

# **Analyse des genom-weiten Bindeverhaltens des Faktors CTCFL**

**Dissertation zur Erlangung des  
naturwissenschaftlichen Doktorgrades (Dr. rer nat.)**

Institut für Genetik  
AG Prof. Dr. Rainer Renkawitz  
Fachbereich 08 (Fachrichtung Biologie)  
Justus-Liebig-Universität Gießen

vorgelegt von  
Dipl. Biol. Philipp Franz Hermann Bergmaier  
aus München

Gießen, April 2016

Dekan: Prof. Dr. Volker Wissemann  
1. Gutachter: Prof. Dr. Rainer Renkawitz  
2. Gutachter: Prof. Dr. Lienhard Schmitz

*Für meine kleine Löwin*

## Kurzfassung

Der wichtige, multifunktionale Faktor CTCF, der als zentrales Element bei der Ausbildung drei-dimensionaler DNA Kontakte und Chromatin-Organisation identifiziert wurde, und sein Testis-spezifischer paraloger Faktor CTCFL zeichnen sich durch eine fast identische ZF-Domäne, die die DNA Bindung vermittelt, aus. Die stark unterschiedlichen Carboxy- und Amino-terminalen Domänen lassen jedoch auf funktionelle Unterschiede der beiden Proteine schließen. Genom-weite Analysen konnten neben einer großen Zahl gemeinsamer, genomischer Bindestellen auch jeweils individuelle Bindestellen identifizieren. Da inzwischen die gemeinsame Präsenz beider Faktoren in Testis-Zellen aber auch in bestimmten Krebsformen bestätigt wurde, kommt der Erforschung der Auswirkungen von CTCF und CTCFL an gemeinsamen Bindestellen eine neue Bedeutung zu.

Im Rahmen dieser Arbeit sollten die möglichen Sequenz-spezifischen oder Chromatin-bedingten Grundlagen der differentiellen Bindemuster von CTCF und CTCFL analysiert werden. Dazu wurde unter Verwendung der umfangreichen ENCODE Datenbank eine bioinformatische Korrelationsanalyse durchgeführt, die zeigen konnte, dass aktive Histon-Modifikationen und Transkriptionsfaktoren an CTCFL Bindestellen, im Vergleich zu CTCF Bindestellen, angereichert vorliegen. Komponenten des Cohesin-Komplexes konnten gleichzeitig als einzige CTCF-spezifische Cofaktoren identifiziert werden. In ChIPseq Experimenten gefolgt von konventionellen ChIPs konnte kein Einfluss von CTCFL Bindung auf CTCF, Histon-Modifikationen oder Cohesin-Komponenten festgestellt werden. Zur weiteren Untersuchung in Zellkultur wurden P19 Zellen mit induzierbarer CTCFL-Expression hergestellt. Vergleichenden Analysen mit einem bereits beschriebenen, weiteren Klon konnten mittels FAIRE Experimenten differentielle DNA-Zugänglichkeit als Basis von Zelltyp-spezifischen CTCFL-Bindemustern identifizieren. Weitere Versuche, anhand von Chromatin-Veränderung in Folge von Differenzierung, die CTCFL-Bindung innerhalb des P19 Klons zu modulieren waren nur an zwei untersuchten Bindestellen (Ptpg & Mzfl) erfolgreich.

Um mögliche weitere Einflüsse auf die CTCF- und CTCFL-Bindung zu untersuchen wurden Analysen der genauen DNA-Bindemotive durchgeführt. Im Zuge dieser Experimente konnten sehr ähnliche DNA-Sequenzen identifiziert werden, die *in vivo* stark abweichende Bindung der beiden Faktoren zeigen. Somit wurde neben dem Chromatin-Einfluss auf die CTCFL Bindung auch eine potenzielle Rolle der gebundenen DNA-Sequenz identifiziert, die es lohnt in weiteren Experimenten zu überprüfen.

## Abstract

The important multifunctional factor CTCF, known to be the central element of three-dimensional DNA contacts and involved in chromatin organisation, and its testis-specific paralogue CTCFL are sharing a nearly identical zinc finger DNA-binding domain. Strong differences in N- and C-termini are probably associated with differential function of the two proteins. Genome-wide analysis revealed a strong overlap of binding sites for both factors but also individual sites. With newer knowledge that both factors actually are expressed in the same cells in testis and also some cancer types, research into the functional consequences of CTCF and CTCFL at shared binding sites becomes more important.

Within this work the possible sequence- and chromatin-driven discrimination in CTCF and CTCFL binding events was examined. Utilizing the extensive ENCODE database bioinformatic correlation analyses were performed and active histone-modifications and transcription factors were enriched at CTCFL binding sites compared to CTCF sites. Only Cohesin-complex components could be identified to be overrepresented at CTCF binding sites. No Influence of CTCFL binding could be identified on CTCF, histone modifications or Cohesin components when performing ChIPseq for CTCF & CTCFL followed by conventional ChIP experiments. For further experiments in cell culture, an inducible CTCFL expression system was also established in P19 cells. Comparative analysis between different cell clones could then show that differential openness of chromatin, as measured by FAIRE, explains cell-type-specific CTCFL binding patterns to a degree. Experiments with differentiating P19 cells could show that CTCFL binding at two selected regions (Ptprg & Mzf1) could be modulated by inducing chromatin changes.

To assess a possible influence of the DNA motif on CTCF & CTCFL binding, detailed analysis of the precise involved DNA sequences were performed. Two highly similar sequences could be identified showing drastically different CTCF and CTCFL binding occupancies *in vivo*. These results in addition to the role of open chromatin provide a hint at a possible role of the DNA sequence on CTCFL binding and it would be worthwhile to test this experimentally in the future.

# Inhaltsverzeichnis

<b>Kurzfassung</b> .....	<b>3</b>
<b>Abstract</b> .....	<b>4</b>
<b>Inhaltsverzeichnis</b> .....	<b>5</b>
<b>Abkürzungsverzeichnis</b> .....	<b>9</b>
<b>1 Einleitung</b> .....	<b>14</b>
1.1 Der multifunktionale Faktor CTCF .....	18
1.2 CTCF und Chromosomenarchitektur .....	20
1.3 CTCFL: ein paraloger Faktor von CTCF .....	22
1.4 CTCFL-Expression in Krebs – Unterschiedliche Meinungen.....	26
1.5 Genomweite Bindemuster von CTCFL.....	27
1.6 Zielsetzung der Arbeit .....	28
<b>2 Material</b> .....	<b>30</b>
2.1 Geräte .....	30
2.2 Verbrauchsmaterialien.....	31
2.3 Chemikalien.....	32
2.4 Enzyme .....	33
2.5 Komplettsysteme .....	33
2.6 Größenstandards .....	34
2.6.1 DNA Längenstandardmarker.....	34
2.6.2 Molekulargewichtsmarker für Proteine .....	34
2.7 Antibiotika.....	34
2.8 Antikörper.....	35
2.9 Eukaryotische Zelllinien.....	35
2.10 Plasmide .....	35
2.11 Oligonukleotide .....	36
2.11.1 Primer für RT- und qRT-PCR (Mensch).....	36
2.11.2 Primer für RT- und qRT-PCR (Maus).....	37
2.11.3 Primer für ChIP/FAIRE-PCR und ChIP/FAIRE-qPCR (Maus) .....	37
2.11.4 Oligonukleotide zur Verwendung in RNAi (Mensch) .....	40
2.12 Software.....	40
<b>3 Methoden</b> .....	<b>41</b>
3.1 Arbeiten mit eukaryotischen Zelllinien .....	41
3.1.1 Ablösen von adhärenenten Zellen .....	41

---

3.1.2	Auftauen und Einfrieren von eukaryotischen Zelllinien .....	41
3.1.3	Kultivierung von eukaryotischen Zelllinien .....	42
3.1.4	Differenzierung von murinen P19 Zellen .....	42
3.1.5	Zellzahlbestimmung .....	43
3.1.6	Transfektion .....	43
3.2	Arbeiten mit DNA .....	43
3.2.1	Lagerung von DNA .....	43
3.2.2	Quantifizierung von DNA .....	44
3.2.3	Minipräparation von Plasmid-DNA .....	44
3.2.4	Midipräparation von Plasmid-DNA .....	45
3.2.5	Phenol-Chloroform-Extraktion .....	45
3.2.6	Ethanol-fällung .....	45
3.2.7	Sequenzierung von DNA .....	46
3.2.8	Gelelektrophorese von DNA-Fragmenten .....	46
3.2.9	DNA-Extraktion aus Agarosegel .....	47
3.3	Enzymatische Modifikationen von DNA .....	47
3.3.1	Schneiden von DNA durch Restriktionsenzyme .....	47
3.3.2	Dephosphorylierung .....	47
3.3.3	Phosphorylierung .....	47
3.3.4	Ligation .....	48
3.3.5	Polymerase-Kettenreaktion (PCR) .....	48
3.3.6	Quantitative-Real-Time-PCR .....	49
3.4	Arbeiten mit RNA .....	49
3.4.1	Umgang mit RNA .....	49
3.4.2	Isolierung von RNA .....	49
3.4.3	Quantifizierung von RNA .....	49
3.4.4	DNase Verdau von RNA .....	50
3.4.5	RT-PCR .....	50
3.5	Arbeiten mit Bakterien .....	50
3.5.1	Vermehrung und Lagerung .....	50
3.5.2	Transformation .....	51
3.6	Arbeiten mit Proteinen .....	51
3.6.1	Proteinextraktion mit RIPA-Puffer .....	51
3.6.2	Proteinbestimmung nach Bradford .....	52
3.6.3	SDS-Polyacrylamidgelelektrophorese .....	53
3.7	Western Blot .....	54
3.7.1	Blotten der Proteine .....	54
3.7.2	Immundetektion .....	55
3.7.3	Strippen von Western-Blot Membranen .....	56
3.8	Chromatin-Immunopräzipitation .....	57
3.8.1	Isolierung des Chromatins .....	57

---

3.8.2	Immunpräzipitation .....	58
3.9	FAIRE-Assay .....	59
3.9.1	Isolierung des Chromatins .....	60
3.9.2	Gewinnung des Input-Materials (=genomische DNA) als Kontrolle.....	60
3.9.3	Gewinnung der spezifischen ungebundenen FAIRE-DNA.....	61
<b>4</b>	<b>Ergebnisse .....</b>	<b>63</b>
4.1	Etablierung und Charakterisierung einer stabilen CTCFL-Überexpression in humanen HeLa-Zellen .....	63
4.1.1	Zwei HeLa-Klone zeigen stabile induzierbare Expression von CTCFL auf RNA-Ebene .....	63
4.1.2	CTCFL lässt sich in zwei HeLa Klonen nicht in voller Länge nachweisen.....	65
4.2	CTCFL-Analysen in humanen K562-Zellen .....	66
4.2.1	Aufeinanderfolgende Transfektion eines spezifischen siRNA-Mixes führt zu Reduktion der CTCFL Expression .....	66
4.2.2	Die Reduktion der CTCFL Expression hat einen Einfluss auf ausgewählte Kandidaten-Gene .....	69
4.2.3	Bioinformatische Untersuchung des Zusammenhangs von CTCFL Bindung mit genomweiten humanen Chromatin-Eigenschaften (ENCODE).....	71
4.3	Untersuchungen des Bindeverhaltens von CTCFL in stabil transfizierten NIH3T3 Zellen .....	79
4.3.1	CTCFL hat keinen klaren Einfluss auf CTCF-Bindung an geteilten Bindestellen .....	79
4.3.2	CTCFL hat keinen klaren Einfluss auf die Anwesenheit bestimmter Histon-Modifikationen an seinen Bindestellen .....	83
4.3.3	CTCFL zeigt keinen klaren Einfluss auf Cohesin-Bindung.....	84
4.4	Untersuchungen des Bindeverhaltens von CTCFL im murinen Differenzierungssystem .....	87
4.4.1	Etablierung einer stabilen Expression von CTCFL in P19 Zellen .....	88
4.4.2	Einfluss von CTCFL auf die genomweite Expression von Genen in Abhängigkeit von Differenzierung (Microarray) .....	90
4.4.3	Identifizierung der genomweiten Bindestellen von CTCFL im Differenzierungssystem (ChIPseq).....	97
4.4.4	Vergleichende Analyse der CTCFL-Bindung zwischen NIH3T3 und P19 Klonen .....	103
4.4.5	Gewinnung eigener Daten für Konformationsänderung des Chromatins in Folge von Differenzierung .....	108
4.5	Untersuchungen des DNA Bindeverhaltens von CTCF und CTCFL auf Grundlage von DNA Sequenz Unterschieden .....	114
4.5.1	Identifikation möglicher DNA Spezifität bei der Bindung von CTCF und CTCFL.....	115
<b>5</b>	<b>Diskussion.....</b>	<b>121</b>
5.1	Versuch der Etablierung stabiler, induzierbarer CTCFL-Expression im humanen System.....	121

---

5.2	Untersuchungen zu CTCFL im humanen System (K562) .....	122
5.2.1	.Reduzierte CTCFL-Expression hat einen Effekt auf ausgesuchte Kandidaten-Gene .....	123
5.2.2	Korrelationsanalysen zu CTCFL anhand der ENCODE Datenbanken .....	124
5.3	Untersuchungen zu CTCFL Bindestellen in NIH3T3 Zellen .....	125
5.3.1	Keine detektierbare CTCF-CTCFL Kooperation im Klon 34 .....	125
5.3.2	CTCFL zeigt keinen Einfluss auf Histon-Modifikationen .....	126
5.3.3	CTCFL zeigt keinen Zusammenhang mit Cohesin-Komponenten .....	127
5.4	CTCFL Überexpression im Differenzierungssystem .....	127
5.4.1	Einfluss von CTCFL auf genom-weite Expression .....	128
5.4.2	Identifikation von CTCFL Bindestellen im P19 Klon .....	129
5.4.3	Vergleichende Analysen zwischen NIH3T3 und P19 Klonen .....	130
5.4.4	Analyse differentieller Chromatin Zugänglichkeit in Bezug auf CTCFL .....	132
5.5	Identifizierung eines möglichen DNA Einflusses auf differentielle CTCF und CTCFL Bindung .....	133
5.6	Ausblick .....	134
<b>6</b>	<b>Literaturverzeichnis .....</b>	<b>136</b>
	<b>Danksagung .....</b>	<b>145</b>
	<b>Eidesstattliche Erklärung .....</b>	<b>Fehler! Textmarke nicht definiert.</b>



## Abkürzungsverzeichnis

%	Prozent
°C	Grad Celsius
μ	mikro- (10 <sup>-6</sup> )
3C	<i>Chromatin conformation capture</i>
5C	<i>Carbon copy chromatin conformation capture</i>
A	Adenin
Abb.	Abbildung
ad.	aufgefüllt auf
Amp	Ampicillin
AmpR	Ampicillin Resistenz
APS	Ammoniumpersulfat
AS	Aminosäure(n)
as	<i>antisense</i>
ATCC	<i>American Type Culture Collection</i>
bp	Basenpaar(e)
bzw.	beziehungsweise
C	Cytosin
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
ca.	circa
cDNA	<i>complementary DNA</i>
ChIP	Chromatin-Immunopräzipitation
ChIPseq	ChIP Sequenzierung
CIAP	<i>Calf Intestine Alkaline Phosphatase</i>
CMV	Cytomegalie-Virus
CO <sub>2</sub>	Kohlenstoffdioxid
CRISPR	<i>Clustered regularly-interspaced short palindromic repeats</i>
CTCF	<i>CCCTC-binding factor</i>
CTCFL	<i>CTCF-like</i>
C-Terminus	Carboxy-Terminus

---

CTS	<i>CTCF target site</i>
d.h.	das heißt
Da	Dalton
DMEM	<i>Dulbecco's Modified Eagle Medium</i>
DMSO	Dimethylsulfoxid
DNA	Desoxyribonukleinsäure
DNase	Desoxyribonuklease
dNTP	Desoxynukleotidtriphosphat
DOC	Deoxycholat
Dox	Doxyzyklin
ds	doppelsträngig
<i>E. coli</i>	<i>Escherichia coli</i>
ECL	<i>Enhanced chemiluminescence</i>
EDTA	Ethylendiamintetraessigsäure
EGFP	<i>Enhanced Green Fluorescent Protein</i>
ENCODE	<i>Encyclopedia of DNA Elements</i>
ES	<i>Embryonic stem</i>
<i>et al.</i>	<i>et alii</i> (und andere)
EtBr	Ethidiumbromid
EtOH	Ethanol
FCS	fötales Kälberserum
G	Guanin
g	Gramm
GAPDH	Glycerinaldehyd-3-phosphat-Dehydrogenase
h	Stunde(n)
H <sub>2</sub> O	Wasser
HCl	Salzsäure
hCTCFL	humanes CTCFL
HRP	Horseradish-Peroxidase
IgG	Immunglobulin G

---

k	kilo- ( $10^3$ )
KaAc	Kaliumacetat
kb	Kilobase (=1000 bp)
kDa	Kilodalton (=1000 Da)
l	Liter
LB-Medium	Luria-Bertani-Medium
Lsg	Lösung
m	Meter oder milli- ( $10^{-3}$ )
M	Molar (mol/l)
MCS	<i>Multiple Cloning Site</i>
mCTCFL	Maus CTCFL
MEM	<i>Minimum Essential Medium</i>
MeOH	Methanol
min	Minute(n)
mQH <sub>2</sub> O	Wasser aus der MilliQ Anlage
mRNA	<i>messenger RNA</i>
MW	Molekulargewicht
n	nano- ( $10^{-9}$ )
NaAc	Natriumacetat
NaHCO <sub>3</sub>	Natriumhydrogencarbonat
NeoR	Neomycin Resistenz
Nr.	Nummer
N-Terminus	Amino-Terminus
OD	Optische Dichte
p	pico
PAGE	Polyacrylamidgelelektrophorese
PBS	<i>Phosphate Buffered Saline</i>
PBS-T	PBS + 0,1% Tween® 20
PCR	<i>Polymerase Chain Reaction</i>
PMSF	Phenylmethylsulfonylfluorid

---

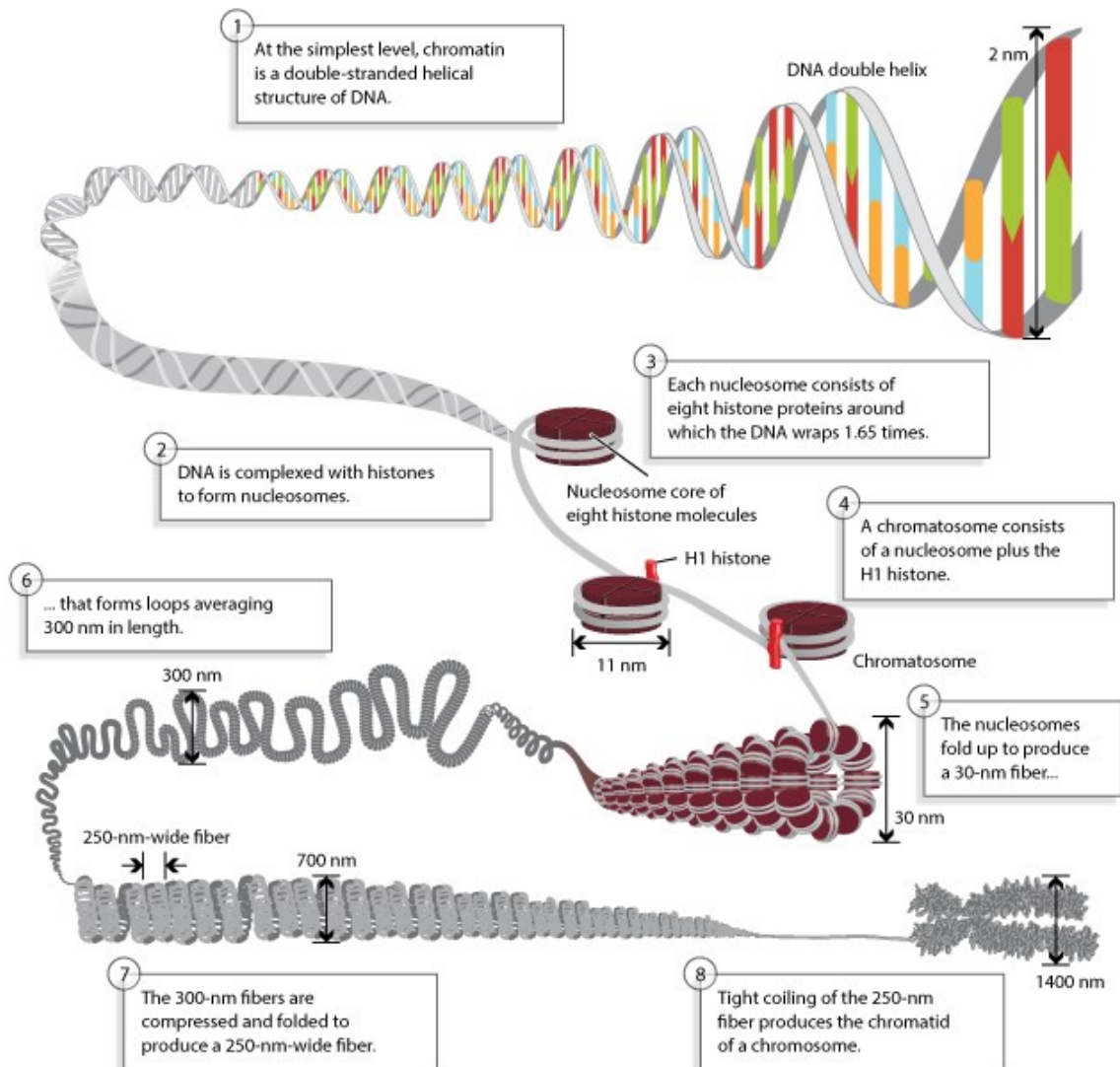
PNK	Polynukleotidkinase
PVDF	Polyvinylidenfluorid
qPCR	quantitative PCR
q-RT-PCR	quantitative RT-PCR
RA	<i>All-Trans</i> -Retinsäure
RNA	Ribonukleinsäure
RNAi	RNA Interferenz
RNase	Ribonuklease
rpm	<i>rounds per minute</i>
RT	Raumtemperatur
RT-PCR	Reverse Transkriptase PCR
s	<i>sense</i>
SDS	Natriumdodecylsulfat
sec	Sekunde(n)
siRNA	<i>Small interfering RNA</i>
sog.	sogenannt
Sol1/2/3	<i>Solution 1/2/3</i>
SV40	Simianes Virus 40
T	Thymin
Tab.	Tabelle
TAD	<i>Topologically associated domain</i>
TAE	Tris-Acetat-EDTA-Puffer
TE	Tris-EDTA-Puffer
TEMED	Tetramethylethylendiamin
TES	<i>transcriptional end site</i>
Tet	Tetracyclin
TRE	Tetracyclin-Response-Element
Tris	Tris-(hydroxymethyl)-aminomethan
TSS	<i>transcriptional start site</i>
U	Uracil

---

UTR	<i>untranslated region</i>
UV	Ultraviolettes Licht ( $\lambda \leq 340$ nm)
V	Volt
v/v	<i>volume/volume</i>
VT	Volumenteil
w/v	<i>weight/volume</i>
ZF	Zinkfinger
$\alpha$	Anti
$\beta$	Beta-
$\lambda$	Wellenlänge

# 1 Einleitung

Im Zuge der fast zeitgleichen Sequenzierung des menschlichen Genoms durch ein öffentlich gefördertes Konsortium (Lander et al. 2001) und eine private Firma unter Führung von Craig Venter (Venter et al. 2001) konnte Anfang des Jahrtausends auch die Gesamtgröße des humanen Genoms identifiziert werden. Das haploide Genom eines Menschen besteht aus ca.  $3 \times 10^9$  Basenpaaren, in somatischen diploiden Zellen liegen somit  $6 \times 10^9$  Basenpaare mit einer Gesamtlänge von ca. 2m vor. Die Verpackung und korrekte Verarbeitung dieser Mengen an Erbinformation im winzigen Raum des Zellkerns, von durchschnittlich  $6 \mu\text{m}$  in Säugern (Alberts et al. 2002), stellt daher eine der größten Herausforderungen an die Zelle dar.



**Abb. 1: Allgemeines schematisches Modell der Ebenen der DNA Verpackung bis hin zum Chromosom.** (Annunziato 2008) Dargestellt ist die DNA ausgehend von ihrer grundsätzlichen Version als Doppelhelix (oben rechts), über die Schritte der Nucleosomen, 30 nm Faser und 300 nm Faser hin bis zur kompaktesten möglichen Variante in Form der Metaphase Chromosomen (unten rechts).

Diese Verpackung der DNA erfolgt in eukaryotischen Zellen auf mehreren Ebenen, die es ermöglichen, dass die DNA trotz starker Komprimierung sinnvoll und schnell abgelesen und genutzt werden kann. Die grundlegendste Form des Chromatins stellt die DNA in Form der Doppelhelix dar, bei der jeweils zwei DNA Stränge über ihre Basen, unter Ausbildung von Wasserstoffbrücken, verbunden sind. Die zweite Ebene der Verpackung stellen die sog. Nukleosomen dar, hierzu ist die DNA um einen Proteinkomplex aus Histonen, das Histon-Oktamer, gewickelt (Abb. 1). Bei den Histonen handelt es sich um wichtige strukturelle Proteine die in Assoziation mit der DNA auch Aufgaben der Expressionsregulation übernehmen können. Jeweils zwei Proteine der Kern-Histone H2A, H2B, H3 und H4 bilden dieses Oktamer. Die um das Histon-Oktamer gewickelte DNA wird durch zusätzliche Bindung von Histon H1 fixiert und bildet so eine stabile Struktur aus, die in mikroskopischen Analysen schon frühzeitig als sogenannte Perlenkette identifiziert wurde (Olins und Olins 1974). Durch die doppelte Windung der DNA um jeweils ein Nukleosom wird eine 5-10 fache Komprimierung erreicht (Kornberg 1974). Je Nukleosom sind 147 Basenpaare fest gebunden und die einzelnen Nukleosomen dann durch freie *Linker* DNA, die keine Bindung an Histone zeigt, verbunden. In einer weiteren Ebene ist das Chromatin in der 30 nm Faser angeordnet. Zur genauen Struktur der 30 nm Faser gibt es auch nach 30 Jahren Forschung keine ganz klaren Ergebnisse (Grigoryev und Woodcock 2012), generell wird jedoch von einer Anordnung der Nukleosomen als Solenoid, d.h als einsträngige eingedrehte Helix, ausgegangen (Kruithof et al. 2009). Es erfolgen weitere Komprimierungsschritte wie das Ausbilden von Chromatin-Schleifen und weiteren Faserstrukturen (300 nm). Die kompakteste Variante des Chromatins stellen die mitotischen Chromosomen während der Metaphase dar. In dieser Form sind jeweils zwei Schwester-Chromatiden über ein Zentromer verbunden. Die kompakte Form der Metaphase Chromosomen erlaubt nur wenig Transkription der verpackten DNA (Christova und Oelgeschläger 2002) und spielt somit nur eine Rolle während der Mitose. In anderen Phasen des Zellzyklus liegt die kompakte Form nicht vor und reguläre Transkriptionsprogramme können geregelt ablaufen.

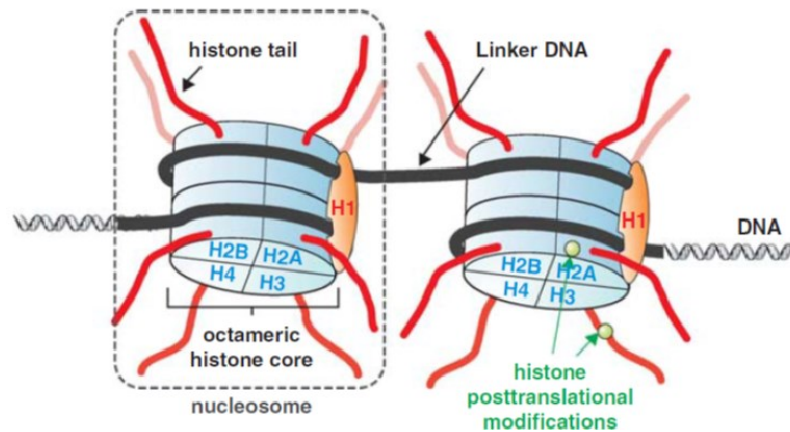
Die Zugänglichkeit des Chromatins kann sich in Abhängigkeit des Verpackungsgrades unterscheiden, es wird hier zwischen zugänglichem Euchromatin und geschlossenem Heterochromatin unterschieden. Bereiche aktiver Transkription befinden sich meist in euchromatischen Bereichen der DNA, da Transkriptionsfaktoren und Polymerasen entsprechend binden können. Bei Heterochromatin unterscheidet man weiterhin zwei Formen,

zum einen fakultativem Heterochromatin, das die inaktiven Gene und das inaktive weibliche X-Chromosom beinhaltet, und bei Bedarf wieder zugänglich werden kann, und zum anderen konstitutivem Heterochromatin, das generell nie exprimiert wird und zu dem die Telomere, das Zentromer und repetitive Sequenzen gezählt werden (Eymery et al. 2009; Schueler und Sullivan 2006). Durch die Ausbildung von Eu- und Heterochromatin ergibt sich die Möglichkeit zur Genregulation unabhängig von der reinen DNA Sequenz, man spricht in diesem Zusammenhang vom Forschungsfeld der Epigenetik. Unter Epigenetik versteht man genauer alle meiotisch und mitotisch vererbaren Genexpressionsmuster die nicht auf der Sequenz beruhen (Egger et al. 2004). Bei der epigenetischen Regulation von Genen kann man zwei Hauptgruppen unterscheiden, so kann zum einen die DNA selbst durch Methylierung modifiziert sein, zum anderen gibt es eine Vielzahl von Regulationsmöglichkeiten durch Modifikation der amino-terminalen Domänen der DNA-assoziierten Histone.

Die Rolle von DNA Methylierung an der C5 Position des Cytosin ist seit vielen Jahren als wichtiger Mechanismus beim epigenetischen Abschalten von Genen bekannt (Holliday und Pugh 1975). Diese Methylierung erfolgt in Säugern durch die Protein-Familie der DNA Methyltransferasen (DNMTs) und 40 % aller Gene weisen methylierbare CpG-Inseln, also größere DNA Bereiche mit erhöhtem G und C Gehalt, auf (Takai und Jones 2002). Abberante Methylierung und damit verbundene Abschaltung wichtiger Gene spielt bei der frühen Entstehung von Krebs eine zentrale Rolle (Jones und Baylin 2002).

Der Großteil der epigenetischen Regulation in Eukaryoten erfolgt jedoch auf Ebene der Histone. So spielen posttranslationale Modifikationen der Histone, wie Acetylierung und Methylierung der Lysine eine wichtige Rolle in der korrekten Regulation von zelltypspezifischen Genexpressionsmustern. Während der Großteil der jeweiligen Histon-Proteine des Histon-Oktamers unzugänglich im Inneren des Nukleosom gebunden liegen, ragen die amino-terminalen Enden der Histone aus dem Nukleosom heraus und können von weiteren Proteinen gebunden und modifiziert werden (Abb. 2).





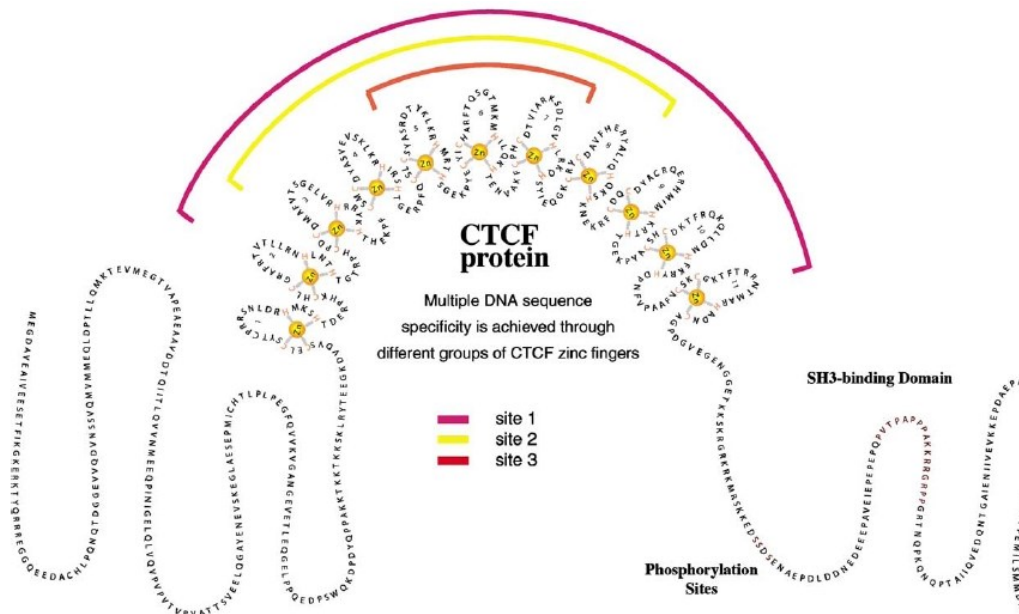
**Abb. 2: Schematische Darstellung von zwei Nucleosomen mit den zugänglichen Histone-Enden.** (Füllgrabe et al. 2011) Dargestellt sind schematisch die Histone und DNA die zwei Nucleosomen ausbilden. In grün sind als Beispiel zwei möglich posttranslationale Modifikationen der freien Histone-Enden angedeutet.

Die Erforschung der existierenden möglichen Histone-Modifikationen und ihrer funktionellen Wirkungen in den Zellen, wurde in den letzten Jahren zu einem der wichtigsten zellbiologischen Forschungsfelder. Neben den Haupt-Modifikationen Methylierung und Acetylierung (Allfrey et al. 1964) konnten weiterhin Ubiquitinierung (Wilkinson 2000), Biotinylierung (Ballard et al. 2002), Sumoylierung (Iniguez-Lluhi 2006), ADP-Ribosylierung (Koch-Nolte et al. 2008), Phosphorylierung (Bannister und Kouzarides 2011), Crotonylierung (Tan et al. 2011) und Citrullinierung (Christophorou et al. 2014) der Histone identifiziert werden. Der Wissensstand zu den jeweiligen Modifikationen variiert hier jedoch immens. Die wichtigsten gut erforschten Histone-Modifikationen lassen sich grob in zwei Gruppen einteilen, nämlich Histone-Modifikationen die aktivierend auf zelluläre Expression wirken und Modifikationen die im Zusammenhang mit Repression stehen. Inaktives Chromatin zeichnet sich meist durch generelle Hypoacetylierung, Methylierung von Argininen und Tri-Methylierung der Lysine bestimmter Histone-Enden (H3K9 und H3K27) aus, während aktives Chromatin oft durch Acetylierung (H3K9 und H4K8) und Di-/Tri-Methylierung (H3K4 und H3K36) von Lysinen charakterisiert wird (Schneider und Grosschedl 2007). Auch im Bereich von Enhancern konnten charakteristische Histone-Modifikations-Muster identifiziert werden und als charakteristische Modifikation aktiver Enhancer wurde Mono-Methylierung von H3K4 festgestellt (Calo und Wysocka 2013). Histone-Modifikationen können ihrerseits wieder von bestimmten Proteinen erkannt werden und weitere funktionelle Auswirkungen haben, so konnte beispielsweise ein funktioneller Zusammenhang zwischen Histone-Methylierung und DNA-Methylierung identifiziert werden der vermutlich auf der Interaktion der jeweiligen SET-Domänen der Histone-Methyltransferasen und DNA-Methyltransferasen basiert (Cedar und Bergman

2009). Zusammengefasst spielen Histon-Modifikationen und spezifische Kombinationen von ihnen die zentrale Rolle bei der Regulation von Chromatin in eukaryotischen Zellen, man spricht in diesem Zusammenhang heute auch vom Histon-Code (Jenuwein und Allis 2001). Anhand von genom-weiten Datensätzen für viele Histon-Modifikationen lassen sich inzwischen Vorhersagen über die funktionelle Rolle von DNA-Bereichen treffen und direkt bestimmten Chromatin-Klassen zuordnen (Filion et al. 2010). Bei der Abgrenzung von Chromatin Bereichen konnte der hoch konservierte, multifunktionale Faktor CTCF als wichtiger Regulator identifiziert werden (Weth et al. 2014), auf den im Folgenden näher eingegangen werden soll.

## 1.1 Der multifunktionale Faktor CTCF

Bei dem Faktor CTCF handelt es sich, wie inzwischen bekannt ist, um einen der wichtigsten strukturellen Faktoren bei der Regulation und Ausbildung von Chromatin-Strukturen. CTCF wurde 1990, im Zuge von Untersuchungen am c-myc Gen des Huhns, entdeckt und erhielt seinen Namen nach der identifizierten CCCTC DNA Sequenz an die er bindet (Lobanekov et al. 1990; Baniahmad et al. 1990). Das Protein CTCF besteht aus 728 Aminosäuren bei einem Molekulargewicht von ca. 82 kDa (Li et al. 2004), zeigt jedoch in SDS-PAGE ein abnormales Laufverhalten und ist bei 130 kDa zu detektieren (Klenova et al. 1997). Neben einer zentralen ZF-Domäne aus 11 Zinkfingern, die die DNA Bindung vermittelt, besitzt das Protein noch flankierende C-terminale und N-terminale Domänen. Diese Domänen spielen eine Rolle bei Protein-Protein Kontakten von CTCF und bieten die Möglichkeit zur Modifikation (z.B. Phosphorylierung, Sumoylierung) des Proteins. Unter Verwendung unterschiedlicher Zinkfinger kann CTCF an variierende DNA-Sequenzen binden (Ohlsson et al. 2001).



**Abb. 3: Strukturelle Darstellung der Domänen des Proteins CTCF** (Klenova et al. 2002). Durch Nutzung unterschiedlicher Zinkfinger Kombinationen (site 1-3) wird Bindung an unterschiedliche DNA Sequenzen vermittelt. Funktionell wichtige Phosphorylierungsstellen (rot) und ein vorhergesagtes SH3-Bindemotiv sind ebenfalls dargestellt.

Bei Zinkfingern handelt es sich um Protein-Strukturen die DNA-Bindung vermitteln können. Sie bestehen aus etwa 30 Aminosäuren und zeichnen sich durch Schleifenbildung um ein zentrales Zink-Ion ( $Zn^{2+}$ ) aus. In CTCF finden sich zwei Typen von Zinkfingern: Die ersten zehn sind vom Typ  $C_2H_2$ , d.h. das Zink-Ion ist durch zwei Cysteine und zwei Histidine gebunden, während der elfte vom Typ  $C_2HC$  ist. Bei diesem Typ ist das Zinkion von drei Cysteinen und nur einem Histidin eingebunden (siehe Abb. 3). Durch Bindung der Cysteine und der Histidine bildet sich eine fingerförmige  $\alpha$ -Helix Aminosäureschleife, die dann an DNA binden kann, aus.

Der Faktor CTCF zeigt eine sehr hohe Konservierung in Vertebraten. So beträgt die Übereinstimmung der Aminosäuresequenz von CTCF in Säugern ca. 90%. Die Zinkfinger-Domäne ist in den Gruppen Fische, Frösche (Burke et al. 2002), Vögel (Klenova et al. 1993), Nager und Mensch sogar zu ca. 100% identisch. Überraschenderweise konnte auch in *Drosophila* ein homologer Faktor identifiziert und gezeigt werden, dass CTCF aus Huhn an alle *Drosophila*-CTCF DNA-Bindestellen bindet (Moon et al. 2005). In Hefe, *C. elegans* und Pflanzen kommt jedoch keine Form von CTCF vor (Heger et al. 2012). Die hohe artübergreifende Homologie spricht für eine wichtige grundlegende Funktion des Proteins in zellulären Vorgängen. Einen weiteren Hinweis auf die essentiellen Funktionen von CTCF liefern Maus-CTCF-Knockout Experimente die embryonale Letalität in Folge des CTCF Verlustes zeigen (Wan et al. 2008; Moore et al. 2012).

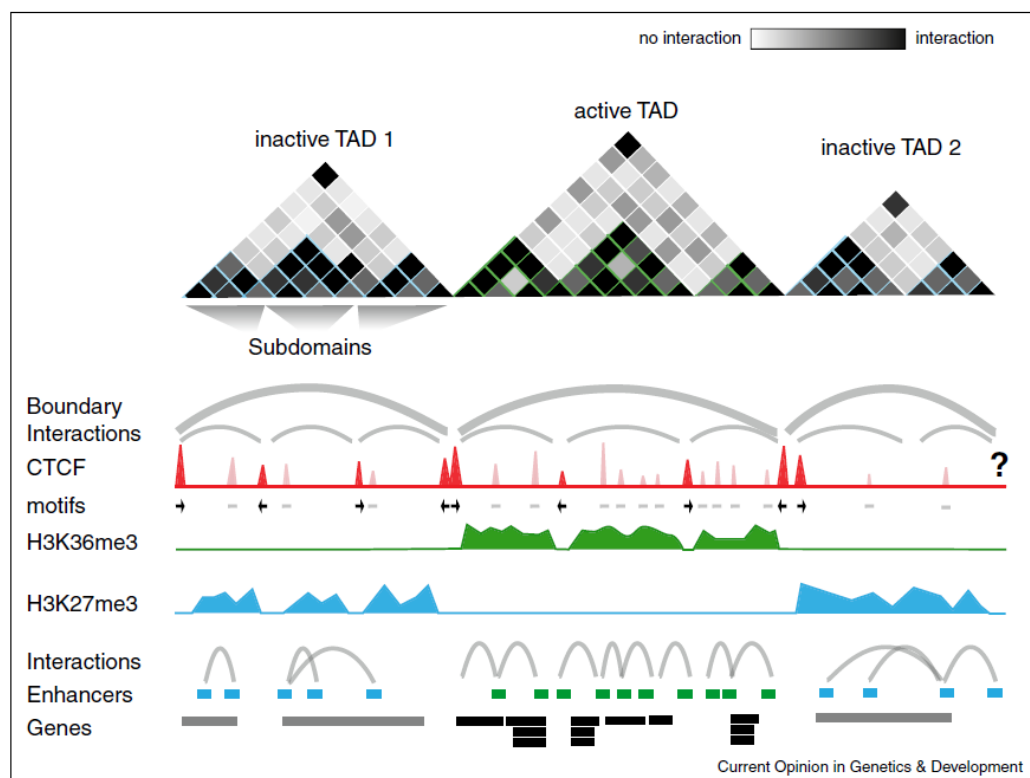
CTCF spielt bei vielen wichtigen Prozessen in der Zelle eine Rolle, so agiert CTCF etwa als genomweites Isolatorprotein das Grenzen zwischen Chromatinbereichen definiert, was sich auf die Expression von Genen auswirkt. Bei Untersuchungen am gut erforschten  $\beta$ -Globin Locus im Huhn konnte gezeigt werden, dass durch Bindung von CTCF an den F2-Isolator ein in der Nähe liegender Enhancer geblockt werden kann, was zur Hemmung der Transkription führt (Bell et al. 1999). Weiterhin spielt CTCF eine wichtige Rolle bei der X-Chromosom Inaktivierung (Chao et al. 2002) und sogar beim alternativen Spleißen (Shukla et al. 2011).

Das Aufkommen von *Next-Generation Sequencing* (NGS) ermöglichte in Kombination mit Chromatin-Immunopräzipitation das Bestimmen genomweiter Bindeprofile für CTCF. So finden sich in Säugern ca. 55.000-65.000 Bindestellen (Chen et al. 2012), wobei zwischen zelltyp-übergreifenden *High-Affinity* Bindestellen und zelltyp-spezifischen Bindestellen, die etwa 30-60 % ausmachen, zu unterscheiden ist (Kim et al. 2007). CTCF Bindestellen lassen sich sowohl in intergenischen Bereichen als auch in Genregionen und Promotorbereichen nachweisen. Ein Phänomen mit dem sich teilweise die Ausbildung zelltyp-spezifischer Bindemuster erklären lässt, ist die Anwesenheit von CpGs in der CTCF Konsensus-Sequenz und eine reduzierte Bindung von CTCF an Bindestellen, wenn diese CpGs sich durch Methylierung auszeichnen (Wang et al. 2012). Der Verlust von CTCF Bindung im Zuge von Methylierung kann gravierende Auswirkungen auf zelluläre Vorgänge haben. So ist beispielsweise der Promotor von PTGS2 in normalen Zellen von CTCF gebunden und das Gen wird transkribiert, in einigen Krebsformen ist diese CTCF Bindung durch DNA Methylierung der Bindestelle verloren und das Gen inaktiv (Kang et al. 2015). Es lassen sich jedoch nicht alle Unterschiede der CTCF-Bindemuster zwischen Zelltypen anhand von differentieller Methylierung erklären, daher müssen weitere Mechanismen vorliegen die zur Ausbildung zelltyp-spezifischer CTCF Bindung führen.

## 1.2 CTCF und Chromosomenarchitektur

Im Zuge der Entwicklung von experimentellen Techniken zur Untersuchung von DNA Interaktionen über große Distanzen, wie etwa 3C (*Chromosome conformation capture*) (Dekker et al. 2002), konnte die Rolle von CTCF bei der Ausbildung solcher Kontakte untersucht werden. Anhand von 3C kann die Interaktion zweier bekannter DNA Abschnitte des Genoms mittels PCR analysiert werden. Weiterentwicklungen der Technik, 5C (Dostie et al. 2006) und Hi-C (Belton et al. 2012), ermöglichten dann einen ungerichteten Blick auf alle DNA Interaktionen im Genom. CTCF konnte als ein zentraler Faktor

identifiziert werden, der an interagierenden DNA Bereichen angereichert vorliegt. Erste Untersuchungen am H19-IGF2 Locus zeigten die Fähigkeit von CTCF durch Dimerisierung DNA Schleifen zu bilden (Pant et al. 2004). Zusammen mit dem Cohesin Protein-Komplex vermittelt CTCF diese DNA Kontakte auch über große Entfernungen hinweg und an vielen genomischen Bindestellen (Hadjur et al. 2009). Cohesin war bis dahin hauptsächlich für seine wichtige Rolle bei der Ausrichtung der Schwester-Chromatiden während der Mitose bekannt. Im Zuge der genomweiten Analyse der Interaktionen konnten die Rolle von CTCF bei der Ausbildung von DNA Interaktionen bestätigt werden (Phillips-Cremins et al. 2013). Weiterhin konnten auch größere, strukturelle Einheiten im Genom, sogenannte TADs (*Topologically associated domains*) identifiziert werden (Dixon et al. 2012). Unter TADs versteht man Abschnitte des Genoms die bevorzugt innerhalb, miteinander interagieren und wenig Interaktion mit weiteren umliegenden DNA Abschnitten zeigen. Gene die zusammen innerhalb einer solchen Domäne vorliegen zeigen häufig Ko-Regulation (Abb. 4). CTCF ist in den Bereichen der TAD Grenzen angereichert (Dixon et al. 2012).

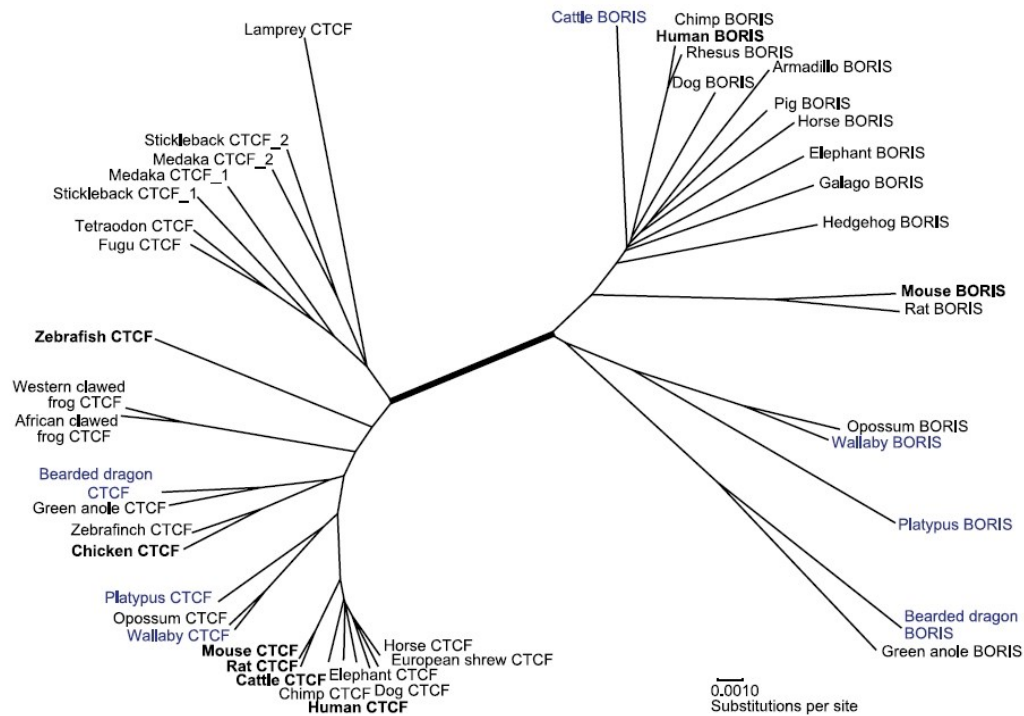


**Abb. 4: Schematische Darstellung von TADs, beispielhafter CTCF Bindung und Anwesenheit aktiver & repressiver Histon-Modifikationen** (Ali et al. 2016). Dargestellt sind drei hypothetische Model-Domänen inklusive der Interaktionen, CTCF-Bindung, Anwesenheit von H3K36me3 & H3K27me3 und assoziierten Genen.

Unter Verwendung der modernen CRISPR Methode (Jinek et al. 2012), zur gezielten Änderung von genomischen Sequenzen *in vivo*, konnte gezeigt werden, dass auch die Orientierung der CTCF Bindestellen von grundlegender Wichtigkeit für die Ausbildung von TADs ist (Guo et al. 2015). So zeigen sich jeweils die äußersten CTCF Bindestellen einer TAD nach innen orientiert und Änderung der Orientierung führt zur Zerrüttung der Domänenstruktur. Zusammengefasst gilt CTCF somit inzwischen als der zentrale verbindende Faktor von zellulärer Genomstruktur und funktionellen Expressionsmustern (Ong und Corces 2014).

### 1.3 CTCFL: ein paraloger Faktor von CTCF

Bis zum Jahre 2002 wurde CTCF als alleinstehender Faktor ohne bekannte vergleichbare Proteine verstanden. Dann wurde ein CTCF nahestehender Faktor in Mensch und Maus identifiziert, der sich über eine große Homologie der ZF-Domäne (Abb. 6) auszeichnet aber komplett verschiedene C- und N-Termini besitzt (Loukinov et al. 2002). Er wurde *CTCF-like* (CTCFL) oder auch *Brother of the regulator of imprinted sites* (BORIS) getauft. Über Analysen der kompletten Chromosomen-Loci in denen die Gene für beide Faktoren liegen, konnte die Abstammung CTCFLs von CTCF dargestellt werden. So entstand der gesamte, das CTCFL-Gen enthaltende, Locus 20q13 wahrscheinlich durch eine Duplikation des CTCF enthaltenden Bereichs 16q22 (Klenova et al. 2002).



**Abb. 5: Verwandtschaftsbeziehungen der Mitglieder der CTCF/BORIS Genfamilie** (Hore et al. 2008). Die Länge der Striche stellt die Zahl der Unterschiede zwischen den einzelnen Proteinen dar. Fett geschriebene Faktoren waren bereits vor dieser Studie bekannt und annotiert. Weiterhin werden die neu identifizierten Faktoren ebenfalls dargestellt: blau steht für experimentellen Nachweis während die schwarzen Faktoren nur durch *in-silico* Analysen identifiziert wurden.

CTCF-L (CTCF-Like) kommt wie auch CTCF in mehreren Spezies vor. Ursprünglich war CTCF-L Expression nur für Mensch und Maus bekannt, inzwischen konnte CTCF-L jedoch in vielen Vertebraten durch *in-silico* Analysen sowie experimentell identifiziert werden (Abb. 5). Man geht davon aus, dass CTCF-L früh in der Evolution, vor dem Abspalten der Säuger von den Reptilien, durch eine vollständige Genduplikation von CTCF entstand (Hore et al. 2008). Der Grad der zwischenartlichen Konservierung ist für den Faktor CTCF heute jedoch größer als bei CTCF-L was evolutiv für eine zentralere Rolle von CTCF spricht.

