DNA migrates faster through native gels than nucleosomes incorporated with the same DNA. Nucleosomes positioned at the end of a DNA fragment migrates faster that a nucleosome formed in the center of a DNA fragment. Native 5% PAA gels (for composition see 6.5) were prepared using the Novex-System (Invitrogen) according to the manufacturers' instructions. The gel cassette was set into the electrophoresis chamber and filled with 0.4xTBE Buffer, before nucleosome samples and a DNA ladder were seeded into the wells. Mononucleosomes were separated at a constant voltage of 100V for 80min. Nucleoprotein complexes were stained with ethidium bromide followed by ultraviolet exposure.

8 Reference List

1. JACOB F, MONOD J (1961) Genetic regulatory mechanisms in the synthesis of proteins. JMolBiol 3: 318–356.

2. Venters BJ, Pugh BF (2009) How eukaryotic genes are transcribed. Crit Rev Biochem Mol Biol 44: 117–141. doi:10.1080/10409230902858785.

3. Dynan WS, Tjian R (1985) Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. , Published online: 29 August 1985; | doi:101038/316774a0 316: 774–778. doi:10.1038/316774a0.

4. Ptashne M, Gann A (1997) Transcriptional activation by recruitment. Nature 386: 569–577. doi:10.1038/386569a0.

5. Hampsey M (1998) Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol Mol Biol Rev 62: 465–503.

6. Young RA (1991) RNA Polymerase II. Annual Review of Biochemistry 60: 689–715. doi:10.1146/annurev.bi.60.070191.003353.

7. Green MR (2005) Eukaryotic transcription activation: right on target. MolCell 18: 399–402.

8. Benoist C, Chambon P (1981) In vivo sequence requirements of the SV40 early promotor region. Nature 290: 304–310.

9. Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. Annu Rev Biochem 72: 449–479. doi:10.1146/annurev.biochem.72.121801.161520.

10. Yuan G-C, Liu Y-J, Dion MF, Slack MD, Wu LF, et al. (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309: 626–630. doi:10.1126/science.1112178.

11. Anderson JD, Widom J (2001) Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. Mol Cell Biol 21: 3830–3839. doi:10.1128/MCB.21.11.3830-3839.2001.

12. Singh H (2009) Teeing up transcription on CpG islands. Cell 138: 14–16. doi:10.1016/j.cell.2009.06.028.

13. Lee TI, Young RA (2000) Transcription of eukaryotic protein-coding genes. Annu Rev Genet 34: 77–137. doi:10.1146/annurev.genet.34.1.77.

14. Glass CK, Rosenfeld MG (2000) The coregulator exchange in transcriptional functions of nuclear receptors. Genes Dev 14: 121–141.

15. Casamassimi A, Napoli C (2007) Mediator complexes and eukaryotic transcription regulation: an overview. Biochimie 89: 1439–1446. doi:10.1016/j.biochi.2007.08.002.

16. Corden JL (1990) Tails of RNA polymerase II. Trends Biochem Sci 15: 383–387.

17. Phatnani HP, Greenleaf AL (2006) Phosphorylation and Functions of the RNA Polymerase II CTD. Genes Dev 20: 2922–2936. doi:10.1101/gad.1477006.

18. Schroeder SC, Schwer B, Shuman S, Bentley D (2000) Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes Dev 14: 2435–2440.

19. Komarnitsky P, Cho E-J, Buratowski S (2000) Different Phosphorylated Forms of RNA Polymerase II and Associated mRNA Processing Factors During Transcription. Genes Dev 14: 2452–2460. doi:10.1101/gad.824700.

20. Mohd-Sarip A, Verrijzer CP (2004) Molecular biology. A higher order of silence. Science 306: 1484–1485. doi:10.1126/science.1106767.

21. Paweletz N (2001) Walther Flemming: pioneer of mitosis research. Nat Rev Mol Cell Biol 2: 72–75. doi:10.1038/35048077.

22. Felsenfeld G, Groudine M (2003) Controlling the double helix. Nature 421: 448–453. doi:10.1038/nature01411.

23. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389: 251–260. doi:10.1038/38444.

24. Pierce BA (2005) Genetics; A conceptional Approach. 2nd ed. New York: W.H.Freeman.

25. Khorasanizadeh S (2004) The nucleosome: from genomic organization to genomic regulation. Cell 116: 259–272.

26. Olins AL, Olins DE (1974) Spheroid chromatin units (v bodies). Science 183: 330–332.

27. Oudet P, Gross-Bellard M, Chambon P (1975) Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell 4: 281–300.

28. Finch JT, Klug A (1976) Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci USA 73: 1897–1901.

29. Bednar J, Horowitz RA, Grigoryev SA, Carruthers LM, Hansen JC, et al. (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. Proc Natl Acad Sci USA 95: 14173–14178.

30. Schalch T, Duda S, Sargent DF, Richmond TJ (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature 436: 138–141. doi:10.1038/nature03686.

31. Robinson PJJ, Fairall L, Huynh VAT, Rhodes D (2006) EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. Proc Natl Acad Sci USA 103: 6506–6511. doi:10.1073/pnas.0601212103.

32. Robinson PJJ, Rhodes D (2006) Structure of the "30 nm" chromatin fibre: a key role for the linker histone. Curr Opin Struct Biol 16: 336–343. doi:10.1016/j.sbi.2006.05.007.

33. Grigoryev SA, Woodcock CL (2012) Chromatin organization — The 30 nm fiber. Experimental Cell Research 318: 1448–1455. doi:10.1016/j.yexcr.2012.02.014.

34. Rattner JB, Lin CC (1987) The higher order structure of the centromere. Genome 29: 588–593. doi:10.1139/g87-099.

35. Sajan SA, Hawkins RD (2012) Methods for identifying higher-order chromatin structure. Annu Rev Genomics Hum Genet 13: 59–82. doi:10.1146/annurev-genom-090711-163818.

36. Adkins NL, Watts M, Georgel PT (2004) To the 30-nm chromatin fiber and beyond. Biochim Biophys Acta 1677: 12–23. doi:10.1016/j.bbaexp.2003.09.013.

37. Zheng C, Hayes JJ (2003) Structures and interactions of the core histone tail domains. Biopolymers 68: 539–546. doi:10.1002/bip.10303.

38. Bird A (2007) Perceptions of epigenetics. Nature 447: 396–398.

39. Barth TK, Imhof A (2010) Fast signals and slow marks: the dynamics of histone modifications. Trends Biochem Sci 35: 618–626. doi:10.1016/j.tibs.2010.05.006.

40. Lee J-S, Smith E, Shilatifard A (2010) The Language of Histone Crosstalk. Cell 142: 682–685. doi:10.1016/j.cell.2010.08.011.

41. Rando OJ (2012) Combinatorial complexity in chromatin structure and function: revisiting the histone code. Curr Opin Genet Dev 22: 148–155. doi:10.1016/j.gde.2012.02.013.

42. Zhou VW, Goren A, Bernstein BE (2011) Charting histone modifications and the functional organization of mammalian genomes. Nat Rev Genet 12: 7–18. doi:10.1038/nrg2905.

43. Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. Cell Research 21: 381–395. doi:10.1038/cr.2011.22.

