

Dissertation der Fakultät für Biologie der Ludwig-Maximilians-Universität München zur Erlangung des Doktorgrades der Naturwissenschaften

Prevention and prediction of production instability of CHO-K1 cell lines by the examination of epigenetic mechanisms

Benjamin Moritz

München

Juli 2015

Eingereicht am 14.07.2015

Mündliche Prüfung am 27.10.2015

1. Gutachter: Prof. Dr. Peter Becker
2. Gutachter: Prof. Dr. Dirk Eick
3. Gutachter: Prof. Dr. Heinrich Leonhardt
4. Gutachter: Prof. Dr. Michael Boshart
5. Gutachter: Prof. Dr. Angelika Böttger
6. Gutachter: Prof. Dr. Wolfgang Enard

Table of Contents

Summary.....	8
Zusammenfassung.....	10
1 Introduction	12
1.1 Chromatin.....	12
1.1.1 Organization of chromatin.....	12
1.1.2 Euchromatin versus heterochromatin.....	13
1.1.3 The dynamic of chromatin structure – epigenetic marks.....	14
1.1.4 Inheritance of epigenetic marks	15
1.2 Histone variants.....	15
1.3 Histone modifications.....	16
1.3.1 Histone methylation	17
1.3.1.1 Location in context with function of specific histone lysine methylations.....	17
1.3.1.2 Establishing of histone lysine methylation and demethylation.....	18
1.3.1.3 Mode of action.....	19
1.3.2 Histone acetylation	19
1.3.2.1 Location in context with function of specific histone lysine acetylations	19
1.3.2.2 Establishment of histone lysine acetylation and deacetylation	20
1.3.2.3 Mode of action.....	21
1.4 DNA methylation	22
1.4.1 Location in context with function of DNA methylations	22
1.4.2 Establishing of DNA methylation and demethylation.....	23
1.4.3 Mode of action.....	25
1.5 Repressive chromatin.....	25
1.5.1 Constitutive heterochromatin	25
1.5.2 Position-effect variegation.....	26
1.5.3 Silencing through polycomb complexes	27
1.5.4 Heritable epigenetic silencing; positive feedback loops between different self-reinforcing repressive principles.....	28
1.5.5 X chromosome inactivation in mammals.....	29
1.5.6 Imprinting in mammals.....	30
1.5.7 Silencing of ribosomal genes in mammals.....	31
1.5.8 Silencing of transposons and retrotransposons	32

1.5.9	Silencing of transgenes in culture cells.....	33
1.6	Cell line development of therapeutic protein producers.....	34
1.6.1	CHO cell line development.....	34
1.6.2	Loss of productivity in CHO cell lines.....	35
1.6.3	Silencing of hCMV-MIE promoter.....	35
1.7	Aims.....	36
2	Materials and methods.....	38
2.1	Material sources.....	38
2.1.1	Laboratory chemicals and biochemicals.....	38
2.1.2	Enzymes.....	39
2.1.3	Antibodies.....	39
2.1.4	Organisms.....	39
2.1.5	Oligonucleotides and plasmids.....	40
2.1.6	Other materials and kits.....	41
2.1.7	Buffers and solutions.....	42
2.2	Methods.....	44
2.2.1	Methods for DNA preparation and analysis.....	44
2.2.1.1	Agarose gel electrophoresis.....	44
2.2.1.2	Gel extraction.....	44
2.2.1.3	DNA quantification.....	44
2.2.1.4	Transformation of competent bacteria.....	44
2.2.1.5	Plasmid preparation.....	45
2.2.1.6	DNA purification.....	45
2.2.1.7	Phenol-chloroform extraction.....	45
2.2.1.8	Transgene copy number determination.....	46
2.2.1.9	Cloning of reporter constructs.....	47
2.2.1.10	Site directed mutagenesis.....	48
2.2.1.11	Chromatin immunoprecipitation.....	50
2.2.1.12	Bisulfite conversion.....	52
2.2.1.13	Inverse PCR.....	54
2.2.1.14	Quantitative real-time polymerase chain reaction.....	56
2.2.1.15	Sequencing methods.....	59
2.2.1.16	Sequencing of amplicons after bisulfite conversion.....	59
2.2.1.17	Analyses of next generation sequencing data.....	60

2.2.2	Methods for protein analysis	62
2.2.2.1	Protein quantification	62
2.2.2.2	Protein fractionation in compartments	62
2.2.2.3	SDS polyacrylamide gel electrophoresis (SDS-PAGE).....	62
2.2.2.4	SDS-PAGE and western blot	63
2.2.3	General methods for working with CHO cell lines	64
2.2.3.1	Cryoconservation of CHO cell lines	64
2.2.3.2	Thawing of CHO cell lines.....	64
2.2.3.3	Cultivation of CHO cell lines	64
2.2.3.4	Transient and stable transfection	65
2.2.3.5	Cell count with Cellavista	65
2.2.3.6	Cell count with Cedex HiRes analyzer	65
2.2.3.7	CHO cell treatment with 5-aza-2deoxycytidine (DAC).....	66
2.2.3.8	Relative quantification of expression with SEAP-assay	66
2.2.3.9	Relative quantification of reporter gene expression with flow cytometry	68
2.2.3.10	Sample preparation for antibody analysis	70
2.2.3.11	Quantification of antibody production with HPLC.....	70
2.2.3.12	Quantification of antibody production with ELISA	70
2.2.3.13	Generation of recombinant CHO cell lines comprising hCMV-MIE promoter variants or the histone acetyltransferase MOF	71
2.2.3.14	Long-term cultivation and production of CHO cell lines comprising hCMV-MIE promoter variants.....	75
2.2.3.15	Treatment of recombinant CHO cell lines, used for identification of epigenetic marker	76
2.2.3.16	Generation of eGFP expressing CHO cell pools and preparation of CHO-K1 M cells for the identification of integration sites and chromatin states.....	78
3	Results	80
3.1	Molecular indicator of production instability.....	80
3.1.1	Cell lines and data of literature.....	80
3.1.2	Validation of ChIP antibodies for both projects with one sample each	83
3.1.3	Validation of reference regions for all cell lines within one project.....	84
3.1.4	Normalization of the different histone modifications at the hCMV-MIE promoter.....	86
3.1.5	Effect analysis of potential prediction markers	88
3.1.6	Exclusion of unstable producers by threshold setting for H3ac/H3 and meC-179.....	92

3.2	Alteration of epigenetic landscape of CHO-K1 cell lines to circumvent the silencing of recombinant genes.....	95
3.2.1	Coupled expression of target gene and the specific histone acetyltransferase MOF, proposing open chromatin formation.	95
3.2.2	Point mutation of CpGs within the hCMV-MIE promoter to improve production stability.....	100
3.2.2.1	Transient transfection: Investigation of promoter strength in CHO pools.....	101
3.2.2.2	Stable transfection: Investigation of stable eGFP expression in CHO pools.....	102
3.2.2.3	Stable transfection: Investigation of stable IgG-Cytokine fusion protein expressing monoclonal CHO cell lines.....	106
3.3	Epigenetic Landscape of CHO-K1-M cells and potential stable integration sites.....	109
3.3.1	Chromatin state of CHO-K1 wild type.....	109
3.3.2	Validation of integration sites.....	112
3.3.3	Combined analysis of chromatin state and integration sites	114
4	Discussion.....	119
4.1	Molecular Indicator of Production Instability	119
4.1.1	Limitation of prediction marker	120
4.1.2	Effect of copy number on stability at begin of cultivation phase	122
4.1.3	Effect of selection agent on stability.....	123
4.1.4	Perspectives	124
4.2	Alteration of epigenetic landscape of CHO-K1 cell lines to circumvent silencing of recombinant genes.....	125
4.2.1	Coupled expression of target gene and the histone acetyltransferase MOF to promote open chromatin formation.	125
4.2.2	Point mutation of CpGs within the hCMV-MIE promoter to improve production stability.....	127
4.2.2.1	Different effect strength of CpG point mutations between CHO pools and CHO cell lines.....	127
4.2.2.2	Copy number and CMV promoter methylation.....	129
4.2.2.3	General different effect strength of point mutations.....	130
4.2.2.4	Presumable influence of G-179 mutation in the cell line development after screening for high producers.	131
4.3	Epigenetic landscape of CHO-K1-M cells and potential stable integration sites	131
4.3.1	Location of integration sites of randomly transfected CHO-K1 pools	131

4.3.2	Discrimination of active and inactive integration sites.....	133
4.3.3	Perspective.....	134
5	Appendix	136
5.1	Supplement tables.....	136
5.2	Abbreviations.....	139
5.3	Table of figures.....	141
5.4	Ehrenwörtliche Versicherung	143
5.5	Erklärung.....	144
5.6	Curriculum vitae	145
5.7	Acknowledgements	147
6	Bibliography	149

Summary

The CHO-K1 cell line is the most common expression system for therapeutic proteins in the pharmaceutical industry. Due to the nature of economics, the cell lines and the vector design are subject to constant change to increase product quality and quantity. During the cultivation, the production cell lines are susceptible to decreasing productivity over time. Often the loss of production can be associated with a reduction of copy number and the silencing of transgenes. During cell line development, the most promising cell lines are cultivated in large batch culture. Consequently, the loss of a stable production cell line can be very cost-intensive. For this reason I developed different strategies to avoid a reduced productivity. Instability of production cell lines can be predicted by the degree of CpG methylation of the driving promoter. Considering that the DNA methylation is at the end of an epigenetic cascade and associated with the maintenance of the repressive state, I investigated the upstream signals of histone modifications with the assumption to obtain a higher predictive power of production instability. For this reason I performed a chromatin immunoprecipitation of the histone modifications H3K9me3 and H3K27me3 as repressive signals and H3ac as well as H3K4me3 as active marks. The accumulations of those signals were measured close to the hCMV-MIE at the beginning of the cultivation and were then compared with the loss of productivity over two month. I found that the degree of the H3 acetylation (H3ac) correlated best with the production stability. Furthermore I was able to identify an H3ac threshold to exclude most of the unstable producers.

In the second project I aimed to improve the vector design by considering epigenetic mechanisms. To this end I designed on the one hand a target-oriented histone acetyltransferase to enforce an open and active chromatin status at the transgene. On the other hand I point-mutated methylation-susceptible CpGs within the hCMV-MIE to impede the maintenance of inactive heterochromatin formation. Remarkably, the C to G mutation located 179 bp upstream of transcription start site resulted in very stable antibody producing cell lines. In addition, the examination of cell pools expressing eGFP showed that G-179 promoter variants were less prone to a general methylation and gene amplification, which illustrates the dominating effect in epigenetic mechanisms of one single CpG.

The last project was performed to localize stable integration sites within the CHO-K1 genome. In so doing I could show that the transfection leads predominantly to integration into inactive regions. Furthermore I identified promising integration sites with a high potential to induce stable expression. However, those results are preliminary and must be viewed with caution. Further examination needs to be done to confirm these results.

Considering the results of all three projects, I propose that the interplay of metabolic burden and selection pressure at an early time point of cultivation plays an important role in cell line development. Small alterations of selection pressure can lead to a decisive change of cell properties. Therefore, stable cells are less susceptible than weak producers. The increase of selection pressure leads to compensatory effect by gene amplification in the instable cell lines. The resulting adjustment of productivity masks the truly stable cells, which precludes the selection of the right cell lines. For this reason the selection pressure, the copy number as well as the growth rate should be considered to minimize repressive effects.

Zusammenfassung

Die CHO-K1 Zelllinie ist das am häufigsten verwendete Expressionssystem für therapeutische Proteine innerhalb der pharmazeutischen Industrie. Aus wirtschaftlichen Gründen wird die verwendete Zelllinie sowie die eingesetzten Vektoren ständig verbessert um die Produktqualität und -quantität zu erhöhen. Während der Kultivierungsphase neigen Produktionszelllinien dazu an Produktivität zu verlieren. Dabei wird der Produktivitätsverlust häufig mit einer Reduktion der Kopienzahl oder dem Silencing von Transgenen assoziiert. Während der Zelllinienentwicklung werden vielversprechende Zelllinien ausgewählt und im großen Ansatz kultiviert. Ein Produktivitätsverlust innerhalb solcher Zellen ist somit sehr kostenintensiv. Um diese Gefahr zu minimieren entwickelte ich unterschiedliche Strategien, welche darauf abzielen den Produktivitätsverlust zu vermeiden.

Produktionsinstabilität konnte von unserer Gruppe schon anhand des CpG Methylierungsgrades am CMV Promoter vorhergesagt werden. Die DNA Methylierung wird wahrscheinlich zur Aufrechterhaltung eines inaktiven Chromatinstatus benötigt und steht am Ende einer epigenetischen Kaskade. Im Gegensatz dazu erscheinen Histonmodifikationen früher in der Signalkaskade und könnten deswegen eine höhere Aussagekraft über die Stabilität haben. Aus diesem Grunde wurden von mir Histonmodifikationen am hCMV-MIE Promoter und *Enhancer* zu Beginn der Kultivierungsphase gemessen. H3K4me₃, H3ac sind Histonmodifikationen die mit Expression assoziiert werden wohingegen H3K27me₃ und H3K9me₃ grundsätzlich mit einem inaktiven Chromatinstatus in Verbindung gebracht werden. Der Grad der unterschiedlichen Modifikationen wurde mit dem über zwei Monate entstehenden Produktivitätsverlust verglichen. Dabei stellte sich heraus, dass der Grad der Histon H3 Acetylierung die höchste Korrelation mit der Stabilität aufwies. Des Weiteren konnte ich einen Grenzwert für die H3 Acetylierung definieren der einen Ausschluss der meisten instabilen Produktionszelllinien ermöglicht.

Im zweiten Projekt wurde das *Vector Design* unter epigenetischen Aspekten verändert. Ich erstellte eine zielgerichtete Histonacetyltransferase, um in dem Chromatinbereich des Transgenes einen offenen und aktiven Status zu induzieren. Desweiteren mutierte ich methylierungsanfällige CpGs des hCMV-MIE Promoters und Enhancers um eine Methylierung und daraus folgend einen inaktiven Chromatinstatus zu verhindern. Die C zu G Konversion an dem 179 Basenpaar oberhalb der Transkriptionsstartstelle führte zu einer bemerkenswert stabilen Antikörperexpression in klonalen Zelllinien. Desweiteren konnte ich bei gleicher Promotervariante in eGFP exprimierenden Zellpools eine geringere Methylierung und Genamplifikation feststellen. Somit konnte zum ersten Mal die Effektsensitivität eines einzelnen CpGs verdeutlicht werden.

Im letzten Projekt wurde die Expressionsstabilität abhängig von der Integrationsstelle des Transgenes untersucht. Dabei konnte ich zeigen, dass die standardmäßig durchgeführte zufällige Integration entweder bevorzugt in inaktiven Bereichen des Euchromatin stattfindet oder dass die Selektionsdruck induzierte Genamplifikation hauptsächlich im Heterochromatin stattfindet. Weiterhin vermute ich, dass beide Ereignisse hintereinander geschaltet sind, bei der die geringe Aktivität des Transgenes im inaktiven Euchromatin die Genamplifikation im Heterochromatin fördert. Bei der Untersuchung der Chromatinlandschaft und den enthaltenen Transgenen konnte ich vielversprechende aktive Regionen identifizieren, die wahrscheinlich die Stabilität der Expression fördern. Jedoch müssten diese Ergebnisse in weiteren Experimenten bestätigt werden.

Bei der Betrachtung der drei Projekte zeigt sich, dass das Wechselspiel zwischen der Belastung des Stoffwechsels der Zelle und dem Selektionsdruck in der frühen Kultivierungsphase ausschlaggebend ist für deren weitere Entwicklung. Dabei können kleine Veränderungen des Selektionsdruckes die Zellen maßgebend beeinflussen. Stabil exprimierende Zellen sind dabei weniger angreifbar als schwach exprimierende Zellen. Bei einer Erhöhung des Selektionsdruckes kompensieren die schlechteren Produktionszelllinien ihren Nachteil durch Genamplifikation. Die Anpassung der Produktivität überdeckt die stabilen Zellen welches die richtige Auswahl erschwert. Aus diesem Grunde sollte der Selektionsdruck, die Kopienzahl, sowie die Wachstumsrate in den Selektionskriterien mit einbezogen werden, um reprimierende Effekte zu minimieren.

1 Introduction

1.1 Chromatin

1.1.1 Organization of chromatin

Deoxyribonucleic acid (DNA) is a macromolecule that encodes the instructions of all known living organisms (Avery, Macleod & McCarty, 1944). In human cells DNA has a length of approximately two meters and is predominantly stored in the nucleus which has a diameter of 10 μm (Turner, 2002). To organize this amount of information the DNA needs to be highly compacted. A group of conserved, small basic proteins called histones in complex with DNA generate an ordered and compact structure termed `chromatin`. Histones were first described in 1911 by Albrecht Kossel. Those positively charged proteins interact with the negatively charged phosphodiester backbone of the DNA double helix (Alberts B, 2002; Meyers, 2012a; Olins, 2003). The four "core" histones H2A, H2B, H3 and H4 combine with DNA to form the basic repeating unit of chromatin, called the nucleosome (Thomas & Kornberg, 1975). The 225 kDA nucleosome core structure consists of approximately 147 bp of DNA wrapped around a histone octamer, comprising two H2A/H2B dimers and a H3/H4 tetramer in 1.67 left-handed superhelical turns (Figure 1) (Arents et al., 1991; Arents & Moudrianakis, 1993; Richmond, 1999). Linker DNA of 10 to 80 bp connect consecutive nucleosomes (Widom, 1992). A fifth histone called linker histone H1 can bind to the linker DNA and thereby induces further compaction by altering the entry/exit angle of the DNA (Allan et al., 1984). Nucleosomes with bound H1 are often referred to as chromatosomes comprising approximately 166 bp of two full turns of DNA (Allan et al., 1984; Wu, Bassett & Travers, 2007).

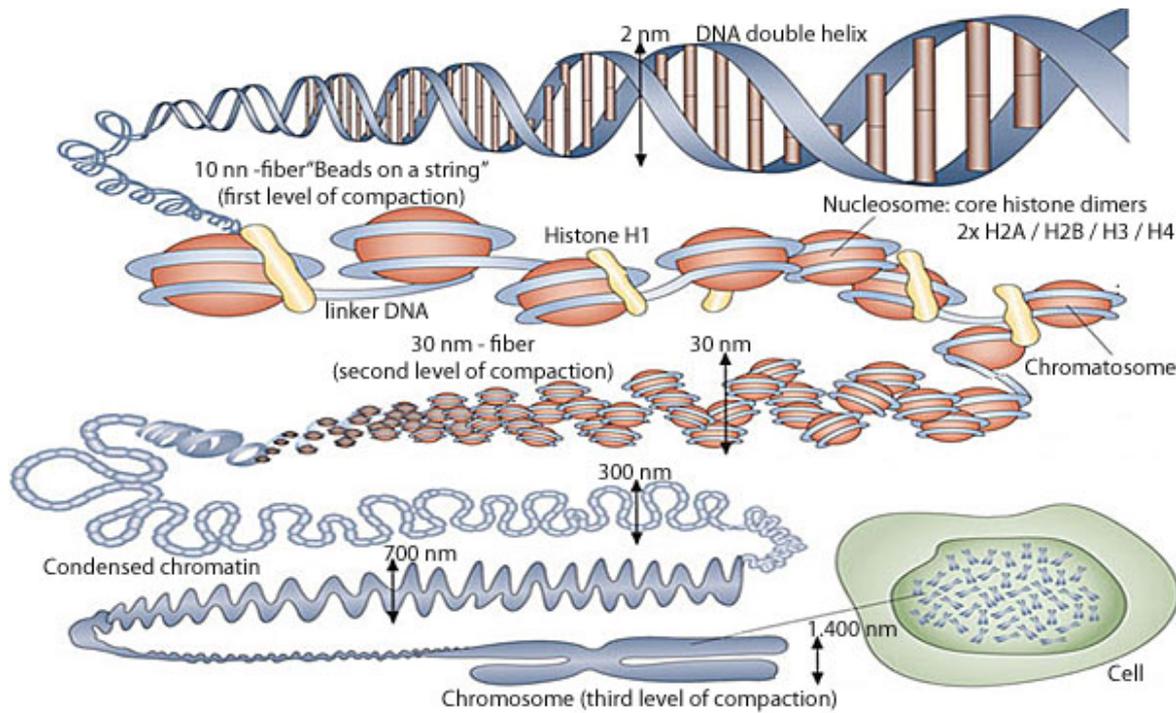


Figure 1: Organization of chromatin: Schematic overview of chromatin structure depicting major levels of histone mediated compaction. Modified from (Tonna et al., 2010).

Repeating nucleosomes build a long chain, which gives the appearance of beads on a string called 10 nm fiber (Olins & Olins, 1974). In addition the linker histone H1 induces further compaction into a secondary structure called 30 nm fiber (Thoma, Koller & Klug, 1979; Wu et al., 2007). The structure of the 30 nm fiber remains unresolved due to the limited visualization of the individual nucleosome within the fiber (Tremethick, 2007). Solenoid and zig-zag structures are reported as models for the 30 nm fiber, whereas the preferred architecture can be influenced by linker length between neighboring nucleosomes (Robinson & Rhodes, 2006; Routh, Sandin & Rhodes, 2008). Recent studies showed that 10 nm fibers were the only regular chromatin structures in vivo and that further compaction was rather achieved by densely packaged 10 nm fiber than by folding into regular structures like the solenoid or zig zag model (Chen & Dent, 2014). Folding at the tertiary level probably requires non-histone proteins known collectively as Chromatin Architectural Proteins (CAPs). The highest level of compaction is represented by mitotic chromosome with 1.4 μm fiber diameter (Woodcock & Ghosh, 2010).

1.1.2 Euchromatin versus heterochromatin

Depending on the accessibility of DNA, chromatin is distinguished into two types. Highly compacted heterochromatin, which is less accessible for transcription, and loosely-packed transcriptionally-

active euchromatin. Condensed chromatin is subdivided into constitutive and facultative (inducible) heterochromatin. Constitutive heterochromatin mainly comprises repetitive genetic elements and is thought to act as a chromosome stabilizer. Heterochromatin formation can spread into non-repetitive, gene-containing adjacent regions – this mechanism is known as position-effect-variegation. Facultative heterochromatin can form anywhere in the nucleus, often localized to promoters, and is established either in a developmentally regulated manner or in response to environmental triggers (Chen & Dent, 2014). Euchromatin, due to its open structure, is more accessible for RNA polymerases. Therefore, genes that are located in euchromatin are often actively transcribed (International Human Genome Sequencing, 2004).

Specific sequence elements known as insulator or boundary elements act either to block the effect of enhancers on promoters, or to prevent heterochromatin spreading into euchromatin areas by forming barriers. Thereby, respective genes are shielded from the influence of their position within a specific nuclear environment (Gaszner & Felsenfeld, 2006; Henikoff, 1990).

1.1.3 The dynamic of chromatin structure – epigenetic marks

During the last decades of research it has been demonstrated that chromatin is highly dynamic. Changes occur at all compaction levels from mitotic chromatin condensation to the individual nucleosome (Meyers, 2012a). Alterations of the chromatin structure can result from the incorporation of histone variants (Greaves et al., 2007; Hake & Allis, 2006; Santenard & Torres-Padilla, 2009), histone modifications (Chen & Dent, 2014), binding of chromatin architectural proteins (Luger & Hansen, 2005) or ATP-dependent remodeling factors (Clapier & Cairns, 2009), non-coding RNAs (Mercer & Mattick, 2013; Peschansky & Wahlestedt, 2014; Zhou, Hu & Lai, 2010) and DNA methylation (Smith & Meissner, 2013). The dynamics of chromatin interaction can be very different. While histone H1 association and dissociation can occur within seconds, histones H3 and H4 typically remain bound over several cell generations (Kimura & Cook, 2001; Kimura, Hieda & Cook, 2004; Lever et al., 2000; Steffen, Fonseca & Ringrose, 2012). Likewise, the post-translational modifications of histones are highly dynamic (Barth & Imhof, 2010). Cytosine methylation of DNA is in contrast considered to be a long-lasting modification (Otto & Walbot, 1990; Smith & Meissner, 2013). Considering the various possibilities of structure modulation at all stages of the cell cycle, it seems plausible why no consistent higher-order structures have been observed in vivo (Woodcock & Ghosh, 2010).

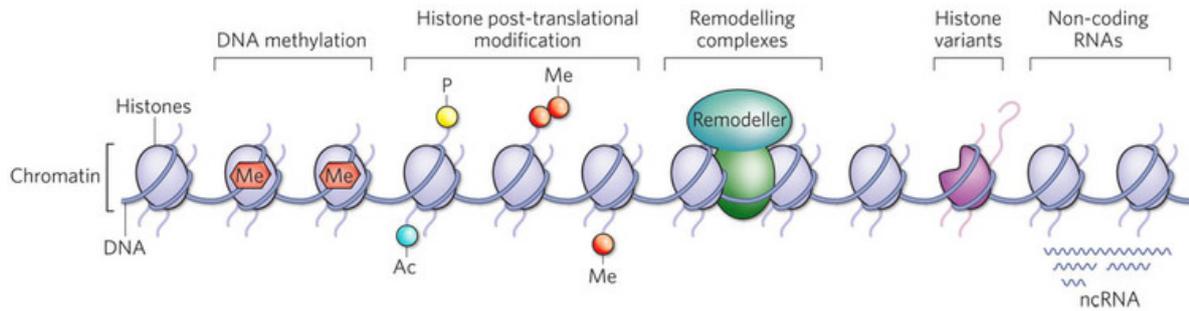


Figure 2: Epigenetic mechanisms involved in chromatin modifications. Five mechanisms for alteration of chromatin structure are known: Chromatin remodeling, insertion of histone variants, histone modification, effects of non-coding RNAs (ncRNAs) and DNA methylation. Ac, acetyl; Me, methyl; P, phosphate: adopted from (Dulac, 2010).

1.1.4 Inheritance of epigenetic marks

Epigenetic marks can be heritable and can be passed to daughter cells to preserve cellular identity (Chen & Dent, 2014). In particular, DNA methylation and H3K9 methylation are stable markers, which for instance can hinder the reprogramming of somatic cells into induced pluripotent stem cells (Papp & Plath, 2013). While it is widely accepted that symmetrical CpG methylation is passed on daughter cells involving semi-conservative segregation and template copying during DNA replication, inheritance mechanisms of histone modifications are still subject to investigation (Chen & Dent, 2014). In at least some cases proteins with specific recognition sites for post translational modifications rather than the histone marks persist through DNA replication at specific chromatin locations to establish histone modifications in daughter cells, which questioned the epigenetic inheritance of histone modifications (Francis et al., 2009; Petruk et al., 2012). However a recent study in fission yeast demonstrates true inheritance of H3K9 methylation independent of sequence specific recruitment (Ragunathan, Jih & Moazed, 2015).

1.2 Histone variants

Canonical histones H2A, H2B, H3 and H4 are conserved from yeast to mammals and are encoded by multiple gene copies, clustered in repeat arrays in higher eukaryotes. The transcription occurs mainly in S-phase during the cell cycle and is tightly coupled to DNA replication (Talbert & Henikoff, 2010). The canonical histones are incorporated in a replication coupled manner, whereas histone variants are expressed throughout the cell cycle and replace canonical histones when nucleosomes are evicted (Weber & Henikoff, 2014). Those histone variants differ in primary amino acid sequence from their canonical paralogues and are usually encoded by a single copy gene (Bernstein & Hake, 2006; Meyers, 2012a). In contrast to canonical histones, the non-canonical histones can be incorporated by replication-coupled and replication independent mechanisms. It has been observed

that the incorporated histone variant composition correlates with the functional status of the chromatin, regardless of the mechanism. Non-canonical histones have roles in DNA repair, meiotic recombination, chromosome segregation, transcription initiation and termination, sex chromosome condensation and sperm chromatin packaging (Talbert & Henikoff, 2010).

The H2A histone variants are well examined. Although the variants can be very different to the canonical H2A on the amino level, the structures are often quite similar. Within the core, the L1 loop as well as the docking domain of the C-terminus are often structurally different, which leads to variation of H3-H4 respectively H2A-H2B interaction (Weber & Henikoff, 2014). In addition internucleosomal and higher-order associations are influenced by alteration of the acidic patch of H2A. For instance H2A.Z and MacroH2A are associated with compaction, whereas H2A.Bbd is associated with decompaction of chromatin structure (Bonisch & Hake, 2012). Furthermore the combinations of non-canonical and canonical subtypes of histones like H2A.Z/H2A can be associated with specific biological processes, which increases complexity of signal pattern (Nekrasov et al., 2013).

1.3 Histone modifications

Histones can be modified in many different ways, mainly occurring along the residues of the N-terminal amino acid sequence which range from amino acids 13 to 40. The large number of modifications even increases due to the fact that some modifications, like lysine methylation, comprise up to three different states (Kouzarides, 2007). More than 100 histone modifications have been discovered and new modifications might be found within the next few years (Zentner & Henikoff, 2013). This complexity provides enormous potential for structural and functional responses. Acetylation and methylation of histone H3 and histone H4 tail residues are the best studied modifications. Overall, 14 distinctive types of modifications have been found (Dawson & Kouzarides, 2012) and are listed in Table 1.

Chromatin Modifications	Modified Residues	Chromatin Modifications	Modified Residues
Acetylation	K-ac	Proline isomerization	P-cis> P-trans
Methylation (lysines)	K-me1, K-me2, K-me3	Crotonylation	K-cr
Methylation (arginines)	R-me1, R-me2a, R-me2s	Propionylation	K-pr
Phosphorylation	S-ph, T-ph	Butyrylation	K-bu
Ubiquitylation	K-ub	Formylation	K-fo
Sumoylation	K-su	Hydroxylation	Y-oh
ADP ribosylation	E-ar	O-GlcNAcylation (serine and threonine)	S-GlcNAc; T-GlcNAc
Deimination	R>Cit		

Table 1: K, lysine; R, arginine; S, serine; T, threonine; E, glutamic acid; P, proline; Ac, acetylation; me, methylation; Cit, citrulline; P, proline; cr, crotonylation; pr, propionylation; bu, butyrylation; fo, formylation; oh, hydroxylation; a, asymmetric; s, symmetric; >, conversion; Cit, citrulline. Adopted from (Dawson & Kouzarides, 2012).

The combination of those modifications, termed as PTM motifs are rather associated with functions than the individual marks. Research into combination of modifications suggests more than 200 different PTM motifs (Feller et al., 2015). Modifications are reversible and controlled by proteins which are according to their function as “writers”, “erasers” and “readers”. Writers are enzymes which modify specific histone loci by adding a post-translational modification, such as histone acetyltransferases (HATs), whereas erasers are enzymes with the opposite function (Baker, Allis & Wang, 2008; Borrelli et al., 2008). Readers are proteins with specific domains which recognize post-translational modifications. For instance bromodomains recognize acetylation whereas chromodomain target the threefold methylation of H3K9 (Taverna et al., 2007). Thus, some histone acetyltransferases can act simultaneously as readers and writers.

1.3.1 Histone methylation

1.3.1.1 *Location in context with function of specific histone lysine methylations*

Methylations of histone amino acids are best-characterized for lysine (K) and arginine (R) residues. The prominent sites for lysine methylation are H3K4, H3K9, H3K27, H3K36, H3K79 and H4K20. The appearance and frequency of methylated residues are dependent on their location and function. H3K4, H3K36 and H3K79 methylated residues are typical for active chromatin, whereas H3K9, H3K27 and H4K20 methylation are often found in heterochromatin (Dawson & Kouzarides, 2012). The active mark H3K4me2/3 is especially enriched at the transcription start sites of active genes (Ruthenburg, Allis & Wysocka, 2007), the appropriate gene bodies are broadly enriched with

H3K36me3 (Zentner & Henikoff, 2013). Repressive marks have different spatial distributions: the H3K27me accumulates on the inactive X-chromosome, whereas H3K9me is enriched in the pericentromeric heterochromatin (Barski et al., 2007; Kouzarides, 2007). Furthermore, H3K9me3 as well as H3K27me3 can be found in the promoter region of silenced genes (Allan et al., 2012; Rougeulle, Navarro & Avner, 2003; Trojer & Reinberg, 2007).

1.3.1.2 Establishing of histone lysine methylation and demethylation

Methylation of histone residues is performed by the writers; lysine methyltransferases (KMTs) or protein arginine methyltransferases (PRMTs). The first lysine methyltransferase to be identified was SUV39H1 (Rea et al., 2000), followed by a varying number of characterized KMTs. The vast majority of KMTs contain a conserved SET domain catalyzing the transfer of a methyl group from S-adenosylmethionine (SAM) to lysine ϵ -amino groups (Meyers, 2012a). There are numerous histone lysine methyltransferases as well as demethylases which are highly specific enzymes that target specific lysines depending on their methylation status (Klose & Zhang, 2007; Xiao et al., 2003). In mammals at least eleven H3K4 methyltransferases have been described (Gu & Lee, 2013). The recruitment of these methyltransferases can be initiated by specific transcription factors and/or RNAs (Figure 3). Both are proposed as initiation mechanisms, followed by maintenance mechanisms (Ruthenburg et al., 2007).

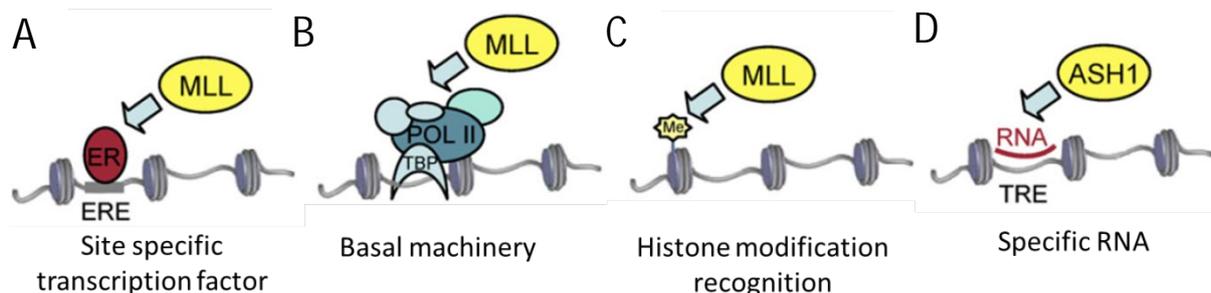


Figure 3 Mechanisms of H3K4 methyltransferase targeting specific genes exemplified by the MLL family. Recruitment is conducted by sequence specific DNA binding factors (A), by basal transcriptional machinery (B), by histone modification readers (C) or by RNA (D). Adopted from (Ruthenburg et al., 2007).

Lysine demethylases were first identified in 2004 (Shi et al., 2004). The lysine specific demethylase 1 (LSD1) is able to demethylate di- and mono-methyllysines via an amine oxide reaction (Klose & Zhang, 2007; Ng et al., 2009). Depending on the associated complex LSD1 can act either as repressor or activator by demethylating H3K4 or H3K9 (Klose & Zhang, 2007). Tri-methyllysines require a specific class of demethylases for degradation to occur. The first enzyme was discovered in 2006 and

termed JMJD2 (Whetstine et al., 2006). Degradation is carried out by an oxidative mechanism in which iron and α -ketoglutarate catalyze hydroxylation, forming a carbinolamine intermediate which readily breaks down into formaldehyde and a lysine lacking one methyl group (Mosammamparast & Shi, 2010; Whetstine et al., 2006).

1.3.1.3 Mode of action

Histone methylation neither influences the overall charge nor is it likely to influence the structure of the molecule. Considering this, methylation of residues indirectly affect the chromatin state by attributing binding sites for functional proteins instead of inducing a direct allosteric change (Meyers, 2012a). A number of protein domains were found which specifically binds to methylated lysines in histones. The Tudor, PWWP and MBT domains comprise a four β -stranded core region while the chromodomain comprises a three-stranded β -barrel with a C-terminal α -helix (Shimojo et al., 2008) as recognition site for methylated lysines within histones (Kim et al., 2006). PHD fingers can also act as recognition sites for methylated lysines, depending on protein context (Champagne & Kutateladze, 2009; Gozani et al., 2003). Those protein domains are part of downstream effector proteins, which are able to execute specific independent functions after binding histone modifications (Kim et al., 2006; Shi et al., 2006; Wysocka et al., 2006). The fact that a variety of proteins with different functions can bind the same histone modification (e.g. 5 of 6 MLL family proteins on H3K4me3), this led to the conclusion that there are primary recruiting events such as site-specific transcription or RNA, followed by the stabilizing effect of specific histone modification (Filippakopoulos et al., 2012; Ruthenburg et al., 2007).

1.3.2 Histone acetylation

1.3.2.1 Location in context with function of specific histone lysine acetylations

Acetylation was discovered in 1961 as the first histone modification (Phillips, 1963). Early studies associate actively transcribed genes with the hyperacetylation of histones, which indicates a function of acetylation in the transcription process (Allfrey, Faulkner & Mirsky, 1964). During S-phase a global increase of acetylation sites, such as H3K56ac, was observed followed by a decrease during G2-phase. This led to the conclusion that histone acetylation might facilitate the incorporation of newly synthesized histones (Miller, Maas & Toczyski, 2006). In addition to their global role during DNA replication, histone acetylation forms specific genomic patterns correlating with active transcription; where heterochromatin regions are hypoacetylated and transcriptionally active genes are highly acetylated (Kouzarides, 2007). Acetylation peaks are found at specific sites in the

promoter close to transcription start sites (TSS) and with tendencies to spread toward the gene body, depending on the acetylation mark and the gene (Wang et al., 2008). The lysine residues of the N-terminal tail of H3 at position 4, 9, 14, 18 and of H4 at lysines 5, 8, 12 and 16 are prone to acetylation (Zentner & Henikoff, 2013). Taken together, regulation of gene expression, DNA replication, repair and recombination are influenced by different acetylation states of histones (Dawson & Kouzarides, 2012).

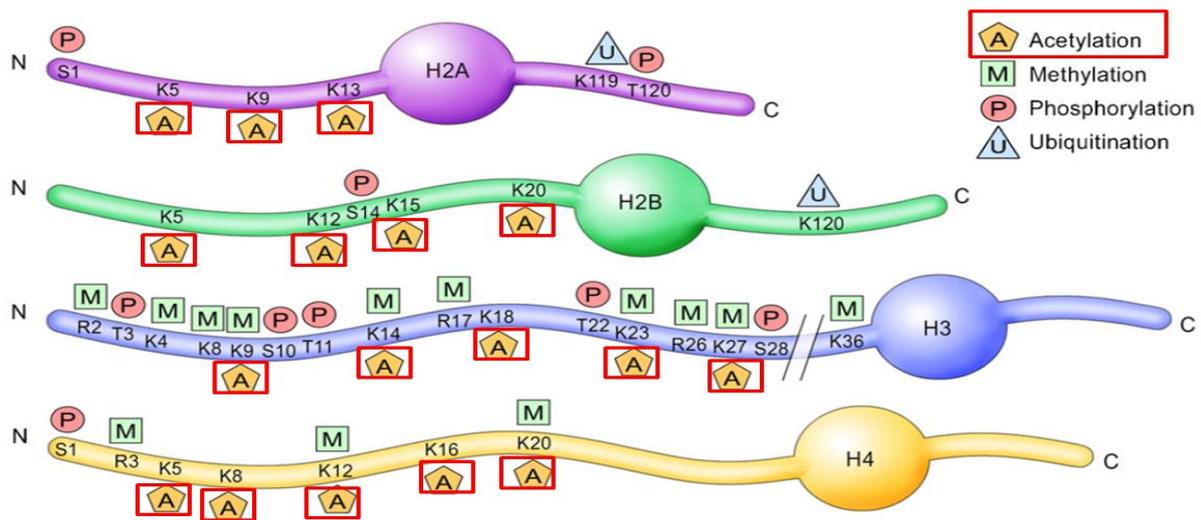


Figure 4: Exemplified modification of histone N-tails. The residues are numbered starting from the N-termini and proceeds towards the core histone body. H3K4 represents the 4th amino acid. Adopted from (Graff & Mansuy, 2008).

1.3.2.2 Establishment of histone lysine acetylation and deacetylation

Histone acetyltransferases (HATs) generally mediate gene expression and transcriptional activation (Cheung, Allis & Sassone-Corsi, 2000). Functional HATs which were examined *in vivo*, can be grouped into three families, based on their structural homology:

- (I) p300-CBP (CREB Binding Protein) coactivator family
- (II) GCN5-related N-acetyltransferases (GNATs); and
- (III) MYSTs (named after its four founding members MOZ, Ybf2, Sas2, and Tip60) (Meyers, 2013; Sun et al., 2012)

The CBP/P300 group only exists in metazoans and regulates a large number of physiological functions (Kalkhoven, 2004; Sun et al., 2012). *In vitro* studies suggested non-specific acetylation of multiple lysines of all core histones (Kouzarides, 2007). Although some *in vivo* studies demonstrate some specificity of CBP (Jin et al., 2011; McManus & Hendzel, 2003; Valor et al., 2011), a recent

study employing mass spectrometry to identify alteration of acetylation code suggests that CBP influences many acetylation sites (Feller et al., 2015).

The GNAT group is a large family of N-acetyltransferases: their conserved structure is a set of three-dimensional (3-D) structural motifs, comprising several α -helices and β -strands (Vetting et al., 2005). PCAF is a well-studied member of the GNAT family which can induce growth arrest by interaction with the cell cycle regulator p53 or to cellular proliferation by interacting with E2F (Timmermann et al., 2001). Furthermore it has been reported that PCAF acetylates specific lysines on H3 (Meyers, 2012b).

The MYST HATs are a closely related family of five human genes termed according to their yeast homologs (MOZ, YBF2 (SAS3), SAS2, and TIP60) (Pillus, 2008). The acetylation of specific histone tail lysines is influenced by MYSTs. For example H4K16 acetylation in MYST1 knockout mice is greatly reduced in early embryo blastocysts, while the acetylation of the remaining H4 and H3 residues remains unaffected, leading to peri-implantation lethality in mouse embryos (Thomas et al., 2008). Furthermore, MOF (the drosophila MYST1 paralog) overexpression suggests that H4K16ac counteracts chromatin compaction mediated by the ISWI ATPase (Corona et al., 2002). MYST3 also acetylates H4K16, whereas MYST4 might only affect H3 acetylation (Kraft et al., 2011; Paggetti et al., 2010).

Three potential mechanisms for HAT enzymatic activity have been proposed: creation of a ternary complex between acetyl-CoA and substrate lysine or generation of an intermediate with the acetyl-CoA through a cysteine followed by attacking the target lysine (Hodawadekar & Marmorstein, 2007) and a more complex "hit and run" mechanism (Liu et al., 2008). HATs are usually found in large macromolecular complexes comprising the additional components that confer specificity and genomic localization; their function have proven very difficult to dissect (Meyers, 2012a).

In contrast to this, the erasers like histone deacetylases (HDACs) remove acetyl groups, which strengthens the histone-DNA interactions and modify the nucleosome fiber to the closed and transcriptionally inactive heterochromatin conformation (Lee & Workman, 2007). The HDACs can be grouped into four classes dependent on the sequence homology. Class I, II and IV share a related catalytic mechanism requiring a zinc metal ion but no cofactors (de Ruijter et al., 2003). The class III enzyme employ a distinct catalytic mechanism that is NAD⁺ dependent (Saunders & Verdin, 2007).

1.3.2.3 Mode of action

The acetylation of histone residues has a direct effect which leads to the neutralization of the positive charge located at the ϵ amino group of lysines, which are located along the histone tails (Brownell & Allis, 1996). The interaction of negatively charged DNA backbone and the N-terminal

histone tails is influenced by the ionization state and acetylated chromatin tends to adopt a relaxed euchromatin structure. For instance, targeting H4K16ac to a promoter in vitro or in yeast can lead to a robust derepression of transcription (Akhtar & Becker, 2000). A recent study propose the redundancy of some acetyltransferases for the overall histone acetylation levels. This suggests a key role in genomic stabilization by maintenance of global charge neutralization (Feller et al., 2015). In addition the alteration of the ionization state, histone acetylation has been shown to exhibit at least two additional functions:

Acetylated lysines can be bound by bromodomains specifically, which enables protein recruitment like the chromatin remodeling complexes (Mujtaba, Zeng & Zhou, 2007). Hence the acetylation motif can influence the binding affinity of bromodomains (Filippakopoulos et al., 2012). Furthermore, the tandem plant homeodomain (PHD) fingers in the BAF chromatin remodeling complex member DPF3b are also able to bind acetyl-lysine motifs (Zeng et al., 2010).

In addition to recruitment of proteins, acetylation of histones can hinder binding of proteins to chromatin, in particular it has been shown that an acetylation of lysine 4 on H3 reduces the affinity of spChp1 to H3K9me (Schalch et al., 2009; Xhemalce & Kouzarides, 2010). Overexpression of the acetyltransferase MOF in *Drosophila melanogaster* leads to a loss of the ISWI chromatin remodeler, indicating that H4K16ac might interfere with ISWI remodeling activity (Corona et al., 2002; Zentner & Henikoff, 2013).

1.4 DNA methylation

1.4.1 Location in context with function of DNA methylations

DNA methylation is a repressive well-studied epigenetic modification and has a profound impact on gene expression, genome stability and development (Jaenisch & Bird, 2003; Smith & Meissner, 2013). Methylation patterns can be associated with the developmental stage and cell type-specific epigenetic memory (Bird, 2002). Methylation of cytosine at the fifth position of the pyrimidine ring is a conserved epigenetic mark from fungi, plants to animals (Feng, Jacobsen & Reik, 2010). In mammals 5-methylcytosine (5mC) is predominantly 5' located to guanine and is primarily restricted to palindromic cytidine-phosphate-guanosine (CpG) dinucleotides (Ramsahoye et al., 2000; Ziller et al., 2011). Due to the mutagenic properties of methylcytosine to deaminate into thymine (Coulondre et al., 1978), CpG dinucleotides are depleted in mammalian genomes (Deaton & Bird, 2011). The human genome has approximately 28 million CpGs, of these 60-80% are generally methylated in somatic cells (Smith & Meissner, 2013).

In pericentromeric repeats the CpG methylation maintains the centromere-proximal heterochromatin, which is essential for cell division. Transposable elements are constitutively hypermethylated to repress the activity of their strong promoters except in specific phases of embryogenesis; Also parent-specific imprints are propagated by CpG methylation of germline imprint control regions (ICRs) (Smith & Meissner, 2013).

Ten percent of the CpGs are accumulated in CpG-dense regions, termed CpG islands. Those islands are frequently located at transcription start sites of housekeeping and developmental regulator genes (Deaton & Bird, 2011), in a predominately nonmethylated state (Figure 5) (Illingworth et al., 2010). Orphan CpG islands are located in intergenic and intragenic regions. They tend to be more frequently methylated, compared to CpG islands at common transcription start sites.

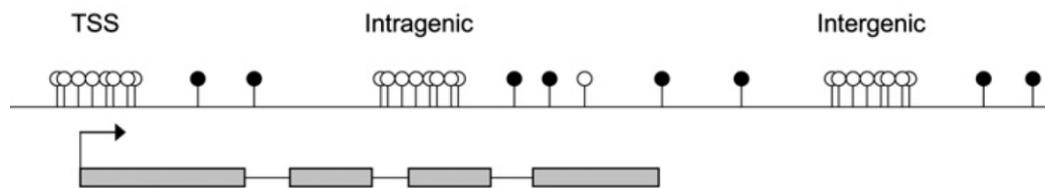


Figure 5: Distribution of CpGs. Empty circles represent unmethylated CpGs whereas filled circles display methylated CpGs. Accumulation of CpGs is termed CpG island which occurs predominately at transcription start sites. Orphan CpG islands which are located at intra and intergenic regions, represent the half of all CpG islands (Illingworth et al., 2010). Adopted from (Deaton & Bird, 2011).

1.4.2 Establishing of DNA methylation and demethylation

DNA methylation is performed by DNA methyltransferases which catalyze the transfer of a methyl group (CH_3) from S-adenosylmethionine (SAM) to the C5 position of cytosine (Wu & Zhang, 2014). Three conserved enzymes, DNA methyltransferase 1 (DNMT1), DNMT3A and DNMT3B are responsible for the establishment and maintenance of methylation and are essential for normal development (Li, Bestor & Jaenisch, 1992; Okano et al., 1999).

De novo DNA methylation patterns are established by the DNA methyltransferases DNMT3A and DNMT3B, which are in turn supported by the DNMT3L (Okano et al., 1999). The maintenance methyltransferase DNMT1 with its obligate partner UHRF1, is responsible for passing the global CpG methylation pattern on to the daughter cells by transferring a methyl group to the hemi-methylated newly replicated DNA (Bostick et al., 2007; Hermann, Goyal & Jeltsch, 2004; Sharif et al., 2007). DNMT1 is the key player of methylation maintenance, but DNMT3A and DNMT3B also contribute (Jones, 2012).

DNMT1 has a low affinity and catalytic activity for nonmethylated DNA. Therefore the *de novo* methylation by DNMT1 is limited (Jeltsch, 2006; Song et al., 2011). In embryonic stem cells (ESCs)

DNMT1 can be supported by DNMT3 to maintain DNA methylation at LINE (long interspersed elements) and LTR (long terminal repeat) element promoter sites (Liang et al., 2002). *De novo* methylation strongly correlates with the activity of close-by transcription factors. Depletion or truncation of the binding site for the common transcription factor SP1 within CpG islands leads to heritable accumulation of CpG methylation (Brandeis et al., 1994; Jones, 2012; Macleod et al., 1994). Furthermore, CpG methylation can influence the binding affinity of appropriate binding factor, as it has been shown for MYC (Jones, 2012). Examination of global transcription factor binding in context of DNA methylation supports the dominant role of transcription factors (Stadler et al., 2011). After transcription factor binding, an order of recruitment is frequently proposed, where chromatin remodeler LSH is followed by linker histone H1, heterochromatin protein HP1, H3K9 methyltransferases and *de novo* DNA methyltransferases (Smith & Meissner, 2013).

Though the mechanism of DNA methylation is well understood, the demethylation mechanism in mammals remains ambiguous. DNA demethylation occurs either passively or actively (Wu & Zhang, 2014). Passive demethylation occurs when maintenance DNA methyltransferase DNMT1 is not able to fulfill its function. 5mC loss is thus achieved by successive cycles of DNA replication (Bostick et al., 2007; Sharif et al., 2007). A main mechanism for active demethylation involves oxidation of 5mC (Wu & Zhang, 2014).

The oxidation of 5mC provides a biochemical mechanism of DNA demethylation (Wu & Zhang, 2014). Here 5mC is oxidized by enzymes of the TET family to 5-hydroxymethylcytosine (5hmC) (Kriaucionis & Heintz, 2009; Tahiliani et al., 2009), which gets successively oxidized to 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) (He et al., 2011; Ito et al., 2011; Pfaffeneder et al., 2011). TET proteins use a base-flipping mechanism to flip the target base out of the duplex DNA into the catalytic site to oxidate the 5-methylgroup in an α -ketoglutarate (α -KG) dependent way (Hashimoto et al., 2014; Hu et al., 2013). From this the CpGs can be demethylated by passive dilution or by active excision of modified 5mC (Wu & Zhang, 2014).

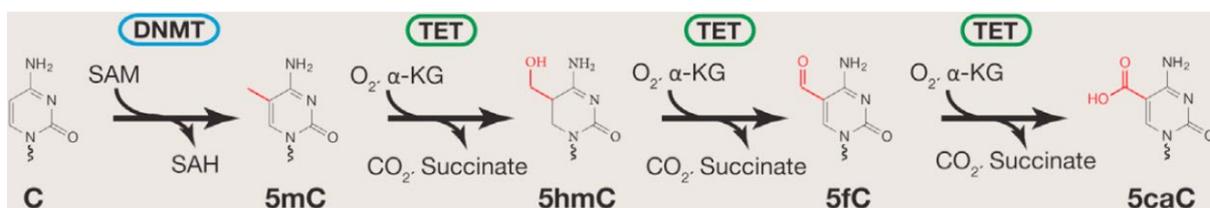


Figure 6: Methylation and demethylation of cytosine. 5-cytosine methylation is catalyzed by DNMTs which uses S-adenosylmethionine (SAM) as methyl donor, resulting in 5mC and S-adenosylhomocysteine (SAH). Methylated C5 can be oxidized by TET enzymes to 5-hydroxymethyl-, 5-formyl- and 5-carboxy-cytosine. Adopted from (Wu & Zhang, 2014).

1.4.3 Mode of action

CpG islands can be methylated during development, leading to a stable silencing of the associated promoter (Jones, 2012; Mohn et al., 2008; Payer & Lee, 2008; Stein, Razin & Cedar, 1982). It has been proposed that the silencing effect is promoted by direct inhibition of transcription factor binding to methylated DNA or by methyl-binding domain (MBD) proteins that recruit chromatin modifying activities (Cedar & Bergman, 2009; De Carvalho, You & Jones, 2010; Klose & Bird, 2006). Recent studies suggest that DNA methylation is not an initial step towards silencing, but rather acts to lock the silent state (Lee & Bartolomei, 2013; Okamoto & Heard, 2009; Payer & Lee, 2008). In mammals it has been proposed that stable promoter silencing requires H3K9 methylation as well as DNA methylation, in which H3K9me might initiate heterochromatin formation and DNA methylation ensures long term stability (Epsztejn-Litman et al., 2008; Myant & Stancheva, 2008; Myant et al., 2011). Also in female X chromosome inactivation the DNA methylation is downstream of a repressive histone modification (H3K27me3) which indicates that DNA methylation predominantly occurs at the end of a silencing cascade as a stabilizing signal (Smith & Meissner, 2013).

1.5 Repressive chromatin

Repressive chromatin can be divided in constitutive and facultative heterochromatin. Constitutive heterochromatin is characterized by consistently inactive regions across all cell types of a species, which indicates a role in the maintenance of genome stability. Sometimes, repressive epigenetic marks like H3K9me3 may spread into adjacent euchromatin regions. This is known as Position Effect Variegation (PEV). The facultative heterochromatin differs between cell types of a species and is therefore associated with differentiation and morphogenesis. Famous key players are the polycomb proteins (Almouzni & Probst, 2011; Oberdoerffer & Sinclair, 2007).