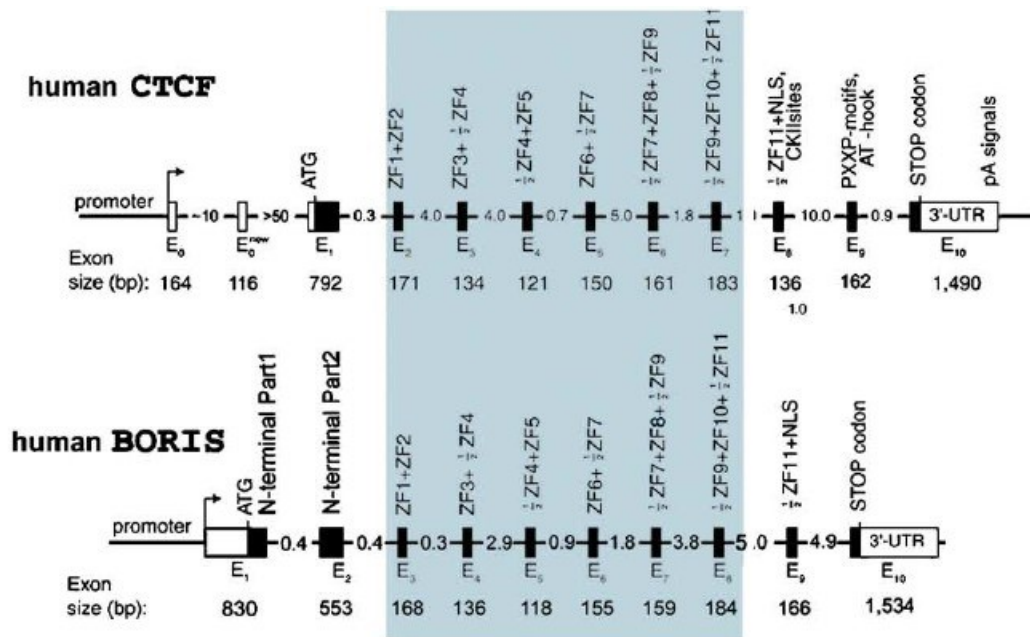
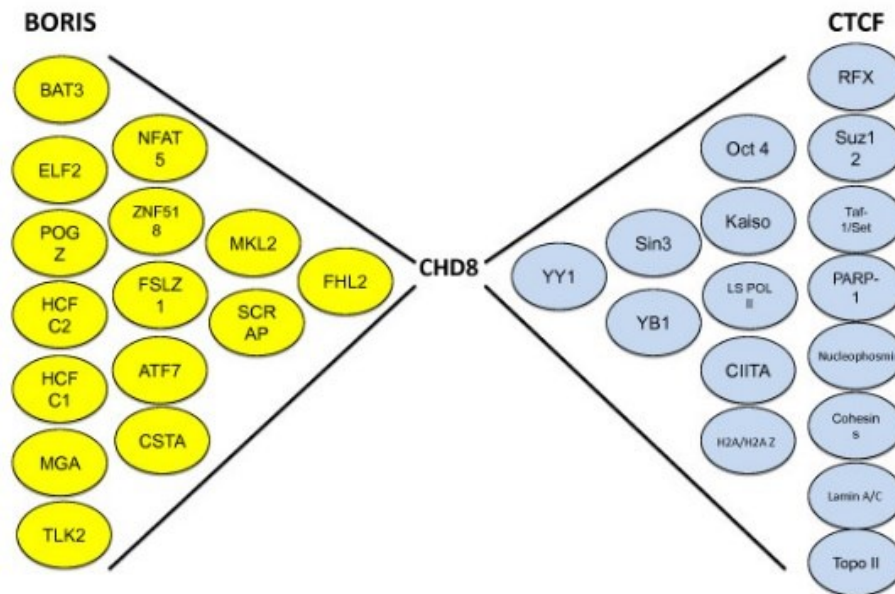


Abb. 6: Vergleich der Exon/Intron-Struktur der Gene für CTCF und CTCFL (BORIS) (Klenova *et al.*, 2002). Der Bereich der Homologie zwischen beiden Genen ist durch eine graue Box dargestellt.

Die fast identischen ZF-Domänen von CTCF und CTCFL (Abb. 6) sprechen für die Fähigkeit beider Proteine an die gleichen DNA-Sequenzen zu binden während die unterschiedlichen C- und N-Termini auf unterschiedliche Funktionen und mögliche unterschiedliche Interaktionspartner hindeuten. CTCFL kann im Gegensatz zu CTCF auf Methylierungs-unabhängige Weise an DNA binden (Nguyen *et al.* 2008b). Die Expression von CTCFL ist im Mensch und auch in der Maus jeweils auf die Hoden beschränkt. Entgegen den ersten Berichten zur Expression von CTCFL, die eine sich gegensätzliche ausschließende Expression zu CTCF beschrieben (Loukinov *et al.* 2002), ist heute die gleichzeitige Expression beider Proteine in Zellen der Keimbahn bestätigt (Sleutels *et al.* 2012). Betrachtet man die bekannten Interaktionspartner der beiden Proteine im menschlichen System so zeigen sich die erwarteten großen Unterschiede und es kann nur ein gemeinsamer Faktor CHD8 identifiziert werden (Abb. 7). Diese Interaktionen erfolgt jedoch bei beiden Proteinen über unterschiedliche Domänen, so interagiert bei CTCFL der N-Terminus mit CHD8 (Nguyen *et al.* 2008a), bei CTCF die ZF-Domäne (Ishihara *et al.* 2006).





**Abb. 7: Bekannte Interaktionspartner von CTCFL (gelb) und CTCF (blau) im Menschen** (Necochea-Campion et al. 2011). Dargestellt sind beschriebene Interaktionspartner der beiden Proteine. Nur für CHD8 ist eine Interaktion mit beiden Faktoren bekannt.

Ein weiterer Faktor der sowohl mit CTCF als auch CTCFL interagiert ist UBC (van de Nobelen et al. 2010). UBC interagiert mit beiden Proteinen über deren ZF-Domäne und kann somit wohl keine funktionellen Unterschiede in der Funktion beider Proteine erklären.

CTCFL *Knockout*-Experimente konnten zeigen, dass CTCFL eine wichtige Funktion bei der Testis Entwicklung spielt (Suzuki et al. 2010; Sleutels et al. 2012). Mäuse ohne CTCFL zeichnen sich durch kleinere Hoden und reduzierte Fertilität aus. Auch wichtige Testis-spezifische Gene wie *Gal3st1* und *Tsp50* zeigen Herabregulation in Folge des CTCFL *Knockout*. Für *Tsp50* konnte auch die direkte positive Regulation durch CTCFL Bindung im Promoter Bereich des Gens identifiziert werden (Kosaka-Suzuki et al. 2011). Somit kann CTCFL als wichtiger Faktor bei der korrekten Testis Entwicklung in der Maus angesehen werden.

Bei der Expression von CTCFL handelt es sich um einen komplizierten, nicht gut verstandenen Vorgang, so gibt es unterschiedliche Angaben über eine mögliche Zahl von Isoformen und unterschiedliche Promotoren. Es wurden drei unterschiedliche Promotoren A, B und C beschrieben die jeweils auf unterschiedliche Transkriptions-Startpunkte wirken und 23 mRNAs mit variierenden 5'-UTRs kodieren (Pugacheva et al. 2010). Alle drei Promotoren spielen eine Rolle bei Aktivierung von CTCFL-Transkription in der

Spermatogenese, während bei der Krebsentstehung vor allem Promotor A und C genutzt werden (Renaud *et al.*, 2007). Diese Existenz von alternativen Promotoren und Transkripten erschwert die Untersuchungen der grundlegenden Mechanismen von CTCFL, da alternative Transkripte unterschiedliche Auswirkungen auf die epigenetische Reprogrammierung in normalem Testisgewebe und die mögliche Tumorentstehung haben könnten.

## 1.4 CTCFL-Expression in Krebs – Unterschiedliche Meinungen

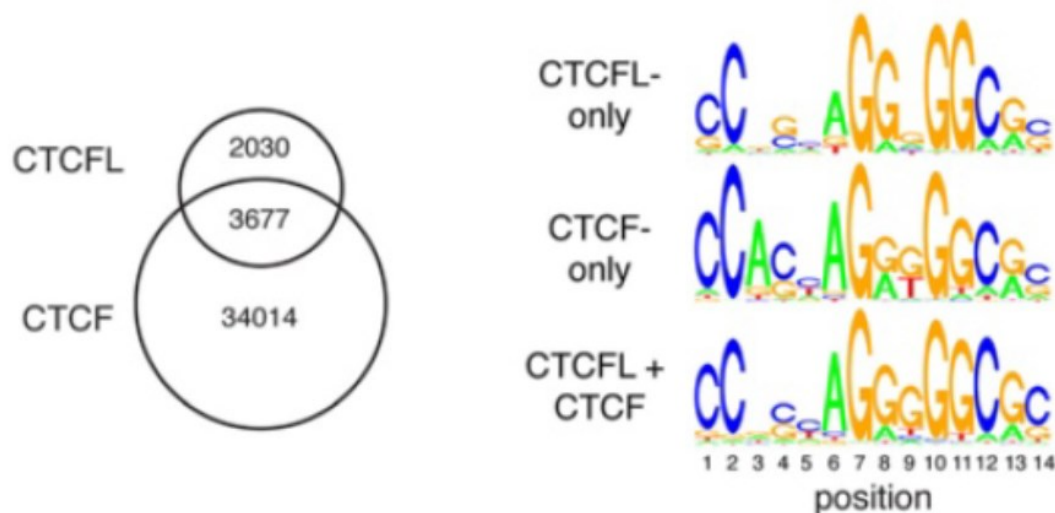
CTCFL wurde ursprünglich zur Klasse der *Cancer-Testis*-Antigene (CTA) gezählt, in der alle Gene zusammengefasst sind, die normale Expression nur in Testisgewebe und aberrante Expression in vielen Krebsarten zeigen. CTAs stellen aufgrund dieser charakteristischen begrenzten Expressionsmuster ein interessantes Ziel für die Entwicklung von krebsspezifischen Immuntherapien dar.

In mehreren Studien wurde ein Zusammenhang zwischen CTCFL-Expression und Krebs angedeutet. Das CTCFL Gen liegt in einem mit bestimmten Krebs-Formen assoziierten „hot-spot“ 20q13, d.h. einem Bereich genomischer DNA der in Krebs häufig amplifiziert wird, und es wurde postuliert, dass in dieser Region ein starker immortalisierender oder transformierender Faktor liegen muss (Tanner *et al.*, 1994). Bei CTCFL könnte es sich um eben diesen Faktor handeln. Bei der Analyse von CTCFL Expression in Krebszelllinien und primären Tumoren gibt es abhängig von der jeweiligen Arbeitsgruppe stark variierende, ja sogar gegensätzliche Ergebnisse. So lässt sich CTCFL einmal in allen untersuchten aus Krebsgeweben hergestellten Zelllinien (Renaud *et al.* 2007) und sogar normalen Zellen (Jones *et al.* 2011) nachweisen und in anderen Fällen wird CTCFL so gut wie gar nicht detektiert (Yoon *et al.* 2011). Greift man veröffentlichten Untersuchungen zu Brustkrebs heraus werden die Unterschiede noch einmal deutlicher. In einer Untersuchung konnte CTCFL in allen untersuchten Brustkrebszelllinien und 70,7% aller untersuchten Gewebe aus Primärtumoren nachgewiesen werden (D'Arcy *et al.*, 2008) während in einer späteren Studie, unter genauer Analyse durch mehrere sensible Methoden, nur eine von zwölf untersuchten Brustkrebszelllinien, und von acht untersuchten Primärtumoren kein einziger, CTCFL Expression aufwies (Hines *et al.* 2010). Unabhängig von diesen ungeklärten gegensätzlichen Resultaten gibt es mehrere Versuche CTCFL als Marker und therapeutisches Ziel in einigen Krebsformen einzusetzen. So zeigte die Analyse einer spezifischen CTCFL Isoform in Kehlkopf-Krebs Patienten die Möglichkeit zur Vorhersage einer Wiederkehr der Krankheit (Novak Kujundžić *et al.* 2014), CTCFL konnte

ebenso im Speiseröhren-Krebs als prognostischer Marker identifiziert werden (Okabayashi et al. 2012) und ein CTCFL-basierter Impfstoff war wirksam in der Bekämpfung einer aggressiven Form von Brustkrebs im Maus Modell (Mkrtichyan et al. 2011). Somit gibt es Hinweise auf eine mögliche Rolle von CTCFL in Krebs, aber abschließende klare Ergebnisse stehen weiterhin aus.

## 1.5 Genomweite Bindemuster von CTCFL

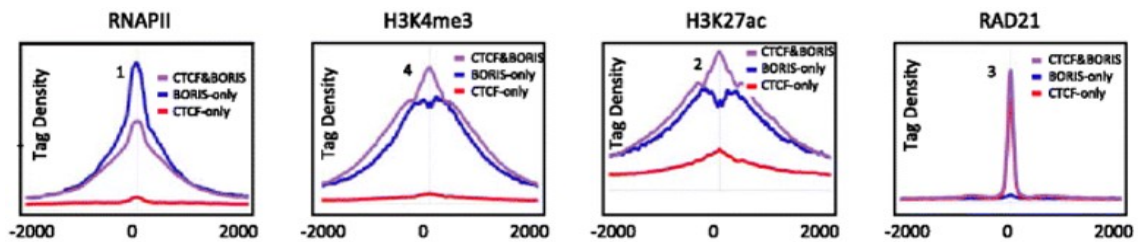
Unter Verwendung spezifischer CTCFL Antikörper gelang es in den letzten Jahren genom-weite Bindemuster für CTCFL zu generieren. In der ersten Studie zur genom-weiten Bindung von CTCFL unserer Arbeitsgruppe wurden mittels CTCFL überexprimierender Maus ES-Zellen und anschließender ChIP-Sequenzierung ca. 5.700 CTCFL Bindestellen identifiziert, von denen zwei Drittel mit CTCF Bindestellen überlappten (Sleutels et al. 2012).



**Abb. 8: Ergebnisse der Analyse genom-weiter Bindung von CTCFL in mES-Zellen.** (Sleutels et al. 2012) Dargestellt ist sowohl ein Venn Diagramm, das die Zahl der identifizierten CTCF- und CTCFL-Bindestellen und den Anteil überlappender Bindestellen zeigt, und jeweils das identifizierte Bindemotiv der einzelnen Gruppen von Bindestellen.

Auch die durchschnittlichen Bindemotive der beiden Proteine wurden errechnet und zeigten keine großen Unterschiede (Abb. 8). Bei genaueren Untersuchungen der Verteilung der CTCFL Bindestellen im Genom zeigte sich eine erhöhte Assoziation mit zugänglichen Chromatin-Bereichen wie etwa Transkriptionsstartstellen und Exons, und auch Korrelations-Analysen mit der aktiven Histon-Modifikation H3K4me3 und elongierender Polymerase II (Serin 5 phosphoryliert) zeigten für CTCFL Bindestellen erhöhte Anreicherung im Vergleich zu CTCF. Weitere genom-weite Datensätze in humanen Zellen konnten

diese beobachtete Präferenz von CTCFL, in zugänglichen Chromatin Bereichen zu binden, bestätigen (Pugacheva et al. 2015). So zeigt CTCFL in K562 Zellen die gleiche Assoziation mit H3K4me3 und Polymerase II. Weiterhin wurden H3K27ac und Rad21, eine Komponente des Cohesin Komplexes, untersucht. H3K27ac zeigt die bekannte positive Korrelation mit alleinigen CTCFL Bindestellen während Rad21 bei diesen, im Vergleich zu den CTCF gebundenen Bereichen, stark unterrepräsentiert ist (Abb. 9).



**Abb. 9: Korrelationsanalysen zu CTCFL Bindestellen in K562 Zellen.** (Pugacheva et al. 2015) Dargestellt ist für vier untersuchte Faktoren/Modifikationen (RNAPII, H3K4me3, H3K27ac und RAD21) die Anreicherung über drei Untergruppen von CTCF und CTCFL Bindestellen. Dabei ist in Rot die Bindung über nur CTCF gebundene Stellen, in Blau über nur CTCFL gebundene und in Lila über von beiden Faktoren gebundenen Bindestellen gezeigt.

Weitergehende bioinformatische Analysen der genom-weiten CTCFL Bindedaten identifizierten eine gleichzeitige Bindung von CTCF und CTCFL an genomische Bereiche die sich durch zwei nah beieinander liegende CTCF-Motive auszeichnen (Pugacheva et al. 2015). Durch diese Doppel-Motive besteht die Möglichkeit, dass überexprimiertes CTCFL keine dramatischen funktionellen Folgen auf die wichtige strukturelle Bindung von CTCF hat, zu einer echten Konkurrenz um Bindestellen kommt es nicht, da weitere Bindemöglichkeiten bestehen. Somit bindet CTCFL an CTCF Bindestellen bevorzugt dabei jedoch in offenen Chromatin Bereichen liegende Bindestellen.

## 1.6 Zielsetzung der Arbeit

Da zum Zeitpunkt des Beginns dieser Arbeit das heutige Wissen über die genomweite Verteilung von CTCFL im Genom noch nicht vorlag sollte im Rahmen dieser Arbeit untersucht werden in welcher Form sich CTCF und CTCFL Bindestellen unterscheiden oder übereinstimmen. Da beide Faktoren sich durch die identische 11-Zinkfinger DNA-Bindedomäne auszeichnen aber die N- und C-terminalen Domänen grundverschieden sind sollte die Hypothese überprüft werden, dass beide Faktoren an identische Bindestellen binden dort aber unterschiedliche Funktionen ausüben und Ko-Faktoren rekrutieren. Hierzu sollten CTCFL Überexpressionssysteme in Mensch und Maus etabliert und dann

die Bindung anhand von ChIP-Experimenten untersucht werden. Mögliche Befunde sollten durch genomweite Korrelation unter Generierung von spezifischen ChIPseq Datensätzen überprüft werden.

Weitere Analysen im murinen Differenzierungssystem sollten der Frage nachgehen, ob CTCFL Bindung möglicherweise anhand von Änderungen des Chromatin-Status modulierbar sein könnte, also z.B. durch induzierte Änderung von Euchromatin zu Heterochromatin ein Verlust von CTCFL Bindung erzielt werden kann. Neben dem möglichen Einfluss von DNA-bindenden Transkriptionsfaktoren und Histon-Modifikationen auf CTCF/CTCFL-Bindung sollten in einem weiteren Ansatz mögliche Unterschiede der DNA-Sequenz der spezifischen CTCF und CTCFL Bindestellen identifiziert werden.

Insgesamt sollte somit ein größeres Verständnis zum Bindeverhalten von CTCFL und dessen bestimmenden Faktoren erlangt werden, um damit eventuell Rückschlüsse auf die zelluläre Funktion von CTCFL, die nach wie vor weitgehend unklar ist, zu ziehen.

## 2 Material

### 2.1 Geräte

Begasungsbrutschrank	Heracell 240, Heraeus
Computer	Acer
Eismaschine	Scotsman
Elektrophoreseapparatur	Peqlab
Entwicklermaschine	Kodak M35 X-Omat Processor
Gefrierschrank (-80°C)	Revco
Kühlschrank	Bosch
Magnetrührer	MAGV Ikamag
Mikroskope	Telaval 31, Zeiss Axiophot, Zeiss
Mikrowelle	Privileg
PCR-Cycler	Mastercycler Gradient, Eppendorf GeneAmp 9600, Perkin Elmer
Pipetten (2µl, 20µl, 200µl, 1000µl)	Gilson
Pipettierhilfe	Pipetboy, Integra Biosciences
Quarzglasküvette	bichrom (50 µl) Hellma (400 µl)
Real-Time-PCR-Cycler	RG-3000, Corbett Research
Rotator	neoLab
Schüttler	Heidolph Instruments
Spannungsgerät	IBI
Spektralphotometer	Ultrospec 3100 pro, Amerham Biosciences
Tubesealer	Beckman
Ultraschall	Sonifier 250, Branson
VersaDoc™ Imaging System	Bio-Rad
Vortexer	Vortex-Genie 2, Scientific Industries

Waage	PM3000, Mettler PRS 320-3, Kern
Wasserbad	Kötterman
Western-Blot Apparatur	Mini Protean II™, Bio-Rad
Zählkammer	Fuchs-Rosenthal
Zentrifugen	Biofuge Pico, Heraeus Multifuge 3, Heraeus J2-MC, Beckman 5417R, Eppendorf Sorvall® MC 12V, Du Pont

## 2.2 Verbrauchsmaterialien

6-well Schalen	Greiner Bio-One
Einfriergefäße (2 ml)	Nunc
Einmalküvetten	Roth
Einwegkanülen (Ø 0,90 x 40 mm)	Braun
Einwegspritzen (5 ml)	Henke Sass Wolf
Filme	BioMax XAR, Kodak
Filtertips	Greiner Bio-One
Gewebekulturschalen (Ø 94 mm & 145 mm)	Greiner Bio-One
Handschuhe	Roth
Immobilon™-P Transfermembran	Millipore
Pasteurpipetten	Roth
PCR Reaktionsgefäße (0,2 ml)	Biozym Scientific
Petrischalen (Ø 145 mm)	Greiner Bio-One
Pipettenspitzen	Greiner Bio-One
Reaktionsgefäße (0,5ml)	Peqlab
Reaktionsgefäße (1,5 ml & 2 ml)	Eppendorf

Real-Time-PCR Reaktionsgefäße (0,1 ml)	Corbett Research
Skalpelle	Braun
Whatman-Papier	Whatman
Zellspachtel	TPP
Zentrifugenröhrchen (15 ml & 50 ml)	Falcon

## 2.3 Chemikalien

Agarose	Invitrogen
APS	Serva
CsCl	Roth
DMEM <sup>1</sup>	Invitrogen
EDTA	Roth
Ethanol	Merck
Ethidiumbromid	Merck
Formaldehyd	Calbiochem
Fötales Kälberserum	Invitrogen
Isopropanol	Merck
MEM <sup>1</sup>	Invitrogen
Methanol	Merck
Milchpulver (Blotting grade)	Roth
NaCl	Roth
NaHCO <sub>3</sub>	Merck
PenStrep-Lösung	Invitrogen
Phenol	Roth
Retinsäure	Merck
Roti®-Load	Roth
Roti®-Quant	Roth

---

<sup>1</sup> Zellkulturmedium



RPMI <sup>2</sup>	Invitrogen
SDS	Roth
TEMED	Sigma
Tris	Merck
Tween® 20	Roth

## 2.4 Enzyme

CIAP	Amersham
<i>Pfu</i> -Polymerase	Fermentas
Proteinase K	Fermentas
Restriktionsendonukleasen	Fermentas
RNase A	Fermentas
T4 Ligase	Fermentas
T4 PNK	Fermentas
Trypsin	Invitrogen

## 2.5 Komplettsysteme

GoTaq® Flexi DNA Polymerase	Promega
High Purity Plasmid Midiprep System	Marligen
jetPEI™ Transfection Reagent	PolyPlus-Transfection
QIAquick® Gel Extraction Kit	Qiagen
RNeasy® Mini Kit	Qiagen
RQ1 RNase-Free DNase	Promega
Transcriptor First Strand cDNA Synthesis Kit	Roche

---

<sup>2</sup> Zellkulturmedium

## 2.6 Größenstandards

### 2.6.1 DNA Längenstandardmarker

GeneRuler™ 100 bp DNA Ladder Plus Fermentas	pUC19 DNA/ <i>MspI</i> ( <i>HpaII</i> ) Marker, 23 Fermentas	λDNA / <i>Eco130I</i> ( <i>StyI</i> ) Marker, 16 Fermentas
bp	bp	bp
3000	501	19329
2000	489	7743
1500	404	6223
1000	331	4254
900	242	3472
800	190	2690
700	147	1882
600	111	1489
500	110	925
400	67	421
300		
200		
100		

### 2.6.2 Molekulargewichtsmarker für Proteine

Prestained Molecular Weight Marker (SDS7B2) Sigma	MagicMark™ XP Western Protein Standard Invitrogen
kDa	kDa
180	220
116	120
90	100
58	80
48,5	60
36,5	50
26,6	40
	30
	20

## 2.7 Antibiotika

Ampicillin	Invitrogen	im Medium: 100 µg/ml
Doxyzyklin	Applichem	im Medium: 1 µg/ml
Kanamycin	Gibco	im Medium: 50 µg/ml
Puromycin	PAA	im Medium: 1 µg/ml

## 2.8 Antikörper

$\alpha$ CTCF (N2.2)	F. Sleutels
$\alpha$ CTCFL (ab18337)	Abcam
$\alpha$ CTCFL (H120)	Santa Cruz
$\alpha$ CTCFL 6	F. Sleutels
$\alpha$ FLAG (M2)	Sigma
$\alpha$ GAPDH (FL-335)	Santa Cruz
$\alpha$ H3 (ab1791)	Abcam
$\alpha$ H3K27me3 (ab6002)	Abcam
$\alpha$ H3K4me3 (ab8580)	Abcam
$\alpha$ H3K9ac (ab4441)	Abcam
$\alpha$ H3K9me3 (ab8898)	Abcam
$\alpha$ RAD21 (ab992)	Abcam
$\alpha$ SA1 (ab4457)	Abcam
$\alpha$ SA2 (ab4463)	Abcam
$\alpha$ Rabbit IgG (ab27472)	Abcam
$\alpha$ Goat IgG	

## 2.9 Eukaryotische Zelllinien

HeLa (ATCC-Nr. CCL-2)
K-562 (ATCC-Nr. CCL-243)
NIH/3T3 (ATCC-Nr. CRL-1658)
P19 (ATCC-Nr. CRL-1825)

## 2.10 Plasmide

pBI-EGFP	Clontech
pBI-EGFP-FLAG-hCTCFL	P. Bergmaier, Diplomarbeit
pBSK	Stratagene
pcDNA3-FLAG-hCTCFL	J. Leers
pTA-N	J. Leers

pBI-EGFPP-FLAG-mCTCF

S. Dienstbach

## 2.11 Oligonukleotide

### 2.11.1 Primer für RT- und qRT-PCR (Mensch)

<i>ACTB</i> as (Vandesompele et al. 2002)	5'-AAG GGA CTT CCT GTA ACA ATG CA-3'
<i>ACTB</i> s (Vandesompele et al. 2002)	5'-CTG GAA CGG TGA AGG TGA CA-3'
<i>AKR1C3</i> as	5'-CCG GTT GAA ATA CGG ATG AC-3'
<i>AKR1C3</i> s	5'-GGA GGC CAT GGA GAA GTG TA-3'
<i>B2M</i> as (Vandesompele et al. 2002)	5'-TCT CTG CTC CCC ACC TCT AAG T-3'
<i>B2M</i> s (Vandesompele et al. 2002)	5'-TGC TGT CTC CAT GTT TGA TGT ATC T-3'
<i>CTCF</i> as	5'-GCT TGT ATG TGT CCC TGC TGG CA-3'
<i>CTCF</i> s	5'-TTA CAC GTG TCC ACG GCG TTC-3'
<i>GAPD</i> as (Vandesompele et al. 2002)	5'-GGC ATG GAC TGT GGT CAT GAG-3'
<i>GAPD</i> s (Vandesompele et al. 2002)	5'-TGC ACC ACC AAC TGC TTA GC-3'
<i>GPR56</i> as	5'-TAG ATG ACG CCC TCT GGA GT-3'
<i>GPR56</i> s	5'-GCC ATC TTC CTG CAC TTC TC-3'
<i>GSTM3</i> as	5'-CCT TCA GGT TTG GGA ACT CA-3'
<i>GSTM3</i> s	5'-TAA TGG ATT TCC GCA CAC AA-3'
<i>RPL13A</i> as (Vandesompele et al. 2002)	5'-TTG AGG ACC TCT GTG TAT TTG TCA A-3'
<i>RPL13A</i> s (Vandesompele et al. 2002)	5'-CCT GGA GGA GAA GAG GAA AGA GA-3'
<i>UBC</i> as (Vandesompele et al. 2002)	5'-TGC CTT GAC ATT CTC GAT GGT-3'
<i>UBC</i> s (Vandesompele et al. 2002)	5'-ATT TGG GTC GCG GTT CTT G-3'

### 2.11.2 Primer für RT- und qRT-PCR (Maus)

<i>Actb</i> as	5'-CCT CGT AGA TGG GCA CAG T-3'
<i>Actb</i> s	5'-CTG TAT TCC CTT CCA TCG TG-3'
<i>Csn3</i> as	5'-TTG TAG GCA TGG CAA GAA AGG-3'
<i>Csn3</i> s	5'-CAA ACC CTA CTG CCA AGC AAG-3'
<i>Ctcf1</i> as	5'-TTC GCG GCT TCC TTC CAT CCC-3'
<i>Ctcf1</i> s	5'-GGC AAA GGC TTT TCC CGC TGG-3'
<i>Foxa1</i> as	5'-TTG GCG TAG GAC ATG TTG AA-3'
<i>Foxa1</i> s	5'-CAT GAG AGC AAC GAC TGG AA-3'
<i>Gpr56</i> as	5'-TTG GTG TTC TGC ACG ACA AT-3'
<i>Gpr56</i> s	5'-AGG CAT ACT CGC TGT TGC TT-3'
<i>Grb10</i> as	5'-CGG TTG CTC CTA GCT CCT TA-3'
<i>Grb10</i> s	5'-CTG CCA TCT CCA GAT CTT CC-3'
<i>Hoxa5</i> as	5'-AGG TAG CGG TTG AAG TGG AA-3'
<i>Hoxa5</i> s	5'-GCG CAA GCT GCA CAT TAG T-3'
<i>Oct4 /Pou5f1</i> as (Palumbo et al. 2002)	5'-GGC GAT GTG GCT GAT CTG CT-3'
<i>Oct4 /Pou5f1</i> s (Palumbo et al. 2002)	5'-ACA CCT GGC TTC GGA TTT CG-3'
<i>Tdh</i> as	5'-GGC CGC TCC TCT TCT TTA GT-3'
<i>Tdh</i> s	5'-CAG ATC ACC TAC TGC GTG GA-3'
<i>Tspan6</i> as	5'-CAG AAA TGG GAG TCG TTG CT-3'
<i>Tspan6</i> s	5'-GGA GAT ATG AGC CCG TGT GT-3'

### 2.11.3 Primer für ChIP/FAIRE-PCR und ChIP/FAIRE-qPCR (Maus)

<i>Abca 5</i> s	5'-GCA ATT CCA GTT CCA GAG GA-3'
<i>Abca5</i> as	5'-GTT AGT GCT TCG CCT TCC TG-3'
<i>Accs</i> as	5'-TGA AAA CCC AAG GTC CAA AG-3'
<i>Accs</i> s	5'-GGC CTC AGA GAC TTC CAG AC-3'
<i>Adcy5</i> as	5'-CCA CAG AAC TGC CTC CAT CC-3'
<i>Adcy5</i> s	5'-GAG GAG CGT CGA GGA AAA GG-3'
<i>Adnp2</i> as	5'-AGA TCC CCT GGA ACT GGA CT-3'

---

Adnp2 s	5'-GAG AAG CCT CGG GGA TAG AG-3'
Alpl as	5'-CAA CCA TCG AGG GAC AAA GT-3'
Alpl s	5'-CGC CTT GCC TCT CTC TTT C-3'
Amylase as (Sleutels et al. 2012)	5'-AGC AAT GAT GTG CAC AGC TGA A-3'
Amylase s (Sleutels et al. 2012)	5'-AAT TCT CCT TGT ACG GCT TCG TG-3'
Atp8b3 as	5'-ATG CTC ACG AGT TGT GCA TC-3'
Atp8b3 s	5'-TAA GCT CCA CCT CGG TGT CT-3'
Cdh22 as	5'-CTC CTG TGG GTC ATC CAT CT-3'
Cdh22 s	5'-CGA GTC CCT GGG AAC ACT TA-3'
Coll2a1 as	5'-ATT CTT GGC TGG CTG AGA TG-3'
Coll2a1 s	5'-GTC TAG TGC GCC ACC TAT CG-3'
Ctgf as	5'-GCT CCG TAC ACA GTT CTC CC-3'
Ctgf s	5'-GCG ATC ACA GAC TGA CCT CC-3'
Ephb1 as	5'-AGC AGC AGC AGC TGT CAT AA-3'
Ephb1 s	5'-CAT GTT TCA GAC CCC AGG AT-3'
Hoxa5 as	5'-AAA ACT CCC TGG GCA ACT CC-3'
Hoxa5 s	5'-CGC ATC CAG GGG TAG ATC TG-3'
Irak3 as	5'-CAC GTC TCA GGT TGC TTT TG-3'
Irak3 s	5'-ACA CCC CAA AAA GGG TCT CT-3'
Jag2 as	5'-CTG TTA ACC CGG GCG AGA AT-3'
Jag2 s	5'-GCT TTA AAG GAG GCG AGG GT-3'
Kremen2 as	5'-ACT TGG GCC TCT CAG AAT GC-3'
Kremen2 s	5'-GAC TTG ACC CAG CTC TGA CC-3'
Lrrc41 as	5'-GTC CAG CCC AGT GTC AGT TC-3'
Lrrc41 s	5'-GAG TGA AGC CTG AGA CCG AG-3'
Meis1 as	5'-GCA TTG TGT AAG ACG CGA CC-3'
Meis1 s	5'-CCG TGC GTG TGT AAA GTG TG-3'
Mzfl as	5'-AGA TGC CAC CAC TCC AGA AT-3'
Mzfl s	5'-CCC TTA GGG CAT CAC ACA CT-3'

---

Nkx3-2 as	5'-GTG GAC TCT GGC CTC AAC TC-3'
Nkx3-2 s	5'-CTA CTC CCA TGC CGA GAG AG-3'
Oct4 as	5'-TCC CCT CAC ACA AGA CTT CC-3'
Oct4 s	5'-TCC TCC TAA TCC CGT CTC CT-3'
Prss50 as	5'-GCT AGG TTC CTC CAC AAT GG-3'
Prss50 s	5'-ATT AGC TGA GGC GAG GCA TA-3'
Ptprg as	5'-AAC ATC CAT TGT GTG GCA GA-3'
Ptprg s	5'-CTT TCT CAC AGC CCT CTT GG-3'
Ptpru as	5'-CAT TGA AGG CAG AAG GGA AG-3'
Ptpru s	5'-CTC ACA GCT CCC AGA TGA CA-3'
Pvalb as	5'-GAC TGA GCT GCT GAG CTG TG-3'
Pvalb s	5'-CGT CTG TCA TCG ACA TCC TG-3'
Rab711 as	5'-CTT TCT CGG ACC GCA CAC-3'
Rab711 s	5'-AAG AGG CAC ACT GGG AAA TG-3'
Sall4 as	5'-CCA GGT GGA CTC TCC AGG TA-3'
Sall4 s	5'-GAG TTA GCA CCC AGG CAG AC-3'
Sars2 as	5'-CAC AAC TGG AAT GCA TCT GG-3'
Sars2 s	5'-GCT TCC CTA CTT GGA CAT GG-3'
Slc48a1 as	5'-GCG CTG TCT CCA GTC TCT CT-3'
Slc48a1 s	5'-GTA GCC GGC TTC TCC ATC TT-3'
Snapin as	5'-AAG TCT AGA TCC GCC CCC TA-3'
Snapin s	5'-AGT GGG ATA GCT GGG AAA TG-3'
Tdh as	5'-AGG GGA GGA GGG CAA GTA TT-3'
Tdh s	5'-AGC CTT GGT TTG ACT CCC AC-3'
Thoc1 as	5'-GTC AGG TTG GGA GCA AAG AGA-3'
Thoc1 down as	5'-GCT GGG ATT TGA ACT CCT GAC-3'
Thoc1 down s	5'-AGG CCA AAC CGA AGA GTG TTA-3'
Thoc1 s	5'-GAG CTG ACA ACA CGA AAC ACC-3'
Trim65 as	5'-GTG ACT TGG ATC CCT GCC AA-3'

Trim65 s 5'-AGC AGA TGG AGC AAG TCA CC-3'

#### 2.11.4 Oligonukleotide zur Verwendung in RNAi (Mensch)

CTCFL siRNA 1 5'-AGU AUG UUU AUA GCG CGU GTT-3'

5'-CAG GCG CUA UAA ACA UAC UTT-3'

CTCFL siRNA 2

5'-UUG AGG AGC AUU UCA CAC CTT-3'

5'-GGU GUG AAA UGC UCC UCA ATT-3'

Unspezifische Kontroll-siRNA

5'-CUA CGA UGA AGC ACU AUU ATT-3'

(Weth et al. 2014)

5'-UAA UAG UGC UUC AUC GUA GTT-3'

## 2.12 Software

Excel 2013

Microsoft

GATC Viewer

GATC Biotech

Oligotech

Oligos Etc. Inc.

QuantityOne

Bio-Rad

Rotor-Gene 6

Corbett-Research

VectorNTI 9

InforMax

Word 2013

Microsoft



## 3 Methoden

### 3.1 Arbeiten mit eukaryotischen Zelllinien

Alle Arbeiten mit Zelllinien wurden unter sterilen Bedingungen an gesonderten Arbeitsplätzen durchgeführt, um die Gefahr der Kontamination der Zellen zu minimieren. Alle verwendeten Materialien wie Zellkulturmedien, Puffer, Glaspipetten und Pipettenspitzen wurden vor ihrer Verwendung autoklaviert. Alle Mengenangaben in diesem Abschnitt beziehen sich auf Maxischalen (Ø145mm).

#### 3.1.1 Ablösen von adhärennten Zellen

Adhärennt wachsende Zellen wurden mit Hilfe von Trypsin/EDTA-Lösung abgelöst. Nach Absaugen des Mediums und Waschen mit 10 ml 1x PBS wurden 2 ml bei 37°C ange-wärmtes Trypsin/EDTA zugegeben und für 5 min inkubiert. Dann wurden die abgelösten Zellen mit 10 ml frischem Medium resuspendiert und die gewünschte Menge auf neue Schalen verteilt. Durch Zugabe von frischem Medium wurde ein finales Volumen von 20 ml erreicht. Die Teilungsverhältnisse betragen je nach Wachstum der einzelnen Zelllinien zwischen 1:2 und 1:20.

#### 3.1.2 Auftauen und Einfrieren von eukaryotischen Zelllinien

Die Lagerung von gefrorenen Zellen erfolgte in flüssigem Stickstoff. Die gefrorenen Zellen wurden in ihrem Cryo-Gefäß bei Raumtemperatur aufgetaut und dann mit 8ml des entsprechenden frischen Mediums in ein 15 ml-Zentrifugenröhrchen überführt. Nach zehnminütiger Zentrifugation bei 800 rpm wurde der Überstand verworfen und das Zell-pellet in 10 ml frischem Medium gelöst. Dieser Schritt ist nötig um das zytotoxische DMSO zu entfernen. Die 10 ml wurden dann mit weiteren 10 ml frischem Medium auf eine Maxischale überführt.

<b>Einfriermedium</b>	
Medium (z.B. DMEM)	45%
FCS	40%
DMSO	15%

Zum Einfrieren wurden die Zellen nach dem Ablösen (siehe 3.1.1) in 8 ml frischem Medium resuspendiert, in ein 15ml-Zentrifugenröhrchen überführt und dann für 10 min bei 800 rpm pelletiert. Der Überstand wurde verworfen, das Pellet in 1 ml frischem Medium gelöst und in ein 2 ml Cryo-Gefäß, in dem 1ml des Einfriermediums vorgelegt war, überführt. Das Gefäß wurde dann in einem Einfrierbehälter bei  $-80^{\circ}\text{C}$  abgekühlt und anschließend in flüssigen Stickstoff überführt.

### 3.1.3 Kultivierung von eukaryotischen Zelllinien

Die Inkubation der Zellen erfolgte im Begasungsbrutschrank bei  $37^{\circ}\text{C}$  und 5%  $\text{CO}_2$  in mit Wasserdampf gesättigter Atmosphäre. Allen verwendeten Medien wurde vor Verwendung 10% FCS und 1% PenStrep-Lösung zugesetzt. Suspensionszellen wuchsen in unbeschichteten Petrischalen während die adhärenent wachsenden Zelllinien in beschichteten Gewebekulturschalen wuchsen. Die Zellen wurden jeweils vor Erreichen der Konfluenz aufgeteilt und in neue Schalen überführt. Bei den Suspensionszellen wurde hierzu ein bestimmter Teil der Zellen in eine neue Schale überführt und mit frischem Medium auf das alte Gesamtvolumen aufgefüllt. Die adhärenenten Zellen wurden wie in 3.1.1 beschrieben, abgelöst und gesplittet.

### 3.1.4 Differenzierung von murinen P19 Zellen

Murine P19 Zellen stellen eine leicht kultivierbare pluripotente Alternative zu embryonalen Stammzellen dar. Sie wurden aus einem Embryonalkarzinom gewonnen und werden heute in einer Vielzahl von Projekten analysiert und verwendet. P19 Zellen besitzen das Potenzial in mehrere Entwicklungslinien zu differenzieren. Die Zellen lassen sich durch Zugabe von *All-trans*-Retinsäure in Neuronalzellen differenzieren (Staines et al. 1996) während Zugabe von DMSO zur Differenzierung in Herz-Muskelzellen führt (Edwards et al. 1983).

#### 3.1.4.1 Neuronale Differenzierung von P19 Zellen

Die zu behandelnden Zellen wurden normal auf adhärenenten Maxi-Schalen kultiviert. Zur Einleitung der Differenzierung wurden dann die Menge an RA zugegeben die einer Endkonzentration von  $10\ \mu\text{M}$  im Medium entsprach.

#### 3.1.4.2 Myokardiale Differenzierung von P19 Zellen

Die zu behandelnden P19 Zellen wurden in Suspensions-petrischalen kultiviert. Dies führte zu Aggregation der Zellen in Embryo-ähnlichen Körperchen. Zur kompletten

Einleitung der Differenzierung wurde dann die Menge an DMSO zugegeben die einer 1%igen Konzentration im Medium entsprach.

### 3.1.5 Zellzahlbestimmung

Um die Zellzahl zu messen wurde eine Fuchs-Rosenthal Zählkammer (Tiefe 0,2 mm; Fläche 0,0625 mm<sup>2</sup>; Umrechnungsfaktor  $8 \cdot 10^4$ ) verwendet. Adhärenente Zellen wurden mit Trypsin abgelöst und in Medium resuspendiert. 20 µl dieser Zellsuspension wurden in die Zählkammer gegeben und unter dem Mikroskop wurden 16 Kleinstquadrate ausgezählt. Durch Multiplikation des Mittelwerts, der 16 Kleinstquadrate, mit dem Umrechnungsfaktor ließ sich dann die Zahl der Zellen pro Milliliter bestimmen.

### 3.1.6 Transfektion

Für Transfektionen wurde das Transfektionsreagens jetPEI™ der Firma Polyplus-Transfection verwendet. jetPEI™ dient zur effektiven und reproduzierbaren *in-vitro* Transfektion bei gleichzeitiger niedriger toxischer Wirkung. DNA wird durch jetPEI™ in positiv geladene Partikel gepackt und diese ermöglichen über Interaktion mit der Zelloberfläche und anschließender Endocytose die Aufnahme in die Zellen. Die Durchführung und die Ansätze der Transfektionen entsprachen den Angaben und Empfehlungen im Handbuch des Herstellers. Für Transfektionen in 6-well-Schalen wurden 3 µg DNA in 100 µl 150 nM NaCl-Lösung, und 6 µl jetPEI™ Reagens in ebenfalls 100 µl 150 nM NaCl-Lösung gelöst. Die jetPEI™-Lösung wurde dann zur DNA-Lösung, auf keinen Fall anders herum, gegeben und 15-30 min bei RT inkubiert bevor die jetPEI™/DNA-Mischung tropfenweise zum Serum gegeben wurde.

## 3.2 Arbeiten mit DNA

### 3.2.1 Lagerung von DNA

Die Lagerung von DNA erfolgte bei 4°C in TE-Puffer oder mQH<sub>2</sub>O. Wurde DNA für einen längeren Zeitraum nicht benötigt erfolgte die Lagerung bei -20°C.

TE-Puffer	
Tris/HCl pH 7,6	10 mM
EDTA	1 mM

### 3.2.2 Quantifizierung von DNA

Zur Bestimmung der Konzentration einer DNA in Lösung wurde die Absorption bei 260 nm in einem Spektralphotometer gemessen. Nach dem Lambert-Beer'schen-Gesetz zeigt bei dieser Wellenlänge eine DNA mit der Konzentration 50 µg/µl, bei einer Schichtdicke der Quarzküvette von 1 cm, eine Absorption von 1. Die Konzentration einer DNA-haltigen Lösung lässt sich dann mit folgender Formel berechnen.

$$c_{\text{DNA}} \left[ \frac{\mu\text{g}}{\mu\text{l}} \right] = \frac{A_{260} \cdot 50 \cdot \text{Verdünnung}}{1000}$$

$c_{\text{DNA}}$ : Konzentration der DNA

$A_{260}$ : Absorption bei 260 nm

Weiterhin lässt sich durch Bestimmung des Verhältnisses der Absorptionen bei 260 nm und 280 nm zueinander, eine Aussage über den Reinheitsgrad der Lösung treffen. Idealerweise liegt dieses Verhältnis ( $A_{260}/A_{280}$ ) bei 1,8 oder höher. Niedrigere Verhältnisse deuten auf das Vorhandensein von Proteinen oder Kontamination mit organischen Stoffen in der Lösung hin.

### 3.2.3 Minipräparation von Plasmid-DNA

Die Minipräparation von Plasmid-DNA wurde nach Prinzip der alkalischen Lyse (Birnboim und Doly 1979) durchgeführt. Sie ermöglicht eine schnelle Isolation von kleinen Mengen Plasmid-DNA aus Bakterienkulturen. Somit lassen sich in kurzer Zeit eine Vielzahl an Klonen untersuchen, und das richtige Plasmid identifizieren. Im ersten Schritt wurden einzelne Klone von Agarplatten mittels autoklavierten Pipettenspitzen gepickt und in 3 ml LB-Medium mit dem entsprechenden Antibiotikum überführt. Es erfolgte eine 12 stündige Inkubation bei 37°C unter ständigem Schütteln, zur Vermehrung der Bakterien. Danach wurden 1,5 ml der Bakterien entnommen und bei 6.000 rpm für 1 min pelletiert. Nun wurde der Überstand verworfen und dem Zellpellet wurden 100 µl Sol1 zugegeben und anschließend gründlich gevortext. Dann wurde, zur alkalischen Lyse der Zellen, 200 µl Sol2 zugegeben, erneut gevortext und das Gemisch für 5min auf Eis inkubiert. Zur Neutralisierung der Lösung und zum Fällen der Proteine und chromosomaler DNA wurde anschließend 150 µl Sol3 zugegeben, gevortext und erneut 5 min auf Eis inkubiert. Die ausgefallenen Bestandteile und Zelltrümmer wurden dann bei 13.000 rpm

für 15 min pelletiert. Der Überstand, der die Plasmide enthält, wurde in ein neues Reaktionsgefäß überführt, und die Plasmid DNA durch Zugabe von 0,8 VT Isopropanol und 10 min Inkubation auf Eis gefällt. Es erfolgte eine weitere Zentrifugation für 10 min bei 13.000 rpm, der Überstand wurde anschließend verworfen und das Pellet mit 500 µl 70% EtOH gewaschen. Nun wurde ein letztes mal zentrifugiert, 5 min bei 13.000 rpm, und der Überstand wieder verworfen. Das erhaltene Pellet wurde nun getrocknet und dann in 50 µl TE + RNase A (10 µg/µl) gelöst.

<b>Sol1</b>		<b>Sol2</b>		<b>Sol3 (pH 4,8)</b>	
Glucose	50 mM	NaOH	0,2 N	KaAc	24,5 g
Tris/HCl pH 8,0	25 mM	SDS	1%	Eisessig	35 ml
EDTA	10 mM	immer frisch angesetzt		mQH <sub>2</sub> O	ad. 100 ml

### 3.2.4 Midipräparation von Plasmid-DNA

Die Midipräparation eignet sich zur schnellen Gewinnung mittlerer Plasmidmengen. Es wurde das „High Purity Plasmid Midiprep System“ der Firma Marligen Bioscience, Inc. verwendet und die Durchführung entsprach den Angaben des Herstellers im Handbuch.

### 3.2.5 Phenol-Chloroform-Extraktion

Mit Hilfe der Phenol-Chloroform-Extraktion wurde DNA aus Protein- und RNA-haltigen Lösungen aufgereinigt. Hierzu wurde die Lösung nach Zugabe von 1 VT Phenol und 1 VT Chloroform gevortext und bei RT zentrifugiert (3 min, 13.000 rpm). Durch die Zentrifugation bildete sich eine Schichtung der Phasen aus. Die phenolhaltige untere Phase wurde vorsichtig entfernt und es wurden 2 VT Chloroform zugegeben, gevortext und zentrifugiert (3 min, 13.000 rpm). Nun wurde die DNA-haltige obere Phase in ein neues Reaktionsgefäß überführt und konnte für die Ethanol-fällung weiter verwendet werden.

### 3.2.6 Ethanol-fällung

Zur Reinigung der isolierten DNA wurde eine Fällung mit Ethanol durchgeführt. Der DNA-Lösung wurden 0,1 VT NaAc und 2,5 VT 100% Ethanol zugegeben, und für 20 min auf Eis gefällt. In einigen Fällen wurde Glykogen als Träger für die Fällung zugegeben. Nach 10minütiger Zentrifugation bei 14.000 rpm bei 4°C wurde der Überstand verworfen

und das Pellet mit 300 µl 70% Ethanol für 3 min bei 14.000 rpm gewaschen. Der Überstand wurde erneut abgesaugt und das Pellet dann bei 37°C angetrocknet. Das Pellet wurde dann in TE-Puffer gelöst. Die zugegebene Menge an TE-Puffer hing von der Menge an eingesetzter DNA ab.

### 3.2.7 Sequenzierung von DNA

Die Sequenzbestimmung von DNA-Fragmenten wurde von der Firma StarSEQ durchgeführt. Der Ansatz der Proben entsprach den Vorgaben (U-Mix) der Firma. So wurden 400-700 ng Plasmid-DNA mit 10 pmol des entsprechenden Primers in 7 µl Gesamtvolumen gemischt und an StarSEQ verschickt. Die zurückerhaltenen Elektropherogramme wurden mit dem Programm GATC Viewer der Firma GATC Biotech betrachtet, und falls nötig manuell korrigiert.

### 3.2.8 Gelelektrophorese von DNA-Fragmenten

Bei der Gelelektrophorese handelt es sich um eine molekularbiologische Methode zur Auftrennung von DNA-Fragmenten nach ihrer Größe. Aufgrund der negativen Ladung des Zucker-Phosphat-Rückgrats wandert DNA im elektrischen Feld zur Anode. Mit Hilfe von Agarosegelen unterschiedlicher Konzentration (0,6 – 2 % (w/v)), die mit DNA beladen wurden, konnten DNA-Fragmente von unterschiedlicher Größe aufgrund des größenabhängigen Widerstands der Gelmatrix, und der daraus resultierenden unterschiedlichen Laufgeschwindigkeit, aufgetrennt werden. Es wurden bekannte Größenstandards eingesetzt um die Fragmentgröße und bei Bedarf auch die Konzentration zu bestimmen. Zur Detektion der DNA-Fragmente im Agarosegel wurde bei dessen Herstellung Ethidiumbromid zugegeben. Ethidiumbromid ist ein giftiger, roter Phenanthridin-Farbstoff der sich in die DNA einlagert und nach Anregung mit UV-Licht sichtbares Licht emittiert. Für die Herstellung der Agarosegele wurde die gewünschte Menge Agarose zu 1x TAE-Puffer gegeben, in der Mikrowelle bis zur vollständigen Lösung aufgeköcht, dann mit 10% Ethidiumbromid versetzt und in die vorbereiteten Elektrophoreseschlitten gegossen.

#### 50x TAE-Puffer

Tris	242 g
Eisessig pH 7,7	57,1 ml
EDTA	18,6 g
mQH <sub>2</sub> O	ad. 1 l

### **3.2.9 DNA-Extraktion aus Agarosegel**

Manche Arbeitsschritte erforderten es, bestimmte Banden von in Agarosegelen aufgetrennter DNA weiterzuverwenden. Nach Ausschneiden der gewünschten Bande mit einem Skalpell bei niedrigenergetischer UV-Strahlung (365 nm) erfolgte die DNA-Extraktion mit Hilfe des Komplettsystems „QIAquick® Gel Extraction Kit“ der Firma Qiagen. Die Durchführung entsprach den Angaben des Herstellers („Animal cells spin“) im Handbuch und die extrahierte DNA wurde je nach erwarteter Menge und gewünschter Konzentration mit 30-50 µl 10 nM Tris eluiert

## **3.3 Enzymatische Modifikationen von DNA**

### **3.3.1 Schneiden von DNA durch Restriktionsenzyme**

Molekularbiologische Experimente verlangen teilweise das Entfernen oder Einfügen von spezifischen DNA-Fragmenten aus/in Plasmiden. Mit Hilfe spezifischer Restriktionsendonukleasen kann doppelsträngige DNA geschnitten werden. Zwei unterschiedliche Klassen dieser Enzyme erzeugen unterschiedliche Enden, überhängende („sticky“) oder stumpfe („blunt“). Alle in dieser Arbeit verwendeten Restriktionsendonukleasen wurden nach den Vorgaben des Herstellers eingesetzt.

### **3.3.2 Dephosphorylierung**

Ein mit einem Restriktionsenzym linearisierter Vektor, erschwert durch seine natürliche Affinität zur Religation das Einbringen von DNA-Fragmenten. Durch Dephosphorylierung, d.h. der hydrolytischen Abspaltung der Phosphatgruppe, am 5'-terminalen Ende des Vektors kann die Religation unterbunden und somit das Einbringen von DNA in den Vektor erleichtert werden. Als Enzym für diese Reaktion diente alkalische Kälberdarm-Phosphatase (CIAP). Die Reaktion erfolgte mit einer Enzymeinheit bei 37°C für 1h im entsprechenden Reaktionspuffer. Zur Aufreinigung nach erfolgter Dephosphorylierung wurde eine Gelelektrophorese mit anschließender Gelextraktion durchgeführt.

### **3.3.3 Phosphorylierung**

Um die Effektivität des Einbringens von DNA in einen Vektor weiter zu erhöhen wurde das verwendete Insert-Fragment mit Hilfe des Enzyms T4 Polynukleotid-Kinase phosphoryliert. Der Ansatz entsprach den Angaben des Herstellers.

### 3.3.4 Ligation

Für die Ligation von linearisiertem Vektor und Insert wurde das Enzym T4-DNA-Ligase verwendet. Es katalysiert die Verbindung von 5'-Phosphatgruppen und 3'-Hydroxylgruppen zu Phosphodiesterbindungen unter Anwesenheit von ATP. Vektor und Insert wurden im Verhältnis 1:4 bei stumpfen Enden und 1:2 bei kohäsiven Enden eingesetzt. Die Inkubation erfolgte in T4-DNA-Ligase-Puffer mit einer Einheit T4-Ligase pro  $\mu\text{g}$  GesamtdNA. Im Falle stumpfer Enden wurde bei 4°C über Nacht, im Falle kohäsiver Enden bei RT für 2 h inkubiert.