44. Kouzarides T (2007) Chromatin Modifications and Their Function. Cell 128: 693–705. doi:10.1016/j.cell.2007.02.005.

45. Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. Cell 128: 707–719.

46. Lee BM, Mahadevan LC (2009) Stability of histone modifications across mammalian genomes: implications for "epigenetic" marking. J Cell Biochem 108: 22–34. doi:10.1002/jcb.22250.

47. Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403: 41–45. doi:10.1038/47412.

48. Henikoff S (2005) Histone modifications: combinatorial complexity or cumulative simplicity? Proc Natl Acad Sci USA 102: 5308–5309. doi:10.1073/pnas.0501853102.

49. Cedar H, Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet 10: 295–304. doi:10.1038/nrg2540.

50. Fuks F, Hurd PJ, Deplus R, Kouzarides T (2003) The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res 31: 2305–2312.

51. Owen-Hughes T (2003) Colworth memorial lecture. Pathways for remodelling chromatin. Biochem Soc Trans 31: 893–905. doi:10.1042/.

52. Saha A, Wittmeyer J, Cairns BR (2006) Chromatin remodelling: the industrial revolution of DNA around histones. Nature Reviews Molecular Cell Biology 7: 437–447. doi:10.1038/nrm1945.

53. Flaus A, Martin DMA, Barton GJ, Owen-Hughes T (2006) Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. Nucleic Acids Res 34: 2887–2905. doi:10.1093/nar/gkl295.

54. Whitehouse I, Stockdale C, Flaus A, Szczelkun MD, Owen-Hughes T (2003) Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme. Mol Cell Biol 23: 1935–1945.

55. Deuring R, Fanti L, Armstrong JA, Sarte M, Papoulas O, et al. (2000) The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. Mol Cell 5: 355–365.

56. Martens JA, Winston F (2003) Recent advances in understanding chromatin remodeling by Swi/Snf complexes. Curr Opin Genet Dev 13: 136–142.

57. Längst G, Becker PB (2004) Nucleosome remodeling: one mechanism, many phenomena? Biochim Biophys Acta 1677: 58–63. doi:10.1016/j.bbaexp.2003.10.011.

58. Satchwell SC, Drew HR, Travers AA (1986) Sequence periodicities in chicken nucleosome core DNA. J Mol Biol 191: 659–675.

59. Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, et al. (2006) A genomic code for nucleosome positioning. Nature 442: 772–778.

60. Gupta S, Dennis J, Thurman RE, Kingston R, Stamatoyannopoulos JA, et al. (2008) Predicting human nucleosome occupancy from primary sequence. PLoSComputBiol 4: e1000134.

61. Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, et al. (2008) The DNA-encoded nucleosome organization of a eukaryotic genome. Nature. Available: PM:19092803.

62. Bai L, Morozov AV (2010) Gene regulation by nucleosome positioning. Trends Genet 26: 476–483. doi:10.1016/j.tig.2010.08.003.

63. Caiafa P, Zampieri M (2005) DNA methylation and chromatin structure: the puzzling CpG islands. JCell Biochem 94: 257–265.

64. Fazzari MJ, Greally JM (2004) Epigenomics: beyond CpG islands. Nat Rev Genet 5: 446–455. doi:10.1038/nrg1349.

65. Bird AP (1980) DNA methylation and the frequency of CpG in animal DNA. Nucleic acids research 8: 1499–1504.

66. Mandrioli M (2007) A new synthesis in epigenetics: towards a unified function of DNA methylation from invertebrates to vertebrates. Cell Mol Life Sci 64: 2522–2524. doi:10.1007/s00018-007-7231-7.

67. Saxonov S, Berg P, Brutlag DL (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. PNAS 103: 1412–1417. doi:10.1073/pnas.0510310103.

68. Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9: 465–476. doi:10.1038/nrg2341.

69. Krieg AM (2006) Therapeutic potential of Toll-like receptor 9 activation. Nat Rev Drug Discov 5: 471–484. doi:10.1038/nrd2059.

70. Illingworth R, Kerr A, Desousa D, Jørgensen H, Ellis P, et al. (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol 6: e22. doi:10.1371/journal.pbio.0060022.

71. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, et al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 38: 1378–1385. doi:10.1038/ng1909.

72. Leeb M, Wutz A (2012) Establishment of epigenetic patterns in development. Chromosoma 121: 251–262. doi:10.1007/s00412-012-0365-x.

73. Morgan HD, Santos F, Green K, Dean W, Reik W (2005) Epigenetic reprogramming in mammals. Hum Mol Genet 14 Spec No 1: R47–58. doi:10.1093/hmg/ddi114.

74. Ji H, Ehrlich LIR, Seita J, Murakami P, Doi A, et al. (2010) Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature 467: 338–342. doi:10.1038/nature09367.

75. Razin A, Riggs AD (1980) DNA methylation and gene function. Science 210: 604–610.

76. Shen L, Kondo Y, Guo Y, Zhang J, Zhang L, et al. (2007) Genome-Wide Profiling of DNA Methylation Reveals a Class of Normally Methylated CpG Island Promoters. PLoS Genet 3: e181. doi:10.1371/journal.pgen.0030181.

77. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, et al. (2010) Dynamic changes in the human methylome during differentiation. Genome Res 20: 320–331. doi:10.1101/gr.101907.109.

78. Ball MP, Li JB, Gao Y, Lee J-H, LeProust EM, et al. (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nature Biotechnology 27: 361–368. doi:10.1038/nbt.1533.

79. Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP (2009) A human B cell methylome at 100-base pair resolution. Proc Natl Acad Sci USA 106: 671–678. doi:10.1073/pnas.0812399106.

80. Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13: 335–340.

81. Smit AF (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev 9: 657–663.

82. Goll MG, Bestor TH (2005) Eukaryotic Cytosine Methyltransferases. Annual Review of Biochemistry 74: 481–514. doi:10.1146/annurev.biochem.74.010904.153721.

83. Jones PA, Liang G (2009) Rethinking how DNA methylation patterns are maintained. Nature Reviews Genetics 10: 805–811. doi:10.1038/nrg2651.

84. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99: 247–257.

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

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

87. Ma DK, Jang M-H, Guo JU, Kitabatake Y, Chang M-L, et al. (2009) Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. Science 323: 1074–1077. doi:10.1126/science.1166859.

88. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16: 6–21.

89. Bird AP (1986) CpG-rich islands and the function of DNA methylation. Nature 321: 209–213. doi:10.1038/321209a0.

90. Antequera F (2003) Structure, function and evolution of CpG island promoters. Cell MolLife Sci 60: 1647–1658.

91. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes Dev 25: 1010–1022. doi:10.1101/gad.2037511.

92. Larsen F, Gundersen G, Lopez R, Prydz H (1992) CpG islands as gene markers in the human genome. Genomics 13: 1095–1107. doi:10.1016/0888-7543(92)90024-M.

93. Davuluri RV, Grosse I, Zhang MQ (2001) Computational identification of promoters and first exons in the human genome. Nat Genet 29: 412–417. doi:10.1038/ng780.

94. Medvedeva YA, Fridman MV, Oparina NJ, Malko DB, Ermakova EO, et al. (2010) Intergenic, gene terminal, and intragenic CpG islands in the human genome. BMC Genomics 11: 48. doi:10.1186/1471-2164-11-48.