1.5.1 Constitutive heterochromatin

A sequential pathway is required to establish constitutive heterochromatin. Although the initial events of constitutive heterochromatin formation still need to be clarified, studies in mouse propose that transcription factors bind major satellite repeats in the pericentric region and attract the histone methyltransferase Suv39h (Dambacher, Hahn & Schotta, 2013; Saksouk N, 2015). This leads to an establishment of H3K9me3 at pericentric heterochromatin. The effect can be strengthened by pre-methylated H3K9me1 transferred from a cytoplasmic pool and generated by histone methyltransferases Prdm3 and Prdm16. The established H3K9me3 marks results in the attraction of heterochromatin protein HP1. HP1 is followed by a Suv4-20h2 accumulation. The histone

methyltransferase Suv4-20h2 recruits cohesin to the pericentric heterochromatin and catalyzes H4K20 trimethylation, which finally results in a highly condensed form of chromatin (Dambacher et al., 2013). In addition to histone methylation, DNA methylation also strongly correlates with the constitutive heterochromatin and is responsible for maintaining genomic integrity (Smith & Meissner, 2013).

1.5.2 Position-effect variegation

A phenomenon of gene silencing adjacent to constitutive heterochromatin was originally observed in *Drosophila* mutants, bearing a radiation-induced chromosome rearrangement that places the white gene close to centromere heterochromatin (Muller, 1930). The mosaic mutants had some white (mutant) and some red (normal) patches, which indicates a variegated gene silencing. Profound examinations showed that chromosomal rearrangements induced the silencing via translocation of white genes close to the pericentric heterochromatin. From this the heterochromatin marks H3K9me3 and H4K20me3 spread into neighboring regions and lead to transcription repression (Almouzni & Probst, 2011; Wang et al., 2014). Depending on the region, different proteins induce the downstream signals of H3K9me3 and H4K20me3 (Dambacher et al., 2013; Wang et al., 2014). Additionally it was observed that the counteracting activation marks H3ac and H4ac correlate negatively with distance of constitutive heterochromatin to transgene, which suggest heterochromatinization-dependent inhibition of histone acetylation (Yin et al., 2012).

Drosophila melanogaster

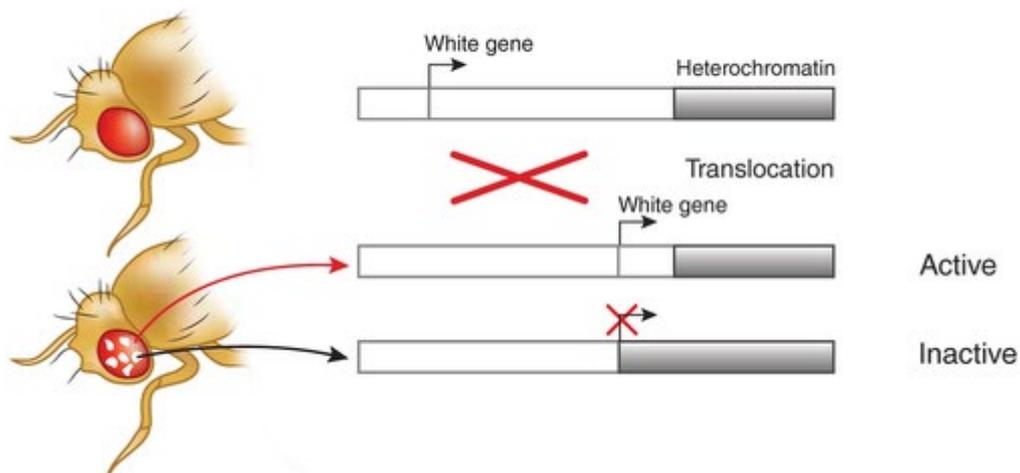


Figure 7: Position-effect variegation. Translocation of white genes close to heterochromatin results in silencing mediated by spreading of heterochromatin marks. Adopted from (Festenstein & Chan, 2012).

Although constitutive heterochromatin is consistent within a species, silenced genes within those regions can be activated rapidly (Bannister, Schneider & Kouzarides, 2002). This indicates the high dynamic nature of chromatin structure. Methylated H3K9 is a permanent and heritable mark which attracts HP1/Swi6 with higher affinity than the unmethylated H3K9. From this it can be determined that HP1/Swi6 is presented in steady state equilibrium of dynamic association and dissociation. Considering this, euchromatin with less methylated H3K9 has lower binding affinity than H3K9me3-rich heterochromatin, leading to increased HP1/Swi6 mediated cross-link of adjacent nucleosomes in heterochromatin. This process is favored by further compaction by enhanced HP1/Swi6 interaction. Therefore a stochastic model for short-term heterochromatin, also known as mass action phenomena is proposed in *Drosophila*, yeast and mammals (Cheutin et al., 2004; Dambacher et al., 2013; Henikoff, 1996). Observed binding of transcription factors in “closed” constitutive heterochromatin regions as well as the need for higher activation signals to derepress genes in constitutive heterochromatin compared to elsewhere repressed regions, support the theory of a stochastic model (Aparicio & Gottschling, 1994; Sekinger & Gross, 2001).

1.5.3 Silencing through polycomb complexes

The polycomb group (PcG) complexes are a family of transcriptional repressors subdivided in two forms of multiprotein complexes, the polycomb repressive complex 1 (PRC1) and 2 (PRC2) (Gil & O’Loghlen, 2014). The PRC2 comprises the core proteins Suz12, Eed, and the lysine methyltransferase Ezh2 or its homolog Ezh1. Both KMTs catalyze the H3K27 methylation, a common hallmark of facultative heterochromatin (Di Croce & Helin, 2013; Mozzetta et al., 2014). Time-dependent expression of differentiation genes is induced by alleviating PC-mediated repression to ensure correct differentiation of adult stem and progenitor cells (Margueron & Reinberg, 2011). During differentiation pluripotency genes are repressed by polycomb. In this way targeted chromatin fibres were folded into small discrete nuclear compartments, termed Pc nuclear foci (Cheutin & Cavalli, 2014). The PRC1 can bind with its subunit chromobox-domain (CBX) protein to H3K27me3. The bound PRC1 induces monoubiquitination of H2A on K119 by activity of E3 ligases RING1a and RING1b (Di Croce & Helin, 2013). From this a complex comprising of the protein RYBP and the catalytic subunit RING1B of PRC1, catalyzes the ubiquitination of H2A (Tavares et al., 2012). In both cases the histone H2A ubiquitination is proposed to repress transcription factor binding and inhibit the RNA polymerase (Zhou et al., 2008). Furthermore, H2A ubiquitination leads to a positive feedback loop by the promotion of H3K27 methylation (Kalb et al., 2014).

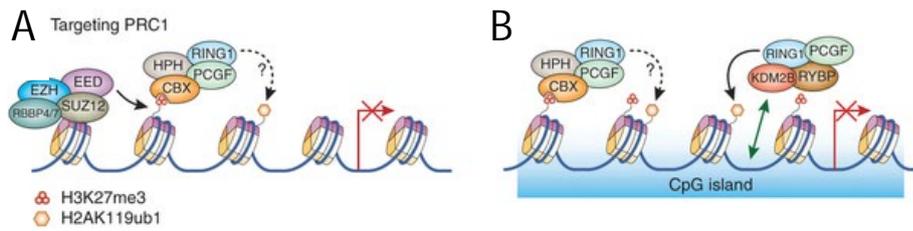


Figure 8: Silencing by polycomb complexes. The canonical pathway requires recognition of H3K27me3 (A) whereas in the noncanonical pathway the affinity of KDM2B for the CpG island enables the binding of the RYBP-PRC1 complex (B). The bound PRC1 induces monoubiquitination of H2A on K119 by activity of E3 ligases RING1a and RING1b. Adopted from (Di Croce & Helin, 2013).

The methylation of H3K9 can be indirectly induced by the interaction of PRC2 with G9a and GLP (H3K9 KMTs), which demonstrates a crosstalk between both epigenetic silencing pathways (Mozzetta et al., 2014).

1.5.4 Heritable epigenetic silencing; positive feedback loops between different self-reinforcing repressive principles

Two major epigenetic marks for the maintenance of the epigenetic status are H3K9 methylation and DNA methylation. Both marks are embedded in positive feedback loops, as exemplified in the fission yeast and Arabidopsis models.

Constitutive heterochromatin in fission yeast is marked with H3K9me3 and comprises various numbers of repeat units. Those repeats can be transcribed into double stranded (ds) RNA and further processed into small interfering (si) RNA by Dcr1 (Volpe et al., 2002). The RNA-dependent RNA polymerase complex (RdRC) produces dsRNA and siRNA from Argonaute (Ago) 1-targeted transcripts by interaction with Dcr1 and Ago1. Also Ago1 is a component of the RNA-induced transcriptional silencing complex (RITS), which guides to nascent pericentromeric non-coding RNA (ncRNA) transcripts if loaded with appropriate siRNA. Furthermore, RITS binds H3K9me3 with the chromodomain protein Chp1 (Verdel et al., 2004). This reinforces H3K9 methylation by recruitment of the cryptic loci regulator complex (CLRC), which contains the H3K9 KMT Clr4 (Zhang et al., 2008a). The target base pairing of siRNA has therefore an important role in the heterochromatinization and creates a positive feedback loop between siRNA generation, RITS localization and H3K9 methylation (Castel & Martienssen, 2013).

In Arabidopsis an analogical reinforcing positive feedback loop was discovered whereas 24 nt small interfering RNAs guide a DNA methyltransferase to a specific genomic loci. Here RNA-dependent RNA polymerase 2 (RDR2) associates with RNA Pol IV on repetitive heterochromatic loci to produce dsRNA (Law et al., 2011). Those dsRNA are cleaved by DICER-LIKE 3 (DCL3) into small interfering RNA (Kasschau et al., 2007). Subsequently, Ago4 is loaded with the siRNA in cytoplasm and transferred

into the nucleus (Wierzbicki, Haag & Pikaard, 2008; Wierzbicki et al., 2009; Ye et al., 2012). In the nucleus, either Ago4 can bind nascent RNA Pol V transcripts to form the RNA-directed DNA methylation complex, or associates with RDM1 to activate the de novo DNA methyltransferase DMR2 (Gao et al., 2010; Wierzbicki et al., 2008; Wierzbicki et al., 2009).

In addition to the self-reinforcing feedback loops, different models propose crosstalks between the epigenetic pathways. Two models are depicted below for the maintenance of genome integrity, mediated by repressive heritable marks.

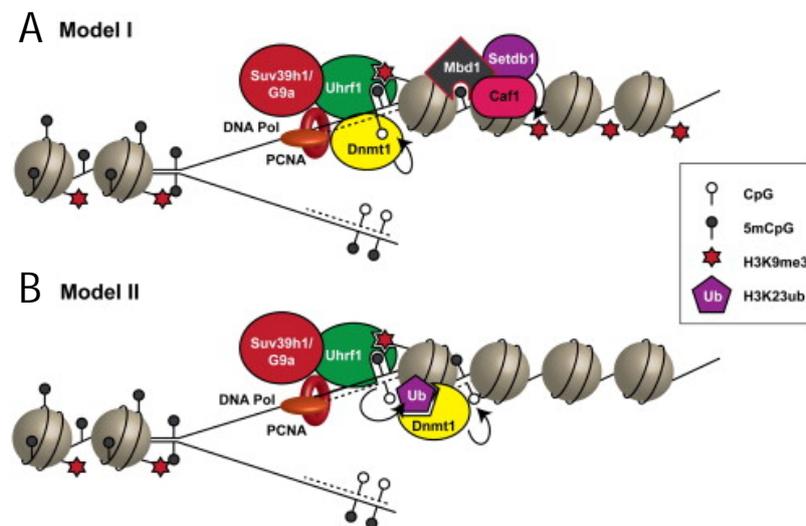


Figure 9: Crosstalk of epigenetic pathways. (A) Model I H3K9me3, hemimethylated DNA and PCNA are recognized by Uhrf1 which recruits Dnmt1 and H3K9 methyltransferase (G9a) directly. In addition Setdb1 is attracted to methylated DNA through interaction with Mbd1, which lead to maintenance of the H3K9 methylation. (B) In model II at first Uhrf1 ubiquitylates H3K23 after PCNA, H3K9me3 and DNA methylation recognition, than Dnmt1 gets recruited.

As described in section 1.4, DNA methylation is maintained through the interaction of Dnmt1 with Uhrf1. In model I PCNA, H3K9me3/me2 and hemimethylated DNA are recognized by Uhrf1 at the replication forks followed by the recruitment of Dnmt1 and H3K9 methyltransferase to copy the methylation pattern onto the newly synthesized daughter strand. In addition Setdb1 is attracted to methylated DNA, maintaining the H3K9 methylation. In model II Uhrf1 induces the ubiquitylation of H3K23, which in turn attracts Dnmt1 (Rose & Klose, 2014).

1.5.5 X chromosome inactivation in mammals

Gene dosage compensation relates to the fact that species with heteromorphic sex chromosomes frequently have different numbers of sex chromosomes. For example, female mammals females two X chromosomes, but males have only one. So the X chromosome inactivation (XCI) compensates for the dosage difference between both sexes. However, recent studies suggest that rather a partial

dosage compensation of dosage sensitive genes occurs in mammals than the observed global dosage compensation of *Drosophila* (Pessia, Engelstadter & Marais, 2014).

The X-inactivation center (Xic) controls XCI, whereby the XCI is randomly chosen in somatic cells of eutherian mammals resulting in mosaic females (Lee & Bartolomei, 2013). The Xic is associated with lncRNAs and the “X-interactive-specific transcript” (XIST/Xist) contains a 17-20 kb RNA, which covers the X-chromosome during XCI (Clemson et al., 1996). Hereby Xist RNA binds PRC2 and guides it to the Xi (Zhao et al., 2008) as long as the cis-expressed Xist antisense partner Tsix is down-regulated (Jeon & Lee, 2011). PRC2 is attracted by strong binding sites along the future inactive X. The resulting H3K27me3 gradually spreads from strong binding sites into neighboring clusters with more moderate binding affinity. Hierarchical spreading of XCI depends on Xist RNA and sites with a high frequency of polycomb proteins (Lee & Bartolomei, 2013). Xist is controlled by other lncRNAs. For instance Tsix acts as a repressor and coordinates X chromosome pairing (Bacher et al., 2006; Xu et al., 2007; Xu, Tsai & Lee, 2006), recruits DNA methyltransferase (Dnmt3a) towards Xist, which leads to DNA methylation (Sado, Hoki & Sasaki, 2005; Sun, Deaton & Lee, 2006; Zhao et al., 2008) and blocks PRC2 binding to Xist. The deletion of the trans-acting Jpx RNA or RepA RNA lead to a reduced Xist activation, which indicates both RNAs as potential activators of Xist (Lee & Bartolomei, 2013).

1.5.6 Imprinting in mammals

The coordinated regulation of a specific gene domain according to the parental origin is termed “genomic imprinting” and affects both male and female offspring. Consequently genomic imprinting is a result of parental inheritance and not of sex (Barlow & Bartolomei, 2014). Around five to ten percent of all genes in the mammalian genome are influenced by imprinting. Imprinted genes are commonly distributed in clusters of 20 to 3,700 kb range (Barlow, 2011). Besides protein coding genes the clusters contain noncoding RNAs (ncRNAs), which can be subdivided in micro-, sno- and lncRNAs oftentimes close to or within imprinting control region (ICR). The ICRs contain the parent-of-origin-specific epigenetic modifications which are responsible for the specified regulation. A well-known mechanism for imprinting includes the expression of lncRNAs in cis following similar schemes as has been observed for X inactivation, although the mechanism of cis-induced silencing by lncRNA remains unclear. A possible mechanism proposes that the lncRNA overlaps in a sense-antisense way with adjacent imprinted genes, which causes transcriptional interference in the overlapped region (Pauler, Barlow & Hudson, 2012). From this an accumulation of repressive chromatin follows and spreads throughout the cluster (Latos et al., 2012). In addition, many imprinted lncRNAs such as Gt2 and Nespas can attract polycomb proteins and act thereby in a manner reminiscent of Xist (Pandey et al., 2008; Zhao et al., 2010). Further experiments with lncRNAs like Airn and Kcnq1ot1 suggest

that besides the attraction of polycomb proteins, also an actively recruitment of repressive histone modifications takes place (Nagano et al., 2008; Pandey et al., 2008; Terranova et al., 2008). In conclusion, genomic imprinting can be a result of direct interaction of lncRNA within the controlled region or the induction of epigenetic signal cascades.

1.5.7 Silencing of ribosomal genes in mammals

Ribosomal gene transcripts are the RNA component of ribosomes and are therefore of pivotal importance in the cell metabolism. These genes are located at the nucleolar organizer regions (NORs) and clustered in tandem repeats on five acrocentric chromosomes (Grummt & Langst, 2013). The transcription of rRNA genes depends on the cell status and is in general active in 50% of all rRNA genes (Conconi et al., 1989; Lucchini & Sogo, 1995).

Silent ribosomal RNA genes are indicated by the typical heterochromatin marks like DNA hypermethylation, H4 hypoacetylation, H3K9me3, H4K20me3 and H3K27me3 (Earley et al., 2006; Lawrence et al., 2004; Santoro, Li & Grummt, 2002; Zentner et al., 2011; Zhou, Santoro & Grummt, 2002). Induced hypomethylation of rRNA genes results in genomic instability. Apparently many rRNA genes are deeply embedded into repressive chromatin structures, which are typically excluded from the cellular recombination machinery (Kobayashi, 2008; Peng & Karpen, 2007). Transcriptional silencing can be promoted by DNA methylation which reduces the protein-DNA binding affinity. For instance, a single methylation located 133 bp upstream of TSS within the upstream control element (UCE) of the rDNA promoter hinders the binding of the basal transcription factor UBF (Santoro & Grummt, 2001). In addition methylation might repress rDNA transcription through attraction of histone deacetylases, as mediated by methyl-CpG-binding proteins or MBD proteins (Brown & Szyf, 2007; Ghoshal et al., 2004; Klose & Bird, 2006). This suggests that only specific positions of DNA methylation have a direct effect on UBF.

In general epigenetic silencing mechanisms for rRNA genes are well known, but the targeting mechanism of specific copies is still unclear. NORs contain in a ratio of 2:1 evenly distributed canonical repeats and non-canonical rDNA repeats, which are indicated by their palindromic structures (Caburet et al., 2005). In addition rRNA genes can differ in number and location of single nucleotide polymorphisms (SNPs) and length of enhancer elements (Guetg et al., 2010; Qu, Nicoloso & Bachellerie, 1991; Santoro et al., 2010; Shiao et al., 2011; Tseng et al., 2008). From this chromatin structure- and transcription- variation might be explained. Once the epigenetic states are accomplished the status is inherited through cell division and stably propagated towards daughter cells (Schlesinger et al., 2009).

A key regulator of rRNA gene transcription is the multifunctional protein TTF-I. TTF-I can bind to specific terminator elements, for instance 170 bp upstream of the transcription start site (To). From this TTF-I can either promote transcription by interaction with activating chromatin modifiers like CSB/G9a (Langst, Becker & Grummt, 1998; Langst et al., 1997) or mediate silencing by attraction of repressive chromatin remodeling complex NoRC. (Santoro et al., 2002; Strohner et al., 2001; Zhou et al., 2002). In addition TTF-I can support DNA loop formation by self-assembly of proteins linked to distant terminator elements. The loop facilitates activation by concentrating relevant transcription factors (Denissov et al., 2011; Nemeth et al., 2008; Nemeth & Langst, 2011; Sander & Grummt, 1997; Shiue, Arabi & Wright, 2010; Shiue, Berkson & Wright, 2009). The repressive epigenetic state is promoted by the ATP-dependent chromatin remodeling machines NoRC, which comprises the DNA-dependent ATPase SNF2h and the TTF-I interacting protein 5 (TIP5) (Strohner et al., 2001). After TTF-I-mediated NoRC-rDNA binding, NoRC induces nucleosome remodeling, histone modification and DNA methylation (Grummt & Langst, 2013; Li, Langst & Grummt, 2006). Nucleosomes are then relocated predominantly to a silent position that does not allow the formation of the transcription complex comprising UBF and TIF-IB/SL1 (Li et al., 2006). In particular, NoRC recruits histone deacetylases (HDAC), H3K9 histone methyltransferases (HMT) SET-DB1 and DNA methyltransferases (DNMT) to establish the heterochromatin state, which is propagated through cell cycle progression (Grummt & Langst, 2013).

1.5.8 Silencing of transposons and retrotransposons

46% of the human and 37.5% of the mouse genome consists of repetitive sequences, located within and between genes (Lander et al., 2001; Mouse Genome Sequencing et al., 2002). These repeat elements can be distinguished in four groups: DNA transposons, long interspersed elements (LINEs), short interspersed elements (SINEs) and long terminal repeats (LTRs) (Ohms & Rangasamy, 2014). Of these the LINE-1 elements are the most active retrotransposons (Brouha et al., 2003). Retrotransposons, as well as DNA transposons, are DNA sequences that can relocate within the genome, mutate and duplicate themselves. Retrotransposons are first transcribed into RNA followed by a reverse transcription into DNA, whereas the underrepresented DNA transposons do not require an RNA intermediate to jump within the genome (Lander et al., 2001; Mouse Genome Sequencing et al., 2002). Retrotransposition affects genomic integrity by generating dsDNA breaks and insertion of pseudo-genes or retro-genes (Esnault, Maestre & Heidmann, 2000; Gasior et al., 2006; Gualtieri et al., 2013). Thus the fitness of the cell depends on the silencing of retrotransposons. Especially the repetitive sequences are mostly silenced in differentiated cells whereas embryonic stem cells harbor non-repressed retrotransposons, which may be explained by an overall open chromatin formation

(Efroni et al., 2008). Silencing of retrotransposons is mediated by DNA methylation, histone modifications and small RNAs (Ohms & Rangasamy, 2014). Small RNAs are integral to various silencing mechanisms and can be grouped into microRNAs (miRNAs), endogenous small interfering RNAs (endo-siRNAs), small nucleolar RNAs (snoRNAs) and Piwi-interacting RNAs (piRNAs) (Ambros & Chen, 2007; Ghildiyal & Zamore, 2009; Peng & Lin, 2013). siRNAs suppresses retrotransposition by binding endogenous mRNAs in a sense-antisense manner (Faulkner et al., 2009; Ghildiyal et al., 2008; Watanabe et al., 2006; Yang & Kazazian, 2006). From this complementary siRNA is generated by folding back of inverted repeats, bidirectional transcription sites and expressed antisense transcripts of retrotransposons and pseudogenes or by multiple Dicer cleavages of long precursor dsRNAs (Okamura et al., 2008; Tam et al., 2008). piRNAs induces retrotransposon silencing by a Piwi-interacting RNA-mediated mechanism and are Dicer independently produced from retrotransposon derived dsRNA (Aravin et al., 2007; Carmell et al., 2007; Girard et al., 2006; Ross, Weiner & Lin, 2014).

After the initial regulation of repetitive elements by RNA in mouse embryos, the loss of activation marks occurs prior to the accumulation of repressive histone modifications (Fadloun et al., 2013).

1.5.9 Silencing of transgenes in culture cells

Integrated recombinant genes tend to get lost by chromosomal disruption/ rearrangement mechanisms (Kim et al., 2011) and are prone to silencing. They are often inserted as multiple copies, which resembles retrotransposon arrangements. Observed epigenetic repressions include promoter methylation (Yang et al., 2010a), histone methylation and histone deacetylation (Mutskov & Felsenfeld, 2004). Especially in CHO cell lines, a simultaneous decrease of H3 acetylation and transgene expression was observed in long term culture (Paredes et al., 2013). The initiation of transgene silencing is proposed to be associated with the transgene sequence, their susceptibility to form tandem repeats, and the integration site into the host genome (Kaufman et al., 2008). Integration site-dependent silencing is mediated by the position effect, which influences the stability of recombinant gene expression (Lattenmayer et al., 2006; Yin et al., 2012). For this reason potentially stable integration sites are extensively studied (Akhtar et al., 2013).

1.6 Cell line development of therapeutic protein producers

1.6.1 CHO cell line development

The mammalian cell culture rose to be the main production system in the pharmaceutical industry, since the first therapeutic protein was expressed from recombinant mammalian cells in 1986 (Kim, Kim & Lee, 2012). Alternative expression systems such as microbial (e.g. E.coli), insect, cells, plant cells or transgenic animals are also in use, but for the expression of therapeutic proteins such as immunoglobulins and other multimeric or glycosylated proteins, mammalian cells are the preferred producers (Osterlehner, Simmeth & Gopfert, 2011; Walsh, 2010).

Despite the variety of available mammalian cell lines, the Chinese hamster ovary (CHO) cells produce nearly 70% of all recombinant therapeutic proteins (Jayapal & Wlaschin, 2007). In regard to the specific demands on therapeutic proteins, the CHO cells have some attributes which make them suitable for recombinant protein expression. CHO cells are a well-established system, supplying therapeutic proteins over the last 20 years which might facilitate the approval by regulatory agencies. Furthermore the usual low specific productivity (q) of protein production in mammalian cells can be overcome by gene amplification in CHO cells. Dihydrofolate reductase (DHFR)-mediated or glutamine synthetase (GS)-mediated (e.g. CHO-K1 SV) selection systems are available in CHO cells to compensate for low q by selective exclusion of low copy number cell lines or to maintain a specific production level, which is dependent on selection agent concentration. Also CHO cells can efficiently conduct post-translational modifications to produce, for instance, glycosylated forms of recombinant proteins, which are essential for biologic functionality and compatibility in human (Wurm, 2004). At last, CHO cells proliferate well in serum-free (SF) suspension, which is preferred from the regulatory authority and therefore used in large scale cultures (Kim et al., 2012; Landauer, 2014).

The demand for biopharmaceuticals is increasing constantly, resulting in the current market with an average yearly growth of 35% since 2001 (Altamirano et al., 2013). Although the development of CHO cell culture over the past two decades has resulted in a 100-fold yield improvement of titer, the highly competitive market of biopharmaceutical drugs requires a constant development of producer cell lines (Kim et al., 2012). Besides the development of serum-free medium and optimized feeding strategies, achievements in cell- and vector-engineering make CHO cell lines highly productive (Kim et al., 2012).

1.6.2 Loss of productivity in CHO cell lines

Economic cell lines are required to provide high productivity and stable production levels during propagation from small to large scale. Decrease of productivity during scale-up constitutes a serious risk during cell line development (Barnes, Bentley & Dickson, 2003). One main reason for production instability is a reduction of copy numbers over time, which might be attributed to chromosomal disruption/rearrangement as an inherent characteristic of CHO cells (Kim et al., 2011) and/or an induction by gene amplification (Kaufman, Sharp & Latt, 1983). The decrease of mRNA at a constant number of copies of recombinant genes is another major cause of productivity drop (Barnes, Bentley & Dickson, 2004; Chusainow et al., 2009; Strutzenberger et al., 1999). A reasonable explanation for this occurrence is epigenetic silencing by promoter methylation (Osterlehner et al., 2011; Yang et al., 2010b) and histone modifications as typified by deacetylation and specific methylation (Mutskov & Felsenfeld, 2004; Paredes et al., 2013). Also, the recombinant sequence itself, the formation of tandem repeats of sequence and the genomic location of transgene are proposed initiators for gene silencing (Kaufman et al., 2008). From this the influence of adjacent chromatin onto integration sites is termed 'position effect' (Lattenmayer et al., 2006; Yin et al., 2012).

1.6.3 Silencing of hCMV-MIE promoter

The promoter upstream of the recombinant gene initiates gene transcription and is able to affect gene expression level and stability (Kaufman et al., 2008). The major immediate early gene promoter of the human cytomegalovirus (hCMV-MIE) is commonly used to drive recombinant expression in mammalian cells for research and manufacturing to obtain high expression levels in transient and stable transfections (Boshart et al., 1985; Chapman et al., 1991; Foecking & Hofstetter, 1986; Wright et al., 2005). Although the hCMV-MIE promoter provides high gene expression levels, many studies have reported a decrease of productivity over long-term culture (Bailey et al., 2012; Barnes, Bentley & Dickson, 2001; He et al., 2012). The silencing of the hCMV-MIE promoter is (in addition to the loss of copy numbers) largely attributed to epigenetic events of promoter DNA methylation and histone modification (Brooks et al., 2004; Kim et al., 2011; Osterlehner et al., 2011; Paredes et al., 2013; Yang et al., 2010b).

1.7 Aims

CHO cell lines are the main production cell lines in the pharmaceutical industry to produce therapeutic/diagnostic proteins (Kim et al., 2012). In addition to the demand for high protein titer, cell lines need to be stable, to maintain the production level during the cell line development. To accomplish this, three independent epigenetic strategies were pursued.

Early exclusion of instable cell lines by an epigenetic prediction marker:

Finding an optimized prediction marker which indicates stable production in cell lines will reduce costs and increase efficiency in the cell line development. The epigenetic modification of CpG methylation of the strong hCMV-MIE promoter was successfully used as an indicator for long-term production stability (Osterlehner et al., 2011). However, other epigenetic marks such as histone modifications are also well known for their association with active or inactive chromatin, frequently proposed as events prior DNA methylation. Depending on epigenetic pathways, the loss of active marks or the accumulation of specific repressive marks are early events of a silencing cascade (Chen & Dent, 2014; Fadloun et al., 2013). In order to cover the most prominent pathways, I examined general lysine acetylation (H3ac) as well as the trimethylation of lysine 4, 9 and 27 of H3 (H3K4me3, H3K9me3, H3K27me3) as potential prediction markers of long-term transgene silencing.

Vector modification to impede silencing of transgenes:

Histone modification and DNA methylation are important steps in the activation and repression of genes. In general, CpG methylation of promoter DNA is proposed to maintain the repressed status, whereas the loss of methylation is associated with activation (Smith & Meissner, 2013). CpGs within the hCMV-MIE promoter were observed to comprise a position-dependent methylation pattern (Osterlehner et al., 2011). Considering this, I point-mutated three predominantly methylated CpGs from cytosine to guanine to investigate the resulting effect on long term production stability. In contrast to DNA methylation, the histone acetylation is associated with an open chromatin formation. H4K16 acetylation especially is a strong activation mark in the dosage compensation of *Drosophila* (Prestel et al., 2010). Thus, I constructed a vector for coexpression of a reporter and a modified histone acetyltransferase MOF, which is target oriented to UAS sites upstream of hCMV-MIE. This should lead to acetylated H4K16 close to the transgene promoter and an activation of the target region. Both strategies were executed to promote stable production in CHO cell line; either by preventing of transgene repression, or by forcing chromatin into an open formation.

Discovery of potential stable integration sites in the CHO-K1 genome:

It is a common approach in the pharmaceutical industry to develop production cell lines by random integration of transgene. However many cell lines need to be discarded during the cultivation phase in regard to their loss of productivity. Although the selection phase prefers cells with transgene expression, the possibility for a position effect like heterochromatinization increases over time. A target integration in constantly active genomic regions minimizes the risk of silencing. The chromatin of CHO-K1 cells was mapped for active (H3ac, H3K4me3, H3K36me3) and inactive (H3K9me3) regions and compared with the integration sites of cell pools of high or low productivity. The analysis of those sites in combination with the epigenetic landscape might lead to the identification of stable integration sites; those stable integration sites can then be used for prospective cell line developments.

2 Materials and methods

2.1 Material sources

2.1.1 Laboratory chemicals and biochemicals

Acrylamide (Rotiphorese Gel [®] 30)	Roth, Karlsruhe
Agar-Agar	Probio, Eggenstein
Agarose (ME, LE GP and low melting)	Biozym, Hessisch Oldendorf
Ampicilin	Roth, Karlsruhe
Bacto Agar	BD, Fance
BSA (Bovine serum albumin)	Sigma, Taufkirchen
β-Mercaptoethanol	Sigma, Taufkirchen, Germany
Casein	Roche, Penzberg, Germany
Chloroform/Isoamylalcohol (24:1)	Roth, Karlsruhe
Chloramphenicol	Roth, Karlsruhe
Coomassie	Serva, Heidelberg
DTT	Roth, Karlsruhe
5-aza-2deoxycytidine (DAC)	Sigma, Taufkirchen
EDTA	Sigma, Taufkirchen
EGTA	Sigma, Taufkirchen
Ethidium bromide	Sigma, Taufkirchen
Fetal bovine serum	Sigma, Taufkirchen
IPTG	Roth, Karlsruhe
NP40 (Igepal CA-630=	Invitrogen, Karlsruhe
Orange G	Sigma, Taufkirchen
Paraformaldehyde	Sigma, Taufkirchen
Phenol	Sigma, Taufkirchen
Phenol-chloroform	Sigma, Taufkirchen
PMSF (Phenylmethansulfonyl fluoride)	Roth, Karlsruhe
Ponceau S solution	Sigma, Taufkirchen, Germany
SDS (Sodium dodecyl sulfat)	Serva, Heidelberg
Roche Complete	Roche, Penzberg
rAPID alkaline phosphatase (Cat.no. 04898133001, Roche, Germany)	Roche, Penzberg
Roti [®] Phenol/Chloroform/Isoamylalcohol (25:24:1)	Roth, Germany
Tris	Invitrogen, Karlsruhe
Triton X-100	Sigma, Taufkirchen
Tween 20	Sigma, Taufkirchen

2.1.2 Enzymes

DNA Polymerase I, Large (Klenow) Fragment	New England BioLabs GmbH, Frankfurt (a.M), Germany
Proteinase K	Roche, Mannheim, Germany
Restriction endonucleases	New England BioLabs GmbH, Frankfurt (a.M), Germany
RNAse, DNase	Roche, Mannheim, Germany
Taq DNA Polymerase	Invitrogen, Darmstadt, Germany
T4 DNA Ligase	New England BioLabs GmbH, Frankfurt (a.M), Germany
ZymoTaq™DNA polymerase premix	ZymoResearch, Freiburg, Germany

2.1.3 Antibodies

Rabbit IgG, Purified, PP64B,	Millipore, Darmstadt, Germany
Rabbit H4K16ac-ab 07-329	Millipore, Darmstadt, Germany
Mouse GFP G 6539	Sigma, Taufkirchen, Germany
Mouse Flag F 3165	Sigma, Taufkirchen, Germany
Rabbit Gal4(DBD+O1679) Sc-577	Santa Cruz Biotech
Mouse Alpha-tubulin T 9026	Sigma, Taufkirchen, Germany
Goat anti rabbit 12-348	Millipore, Darmstadt, Germany
Goat anti mouse	Millipore, Darmstadt, Germany
ChIP grade Antibodies	
ChIPAb+ Trimethyl-Histone H3 (Lys4)	Millipore, Darmstadt, Germany
ChIPAb+ Trimethyl-Histone H3 (Lys27)	Millipore, Darmstadt, Germany
Anti-Histone H3 (tri methyl K9) antibody	Abcam
Anti-Histone H3 antibody	Abcam
Anti-acetyl-Histone H3	Millipore, Darmstadt, Germany
Histone H3K4me3 antibody (pAb)	Activemotif, La Hulpe, Belgium
Anti-acetyl-Histone H3 Antibody	Millipore, Darmstadt, Germany
ChIPAb+ Trimethyl-Histone H3 (Lys9)	Millipore, Darmstadt, Germany
ChIPAb+ Trimethyl-Histone H3 (Lys27)	Millipore, Darmstadt, Germany
Anti-Histone H3 antibody	Abcam

2.1.4 Organisms

E.coli NEB 5 alpha (NEB, Germany)	New England BioLabs GmbH, Frankfurt (a.M), Germany
XL10 Gold ultracompetent cells	Agilent Technologies, Oberhaching Germany

2.1.5 Oligonucleotides and plasmids

Primers:	Sequence:
Primer for bisulfite converted CMV	
UG_730_Bisul_CMV1 F1	GATATTGATTATTGATTAGTTATTAATAGTAATTA
UG_731_Bisul_CMV1 R1	CAAATAAAAAAATCCATAAAAATCATATACTAA
UG_732_Bisul_CMV2 F2	TTAGTATATGATTTTATGGGATTTTTTTATTG
UG_733_Bisul_CMV2 R2	TTCTAATACTAAACTCCTCTCCCAA
iPCR primer	
UG_574_16110 BtsI_rev	CTGTCATGCCATCCGTAAGATGCT
UG_575_16110 BtsI_for	GCGGCCAACTTACTTCTGACAACG
UG_576_16110 AceIII_rev	AAGGCGAGTTACATGATCCCCCAT
UG_577_16110 AceIII_for	CACCACGATGCCTGTAGCAATGG
Vector design primer	
UG_362_CMV_G-508F	CGTTACATAACTTAGGGTAAATGGCCCGCC
UG_363_CMV_G-179F	AATGGGCGTGGATAGGG TTTGACTCACG
UG_364_CMV_G-41F	GTAGGCGTGTAAGGTGGGAGGTCTATATAAG
UG_404_Linker_ClaI-UAS-SpeIF	AAAAAATCGATCAATTGGATATCGTCGACGGATCGGAGTACTGT
UG_405_Linker_ClaI-UAS-SpeIR	TTTTTTACTAGTCAATAATCAATGTCAACGCGTGGATCCGCTCGGAGGAC
UG_406_Linker_AgeI-UAS-SspIF	AAAAAAACCGGTGTCGACGGATCGGAGTACT
UG_407_Linker_AgeI-UAS-SspIR	TTTTTTAATATTAGCGCTGATCCGCTCGGAGGACA
UG_408_Linker_BclI-GFP-BsiWIF	AAAAAATGATCAAGGCTAGCGCCGCCACCATGGTGAGCAA
UG_409_Linker_BclI-GFP-BsiWIR	TTTTTTCGTACGCTCCTTAGGTTAACGGTCCATCCCGCTCTCCTG
UG_415_Rec._GFP-PolyAF	TGTCTTGTGCCCAGGAGAG
UG_416_Rec._GFP-PolyAR	TGAGTTTGGACAAACCACAAC
UG_417_Linker_BclI-GFP-BsiWIF	AAAAAATGATCAGCACTGAACACAGAGGAAGCTAGCCGCCACCATGGTG AG
UG_418_Linker_Bsu36I-IRESF	AAAACCTAAGGAGGGATATCGCGGCCGCG
UG_419_Linker_IRES-BsiWIR	TTTCGTACGATATTTAAATCTGGCCGAGGCGGCCAGGATCCATCGTGTTT TTCAAAGG
UG_420_Linker_SfiI-MOF F	AAAAAAGGCCGCTCGGCCTCTGA
UG_421_Linker_MOF-SwaIR	TTTTTTATTTAAATCTATTTGTGTCGTCGTCT
UG_422_Linker_MluI-CMVF	AAAAAACGCGTTGACATTGATTATTGACTAGTTAT
UG_423_Linker_CMV-NheIR	AAAAAAGCTAGCTTCTCTGTGTTTCAGT
ChIP primer	
UG_386_Chip_GusbF2	CAGGGTGGGATGCTCTTC
UG_387_Chip_GusbR2	GCCGGTTTTCCGAGAAGT

UG_388_Chip_RhoF1	AGCCTCGGTCTCTATTGACG
UG_389_Chip_RhoR1	CGTTGGAGAAGGGCACATAA
UG_396_Chip_CMVF3	TACATCAATGGGCGTGGATA
UG_397_Chip_CMVR3	AAGTCCCGTTGATTTTGGTG
UG_431_CHIP_Eif3iF1	GTTCCCGGCACTGACACT
UG_432_CHIP_Eif3iR1	ACTTGATCTGCGTGATGGAC
UG_445_CHIP_CMVF7	CATTATGCCAGTACATGACCTT
UG_446_CHIP_CMVR7	AAACCGCATCACCATGCTA
UG_468_CHIP_Gata5 F1	CACCTACCCATCCTGTCTG
UG_469_CHIP_Gata5 R1	GAGGAGGTGAAGGCAAAGTCT
UG_474_CHIP_UNC13c F1	GGGTGCTTTACGGAAACTGA
UG_475_CHIP_UNC13c R1	GCTTCTTATGCCCCAGGTTT
UG_484_CHIP_Foxa2 F1	ATCACCCGTA CTGCTGCTCT
UG_485_CHIP_Foxa2 R1	GAGGCTTCTGGGGATCTCTT

2.1.6 Other materials and kits

Kits	
AllPrep DNA/RNA Mini Kit (50)	Qiagen, Hilden, Germany
Dnease Blood & Tissue Kit	Qiagen, Hilden, Germany
EZ-96 DNA Methylation-Lightning™ TM Kit (deep-well format)	ZymoResearch, Freiburg, Germany
High Pure PCR Product Purification Kit	Roche, Penzberg, Germany
LightCycler®480 SYBR Green I Master	Roche, Penzberg, Germany
Lock Gel™ Light tube (Cat.no. 0032005.101.)	Eppendorf, Hamburg, Germany
Lumi-LightPLUSWester blotting system	Roche, Penzberg, Germany
(NuPAGE®Novex® 10% Bis-Tris Gel 1.0mm x 12 well, Cat. No. NP0302BOX	Invitrogen, Darmstadt, Germany
20x NuPAGE MOPS SDS Running Buffer, Cat.No. NP0001,	Invitrogen, Darmstadt, Germany
4xNuPAGE LDS-Samplebuffer	Invitrogen, Darmstadt, Germany
Pierce®BCA Protein Assay Kit,	Thermo Scientific, Rockford, USA
Protein A Agarose slurry	Roche, Mannheim, Germany
QIAquick® Gel Extraction Kit	Qiagen, Hilden, Germany
Quiagen Blood & Cell culture DNA prep. Midi Kit	Qiagen, Hilden, Germany
Quiagen Plasmid Maxi Kit	Qiagen, Hilden, Germany
Quiagen Plasmid Mini Kit	Qiagen, Hilden, Germany
QuikChange Multi Site-Directed Mutagenesis Kit	Agilent Technology, Waldbronn, Germany
Rapid DNA Ligation Kit	Roche, Mannheim, Germany)
Platinum PCR SuperMix High Fidelity	Invitrogen, Germany
Other materials	
Branson Sonifier B15	Branson, Dietzenbach, Germany
Branson Sonifier 45,	Branson, Dietzenbach, Germany

Flash-gel, FlashGel™systems	Lonza, Cologne, Germany
Cellavista CV3.1	Synentech Bio Services GmbH, Munich, Germany
Centrifuge 5415D	Eppendorf, Hamburg, Germany
Centrifuge 5417 R	Eppendorf, Hamburg, Germany
Centrifuge HeraeusMultifuge 35R+	Thermo Scientific
Flowcytometer FACSCalibur	BD-Bioscience, Heidelberg, Germany
Flowcytometer BD-FACSCanto II	BD-Bioscience, Heidelberg, Germany
gel documentation system INTAS	Intas Science Imaging Instruments GmbH, Göttingen, Germany
gel documentation system Roche Lumi-Imager F1	Roche, Mannheim, Penzberg
Incubator LabTerm-LT-XC	Kuhner AG, Birsfelden, Switzerland
Nanodrop 2000	PEQLAB Biotechnologie GmbH, Erlangen, Germany
PCR cycler GeneAmp(r) PCR system 9700	PE Applied Biosystems, Darmstadt, Germany
PCR cycler LightCycler®480 Instrument II System	Roche, Mannheim, Germany
Plate reader Tecan infinite M200 Pro	Tecan Group Ltd., Crailsheim, Germany
Power device Consort E802	Sigma, Taufkirchen, Germany
Mastercycler nexus X1	Eppendorf, Hamburg, Germany
Marker SeeBlue Marker prestained,	Invitrogen, Darmstadt, Germany
Marker MagicMark XP anti IgG	Invitrogen, Darmstadt, Germany
Salmon Sperm DNA	Invitrogen, Darmstadt, Germany
Xcell II Blot-Modul	Invitrogen, Darmstadt, Germany

2.1.7 Buffers and solutions

CHIP Cell lysis buffer (Gonzani lab.)	Stock	Vol for 1 L
20 mM TrisHCl pH 8.0	1M	20 ml
85 mM KCl	3 M	28 ml
0.5% NP40 (Nonident P-40/ octylphenolpolyethoxyethanol)	10%	50 ml
ddH2O		883 ml
Nuclei Lysis buffer (Gonzani lab.)	Stock	Vol for 1L
50 mM TrisHCl pH 8.0	1M	50 ml
10 mM EDTA pH 8.0	0.5 M	20 ml
1% SDS	10%	100 ml
ddH2O		830 ml
1 tablet Roche Complete per 10 ml		
Low salt wash buffer (Gonzani lab.)	Stock	Vol. for 1 L
0.1% SDS	10%	10 ml
1% Triton X 100	10%	100 ml
2 mM EDTA	0.5 M	4 ml
20 mM Tris-CL pH 8.0	1 M	20 ml

150 mM NaCl	4 M	37.5 ml
ddH2O		828.5 ml
High salt wash buffer (Gonzani lab.)	Stock	Vol. for 1 L
0.1% SDS	10%	10 ml
1% Triton X 100	10%	100 ml
2 mM EDTA	0.5 M	4 ml
20 mM Tris-CL pH 8.0	1 M	20 ml
500 mM NaCl	4 M	125 ml
LiCl wash buffer (Gonzani lab.)	Stock	Vol. for 1 L
0.25 M LiCl (Villagra et al. CHIP)	1 M	250 ml
1% NP40	10%	100 ml
1% docycolate (deoxycholic acid)		10 g
1 mM EDTA	0.5 M	2 ml
20 mM Tris, pH 8.0	1 M	20 ml
ddH2O		ad to 1 L
TE buffer	Stock	Vol. 1 L
10 mM TrisHCl pH 8.0	1 M	10 ml
1 mM EDTA	0.5 M	2 ml
(IP) Elution buffer (Farnham lab.)	Stock	Vol for 5 ml
50 mM NaHCO ₃	1 M	0.25 ml (or 21 mg direct)
1% SDS	10%	0.5 ml
ddH2O		Ad to 5 ml
IP Dilution buffer (Farnham lab.)	Stock	Vol for 5 ml
0.01% SDS	10% SDS	0.5 ml
1.1% Triton X 100	10% Triton X 100	0.5 ml
1.2 mM EDTA	0.5 M	12 µl
16.7 mM Tris-HCl pH 8.1 (8.0)	1 M Tris-HCl pH 8.0	83.5 µl
167 mM NaCl	4 M NaCl	209 µl
H2O		4.196 ml
IP Blocking buffer		Volume for 10 ml
IP Dilution buffer		10 ml
BSA		50 mg
Salmon Sperm DNA		1 mg
Roche Complete Protease Inhibitor		1 tablet

Triton X-100 lysis buffer	Concentration
Tris-HCl (pH 7.5)	50 mM
Triton X-100	0.5%
NaCl	137.5 mM
Glycerol	10%
Roche Complete EDTA (1 tab/10ml)	

2.2 Methods

2.2.1 Methods for DNA preparation and analysis

2.2.1.1 *Agarose gel electrophoresis*

Agarose gel electrophoresis was performed to analyze the quality, size and amount of linear DNA fragments (Chong, 2001). According to the size of DNA fragments, agarose solutions of 1% w/v were dissolved in 1 × TAE by boiling. Gels contained a final concentration of 0.5 µg/ml ethidium bromide. Samples were prepared by adding 1/6 v/v of 6 × DNA loading dye. A DNA ladder was used as a size standard. Electrophoresis was performed in 1 × TAE by applying 10 V/cm gel length (not more than 250 V). After separation, DNA was examined under UV light (254-366 nm) in a gel documentation system (Intas Science Imaging Instruments GmbH, Göttingen, Germany).

2.2.1.2 *Gel extraction*

After Agarose Gel Electrophoresis, the gel slice with the favored band was excised by scalpel and the gel slice was further extracted by QIAquick® Gel Extraction Kit (Qiagen, Hilden, Germany). Following this, three volumes of Buffer QG were added to one volume of gel (100 mg ~ 100 µl) and incubated for ten minutes at 50°C. One gel volume of isopropanol was added to the dissolved gel and mixed. DNA fragments adhered to QIAquick spin column by one minute centrifugation at 13,000 rpm and were washed by 0.75 ml Buffer PE and subsequent centrifugation (13,000 rpm, one minute). Second wash in another collection tube removed the residual wash buffer. Finally DNA was eluted in 30 µl RNase-free H₂O into a sample tube by centrifugation (13,000 rpm, one minute). DNA can be checked on Flash-gel (FlashGel™systems, Lonza, Cologne, Germany).

2.2.1.3 *DNA quantification*

NanoDrop 2000 (PEQLAB Biotechnologie GmbH, Erlangen, Germany) was used to quantify DNA concentration by measuring the optical density (OD) at a wavelength of 260 nm. The purity of the DNA can be judged by the ratio OD 260/280 and 260/230. Pure DNA preparations should possess a ratio ≥ than 1.8 and 2.0.

2.2.1.4 *Transformation of competent bacteria*

45 µl of chemically competent E.coli NEB 5 alpha (NEB, Germany) were thawed on ice for ten minutes and incubated with 1 pg – 100 ng of plasmid DNA (1 -5 µl) for 30 minutes on ice. The cell suspension was heat shocked at 42°C for 45 seconds and immediately chilled on ice for five minutes.

950 µl of room tempered SOC or LB was added and suspension was incubated at 37° for 60 minutes being agitated (250 rpm). Transformed bacteria were plated on pre-warmed agar plates containing ampicillin. Plates were incubated overnight at 37°C.

2.2.1.5 Plasmid preparation

Plasmids were prepared using the Qiagen Plasmid Mini and Maxi Kits (Qiagen, Hilden, Germany) following the manufacturer's instructions.

2.2.1.6 DNA purification

Genomic DNA was isolated and purified with the AllPrep DNA/RNA Mini Kit (50) (Cat.no.80204, Qiagen, Hilden, Germany) and the Dnease Blood & Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's recommendations. Purification of fragments was done with the High Pure PCR Product Purification Kit (Roche, Penzberg, Germany) by following recommended protocol.

2.2.1.7 Phenol-chloroform extraction

Plasmid concentration and purification was done by Phenol/Chloroform extraction. All steps were performed under a chemical hood. 500 µl of linearized plasmid was mixed with 1 volume Roti®Phenol/Chloroform/Isoamylalcohol (25:24:1) (Cat.no.A156, Roth, Germany) and vortexed for 30 seconds. Short spin removes liquid from lid and emulsion was transferred in precentrifuged Phase Lock Gel™ Light tube (Cat.no. 0032005.101, Eppendorf, Hamburg, Germany) for centrifugation (13,000 rpm, one minute, Eppendorf table centrifuge). Upper aqueous phase was transferred into new tube and extraction was repeated. Thereafter the upper aqueous phase was mixed with 500 µl Chloroform/Isoamylalcohol (24:1) and extraction steps were repeated once. The upper phase was transferred into a new tube and mixed with 0.1 start volume of 3 M NaAc (pH 4.8-5.2) and 0.7 start volume of 100% isopropanol (20-30°C) by inverting tube 4-6 times. 20 minutes incubation at -80°C followed by centrifugation (13,000 rpm, 30 minutes at 4°C) resulted in DNA precipitation and sedimentation. Supernatant was discarded. Pellet was washed with chilled 1 ml 70% EtOH, centrifuged (13,000 rpm, five minutes at 4°C) and supernatant was discarded. Wash step was repeated under sterile conditions and residual supernatant was completely removed. Pellet was dried for five to ten minutes on sterile air and subsequently up taken in 0.5 start volume ddH₂O.

2.2.1.8 Transgene copy number determination

Determinations of transgene copy number as well as methylation rate of C-179 of hCMV-MIE promoter were performed by Silke Simmeth and Tobias Scherzinger according to the recommendations of Osterlehner (Osterlehner et al., 2011).

Relative Copy number (rCN) was calculated by two to the power of the subtraction of the Cq value of the reference input sample (e.g. Gusb) from the Cq value of the target input sample hCMV-MIE promoter.

$$\text{Relative Copynumber} = 2^{(C_q \text{ Input sample}_{\text{target}} - C_q \text{ Input sample}_{\text{reference}})}$$

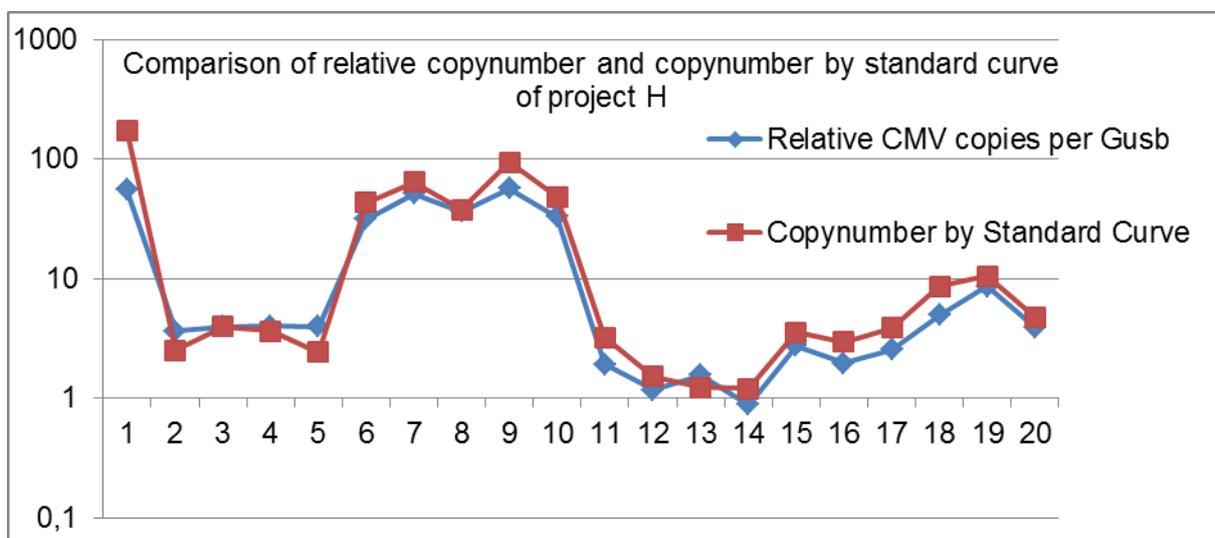


Figure 10: Comparison of relative copy number of transgene and the calculated copy number by standard curve. Y-axis represents the relative copy number and x-axis displays the sample numbers.