### 3.3.5 Polymerase-Kettenreaktion (PCR)

Die Polymerase-Kettenreaktion stellt ein molekularbiologisches Verfahren dar mit dem sehr spezifisch gewünschte DNA-Fragmente in einer exponentiell verlaufenden Polymerisationsreaktion amplifiziert werden können. Für die PCR werden drei grundlegende Komponenten benötigt: eine hitzestabile DNA-abhängige DNA-Polymerase, ein Plasmid oder DNA-Fragment das als Matrize für die Polymerase dient und zwei Oligonukleotid-Primer, die komplementär zu den flankierenden Regionen der zu amplifizierenden Sequenz sind. Die Primer hybridisieren antiparallel und ihre freie 3'-OH-Gruppe dient der Polymerase als Ansatzpunkt für die Addition von freien dNTPs.

Die Reaktionen der PCR verlaufen, nach einem 10 minütigen Anfangsdenaturierungsschritt (94°C), zyklisch und die einzelnen Reaktionen sind temperaturabhängig und verlaufen in drei Schritten. Auf die Denaturierung der doppelsträngigen DNA bei 94°C folgt jeweils die Anlagerung der Primer an ihre komplementäre Sequenz bei spezifischer Temperatur und dann die Elongation bei 72°C, dem Temperaturoptimum der verwendeten Polymerase. Nach einer, durch Test bestimmten, Zahl Zyklen (25 – 40) erfolgte ein abschließender Elongationsschritt (5 min, 72°C) und das PCR Produkt wurde bis zur Verwendung in der Gelelektrophorese bei 4°C gelagert.

Es wurden je nach Experiment unterschiedliche Polymerasen verwendet. *Taq*-Polymerase für gewöhnliche Amplifizierungen und *Pfu*-Polymerase für Klonierungen, da sie durch ihre Möglichkeit zur Fehlerkorrektur eine geringere Fehlerrate aufweist. Die Primer wurden nach bestimmten Kriterien ausgewählt. Es wurde darauf geachtet, dass sie eine Länge zwischen 17-25 Nukleotiden aufwiesen, der GC-Gehalt lag zwischen 50-60% und die Schmelztemperatur idealerweise um die 60°C. Die ausgewählten Primer wurden



auch auf Ausbildung von Loop-Strukturen oder von Homo- und Hetero-Dimeren, unter Verwendung der Software Oligotech der Firma Oligos Etc. Inc., untersucht.

### **3.3.6 Quantitative-Real-Time-PCR**

Die Quantitative-Real-Time-PCR erlaubt auf Basis der normalen PCR-Methode eine Quantifizierung des amplifizierten Produktes. Nach jedem Amplifikationsschritt wird mit Hilfe eines Fluoreszenzfarbstoffes, der sich in die kleine Furche der DNA-Helix einlagert und im gebundenen Zustand Licht emittiert, die DNA Menge quantifiziert. Mit der Zunahme an DNA nimmt somit auch die Fluoreszenz des Farbstoffes zu, und ermöglicht die relative Quantifizierung zu einer internen Kontrolle. In dieser Arbeit wurde mit dem Platinum SYBR Green qPCR SuperMix UDG-Kit der Firma Invitrogen anhand der Angaben im Handbuch gearbeitet. Als Fluoreszenzfarbstoff diente SYBR Green I.

## **3.4 Arbeiten mit RNA**

### **3.4.1 Umgang mit RNA**

Bei Arbeit mit RNA ist besondere Vorsicht geboten, da RNA leicht von den fast überall vorhandenen RNasen zersetzt wird. Daher wurden alle Arbeitsschritte mit RNA nur mit Filtertips und besonderer Vorsicht durchgeführt. Vor dem Umgang mit RNA wurde weiterhin der komplette Arbeitsbereich und teilweise auch die Utensilien mit 70% EtOH gereinigt. Die Lagerung der RNA erfolgte bei -20°C in RNase-freiem Wasser.

### **3.4.2 Isolierung von RNA**

RNA wurde durch Nutzung des RNeasy® Mini Kits der Firma Qiagen gewonnen. Die Durchführung entsprach den Herstellerangaben im Handbuch (Programm: „animal cells spin“). Vor Nutzung des Kits mussten die Zellen, aus denen die RNA isoliert werden sollte, bei 2000 rpm für 5 min pelletiert werden. Handelte es sich um adhären wachsende Zellen wurden diese mit Trypsin von ihrer Wachstumsoberfläche gelöst und dann pelletiert.

### **3.4.3 Quantifizierung von RNA**

Die Quantifizierung von RNA basiert wie die Quantifizierung von DNA auf dem Lambert-Beer'schen Gesetz. (siehe 3.2.2). Es wurde jedoch ein Absorptionsverhältnis

A260/A280 von 2,0 angestrebt. Niedrigere Werte waren ein Anzeichen für Kontamination mit DNA oder Proteinen und höhere Werte deuteten auf Degradierung der RNA hin. Zur Qualitätsprüfung der RNA wurde weiterhin ein Testgel mit 1 µl RNA durchgeführt.

#### **3.4.4 DNase Verdau von RNA**

Wenn mit RNA aus transfizierten Zellen gearbeitet wurde oder wenn die Quantifizierung der RNA niedrige Werte beim A260/A280-Verhältnis ( $< 1,9$ ) zeigte, wurde ein optionaler DNase Verdau durchgeführt. Hierzu wurde das Komplettsystem „RQ1 RNase-Free DNase“ der Firma Promega verwendet. Die Vorgehensweise entsprach dem Handbuch und je Ansatz wurde die 1 µg RNA entsprechende Menge RNA-Lösung eingesetzt. Für einige Vorversuche wurde eine weitere DNase verwendet, ... Auch hier entsprach das Vorgehen den Angaben des Herstellers.

#### **3.4.5 RT-PCR**

Die RT-PCR stellt eine gute Möglichkeit dar um den RNA-Gehalt in Zellen zu untersuchen. Vor der PCR muss die mRNA mit Hilfe des Enzyms Reverse Transkriptase in stabile, doppelsträngige cDNA umgeschrieben werden. Zur Herstellung der cDNA aus mRNA wurden die Komplettsysteme „Transcriptor First Strand cDNA Synthesis Kit“ der Firma Roche verwendet. Die Vorgehensweise entsprach dem Handbuch und je Ansatz wurde die 1 µg RNA entsprechende Menge RNA-Lösung eingesetzt. In die anschließende PCR wurde jeweils 1 µl der cDNA-Lösung als Template eingesetzt. Alle Primerpaare für die Analyse von RNA-Gehalt wurden so gewählt, dass das Amplikon mindestens ein Intron überspannt um Kontaminationen mit genomischer DNA ausschließen zu können.

### **3.5 Arbeiten mit Bakterien**

#### **3.5.1 Vermehrung und Lagerung**

Die Vermehrung der Bakterien erfolgte bei 37°C in LB-Medium als Schüttelkultur oder auf LB-Agarplatten. Maxi-Kulturen wurden in TB-Medium kultiviert. Zur Selektion wurde dem Medium jeweils das Antibiotikum zugegeben gegen das die transformierten Plasmide eine Resistenz vermittelten. Nach Anzucht der Bakterien auf Agarplatten konnten diese für maximal 4 Wochen bei 4°C im Kühlschrank gelagert werden. Für eine längere Lagerung wurden sog. Glycerolstocks angelegt. Zu 200 µl Bakterien in LB-Medium

wurden 200 µl frisches LB-Medium und 400 µl 87% Glycerin gegeben und in flüssigem Stickstoff schockgefroren. Die anschließende Lagerung erfolgte bei -80°C.

<b>LB-Agar</b>	<b>LB-Medium</b>	<b>TB-Medium</b>
Bacto-Trypton 10 g/l	Bacto-Trypton 10 g/l	Casein (verdaut) 12 g/l
Hefe-Extrakt 5 g/l	Hefe-Extrakt 5 g/l	Hefe-Extrakt 24 g/l
NaCl 10 g/l	NaCl 10 g/l	K <sub>2</sub> HPO <sub>4</sub> 9,4 g/l
Agar 15 g/l	pH 7,5; autoklaviert	KH <sub>2</sub> PO <sub>4</sub> 2,2 g/l
autoklaviert		autoklaviert

### 3.5.2 Transformation

Unter Transformation versteht man das Einbringen von Plasmid-DNA in entsprechend vorbereitete, sog. kompetente, Bakterien. Durch Behandlung mit Kalziumchlorid (= chemische Kompetenz) wurde die Zellmembran der Zellen so verändert, dass die Aufnahme von negativ geladener DNA in die Bakterienzellen erleichtert war. Es wurden 100-500 ng Plasmid-DNA zu 200 µl kompetenten *E. coli* Bakterien gegeben und für 30min auf Eis inkubiert. Danach erfolgte ein Hitzeschock für 45 sec bei 42°C, der die DNA-Aufnahme in die Bakterien vermittelte. Nun wurden der Bakteriensuspension 100 µl LB-Medium zugegeben, kurz geschüttelt und anschließend auf eine Agarplatte mit zur Selektion geeignetem Antibiotikum ausplattiert. Nach 10-16 h Inkubation bei 37°C konnten Einzelkolonien isoliert und damit weitergearbeitet werden.

## 3.6 Arbeiten mit Proteinen

Alle Arbeitsschritte mit Proteinen wurden bei 4°C oder auf Eis durchgeführt und die Lagerung erfolgte bei -20°C.

### 3.6.1 Proteinextraktion mit RIPA-Puffer

Zur Gesamtprotein-Extraktion wurde die Methode der Proteinextraktion mit RIPA-Puffer genutzt. RIPA stellt einen der verlässlichsten Puffer zur Lyse von Säugerzellen dar und er ermöglicht die schnelle Extraktion von cytoplasmatischen, membrangebundenen und Zellkern lokalisierten Proteinen.

<b>RIPA-Puffer (2x)</b>		<b>Immer frisch angesetzt (4°C)</b>	
NaCl	280 mM	RIPA (2x)	1 VT
Tris/HCl pH8	20 mM	Complete Mix Mini (7x)	0,1 VT
EDTA pH8	2 mM	mQH <sub>2</sub> O	0,9 VT
10% Triton-X-100	2%		
10% SDS	0,2%		
10% DOC	0,2%		

Zur Extraktion von Proteinen aus adhären wachsenden Zellen wurden diese nach Waschen mit 1x PBS mit einem Zellschaber in 10 ml frischem Medium abgeschabt, in der Zentrifuge bei 900 rpm pelletiert und anschließend auf Eis gestellt. Nun erfolgte die Zugabe von frisch angesetztem und vorgekühltem Lyse-Mix (s.o). Je nach Konfluenz und Schalengröße wurde die Menge variiert, so wurden bei einer dichten Maxi-Schale 100 µl zugegeben. Nach 10 minütiger Inkubation auf Eis wurden die Ansätze bei 4°C für 10 min bei 14.000 rpm pelletiert. Der proteinhaltige Überstand wurde abgenommen und in ein neues Reaktionsgefäß überführt. 5 µl wurden für die Proteinbestimmung nach Bradford abgenommen und der Rest 1:1 mit dem Auftragspuffer Roti®-Load versetzt.

### 3.6.2 Proteinbestimmung nach Bradford

Basierend auf photometrischer Messung stellt der Bradford-Test eine sehr genaue Methode zur quantitativen Bestimmung von Proteinen dar. Der Test beruht auf der Bindung von Proteinen an den Farbstoff Coomassie Brilliant Blue G-250 in saurer Lösung. In freier Form liegt Coomassie Brilliant Blue G-250 in roter Form vor, die Farbe ändert sich nach Komplexbildung mit Proteinen zu blau und das Absorptionsmaximum verschiebt sich von 470 nm hin zu 595 nm (Bradford 1976). Mit Hilfe von Eichkurven oder Kalibrationslösungen können durch die Absorptionsmessung bei 595 nm Aussagen über die Proteinkonzentrationen getroffen werden. Da das Ausmaß der Farbreaktion von Protein zu Protein variiert, sind die Ergebnisse der Konzentrationsbestimmung von Proteingemischen nur semi-quantitativ.

In der vorliegenden Arbeit wurde folgendes Protokoll verwendet: 5 µl des Proteingemisches und 5 µl des zur Extraktion genutzten RIPA-Puffers (als Nullwert) wurden 1:10 mit

H<sub>2</sub>O verdünnt. 5 µl dieser Verdünnungen wurden zusammen mit 200 µl Roti®-Quant und 800 µl Tris (250 mM, pH 6,8) für 10 min bei RT inkubiert und dann die Absorption bei 595 nm im Photometer bestimmt. Mit folgender Formel konnte dann die Konzentration der Probe berechnet werden.

$$c = \left( \frac{E \times 19,89}{x \mu l} \right) \times V$$

E = Absorption bei 595 nm

x = eingesetzte Probenmenge

V = Verdünnungsfaktor

### 3.6.3 SDS-Polyacrylamidgelelektrophorese

Die SDS-PAGE stellt ein Verfahren zur schnellen Auftrennung von komplexen Proteingemischen entsprechend ihrer Größe, unter Verwendung der denaturierenden Chemikalie Natriumdodecylsulfat (SDS), dar. SDS zerstört die räumliche Struktur und überdeckt durch Bindung an die Aminosäuren die Eigenladung der Proteine. Durch Bindung von je zwei SDS-Molekülen pro AS entsteht ein gleichmäßiges Ladung-Massen-Verhältnis. Ebenfalls denaturierend wirkt das im Ladepuffer enthaltene β-Mercaptoethanol, welches die Disulfidbrücken, die eine Rolle bei der Stabilisierung der Sekundär- und Tertiärstruktur der Proteine spielen, spaltet. Nach Zugabe des Ladepuffers wurde das Proteingemisch im Wasserbad bei 80-100°C für 10 min erhitzt um die Denaturierung zu beschleunigen.

Die Auftrennung des Proteingemisches erfolgte über eine Gelmatrix aus polymerisiertem Acrylamid/Bisacrylamid. Nach Auftragen der Proteinproben und des Einbringens in ein elektrisches Feld wanderten die negativ aufgeladenen Proteine in größenabhängiger Weise durch das Gel zur Anode hin. Das Gel bestand aus zwei Abschnitten: einem Sammelgel mit niedriger Acrylamid Konzentration in dem sich alle Proteine in einer dünnen Lauffront sammelten und dem Trenngel mit höherer Konzentration das die Trennung der Proteine ermöglichte. Zur Quantifizierung der Molekulargewichte der Proteine wurden immer Molekulargewichtsmarker mit bekannten Proteingrößen parallel verwendet.

<b>Sammelgel (6%)</b>		<b>Trenngel (7,5%)</b>	
Acrylamid/Bis	1,3ml	Acrylamid/Bis	2,5ml
H <sub>2</sub> O	6,1ml	H <sub>2</sub> O	4,9ml
Tris/HCl pH6,8 (0,5 M)	2,5ml	Tris/HCl pH 8,8 (1,5 M)	2,5ml
10% SDS	100µl	10% SDS	100µl
40% APS	50µl	40% APS	50µl
TEMED	10µl	TEMED	5µl

Bei der Gelherstellung wurde zuerst das Trenngel und danach das Sammelgel in einem Gießstand polymerisiert. Der Beginn der Polymerisation wurde jeweils von APS und TEMED vermittelt. Nach erfolgter Polymerisation wurde das Gel in eine entsprechende Laufapparatur eingespannt und diese mit 1x Lämmli-Puffer befüllt. Bei einer Spannung von 100V erfolgte dann die Trennung der Proteine.

<b>Lämmli-Puffer (10x)</b>	
Tris	250 mM
Glycin	1920 mM
SDS	1%

### 3.7 Western Blot

Der Western Blot stellt eine Methode zur spezifischen Detektion von Proteinen mit Hilfe von entsprechenden Antikörpern dar. Nach einer SDS-PAGE (siehe 3.6.3) werden die Proteine geblottet, d.h. auf eine Transfermembran übertragen. Auf dieser Membran kann dann im zweiten Schritt die Detektion durch Antigen-Antikörper-Reaktion erfolgen.

#### 3.7.1 Blotten der Proteine

Das Übertragen der Proteine auf eine PVDF-Membran erfolgte mit Hilfe des „Wet Blot“-Verfahrens. Die Membran wurde hierzu in 100% Methanol aktiviert und dann mit aufgelegtem Gel zwischen zwei Schwammplatten und je zwei Whatman-Papieren in eine Wet-

Blot-Apparatur eingebaut. Die Apparatur wurde mit vorgekühltem Blot-Puffer gefüllt und ein Kühlpaket zur Abschwächung der Wärmeentwicklung wurde zugegeben. Die PVDF-Membran wurde so ausgerichtet, daß sie in Richtung der Anode orientiert war. Somit wurden bei einstündigem Anlegen eines elektrischen Feldes (100 V) die negativ geladenen Proteine von dem Gel auf die Membran übertragen.

<b>Blot-Puffer (1x)</b>	
Tris	25 mM
Glycin	192 mM

### 3.7.2 Immundetektion

Der spezifische und hochsensitive Nachweis von Proteinen auf der Membran erfolgte durch Detektion mit entsprechenden Antikörpern. Die Methode der Detektion basiert auf zwei nacheinander durchgeführten Antigen-Antikörper-Reaktionen. Im ersten Schritt wird das Protein von einem Primärantikörper gebunden, im zweiten werden Epitope dieses Antikörpers von einem enzymgekoppelten Sekundärantikörper erkannt und gebunden. Im Falle dieser Arbeit war das gekoppelte Enzym die Meerrettichperoxidase und über sie erfolgte der Nachweis des Proteins über Chemilumineszenz. Die hohe Empfindlichkeit dieser Methode wird durch die Bindung mehrerer Sekundärantikörper-Moleküle an ein Primärantikörpermolekül erreicht.

Nach der Durchführung des Wet-Blots wurde die PVDF-Membran über Nacht bei 4°C in einer Lösung aus 5% Milchpulver in PBS-T geblockt, d.h. unspezifische Bindestellen der Membran wurden abgesättigt. Die Membran wurde dann für 1 h mit dem Primärantikörper in entsprechender Verdünnung in PBS-T (+ evtl. Milchpulver) bei RT inkubiert. Nach dreimaligem 10minütigem Waschen mit PBS-T, zur Entfernung von überschüssigem Antikörper, erfolgte die Inkubation mit dem Sekundärantikörper, ebenfalls verdünnt in PBS-T (+ evtl. Milchpulver). Anschließend wurde erneut dreimal mit PBS-T gewaschen und dann detektiert. Die Detektion erfolgte durch Zugabe von ECL-Lösung. Die Lösung wurde aus zwei Bestandteilen im Verhältnis 1:1 frisch angesetzt, und auf den Blot pipettiert. Nach fünfminütiger Inkubation bei RT konnte die entstehende Lumineszenz durch Auflage von Röntgenfilmen und anschließender Entwicklung detektiert werden. Zum

Ende der Arbeit erfolgte die Detektion der Lumineszenz mit dem neuerworbenen Imaging-System VersaDoc der Firma BioRad. Die Lumineszenz entsteht durch katalytische Umsetzung des in der ECL-Lösung enthaltenen Luminols durch das Enzym Meerrettichperoxidase.

<b>PBS-T (pH 7,4)</b>	
Na <sub>2</sub> HPO <sub>4</sub>	10 mM
KH <sub>2</sub> PO <sub>4</sub>	1,7 mM
NaCl	137 mM
KCl	2,7 mM
Tween® 20	0,1%

### 3.7.3 Strippen von Western-Blot Membranen

Teilweise war es nötig eine Western-Blot Membran nacheinander mit mehreren Antikörpern zu behandeln, um zum Beispiel eine Protein-Ladekontrolle durchzuführen. Zum Strippen, d.h. Ablösen der vorherigen Primär- und Sekundärantikörper, wurde die Membran für eine halbe Stunde bei RT in Strippuffer inkubiert, dann dreimal für 10 min mit PBS-T gewaschen und mit 5% Milch/PBS-T für 1 h bei RT oder über Nacht bei 4°C geblockt. Dann konnte eine neue Immundetektion (siehe 3.7.2) durchgeführt werden. Eine Membran konnte bis zu dreimal gestriipt und somit bis zu vier unterschiedliche Immundetektionen durchgeführt werden.

<b>Strippuffer</b>	
Glycin	25 mM
SDS	1%
mit HCl auf pH2,0 eingestellt	



## 3.8 Chromatin-Immunopräzipitation

Die Chromatin-Immunopräzipitation dient zum Nachweis von Protein-DNA-Interaktionen *in-vivo*. Sie besteht aus zwei aufeinanderfolgenden Arbeitsschritten, dem DNA/Protein-Vernetzung mit anschließender Chromatingewinnung und der Präzipitation der gewünschten DNA/Protein-Komplexe mit Hilfe von Antikörpern. Alle verwendeten Flüssigkeiten wurden vor Verwendung sterilfiltriert und es wurde mit Filtertips gearbeitet.

### 3.8.1 Isolierung des Chromatins

Die Isolierung des Chromatins erfolgte aus Maxischalen. Zum Vernetzen der DNA mit den Proteinen wurde 37% Formaldehyd in einer Endkonzentration von 1% tropfenweise direkt ins Kulturmedium gegeben und die Zellen für 10min bei RT inkubiert. Zur Beendigung der Kreuzvernetzung wurde dann 1/7 VT 1M Glycin zugegeben und für 5 min bei RT inkubiert. Nun wurden die Zellen, nach zweimaligem Waschen mit je 10ml eiskaltem 1x PBS, unter Verwendung eines Zellschabers in 1 ml 1xPBS + 1 mM PMSF geerntet. Die Zellen wurden dann nach 5min Zentrifugation (2000rpm, 4°C), und Verwerfen des Überstandes, in 1ml SDS-Lysepuffer resuspendiert und für 10min auf Eis inkubiert. Zur vollständigen Lyse der Zellen und der mechanischen Scherung des Chromatins wurde die Zelllösung anschließend in einem 15ml Zentrifugenröhrchen im Bioruptor sonifiziert (15x 30 sek On/Off). Es entstanden dadurch Chromatinfragmente mit einer durchschnittlichen Länge von 500 bp, was über eine Agarosegel kontrolliert wurde.. Zur Abtrennung der Zellbruchstücke erfolgten nun zwei weitere Zentrifugationsschritte bei 4°C, für 2min bei 5.000rpm und nach Überführen des Überstandes in ein 2ml Reaktionsgefäß für 10min bei 14.000rpm. Das im Überstand enthaltene Chromatin konnte dann direkt für die Immun-Präzipitation eingesetzt werden oder nach Zugabe von 5% Glycerin bei -80°C gelagert werden.

<b>SDS-Lysepuffer</b>	
SDS	1%
EDTA	10 mM
Tris/HCl pH 8,1	50 mM
immer frisch:	
PMSF	1 mM
Complete Mix Mini (7x)	100 µl/ml

### 3.8.2 Immunpräzipitation

Alle Arbeitsschritte der Immunpräzipitation fanden, wenn nicht anders vermerkt, bei 4°C statt. Von jedem Chromatinüberstand wurden 20 µl zur Weiterverwendung als Inputmaterial abgenommen und bei -20°C gelagert. Je Immunpräzipitation wurden 200 µl Überstand mit 1,8 ml Dilutionspuffer versetzt. Vor der eigentlichen Präzipitation wurde der Zellextrakt vorgereinigt, um unspezifisch an die Beads bindende Proteine zu entfernen. Hierfür wurden je 30 µl Protein-A/G-Agarose Beads (50% slurry) in 1 ml Dilutionspuffer äquilibriert und für 2 min bei 2.000 rpm zentrifugiert. Nach vorsichtigem Abnehmen des Überstands wurden der verdünnte Zellextrakt (2 ml) und 5 µl IgG zugegeben und alles für 2 h rotierend inkubiert. Nach Zentrifugation (5 min, 2.000 rpm) wurde der Überstand in ein neues 2 ml Reaktionsgefäß überführt und nach Zugabe des entsprechenden Antikörpers erfolgte die Bildung von Antikörper/Protein-Komplexen rotierend über Nacht. Als Negativkontrolle wurde eine Probe mit 5 µl IgG inkubiert. Die entstandenen Komplexe wurden dann nach Zugabe von 30 µl Protein-A/G-Beads für 2 h rotierend inkubiert und somit gebunden. Anschließend erfolgte eine Zentrifugation (2 min, 2.000 rpm) und der Überstand wurde vorsichtig abgenommen und verworfen. Nun erfolgten mehrere Waschschrte mit jeweils 1 ml Niedrigsalzpuffer, Hochsalzpuffer, LiCl-Puffer und Zwei mit TE-Puffer. Ein Waschschrte bestand aus der Zugabe des entsprechenden Puffers, einer 5minütigen rotierenden Inkubation mit darauffolgender Zentrifugation (2 min, 2.000 rpm) und Verwerfen des Überstandes. Die Ablösung der Protein/DNA-Komplexe erfolgte dann durch zweimalige Zugabe von je 250 µl Elutionspuffer und 15 minütiger rotierender Inkubation bei RT. Die Beads wurden jeweils abzentrifugiert (2 min, 2.000 rpm) und die abgenommenen Überstände in einem neuen 1,5 ml Reaktionsgefäß vereinigt. Der Input (20 µl) wurde durch Zugabe von Elutionspuffer ebenfalls auf 500 µl aufgefüllt und gleich bearbeitet. Durch Zugabe von 20 µl 5 M NaCl und Inkubation bei 65°C im Wasserbad über Nacht erfolgte nun die Lösung der Formaldehyd-vermittelten Kreuzvernetzungen zwischen DNA und Proteinen. Dann erfolgte die Zugabe von 10 µl 0,5 M EDTA und 100 µl 0,5 M Tris/HCl (pH 6,8) zur Schaffung der nötigen Pufferbedingungen für den folgenden RNA- und Protein-Verdau. Zur Degradierung der RNA wurde anschließend 5 µg RNase A zugegeben und 2 h bei 37°C inkubiert. Nach Zugabe von 20 µg Proteinase K wurde noch einmal für 2 h bei 45°C inkubiert. Nun konnte die verbleibende DNA mittels Phenol/Chloroform-Extraktion (siehe 3.2.5) und anschliessender Ethanolfällung (siehe 3.2.6) extrahiert werden. Die Fällung erfolgte unter Verwendung von 20 µl Glykogen als Carrier und der eigentliche Fällungsschrte erfolgte für

30 min bei  $-20^{\circ}\text{C}$ . Das getrocknete DNA-Pellet wurde in  $43\ \mu\text{l}$  mQH<sub>2</sub>O und  $7\ \mu\text{l}$  TE-Puffer gelöst und für die anschließenden PCRs wurde je  $1\ \mu\text{l}$  als Template verwendet.

<b>Dilutionspuffer</b>		<b>Niedrigsalzpuffer</b>	
SDS	0,01%	SDS	0,1%
Triton X100	1,1%	Triton X100	1%
EDTA	1,2 mM	EDTA	2 mM
Tris/HCl pH 8,1	16,7 mM	Tris/HCl pH 8,1	20 mM
NaCl	167 mM	NaCl	150 mM

<b>Hochsalzpuffer</b>		<b>LiCl-Puffer</b>	
SDS	0,1%	LiCl	0,25 M
Triton X100	1%	NP40	1%
EDTA	2 mM	DOC	1%
Tris/HCl pH 8,1	20 mM	EDTA	1 mM
NaCl	500 mM	Tris/HCl pH 8,1	10 mM

### 3.9 FAIRE-Assay

Der FAIRE-Assay (abgekürzt von *Formaldehyd-assisted isolation of regulatory elements*) stellt eine Methode zur Untersuchung der generellen „Zugänglichkeit“ von Chromatin-bereichen dar. Die Methode basiert nicht wie die ChIP auf bestimmten Antikörpern sondern analysiert unabhängig von bestimmten Faktoren wie „offen“ sich das Chromatin darstellt. Beim FAIRE-Assay werden wie bei der ChIP, durch Formaldehyd Quervernetzungen zwischen DNA und ihrer assoziierten Proteinen erzeugt. Allerdings arbeitet man bei FAIRE mit kürzeren Zeiten der Formaldehyd Behandlung, um nicht in allen DNA Bereichen starke Querverbindungen zu erzeugen. Diese festen DNA-Protein Verbindungen werden im weiteren Verlauf der Methode durch eine Phenol-Chloroform-Extraktion von den wenig oder komplett ungebundenen DNA Fragmenten abgetrennt. Nach erfolgter DNA Aufreinigung lässt sich dann per qPCR oder NGS die relative „Offenheit“ von Chromatinbereichen untereinander darstellen.

Wie bei der ChIP wurden auch bei dem FAIRE-Assay alle verwendeten Lösungen sterilfiltriert und ausschließlich mit Filter-Tips gearbeitet.

### 3.9.1 Isolierung des Chromatins

Die Isolierung des Chromatins erfolgte aus Maxischalen. Adhärenente Zellen wurden gewaschen, abtrypsiniert, in 20 ml frischem Medium resuspendiert und in ein 50 ml Falcon überführt. Zur Vernetzung der Proteine mit der DNA wurden 555 µl 37%iges Formaldehyd für eine Endkonzentration von 1% zugegeben. Nun erfolgte eine 5-7 minütige Inkubation auf dem Rotator bei RT. Zum Abstoppen der Vernetzungs-Reaktion wurden dann 2 ml 1,25 M Glycin zugegeben und weitere 5 Minuten drehend inkubiert. Die Zellen wurden dann pelletiert (5 min, 2.000 rpm) der Überstand verworfen und zweimal mit 10 ml eiskaltem PBS gewaschen. An diesem Punkt konnten die Zellen für späteren Gebrauch mit flüssigem Stickstoff schockgefroren werden und bei -80° C gelagert werden. Bei direkter Verwendung wurden je  $1 \cdot 10^7$  Zellen in 1,5 ml Lysis Puffer resuspendiert und in ein sonifizierbares Hartplastik-Falcon überführt. Um eine DNA-Fragment Größe von 200-500 bp zu erhalten wurden die Zellen im Bioruptor für 15 Zyklen (high energy, 30 Sekunden On/Off) sonifiziert.

<b>FAIRE Lysis Puffer</b>	
SDS	1%
Triton X100	2%
EDTA	1 mM
Tris/HCl pH 8,0	10 mM
NaCl	100 mM

### 3.9.2 Gewinnung des Input-Materials (=genomische DNA) als Kontrolle

Um später eine Referenz-Probe für die FAIRE DNA zu haben wurden 100 µl des Chromatins, was einem 20%-igen Input entspricht, aufgearbeitet. Hierzu wurden die 100 µl mit 1 µl RNase A versetzt und 30 Minuten bei 37°C inkubiert. Danach erfolgte eine einstündige Behandlung mit 1 µl Proteinase K bei 55°C, gefolgt von Inkubation bei 65°C über Nacht. Am nächsten Tag erfolgte dann eine Phenol-Chloroform-Extraktion. Hierfür wurde die Probe mit 200 µl 10 mM TRIS versetzt, dann jeweils 0,5 VT Phenol und 0,5

VT Chloroform/Isoamylalkohol zugegeben und 10 Sek. gevortext. Nach 5 Minuten Zentrifugation bei 13.000 rpm wurde der Überstand abgenommen und in ein neues Reaktionsgefäß überführt. Zum verbliebenen Rest wurden 150 µl 10 mM TRIS zugegeben und 10 Sek. gevortext. Nach erneuter 5 minütiger Zentrifugation wurde wieder der Überstand abgenommen und mit dem vorherigen vereinigt. Nun erfolgte eine identische zweite Runde der Extraktion um möglichst wenig/keine Proteinrückstände im Überstand zu behalten. Nach dieser zweiten Runde wurden die Überstände in einem 2 ml Reaktionsgefäß gesammelt und durch Zugabe von 1/10 VT Natriumacetat, 2 VT 95% Ethanol und 1 µl Glykogen (20 mg/ml) bei -80°C für mehr als zwei Stunden, meist über Nacht, gefällt. Die gefällte DNA wurde dann bei 13.000 rpm und 4°C für 15 Minuten zentrifugiert. Das DNA Pellet wurde anschließend mit 500 µl eiskaltem 70%igen Ethanol gewaschen und erneut für 5 Minuten zentrifugiert. Nun wurde der Überstand vorsichtig abgenommen und das Pellet kurz in der SpeedVac getrocknet und abschließend in 40 µl 10 mM TRIS/HCl pH 7,4 gelöst. Die gelöste Input DNA konnte nun als Referenz zur FAIRE DNA z. B. in qPCR eingesetzt werden.

### **3.9.3 Gewinnung der spezifischen ungebundenen FAIRE-DNA**

Zur Aufreinigung der spezifischen FAIRE-Probe wurden 500 µl des Chromatins mit je 0,5 VT Phenol und 0,5 VT Chloroform/Isoamylalkohol versetzt und 10 Sek. gevortext. Nach 5 Minuten Zentrifugation bei 13.000 rpm wurde der Überstand in ein neues Reaktionsgefäß überführt und der Rest erneut mit 150 µl 10 mM TRIS/HCl versetzt. Der Rest wurde dann erneut gevortext, zentrifugiert und die Überstände vereinigt. Dann erfolgte eine komplette weitere Runde der Phenol-Chloroform-Extraktion. Nachdem die beiden Überstände vereinigt wurden wurde noch einmal 200 µl Chloroform/Isoamylalkohol dazu gegeben, gevortext, zentrifugiert und der wässrige Überstand in ein 2 ml Reaktionsgefäß überführt. Dieser Schritt sollte sicherstellen, dass keine Phenolrückstände mehr in der Probe enthalten sind. Dann erfolgte die Fällung der DNA durch Zugabe von 1/10 VT Natriumacetat, 2 VT 95% Ethanol und 1 µl Glykogen (20 mg/ml) bei -80°C für mehr als zwei Stunden. Die gefällte DNA wurde dann bei 13.000 rpm und 4°C für 15 Minuten zentrifugiert. Das DNA Pellet wurde anschließend mit 500 µl eiskaltem 70%igen Ethanol gewaschen und erneut für 5 Minuten zentrifugiert. Nun wurde der Überstand vorsichtig abgenommen und das Pellet kurz in der SpeedVac getrocknet und abschließend in 50 µl 10 mM TRIS/HCl pH 7,4 gelöst. Dann wurde 1 µl DNase-freie RNase A zugegeben und

bei 37°C für 30 Minuten inkubiert. Anschließend erfolgte ein weiterer Verdau durch Zugabe von 1 µl Proteinase K und Inkubation erst für eine Stunde bei 55°C und dann über Nacht bei 65°C. Am nächsten Tag erfolgte eine DNA-Aufreinigung unter Verwendung des *illustra™ GFX™ PCR DNA and Gel Band Purification Kit* nach Anleitung des Herstellers an dessen Ende die DNA, wie die Input DNA auch, in 40 µl 10 mM TRIS/HCl pH 7,4 gelöst wurde.

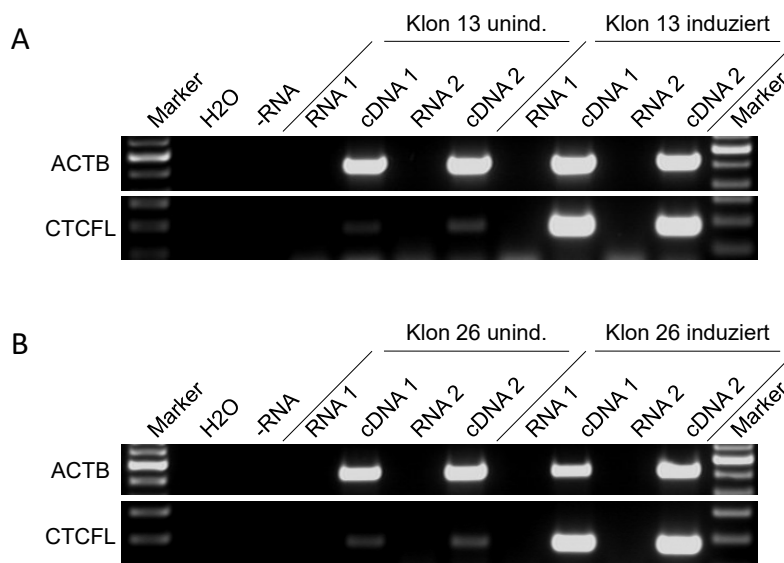
## 4 Ergebnisse

### 4.1 Etablierung und Charakterisierung einer stabilen CTCFL-Überexpression in humanen HeLa-Zellen

Zur Analyse des Bindevhaltens von CTCFL und der eventuellen Effekte auf die Expression von weiteren Genen sollte ein stabiles Überexpressions-System dienen. Als Zelllinie, zur Etablierung dieses Systems, bot sich die humane HeLa-Linie an, die sich durch gute Transfizierbarkeit und einfache Kultur-Ansprüche auszeichnet. HeLa stellt weiterhin eine der am weitest verbreiteten und dadurch auch am besten charakterisierten Zelllinien dar, und somit ist für weitergehende Analysen eine Vielzahl an öffentlichen genomweiten Daten verfügbar. Für die Überexpression wurde ein induzierbares System gewählt, um den Zeitpunkt und gegebenenfalls die Menge der CTCFL-Expression regulieren zu können. Hierzu wurden die Zellen mit dem Zwei-Vektor-System pBI-EGFP-FLAG-hCTCFL und pTA-N transfiziert. Dabei trägt der eine Vektor neben eGFP (enhanced green fluorescent protein), das zur schnellen positiven Selektion der Klone mittels Fluoreszenz-Mikroskopie dient, auch eine Expressionskassette für das mit einem FLAG-Tag versehene humane CTCFL, die beide unter dem Einfluss eines bidirektionalen Promotors stehen. Der Vektor pTA-N fungiert als Regulator-Plasmid, auf diesem ist neben einer Puromycin-Resistenz ein Fusionsprotein aus dem Tetrazyklin-Repressor und der stark aktivierenden VP16-Domäne codiert. Dieses Transaktivatorprotein kann nur bei Abwesenheit von Tetrazyklin beziehungsweise Doxyzyklin auf dem anderen Vektor binden und den bi-direktionalen Promotor aktivieren. Dieses System wird als Tet-Off bezeichnet.

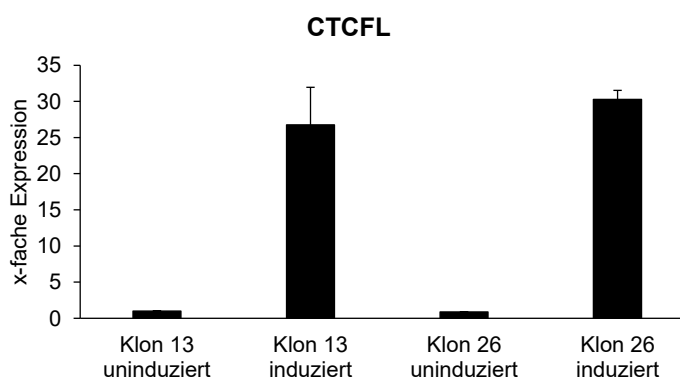
#### 4.1.1 Zwei HeLa-Klone zeigen stabile induzierbare Expression von CTCFL auf RNA-Ebene

Nach erfolgter Transfektion mit den beiden Vektoren pBI-EGFP-FLAG-hCTCFL und pTA-N unter Verwendung von jetPEI wurden die Zellen unter Puromycin Selektionsdruck gesetzt. Erfolgreich transfizierte Zellen sollten somit durch die pTA-N vermittelte Puromycin Resistenz vereinzelt werden. Nach zwei Wochen Wachstum konnten einzelne Kolonien auf den Maxi-Schalen identifiziert und in separate neue Schalen übertragen werden. Nach entsprechender weiterer Kultivierung dieser Klone wurde jeweils mit oder ohne Induktion RNA isoliert. Diese RNA wurde in cDNA umgeschrieben und dann per RT-PCR und RT-qPCR auf CTCFL-Induktion überprüft.



**Abb. 10: hCTCFL ist als induzierbare RNA in zwei Klonen nachweisbar.** HeLa Zellen wurden mit pBI-EGFP-FLAG-hCTCFL und pTA-N transfiziert und mit Puromycin selektiert. Zell-Klone wurden für 48 h uninduziert (mit Doxyzyklin) oder induziert (ohne Doxyzyklin) kultiviert, anschließend wurde jeweils die RNA isoliert und in cDNA umgeschrieben. RT-PCR mit Oligonukleotidprimern für  $\beta$ -Aktin (ACTB) und CTCFL gegen RNA und cDNA von Klon 13 (A) und Klon 26 (B) in biologischen Replikaten (1&2). Als Negativkontrollen dienten Wasser (H<sub>2</sub>O) und cDNA-Synthese Ansatz ohne zugegebene RNA (-RNA).

Die RT-PCR zeigt unter Verwendung von  $\beta$ -Aktin-spezifischen Oligonukleotidprimern, dass bei Klon 13 und Klon 26 jeweils gleiche Mengen an RNA in die cDNA-Synthese eingesetzt wurden (Abb. 10 A und B). Die Verwendung von hCTCFL-spezifischen Oligonukleotidprimern weist eine deutliche Expression von hCTCFL nach Induktion in beiden Klonen auf. Demgegenüber steht ein sehr schwaches PCR-Produkt im uninduzierten Zustand. Sämtliche Negativkontrollen (H<sub>2</sub>O und -RNA) sind frei von Verunreinigungen und weisen kein Signal in der PCR auf.



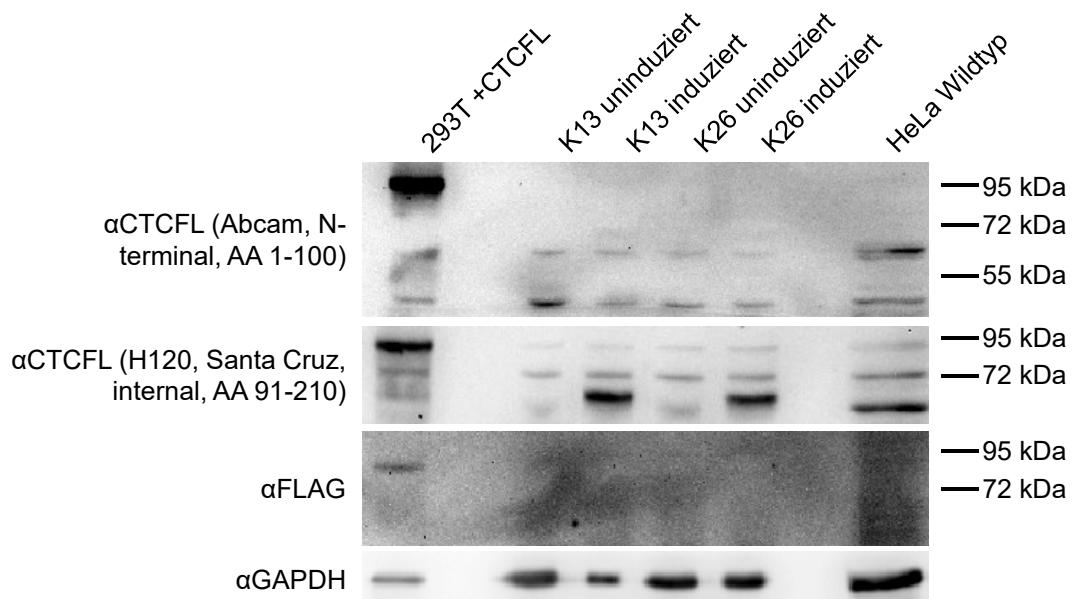
**Abb. 11: Beide Klone zeigen quantifizierbare Induktion von hCTCFL in der gleichen Größenordnung.** HeLa Zellen wurden mit pBI-EGFP-FLAG-hCTCFL und pTA-N transfiziert und mit Puromycin selektiert. Zell-Klone wurden für 48 h uninduziert (mit Doxyzyklin) oder induziert (ohne Doxyzyklin) kultiviert, anschließend wurde jeweils die RNA isoliert und in cDNA umgeschrieben. qRT-PCR mit Oligonukleotidprimern für CTCFL gegen cDNA von Klon 13 und Klon 26. Dargestellt ist jeweils der Mittelwert von zwei biologischen Replikaten inklusive Standardabweichung. Die Expression wurde gegen das Housekeeping-Gen  $\beta_2$ -Mikroglobulin normalisiert und dann der Expressionswert von uninduziertem Klon 13 als 1 gesetzt.



Bei genauer Quantifizierung dieser Induktion mittels RT-qPCR erhält man für beide Klone Werte im Bereich von 25-30-facher Zunahme wobei Klon 26 die höchste Expression von hCTCFL besitzt (Abb. 11). Somit konnte eine induzierbare Expression von CTCFL-RNA sowohl in Klon 13 als auch in Klon 26 nachgewiesen werden.

#### 4.1.2 CTCFL lässt sich in zwei HeLa Klonen nicht in voller Länge nachweisen

Nachdem der Nachweis auf RNA Ebene erfolgreich war, sollte der Nachweis auf Proteinebene erfolgen um die Herstellung des korrekten, funktionalen CTCFL Proteins abzusichern. Hierzu wurde jeweils Ganzzellextrakt mittels RIPA-Puffer für induzierte und uninduzierte Zellen der beiden Klone 13 und 26 gewonnen. Das erhaltene Protein wurde dann in einem SDS-PAGE-Gel aufgetrennt und im Western Blot verwendet. Der Nachweis von FLAG-hCTCFL erfolgte mit mehreren Antikörpern. Neben zwei CTCFL spezifischen Antikörpern, die gegen unterschiedliche Abschnitte des CTCFL-Proteins gerichtet sind, kam auch ein gegen das FLAG-Tag gerichteter Antikörper zum Einsatz.



**Abb. 12: Klon 13 und Klon 26 zeigen Expression einer verkürzten Variante von hCTCFL.** Zell-Klone wurden jeweils für 2 Tage durch Kultivierung ohne Doxyzyklin induziert und anschließend wurde mittels RIPA-Puffer Protein extrahiert. Als Positivkontrolle dienten transient mit pcDNA3-FLAG-hCTCFL transfizierte 293T Zellen, als Negativkontrolle HeLa Wildtyp Zellen, die keine CTCFL Expression aufweisen. Es wurden jeweils 20 µg Protein der einzelnen Proben elektrophoretisch in einem 10% SDS-PAGE-Trenngel aufgetrennt und mittels Wet-Blot auf eine PVDF-Membran übertragen. Die Immundetektion erfolgte mit Antikörpern gegen GAPDH, als Ladekontrolle, gegen das FLAG-Tag und zwei spezifischen Antikörpern gegen unterschiedliche Bereiche von hCTCFL. Detektion der Antikörper-Reaktion erfolgte unter Verwendung von chemilumineszenten HRP Substrates.

Unter Verwendung der gegen CTCFL gerichteten Antikörper zeigt die Positivkontrolle, bei der es sich um transient mit FLAG-hCTCFL transfizierte HEK293T Zellen handelt, starke Expression von CTCFL (Abb. 12, Spur 1). Dieses Protein, welches durch eine Molekülmasse von 95 kDa charakterisiert ist, kann in keiner der HeLa-Klon Proben detektiert werden (Spur 4 & 6). Unter Verwendung des gegen die CTCFL Aminosäuren 91 bis 210 gerichteten Antikörpers (H120) ist ein spezifisches, induzierbares Protein in Klon 13 und 26 detektierbar das jedoch eine niedrigere Laufhöhe von unter 72 kDa aufweist. Dieses Protein konnte mit dem gegen den N-Terminus gerichteten Antikörper der Firma Abcam nicht nachgewiesen werden (Abb. 12). Die FLAG Detektion, gegen das N-terminal ständige FLAG-Tag von FLAG-hCTCFL, ist nur für die Positivkontrolle feststellbar. Die Detektion mit GAPDH Antikörper zeigt, dass in allen untersuchten Proben Protein vorliegt, wobei die jeweilige geladene Proteinmenge leichte Unterschiede aufweist. Die höchste Proteinmenge liegt bei der Negativkontrolle (HeLa Wildtyp) vor, die niedrigste bei der Positivkontrolle (293T +CTCFL). Somit ist in den Klonen von der Expression einer verkürzten Version von FLAG-hCTCFL auszugehen, bei der Bereiche des N-Terminus inklusive des FLAG-Tags im Zuge der stabilen Transfektion verloren gingen. Für weitere Experimente zur Untersuchung der zellulären CTCFL Funktion sind diese Klone somit nicht nutzbar, da das Protein nicht in voller Länge vorlag und somit vollständige korrekte Funktionalität nicht gegeben ist.

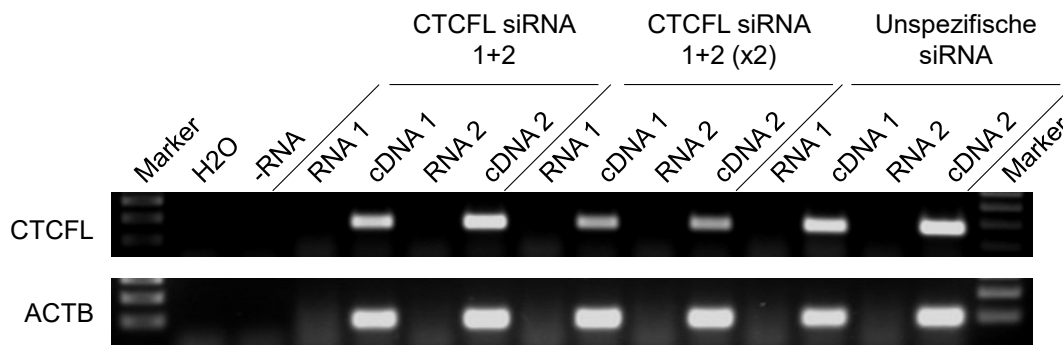
## **4.2 CTCFL-Analysen in humanen K562-Zellen**

K562 stellt die menschliche Zelllinie mit der höchsten bekannten und am besten nachgewiesenen CTCFL Expression dar. Somit bietet sie sich für Experimente im Zusammenhang mit CTCFL an. Wie auch bei HeLa handelt es sich bei K562 um eine sehr gut charakterisierte humane Zelllinie. Eine Vielzahl von öffentlichen Datensätzen ist frei verfügbar und somit für Korrelations-Analysen verwendbar. Weiterhin lassen sich die in Suspension wachsenden Zellen, die ursprünglich von einer an chronisch myeloider Leukämie erkrankten Patientin gewonnen wurden, schnell und leicht kultivieren.

### **4.2.1 Aufeinanderfolgende Transfektion eines spezifischen siRNA-Mixes führt zu Reduktion der CTCFL Expression**

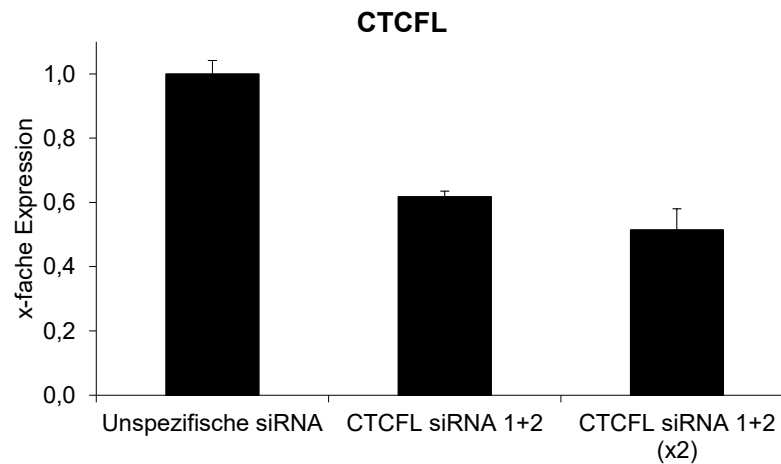
Da in K562 CTCFL bereits in ausreichenden Mengen exprimiert ist boten sich Überexpressionsexperimente hier nicht an. Stattdessen sollte die Herabregulierung von CTCFL, unter Verwendung von siRNA, Aufschluss über funktionale Einflüsse von CTCFL geben.

Zur Charakterisierung der Auswirkung der Depletion von CTCFL auf die Expression anderer Gene wurde eine Kombination von zwei, spezifisch gegen CTCFL gerichteten, siRNAs verwendet. Nach erfolgter Transfektion der K562 Zellen mit entweder unspezifischer siRNA, oder der Kombination von CTCFL siRNA 1+2 in einfacher Transfektion oder an zwei aufeinander folgenden Tagen, wurde RNA isoliert in cDNA umgeschrieben und mittels RT-PCR analysiert.



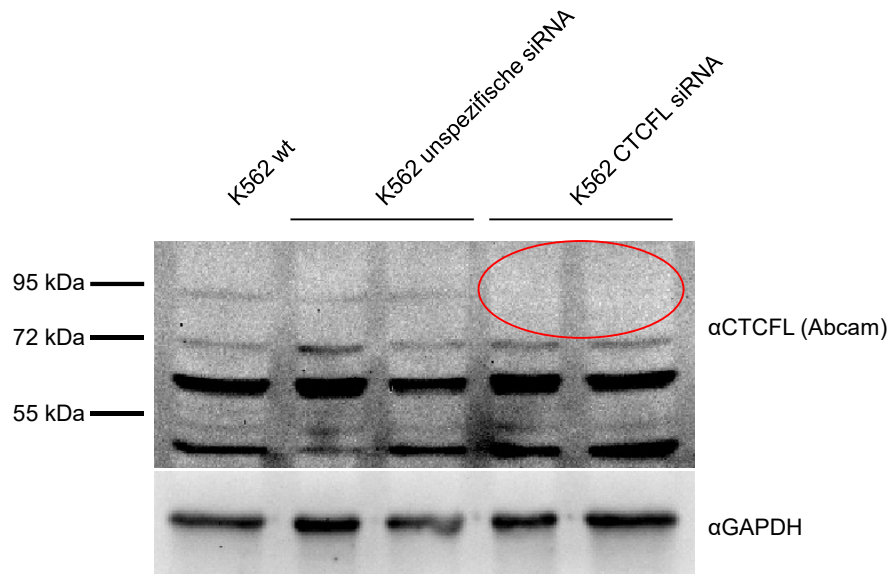
**Abb. 13: Die Kombination von siRNA 1 und 2 in wiederholter Transfektion führt zu stärksten Abnahme des hCTCFL Signals.** K562 Zellen wurden entweder mit unspezifischer siRNA oder einem Mix aus zwei spezifischen siRNAs gegen CTCFL mittels RNAiMAX transfiziert. Der spezifische siRNA Mix wurde sowohl in einfacher Transfektion als auch in zweifacher Transfektion (x2) an aufeinanderfolgenden Tagen eingesetzt. Nach der finalen Transfektion erfolgte nach 12 h ein Mediumswechsel und nach 48 h wurde die RNA isoliert und in cDNA umgeschrieben RT-PCR mit Oligonukleotidprimern für  $\beta$ -Aktin (ACTB) und CTCFL in biologischen Replikaten (1&2). Als Negativkontrollen dienten Wasser (H<sub>2</sub>O), cDNA-Synthese Ansatz ohne zugegebene RNA (-RNA) und die jeweiligen in die cDNA-Synthese eingesetzte RNA.