95. Blackledge NP, Klose RJ (2011) CpG island chromatin. Epigenetics 6: 147–152. doi:10.4161/epi.6.2.13640.

96. Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, et al. (1994) Sp1 elements protect a CpG island from de novo methylation. Nature 371: 435–438. doi:10.1038/371435a0.

97. Mancini D (1999) Site-specific DNA methylation in the neurofibromatosis (NF1) promoter interferes with binding of CREB and SP1 transcription factors. , Published online: 15 July 1999; | doi:101038/sj.onc1202764 18. Available: http://www.nature.com/onc/journal/v18/n28/full/1202764a.html. Accessed 10 October 2012.

98. Renda M, Baglivo I, Burgess-Beusse B, Esposito S, Fattorusso R, et al. (2007) Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. J Biol Chem 282: 33336–33345. doi:10.1074/jbc.M706213200.

99. Fisscher U, Weisbeek P, Smeekens S (1996) A tobacco nuclear protein that preferentially binds to unmethylated CpG-rich DNA. EurJBiochem 235: 585–592.

100. Voo KS, Carlone DL, Jacobsen BM, Flodin A, Skalnik DG (2000) Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. MolCell Biol 20: 2108–2121.

101. Lee JH, Skalnik DG (2005) 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. JBiolChem 280: 41725–41731.

102. Lee JH, Skalnik DG (2002) CpG-binding protein is a nuclear matrix- and euchromatin-associated protein localized to nuclear speckles containing human trithorax. Identification of nuclear matrix targeting signals. JBiolChem 277: 42259–42267.

103. Bestor TH, Verdine GL (1994) DNA methyltransferases. Curr Opin Cell Biol 6: 380–389.

104. Cross SH, Meehan RR, Nan X, Bird A (1997) A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nat Genet 16: 256–259. doi:10.1038/ng0797-256.

105. Hughes CM, Rozenblatt-Rosen O, Milne TA, Copeland TD, Levine SS, et al. (2004) Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. Mol Cell 13: 587–597.

106. Butler JS, Lee JH, Skalnik DG (2008) CFP1 interacts with DNMT1 independently of association with the Setd1 Histone H3K4 methyltransferase complexes. DNA Cell Biol 27: 533–543.

107. Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, et al. (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 464: 1082–1086. doi:10.1038/nature08924.

108. Vermeulen M, Mulder KW, Denissov S, Pijnappel WWMP, van Schaik FMA, et al. (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell 131: 58–69. doi:10.1016/j.cell.2007.08.016.

109. Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, et al. (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature 442: 86–90. doi:10.1038/nature04815.

110. Blackledge NP, Zhou JC, Tolstorukov MY, Farcas AM, Park PJ, et al. (2010) CpG islands recruit a histone H3 lysine 36 demethylase. Mol Cell 38: 179–190. doi:10.1016/j.molcel.2010.04.009.

111. Tazi J, Bird A (1990) Alternative chromatin structure at CpG islands. Cell 60: 909–920.

112. Croston GE, Kerrigan LA, Lira LM, Marshak DR, Kadonaga JT (1991) Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription. Science 251: 643–649.

113. Thoma F, Koller T, Klug A (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J Cell Biol 83: 403–427.

114. Jones PA, Laird PW (1999) Cancer epigenetics comes of age. Nat Genet 21: 163–167. doi:10.1038/5947.

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

116. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, et al. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393: 386–389. doi:10.1038/30764.

117. Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, et al. (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448: 714–717. doi:10.1038/nature05987.

118. Sugiyama T, Cam H, Verdel A, Moazed D, Grewal SIS (2005) RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. Proc Natl Acad Sci USA 102: 152–157. doi:10.1073/pnas.0407641102.

119. Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, et al. (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev 19: 489–501. doi:10.1101/gad.1248505.

120. Fukagawa T, Nogami M, Yoshikawa M, Ikeno M, Okazaki T, et al. (2004) Dicer is essential for formation of the heterochromatin structure in vertebrate cells. Nature Cell Biology 6: 784–791. doi:10.1038/ncb1155.

121. Brondyk WH (2009) Selecting an appropriate method for expressing a recombinant protein. Meth Enzymol 463: 131–147. doi:10.1016/S0076-6879(09)63011-1.

122. Schirrmann T, Al-Halabi L, Dübel S, Hust M (2008) Production systems for recombinant antibodies. Front Biosci 13: 4576–4594.

123. Wurm FM (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 22: 1393–1398. doi:10.1038/nbt1026.

124. Barnes LM, Dickson AJ (2006) Mammalian cell factories for efficient and stable protein expression. Current Opinion in Biotechnology 17: 381–386. doi:10.1016/j.copbio.2006.06.005.

125. Sheridan C (2011) Gene therapy finds its niche. Nature Biotechnology 29: 121–128. doi:10.1038/nbt.1769.

126. St George JA (2003) Gene therapy progress and prospects: adenoviral vectors. Gene Ther 10: 1135–1141. doi:10.1038/sj.gt.3302071.

127. Lech P, Somia NV (2008) Retrovirus vectors. Contrib Nephrol 159: 30–46. doi:10.1159/000125574.

128. Frankel AD, Young JA (1998) HIV-1: fifteen proteins and an RNA. Annu Rev Biochem 67: 1–25. doi:10.1146/annurev.biochem.67.1.1.

129. Naldini L, Blömer U, Gallay P, Ory D, Mulligan R, et al. (1996) In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science 272: 263–267.

130. Mátrai J, Chuah MKL, VandenDriessche T (2010) Recent advances in lentiviral vector development and applications. Mol Ther 18: 477–490. doi:10.1038/mt.2009.319.

131. Zufferey R, Dull T, Mandel RJ, Bukovsky A, Quiroz D, et al. (1998) Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. J Virol 72: 9873–9880.

132. Klug CA, Cheshier S, Weissman IL (2000) Inactivation of a GFP retrovirus occurs at multiple levels in long-term repopulating stem cells and their differentiated progeny. Blood 96: 894–901.

133. Bestor TH (2000) Gene silencing as a threat to the success of gene therapy. JClinInvest 105: 409–411.

134. Mok HP, Javed S, Lever A (2007) Stable gene expression occurs from a minority of integrated HIV-1-based vectors: transcriptional silencing is present in the majority. Gene Ther 14: 741–751. doi:10.1038/sj.gt.3302923.

135. Antoniou M, Harland L, Mustoe T, Williams S, Holdstock J, et al. (2003) Transgenes encompassing dual-promoter CpG islands from the human TBP and HNRPA2B1 loci are resistant to heterochromatin-mediated silencing. Genomics 82: 269–279.

136. Williams S, Mustoe T, Mulcahy T, Griffiths M, Simpson D, et al. (2005) CpG-island fragments from the HNRPA2B1/CBX3 genomic locus reduce silencing and enhance transgene expression from the hCMV promoter/enhancer in mammalian cells. BMC Biotechnol 5: 17. doi:10.1186/1472-6750-5-17.

137. Gill DR, Pringle IA, Hyde SC (2009) Progress and prospects: the design and production of plasmid vectors. Gene Ther 16: 165–171. doi:10.1038/gt.2008.183.

138. Wolff JA, Malone RW, Williams P, Chong W, Acsadi G, et al. (1990) Direct gene transfer into mouse muscle in vivo. Science 247: 1465–1468.