2.2.1.9 Cloning of reporter constructs

All transfection vectors were designed with the p5532 backbone as exemplified above. Depending on assay UAS sites, GFP-PEST, IRES and Gal4BD-MOF were integrated into backbone (Figure 12).

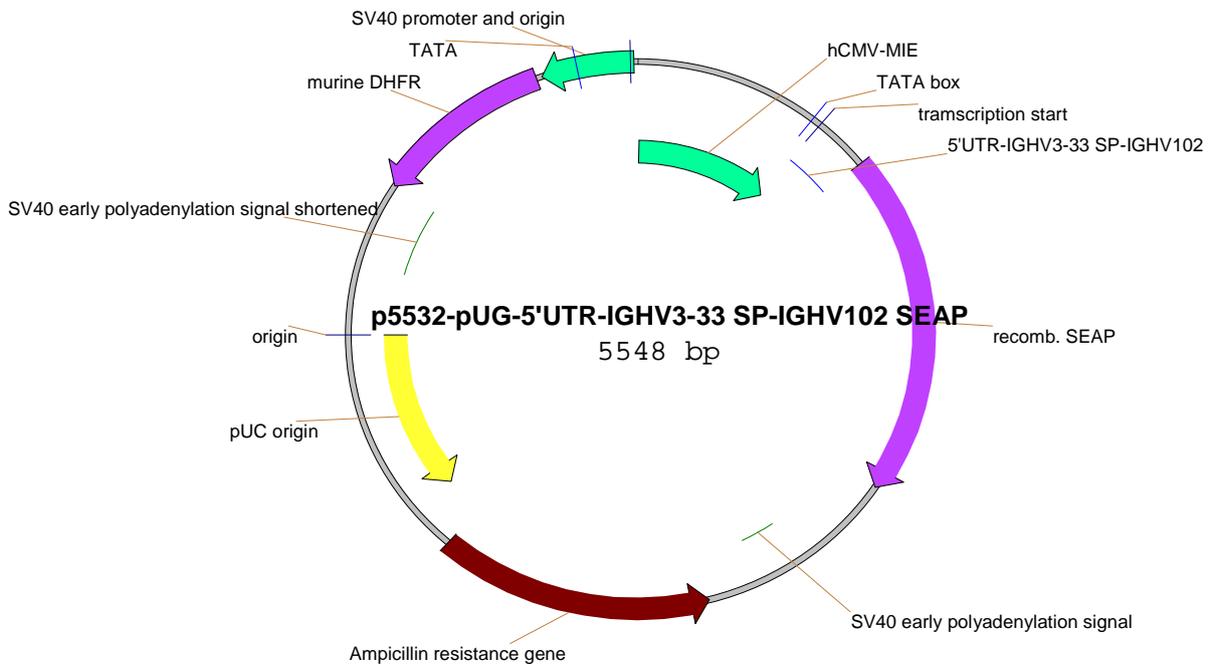


Figure 11: Plasmid design for measuring hCMV-MIE promoter strength by expression of secretory alkaline phosphatase (SEAP).

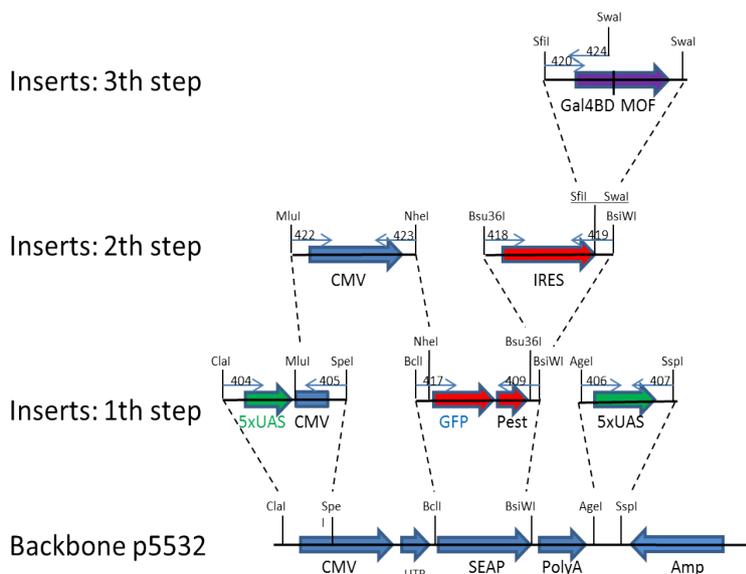


Figure 12: Cloning strategy for transfection vectors. UAS sites, GFP-Pest, IRES and Gal4BD-MOF were integrated into vector backbone. Testing CpG point mutations of hCMV-MIE was done in a vector without IRES and Gal4BD-MOF.

The sequences for UAS and Gal4BD-MOF were kindly provided by Prof. Dr. Peter Becker (Adolf Butenandt Institute, Munich, Germany). The EV71-IRES DNA as well as the eGFP and Pest DNA were kindly provided by the in-house groups of Dr. Peter Hülsmann and Dr. Gwendlyn Kollmorgen (Roche, Penzberg, Germany). Fragments were amplified by PCR, according to integration sites, primers were equipped with appropriate cutting sites. Digestion of integrates and backbone was done with the appropriate restriction enzymes (NEB, Germany) according to manufacturer's protocol. Dephosphorylation of vectorbackbones was done with rAPID alkaline phosphatase (Cat.no. 04898133001, Roche, Germany). Thereby 30 µl linearized vector (< 1 pmol) were incubated with 3.6 µl 10x Dephosphorylation buffer and 2 µl (1 Unit) alkaline Phosphatase for ten minutes at 37°C, followed by 15 minutes at 65°C and chilling on ice. Ligation of fragments was done with Rapid DNA Ligation Kit (Cat.no. 1635379, Roche, Germany) according to manufacturer's protocol. Ligation was then performed at a ratio 1:3 (backbone:insert) for standard and 1:5 for sticky ends. The number of copies was calculated with the following formula:

$$\text{Number of copies} = \frac{[\text{amount (ng)} * \text{avrogrado constant}]}{[\text{length (bp)} * 10^9 \left(\frac{\text{ng}}{\text{g}}\right) * 650 \left(\frac{\text{g}}{\text{mole of bp}}\right)]}$$

All constructs were verified by digestion analysis and sequencing (Sequiserve GmbH, Vaterstetten, Germany). After verification of correct sequence, plasmids were amplified in E.coli and purified with Qiagen Plasmid Mini and Maxi Kits (Qiagen, Hilden, Germany).

2.2.1.10 Site directed mutagenesis

Influence of major methylated CpG sites of the hCMV-MIE promoter was investigated by C to G point mutation of predominately methylated CpG sites in various combinations.

Mutation of CpG sites of the hCMV-MIE promoter was done by QuikChange Multi Site-Directed Mutagenesis Kit (Cat.no 200514-12, Agilent Technology, Waldbronn, Germany). The hCMV-MIE promoter was mutated at C-508, C-179 and C-41 upstream of transcription start site with following primers in various combinations.

Primer number	Position of point mutation	Primer sequence
1	C-508	CGTTACATAACTTAGGGTAAATGGCCCGCC
2	C-179	AATGGGCGTGGATAGGGG TTTGACTCACG
3	C-41	GTAGGCGTGTA GG GTGGGAGGTCTATATAAG

Table 1: Primer sequences for C to G mutation of hCMV-MIE promoters to investigate DNA methylation dependent silencing effects

Mutation was done according to manufacturer's protocol:

Mutant strand synthesis was performed by thermal cycling to denature DNA template, anneal the mutagenic primers (primers bind same strand) and extend primers and ligate nicks with the QuikChange Multi enzyme blend. The following digestion of template was executed to digest methylated and hemimethylated DNA with restriction enzyme DpnI. The mutated ssDNA was transformed into XL10 Gold ultracompetent cells.

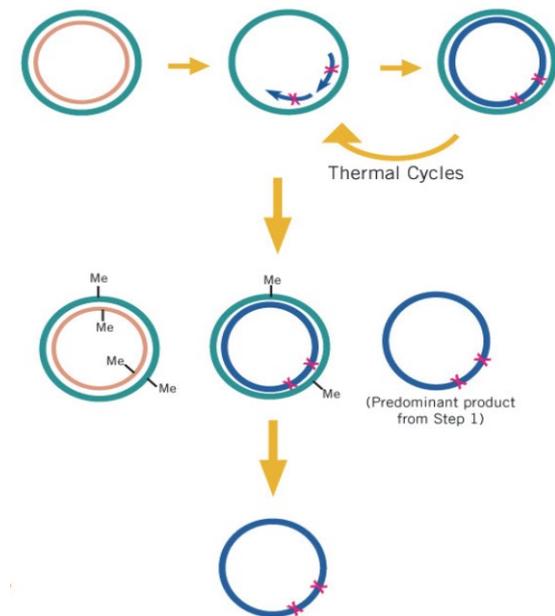


Figure 13: Overview of the QuikChange Multi Site-Directed Mutagenesis method. Adapted from Agilent technologies protocol.

Position of C to G mutations upstream of TSS	Code
-508, -179, -41	GGG
-508, -41	GCG
-179	CGC
-508	GCC
-41	CCG

Table 2: Combinations of C to G point mutations of hCMV-MIE. Those variations were used to investigate DNA methylation dependent silencing effects.

Transformed E.coli colonies were picked, incubated and plasmids were prepared. Finally plasmids were sequenced by Sequiserve (Sequiserve GmbH, Vaterstetten, Germany) and hCMV-MIE promoters with desired combinations (displayed in Table 2) of point mutations were used for further investigations of silencing effect.

2.2.1.11 Chromatin immunoprecipitation

CHO cells lines were harvested with a viability of greater than 97% to obtain comparable results. Before fixation, agarose beads for preclearing and precipitation were prepared. Preclearing beads were generated by washing (one hour, RT while rotation) of 15 µl Protein A Agarose slurry (Roche, Mannheim, Germany) two times in 120 µl Dilution buffer followed by incubation with 2 µl Rabbit IgG, Purified (PP64B, Millipore, Germany) in 120 µl ChIP Dilution Buffer per sample. Two Wash steps with 120 µl of ChIP Dilution Buffer were done for purification and agarose beads were resuspended in 85 µl ChIP Dilution buffer. In addition 15 µl Agarose A Beads slurry for precipitation were blocked in 1 ml ChIP Dilution buffer with 5 mg/ml BSA (Roche, Penzberg) and 100 µg/ml of preheated Salmon Sperm DNA (ten minutes at 95°C and five minutes on ice; Cat no. 15632-011, Invitrogen, Germany). Resuspended pellet were incubated for four to five hours at 4°C while rotation and subsequently washed three times and mounted in 40 µl of ChIP Dilution Buffer.

1.10⁷ cells per sample were fixed by adding formaldehyde into media up to a final concentration of 3.7% and incubated for ten minutes at RT as described previously (Beneke et al., 2012). Fixation was stopped by adding glycine up to a one fold solution. After two wash steps with ice cold PBS, and centrifugation at 2,000 g for three minutes, pellet was resuspended in 1 ml PBS plus protease inhibitor Roche Complete (Roche, Penzberg, Germany). The suspension was pelleted by centrifugation (3,000 g, five minutes, 4°C) and supernatant was discarded.

Lysation of pellet was done by adding 1 ml cell lysis buffer plus Roche Complete (Roche, Penzberg, Germany) and incubation on ice for ten minutes. Centrifugation (2,300 g, four minutes, 4°C) resulted in nuclei pellet which was resuspended in 300 µl Nuclei Lysis Solution and sonicated (Output 5, Duty cycle 90%, 15 seconds sonication followed by two to three minutes incubation on ice for six cycles; Branson Sonifier B15, Dietzenbach, Germany). 700 µl of Nuclei Lysis Solution were added to sonicated nuclei followed by centrifugation (13,000 rpm, 15 minutes, 4°C). The supernatant contains the chromatin and was transferred into a new tube for storage at -80°C until use. Sonification grade was tested after protein digestion on an agarose gel (Figure 14).

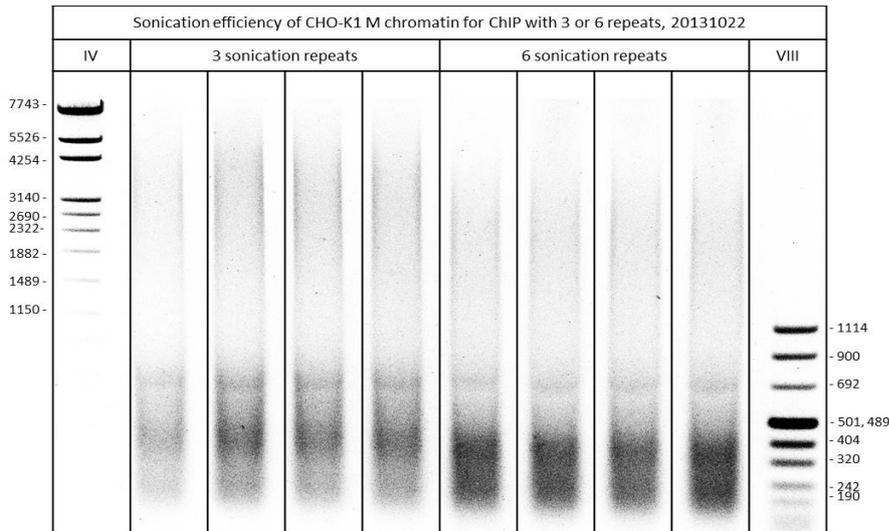


Figure 14: Sonification efficiency test. CHO-K1 M chromatin was sheared with three and six repeats to obtain the optimal fragment size.

Chromatin was precleared with 80 μ l prepared protein A agarose slurry by incubation (two hours at 4°C while rotation) and centrifugation (13,000 rpm, six minutes, 4°C, Eppendorf 5417R, Eppendorf, Hamburg, Germany). Supernatant was transferred into new tube and protein concentration was determined (Pierce®BCA Protein Assay Kit, Thermo Scientific, Rockford, USA). 25-100 μ g of chromatin per immunoprecipitation (IP) were mounted in 200 μ l Nuclei Lysis buffer and added to 300 μ l ChIP Dilution buffer comprising 3 μ g of appropriate antibody. Immunoprecipitation was done overnight at 4°C while under rotation. Undiluted input samples were stored at -20°C. After overnight incubation, 40 μ l of blocked Protein A Agarose beads were added to IP solution and incubated (one hour at 4°C while rotating). Precipitation was washed with 2 x low-, 1x with high-salt wash buffer, 1 x with LiCl wash buffer and additional with 1x TE wash buffer. Each wash was done in 1 ml buffer for five minutes on rotating platform followed by centrifugation (500 g, 30 seconds, 4°C). After the last wash step beads were centrifuged with 500 g for one minute. Bead pellets were combined with 200 μ l IP Elution buffer, while simultaneously 25-100 μ g chromatin input sample were filled up to 200 μ l with IP Elution buffer and incubated for 30 minutes at 65°C while shaking. Subsequently tubes were incubated for 30 minutes at 37°C after adding 0.5 μ l RNase DNase free (Roche, Germany) to each tube. For final protein digestion 10 μ l 4 M NaCl (final conc. 0.2 M) and 2 μ l Proteinase K (final conc. 100-200 μ g/ml, Roche Germany) were added to the reaction samples and incubated for 1.5 hours at 65°C. DNA was purified by Roche PCR purification Kit (Roche, Germany).

2.2.1.12 Bisulfite conversion

For conversion of unmethylated cytosines into uracil the EZ-96 DNA Methylation-Lightning™ Kit (deep-well format) (ZymoResearch, Freiburg, Germany) was used according to manufacturer's instructions. In brief, gDNA was extracted with Dnease Blood & Tissue Kit (Qiagen, Hilden, Germany) and concentration was measured with NanoDrop 2000 (PEQLAB Biotechnologie GmbH, Erlangen, Germany) to set concentration at 350 ng gDNA in 20 µl ddH₂O. 20 µl gDNA were mixed with 130 µl of Lightning Conversion Reagent in a Conversion Plate, afterwards were incubated at 98°C for eight minutes, then at 54°C for 60 minutes and finally temporarily stored at 4°C. A Zymo-Spin™ I-96 Binding Plate containing 600 µl of M-Binding Buffer per well, was mounted on a Collection Plate. Samples from the Conversion Plate were added to the Zymo-Spin™ I-96 Binding Plate, mixed and centrifuged (3,000 g for five minutes). Plates were washed with 400 µl fo M-Wash Buffer and centrifuged (3,000 g, five minutes). 200 µl of L-Desulphonation Buffer were added to each well, incubated (20-30°C, 20 minutes) and subsequently centrifuged (3,000 g for five minutes). An additional wash step was completed with ten minutes centrifugation at 3,000 g and converted DNA was eluted in 30 µl M-Elution buffer. Converted DNA was stored at -20°C until use.

Amplification of converted and integrated hCMV-MIE promoter:

For investigation of CpG methylation of hCMV-MIE promoter, primer pairs for distal (F1 & R1) and proximal (F2 & R2) promoter region were designed as follows:

UG_730_Bisul_CMV1 F1	gatattgattattgattagttattaatagtaattaa
UG_731_Bisul_CMV1 R1	caaataaaaaaatcccataaaatcatatactaa
UG_732_Bisul_CMV2 F2	ttagtatatgattttatggattttttatttg
UG_733_Bisul_CMV2 R2	Ttctaactaaactcctctcccaa

Table 3: Primers for amplification of distal and proximal hCMV-MIE promoter region.

10 pmol of each primer (0.5 µl) were added to 12.5 µl ZymoTaq™DNA polymerase premix (ZymoResearch, Germany), 3 µl gDNA template and filled up to 25 µl with ddH₂O. Polymerase chain reaction (PCR) was done with Mastercycler nexus X1 (Eppendorf, Hamburg, Germany), PCR conditions were as follows:

95°C	10 min	2 cycles
94°C	30 sec	
49°C	2 min	
68°C	2 min	
94°C	30 sec	34-43 cycles
54°C	2 min	
68°C	2 min	
72°C	10 min	
4°C	∞	

Table 4: Adjustment of Mastercycler nexus X1 conditions.

Amplicons were purified with Roche PCR Purification Kit (Roche, Germany) in 30-40 µl Elution buffer. Concentration was defined with NanoDrop 2000 (PEQLAB Biotechnologie GmbH, Erlangen, Germany) and size of amplicons were displayed by agarose gel electrophoresis.

2.2.1.13 Inverse PCR

The inverse Polymerase Chain Reaction (iPCR) is a method to determine integration sites of a randomly integrated vector. From this the known vector-sequence near the cutting site will be used to generate primers of at least 20 bp length. Those primers are oriented in opposite directions.

The genome was digested with frequently cutting enzymes (4 bp cutter) which did not cut between the primer binding sites and at the linearization site of the vector. The resulting fragments are ligated at very dilute conditions so as to favor intramolecular ligation. The circularized sequence can now be amplified with the inverse primers.

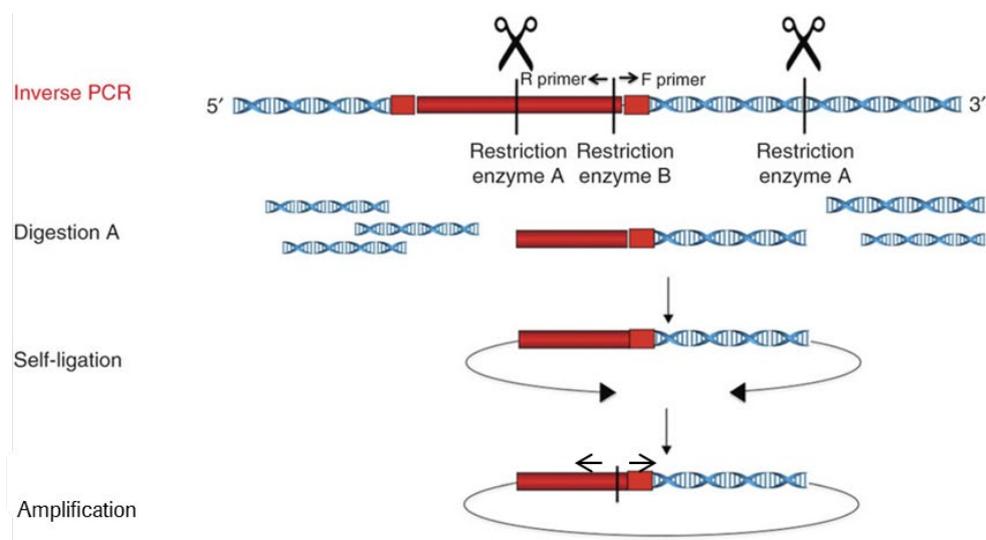


Figure 15: Schema of inverse PCR. Adapted from (Papapetrou & Sadelain, 2011). Well-chosen restriction enzymes cut in a defined region of vector sequence near the integrations site and in the unknown area of genome. Self-ligation under highly diluted condition is favored. Primers bind the known vector sequence and amplify the circularized fragment. Sequencing of amplicons leads to the identification of integration sites.

Genomic DNA was isolated with Quiagen Blood & Cell culture DNA prep. Midi Kit (Qiagen, Hilden, Germany) out of 2×10^7 cells ($\sim 90 \mu\text{g}$ DNA) per condition (genomic DNA can be stored at -20°C for several months) and dissolved in $90 \mu\text{l}$ ddH₂O.

The isolated genomic DNA was digested with differing restriction endonucleases at the appropriate working temperature for a minimum of 16 hours (overnight). For the digestion of gDNA the Enzymes CviQI, MseI and MspI for the iPCR upstream of PvuI integration site and the enzymes MseI, BfaI and MluCI for the iPCR downstream of the PvuI integration site were used. $10 \mu\text{g}$ genomic DNA were used per digestion ($3,8 \times 10^6$ copies).

Digestion:

Component	Volume / 80µl	Final amount
gDNA template	x µl	10 µg
NEBuffer 10×	8 µl	
BSA 10× (if required)	8 µl	
ddH2O	Add to 80 µl	
Restriction endonuclease	x µl	50 U*

Table 5: Digestion of gDNA for inverse PCR. Long incubation increases change of complete digestion.

The digested DNA was purified with the Roche PCR-purification kit and eluted in x µl (200 µl) Elution buffer. The high dilution grade of DNA favors intramolecular ligation.

Ligation:

Neb T4 DNA Ligase (M0202S) for iPCR	
Digested DNA	20 µl (1 µg) (3,8x10 ⁵ copies)
Ligation buffer, 10x	50 µl
T4 DNA ligase (400 Uµl ⁻¹)	1 µl
WFI (ddH2O)	429 µl

Table 6: Ligation of digested gDNA. High Dilution leads to preferred self-ligation.

In order to generate circular DNA fragments, the digested DNA was ligated overnight at 16°C using T4 DNA ligase (NEB, Frankfurt/Main, Germany). Ligated DNA was eluted in 50 µl Elution buffer of the High PurePCR Purification Kit (Roche, Penzberg, Germany). *Phenol-chloroform extraction is not recommended by GATC guidelines.* For iPCR of circularized fragments, 10 µM of each primer (574 & 575 or 576 & 577) and 45 µl of Platinum PCR SuperMix High Fidelity (Invitrogen, Germany) were added to 10 ng template DNA. As a control group, untransfected gDNA for CHO cell line was investigated.

Primer	Sequence	Location
UG_574_16110 BtsI_rev	CTGTCATGCCATCCGTAAGATGCT	5' PvuI Integration Site
UG_575_16110 BtsI_for	GCGGCCAACTTACTTCTGACAACG	
UG_576_16110 AclIII_rev	AAGGCGAGTTACATGATCCCCCAT	3' PvuI Integration Site
UG_577_16110 AclIII_for	CACCACGATGCCTGTAGCAATGG	

Table 7: Primer for inverse PCR. Appropriate forward and reverse Primer are in opposite direction. Circularization of template DNA results in amplification.

The iPCR for 5' iPCR conditions are described as follows:

Initial Denaturation	94°C	7 min	28 cycles
Denaturation	94°C	45 sec	
Annealing	60°C	45 sec	
Elongation	68°C	4 min	
Final Elongation	68°C	7 min	
Short storage	4°C	∞	

Table 8: iPCR conditions. Extended elongation time minimize the loss of amplification according to fragment length.

5 µl of PCR fragment was mixed with Orange G (5 µl) and ddH₂O (10 µl) and loaded on agarose-gel to verify the distribution of fragments. For 5' iPCR of MseI and MluCI cutted fragments good results were obtained and no unspecific band in control lanes was detected.

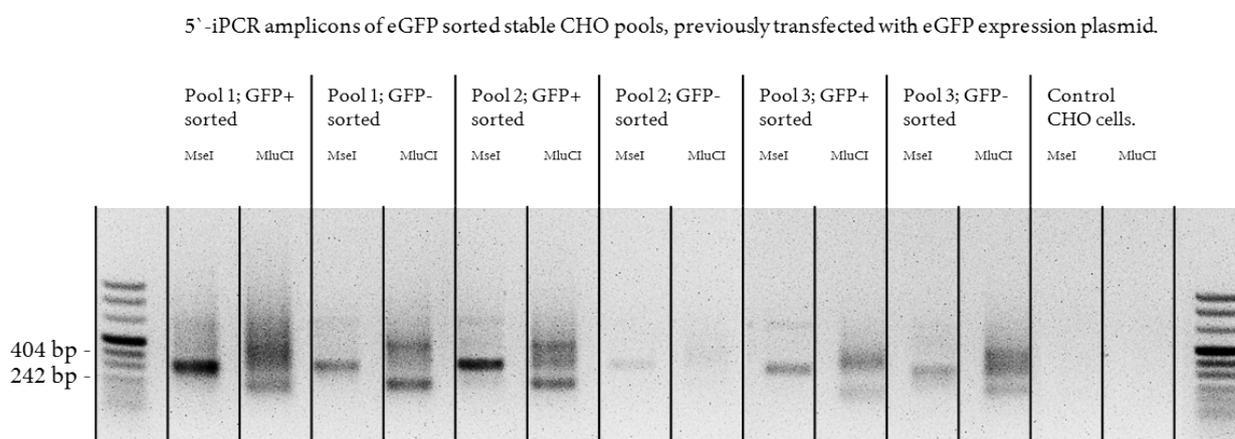


Figure 16: 5' iPCR of stable CHO pools after eGFP sort. Pools were randomly transfected with eGFP reporter plasmid and MTX selected. After one month of cultivation, pools were positive and negative eGFP sorted and integration sites were examined with iPCR. No unspecific amplifications were obtained in untransfected CHO control cells using 5' integration site primers.

5' amplicons of each sample (cut with MseI and MluCI) were combined and purified with High Pure PCR Purification Kit (Roche, Penzberg, Germany) in 40 µl Elution buffer. Sequencing of fragments was done by GATC-Biotech (Konstanz, Germany).

2.2.1.14 Quantitative real-time polymerase chain reaction

For detection of relative amounts of a target sequence the quantitative real-time polymerase chain reaction (qPCR) is the method of choice. Based on the principle of PCR, a target sequence is amplified by target specific primers. In addition, each amplification cycle is measured in real time. For each qPCR the non-specific fluorescent dye LightCycler®480 SYBR Green I Master in the LightCycler®480 Instrument II System (Roche, Penzberg, Germany) was used. The LightCycler®480

System is a high-performance, medium- to high-throughput PCR platform (96- or 384-well plates) that can be integrated as a robotically controlled, automated high-throughput solution. SYBR Green is excited using blue light ($\lambda_{\max} = 488 \text{ nm}$) and it emits green light ($\lambda_{\max} = 522 \text{ nm}$). The SYBR Green dye binds to all double-stranded DNA in PCR. As a result the fluorescence intensity and the amount of DNA product increase simultaneously, which can be detected by the LightCycler® System after each cycle. For quantification the fluorescence was plotted against the number of cycles on a logarithmic scale. Slightly above the emitted background the threshold for detection of DNA-based fluorescence was set by the LightCycler® System. The number of cycles at which the fluorescence pass the threshold is termed quantification cycle (Cq) (Bustin et al., 2009). During the exponential amplification phase a doubling of target DNA is expected in every new cycle. However, the efficiency of amplification is often variable among primers and templates. Analysis of the melting temperature of amplified DNA fragments gives first hints of the specificity of used primer pairs and the amount of primer dimers.

An efficiency of primer template combination can be assessed by a titration experiment to create a standard curve or by efficiency calculation program like LinRegPCR. The program uses non-baseline corrected data from the LightCycler® System, performs a baseline correction on each sample separately, determines a window-of-linearity and then uses linear regression analysis to fit a straight line through the PCR data set. From the slope of this line the PCR efficiency of each individual sample is calculated (Ruijter, Velden & Ilgun, 2009).

For relative quantification the amount of target sequence were compared to a reference sequence, by subtracting Cq (target) from Cq (reference). This normalization is termed ΔCq -method (Scheffe et al., 2006). The reference sequence has to have very stable values, considerably higher than the background values (Mock). Therefore for each reference sequence, the average and the coefficient of variation of all samples within one project was calculated and compared to background levels. Cq values of the different conditions were displayed as % of the Input Sample, visualizing the distance to mock control.

$$\% \text{ of Input Sample} = 100 * 2^{\Delta\text{Cq}(\text{Input Sample} - \text{ChIP Sample})}$$

Regarding the large number of samples, all qPCR were accomplished in 384 well plates. In each well following ingredients were transferred.

Volume (μ l) 384 wellplate	Solution
5	Sybr Green MasterMix
0.5	Primer (10 pM) for.
0.5	Primer (10 pM) rev.
1.5	H2O
2.5	DNA (X ng/ μ l)
10	Total

Table 9: qPCR mix. DNA template has normally a concentration around 1 ng/ μ l. ChIP samples have a low number of target sequences. Therefor concentration could not be estimated.

The qPCR program was designed as follows:

Appliance	Temperature ($^{\circ}$ C)	Time	Degree ($^{\circ}$ C/sec)	Cycles
Pre-Denaturation	95	10 min	4.8	1
Amplification	95	10 sec	4.8	40
	60	5 sec (single)	2.5	
	72	15 sec	4.8	
Melt curve	95	5 sec	4.8	1
	70	1 min	2.5	
	95	(continuously)		
Cooling				

Table 10: qPCR program of ChIP samples accomplished in the LightCycler[®]System. Number of amplification cycles can vary between 40-45. All Cq values scratching 40 cycles cannot be trusted anymore, because of the increase of unspecific amplification.

In the special case of the relative quantification of histone modifications, the normalization of sample (treated) to input sample (untreated) is directly replaced by the normalization of treated target sequence to treated reference sequence. Further calculation of histone modification per histone makes the normalization to input sample redundant.

The relative amount of specific histone modification close to human CMV major immediate-early promoter/enhancer fragment were estimated with the Livak method, also known as delta delta Cq ($\Delta\Delta$ Cq) method (Livak & Schmittgen, 2001), as long as primer-template efficiencies are close to 2 and close to each other (< 2% difference of Cq mean).

The method can be used for relative quantification of a target regarding to a reference. In this example the relative quantification of H3 acetylation and H3 levels close to human CMV major immediate-early promoter/enhancer fragment were normalized to the histone 3 acetylation and the histone 3 levels close to reference gene in two steps. First ΔCq was calculated as follows:

$$\Delta Cq = \text{control} - \text{sample}$$

ΔCq is the distance of quantification cycles (Cq) between control (e.g. ChIP sample value of Gusb) and target (e.g. ChIP sample value of hCMV-MIE) of the same sample, amplified with different primer pairs. To identify the level of H3 modification per H3, the $\Delta\Delta Cq$ method was used.

$$\Delta\Delta Cq \rightarrow \text{ratio} = \frac{2^{\Delta Cq_{\text{target}}(\text{control-sample})}}{2^{\Delta Cq_{\text{ref}}(\text{control-sample})}}$$

2.2.1.15 Sequencing methods

Sanger sequencing was performed by the company Sequiserve (Sequiserve GmbH, Vaterstetten, Germany). Next generation sequencing was performed by GATC (GATC-biotech, Konstanz, Germany) and Active Motif (La Hulpe, Belgium) and in house by myself using Illumina Sequencing technology.

2.2.1.16 Sequencing of amplicons after bisulfite conversion

I performed sequencing of bisulfite converted amplicons by employing the MiSeq system (Illumina, Inc., San Diego, USA). At first amplicon concentration was measured with the Qubit system (Life Technologies GmbH, Darmstadt, Germany) and library was generated according to the TruSeq Nano DNA Library Prep Guide (Part # 15041110 Rev. C, Illumina, Inc., San Diego, USA) using 100 ng per amplicon. Library quality was checked with the Agilent 2100 bioanalyzer (Agilent Technologies, Waldbronn, Germany) using a 1/100 dilution of dsDNA with the High Sensitivity DNA Assay (Agilent Technologies, Waldbronn, Germany) (Figure 17).

Hence in each sample two clear peaks were observed. Those peaks are the result of the combined distal and proximal hCMV-MIE amplification and demonstrate the quality of each sample. After verification of quality, all 56 amplicons were pooled in equal shares and mixed with 10% PhiX before run.

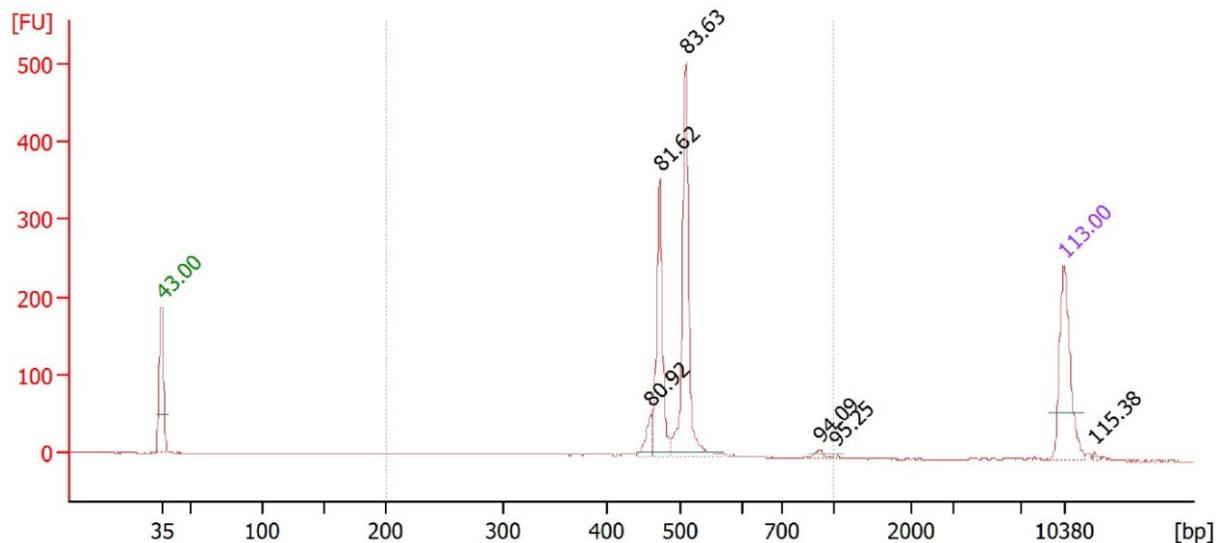


Figure 17: Electropherogram of one sample including the proximal and distal amplicons of bisulfite converted hCMV-MIE. The clear peaks between 400 and 600 base pairs confirm the good quality of the library.

Reads were mapped with the Bismark software (Krueger & Andrews, 2011) which uses Bowtie2 (Langmead & Salzberg, 2012) for alignment. Bismark was performed with the following commands:

```
/data64/sequencing/LMR_IntegrationSite_BenjaminMoritz/bisulfite_data/bismark/bismark_v0.14.3/  
bismark --non_directional --bowtie2 -o $(dirname $file1) -p 8 $reference -1 $file1 -2 $file2
```

For this purpose following alignment parameters for non-directional reads were used according to the manual and calculation was performed by Matthias Barann. The output of the bismark_methylation_extractor includes --include_overlap, --bedGraph, --cytosine_report and --report

2.2.1.17 Analyses of next generation sequencing data

Alignment of reads and mapping was performed by Matthias Barann. Reads of the integration sites, generated by inverse PCR, and the ChIP sequencing reads for H3K36me3, H3ac, H3K4me3 and H3K9me3 were aligned to the CHO-k1 genome (genBank Assembly ID GCA_000223135.1), employing the Burrow-Wheeler alignment tool BWA (Li & Durbin, 2009)

BWA was performed with the following commands:

```
reference=/data64/sequencing/LMR_IntegrationSite_BenjaminMoritz/reference/cricetulus_griseus_
merged.fasta
/apps64/bidev/ngs/bwa/bwa-0.7.10/bwa aln -t 12 $reference $file1 > $(basename $file1).sai
/apps64/bidev/ngs/bwa/bwa-0.7.10/bwa samse $reference $(basename $file1).sai $file1 >
$(basename $file1).sam
samtools view -b $(basename $file1).sam > $(basename $file1).bam
samtools sort $(basename $file1).bam $(basename $file1)_sorted
samtools index $(basename $file1)_sorted.bam
```

The analysis of ChIP peaks was performed with MACS2 (Zhang et al., 2008b) and the subsequent commands were given:

```
callpeak-t /path...file.name(H3K36me3) .bam -c /path...file.name(Input control).bam -f BAM-g hs
(genome size) -n file.name(output H3K36me3).bam -outdir /path.../(broad)—buffer-size 10000
```

The command broad was only added for H3K36me3 and H3K9me3 reads.

The arguments were as follows:

- name = 2_2578Roche_CHO_K1_M_rep_H3K36me3_i81_sorted.bam
- format = AUTO
- ChIP-seq file = [`/path...file.name(H3K36me3).bam`]
- control file = [`/path...file.name(Input control).bam`]
- effective genome size = 2.7e+09
- band width = 300
- model fold = [5,50]
- and the quality values were
- qvalue (minimum false discovery rate FDR) cutoff for narrow/strong regions = 5.00e-02, or 1.00e-01 for broad/weak regions respectively.
- Larger dataset will be scaled towards smaller dataset
- Broad region calling is on for H3K36me3 and H3K9me3 whereas it is off for H3ac and H3K4me3.

2.2.2 Methods for protein analysis

2.2.2.1 Protein quantification

Protein concentrations were estimated in comparison to a protein standard with the Pierce®BCA Protein Assay Kit (Thermo Scientific, Rockford, USA).

2.2.2.2 Protein fractionation in compartments

Proteins were fractionated in membrane/cytoplasmic and nuclear proteins (Misawa Y, 2006). Following this CHO cells (1×10^7 cells per sample) were sedimented by centrifugation (300 g, three minutes at 4°C) and the pellet was washed twice in ice-cold PBS + Roche Complete (Roche, Penzberg, Germany). Washed pellet was lysed with 1 ml 0.5% Triton X-100 lysis buffer and incubated on ice for 15 minutes. Insoluble nuclei were separated by centrifugation (13,000 rpm, 15 minutes at 4°C) with Eppendorf table centrifuge (Eppendorf, Hamburg, Germany). Membrane/cytoplasma containing supernatant was transferred into a new tube and chilled on ice. Nuclear pellet was rinsed with lysis buffer once and resuspended in 300 µl lysis buffer containing 0.5% SDS followed by sonication (five seconds, Output 2, duty cycle 90%, Branson Sonifier 450, Dietzenbach, Germany). 700 µl lysis buffer were added and sample was centrifuged at 13,000 rpm, 15 minutes at 4°C with an Eppendorf table centrifuge. Nuclei containing supernatant were transferred into new tube and protein fractions were stored at -80°C.

2.2.2.3 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

Denaturing SDS polyacrylamide gels were used for protein separation according to their molecular weight. To accomplish this, the NuPAGE gels (4-12%) were used (NuPAGE®Novex® 10% Bis-Tris Gel 1.0 mm x 12 well, Cat. No. NP0302BOX, Invitrogen, Germany) in combination with the NuPAGE® electrophoresis system. According to Quick Reference Card (NuPAGE® Bis-Tris Mini Gels), gel was fixed in chamber and the inner cassette was filled with 1x MOPS running buffer (20x NuPAGE MOPS SDS Running Buffer, Cat. No. NP0001, Invitrogen, Germany). The outer chamber was also filled to two-thirds of its volume. The comb was removed and the lanes were purged with a small pipette. 30 µl of sample was added to 10 µl 4xNuPAGE LDS-Samplebuffer, (Cat. No. NP0007, Invitrogen, Germany), 2 µl 1 M DTT (final 50 mM) and incubated at 95°C for 5 minutes. Samples (20 µl/lane) and markers ((10 µl/lane) SeeBlue Marker prestained, Cat. No. LC 5625, Invitrogen; (5 µl/lane) MagicMark XP anti IgG, Cat.No. LC5602, Invitrogen, Germany) were transferred into the lanes and gel was applied at 30 volts for 50 minutes with NuPAGE®electrophoresis system.

2.2.2.4 SDS-PAGE and western blot

SDS-PAGE induced separated proteins were blotted to a methanol-activated nitrocellulose membrane according to NuPAGE protocol for denaturing electrophoresis (NuPAGE® Technical Guide, Manual part no. IM1001, Invitrogen, Germany) in a Xcell II Blot-Modul (Invitrogen, Germany) with the appropriate reagents. Blotting was conducted in the SDS-PAGE chamber at 30 volt for 50 minutes. Blotted protein membrane was transferred into a new chamber and washed three times for ten minutes in one fold TBS wash buffer (10 fold TBS: 0.5 M Tris-Base, 1.5 M NaCl, pH 7.5) followed by five minutes incubation in Ponceau solution (0.1%) for general protein detection. The Ponceau solution was discarded and residues were washed away with two times one minute TBS wash steps.

Membrane was blocked for one hour at 20-30°C in 1% blocking solution (Blocking solution: 1/10 Casein (Roche, Penzberg, Germany), 1x TBS). Subsequently membrane was incubated in primary antibody solution (1% Blocking solution + Antibody) at 4°C overnight while shaking. The membrane was washed three times in 1x TBST for ten minutes at 20-30°C followed by one minute wash step in 1x TBS. Membrane was incubated in secondary antibody solution (1% Blocking solution + antibody) for 45 minutes at 20-30°C or at 4°C overnight while being agitated. The membrane was washed three times in 1x TBST for ten minutes at 20-30°C followed by two times one minute wash step in 1x TBS to remove Tween.

Antibody name	Dilution (WB)	Host	Company	Cat.no
H4K16ac-ab	1:5000	Rabbit	Millipore	07-329
Anti GFP	1:2000	Mouse	SIGMA	G 6539
Anti Flag	1:1000-10000	Mouse	SIGMA	F 3165
Gal4(DBD+O1679	1:100-1000	Rabbit	Santa Cruz Biotech	Sc-577
Alpha-tubulin	1:10000	Mouse	SIGMA	T 9026
Goat anti rabbit	1:2000	Goat	Millipore	12-348
Goat anti mouse	1:2500	Goat	Millipore	

Table 11: Western blot antibodies with the used concentration to detect specific proteins.

Detection of Antibody-binding was done with the Lumi-LightPLUS western blotting system (Roche, Penzberg, Germany).

2.2.3 General methods for working with CHO cell lines

2.2.3.1 Cryoconservation of CHO cell lines

1x 10⁷ viable cells per sample were centrifuged (500 g, three minutes at 20-30°C) to discard the supernatant. Pellet was resuspended in pre chilled 1 ml Cryomedium (Cdm2 Opt 1.1, 10 mM L-glutamine, 7.5% DMSO), transferred into cryovial and immediately stored at -80°C in Mr.Frosty racks (filled with Isopropanol). Two days later Cryovials were transferred to a liquid nitrogen tank.

2.2.3.2 Thawing of CHO cell lines

15 ml falcon tubes were prepared with 5 ml of an appropriate medium (Cdm2 Opt 1.1, 10 mM L-Glutamine). Cryovials were stored on dry ice until thawing for a maximum of 10 minutes in water bath (37°C). Afterwards CHO cells were taken up into prepared falcon tubes and centrifuged for three minutes at 500 g to remove residual DMSO. Pellet was resuspended in 5 ml medium and propagated in disposable 125 ml vented shake flasks (contain 20 ml appropriate medium) under standard humidified conditions (95% rH, 37°C, and 5% to 8% CO₂) at a constant agitation rate of 120 rpm/min to 150 rpm/min. Viability (> 95%) and cell concentration can be tested with Cedex HiRes Analyzer (Roche, Germany).

2.2.3.3 Cultivation of CHO cell lines

Untransfected CHO cells were split every three to four days and seeded with a concentration of 2-3x 10⁵ cells/ml in Cdm2 Opt 1.1 medium (10 mM L-Glutamine). Transfected cells were seeded with various concentrations of methotrexate (MTX) or methionine sulfoximine (MSX) as the selection agent. After recovery of cell viability (three to four weeks), stably transfected cells were selected in thymidine and protein-free medium containing 20 nM to 1,200 nM Methotrexate (MTX) or 140-160 μM methionine sulfoximine (MSX) as the selection agent. The cells were propagated in disposable 125 ml vented shake flasks under standard humidified conditions (95% rH, 37°C, and 5% to 8% CO₂) at a constant agitation rate of 120 rpm/min to 150 rpm/min. Every three to four days the cells were split into fresh mediums with a cell concentration of 2-3x 10⁵ cells/ml. Density and viability of the cultures were determined using the CASY TT or Cedex HiRes cell counter (Roche Innovates AG, Bielefeld, Germany).

2.2.3.4 Transient and stable transfection

Transfection was achieved with the Amaxa® Cell line Nucleofector® Kit V (Lonza, Cologne, Germany) and the transfection platforms Nucleofector™2b Device and 96-well Shuttle™System (Lonza, Cologne, Germany).

5x 10⁶ CHO cells per transfection were centrifuged (200 g, five minutes) and supernatant was discarded. Pellet was resuspended in 100 µl of supplemented Nucleofector Solution V and 1.2 pmol of sterile plasmid was added by pipetting up and down. Plasmid was linearized for stable transfection, for transient transfection the circular form was maintained. Suspension was mounted into bubble-free cuvette and program U-24 was activated for CHO cell line transfection. After pulse 500 µl of prewarmed media (10 mM L-Glutamine, 0 nM MTX, Mafog In-house platform) was mounted into cuvette and whole suspension was transferred in 8 ml prewarmed media. Cells were transferred into incubator and cells were ready for first examinations two days after.

2.2.3.5 Cell count with Cellavista

60 µl of CHO cell suspension was mixed with 60 µl Trypan blue (0.1 µm) in 96 round bottom well and incubated at 20-30°C for five minutes. Afterwards treated CHO cell suspension was 1:50 diluted with Mafog platform medium and 200 µl were transferred into 96 flat bottom wells (Greiner, Germany). Slow centrifugation (500 g, five minutes) resulted in fast sedimentation of cells.

For a precise cell count pictures from the middle of the well were taken, to avoid display error of pictures approaching the edge of each well. The cellavista cell imager (SynenTec, Germany) was used for the purpose of cell count with the following parameters. Taken pictures for cell count calculation are displayed in green. Leaving out pictures which scraping edge of well will minimize reading error. Each sample was done in replicate and the entire ten pictures were used to calculate cell count.

2.2.3.6 Cell count with Cedex HiRes analyzer

For cell count of a sample number up to 40, the Cedex HiRes Analyzer (Roche, Germany) was used. Following this, the CHO cells were diluted at a ratio of 1:5 in HiRes medium and 300 µl were transferred to tubes. Cell calculation was done according to manufacturer's recommendations. Samples were analyzed in duplicates to verify the cell number.

Doubling time (Korzynska & Zychowicz, 2008) of cell culture was calculated as follows:

$$T2 = \frac{\Delta t \ln 2}{\ln(N_1 - N_0)}$$

Δt is the time period between split and measurement (~ three days); N is the seeded viable cell concentration (N_0) and the measured viable cell concentration (N_1).

2.2.3.7 CHO cell treatment with 5-aza-2deoxycytidine (DAC)

Cell cultures were seeded into 96 well plates and DAC dissolved in phosphate buffered saline (PBS) were added up to final concentration of 1 μ M. Cells were examined on day three after DAC addition, comparing reporter gene expression with non-treated cells.

2.2.3.8 Relative quantification of expression with SEAP-assay

The secreted embryonic alkaline phosphatase (SEAP) catalyzes para-nitrophenylposphate (nNPP) to para-nitrophenole (pNP) which results in color changing of the medium from blue to yellow, measured at 405 nm. From this it was determined that, as more SEAP is expressed, the faster is the color change.

Transient transfected cell suspensions were seeded in 96 well plates and incubated for five days. SEAP concentration was examined in a Tecan-Reader Infinite M200 Pro (Tecan Deutschland GmbH, Crailsheim, Germany) by color change of a chemical reaction.

CHO-K1 M suspension was transfected with circular plasmid DNA for transient expression of SEAP, using the Nucleofector device in combination with the Nucleofector Kit V (Lonza Cologne GmbH, Cologne, Germany) in the Amaxa Shuttle 96 well plate (Figure 18) according to the manufacturer's protocols. Twelve transient transfected cell suspensions per plasmid were seeded in 96 well plates and incubated for five days. On day two, 5'-Aza-2'-Deoxycytidine (DAC) was added in four replicates per plasmid to a final concentration of 1 μ M to demethylate DNA. SEAP concentration was examined using the Tecan-Reader Infinite M200 Pro (Tecan Deutschland GmbH, Crailsheim, Germany).

Code	Line	1	2	3	4	5	6	7	8	9	10	11	12	
-	A													
GGG	B													
GCG	C													Transfection with gene of interest (plasmid)
CGC	D													Control-transf. with zero volt
GCC	E													Control-transf. with H2O probe
CCG	F													Control-transf. with GFP (Lonza)
CCC	G													
Contr.	H									DAC				

Figure 18: Amaxa Shuttle 96 well plate for transient transfection of CHO-K1 cell suspension.

Relative concentration of SEAP was examined by the metabolic rate of pNPP (para-nitrophenylphosphate) to pNP (para-Nitrophenole) according to the following protocol:

Diluent	
Charge	Amount for 1 assay
1 mM MgCl₂*6H₂O (Sigma, Cat.No. HM2670-500g MW=203.31 g/mol)	100 μ l 1 M MgCl ₂ *6H ₂ O
10 mM L-Homoarginine-HCl (Sigma, Cat.No H1007-5G, MW 224.69 g/mol)	224.69 mg
1 M Diethanolamine (Sigma, Cat.No. D-8885, min. 98%, MW 105.1 g/mol, conc. 1.1 g/ml)	9.75 ml 98%
	ad 90 ml Bidest H2O à pH to 9.8 with 25-37% HCl à ad 100 ml Bidest H2O
Substrate solution	
Charge	Amount for 1 assay
pNPP (4-Nitrophenylphosphate, Roche, Cat.No. 107905, Lot.No. 10030536)	74.2 mg in 10 ml bidest H2O

Table 12: Solutions for SEAP Assay.

150 μ l of centrifuged cell culture supernatant was diluted 1:3 in multiple steps (Figure 19). 50 μ l dilutions were added in new 96 well plates and combined with a 50 μ l substrate solution. The reaction was measured every five minutes at 405 nm wave length with the Tecan-Infinite M200 Pro (Tecan Deutschland GmbH, Crailsheim, Germany) until the optical density came up to a value of 2.

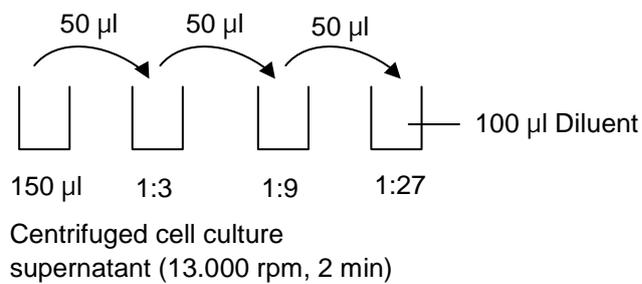


Figure 19: Procedure of dilution of cell culture supernatant.

The average of eight replicates per plasmid without DAC treatment and the average of four replicates per plasmid (Figure 18) with DAC treatment were normalized to control plasmid (p5532).

2.2.3.9 Relative quantification of reporter gene expression with flow cytometry

Stable transfected cell suspensions expressing eGFP were cultivated over a period of one to three months. Intensity of eGFP expression per cell was measured with the Fluorescence Activated Cell Sort system (FACS) BD FACS Canto II or BD FACS Calibur (BD, Heidelberg, Germany). Activation, inactivation and washing routine were performed according to the manufacturer's recommendations.

The flow cytometry allows simultaneous multi-parameter analysis of single cells by separating cells in a tiny stream of fluid, where cells pass the laser light one cell at a time. 10,000 events per sample were measured. The cell size was detected as forward scatter (FSC), the granularity of the cell was detected as side scatter (SSC). Specific values of FSC and SSC combined, indicated the normal living cell population (Figure 20). Gates for living CHO cell suspensions were set in untransfected control CHO-K1 cells and transferred to all measured samples.

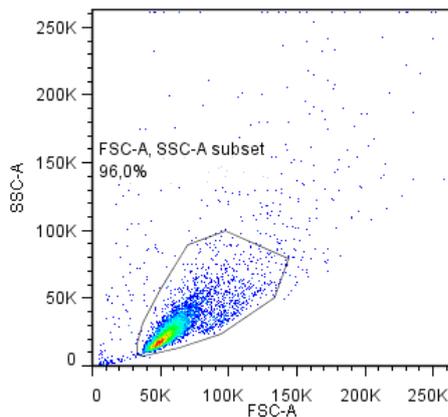


Figure 20: Living gate of CHO-K1 cells. Granularity (SSC) and cell size (FSC) were plotted in diagram. Living Gate of untransfected control CHO-K1 cells was transferred to all examined samples of the same FACS assay.