Die RT-PCR mit  $\beta$ -Aktin-spezifischen Oligonukleotidprimern weist in allen Ansätzen auf das Vorliegen von vergleichbaren Mengen an cDNA hin (Abb. 13). Die Verwendung von CTCFL-spezifischen Oligonukleotidprimern zeigt klar reduzierte Mengen an cDNA bei K562 Zellen die zweimal mit CTCFL siRNA transfiziert wurden. Die untersuchten Negativkontrollen H<sub>2</sub>O, -RNA und die jeweiligen zur cDNA-Synthese eingesetzten RNAs sind frei von Verunreinigungen und führen zu keinem Signal in der PCR. Zur genauen Bestimmung der Abnahme der CTCFL Expression wurde anschließend eine RT-qPCR durchgeführt.



**Abb. 14: Die Kombination von siRNA 1 und 2 in wiederholter Transfektion reduziert die CTCFL Expression in K562 Zellen um die Hälfte.** K562 Zellen wurden entweder mit unspezifischer siRNA oder einem Mix aus zwei spezifischen siRNAs gegen CTCFL mittels RNAiMAX transfiziert. Der spezifische siRNA Mix wurde sowohl in einfacher Transfektion als auch in zweifacher Transfektion an aufeinanderfolgenden Tagen eingesetzt. Nach der finalen Transfektion erfolgte nach 12 h ein Mediumswechsel und nach 48 h wurde die RNA isoliert und in cDNA umgeschrieben. Die qRT-PCR wurde mit Oligonukleotidprimern für CTCFL durchgeführt. Dargestellt ist jeweils der Mittelwert von zwei biologischen Replikaten inklusive Standardabweichung. Die Expression wurde unter Verwendung der GeNorm Software gegen vier Housekeeping Gene (ACTB, RPL13, GAPDH & UBC) normalisiert und dann der Expressionswert von mit unspezifischer RNA transfizierten K562 Zellen als 1 gesetzt.

Die Quantifizierung des Knockdowns zeigt bei beiden mit spezifischer CTCFL siRNA transfizierten Ansätzen eine reduzierte Expression von CTCFL (Abb. 14). Während eine einmalige Transfektion die Expression auf ca. 60% reduziert führt die zweifache Transfektion zum größten Effekt und halbiert die Expression auf ca. 50%. Um die Abnahme der CTCFL Expression final abzusichern sollte der Knockdown auch auf Proteinebene gezeigt werden. Hierzu wurde Protein von doppelt transfizierten CTCFL siRNA K562 Zellen sowie entsprechenden Kontrollzellen mittels RIPA Extraktion gewonnen und mit dem entsprechenden CTCFL Antikörper (Abcam) im Western Blot untersucht.



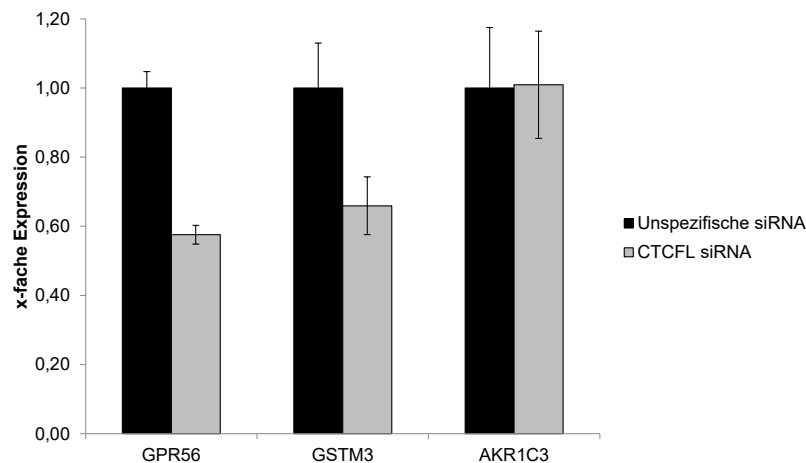
**Abb. 15: Der CTCFL Knockdown kann auf Proteinebene in biologischen Duplikaten bestätigt werden.** K562 Zellen wurden an aufeinanderfolgenden Tagen zweimal mit unspezifischer siRNA oder einem Mix aus zwei spezifischen siRNAs gegen CTCFL mittels RNAiMAX transfiziert. Nach der zweiten Transfektion erfolgte nach 12 h ein Mediumswechsel und nach 48 h wurde Protein mit RIPA-Puffer isoliert. Als weitere Kontrolle wurde Protein aus untransfizierten K562 Wildtyp-Zellen isoliert. Es wurden jeweils 20 µg Protein der einzelnen Proben elektrophoretisch in einem 10% SDS-PAGE-Trenngel aufgetrennt und mittels Wet-Blot auf eine PVDF-Membran übertragen. Die Immundetektion erfolgte mit Antikörpern gegen GAPDH, als Ladekontrolle, und unter Verwendung des Abcam Antikörpers gegen CTCFL. Detektion der Antikörper-Reaktion erfolgte unter Verwendung von chemilumineszenten HRP Substrates.

Der Western Blot zeigt bei Detektion mit CTCF-spezifischem Antikörper in biologischen Replikaten, dass in den, mit spezifischer CTCFL siRNA transfizierten, Zellen, das Protein auf Höhe von 95 kDa nicht detektierbar ist (Abb. 15, roter Kreis). Sowohl in den K562 Wildtyp Zellen als auch in den mit unspezifischer siRNA transfizierten K562 Zellen ist ein Signal für CTCFL auf Höhe von 95 kDa nachzuweisen. Bei allen weiteren detektierbaren Signalen handelt es sich um unspezifischen Hintergrund, der keine Unterschiede zwischen den Proben zeigt. Unter Verwendung eines GAPDH-spezifischen Antikörpers sind gleiche Mengen an Protein in allen untersuchten Proben festzustellen. Somit konnte die Herabregulation von CTCFL auch auf Proteinebene bestätigt werden.

#### 4.2.2 Die Reduktion der CTCFL Expression hat einen Einfluss auf ausgewählte Kandidaten-Gene

Um den Einfluss der CTCFL Abnahme auf die Expression weiterer Gene zu untersuchen, wurden einige Kandidaten-Gene untersucht. Es wurde angenommen, dass humane homologe Gene zu Genen aus der Maus, für die ein Einfluss von CTCFL bekannt war, am ehesten auch im menschlichen System eine Änderung zeigen könnten. Bei den untersuchten Genen handelt es sich um GPR56 (Maus: *Gpr56*), GSTM3 (Maus: *Gstm5*) und

AKR1C3 (Maus: *Akr1c18*). Für diese Gene konnte ein Einfluss von *Ctcf* Überexpression im Maussystem gezeigt werden (Dissertation, S. Dienstbach), *Gstm5* und *Gpr56* wurden dabei positiv reguliert, während *Akr1c18* reprimiert wurde. Zur Analyse der Expression der humanen Homologe, wurde die vorher verwendete cDNA (4.2.1) in RT-qPCR-Experimenten, unter Verwendung von jeweils spezifischen Oligonukleotidprimern, ausgewertet.



**Abb. 16: GPR56 und GSTM3 zeigen eine geringere Expression nach CTCFL Knockdown.** K562 Zellen wurden entweder mit unspezifischer siRNA oder einem Mix aus zwei spezifischen siRNAs gegen CTCFL zweimal an aufeinanderfolgenden Tagen mittels RNAiMAX transfiziert. Nach der zweiten Transfektion erfolgte nach 12 h ein Mediumwechsel und nach 48 h wurde die RNA isoliert und in cDNA umgeschrieben. Die qRT-PCR wurde mit spezifischen exonübergreifenden Oligonukleotidprimern für die drei Gene GPR56, GSTM3 und AKR1C3 durchgeführt. Dargestellt ist jeweils der Mittelwert von zwei biologischen Replikaten inklusive Standardabweichung. Die Expression wurde unter Verwendung der GeNorm Software gegen vier Housekeeping Gene (*ACTB*, *RPL13*, *GAPDH* & *UBC*) normalisiert und dann jeweils der Expressionswert der Kontroll-siRNA-Behandlung als 1 gesetzt.

Die RT-qPCR zeigt für zwei (GPR56 und GSTM3) der drei untersuchten Gene eine verminderte Expression, während der CTCFL Knockdown keinen nachweisbaren Einfluss auf die Expression von AKR1C3 hat (Abb. 16). Dieser Effekt fällt am stärksten bei GPR56 mit einer Abnahme um ca. 40% aus. Somit zeigen sowohl GPR56 und GSTM3 den gegensätzlichen Effekt zu den Experimenten mit CTCFL Überexpression. Um einen Hinweis auf einen funktionalen Zusammenhang zwischen CTCFL und GPR56 Expression zu erhalten wurde der genomische DNA-Bereich bezüglich CTCF und CTCFL Bindung betrachtet. Hierzu wurden die öffentlichen genomweiten K562-ENCODE-Bindedaten genutzt. Diese ChIPseq Daten wurden im UCSC Genome Browser geladen und ausgewertet.

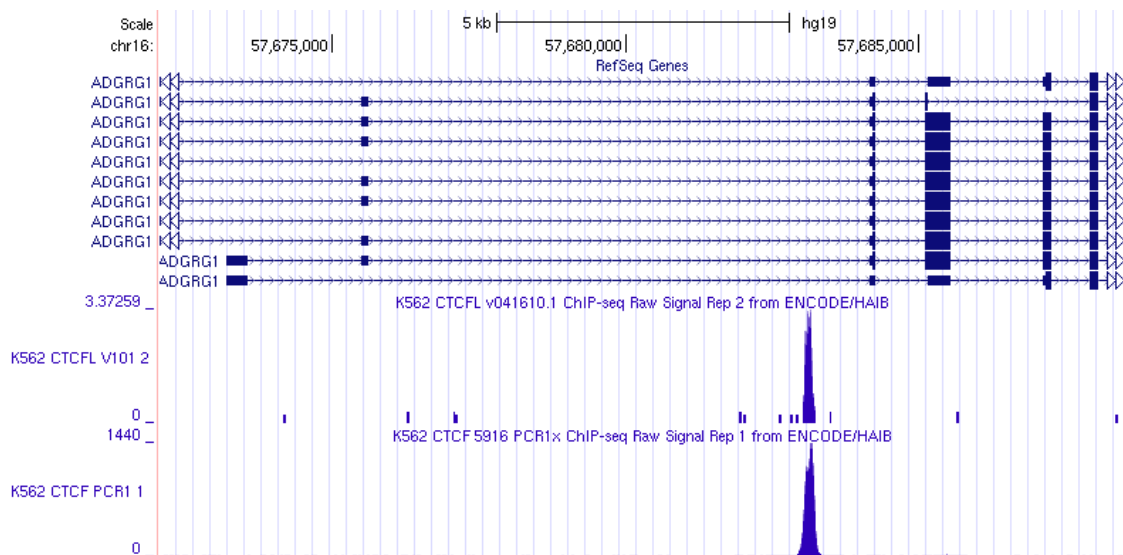


Abb. 17: **GPR56 zeigt eine Bindung von CTCF und CTCFL im Intron vor dem translatieren Bereich.** Dargestellt ist ein Ausschnitt des Genbereiches von GPR56 (seit 2015 offiziell als ADGRG1 bezeichnet) mit mehreren bekannten Isoformen des Gens. Exons sind als blaue Blöcke angegeben, der translatierte Bereich jeweils als dickerer Block. Darunter sind jeweils die Bindeprofile von CTCFL und CTCF in K562 Zellen angegeben. Bei diesen Profilen handelt es sich um die öffentlichen Daten des ENCODE Konsortiums.

Sowohl CTCF als auch CTCFL zeigen Bindung im intronischen Bereich des GPR56-Gens (Abb. 17). Diese Bindung befindet sich für alle annotierten Isoformen von GPR56 vor dem Translationsstart des Proteins, was für eine Rolle dieser Bindung bei der Bildung von funktionalem Protein sprechen könnte. Somit konnte ein möglicher Zusammenhang für die Co-Regulation von GPR56 und CTCFL identifiziert werden.

### 4.2.3 Bioinformatische Untersuchung des Zusammenhangs von CTCFL Bindung mit genomweiten humanen Chromatin-Eigenschaften (ENCODE)

#### 4.2.3.1 Vorstellung ENCODE Project

Im Rahmen des ENCODE (ENCyclopedia Of DNA Elements) Projektes schlossen sich 2003 unter Führung des amerikanischen National Human Genome Research Institute (NHGRI) mehrere Institutionen zusammen um, in einem großen genom-weiten Ansatz, alle funktionalen DNA Bereiche des menschlichen Genomes zu identifizieren (The ENCODE Project Consortium 2004). Im Zuge der damals neu verfügbaren gesamten menschlichen Genomsequenz sollten nun also alle Protein-codierenden Gene, Nicht-kodierende Gene, transkriptionell regulative Elemente und Bereiche die für den Aufbau und die Dynamik der Chromosomen-Struktur von Bedeutung sind, analysiert werden. Das Projekt war in zwei zeitliche Phasen gegliedert. Eine Pilot-Phase in der 1% des Genomes, ca. 30 Megabasen, untersucht und optimale experimentelle Strategien entwickelt werden

sollten, parallel zu dieser lief eine Technologie-Entwicklungs-Phase, in der nach neuen technischen Möglichkeiten zur Analyse regulatoriver Elemente im Genom gesucht wurde. Nach Abschluss dieser Phasen sollte in der Hauptphase die restlichen 99% des Genomes analysiert werden.

Im Rahmen der Pilot-Phase wurden insgesamt 44 Bereiche des Genomes untersucht. Die Abschnitte wurden teilweise gezielt, teilweise randomisiert ausgewählt um möglichst abwechslungsreiche Bereiche des Genomes abzudecken. Die Größe der analysierten Bereiche schwankte zwischen 0,5 und 2 Megabasen. Die Pilot-Phase konnte im Jahr 2007 erfolgreich abgeschlossen werden (The ENCODE Project Consortium 2007).

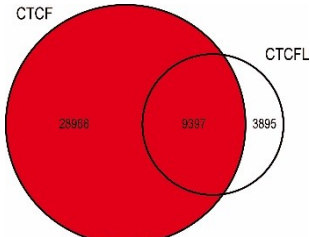
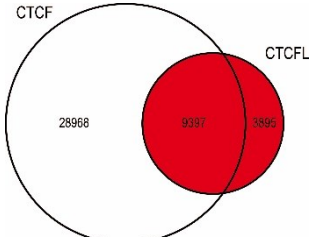
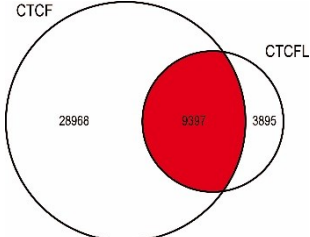
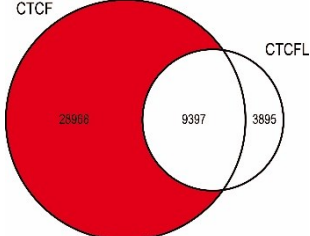
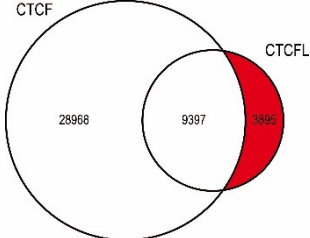
Im Rahmen dieser Hauptphase wurden riesige Mengen an genomweiten Daten produziert und ausgewertet (The ENCODE Project Consortium 2012). Alle Daten wurden öffentlich zugänglich online hinterlegt und für weitere Analysen bereitgestellt. Die Daten sind über das Web-Portal [www.encodeproject.org](http://www.encodeproject.org) abrufbar.

#### 4.2.3.2 Korrelation der CTCFL Bindedaten in K562 Zellen

Eine der drei, am besten untersuchten, Zelllinien des ENCODE Projektes stellt die Leukämie-Linie K562 dar. Im Rahmen der Identifikation von Transkriptionsfaktor-Bindestellen wurden unter anderem auch die genomweiten Bindestellen von CTCFL in diesen Zellen identifiziert. Anhand dieser öffentlichen Daten sollte dann bestimmt werden welche Chromatin-Merkmale und andere DNA-Bindefaktoren die beste Übereinstimmung mit CTCFL und/oder CTCF zeigen. Anhand dieser bioinformatischen Analyse sollten auch Faktoren und Merkmale identifiziert werden die jeweils nur für CTCF- oder CTCFL-gebundene Bereiche charakteristisch sind. Mit Hilfe von M. Bartkuhn wurden die bioinformatischen Analysen durchgeführt. Dazu wurden alle benötigten Datensätze von der zentralen ENCODE Webseite ([www.encodeproject.org](http://www.encodeproject.org)) heruntergeladen und lokal weiter verarbeitet. Als erstes wurden Klassen von CTCF und CTCFL Bindestellen festgelegt. Durch Verrechnung der Binde-Datensätze von CTCF und CTCFL wurden die Bindestellen in Untergruppen eingeteilt (Tab. 1). Dargestellt sind mit VennDis erstellte Venn-Diagramme die den Überlapp der CTCF und CTCFL Datensätze zeigen, und in Rot jeweils der Bereich der Daten der in jeder Gruppe repräsentiert ist. Insgesamt zeigen ca. 70,7% aller CTCFL Bindestellen in K562 auch Bindung von CTCF.

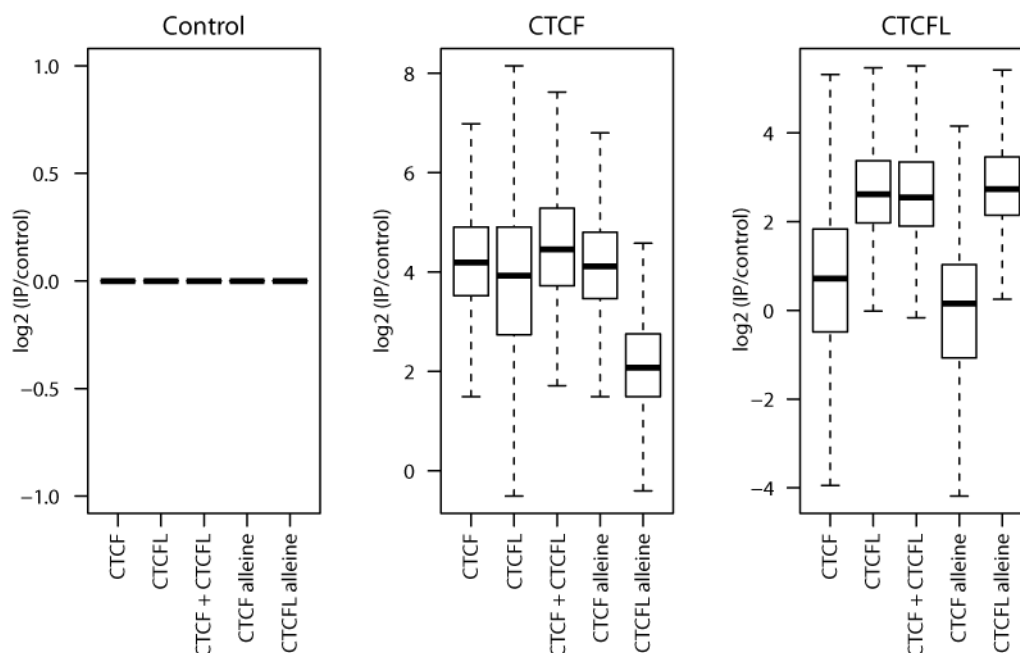


Tab. 1: Einteilung der Unterklassen von CTCF/CTCFL Bindestellen in K562 Zellen

Name	Venn-Diagramm	Zahl der Instanzen
CTCF		38.365 Bindestellen
CTCFL		13.292 Bindestellen
CTCF + CTCFL		9.397 Bindestellen
CTCF alleine		28.968 Bindestellen
CTCFL alleine		3.895 Bindestellen

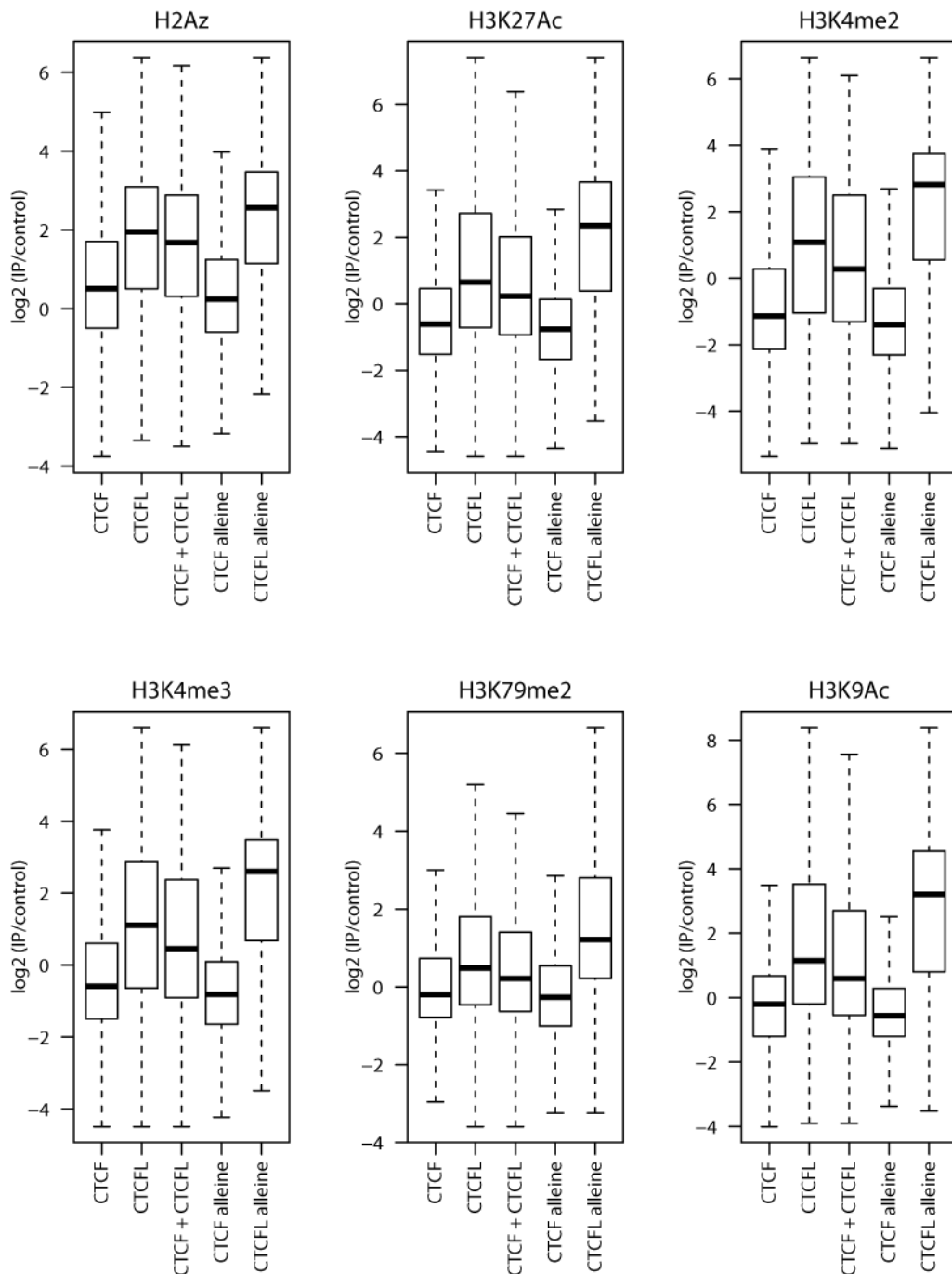
Die Betrachtung der Daten zeigt, dass ein großer Teil (70,7%) der CTCFL Bindestellen mit CTCF korreliert, aber auch ein Teil von nur CTCFL gebundenen Bereichen vorhanden ist. Somit sollte nun vor allem untersucht werden, was den Unterschied zwischen CTCF alleine gebundenen und CTCFL alleine gebundenen DNA Bereichen ausmacht. Als erste Kontrolle wurden die festgelegten Gruppen, nach Normalisierung über einen Kontroll-Datensatz, mit den zugrundeliegenden CTCF und CTCFL Datensätzen korre-

liert. Bioinformatisch wurden jeweils die Bereiche +/- 500 bp um die identifizierten Bindestellen betrachtet, jeweils alle Reads in diesem Bereich addiert und dann mittels Boxplots als  $\log_2$  transformierte IP Bindung über Kontrolle dargestellt.



**Abb. 18: Die gewählten CTCF/CTCF-Bindeklassen zeigen die jeweilige erwartete Bindungsverteilung von CTCF und CTCFL.** Dargestellt sind jeweils Boxplots der  $\log_2$  transformierten Bindung über die gewählten Untergruppen von CTCF/CTCF Bindestellen. Control zeigt hier die Untergruppen normalisiert gegen einen IP-Datensatz der mit unspezifischem Antikörper generiert wurde. In den Boxen sind jeweils die Datenbereiche von 25% bis 75% inklusive dem Median angezeigt, die Whisker geben dann Auskunft über die gesamte Spanne der Datenpunkte.

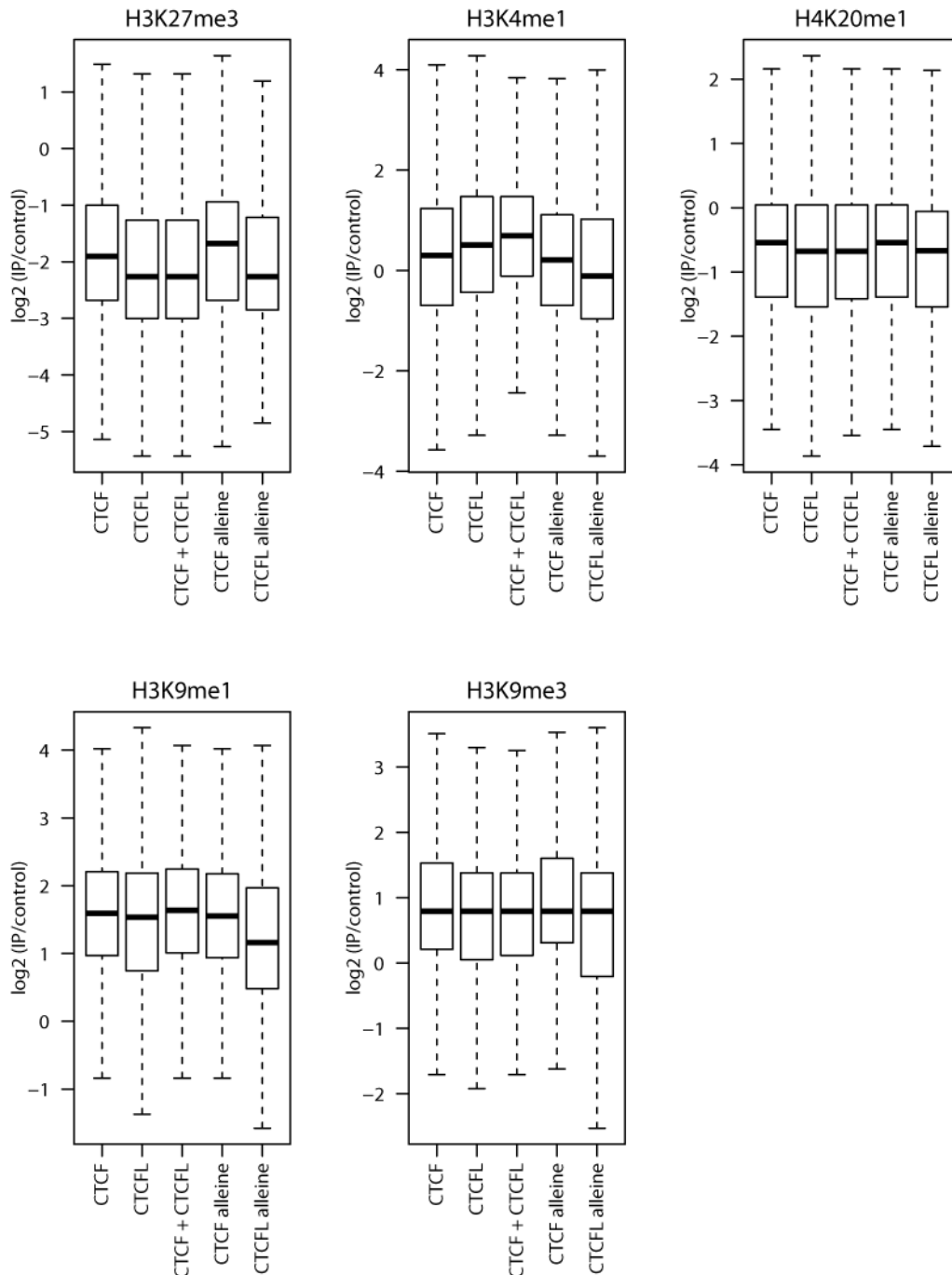
Sowohl CTCF als auch CTCFL zeigen eine Anreicherung in ihren spezifischen Bindeklassen. CTCF die schwächste Anreicherung in der Klasse *CTCF alleine*, CTCFL umgekehrt in der Klasse *CTCF alleine*. Während bei CTCFL kein Unterschied zwischen den Klassen *CTCF* und *CTCF + CTCFL* erkennbar ist, zeigt *CTCF + CTCFL* eine höhere Anreicherung von CTCF als die Gesamtgruppe *CTCF*. Somit waren die Gruppen ausreichend kontrolliert und es konnte die eigentliche Korrelationsanalyse durchgeführt werden. Dafür wurden die entsprechenden Datensätze für Histon-Modifikationen und Transkriptionsfaktoren aus der ENCODE Datenbank heruntergeladen und jeweils die gleiche Analyse wie für die Kontrollen durchgeführt. Als erstes sind die Ergebnisse für die Histon-Modifikationen dargestellt, die eine positive Korrelation mit CTCFL Bindung aufweisen.



**Abb. 19: Mehrere aktive Histonmodifikationen zeigen eine positive Korrelation mit CTCFL Bind-einstanzen.** Dargestellt sind jeweils Boxplots für aktive Histonmodifikationen der log<sub>2</sub> transformierten Bindungen über die gewählten Untergruppen von CTCF/CTCFL Bindestellen. In den Boxen sind jeweils die Datenbereiche von 25% bis 75% inklusive dem Median angezeigt, die Whisker geben dann Auskunft über die gesamte Spanne der Datenpunkte.

Insgesamt zeigen 5 der untersuchten Histon-Modifikationen und die Histon Variante H2A.Z eine klare Spezifität für CTCFL gebundene Regionen denn alle Modifikationen sind jeweils in der Klasse *CTCFL alleine* am stärksten angereichert. Bei diesen Modifikationen handelt es sich um die klassischen mit offenem, aktivem Chromatin assoziierten

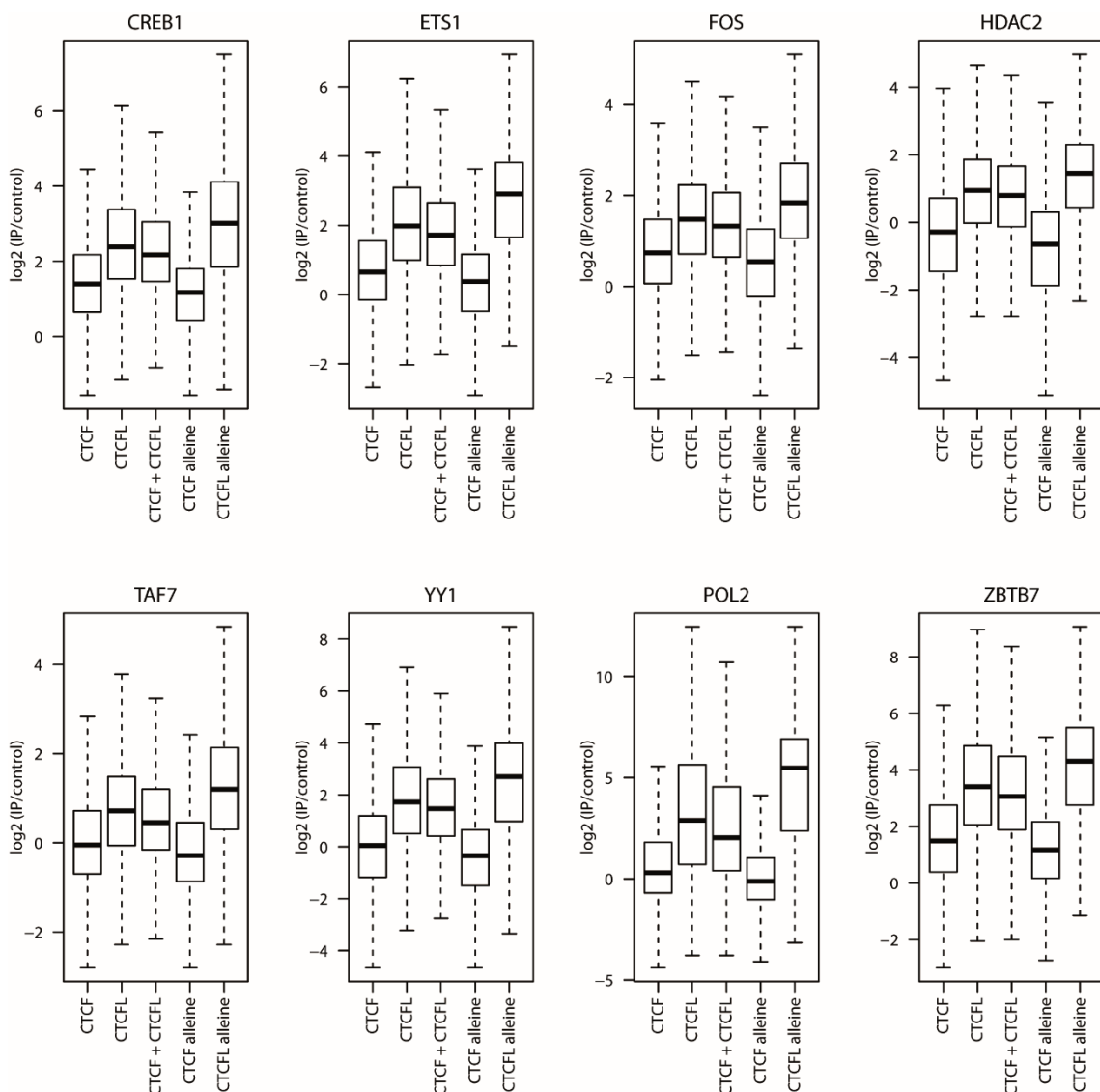
Modifikationen H3K4me2, H3K4me3, H3K9Ac, H3K27ac sowie H3K79me2. Weitere untersuchte Modifikationen zeigten diese Korrelation mit CTCFL Bindung nicht.



**Abb. 20: Untersuchte repressive Histon-Modifikationen zeigen keine positive Korrelation mit CTCFL Bindung.** Dargestellt sind jeweils Boxplots für repressive, nicht speziell mit aktiver Transkription assoziierten, Histonmodifikationen der  $\log_2$  transformierten Bindung über die gewählten Untergruppen von CTCF/CTCFL Bindestellen. In den Boxen sind jeweils die Datenbereiche von 25% bis 75% inklusive dem Median angezeigt, die Whisker geben dann Auskunft über die gesamte Spanne der Datenpunkte.

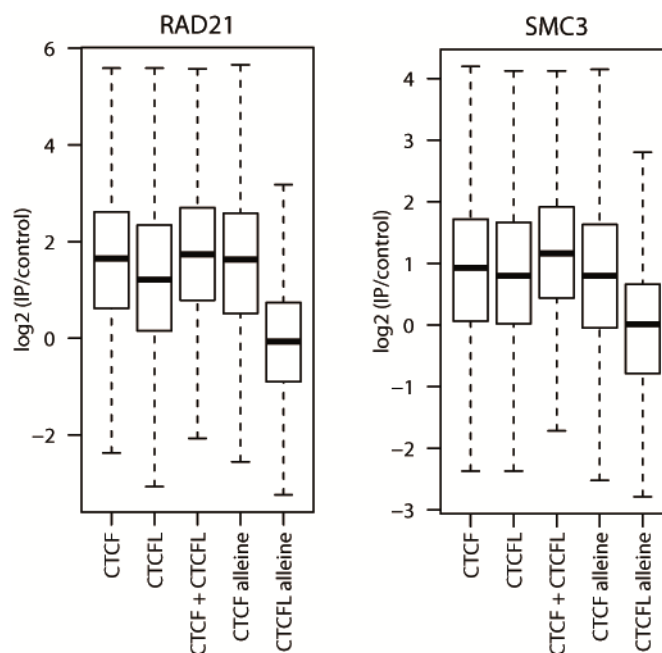
Fünf untersuchte Histon-Modifikationen, für die keine Verbindung zu aktiver Transkription bekannt ist, zeigen keine Spezifität für einen der beiden Faktoren CTCF oder CTCFL (Abb. 20). Bei diesen Modifikationen handelte es sich um H3K27me3, H3K4me1,

H4K20me1, H3K9me1 und H3K9me3, die meist mit repressivem Chromatin in Verbindung gebracht werden. Die fünf Untergruppen der CTCF/CTCFL-Bindestellen zeigen für diese fünf Modifikationen einen vergleichbaren Wert für die IP-Bindung. Diese IP-Bindungswerte bewegen sich auch in einem sehr niedrigen Bereich, was für kein Vorhandensein der Histon-Modifikationen an CTCF und CTCFL Bindestellen spricht. Da bei den untersuchten Histon-Modifikationen diese klare Präferenz für aktive Merkmale zu erkennen ist, wurden auch die öffentlichen Daten für die Bindung von bekannten Transkriptionsfaktoren genutzt, um zu überprüfen ob ein ähnliches differentielles Bindeverhalten in Bezug auf CTCF und CTCFL vorliegen könnte.



**Abb. 21: Untersuchte Transkriptions-Faktoren zeigen eine positive Korrelation mit CTCFL Bindung.** Dargestellt sind jeweils Boxplots für in den ENCODE Datenbanken erhobene Transkriptionsfaktoren, der  $\log_2$  transformierten Bindung über die gewählten Untergruppen von CTCF/CTCFL Bindestellen. In den Boxen sind jeweils die Datenbereiche von 25% bis 75% inklusive dem Median angezeigt, die Whisker geben dann Auskunft über die gesamte Spanne der Datenpunkte.

Die untersuchten Transkriptionsfaktoren (CREB1, ETS1, FOS, HDAC2, TAF7, YY1, POLII & ZBTB7) zeigen jeweils eine höhere Bindung an die CTCFL beinhaltenden Untergruppen (Abb. 21). Die höchste Bindung ist für alle Faktoren jeweils für die *CTCFL alleine* Gruppe zu erkennen. Vor allem bei Betrachtung der Auswertung für RNA Polymerase II (POL2) zeigen die CTCF-Untergruppen eine Bindung um 0 während die CTCFL-Untergruppen eine sehr starke Bindung aufweisen. Abschließend sollte noch die Assoziation von Bestandteilen des Cohesin-Komplexes mit CTCF und CTCFL untersucht werden. Cohesin interagiert mit CTCF im Rahmen von grundlegenden Zellfunktionen und räumlicher Organisation im Zellkern. Für CTCFL ist aktuell keine funktionelle Interaktion mit Cohesin beschrieben.



**Abb. 22: Komponenten des Cohesin-Komplexes zeigen negative Korrelation mit CTCFL Bindung.** Dargestellt sind jeweils Boxplots für Komponenten des Cohesin-Komplexes, der log<sub>2</sub> transformierten Bindung über die gewählten Untergruppen von CTCF/CTCFL Bindestellen. Neben In den Boxen sind jeweils die Datenbereiche von 25% bis 75% inklusive dem Median angezeigt, die Whisker geben dann Auskunft über die gesamte Spanne der Datenpunkte.

Die bioinformatische Analyse zeigt für die beiden untersuchten Cohesin-Komponenten, RAD21 und SMC3, jeweils die niedrigste Bindung für die CTCFL alleine Untergruppe (Abb. 22). Somit sind RAD21 und SMC3 die einzigen Faktoren der durchgeführten Analyse, die eine CTCF-spezifische Bindung aufweisen. Alle anderen untersuchten Faktoren und Modifikationen zeigen entweder spezifische Anreicherung an CTCFL Bindestellen oder keine Präferenz/keine Bindung über CTCF/CTCFL-Bindestellen.

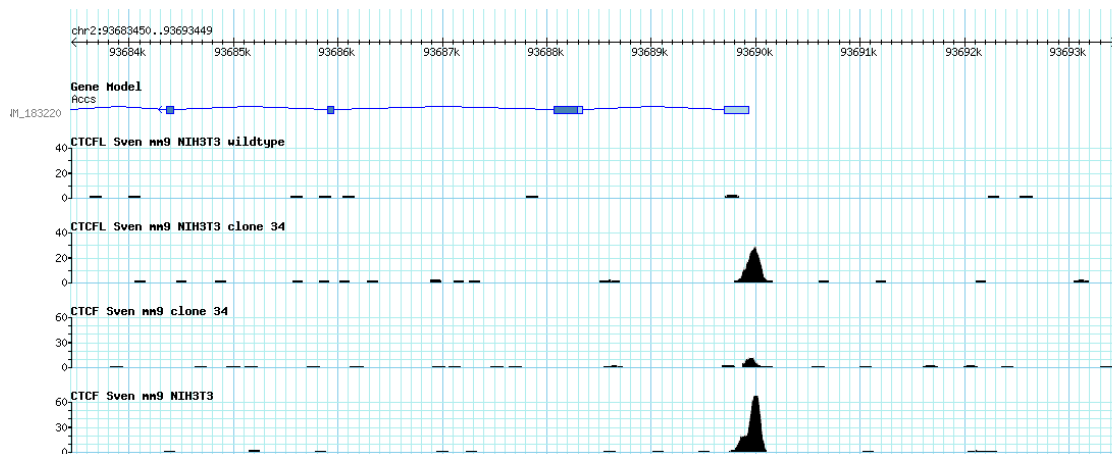
### 4.3 Untersuchungen des Bindeverhaltens von CTCFL in stabil transfizierten NIH3T3 Zellen

Bei dem NIH3T3 Klon 34 handelt es sich um einen von S. Dienstbach im Rahmen seiner Promotion etablierten stabilen Zell-Klon mit induzierbarer muriner CTCFL Expression. Hierzu wurden NIH3T3 Zellen mit dem Zwei-Vektor-System pBI-EGFP-FLAG-m*Ctcf* und pTA-N, die zusammen als Tet-off-System wirken, transfiziert, selektioniert und charakterisiert. Klon 34 zeichnete sich durch niedrige Expression von CTCFL im uninduzierten Zustand aus und zeigte eine vielfache Induktion von CTCFL nach Kultivierung der Zellen ohne Doxyzyklin. Dieser Klon wurde dann bezüglich genomweiter Gen-Expressionseinflüsse von CTCFL analysiert. Im Rahmen dieser Analyse trat *Gpr56* als, durch CTCFL positiv reguliertes Gen, auf. Auch die genomweiten Bindungsmuster von CTCF und CTCFL wurden mittels ChIPseq erhoben und ausgewertet (Dissertation S. Dienstbach). Im Rahmen weitergehender Analysen sollten nun in diesem System weitere Fragestellungen bezüglich CTCF und CTCFL bearbeitet werden. Diese Fragestellungen betrafen den möglichen Einfluss von CTCFL auf die Bindung von CTCF an geteilten Bindestellen, CTCFL vermittelte Änderungen am Histon-Modifikations-Besatz der Bindestellen und die Untersuchung des Zusammenhanges von CTCFL- und Cohesin-Bindung.

#### 4.3.1 CTCFL hat keinen klaren Einfluss auf CTCF-Bindung an geteilten Bindestellen

Im Klon 34 sollte untersucht werden, ob durch die induzierbare Expression und die dadurch vermittelte DNA-Bindung von CTCFL eine Änderung der CTCF Bindung an bestimmte Stellen des Genoms einhergeht. Ein erster Ansatz sollte eine Überprüfung von Kandidaten-Bereichen darstellen. Hierbei wurden, entsprechend ihrer Änderung der CTCF Bindung, in Antwort auf CTCFL Überexpression, ausgewählte genomische Bereiche mittels individueller ChIP Experimente ausgewertet. Die Datengrundlage war hierbei die Analyse der CTCF und CTCFL ChIPseq Daten aus den NIH3T3-Zellen. Nach eingehender Filterung der Daten nach folgenden Kriterien: False Discovery Rate kleiner 5%, Entfernung der Bindestellen zu Promotoren von weniger als 2000 bp und Änderung der CTCF Bindung größer der Faktor 1,4, blieben 176 CTCF Bindestellen übrig, die Änderung ihrer Bindung nach Überexpression von CTCFL zeigen. Von diesen 176 Stellen nimmt jedoch nur an 91 die Bindung von CTCF ab, an den anderen 85 nimmt die CTCF Bindung zu. Ein genereller reprimierender Einfluss von CTCFL auf die CTCF Bindung

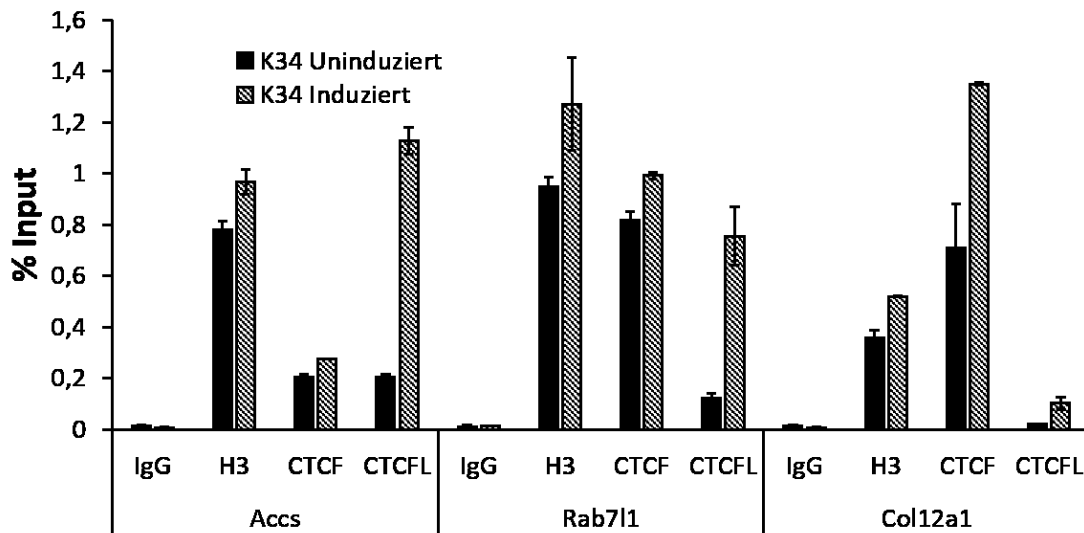
ist somit nicht erkennbar. Aus der Liste der Bereiche mit abnehmender CTCF Bindung wurden 3 Bereiche ausgewählt und der genomische Bereich genauer betrachtet. Es handelt sich dabei um die CTCFL Bindestellen in den Promotoren der deregulierten Gene Accs, Rab711 und Col12a1. Beispielhaft ist im folgenden ein genomischer Ausschnitt im Bereich des Accs Gens inklusive der CTCF und CTCFL ChIPseq Bindedaten dargestellt (Abb. 23).



**Abb. 23: Genomischer Bereich des Gens Accs der die abnehmende Bindung von CTCF in Folge von CTCFL Expression zeigt.** Gezeigt ist ein Teil des Genbereiches von Accs, Exons sind in blau angegeben, der translatierte Bereich in dunklerem Blau, untranslatierte Bereiche in hellerem. Darunter sind jeweils die CTCF- und CTCFL-Bindeprofile in induziertem Klon 34 und untransfizierten NIH3T3-Zellen dargestellt.

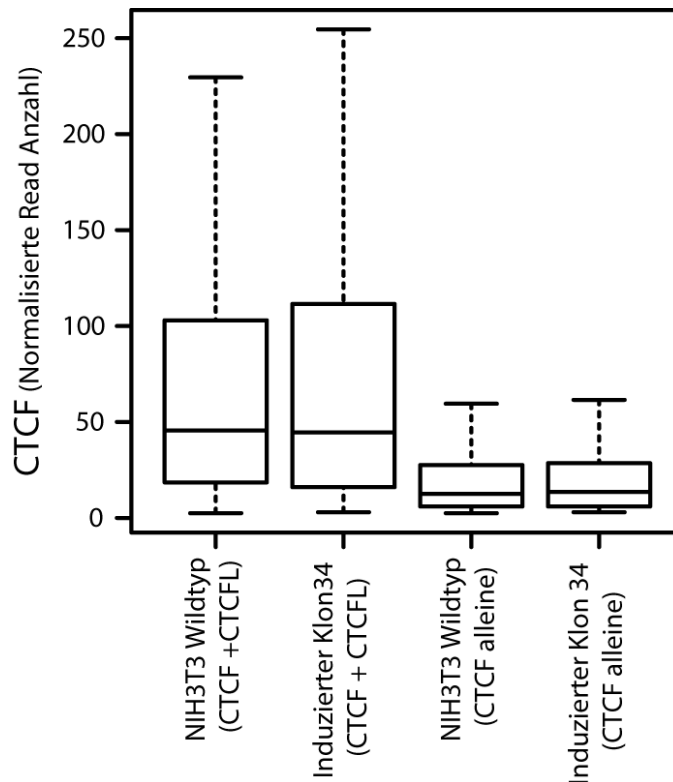
Der genomische Beispielbereich Accs zeigt Bindung von CTCF und CTCFL im Promoter des Accs Gens. Für CTCFL lässt sich nur in induziertem Klon 34 Bindung von CTCFL erkennen, in NIH3T3 Wildtyp Zellen ist keine Bindung detektierbar. CTCF zeigt eine Abnahme der Bindung von den Wildtyp-Zellen zum Klon 34. Somit weisen die genomweiten Datensätze auf eine mögliche Konkurrenz der beiden Faktoren an ihren geteilten Bindestellen hin. Um diesen Befund der genomweiten Analysen zu validieren wurde aus Klon 34 Zellen in biologischen Triplikaten Chromatin, jeweils im uninduzierten und induzierten Zustand, gewonnen. Damit wurden entsprechende ChIP Experimente mit un-spezifischen-, H3-, CTCF- und CTCFL-Antikörpern durchgeführt. Die resultierende ChIP-DNA wurde mittels qPCR analysiert und gegen Input DNA normalisiert.





**Abb. 24: CTCFL Bindung führt an untersuchten Kandidaten-Bereichen nicht zur Abnahme von CTCF Bindung.** Zellextrakt von uninduzierten und für 48 Stunden induzierten Klon 34-Zellen wurde mit Antikörpern gegen H3, CTCF, CTCFL oder zur Kontrolle mit unspezifischen IgG inkubiert. Die Auswertung erfolgte mittels qPCR mit Oligonukleotidprimern gegen die entsprechenden Regionen, welche jeweils nach dem nächstgelegenen Gen benannt und relativ zur eingesetzten Input-DNA dargestellt sind. Die Fehlerbalken stellen die Standardabweichung von drei unabhängigen Ansätzen dar.

Für alle drei untersuchten Bindestellen, ist die Bindung von CTCFL im induzierten Klon nachweisbar (Abb. 24). Zwar ist diese Bindung an der Col12a1 Stelle nur bei ca. 0,1% Input, jedoch stellt das eine klare Zunahme der CTCFL Bindung im Vergleich zum uninduzierten Klon dar. Bei den anderen beiden Bereichen ist diese Zunahme ebenfalls zu erkennen und die Bindung von CTCFL ist insgesamt stärker. Bei Betrachtung der CTCF Bindung im induzierten und uninduzierten Fall zeigt sich keine Abnahme der Bindung an einer der untersuchten Bindestellen. In allen drei Fällen nimmt die Bindung leicht (Accs) bis stärker (Col12a1) zu. Bei Auswertung der Ergebnisse der ChIPs mit Antikörper gegen Histon H3, ist auch jeweils eine leichte Zunahme der Bindung im induzierten Klon erkennbar. Eine Konkurrenz der CTCF Bindung durch Expression von CTCFL lässt sich somit an drei ausgesuchten Kandidaten-Bereichen nicht bestätigen. Um weiterhin zu untersuchen ob es Hinweise auf eine Konkurrenz der beiden Faktoren in genomweiten Analysen geben könnte wurden die ChIPseq Bindedaten der beiden Proteine bioinformatisch ausgewertet. Es wurden hierzu die Bindestellen in zwei Gruppen aufgeteilt, einmal die doppelt von CTCF und CTCFL gebunden Bereiche und einmal die Bindestellen die nur Bindung von CTCF aufweisen. Die Klassifizierung erfolgte aus dem Datensatz für den induzierten Klon. Dann wurden im Bereich +/- 500 bp um die Bindestellen die Reads für CTCF und CTCFL im uninduzierten als auch induzierten Zustand addiert und als Boxplots dargestellt.