139. Choate KA, Khavari PA (1997) Direct cutaneous gene delivery in a human genetic skin disease. Hum Gene Ther 8: 1659–1665. doi:10.1089/hum.1997.8.14-1659.

140. Hickman MA, Malone RW, Lehmann-Bruinsma K, Sih TR, Knoell D, et al. (1994) Gene expression following direct injection of DNA into liver. Hum Gene Ther 5: 1477–1483. doi:10.1089/hum.1994.5.12-1477.

141. Habib NA, Ding SF, el-Masry R, Mitry RR, Honda K, et al. (1996) Preliminary report: the short-term effects of direct p53 DNA injection in primary hepatocellular carcinomas. Cancer Detect Prev 20: 103–107.

142. Prausnitz MR, Mikszta JA, Cormier M, Andrianov AK (2009) Microneedle-based vaccines. Curr Top Microbiol Immunol 333: 369–393. doi:10.1007/978-3-540-92165-3_18.

143. Neumann E, Schaefer-Ridder M, Wang Y, Hofschneider PH (1982) Gene transfer into mouse lyoma cells by electroporation in high electric fields. EMBO J 1: 841–845.

144. O'Brien J, Lummis SCR (2002) An improved method of preparing microcarriers for biolistic transfection. Brain Res Brain Res Protoc 10: 12–15.

145. Kim HJ, Greenleaf JF, Kinnick RR, Bronk JT, Bolander ME (1996) Ultrasound-mediated transfection of mammalian cells. Hum Gene Ther 7: 1339–1346. doi:10.1089/hum.1996.7.11-1339.

146. Tata DB, Dunn F, Tindall DJ (1997) Selective clinical ultrasound signals mediate differential gene transfer and expression in two human prostate cancer cell lines: LnCap and PC-3. Biochem Biophys Res Commun 234: 64–67. doi:10.1006/bbrc.1997.6578.

147. Kamimura K, Suda T, Zhang G, Liu D (2011) Advances in Gene Delivery Systems. Pharmaceut Med 25: 293–306. doi:10.2165/11594020-000000000-00000.

148. Felgner PL, Gadek TR, Holm M, Roman R, Chan HW, et al. (1987) Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. Proc Natl Acad Sci USA 84: 7413–7417.

149. Boussif O, Lezoualc'h F, Zanta MA, Mergny MD, Scherman D, et al. (1995) A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. Proc Natl Acad Sci USA 92: 7297–7301.

150. Haensler J, Szoka FC Jr (1993) Polyamidoamine cascade polymers mediate efficient transfection of cells in culture. Bioconjug Chem 4: 372–379.

151. Nimesh S, Kumar R, Chandra R (2006) Novel polyallylamine-dextran sulfate-DNA nanoplexes: highly efficient non-viral vector for gene delivery. Int J Pharm 320: 143–149. doi:10.1016/j.ijpharm.2006.03.050.

152. Erbacher P, Zou S, Bettinger T, Steffan AM, Remy JS (1998) Chitosan-based vector/DNA complexes for gene delivery: biophysical characteristics and transfection ability. Pharm Res 15: 1332–1339.

153. Wang Z, Jin L, Yuan Z, Wegrzyn G, Wegrzyn A (2009) Classification of plasmid vectors using replication origin, selection marker and promoter as criteria. Plasmid 61: 47–51. doi:10.1016/j.plasmid.2008.09.003.

154. Mayrhofer P, Blaesen M, Schleef M, Jechlinger W (2008) Minicircle-DNA production by site specific recombination and protein-DNA interaction chromatography. J Gene Med 10: 1253–1269. doi:10.1002/jgm.1243.

155. Hagedorn C, Wong S-P, Harbottle R, Lipps HJ (2011) Scaffold/Matrix Attached Region-Based Nonviral Episomal Vectors. Human Gene Therapy 22: 915–923. doi:10.1089/hum.2011.084.

156. Poulsen TT, Pedersen N, Juel H, Poulsen HS (2008) A chimeric fusion of the hASH1 and EZH2 promoters mediates high and specific reporter and suicide gene expression and cytotoxicity in small cell lung cancer cells. Cancer Gene Ther 15: 563–575. doi:10.1038/cgt.2008.24.

157. Izumo T, Ohtsuru A, Tokunaga Y, Namba H, Kaneda Y, et al. (2007) Epstein-Barr virus-based vector improves the tumor cell killing effect of pituitary tumor in HVJ-liposome-mediated transcriptional targeting suicide gene therapy. Int J Oncol 31: 379–387.

158. Fath S, Bauer AP, Liss M, Spriestersbach A, Maertens B, et al. (2011) Multiparameter RNA and codon optimization: a standardized tool to assess and enhance autologous mammalian gene expression. PLoS ONE 6: e17596. doi:10.1371/journal.pone.0017596.

159. Graf M, Deml L, Wagner R (2004) Codon-optimized genes that enable increased heterologous expression in mammalian cells and elicit efficient immune responses in mice after vaccination of naked DNA. Methods Mol Med 94: 197–210.

160. Sharp PM, Li WH (1987) The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15: 1281–1295.

161. Hyde SC, Pringle IA, Abdullah S, Lawton AE, Davies LA, et al. (2008) CpG-free plasmids confer reduced inflammation and sustained pulmonary gene expression. Nat Biotechnol 26: 549–551. doi:10.1038/nbt1399.

162. Kosovac D, Wild J, Ludwig C, S M, Bauer AP, et al. (2010) Minimal doses of a sequence-optimized transgene mediate high level and longterm EPO expression in vivo: challenging CpG-free gene design. Gene Ther.

163. Jechlinger W (2006) Optimization and delivery of plasmid DNA for vaccination. Expert Rev Vaccines 5: 803–825. doi:10.1586/14760584.5.6.803.

164. Carpentier E, Paris S, Kamen AA, Durocher Y (2007) Limiting factors governing protein expression following polyethylenimine-mediated gene transfer in HEK293-EBNA1 cells. J Biotechnol 128: 268–280. doi:10.1016/j.jbiotec.2006.10.014.

165. Karimi M, Goldie LC, Ulgiati D, Abraham LJ (2007) Integration site-specific transcriptional reporter gene analysis using Flp recombinase targeted cell lines. BioTechniques 42: 217–224.

166. Wiberg FC, Rasmussen SK, Frandsen TP, Rasmussen LK, Tengbjerg K, et al. (2006) Production of target-specific recombinant human polyclonal antibodies in mammalian cells. Biotechnol Bioeng 94: 396–405. doi:10.1002/bit.20865.

167. Huang Y, Li Y, Wang YG, Gu X, Wang Y, et al. (2007) An efficient and targeted gene integration system for high-level antibody expression. J Immunol Methods 322: 28–39. doi:10.1016/j.jim.2007.01.022.

168. Lu H, Khurana S, Verma N, Manischewitz J, King L, et al. (2011) A Rapid Flp-In System for Expression of Secreted H5N1 Influenza Hemagglutinin Vaccine Immunogen in Mammalian Cells. PLoS ONE 6: e17297. doi:10.1371/journal.pone.0017297.