Relative eGFP expression was compared with the autofluorescence levels of untransfected control CHO-K1 cells. Fluorescence of eGFP of the living gate was measured at excitation wavelength of 488 nm (Alexa Fluor 488 channel). The fluorescent light is filtered so each sensor will detect fluorescence only at a specified wavelength. These sensors are called photomultiplier tubes (PMTs). The PMT detect light emitted from eGFP at approximately 516 nm wavelength and converted this information into an electronic signal. This signal is plotted in a histogram of living gate events (Figure 21).

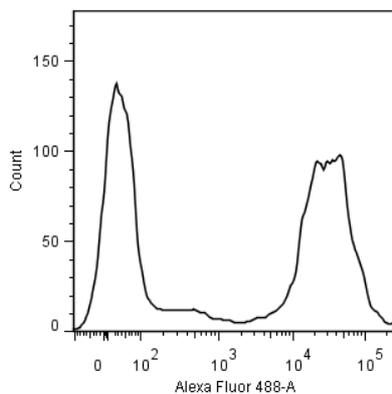


Figure 21: Living gate histogram of eGFP expressing CHO cell population. Values measured at excitation wavelength 488 nm with the Alexa Fluor 488 channel. Two peaks are displayed representing the two main subpopulation of non eGFP expressers (left peak), which is comparable to autofluorescence of untransfected control CHO cells, and high eGFP expressers (right peak).

Data collection was performed with BD FACS Diva Software v6.12 or Cell Quest Pro Software (BD, Heidelberg, Germany). Primary data analysis was performed with FlowJo 7.6.5 EN software (TreeStar, Olten, Switzerland).

2.2.3.10 Sample preparation for antibody analysis

Cell concentration was calculated and 2 ml per sample were centrifuged (500 g, five minutes at 20-30°C). Supernatant was transferred to new 96 deep well plates and stored at -20°C until use. Frozen supernatant was thawed overnight at 4°C, 6x inverted and centrifuged (4,000 rpm, 30 minutes at 20-30°C). 310 µl were filtered with a multiscreen Millipore plate atop a barcoded 96 roundwell plate by centrifugation (1,200 rpm, three minutes at 20-30°C).

2.2.3.11 Quantification of antibody production with HPLC

A chromatographic method was used to quantify the amount of antibody present in a sample. A PorosA column was used that binds the Fc-region of the antibody. The antibody binds to the column and is subsequently eluted by low pH conditions. Protein concentration was established by determining the optical density (OD) at 280 nm, with a reference wavelength of 320 nm, using the molar extinction coefficient calculated on the basis of the amino acid sequence. Quantification of antibody amount with the chromatographic method was performed in-house by Uta Werner.

2.2.3.12 Quantification of antibody production with ELISA

The ELISA (Enzyme Linked Immunosorbent Assay) technique is based on the antibody sandwich principle. A capture antibody specific to the analyte of interest for instance the Fc part of IgG is bound to a microtiter plate (Maxisorp, Inhouse, Roche) to create the solid phase. Following the blocking and washing steps, samples, standards (dilution series of reference Ab), and controls are then incubated with the solid phase antibody, which captures the analyte. After washing away unbound analyte, a conjugated detection antibody (e.g. POD conjugated) is added. This detection antibody binds to a different epitope of the molecule being measured, completing the sandwich. The BM Chemiluminescence ELISA Substrate POD (Roche, Penzberg, Germany) provides a substrate solution of peroxidase-based (POD, HRP) secondary detection system.

The rate of signal generation in an immunoassay is directly proportional to the amount of marker enzyme bound to the solid phase. Antibody concentration was calculated by the slope of standard dilution. ELISA assays were performed by Manuela Bernhard.

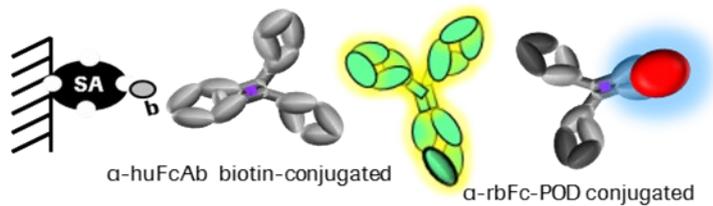


Figure 22: ELISA principle with POD conjugated detection antibody.

2.2.3.13 Generation of recombinant CHO cell lines comprising hCMV-MIE promoter variants or the histone acetyltransferase MOF

Recombinant cell lines expressing Secreted Alkaline Phosphatase (SEAP: Figure 23), enhanced Green Fluorescence Protein (eGFP: Figure 24 & Figure 25) or human antibody constructs of class IgG (Figure 26) were generated by transient or stable transfection of CHO-K1 suspension growing cells. Used vectors expressing reporter genes under control of hCMV-MIE promoter, comprising C to G point mutations, or under control of unmutated hCMV-MIE promoter with an internal ribosomal entry sites (IRES EV71), coupled expression of the histone acetyltransferase MOF (Figure 25). Light and heavy chain expression cassettes of human immunoglobulin were both under the control of a human CMV major immediate-early promoter and enhancer comprising C to G point mutations (Figure 27: Seq ID NO: 01).

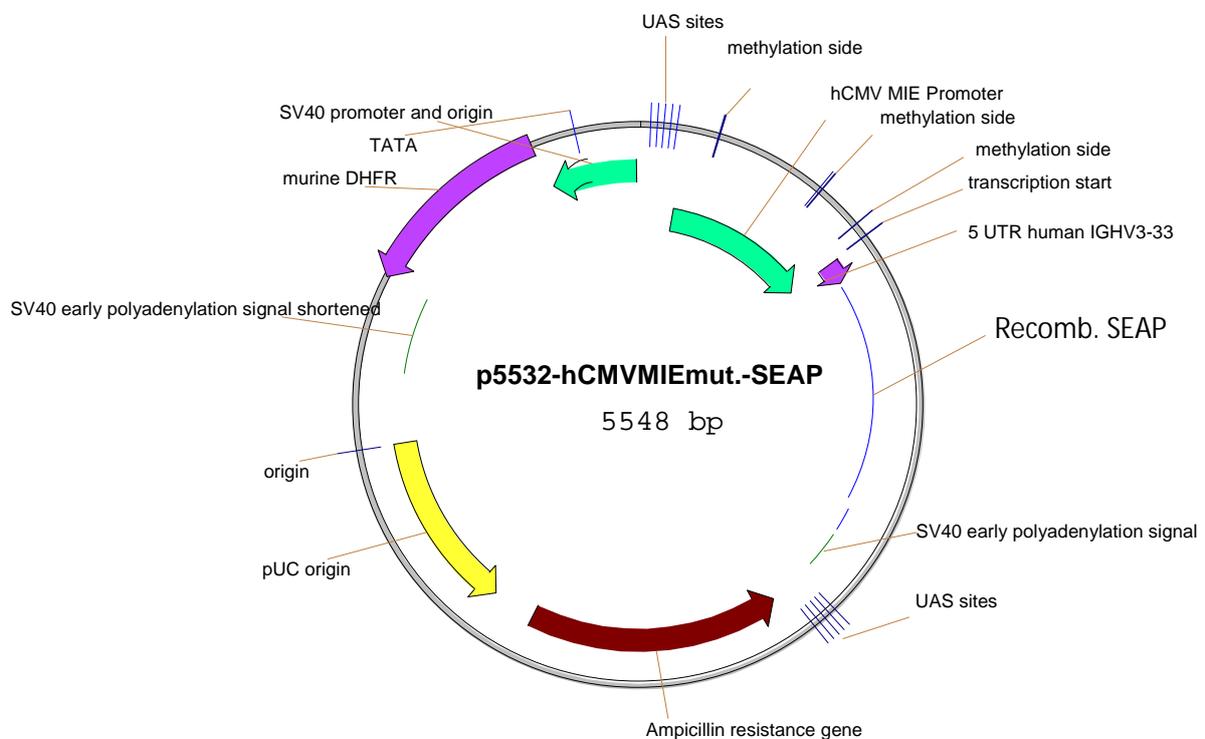


Figure 23: Secreted alkaline phosphatase (SEAP) expressing plasmid. SEAP expression is under control of human CMV major immediate-early promoter/enhancer. The C to G mutations C-508, C-179 and C-41 of CpG dinucleotides (SEQ ID NO 1) within the hCMV-MIE fragment were inserted in various combinations to increase long term stability (Table 13). The C to G mutations are identified by their distance to transcription start site.

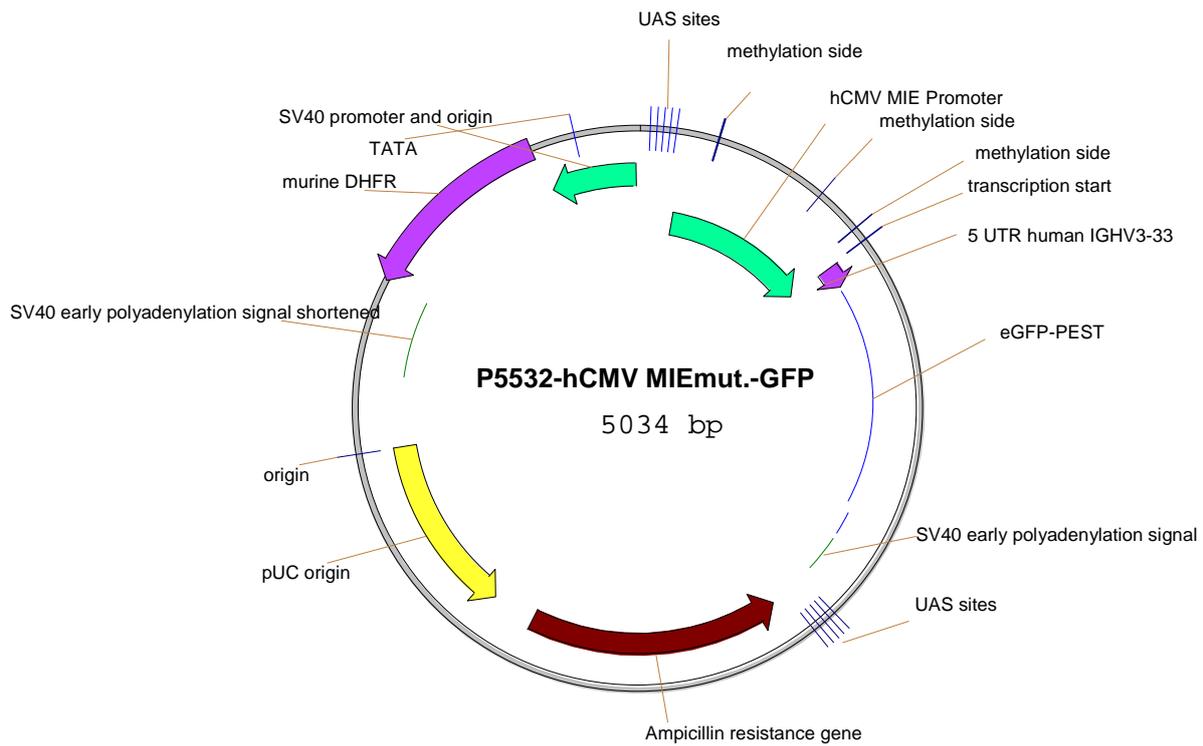


Figure 24: Enhanced Green Fluorescent Protein (eGFP) expressing plasmid. GFP expression is under control of human CMV major immediate-early promoter/enhancer (hCMV-MIE). The C to G mutations C-508, C-179 and C-41 of CpG dinucleotides (SEQ ID NO 1) within the hCMV-MIE fragment were inserted in various combinations to increase stability (Table 13). The C to G mutations are identified by their distance to transcription start site.

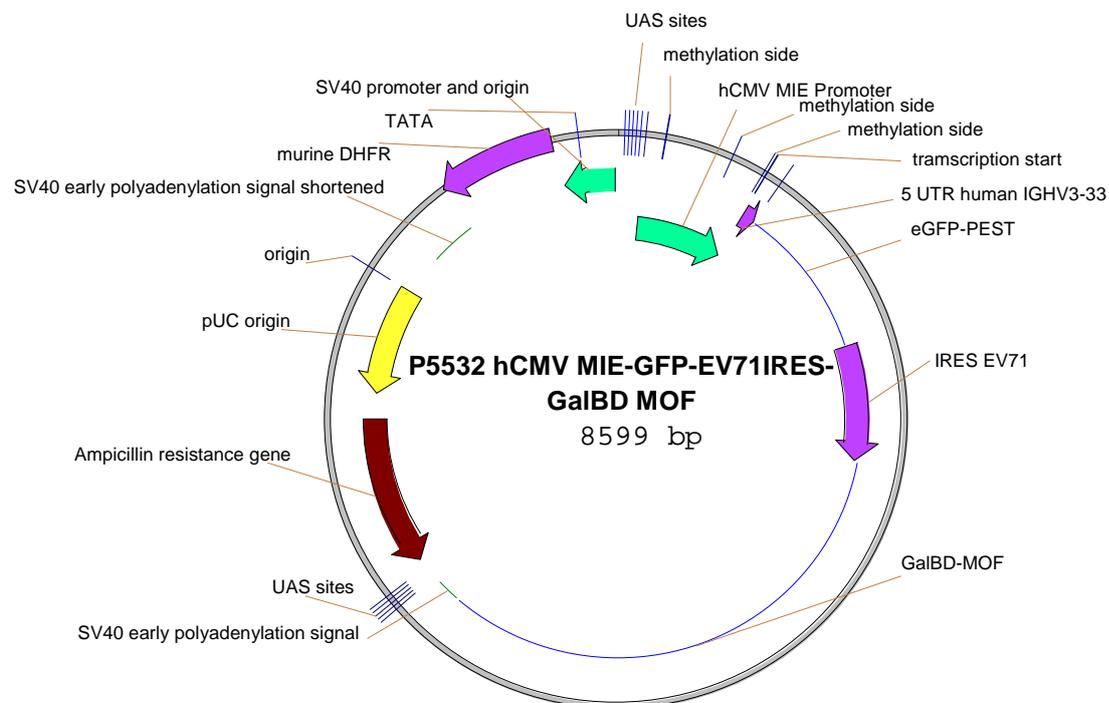


Figure 25: Enhanced Green Fluorescent Protein (eGFP) expressing plasmid. GFP expression is under control of human CMV major immediate-early promoter/enhancer (hCMV-MIE). Histone acetyltransferases MOF is additionally expressed to increase stability (Table 13).

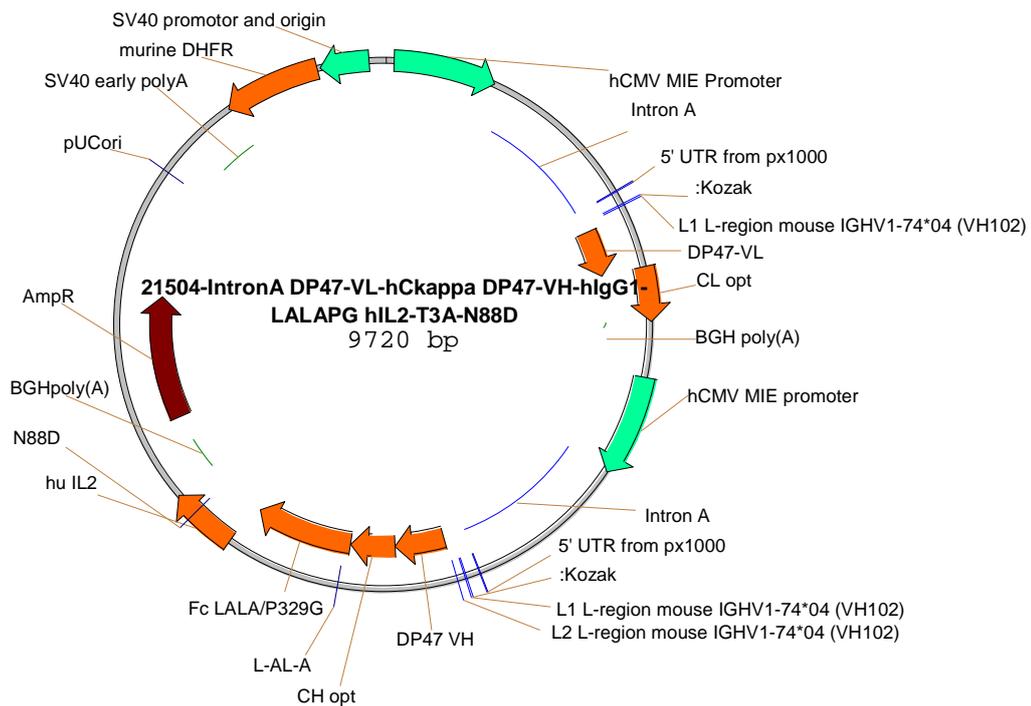


Figure 26: Human antibody of class IgG expressing plasmid. Light and heavy chain expression cassette of human immunoglobulin were both under the control of a human CMV major immediate-early promoter and enhancer (Figure 27: Seq ID NO: 01) (hCMV-MIE). The C to G mutations C-508, C-179 and C-41 of CpG dinucleotides (SEQ ID NO 1) within the hCMV-MIE fragment were inserted in various combinations to increase the stability (Table 13). The C to G mutations are identified by their distance to transcription start site.

The three C to G point mutations C-508, C-179 and C-41 of CpG dinucleotides (Figure 27: SEQ ID NO: 1) are identified by their distance to transcription start site (TSS) and inserted in various combinations (Table 13) with QuikChange Multi-Site-Directed Mutagenesis Kit (Agilent Technologies, Waldbronn, Germany). The vector also comprised a nucleic acid sequence encoding murine dihydrofolate reductase (DHFR) (Figure 23 - Figure 26). Transfection of cells was performed by Amaxa nucleofection system (Lonza Cologne GmbH, Cologne, Germany).

```
GTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGA
GTTCCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCAACGACCCCGCCATTGACGTCA
ATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGG
TAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTA
AATGGCCCGCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTTCTACTTGGCAGTACATCTACGTATT
AGTCATCGCTATTAGCATGGTGTATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGGGTGGTGGTACCG
GGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTGTGGTGGCACCAAAATCAACGGGACTTTCCA
AAATGTCGTAACAACCTCCGCCCATTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAG
CAGAGCTCCGTTTGTAGTGAACGTCAGATC
```

Figure 27: SEQ ID NO 1. Human CMV major immediate-early promoter/enhancer fragment: Grey marked cytosines are potentially C to G mutated. CpGs dinucleotides are identified by their distance upstream of transcription start site. The names of the three potentially mutated C's are C-508, C-179 and C-41. The red highlighted tyrosine represents the transcription start site.

C to G point mutations in human CMV-major immediate-early promoter/enhancer fragment of SEAP expressing plasmids				
Plasmid NO.:	Code of point mutations	Name of point mutations	Position upstream of TSS	
16100	GGG	C-508, C-179, C-41	508, 179, 41	
16101	GCG	C-508, C-41	508, 41	
16102	CGC	C-179	179	
16103	GCC	C-508	508	
16104	CCG	C-41	41	
p5532	CCC	No mutation		
C to G point mutations in human CMV-major immediate-early promoter/enhancer fragment of eGFP expressing plasmids and eGFP expressing plasmids comprising Histone acetyltransferases MOF				
Plasmid NO.:	Code of point mutations	Name of point mutations	Position upstream of TSS	Feature
16105	GGG	C-508, C-179, C-41	508, 179, 41	Gal4-BD binding sequence UAS upstream of hCMV-MIE
16106	GCG	C-508, C-41	508, 41	
16107	CGC	C-179	179	
16108	GCC	C-508	508	
16109	CCG	C-41	41	
16111	CCC	No mutation		
13134	CCC	GalBD control		
13135	CCC	GalBD-MOF		
16110	CCC	No mutation		
C to G point mutations in human CMV-major immediate-early promoter/enhancer fragment of human antibody of class IgG expressing plasmids				
Plasmid NO.:	Code of point mutations	Name of point mutations	Position upstream of TSS	
16134	CGG	C-179, C-41	179, 41	
16135	CGC	C-179	179	
16136	CCG	C-41	41	
21504	CCC	No mutation		

Table 13: Combinations of C to G point mutations of CpG dinucleotides within the human CMV major immediate-early promoter/enhancer fragment in the appropriate reporter vectors. GFP expression plasmid with IRES EV71 coupled expression of Mof and control highlighted in light grey. Target oriented Mof towards UAS_{Gal} sites.

CHO-K1 suspension was transfected with circular plasmid DNA for transient expression of SEAP or with linearized plasmid DNA for stable expression of antibody or eGFP, using the Nucleofector device in combination with the Nucleofector Kit V (Lonza Cologne GmbH, Cologne, Germany) according to the manufacturer's protocols. Transient transfected cell suspensions were seeded in 96 well plates and incubated for five days. SEAP concentration was examined in a Tecan-Reader Infinite M200 Pro (Tecan Deutschland GmbH, Crailsheim, Germany) by color change of a chemical reaction.

Stable transfected cell suspensions were seeded in 384 or 6-well plates containing thymidine-free medium with 250–1,600 nM methotrexate (MTX) as the selection agent. After three to four weeks, eGFP or antibody-expressing cell pools were examined for long term stability over a period of one to three months. Intensity of eGFP expression was examined by flow cytometry. Antibodies expressing single cell clones were seeded in 384 and 96 well plates. After three weeks, antibody-expressing cell lines were identified by measuring antibody titers in the culture medium by ELISA. Growing wells were randomly picked and in the interests of long term stability assay cell clones were expanded in higher volumes (3 ml per well in six well plates) and antibody concentration was determined by protein A HPLC and ELISA at the end of each passage.

The cells were propagated in disposable 125 ml vented shake flasks or six well plates under standard humidified conditions (95% rH, 37°C, and 5% to 8% CO₂) at a constant agitation rate of 120 rpm/min to 150 rpm/min. Every three to four days the cells were split into fresh mediums. Density and viability of the cultures were determined using the Cedex HiRes cell counter (Roche Innovates AG, Bielefeld, Germany) or cellavista CV3.1 (Synentech Bio Services GmbH, Munich, Germany).

2.2.3.14 Long-term cultivation and production of CHO cell lines comprising hCMV-MIE promoter variants

Various CHO cell pools and clones containing CpG point mutations within the hCMV-MIE promoter or GalBD-MOF (Table 13) were investigated for long-term productivity. The cells were tested for phenotypic (i.e. production), stability for two to three months after transfection in the presence of selection agent MTX. The cells were continuously cultivated in vented 125 ml shake flasks containing 20-40 ml medium or six well plates containing 2-4 ml medium with selection agent and diluted twice a week with fresh medium. Seeding density was 2 to 3 *10⁵ cells/ml. Prior to passage, viable cell density and viability were determined.

Antibody concentration of the supernatant (antibody titer) was determined by protein A HPLC and ELISA at the end of each passage. From these data, the cell specific productivity (qP) for each passage was calculated using the following formula:

$$qP = \frac{P_2 - P_1}{(D_2 - D_1) / 2 * \Delta t}$$

qP [pg/cell/d]: cell specific productivity,

P₁ [µg/ml]: antibody titer at the beginning of the passage,

P₂ [µg/ml]: antibody titer at the end of the passage,

D₁ [cells/ml]: viable cell density at the beginning of the passage,

D₂ [cells/ml]: viable cell density at the end of the passage,

Δt [d]: duration of the passage.

The qP values were plotted against the age of culture at the end of the respective passage in generations. A linear trend line was calculated over all qP data points and the alteration of the qP (in percent) over the period was calculated in house, according to the following equation:

$$\Delta qP = \frac{m * a}{qP_0 * 100}$$

ΔqP : percentage alteration of qP,
 m [pg/cell/d/generation]: slope of linear trend line,
 a [no. of generations]: age of culture,
 qP_0 : y-axis intercept of linear trend line.

In regard to the lower number of data points obtained in the stability assay of hCMV-MIE promoter variants, for each sample the average of the last three qP values was divided by the average of the first two qP values and displayed in percent to obtain ΔqP . ΔqP being the percentual alteration of qP, qP_{start} being the specific productivity at the beginning of the cultivation period and qP_{end} being the specific productivity at the end of the cultivation period.

$$\Delta qP = \left(\frac{qP_{end}}{qP_{start}} - 1 \right) * 100$$

2.2.3.15 Treatment of recombinant CHO cell lines, used for identification of epigenetic marker

32 CHO cell lines of two projects were used to identify epigenetic markers, predicting target gene expression two months in advance. Both projects were already performed in-house and data including antibody concentration, C-179 methylation status of hCMV-MIE promoter and copy number had been collected over a period of 60 generations (~ two months).

Project H comprises 20 CHO-K1 M cell lines expressing a human antibody of class IgG under control of hCMV-MIE promoter. The vector further comprises a nucleic acid sequence encoding glutamine synthetase, making cells susceptible to methionine sulfoximine (MSX) selection.

Project T comprises 12 CHO-K1 SV cell lines expressing a human antibody of class IgG under control of hCMV-MIE promoter. Those cell lines are prone to methotrexate (MTX) selection because of the transgene of the murine dihydrofolate reductase. Frozen start cultures were thawed and cultivated in appropriate mediums comprising 250 nM MTX or 140 μ M MSX. After two weeks (time point PSB) of cultivation cells were harvested and chromatin immunoprecipitation with following antibodies was carried out (Table 14).

List for project T

Target	Name	Company	ChIP grade	Host	Cat.No.	Lot
H3K4me3	ChIPAb+ Trimethyl-Histone H3 (Lys4)	millipore	v	rabbit monoclonal	17-614	ng1848343
H3K27me3	ChIPAb+ Trimethyl-Histone H3 (Lys27)	millipore	v	rabbit polyclonal	17-622	dam1850974
H3K9me3	Anti-Histone H3 (tri methyl K9) antibody	abcam	v	rabbit polyclonal	ab8898	gr58934-1
H3	Anti-Histone H3 antibody	abcam	v	rabbit polyclonal	ab1791	GR103799
H3ac	Anti-acetyl-Histone H3	millipore	v	rabbit polyclonal	06-599	2068182

List for project H

Target	Name	Company	ChIP grade	Host	Cat.No.	Lot
H3K4me3	Histone H3K4me3 antibody (pAb)	activemotif	v	rabbit Polyclonal	39915	14013006
H3ac	Anti-acetyl-Histone H3 Antibody	millipore	v	rabbit polyclonal	06-599	2153150
H3K9me3	ChIPAb+ Trimethyl-Histone H3 (Lys9)	millipore	v	polyclonal	17-625	2190636
H3K27me3	ChIPAb+ Trimethyl-Histone H3 (Lys27)	millipore	v	rabbit Polyclonal	17-622	2034145
H3	Anti-Histone H3 antibody	abcam	v	Rabbit polyclonal	ab1791	GR135321

Table 14: ChIP grade antibodies specific for Histone or Histone modifications. Antibodies were used for immunoprecipitation of specific chromatin fragments.

Quantitative PCR was performed with the following primers:

SEQ ID NO:	Primer NO:	Target	Sequence
08	396	hCMV-MIE forward 1	TACATCAATGGGCGTGGATA
09	397	hCMV-MIE reverse 1	AAGTCCCCTTGATTTTGGTG
10	386	Gusb forward 1	CAGGGTGGGATGCTCTTC
11	387	Gusb reverse 1	GCCGGTTTTCCGAGAAGT
12	431	Eif3i forward 1	GTTCCCGGCACTGACACT
13	432	Eif3i reverse 2	ACTTGATCTGCGTGATGGAC
14	484	Foxa2 forward 1	ATCACCCGTA CTGCTGCTCT
15	485	Foxa2 reverse	GAGGTTCTGGGGATCTCTT
16	468	Gata5 forward 1	CACCTACCCCATCCTGTCTG
17	469	Gata5 reverse 1	GAGGAGGTGAAGCAAAGTCT
18	388	Rho forward 1	AGCCTCGGTCTCTATTGACG
19	389	Rho reverse 1	CGTTGGAGAAGGGCACATAA
20	474	UNC13c forward 1	GGTGCTTTACGAAACTGA
21	475	UNC13c reverse 1	GCTTCTTATGCCCCAGGTTT

Table 15: Primer sequences for real time quantitative PCR. Gusb and Eif3i sequences are close to stable active marks.

2.2.3.16 Generation of eGFP expressing CHO cell pools and preparation of CHO-K1 M cells for the identification of integration sites and chromatin states

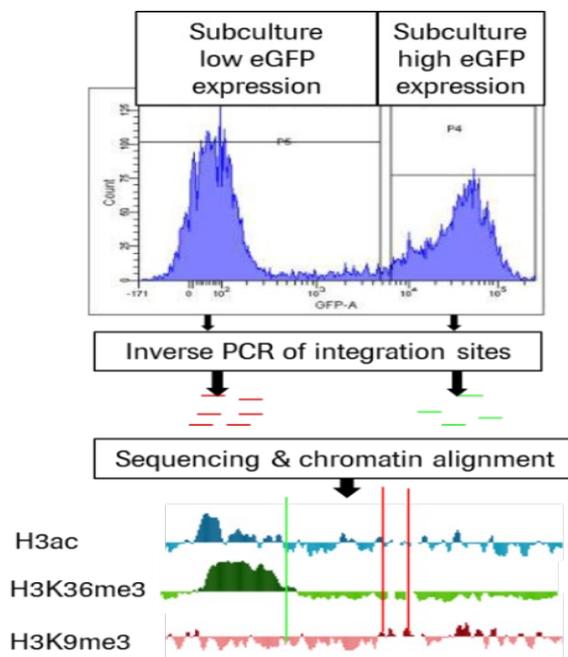


Figure 28: Exemplified identification of stable integration sites. Combination of chromatin state and integration sites to distinguish active and inactive areas of CHO-K1 cell pools.

Histone 3 acetylation, H3K4me3, and H3K36me3 as active marker and H3K9me3 as inactive marker of wild type CHO-K1-M chromatin were examined by ChIP sequencing. CHO-K1 cell pools were randomly transfected with eGFP expressing reporter plasmids, MTX selected and cultivated. After one month of cultivation, three independent recombinant cell pools were sorted into subcultures, distinguished by high and low levels of eGFP. Inverse PCR of subcultures was conducted to locate transgene integration sites. Next generation sequencing of immunoprecipitated chromatin of CHO-K1 wild type and the integration sites of sorted eGFP pools was performed.

ChIP sequencing protocol was carried out as described previously in ChIP protocol with the following modifications:

50 μ l resuspended dynabeads (Dynabeads[®] Protein A, Life Technologies GmbH, Darmstadt, Germany) per IP were blocked in 1 ml icy PBS (5 mg/ml BSA) for five minutes. Separation of beads from liquid was performed by magnets. Blocking step was repeated three times. Beads were resuspended in 200 μ l PBS/BSA containing 6 μ g appropriate antibodies for saturation and incubated overnight at 4°C while rotating. Beads were washed three times and resuspended in 100 μ l PBS/BSA. 3.5x 10⁷ CHO-K1-M wild-type cells (passage 4) were fixed and sonicated in 3 x 300 μ l Nuclei Lysis Solution. 700 μ l of Nuclei Lysis Solution were added to sonicated cells and cells were microcentrifuged for 15 min 4°C at 14,000 rpm. The supernatant contains the chromatin and can be stored at -80° for up to two months. Sonification efficiency was tested and protein concentration was specified. 200 μ l of chromatin solution (1 μ g/ μ l) and 300 μ l Dilution buffer were added to 50 μ l of antibody loaded dynabeads. Incubation was done overnight at -4°C while rotating. After Wash and elution steps, crosslink was reversed overnight at 65°C. 200 μ l TE buffer were added to each tube for SDS dilution. For RNase digestion, 4 μ l RNase, DNase (Roche, Mannheim, Germany) free solution were added and incubated at 37°C for two hours. Finally 4 μ l EDTA (0.5 M Stock) and 4 μ l Proteinase K (100-200 μ g/ml final concentration, Roche, Mannheim, Germany) were added and proteins were digested for 1.5 hours at 65°C. DNA was recovered with High Pure PCR Product Purification Kit. DNA was stored at -20°C and transferred to GATC for sequencing.

Antibodies

Target	Name	Company	Cat.No.
H3K4me3	Anti-trimethyl Histone H3 (Lys4)	Millipore	CS200580
H3ac	Anti-acetyl Histone H3	Millipore	06-599
H3K9me3	pAb to histone h3 (tri methyl k9)	Abcam	Ab8898

Table 16: ChIP grade antibodies. Fishing of active and inactive chromatin areas by immunoprecipitation of specific histone modifications.

3 Results

3.1 Molecular indicator of production instability

In order to check whether the degree of histone modifications have similar or even enhanced predictive power for the stability of recombinant CHO cell lines, than the DNA methylation (Osterlehner et al., 2011), H3ac and H3K4me3 as activation marks and H3K9me3 as well as H3K27me3 as repression marks were examined at the sequence from -97 to -200 bp upstream of transcription start site of hCMV-MIE to enable the direct comparison with the DNA methylation degree at position -179 (Figure 29).

```
-603 ATGTTGACAT TGATTATTGA CTAGTTATTA ATAGTAATCA ATTACGGGGT
-553 CATTAGTTCA TAGCCCATAT ATGGAGTTCC GCGTTACATA ACTTACGGTA
-503 AATGGCCCGC CTGGCTGACC GCCCAACGAC CCCC GCCCAT TGACGTCAAT
-453 AATGACGTAT GTTCCCATAG TAACGCCAAT AGGGACTTTC CATTGACGTC
-403 AATGGGTGGA GTATTTACGG TAAACTGCC ACTTGGCAGT ACATCAAGTG
-353 TATCATATGC CAAGTACGCC CCCTATTGAC GTCAATGACG GTAAATGGCC
-303 CGCCTGGCAT TATGCCCAGT ACATGACCTT ATGGGACTTT CCTACTTGGC
-253 AGTACATCTA CGTATTAGTC ATCGCTATTA GCATGGTGAT GCGGTTTTGG
-203 CAGTACATCA ATGGGCGTGG ATAGCGTTT GACTCACGGG GATTCCAAG
-153 TCTCCACCCC ATTGACGTCA ATGGGAGTTT GTTTTGGCAC CAAAATCAAC
-103 GGGACTTTC AAAATGTCGT AACAACTCCG CCCATTGAC GCAAATGGGC
-53 GGTAGCGGTG TACGGTGGGA GGTCTATATA AGCAGAGCTC CGTTTAGTGA
-3 ACG
```

Figure 29: hCMV-MIE sequence. Underlined sequence spanning a region of -97 to -200 bp upstream of transcription start site, which is amplified with the primer pair 396/397 to quantify the degree of histone modifications after chromatin immunoprecipitation. Highlighted is the CpG -179 bp upstream of transcription start site, which had been successfully used as a prediction marker for stability (Osterlehner et al., 2011).

3.1.1 Cell lines and data of literature

Two Roche cell line development projects (project T and H) were chosen for the identification of potential prediction markers. For this purpose frozen cell lines from the start point (PSB) of cultivation were thawed and the relative amounts of specified histone modifications close to the hCMV-MIE promoter were determined. Project T (Table 17) and project H (Table 18) were independently executed and each cell line was cultivated for at least 60 generations under the

conditions with (+) and without (-) selection reagent (250 nM methotrexate). Project H ChIP analysis was performed by Laura Woltering under my supervision. The data of the hCMV-MIE promoter methylation degree (meC-179), the copy number per cell from light chain IgG transgene (PSB) as well as the alteration over 60 generations of specific productivity per cell per day (ΔqP) were kindly provided by the PLPM department. Each cell line was examined once as described previously (Osterlehner et al., 2011).

Project T: 12 CHO antibody producing cell lines Generated by transfection of CHO-K1-M with Plasmid p5057				
Sample No.	meC-179 [%]	ΔqP (60 generations) in %		LC-transgene Copies/Cell
	PSB	+MTX	-MTX	PSB
T-1	0.44	-35.40	-34.20	4.4
T-2	2.20	-30.60	-72.00	11.5
T-3	3.54	-81.60	-96.60	9.4
T-4	0.92	-23.40	-47.40	4.6
T-5	4.41	-15.60	-22.20	5.3
T-6	0.47	-14.40	-20.40	2.6
T-7	0.22	-16.20	-31.80	1.6
T-8	52.40	-64.80	-79.20	64.3
T-9	52.84	-27.60	-49.20	67.5
T-10	49.61	3.60	1.20	70.5
T-11	2.02	-10.20	-21.60	4.6
T-12	61.34	-32.40	-22.80	11.5

Table 17: In house data of 12 cell lines from project T. meC-179 is the mean of percentage methylation of cytosine at position -179 upstream of TSS of hCMV-MIE at begin of cultivation phase (PSB). ΔqP is the percentage alteration of specific productivity (qP) over a time period of 60 cell generations. The average copy number of the light chain (LC) of the IgG transgene was measured at begin of cultivation (PSB).

Project H: 20 CHO antibody producing cell lines Generated by transfection of CHO-K1-M with Plasmid p7672				
Sample No.	meC-179 (%)	Δ qP (60 generations)(%)		LC-transgene Copies/Cell
	PSB	+MSX	-MSX	PSB
H-1	15.7	-46.80	-80.20	168.8
H-2	9.9	-71.00	-70.30	2.0
H-3	16.5	-73.10	-89.60	3.2
H-4	10.5	-69.30	-69.50	2.9
H-5	11.0	-40.40	-48.80	1.8
H-6	7.1	-79.50	-71.70	37.6
H-7	4.8	-66.20	-81.50	53.0
H-8	4.4	-67.40	-75.70	30.6
H-9	23.4	-87.00	-77.00	95.1
H-10	26.2	-16.40	-73.30	48.8
H-11	0.2	-43.10	-96.10	3.3
H-12	0.1	-42.00	-33.00	1.5
H-13	1.3	8.00	-18.50	1.1
H-14	0.0	6.80	5.40	1.3
H-15	0.1	-16.50	-40.40	3.5
H-16	0.4	16.20	-32.80	3.0
H-17	1.0	-13.20	-31.30	4.2
H-18	46.3	-67.50	319.40	8.8
H-19	16.7	-11.20	2.70	10.1
H-20	21.8	-40.30	-80.60	4.8

Table 18: In house data of 20 project H cell lines. meC-179 is the mean of percentage methylation of cytosine at position -179 upstream of TSS of hCMV-MIE at begin of cultivation phase (PSB). Δ qP is the percentage alteration of specific productivity (qP) over a time period of 60 cell generations. The average copy number of the light chain (LC) of the IgG transgene was measured at begin of cultivation (PSB).

3.1.2 Validation of ChIP antibodies for both projects with one sample each

In order to ensure the specific immunoprecipitation of histone-DNA complexes, only commercial ChIP grade antibodies were used and tested in one randomly chosen cell line for each project. For the chromatin immunoprecipitation, viable CHO cells which bear a recombinant gene driven by the hCMV-MIE promoter were harvested and protein-DNA interactions were fixed in 3.7% formaldehyde. Cells were lysed and chromatin was sonicated up to a fragment size of approximately 300 bp. Cross-linked histone-DNA fragments were purified with unspecific control beads (Mock) and with specific ChIP grade antibodies according to the appropriate histone modification (Table 14). Subsequently the accumulated histone-DNA complexes were degraded with proteinase K and DNA fragments were eluted.

To verify the binding of the used antibodies in CHO cells, the accumulation of potential reference sequences was compared between the antibody-treated samples and the mock control. The active genes *Gusb* (Abruzzo et al. 2005) and *Eif3i* (Bahr et al. 2009) were chosen as potential references for the expression marks H3ac and H3K4me3. As potential references for H3K9me3 and H3K27me3 the assumed inactive genes *Foxa2* and *Gata5* were investigated. Both genes were not expressed in ovary, according to the Tissue-specific Gene Expression and Regulation (TiGER) database (Liu et al. 2008).

DNA fragments of antibody-purified samples, mock and untreated input sample were quantified by qPCR (2.2.1.11). Primer pairs of potential reference regions (Table 15) were added to each sample and qPCR was performed in triplicates. The obtained Cq values were displayed as percentage of the input sample (Figure 30) In both projects, the activation marks H3ac and H3K4me3 accumulated at the active regions *Eif3i* and *Gusb*. In contrast, the silenced region *Foxa2* had higher concentration of H3K9me3 whereas H3K27me3 accumulated at the *Gata5* region. The differences of the control regions to each other and to the mock control in project H were not as distinctive as in project T. Nevertheless, the different histone modifications matched to the corresponding control regions in both projects.

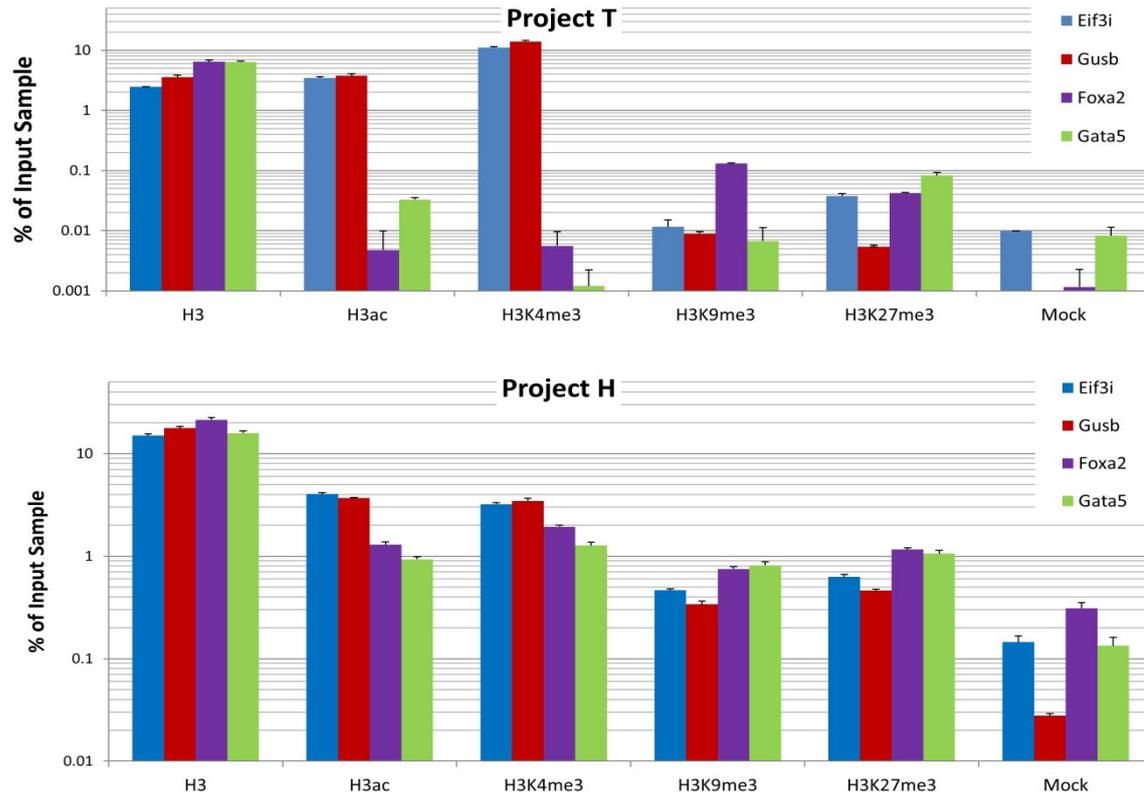


Figure 30: Antibody validation within project T and H in one sample each. Reference regions Eif3i, Gusb, Foxa2 and Gata5 are displayed as percentage of input sample and are plotted against the specific histone purified with the appropriate antibodies. Antibodies for total H3, acetylated H3 (panH3ac), the H3K4me3, H3K9me3 and the H3K27me3 were tested. Input sample precipitation without antibody is termed mock. For each project, the following could be observed: Eif3i and Gusb as proposed active regions, comprising higher levels of H3ac and H3K4me3 compared to mock and the Foxa2/Gata5 regions. In contrast the proposed silenced region Foxa2 has higher H3K9me3 values than mock and the active regions. The same is true for Gata5 and the H3K27me3. As expected the total H3 level was always high compared to mock. Therefore the specificity of each antibody was confirmed. Error bars represent the standard deviation of mean of triplicates.

Considering the lower values obtained under mock condition as well as the higher concentration of active histone marks in active regions compared to the inactive regions and vice versa, I assume a proper binding of the ChIP grade antibodies in both projects.

3.1.3 Validation of reference regions for all cell lines within one project

In order to enable quantification of the histone modifications and the comparison between cell lines, the reference genes need to stably display these modifications at high levels. For this purpose the reference regions were checked for the consistency of epigenetic marks. The average Cq values and the corresponding coefficients of variation (CV) were calculated for 20 cell lines in project H and for twelve cell lines in project T. Calculations were performed for each ChIP condition, the mock precipitation and the untreated input sample (Table 3).

In the following table the average Cq values and the coefficient of variation (CV) of all cell lines, including the three biological replicates are displayed for project T and project H:

Project T: Cq values of ChIP Samples														
Target	H3		H3ac		H3K4me3		H3K9me3		H3K27me3		Mock		Input Sample	
	mean	CV	mean	CV	mean	CV	mean	CV	mean	CV	mean	CV	mean	CV
Eif3i 431	28.24	0.01	27.57	0.02	25.45	0.02	34.35	0.03	35.11	0.03	35.80	0.03	22.38	0.01
Gusb 386	27.16	0.01	26.73	0.02	24.64	0.02	33.38	0.03	33.90	0.03	35.76	0.04	22.32	0.02
Foxa2 484	26.32	0.01	36.43	0.05	36.20	0.05	31.54	0.02	33.60	0.02	34.64	0.04	21.31	0.03
Gata5 468	25.49	0.02	33.34	0.03	34.98	0.03	32.96	0.04	30.70	0.03	34.61	0.04	21.72	0.01
Rho 388	27.29	0.02	35.56	0.02	36.66	0.03	35.17	0.03	35.15	0.02	36.26	0.02	22.91	0.02
Unc13c 474	26.81	0.01	35.40	0.03	36.35	0.05	32.01	0.02	34.03	0.02	34.73	0.18	22.50	0.01
Project H: Cq values of ChIP Samples														
Target	H3		H3ac		H3K4me3		H3K9me3		H3K27me3		Mock		Input Sample	
	mean	CV	mean	CV	mean	CV	mean	CV	mean	CV	mean	CV	mean	CV
Eifi 431	26.41	0.02	27.81	0.02	27.88	0.03	35.66	0.06	34.34	0.05	35.55	0.06	22.94	0.02
Gusb 386	25.98	0.02	27.65	0.03	27.85	0.03	34.96	0.06	33.69	0.04	35.01	0.06	22.63	0.03
Foxa2 484	25.96	0.02	34.98	0.06	35.45	0.07	34.69	0.06	30.86	0.03	35.46	0.08	22.96	0.03
Gata5 468	25.75	0.02	33.61	0.05	35.30	0.07	35.16	0.07	30.09	0.03	34.88	0.06	22.62	0.03
Rho 388	27.08	0.02	34.79	0.04	35.91	0.06	35.82	0.06	33.05	0.03	36.10	0.06	23.90	0.03
Unc13c 474	26.70	0.02	35.02	0.05	35.03	0.06	34.79	0.05	31.50	0.02	35.12	0.05	23.55	0.03

Table 19: Test of reference regions in project T and H. The averages and the appropriate coefficient of variation (CV) of Cq values of all ChIP Samples including the biological triplicates were calculated for both projects. The rows display the different reference regions within the genome. The columns contain the specific histone modifications. The chosen control regions for the appropriate epigenetic modifications are marked as green cells. In both projects, the activation marks H3ac, H3K4me3 as well as total histone 3 are stable at the Gusb and the Eif3i regions (CV of 2 %). The lower Cq values suggests Gusb as the reference for activation marks. Also the repression mark H3K27me3 was in both projects stable at the Gata5 control region. In contrast, the investigation of project H cell lines leads to the conclusion that the Foxa2 region has different H3K9me3 amounts as it was observed in Project T. Unc13c and Rho were potential inactive regions, but both sequences obtained higher Cq and CV values compared to Foxa2 and Gata5.

The active regions Gusb and Eif3i consistently displayed high levels (i.e. low Cp values) of H3ac and H3K4me3 whereas H3K9me3 and H3K27me3 were at background level. The inactive region Gata5 was consistently marked with H3K27me3 whereas the activity markers H3ac and H3K4me3 were low. Surprisingly the reference region Foxa2 of project T loses the H3K9me3 mark in project H. Therefore the H3K9me3 levels can only be analyzed in project T through the normalization to Foxa2 whereas in both projects Gusb and Eif3i are suitable reference genes for H3ac and H3K4me3 and Gata5 can be used to normalize H3K27me3 levels .

3.1.4 Normalization of the different histone modifications at the hCMV-MIE promoter

The absolute amount of a certain histone modification at the hCMV-MIE promoter/enhancer is influenced by the number of transgene copies, e.g. a cell line with two transgene copies will display twice the amount of H3ac than a cell line with just one copy of the same status. Furthermore, the H3 density can differ between the cell lines and therefore the non-normalized modifications are also influenced by the histone density per promoter. In order to compensate for these biases we normalized H3 modifications to the level of H3 and looked for a reference gene for H3 as well. However H3 levels were constant in all cell lines with all reference genes suggesting that each of them can be used as reference gene for H3 (Table 19).

For a robust relative quantification of a target DNA by qPCR it is desirable to use target and reference primer pairs that have high and comparable amplification efficiencies. Therefore, the amplification efficiencies of hCMV-MIE and reference gene primers were determined by employing the LinReg PCR software, version 2014.5 (Ruijter et al., 2009). All amplification efficiencies were greater than 1.9 and differed by less than 2% (Table 20). Consequently, we were able to apply the simple $\Delta\Delta C_p$ or Livak method (2.2.1.11) to calculate relative levels of H3 modifications at the hCMV-MIE promoter enhancer. H3 modifications were normalized to H3 levels (H3ac/H3; H3K4me3/H3; H3K27me3/H3 and H3K9me3/H3).

Amplification efficiency with LinReg PCR					
Project T			Project H		
Primer pairs	mean	stdev	Primer pairs	mean	stdev
CMV 396/397	1.909	0.033	CMV 396/397	1.919	0.032
Gusb 386/387	1.923	0.036	Gusb 386/387	1.935	0.037
Eif3i 431/432	1.893	0.036	Eif3i 431/432	1.923	0.034
Foxa2 484/485	1.925	0.029	Foxa2 484/485	1.946	0.031
Gata5 468/467	1.914	0.033	Gata5 468/467	1.934	0.034
Rho 388/389	1.876	0.036	Rho 388/389	1.902	0.037
Unc13c 474/4745	1.904	0.032	Unc13c 474/4745	1.927	0.034

Table 20: Determination of primer efficiencies in project T and H. Amplification Efficiencies in project T and H were determined for all cell lines including biological triplicates. Because of a later detected point mutation within hCMV-MIE of project H cell lines a new primer pair (396/397) was used. For this purpose the qPCR raw data were applied to the LinRegPCR software, version 2014.5 and efficiencies were calculated. Rows are distinguished by the different primer pairs. The mean column contains the average amplification efficiencies of each primer-template pair and the stdev column contains the appropriate standard deviation. Distances between all amplification efficiencies are less than 2%. Thus the Livak method was used for normalization.

The normalization by Livak comprises two normalization steps:

- In the first step, the amount (Cq) of H3 modification close to the hCMV-MIE promoter is normalized to the amount (Cq) of H3 modification at the reference region. ΔCq
- In the second step, the normalized H3 modification is set to the normalized total H3 close to the hCMV-MIE promoter. $\Delta\Delta Cq$

This finally displays the relative amount of modification per H3 close to hCMV-MIE promoter!

In any case both parameters, H3 and H3 modification, were determined relative to the same reference gene. This minimizes variation within the normalization process. For all further studies Gusb was used as reference for H3ac and H3K4me3 whereas Gata5 was used for H3K27me3. Foxa2 was used for H3K9me3 in project T only. For instance, the relative histone acetylation per histone 3 can be normalized to the control region Gusb, as long as the Gusb region comprises a stable amount of the modified histone and the total H3. The following formula was used for calculation:

$$\Delta\Delta Cq \rightarrow ratio = \frac{2^{\Delta Cq_{H3_acetylation}(Cq_{Gusb} - Cq_{CMV})}}{2^{\Delta Cq_{H3_total}(Cq_{Gusb} - Cq_{CMV})}}$$

3.1.5 Effect analysis of potential prediction markers

After establishment of the CHIP assay to measure histone 3 modifications proximal to the hCMV-MIE, the abundance of these epigenetic marks and the alteration of their specific productivity ΔqP over 60 generations of the affected cell lines were correlated. In addition the recorded data of methylation of hCMV-MIE at C-179 (meC-179) and the plasmid copy numbers of one biological replicate were also analyzed. Before the correlations were performed, all values were screened for outliers employing the jackknife technique (Riu and Bro 2003). The distance for each observation is calculated with estimates of the mean, standard deviation, and correlation matrix that do not include the observation itself. By performing the analysis, two outliers were found above the upper confidence limit (UCL) in project H (Figure 31). ΔqP of cell line H-18 (Figure 31 A) in the absence of MSX was strikingly different as compared to all other cell lines (threefold increase in ΔqP) and one biological replicate of cell line H-3 (Figure 31 B) showed unusual high H3ac. Consequently, the third biological replicate of cell line H-3 as well as cell line H-18 were excluded from further effect measurements.