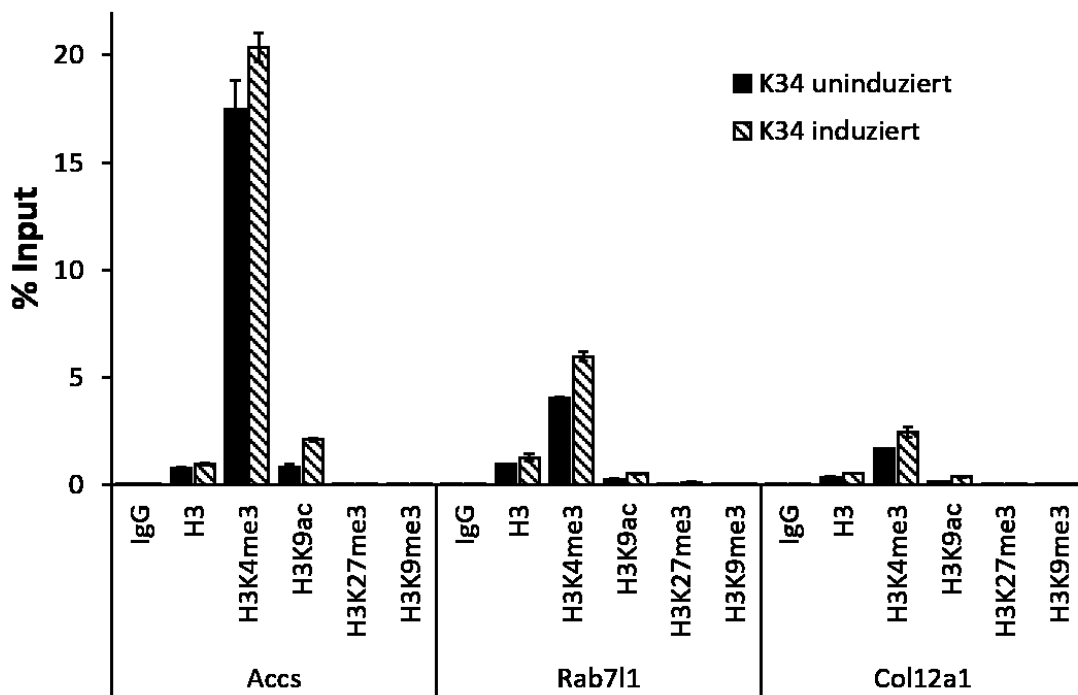


**Abb. 25: CTCFL Bindung zeigt keinen genomweiten Einfluss auf CTCF Anwesenheit an den geteilten Bindestellen.** Dargestellt ist die normalisierte CTCF Read-Anzahl der CTCF-ChIPseq Datensätze aus NIH3T3-Zellen und induziertem Klon 34 über die doppelt gebundenen Bindestellen (CTCF + CTCFL) als auch über den Anteil der nur CTCF gebundenen Bindestellen (CTCF alleine). In den Boxplots sind die CTCF Reads über alle Bindestellen +/- 500 bp der beiden Untergruppen zusammengefasst. In den Boxen sind jeweils die Datenbereiche von 25% bis 75% inklusive dem Median angezeigt, die Whisker geben dann Auskunft über die gesamte Spanne der Datenpunkte.

Der Teil der CTCF Bindestellen, der sich in induziertem Klon 34 durch zusätzliche CTCFL Bindung auszeichnete, zeigt bei genom-weiter Zusammenfassung der Daten keine Abnahme des CTCF Signals (Abb. 25). So ist der Median der CTCF Bindung in den Wildtyp-Zellen, die keine CTCFL-Expression und –Bindung aufweisen, als auch in den Klon 34 Zellen, identisch. Auch die weitere Verteilung der Daten ist für diese zwei untersuchten Gruppen sehr ähnlich. Bei Betrachtung der nur CTCF gebundenen Stellen, fällt die generelle niedrigere Bindung von CTCF an diesen Bereichen auf. Auch hier ist kein Unterschied zwischen der Bindung in Wildtyp-Zellen und den CTCFL-positiven Klon 34 Zellen sichtbar. Somit konnte zusammengefasst die in wenigen einzelnen Fällen, bei Auswertung der genomweiten Daten, detektierte Beeinflussung der CTCF Bindung durch CTCFL nicht validiert und nicht als genereller Effekt von CTCFL beschrieben werden.

### 4.3.2 CTCFL hat keinen klaren Einfluss auf die Anwesenheit bestimmter Histon-Modifikationen an seinen Bindestellen

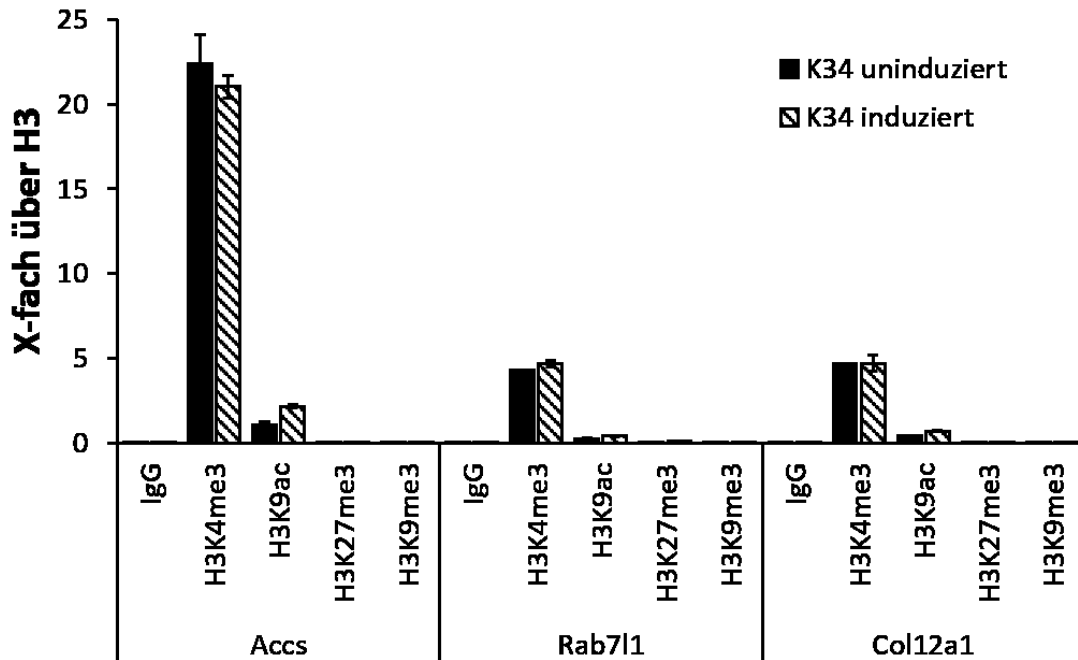
Eine weitere zu untersuchende Möglichkeit bestand in dem Einfluss der CTCFL Expression und Bindung auf Chromatin-bestimmende Histon-Modifikationen. Hierzu wurden ChIP-Experimente mit entsprechenden Antikörpern durchgeführt und mittels qPCR ausgewertet. Insgesamt wurden vier klassische Histon-Modifikationen für die Analyse ausgewählt, zwei die im Zusammenhang mit aktivem Chromatin stehen, H3K4me3 und H3K9ac, und zwei die charakteristisch für Hetero-Chromatin sind, H3K27me3 und H3K9me3. Bei den untersuchten Bindestellen handelte es sich um die bereits in den Kompletions-Analysen verwendeten Accs, Rab711 und Col12a1, die jeweils im Zusammenhang mit Aktivierung des entsprechenden Gens bei CTCFL Überexpression stehen.



**Abb. 26: Aktive Histon-Modifikationen zeigen leichte erhöhte Anwesenheit in Folge von CTCFL Expression an den untersuchten Bindestellen.** Zellextrakt von uninduzierten und für 48 Stunden induzierten Klon 34-Zellen wurde mit Antikörpern gegen H3, H3K4me3, H3K9ac, H3K27me3, H3K9me3 oder zur Kontrolle mit unspezifischen IgG inkubiert. Die Auswertung erfolgte mittels qPCR mit Oligonukleotidprimern gegen die entsprechenden Regionen, welche jeweils nach dem nächstgelegenen Gen benannt und relativ zur eingesetzten Input-DNA dargestellt sind. Die Fehlerbalken stellen die Standardabweichung von zwei unabhängigen Ansätzen dar.

Für die drei untersuchten Bindestellen kann für die beiden aktiven Histon-Modifikationen H3K4me3 und H3K9ac jeweils leichte Zunahmen der Bindung nach Expression von CTCFL beobachtet werden (Abb. 26). H3K4 Trimethylierung zeigt jeweils die höchste Anwesenheit, während die reprimierenden Modifikationen H3K27me3 und H3K9me3 an den Bindestellen überhaupt nicht nachgewiesen werden können. Da an allen Bindestellen

auch eine leichte Zunahme der Gesamtmenge an Histon H3 erkennbar ist wurde durch eine weitere Normalisierung untersucht, ob die Zunahme der aktiven Modifikationen, nur eine Folge der höheren Menge an zu modifizierendem Histon H3 darstellt.



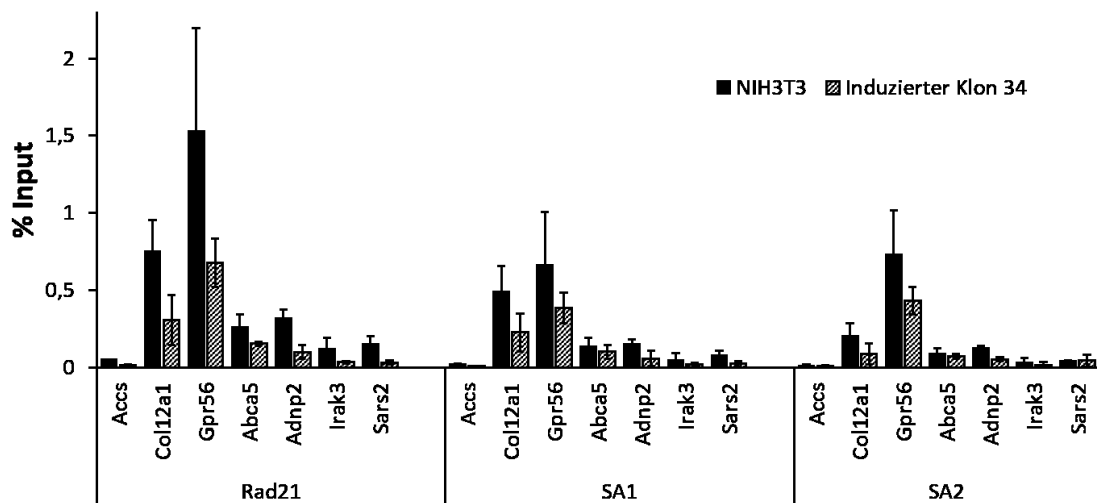
**Abb. 27:** Nach Normalisierung über die H3 Menge zeigen die aktiven Histon-Modifikationen fast keine Anreicherung an den untersuchten Bindestellen. Zellextrakt von uninduzierten und für 48 Stunden induzierten Klon 34-Zellen wurde mit Antikörpern gegen H3, H3K4me3, H3K9ac, H3K27me3, H3K9me3 oder zur Kontrolle mit unspezifischen IgG inkubiert. Die Auswertung erfolgte mittels qPCR mit Oligonukleotidprimern gegen die entsprechenden Regionen, welche jeweils nach dem nächstgelegenen Gen benannt und relativ zur detektierten Histon H3 Menge an den Bindestellen dargestellt sind. Die Fehlerbalken stellen die Standardabweichung von zwei unabhängigen Ansätzen dar.

In Folge der Normalisierung sind die leichten Zunahmen der aktiven Modifikationen nicht mehr erkennbar (Abb. 27). Einzig H3K9ac zeigt an der Bindestelle Accs eine höhere Bindung im induzierten Klon 34. H3K4me3 dagegen zeigt hier nun eine schwache Abnahme, die sich jedoch im Bereich des Fehlers befindet. Somit können keine klaren Effekte von CTCFL auf die vier untersuchten Histon-Modifikationen nachgewiesen werden.

#### 4.3.3 CTCFL zeigt keinen klaren Einfluss auf Cohesin-Bindung

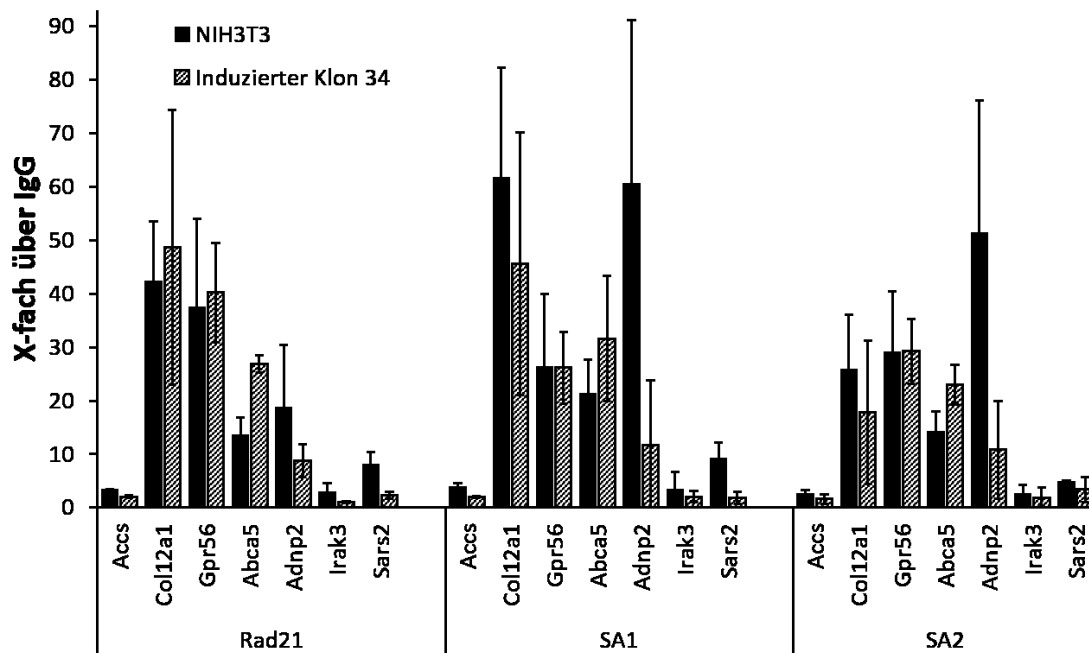
In abschließenden Korrelationsanalysen wurde untersucht, welchen Einfluss die Bindung von CTCFL auf Komponenten des Cohesin Komplexes hat. Bei dem Cohesin Komplex handelt es sich um einen sehr wichtigen CTCF Interaktionspartner. CTCF und Cohesin spielen zusammen die zentrale Rolle bei der Ausbildung von 3-dimensionalen DNA

Schleifen. Hierzu wurden mögliche Kandidaten-Bindestellen in ChIP-qPCR Experimenten untersucht. Neben den drei Bindestellen im Promotor der Gene Accs, Gpr56 und Col12a1, die bereits für die Kompetitions- als auch Histon-Modifikation-Experimente genutzt wurden, sollten weitere doppelt gebundene Bindestellen, unabhängig von beobachteter Konkurrenz in den ChIPseq Daten, analysiert werden. Somit wurden vier weitere Bindestellen ausgewählt die, analog zu vorherigem Vorgehen, nach ihrem nächstgelegenen Gen benannt wurden: Abca5, Adnp2, Irak3 und Sars2. Die entsprechenden ChIPs mit Antikörpern gegen Rad21, SA1, und SA2 wurden in biologischen Triplikaten in Wildtyp NIH3T3 Zellen und induziertem Klon 34 durchgeführt und die erhaltene ChIP-DNA mittels qPCR analysiert.



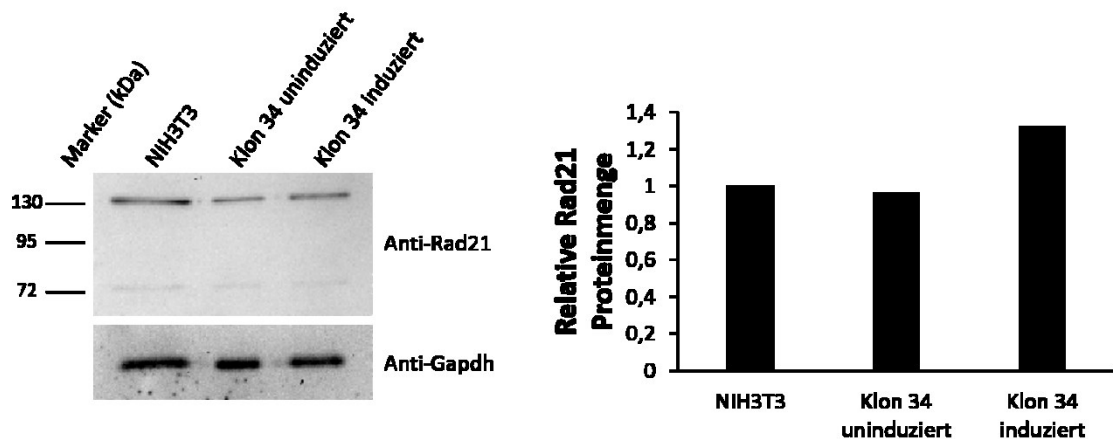
**Abb. 28: Komponenten des Cohesin-Komplexes zeigen reduzierte Präzipitation in induziertem Klon 34.** Zellextrakt von NIH3T3 Wildtyp-Zellen und für 48 Stunden inkubierten Klon 34-Zellen wurde mit Antikörpern gegen Rad21, SA1 oder SA2 inkubiert. Die Auswertung erfolgte mittels qPCR mit Oligonukleotidprimern gegen die entsprechenden Regionen, welche jeweils nach dem nächstgelegenen Gen benannt und relativ zur eingesetzten Input-DNA dargestellt sind. Die Fehlerbalken stellen die Standardabweichung von drei unabhängigen Ansätzen dar.

Die Präzipitation aller drei untersuchten Faktoren ist im induzierten Klon 34 im Vergleich zu den untransfizierten NIH3T3 Zellen reduziert (Abb. 28). Die einzelnen Bindestellen zeigen dabei unterschiedliche Stärken der Bindung, wobei Gpr56 jeweils am stärksten gebunden ist. Bei Gpr56 und Col12a1 ist für alle drei Faktoren klar schwächere Bindung im induzierten Klon 34 zu erkennen. Die Bindestelle Accs zeigt die schwächste, fast nicht detektierbare Bindung des Cohesin Komplexes. Da aber auch bei dieser sehr schwach gebundenen Stelle, zumindest für Rad21, eine niedrigere Präzipitation im Klon 34 ersichtlich ist, wurden die Daten zur Absicherung gegen ihre jeweilige unspezifische IgG-Kontrolle verrechnet. Als unspezifische IgG Kontrollen diente für Rad21 Rabbit IgG und für SA1 und SA2 Goat IgG.



**Abb. 29:** Nach Normalisierung zeigen viele Bindestellen kein differentielles Bindeverhalten für Komponenten des Cohesin Komplexes. Zellextrakt von NIH3T3 Wildtyp-Zellen und für 48 Stunden inkubierten Klon 34-Zellen wurde mit Antikörpern gegen Rad21, SA1 oder SA2 inkubiert. Die Auswertung erfolgte mittels qPCR mit Oligonukleotidprimern gegen die entsprechenden Regionen, welche jeweils nach dem nächstgelegenen Gen benannt und relativ zur jeweiligen Präzipitation von Rabbit IgG für RAD21, oder Goat IgG für SA1 und SA2 dargestellt sind. Die Fehlerbalken stellen die Standardabweichung von drei unabhängigen Ansätzen dar.

Nach der Normalisierung zeigen die meisten Bindestellen keine unterschiedliche Bindung der drei untersuchten Faktoren zwischen den beiden untersuchten Zelltypen (Abb. 29). Die vorher klar differentiell gebundenen Bindestellen Col12a1 und Gpr56 zeigen diese nun nicht mehr, und wenn sie doch schwach sichtbar ist, wie z.B. für SA2 über Col12a1, dann befindet sich diese Abnahme klar im Bereich der Fehlerbalken. An der Abca5 Stelle dreht sich der vorher betrachtete Trend sogar um und Klon 34 zeigt nun höhere Bindung von Rad21, SA1 und SA2. Nur für die Bindestelle Adnp2 ist weiterhin für alle drei Komponenten eine Abnahme der Bindung nach Expression von CTCFL erkennbar, wobei hier sehr große Fehlerbalken, aufgrund der niedrigen Bindung der Faktoren im Vergleich zur Input-DNA (Abb. 28), auftreten. Zur Kontrolle der jeweils vorliegenden Proteinmenge von Rad21, SA1 und SA2 in den untersuchten Zellen wurden Western Blot-Analysen mit Proteinextrakt der NIH3T3- und Klon 34-Zellen und den entsprechenden Antikörpern, die auch in den CHIP-Experimenten Verwendung fanden, durchgeführt. Mit den SA1 und SA2 Antikörpern schlugen jeweils alle Versuche den Western Blot zu etablieren fehl, somit konnten nur für Rad21 auswertbare Ergebnisse erzielt werden.



**Abb. 30: Rad21 zeigt gleiche Proteinmenge in den analysierten Zellen.** Ganzzell-Proteinextrakt aus NIH3T3 Zellen, uninduzierten- und für 48 Stunden induzierten Klon 34 Zellen wurde mittels RIPA-Puffer extrahiert. Es wurden jeweils 20  $\mu$ g Protein der einzelnen Proben elektrophoretisch in einem 10% SDS-PAGE-Trenngel aufgetrennt und mittels Wet-Blot auf eine PVDF-Membran übertragen. Die Immundetektion erfolgte mit Antikörpern gegen GAPDH, als Ladekontrolle, und unter Verwendung des Abcam Antikörpers gegen Rad21. Detektion der Antikörper-Reaktion erfolgte unter Verwendung von chemilumineszenten HRP Substrates. Zur Quantifizierung der Rad21 wurde das Programm Quantity-One des VersaDocks verwendet und jeweils die Rad21 Banden über das GAPDH Signal normalisiert

Unter Verwendung von GAPDH-spezifischem Antikörper zeigen alle untersuchten Proben vergleichbare Mengen an geladenem Protein (Abb. 30). Die Detektion mit Rad21 spezifischem Antikörper zeigt in allen Proben Expression von Rad21-Protein. Das spezifische Signal ist bei einer Molekülmasse von 130 kDa zu detektieren. Dies kann durch quantitative Auswertung des Rad21 Western Blots unter Verwendung der QuantityOne Software bestätigt werden. Für den uninduzierten Klon 34 ist die Rad21 Menge mit 0,96 fach fast identisch zu den untransfizierten NIH3T3 Zellen. Der induzierte Klon 34 zeigt sich leicht erhöht mit 1,32 facher Expression von Rad21 Protein. Starke generelle Einflüsse der Proteinmenge von Rad21 auf die DNA-Bindung des Cohesin-Komplexes können somit also ausgeschlossen werden.

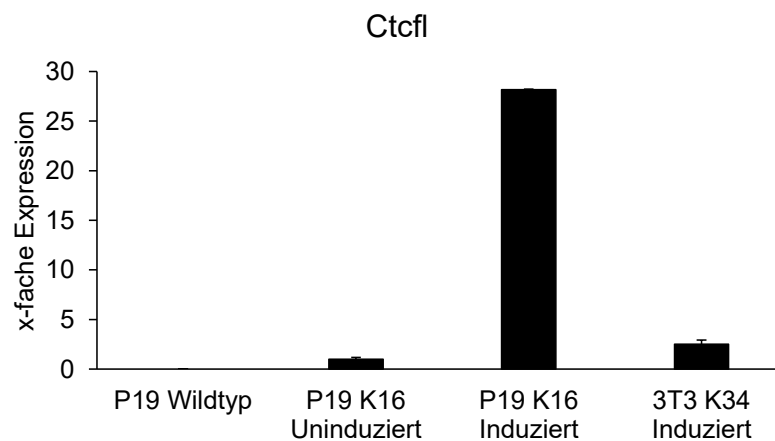
#### 4.4 Untersuchungen des Bindeverhaltens von CTCFL im murinen Differenzierungssystem

Im Rahmen der Experimente im NIH3T3-System konnten keine klaren Korrelationen oder funktionale Aussagen über das Bindeverhalten von CTCFL getroffen werden. Um der Frage nach Grund und Folge der Bindung nachzugehen sollte daher eine stabile, induzierbare CTCFL-Expression im differenzierbaren Zellsystem etabliert und untersucht werden. Als Zelllinie für diese Untersuchungen bot sich hier die murine Linie P19 an. Bei dieser handelt es sich um eine Teratokarzinom-Linie mit Eigenschaften embryonaler

Stammzellen (McBurney 1993). Analysiert werden sollte, ob die Differenzierung der Zellen und die damit einhergehende Änderung des Chromatin-Kontextes einen Einfluss auf die Bindung von CTCFL haben würde. Eine andere Möglichkeit die untersucht werden sollte ist der eventuelle Einfluss von CTCFL auf differenzierungsbedingte Änderungen am Chromatin und dem genomweiten Gen-Expressionsmuster der Zellen. Da im Rahmen dieser Fragestellung genomweite Bindemuster von CTCF und CTCFL erhoben wurden boten sich auch vergleichende Analysen zum NIH3T3-Klon 34 an, um Aufschluss über die Konservierung von CTCFL Bindevorgängen zu erhalten.

#### 4.4.1 Etablierung einer stabilen Expression von CTCFL in P19 Zellen

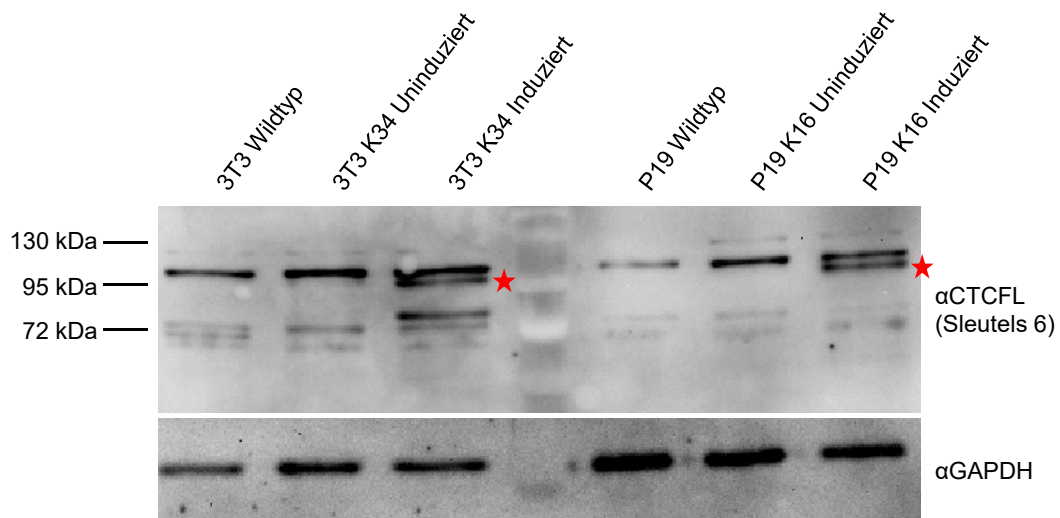
Die P19 Zellen wurden zur Herstellung stabiler Klone mit den Vektoren pBI-EGFP-FLAG-mCtcf1 & pTA-N unter Verwendung von jetPEI transfiziert. Nach 24 h erfolgte ein Mediumswechsel und nach 72 h wurde dem Medium Puromycin zur Selektion der transfizierten Zellen zugegeben. Nach einer weiteren Kultivierung von 2 Wochen mit Mediumswechsel alle zwei Tage, wurden Klone gepickt und gesondert weiter kultiviert. Nach entsprechendem Wachstum wurden die Klone sowohl eingefroren als auch zur Überprüfung induziert. Induktion erfolgte für 48 h durch ausbleibende Zugabe von Doxzyklin zum Nährmedium. Aus den Klonen wurde dann jeweils RNA isoliert, in cDNA umgeschrieben und mittels RT-qPCR auf CTCFL Expression überprüft.



**Abb. 31: P19 Klon 16 zeigt starke induzierbare Expression von CTCFL.** P19 Zellen wurden mit pBI-EGFP-FLAG-mCTCFL und pTA-N transfiziert und mit Puromycin selektiert. Es wurde sowohl von P19 Wildtyp-Zellen, als auch uninduziertem und für 48 h induziertem Klon 16 RNA isoliert und in cDNA umgeschrieben. Als Positivkontrolle diente cDNA von induziertem NIH3T3 Klon 34. Die Auswertung erfolgte mittels qRT-PCR unter Verwendung spezifischer exonübergreifender Oligonukleotidprimer gegen Ctcf1. Die Expression wurde gegen das Housekeeping-Gen  $\beta$ -Aktin normalisiert und dann der Expressionswert von uninduziertem Klon 19 als 1 gesetzt.



Die Auswertung der RT-qPCR zeigt induzierbare Expression von CTCFL im Klon 16 (Abb. 31). Die Expression wird durch die Induktion ca. 25-fach verstärkt. Bei Vergleich mit induziertem Klon 34 zeigt induzierter Klon 16 einen um ca. 11-fach höheren Expressionswert. Der uninduzierte Klon 16 zeigt ebenfalls eine klar detektierbare Expression von CTCFL, die etwa bei der Hälfte vom induzierten Klon 34 liegt. Die induzierbare CTCFL Expression sollte dann ebenso auf Proteinebene überprüft werden um sicherzustellen, dass funktionales Protein von richtiger Länge gebildet wird. Hierzu wurde ein Western-Blot mit entsprechenden Antikörpern und NIH3T3 Klon 34 als Positivkontrolle durchgeführt.



**Abb. 32: Klon 16 zeigt induzierbare Expression von CTCFL auf Proteinebene.** Ganzzell-Proteinextrakt aus NIH3T3 Zellen, uninduzierten- und für 48h induzierten Klon 34 Zellen, P19 Zellen, uninduzierten und für 48h induzierten Klon 16 Zellen wurde mittels RIPA-Puffer extrahiert. Es wurden jeweils 20 µg Protein der einzelnen Proben elektrophoretisch in einem 10% SDS-PAGE-Trenngel aufgetrennt und mittels Wet-Blot auf eine PVDF-Membran übertragen. Die Immundetektion erfolgte mit Antikörpern gegen GAPDH, als Ladekontrolle, und unter Verwendung des sog. Sleutels6 Antikörpers gegen CTCFL. Detektion der Antikörper-Reaktion erfolgte unter Verwendung von chemilumineszenten HRP Substrates.

Im Western Blot zeigt sich unter Verwendung von GAPDH-spezifischem Antikörper, dass in allen Proben Protein vorliegt und starke Abweichungen in den Proteinmengen sind nicht erkennbar. Detektion mit CTCFL-spezifischem Antikörper zeigt Expression von CTCFL in den induzierten Klonen auf Höhe von ca. 95 kDa (Abb. 32, rote Sterne). In den uninduzierten Klonen ist diese Expression nicht erkennbar. Zusätzliche unspezifische Signale treten in allen geladenen Proben auf. Somit kann die induzierbare Expression von CTCFL im Klon 16 auf RNA- und Protein-Ebene belegt werden und der Klon für weitergehende Experimente genutzt werden.

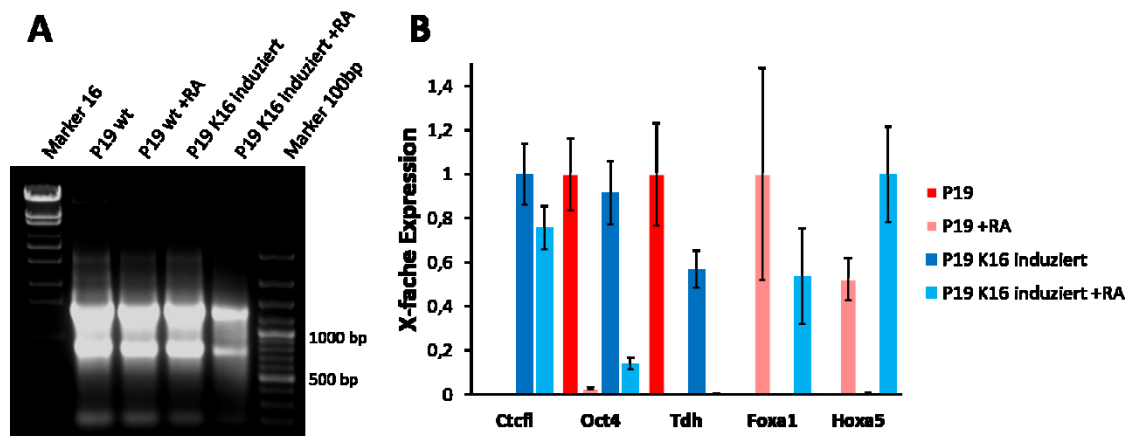
#### 4.4.2 Einfluss von CTCFL auf die genomweite Expression von Genen in Abhängigkeit von Differenzierung (Microarray)

Zur Untersuchung des Einflusses der CTCFL Expression auf die genomweiten Genexpressionsmuster wurden *Microarray* Experimente durchgeführt. Um den generellen Einfluss von CTCFL, als auch den speziellen Einfluss auf differenzierungsspezifische Expressionsmuster, zu untersuchen wurden insgesamt vier unterschiedliche Ansätze analysiert (Tab. 2).

**Tab. 2: Auflistung der bezüglich genomweiter Expressionsmuster analysierten Proben.** Es wurde jeweils RNA von undifferenzierten und differenzierte P19 Wildtyp Zellen als auch von undifferenzierten und differenzierten, induzierten Klon 16 Zellen. Die Differenzierung erfolgte durch Retinsäure Zugabe zum Nährmedium für 48 h.

Nummer	Analysierte Probe
1	P19 Wildtyp
2	P19 Wildtyp differenziert (2 Tage RA)
3	P19 K16 induziert
4	P19 K16 induziert und differenziert (2 Tage RA)

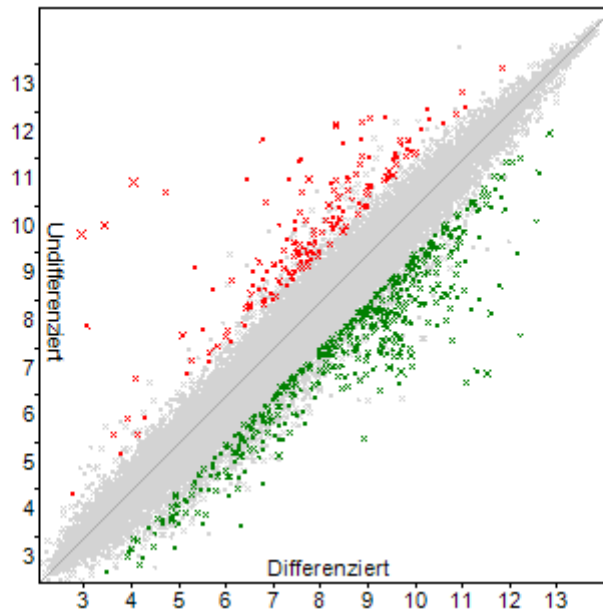
Da für *Microarray*-Genexpressionsanalysen hochwertige RNA benötigt wird wurden vor der eigentlichen Durchführung entsprechende Kontroll-Experimente durchgeführt. So wurde je ein kleiner Teil der isolierten RNA in einem Agarose Gel aufgetrennt um die 18S und 28S RNA sichtbar zu machen. Gute nutzbare RNA zeichnet sich durch klare Sichtbarkeit der beiden ribosomalen RNA Moleküle aus und das ist für die untersuchten Proben jeweils der Fall (Abb. 33, A).



**Abb. 33: Die für die Genexpressionsanalyse isolierte RNA erfüllt die qualitativen Anforderungen.** (A) Je 1  $\mu$ l der isolierten RNA wurde mit 4  $\mu$ l H<sub>2</sub>O und 1  $\mu$ l 6x Ladepuffer in einem 1% Agarose-Gel aufgetragen und durch Verwendung von EtBr unter UV-Licht visualisiert. (B) P19 Wildtyp Zellen und induzierte Klon 16 Zellen wurden mit und ohne RA Einfluss kultiviert und jeweils RNA isoliert. Je 1  $\mu$ g der isolierten RNA wurde in cDNA umgeschrieben. Die Auswertung erfolgte mittels qRT-PCR unter Verwendung spezifischer exonübergreifender Oligonukleotidprimer gegen Ctcfl, Oct4, Tdh, Foxa1 und Hoxa5. Die Expression wurde gegen das Housekeeping-Gen  $\beta$ -Aktin normalisiert und dann jeweils der höchste Expressionswert der 4 Proben als 1 gesetzt.

Neben der generellen RNA-Qualität sollte vor Durchführung des genomweiten Experimentes auch die Induktion von CTCFL im Klon 16 und die erfolgreiche Retinsäure-vermittelte Differenzierung überprüft werden. Hierzu wurde ein kleiner Teil der isolierten RNA in cDNA umgeschrieben und mittels qRT-PCR überprüft. Sowohl die CTCFL Expression in den Klon 16 Zellen im Vergleich zum Wildtyp als auch die differentielle Expression von Marker-Genen für Differenzierung kann anhand der isolierten RNA bestätigt werden (Abb. 33, B). So ist CTCFL nur in den Proben des induzierten Klon 16 detektierbar und die zwei Gene Oct4 und Tdh, die charakteristisch für undifferenzierte Zellen sind, jeweils hohe Expression nur in den nicht RA-behandelten Proben. Umgekehrt zeigen die Gene Foxa1 und Hoxa5, die normalerweise in differenzierten Zellen exprimiert sind, Expression in genau diesen RA-behandelten und damit differenzierten Zellen. Somit war die RNA bereit für die Verwendung im Microarray Experiment. Die Genexpressionsanalyse selbst wurde anschließend von T. Böttger (MPI Bad Nauheim) auf der Affymetrix Gene 2.0 ST Plattform durchgeführt. Nach Erhalt der Rohdaten wurden die weiteren Analysen mit der jeweiligen von Affymetrix angebotenen Software (Expression Console & Transcriptome Analysis Console) durchgeführt. In der initialen Analyse sollten genomweit die in Folge der Differenzierung deregulierten Gene erfasst werden. Dazu wurden die beiden undifferenzierten Datensätze (P19 Wildtyp & P19 K16 induziert) mit den beiden differenzierten (P19 Wildtyp +RA & P19 K16 +RA) paarweise verglichen.

Ein dereguliertes Gen wurde in dieser Analyse nur erfasst wenn es sich in der Expression zu den jeweils beiden anderen Datensätzen unterschied. Zeigte beispielsweise nur einer der differenzierten Datensätze eine Zu-/Abnahme eines Genes, der andere jedoch nicht, so ging dieses Gen nicht in die Ergebnisse ein.



**Abb. 34: 623 Gene zeigen Deregulation in P19 Zellen im Zuge der RA-vermittelten Differenzierung.** Dargestellt ist eine mit der Transcriptome Analysis Console produzierte Darstellung der Expressionswerte aller Gene in Gegenüberstellung der undifferenzierten und differenzierten Datensätze. Statistisch signifikant hochregulierte Gene sind als grüne, signifikant herab regulierte Gene als rote und nicht dereguliert oder nicht signifikant deregulierte Gene als graue Datenpunkte dargestellt.

Mit dieser Analyse wurden 623 deregulierte Gene, mit den Charakteristika mindestens zweifach dereguliert und einem p-Wert kleiner als 0,05, identifiziert (Abb. 34). Von den 623 deregulierten Genen sind 458 (Top 20 in Tab. 3) hochreguliert und 174 (Top 20 in Tab. 4) herunterreguliert.

**Tab. 3: Die 20 am stärksten induzierten Gene nach Differenzierung der P19 Zellen durch Zugabe von Retinsäure.** Neben dem offiziellen Gensymbol sind der vollständige Name des Genes, die Stärke der Deregulation und der entsprechende p-Wert angegeben.

Gen	Vollständiger Name	x-fache Deregulation	p-Wert
Tshz1	teashirt zinc finger family member 1	33,26	0,00052
Hoxa1	homeobox A1	31,09	0,018592
Hoxa5	homeobox A5	28,04	0,004093
Meis1	Meis homeobox 1	27,97	0,004004
Hoxb1	homeobox B1	14,51	0,003686
Hoxc5	homeobox C5	14,15	0,032364
Ednrb	endothelin receptor type B	12,41	0,006005
Dhrs3	dehydrogenase/reductase (SDR family) member 3	11	0,005

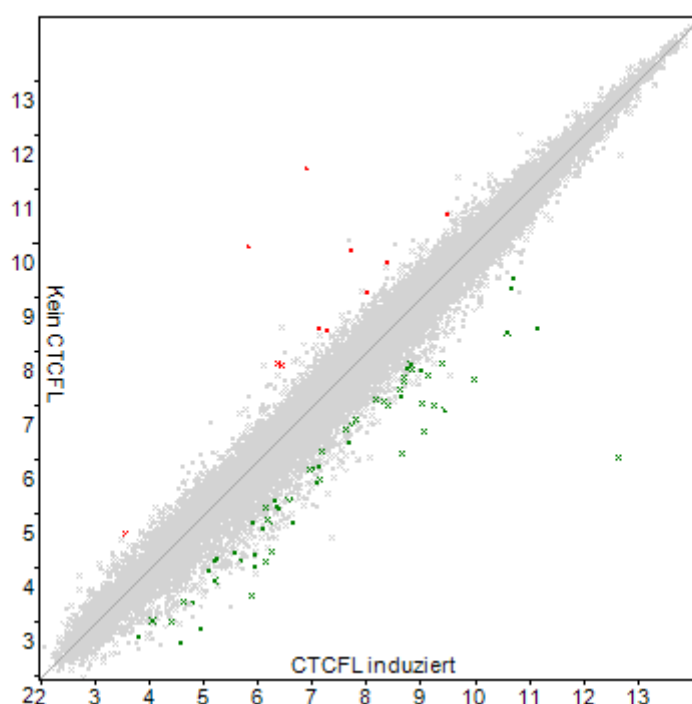
Rasgrp3	RAS, guanyl releasing protein 3	10,67	0,013889
Npnt	nephronectin	10,21	0,008832
Rarb	retinoic acid receptor, beta	10,11	0,002493
Plxna2	plexin A2	10,04	0,004698
Crabp1	cellular retinoic acid binding protein I	8,91	0,040855
Gata6	GATA binding protein 6	8,47	0,044787
Dach1	dachshund 1 (Drosophila)	8,36	0,0008
Mir10a, Hoxb4	microRNA 10a, homeobox B4	8,34	0,011107
Foxa1	forkhead box A1	7,92	0,020192
Cyp26b1	cytochrome P450, family 26, subfamily b, polypeptide 1	7,55	0,014164
Hoxa2	homeobox A2	7,53	0,007665
Flrt3	fibronectin leucine rich transmembrane protein 3	7,43	0,000745

**Tab. 4: Die 20 am stärksten herunterregulierten Gene nach Differenzierung der P19 Zellen durch Zugabe von Retinsäure.** Neben dem offiziellen Gensymbol sind der vollständige Name des Genes, die Stärke der Deregulation und der entsprechende p-Wert angegeben.

Gen	Vollständiger Name	x-fache Deregulation	p-Wert
Mir302a	microRNA 302a	-89,31	0,000072
Mir302c	microRNA 302c	-88,84	0,000026
Mir367	microRNA 367	-71,16	0,00047
Mir302d	microRNA 302d	-47,27	0,000771
Enpp2	ectonucleotide pyrophosphatase/phosphodiesterase 2	-24,69	0,005634
Mir302b	microRNA 302b	-21,38	0,00855
Tdglf1	teratocarcinoma-derived growth factor 1	-17,52	0,039555
Otx2	orthodenticle homolog 2 (Drosophila)	-10,71	0,022051
Utf1	undifferentiated embryonic cell transcription factor 1	-10,65	0,005225
Foxi3	forkhead box I3	-10,6	0,007122
Mir290	microRNA 290	-10,23	0,011423
Pim2	proviral integration site 2	-10,19	0,002514
Tdh	L-threonine dehydrogenase	-9,44	0,022433
Efh1	EF hand domain containing 1	-9,43	0,016778
Cyp2b23	cytochrome P450, family 2, subfamily b, polypeptide 23	-7,47	0,012784
Slc7a7	solute carrier family 7 (cationic amino acid transporter, y <sup>+</sup> system), member 7	-7,24	0,007328
Dnmt3b	DNA methyltransferase 3B	-7,04	0,000812
Pdzd4	PDZ domain containing 4	-6,99	0,000594
Epha1	Eph receptor A1	-6,48	0,007141
Fst	follistatin	-5,91	0,012438

In den Ergebnislisten können bereits bekannte Gene identifiziert werden, die vorher zur Kontrolle der Differenzierung genutzt wurden. So ist beispielsweise Hoxa5 das am dritt stärksten induzierte Gen nach Differenzierung. Auch bei den herunterregulierten Genen tauchten bekannte Marker für Pluripotenz auf, so zum Beispiel die microRNA Familie 302. Somit ist die erfolgreiche Differenzierung der entsprechenden Zellen abgesichert

und es konnte der Einfluss von CTCFL im P19-System untersucht werden. Hierzu wurden analog zur vorherigen Analyse je zwei Datensätze zusammengefasst und die gemeinsamen deregulierten Gene mittels der *Transcriptome Analysis Console* identifiziert. Zum einen die CTCFL-negativen Proben P19 Wildtyp und P19 Wildtyp +RA zum anderen P19 K16 induziert und P19 K16 induziert +RA. Im Zuge dieser Analyse können nur 82 deregulierte Gene identifiziert werden (Abb. 35). Auch hier ist die Mindestanforderung eine Deregulation größer zweifach und ein p-Wert von kleiner 0,05.



**Abb. 35: 82 Gene zeigen differenzierungsunabhängige Deregulation in P19 Zellen nach Induktion von CTCFL.** Dargestellt ist eine mit der Transcriptome Analysis Console produzierte Darstellung der Expressionswerte aller Gene in Gegenüberstellung der CTCFL-negativen und CTCFL-positiven Datensätze. Statistisch signifikant hochregulierte Gene sind als grüne, signifikant herab regulierte Gene als rote und nicht deregulierte oder nicht signifikant deregulierte Gene als graue Datenpunkte dargestellt.

Nach Induktion von CTCFL ist mit 71 Genen (Top 20 in Tab. 5) der Hauptteil der deregulierten Gene induziert. Nur 11 Gene (Tab. 6) zeigen reduzierte Expression im vorliegenden Datensatzvergleich.

**Tab. 5: Die 20 am stärksten hochregulierten Gene nach CTCFL Induktion in P19 Zellen.** Neben dem offiziellen Gensymbol sind der vollständige Name des Gens, die Stärke der Deregulation und der entsprechende p-Wert angegeben.

Gen	Vollständiger Name	x-fache Deregulation	p-Wert
Grb10	growth factor receptor bound protein 10	6,56	0,0129
L1Md-Gf21, Tex16	L1 repeat, Gf subfamily, member 21, testis expressed gene 16	5,83	0,0478

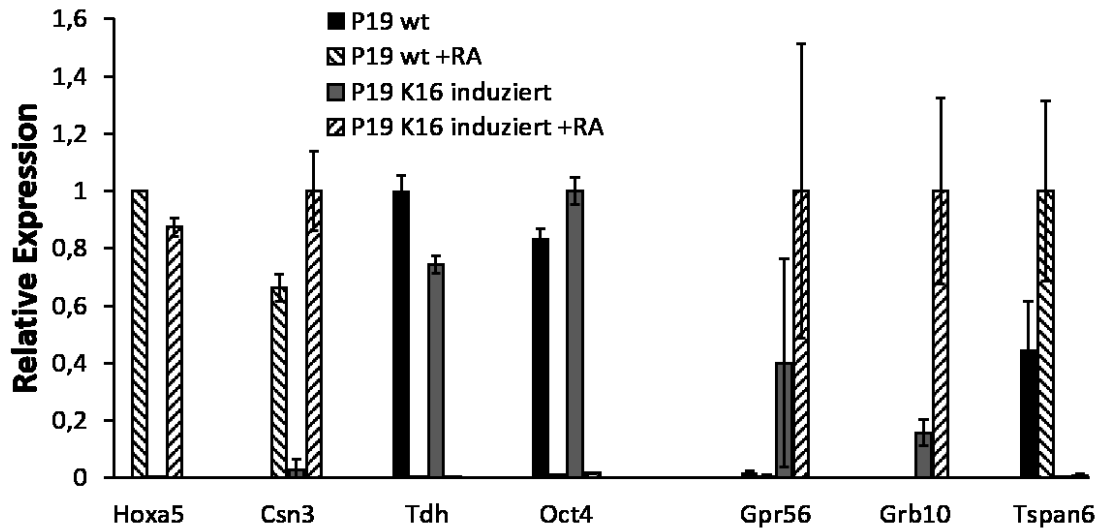
Ctcf1	CCCTC-binding factor (zinc finger protein)-like	5,6	0,0475
Gpr56	G protein-coupled receptor 56	4,78	0,0062
Gabra5	gamma-aminobutyric acid (GABA) A receptor, subunit alpha 5	4,71	0,0216
Zfp640	zinc finger protein 640	4,12	0,038
Gm9805	predicted gene 9805	3,99	0,0496
Tex16	testis expressed gene 16	3,97	0,0309
Zewpw1	zinc finger, CW type with PWWP domain 1	3,07	0,0292
Ctsf	cathepsin F	2,95	0,0142
Nodal	nodal	2,78	0,0459
Mir669k	microRNA 669k	2,58	0,0346
Mir669k	microRNA 669k	2,58	0,0346
Rpp25	ribonuclease P 25 subunit (human)	2,54	0,0227
Pdgfd	platelet-derived growth factor, D polypeptide	2,51	0,0223
Gm7173	predicted gene 7173	2,45	0,0007
Kbtbd7	kelch repeat and BTB (POZ) domain containing 7	2,35	0,0076
9230114K14Rik	RIKEN cDNA 9230114K14 gene	2,27	0,0415
Slc46a1	solute carrier family 46, member 1	2,23	0,0115
Fzd10	frizzled homolog 10 (Drosophila)	2,22	0,0218

**Tab. 6: Die 11 herunterregulierten Gene nach CTCFL Induktion in P19 Zellen.** Neben dem offiziellen Gensymbol sind der vollständige Name des Genes, die Stärke der Deregulation und der entsprechende p-Wert angegeben.

Gen	Vollständiger Name	x-fache Deregulation	p-Wert
Tspan6	tetraspanin 6	-22,75	0,0109
Cftr	cystic fibrosis transmembrane conductance regulator	-17,58	0,0151
Jmjd8	jumonji domain containing 8	-4,44	0,0082
Pgcp	plasma glutamate carboxypeptidase	-2,7	0,0422
C330024D21Rik	RIKEN cDNA C330024D21 gene	-2,49	0,0183
Maob	monoamine oxidase B	-2,46	0,0085
Rab6b	RAB6B, member RAS oncogene family	-2,45	0,02
Timp1	tissue inhibitor of metalloproteinase 1	-2,18	0,0285
nA	nicht angegeben	-2,13	0,0061
Sumf2	sulfatase modifying factor 2	-2,11	0,0121
Ppap2c	phosphatidic acid phosphatase type 2C	-2,05	0,0359

Einige der identifizierten Gene wurden anschließend zur Validierung der genom-weiten Analyse ausgewählt. Dies waren, zur Überprüfung der abnehmenden Expression nach Differenzierung, die Gene Tdh und Oct4 (auch Pou5f1), zur Überprüfung der Zunahme, die Gene Hoxa5 und Csn3. Einige dieser Gene waren bereits von der Initial-Kontrolle der RNA vor dem Microarray-Experiment bekannt (Abb. 33), sollten nun jedoch in einer un-

abhängigen Wiederholung des Experimentes in vier biologischen Replikaten erneut überprüft werden. Zur Validierung des CTCFL-Einflusses wurden die induzierten Gpr56 und Grb10, sowie das reprimierte Gen Tspan6 gewählt. Diese Gene wurden ebenfalls mittels RT-qPCR in jeweils vier biologischen Replikaten analysiert.



**Abb. 36: Die Deregulation der gewählten Gene konnte mittels qRT-PCR bestätigt werden.** Die RNA von P19 Wildtyp Zellen, P19 Wildtyp Zellen die 2 Tage differenziert wurden, induzierten Klon 16 Zellen und induzierten Klon 16 Zellen die für 2 Tage differenziert wurden, wurde isoliert und in cDNA umgeschrieben. Die Auswertung erfolgte mittels qRT-PCR unter Verwendung spezifischer exonübergreifender Oligonukleotidprimer gegen Hoxa5, Csn3, Tdh, Oct4, Gpr56, Grb10 und Tspan6. Die Expression wurde gegen das Housekeeping-Gen  $\beta$ -Aktin normalisiert und dann für jedes Gen jeweils der höchste Expressionswert der vier cDNAs als 1 gesetzt. Die Fehlerbalken stellen die Standardabweichung von vier unabhängigen Ansätzen dar.