169. Leikam D (2006) Die Rolle von CpG-Dinukleotiden bei der Regulation der Transgenexpression am Beispiel verschiedener Reportergene Universit„tsbibliothek der Universit„t Regensburg. Available: http://www.opus-bayern.de/uni-regensburg/volltexte/2006/696.

170. Bauer AP (2008) Die Bedeutung intragenischer CpG Dinukleotide für die in vitro Expression von Immunmodulatoren Regensburg.

171. Zolotukhin S, Potter M, Hauswirth WW, Guy J, Muzyczka N (1996) A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells. JVirol 70: 4646–4654.

172. Bauer AP, Leikam D, Krinner S, Notka F, Ludwig C, et al. (2010) The impact of intragenic CpG content on gene expression. Nucleic Acids Res 38: 3891–3908.

173. Braund R, Hook S, Medlicott NJ (2007) The role of topical growth factors in chronic wounds. Curr Drug Deliv 4: 195–204.

174. Taylor-Robinson A (2001) Schistosomiasis-induced IL-10 suppresses allergy prevalence. Trends Parasitol 17: 62.

175. Andrew ME (1991) Cytokines--a way towards better animal health. Aust Vet J 68: 81.

176. Karin N (2004) Induction of protective therapy for autoimmune diseases by targeted DNA vaccines encoding pro-inflammatory cytokines and chemokines. Curr Opin Mol Ther 6: 27–33.

177. Edwards CJ, Feldman JL, Beech J, Shields KM, Stover JA, et al. (2007) Molecular profile of peripheral blood mononuclear cells from patients with rheumatoid arthritis. Mol Med 13: 40–58. doi:10.2119/2006-000056.Edwards.

178. O'Gorman S, Fox DT, Wahl GM (1991) Recombinase-mediated gene activation and site-specific integration in mammalian cells. Science 251: 1351–1355.

179. Baldi L, Hacker DL, Adam M, Wurm FM (2007) Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. BiotechnolLett 29: 677–684.

180. Hunt AG, Xu R, Addepalli B, Rao S, Forbes KP, et al. (2008) Arabidopsis mRNA polyadenylation machinery: comprehensive analysis of protein-protein interactions and gene expression profiling. BMC Genomics 9: 220. doi:10.1186/1471-2164-9-220.

181. Clark SJ, Harrison J, Paul CL, Frommer M (1994) High sensitivity mapping of methylated cytosines. Nucleic Acids Res 22: 2990–2997.

182. Jiang M, Zhang Y, Fei J, Chang X, Fan W, et al. (2010) Rapid quantification of DNA methylation by measuring relative peak heights in direct bisulfite-PCR sequencing traces. Laboratory Investigation 90: 282–290. doi:10.1038/labinvest.2009.132.

183. Schneider R, Bannister AJ, Myers FA, Thorne AW, Crane-Robinson C, et al. (2004) Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nat Cell Biol 6: 73–77. doi:10.1038/ncb1076.

184. Kaplan N, Hughes TR, Lieb JD, Widom J, Segal E (2010) Contribution of histone sequence preferences to nucleosome organization: proposed definitions and methodology. Genome Biology 11: 140. doi:10.1186/gb-2010-11-11-140.

185. Rhodes D, Laskey RA (1989) Assembly of nucleosomes and chromatin in vitro. Methods Enzymol 170: 575–585.

186. Krinner S (2009) The influence of intragenic CpG dinucleotides on gene expression and chromatin structure.

187. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, Gross F, Yvon E, et al. (2000) Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. Science 288: 669–672.

188. Ott MG, Schmidt M, Schwarzwaelder K, Stein S, Siler U, et al. (2006) Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat Med 12: 401–409. doi:10.1038/nm1393.

189. Randell SH, Fulcher ML, O'Neal W, Olsen JC (2011) Primary epithelial cell models for cystic fibrosis research. Methods Mol Biol 742: 285–310. doi:10.1007/978-1-61779-120-8_18.

190. Isacson O, Costantini L, Schumacher JM, Cicchetti F, Chung S, et al. (2001) Cell implantation therapies for Parkinson's disease using neural stem, transgenic or xenogeneic donor cells. Parkinsonism Relat Disord 7: 205–212.

191. Deacon T, Dinsmore J, Costantini LC, Ratliff J, Isacson O (1998) Blastula-stage stem cells can differentiate into dopaminergic and serotonergic neurons after transplantation. Exp Neurol 149: 28–41. doi:10.1006/exnr.1997.6674.

192. Brüstle O, Jones KN, Learish RD, Karram K, Choudhary K, et al. (1999) Embryonic stem cell-derived glial precursors: a source of myelinating transplants. Science 285: 754–756.

193. Humpherys D, Eggan K, Akutsu H, Hochedlinger K, Rideout WM 3rd, et al. (2001) Epigenetic instability in ES cells and cloned mice. Science 293: 95–97. doi:10.1126/science.1061402.

194. Cherry SR, Biniszkiewicz D, van Parijs L, Baltimore D, Jaenisch R (2000) Retroviral expression in embryonic stem cells and hematopoietic stem cells. Mol Cell Biol 20: 7419–7426.

195. Zhang F, Frost AR, Blundell MP, Bales O, Antoniou MN, et al. (2010) A ubiquitous chromatin opening element (UCOE) confers resistance to DNA methylation-mediated silencing of lentiviral vectors. Mol Ther 18: 1640–1649. doi:10.1038/mt.2010.132.

196. Gardner RS, Wahba AJ, Basilio C, Miller RS, Lengyel P, et al. (1962) SYNTHETIC POLYNUCLEOTIDES AND THE AMINO ACID CODE, VII*. Proc Natl Acad Sci U S A 48: 2087–2094.

197. Bird A, Taggart M, Frommer M, Miller OJ, Macleod D (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell 40: 91–99.

198. Kundu TK, Rao MR (1999) CpG islands in chromatin organization and gene expression. J Biochem 125: 217–222.

199. Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. J Mol Biol 196: 261–282.

200. Bird AP, Wolffe AP (1999) Methylation-induced repression--belts, braces, and chromatin. Cell 99: 451–454.

201. de Wolf HK, Johansson N, Thong A-T, Snel CJ, Mastrobattista E, et al. (2008) Plasmid CpG Depletion Improves Degree and Duration of Tumor Gene Expression After Intravenous Administration of Polyplexes. Pharm Res 25: 1654–1662. doi:10.1007/s11095-008-9558-7.

202. Yew NS, Zhao H, Przybylska M, Wu IH, Tousignant JD, et al. (2002) CpG-depleted plasmid DNA vectors with enhanced safety and long-term gene expression in vivo. MolTher 5: 731–738.

203. Yew NS (2005) Controlling the kinetics of transgene expression by plasmid design. AdvDrug DelivRev 57: 769–780.

204. Chevalier-Mariette C, Henry I, Montfort L, Capgras S, Forlani S, et al. (2003) CpG content affects gene silencing in mice: evidence from novel transgenes. Genome Biol 4: R53.

205. Dalle B, Rubin JE, Alkan O, Sukonnik T, Pasceri P, et al. (2005) eGFP reporter genes silence LCRbeta-globin transgene expression via CpG dinucleotides. MolTher 11: 591–599.

206. Duan J, Antezana MA (2003) Mammalian mutation pressure, synonymous codon choice, and mRNA degradation. J Mol Evol 57: 694–701. doi:10.1007/s00239-003-2519-1.