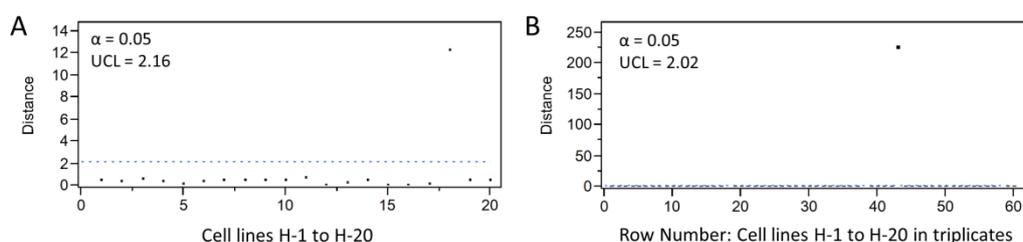


Figure 31: Jackknife outlier analysis of project H. The distance for each observation was calculated with estimates of the mean, standard deviation, and correlation matrix that do not include the observation itself. ΔqP - of cell line H-18 (A) and H3ac/H3 of the third biological replicate of cell line H-3 (B) were way above the upper confidence limit (UCL) and therefore excluded from the following effect measurements

The magnitude of the effects of epigenetic marks proximal to hCMV-MIE as well as the plasmid copies on stability were examined with a standard least squares fit model. To this end the software JMP10 (JMP®10.0.1 Release: 2, 64-bit edition; SAS Institute Inc.) was used. The significance levels were set to 0.05 (5%), 0.005 (0.5%) and 0.0005 (0.05%) and effects with lower p-values are described as significant, very significant and extremely significant inspired by the definition used in the GraphPad software. In addition to the copy number and the DNA methylation, the histone modifications H3ac, H3K4me3, H3K9me3 and H3K27me3 normalized to H3 as well as H3 alone were tested separately (Table IV and Figure 2).

In addition mock per H3 (Mock/H3) was considered as the unspecific effect of background noise per H3. To obtain a linear relation between stability and markers, all marker values were log-transformed. Thus, the adjusted R_squares were increased in each case and a higher

symmetrical distribution for the residuals was observed. The adjusted R_square is also termed the coefficient of determination and describes the strength of the effect.

Significant effects were detected in project H but not in project T. Project T comprises a smaller number of cell lines, with a lower variability of ΔqP values accompanied by a general higher stability than the cell lines of project H. Conceivably, the used prediction markers might be not accurate enough to distinguish between cell lines with almost similar values. In addition the project T cell lines were quite stable and a certain degree of instability might be needed for the correlation with specific histone modifications. Therefore the focus was set on project H to identify the best prediction mark.

In project H the most significant effect on ΔqP with and without selection agent was observed for H3ac normalized to H3 (H3ac/H3). The adjusted R_square also indicates H3ac/H3 as the most powerful effect (Table 21). The second strongest effect was observed for H3K4me3/H3, which confirms the coincidence of positive markers. Significant effects were observed for all histone markers. However I propose that H3 has a dominant influence on the combined effect, through the influence of the plasmid copy number as well as the histone density. The significant p-values for Mock/H3 support this hypothesis. In line with this assumption only H3ac/H3 as the most significant and strongest effector was taken into account as a potential prediction marker. In addition the effects obtained for the plasmid copy number and the DNA methylation were only significant for plasmid copy number in project H. Nevertheless, the DNA methylation correlates notably with the stability of cell lines in the presence of selection pressure in project H. The data for transgene copies/cell and meC-179 were compiled once. Because only one biological replicate was measured, the lower number of observations led to higher p-values.

Single effect leverage model: P-values for an epigenetic effect						
	Project T		Project H			
	$\Delta qP+$	$\Delta qP-$	$\Delta qP+$		$\Delta qP-$	
Epigenetic marks	P-values	P-values	P-values	Rsquare Adj	P-values	Rsquare Adj
log H3	0.6078	0.7737	0.0005	0.18	0.0015	0.15
log Mock/H3	0.7050	0.4594	0.0020	0.15	0.0034	0.13
log H3ac/H3	0.9801	0.7571	<0.0001	0.35	<0.0001	0.24
log H3K4me3/H3	0.6993	0.4909	<0.0001	0.30	0.0013	0.16
log H3K27me3/H3	0.9843	0.8354	<0.0001	0.27	0.0009	0.17
log H3K9me3/H3	0.3664	0.3639	n.d.			
log copies/cell	0.2806	0.2800	0.0404	0.19	0.0289	0.21
log me C-179 (%)	0.3885	0.6994	0.0743	0.13	0.1779	0.05

Table 21: Effect measurement of epigenetic modifications on long term stability in project T and H. P-values as indicator of significance and R_square adjusted as indicator of strength of epigenetic effects on ΔqP with (+) and without (-) selection pressure (250 nM MTX) are displayed in the columns. All epigenetic effects were log normalized and measured close to hCMV-MIE promoter. Effects of normalized H3, background noise per histone 3 (Mock/H3) and modifications per histone 3 (H3ac/H3, H3K4me3/H3, H3K27me3/H3 and H3K9me3/H3) were calculated in a standard least square model with software JMP10. Significant effects in project H cell lines but not in project T were observed. In project H, acetylation per histone 3 (H3ac/H3) has a highly significant effect on ΔqP . Also Mock/H3, H3K4me3, H3K27me3 and H3 events resulted in moderate to extremely significant effects.

It had been reported previously that hCMV-MIE methylation at C-179 and the number of transgene copies predicted a subsequent decrease in productivity (Osterlehner et al., 2011). Both parameters were measured in this study in order to determine their effect on production stability and to use it as a benchmark for the suitability of histone marker based prediction. It was observed that the plasmid copies and the DNA methylation correlate notably well with the loss of stability. Hence the strong effect of H3ac/H3 was remarkably significant (Figure 32 A-B), whereas the high copy number and high DNA methylation occurs predominantly in unstable cell lines (Figure 32 C-F).

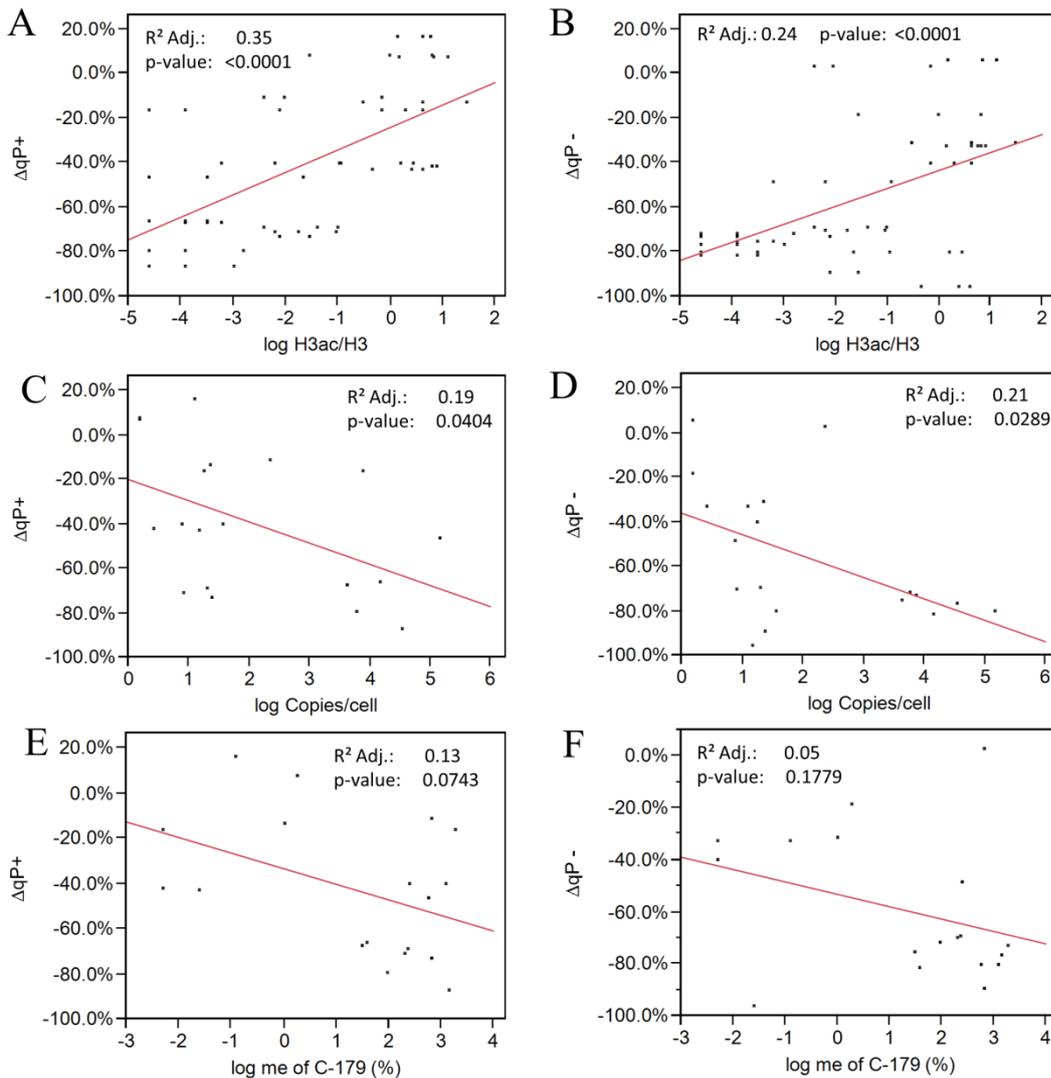


Figure 32: Correlation studies of log transformed plasmid copy number and epigenetic effects on ΔqP with and without selection agent. The alteration of qP over 60 generations (ΔqP) with (left) or without (right) MSX were plotted against H3 acetylation levels normalized to total H3 (A and B), copy number (C and D) as well as the DNA methylation at position C-179 (E and F) at hCMV-MIE. Production stability data from 19 cell lines and three biological ChIP replicates are shown. The p-values of the effect leverage plot are indicated.

Several studies propose an interplay between epigenetic marks and mechanisms (Feller et al., 2015; Murr, 2010). For this reason the correlation of plasmid copy number, histone 3 acetylation and DNA methylation at the hCMV-MIE promoter/enhancer was investigated. As expected, the degree of DNA methylation contrasted to histone 3 acetylation (Figure 33 A). This is plausible since acetylation of H3 is an activation mark, whereas DNA methylation is associated with inactivation. It has been reported previously that CMV-MIE methylation often coincides with elevated transgene copies, which could be confirmed in this study (Figure 33 B). Consequently histone 3 acetylation was increased in cell lines that have less transgene copies (Figure 33 C). Summing up, these correlations demonstrate the consistency of the data.

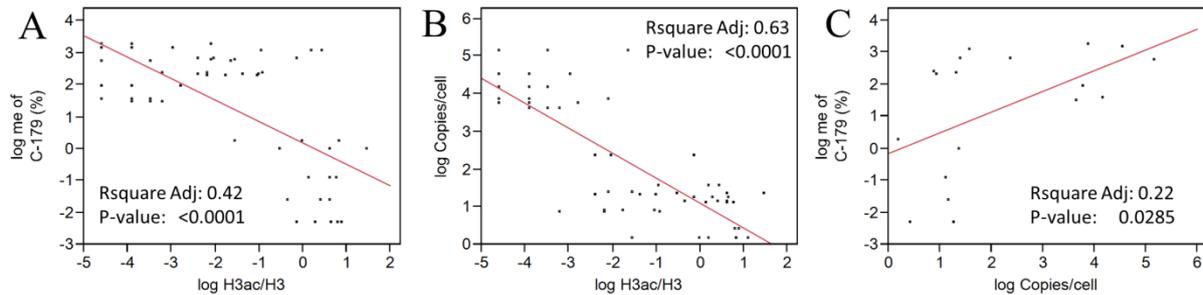


Figure 33: Interplay of epigenetic effects and the plasmid copy number. Histone 3 acetylation inversely correlates with promoter methylation and integrated plasmid copies. 19 cell lines were analyzed and p-value was calculated for the linear regression model of the effect leverage plot. (A) Methylation at C-179 (me of C-179 (%)) of hCMV-MIE was plotted against the H3ac/H3 values for three biological replicates per cell line. (B) The number of integrated plasmid copies was plotted against percent methylation at C-179. (C) The number of integrated plasmid copies was plotted against H3ac/H3. The degree of DNA methylation and copy number were contrary to histone 3 acetylation, which demonstrates the reliability of our data.

3.1.6 Exclusion of unstable producers by threshold setting for H3ac/H3 and meC-179

Amongst all parameters, the strongest correlation was observed in project H for Δ qP and H3ac/H3 in the presence of methotrexate. From this, H3ac/H3 was re-analyzed to identify a threshold that allows distinguishing between stable and unstable cell lines most effectively. To this end a LogWorth statistic was performed using the Jmp 10 software (Sall 2002), which led to the identification of the best split node. The LogWorth is calculated as $\log_{10}(\text{p-value})$, where the adjusted p-value is calculated regarding the myriad number of different ways splits can occur (Sall, 2002). In this case the best split was found at 0.47 acetylation per H3. To minimize false positive values, filter was set to 0.5 H3ac/H3 and this filter setting was applied in project H and T to exclude bad producers.

In addition, the published threshold of 5% for hCMV-MIE promoter methylation (meC-179 (%)) was also applied to the recorded methylation data (Table 17 and Table 18) as a benchmark for the predictive power of H3ac/H3. In this way samples with more than 5% DNA methylation or less than 0.5 acetylation per histone 3 at the hCMV-MIE promoter were excluded in separate examinations.

The average of ΔqP s of the remaining cell lines was calculated for both filter settings separately and compared with the unfiltered samples (Figure 34).

For project H, the average loss in qP over 60 generations was significantly decreased from 39% to 15% with MSX and from 55% to 38% without MSX employing the H3ac/H3 filter, whereas for the meC-179 [%] filter the average loss in qP decreased significantly from 39% to 24% with MSX. Also at the absence of selection agent the ΔqP decreased from 55% to 45% but without reaching significant difference between the two subpopulations (Figure 34). Therefore, H3ac/H3 appears to be more effective than meC-179 to pick cell lines with stable productivity.

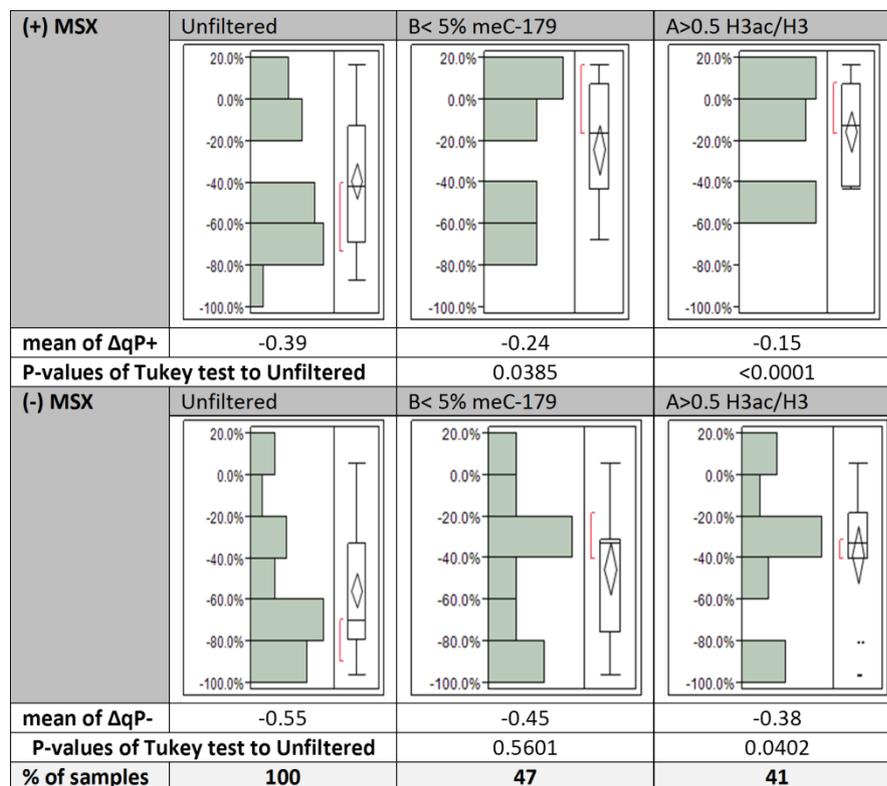


Figure 34: Application of filter settings. 0.5 is a suitable threshold for H3ac/H3 to exclude unstable producer cell lines in project H. The bar charts show the distribution of biological samples by ΔqP over 60 generations with (upper row) or without MSX (lower row), either without filtering (left column), after filtering for samples with less than 5% methylation at C-179 (meC-179: middle column) or after filtering for samples with more than 0.5 H3ac/H3 (right column). The average ΔqP of unfiltered and filtered samples is indicated below the charts. The percentage of samples before and after filtering is indicated in the lower row. The highest increase of stability was obtained for the H3ac/H3 filter settings. The H3ac/H3 filter results in a significant improved stability of sorted cell lines. Also the DNA methylation filter results in a significant accumulation of stable cell lines at the presence of MSX. The box plot shows the 1st and 3rd quartile indicated by the ends of the box, whereby the horizontal line within the box represents the median sample value. The end of the box has extended lines which represent the whiskers. The confidence diamond contains the sample means in the middle and the upper and lower 95% of the mean indicated by the top and bottom points. The bracket outside identifies the most dense 50% of the observations.

The application of the same filter settings in project T results in almost the same improved stabilities after H3ac/H3-dependent cell sort, which was significant in the presence of selection (Figure 35).

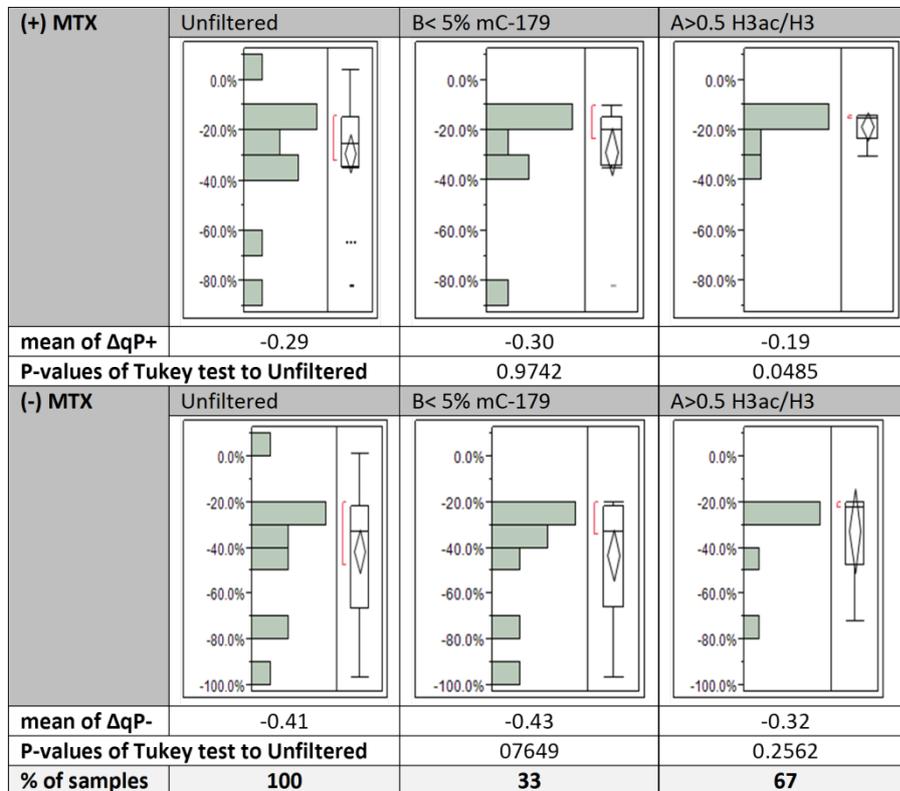


Figure 35: Transferred threshold settings from project H to project T. The bar charts show the distribution of biological samples by ΔqP over 60 generations with (upper row) or without MSX (lower row), either without filtering (left column), after filtering for samples with less than 5% methylation at C-179 (mC-179: middle column) or after filtering for samples with more than 0.5 H3ac/H3 (right column). The average ΔqP of unfiltered and filtered samples is indicated below the charts. The number of samples before and after filtering is also given. The H3ac/H3 filter at the presence of selection agent results in a significant improved stability of sorted cell lines. The box plot shows the 1st and 3rd quartile indicated by the ends of the box, whereby the horizontal line within the box represents the median sample value. The end of the box has extended lines which represent the whiskers. The confidence diamond contains the sample means in the middle and the upper and lower 95% of the mean indicated by the top and bottom points. The bracket outside identifies the most dense 50% of the observations.

3.2 Alteration of epigenetic landscape of CHO-K1 cell lines to circumvent the silencing of recombinant genes

Loss of productivity in CHO cell lines poses a serious risk in cell line development. In addition to the loss of copy number, epigenetic mechanisms like promoter methylation (Osterlehner et al., 2011; Yang et al., 2010b) or histone modifications (Mutskov & Felsenfeld, 2004; Paredes et al., 2013) are proposed repressors of productivity. Following this, reducing the accumulation of repression signals or an accumulation of activation marks close to the gene of interest might hinder epigenetic silencing effects. As described in the introduction, the acetylation of histone H3 and H4 is mainly associated with the activation of transcription. In particular, acetylation of lysine 16 at Histone 4 (H4K16ac) promotes an open chromatin formation (Corona et al., 2002). Inversely, CpG promoter methylation correlates strongly with transcriptional silencing of production cell lines (Smith & Meissner, 2013). Consequentially, targeting H4K16 acetylation to the promoter that drives the gene of interest and reduction of CpG methylation sites might stabilize the productivity of recombinant CHO cells.

In order to check whether H4K16 hypoacetylation results in higher productivity, the histone acetyltransferase MOF was fused to a DNA binding domain. MOF was coexpressed with eGFP reporter and targeted to the corresponding cis element upstream of the promoter. Stably transfected CHO pools were cultivated and fluorescence intensity was measured.

Furthermore, I generated various CHO cell lines which contain reporter genes driven by hCMV-MIE variants. The promoter variants contain up to three C to G conversions of specific CpG sites. The influence of the mutations on promoter strength and expression stability was assessed using reporter gene assays after transient and stable transfection of suitable constructs.

3.2.1 Coupled expression of target gene and the specific histone acetyltransferase MOF, proposing open chromatin formation.

It has been shown in *Drosophila* that Gal4-BD MOF can be targeted to the UAS binding sites, which leads to H4K16 acetylation of adjacent regions (Corona et al., 2002; Prestel et al., 2010), which in turn promotes the expression of the adjacent reporter gene (Schiemann et al., 2010). I aimed for evaluating this concept in CHO cells. Therefore I fused the histone acetyltransferase MOF to a Gal4-BD, which is able to bind the yeast upstream activating sequence (UAS_{Gal}). A plasmid was generated, comprising 5x UAS sites in front of the hCMV-MIE promoter. This promoter is supported by the IRES element EV71 to drive the simultaneous expression of the reporter gene eGFP and Gal4-BD MOF.

In addition a control plasmid expressing unfused Gal4-BD was also generated. CHO cells were transfected with the appropriate plasmids in triplicates and stable transfectants were selected using the DHFR/MTX system. After 34 days of selection, the CHO pools were cultivated for additional 90 days at a constant MTX concentration (250 nM).

During the cultivation period cell pools were examined every week for their eGFP expression by flow cytometry with the BD FACS Canto II (BD, Heidelberg, Germany) (2.2.3.9). The geometrical mean of fluorescence intensity of living cell pool was calculated for each measurement and plotted against the time of cultivation (Figure 37).

During the cultivation the CHO pools with GalBD-MOF plasmid (plasmid no. 16125) had lower eGFP expression than the control pools (plasmid no. 16124) (Figure 36 and Figure 37).

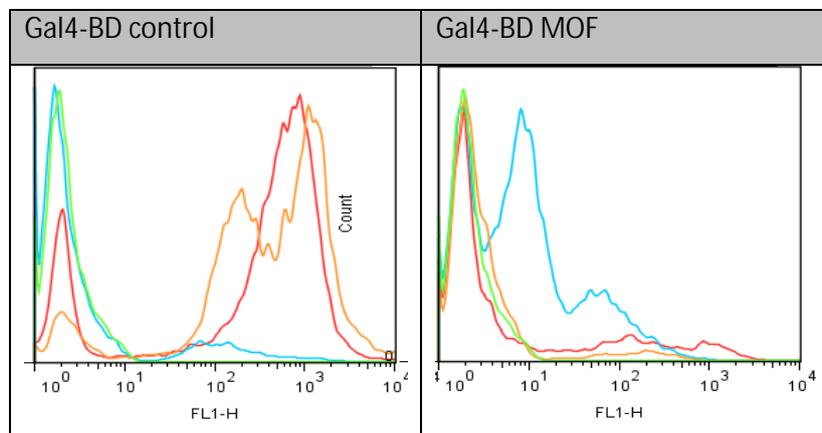


Figure 36: Histogram of fluorescence intensity (x-axis) of CHO pools, 52 days after transfection. 10,000 events (y-axis) per pools were measured. Gal4-BD MOF expressing CHO pools one (red), two (blue) and three (orange) had fewer events with high fluorescence intensity than the Gal4-BD control pools one (red), two (blue) and three (orange). Autofluorescence of CHO-K1 wild type cells are displayed in green.

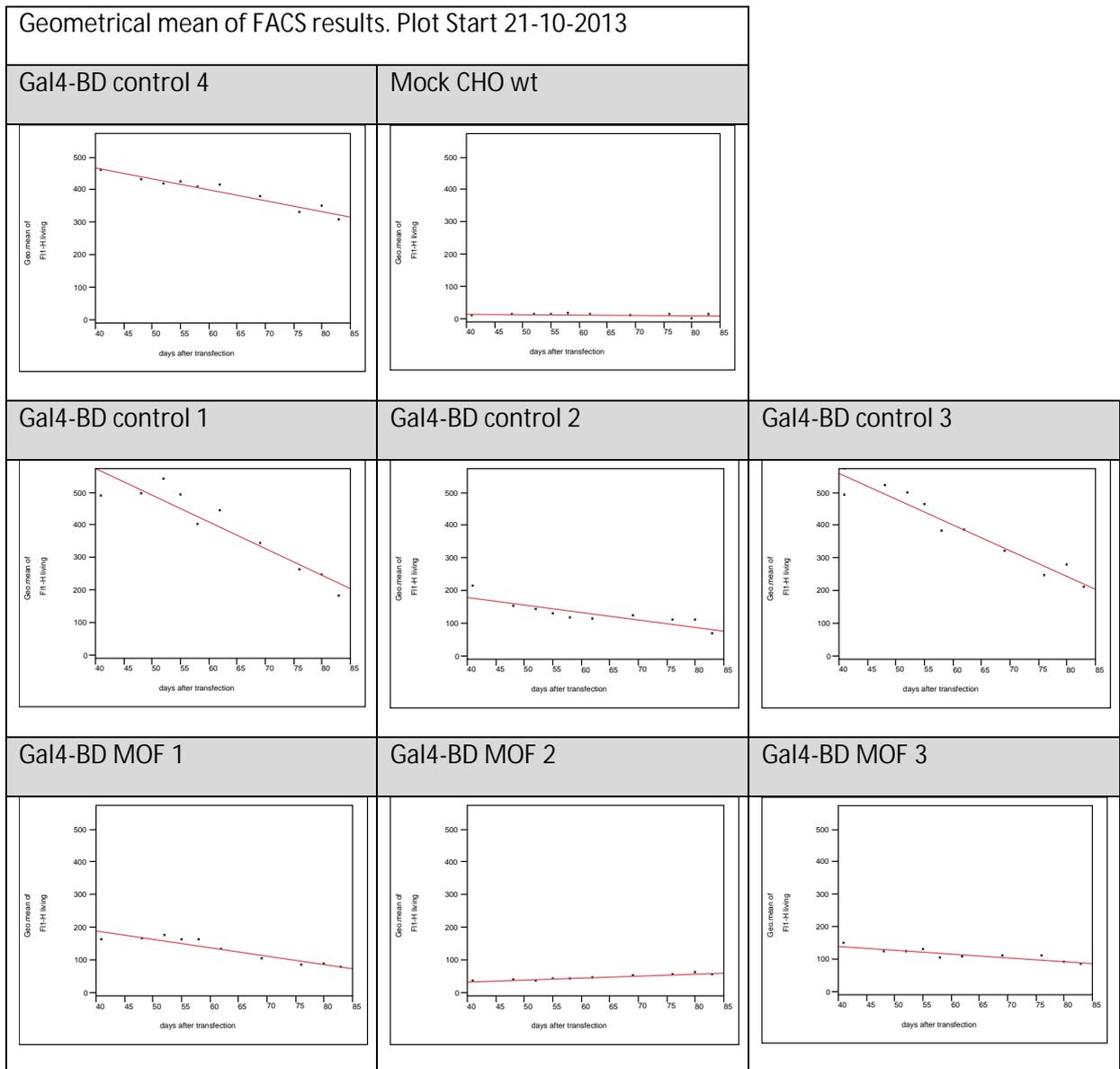


Figure 37: Geometrical means of FACS measurements over time. Gal4-BD MOF pools have lower eGFP signals overall compared to control pools 16124. High expressors of Gal4-BD control have a strong loss of eGFP signal over time, but expression was still higher than in the Gal4-BD MOF pools (plasmid no. 16125).

In addition to the eGFP expression the global H4K16 acetylation was measured. To this end, the membrane/cytoplasmic and nuclear fraction of all cell pools were prepared 52 days after transfection (Figure 38) according to Misawa et al. (2006).

After transfection the same amount of nuclear proteins of each cell pool was analyzed in two independent SDS-PAGE assays. The Western blot was conducted with three independent MOF and control CHO cell pools and the global acetylation of histone 4 at lysine 16 was measured with the appropriate antibody.

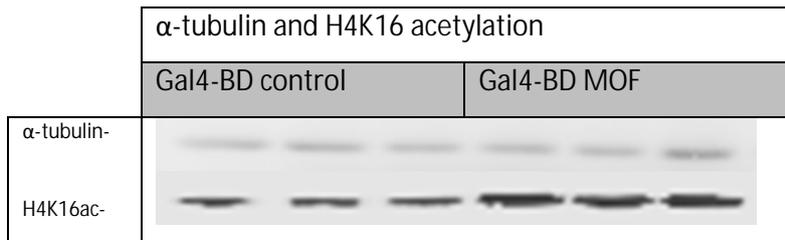


Figure 38: Western blot of the nuclear fraction of three independent MOF and control CHO pools. Acetylation of H4K16 was measured with an appropriate antibody, 52 days after transfection. Thereby global H4K16 acetylation was noticeably higher in all MOF expressing pools than in the control pools.

Although the eGFP expression of MOF comprising cell pools was very low (Figure 36), the global H4K16 acetylation was noticeably higher compared to controls (Figure 38). To avoid over-interpretations of H4K16 acetylation by a small fraction of high producers, pools were sorted 50 to 57 days after transfection for high eGFP expressers. Sorted pools were cultivated for 45 days and eGFP expression was recorded. A shift of eGFP expression was observed in most of the CHO pools, especially one MOF pool did not remain stable (Figure 39).

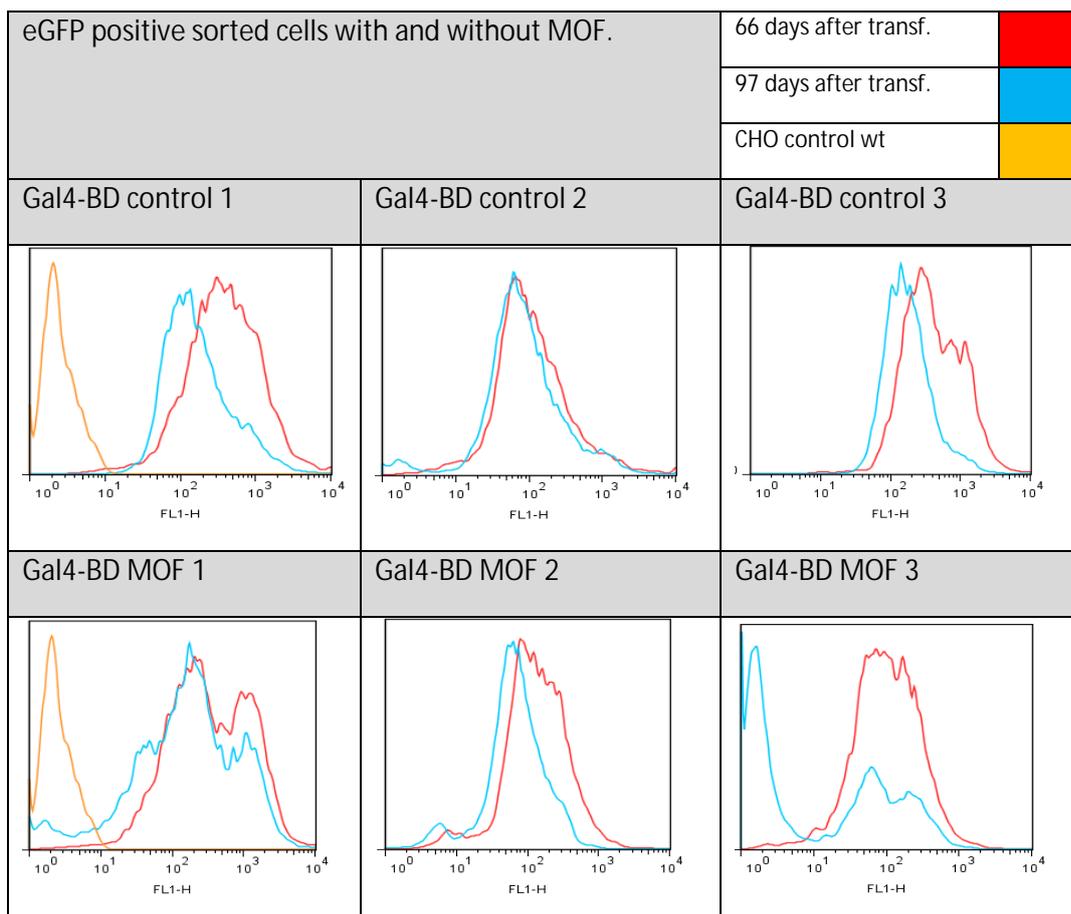


Figure 39: Histogram of fluorescence intensity (x-axis) of eGFP positive sorted CHO pools. 10,000 events (y-axis) per pool were measured. After recovery of sort, pools were cultivated for one month and start (red) as well as end (blue) of new cultivation was displayed in the histograms. Autofluorescence signal of CHO-K1 wild type cells is displayed in orange. Gal4-BD MOF and Gal4-BD control CHO pools tend to lose eGFP expression over time. Particularly Gal4-BD MOF CHO pools have a rapid decrease of eGFP signal.

The concurrent appearance of eGFP and Gal4-BD or Gal4-BD MOF transcripts is ensured by the IRES-coupled expression. Furthermore, the global increase of H4K16 acetylation demonstrates the presence of a functional expressed MOF protein. Nevertheless, low eGFP expression of MOF pools was observed. To exclude the possibility of over interpretation of acetylation signal by a small fraction of highly hyperacetylated cells, pools were eGFP positive sorted and cultivated. Thereby a MOF dependent effect on stability was not detected. Unfortunately, I wasn't able to detect MOF either in a western blot or in a chromatin immunoprecipitation assay. Considering this, the effect of MOF close to the reporter gene could not be analyzed.

However, these preliminary results suggest that at least a global hyperacetylation of H4K16 does not promote stable expression of recombinant genes. In regard to the strength of the used IRES and the eGFP positive sort the presence of MOF is very probably but the hypoacetylation of H4K16 might not have the assumed effect. The additional expression of MOF might rather hinder stable expression by overburden of cell metabolism by his additional expression.

3.2.2 Point mutation of CpGs within the hCMV-MIE promoter to improve production stability

Osterlehner et.al. (2011) have reported that specific CpG sites within the hCMV-MIE promoter are frequently methylated in unstable CHO cell lines. This suggested that these sites may play a prominent role in silencing of the hCMV-MIE promoter. To test this hypothesis different hCMV-MIE promoter variants which comprise cytosine to guanine mutation of those specific CpG sites were generated. The resulting effects were measured in different reporter plasmids by the expression of secreted embryonic alkaline phosphatase (SEAP), eGFP and an IgG antibody-cytokine fusion protein (Figure 40). The cytosine to guanine conversions were performed at location -508, -179 and -41 upstream of transcription start site of hCMV-MIE. In doing so I chose methylation-susceptible CpGs from the enhancer (-508 bp), the proximal promoter (-179 bp) and the core promoter (-41 bp) region, which in addition are not involved in known transcription factor binding. Furthermore, I engineered C to G conversions to maintain the overall GC content without the recreation of new CpG sites (Figure 40 B). The used promoter variants were abbreviated with a three letter code, starting from C-508 over C-179 to C-41 (Figure 40 C). Each point mutation within the hCMV-MIE is indicated by a G instead of a C. All point-mutated constructs were compared to the unmutated control CCC. Plasmid variants were used for transient and stable transfection of CHO-K1-M cells.

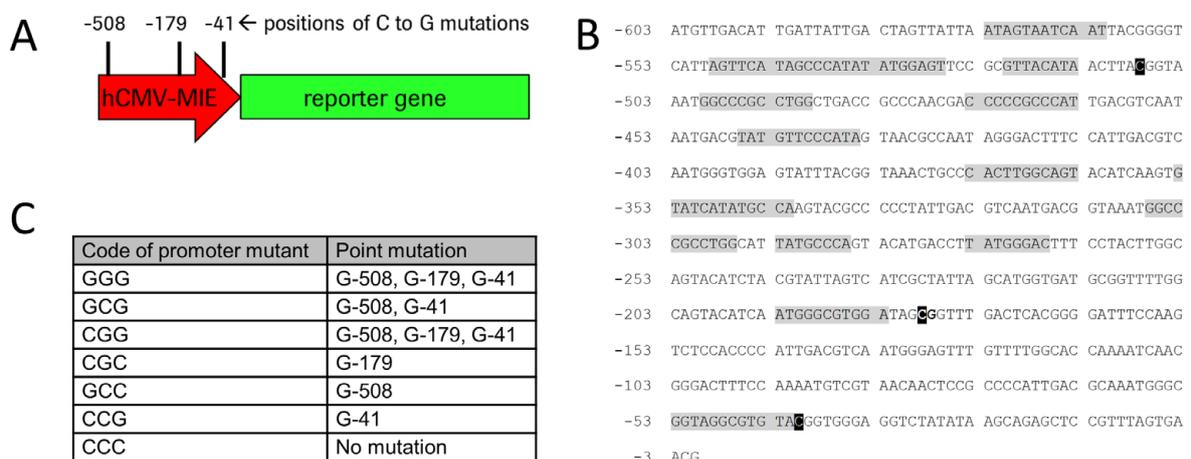


Figure 40: hCMV-MIE: Effects of C to G conversions at position-508, -179 and -41 upstream of transcription start site was measured with the expression pattern of the reporters SEAP, eGFP and IgG-cytokine fusion protein (A). CpG sites susceptible for methylation were mutated without the changing known transcription factor binding sites marked in grey and without creating of new CpG sites. C to G mutations are highlighted with a black background. putative transcription factor binding sites are highlighted with a grey background. Those sites were identified by employing software PROMO 3.0 for species cricetidae which uses the TRANSFAC database version 8.3 (B). Various point mutations were combined in several hCMV-MIE promoter variants and examined in different reporter vectors. Promoter variants are identified by a three letter referring to the nucleotides at positions -508, -179 and -41 (C).

3.2.2.1 Transient transfection: Investigation of promoter strength in CHO pools

During the selection phase, CHO cells are highly stressed (Chusainow et al., 2009), which results in unreliable expression of target genes. Considering that the expression of transgenes in the early selection phase is not consistent, the direct promoter strength of hCMV-MIE variants were investigated to ensure comparability of constructs.

To this end CHO cells were transiently transfected with reporter vectors carrying SEAP under the control of a specific hCMV-MIE promoter variant. The expression level of SEAP was examined by measuring the metabolic rate of para-nitrophenylphosphate to para-nitrophenole on day five after transfection, (2.2.3.8).

The averages of the resulting optical densities were normalized to those obtained after transfection of the unmutated control plasmid (CCC). The mock controls are duplicates of water-transfected CHO-K1 cells. Two independent experiments were performed, denoted in Figure 41 as assay 1 and 2.

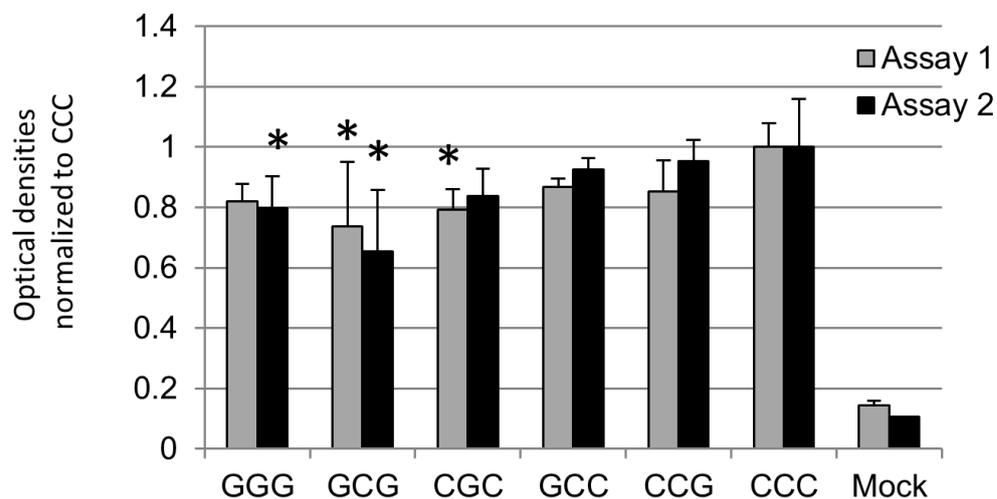


Figure 41: Transient transfection of CHO cells with SEAP expression plasmids under the control of hCMV-MIE promoter variants. The optical densities obtained by ELISA assay were normalized to unmutated control CCC. Mock is the untransfected control. Error bars represent the standard deviation of eight replicates. The diagram shows the result of two independent transient transfection experiments. Differences between mutants and control CCC were tested with the Tukey HSD test and significance was set to $p < 0.05$. Significant differences are marked with an asterisk.

The hCMV-MIE promoter variants especially with multiple mutations showed noticeable differences to the OD of the unmutated control (CCC) (Figure 41). For a statistical evaluation, means of replicates were tested with the Tukey honest significant difference (HSD) test (Table 22) (Abdi, 2010). The significance level was set to 0.05 (Figure 41). The SEAP expression of GGG and CGC mutants was significantly lower in one experiment whereas expression by GCG was significantly lower in both assays. In general the CpG mutations rather led to a reduced than enhanced direct promoter strength.

P values of Tukey HSD test, comparing SEAP expression of mutants to control				
Code of promoter mutant	DAC untreated		DAC treated	
	Assay 1	Assay 2	Assay 1	Assay 2
GGG	0.126	0.016	0.985	0.999
GCG	0.006	<0.0001	0.348	0.76
CGC	0.049	0.424	0.098	0.578
GCC	0.446	0.856	0.076	0.521
CCG	0.334	0.984	0.578	0.989

Table 22: P-values of Tukey HSD Test as an indicator of different promoter strength of mutants to the unmutated control. Normalized OD's of the SEAP assays were used for comparison. Point mutations are ordered in rows and the average of eight replicates in DAC untreated and the average of four replicates in DAC treated condition were compared with the average of unmutated control. The columns comprise the Tukey HSD calculated P-values of the two assays with and without DAC treatment. The significance level was set to 0.05 (5%). Only GCG point mutations results in significant lower OD values in both assays in absence of DAC.

3.2.2.2 Stable transfection: Investigation of stable eGFP expression in CHO pools

With the knowledge about the direct promoter strength of promoter mutants I was able to analyze the influence on long term stability. Consequently I examined, whether hCMV-MIE promoter variants influence the expression stability of stable transfected CHO cells during long term cultivation. For this purpose eGFP was chosen as reporter gene to enable the measurement of fluorescence intensity of each individual cell in the suspension (Table 23).

Four replicates per hCMV-MIE promoter variant were generated by transfection of CHO-K1-M cells. Two days following transfection, selection was started by adding MTX. After the reduction of the overall cell viability to < 20 %, the value was maintained for one week followed by slow recovery over one month until a viability of > 97% was obtained. Stable transfected cell suspensions expressing eGFP were cultivated over a period of one to three months. The intensity of eGFP expression was measured with BD FACS Canto II (BD, Heidelberg, Germany) and primary data analysis was performed with the FlowJo 7.6.5 EN software (TreeStar, Olten, Switzerland). The geometrical means of the fluorescence intensity of 10,000 events per measurement were calculated

for each living cell pool and plotted on a logarithmic scale against the time of cultivation (2.2.3.9). The start and end point of cultivation was compared to identify high and stable expressers (Figure 42 A).

hCMV-MIE promoter variants in eGFP reporter gene plasmids (Figure 24)			
Plasmid No.:	Code of promoter mutant	Point mutations	Feature
16105	GGG	G-508, G-179, G-41	Gal4-BD binding sequence UAS
16106	GCG	G-508, G-41	
16107	CGC	G-179	
16108	GCC	G-508	
16109	CCG	G-41	
16111	CCC	No mutation	

Table 23: eGFP reporter plasmid list: Code of promoter mutant describes the C to G point mutations within the hCMV-MIE promoter located -508, -179 and -41 base pairs upstream of transcription start site. CCC is the code for the unmutated control with UAS sites.

I found that the CHO pools with the CGC or CCG mutation yielded distinctively higher eGFP values than the control at the beginning of the cultivation phase. Although consistent eGFP expression over time suggested an enhanced stability of CGC pools, no significant difference (significance level 0.05) to control was determined with the Tukey HSD test. Therefore, I performed a power analysis in order to calculate the minimal number of samples required to obtain a significant difference of promising CGC (G-179) and CCG (G-41) mutants to the control. At least eight replicates per construct were suggested and a second experiment was pursued.

Stable transfected CHO cell pools with eight replicates of CGC promoter mutants (two losses by contamination) and ten replicates of CCG and CCC promoter mutants were selected, cultivated and analyzed as described previously (Figure 42 B). The CGC and CCG mutants showed moderately to extensively higher eGFP values than the control at the beginning of the cultivation phase. During cultivation an increase of eGFP expression was observed in some CCG mutants, suggesting higher stability. The Tukey HSD test showed that the eGFP expression in CCG mutants were significantly higher than in the control on day 42 after transfection. However a decrease of eGFP expression over time was seen in CCG as well as CGC mutants, which might be caused by evolutionary pressure within the complex CHO pools.

In addition to eGFP measurements, the copy number per cell as well as the hCMV-MIE promoter methylation was determined for each CHO pool 42 and 69 days after transfection (Figure 42 C and D). The average copy number per pool was calculated as described previously (Osterlehner et al., 2011) and were compared between mutants and control by employing the Tukey HSD test. Although

no significant difference was obtained, the G-179 (CGC) mutants have less transgenes compared to the control and the G-41 (CCG) mutants. The promoter methylation was examined after bisulfite conversion of gDNA, the amplification of hCMV-MIE and the library preparation followed by sequencing employing the TrueSeq Nano Kit and the MiSeq system (Illumina, Inc., San Diego, USA). The average CpG methylation of hCMV-MIE was calculated with the Bismark software (Krueger & Andrews, 2011). Similar to the tendencies obtained for the copy numbers, the G-179 (CGC) mutants are less methylated compared to the control and the G-41 (CCG) mutants.

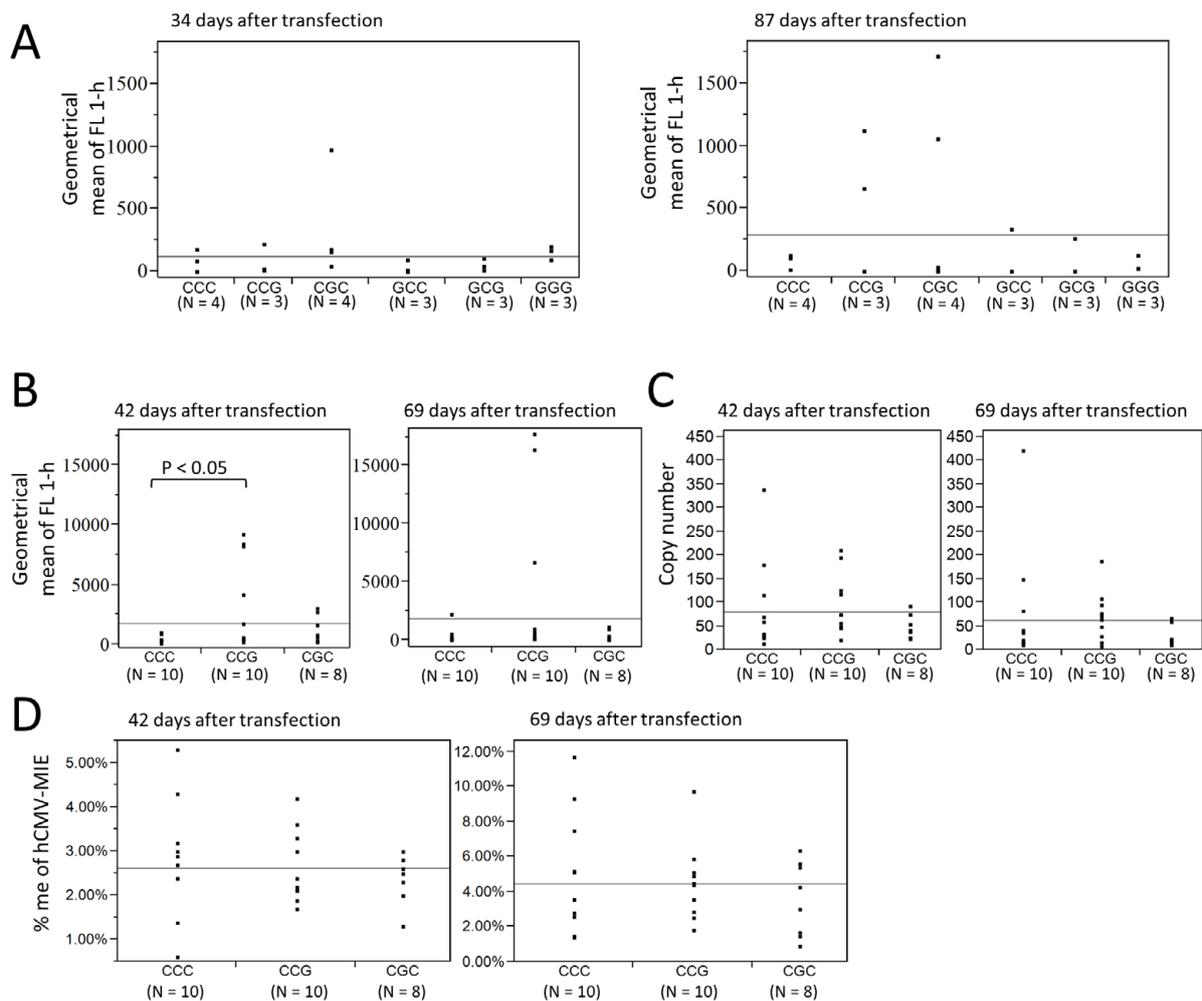


Figure 42: eGFP expression of stable transfected CHO cell pools. Promoter mutants were compared to unmutated control (CCC). The progression of eGFP expression is exemplified for begin and end of the cultivation phase. N indicates the number of replicates. (A) Prominent eGFP expression was observed in CGC and CCG pools at begin of cultivation phase. However significant difference to control was not obtained. (B) Assay was repeated for CCC, CCG and CGC with a higher replicate number indicated by the symbol N. Again higher eGFP expression was observed in CCG and CGC mutants at begin of cultivation phase, whereat the eGFP expression of CCG pools seems to be maintained over time. In addition the CCG pools had significantly higher eGFP expression than the control 42 days after transfection. Neither the copy numbers (C) nor the methylation of hCMV-MIE (D) of the appropriate eGFP pools from (B) were not significantly different between the promoter variants.

The methylation coverage for each CpG was also calculated with the Bismarck software and the methylation patterns for all pools are displayed in Figure 43. I found the unmethylated CpG at position -179 upstream of TSS prominently methylated in comparison to all other CpG sites. Furthermore, it could be observed that the mutation of CpG -179 and -41 did not result in an altered methylation pattern across the CMV promoter. However, a general lower methylation of methylation-susceptible CpGs could be observed in both mutants and in particular for the CGC promoter variant. For this reason the average methylation of hCMV-MIE (Figure 42 D) describes a reliable value for further analysis.

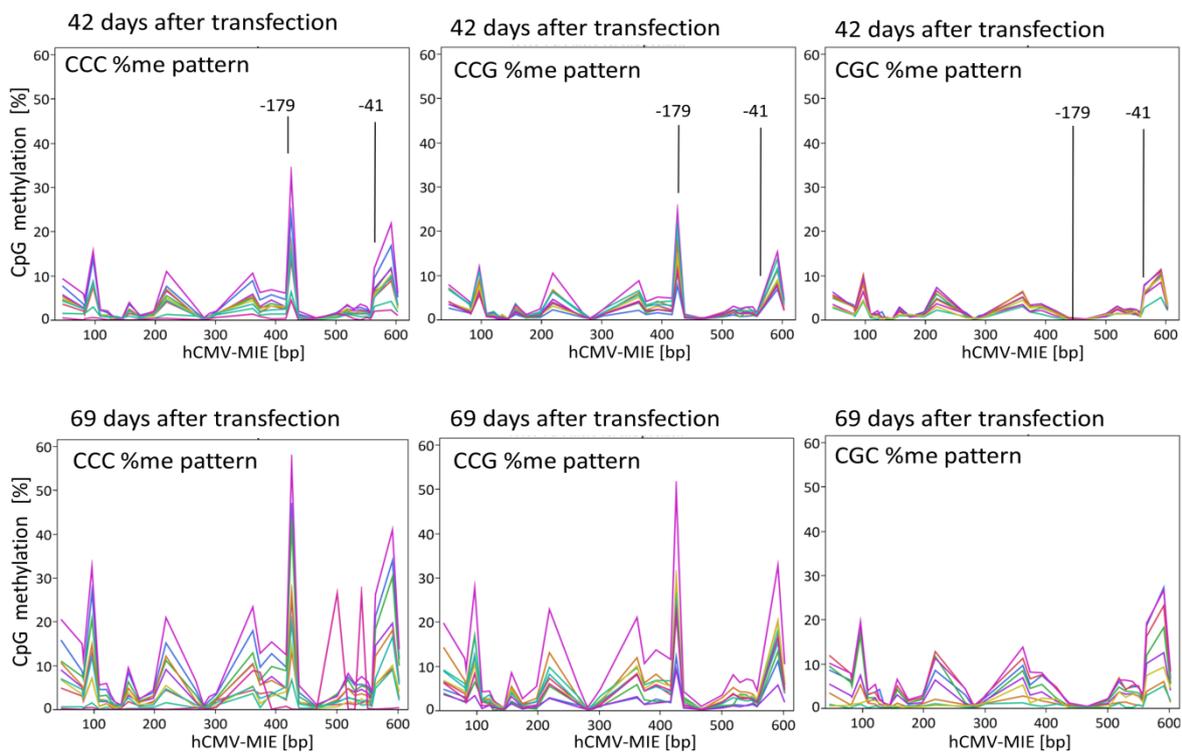


Figure 43: hCMV-MIE methylation pattern of CpGs. Methylation of GFP expressing pools were examined for promoter variants CCG, CGC and control CCC 42 and 69 days after transfection.