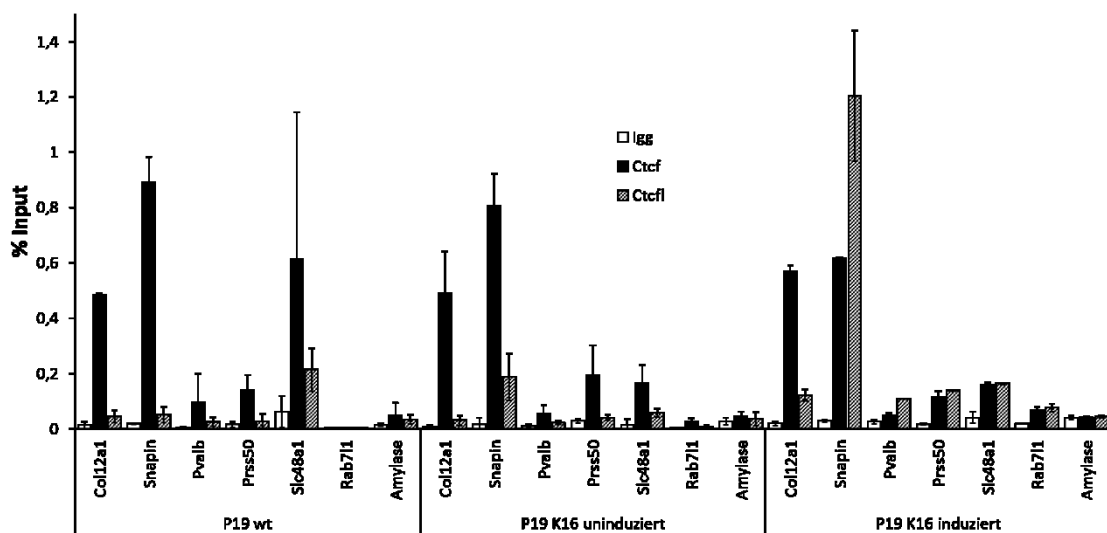
Die Auswertung mittels qRT-PCR unter Verwendung der entsprechenden spezifischen Oligonukleotidprimern zeigt für die Pluripotenz-assoziierten Gene Oct4 und Tdh nur Expression in den Zellen ohne RA Zugabe, für die Differenzierungsgene Hoxa5 und Csn3 ist genau das reziproke Muster an Expression zu beobachten. Diese beiden Gen-Paare sind somit in keiner der Zellen gleichzeitig exprimiert. Bei Betrachtung der CTCFL beeinflussten Gene zeigen Gpr56 und Grb10 jeweils Expression in den Klon 16 Zellen, wobei die Expression für Beide in den zusätzlich RA-behandelten Zellen höher als in den unbehandelten Zellen ist. Tspan6, das als einziges der 11 durch CTCFL Expression reprimierten Gene untersucht wurde, zeigt in den Wildtyp Zellen Expression, in den Klon 16 Zellen dagegen ist überhaupt keine Expression nachweisbar. Auch Tspan6 zeigt in den RA-behandelten Zellen eine etwas höhere Expression als in den undifferenzierten Wildtyp Zellen. Für alle drei der CTCFL-beeinflussten Gene ist jedoch auch die Varianz innerhalb der Replikate eines Zelltyps größer als bei den untersuchten Differenzierungsgenen, was sich in den um einiges größeren Standardabweichungen darstellt. Für alle sieben



untersuchten Gene können somit aber die Ergebnisse der genomweiten Microarray-Experimente und Auswertungen bestätigt werden.

#### 4.4.3 Identifizierung der genomweiten Bindestellen von CTCFL im Differenzierungssystem (ChIPseq)

Nach Identifizierung der deregulierten Gene im P19 System sollten darauffolgend Analysen zur DNA Bindung von CTCFL durchgeführt werden. Ziel war der Erhalt eines genomweiten Bindemusters für CTCFL als auch CTCF und die eventuelle Identifikation eines Einflusses der RA-vermittelten Differenzierung auf diese Muster. Bevor genomweite Analysen durchgeführt werden konnten, mussten die entsprechenden experimentellen Bedingungen der Chromatin-Immunpräzipitation für die P19 Zellen angepasst werden und positive Kontroll-Bindestellen von CTCFL identifiziert werden. Anhand des bereits bekannten Bindepfiles von CTCFL in NIH3T3 Zellen wurden sechs mögliche Bindestellen ausgewählt. Neben den neuen Bindestellen Snapin, Pvalb, Prss50, Slc48a1 sollten auch die bereits in den Kompetitions-Experimenten untersuchten Bindestellen Col12a1 und Rab711 (4.3) bezüglich der CTCFL Bindung im P19 Klon 16 überprüft werden. Dazu wurden, nach Optimierung der Methode, ChIPs mit den Antikörpern Sleutels6 (gegen CTCFL) und N2.2 (gegen CTCF) durchgeführt.



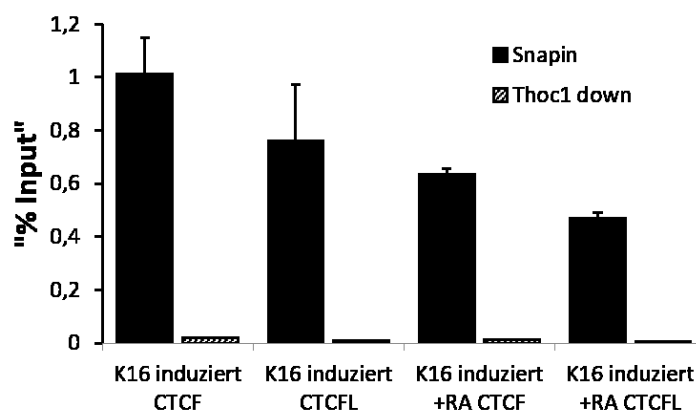
**Abb. 37: Die Bindestelle Snapin zeigt starke CTCFL Bindung in induzierten Klon 16 Zellen.** Zellextrakt von P19 Wildtyp-Zellen, uninduzierten P19 Klon 16 Zellen und für 48 Stunden inkubierten Klon 16 Zellen wurde mit Antikörpern gegen CTCF, CTCFL oder zur Kontrolle mit unspezifischen IgG inkubiert. Die Auswertung erfolgte mittels qPCR mit Oligonukleotidprimern gegen die entsprechenden Regionen, welche jeweils nach dem nächstgelegenen Gen benannt und relativ zur eingesetzten Input-DNA dargestellt sind. Als für beide Faktoren ungebundene Kontrolle wurde eine Bindestelle im genomischen Bereich des Amylase Gens verwendet. Die Fehlerbalken stellen die Standardabweichung von zwei unabhängigen Ansätzen dar.

Nach Auswertung der ChIP-DNA mittels qPCR kann die Snapin Bindestelle als CTCFL gebunden identifiziert werden (Abb. 37). So ist im induzierten Klon 16 hier eine Bindung von über 1% Input detektierbar, während diese im uninduzierten Klon bei nur ca. 0,2% liegt. Die Wildtyp Zellen zeigen jeweils nur Bindung von CTCFL im Bereich der Negativkontrolle Amylase. Einzige Ausnahme stellt hier die Slc48a1 Bindestelle dar, die eine Bindung von ca. 0,2% zeigt. Diese Bindestelle zeichnet sich jedoch durch schlechte Amplifikation in der qPCR aus was die Varianz der PCR Signale stark erhöht. In den beiden anderen Chromatin-Proben der Klone ist Slc48a1 nicht (K16 uninduziert), oder nur schwach (K16 induziert) von CTCFL gebunden. Im induzierten Klon 16 kann weiterhin schwache Bindung von CTCFL an die Bindestellen Col12a1, Pvalb und Prss50 detektiert werden. Die Bereiche Col12a1 und Snapin sind in allen drei untersuchten Proben jeweils im Bereich von 0,5% bis 1% gebunden und werden durch die Bindung von CTCFL augenscheinlich nicht beeinträchtigt. Anhand der Identifikation von Snapin als robuster CTCFL Bindestelle im P19 System konnten darauffolgende ChIPs nun leicht bezüglich ihrer Effizienz der Präzipitation analysiert und verglichen werden. Zur genomweiten Bestimmung der Bindestellen von CTCFL im P19 System sollte dann eine ChIP-Sequenzierung durchgeführt werden. Dafür wurde eine entsprechende Zahl an individuellen ChIPs durchgeführt und die erhaltene DNA gesammelt um 10 ng, die benötigte Menge für die Library-Präparation, zu erhalten. Um einen eventuellen Einfluss der Differenzierung auf das Bindeverhalten von CTCF und CTCFL zu identifizieren wurde auch ChIP DNA von differenziertem Klon 16 Zellen gewonnen und sequenziert. Diese Zellen wurden für zwei Tage unter Verwendung von Retinsäure differenziert und dann für zwei weitere Tage induziert um CTCFL zu produzieren. Zur späteren Kontrolle und Verrechnung der ChIP-seq Daten wurde auch jeweils eine Probe der Input-DNA vom induzierten, undifferenzierten und vom induzierten, differenzierten Klon sequenziert. Vor Versand der DNA-Proben wurden jeweils qualitative und quantitative Analysen des Materials durchgeführt, damit sollte einerseits sichergestellt werden, dass mindestens eine 10-fache Anreicherung einer Bindestelle gegenüber einer nicht-Bindestellen vorlag, andererseits, dass die benötigte minimale DNA-Menge von 10 ng vorlag, die für den Sequenzierungsablauf nötig war. Die genaue Quantifizierung der DNA Mengen erfolgte mittels einem Qubit (ThermoFisher Scientific) Fluorometer und für alle Proben kann eine ausreichende Menge an DNA, von mehr als 10 ng, nachgewiesen werden (Tab. 7).

**Tab. 7: Auflistung der sequenzierten Proben und jeweiligen DNA Mengen.** Chromatin Immunopräzipitationen erfolgten entweder mit spezifischem Antikörper N2.2 gegen CTCF oder Sleutels6 gegen CTCFL. Alle IPs mit einem Antikörper wurden vereinigt und das Volumen dann mittels SpeedVac auf 40  $\mu$ l reduziert. Vor Versand der DNA erfolgte die Quantifizierung unter Verwendung eines Qubit Flurometers

Nummer	Analysierte Probe	DNA Menge (ng)
1	P19 Klon 16 induziert – Input-DNA	23,76
2	P19 Klon 16 induziert – CTCF IP	54,00
3	P19 Klon 16 induziert – CTCFL IP	31,12
4	P19 Klon 16 induziert & differenziert – Input-DNA	25,68
5	P19 Klon 16 induziert & differenziert – CTCF IP	44,40
6	P19 Klon 16 induziert & differenziert – CTCFL IP	23,28

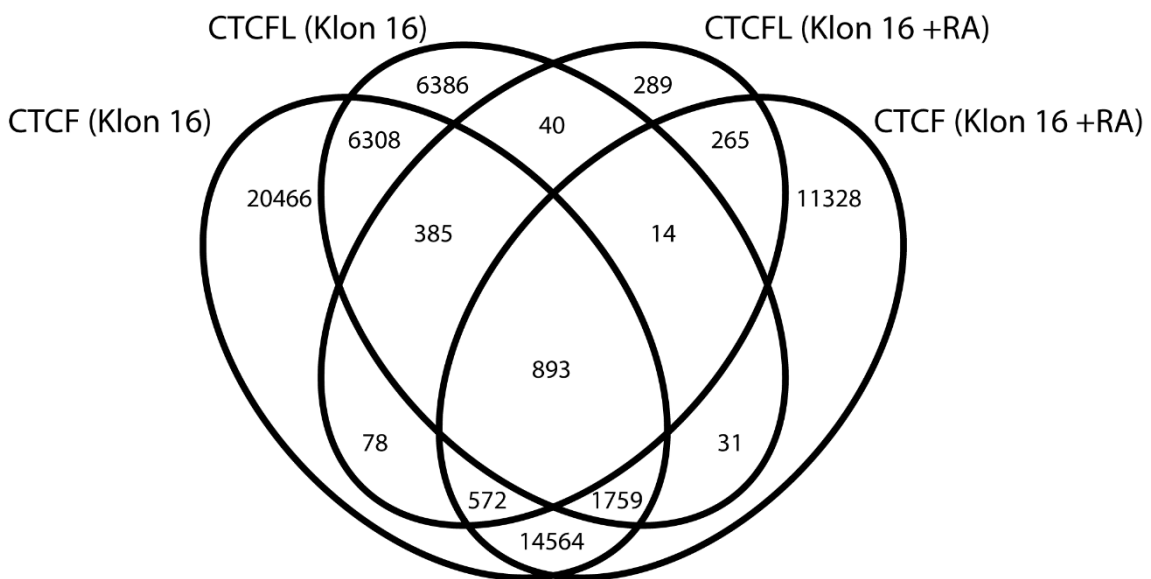
Zur Kontrolle der Anreicherung der Bindestellen wurde die zu sequenzierende DNA mittels qPCR überprüft. Die hierzu ausgewählten genomischen Bereiche waren einerseits Snapin als positive Bindestelle und andererseits Thoc1 down als Bereich für den keine Bindung von CTCF oder CTCFL bekannt war.



**Abb. 38: CTCF und CTCFL sind in den ChIPseq-Proben an der Snapin Bindestelle stark angereichert.** Chromatin Immunopräzipitationen erfolgten entweder mit spezifischem Antikörper N2.2 gegen CTCF oder Sleutels6 gegen CTCFL. Die Auswertung erfolgte mittels qPCR mit Oligonukleotidprimern gegen die entsprechenden Regionen, welche jeweils nach dem nächstgelegenen Gen benannt und relativ zum qPCR Signal der Klon 16 induziert CTCF Probe dargestellt sind. Die Fehlerbalken stellen die Standardabweichung von je zwei pPCR Ansätzen dar.

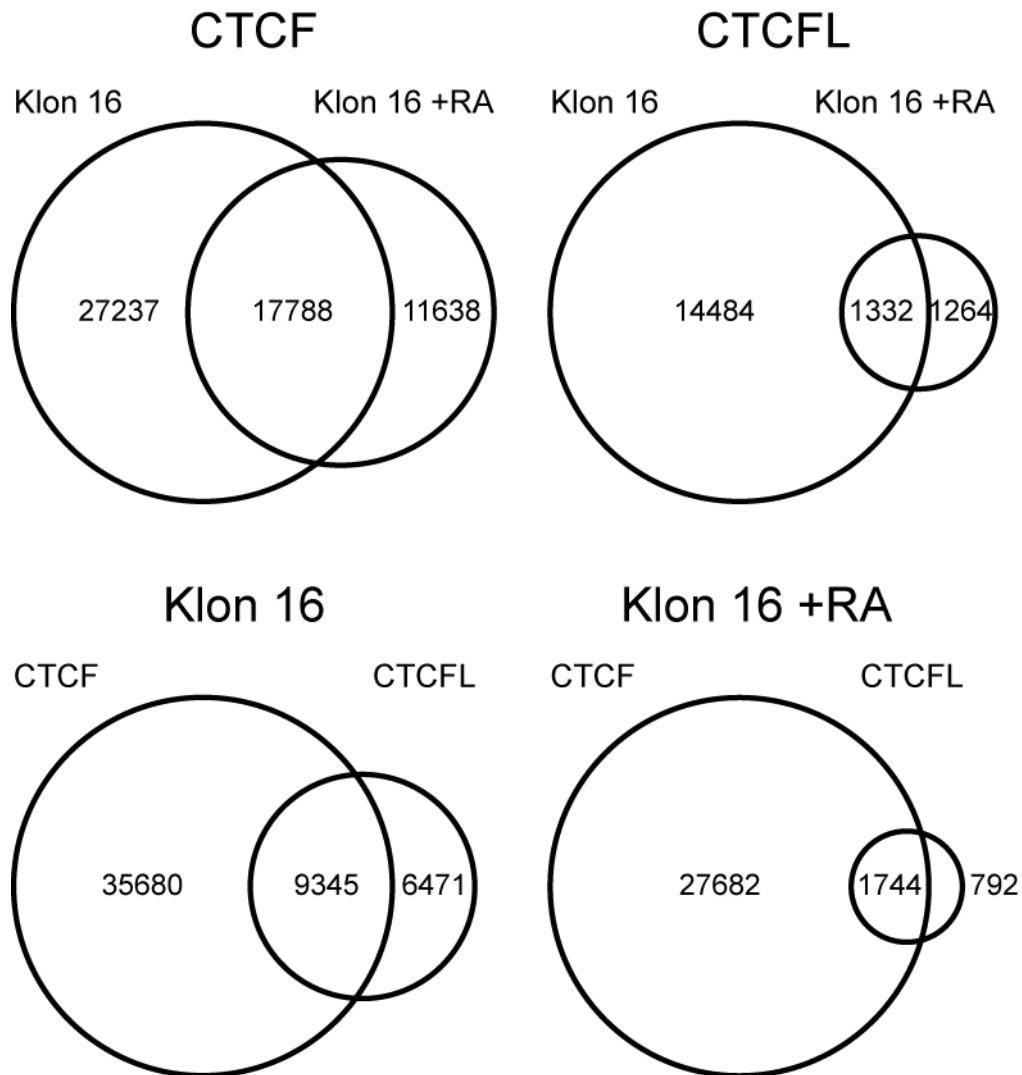
Nach Auswertung der ChIP-DNA mittels qPCR zeigen CTCF und CTCFL eine stark erhöhte Bindung an die Snapin Stelle im Vergleich zur Thoc1 down Stelle (Abb. 38). Diese Anreicherung bewegt sich zwischen 47-fach für die CTCF-IP im induzierten Klon 16 bis hin zu 101-fach für die CTCFL-IP im induzierten und differenzierten Klon 16. Somit ist in allen, zur Sequenzierung bestimmten, Proben eine Anreicherung einer Bindestelle über

eine Nicht-Bindestelle von weit über 10-fach detektierbar und die DNA damit für die weiteren Experimente nutzbar. Die Sequenzierung selbst, inklusive der Library-Präparation, wurde dann von S. Bhujy und M. Jarek (HZI Braunschweig) durchgeführt und die bioinformatische Auswertung der Rohdaten erfolgte durch M. Bartkuhn (JLU Gießen). Nach der Identifikation der jeweiligen Bindestellen, für die Datensätze, wurde eine Korrelationsanalyse durchgeführt und der Grad der Überlappung von Bindestellen zwischen den einzelnen Datensätzen in einem Venn Diagramm dargestellt.



**Abb. 39:** Der systematische Vergleich der CTCF und CTCFL-Bindestellen des undifferenzierten und differenzierten Klons 16 identifiziert überlappende aber auch jeweils individuelle Bindestellen. Dargestellt ist ein Venn-Diagramm das die Zahl der identischen Bindestellen zwischen den Datensätzen anzeigt. Alle Bindestellen eines Datensatzes sind dabei in einer Ellipse angezeigt und die sich überschneidende Bereiche zeigen gemeinsame Bindestellen.

Die Datensätze für den undifferenzierten Klon 16 weisen sowohl für CTCF (45.025 zu 29.426) als auch CTCFL (15.816 zu 2.536) eine höhere Zahl an identifizierter Bindestellen auf als die entsprechenden Sequenzierungen im differenzierten Klon 16. Bei Betrachtung der Überlappung der Bindestellen zeigen die 4 analysierten Datensätze eine große Diversität untereinander, nur 893 Bindestellen sind in allen Fällen von CTCF bzw. CTCFL gebunden (Abb. 39). Somit zeigen sowohl CTCF als auch CTCFL einen Grad an differentieller DNA Bindung auf genomweiter Ebene in Folge von Zelldifferenzierung. Zur genaueren Betrachtung und Einordnung der Daten wurden auch einzelne paarweise Vergleiche der überlappenden Bindestellen in Venn Diagrammen dargestellt.

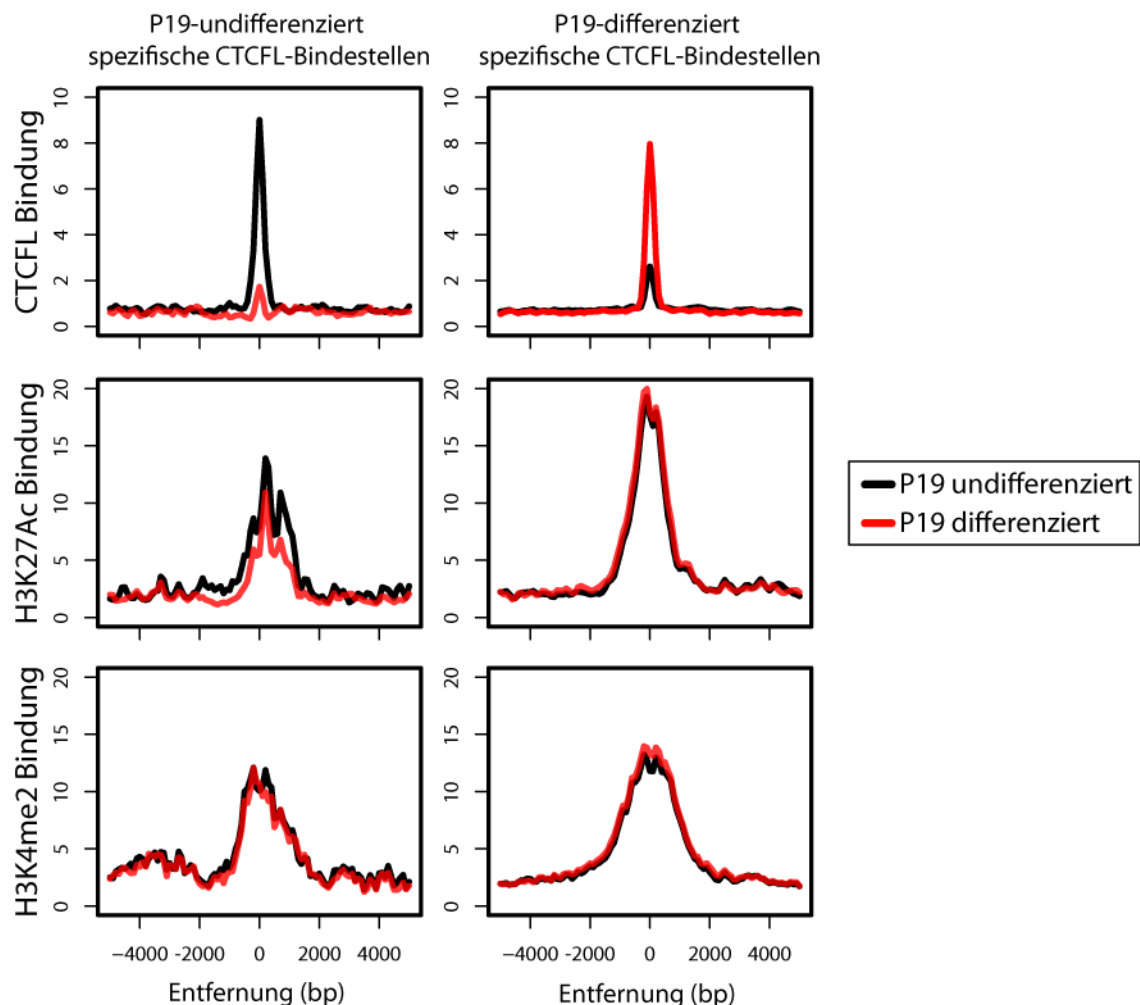


**Abb. 40: Paarweise Vergleiche der Bindestellen zeigen unterschiedliche starke Überlappung.** Dargestellt sind Venn-Diagramme der paarweisen Vergleiche der genom-weiten Datensätze, die die Zahl der jeweils identischen Bindestellen anzeigen. Alle Bindestellen eines Datensatzes sind dabei in einem Kreis angezeigt und die sich überschneidende Bereiche zeigen gemeinsame Bindestellen.

Die paarweise Betrachtung der Überlappung der Bindestellen zeigt das 60% der CTCF Bindestellen (17.788 von 29.426) und 52 % der CTCFL Bindestellen (1.332 von 2.536) die im differenzierten Klon identifiziert wurden auch im uninduzierten Fall, von den beiden Faktoren gebunden, vorliegen (Abb. 40, obere Reihe). CTCFL zeigt somit eine etwas höhere Variabilität der Bindestellen in Folge der Differenzierung. Analysiert man den Grad an Überlappung von CTCF und CTCFL jeweils innerhalb eines Zelltyps so sind im undifferenzierten Klon 59% der CTCFL Bindestellen (9.345 von 15.816) auch CTCF Bindestellen während im differenzierten Klon 68% der CTCFL Bindestellen (1.744 von 2.536) CTCF Bindung aufweisen (Abb. 40, untere Reihe). Somit zeigen die Faktoren in

beiden untersuchten Zuständen der P19 Zellen einen deutlichen Teil an identischer Bindestellen, aber es lassen sich jeweils auch nur von einem Faktor gebundene genomische Bereiche feststellen.

Zur Identifizierung des möglichen Einflusses von Änderungen der Histon-Modifikationen im Bereich der CTCFL-Bindestellen auf die CTCFL Bindung, wurden differenzielle CTCFL-Stellen identifiziert und dann mit frei verfügbaren Datensätzen für H3K4me2 und H3K27Ac verglichen, die ebenfalls in P19 Zellen mit identischem Differenzierungsprotokoll gewonnen wurden (Sérandour et al. 2012). Dieser Vergleich erfolgte anhand einer kumulativen Zusammenfassung aller Bindesignale der beiden Histon-Modifikationen über die beiden Gruppen der identifizierten P19-differenziert spezifischen und P19-undifferenziert spezifischen Bindestellen.

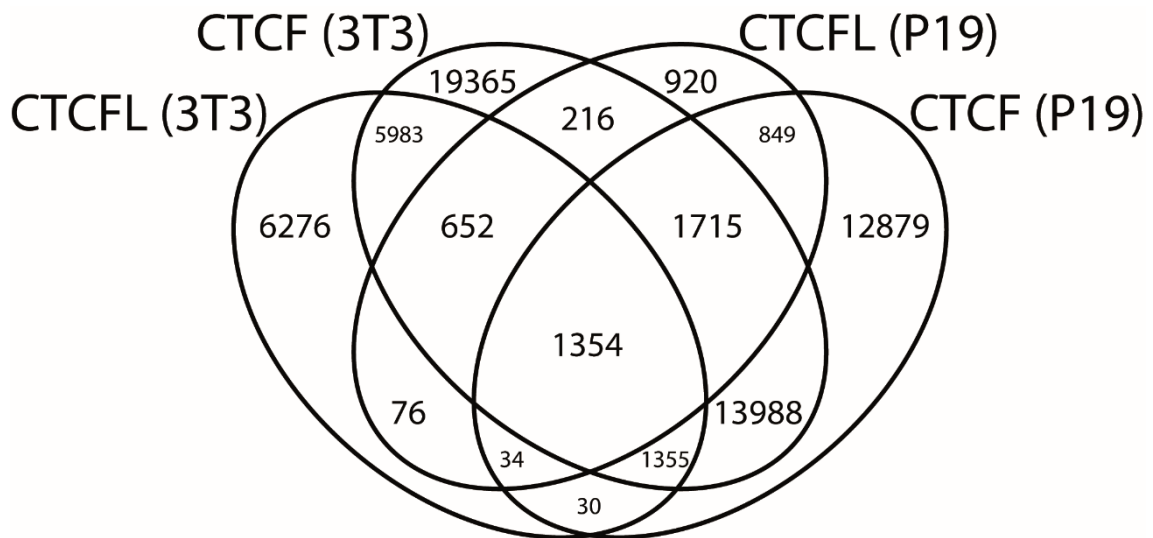


**Abb. 41: Differentielle CTCFL Bindemuster korrelieren nicht mit zwei untersuchten Histonmodifikationen.** Differenzielle CTCFL-Bindestellen wurden identifiziert und dann die Bindedaten der entsprechenden Datensätze für H3K4me3 und H3K27Ac, in einem genomischen Fenster von +/-5000 bp darüber dargestellt.

Die bioinformatische Auswertung zeigt einerseits, dass die Gruppen der Bindestellen korrekt gewählt sind, da bei Betrachtung der CTCFL Bindung nur für die jeweils zugrunde liegende Gruppe eine starke Bindung detektierbar ist (Abb. 41), andererseits ist für die Modifikationen kein Unterschied erkennbar. So zeigen sowohl die P19-undifferenziert spezifischen als auch die P19-differenziert spezifischen Bindestellen in beiden Fällen Anwesenheit von H3K4me3 und H3K27ac die nicht zur beobachteten CTCFL-Bindung korreliert. Einzig für die H3K27 Acetylierung über die P19-undifferenziert spezifischen Bindestellen ist eine leicht erniedrigte Anwesenheit in den differenzierten P19 Zellen ersichtlich, diese Korrelation lässt sich jedoch für den umgedrehten Fall nicht bestätigen. Somit konnte im Zuge dieser Analyse kein genereller Zusammenhang zwischen differenziellen CTCFL-Bindevorgängen und differenziellen Histon-Modifikations-Mustern gefunden werden.

#### **4.4.4 Vergleichende Analyse der CTCFL-Bindung zwischen NIH3T3 und P19 Klonen**

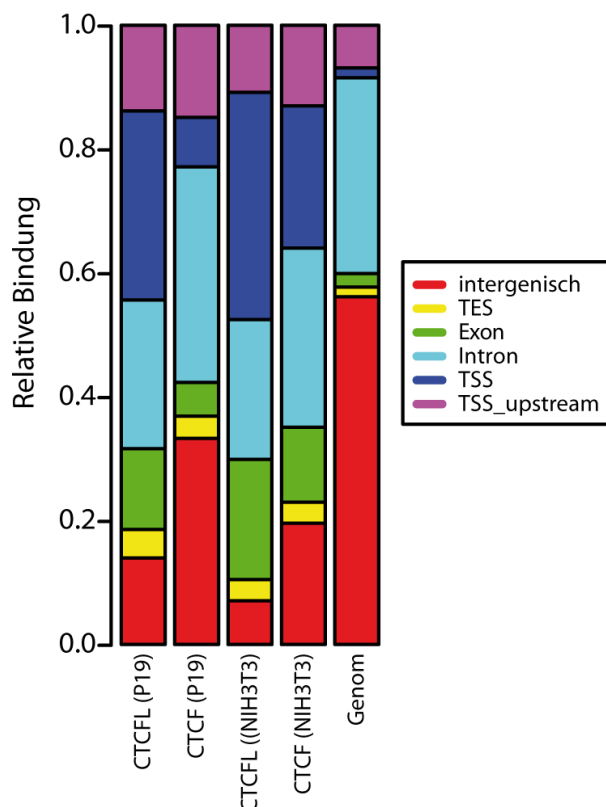
Da bei den Analysen innerhalb des Klons 16 keine Aussagen über den Zusammenhang zwischen CTCFL-Bindung und der zwei untersuchten Histon-Modifikationen H3K4me3 und H3K27ac gemacht werden konnten sollten Zelltyp übergreifende vergleichende Analysen durchgeführt werden. Zur Untersuchung zelltyp-spezifischer Bindemuster von CTCFL wurden die Bindestellen der NIH3T3- und P19-Zellklone systematisch miteinander verglichen. Für den P19-Klon wurden jeweils die Datensätze der Zellen im undifferenzierten Zustand verwendet. Die Gesamtzahlen der Bindestellen der Datensätze waren, bei stringenter Filterung, für NIH3T3 44.628 CTCF und 15.760 CTCFL Bindestellen und für P19 32.204 CTCF und 5.816 CTCFL Bindestellen. Für diese Datensätze wurde der Anteil an geteilten Bindestellen in einem Venn-Diagramm dargestellt.



**Abb. 42:** Der systematische Vergleich der CTCF und CTCFL-Bindestellen der beiden Klone identifiziert überlappende aber auch jeweils individuelle Bindestellen. Dargestellt ist ein Venn-Diagramm das die Zahl der identischen Bindestellen zwischen den Datensätzen anzeigt. Alle Bindestellen eines Datensatzes sind dabei in einer Ellipse angezeigt und die sich überschneidende Bereiche zeigen gemeinsame Bindestellen.

Die Analyse belegt, dass alle Datensätze gemeinsame Bindestellen zeigen. So sind beispielsweise 1.354 Bindestellen in beiden Zellklonen sowohl CTCF als auch CTCFL gebunden (Abb. 42). Andererseits zeigen beide Klone auch jeweils einzigartige Bindestellen für beide Proteine, so sind etwa 6.276 Bindestellen nur von CTCFL im NIH3T3-Datensatz und sonst keinem anderen untersuchten Datensatz gebunden. Betrachtet man den Anteil identischer CTCF Bindestellen zwischen den Klonen so zeigt sich eine 57,2%ige Überlappung (18.412 von 32.204) der P19 Bindestellen mit den NIH3T3 Bindestellen. Für CTCFL ist eine niedrigere Überlappung zwischen den Zelltypen zu betrachten, hier sind nur 36,4% (2.116 von 5.816) der P19-Stellen auch in NIH3T3 Zellen gebunden, was somit einen Hinweis auf variabelere Bindemuster von CTCFL im Vergleich zu CTCF darstellt. Im nächsten Schritt wurde die genomweite Verteilung der Bindestellen in Bezug auf funktionelle genomische Klassen analysiert. Dazu wurde dargestellt welcher Anteil der Bindestellen der einzelnen Datensätze welcher genomischen Klasse zugeordnet werden konnte.

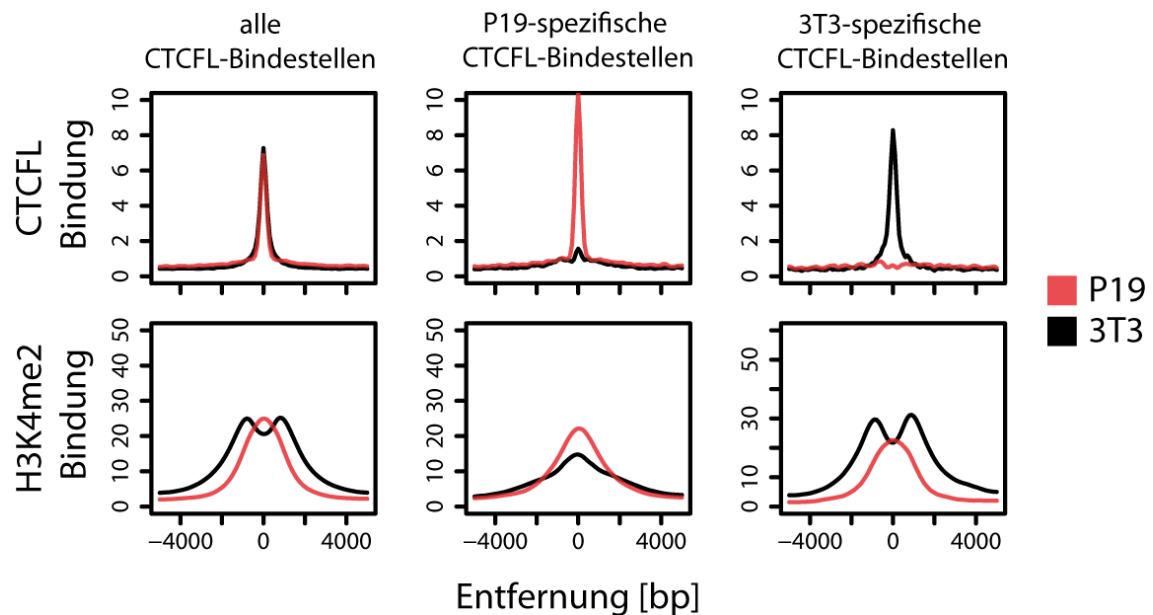




**Abb. 43: CTCF- und CTCFL-Bindestellen sind sowohl in P19 als auch NIH3T3 Zellen in Gen-assoziierten Bereichen angereichert.** Der Anteil der jeweiligen Bereiche am Genom ist in der rechten Säule angegeben. Links davon sind die tatsächlichen Verteilungen der CTCF und CTCFL Bindestellen in den P19- und NIH3T3-Zellen dargestellt.

Bei der Verteilung der jeweiligen Bindestellen im Genom sind Unterschiede zu betrachten. So zeigt CTCF, bei Analyse der Verteilung der Bindestellen im Genom, in NIH3T3-Zellen eine größere Anreicherung in Gen-assoziierten Bereichen als in P19-Zellen (Abb. 43). Diese zeigen zwar auch eine Anreicherung im Vergleich zum Anteil im Genom, aber in den NIH3T3 Zellen sind insgesamt mehr CTCF Bindestellen in Exons, TESs, TSSs und in den upstream-Bereichen der TSSs vorhanden. CTCFL zeigt in beiden Klonen eine größere Präferenz für offene Chromatin-Bereiche, so zeigt CTCFL jeweils einen größeren Anteil seiner Bindestellen in Exons und TSSs als der vergleichbare CTCF Datensatz. Somit konnten in beiden Zell-Klonen die gleichen zugrundeliegenden genomischen Verteilungsmuster von CTCFL-Bindestellen identifiziert werden, obwohl insgesamt die Überlappung dieser Bindestellen nicht sehr groß war. Um den Zusammenhang dieser zelltypspezifischer Bindevorgänge von CTCFL mit möglichen Unterschieden im Aufbau des Chromatins in den jeweiligen DNA-Abschnitten zu untersuchen wurden weitere bioinformatische Korrelationsanalysen durchgeführt. Dazu wurden die spezifischen CTCFL-

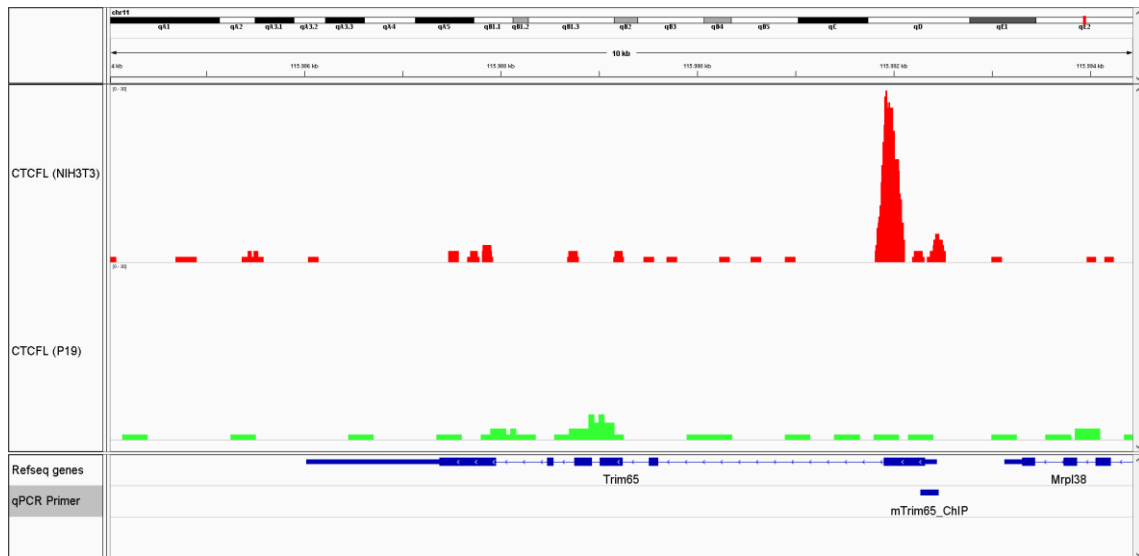
Bindestellen identifiziert und dann mit frei verfügbaren Datensätze für die Histon-Modifikation H3K4me2 aus P19- und NIH3T3-Zellen korreliert. Weitere Datensätze anderer Histon-Modifikationen standen zum Zeitpunkt der Analysen nicht zur Verfügung.



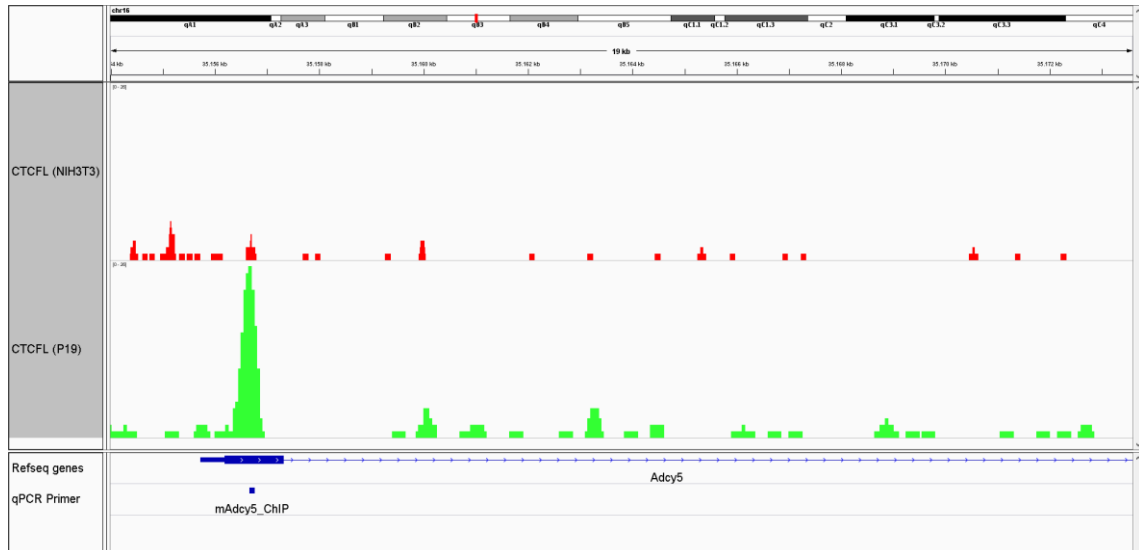
**Abb. 44: Zelltypspezifische CTCFL-Bindestellen zeichnen sich durch differenzielle Anwesenheit von H3K4me2 aus.** Differenzielle CTCFL-Bindestellen wurden identifiziert und dann die Bindedaten der entsprechenden Datensätze für H3K4me2, in einem genomischen Fenster von +/-5000 bp darüber dargestellt.

Die bioinformatische Analyse zeigt jeweils für die spezifischen Bindestellen einhergehende verstärkte Anwesenheit von H3K4me2, so zeichnen sich die in P19-Zellen von CTCFL gebundenen Bindestellen durch mehr H3K4me2 aus, als die gleichen, nicht gebundenen Bereiche, in NIH3T3 Zellen (Abb. 44). Anders herum lässt sich die gleiche Korrelation betrachten. Bei Betrachtung der Ergebnisse für CTCFL Bindung zeigt sich, das aufgrund der Auswahl der Klassen erwartete Ergebnis, so zeigen beispielsweise P19 spezifische Bindestellen auch wirklich nur ein Signal in P19 Zellen. Somit können bei der zelltyp-übergreifenden Analyse differenzielle CTCFL-Bindestellen mit einhergehender Anwesenheit einer Histon-Modifikation, H3K4me2, positiv korreliert werden. Weitere Untersuchungen zu anderen Histon-Modifikationen waren auf einer genom-weiten Ebene, wegen dem Fehlen entsprechender Datensätze, nicht möglich. Um dennoch Aussagen über differenzielle CTCFL-Bindemuster in Abhängigkeit von unterschiedlicher DNA-Zugänglichkeit in den Klonen treffen zu können wurden FAIRE Experimente durchgeführt. Hierzu wurde sowohl in induziertem NIH3T3 Klon 34 und induziertem P19 Klon 16 der Anteil an offener, zugänglicher DNA mittels FAIRE isoliert und durch qPCR ausgewertet. Als Bereiche die in den PCRs überprüft werden sollten, wurden jeweils 3 Bereiche mit NIH3T3 spezifischer CTCFL-Bindung (Beispiel Abb. 45) und 3 mit P19-

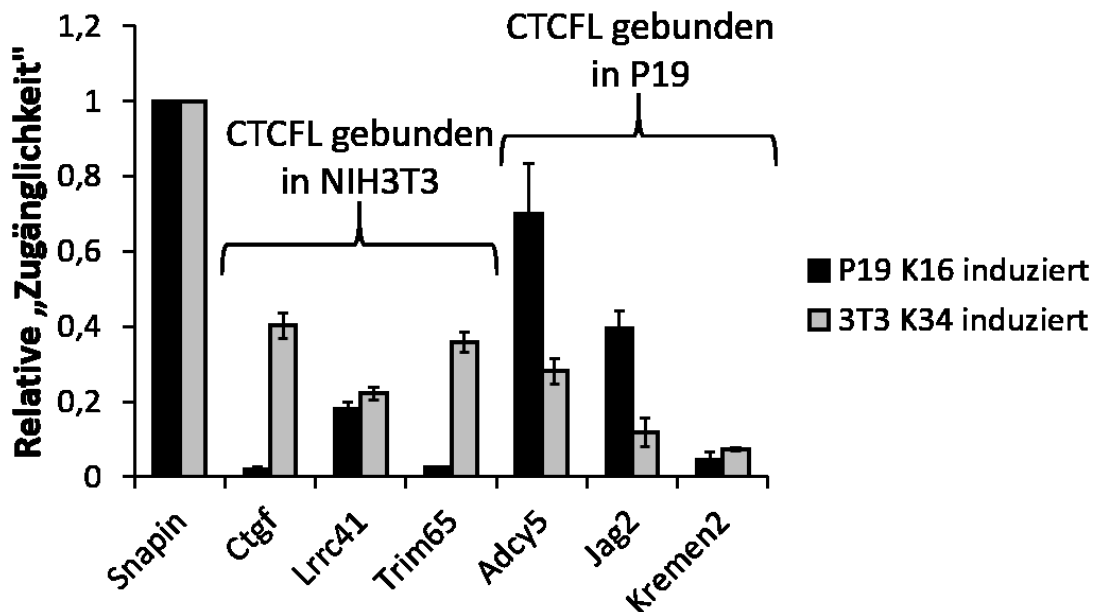
spezifischer (Beispiel Abb. 46) anhand der vorhandene CTCFL ChIPseq Daten ausgewählt.



**Abb. 45: Genomischer Bereich des Gens *Trim65* der die NIH3T3 spezifische Bindung von CTCFL zeigt.** Dargestellt der Genbereiches von *Trim65*, Exons sind in blauen Boxen angegeben, der translatierte Bereich dabei in dickeren, der untranslatierte Bereiche in schmalen Boxen. Darüber sind jeweils die CTCFL-ChIPseq-Profile in NIH3T3-Zellen (rot) und P19-Zellen (grün) angegeben. Die genaue Lokalisation des entsprechenden CHIP-PCR Produktes ist ebenfalls dargestellt.



**Abb. 46: Ausschnitt des genomischen Bereichs des Gens *Adcy5* der die P19 spezifische Bindung von CTCFL zeigt.** Dargestellt ist ein Teil des Genbereiches von *Adcy5*, Exons sind in blauen Boxen angegeben, der translatierte Bereich dabei in dickeren, der untranslatierte Bereiche in schmalen Boxen. Darüber sind jeweils die CTCFL-ChIPseq-Profile in NIH3T3-Zellen (rot) und P19-Zellen (grün) angegeben. Die genaue Lokalisation des entsprechenden CHIP-PCR Produktes ist ebenfalls dargestellt.



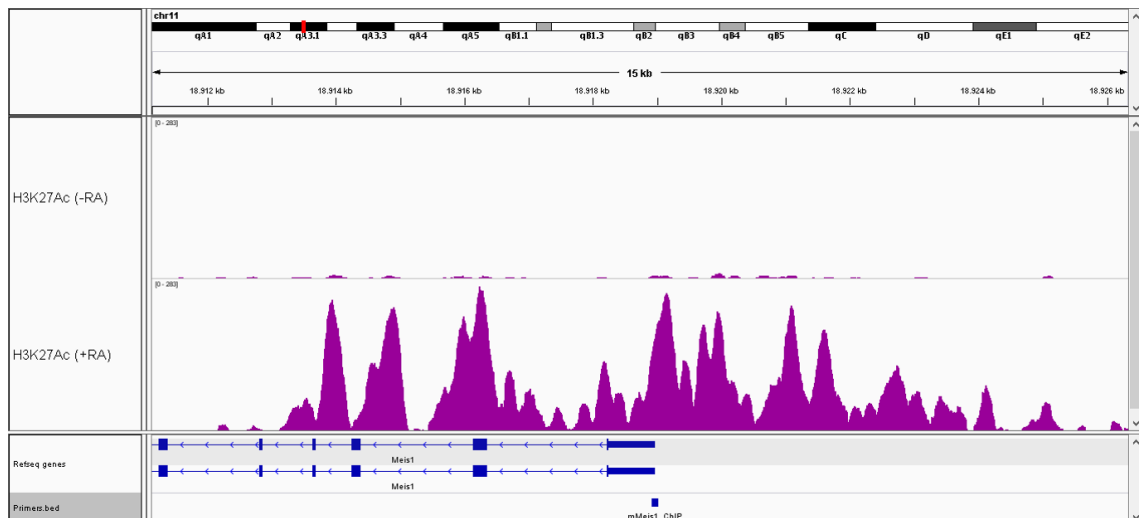
**Abb. 47:** Für 4 der 6 untersuchten Bereiche kann differenzielle Offenheit des Chromatins zwischen den Klonen detektiert werden. FAIRE-DNA von induzierten P19 Klon 16 und induzierten NIH3T3 Klon 34 Zellen wurde gewonnen und mittels qPCR, unter Verwendung entsprechender Oligonukleotidprimer gegen genomische Regionen, welche jeweils nach dem nächstgelegenen Gen benannt und relativ zum FAIRE qPCR Signal der Snapin Bindestelle in den jeweiligen Proben dargestellt sind, analysiert. Die Fehlerbalken stellen die Standardabweichung von je drei unabhängigen Ansätzen dar.

Die Auswertung der qPCR zeigt für vier der sechs ausgewählten Bereiche eine differenzielle Zugänglichkeit in den FAIRE-Experimenten (Abb. 47). So zeigen zwei (Ctgf & Trim65) der drei, in NIH3T3 CTCFL gebundenen Bereiche, auch eine NIH3T3 spezifische Offenheit des Chromatins. Umgekehrt zeichnen sich von den drei, in P19 Zellen CTCFL gebundenen Bereichen, zwei (Adcy5 & Jag2) auch durch eine höhere Zugänglichkeit in P19 Zellen aus. Die Bereiche Lrrc41 und Kremen2 zeigen keine differenzielle Offenheit zwischen den untersuchten Zellen. Somit können für einige Bindestellen die zelltyp-spezifische CTCFL Bindung zeigen einhergehende Unterschiede in der Zugänglichkeit der DNA in diesen Zellen festgestellt werden.

#### 4.4.5 Gewinnung eigener Daten für Konformationsänderung des Chromatins in Folge von Differenzierung

Da die genomweiten Daten für Histon-Modifikationen in P19 Zellen aus öffentlichen Datenbanken stammten unterliegen sie möglichen Fehlerquellen und experimentellen Abweichungen zu den eigenen analysierten Zellen. Auch lagen nur Daten für wenige aktive Modifikationen vor. Daher sollten eigene Daten für die generelle Zugänglichkeit des Chromatins im P19 Klon 16 in Abhängigkeit von CTCFL Expression erzeugt werden.

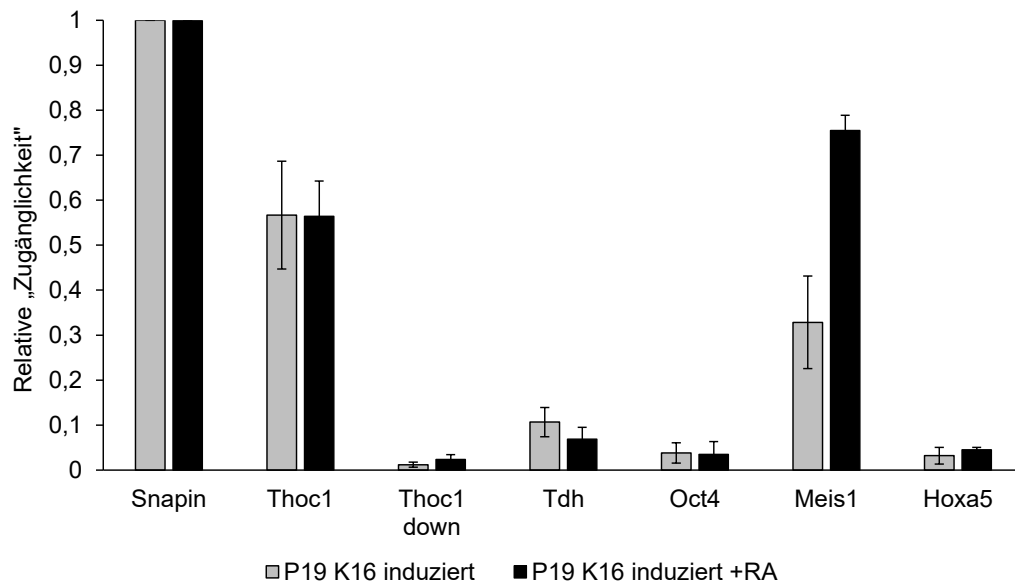
Hierzu bot sich die FAIRE Methode an, die ohne Berücksichtigung von bestimmten Transkriptionsfaktoren oder Histon-Modifikationen ein Maß an Offenheit von DNA Bereichen liefert. Um einen möglichst großen Einfluss der Differenzierung auf das Chromatin zu erhalten wurden die P19 Klon 16 Zellen für 9 Tage durch Zugabe von RA differenziert und anschließend für 2 Tage für CTCFL Expression induziert. Als Kontrolle wurden P19 Klon 16 Zellen für die gleiche Dauer normal kultiviert und dann auch 2 Tage induziert. Anschließend wurde die FAIRE Methode in biologischen Triplikaten durchgeführt. Zur Überprüfung der erfolgreichen Durchführung der FAIRE-Methode, wurden 4 Kandidaten-Bereiche Tdh, Oct4, Meis1 und Hoxa5 ausgesucht, die bei Betrachtung von öffentlichen Daten für H3K27Ac eine Änderung dieser Modifikation in Antwort auf RA vermittelte Differenzierung zeigen (Beispielhaft Meis1, Abb. 48). Diese Bereiche wurden jeweils nach dem nächstgelegenen Gen benannt.



**Abb. 48: Genomischer Bereich des Gens *Meis1* der das differenzielle Muster von H3K27 Acetylierung im Zuge RA vermittelter Differenzierung zeigt.** Dargestellt ist ein Teil des Genbereiches von *Meis1*, Exons sind in blauen Boxen angegeben, der translatierte Bereich dabei in dickeren, der untranslatierte Bereiche in schmaleren Boxen. Darüber sind jeweils die H3K27Ac-ChIPseq-Profile in undifferenzierten (-RA) und differenzierten (+RA) P19 Zellen angegeben.

Die ausgewählten Kandidaten-Bereiche zeigen bei Betrachtung der H3K27Ac Chip-seq-Profile sehr starke Unterschiede zwischen undifferenzierten und differenzierten Zellen. So zeigt exemplarisch die *Meis1* Stelle fast komplette Abwesenheit dieser Modifikation welche dann nach 2 Tagen RA-vermittelter Differenzierung sehr breit über den gesamten Promotorbereich des *Meis1*-Gens zu detektieren ist (Abb. 48). Diese Änderung im Modifikationsmuster geht mit Hochregulation von *Meis1*, einem wichtigen Faktor in zellulären Differenzierungsabläufen, einher. Somit stellten die ausgesuchten DNA-Bereiche

gute Ziele zur Detektion von differenziellem FAIRE-Signal dar. Dazu wurde die gewonnene FAIRE-DNA mittels qPCR für diese DNA Bereiche untersucht.