207. Mitchell P, Tollervey D (2000) mRNA stability in eukaryotes. Current Opinion in Genetics & Development 10: 193–198. doi:10.1016/S0959-437X(00)00063-0.

208. Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M (2006) High Guanine and Cytosine Content Increases mRNA Levels in Mammalian Cells. PLoS Biol 4. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1463026/. Accessed 11 October 2012.

209. Hammill L, Welles J, Carson GR (2000) The gel microdrop secretion assay: Identification of a low productivity subpopulation arising during the production of human antibody in CHO cells. Cytotechnology 34: 27–37. doi:10.1023/A:1008186113245.

210. Rodolosse A, Barbat A, Chantret I, Lacasa M, Brot-Laroche E, et al. (1998) Selecting agent hygromycin B alters expression of glucose-regulated genes in transfected Caco-2 cells. Am J Physiol 274: G931–938.

211. Kaufman WL, Kocman I, Agrawal V, Rahn H-P, Besser D, et al. (2008) Homogeneity and persistence of transgene expression by omitting antibiotic selection in cell line isolation. Nucleic Acids Res 36: e111. doi:10.1093/nar/gkn508.

212. Tollefsbol TO, Hutchison III CA (1997) Control of methylation spreading in synthetic DNA sequences by the murine DNA methyltransferase. Journal of Molecular Biology 269: 494–504. doi:10.1006/jmbi.1997.1064.

213. Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, et al. (2009) A Unifying Model for the Selective Regulation of Inducible Transcription by CpG Islands and Nucleosome Remodeling. Cell 138: 114–128. doi:10.1016/j.cell.2009.04.020.

214. Widom J (2001) Role of DNA sequence in nucleosome stability and dynamics. Q Rev Biophys 34: 269–324.

215. Lowary PT, Widom J (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol 276: 19–42. doi:10.1006/jmbi.1997.1494.

216. Gencheva M, Boa S, Fraser R, Simmen MW, A Whitelaw CB, et al. (2006) In Vitro and in Vivo nucleosome positioning on the ovine beta-lactoglobulin gene are related. J Mol Biol 361: 216–230. doi:10.1016/j.jmb.2006.06.039.

217. Hertel CB, Längst G, Hörz W, Korber P (2005) Nucleosome stability at the yeast PHO5 and PHO8 promoters correlates with differential cofactor requirements for chromatin opening. Mol Cell Biol 25: 10755–10767. doi:10.1128/MCB.25.24.10755-10767.2005.

218. Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, et al. (2008) Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. PLoS Comput Biol 4: e1000216. doi:10.1371/journal.pcbi.1000216.

219. Kininis M, Isaacs GD, Core LJ, Hah N, Kraus WL (2009) Postrecruitment Regulation of RNA Polymerase II Directs Rapid Signaling Responses at the Promoters of Estrogen Target Genes. Mol Cell Biol 29: 1123–1133. doi:10.1128/MCB.00841-08.

220. Choi JK, Bae J-B, Lyu J, Kim T-Y, Kim Y-J (2009) Nucleosome deposition and DNA methylation at coding region boundaries. Genome Biol 10: R89. doi:10.1186/gb-2009-10-9-r89.

221. Lei H, Oh SP, Okano M, Jüttermann R, Goss KA, et al. (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development 122: 3195–3205.

222. Modlich U, Bohne J, Schmidt M, von Kalle C, Knöss S, et al. (2006) Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity. Blood 108: 2545–2553. doi:10.1182/blood-2005-08-024976.

223. McBurney MW, Jones-Villeneuve EM, Edwards MK, Anderson PJ (1982) Control of muscle and neuronal differentiation in a cultured embryonal carcinoma cell line. Nature 299: 165–167.

224. Jones-Villeneuve EM, McBurney MW, Rogers KA, Kalnins VI (1982) Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. J Cell Biol 94: 253–262.

225. Astigiano S, Damonte P, Fossati S, Boni L, Barbieri O (2005) Fate of embryonal carcinoma cells injected into postimplantation mouse embryos. Differentiation 73: 484–490. doi:10.1111/j.1432-0436.2005.00043.x.

226. Keating A, Horsfall W, Hawley RG, Toneguzzo F (1990) Effect of different promoters on expression of genes introduced into hematopoietic and marrow stromal cells by electroporation. Exp Hematol 18: 99–102.

227. Muller SR, Sullivan PD, Clegg DO, Feinstein SC (1990) Efficient transfection and expression of heterologous genes in PC12 cells. DNA Cell Biol 9: 221–229.

228. Qin JY, Zhang L, Clift KL, Hulur I, Xiang AP, et al. (2010) Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. PLoS ONE 5: e10611. doi:10.1371/journal.pone.0010611.

229. Chung S, Andersson T, Sonntag K-C, Björklund L, Isacson O, et al. (2002) Analysis of Different Promoter Systems for Efficient Transgene Expression in Mouse Embryonic Stem Cell Lines. STEM CELLS 20: 139–145. doi:10.1634/stemcells.20-2-139.

230. Varma NRS, Janic B, Ali MM, Iskander A, Arbab AS (2011) Lentiviral Based Gene Transduction and Promoter Studies in Human Hematopoietic Stem Cells (hHSCs). J Stem Cells Regen Med 7: 41–53.

231. Ovchinnikov DA, Turner JP, Titmarsh DM, Thakar NY, Sin DC, et al. (2012) Generation of a human embryonic stem cell line stably expressing high levels of the fluorescent protein mCherry. World J Stem Cells 4: 71–79. doi:10.4252/wjsc.v4.i7.71.

232. Sinn PL, Sauter SL, McCray PB Jr (2005) Gene therapy progress and prospects: development of improved lentiviral and retroviral vectors--design, biosafety, and production. Gene Ther 12: 1089–1098. doi:10.1038/sj.gt.3302570.

233. Zhang F, Thornhill SI, Howe SJ, Ulaganathan M, Schambach A, et al. (2007) Lentiviral vectors containing an enhancer-less ubiquitously acting chromatin opening element (UCOE) provide highly reproducible and stable transgene expression in hematopoietic cells. Blood 110: 1448–1457. doi:10.1182/blood-2006-12-060814.

234. Zhou Q, Li T, Price DH (2012) RNA polymerase II elongation control. Annu Rev Biochem 81: 119–143. doi:10.1146/annurev-biochem-052610-095910.

235. Lee JH, Voo KS, Skalnik DG (2001) Identification and characterization of the DNA binding domain of CpG-binding protein. JBiolChem 276: 44669–44676.

236. Carlone DL, Skalnik DG (2001) CpG binding protein is crucial for early embryonic development. MolCell Biol 21: 7601–7606.

237. Tate CM, Lee J-H, Skalnik DG (2010) CXXC finger protein 1 restricts the Setd1A histone H3K4 methyltransferase complex to euchromatin. FEBS J 277: 210–223. doi:10.1111/j.1742-4658.2009.07475.x.

238. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. AnalBiochem 72: 248–254.

239. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD (2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res 17: 877–885. doi:10.1101/gr.5533506.