Although only CCG pools yielded significantly higher eGFP values, both point mutations followed the trend towards higher eGFP expressions observed, in two independent experiments (Figure 42 A and B). Considering the high variability of pools I assume that the mutation effect might be masked by the increased evolutionary competition within the cell suspensions. As a consequence, the promoter mutants CGC and CCG were investigated in clonal cell lines to reduce variability. Furthermore, these effects were measured in conditions as close as possible to the standard cell line development (Figure 44). Therefore, both mutations alone as well as in combination (CGG) were examined as IgG-cytokine fusion protein expressing clonal CHO cell lines. In consideration of the copy number similarity of the different CHO pools, I propose the same bias in clonal cell lines.

3.2.2.3 Stable transfection: Investigation of stable IgG-Cytokine fusion protein expressing monoclonal CHO cell lines

IgG plasmids with CGG, CGC, CCG promoter mutants and the unmutated control of hCMV-MIE promoter were generated (Table 24 and Figure 44). The CGG mutation was constructed to assess potential additive effects of both single mutations. As reporter the IgG-cytokine fusion protein (Table 23) was chosen to measure expression and stability. Furthermore, the IgG-cytokine fusion protein is a pipeline product of Roche, which enables effect measurements close to the standard cell line development. To generate monoclonal cell lines, CHO-K1-M cells were transfected with the appropriate plasmid, isolated in 384 well plates and kept under high selection pressure for three weeks. In order to assess cell viability, cultures were continuously expanded until 41 days after transfection. 30 clonal cell lines per construct were randomly picked and transferred into six well plates for long term cultivation. During the cultivation phase some cell lines were discarded as a result of contamination or reduced viability. 22 to 30 clones per construct were examined for two months.

C to G point mutations in human CMV-major immediate-early promoter/enhancer fragment of human antibody of class IgG expressing plasmids (Figure 26)		
Plasmid No.:	Code of promoter mutant	Point mutations
16134	CGG	G-179, G-41
16135	CGC	G-179
16136	CCG	G-41
21504	CCC	No mutation

Table 24: IgG reporter plasmid list: Code of promoter mutant describes the C to G point mutations within the hCMV-MIE promoter located -179 and -41 base pairs upstream of transcription start site. CCC is the code for the unmutated control. IgG expression is driven by the appropriate hCMV-MIE promoter.

The antibody concentration was kindly examined by Manuela Bernhard using ELISA technology and antibody titer ($\mu\text{g/ml}$) and qP was calculated (2.2.3.12). Measurements of the beginning and the end of the cultivation phase were displayed to visualize the difference of mutants to control and alteration of expression during cultivation. The titers at the end of cultivation phase were found noticeably higher in single mutants than in the control cell lines (Figure 44 A). In particular the CGC (G-179) mutant obtained significantly higher titer during the cultivation phase than the control. Significant difference was tested again with the Tukey HSD test. To reduce potential propagation effects, the time of cultivation and the concentration of cells were considered. For this purpose the qP of mutants and control were compared during the whole cultivation phase (Figure 44 B). Furthermore to reduce errors which might occur during the estimation of qPs like miscounting of cells or dilution errors, the specific productivity of the first two (Start = 68 & 83 days after

transfection) and the last three (End = 124, 127 & 131 days after transfection) qP values of each cell line were averaged. The performance of qP values at the beginning and the end of the cultivation phase matched the respective titers. The difference of CGC mutants to unmutated control increases over time and at the end of cultivation the specific productivity of CGC mutants were tremendously higher than the control.

Taking into account the slightly reduced direct promoter strength of mutants, the comparable copy number values obtained in the previous assay as well as the higher specific productivity of mutants, I conclude about a stabilizing effect of the point mutations G-179 and G-41. To verify this hypothesis, I calculated the alteration of the specific productivity ΔqP of antibody expressing cell lines during the cultivation phase (Figure 44 C). Therefore, I only chose cell lines which produce antibodies from the beginning of the cultivation phase. In the following formula the ΔqP being the percental alteration of qP, qP_{start} being the specific productivity at the beginning of the cultivation period and qP_{end} being the specific productivity at the end of the cultivation period.

$$\Delta qP = \left(\frac{qP_{end}}{qP_{start}} - 1 \right) * 100$$

The comparison of all ΔqP values shows that one control sample had an uncommonly high ΔqP value, indicating a more than six fold higher increased productivity than the average (Figure 44 C: left plot). In light of this exceptional high ΔqP value of one control, a jackknife outlier analysis was performed, which led me to exclude sample CCC-13. qP values were plotted again and the averages were compared with the Tukey HSD Test. Hence the monoclonal cell lines with the CGC and CCG promoter mutants obtained a significantly higher stability than the unmutated control cell lines (Figure 2 C right). In particular the G-179 point mutation results in remarkable expression stability over time.

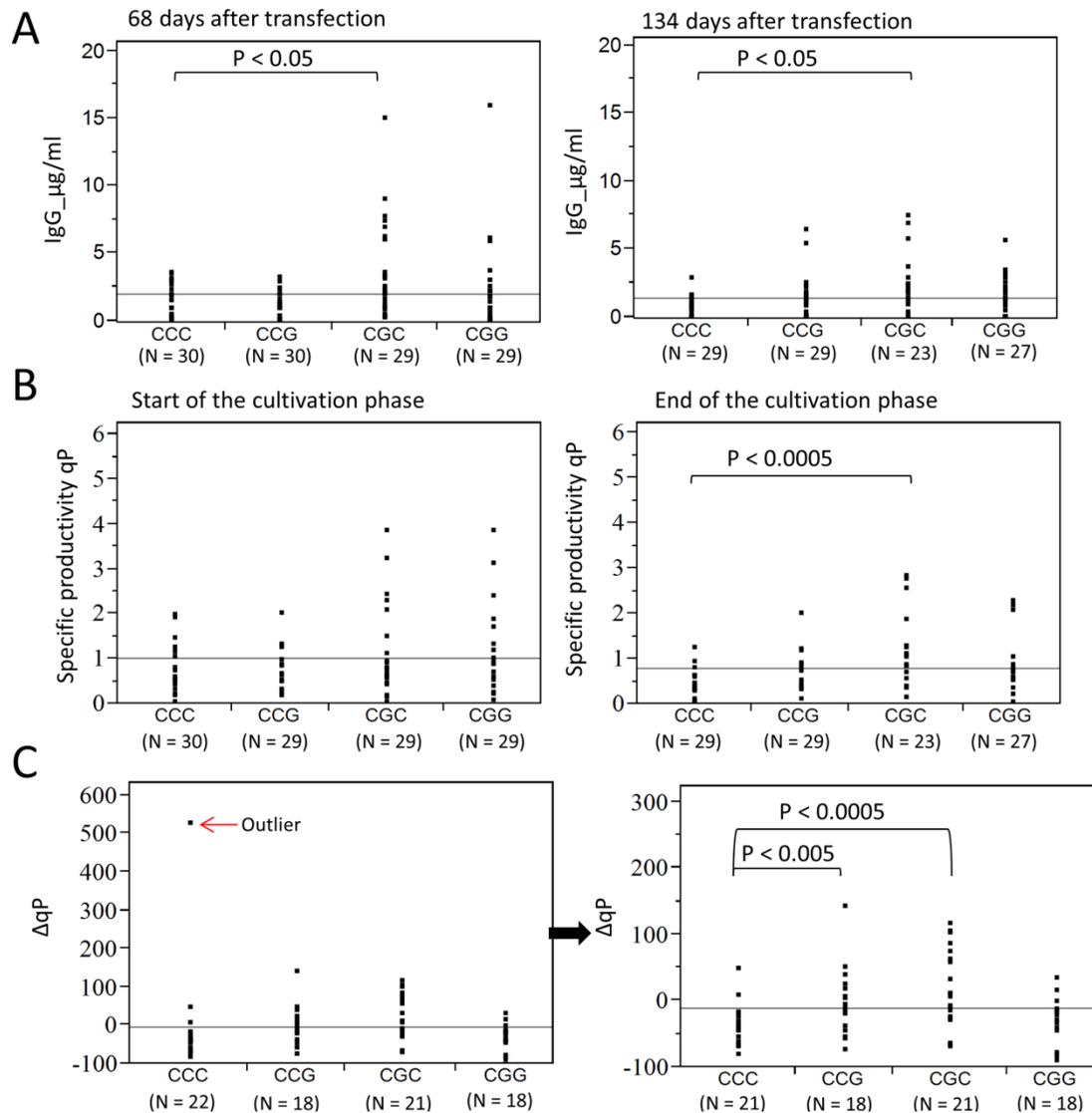


Figure 44: Generation of stable transfected clonal CHO cell lines which express antibodies under the control of CCC, CCG, CGC and CGG promoter variants. Comparisons of mutants and control were performed with the Tukey HSD Test. A) Antibody titer ($\mu\text{g/ml}$) of CCG, CGC and CGG cell lines were compared to unmutated control (CCC), employing the Tukey HSD test. The titer from the beginning (left) and the end (right) of the cultivation phase are displayed for all cell lines. Higher titers were obtained for the CGC and the CGG mutant at the start of cultivation. At the end of cultivation all mutants obtained higher antibody titer than the control. B) The specific productivities (qP) of all cell lines were calculated as pg/cell per day and were averaged for the beginning (left) and the end (right) of the cultivation phase. CGC cell lines have significantly higher qP values at the end of cultivation phase. C) The percentage alteration of specific productivities was compared between control and mutants. After exclusion of outlier cell line CCC-13 by employing the jackknife technique, the CGC and CCG mutation resulted in significant higher stability compared to the unmutated control cell lines (CCC).

Taken together, the single point mutations G-41 and G-179 have a positive effect on expression stability in CHO pools as well as in clonal cell lines. The observed effect in CHO pools was not as distinctive as in clonal cell lines, which might be linked to the high variability of pools and the subsequent evolutionary competition within the suspension. However, the examination of antibody-expressing clonal cell lines shows the tremendous effect of G-179. Interestingly combined mutations enhance the stability neither in eGFP cell pools (e.g. GGG) nor in antibody clonal cell lines (CGG). This

indicates that the number and the position of point mutations are more important than a general depletion of CpG sites within the hCMV-MIE. Considering that the initiation of transgene silencing depends not only on the sequence of transgene, but also on the tendency to form tandem repeats upon integration, and the integration site into the host genome (Kaufman et al., 2008) I propose a dominant effect of DNA methylation on gene expression.

3.3 Epigenetic Landscape of CHO-K1-M cells and potential stable integration sites

Transfection of CHO cells with expression vectors followed by selection and cultivation phases is a common method to obtain high titer of a recombinant protein. However, many cell lines are prone to lose productivity over time. During the transfection and the selection phase the expression vectors integrate randomly into the genome. It seems plausible that the integration of transgenes within active, highly transcribed regions promotes the transcription of the transgene. A randomly transfected cell pool most likely contains high expressers with frequently integrated transgenes in active regions and low expressers with transgenes predominantly located in inactive regions of the genome.

3.3.1 Chromatin state of CHO-K1 wild type

In order to enable the segmentation of the CHO-K1 genome in active and inactive regions, a ChIP sequencing experiment, monitoring of well-known marks was performed. To this end, chromatin was immunoprecipitated with chip grade antibodies against specific diagnostic histone modifications and the enrichment over input sample was determined by qPCR. The H3K4me3, H3ac and H3K36me3 modifications are mainly present in active euchromatin whereas H3K9me3 is known as a repressive signal (Zentner & Henikoff, 2013; Zhou, Goren & Bernstein, 2011). I performed the ChIP to enriched chromatin for H3K4me3 and H3K9me3 with two biological replicates (2.2.3.16). For qPCR the control primers for Gusb, Gata5 and Foxa2 sites were used and enrichment was normalized to the input sample. Thereby a strong accumulation of H3K4me3 on Gusb site and moderate enrichment H3K9me3 at the Foxa2 site was observed in both biological replicates (Figure 45 A and B). Consequently, samples were transferred to GATC (GATC-biotech, Konstanz, Germany) for sequencing. The ChIP sequencing for H3ac as well as H3K36me3 was performed by Active motif (La Hulpe, Belgium) from a provided sample of CHO-K1 chromatin. For this purpose I prepared chromatin as described for ChIP sequencing and send the sample to Active motif on dry ice. Active

motif completed the H3Ac and H3K36me3 ChIP reactions using 25 μg of Chinese hamster ovary cell chromatin and 5 μl of H3Ac antibody (Active Motif, cat # 39139) and 4 μg of H3K36me3 antibody (Active Motif, cat # 61101). The qPCR was performed using positive control primers that were designed for CHO cells based on conserved sites in human and mouse, as well two negative control primer pairs that amplify regions in gene deserts (Untr87, Untr99). Results were normalized to input values and to 1000 cells. The signal for H3Ac binding is strong at the Gapdh site. In the case of H3K36me3, signals were strong on Gapdh and Actb in both biological samples (Figure 45 C and D).

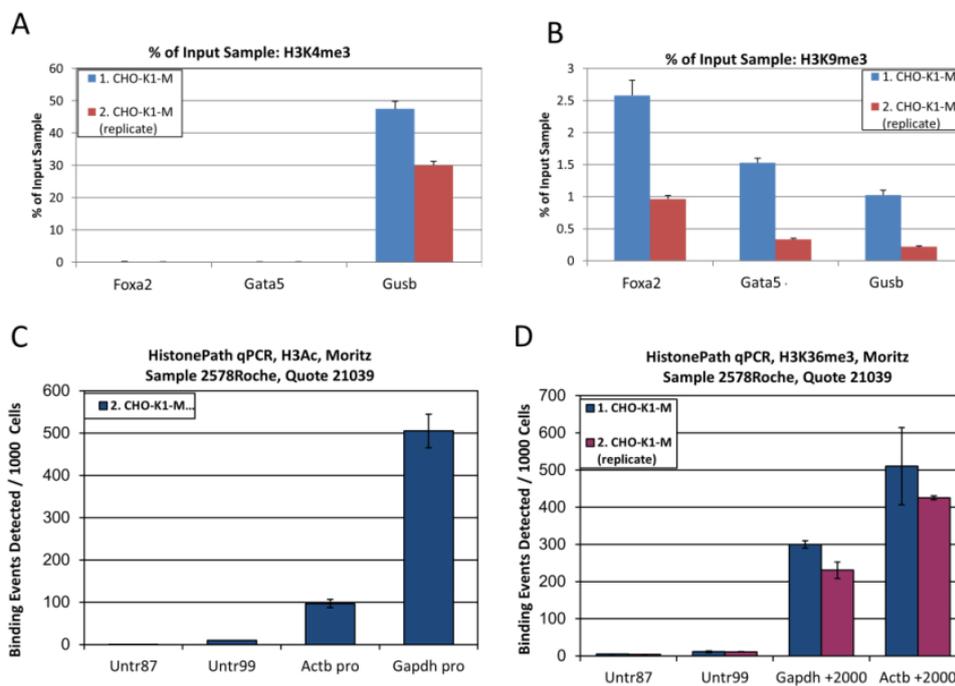


Figure 45: Enrichment of histone modification on specific genomic regions analyzed by qPCR. Accumulation of H3K4me3 and H3K9me3 were examined in house as % of Input sample (A and B). In contrast, Active Motif normalized the enrichment to the input values and the number of cells (C and D). Strong enrichment was observed for H3K4me3 on Gusb, for H3ac on Gapdh and for H3K36me3 on Actb. Moderate enrichment was observed for H3K9me3 on Foxa2.

Sequencing was done with the Illumina, HiSeq (Single Read, 50 bp; GATC) and Illumina, NextSeq 500 instruments (Single Read, 75 bp; Active Motif).

Encode suggests for mammalian broad-source histone marks more than 20 million uniquely mapped reads with a non-redundant fraction of greater than 0.8 million (Landt et al., 2012). Alignment of reads and mapping was kindly performed by Matthias Barann. Reads of the integration sites, generated by inverse PCR, and the ChIP sequencing reads for H3K36me3, H3ac, H3K4me3 and H3K9me3 were aligned to the CHO-k1 genome (genBank Assembly ID GCA_000223135.1), employing the Burrow-Wheeler alignment tool BWA (Li & Durbin, 2009). In doing so, the quality criteria of Encode could be respected for the ChIP sequencing results.

Histone modification	Sequencing Service	# of all reads	# of mapped reads	% of all reads
1_H3K36me3	Active Motif	33167150	26117120	78.74%
2_H3K36me3	Active Motif	40764874	23992326	58.86%
1_H3ac pan	Active Motif	31233220	21303019	68.21%
InputControl A	Active Motif	29208305	22008294	75.35%
1_H3K4me3	GATC	51681788	27614201	53.43%
1_H3K9me3	GATC	34733630	26931375	77.54%
2_H3K4me3	GATC	51434453	33773785	65.66%
2_H3K9me3	GATC	37975834	28716974	75.62%
InputControl B	GATC	65239893	55050603	84.38%

Table 25: BWA mapping report of ChIP sequencing reads. More than 20 million reads were obtained and mapped to the CHO-K1 genome.

Peak calling of the aligned reads was performed with the Model-based Analysis tool for ChIP Sequencing termed MACS2 (Zhang et al., 2008b). Hence H3K36me3 and H3K9me3 were investigated with the broad peak model, whereas narrow peaks were called for H3ac and H3K4me3.

3.3.2 Validation of integration sites

For the investigation of location-dependent effects on integrates, I examined transfected CHO pools with random integration sites of transgenes and compared their productivity with the location of the transgenes and the surrounding chromatin state of the CHO-K1 wild type. CHO cells were transfected with an eGFP reporter plasmid and three independent high variable cell pools were selected for one month. After selection phase, eGFP expressing CHO-K1 cell pools were sorted into two subcultures of high and low eGFP expressers. Cells were harvested and stored two weeks after sort while high eGFP expressing pools were still cultivated to verify the long term stability of eGFP expression (Figure 46). Sorted cell pools 2.3 and in particular 1.7 could be clearly distinguished by their eGFP expression. Even five weeks after sorting the eGFP expression of high eGFP expressing pools remained stable as it can be seen in the stability control. The eGFP expression in the low eGFP expressing pool 4.6 (-) might compromise the discrimination of integration sites that support stable or variegated expression in these pools. Therefore further analysis of pool 4.6 was done with a note of caution.

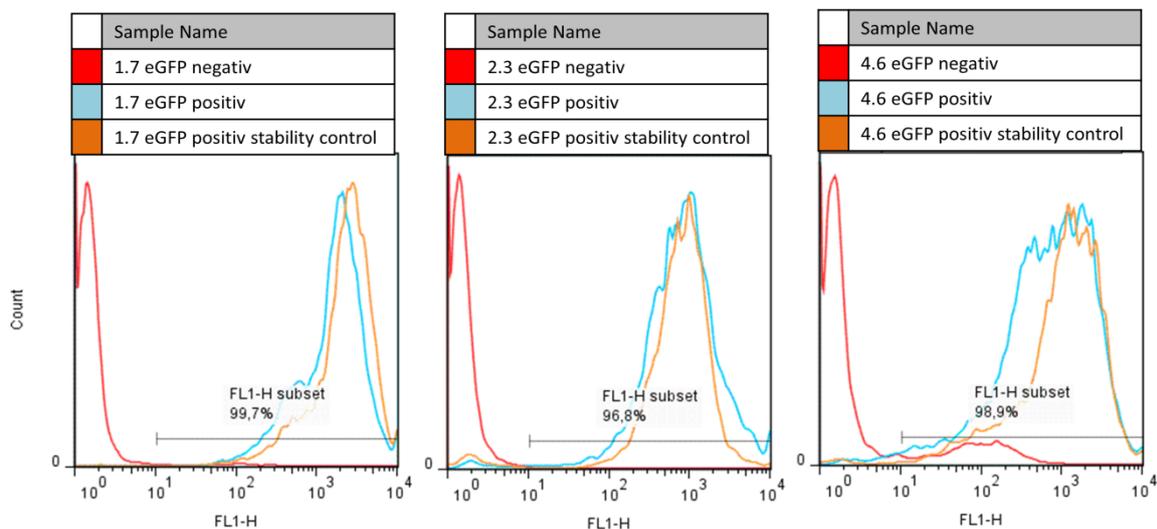


Figure 46: Subcultures of eGFP expressing CHO-K1 pools. Two weeks after sorting into high and low eGFP expressing pools, cells were analyzed employing flow cytometry and genomic DNA was stored for inverse PCR. The stability control of high eGFP expressers was analyzed five weeks after sort to confirm stability of expression.

Inverse PCR of subcultures was conducted to locate transgene integration sites of sorted cell pools. For this purpose the stored genomic DNA of CHO pools was digested with the 4 bp cutting enzymes. The chosen enzymes did not cut between the primer binding sites and the linearization site of the vector. The enzymes cut statistically every 256 bp in the genome, which should result in small fragmented DNA. After self-ligation only fragments with the appropriate primer binding sites are

able to be amplified. Only the 5' iPCR of MseI and MluCI cutted fragments results in amplification without an unspecific band in control CHO-K1 wild type (Figure 47). For this reason the 5' amplicons of each sample were combined and purified with High Pure PCR Purification Kit (Roche, Penzberg, Germany) in 40 µl Elution buffer. The sequencing of fragments was done by GATC-Biotech (Konstanz, Germany).

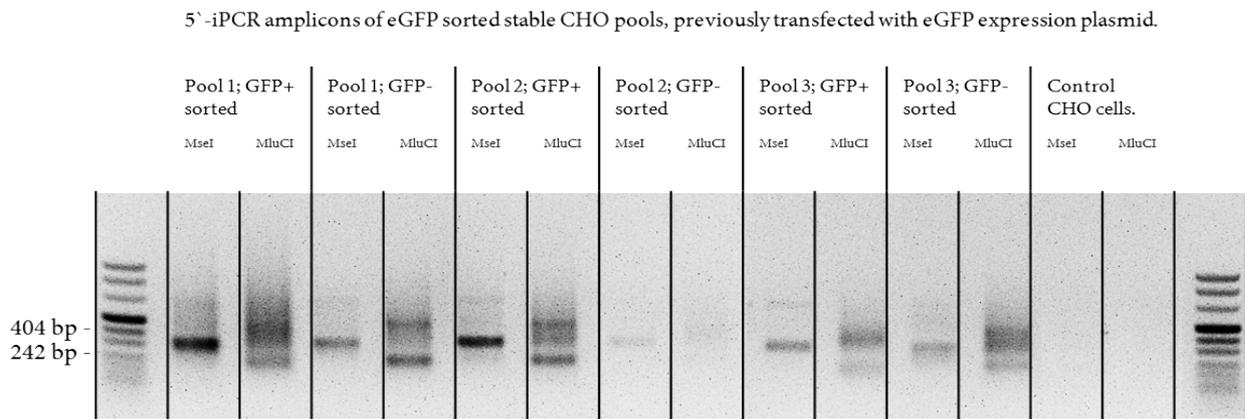


Figure 47: 5' iPCR of stable CHO pools after eGFP sort. Pools were randomly transfected with eGFP reporter plasmid and MTX selected. After one month of cultivation, pools were positive and negative eGFP sorted and integration sites were examined with iPCR. No unspecific amplifications were obtained in untransfected CHO control cells using 5' integration site primers.

For the alignment of the amplicons the BWA software (Li & Durbin, 2009) was used. Unfortunately only a small proportion of reads could be aligned to the CHO-K1 genome (Table 26).

Pools	Sequencing Service	# of all reads	# of mapped reads	% of all reads
1,7 minus	GATC	3114808	31670	1.01%
1,7 plus	GATC	10680020	128682	1.20%
2,3 minus	GATC	4344334	105701	2.43%
2,3 plus	GATC	1964838	26851	1.36%
4,6 minus	GATC	2786248	151038	5.42%
4,6 plus	GATC	3916138	167914	4.28%

Table 26: BWA report of amplicon alignment on CHO-K1 genome. Only a small fraction of reads could be aligned.

In light of the low number of aligned reads, the specific amplification by inverse PCR primers was verified (Table 27). To this end all reads were scanned for the primer sequence and known vector sequences, which revealed that 80% to 96% of all reads containing the primer sequence also contained the adjacent vector sequence. This implies in the most cases a correct primer annealing and amplification. The low number of mapped reads might be a result of a large number of tandem

repeats or viral infection, which suggests that further analysis should be interpreted with caution. The mapped integration sites were compared with activation and inactivation marks of the ChIP sequencing.

Sequence files of sorted eGFP pools	% of reads containing primer sequence	% of reads containing primer & adjacent vector sequence
1_7_gfp_minus_lib32736_1894_1_1.fastq.gz	35%	85%
1_7_gfp_minus_lib32736_1894_1_2.fastq.gz	27%	80%
1_7_gfp_plus_lib34418_1923_1_1.fastq.gz	47%	85%
1_7_gfp_plus_lib34418_1923_1_2.fastq.gz	40%	83%
2_3_gfp_minus_lib32738_1862_1_1.fastq.gz	32%	95%
2_3_gfp_minus_lib32738_1862_1_2.fastq.gz	34%	95%
2_3_gfp_plus_lib32737_1923_1_1.fastq.gz	44%	94%
2_3_gfp_plus_lib32737_1923_1_2.fastq.gz	38%	91%
4_6_gfp_minus_lib32740_1862_1_1.fastq.gz	40%	96%
4_6_gfp_minus_lib32740_1862_1_2.fastq.gz	36%	95%
4_6_gfp_plus_lib32739_1862_1_1.fastq.gz	46%	95%
4_6_gfp_plus_lib32739_1862_1_2.fastq.gz	41%	94%

Table 27: Specificity of primer binding: Most of the reads did not contain the primer sequence. However most of the primer sequence containing reads have also the following 16 bp vector sequence before the linearization site PvuI.

3.3.3 Combined analysis of chromatin state and integration sites

Until now the CHO-K1 genome has not been fully annotated. For this reason all investigated software tools such as ChromHMM are not able to cluster the genome in predominately active and inactive regions with respect to the histone modifications H3K4me3, H3ac, H3K36me3 and H3K9me3. Furthermore, the most overlapping sequence data (contigs) are very small, which impedes the full analyses of the regions. Accordingly, I was not able to perform a general analysis of all integration sites with a chromatin map, segmented in active and inactive areas.

To circumvent this problem, Matthias Barann scripted a small program, in which all peaks within 6,500 bp around an integration site were counted and listed. The table 28 displays for every pool the number of peaks of the investigated histone modification 6,500 bp around each integration site. The coverage of an integration site is the average number of reads representing a given nucleotide in the reconstructed sequence. This read depth should correlate with the frequency of the specific integration site within the examined pool.

To assure comparability I excluded all contigs, which are smaller than 13,000 bp, and all Integration sites, which are closer than 6,500 bp to one end of the appropriate contig. Taking into consideration that the H3K36me3 and H3K9me3 are the most promising histone modifications to distinguish between active and inactive chromatin states, I excluded all contigs, which do not contain at least one of those marks. In regard to the poor mapping results for integration sites, I only analyzed integration sites with coverage of more than 100x. Unfiltered integration sites with coverage of >0 are depicted as control. With the threshold of >100x coverage, the total number of integration sites and the appropriate contigs reduces drastically to 57 integration sites for low eGFP expressers and 33 integration sites for high eGFP expressers. Following the assumption that silenced integration sites should mainly occur in inactive regions and vice versa, I calculated the ratio of H3K9me3 to H3K36me3 peaks for the integration sites of low and high eGFP expressers, and the statistical ratio for the whole genome of CHO-K1 wild type as control (Table 28). This analyses show that the portion of H3K9me3 increases close to integration sites compared to the statistical ratio of 3.25 H3K9me3/H3K36me3. In particular the low eGFP expressers obtained a notably higher value for H3K9me3 close to their highly covered integration sites. However, a significant difference of unfiltered negative and positive sites with respect to the number of H3K9me3 peaks could not be demonstrated. For this reason those data must be treated with caution.

Integration sites	Peaks of histone modifications				
	H3K9me3	H3K36me3	H3panAc	H3K4me3	H3K9me3/ H3K36me3
low (-) eGFP pools with 100x coverage	114	18	5	2	6.33
high (+) eGFP pools with 100x coverage	71	17	3	2	4.17
(-) eGFP pools with coverage >0	920	215	82	37	4.27
(+) eGFP pools with coverage >0	3182	741	244	120	4.29
close to all integration sites	4000	943	321	154	4.24
global	150496	46214	37998	24822	3.25

Table 28: Peaks count of histone modifications located 6,500 bps around integration sites. Peaks were counted for high covered (>100x coverage), low covered (>0x coverage) and all integration sites. In addition all peaks were counted independent of integration sites. The ratio of the most prominent inactive marker H3K9me3 and active marker H3K36me3 were calculated.

Due to the unreliability of a general analysis and under the assumption that the coverage correlates with the frequency of the integration sites within the examined pool, single integration sites with the highest coverage were examined. The three most prominent integration sites for each pool were chosen (Table 29) and displayed with the IGV genome browser (Figure 48) (Robinson et al., 2011). To

facilitate discrimination of active and inactive regions, the integration sites of low eGFP (-) expressing pools are red and the integration sites of high eGFP (+) expressing pools are green.

In addition the inactivation mark H3K9me3 is painted in purple whereas the activation marks H3K4me4, H3ac and H3K36me3 are painted in green (Figure 48). Hence it could be observed that all integration sites of low eGFP expressing pools are surrounded by the inactivation mark H3K9me3 (Figure 48 A-C). In contrast to this the integration sites of the high eGFP expressing pools were encircled by active as well as inactive marks (Figure 48 D-F). The 1.7 (+) pool obtained the best association of integration sites and activation marks followed by the 2.3 (+) pool, whereas no activation mark could be observed close to the most prominent integration sites of the high eGFP expressing 4.6 (+) pool. This suggests that many silenced transgenes also exist in high expressers. The lower eGFP expression and the reduction over time in the 4.6 (+) pool support this theory (Figure 46).

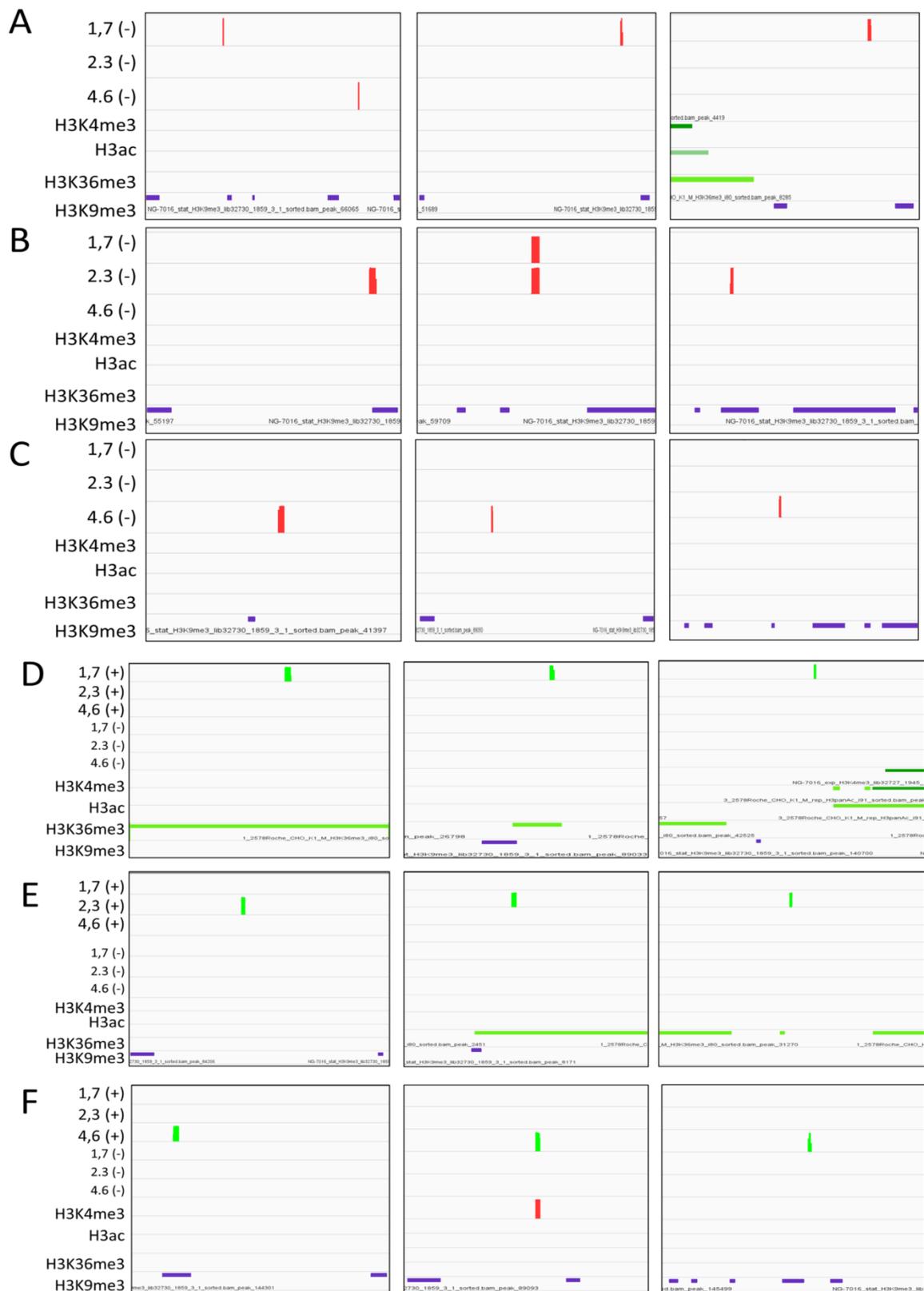


Figure 48: IGV genome browser pictures of the three highest covered integration sites of each pool. The integration sites of the low eGFP expressing pools 1.7 (-), 2.3 (-), and 4.6 (-) are displayed as red columns in A, B and C whereas the integration sites of the high eGFP expressing pools 1.7 (+), 2.3 (+), and 4.6 (+) are pictured in D, E, and F as green columns. The inactivation peaks of H3K9me3 are portrayed as purple bars. The activation peaks for H3K4me3, H3ac and H3K36me3 are portrayed as green bars. All integration sites of low eGFP expressing pools are surrounded by H3K9me3 peaks. The integration sites of high eGFP expressing pools predominantly encircled by activation marks in particular in 1.7 (+) pool.

Pools	Integration site	start	end	contig	contig length
1.7 (-)	1	33710	33879	gi 529736278 gb APMK01209506.1	61339
	2	27820	28038	gi 529632532 gb APMK01259496.1	34698
	3	17844	18011	gi 529542894 gb APMK01299512.1	28254
2.3 (-)	1	9560	9732	gi 529755684 gb APMK01200188.1	18950
	2	33177	33363	gi 530129180 gb APMK01019987.1	48062
	3	12592	12772	gi 529840071 gb APMK01159973.1	24537
4.6 (-)	1	9329	9392	gi 529543284 gb APMK01299321.1	44147
	2	15923	16036	gi 529899463 gb APMK01129909.1	29170
	3	8008	8136	gi 529695444 gb APMK01229803.1	24501
1.7 (+)	1	11339	11522	gi 529983619 gb APMK01090275.1	21655
	2	14892	14989	gi 529899456 gb APMK01129916.1	28873
	3	28880	28977	gi 530129135 gb APMK01020008.1	43344
2.3 (+)	1	14549	14753	gi 529879688 gb APMK01139946.1	26816
	2	7600	7740	gi 529542165 gb APMK01299870.1	19894
	3	33062	33224	gi 529961879 gb APMK01099785.1	52446
4.6 (+)	1	12833	12999	gi 529899487 gb APMK01129885.1	30233
	2	39673	39872	gi 530149103 gb APMK01009994.1	56793
	3	9333	9506	gi 530147711 gb APMK01010668.1	18622

Table 29: Position of the three highest covered single integration sites of each pool. High eGFP expressing pools are indicated by (+) and low eGFP expressing by (-).

4 Discussion

4.1 Molecular Indicator of Production Instability

Epigenetic marks are well known to alter the expression of genes and specific marks correlate with an open or a closed formation of specific chromatin regions (Chen & Dent, 2014). Most therapeutic proteins are produced in large fed-batch cultures of CHO cells. Late stage reduction of therapeutic protein expression levels, caused by epigenetic alterations can therefore be very cost intensive. To minimize that risk, epigenetic markers were investigated at an early stage to distinguish stable from unstable production cell lines. On account of this, the epigenetic modifications of cell lines were measured at the start point of cultivation and compared to the alteration of stability over 60 generations.

It was demonstrated that the level of histone H3 acetylation at hCMV-MIE can be used to screen for stable cell lines and exclude unstable ones from further development early on. The results suggest that cell lines, which are highly acetylated at H3, are less prone to transgene silencing than others. In line with this, hypoacetylation of H3 has been shown to accompany loss of productivity in unstable CHO cell lines (Paredes et al. 2013; Spencer et al. 2015) suggesting that H3 hypoacetylation plays a major role in silencing of transgenes in CHO cells.

Considering the moderate predictive power indicated by the adjusted R^2 , I aimed to identify a threshold of H3ac/H3 which appears to be optimally suited to separate unstable from stable producers. It appeared that most cell lines with $H3ac/H3 > 0.5$ have above average production stability whereas most cell lines with $H3ac/H3 < 0.5$ have below average production stability. After applying this filter most selected cell lines lose not more than 40% of their specific productivity over 60 generations with and without selection agent, which we consider as the critical limit. Furthermore I was able to demonstrate in two independent cell line development projects that the subpopulations filtered by H3 acetylation are significantly more stable than the unfiltered populations, employing the Tukey HSD test.

The selection of cell lines by acetylation of H3 at the heterologous hCMV major immediate early promoter and enhancer appears to be more reliable and effective than using DNA methylation as a marker. According to the general model of gene silencing, activating histone modifications disappear and inactivating histone modifications occur before the DNA is permanently 'locked' into the silent state by methylation (Smith and Meissner 2013; Yin et al. 2012). This sequence of events may explain why the early occurring H3 acetylation is better suited as early indicator of production

stability or instability than promoter methylation. Both, H3 acetylation and production stability generally anti-correlate with methylation of hCMV-MIE as well as the number of transgene copies, which demonstrates the consistency of the data.

It seems obvious that the H3 acetylation levels of low copy number cell lines might be primarily influenced by the location of transgenes, particularly by the position effect, which can influence the stability of recombinant gene expression (Lattenmayer et al. 2006; Yin et al. 2012). For this reason I propose that the degree of H3 acetylation close to the transgene is not only a sign of activation but also an indicator of stable integration sites.

4.1.1 Limitation of prediction marker

As already noted in the result section, the differences of the effect analyses between project H and project T indicates the limitations of H3 acetylation as a predictive marker. The high stability combined with the low variability between cell lines in project T demonstrates that the marker is able to distinguish roughly between stable and unstable subpopulations, but does not allow for a more precise prediction. For this reason I aimed to identify a general threshold. Project H outperforms project T in terms of the number of cell lines and the variance of expression changes and was therefore chosen for the definition of the threshold. Applying the threshold of 0.5 H3ac to project T led to a significant discrimination of stable and unstable cell lines. However, the suitability of this marker might diminish with the development of more stable cell lines. To this end I compared the average stability of the examined projects H and T with data from other in-house projects (Figure 49). The average stability in project T was noticeably higher than in most other projects. From this I assume that the prediction marker H3ac will also be of great benefit in the current development of CHO cell lines.

	Cell line projects							
	H	T		CT	CI	DF	FI	V
Mean $\Delta qP+$	-0.39	-0.29		-0.45	0.46	-0.30	-0.43	-0.33
Mean $\Delta qP-$	-0.55	-0.41		-0.59	-0.39	-0.40	-0.64	-0.49

Figure 49: Comparison of differences of in-house cell line projects according to their average stability. Stability (ΔqP) was examined over 60 generations in the presence (+) and absence (-) of selection pressure. Project T cell lines were more stable compared to the other projects.

On closer inspection of each examined cell line, there were some which were remarkably stable with H3ac/H3 values below the threshold (false negatives) and others with H3ac/H3 values above the

threshold that were unstable (false positives). The unstable cell lines with high H3ac/H3 values were predominantly observed in absence of selection agent. This leads to the assumption that the alteration of selection condition influences the chromatin stability, which will be discussed in more detail later.

Histone acetylation and deacetylation are highly dynamic processes (Zentner & Henikoff, 2013), which is why the status of H3 acetylation at a certain point in time cannot absolutely anticipate the future evolution of productivity.

One source of error is the fact that our assays measure the average modification of all transgenes of a specific cell line. The same is true for DNA methylation measurement. A population with mixed modification status may gradually shift towards deacetylation. Otherwise, in a stable cell line, strong acetylation of one active transgene can be masked by multiple non-acetylated inactive transgenes. Similarly, DNA methylation of multiple silenced transgenes can conceal one unmethylated active copy. This underestimation of activation signals and the overestimation of repression signals likely results in a decreased predictive power as the number of copies increases. For instance, cell line H-19 is quite stable although it contains ten integrated plasmid copies, H3ac/H3 is very low and DNA methylation is above five percent. Similar effects were observed for cell line H-10 with 50 plasmid copies at the presence of selection agent (Supplement Table 18). Conceivably, a minority of these copies is responsible for most of the production, whereas the majority contributes to the modification status. In this context it may be quite relevant whether all copies are integrated in one locus as tandem repeats or present as several independent integrates. Furthermore, we assume that some false predictions may result from subsequent alterations of plasmid copy number or from posttranscriptional mechanisms (Kim et al., 2011; Li & Rana, 2014).

The variety of different lysines per H3 which can be acetylated might also decrease the discrimination by the marker. The chip-graded antibodies for H3ac recognize all acetylated lysines at the N terminus of histone three. Although H3 can be acetylated on the lysines 4, 9, 14, 18, 23 and 27 in different combinations, it was shown that the activation effect of these residues is in general redundant (Martin et al., 2004). Consequently the acetylation of each residue is predominantly associated with transcription (Dawson & Kouzarides, 2012). Furthermore it was shown recently that the histone acetylation degree remained constant even upon depletion of individual acetyltransferases. This indicates the global role of histone acetylation in the stability of nuclei (Feller et al., 2015). From this I conclude that the general acetylation of H3 is a reliable and consistent mark for activation

4.1.2 Effect of copy number on stability at begin of cultivation phase

It has been demonstrated previously that multi-copy integrates are more susceptible to methylation and silencing (Osterlehner et al., 2011). In addition to the absolute copy number, the different location of copies and the number of tandem repeats influences the expression (Garrick et al., 1998; He et al., 2012; Mauro et al., 2015). The phenomenon of repeat-induced gene silencing was first observed in transgenic organisms as a highly conserved mechanism from plants to mammals. (Rosser & An, 2010). A recent study has shown the association of expanding natural tandem repeats, with DNA methylation and repression of adjacent genes, which suggest that repeat-induced gene silencing (RIGS) also occurs as a natural mechanism (Brahmachary et al., 2014). Cell lines with a high copy number have a naturally increased chance for tandem repeat formation, which in turn increases the possibility for repeat-induced gene silencing. Our studies are consistent with the idea that RIGS is a dominant inducer of transgene silencing in recombinant mammalian cell lines and confirms the value of strategies that aim to avoid multiple tandem integrates.

In addition I propose that the location of the first integrated transgenes influence the copy number variation during selection phase and in consequence the stability of selected cell lines. Taking into account that transfected CHO cells contain the same transgenic sequence, the copy number variation and location of transgenes might be the major triggers. A recent study has asserted that the acetylation in single copy CHO cell lines is associated with the expression of transgenes (Spencer et al. 2015). This leads to the assumption that the site-specific effects are dominant at low copy number. Due to the nature of random integration during transfection, plasmids may integrate in active and less active regions. The subsequent selection pressure forces cell lines with less active transcription of transgenes to compensate their lack of resistance, for example through the amplification of transgenes (Chusainow et al., 2009; Schimke et al., 1978).

Although a strong gene amplification is no longer preferred in the cell line development, a recent study has shown that even a low selection pressure of 250 nM MTX can lead to notably higher copy numbers of control plasmid when compared with a plasmid with ubiquitous chromatin opening elements which are proposed to maintain the open state of chromatin (Betts, Croxford & Dickson, 2015). Furthermore, the copy number variation was frequently found in heterochromatin regions (Aldrich & Maggert, 2014; Hastings et al., 2009). Considering the high selection pressures present during the selection phase, it seems plausible that gene amplification occurs predominantly in less active transgenes, probability accumulated in heterochromatin regions. After the primary selection phase cells are selected by their ability to produce antibody levels above a specific detection threshold. In doing so, cell lines with a low copy number need to be predominately active whereas producers with a high copy number may have a plethora of different epigenetic states that define a

wide range of activities of each individual transgene, provided that the sum of the transcripts exceeds the detection level.

Consequently, the interplay between the activity of the first integration sites and the early selection steps during the cell line development supports the formation of two different cell line types; either predominantly active at low copy number or with tendencies to silencing at high copy number. Hence cell lines with a high copy number have an increased chance of originating from weak producers, which were forced to compensate the evolutionary drawback as a result of the high selection pressure.

Assuming that the moderate selection pressure during cultivation phase still influences and maintain the copy number, those copies should have a greater effect on the stability in the absence of selection agent than in the presence of selection agent. To verify whether this speculation bears substance, the copy number from the beginning of the cultivation phase was plotted against the production stability of cell lines in the presence or absence of selection agent in project H. The predictive power of copy number was slightly better in the absence of selection agent (Figure 32 A and B), which is different from the case of epigenetic marks.

4.1.3 Effect of selection agent on stability

In general the examined cell lines showed an accelerated decline of stability in the absence of selection agent. By contrast, the cell line H-18 showed a different behavior (Table 2). A possible explanation might be a major structural alteration of chromosomes. Given the fact that the expression of transgenes leads to a higher metabolic burden, subpopulations of cells with lower transgene expression are proposed to have an evolutionary benefit. During constant selection pressure this evolutionary benefit is abrogated by the vital need to express sufficient amounts of glutamine synthetase. The subsequent absence of selection agent promotes a selective amplification of subpopulations with susceptibility to silencing depending on their evolutionary benefit. The reduction of transgene mRNA is either promoted by loss of copy number, epigenetic silencing or both (Chusainow et al. 2009; Kim et al. 2011; Paredes et al. 2013). Local epigenetic silencing or activation can influence adjacent or even distant regions (Almouzni and Probst 2011; Cheutin and Cavalli 2014; Vempati et al. 2010; Wang et al. 2014) which might lead to formation of new foci whereas the copy number variation can result into a chromosomal rearrangements, resulting in juxtaposition of previously separated chromosomal regions (Hastings et al. 2009). Considering this I assume that structural changes can lead to exceptional reactivation of previously silenced transgene, which would explain the tremendous increased specific productivity of cell line H-18.

4.1.4 Perspectives

My study proposes that a high plasmid copy number correlates with transgene silencing in recombinant mammalian cell lines. Hence the high plasmid copy number might be induced by location dependent effect of the first integrated transgenes or vice versa. This confirms the value of strategies that aim to avoid tandem integrates and multiple copies in favor of single integrates. Lentiviral vectors (Overbeck et al., 2011), transposase-mediated transfection (Matasci et al. 2011) and targeted integration approaches such as genome editing (Krämer et al., 2010) and recombinase-mediated integration or cassette exchange (Kito et. al., 2002; Zhou et al. 2010; Turan et al., 2011; Crawford et al., 2013) have been applied and discussed as alternatives to random integration of transfected plasmids. The recent development of the CRISPR/Cas9 system for genome editing will offer exciting new opportunities in this field (Mali et al., 2013; Ran et al., 2013). However, there are applications where cell lines containing several copies of transfected plasmids are desirable, e.g. if proteins with several different subunits such as asymmetric, bispecific antibodies need to be expressed (Schäfer et al., 2011). For proper composition of these products the subunits need to be expressed at the optimal molar ratio. This can easily be accomplished by adjusting gene doses if the different subunits are expressed from different plasmids. The plasmids are mixed at the optimal molar ratio and the mixture is transfected into cells. This is usually more convenient than engineering a multi-gene plasmid with each gene under the control of a promoter of distinct strength.

Nevertheless, due to the evolutionary principles within cell cultures silencing remains a serious risk even in improved cell lines. Especially the unstable genome of CHO cells, which has a tendency for chromosome rearrangements and loss (Cao et al. 2012; Derouazi et al. 2006) might promote an accelerated selection of cell lines towards producers, which are stable, primarily under selection pressure. Considering this, the ability to predict stability is important, regardless of enhanced vector design or integration sites. Screening generated cell lines for production stability using a predictive marker can reduce the drawback of instability with this approach.

Histone acetylation is performed by a variety of histone acetyltransferases (Feller et al., 2015). Therefore the localization of those HATs close to the hCMV-MIE might be an even better prediction marker, because of their upstream position in the epigenetic cascade. However, the large number of potential targets makes it extremely difficult to choose a general signal of activation. Furthermore, the depletion of HATs did not result in an altered acetylation pattern (Feller et al., 2015). This obvious redundancy of some lysine acetyltransferases even increases the complexity of target definition.

Considering the fast turnover of histone acetylation at promoters (Zheng, Thomas & Kelleher, 2013), even the observation of a HAT, which is consistently related to gene activity might not increase the predictive power remarkably. Therefore the detection of the general histone acetylation appears to be the most reliable prediction marker in a variable system.

Finally I suggest two modifications to the standard procedure of cell line selection. In the first step cells should be selected based on their production titer followed by the expansion of low copy number cell lines as long as quality and titer are not affected. In the second step the H3 acetylation should be determined once the number of candidate cell lines has been narrowed down to about 20 based on productivity and product quality. 20 cell lines can be evaluated within two days so that timelines are not affected too much. The cell lines that pass the filter would proceed to fermentation evaluation.

4.2 Alteration of epigenetic landscape of CHO-K1 cell lines to circumvent silencing of recombinant genes

Different methods are published to overcome the location-dependent effect by targeted integrations (Inao et al., 2015; Lee et al., 2015; Matasci et al., 2011). Nevertheless the random integration of transfected plasmids is a common method to obtain high titer of recombinant proteins. For this reason I aimed to identify vector modifications, which are able to overrule the location and copy number effects.

4.2.1 Coupled expression of target gene and the histone acetyltransferase MOF to promote open chromatin formation.

To improve long-term stability of different production cell lines I aimed to maintain the open chromatin state at the integration sites. For this purpose a plasmid was designed, which contains a IRES EV 71 coupled expression of the reporter eGFP and the histone acetyltransferase MOF or a control. The histone acetyltransferase MOF was fused to Gal4-BD, a yeast protein which binds to a yeast upstream activation sequence (UAS). This enables the recruitment targeting of UAS sites located upstream of the hCMV-MIE. MOF is well known for his role as a transcription activator and it was shown that the hyperacetylation of H4K16 by MOF results in an open chromatin formation (Corona et al., 2002; Prestel et al., 2010; Schiemann et al., 2010). Furthermore, MOF is the enzymatic subunit of the MSL and the NSL complex. The MSL complex in mammals is the main HAT acetylating H4K16 in murine embryonic stem cells. NSL binds exclusively at promoters mostly of housekeeping

genes and regulates proliferation and cellular homeostasis (Ravens et al., 2014). Therefore I assumed activating effects if MOF is directed to the hCMV-MIE.

CHO pools were generated and the expression of eGFP was measured over three months. The results did not show an enhanced expression or an improved stability in MOF CHO pools. To exclude the possibility that MOF containing cells are only the subpopulation of high eGFP producers which tend to be overgrown, all pools were sorted for high eGFP expression and production stability was recorded. However, the additional sort did not result in enhanced production stability.

Unfortunately, neither the binding of MOF to hCMV-MIE by ChIP nor its general presence by Western blot could be detected. However, the present eGFP signal in combination with the nature of IRES coupled expression strongly suggests the expression of MOF. In addition, a global hyperacetylation within the unsorted MOF pools was observed. Both facts lead to the assumption that MOF is expressed and induces the acetylation of H4K16. From this it can be concluded that a global hyperacetylation of H4K16 did not influence the expression stability of recombinant genes in CHO cells. On the other hand I wasn't able to examine a direct interaction of MOF and hCMV-MIE and a conclusion about target oriented H4K16 acetylation cannot be derived at this point.

Paredes et al. (2013) showed that the general H4 acetylation degrees did not differ between stable and unstable clones, producing IgG. This suggests that the level of general H4ac is not correlated with production stability. Furthermore, it was observed that the H4K16 acetylation level close to the CMV promoter in eGFP expressing transgenic pigs did not correlate with the transgene expression. Even after combined treatment with 5-Aza-dC and TSA no significant alteration of H4K16 acetylation close to CMV promoter could be observed (Yin et al., 2012). In addition it was shown that the functionality of H4K16ac differs according to the model system (Horikoshi et al., 2013). Taken together, the effect of H4K16 acetylation in mammals depends on the cell type and the genomic location. Considering the higher eGFP expression levels of the Gal4-BD control, MOF might perturb balanced acetylation levels and thus lead to a decreased proliferation of the cells that contain the expression vector (Shogren-Knaak et al., 2006).