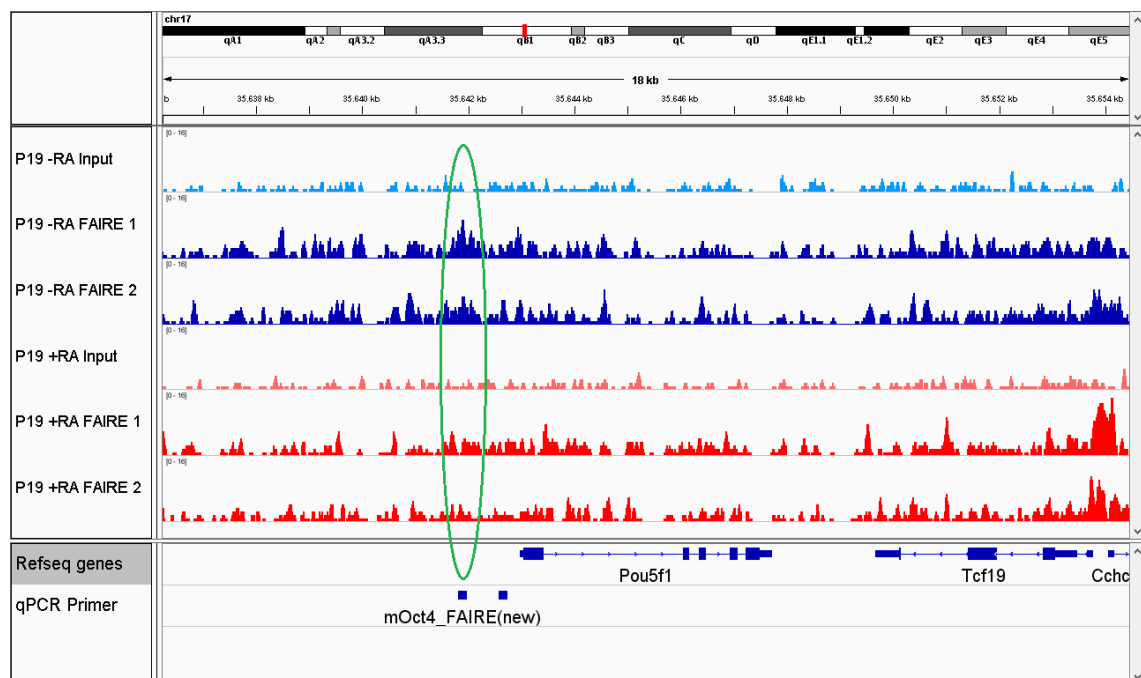
**Abb. 49: Meis1 zeigt differentielle FAIRE-Zugänglichkeit in Abhängigkeit von RA vermittelter Differenzierung.** FAIRE-DNA undifferenzierter und differenzierter Klon 16 Zellen wurde gewonnen und mittels qPCR, unter Verwendung entsprechender Oligonukleotidprimer gegen genomische Regionen, welche jeweils nach dem nächstgelegenen Gen benannt und relativ zum FAIRE qPCR Signal der Snapin Bindestelle in den jeweiligen Proben dargestellt sind, analysiert. Die Fehlerbalken stellen die Standardabweichung von je drei unabhängigen Ansätzen dar.

Für einen der vier untersuchten Bereiche, Meis1, ist in der FAIRE qPCR eine klare differentielle Zugänglichkeit des Chromatins erkennbar (Abb. 49). Meis1 zeichnet sich durch Zunahme der Zugänglichkeit im Zuge der Differenzierung der Zellen aus. Für die drei anderen Kandidaten-Bereiche, Tdh, Oct4 und Hoxa5, sind keine klaren Unterschiede der Zugänglichkeit des Chromatins zu erkennen, und generell zeigen diese Bereiche auch ein sehr niedriges Signal in den qPCRs. Somit kann zumindest für einen ausgewählten genomischen Bereich differenzielles FAIRE-Signal detektiert werden und der prinzipielle Erfolg der Methode bestätigt werden. Um anschließend die genom-weiten Änderungen des Chromatins anhand von FAIREseq zu identifizieren wurden, jeweils von 2 Proben der differenzierten und undifferenzierten K16 Zellen die FAIRE-DNA als auch jeweils eine entsprechenden Input-Probe von S. Bhuju und M. Jarek (HZI Braunschweig) sequenziert. Auch die Präparation der Sequenzier-Libraries erfolgte extern in Braunschweig. Da man bei der FAIRE Methode, im Vergleich zur ChIP größere Mengen an DNA erhält konnten einzelne Ansätze sequenziert werden und es war nicht nötig mehrere Experimente zu vereinigen.

**Tab. 8: Auflistung der sequenzierten FAIRE Proben und der jeweiligen DNA Mengen.** FAIRE wurde sowohl mit undifferenzierten als auch differenzierten, induzierten Klon 16 Zellen durchgeführt. Die gewonnene DNA als auch jeweils eine Probe der Input-DNA wurden jeweils vor Versand der unter Verwendung eines Qubit Fluorimeters quantifiziert.

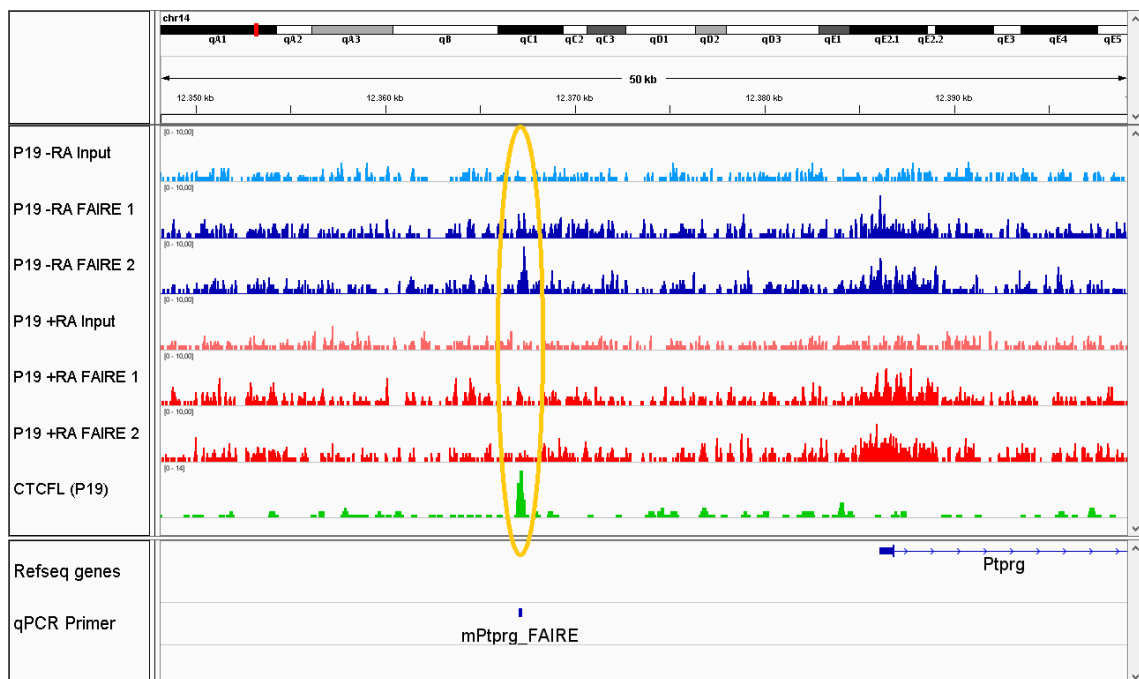
Nummer	Analysierte Probe	DNA Menge (ng)
1	P19 K16 induziert Input	1120
2	P19 K16 induziert FAIRE 1	192
3	P19 K16 induziert FAIRE 2	231
4	P19 K16 induziert +RA Input	248
5	P19 K16 induziert +RA FAIRE 1	26,7
6	P19 K16 induziert +RA FAIRE 2	36

Nach Sequenzierung der Proben erfolgte die bioinformatische Rohdaten-Auswertung durch M. Bartkuhn. Im Zuge dieser Auswertung konnten genomische Bereiche mit differentieller Zugänglichkeit der DNA identifiziert werden. Beispielhaft ist im Folgenden der genomische Bereich des wichtigen Pluripotenz-Gens *Oct4* (auch *Pou5f1*) dargestellt.



**Abb. 50: Im Bereich des Gens *Oct4* lässt sich differentielle Zugänglichkeit der DNA nach Differenzierung detektieren.** Dargestellt ist der Genbereich von *Oct4* (*Pou5f1*), Exons sind in blauen Boxen angegeben, der translatierte Bereich dabei in dickeren, der untranslatierte Bereiche in schmaleren Boxen. Darüber sind jeweils die FAIREseq-Profile in induzierten (-RA, Blau) und induzierten, differenzierten (+RA, Rot) P19 Zellen angegeben. Die entsprechenden Daten der Input-Sequenzierungen sind jeweils in schwächerem Blau bzw. Rot dargestellt.

Bei Betrachtung der FAIREseq Daten für den Oct4 Bereich zeigen beide undifferenzierten Proben eine höhere Zugänglichkeit als die differenzierten (Abb. 50). Somit können unter Nutzung der genomweiten FAIRE-Daten genomische Bereiche differentieller Zugänglichkeit identifiziert werden. Zur Untersuchung von CTCFL Bindung in entsprechenden Bereichen differentieller Offenheit wurden diese Stellen mit den entsprechenden CTCFL-Bindestellen in induziertem, undifferenziertem Klon verglichen. Auf diese Weise sollten Bereiche identifiziert werden an denen die detektierte CTCFL-Bindung durch die differentielle Offenheit im Zuge der Differenzierung beeinflusst wird. Es wurden dann anhand dieses Vergleiches 9 Kandidaten-Bereiche (*Ptprg*, *Sall4*, *Ptpru*, *Atp8b3*, *Ephb1*, *Cdh22*, *Mzf1*, *Alpl* & *Nkx3-2*) ausgewählt und mittels qPCR bezüglich ihrer DNA-Zugänglichkeit in Folge der Differenzierung analysiert. Eine dieser gewählten Bindestellen ist im Folgenden exemplarisch dargestellt.

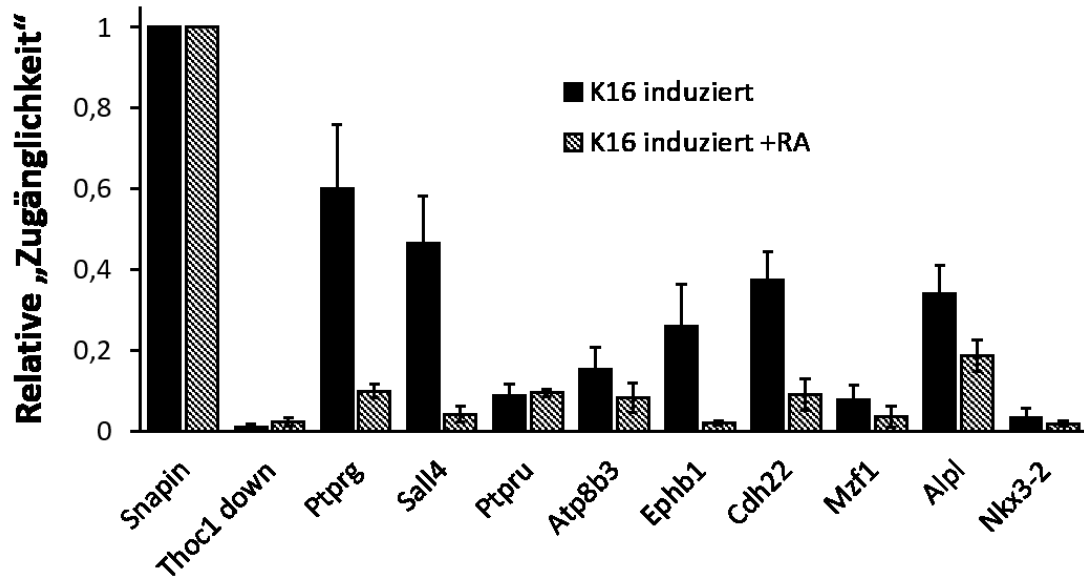


**Abb. 51: Die CTCFL-Bindestelle im Bereich des *Ptprg*-Gens zeigt differentielle Zugänglichkeit im Zuge der Differenzierung.** Dargestellt ist ein Teil des Genbereiches von *Ptprg*, Exons sind in blauen Boxen angegeben, der translatierte Bereich dabei in dickeren, der untranslatierte Bereiche in schmaleren Boxen. Darüber sind jeweils die FAIREseq-Profile in induzierten (-RA, Blau) und induzierten, differenzierten (+RA, Rot) P19 Zellen angegeben. Die entsprechenden Daten der Input-Sequenzierungen sind jeweils in schwächerem Blau bzw. Rot dargestellt. In grün ist weiterhin die CTCFL-Bindung in undifferenzierten P19 Zellen angegeben.

So zeigt die Betrachtung des genomischen Bereiches der CTCFL-Bindestelle vor dem *Ptprg*-Gen differentielle Zugänglichkeit der DNA anhand den FAIREseq Daten (Abb. 51). Es sind stärkere Peaks in den P19 -RA FAIRE Proben zu erkennen, diese nehmen in den P19 +RA Proben ab, was dann für eine reduzierte Zugänglichkeit spricht. Diese und

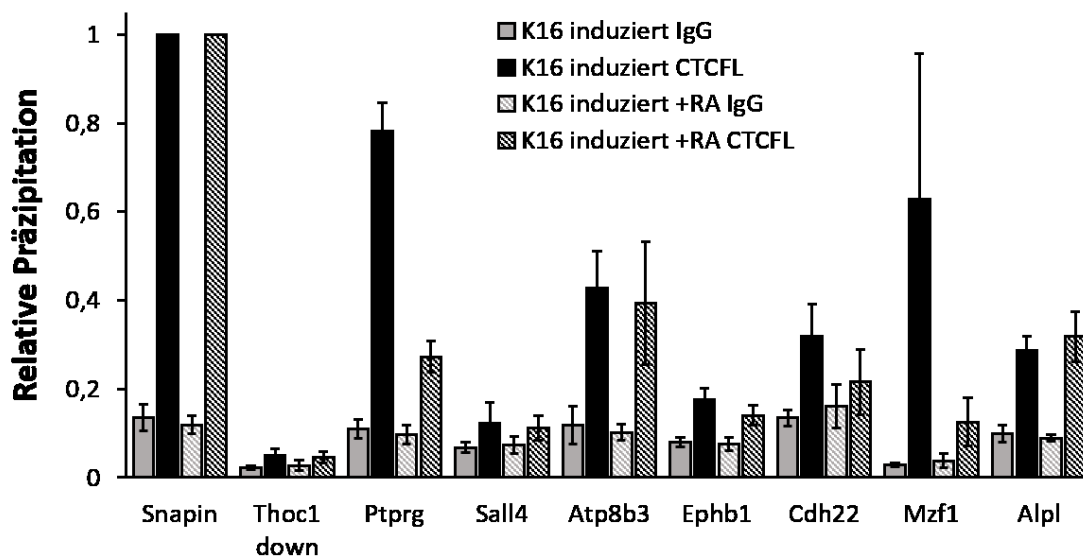


die 8 weiteren CTCFL-Bindestellen, die Überlappung mit Bereichen differentieller Offenheit zeigten, wurden in unabhängigen Experimenten in biologischen Triplikaten überprüft.



**Abb. 52:** Die meisten untersuchten Bereiche zeigen die erwartete differentielle Zugänglichkeit nach Differenzierung. FAIRE-DNA undifferenzierter und differenzierter Klon 16 Zellen wurde gewonnen und mittels qPCR, unter Verwendung entsprechender Oligonukleotidprimer gegen genomische Regionen, welche jeweils nach dem nächstgelegenen Gen benannt und relativ zum FAIRE qPCR Signal der Snapin Bindestelle in den jeweiligen Proben dargestellt sind, analysiert. Die Fehlerbalken stellen die Standardabweichung von je drei unabhängigen Ansätzen dar.

Die Auswertung der qPCR der FAIRE DNA zeigt für 7 der 9 untersuchten Bindestellen einen Unterschied in der Zugänglichkeit der DNA (Abb. 52). Für diese Bereiche ist jeweils ein reduziertes PCR-Signal in den FAIRE-Proben des differenzierten Klon 16 erkennbar. Die zwei Bindestellen Ptpru und Nkx3-2 zeigen diesen Trend nicht, wobei sich Nkx3-2 generell durch sehr schwache Signale in der PCR auszeichnet und somit Aussagen über Ab- oder Zunahmen der Offenheit unmöglich macht. Als Kontrollbereiche dienen in der Auswertung der geschlossene DNA-Bereich Thoc1 down und die offene CTCFL-Bindestelle Snapin, die auch zur Normalisierung zwischen den einzelnen FAIRE-Proben herangezogen wird. Somit kann für die meisten untersuchten Bindestellen der Befund der FAIREseq Daten bestätigt werden und differentielle Zugänglichkeit lässt sich nachweisen. Für die 7 bestätigten Bereiche wurde dann mittels ChIP untersucht, ob sich ein Einfluss der reduzierten Offenheit des Chromatins auf die Bindung von CTCFL nachweisen lässt. Hierzu wurde wie für die FAIREseq Experimente eine 9-tägige Differenzierung der Zellen mittels RA-Zugabe erzielt und anschließend ausgewertet.



**Abb. 53: Die Bereiche Ptpg und Mzf1 zeigen differentielle CTCFL Bindung in Abhängigkeit von Differenzierung.** Chromatin Immunopräzipitationen erfolgten entweder mit spezifischem Antikörper Sleutels6 gegen CTCFL oder zur Kontrolle mit unspezifischen IgG inkubiert. Die Auswertung erfolgte mittels qPCR mit Oligonukleotidprimern gegen die entsprechenden Regionen, welche jeweils nach dem nächstgelegenen Gen benannt und relativ zum qPCR Signal der jeweiligen CTCFL-Präzipitation an der Snapin-Bindestelle dargestellt sind. Die Fehlerbalken stellen die Standardabweichung von je drei unabhängigen Ansätzen dar.

Nach Auswertung der Chromatin-Immunopräzipitationen mittels qPCR zeigen Ptpg und Mzf1 eine differentielle Bindung von CTCFL (Abb. 53). An diesen beiden Bindestellen ist CTCFL in Folge der Differenzierung reduziert. Die anderen 5 untersuchten Bindestellen zeigen keinen Effekt der Differenzierung auf die Bindung von CTCFL. CTCFL Bindung kann an allen getesteten Bindestellen, im Vergleich zur unspezifischen Kontrolle angereichert, detektiert werden. Somit zeigen einzelne CTCFL-Bindestellen in Folge von zellulärer Differenzierung, und damit einhergehendem Verlust der DNA Zugänglichkeit, eine Abnahme der CTCFL-Bindung. Dieser Befund konnte aufgrund des Fehlens von genom-weiten CTCFL-Bindedaten im Langzeit Differenzierungs-Fall nicht generell bestätigt werden.

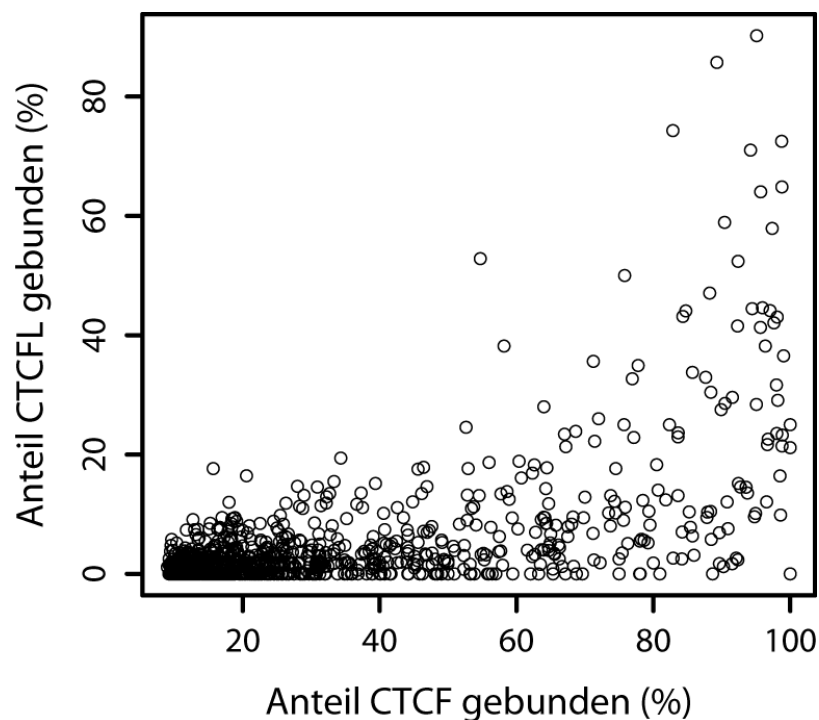
#### 4.5 Untersuchungen des DNA Bindeverhaltens von CTCF und CTCFL auf Grundlage von DNA Sequenz Unterschieden

Im Rahmen dieser Arbeit sollte ebenso untersucht werden welchen Einfluss DNA Sequenz Unterschiede auf die Bindung von CTCF und CTCFL haben. Auch wenn aufgrund der fast absolut identischen 11 Zink-Finger eine solche DNA Spezifität der beiden Fak-

toren lange vernachlässigt oder nicht vermutet wird. Anhand neuerer Analysen zum Einfluss der genauen DNA Sequenz auf die CTCF Bindung und Funktion (Liu et al. 2015) sollten Aussagen auf CTCFL Bindung übertragen werden.

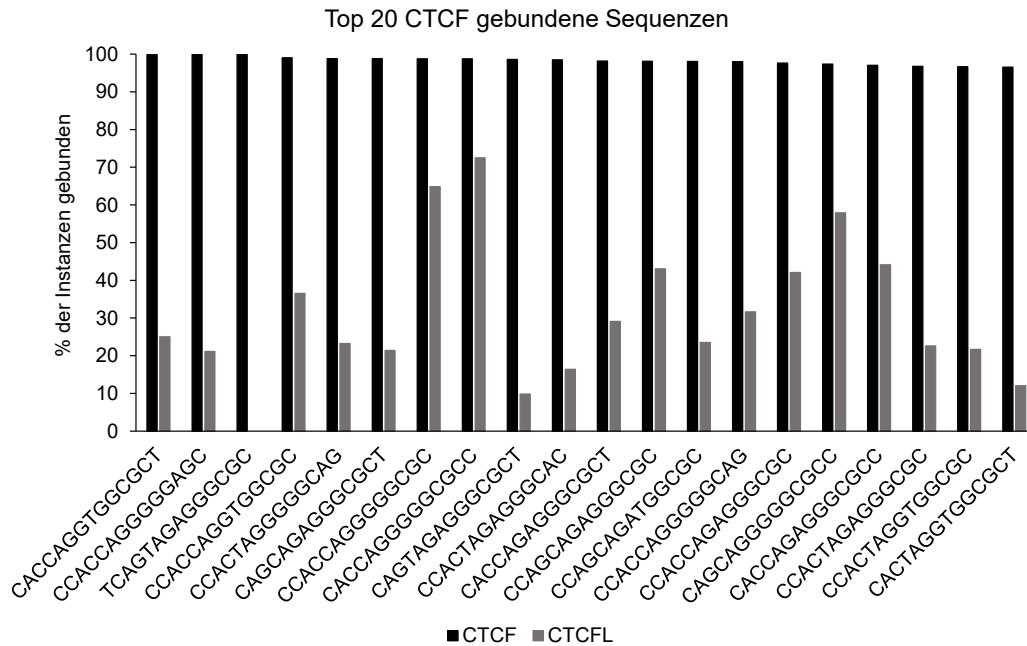
#### 4.5.1 Identifikation möglicher DNA Spezifität bei der Bindung von CTCF und CTCFL

Die in Liu *et al.* identifizierten Top 1000 14 bp DNA CTCF-Bindesequenzen wurden bioinformatisch unter Verwendung der öffentlichen ENCODE K562 Daten bezüglich dem Anteil ihrer CTCFL Bindung charakterisiert. Hierzu wurde bioinformatisch für jede der Top 1000 CTCF gebundenen Sequenzen der Anteil an Instanzen im Genom bestimmt der auch Bindung von CTCFL zeigte.



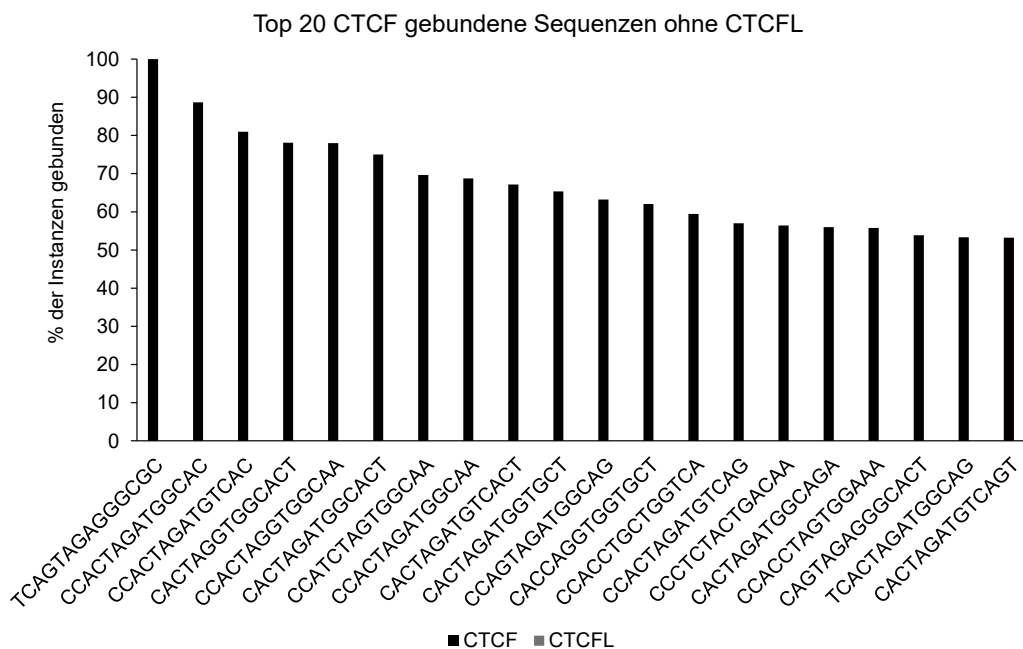
**Abb. 54: Viele der Top 1000 CTCF Sequenzen zeichnen sich auch durch CTCFL-Bindung aus.** Für jede der Top 1000 CTCF-Bindesequenzen wurde der Anteil an CTCF- und CTCFL-Bindung ermittelt und gegenübergestellt. Als Datengrundlage dienten genomweite ChIPseq Bindedaten für CTCF und CTCFL aus K562 Zellen. Jeder Datenpunkt gibt für eine individuelle DNA-Sequenz den prozentualen Anteil an CTCF- und CTCFL-Bindung über diese Sequenz im Genom an.

Bei Betrachtung der grafischen Auswertung der CTCFL-Bindung über die CTCF-Bindesequenzen zeigt sich ein linearer Trend (Abb. 54). So zeigen die meisten DNA-Sequenzen die einen hohen Anteil an CTCF-Bindung aufweisen auch einen stärkeren Anteil an CTCFL-Bindung. Im nächsten Schritt sollten die CTCFL-Bindung über die Top 20 CTCF-Bindesequenzen dargestellt werden.



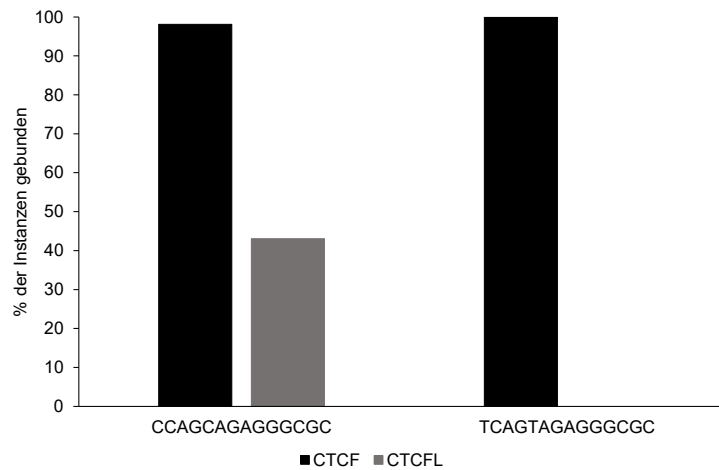
**Abb. 55: Die Top 20 CTCF gebundenen Sequenzen zeigen unterschiedliche Bindung von CTCFL.** Dargestellt sind, in absteigender Reihenfolge, die 20 CTCF-Bindesequenzen die den höchsten Grad an CTCF-Bindung in K562-Zellen aufweisen. Zusätzlich ist jeweils der Anteil an CTCFL-Bindung angegeben.

Es zeigen sich bei Betrachtung der am besten gebundenen CTCF Sequenzklassen unterschiedliche Level an CTCFL Bindung (Abb. 55). Der Grad der CTCF Bindung nimmt innerhalb der Top 20 Gruppe nur langsam von 100%, also der Fall in dem jede Instanz der Sequenz im Genom ist auch wirklich von CTCF gebunden ist, auf ca. 96% ab. CTCFL zeigt dabei hohe Variabilität, während die Sequenz CACCAGGGGGCGCC in mehr als 70% der im Genom vorkommenden Fälle von CTCFL gebunden ist, findet sich mit TCAGTAGAGGGCGC auch eine Sequenz die nie von CTCFL gebunden vorliegt. Weiterhin sollten die besten CTCF Motif-Sequenzen identifiziert werden, für die überhaupt keine Bindung von CTCFL vorliegt.



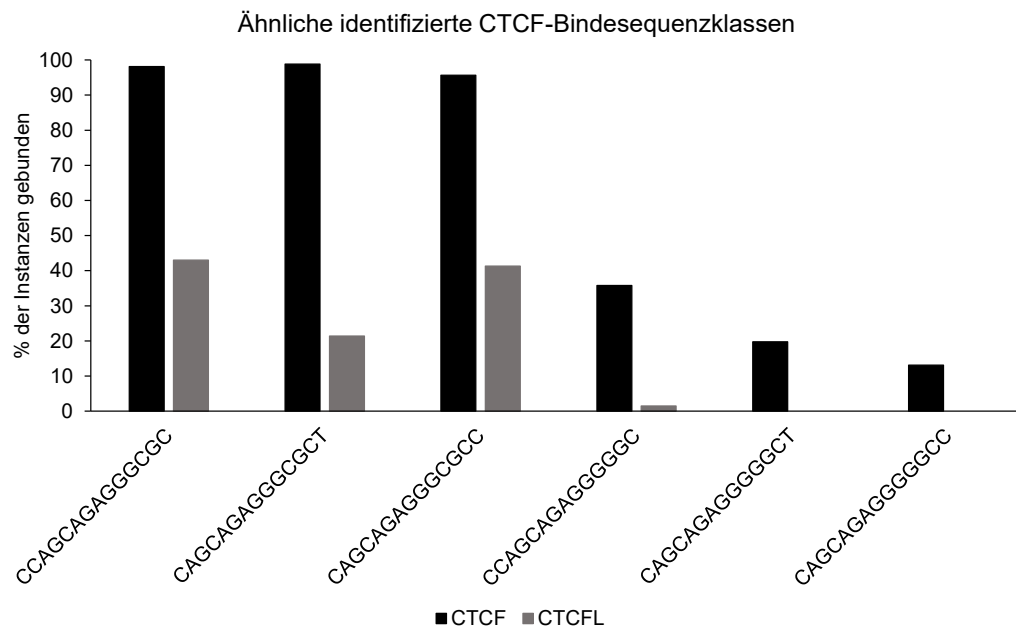
**Abb. 56: Die Top 20 CTCF gebunden Sequenzen, die keine CTCFL Bindung aufweisen, bewegen sich im Bereich mittlerer CTCF Bindung.** Dargestellt sind, in absteigender Reihenfolge, die 20 CTCF-Bindesequenzen die den höchsten Grad an CTCF-Bindung, bei gleichzeitiger Abwesenheit von CTCFL, in K562-Zellen aufweisen. Zusätzlich ist jeweils der Anteil an CTCFL-Bindung angegeben, der hier jedoch jeweils 0 beträgt.

Bei diesen Top 20 CTCF, aber nicht CTCFL, gebundenen Sequenzen zeigt sich eine durchschnittlich niedrigerer Bindungsanteil von CTCF als bei Sequenzen mit vorhandener CTCFL Bindung (Abb. 56). Während eine Sequenz (TCAGTAGAGGGCGC) zu 100% Prozent von CTCF gebunden vorliegt, weisen die anderen Sequenzen einen deutlich niedrigeren Anteil an gebundenem CTCF auf. So liegt der Prozentsatz bei der zwanzigsten Sequenz (CACTAGATGTCAGT) nur noch bei knapp über 50%. Um zu untersuchen ob es in beiden Gruppen ähnliche, sich nur wenig unterscheidende, Sequenzen gibt, die Hinweise auf Sequenzspezifität der beiden Proteine CTCF und CTCFL liefern könnten wurde eine webbasierte DNA-Sequenz-Alignierung unter Verwendung des Clustal Omega Tools des European Bioinformatics Institute (<http://www.ebi.ac.uk/Tools/msa/clustalo>) durchgeführt.



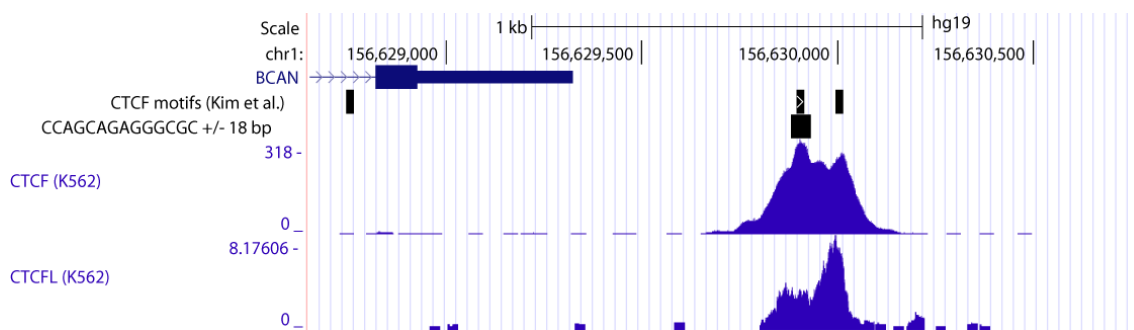
**Abb. 57:** Die zwei sehr ähnlichen CTCF Bindemotive CCAGCAGAGGGCGC und TCAGTAGAGGGCGC unterscheiden sich stark im Grad ihrer CTCFL Bindung. Dargestellt sind die zwei sehr ähnliche CTCF-Bindesequenzen CCAGCAGAGGGCGC und TCAGTAGAGGGCGC und jeweils ihr Anteil an CTCF und CTCFL gebundenen Instanzen in K562 Zellen.

Anhand dieser Korrelations-Analyse können zwei sehr ähnliche Sequenzen identifiziert werden, die sich nur an 2 der 14 Basenpositionen der Motife unterscheiden, jedoch ein stark unterschiedliches Bindevverhalten für CTCFL aufweisen (Abb. 57). Bei diesen identifizierten DNA Sequenzen handelt es sich um CCAGCAGAGGGCGC und TCAGTAGAGGGCGC, mit den Sequenzunterschieden C zu T an Position 1 und 5. Während die Sequenz, die ein Cytosin an diesen Positionen besitzt, eine Bindung von CTCFL in über 40% der vorkommenden Instanzen zeigt, ist in K562 Zellen keine einzige vorliegende Instanz, der Sequenz mit Thymin an 1 und 5, von CTCFL gebunden. Beide Sequenzen zeigen dabei aber eine identische Bindung von CTCF von jeweils fast 100%. Anhand der identifizierten Sequenzen wurden dann weitere Sequenz-Korrelationsanalysen unter Verwendung des Clustal Omega Tools durchgeführt. Diesmal wurde die Ausgangs-Sequenz CCAGCAGAGGGCGC gegen die gesamte Gruppe der Top 1000 CTCF-Motivklassen aligniert und versucht weitere ähnliche DNA-Sequenzen zu identifizieren.



**Abb. 58:** Die fünf ähnlichsten DNA Sequenzen zu CCAGCAGAGGGCGC zeigen stark unterschiedliche Bindung von CTCF. Dargestellt ist die CTCF-Bindesequenz CCAGCAGAGGGCGC und die 5 ähnlichsten identifizierten Sequenzen der Top 1000 Gruppe. Für die Sequenzen ist jeweils der Anteil an CTCF und CTCFL gebundenen Instanzen in K562 Zellen angegeben.

Bei Betrachtung der fünf, als am ähnlichsten zu CCAGCAGAGGGCGC identifizierten DNA Sequenzen lassen sich zwei funktional unterschiedliche Gruppen erkennen. Zum einen, 2 Sequenzen mit fast 100% Bindung von CTCF, und 3 Sequenzen mit stark reduzierter CTCF- und fast komplett eingebüßter CTCFL-Bindung (Abb. 58). Bei genauerer Betrachtung unterscheiden sich diese beiden Sequenzgruppen vor allem durch das Vorhandensein eines GGGGG Blockes bei den schwächeren gebundenen Sequenzen. Die stark gebundenen Sequenzen zeichnen sich an diesen Positionen stattdessen durch die Abfolge GGGCG aus. Für die weiteren Experimente, zur Untersuchung der Bindespezifitäten von CTCF und CTCFL, wurden dann entsprechende Oligonukleotide bestellt. Dazu wurden jeweils die Instanzen der 14 bp Bindesequenzen per Hand im UCSC Genome Browser untersucht und geeignete Stellen als 51-Mer herausgesucht. Zu den *in vivo* vorkommenden Sequenzen wurden auch mutierte Versionen produziert, die über die eventuell factorspezifischen DNA Unterschiede verfügten. Es wurde jeweils auch eine Variante genutzt, die gegebener Weise die Bindung beider Faktoren verloren hat. All diese Sequenzen sind in den folgenden Tabellen zusammengefasst. Die Oligonukleotide wurden jeweils nach dem nächstgelegenen Gen benannt. Beispielhaft ist im Folgenden, mit Bcan, einer dieser genomischen Bereiche dargestellt.



**Abb. 59: Genomischer Bereich in der Nähe des Gens Bcan zeigt eine Bindung von CTCF und CTCFL über der Motifsequenz CCAGCAGAGGGCGC.** Dargestellt ist ein Ausschnitt des Genbereiches von Bcan. Exons sind als blaue Blöcke angegeben, der translatierte Bereich jeweils als dickerer Block. Darunter sind neben bekannten CTCF-Motiven und der identifizierten Sequenz CCAGCAGAGGGCGC +/-18 bp jeweils die Bindeprofile von CTCF und CTCFL in K562 Zellen angegeben. Bei diesen Profilen handelt es sich um die öffentlichen Daten des ENCODE Konsortiums.

Aufgrund von zeitlicher Limitierung konnten die *in vitro* Experimente nicht mehr durchgeführt werden und wurden in die Hände eines Kollegen übergeben. Im Folgenden sind die identifizierten Sequenzen und erwartete *in vitro* Bindung in einer Tabelle zusammengefasst.

**Tab. 9: Identifizierte Sequenzen und Hypothese bezüglich CTCF und CTCFL Bindung.** Dargestellt sind neben der Ausgangssequenz zwei mutierte Formen für die ein verändertes Bindevverhalten von CTCF und CTCFL erwartet wird. Diese Erwartung ist in der rechten Spalte angegeben. Veränderte Basen sind in dick und rot angegeben.

Sequenz	Erwartetes Bindevverhalten
CCAGCAGAGGGCGC	Bindung von CTCF und CTCFL
<b>TCAGTAGAGGGCGC</b>	Bindung von CTCF
<b>TCAGTAGAGGGGCGC</b>	Keine Bindung

Somit kann im Zuge der bioinformatischen Untersuchungen eine mögliche Sequenz-Spezifität von CTCF und CTCFL identifiziert werden. Diese muss nun in folgenden Experimenten entsprechend überprüft werden.



## 5 Diskussion

Bei CTCF und CTCFL handelt es sich um zwei sehr nah verwandte Proteine, wobei CTCFL auf eine Genduplikation von CTCF im Laufe der Evolution zurückgeht (Hore et al. 2008). Während der Wissenstand über CTCF inzwischen sehr reichhaltig ist sind genaue Funktionen und Wirkweisen von CTCFL weiter ungeklärt. Aufgrund der identischen DNA-Bindedomäne der beiden Proteine geht man von ähnlichen genomischen Bindestellen aus, während die stark variierenden N- und C-terminalen Domänen auf unterschiedliche Funktionen und Interaktionspartner schließen lassen. Weiterhin kann aufgrund der gemeinsamen DNA-Bindedomäne die Möglichkeit der Konkurrenz der beiden Proteine um einzelne Bindestellen nicht ausgeschlossen werden. Lange wurde für beide Proteine im normalen Organismus von sich ausschließenden Expressionsmustern ausgegangen, d.h. man vermutete, dass CTCFL die Rolle von CTCF in einigen Zelltypen der Keimbahn übernimmt in denen CTCF selbst nicht exprimiert wird. Neuere Untersuchungen konnten dann jedoch belegen, dass auch im normalen Organismus Expression beider Proteine in der gleichen Zelle vorliegen kann und diese Art der gleichzeitigen Expression nicht alleine auf abberante Krebszellen beschränkt ist (Sleutels et al. 2012). Über die Rolle von CTCFL in Krebszellen besteht bis heute weiterhin Unklarheit. Während einige Veröffentlichungen (Novak Kujundžić et al. 2014; Mkrtychyan et al. 2011) und auch die Lokalisation des CTCFL-Gens in einem Tumor Hot-Spot für eine mögliche Rolle von CTCFL bei Krebs-Entstehung und –Progression sprechen, können andere Arbeitsgruppen keinen Zusammenhang und so gut wie keine Expression von CTCFL in Krebs erkennen (Hines et al. 2010; Yoon et al. 2011). Eigene Analysen zur Expression von CTCFL in humanen und murinen Zelllinien belegen eher den Befund der seltenen/nicht vorhandenen CTCFL-Expression (nicht gezeigt). Um genaueren Aufschluss über die funktionelle Rolle von CTCFL, vor allem auch im Vergleich zu CTCF, zu erhalten wurden CTCFL-Überexpressionsexperimente und auch Korrelationsanalysen mit öffentlich verfügbaren Datensätzen durchgeführt.

### 5.1 Versuch der Etablierung stabiler, induzierbarer CTCFL-Expression im humanen System

Zur Untersuchung der funktionellen Wirkung von CTCFL im humanen System sollte stabile, induzierbare Expression in humanen Zellen erzielt werden. Als sinnvolle Zelllinie, da leicht zu transfizieren und kultivieren, bot sich hier die weit verbreitete Linie HeLa

an. HeLa Zellen exprimieren im Normalzustand kein CTCFL (Renaud et al. 2011). Nach Transfektion der Zellen mit einem auf zwei Vektoren basierenden System zur Erzeugung stabiler Klone mit induzierbarer Expression des gewünschten Gens, in diesem Fall humanes CTCFL, das mit einem FLAG-Tag versehen war, erfolgte die entsprechende Selektion mit Puromyzin und mehrere einzelne Klone konnten selektioniert werden. Zwei von 39 Klonen zeigten eine induzierbare Expression von CTCFL auf RNA Ebene. Diese Klone zeichneten sich auch durch niedrige Expression von CTCFL im uninduzierten Zustand, aus und die Induktion lag in der Größenordnung 25- bis 30-fach. Durch den Einsatz von geeigneten Antikörpern (ab18337 von Abcam & H120 von Santa Cruz Biotechnology) konnte im Western Blot transient überexprimiertes CTCFL klar detektiert werden. Unter Verwendung des einen Antikörpers (ab18337), der gegen den N-Terminus von CTCFL gerichtet ist, konnte kein CTCFL Protein in den induzierten Klonen nachgewiesen werden, der weitere Antikörper (H120) der gegen einen internen Bereich von CTCFL gerichtet ist detektierte spezifisches induzierbares Protein, das jedoch eine niedrigere Laufhöhe im SDS-Gel aufwies. Im Allgemeinen ist die Versorgungslage mit gut abgesicherten CTCFL Antikörpern schlecht und es mussten eine Vielzahl von Antikörpern und Bedingungen ausprobiert werden bevor reproduzierbare Ergebnisse vorlagen, ein Problem von dem auch weitere Arbeitsgruppen berichten (Woloszynska-Read et al. 2011). Da auch der Western Blot gegen das N-terminal-ständige FLAG-Tag in den Klonen negativ verlief ist von einer Kappung eines Teils des N-Terminus von CTCFL auszugehen. In weiteren, in dieser Arbeit nicht gezeigten, Experimenten wurde die in den Klonen stabil integrierte DNA für FLAG-CTCFL mittels PCR überprüft und es konnte festgestellt werden, dass auf dem Level der DNA diese Verkürzung nicht vorlag. Somit muss diese Verkürzung des Proteins durch einen nicht geklärten Mechanismus nach der Transkription erfolgen. Da die Art des Mechanismus jedoch keine Auswirkung auf das Ergebnis, nämlich, dass diese Klone nicht für weitere sinnvollen Experimente eingesetzt werden können, hat, wurden keine weiteren Experimente mit diesen Klonen durchgeführt.

## **5.2 Untersuchungen zu CTCFL im humanen System (K562)**

Um weiter im humanen System zu arbeiten und nach dem Fehlschlag der induzierbaren Klone dennoch Ergebnisse zur Wirkweise von CTCFL im Menschen zu erhalten, wurden die gut untersuchten K562 Zellen herangezogen. Bei diesen Zellen handelt es sich um eine Leukämie-Linie die relativ einfach in Suspension kultiviert werden kann und für die robuste Expression von CTCFL beschrieben ist und auch genom-weite Bindemuster von

CTCFL wurden in K562 Zellen im Rahmen des ENCODE Projektes erhoben. Unter diesen Voraussetzungen wurden einerseits ein CTCFL-*Knockdown* etabliert und andererseits mit den erwähnten genom-weiten Daten des ENCODE Konsortiums eine breit angelegte Korrelationsanalyse durchgeführt.

### **5.2.1 .Reduzierte CTCFL-Expression hat einen Effekt auf ausgesuchte Kandidaten-Gene**

Die Reduktion der CTCFL Expression erfolgte mittels einer Mischung zweier spezifischer siRNAs gegen CTCFL. Die siRNAs zeigten die beste Wirkung wenn die K562 Zellen zweimal an aufeinanderfolgenden Tagen transfiziert wurden, so erhielt man eine 50 %-ige Abnahme der CTCFL mRNA und auf Proteinebene war CTCFL unter Verwendung des spezifischen Antikörpers von Abcam gar nicht mehr nachzuweisen. Um mögliche beeinflusste Zielgene von CTCFL zu identifizieren ohne eine genom-weite Analyse durchführen zu müssen, wurden die humanen Homologe von bekannten murinen, von CTCFL regulierten, Genen (S. Dienstbach, Dissertation) untersucht. Hierzu wurden drei Gene (Gpr56, Gstm5 & Akr1c18) ausgesucht die die stärkste Deregulation in Folge von CTCFL Überexpression gezeigt hatten. Bei den human Homologen handelte es sich um GPR56 (2015 im Zuge einer Neuordnung der Gruppe von G-Protein gekoppelten Rezeptoren umbenannt in ADGRG1 (Hamann et al. 2015)), GSTM5 und AKR1C3. Von diesen drei Genen zeigten GPR56 und GSTM5 eine Deregulation in den K562 Zellen nach CTCFL-*Knockdown*, während AKR1C3 nicht beeinflusst war. Die Richtung der Deregulation war für beide beeinflussten Gene genau gegensätzlich zu den nach CTCFL-Überexpression betrachteten Effekten. So zeigte GPR56 in K562 Zellen eine ca. 40 %-ige Abnahme der Expression während Gpr56 in Maus-Zellen (NIH3T3) nach CTCFL Induktion 13,7-fach induziert wurde (S. Dienstbach, Dissertation). CTCFL scheint somit ein direkter positiver Regulator der beiden Gene, GPR56 und GSTM5 zu sein. GPR56 ist im Zusammenhang mit CTCFL weiterhin interessant, da eine wichtige funktionelle Rolle von GPR56 bei der Testis-Entwicklung beschrieben ist (Chen et al. 2010). Bei Betrachtung öffentlicher Bindedaten von CTCFL im Bereich des GPR56 Gens zeigt sich eine starke Bindestelle im intronischen Bereich vor dem ersten translatierten Exon des Gens. Somit lässt sich über eine Rolle der CTCFL Bindung in diesem Bereich bei der Aktivierung von GPR56 spekulieren.

### 5.2.2 Korrelationsanalysen zu CTCFL anhand der ENCODE Datenbanken

Die umfangreichen Datenbanken des ENCODE-Konsortiums mit genom-weiten Datensätzen zu vielen DNA-bindenden Faktoren und Histon-Modifikationen ermöglichen groß angelegte Korrelationsanalysen mit CTCF und CTCFL Bindestellen, da auch hier Datensätze für K562 Zellen vorliegen. Im ersten Schritt wurden die CTCF und CTCFL Bindestellen bezüglich ihrem Grad der Überlappung analysiert und dann die einzelnen Bindestellen jeweils Untergruppen zugeordnet. So zeigen in K562 Zellen 70,7 % der CTCFL Bindestellen CTCF Bindung, umgekehrt sind nur 24,5 % der CTCF Bindestellen von CTCFL gebunden. Diese Diskrepanz lässt sich durch die generell niedrigere Zahl an identifizierten CTCFL Bindestellen erklären. Bei den jeweiligen Untergruppen handelte es sich um alle CTCF Bindestellen, alle CTCFL Bindestellen, alle von beiden Faktoren gebundenen Bindestellen (CTCF + CTCFL), nur CTCF gebundene Bindestellen und nur CTCFL gebundene Bindestellen. Über diese Gruppen wurde jeweils die Anreicherung von bestimmten Transkriptionsfaktoren und Histon-Modifikationen untersucht. Dabei konnten viele aktive Histon-Modifikationen (z.B. H3K9Ac & H3K4me3) und Transkriptionsfaktoren (z.B. HDAC2, POLII & YY1) identifiziert werden, die eine verstärkte Anreicherung an den Gruppen CTCF+ CTCFL und nur CTCFL gebunden, und somit generell für CTCFL Bindestellen zeigten. Eher reprimierende Histon-Modifikationen (z.B. H3K27me3) zeigten keine spezifische Anreicherung für eine der analysierten Gruppen von Bindestellen. Somit zeigt CTCFL im Vergleich zu CTCF bevorzugt Bindung an Bereiche aktiver Transkription die sich auch durch eine höhere Zugänglichkeit der entsprechenden DNA auszeichnet (Sleutels et al. 2012). Auch weitere Publikationen anderer Arbeitsgruppen die im Zeitraum dieser Arbeit veröffentlicht wurden zeigten die Präferenz von CTCFL für offenes Chromatin (Pugacheva et al. 2015). Im Zuge der Analyse konnten nur die Komponenten des Cohesin-Komplexes RAD21 und SMC3 als Faktoren identifiziert werden die spezifische Anreicherung für die CTCF gebundenen Bindestellen aufweisen. Bindestellen die CTCFL alleine aufweisen zeigen die niedrigsten Anreicherungs-Werte für diese beiden Faktoren. Da CTCF mit dem Cohesin Komplex über die C-terminal vermittelte Bindung an SA2 interagiert (Xiao et al. 2011), ist durch die stark variierenden C-Termini von CTCF und CTCFL davon auszugehen, dass CTCFL wenn es alleine bindet nicht die Möglichkeit besitzt mit SA2 zu interagieren. Somit findet sich auch keine Korrelation mit den anderen Komponenten des Cohesin-Komplexes. Zusammengefasst konnte durch die Korrelationsanalyse die starke Präferenz von CTCFL für offene, zugängliche Bereiche des Genoms, im Vergleich zu CTCF belegt werden.

### 5.3 Untersuchungen zu CTCFL Bindestellen in NIH3T3 Zellen

Von S. Dienstbach wurde bereits im Rahmen seiner Dissertation ein muriner Zellklon (Klon 34) mit stabiler, induzierbarer mCTCFL Expression etabliert und analysiert. So wurden etwa auch durch ChIP-Sequenzierung die genom-weiten Bindestellen von CTCF und CTCFL in diesem Klon identifiziert. Viele weitergehende Analysen wurden mit diesem Klon und den Datensätzen jedoch noch nicht durchgeführt, daher sollten im Rahmen dieser Arbeit versucht werden anhand der verfügbaren Daten und neuer Experimente Hinweise auf mögliche CTCFL Funktionen zu erhalten. Im speziellen sollte der Einfluss von CTCFL auf CTCF Bindung, Anwesenheit von Histon-Modifikationen und Cohesin Bindung untersucht werden.

#### 5.3.1 Keine detektierbare CTCF-CTCFL Konkurrenz im Klon 34

Zur Untersuchung der Auswirkung von CTCFL auf CTCF an gemeinsamen Bindestellen wurde bioinformatisch untersucht wie viele ursprüngliche CTCF Bindestellen die in NIH3T3 Wildtyp Zellen identifiziert wurden eine differentielle CTCF Bindung nach Überexpression von CTCFL aufwiesen. Nach eingehender entsprechender Filterung der Daten konnten 176 Bindestellen mit differentieller CTCF Bindung identifiziert werden. Von diesen zeigten 91 eine niedrigere Bindung von CTCF, 85 aber auch eine Zunahme von CTCF Bindung. Somit kann in dieser Analyse keine generelle Konkurrenz von CTCFL mit CTCF an geteilten Bindestellen festgestellt werden, da an fast 50% der differentiellen CTCF Bindestellen CTCFL Bindung sogar zu einem verstärkten CTCF Signal führt. Da die Zahl identifizierten differentieller CTCF Bindestellen im Vergleich zur Anzahl der gesamten CTCF und CTCFL Bindestellen auch sehr niedrig ist kann es sich bei den Schwankungen auch um unabhängige klonale Schwankungen oder Sequenzierungs-Artefakte handeln. Zur Überprüfung einiger Bindestellen in unabhängigen Experimenten wurden drei DNA-Bereiche ausgewählt (Accs, Rab711 & Col12a1) und überprüft. So wurden in biologischen Triplikaten ChIPs für CTCF und CTCFL in uninduziertem und induziertem Klon 34 durchgeführt, und es konnte für keine der drei untersuchten Bindestellen die erwartete Abnahme der CTCF Bindung an diesen Bereichen festgestellt werden. Während die höhere Bindung von CTCFL im induzierten Klon jeweils klar erkennbar war änderte sich die CTCF Bindung nicht, oder nahm sogar leicht zu. Da die Analyse individueller Bindestellen keinen Hinweis auf *in vivo* Konkurrenz zwischen CTCF und CTCFL lieferte wurde in einem abschließendem Experiment versucht bioinformatisch doch den, evtl. auch sehr schwachen, Trend der Konkurrenz zu identifizieren.