9 Appendix

9.1 List of abbreviations

A	
5'aza	5-Azacytidine -2'-deoxycytidine
A2UCOE	heterogeneous nuclear ribonucleoprotein A2/B1-chromobox homolog 3 ubiquitously chromatin opening element
ApG	adenin guanine dinucleotide
APS	ammoniumperoxidsulfat
B	
BB	bacterial backbone
BGH	bovine growth hormone
bp	base pair
BSA	bovine serum albumin
C	
CAI	codon adaptation index
CAP	chromatin associated protein
CBX3	chromobox homolog 3
CFP-1	CXXC finger protein 1
CGBP-1	CpG-binding protein 1
CHD	chromodomain-helicase-DNA binding
ChIP	chromatin immunoprecipitation
CHO	chinese hamster ovaries
CMV	cytomegalovirus
CP	crossing point
CpG	cytosine guanine dinucleotides
CpT	cytosine thymine dinucleotides
CREB	cAMP response element-*binding* protein
CTCF	CCCTC binding factor
CTD	Carboxy-terminal domain
D	
DBD	DNA binding domain
DMEM	Dullbecco's modified eagle medium
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DPE	downstream promoter element
DRB	5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole

DRE	distal regulatory element
DSIF	DRB sensitivity-inducing factor
E	
E	efficiency
EC	expression cassette
EDTA	ethylenediaminetetraacetic acid
EF-1α	elongation factor 1α
EGTA	ethylene glycol tetraacetic acid
ELISA	enzyme linked immunosorbent assay
ELL	eleven-nineteen lysine-rich
F	
FACS	fluorescence activated cell sorting
FAIRE	formaldehyde assisted isolation of regulatory elements
FBS	fetal bovine serum
Flp	flippase
FRT	flp recombination target
G	
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GFP	green fluorescent protein
GTF	general transcription factor
H	
HAT	histone acetyl transferase
hCGBP	human CpG binding protein
HDAC	histone deacetylase
HEK 293	human embryonic kidney cells 293
hGFP	humanized green fluorescent protein
HIV	human immunodeficiency virus
HNRPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1
hph	hygromycin resistance gene
HRP	horseradish peroxidase
I	
IgG	immunglobulin G
IGR	intergenic region
Inr	initiator
IOD	integrated optical density
ISWI	imitation switch

K	
kb	kilobase
kDa	kilo Dalton
KDM2A	lysine demethylase enzyme 2A
L	
lacZ	encodes ß-galactosidase
LB	Luria Bertani
LINE	long interspersed nuclear elements
LTR	long terminal repeat
LV	lentivirus
M	
M	molar
MBD	methyl CpG-binding domain proteins
MBD	methyl-DNA binding domain
ME	mercaptoethanol
MeCP	methyl CpG binding protein
MED	mediator complex
mEPO	murine erythropoietin
MFI	mean fluorescence intensity
MHC	major histocompatibility complex
min	minutes
mMIP-1α	murine macrophage inflammatory protein 1α
MOI	multiplicity of infection
mRNA	messenger RNA
mTERT	endogenous mouse telomerase reverse transcriptase
N	
NaB	natrium butyrate
NELF	negative elongation factor
NFR	nucleosome free region
nm	nanometer
NURD	nucleosome remodeling and histone deacetylation
NURF	nucleosome remodeling factor
O	
ORF	open reading frame
P	
pA	polyadenylation
PAA	polyacrylamide
PAGE	polyacrylamide gel electrophoresis

PBS	phosphate buffered saline
PBS-T	phosphate buffered saline Tween-20
PCR	polymerase chain reaction
PEI	polyethylenimine
PHD	plant homeodomain
PIC	preinitiation complex
Pol	polymerase
Q	
qPCR	quantitative PCR
R	
rdna	DNA coding for ribosomal RNA
RIPA	radioimmunoprecipitation analysis
RNA	ribonucleic acid
RNAPII	RNA Polymerase II
rRNA	ribosomal RNA
RT	room temperature
rtPCR	Real-time PCR
S	
S/MAR	scaffold matrix attachment region
SDS	sodium dodecyl sulfate
sec	seconds
Ser2P	phosphorylation of serine 2 within the RNA polymerase II C-terminal domain
Ser5P	phosphorylation of serine 5 within the RNA polymerase II C-terminal domain
SFII	superfamily II
SINE	long interspersed nuclear elements
SIN-LV	self inactivating lentiviral vector
Sp1	specificity Protein 1
ß2-M	ß2-Microglobulin
SUV39H1	histone-lysine N-methyltransferase
SV40	simian virus 40
SWI/SNF	switch 2/sucrose-non-fermenting
T	
TF	transcription factor
TLR9	toll like receptor
TpA	thymine adenine dinucleotide
TpC	thymine cytosine dinucleotide
TRAP	thyroid hormone receptor associated protein
tRNA	transfer RNA

TSS	transcription start site
TTS	transcription termination site
UCOE	ubiquitously chromatin opening element
UTR	untranslated region
VCN	vector copy number
WHP	woodchuck hepatitis virus
WPRE	woodchuck hepatitis virus (WHP) posttranscriptional regulatory element
X	
X-Gal	5-Bromo-4-chloro-3-indolyl-ß-D-Galactopyranosid

9.2 Sequences

9.2.1 Murine MIP-1α variants

mmip-1α 7CpG (mMIP-wt)

```
1  ATGAAGGTCT CCACCACTGC CCTTGCTGTT CTTCTCTGTA CCATGACACT CTGCAACCAA
61 GTCTTCTCAG CGCCATATGG AGCTGACACC CCGACTGCCT GCTGCTTCTC CTACAGCCGG
121 AAGATTCCAC GCCAATTCAT CGTTGACTAT TTTGAAACCA GCAGCCTTTG CTCCCAGCCA
181 GGTGTCATTT TCCTGACTAA GAGAAACCGG CAGATCTGCG CTGACTCCAA AGAGACCTGG
241 GTCCAAGAAT ACATCACTGA CCTGGAACTG AATGCCTAG
```

mmip-1α 0CpG (mMIP-0)

```
1  ATGAAGGTGA GCACAACAGC TCTGGCTGTG CTGCTGTGTA CCATGACCCT GTGCAACCAG
61 GTGTTCTCTG CCCCTTATGG AGCAGATACC CCTACAGCCT GCTGTTTCAG CTACAGCAGG
121 AAGATCCCCA GGCAGTTCAT TGTGGACTAC TTTGAGACCA GCAGCCTGTG TTCTCAGCCT
181 GGGGTGATCT TTCTGACCAA GAGGAACAGG CAGATCTGTG CAGACAGCAA GGAGACATGG
241 GTGCAGGAGT ACATCACAGA CCTGGAGCTG AATGCCTAG
```

mmip-1α 13CpG (mMIP-13

```
1  ATGAAGGTGA GCACCACAGC TCTGGCTGTG CTGCTGTGCA CCATGACCCT GTGCAACCAG
61 GTGTTCAGCG CTCCTTACGG CGCCGATACC CCTACAGCCT GCTGCTTCAG CTACAGCAGG
121 AAGATCCCCA GGCAGTTCAT CGTGGACTAC TTCGAGACCA GCAGCCTGTG TTCTCAGCCC
181 GGCGTGATCT TCCTGACCAA GCGGAACAGA CAGATCTGCG CCGACAGCAA GGAGACATGG
241 GTGCAGGAGT ACATCACCGA CCTGGAGCTG AACGCCTAG
```

mmip-1α 42CpG (mMIP-42)

```
1  ATGAAGGTGT CGACGACCGC GCTCGCCGTG CTGCTGTGCA CGATGACGCT GTGCAACCAG
61 GTGTTCAGCG CCCCGTACGG CGCCGACACG CCGACCGCGT GCTGCTTCTC GTACTCGCGG
121 AAGATCCCGC GGCAGTTCAT CGTCGACTAC TTCGAAACGT CGTCGCTGTG CTCGCAGCCC
181 GGCGTGATCT TCCTCACGAA GCGGAACCGG CAGATCTGCG CCGACTCGAA GGAAACGTGG
241 GTGCAGGAGT ACATCACCGA CCTCGAACTG AACGCGTAG
```