Although it seems that MOF might not have an activation potential for hCMV-MIE under the chosen conditions other target oriented activators were shown to be effective. Gal4-VP16 fusion protein was described over two decades ago as an potent activator of transcription in CHO cells (Sadowski et al., 1988). VP16 recruits interacting proteins which lead to large scale decondensation of chromatin and stable transcription (Carpenter et al., 2005). Also the overexpression of targeted histone acetyltransferase p300 resulted in higher expression. However, the transgene obtained only a slightly increased stability compared to control (Kwaks et al., 2005). A recent study presents a CRISPR-Cas9-p300core fusion protein, which enables highly specific, targeted activation of

promoters (Hilton et al., 2015). Hence they showed a promising tool for epigenome editing. Complementary to the methods which aimed at hyperacetylation, methods were developed to avoid histone deacetylation. For instance the compound sodium butyrate is well known for his role as histone deacetylase inhibitor. For this reason butyrate was extensively investigated in different cell lines (Choi, 2006; De Leon Gatti et al., 2007; Li & Li, 2006; Lu et al., 2008; Yee et al., 2008) often resulting in a down-regulation of cell cycle proteins and an arrest in the G1 or G2 phase (Hendrick et al., 2001). At a concentration of 1 mM sodium butyrate a decline of cell viability was observed accompanied by a decreased productivity, whereas a lower concentration of 0.5 mM induces a slightly increased productivity. The usage of a general and unspecific inhibitor like butyrate can result in the deregulation of many genes (Birzele et al., 2010).

Considering the variety of effects and the amount of sodium butyrate which will be needed in large batch culture, the usage in the standard cell line project might be complicated.

4.2.2 Point mutation of CpGs within the hCMV-MIE promoter to improve production stability

4.2.2.1 Different effect strength of CpG point mutations between CHO pools and CHO cell lines

To improve transgene expression over time I aimed to diminish repressive chromatin marks by the reduction of methylation-susceptible CpG islands within the promoter. In so doing, I observed notably effects in CHO pools and in CHO cell lines after mutation of cytosines to guanine located 179 and 41 bp upstream of transcription start site of the hCMV-MIE. Especially in clonal CHO cell lines a tremendous effect was observed for the G-179 constructs. By contrast, in CHO pools the highest reporter gene expression was obtained for the G-41 mutation.

However the observed differences in CHO pools were not as significant as observed in CHO cell lines. In contrast to clonal cell lines, CHO pools contain cells with broadly different characteristics after transfection (Mariati et al., 2014). The higher variability and consequently the increased evolutionary pressure in CHO pools might explain the difference between the G-179 and G-41 expression levels as well as the lower difference to control. In turn the effect measurement in clonal cell lines is accompanied by less disturbing variability, which makes the results more reliable. For this reason I propose that the G-179 mutation is more potent than the G-41 mutation.

The notably lower copy number of G-179 constructs in CHO pools might mask the higher activity per each transgene considering that the number of copies influences the expression levels (Mauro et al., 2015; Pruner et al., 2014). At the same time the lower CMV methylation for G-179 constructs (Figure 42 D) suggests a higher activity per transgene. Considering this, the low number of transgenes might be a direct result of the potent G-179 mutation. A functional crosstalk between promoters is well known and can be divided into activating promoter-promoter interaction (Li et al., 2012) and repressive promoter interference. Bidirectional promoter interference is accompanied by an impaired expression of both promoters (Curtin et al., 2008). Promoter interference is a well-studied phenomenon and was also found between the CMV promoter and the Simian virus 40 early promoter (West, 2014). The SV40 originated from polyomavirus and was firstly discovered in Rhesus monkey cells (Sweet & Hilleman, 1960). In my vector, the part that contributes the promoter variants contains the SV40 which is located 11 bp upstream of hCMV-MIE and drives the DHFR expression in antisense direction (Figure 11). Therefore, it seems plausible that the activity of hCMV-MIE might influence the activity of SV40 by positive neighboring effects like a regional decondensation.

As discussed for the prediction marker (4.1.2), I propose that the degree of DHFR expression at constant selection pressure directly influences the evolutionary benefit of each cell. Hence cells with lower expression of DHFR need to overcome this disadvantage by gene amplification, as it was observed in Betts et al., (2015). Following the assumption that G-179-containing CHO pools have an improved activity and stability per transgene as was observed for the cell lines, the lower copy number might be a consequence of lower evolutionary pressure. However in addition to the compensatory effect of gene amplification, the increase of copy numbers correlates with the reduction of transgene mRNA either by loss of copies or by epigenetic silencing mechanisms (Mauro et al., 2015; Osterlehner et al., 2011). I propose that the constant selection pressure decelerates the growth of subpopulations with reduced amount of transgene mRNA.

In conclusion, cells with a higher copy number should have a generally higher doubling time, balancing the adverse loss of transgene mRNA. To this end I performed the Pearson and the Spearman correlation, employing the Jmp10 software (Figure 50). This was done for all eGFP-expressing CHO cell pools at the beginning of the cultivation with the copy number and the doubling time T2 of all cell cultures (Korzynska & Zychowicz, 2008).

The resulting low correlation was significant in the parametric and the non-parametric model. This led to the assumption that high copy number pools partly tend to lose transgene mRNA which in turn results in an overall higher doubling time of the stable high expressers.

Pairwise Correlations															
Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob	-.8	-.6	-.4	-.2	0	.2	.4	.6	.8
Copy number	Doubling time	0.3746	28	0.0017	0.6560	0.0495*									

Nonparametric: Spearman's ρ												
Variable	by Variable	Spearman ρ	Prob> ρ	-.8	-.6	-.4	-.2	0	.2	.4	.6	.8
Copy number	Doubling time	0.5014	0.0066*									

Figure 50: Correlation study of copy number and doubling time of cell culture. Pairwise correlation was calculated with the parametric Pearson test, whereas the Spearman calculation was used for the non-parametric test. Both correlations were low to moderate but significant.

4.2.2.2 Copy number and CMV promoter methylation

With regard to the correlation of copy number and CMV promoter methylation (Figure 42 C and D), I investigated the time-dependent alteration of both markers. For this purpose the methylation and copy number values obtained 42 and 69 days after transfection were plotted for each promoter variant in an outlier boxplot (Figure 51). It could be clearly observed that in general the copy number declines whereas the methylation of hCMV-MIE increases over time. This confirms the association of high copy numbers with the reduction of copies and the increase of DNA methylation (Mauro et al., 2015; Osterlehner et al., 2011). The coincidence of these events led to the assumption that primarily unmethylated transgenes are even more prone for loss of copies by their higher accessibility (Lay et al., 2015).

Following this I propose that G-179 results in a stable expression over time whereas the less stable promoter variants compensate the selection disadvantage with gene amplification. This gene amplification in turn promotes the promoter methylation and a declined copy number. Hence the inverse behavior of both marks might be enforced by a primarily reduction of unmethylated transgenes.

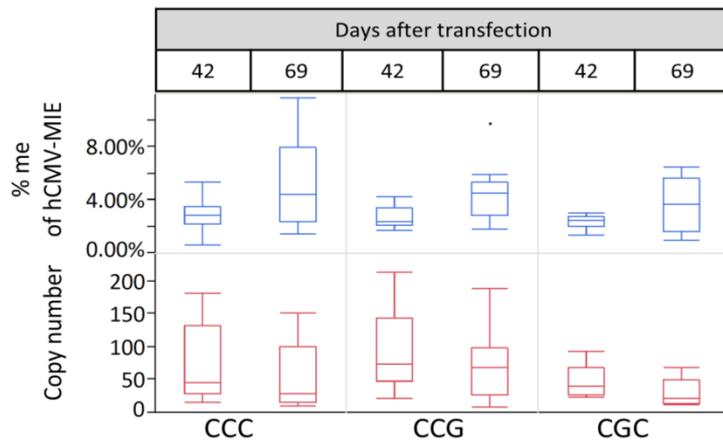


Figure 51: Box plot of the average CMV methylation and transgene copy number of GFP expressing pools. The alteration of both markers were compared for promoter variants CCG (N=10), CGC (N=8) and the control CCC (N=10). Hence a general reduction of copy numbers and an increase of CMV methylation could be observed. The box plot shows the 1st and 3rd quartile indicated by the ends of the box, whereby the horizontal line within the box represents the median sample value. The end of the box has extended lines which represent the whiskers.

Besides the point mutations, no general alteration of the methylation patterns was observed between the promoter variants. However the G-179 construct tended to have a general lower CMV promoter methylation, which can be seen at the other predominately methylated CpGs. This might be either a direct effect of the point mutation or an indirect effect due to the lower number of transgene copies. Control and G-41 pools are highly methylated in particular at position -179 upstream of TSS. This confirms previous results (Osterlehner et al., 2011) and indicates the importance of this specific CpG site within the silencing mechanisms of transgenes.

4.2.2.3 General different effect strength of point mutations

Osterlehner et al (2011) proposed that the degree of methylation of single CpGs is an indicator of stability. Especially the methylation degree of C-179 can be used to predict stability of monoclonal cell lines. It seems plausible that the mutation of the CpG site with the highest susceptibility for methylation also results in the most prominent effect compared to control. In fact, the expression and stability was outstanding in clonal cell lines containing the G-179 constructs. Among the various factors that affect transgene silencing like the copy number, the number of tandem repeats, the sequence and location of the transgene, I propose that G-179 mutation elicits a dominant role.

In addition to the possibility of CpG methylation (Brooks et al., 2004; Kim et al., 2011; Yang et al., 2010a), CpG richness itself may attract protein complexes which promote accumulation of H3K4me3 (Thomson et al., 2010) and H3K27me3 (Mendenhall et al., 2010). Furthermore, a gradually increasing GC skew might promote stable expression as it was observed in highly transcribed prototypical

housekeeping genes (Ginno et al., 2013). It was shown that a general depletion of CpGs around the transcription start site of a promoter can reduce silencing (Swindle, Kim & Klug, 2004). However my results indicate that the combined mutations do not enhance the stability in eGFP cell pools (GGG) nor in antibody clonal cell lines (CGG) at least not to levels that were observed for single mutations, in particular the C to G conversion of the most methylated cytosine located -179 bp upstream of TSS resulted in the most stable cell lines. This illustrates the importance of the methylation degree at each individual CpG site as it was proposed previously (Nile et al., 2008). Furthermore, I propose that the underlying epigenetic mechanism is highly sensible for alterations of each CpG and precise adjustment of the promoter methylation pattern might be more effective than a general depletion of CpGs.

4.2.2.4 Presumable influence of G-179 mutation in the cell line development after screening for high producers.

The comparison of CHO pools and clonal cell lines indicate that factors like copy number can influence the expression to compensate lower activity per transgene even at low selection pressure (Figure 42). Depending on the intensities of the selection pressure this effect might be also seen in clonal cell lines, which will reduce the direct benefit of G-179 mutation over unmutated control. The additional screening step for high producers will further reduce the direct advantage of the promoter variant. However, these selection steps only mask the advantage of G-179 by enforcing the accumulation of unmutated producers with higher copy numbers. In large batch cultures the selection agent will be removed and consequently the evolutionary pressure which keeps unstable producers stable disappears. From the previous results I presume that the G-179 constructs might lead to more stable producers with a lower number of copies after screening.

4.3 Epigenetic landscape of CHO-K1-M cells and potential stable integration sites

4.3.1 Location of integration sites of randomly transfected CHO-K1 pools

The mapping of histone modification with the broad peak modus of MACS2 led to a coverage of 8.2% for H3K36me3 and 6.7% for H3K9me3 all over the CHO-K1 genome. However more peaks were detected for H3K9me3 with 150496 peaks compared to the 46214 peaks of the H3K36me3. This implies that the H3K36me3 modification covers larger areas than H3K9me3 and has broader peaks. For this reason the peak counting around integration sites must be considered with caution and the

inspection of single integration sites and the appropriate chromatin marks was essential to estimate the activity of the regions. The analysis of the H3K9me3/H3K36me3 ratios was done to enable a generalized statement about predominantly active and inactive regions. For this reason the statistical ratio of 3.25x H3K9me3 per H3K36me3 peaks was calculated by the inclusion of all mapped H3K9me3 and H3K36me3 peaks and compared to the ratios obtained for peaks +/- 6500 bps around of integration sites. Hence a general increase of H3K9me3 close to the integration sites could be observed (Table 28). From this I propose that the integration or that the following gene amplification is not truly random and occurs frequently in H3K9me3 dense regions (Aldrich & Maggert, 2014). To enable stable integration, plasmids were enzymatically linearized. The resulting non-homologous ends of the linearized plasmids should be consequently integrated by the non-homologous end joining mechanism (NHEJ). From this it was proposed that in euchromatin the NHEJ preferably occurs in inactive H3K36me3 depleted regions whereas homologous recombination (HR) was predominantly found in actively transcribed regions (Aymard et al., 2014). For this reason the integration might preferably occur in accessible but inactive euchromatin regions, whereupon the moderate activity might be compensated by gene amplification and subsequent heterochromatinization.

Without threshold settings for the integration sites, the histone modification ratios were higher than the statistical ratio but similar between high eGFP and low eGFP expressing cell pools. Considering that the low eGFP expressing pools should mainly correspond to inactive integration sites, the relatively high proportion of H3K36me3 indicates the presence a second class of active integration sites. The heterogeneity lead to an implausible modification ratio. For this reason, integration sites were sorted for their read coverage and those sites with less than 100x coverage were excluded. As a result, integration sites of the eGFP negative sorted pools, obtained a notably higher ratio compared to the positive sorted pools. The fraction of H3K9me3 was twice as much as was calculated for the general statistical ratio.

Furthermore the counting of all integration sites showed that the high eGFP expressing pools (except 4.6), yielded more integration sites than the low eGFP expressing pools. This seems plausible considering that a loss of copy number is associated with a declined productivity (Osterlehner et al., 2011). The higher eGFP intensity observed in the 4.6 (-) pool might explain the contrasty high number of integration sites within this sort. For comparison, the copy number per pool was determined as described previously. Hence the same tendencies were observed within each pool sort (1.7 +/-) as was seen for all unsorted integration sites. This suggests that the most of the integration sites are not unspecific background reads. However most of the integration sites are infrequently as it is indicated by the coverage. Those integration sites might not have a great

influence on the general expression level of eGFP, which would explain the similar H3K9me3/H3K36me3 ratios although the eGFP expression levels between low and high eGFP expressing pools were different.

pool identification	Positive sorted eGFP pools			Negative sorted eGFP pools		
	1.7 (+)	2.3 (+)	4.6 (+)	1.7 (-)	2.3 (-)	4.6 (-)
# of Integration sites	8452	1855	1621	976	502	1945
# of Integration sites >100x coverage	19	4	11	4	13	39
copy number per cell	50	13	5	4	5	9

Table 30: Number of integration sites per pool before and after sort for integration sites with a coverage of higher than 100x. Copy number of transgenes per cell was calculated. The same trend between negative and positive sorted pools was seen for copy number as well as for all integration sites.

4.3.2 Discrimination of active and inactive integration sites

The ratio H3K9me3/H3K36me3 for the unfiltered integration sites (coverage > 0) was similar between high eGFP and low eGFP expressing pools (Table 28). This illustrates the challenge of defining the right and plausible threshold for the integration sites. Similar to “omics” studies the small number of samples and the high number of analytes can result in a high number of false positive or negatives (Dietmair, Nielsen & Timmins, 2012). However, all mapped integration sites might be specific as the similarity between the alteration of copy numbers and the number of unfiltered integration sites within one pool sort suggests. But under the chosen experimental conditions I was not able to correlate any single integration site to the appropriate expression stability. However the simultaneous investigation of many different cells enables the identification of overall histone modification patterns which describe general associations between expression and location of transgenes.

Following the assumption that the integration sites that were covered most are also the most prominent integration sites, the three single integration sites per pool with the highest coverage were chosen for the visual analyses of the adjacent region (Figure 48). This way the background integration sites were excluded as well as those integration sites which might only have minor contributions to the overall transgene expression. For all low eGFP expressing pools the most prominent single integration sites were surrounded by the inactivation mark H3K9me3. This seems plausible considering that the absence of eGFP signal follows the complete silencing of all transgenes. The most prominent integration sites of the high eGFP expressing pools are surrounded by H3K36me3 in five out of nine cases. The Pool 1.7 with the clearest difference between the high and low eGFP expressing subpopulations (Figure 46) also showed the strongest association of active

chromatin status and integration sites that support transgene expression (Figure 48). In contrast, the 2.3 (+) in one case and in particular the high eGFP expressing pool 4.6 (+) revealed integration sites with repressive H3K9me3 marks. However both pools neither reached the same eGFP expression levels nor a comparable sharpness as was seen for the high eGFP expressing pool 1.7. Considering this I propose that the broader peaks correspond to a higher variability of cells with different integration sites. In conclusion the highest coverage might not belong to most stable integration site within the variable pools. In addition, not all the integration sites of high eGFP expressing pools are necessarily in active regions as long as some integration sites are active. Hence the chance that the prominent integration sites of high eGFP expressing pools support stable expression, increases with the number of adjacent active marks.

4.3.3 Perspective

My preliminary results lead me to propose an influence of the chromatin state on the stability of the integration site. To confirm this hypothesis I suggest two different approaches. Initially and directly I suggest to explore the targeted integration of transgenes expressing the reporter eGFP by employing the CRISPR/Cas9 technology (Lee et al., 2015). If the results are promising the assay should be repeated with an antibody reporter.

For this purpose the integration site must be well chosen. Many of the aligned integration sites were found in small contigs of the CHO-K1 genome, closer than 6,500 bps to one contig end or in the absence of any histone modification. Therefore the chosen filter settings exclude many potentially stable integration sites. For this reason I reviewed whether the three highest covered integration sites per pool are also highly covered compared to all possible integration sites. I found that in the high eGFP expressing pool 1.7 (+), the highest covered integration site close to histone modifications in contig gi|529983619|gb|APMK01090275.1| was ranked place four according to the coverage of all integration sites (Table 31). As a consequence both sites should be investigated as potentially active ones. The integration site of low eGFP expressing pool 1.7 (-) within the contig gi|529840071|gb|APMK01159973.1| is surrounded by heterochromatin and therefore is suitable for use as a negative control. From this approach I expect production stabilities and expression degrees which confirm the selection of the chosen locations.

contig	start	end	cov	chromatin state	expression
gi 529983619 gb APMK01090275.1	11339	11522	1640	active	yes
gi 529537223 gb APMK01302182.1	2250	2473	3106	no mark	yes
gi 529840071 gb APMK01159973.1	24537	12592	1160	inactive	no

Table 31: Integration sites or target oriented integration. First integration site is encircled by active histone marks in an eGFP expressing pool. The second integration sites is the highest covered site, but without any surrounding marks. The third negative sorted integration site is in a heterochromatin region and can be used as negative control.

In the second approach the identification of integration sites of random integrations after transfection of CHO-K1 cells should be repeated. At least five replications should be investigated to increase the significance. Furthermore, the eGFP sorting window should be tightly focused to promote the development of a well-defined subpopulation, comparable to the high eGFP expressing pool 1.7 (+). From this I expect the reduction of background integration sites and facilitated analyses of stable and unstable locations. The extension of cultivation time, followed by resorting will also support the formation of an isolated subpopulation and is strongly recommended.

5 Appendix

5.1 Supplement tables

Cell line	me of C-179 in %	Δ qP (60 generations) in %		Plasmid copies/cell	H3			H3ac/H3			H3K4me3/H3			H3K27me3/H3		
					biol. replicate			biol. replicate			biol. replicate			biol. replicate		
					1	2	3	1	2	3	1	2	3	1	2	3
	PSB	+MSX	-MSX	PSB												
H-1	15.7	-46.80	-80.20	168.8	54.44	53.69	59.58	0.19	0.01	0.03	0.02	0.02	0.03	1.35	1.72	1.28
H-2	9.9	-71.00	-70.30	2.0	2.41	1.08	1.55	0.17	0.11	0.35	0.19	0.19	0.38	3.11	2.07	5.53
H-3	16.5	-73.10	-89.60	3.2	2.08	2.15	2.27	0.21	0.12	222.35	1.02	1.75	0.35	2.36	2.17	4.88
H-4	10.5	-69.30	-69.50	2.9	2.71	2.54	2.23	0.36	0.25	0.09	0.28	0.05	0.09	7.08	3.29	2.62
H-5	11.0	-40.40	-48.80	1.8	2.09	2.02	2.21	0.11	0.04	0.39	0.33	0.05	0.07	5.92	2.24	7.33
H-6	7.1	-79.50	-71.70	37.6	27.60	32.60	30.63	0.01	0.06	0.02	0.03	0.01	0.02	1.11	2.64	2.04
H-7	4.8	-66.20	-81.50	53.0	44.84	44.12	49.98	0.01	0.02	0.03	0.02	0.03	0.02	1.21	2.43	2.03
H-8	4.4	-67.40	-75.70	30.6	37.44	28.77	36.50	0.02	0.03	0.04	0.01	0.03	0.10	1.57	1.56	1.84
H-9	23.4	-87.00	-77.00	95.1	67.96	60.83	60.13	0.01	0.05	0.02	0.01	0.04	0.06	1.71	2.91	2.16
H-10	26.2	-16.40	-73.30	48.8	46.31	45.05	36.50	0.01	0.12	0.02	0.01	0.09	0.04	1.31	3.97	1.51
H-11	0.2	-43.10	-96.10	3.3	0.21	9.43	0.54	0.70	1.48	1.82	15.62	1.40	3.47	4.30	10.14	13.78
H-12	0.1	-42.00	-33.00	1.5	0.11	4.95	0.19	2.41	2.22	2.42	6.54	3.72	1.44	7.34	17.07	8.28
H-13	1.3	8.00	-18.50	1.1	1.14	4.90	0.35	0.21	0.98	2.23	0.29	3.92	9.21	9.92	103.97	16.00
H-14	0.0	6.80	5.40	1.3	0.12	8.04	0.25	2.29	1.19	2.98	2.46	4.12	1.16	200.16	15.33	2.65
H-15	0.1	-16.50	-40.40	3.5	0.29	9.83	0.45	0.85	1.33	1.87	0.40	0.97	1.47	2.44	3.71	38.23
H-16	0.4	16.20	-32.80	3.0	0.12	6.12	0.50	1.14	2.12	1.84	1.88	3.93	4.85	25.78	7.35	12.35
H-17	1.0	-13.20	-31.30	4.2	0.42	5.19	0.29	0.59	1.84	4.31	1.47	2.15	0.56	7.03	8.33	4.06
H-18	46.3	-67.50	319.40	8.8	3.08	6.98	2.65	0.37	2.28	0.17	0.21	1.88	0.12	2.64	11.38	2.17
H-19	16.7	-11.20	2.70	10.1	4.44	6.93	4.36	0.13	0.85	0.09	0.19	1.13	0.15	2.51	3.67	2.27
H-20	21.8	-40.30	-80.60	4.8	0.61	3.82	0.32	0.38	1.55	1.20	0.70	4.14	1.78	22.21	43.51	1.62

Table 32: Data of normalized histone3 and histone3 modification, DNA methylation, plasmid number and stability of 20 CHO cell lines of project H. Transgene LC and HC copy numbers are comparable. LC transgene copy number was used for the effect measurements and is here described as plasmid copies/cell. Highlighted cells are described in results or discussion.

Cell line	me of C-179 in %	Δ qP (60 generations) in %		Plasmid copies/cell	H3			H3ac/H3			H3K4me3/H3			H3K27me3/H3			H3K9me3/H3		
					biol. replicate			biol. replicate			biol. replicate			biol. replicate			biol. replicate		
					1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
PSB	+MSX	-MSX	PSB																
T-1	0.44	-35.40	-34.20	4.4	0.70	0.38	0.33	0.27	0.19	0.47	0.09	0.22	0.15	8.30	14.04	19.74	151.52	19.38	20.92
T-2	2.20	-30.60	-72.00	11.5	0.98	1.38	1.16	0.17	0.12	1.40	0.09	0.05	0.05	11.75	3.16	6.13	8.72	4.62	8.66
T-3	3.54	-81.60	-96.60	9.4	0.87	0.61	0.77	0.20	0.24	0.21	0.12	0.13	0.38	6.48	6.30	16.60	8.03	3.66	5.66
T-4	0.92	-23.40	-47.40	4.6	0.48	0.27	0.41	0.17	0.47	1.79	0.15	0.19	0.15	17.19	45.25	195.81	23.86	530.06	94.35
T-5	4.41	-15.60	-22.20	5.3	0.11	0.06	0.40	0.99	2.66	0.59	0.18	0.32	0.28	38.23	6.62	16.49	321.42	13.96	10.27
T-6	0.47	-14.40	-20.40	2.6	0.26	0.09	0.36	0.51	1.24	0.33	0.12	0.23	0.49	25.99	7.62	44.74	14.83	4.29	34.78
T-7	0.22	-16.20	-31.80	1.6	0.51	0.58	0.88	0.32	0.31	0.11	0.13	0.03	0.06	193.56	7.43	11.63	9.96	4.78	7.98
T-8	52.40	-64.80	-79.20	64.3	265.64	186.97	212.31	0.00	0.00	0.00	0.00	0.00	0.00	0.63	0.67	0.96	1.62	1.56	1.74
T-9	52.84	-27.60	-49.20	67.5	142.68	91.14	83.87	0.00	0.00	0.00	0.00	0.00	0.00	0.40	0.54	1.03	0.98	1.81	1.84
T-10	49.61	3.60	1.20	70.5	82.71	109.39	75.06	0.00	0.00	0.01	0.00	0.00	0.00	0.85	0.43	1.21	1.79	1.51	1.47
T-11	2.02	-10.20	-21.60	4.6	3.46	5.07	3.57	0.06	0.02	0.06	0.01	0.01	0.01	5.19	0.96	2.14	156.86	3.67	1.90
T-12	61.34	-32.40	-22.80	11.5	15.60	13.77	11.06	0.01	0.01	0.02	0.00	0.00	0.01	0.87	1.78	2.07	1.59	3.00	2.70

Table 33: Data of normalized histone3 and histone3 modification, DNA methylation, plasmid number and stability of twelve CHO cell lines of project T. Transgene LC and HC copy numbers are comparable. LC transgene copy number was used for the effect measurements and is here described as plasmid copies/cell

5.2 Abbreviations

A	Adenine
Amp	Ampicillin
BSA	Bovine Serum Albumin
C	Cytosine
ChIP	Chromatin Immunoprecipitation
CHO	Chinese Hamster Ovary
CHO-K1 cell	Chinese Hamster Ovary cell line K1
CMV	Cytomegalovirus-Promoter
CpG	Cytosine-phosphat-Guanine
DAC	5-Aza-2-Deoxycytidine
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic Acid
DNMT	DNA Methyltransferase
dNTP	Deoxynucleotidethriphosphate
DTT	Dithiothreitol
EDTA	Ethylendiaminetetraacetate
eGFP	enhanced Green Fluorescent Protein
EGTA	Ethylenglycol-bis(2-aminethyl)-N,N,N,N-tetraacetic acid
Eif3i	Eukaryotic translation initiation factor 3, subunit I
EtBr	Ethidiumbromide
F	Forward
Foxa2	Forkhead box A2
Gata5	Gata binding protein 5
Gusb	Glucuronidase, beta
G	Guanine
h	hour
H2A	Histone 2 A
H3	Histone 3
H3ac	acetylated Histone 3
H3K4me3	trimethylated lysine 4 at Histone 3
H3K9me	trimethylated lysine 9 at Histone 3
H3K27me3	trimethylated lysine 27 at Histone 3
HAT	Histone Acetyltransferases
HDAC	Histone Deacetylase
HEPES	(N-(2-Hydroxyethyl)piperazine-H ⁺)-(2-ethanesulfonic acid)
HMT	Histone Methyltransferase
HP1	Heterochromatin protein 1
HRP	Horseradish Peroxidase
Ig	Immunoglobulin
IP	Immunoprecipitation
kb	Kilobase

kDA	Kilo Dalton
M	Molar
min	minute
ml	milliliter
mM	Millimolar
MNase	Micrococcal Nuclease
mRNA	Messenger RNA
OD	Optical Density
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PcG	Polycomb Group
PCR	Polymerase Chain Reaction
PEST	Proline (P), Glutamic acid (E), Serine (S), and Threonine (T) rich
PFA	Paraformaldehyde
PMSF	Phenylmethanesulfonyl Fluoride
PTM	Posttranslational Modification
PVDF	Polyvinylidene Fluoride
qP	specific Productivity (pg/cell/day)
Δ qP	percentual alteration of qP
rCN	relative Copy Number
R	Reverse
Rho	Rhodopsin
RNA	Ribonucleic acid
RT	Room Temperature
SDS	Sodiumdodecylsulfate
T	Thymine
Temed	N,N,N,N-Tetramethylethylenediamine
UAS	Upstream Activating Sequence
UHRF1	Ubiquitin-like plant Homeodomain and RING Finger domain 1
Unc13c	Unc-13 homolog C (C. elegans), transcript variant X1
UV	Ultraviolet
V	Volt
v/v	Volume per volume
WT	Wild Type
w/v	Weight per volume
μ g	Microgramm
μ l	Microliter
μ M	Micromolar

5.3 Table of figures

Figure 1: Organization of chromatin:.....	13
Figure 2: Epigenetic mechanisms involved in chromatin modifications.	15
Figure 3: Mechanisms of H3K4 methyltransferase targeting specific genes exemplified by the MLL family.....	18
Figure 4: Exemplified modification of histone N-tails.....	20
Figure 5: Distribution of CpGs.....	23
Figure 6: Methylation and demethylation of cytosine.	24
Figure 7: Position-effect variegation.....	26
Figure 8: Silencing by polycomb complexes.	28
Figure 9: Crosstalk of epigenetic pathways.	29
Figure 10: Comparison of relative copy number of transgene and the calculated copy number by standard curve.....	46
Figure 11: Plasmid design for measuring hCMV-MIE promoter strength by expression of secretory alkaline phosphatase (SEAP).....	47
Figure 12: Cloning strategy for transfection vectors	47
Figure 13: Overview of the QuikChange Multi Site-Directed Mutagenesis method.	49
Figure 14: Sonification efficiency test.....	51
Figure 15: Schema of inverse PCR.....	54
Figure 16: 5' iPCR of stable CHO pools after eGFP sort.....	56
Figure 17: Electropherogram of one sample including the proximal and distal amplicons of bisulfite converted hCMV-MIE.....	60
Figure 18: Amaxa Shuttle 96 well plate for transient transfection of CHO-K1 cell suspension.....	67
Figure 19: Procedure of dilution of cell culture supernatant	68
Figure 20: FACS: living gate of CHO-K1 cells.....	69
Figure 21: FACS: histogram of eGFP expressing CHO cell population.....	69
Figure 22: ELISA principle with POD conjugated detection antibody.....	71
Figure 23: Secreted alkaline phosphatase (SEAP) expressing plasmid.	71
Figure 24: Enhanced Green Fluorescent Protein (eGFP) expressing plasmid.....	72
Figure 25: Enhanced Green Fluorescent Protein (eGFP) expressing plasmid with MOFi.....	72
Figure 26: Human antibody of class IgG expressing plasmid.....	73
Figure 27: SEQ ID NO 1. Human CMV major immediate-early promoter/enhancer fragment.....	73
Figure 28: Exemplified identification of stable integration sites.....	78
Figure 29: hCMV-MIE sequence.....	80
Figure 30: Antibody validation within project T and H in one sample each.....	84
Figure 31: Jackknife outlier analysis of project H.	88
Figure 32: Correlation studies of log transformed plasmid copy number and epigenetic effects on Δ qP with and without selection agent.....	91
Figure 33: Interplay of epigenetic effects and the plasmid copy number.....	92
Figure 34: Application of filter settings to exclude low producers.....	93
Figure 35: Transferred threshold settings from project H to project T.	94
Figure 36: Histogram of fluorescence intensity (x-axis) of CHO pools, 52 days after transfection.....	96
Figure 37: Geometrical means of FACS measurements over time.	97
Figure 38: Western blot of the nuclear fraction of three independent MOF and control CHO pools .	98

Figure 39: Histogram of fluorescence intensity (x-axis) of eGFP positive sorted CHO pools.	98
Figure 40: hCMV-MIE: Effects of C to G conversions.	100
Figure 41: Transient transfection of CHO cells with SEAP expression plasmids under the control of hCMV-MIE promoter variants.	101
Figure 42: eGFP expression of stable transfected CHO cell pools.	104
Figure 43: hCMV-MIE methylation pattern of CpG's.	105
Figure 44: Generation of stable transfected clonal CHO cell lines.	108
Figure 45: Enrichment of histone modification on specific genomic regions.	110
Figure 46: Subcultures of eGFP expressing CHO-K1 pools.	112
Figure 47: 5`iPCR of stable CHO pools after eGFP sort.	113
Figure 48: IGV genome browser pictures of the three highest covered integration sites.	117
Figure 49: Comparison of differences of in-house cell line projects according to their average stability.	120
Figure 50: Correlation study of copy number and doubling time of cell culture.	129
Figure 51: Box plot of the average CMV methylation and transgene copy number of GFP expressing pools.	130

5.4 Eidesstattliche Versicherung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt ist.

München, den

.....

(Benjamin Moritz)

5.5 Erklärung

Hiermit erkläre ich, dass ich mich anderweitig einer Doktorprüfung ohne Erfolg nicht unterzogen habe und dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.

München, den

.....

(Benjamin Moritz)

5.6 Curriculum vitae

Benjamin Moritz

Birth name: Moritz
Date of birth: July 2nd 1982
Place of birth: Waldbröl

Education

- 11/2011 – present **PhD thesis at the Ludwig-Maximilians-University of Munich**
Center for Integrated Protein Science, Munich, Germany &
Roche Innovation Center Penzberg, Germany
Professor Dr. Peter Becker, Group of Dr. Ulrich Goepfert
Project: Prevention and prediction of production instability of CHO-K1 cell
lines by the examination of epigenetic mechanisms.
- 11/2010 – 07/2011 **Master thesis at the Technical University of Munich**
Institute of Experimental Genetics
Professor Dr. Martin Hrabě de Angelis, Advisor Dr. Gerhard Przemecik
Project: Overexpression of Delta-like 1 with and without PDZ binding motif
in the Insulinoma cell line INS-1E
- 10/2009 – 07/2011 **Master studies in molecular biology at the Technical University of Munich**
- 05/2009 – 07/2009 **Bachelor thesis at the University of Konstanz**
Professor Dr. Elke Deuerling
Project: Characterization of NAC (nascent polypeptide-associated complex)
chaperone– ribosome interaction.
- 10/2006 – 07/2009 **Bachelor studies in biology at the University of Konstanz**

Publications

MORITZ, B., WOLTERING, L., Becker, P.B., Goepfert, U. (2015). Histone Mark Enables Rejection of Unstable Chinese Hamster Ovary Cell Lines. *Biotechnology and Bioengineering*. *Manuscript Id 15-525, Under review.*

EZAZ, T., MORITZ, B., WATERS, P., GRAVES, J. A. M., GEORGES, A. & SARRE, S. D. (2009). The ZW sex microchromosomes of an Australian dragon lizard share no homology with those of other reptiles or birds. *Chromosome Research* **17**, 965-973.

Patents

Moritz, B., Goepfert, U., Becker, P.B., 2014. Method for the selection of a long-term producing cell using histone acylation as marker. EP14168896.0, filed May 2014. Patent Pending

Moritz, B., Goepfert, U., 2014. Method for production of polypeptides. EP14168901.8, filed May 2014. Patent Pending

5.7 Acknowledgements

Diese Arbeit wurde von Oktober 2011 bis Juni 2015 am Roche Innovation Center Penzberg von Roche Pharmaceutical Research and Early Development angefertigt.

Zuallererst möchte ich meinem Betreuer Dr. Ulrich Göpfert für die Gelegenheit danken, dass ich das spannende Forschungsfeld der Epigenetik in Produktionszelllinien näher beleuchten konnte. Ich danke dir dafür, dass du mir die Freiräume gelassen hast, mich als Wissenschaftler selbstständig zu entwickeln und mir doch gleichzeitig an kritischen Punkten mit deinem Rat zur Seite standest. Ich hätte mir keinen besseren Betreuer für meine Arbeit wünschen können.

Im Besonderen möchte ich mich bei Prof. Dr. Peter Becker bedanken, der sich immer Zeit für meine Forschungsarbeit genommen hat. Vielen Dank, dass du so viel Interesse für dieses industrienaher Projekt gezeigt hast und ich immer das Gefühl hatte, willkommen zu sein. Danke für deine Ideen und kritischen Anmerkungen sowie deine große Expertise, die mich stets dazu angetrieben hat, etwas mehr zu lernen.

Danke auch an Dr. Silke Hansen, die es mir ermöglichte in ihrer Abteilung zu arbeiten. Vielen Dank für dein Vertrauen in mich und für deine stets offene Tür.

Herzlichen Dank auch an meine Kollegen Carina Post, Andrea Osterlehner, Laura Woltering und Tobias Scherzinger. Die guten Gespräche und die freundschaftliche Atmosphäre haben mir über so manche Durststrecke hinweg geholfen. Eure Unterstützung mit Rat und Tat haben viele meiner Projekte erst möglich gemacht.

Ferner möchte ich mich bei Dr. Florian Lipsmeier, Dr. Fabian Birzele und Dr. Matthias Barann bedanken, die mir in den biostatistischen und bioinformatischen Fragestellungen immer unter die Arme gegriffen haben. Ich weiß, dass es nicht immer leicht war und umso mehr bin ich euch dankbar für die viele Zeit, die ihr geopfert habt.

Auch allen anderen in der Abteilung möchte ich danken für die 3 ½ spannenden Jahre. Für die guten Ratschläge oftmals zwischen Tür und Angel und doch so wertvoll. Danke für die gemeinsamen Kaffees, die Essen und Seminare. Ich habe mich sehr wohl bei euch gefühlt.

Allen Dank an meine Eltern Elke und Jürgen Moritz. Ich danke euch, dass ihr immer für mich da wart, mich unterstützt und motiviert habt. Ich danke euch, für euch.

Am meisten möchte ich meiner Frau Claudia danken. Du bist der Schinken meines Lebens und die Quelle meiner Energie. Danke, dass du mein zeitraubendes Hobby und mich so tolerierst. Danke dafür, dass ich an deiner Seite sein darf.

6 Bibliography

- ABDI, H. (2010). Tukey's Honestly Significant Difference (HSD) Test. *Encyclopedia of Research Design*.
- AKHTAR, A. & BECKER, P. B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in *Drosophila*. *Mol Cell* **5**, 367-75.
- AKHTAR, W., DE JONG, J., PINDYURIN, A. V., PAGIE, L., MEULEMAN, W., DE RIDDER, J., BERNS, A., WESSELS, L. F. A., VAN LOHUIZEN, M. & VAN STEENSEL, B. (2013). Chromatin Position Effects Assayed by Thousands of Reporters Integrated in Parallel. *Cell* **154**, 914-927.
- ALBERTS B, J. A., LEWIS J, ET AL. . (2002). Molecular Biology of the Cell. Chromosomal DNA and Its Packaging in the Chromatin Fiber. . **4th edition**.
- ALDRICH, J. C. & MAGGERT, K. A. (2014). Simple quantitative PCR approach to reveal naturally occurring and mutation-induced repetitive sequence variation on the *Drosophila* Y chromosome. *PLoS One* **9**, e109906.
- ALLAN, J., RAU, D. C., HARBORNE, N. & GOULD, H. (1984). Higher order structure in a short repeat length chromatin. *J Cell Biol* **98**, 1320-7.
- ALLAN, R. S., ZUEVA, E., CAMMAS, F., SCHREIBER, H. A., MASSON, V., BELZ, G. T., ROCHE, D., MAISON, C., QUIVY, J. P., ALMOUZNI, G. & AMIGORENA, S. (2012). An epigenetic silencing pathway controlling T helper 2 cell lineage commitment. *Nature* **487**, 249-53.
- ALLFREY, V. G., FAULKNER, R. & MIRSKY, A. E. (1964). Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A* **51**, 786-94.
- ALMOUZNI, G. & PROBST, A. V. (2011). Heterochromatin maintenance and establishment: lessons from the mouse pericentromere. *Nucleus* **2**, 332-8.
- ALTAMIRANO, C., BERRIOS, J., VERGARA, M. & BECERRA, S. (2013). Advances in improving mammalian cells metabolism for recombinant protein production. *Electronic Journal of Biotechnology* **16**.
- AMBROS, V. & CHEN, X. (2007). The regulation of genes and genomes by small RNAs. *Development* **134**, 1635-41.
- APARICIO, O. M. & GOTTSCHLING, D. E. (1994). Overcoming telomeric silencing: a trans-activator competes to establish gene expression in a cell cycle-dependent way. *Genes Dev* **8**, 1133-46.
- ARAVIN, A. A., SACHIDANANDAM, R., GIRARD, A., FEJES-TOTH, K. & HANNON, G. J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**, 744-7.
- ARENTS, G., BURLINGAME, R. W., WANG, B. C., LOVE, W. E. & MOUDRIANAKIS, E. N. (1991). The Nucleosomal Core Histone Octamer at 3.1-a Resolution - a Tripartite Protein Assembly and a Left-Handed Superhelix. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 10148-10152.
- ARENTS, G. & MOUDRIANAKIS, E. N. (1993). Topography of the histone octamer surface: repeating structural motifs utilized in the docking of nucleosomal DNA. *Proc Natl Acad Sci U S A* **90**, 10489-93.
- AVERY, O. T., MACLEOD, C. M. & MCCARTY, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *J Exp Med* **79**, 137-58.
- AYMARD, F., BUGLER, B., SCHMIDT, C. K., GUILLOU, E., CARON, P., BRIOIS, S., IACOVONI, J. S., DABURON, V., MILLER, K. M., JACKSON, S. P. & LEGUBE, G. (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat Struct Mol Biol* **21**, 366-74.
- BACHER, C. P., GUGGIARI, M., BRORS, B., AUGUI, S., CLERC, P., AVNER, P., EILS, R. & HEARD, E. (2006). Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat Cell Biol* **8**, 293-9.
- BAILEY, L. A., HATTON, D., FIELD, R. & DICKSON, A. J. (2012). Determination of Chinese hamster ovary cell line stability and recombinant antibody expression during long-term culture. *Biotechnol Bioeng* **109**, 2093-103.
- BAKER, L. A., ALLIS, C. D. & WANG, G. G. (2008). PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. *Mutat Res* **647**, 3-12.
- BANNISTER, A. J., SCHNEIDER, R. & KOUZARIDES, T. (2002). Histone methylation: dynamic or static? *Cell* **109**, 801-6.
- BARLOW, D. P. (2011). Genomic imprinting: a mammalian epigenetic discovery model. *Annu Rev Genet* **45**, 379-403.
- BARLOW, D. P. & BARTOLOMEI, M. S. (2014). Genomic imprinting in mammals. *Cold Spring Harb Perspect Biol* **6**.

- BARNES, L. M., BENTLEY, C. M. & DICKSON, A. J. (2001). Characterization of the stability of recombinant protein production in the GS-NS0 expression system. *Biotechnol Bioeng* **73**, 261-70.
- BARNES, L. M., BENTLEY, C. M. & DICKSON, A. J. (2003). Stability of protein production from recombinant mammalian cells. *Biotechnol Bioeng* **81**, 631-9.
- BARNES, L. M., BENTLEY, C. M. & DICKSON, A. J. (2004). Molecular definition of predictive indicators of stable protein expression in recombinant NS0 myeloma cells. *Biotechnol Bioeng* **85**, 115-21.
- BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T.-Y., SCHONES, D. E., WANG, Z., WEI, G., CHEPELEV, I. & ZHAO, K. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**, 823-837.
- BARTH, T. K. & IMHOF, A. (2010). Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem Sci* **35**, 618-26.
- BENEKE, S., MEYER, K., HOLTZ, A., HUTTNER, K. & BURKLE, A. (2012). Chromatin composition is changed by poly(ADP-ribose)ylation during chromatin immunoprecipitation. *PLoS One* **7**, e32914.
- BERNSTEIN, E. & HAKE, S. B. (2006). The nucleosome: a little variation goes a long way. *Biochem Cell Biol* **84**, 505-17.
- BETTS, Z., CROXFORD, A. S. & DICKSON, A. J. (2015). Evaluating the interaction between UCOE and DHFR-linked amplification and stability of recombinant protein expression. *Biotechnol Prog*.
- BIRD, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21.
- BIRZELE, F., SCHAUB, J., RUST, W., CLEMENS, C., BAUM, P., KAUFMANN, H., WEITH, A., SCHULZ, T. W. & HILDEBRANDT, T. (2010). Into the unknown: expression profiling without genome sequence information in CHO by next generation sequencing. *Nucleic acids research* **38**, 3999-4010.
- BONISCH, C. & HAKE, S. B. (2012). Histone H2A variants in nucleosomes and chromatin: more or less stable? *Nucleic Acids Res* **40**, 10719-41.
- BORRELLI, E., NESTLER, E. J., ALLIS, C. D. & SASSONE-CORSI, P. (2008). Decoding the epigenetic language of neuronal plasticity. *Neuron* **60**, 961-74.
- BOSHART, M., WEBER, F., JAHN, G., DORSCHHASLER, K., FLECKENSTEIN, B. & SCHAFFNER, W. (1985). A Very Strong Enhancer Is Located Upstream of an Immediate Early Gene of Human Cytomegalo-Virus. *Cell* **41**, 521-530.
- BOSTICK, M., KIM, J. K., ESTEVE, P. O., CLARK, A., PRADHAN, S. & JACOBSEN, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**, 1760-4.
- BRAHMACHARY, M., GUILMATRE, A., QUILEZ, J., HASSON, D., BOREL, C., WARBURTON, P. & SHARP, A. J. (2014). Digital genotyping of macrosatellites and multicopy genes reveals novel biological functions associated with copy number variation of large tandem repeats. *PLoS Genet* **10**, e1004418.
- BRANDEIS, M., FRANK, D., KESHET, I., SIEGFRIED, Z., MENDELSON, M., NEMES, A., TEMPER, V., RAZIN, A. & CEDAR, H. (1994). Sp1 elements protect a CpG island from de novo methylation. *Nature* **371**, 435-8.
- BROOKS, A. R., HARKINS, R. N., WANG, P., QIAN, H. S., LIU, P. & RUBANYI, G. M. (2004). Transcriptional silencing is associated with extensive methylation of the CMV promoter following adenoviral gene delivery to muscle. *J Gene Med* **6**, 395-404.
- BROUHA, B., SCHUSTAK, J., BADGE, R. M., LUTZ-PRIGGE, S., FARLEY, A. H., MORAN, J. V. & KAZAZIAN, H. H., JR. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* **100**, 5280-5.
- BROWN, S. E. & SZYF, M. (2007). Epigenetic programming of the rRNA promoter by MBD3. *Mol Cell Biol* **27**, 4938-52.
- BROWNELL, J. E. & ALLIS, C. D. (1996). Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Curr Opin Genet Dev* **6**, 176-84.
- BUSTIN, S. A., BENES, V., GARSON, J. A., HELLEMANS, J., HUGGETT, J., KUBISTA, M., MUELLER, R., NOLAN, T., PFAFFL, M. W., SHIPLEY, G. L., VANDESOMPELE, J. & WITTEWER, C. T. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**, 611-22.
- CABURET, S., CONTI, C., SCHURRA, C., LEBOFKY, R., EDELSTEIN, S. J. & BENSIMON, A. (2005). Human ribosomal RNA gene arrays display a broad range of palindromic structures. *Genome Res* **15**, 1079-85.
- CARMELL, M. A., GIRARD, A., VAN DE KANT, H. J., BOURC'HIS, D., BESTOR, T. H., DE ROOIJ, D. G. & HANNON, G. J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* **12**, 503-14.
- CARPENTER, A. E., MEMEDULA, S., PLUTZ, M. J. & BELMONT, A. S. (2005). Common effects of acidic activators on large-scale chromatin structure and transcription. *Mol Cell Biol* **25**, 958-68.
- CASTEL, S. E. & MARTIENSSON, R. A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nature Reviews Genetics* **14**, 100-112.

- CEDAR, H. & BERGMAN, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* **10**, 295-304.
- CHAMPAGNE, K. S. & KUTATELADZE, T. G. (2009). Structural insight into histone recognition by the ING PHD fingers. *Curr Drug Targets* **10**, 432-41.
- CHAPMAN, B. S., THAYER, R. M., VINCENT, K. A. & HAIGWOOD, N. L. (1991). Effect of Intron-a from Human Cytomegalovirus (Towne) Immediate-Early Gene on Heterologous Expression in Mammalian-Cells. *Nucleic Acids Research* **19**, 3979-3986.
- CHEN, T. & DENT, S. Y. (2014). Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet* **15**, 93-106.
- CHEUNG, P., ALLIS, C. D. & SASSONE-CORSI, P. (2000). Signaling to chromatin through histone modifications. *Cell* **103**, 263-71.
- CHEUTIN, T. & CAVALLI, G. (2014). Polycomb silencing: from linear chromatin domains to 3D chromosome folding. *Curr Opin Genet Dev* **25**, 30-7.
- CHEUTIN, T., GORSKI, S. A., MAY, K. M., SINGH, P. B. & MISTELI, T. (2004). In vivo dynamics of Swi6 in yeast: evidence for a stochastic model of heterochromatin. *Mol Cell Biol* **24**, 3157-67.
- CHOI, Y. H. (2006). Apoptosis of U937 human leukemic cells by sodium butyrate is associated with inhibition of telomerase activity. *Int J Oncol* **29**, 1207-13.
- CHONG, L. (2001). Molecular cloning - A laboratory manual, 3rd edition. *Science* **292**, 446-446.
- CHUSAINOW, J., YANG, Y. S., YEO, J. H., TOH, P. C., ASVADI, P., WONG, N. S. & YAP, M. G. (2009). A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer? *Biotechnol Bioeng* **102**, 1182-96.
- CLAPIER, C. R. & CAIRNS, B. R. (2009). The biology of chromatin remodeling complexes. *Annu Rev Biochem* **78**, 273-304.
- CLEMSON, C. M., MCNEIL, J. A., WILLARD, H. F. & LAWRENCE, J. B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol* **132**, 259-75.
- CONCONI, A., WIDMER, R. M., KOLLER, T. & SOGO, J. M. (1989). Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. *Cell* **57**, 753-61.
- CORONA, D. F., CLAPIER, C. R., BECKER, P. B. & TAMKUN, J. W. (2002). Modulation of ISWI function by site-specific histone acetylation. *EMBO Rep* **3**, 242-7.
- COULONDRE, C., MILLER, J. H., FARABAUGH, P. J. & GILBERT, W. (1978). Molecular basis of base substitution hotspots in Escherichia coli. *Nature* **274**, 775-80.
- CURTIN, J. A., DANE, A. P., SWANSON, A., ALEXANDER, I. E. & GINN, S. L. (2008). Bidirectional promoter interference between two widely used internal heterologous promoters in a late-generation lentiviral construct. *Gene Ther* **15**, 384-90.
- DAMBACHER, S., HAHN, M. & SCHOTTA, G. (2013). The compact view on heterochromatin. *Cell Cycle* **12**, 2925-2926.
- DAWSON, M. A. & KOUZARIDES, T. (2012). Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12-27.
- DE CARVALHO, D. D., YOU, J. S. & JONES, P. A. (2010). DNA methylation and cellular reprogramming. *Trends Cell Biol* **20**, 609-17.
- DE LEON GATTI, M., WLASCHIN, K. F., NISSOM, P. M., YAP, M. & HU, W. S. (2007). Comparative transcriptional analysis of mouse hybridoma and recombinant Chinese hamster ovary cells undergoing butyrate treatment. *J Biosci Bioeng* **103**, 82-91.
- DE RUIJTER, A. J., VAN GENNIP, A. H., CARON, H. N., KEMP, S. & VAN KUILENBURG, A. B. (2003). Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem J* **370**, 737-49.
- DEATON, A. M. & BIRD, A. (2011). CpG islands and the regulation of transcription. *Genes Dev* **25**, 1010-22.
- DENISSOV, S., LESSARD, F., MAYER, C., STEFANOVSKY, V., VAN DRIEL, M., GRUMMT, I., MOSS, T. & STUNNENBERG, H. G. (2011). A model for the topology of active ribosomal RNA genes. *EMBO Rep* **12**, 231-7.
- DI CROCE, L. & HELIN, K. (2013). Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol* **20**, 1147-55.
- DIETMAIR, S., NIELSEN, L. K. & TIMMINS, N. E. (2012). Mammalian cells as biopharmaceutical production hosts in the age of omics. *Biotechnology Journal* **7**, 75-89.
- DULAC, C. (2010). Brain function and chromatin plasticity. *Nature* **465**, 728-35.
- EARLEY, K., LAWRENCE, R. J., PONTES, O., REUTHER, R., ENCISO, A. J., SILVA, M., NEVES, N., GROSS, M., VIEGAS, W. & PIKAARD, C. S. (2006). Erasure of histone acetylation by Arabidopsis HDA6 mediates large-scale gene silencing in nucleolar dominance. *Genes Dev* **20**, 1283-93.