9.2.2 Humanized GFP variants

hgfp oCpG (hGFP-o)

```
1  ATGGTGTCCA AGGGGGAGGA GCTGTTCACA GGGGTGGTGC CCATCCTGGT GGAGCTGGAT
61 GGGGATGTGA ATGGCCACAA GTTCTCTGTG TCTGGGGAGG GGGAGGGGGA TGCCACCTAT
121 GGCAAGCTCA CCCTGAAGTT CATCTGCACC ACAGGCAAGC TGCCAGTGCC CTGGCCCACC
181 CTGGTGACCA CCTTCACCTA TGGGGTGCAG TGCTTCAGCA GATACCCAGA CCACATGAAG
241 CAGCATGACT TCTTCAAGTC TGCCATGCCT GAGGGCTATG TGCAGGAGAG GACCATCTTC
301 TTCAAGGATG ATGGCAACTA CAAGACCAGG GCTGAGGTGA AGTTTGAGGG GGATACCCTG
361 GTGAACAGGA TTGAGCTGAA GGGCATTGAC TTTAAGGAGG ATGGCAATAT CCTGGGCCAC
421 AAGCTGGAGT ACAACTACAA CAGCCACAAT GTGTACATCA TGGCAGACAA GCAGAAGAAT
481 GGCATCAAGG TGAACTTCAA GATCAGGCAC AACATTGAGG ATGGCTCTGT GCAGCTGGCA
541 GACCACTACC AGCAGAACAC CCCCATTGGA GATGGCCCTG TCCTGCTGCC AGACAACCAC
601 TACCTGAGCA CCCAGTCTGC CCTGAGCAAG GACCCCAATG AGAAGAGGGA CCACATGGTG
661 CTGCTGGAGT TTGTGACAGC TGCTGGCATC ACCCTGGGCA TGGATGAGCT GTACAAGTGA
```

hgfp 60CpG (hGFP-60)

```
1  ATGGTGAGCA AGGGCGAGGA GCTGTTCACC GGGGTGGTGC CCATCCTGGT CGAGCTGGAC
61 GGCGACGTAA ACGGCCACAA GTTCAGCGTG TCCGGCGAGG GCGAGGGCGA TGCCACCTAC
121 GGCAAGCTGA CCCTGAAGTT CATCTGCACC ACCGGCAAGC TGCCCGTGCC CTGGCCCACC
181 CTCGTGACCA CCTTCACCTA CGGCGTGCAG TGCTTCAGCC GCTACCCCGA CCACATGAAG
241 CAGCACGACT TCTTCAAGTC CGCCATGCCC GAAGGCTACG TCCAGGAGCG CACCATCTTC
301 TTCAAGGACG ACGGCAACTA CAAGACCCGC GCCGAGGTGA AGTTCGAGGG CGACACCCTG
361 GTGAACCGCA TCGAGCTGAA GGGCATCGAC TTCAAGGAGG ACGGCAACAT CCTGGGGCAC
421 AAGCTGGAGT ACAACTACAA CAGCCACAAC GTCTATATCA TGGCCGACAA GCAGAAGAAC
481 GGCATCAAGG TGAACTTCAA GATCCGCCAC AACATCGAGG ACGGCAGCGT GCAGCTCGCC
541 GACCACTACC AGCAGAACAC CCCCATCGGC GACGGCCCCG TGCTGCTGCC CGACAACCAC
601 TACCTGAGCA CCCAGTCCGC CCTGAGCAAA GACCCCAACG AGAAGCGCGA TCACATGGTC
661 CTGCTGGAGT TCGTGACCGC CGCCGGGATC ACTCTCGGCA TGGACGAGCT GTACAAGTAA
```

10 Danksagung

An erster Stelle möchte ich mich sehr herzlich bei Herrn Prof. Dr. Ralf Wagner für die Bereitstellung und die kompetente Betreuung dieser Promotionsarbeit, die fachlichen Ratschläge und großzügige Unterstützung bezüglich Kongressreisen, Teilnahme an Fortbildungen oder ähnlichen Veranstaltungen bedanken.

Ein ebenso herzliches Dankeschön gilt meinen Mentoren Prof. Dr. Gernot Längst und Prof. Dr. Christopher Baum für das große Interesse an meiner Arbeit und die sehr hilfreichen Anregungen die maßgeblich zum Gelingen dieser Promotion beigetragen haben. Vielen Dank an Herrn Prof. Baum für die bereitwillige Übernahme des Zweitgutachtens, und an Herrn Prof. Längst für die zahlreichen fachlichen Diskussionen und Korrekturen von Manuskripten.

Herrn Prof. Dr. Hans Wolf und Herrn Prof. Dr. Dr. André Gessner möchte ich für die Bereitstellung des Arbeitsplatzes am Institut für Medizinische Mikrobiologie und Hygiene am Universitätsklinikum Regensburg danken.

Ganz besonders danken möchte ich Dr. Asli Heizer für ihre fachliche Unterstützung und den moralischen Beistand während dieser Arbeit, der auch über eine Distanz von 7000 km und teilweise suboptimalen afrikanischen Empfangsbedingungen nicht abriss. Vielen Dank an Dr. Asli Heizer, Dr. Philip Rosenstock, Dr. Ingrid Obermeier und Dr. Benedikt Asbach für das Korrekturlesen dieser Arbeit.

Chicas y chico del laboratorio 66, ich habe euch jeden Tag seit eurem sukzessiven Auszug aus dem Labor des Wahnsinns vermisst. Danke für die unvergessliche Zeit! Auch allen übrigen (Ex-)Kollegen der AG Wagner möchte ich danken für das unkomplizierte und freundliche Arbeitsklima, die vielen lustigen Unternehmungen und ausgelassenen Momente in unserem „Sozialflur“ und auf der „Beauty Farm“.

Liebe (Ex-)Tannis & Friends, beste WG der Welt! Muchas gracias dass ihr mich mit eurer unverbesserlichen Art das Leben nicht ernst zu nehmen die bisweilen auftretenden Frustrationen immer ganz schnell vergessen habt lassen.

Vielen herzlichen Dank an meine Eltern, dass ihr mich bedingungslos auf meinem Lebensweg unterstützt und mir in schwierigen Phasen den Rücken freigehalten habt.

Lieber Jürgen, dir möchte ich ganz aufrichtig danken für deine Geduld und Unterstützung, die aufmunternden Gesten in Momenten der Verzweiflung und dafür dass du immer an mich geglaubt hast.