- EFRONI, S., DUTTAGUPTA, R., CHENG, J., DEGHANI, H., HOEPFNER, D. J., DASH, C., BAZETT-JONES, D. P., LE GRICE, S., MCKAY, R. D., BUETOW, K. H., GINGERAS, T. R., MISTELI, T. & MESHORER, E. (2008). Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2**, 437-47.
- EPSZTEJN-LITMAN, S., FELDMAN, N., ABU-REMAILEH, M., SHUFARO, Y., GERSON, A., UEDA, J., DEPLUS, R., FUKS, F., SHINKAI, Y., CEDAR, H. & BERGMAN, Y. (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nat Struct Mol Biol* **15**, 1176-83.
- ESNAULT, C., MAESTRE, J. & HEIDMANN, T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* **24**, 363-7.
- FADLOUN, A., LE GRAS, S., JOST, B., ZIEGLER-BIRLING, C., TAKAHASHI, H., GORAB, E., CARNINCI, P. & TORRES-PADILLA, M. E. (2013). Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat Struct Mol Biol* **20**, 332-8.
- FAULKNER, G. J., KIMURA, Y., DAUB, C. O., WANI, S., PLESSY, C., IRVINE, K. M., SCHRODER, K., CLOONAN, N., STEPTOE, A. L., LASSMANN, T., WAKI, K., HORNIG, N., ARAKAWA, T., TAKAHASHI, H., KAWAI, J., FORREST, A. R. R., SUZUKI, H., HAYASHIZAKI, Y., HUME, D. A., ORLANDO, V., GRIMMOND, S. M. & CARNINCI, P. (2009). The regulated retrotransposon transcriptome of mammalian cells. *Nature Genetics* **41**, 563-571.
- FELLER, C., FORNE, I., IMHOF, A. & BECKER, P. B. (2015). Global and specific responses of the histone acetylome to systematic perturbation. *Mol Cell* **57**, 559-71.
- FENG, S., JACOBSEN, S. E. & REIK, W. (2010). Epigenetic reprogramming in plant and animal development. *Science* **330**, 622-7.
- FESTENSTEIN, R. & CHAN, J. P. K. (2012). Context is everything: activators can also repress. *Nature Structural & Molecular Biology* **19**, 973-975.
- FILIPPAKOPOULOS, P., PICAUD, S., MANGOS, M., KEATES, T., LAMBERT, J. P., BARSYTE-LOVEJOY, D., FELLETER, I., VOLKMER, R., MULLER, S., PAWSON, T., GINGRAS, A. C., ARROWSMITH, C. H. & KNAPP, S. (2012). Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**, 214-31.
- FOECKING, M. K. & HOFSTETTER, H. (1986). Powerful and versatile enhancer-promoter unit for mammalian expression vectors. *Gene* **45**, 101-5.
- FRANCIS, N. J., FOLLMER, N. E., SIMON, M. D., AGHIA, G. & BUTLER, J. D. (2009). Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. *Cell* **137**, 110-22.
- GAO, Z. H., LIU, H. L., DAXINGER, L., PONTES, O., HE, X. J., QIAN, W. Q., LIN, H. X., XIE, M. T., LORKOVIC, Z. J., ZHANG, S. D., MIKI, D., ZHAN, X. Q., PONTIER, D., LAGRANGE, T., JIN, H. L., MATZKE, A. J. M., MATZKE, M., PIKAARD, C. S. & ZHU, J. K. (2010). An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* **465**, 106-U118.
- GARRICK, D., FIERING, S., MARTIN, D. I. & WHITELAW, E. (1998). Repeat-induced gene silencing in mammals. *Nat Genet* **18**, 56-9.
- GASIOR, S. L., WAKEMAN, T. P., XU, B. & DEININGER, P. L. (2006). The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol* **357**, 1383-93.
- GASZNER, M. & FELSENFELD, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* **7**, 703-13.
- GHILDIAL, M., SEITZ, H., HORWICH, M. D., LI, C., DU, T., LEE, S., XU, J., KITTLER, E. L., ZAPP, M. L., WENG, Z. & ZAMORE, P. D. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science* **320**, 1077-81.
- GHILDIAL, M. & ZAMORE, P. D. (2009). Small silencing RNAs: an expanding universe. *Nat Rev Genet* **10**, 94-108.
- GHOSHAL, K., MAJUMDER, S., DATTA, J., MOTIWALA, T., BAI, S., SHARMA, S. M., FRANKEL, W. & JACOB, S. T. (2004). Role of human ribosomal RNA (rRNA) promoter methylation and of methyl-CpG-binding protein MBD2 in the suppression of rRNA gene expression. *J Biol Chem* **279**, 6783-93.
- GIL, J. & O'LOGHLEN, A. (2014). PRC1 complex diversity: where is it taking us? *Trends Cell Biol* **24**, 632-641.
- GINNO, P. A., LIM, Y. W., LOTT, P. L., KORF, I. & CHEDIN, F. (2013). GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Research* **23**, 1590-1600.
- GIRARD, A., SACHIDANANDAM, R., HANNON, G. J. & CARMELL, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199-202.
- GOZANI, O., KARUMAN, P., JONES, D. R., IVANOV, D., CHA, J., LUGOVSKOY, A. A., BAIRD, C. L., ZHU, H., FIELD, S. J., LESSNICK, S. L., VILLASENOR, J., MEHROTRA, B., CHEN, J., RAO, V. R., BRUGGE, J. S., FERGUSON, C. G., PAYRASTRE, B., MYSZKA, D. G., CANTLEY, L. C., WAGNER, G., DIVECHA, N., PRESTWICH, G. D. & YUAN, J. (2003). The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. *Cell* **114**, 99-111.
- GRAFF, J. & MANSUY, I. M. (2008). Epigenetic codes in cognition and behaviour. *Behav Brain Res* **192**, 70-87.

- GREAVES, I. K., RANGASAMY, D., RIDGWAY, P. & TREMETHICK, D. J. (2007). H2A.Z contributes to the unique 3D structure of the centromere. *Proc Natl Acad Sci U S A* **104**, 525-30.
- GRUMMT, I. & LANGST, G. (2013). Epigenetic control of RNA polymerase I transcription in mammalian cells. *Biochim Biophys Acta* **1829**, 393-404.
- GU, B. & LEE, M. G. (2013). Histone H3 lysine 4 methyltransferases and demethylases in self-renewal and differentiation of stem cells. *Cell Biosci* **3**, 39.
- GUALTIERI, A., ANDREOLA, F., SCIAMANNA, I., SINIBALDI-VALLEBONA, P., SERAFINO, A. & SPADAFORA, C. (2013). Increased expression and copy number amplification of LINE-1 and SINE B1 retrotransposable elements in murine mammary carcinoma progression. *Oncotarget* **4**, 1882-93.
- GUETG, C., LIENEMANN, P., SIRRI, V., GRUMMT, I., HERNANDEZ-VERDUN, D., HOTTIGER, M. O., FUSSENEGGER, M. & SANTORO, R. (2010). The NoRC complex mediates the heterochromatin formation and stability of silent rRNA genes and centromeric repeats. *EMBO J* **29**, 2135-46.
- HAKE, S. B. & ALLIS, C. D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". *Proc Natl Acad Sci U S A* **103**, 6428-35.
- HASHIMOTO, H., PAIS, J. E., ZHANG, X., SALEH, L., FU, Z. Q., DAI, N., CORREA, I. R., JR., ZHENG, Y. & CHENG, X. (2014). Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* **506**, 391-5.
- HASTINGS, P. J., LUPSKI, J. R., ROSENBERG, S. M. & IRA, G. (2009). Mechanisms of change in gene copy number. *Nat Rev Genet* **10**, 551-64.
- HE, L., WINTERROWD, C., KADURA, I. & FRYE, C. (2012). Transgene copy number distribution profiles in recombinant CHO cell lines revealed by single cell analyses. *Biotechnol Bioeng* **109**, 1713-22.
- HE, Y. F., LI, B. Z., LI, Z., LIU, P., WANG, Y., TANG, Q., DING, J., JIA, Y., CHEN, Z., LI, L., SUN, Y., LI, X., DAI, Q., SONG, C. X., ZHANG, K., HE, C. & XU, G. L. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-7.
- HENDRICK, V., WINNEPENNINGKX, P., ABDELKAFI, C., VANDEPUTTE, O., CHERLET, M., MARIQUE, T., RENEMANN, G., LOA, A., KRETZMER, G. & WERENNE, J. (2001). Increased productivity of recombinant tissular plasminogen activator (t-PA) by butyrate and shift of temperature: a cell cycle phases analysis. *Cytotechnology* **36**, 71-83.
- HENIKOFF, S. (1990). Position-effect variegation after 60 years. *Trends Genet* **6**, 422-6.
- HENIKOFF, S. (1996). Dosage-dependent modification of position-effect variegation in *Drosophila*. *Bioessays* **18**, 401-9.
- HERMANN, A., GOYAL, R. & JELTSCH, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem* **279**, 48350-9.
- HILTON, I. B., D'IPPOLITO, A. M., VOCKLEY, C. M., THAKORE, P. I., CRAWFORD, G. E., REDDY, T. E. & GERSBACH, C. A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* **33**, 510-7.
- HODAWADEKAR, S. C. & MARMORSTEIN, R. (2007). Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. *Oncogene* **26**, 5528-40.
- HORIKOSHI, N., KUMAR, P., SHARMA, G. G., CHEN, M., HUNT, C. R., WESTOVER, K., CHOWDHURY, S. & PANDITA, T. K. (2013). Genome-wide distribution of histone H4 Lysine 16 acetylation sites and their relationship to gene expression. *Genome Integr* **4**, 3.
- HU, L., LI, Z., CHENG, J., RAO, Q., GONG, W., LIU, M., SHI, Y. G., ZHU, J., WANG, P. & XU, Y. (2013). Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* **155**, 1545-55.
- ILLINGWORTH, R. S., GRUENEWALD-SCHNEIDER, U., WEBB, S., KERR, A. R., JAMES, K. D., TURNER, D. J., SMITH, C., HARRISON, D. J., ANDREWS, R. & BIRD, A. P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* **6**, e1001134.
- INAO, T., KAWABE, Y., YAMASHIRO, T., KAMEYAMA, Y., WANG, X., ITO, A. & KAMIHIRA, M. (2015). Improved transgene integration into the Chinese hamster ovary cell genome using the Cre-loxP system. *J Biosci Bioeng*.
- INTERNATIONAL HUMAN GENOME SEQUENCING, C. (2004). Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-45.
- ITO, S., SHEN, L., DAI, Q., WU, S. C., COLLINS, L. B., SWENBERG, J. A., HE, C. & ZHANG, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-3.
- JAENISCH, R. & BIRD, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**, 245-54.
- JAYAPAL, K. P. & WLASCHIN, K. F. (2007). Recombinant Protein Therapeutics from CHO Cells - 20 Years and Counting.

- JELTSCH, A. (2006). On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. *Epigenetics* **1**, 63-6.
- JEON, Y. & LEE, J. T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146**, 119-33.
- JIN, Q., YU, L. R., WANG, L., ZHANG, Z., KASPER, L. H., LEE, J. E., WANG, C., BRINDLE, P. K., DENT, S. Y. & GE, K. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J* **30**, 249-62.
- JONES, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**, 484-92.
- KALB, R., LATWIEL, S., BAYMAZ, H. I., JANSEN, P. W., MULLER, C. W., VERMEULEN, M. & MULLER, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol* **21**, 569-71.
- KALKHOVEN, E. (2004). CBP and p300: HATs for different occasions. *Biochem Pharmacol* **68**, 1145-55.
- KASSCHAU, K. D., FAHLGREN, N., CHAPMAN, E. J., SULLIVAN, C. M., CUMBIE, J. S., GIVAN, S. A. & CARRINGTON, J. C. (2007). Genome-wide profiling and analysis of Arabidopsis siRNAs. *Plos Biology* **5**, 479-493.
- KAUFMAN, R. J., SHARP, P. A. & LATT, S. A. (1983). Evolution of chromosomal regions containing transfected and amplified dihydrofolate reductase sequences. *Mol Cell Biol* **3**, 699-711.
- KAUFMAN, W. L., KOCMAN, I., AGRAWAL, V., RAHN, H. P., BESSER, D. & GOSSEN, M. (2008). Homogeneity and persistence of transgene expression by omitting antibiotic selection in cell line isolation. *Nucleic Acids Res* **36**, e111.
- KIM, J., DANIEL, J., ESPEJO, A., LAKE, A., KRISHNA, M., XIA, L., ZHANG, Y. & BEDFORD, M. T. (2006). Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep* **7**, 397-403.
- KIM, J. Y., KIM, Y. G. & LEE, G. M. (2012). CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl Microbiol Biotechnol* **93**, 917-30.
- KIM, M., O'CALLAGHAN, P. M., DROMS, K. A. & JAMES, D. C. (2011). A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnol Bioeng* **108**, 2434-46.
- KIMURA, H. & COOK, P. R. (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J Cell Biol* **153**, 1341-53.
- KIMURA, H., HIEDA, M. & COOK, P. R. (2004). Measuring histone and polymerase dynamics in living cells. *Methods Enzymol* **375**, 381-93.
- KLOSE, R. J. & BIRD, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* **31**, 89-97.
- KLOSE, R. J. & ZHANG, Y. (2007). Regulation of histone methylation by demethylination and demethylation. *Nat Rev Mol Cell Biol* **8**, 307-18.
- KOBAYASHI, T. (2008). A new role of the rDNA and nucleolus in the nucleus--rDNA instability maintains genome integrity. *Bioessays* **30**, 267-72.
- KORZYNSKA, A. & ZYCHOWICZ, M. (2008). A Method of Estimation of the Cell Doubling Time on Basis of the Cell Culture Monitoring Data. *Biocybernetics and Biomedical Engineering* **28**, 75-82.
- KOUZARIDES, T. (2007). Chromatin modifications and their function. *Cell* **128**, 693-705.
- KRAFT, M., CIRSTEAN, I. C., VOSS, A. K., THOMAS, T., GOEHRING, I., SHEIKH, B. N., GORDON, L., SCOTT, H., SMYTH, G. K., AHMADIAN, M. R., TRAUTMANN, U., ZENKER, M., TARTAGLIA, M., EKICI, A., REIS, A., DORR, H. G., RAUCH, A. & THIEL, C. T. (2011). Disruption of the histone acetyltransferase MYST4 leads to a Noonan syndrome-like phenotype and hyperactivated MAPK signaling in humans and mice. *J Clin Invest* **121**, 3479-91.
- KRIAUCIONIS, S. & HEINTZ, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929-30.
- KRUEGER, F. & ANDREWS, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-2.
- KWAKS, T. H., SEWALT, R. G., VAN BLOKLAND, R., SIERSMA, T. J., KASIEM, M., KELDER, A. & OTTE, A. P. (2005). Targeting of a histone acetyltransferase domain to a promoter enhances protein expression levels in mammalian cells. *J Biotechnol* **115**, 35-46.
- LANDAUER, K. (2014). Designing media for animal cell culture: CHO cells, the industrial standard. *Methods Mol Biol* **1104**, 89-103.
- LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, N., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N.,

- COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., McMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.
- LANDT, S. G., MARINOV, G. K., KUNDAJE, A., KHERADPOUR, P., PAULI, F., BATZOGLOU, S., BERNSTEIN, B. E., BICKEL, P., BROWN, J. B., CAYTING, P., CHEN, Y., DESALVO, G., EPSTEIN, C., FISHER-AYLOR, K. I., EUSKIRCHEN, G., GERSTEIN, M., GERTZ, J., HARTEMINK, A. J., HOFFMAN, M. M., IYER, V. R., JUNG, Y. L., KARMAKAR, S., KELLIS, M., KHARCHENKO, P. V., LI, Q., LIU, T., LIU, X. S., MA, L., MILOSAVLJEVIC, A., MYERS, R. M., PARK, P. J., PAZIN, M. J., PERRY, M. D., RAHA, D., REDDY, T. E., ROZOWSKY, J., SHORESH, N., SIDOW, A., SLATTERY, M., STAMATOYANNOPOULOS, J. A., TOLSTORUKOV, M. Y., WHITE, K. P., XI, S., FARNHAM, P. J., LIEB, J. D., WOLD, B. J. & SNYDER, M. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**, 1813-31.
- LANGMEAD, B. & SALZBERG, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-U54.
- LANGST, G., BECKER, P. B. & GRUMMT, I. (1998). TTF-I determines the chromatin architecture of the active rDNA promoter. *EMBO J* **17**, 3135-45.
- LANGST, G., BLANK, T. A., BECKER, P. B. & GRUMMT, I. (1997). RNA polymerase I transcription on nucleosomal templates: the transcription termination factor TTF-I induces chromatin remodeling and relieves transcriptional repression. *EMBO J* **16**, 760-8.
- LATOS, P. A., PAULER, F. M., KOERNER, M. V., SENERGIN, H. B., HUDSON, Q. J., STOCSITS, R. R., ALLHOFF, W., STRICKER, S. H., KLEMENT, R. M., WARCZOK, K. E., AUMAYR, K., PASIERBEK, P. & BARLOW, D. P. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* **338**, 1469-72.
- LATTENMAYER, C., LOESCHEL, M., STEINFELLNER, W., TRUMMER, E., MUELLER, D., SCHRIEBL, K., VORAUER-UHL, K., KATINGER, H. & KUNERT, R. (2006). Identification of transgene integration loci of different highly expressing recombinant CHO cell lines by FISH. *Cytotechnology* **51**, 171-82.
- LAW, J. A., VASHISHT, A. A., WOHLSCHLEGEL, J. A. & JACOBSEN, S. E. (2011). SHH1, a Homeodomain Protein Required for DNA Methylation, As Well As RDR2, RDM4, and Chromatin Remodeling Factors, Associate with RNA Polymerase IV. *Plos Genetics* **7**.
- LAWRENCE, R. J., EARLEY, K., PONTES, O., SILVA, M., CHEN, Z. J., NEVES, N., VIEGAS, W. & PIKAARD, C. S. (2004). A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance. *Mol Cell* **13**, 599-609.
- LAY, F. D., LIU, Y., KELLY, T. K., WITT, H., FARNHAM, P. J., JONES, P. A. & BERMAN, B. P. (2015). The role of DNA methylation in directing the functional organization of the cancer epigenome. *Genome Res* **25**, 467-77.
- LEE, J. S., KALLEHAUGE, T. B., PEDERSEN, L. E. & KILDEGAARD, H. F. (2015). Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci Rep* **5**, 8572.
- LEE, J. T. & BARTOLOMEI, M. S. (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* **152**, 1308-23.
- LEE, K. K. & WORKMAN, J. L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. *Nat Rev Mol Cell Biol* **8**, 284-95.
- LEVER, M. A., TH'NG, J. P., SUN, X. & HENDZEL, M. J. (2000). Rapid exchange of histone H1.1 on chromatin in living human cells. *Nature* **408**, 873-6.
- LI, E., BESTOR, T. H. & JAENISCH, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-26.
- LI, G., RUAN, X., AUERBACH, R. K., SANDHU, K. S., ZHENG, M., WANG, P., POH, H. M., GOH, Y., LIM, J., ZHANG, J., SIM, H. S., PEH, S. Q., MULAWADI, F. H., ONG, C. T., ORLOV, Y. L., HONG, S., ZHANG, Z., LANDT, S., RAHA, D., EUSKIRCHEN, G., WEI, C. L., GE, W., WANG, H., DAVIS, C., FISHER-AYLOR, K. I., MORTAZAVI, A., GERSTEIN, M., GINGERAS, T., WOLD, B., SUN, Y., FULLWOOD, M. J., CHEUNG, E., LIU, E., SUNG, W. K., SNYDER, M. & RUAN, Y. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84-98.
- LI, H. & DURBIN, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60.
- LI, J., LANGST, G. & GRUMMT, I. (2006). NoRC-dependent nucleosome positioning silences rRNA genes. *EMBO J* **25**, 5735-41.

- LI, R. W. & LI, C. (2006). Butyrate induces profound changes in gene expression related to multiple signal pathways in bovine kidney epithelial cells. *BMC Genomics* **7**, 234.
- LI, Z. & RANA, T. M. (2014). Therapeutic targeting of microRNAs: current status and future challenges. *Nat Rev Drug Discov* **13**, 622-38.
- LIANG, G., CHAN, M. F., TOMIGAHARA, Y., TSAI, Y. C., GONZALES, F. A., LI, E., LAIRD, P. W. & JONES, P. A. (2002). Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Mol Cell Biol* **22**, 480-91.
- LIU, X., WANG, L., ZHAO, K., THOMPSON, P. R., HWANG, Y., MARMORSTEIN, R. & COLE, P. A. (2008). The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. *Nature* **451**, 846-50.
- LIVAK, K. J. & SCHMITTGEN, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* **25**, 402-408.
- LU, R., WANG, X., SUN, D. F., TIAN, X. Q., ZHAO, S. L., CHEN, Y. X. & FANG, J. Y. (2008). Folic acid and sodium butyrate prevent tumorigenesis in a mouse model of colorectal cancer. *Epigenetics* **3**, 330-5.
- LUCCHINI, R. & SOGO, J. M. (1995). Replication of transcriptionally active chromatin. *Nature* **374**, 276-80.
- LUGER, K. & HANSEN, J. C. (2005). Nucleosome and chromatin fiber dynamics. *Curr Opin Struct Biol* **15**, 188-96.
- MACLEOD, D., CHARLTON, J., MULLINS, J. & BIRD, A. P. (1994). Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. *Genes Dev* **8**, 2282-92.
- MARGUERON, R. & REINBERG, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-9.
- MARIATI, YEO, J. H., KOH, E. Y., HO, S. C. & YANG, Y. (2014). Insertion of core CpG island element into human CMV promoter for enhancing recombinant protein expression stability in CHO cells. *Biotechnol Prog* **30**, 523-34.
- MARTIN, A. M., POUCHNIK, D. J., WALKER, J. L. & WYRICK, J. J. (2004). Redundant roles for histone H3 N-terminal lysine residues in subtelomeric gene repression in *Saccharomyces cerevisiae*. *Genetics* **167**, 1123-32.
- MATASCI, M., BALDI, L., HACKER, D. L. & WURM, F. M. (2011). The PiggyBac transposon enhances the frequency of CHO stable cell line generation and yields recombinant lines with superior productivity and stability. *Biotechnol Bioeng* **108**, 2141-50.
- MAURO, J. A., BUTLER, S. N., RAMSAMOOJ, M. & BLANCK, G. (2015). Copy number loss or silencing of apoptosis-effector genes in cancer. *Gene* **554**, 50-7.
- McMANUS, K. J. & HENDZEL, M. J. (2003). Quantitative analysis of CBP- and P300-induced histone acetylations in vivo using native chromatin. *Mol Cell Biol* **23**, 7611-27.
- MENDENHALL, E. M., KOCHER, R. P., TRUONG, T., ZHOU, V. W., ISSAC, B., CHI, A. S., KU, M. & BERNSTEIN, B. E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet* **6**, e1001244.
- MERCER, T. R. & MATTICK, J. S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol* **20**, 300-7.
- MEYERS, R., A. (2012a). Epigenetic Regulation and Epigenomics. *WILEY-BLACKWELL* **1**.
- MEYERS, R., A. (2012b). Epigenetic Regulation and Epigenomics. *WILEY-BLACKWELL*, 524.
- MEYERS, R., A. (2013). Epigenetic Regulation and Epigenomics. *WILEY-BLACKWELL* **1**.
- MILLER, K. M., MAAS, N. L. & TOCZYSKI, D. P. (2006). Taking it off: regulation of H3 K56 acetylation by Hst3 and Hst4. *Cell Cycle* **5**, 2561-5.
- MISAWA Y. (2006). Western blot analysis of sub-cellular fractionated samples using the Odyssey Infrared Imaging System.
- MOHN, F., WEBER, M., REBHAN, M., ROLOFF, T. C., RICHTER, J., STADLER, M. B., BIBEL, M. & SCHUBELER, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* **30**, 755-66.
- MOSAMMAPARAST, N. & SHI, Y. (2010). Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. *Annu Rev Biochem* **79**, 155-79.
- MOUSE GENOME SEQUENCING, C., WATERSTON, R. H., LINDBLAD-TOH, K., BIRNEY, E., ROGERS, J., ABRIL, J. F., AGARWAL, P., AGARWALA, R., AINSCOUGH, R., ALEXANDERSSON, M., AN, P., ANTONARAKIS, S. E., ATTWOOD, J., BAERTSCH, R., BAILEY, J., BARLOW, K., BECK, S., BERRY, E., BIRREN, B., BLOOM, T., BORK, P., BOTCHERBY, M., BRAY, N., BRENT, M. R., BROWN, D. G., BROWN, S. D., BULT, C., BURTON, J., BUTLER, J., CAMPBELL, R. D., CARNINCI, P., CAWLEY, S., CHIAROMONTE, F., CHINWALLA, A. T., CHURCH, D. M., CLAMP, M., CLEE, C., COLLINS, F. S., COOK, L. L., COPLEY, R. R., COULSON, A., COURONNE, O., CUFF, J., CURWEN, V., CUTTS, T., DALY, M., DAVID, R., DAVIES, J., DELEHAUNTY, K. D., DERI, J., DERMITZAKIS, E. T., DEWEY, C., DICKENS, N. J., DIEKHANS, M., DODGE, S., DUBCHAK, I., DUNN, D. M., EDDY, S. R., ELNITSKI, L., EMES, R. D., ESWARA, P., EYRAS, E., FELSENFELD, A., FEWELL, G. A., FLICEK, P., FOLEY, K., FRANKEL, W. N., FULTON, L. A., FULTON, R. S., FUREY, T. S., GAGE, D., GIBBS, R. A., GLUSMAN, G., GNERRE, S., GOLDMAN, N., GOODSTADT, L., GRAFHAM, D., GRAVES, T. A., GREEN, E. D., GREGORY, S., GUIGO, R., GUYER, M., HARDISON, R. C., HAUSSLER, D., HAYASHIZAKI, Y., HILLIER, L. W., HINRICHS, A., HLAVINA, W., HOLZER, T., HSU, F.,

- HUA, A., HUBBARD, T., HUNT, A., JACKSON, I., JAFFE, D. B., JOHNSON, L. S., JONES, M., JONES, T. A., JOY, A., KAMAL, M., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-62.
- MOZZETTA, C., PONTIS, J., FRITSCH, L., ROBIN, P., PORTOSO, M., PROUX, C., MARGUERON, R. & AIT-SI-ALI, S. (2014). The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. *Mol Cell* **53**, 277-89.
- MUJTABA, S., ZENG, L. & ZHOU, M. M. (2007). Structure and acetyl-lysine recognition of the bromodomain. *Oncogene* **26**, 5521-7.
- MULLER, H. J. (1930). TYPES OF VISIBLE VARIATIONS INDUCED BY X-RAYS IN DROSOPHILA. *Journal of genetics* **III**.
- MURR, R. (2010). Interplay Between Different Epigenetic Modifications and Mechanisms. *Epigenetics and Cancer, Pt A* **70**, 101-141.
- MUTSKOV, V. & FELSENFELD, G. (2004). Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J* **23**, 138-49.
- MYANT, K. & STANCHEVA, I. (2008). LSH cooperates with DNA methyltransferases to repress transcription. *Mol Cell Biol* **28**, 215-26.
- MYANT, K., TERMANIS, A., SUNDARAM, A. Y., BOE, T., LI, C., MERUSI, C., BURRAGE, J., DE LAS HERAS, J. I. & STANCHEVA, I. (2011). LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. *Genome Res* **21**, 83-94.
- NAGANO, T., MITCHELL, J. A., SANZ, L. A., PAULER, F. M., FERGUSON-SMITH, A. C., FEIL, R. & FRASER, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717-20.
- NEKRASOV, M., SOBOLEVA, T. A., JACK, C. & TREMETHICK, D. J. (2013). Histone variant selectivity at the transcription start site: H2A.Z or H2A.Lap1. *Nucleus* **4**, 431-8.
- NEMETH, A., GUIBERT, S., TIWARI, V. K., OHLSSON, R. & LANGST, G. (2008). Epigenetic regulation of TTF-I-mediated promoter-terminator interactions of rRNA genes. *EMBO J* **27**, 1255-65.
- NEMETH, A. & LANGST, G. (2011). Genome organization in and around the nucleolus. *Trends Genet* **27**, 149-56.
- NG, S. S., YUE, W. W., OPPERMANN, U. & KLOSE, R. J. (2009). Dynamic protein methylation in chromatin biology. *Cell Mol Life Sci* **66**, 407-22.
- NILE, C. J., READ, R. C., AKIL, M., DUFF, G. W. & WILSON, A. G. (2008). Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis. *Arthritis and Rheumatism* **58**, 2686-2693.
- OBERDOERFFER, P. & SINCLAIR, D. A. (2007). The role of nuclear architecture in genomic instability and ageing. *Nat Rev Mol Cell Biol* **8**, 692-702.
- OHMS, S. & RANGASAMY, D. (2014). Silencing of LINE-1 retrotransposons contributes to variation in small noncoding RNA expression in human cancer cells. *Oncotarget* **5**, 4103-17.
- OKAMOTO, I. & HEARD, E. (2009). Lessons from comparative analysis of X-chromosome inactivation in mammals. *Chromosome Res* **17**, 659-69.
- OKAMURA, K., BALLA, S., MARTIN, R., LIU, N. & LAI, E. C. (2008). Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. *Nat Struct Mol Biol* **15**, 998.
- OKANO, M., BELL, D. W., HABER, D. A. & LI, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-57.
- OLINS, A. L. & OLINS, D. E. (1974). Spheroid chromatin units (v bodies). *Science* **183**, 330-2.
- OLINS, E. D. (2003). Chromatin history: Our view from the bridge. *Nature Reviews. Molecular Cell Biology*.
- OSTERLEHNER, A., SIMMETH, S. & GOPFERT, U. (2011). Promoter methylation and transgene copy numbers predict unstable protein production in recombinant Chinese hamster ovary cell lines. *Biotechnol Bioeng* **108**, 2670-81.
- OTTO, S. P. & WALBOT, V. (1990). DNA methylation in eukaryotes: kinetics of demethylation and de novo methylation during the life cycle. *Genetics* **124**, 429-37.
- PAGGETTI, J., LARGEOT, A., AUCAGNE, R., JACQUEL, A., LAGRANGE, B., YANG, X. J., SOLARY, E., BASTIE, J. N. & DELVA, L. (2010). Crosstalk between leukemia-associated proteins MOZ and MLL regulates HOX gene expression in human cord blood CD34+ cells. *Oncogene* **29**, 5019-31.
- PANDEY, R. R., MONDAL, T., MOHAMMAD, F., ENROTH, S., REDRUP, L., KOMOROWSKI, J., NAGANO, T., MANCINI-DINARDO, D. & KANDURI, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* **32**, 232-46.

- PAPAPETROU, E. P. & SADELAIN, M. (2011). Derivation of genetically modified human pluripotent stem cells with integrated transgenes at unique mapped genomic sites. *Nat Protoc* **6**, 1274-89.
- PAPP, B. & PLATH, K. (2013). Epigenetics of reprogramming to induced pluripotency. *Cell* **152**, 1324-43.
- PAREDES, V., PARK, J. S., JEONG, Y., YOON, J. & BAEK, K. (2013). Unstable expression of recombinant antibody during long-term culture of CHO cells is accompanied by histone H3 hypoacetylation. *Biotechnol Lett* **35**, 987-93.
- PAULER, F. M., BARLOW, D. P. & HUDSON, Q. J. (2012). Mechanisms of long range silencing by imprinted macro non-coding RNAs. *Curr Opin Genet Dev* **22**, 283-9.
- PAYER, B. & LEE, J. T. (2008). X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet* **42**, 733-72.
- PENG, J. C. & KARPEN, G. H. (2007). H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat Cell Biol* **9**, 25-35.
- PENG, J. C. & LIN, H. (2013). Beyond transposons: the epigenetic and somatic functions of the Piwi-piRNA mechanism. *Curr Opin Cell Biol* **25**, 190-4.
- PESCHANSKY, V. J. & WAHLESTEDT, C. (2014). Non-coding RNAs as direct and indirect modulators of epigenetic regulation. *Epigenetics* **9**, 3-12.
- PESSIA, E., ENGELSTADTER, J. & MARAIS, G. A. (2014). The evolution of X chromosome inactivation in mammals: the demise of Ohno's hypothesis? *Cell Mol Life Sci* **71**, 1383-94.
- PETRUK, S., SEDKOV, Y., JOHNSTON, D. M., HODGSON, J. W., BLACK, K. L., KOVERMANN, S. K., BECK, S., CANAANI, E., BROCK, H. W. & MAZO, A. (2012). TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. *Cell* **150**, 922-33.
- PFAFFENEDER, T., HACKNER, B., TRUSS, M., MUNZEL, M., MULLER, M., DEIML, C. A., HAGEMEIER, C. & CARELL, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl* **50**, 7008-12.
- PHILLIPS, D. M. (1963). The presence of acetyl groups of histones. *Biochem J* **87**, 258-63.
- PILLUS, L. (2008). MYSTs mark chromatin for chromosomal functions. *Curr Opin Cell Biol* **20**, 326-33.
- PRESTEL, M., FELLER, C., STRAUB, T., MITLOHNER, H. & BECKER, P. B. (2010). The activation potential of MOF is constrained for dosage compensation. *Mol Cell* **38**, 815-26.
- PRUNER, I., DJORDJEVIC, V., GVOZDENOV, M., TOMIC, B. & RADOJKOVIC, D. (2014). Determination of transgene copy number in stably transfected mammalian cells by PCR-capillary electrophoresis assay. *Biochem Genet* **52**, 159-65.
- QU, L. H., NICOLOSO, M. & BACHELLERIE, J. P. (1991). A sequence dimorphism in a conserved domain of human 28S rRNA. Uneven distribution of variant genes among individuals. Differential expression in HeLa cells. *Nucleic Acids Res* **19**, 1015-9.
- RAGUNATHAN, K., JIH, G. & MOAZED, D. (2015). Epigenetics. Epigenetic inheritance uncoupled from sequence-specific recruitment. *Science* **348**, 1258699.
- RAMSAHOYE, B. H., BINISZKIEWICZ, D., LYKO, F., CLARK, V., BIRD, A. P. & JAENISCH, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S A* **97**, 5237-42.
- RAVENS, S., FOURNIER, M., YE, T., STIERLE, M., DEMBELE, D., CHAVANT, V. & TORA, L. (2014). Mof-associated complexes have overlapping and unique roles in regulating pluripotency in embryonic stem cells and during differentiation. *Elife* **3**.
- REA, S., EISENHABER, F., O'CARROLL, D., STRAHL, B. D., SUN, Z. W., SCHMID, M., OPRAVIL, S., MECHTLER, K., PONTING, C. P., ALLIS, C. D. & JENUWEIN, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593-9.
- RICHMOND, T. J. (1999). Hot papers - Crystal structure - Crystal structure of the nucleosome core particle at 2.8 angstrom resolution by K. Luger, A.W. Mader, R.K. Richmond, D.F. Sargent, T.J. Richmond - Comments. *Scientist* **13**, 15-15.
- ROBINSON, J. T., THORVALDSDOTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. (2011). Integrative genomics viewer. *Nat Biotechnol* **29**, 24-6.
- ROBINSON, P. J. & RHODES, D. (2006). Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Curr Opin Struct Biol* **16**, 336-43.
- ROSE, N. R. & KLOSE, R. J. (2014). Understanding the relationship between DNA methylation and histone lysine methylation. *Biochim Biophys Acta* **1839**, 1362-1372.
- ROSS, R. J., WEINER, M. M. & LIN, H. (2014). PIWI proteins and PIWI-interacting RNAs in the soma. *Nature* **505**, 353-9.
- ROSSER, J. M. & AN, W. (2010). Repeat-induced gene silencing of L1 transgenes is correlated with differential promoter methylation. *Gene* **456**, 15-23.

- ROUGEULLE, C., NAVARRO, P. & AVNER, P. (2003). Promoter-restricted H3 Lys 4 di-methylation is an epigenetic mark for monoallelic expression. *Hum Mol Genet* **12**, 3343-8.
- ROUTH, A., SANDIN, S. & RHODES, D. (2008). Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proc Natl Acad Sci U S A* **105**, 8872-7.
- RUIJTER, J. M., VELDEN, S. v. D. & ILGUN, A. (2009). LinRegPCR Manual 11.0.
- RUTHENBURG, A. J., ALLIS, C. D. & WYSOCKA, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell* **25**, 15-30.
- SADO, T., HOKI, Y. & SASAKI, H. (2005). Tsix silences Xist through modification of chromatin structure. *Dev Cell* **9**, 159-65.
- SADOWSKI, I., MA, J., TRIEZENBERG, S. & PTASHNE, M. (1988). Gal4-Vp16 Is an Unusually Potent Transcriptional Activator. *Nature* **335**, 563-564.
- SAKSOUK N. (2015). Constitutive heterochromatin formation and transcription in mammals. *Epigenetics & Chromatin*.
- SALL, J. (2002). Monte Carlo Calibration of Distributions of Partition Statistics. *SAS white paper*.
- SANDER, E. E. & GRUMMT, I. (1997). Oligomerization of the transcription termination factor TTF-I: implications for the structural organization of ribosomal transcription units. *Nucleic Acids Res* **25**, 1142-7.
- SANTENARD, A. & TORRES-PADILLA, M. E. (2009). Epigenetic reprogramming in mammalian reproduction: contribution from histone variants. *Epigenetics* **4**, 80-4.
- SANTORO, R. & GRUMMT, I. (2001). Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. *Mol Cell* **8**, 719-25.
- SANTORO, R., LI, J. & GRUMMT, I. (2002). The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. *Nat Genet* **32**, 393-6.
- SANTORO, R., SCHMITZ, K. M., SANDOVAL, J. & GRUMMT, I. (2010). Intergenic transcripts originating from a subclass of ribosomal DNA repeats silence ribosomal RNA genes in trans. *EMBO Rep* **11**, 52-8.
- SAUNDERS, L. R. & VERDIN, E. (2007). Sirtuins: critical regulators at the crossroads between cancer and aging. *Oncogene* **26**, 5489-504.
- SCHALCH, T., JOB, G., NOFFSINGER, V. J., SHANKER, S., KUSCU, C., JOSHUA-TOR, L. & PARTRIDGE, J. F. (2009). High-affinity binding of Chp1 chromodomain to K9 methylated histone H3 is required to establish centromeric heterochromatin. *Mol Cell* **34**, 36-46.
- SCHEFE, J. H., LEHMANN, K. E., BUSCHMANN, I. R., UNGER, T. & FUNKE-KAISER, H. (2006). Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's CT difference" formula. *J Mol Med (Berl)* **84**, 901-10.
- SCHIEHMANN, A. H., LI, F., WEAKE, V. M., BELIKOFF, E. J., KLEMMER, K. C., MOORE, S. A. & SCOTT, M. J. (2010). Sex-biased transcription enhancement by a 5' tethered Gal4-MOF histone acetyltransferase fusion protein in *Drosophila*. *BMC molecular biology* **11**, 80.
- SCHIMKE, R. T., KAUFMAN, R. J., ALT, F. W. & KELLEMS, R. F. (1978). Gene amplification and drug resistance in cultured murine cells. *Science* **202**, 1051-5.
- SCHLESINGER, S., SELIG, S., BERGMAN, Y. & CEDAR, H. (2009). Allelic inactivation of rDNA loci. *Genes Dev* **23**, 2437-47.
- SEKINGER, E. A. & GROSS, D. S. (2001). Silenced chromatin is permissive to activator binding and PIC recruitment. *Cell* **105**, 403-414.
- SHARIF, J., MUTO, M., TAKEBAYASHI, S., SUETAKE, I., IWAMATSU, A., ENDO, T. A., SHINGA, J., MIZUTANI-KOSEKI, Y., TOYODA, T., OKAMURA, K., TAJIMA, S., MITSUYA, K., OKANO, M. & KOSEKI, H. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**, 908-12.
- SHI, X., HONG, T., WALTER, K. L., EWALT, M., MICHISHITA, E., HUNG, T., CARNEY, D., PENA, P., LAN, F., KAADIGE, M. R., LACOSTE, N., CAYROU, C., DAVRAZOU, F., SAHA, A., CAIRNS, B. R., AYER, D. E., KUTATELADZE, T. G., SHI, Y., COTE, J., CHUA, K. F. & GOZANI, O. (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**, 96-9.
- SHI, Y., LAN, F., MATSON, C., MULLIGAN, P., WHETSTINE, J. R., COLE, P. A., CASERO, R. A. & SHI, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941-53.
- SHIAO, Y. H., LEIGHTY, R. M., WANG, C., GE, X., CRAWFORD, E. B., SPURRIER, J. M., MCCANN, S. D., FIELDS, J. R., FORNWALD, L., RIFFLE, L., DRIVER, C., QUINONES, O. A., WILSON, R. E., KASPRZAK, K. S., TRAVLOS, G. S., ALVORD, W. G. & ANDERSON, L. M. (2011). Ontogeny-driven rDNA rearrangement, methylation, and transcription, and paternal influence. *PLoS One* **6**, e22266.
- SHIMOJO, H., SANO, N., MORIWAKI, Y., OKUDA, M., HORIKOSHI, M. & NISHIMURA, Y. (2008). Novel structural and functional mode of a knot essential for RNA binding activity of the Esa1 presumed chromodomain. *J Mol Biol* **378**, 987-1001.

- SHIUE, C. N., ARABI, A. & WRIGHT, A. P. (2010). Nucleolar organization, growth control and cancer. *Epigenetics* **5**, 200-5.
- SHIUE, C. N., BERKSON, R. G. & WRIGHT, A. P. (2009). c-Myc induces changes in higher order rDNA structure on stimulation of quiescent cells. *Oncogene* **28**, 1833-42.
- SHOGREN-KNAAK, M., ISHII, H., SUN, J. M., PAZIN, M. J., DAVIE, J. R. & PETERSON, C. L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844-7.
- SMITH, Z. D. & MEISSNER, A. (2013). DNA methylation: roles in mammalian development. *Nat Rev Genet* **14**, 204-20.
- SONG, J., RECHKOBLIT, O., BESTOR, T. H. & PATEL, D. J. (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331**, 1036-40.
- STADLER, M. B., MURR, R., BURGER, L., IVANEK, R., LIENERT, F., SCHOLER, A., VAN NIMWEGEN, E., WIRBELAUER, C., OAKELEY, E. J., GAIDATZIS, D., TIWARI, V. K. & SCHUBELER, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490-5.
- STEFFEN, P. A., FONSECA, J. P. & RINGROSE, L. (2012). Epigenetics meets mathematics: towards a quantitative understanding of chromatin biology. *Bioessays* **34**, 901-13.
- STEIN, R., RAZIN, A. & CEDAR, H. (1982). In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. *Proc Natl Acad Sci U S A* **79**, 3418-22.
- STROHNER, R., NEMETH, A., JANSÁ, P., HOFMANN-ROHRER, U., SANTORO, R., LANGST, G. & GRUMMT, I. (2001). NoRC--a novel member of mammalian ISWI-containing chromatin remodeling machines. *EMBO J* **20**, 4892-900.
- STRUTZENBERGER, K., BORTH, N., KUNERT, R., STEINFELLNER, W. & KATINGER, H. (1999). Changes during subclone development and ageing of human antibody-producing recombinant CHO cells. *J Biotechnol* **69**, 215-26.
- SUN, B. K., DEATON, A. M. & LEE, J. T. (2006). A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol Cell* **21**, 617-28.
- SUN, W. J., ZHOU, X., ZHENG, J. H., LU, M. D., NIE, J. Y., YANG, X. J. & ZHENG, Z. Q. (2012). Histone acetyltransferases and deacetylases: molecular and clinical implications to gastrointestinal carcinogenesis. *Acta Biochim Biophys Sin (Shanghai)* **44**, 80-91.
- SWEET, B. H. & HILLEMANN, M. R. (1960). The vacuolating virus, S.V. 40. *Proc Soc Exp Biol Med* **105**, 420-7.
- SWINDLE, C. S., KIM, H. G. & KLUG, C. A. (2004). Mutation of CpGs in the murine stem cell virus retroviral vector long terminal repeat represses silencing in embryonic stem cells. *J Biol Chem* **279**, 34-41.
- TAHILIANI, M., KOH, K. P., SHEN, Y., PASTOR, W. A., BANDUKWALA, H., BRUDNO, Y., AGARWAL, S., IYER, L. M., LIU, D. R., ARAVIND, L. & RAO, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-5.
- TALBERT, P. B. & HENIKOFF, S. (2010). Histone variants--ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* **11**, 264-75.
- TAM, O. H., ARAVIN, A. A., STEIN, P., GIRARD, A., MURCHISON, E. P., CHELOUFI, S., HODGES, E., ANGER, M., SACHIDANANDAM, R., SCHULTZ, R. M. & HANNON, G. J. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**, 534-8.
- TAVARES, L., DIMITROVA, E., OXLEY, D., WEBSTER, J., POOT, R., DEMMERS, J., BEZSTAROSTI, K., TAYLOR, S., URA, H., KOIDE, H., WUTZ, A., VIDAL, M., ELDERKIN, S. & BROCKDORFF, N. (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664-78.
- TAVERNA, S. D., LI, H., RUTHENBURG, A. J., ALLIS, C. D. & PATEL, D. J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14**, 1025-40.
- TERRANOVA, R., YOKOBAYASHI, S., STADLER, M. B., OTTE, A. P., VAN LOHUIZEN, M., ORKIN, S. H. & PETERS, A. H. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev Cell* **15**, 668-79.
- 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**, 403-27.
- THOMAS, J. O. & KORNBERG, R. D. (1975). An octamer of histones in chromatin and free in solution. *Proc Natl Acad Sci U S A* **72**, 2626-30.
- THOMAS, T., DIXON, M. P., KUEH, A. J. & VOSS, A. K. (2008). Mof (MYST1 or KAT8) is essential for progression of embryonic development past the blastocyst stage and required for normal chromatin architecture. *Mol Cell Biol* **28**, 5093-105.
- THOMSON, J. P., SKENE, P. J., SELFRIDGE, J., CLOUAIRE, T., GUY, J., WEBB, S., KERR, A. R., DEATON, A., ANDREWS, R., JAMES, K. D., TURNER, D. J., ILLINGWORTH, R. & BIRD, A. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082-6.

- TIMMERMANN, S., LEHRMANN, H., POLESSKAYA, A. & HAREL-BELLAN, A. (2001). Histone acetylation and disease. *Cell Mol Life Sci* **58**, 728-36.
- TONNA, S., EL-OSTA, A., COOPER, M. E. & TIKELLIS, C. (2010). Metabolic memory and diabetic nephropathy: potential role for epigenetic mechanisms. *Nat Rev Nephrol* **6**, 332-41.
- TREMETHICK, D. J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* **128**, 651-4.
- TROJER, P. & REINBERG, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? *Mol Cell* **28**, 1-13.
- TSENG, H., CHOU, W., WANG, J., ZHANG, X., ZHANG, S. & SCHULTZ, R. M. (2008). Mouse ribosomal RNA genes contain multiple differentially regulated variants. *PLoS One* **3**, e1843.
- TURNER, B. M. (2002). Cellular memory and the histone code. *Cell* **111**, 285-91.
- VALOR, L. M., PULOPULOS, M. M., JIMENEZ-MINCHAN, M., OLIVARES, R., LUTZ, B. & BARCO, A. (2011). Ablation of CBP in forebrain principal neurons causes modest memory and transcriptional defects and a dramatic reduction of histone acetylation but does not affect cell viability. *J Neurosci* **31**, 1652-63.
- VERDEL, A., JIA, S., GERBER, S., SUGIYAMA, T., GYGI, S., GREWAL, S. I. & MOAZED, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**, 672-6.
- VETTING, M. W., LP, S. D. C., YU, M., HEGDE, S. S., MAGNET, S., RODERICK, S. L. & BLANCHARD, J. S. (2005). Structure and functions of the GNAT superfamily of acetyltransferases. *Arch Biochem Biophys* **433**, 212-26.
- VOLPE, T. A., KIDNER, C., HALL, I. M., TENG, G., GREWAL, S. I. & MARTIENSSEN, R. A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833-7.
- WALSH, G. (2010). Biopharmaceutical benchmarks 2010.
- WANG, S. H., NAN, R., ACCARDO, M. C., SENTMANAT, M., DIMITRI, P. & ELGIN, S. C. (2014). A distinct type of heterochromatin at the telomeric region of the *Drosophila melanogaster* Y chromosome. *PLoS One* **9**, e86451.
- WANG, Z., ZANG, C., ROSENFELD, J. A., SCHONES, D. E., BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., PENG, W., ZHANG, M. Q. & ZHAO, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**, 897-903.
- WATANABE, T., TAKEDA, A., TSUKIYAMA, T., MISE, K., OKUNO, T., SASAKI, H., MINAMI, N. & IMAI, H. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev* **20**, 1732-43.
- WEBER, C. M. & HENIKOFF, S. (2014). Histone variants: dynamic punctuation in transcription. *Genes & Development* **28**, 672-682.
- WEST, N. R. (2014). Development of a Tunable Mammalian Protein Expression System and an Investigation of Promoter Interference in Three Promoters Often Utilized in the Production of Biopharmaceuticals. *White Rose eTheses Online*.
- WHETSTONE, J. R., NOTTKE, A., LAN, F., HUARTE, M., SMOLIKOV, S., CHEN, Z., SPOONER, E., LI, E., ZHANG, G., COLAIACOVO, M. & SHI, Y. (2006). Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**, 467-81.
- WIDOM, J. (1992). A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells. *Proc Natl Acad Sci U S A* **89**, 1095-9.
- WIERZBICKI, A. T., HAAG, J. R. & PIKAARD, C. S. (2008). Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. *Cell* **135**, 635-648.
- WIERZBICKI, A. T., REAM, T. S., HAAG, J. R. & PIKAARD, C. S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* **41**, 630-4.
- WOODCOCK, C. L. & GHOSH, R. P. (2010). Chromatin higher-order structure and dynamics. *Cold Spring Harb Perspect Biol* **2**, a000596.
- WRIGHT, A., SEMYONOV, A., DAWES, G., CRAMERI, A., LYONS, R., STEMMER, W. P. C., APT, D. & PUNNONEN, J. (2005). Diverse plasmid DNA vectors by directed molecular evolution of cytomegalovirus promoters. *Human Gene Therapy* **16**, 881-892.
- WU, C., BASSETT, A. & TRAVERS, A. (2007). A variable topology for the 30-nm chromatin fibre. *EMBO Rep* **8**, 1129-34.
- WU, H. & ZHANG, Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156**, 45-68.
- WURM, F. M. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol* **22**, 1393-8.
- WYSOCKA, J., SWIGUT, T., XIAO, H., MILNE, T. A., KWON, S. Y., LANDRY, J., KAUER, M., TACKETT, A. J., CHAIT, B. T., BADENHORST, P., WU, C. & ALLIS, C. D. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86-90.

- XHEMALCE, B. & KOUZARIDES, T. (2010). A chromodomain switch mediated by histone H3 Lys 4 acetylation regulates heterochromatin assembly. *Genes Dev* **24**, 647-52.
- XIAO, B., JING, C., WILSON, J. R., WALKER, P. A., VASISHT, N., KELLY, G., HOWELL, S., TAYLOR, I. A., BLACKBURN, G. M. & GAMBLIN, S. J. (2003). Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* **421**, 652-6.
- XU, N., DONOHOE, M. E., SILVA, S. S. & LEE, J. T. (2007). Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. *Nat Genet* **39**, 1390-6.
- XU, N., TSAI, C. L. & LEE, J. T. (2006). Transient homologous chromosome pairing marks the onset of X inactivation. *Science* **311**, 1149-52.
- YANG, N. & KAZAZIAN, H. H. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nature Structural & Molecular Biology* **13**, 763-771.
- YANG, Y., MARIATI, CHUSAINOW, J. & YAP, M. G. (2010a). DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. *Journal of biotechnology* **147**, 180-5.
- YANG, Y., MARIATI, CHUSAINOW, J. & YAP, M. G. (2010b). DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. *J Biotechnol* **147**, 180-5.
- YE, R., WANG, W., IKI, T., LIU, C., WU, Y., ISHIKAWA, M., ZHOU, X. & QI, Y. (2012). Cytoplasmic assembly and selective nuclear import of Arabidopsis Argonaute4/siRNA complexes. *Mol Cell* **46**, 859-70.
- YEE, J. C., DE LEON GATTI, M., PHILP, R. J., YAP, M. & HU, W. S. (2008). Genomic and proteomic exploration of CHO and hybridoma cells under sodium butyrate treatment. *Biotechnol Bioeng* **99**, 1186-204.
- YIN, Z., KONG, Q. R., ZHAO, Z. P., WU, M. L., MU, Y. S., HU, K. & LIU, Z. H. (2012). Position effect variegation and epigenetic modification of a transgene in a pig model. *Genet Mol Res* **11**, 355-69.
- ZENG, L., ZHANG, Q., LI, S., PLOTNIKOV, A. N., WALSH, M. J. & ZHOU, M. M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* **466**, 258-62.
- ZENTNER, G. E. & HENIKOFF, S. (2013). Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol* **20**, 259-66.
- ZENTNER, G. E., SAIKHOVA, A., MANAENKOV, P., ADAMS, M. D. & SCACHERI, P. C. (2011). Integrative genomic analysis of human ribosomal DNA. *Nucleic Acids Res* **39**, 4949-60.
- ZHANG, K., MOSCH, K., FISCHLE, W. & GREWAL, S. I. (2008a). Roles of the Ctr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nat Struct Mol Biol* **15**, 381-8.
- ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. (2008b). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137.
- ZHAO, J., OHSUMI, T. K., KUNG, J. T., OGAWA, Y., GRAU, D. J., SARMA, K., SONG, J. J., KINGSTON, R. E., BOROWSKY, M. & LEE, J. T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* **40**, 939-53.
- ZHAO, J., SUN, B. K., ERWIN, J. A., SONG, J. J. & LEE, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750-6.
- ZHENG, Y. P., THOMAS, P. M. & KELLEHER, N. L. (2013). Measurement of acetylation turnover at distinct lysines in human histones identifies long-lived acetylation sites. *Nature Communications* **4**.
- ZHOU, H., HU, H. & LAI, M. (2010). Non-coding RNAs and their epigenetic regulatory mechanisms. *Biol Cell* **102**, 645-55.
- ZHOU, V. W., GOREN, A. & BERNSTEIN, B. E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* **12**, 7-18.
- ZHOU, W., ZHU, P., WANG, J., PASCUAL, G., OHGI, K. A., LOZACH, J., GLASS, C. K. & ROSENFELD, M. G. (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Mol Cell* **29**, 69-80.
- ZHOU, Y., SANTORO, R. & GRUMMT, I. (2002). The chromatin remodeling complex NoRC targets HDAC1 to the ribosomal gene promoter and represses RNA polymerase I transcription. *EMBO J* **21**, 4632-40.
- ZILLER, M. J., MULLER, F., LIAO, J., ZHANG, Y., GU, H., BOCK, C., BOYLE, P., EPSTEIN, C. B., BERNSTEIN, B. E., LENGAUER, T., GNIRKE, A. & MEISSNER, A. (2011). Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet* **7**, e1002389.