Hierzu wurden die CTCF Bindestellen in zwei Gruppen eingeteilt, zum einen Bindestellen die im induzierten Klon zusätzliche CTCFL Bindung aufweisen und zum anderen alle übrigen Bindestellen die sowohl in Wildtyp-Zellen als auch im Klon nur Bindung von CTCF alleine aufweisen. Über diese Gruppen an Bindestellen wurden dann die CTCF Signale im Wildtyp und induzierten Klon in Boxplots zusammengefasst um schwache Abweichungen zu erkennen. Im Falle einer Kompetition von CTCFL hätte man erwartet, dass die CTCF Signale über den geteilten Bindestellen im induzierten Klon schwächer wären. Die Boxplots zeigten jedoch keine Unterschiede der CTCF Bindestärke über die Klassen. Somit konnte zusammengefasst auch in den individuellen Experimenten und einer genom-weiten Auswertung die anhand der ChIPseq Daten detektierte Kompetition nicht bestätigt werden und es ist von Schwankungen bei der Sequenzierung auszugehen.

### 5.3.2 CTCFL zeigt keinen Einfluss auf Histon-Modifikationen

Um einen möglichen Einfluss von CTCFL Bindung auf die Anwesenheit von Histon-Modifikationen zu untersuchen wurden ChIPs mit Antikörpern gegen zwei aktive Modifikationen (H3K4me3 & H3K9ac) und zwei repressive Modifikationen (H3k27me3 & H3K9me3) durchgeführt und ausgewertet. Untersuchte genomische Bereiche waren auch in diesem Fall Accs, Rab711 und Col12a1, da die jeweiligen mit den Bindestellen assoziierten Gene eine Deregulation in Folge der CTCFL Expression zeigten. Im Zuge der durchgeführten qPCRs konnte jedoch an keiner der drei Bindestellen eine Änderung der Histon-Modifikation-Muster oder jeweiligen Bindestärke festgestellt werden. Wenn es relativ zum ChIP Input teilweise so aussah als ob Änderungen vorlagen, so führte eine nötige Normalisierung der Daten über H3 ChIP-Signal dazu, dass diese Effekte verschwanden. Somit konnte an drei CTCFL-Bindestellen kein Einfluss von CTCFL auf Anwesenheit von vier untersuchten Histon-Modifikationen festgestellt werden. Im Zuge der anderen Ergebnisse aus K562 Zellen drängt sich das Bild auf, dass CTCFL offene Chromatin Bereiche zur Bindung benötigt und nicht, dass durch CTCFL Bindung Bereiche geöffnet werden können. Somit verhält sich CTCFL wie der Großteil der bekannten Transkriptionsfaktoren, für die im Rahmen der ENCODE Experimente eine durchschnittliche Assoziation von 98,2 % ihrer Bindestellen mit offenem Chromatin gezeigt wurde (Thurman et al. 2012).

### 5.3.3 CTCFL zeigt keinen Zusammenhang mit Cohesin-Komponenten

Um einem möglichen wichtigen Einfluss von CTCFL auf die Bindung des Cohesin Komplexes an CTCF Bindestellen nachzugehen wurden ChIP Experimente mit spezifischen Antikörpern gegen RAD21, SA1 und SA2 durchgeführt. Sollte CTCFL einen negativen Einfluss auf Cohesin Bindung haben wären möglicherweise wichtige strukturelle Domänen im Genom betroffen und es wäre mit massiven Effekten auf Genexpressionsmuster zu rechnen. Neben den drei bekannten Bindestellen Accs, Rab711 und Col12a1 wurden vier weitere doppelt gebundene Bindestellen zur qPCR Analyse herangezogen. Für alle konnte bei Darstellung gegen den ChIP Input eine Abnahme der Cohesin Bindung im induzierten Klon beobachtet werden. Verrechnet man diese Signale jedoch gegen die jeweiligen unspezifischen IgG Präzipitationen, eine Art der Normalisierung die vor allem bei stark abweichenden IgG Werten vorzuziehen ist, war diese Abnahme in den meisten Fällen nicht mehr zu beobachten. Eine Bindestelle (Adnp2) zeigte weiterhin eine Abnahme aller drei Faktoren im induzierten Klon, diese Bindestelle war jedoch generell nur sehr schwach von den Cohesin Komponenten gebunden und somit könnte diese Abnahme auch durch einfache Schwankung der experimentellen Durchführung und/oder Variation der PCRs erklärt werden. Ein genereller Einfluss von CTCFL auf Anwesenheit von Cohesin an CTCF Bindestellen ließ sich somit nicht beweisen. Ein Befund der inzwischen von anderen Publikationen bestätigt wird (Pugacheva et al. 2015).

## 5.4 CTCFL Überexpression im Differenzierungssystem

Da die experimentellen Ergebnisse der CTCFL Funktion im NIH3T3 Klon keine klaren Ergebnisse lieferten wurde ein weiteres stabiles Expressionssystem etabliert. Diesmal wurden murine P19 Zellen gewählt, da sich diese durch ihre Nähe zu embryonalen Stammzellen auszeichnen, im Gegensatz zu diesen aber weit komplikationsfreier kultivieren lassen. P19 zeichnete sich in selbst durchgeführten Experimenten immer durch keine detektierbare Expression von CTCFL aus, auch wenn eine etwas zweifelhafte Veröffentlichung von CTCFL Expression in P19 Zellen berichtet (Soltanian et al. 2014). Durch Zugabe von Retinsäure (RA) lassen sich P19 Zellen auch schnell und einfach differenzieren und somit ergeben sich neue Möglichkeiten zur Beantwortung CTCFL-spezifischer Fragestellungen. Zur Erzeugung stabiler Überexpression von CTCFL kam das gleiche Zwei-Vektoren-System zum Einsatz, das schon bei der Erzeugung der NIH3T3 Klone genutzt wurde. Nach Transfektion und Selektion mit Puromycin konnten mehrere Klone selektiert werden. Von diesen zeigte nach eingehender Charakterisierung nur ein einziger

eine detektierbare induzierbare Expression von CTCFL. Diese Expression konnte sowohl auf RNA- als auch auf Protein-Ebene demonstriert werden. Die Induktion von CTCFL lag mit einem Wert von ca. 30-fach in einer ähnlichen Größenordnung wie bei dem NIH3T3 Klon, wohingegen die gemessene Hintergrund-Aktivität von CTCFL im P19 Klon größer war. Somit zeigte der uninduzierte P19 Klon auch eine höhere CTCFL Expression als bei vorherigen Klonen beobachtet. Zusammengefasst konnte jedoch ein neues stabiles Expressionssystem in P19 Zellen etabliert werden und dieser bot sich durch den möglichen Vergleich zu P19 Wildtyp Zellen trotz der festgestellten Hintergrund-Aktivität, die im Übrigen auf Protein-Ebene nicht so dramatisch detektiert werden konnte, für weitere Experimente an.

#### 5.4.1 Einfluss von CTCFL auf genom-weite Expression

Um die Wirkung von CTCFL auf die Expression von Zielgenen zu erfassen wurden in dem P19 Klon genom-weite Expressionsanalysen mittels *Microarray* durchgeführt. Es wurden jeweils nach den entsprechenden erfolgreichen Kontrollexperimenten sowohl von P19 Wildtyp Zellen als auch von induzierten Klon 16 Zellen undifferenzierte und mittels RA differenzierte RNA Proben analysiert. Durch qRT-PCR Experimente konnte auch die erfolgreiche Differenzierung der Zellen anhand von bekannten Differenzierungsmarkern validiert werden. Im Zuge der genom-weiten Auswertung konnten spezifische Zielgene von CTCFL identifiziert werden. Wurden die jeweiligen Ergebnisse der P19 Wildtyp Zellen und der induzierten Klon 16 Zellen zusammengefasst und verglichen so konnte nur eine kleine Zahl von 82 deregulierten Genen identifiziert werden. Von diesen 82 Genen war mit 71 Genen der Großteil induziert was für eine Rolle von CTCFL als Aktivator spricht. Herabreguliert wurden in diesem Zusammenhang nur 11 Gene. Eine Durchführung von *gene ontology* Analysen, zur Identifizierung bestimmter regulierter Untergruppen von Genen oder zur Identifizierung von beeinflussten Signalwegen bot sich aufgrund der kleinen Zahl an deregulierten Genen hier nicht an. Die Rolle von CTCFL bei der Aktivierung von Zielgenen wird durch CTCFL *knockout* Daten aus *in vivo* Experimenten gestützt, so fand eine Arbeitsgruppe in Folge der Herabregulierung in Testis 3 signifikant deregulierte Gene und alle drei zeigten die gleiche Tendenz wie CTCFL (Suzuki et al. 2010). Es ist jedoch zu erwähnen, dass es auch gegensätzliche Befunde gibt, so zeigten neuste Daten zur Deregulation von Genen in Folge von CTCFL Herabregulierung in K562 Zellen ein ausgeglichenes Verhältnis zwischen induzierten und reprimierten Genen (Pugacheva et al. 2015). Interessanterweise konnte in der Gruppe der induzierten Gene



das bereits bekannte GPR56 identifiziert werden, womit ein weiterer Hinweis auf einen funktionellen Zusammenhang in der Regulation von GPR56 durch CTCFL besteht. Im Zuge der Differenzierung der Zellen konnten anhand bioinformatischer Auswertung 623 deregulierte Gene identifiziert werden. Von diesen zeigte ebenso die Mehrheit (458) eine Induktion und nur ein kleinerer Teil (174) Herabregulierung. Bei den deregulierten Genen handelte es sich um bereits bekannte Differenzierungs-assoziierte Gene in P19 Zellen (Sérandour et al. 2012). Hier lässt sich die höhere Anteil an heraufregulierten Genen eventuell mit den komplexeren Anforderungen der differenzierten Zellen an die Expression bestimmter differenzierungsspezifischen Gene erklären. Andere Arbeitsgruppen konnten im Zusammenhang mit anderen Formen von Differenzierung auch jeweils mehr induzierte als reprimierte Gene identifizieren (Arenzana et al. 2015). Führte man an Hand der deregulierten Gene eine *gene ontology* Analyse durch, konnten erwartete Gengruppen, wie z.B. Gene der embryonischen Organ-Entwicklung in induzierten Genen oder Gene der Blastocysten Entwicklung in reprimierten Genen identifiziert werden (Nicht gezeigt). Im Zuge der Analyse, ob möglicherweise CTCFL Expression einen Einfluss auf die Expression von Differenzierungs-Genen hat, konnten zwar eine kleine Zahl an möglichen beeinflussten Genen identifiziert werden (z.B. Brachyury), für diese gelang jedoch keine unabhängige Bestätigung in Wiederholungs-Experimenten (Nicht gezeigt). Speziell im Fall von Brachyury könnte sich die initiale Identifikation durch spezifische Fluktuation von  $\beta$ -Catenin einem Regulator von Brachyury, die in ES-Zellen beobachtet werden kann, erklären (Marucci et al. 2014). Somit konnten in P19 Zellen Zielgene von CTCFL identifiziert jedoch kein klarer Einfluss von CTCFL auf Genexpressionsmuster der Differenzierung gefunden werden. Diese Erwartung bestand jedoch auch nicht unbedingt, da bei der Kultivierung der P19 Zellen kein CTCFL Einfluss auf den Phänotyp oder die Wachstum-Geschwindigkeit der differenzierenden Zellen zu beobachten war.

#### **5.4.2 Identifikation von CTCFL Bindestellen im P19 Klon**

Zur Untersuchung der möglichen Modulierbarkeit von CTCFL Bindemustern durch Veränderung des Chromatins wurde anhand von ChIP-Sequenzierung die genom-weite Bindung von CTCFL und auch CTCF in undifferenzierten und differenzierten, induzierten Klon 16 Zellen erhoben. Zur Etablierung der CTCFL ChIP-Methode in P19 Zellen unter Verwendung des Sleutels6 Antikörpers diente eine bereits in NIH3T3 Zellen identifizierte Bindestelle im Bereich des Snapin Gens. Diese Bindestelle zeigte auch in allen durchgeführten Experimenten nur sehr kleine Schwankungen für CTCFL Präzipitation, auch bei

weitergehenden Behandlungen der Zellen wie etwa durch Differenzierung. Daher wurde sie in einigen Experimenten auch zur Normalisierung der CTCFL Bindung zwischen variierenden Replikaten der ChIPs herangezogen. Für die Sequenzierung der ChIP DNA gab es die Mindestmengen-Anforderung von 10 ng und zum Erreichen dieser Menge war es im Fall von CTCFL nötig über 10 IPs zu vereinigen und dann im Volumen zu reduzieren. Die eigentliche Sequenzierung der DNA erfolgte am HZI in Braunschweig. Die bioinformatische Bearbeitung der Rohdaten wurde in Kooperation mit M. Bartkuhn durchgeführt. Im Zuge der Sequenzierung konnten jeweils die genom-weiten Bindestellen für CTCF und CTCFL identifiziert werden. Der erste Blick auf die Zahl der identifizierten Bindestellen zeigte für beide Faktoren jeweils eine höhere Anzahl in den undifferenzierten Klon Zellen, für CTCF 45.025 zu 29.426 Stellen und für CTCFL 15.816 zu 2.536. In den jeweiligen Proben des differenzierten Klons wurde weniger DNA nach Braunschweig übersendet, was möglicherweise einer der Gründe für die niedrigere Komplexität der Datensätze des differenzierten Klons sein kann. Bei Betrachtung der paarweisen Überlappung von CTCF und CTCFL zeigte CTCF eine größere Übereinstimmung der Bindestellen in Abhängigkeit der Differenzierung als CTCFL. Bei Betrachtung der Überlappung von CTCF und CTCFL innerhalb eines Zelltyps zeigten im undifferenzierten Klon 59 % der CTCFL Bindestellen auch Bindung von CTCF, im differenzierten Klon etwas höhere 68 %. Zum Vergleich können hier die humanen K562 Daten herangezogen werden die mit einer CTCFL-CTCF Überlappung von 70,7 % einen ähnlichen Grad der Überlappung zeigen. Insgesamt ist somit für die vier gewonnen Datensätze eine Variabilität der Bindestellen detektierbar und es konnten nur 893 Bindestellen gefunden werden die stabile Bindung beider Faktoren in beiden untersuchten Zell-Zuständen zeigten.

#### **5.4.3 Vergleichende Analysen zwischen NIH3T3 und P19 Klonen**

Um anhand möglicher Zelltyp-spezifischer CTCFL Bindemuster Hinweise auf funktionelle Zusammenhänge zu untersuchen wurden die neu erhobenen Bindestellen von P19 mit den bereits bekannten Datensätzen aus NIH3T3 Zellen systematisch verglichen. Für diesen Vergleich wurden die undifferenzierten P19 Klon Daten verwendet da sie über mehr identifizierte Bindestellen, sowohl für CTCF als auch CTCFL verfügen. Im Zuge der Analysen konnte die Überlappung der Bindestellen dargestellt und berechnet werden. So betrug der Grad der Überlappung für CTCF 57 % der in P19 Zellen gebundenen Bindestellen und somit nur minimal geringer als die 60 % identische Bindestellen zwischen

differenzierten und undifferenzierten P19 Zellen. Diese Werte bewegen sich im vergleichbaren Bereich zu veröffentlichten Werten für CTCF Bindestellen in unterschiedlichen Zelltypen, so konnten beispielsweise 67 % der in IMR90 Zellen identifizierten CTCF Bindestellen auch in U937 Zellen gefunden werden (Kim et al. 2007). Für CTCFL lag der Prozentsatz beobachteter gemeinsamer Bindestellen zwischen P19 und NIH3T3 Zellen mit 36 % im Vergleich zu CTCF niedriger. Auch im Vergleich zur CTCFL Überlappung zwischen den zwei Zuständen des P19 Klonen waren diese Zahlen reduziert. Somit zeigt CTCFL die größere Varianz bei Betrachtung zelltyp-spezifischer Bindemuster. Eine mögliche Erklärung liefert hier die im Vergleich zu CTCF nicht bekannte Beteiligung von CTCFL an der Ausbildung wichtiger struktureller DNA-Kontaktpunkte z.B. bei der Bildung von TADs. Für diese TADs ist nur wenig Variation zwischen Zelltypen beschrieben und somit bleibt auch der Teil des strukturell wirkenden CTCFs in seiner DNA Lokalisierung stabil (Pope et al. 2014).

Bei der Betrachtung der Verteilung der CTCF und CTCFL Bindestellen in beiden Zelltypen relativ zu genomischen Merkmalen konnten vergleichbare Verteilungsmuster identifiziert werden. Sowohl in P19 als auch in NIH3T3 Zellen zeigte CTCFL die größte Anreicherung im Bereich von TSSs und diese Präferenz war jeweils noch im Vergleich zu CTCF erhöht. Somit zeigt CTCFL trotz der eher niedrigen Überlappung der Bindestellen ein vergleichbares Verteilungsmuster zwischen den untersuchten Klonen. Somit lässt sich spekulieren, dass CTCFL aufgrund differentieller Zugänglichkeit des Chromatins, möglicherweise vermittelt durch differentielle Genexpressionsmuster, an unterschiedlichen Bereichen bindet. Um diese Hypothese zu überprüfen wurden öffentliche Bindedaten für die aktive Histon-Modifikation H3K4me2 genutzt um eine Korrelationsanalyse mit beobachteter CTCFL Bindung durchzuführen. Hierzu wurden jeweils die spezifisch differentiell gebunden CTCFL Bindestellen zusammengefasst und die kumulative Bindung der Modifikation über diese Gruppen in beiden Zelltypen dargestellt. So konnte für die zelltyp-spezifischen Bindestellen differentielle Bindung von H3K4me2 detektiert werden, CTCFL gebundene Stellen zeigten jeweils in ihren Zelltypen eine höhere Anwesenheit der Modifikation als an den ungebundenen Bereichen im anderen Zelltyp. Um diesen Befund differentieller Zugänglichkeit an zelltyp-spezifischen CTCFL Bindestellen weiter zu untermauern wurden zusätzlich vergleichende FAIRE Experimente durchgeführt. FAIRE erfasst unabhängig von bestimmten Modifikationen oder Faktoren die Zugänglichkeit von DNA und bot sich somit als Technik an. Ein Problem der zuvor durchgeführten Analysen hatten vermutlich auch die unterschiedlichen Quellen der H3K4me2

Daten gespielt, die in unterschiedlichen Arbeitsgruppen mit unterschiedlichen Antikörpern erhoben wurden. Somit wurde FAIRE sowohl für den P19 Klon als auch den NIH3T3 Klon parallel in drei biologischen Replikaten durchgeführt um die experimentelle Varianz so gering wie möglich zu halten. Anhand der genom-weiten CTCFL Binde-daten wurden jeweils drei zelltyp-spezifische Bindestellen ausgesucht und dann mittels FAIRE-qPCR überprüft. Jeweils zwei der drei ausgesuchten Bindestellen zeigten ein sehr klares differentielles FAIRE-Muster, dass die CTCFL Verteilung nachbildete. Nur die zwei Bindestellen *Lrrc41* und *Kremen2* zeigten diesen Effekt nicht, wobei *Kremen2* in beiden Zelltypen als sehr schwach zugänglich detektiert wurde. Somit konnte anhand der FAIRE Daten für die Mehrheit der analysierten Bindestellen eine differentielle Zugänglichkeit des Chromatins detektiert werden die möglicherweise eine Rolle bei der Ausbildung der beobachteten differentiellen CTCFL Bindung spielt.

#### **5.4.4 Analyse differentieller Chromatin Zugänglichkeit in Bezug auf CTCFL**

Um der Frage nachzugehen, ob die beobachtete Korrelation von CTCFL Bindung mit zugänglichem Chromatin auch innerhalb eines Zelltyps, bei weitreichender Änderung des Chromatin Status, bestehen bleibt, wurden genomweite FAIRE Analysen im P19 Klon durchgeführt. Um den Grad der höchsten Veränderung des Chromatins durch Zugabe von RA zu erreichen erfolgte die Differenzierung diesmal in einem längeren Zeitraum von 9 Tagen, da die Auswertung öffentlicher Genexpressionsprofile nach diesem Zeitraum die höchste Zahl an deregulierten Genen zeigte (Suzuki et al. 2011). Um eine bessere statistische Auswertbarkeit der Daten zu erreichen wurde die Sequenzierung in biologischen Replikaten für den undifferenzierten und differenzierten Klon durchgeführt. Im Zuge der Sequenzierung, die vor der Durchführung entsprechend kontrolliert wurde, konnten DNA Bereiche mit differentieller Zugänglichkeit identifiziert werden. Zur Validierung der Sequenzier-Ergebnisse wurden einige Bereiche differentieller Zugänglichkeit ausgewählt und in unabhängigen Experimenten überprüft (Daten nicht gezeigt). So konnte beispielsweise im Promotor Bereich des wichtigen Pluripotenz-Gens *Pou5f1* (auch *Oct4*) (Niwa et al. 2000) abnehmende Zugänglichkeit der DNA, die mit einer Herabregulierung des Gens einhergeht, beobachtet werden. Zur Analyse der Auswirkung der Änderungen der Chromatin Offenheit auf CTCFL Bindung wurden neun DNA Bereiche ausgewählt, die differentielles FAIRE Signal sowie Bindung von CTCFL in undifferenzierten Klon Zellen zeigten. Eine genom-weite Auswertung konnte hier aufgrund von fehlender CTCFL

Bindedaten für die Langzeit-Differenzierung nicht durchgeführt werden. Von den neun gewählten Bindestellen konnte für sieben die differentielle Zugänglichkeit auch in unabhängigen FAIRE Experimenten bestätigt werden. Zwei Bindestellen zeigten keine Änderung (Ptpu) oder generell keine feststellbare Zugänglichkeit (Nkx3-2). Die sieben bezüglich FAIRE validierten Bindestellen wurden dann in ChIP Experimenten auf ihre CTCFL Bindung hin untersucht. In diesem Zuge konnten zwei Bindestellen (Ptprg & Mzfl) mit differentieller CTCFL Bindung in Abhängigkeit der Zugänglichkeit identifiziert werden. Für die anderen Bindestellen konnte zwar CTCFL Bindung beobachtet werden, diese zeigte jedoch keinen Effekt im Zuge der Differenzierung. Die Möglichkeit der umgekehrten Suche, also Bindestellen zu analysieren die zunehmende Offenheit und einhergehend auftretende CTCFL Bindung zeigen, bot sich auch hier aufgrund fehlender genom-weiter Bindedaten für CTCFL, im Falle der längeren Differenzierung, nicht. Somit konnte zumindest für zwei CTCFL Bindestellen eine klare differentielle Bindung in Folge veränderter Zugänglichkeit gezeigt werden. Die RA-vermittelte Differenzierung führte zur Kompaktierung der Bindestellen und gebundenes CTCFL ging verloren. An den anderen fünf untersuchten Bindestellen wirkten möglicherweise weitergehende Mechanismen die es CTCFL erlaubten weiter in diesen Bereichen zu binden oder die CTCFL Bindung an die DNA war in diesen Bereichen generell höher und die Änderung der Zugänglichkeit wirkte sich daher nicht auf sie aus.

## **5.5 Identifizierung eines möglichen DNA Einflusses auf differentielle CTCF und CTCFL Bindung**

Da die geteilte Zinkfinger-Domäne zwischen CTCF und CTCFL zwar eine sehr hohe aber keine 100 %-ige Übereinstimmung auf Aminosäure-Ebene aufweist sollte im Rahmen dieser Arbeit auch der Möglichkeit eines Einflusses der DNA Sequenz auf bevorzugte Bindung eines der beiden Proteine untersucht werden. Im Zuge neuer Daten zur genauen DNA Sequenz von CTCF Bindestellen (Liu et al. 2015) bot es sich an einen ähnlichen experimentellen Ansatz auch für CTCFL zu verfolgen. So wurden über die Top 1000, am stärksten gebundenen, 14 bp langen CTCF Bindesequenzen zusätzlich die CTCFL Bindung analysiert. Hier zeigte es sich von sehr großem Vorteil, dass die ursprünglichen Analysen auf den ENCODE CTCF Bindedaten in K562 Zellen beruhten, für die auch CTCFL Daten vorliegen. Im Zuge der Analyse konnten Bindesequenzen mit unterschiedlichen Mustern an CTCF und CTCFL Bindung identifiziert werden. Anhand von DNA Alignierungsanalysen wurden zwei stark unterschiedliche Untergruppen an Bindestellen

miteinander verglichen um möglicherweise kleine Unterschiede in den Sequenzen zu identifizieren die eine große Auswirkung auf die differentielle Bindung der beiden Proteine haben könnte. Bei diesen Gruppen handelte es sich zum einen um die 20 am stärksten gebundenen CTCF Sequenzen, unabhängig vom Grad der CTCFL Bindung, und zum anderen um die 20 am stärksten CTCF gebundenen Sequenzen die gleichzeitig überhaupt keine CTCFL Bindung zeigen. Wieder erwarten konnten zwei sehr ähnliche Sequenzen mit nur zwei Basen Unterschied zwischen den Gruppen identifiziert werden. Während die Sequenz CCAGCAGAGGGCGC in allen vorkommenden Instanzen im Genom von CTCF und in ca. 40 % der Fälle CTCFL gebunden vorlag konnte für die Sequenz TCA-GTAGAGGGCGC nur Bindung von CTCF, dies aber in allen im Genom vorhandenen Instanzen, detektiert werden. In weiteren Alignierungs-Analysen zur Identifizierung der jeweils nächstverwandten Sequenzen zu CCAGCAGAGGGCGC konnte eine weitere Base in der Bindesequenz entdeckt werden die einen entscheidenden Einfluss auf die Bindung von sowohl CTCF als auch CTCFL hat. So zeigt die Sequenz CCAGCAGAGGGGGC, die sich nur durch einen C zu G Tausch an Position 12 der Sequenz unterscheidet, stark reduzierte Bindung von CTCF und überhaupt keine Bindung von CTCFL an im Genom vorliegenden Instanzen. Position 12 des CTCF Motivs wurde unabhängig von dieser Analyse in einer kürzlich veröffentlichten Studie als eminent wichtige Position für CTCF Bindung identifiziert, weiterhin konnten in besagter Studie auch funktionelle Wirkunterschiede von CTCF anhand der gebundenen Motive festgestellt werden (Fang et al. 2015). Somit erlangen die Untersuchungen zur DNA Präferenz der beiden Proteine eine neue Aktualität und es wurden im letzten Schritt der vorliegenden Arbeit mögliche genomische Bereiche zur Überprüfung in *in silico* Analysen identifiziert. Anhand dieser Daten kann mittels EMSA Experimenten die Hypothese überprüft werden, ob die identifizierten DNA Sequenz-Unterschiede es wirklich erlauben zwischen CTCF und CTCFL Bindung zu diskriminieren. Die praktische Durchführung wurde in Zusammenarbeit mit O. Weth durchgeführt und die Daten fließen in ein gemeinsames Manuskript (Bergmaier et al. 2016) ein.

## 5.6 Ausblick

In der vorliegenden Arbeit wurde der Frage nachgegangen welche Faktoren die genomweiten Bindungsmuster des Faktors CTCFL bestimmen und es konnte gezeigt werden, dass CTCFL im Vergleich zu CTCF eine gesteigerte Präferenz für aktive Bereiche und somit offenes Chromatin besitzt. In der Frage der Modulierbarkeit der CTCFL Bindung

durch Änderung der Chromatin-Zugänglichkeit wurden genom-weite FAIRE Daten erhoben die jedoch nicht mit genom-weiten Daten zur CTCFL Bindung korreliert werden konnten. Hier wäre es interessant den weiteren Arbeitsaufwand aufzunehmen und einen Datensatz für CTCFL Bindung nach 9-tägiger Differenzierung im P19 Klon zu generieren. Somit könnte beantwortet werden ob es sich bei dem an zwei Bindestellen beobachteten Effekt des Verlustes der CTCFL Bindung in Folge verringerter Zugänglichkeit um einen generellen Mechanismus in Bezug auf CTCFL handeln könnte.

Im Zuge der neuen sich bietenden Möglichkeiten durch die CRISPR/CAS9 Technologie wären auch CRISPR *interference* Experimente in K562 Zellen oder dem therapeutischen Ansatz den einige Arbeitsgruppen verfolgen nachgehend in CTCFL exprimierenden weiteren Krebs-Zellen denkbar um einen fast vollständigen Knockout des CTCFL Gens zu erhalten. Eine weitere mögliche Anwendung der CRISPR Technologie wäre die Zerstörung von bekannten CTCFL Bindestellen, so könnte z.B. die CTCFL Bindestelle im Bereich des GPR56 Gens durch CRISPR gezielt entfernt/beeinträchtigt und untersucht werden ob in diesem Zuge die Regulierbarkeit des GPR56 Gens durch CTCFL verloren geht. Neben weiteren Experimenten zur Bindung von CTCFL, und dem transkriptionellen Einfluss auf Zielgene sollte auch die Suche nach möglichen Interaktionspartnern von CTCFL nicht vernachlässigt werden. hierfür würde sich die Überexpression von CTCFL-Domänen, die jeweils mit einem entsprechenden Tag-versehenen wurden, gefolgt von Immunopräzipitation und Massenspektrometrie anbieten. Die Identifizierung von interagierenden Faktoren könnte dann wieder Rückschlüsse auf die Funktion von CTCFL an seinen Bindestellen liefern und weitere Richtungen für neue Experimente liefern.

Abschließend sind noch die Studien zur möglichen Sequenz-Spezifität der CTCF und CTCFL Bindung zu nennen. Hier müssen die gestarteten Experimente weitergeführt und mit Hilfe von *in vitro* Shift-Versuchen geklärt werden ob sehr kleine Änderungen der CTCF Motiv Sequenz große Auswirkung auf das Bindeverhalten von CTCF als auch CTCFL haben könnten.

## 6 Literaturverzeichnis

- Alberts, Bruce; Johnson, Alexander; Lewis, Julian; Raff, Martin; Roberts, Keith; Walter, Peter (2002): *Molecular biology of the cell*. 4. ed. New York, NY: Garland Science.
- Ali, Tamer; Renkawitz, Rainer; Bartkuhn, Marek (2016): Insulators and domains of gene expression. In: *Current opinion in genetics & development* 37, S. 17–26. DOI: 10.1016/j.gde.2015.11.009.
- Allfrey, V. G.; Faulkner, R.; Mirsky, A. E. (1964): Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. In: *Proceedings of the National Academy of Sciences of the United States of America* 51, S. 786–794.
- Annunziato, Anthony T. (2008): DNA Packaging: Nucleosomes and Chromatin. In: *Nature Education* 1 (1), S. 26.
- Arenzana, T. L.; Schjerven, H.; Smale, S. T. (2015): Regulation of gene expression dynamics during developmental transitions by the Ikaros transcription factor. In: *Genes & development* 29 (17), S. 1801–1816. DOI: 10.1101/gad.266999.115.
- Ballard, T. D.; Wolff, J.; Griffin, J. B.; Stanley, J. S.; van Calcar, S.; Zemleni, J. (2002): Biotinidase catalyzes debiotinylation of histones. In: *European journal of nutrition* 41 (2), S. 78–84.
- Baniahmad, A.; Steiner, C.; Köhne, A. C.; Renkawitz, R. (1990): Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. In: *Cell* 61 (3), S. 505–514.
- Bannister, A. J.; Kouzarides, T. (2011): Regulation of chromatin by histone modifications. In: *Cell research* 21 (3), S. 381–395. DOI: 10.1038/cr.2011.22.
- Bell, A. C.; West, A. G.; Felsenfeld, G. (1999): The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. In: *Cell* 98 (3), S. 387–396.
- Belton, Jon-Matthew; McCord, Rachel Patton; Gibcus, Johan Harmen; Naumova, Natalia; Zhan, Ye; Dekker, Job (2012): Hi-C: a comprehensive technique to capture the conformation of genomes. In: *Methods (San Diego, Calif.)* 58 (3), S. 268–276. DOI: 10.1016/j.ymeth.2012.05.001.
- Bergmaier, Philipp; Weth, Oliver; Dienstbach, Sven; Leers, Joerg; Sleutels, Frank; Galjart, Niels et al. (2016): Determinants of CTCFL binding in contrast to CTCF. In: *in preparation*.
- Birnboim, H. C.; Doly, J. (1979): A rapid alkaline extraction procedure for screening recombinant plasmid DNA. In: *Nucleic acids research* 7 (6), S. 1513–1523.
- Bradford, M. M. (1976): A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. In: *Analytical biochemistry* 72, S. 248–254.
- Burke, Les J.; Hollemann, Thomas; Pieler, Tomas; Renkawitz, Rainer (2002): Molecular cloning and expression of the chromatin insulator protein CTCF in *Xenopus laevis*. In: *Mechanisms of development* 113 (1), S. 95–98.
- Calo, Eliezer; Wysocka, Joanna (2013): Modification of Enhancer Chromatin. What, How, and Why? In: *Molecular cell* 49 (5), S. 825–837. DOI: 10.1016/j.molcel.2013.01.038.



- Cedar, Howard; Bergman, Yehudit (2009): Linking DNA methylation and histone modification. Patterns and paradigms. In: *Nat Rev Genet* 10 (5), S. 295–304. DOI: 10.1038/nrg2540.
- Chao, Wendy; Huynh, Khanh D.; Spencer, Rebecca J.; Davidow, Lance S.; Lee, Jeannie T. (2002): CTCF, a candidate trans-acting factor for X-inactivation choice. In: *Science (New York, N.Y.)* 295 (5553), S. 345–347. DOI: 10.1126/science.1065982.
- Chen, Guangchun; Yang, Liquan; Begum, Shahinoor; Xu, Lei (2010): GPR56 is essential for testis development and male fertility in mice. In: *Developmental dynamics : an official publication of the American Association of Anatomists* 239 (12), S. 3358–3367. DOI: 10.1002/dvdy.22468.
- Chen, Hebing; Tian, Yao; Shu, Wenjie; Bo, Xiaochen; Wang, Shengqi (2012): Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. In: *PloS one* 7 (7), S. e41374. DOI: 10.1371/journal.pone.0041374.
- Christophorou, Maria A.; Castelo-Branco, Gonçalo; Halley-Stott, Richard P.; Oliveira, Clara Slade; Loos, Remco; Radzisheuskaya, Aliaksandra et al. (2014): Citrullination regulates pluripotency and histone H1 binding to chromatin. In: *Nature* 507 (7490), S. 104–108. DOI: 10.1038/nature12942.
- Christova, Rossitza; Oelgeschläger, Thomas (2002): Association of human TFIID-promoter complexes with silenced mitotic chromatin in vivo. In: *Nature cell biology* 4 (1), S. 79–82. DOI: 10.1038/ncb733.
- Dekker, Job; Rippe, Karsten; Dekker, Martijn; Kleckner, Nancy (2002): Capturing chromosome conformation. In: *Science (New York, N.Y.)* 295 (5558), S. 1306–1311. DOI: 10.1126/science.1067799.
- Dixon, Jesse R.; Selvaraj, Siddarth; Yue, Feng; Kim, Audrey; Li, Yan; Shen, Yin et al. (2012): Topological domains in mammalian genomes identified by analysis of chromatin interactions. In: *Nature* 485 (7398), S. 376–380. DOI: 10.1038/nature11082.
- Dostie, Josée; Richmond, Todd A.; Arnaout, Ramy A.; Selzer, Rebecca R.; Lee, William L.; Honan, Tracey A. et al. (2006): Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. In: *Genome research* 16 (10), S. 1299–1309. DOI: 10.1101/gr.5571506.
- Edwards, M. K.; Harris, J. F.; McBurney, M. W. (1983): Induced muscle differentiation in an embryonal carcinoma cell line. In: *Mol Cell Biol* 3 (12), S. 2280–2286.
- Egger, Gerda; Liang, Gangning; Aparicio, Ana; Jones, Peter A. (2004): Epigenetics in human disease and prospects for epigenetic therapy. In: *Nature* 429 (6990), S. 457–463. DOI: 10.1038/nature02625.
- Eymery, Angeline; Callanan, Mary; Vourc'h, Claire (2009): The secret message of heterochromatin. New insights into the mechanisms and function of centromeric and pericentric repeat sequence transcription. In: *Int. J. Dev. Biol.* 53 (2-3), S. 259–268. DOI: 10.1387/ijdb.082673ae.
- Fang, Rongxin; Wang, Chengqi; Skogerbo, Geir; Zhang, Zhihua (2015): Functional diversity of CTCFs is encoded in their binding motifs. In: *BMC Genomics* 16 (1), S. 520. DOI: 10.1186/s12864-015-1824-6.
- Filion, Guillaume J.; van Bemmelen, Joke G.; Braunschweig, Ulrich; Talhout, Wendy; Kind, Jop; Ward, Lucas D. et al. (2010): Systematic Protein Location Mapping Reveals

- Five Principal Chromatin Types in *Drosophila* Cells. In: *Cell* 143 (2), S. 212–224. DOI: 10.1016/j.cell.2010.09.009.
- Füllgrabe, J.; Kavanagh, E.; Joseph, B. (2011): Histone onco-modifications. In: *Oncogene* 30 (31), S. 3391–3403. DOI: 10.1038/onc.2011.121.
- Grigoryev, Sergei A.; Woodcock, Christopher L. (2012): Chromatin organization — The 30nm fiber. In: *Experimental cell research* 318 (12), S. 1448–1455. DOI: 10.1016/j.yexcr.2012.02.014.
- Guo, Ya; Xu, Quan; Canzio, Daniele; Shou, Jia; Li, Jinhuan; Gorkin, David U. et al. (2015): CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. In: *Cell* 162 (4), S. 900–910. DOI: 10.1016/j.cell.2015.07.038.
- Hadjur, Suzana; Williams, Luke M.; Ryan, Natalie K.; Cobb, Bradley S.; Sexton, Tom; Fraser, Peter et al. (2009): Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. In: *Nature* 460 (7253), S. 410–413. DOI: 10.1038/nature08079.
- Hamann, J.; Aust, G.; Arac, D.; Engel, F. B.; Formstone, C.; Fredriksson, R. et al. (2015): International Union of Basic and Clinical Pharmacology. XCIV. Adhesion G Protein-Coupled Receptors. In: *Pharmacological Reviews* 67 (2), S. 338–367. DOI: 10.1124/pr.114.009647.
- Heger, Peter; Marin, Birger; Bartkuhn, Marek; Schierenberg, Einhard; Wiehe, Thomas (2012): The chromatin insulator CTCF and the emergence of metazoan diversity. In: *Proceedings of the National Academy of Sciences of the United States of America* 109 (43), S. 17507–17512. DOI: 10.1073/pnas.1111941109.
- Hines, William C.; Bazarov, Alexey V.; Mukhopadhyay, Rituparna; Yaswen, Paul (2010): BORIS (CTCF-L) is not expressed in most human breast cell lines and high grade breast carcinomas. In: *PloS one* 5 (3), S. e9738. DOI: 10.1371/journal.pone.0009738.
- Holliday, R.; Pugh, J. E. (1975): DNA modification mechanisms and gene activity during development. In: *Science (New York, N.Y.)* 187 (4173), S. 226–232.
- Hore, Timothy A.; Deakin, Janine E.; Marshall Graves, Jennifer A. (2008): The evolution of epigenetic regulators CTCF and BORIS/CTCF-L in amniotes. In: *PLoS genetics* 4 (8), S. e1000169. DOI: 10.1371/journal.pgen.1000169.
- Iniguez-Lluhi, J. A. (2006): For a healthy histone code, a little SUMO in the tail keeps the acetyl away. In: *ACS chemical biology* 1 (4), S. 204–206. DOI: 10.1021/cb600188m.
- Ishihara, Ko; Oshimura, Mitsuo; Nakao, Mitsuyoshi (2006): CTCF-dependent chromatin insulator is linked to epigenetic remodeling. In: *Molecular cell* 23 (5), S. 733–742. DOI: 10.1016/j.molcel.2006.08.008.
- Jenuwein, T.; Allis, C. D. (2001): Translating the histone code. In: *Science (New York, N.Y.)* 293 (5532), S. 1074–1080. DOI: 10.1126/science.1063127.
- Jinek, Martin; Chylinski, Krzysztof; Fonfara, Ines; Hauer, Michael; Doudna, Jennifer A.; Charpentier, Emmanuelle (2012): A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. In: *Science (New York, N.Y.)* 337 (6096), S. 816–821. DOI: 10.1126/science.1225829.
- Jones, Peter A.; Baylin, Stephen B. (2002): The fundamental role of epigenetic events in cancer. In: *Nature reviews. Genetics* 3 (6), S. 415–428. DOI: 10.1038/nrg816.

- Jones, Tania A.; Ogunkolade, Babatunji W.; Szary, Jaroslaw; Aarum, Johan; Mumin, Muhammad A.; Patel, Shyam et al. (2011): Widespread expression of BORIS/CTCF in normal and cancer cells. In: *PloS one* 6 (7), S. e22399. DOI: 10.1371/journal.pone.0022399.
- Kang, J. Y.; Song, S. H.; Yun, J.; Jeon, M. S.; Kim, H. P.; Han, S. W.; Kim, T. Y. (2015): Disruption of CTCF/cohesin-mediated high-order chromatin structures by DNA methylation downregulates PTGS2 expression. In: *Oncogene* 34 (45), S. 5677–5684. DOI: 10.1038/onc.2015.17.
- Kim, Tae Hoon; Abdullaev, Ziedulla K.; Smith, Andrew D.; Ching, Keith A.; Loukinov, Dmitri I.; Green, Roland D. et al. (2007): Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. In: *Cell* 128 (6), S. 1231–1245. DOI: 10.1016/j.cell.2006.12.048.
- Klenova, E. M.; Nicolas, R. H.; Paterson, H. F.; Carne, A. F.; Heath, C. M.; Goodwin, G. H. et al. (1993): CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. In: *Molecular and cellular biology* 13 (12), S. 7612–7624.
- Klenova, E. M.; Nicolas, R. H.; U, S.; Carne, A. F.; Lee, R. E.; Lobanekov, V. V.; Goodwin, G. H. (1997): Molecular weight abnormalities of the CTCF transcription factor: CTCF migrates aberrantly in SDS-PAGE and the size of the expressed protein is affected by the UTRs and sequences within the coding region of the CTCF gene. In: *Nucleic acids research* 25 (3), S. 466–474.
- Klenova, Elena M.; Morse, Herbert C.; Ohlsson, Rolf; Lobanekov, Victor V. (2002): The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer. In: *Seminars in cancer biology* 12 (5), S. 399–414.
- Koch-Nolte, F.; Kernstock, S.; Mueller-Dieckmann, C.; Weiss, M. S.; Haag, F. (2008): Mammalian ADP-ribosyltransferases and ADP-ribosylhydrolases. In: *Frontiers in bio-science : a journal and virtual library* 13, S. 6716–6729.
- Kornberg, R. D. (1974): Chromatin structure: a repeating unit of histones and DNA. In: *Science (New York, N.Y.)* 184 (4139), S. 868–871.
- Kosaka-Suzuki, Natsuki; Suzuki, Teruhiko; Pugacheva, Elena M.; Vostrov, Alexander A.; Morse, Herbert C.; Loukinov, Dmitri; Lobanekov, Victor (2011): Transcription factor BORIS (Brother of the Regulator of Imprinted Sites) directly induces expression of a cancer-testis antigen, TSP50, through regulated binding of BORIS to the promoter. In: *The Journal of biological chemistry* 286 (31), S. 27378–27388. DOI: 10.1074/jbc.M111.243576.
- Kruithof, Maarten; Chien, Fan-Tso; Routh, Andrew; Logie, Colin; Rhodes, Daniela; van Noort, John (2009): Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. In: *Nat Struct Mol Biol* 16 (5), S. 534–540. DOI: 10.1038/nsmb.1590.
- Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J. et al. (2001): Initial sequencing and analysis of the human genome. In: *Nature* 409 (6822), S. 860–921. DOI: 10.1038/35057062.
- Li, Tie; Lu, Zhenyu; Lu, Luo (2004): Regulation of eye development by transcription control of CCCTC binding factor (CTCF). In: *The Journal of biological chemistry* 279 (26), S. 27575–27583. DOI: 10.1074/jbc.M313942200.

- Liu, Mingdong; Maurano, Matthew T.; Wang, Hao; Qi, Heyuan; Song, Chao-Zhong; Navas, Patrick A. et al. (2015): Genomic discovery of potent chromatin insulators for human gene therapy. In: *Nature biotechnology* 33 (2), S. 198–203. DOI: 10.1038/nbt.3062.
- Lobanenkova, V. V.; Nicolas, R. H.; Adler, V. V.; Paterson, H.; Klenova, E. M.; Polotskaja, A. V.; Goodwin, G. H. (1990): A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. In: *Oncogene* 5 (12), S. 1743–1753.
- Loukinov, Dmitri I.; Pugacheva, Elena; Vatolin, Sergei; Pack, Svetlana D.; Moon, Hanlim; Chernukhin, Igor et al. (2002): BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. In: *Proceedings of the National Academy of Sciences of the United States of America* 99 (10), S. 6806–6811. DOI: 10.1073/pnas.092123699.
- Marucci, Lucia; Pedone, Elisa; Di Vicino, Umberto; Sanuy-Escribano, Blanca; Isalan, Mark; Cosma, Maria Pia (2014):  $\beta$ -Catenin Fluctuates in Mouse ESCs and Is Essential for Nanog-Mediated Reprogramming of Somatic Cells to Pluripotency. In: *Cell reports* 8 (6), S. 1686–1696. DOI: 10.1016/j.celrep.2014.08.011.
- McBurney, M. W. (1993): P19 embryonal carcinoma cells. In: *Int J Dev Biol* 37 (1), S. 135–140.
- Mkrtichyan, Mikayel; Ghochikyan, Anahit; Davtyan, Hayk; Movsesyan, Nina; Loukinov, Dmitry; Lobanenkova, Victor et al. (2011): Cancer-testis antigen, BORIS based vaccine delivered by dendritic cells is extremely effective against a very aggressive and highly metastatic mouse mammary carcinoma. In: *Cellular immunology* 270 (2), S. 188–197. DOI: 10.1016/j.cellimm.2011.05.007.
- Moon, Hanlim; Filippova, Galina; Loukinov, Dmitry; Pugacheva, Elena; Chen, Qi; Smith, Sheryl T. et al. (2005): CTCF is conserved from *Drosophila* to humans and confers enhancer blocking of the Fab-8 insulator. In: *EMBO reports* 6 (2), S. 165–170. DOI: 10.1038/sj.embor.7400334.
- Moore, James M.; Rabaia, Natalia A.; Smith, Leslie E.; Fagerlie, Sara; Gurley, Kay; Loukinov, Dmitry et al. (2012): Loss of maternal CTCF is associated with peri-implantation lethality of *Ctcf* null embryos. In: *PloS one* 7 (4), S. e34915. DOI: 10.1371/journal.pone.0034915.
- Necochea-Campion, Rosalia de; Ghochikyan, Anahit; Josephs, Steven F.; Zacharias, Shelly; Woods, Erik; Karimi-Busheri, Feridoun et al. (2011): Expression of the epigenetic factor BORIS (CTCF-L) in the human genome. In: *Journal of translational medicine* 9, S. 213. DOI: 10.1186/1479-5876-9-213.
- Nguyen, Phuongmai; Bar-Sela, Gil; Sun, Lunching; Bisht, Kheem S.; Cui, Hengmi; Kohn, Elise et al. (2008a): BAT3 and SET1A form a complex with CTCFL/BORIS to modulate H3K4 histone dimethylation and gene expression. In: *Molecular and cellular biology* 28 (21), S. 6720–6729. DOI: 10.1128/MCB.00568-08.
- Nguyen, Phuongmai; Cui, Hengmi; Bisht, Kheem S.; Sun, Lunching; Patel, Krish; Lee, Richard S. et al. (2008b): CTCFL/BORIS is a methylation-independent DNA-binding protein that preferentially binds to the paternal H19 differentially methylated region. In: *Cancer research* 68 (14), S. 5546–5551. DOI: 10.1158/0008-5472.CAN-08-1005.

- Niwa, H.; Miyazaki, J.; Smith, A. G. (2000): Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. In: *Nature genetics* 24 (4), S. 372–376. DOI: 10.1038/74199.
- Novak Kujundžić, Renata; Grbeša, Ivana; Ivkić, Mirko; Krušlin, Božo; Konjevoda, Paško; Gall Trošelj, Koraljka (2014): Possible prognostic value of BORIS transcript variants ratio in laryngeal squamous cell carcinomas - a pilot study. In: *Pathology oncology research : POR* 20 (3), S. 687–695. DOI: 10.1007/s12253-014-9749-y.
- Ohlsson, R.; Renkawitz, R.; Lobanenko, V. (2001): CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. In: *Trends in genetics : TIG* 17 (9), S. 520–527.
- Okabayashi, Koji; Fujita, Tomonobu; Miyazaki, Junichiro; Okada, Tsutomu; Iwata, Takashi; Hirao, Nobumaru et al. (2012): Cancer-testis antigen BORIS is a novel prognostic marker for patients with esophageal cancer. In: *Cancer science* 103 (9), S. 1617–1624. DOI: 10.1111/j.1349-7006.2012.02355.x.
- Olins, A. L.; Olins, D. E. (1974): Spheroid Chromatin Units (ngr Bodies). In: *Science* 183 (4122), S. 330–332. DOI: 10.1126/science.183.4122.330.
- Ong, Chin-Tong; Corces, Victor G. (2014): CTCF. An architectural protein bridging genome topology and function. In: *Nat Rev Genet* 15 (4), S. 234–246. DOI: 10.1038/nrg3663.
- Palumbo, Camilla; van Roozendaal, Kees; Gillis, Ad J.; van Gurp, Ruud H.; Munnik, Hannie de; Wolter Oosterhuis, J. et al. (2002): Expression of the PDGF  $\alpha$ -receptor 1.5 kb transcript, OCT-4, and c-KIT in human normal and malignant tissues. Implications for the early diagnosis of testicular germ cell tumours and for our understanding of regulatory mechanisms. In: *The Journal of pathology* 196 (4), S. 467–477. DOI: 10.1002/path.1064.
- Pant, Vinod; Kurukuti, Sreenivasulu; Pugacheva, Elena; Shamsuddin, Shaharum; Mariano, Piero; Renkawitz, Rainer et al. (2004): Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. In: *Molecular and cellular biology* 24 (8), S. 3497–3504.
- Phillips-Cremins, Jennifer E.; Sauria, Michael E. G.; Sanyal, Amartya; Gerasimova, Tatiana I.; Lajoie, Bryan R.; Bell, Joshua S. K. et al. (2013): Architectural protein subclasses shape 3D organization of genomes during lineage commitment. In: *Cell* 153 (6), S. 1281–1295. DOI: 10.1016/j.cell.2013.04.053.
- Pope, Benjamin D.; Ryba, Tyrone; Dileep, Vishnu; Yue, Feng; Wu, Weisheng; Denas, Olgert et al. (2014): Topologically associating domains are stable units of replication-timing regulation. In: *Nature* 515 (7527), S. 402–405. DOI: 10.1038/nature13986.
- Pugacheva, Elena M.; Rivero-Hinojosa, Samuel; Espinoza, Celso A.; Méndez-Catalá, Claudia Fabiola; Kang, Sungyun; Suzuki, Teruhiko et al. (2015): Comparative analyses of CTCF and BORIS occupancies uncover two distinct classes of CTCF binding genomic regions. In: *Genome biology* 16, S. 161. DOI: 10.1186/s13059-015-0736-8.
- Pugacheva, Elena M.; Suzuki, Teruhiko; Pack, Svetlana D.; Kosaka-Suzuki, Natsuki; Yoon, Jeongheon; Vostrov, Alexander A. et al. (2010): The structural complexity of the human BORIS gene in gametogenesis and cancer. In: *PloS one* 5 (11), S. e13872. DOI: 10.1371/journal.pone.0013872.

- Renaud, S.; Loukinov, D.; Abdullaev, Z.; Guilleret, I.; Bosman, F. T.; Lobanenkov, V.; Benhattar, J. (2007): Dual role of DNA methylation inside and outside of CTCF-binding regions in the transcriptional regulation of the telomerase hTERT gene. In: *Nucleic acids research* 35 (4), S. 1245–1256. DOI: 10.1093/nar/gkl1125.
- Renaud, Stéphanie; Loukinov, Dmitri; Alberti, Loredana; Vostrov, Alexander; Kwon, Yoo-Wook; Bosman, Fred T. et al. (2011): BORIS/CTCF-mediated transcriptional regulation of the hTERT telomerase gene in testicular and ovarian tumor cells. In: *Nucleic acids research* 39 (3), S. 862–873. DOI: 10.1093/nar/gkq827.
- Schneider, Robert; Grosschedl, Rudolf (2007): Dynamics and interplay of nuclear architecture, genome organization, and gene expression. In: *Genes & development* 21 (23), S. 3027–3043. DOI: 10.1101/gad.1604607.
- Schueler, Mary G.; Sullivan, Beth A. (2006): Structural and functional dynamics of human centromeric chromatin. In: *Annual review of genomics and human genetics* 7, S. 301–313. DOI: 10.1146/annurev.genom.7.080505.115613.
- Sérandour, Aurélien A.; Avner, Stéphane; Oger, Frédéric; Bizot, Maud; Percevault, Frédéric; Lucchetti-Miganeh, Céline et al. (2012): Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. In: *Nucleic acids research* 40 (17), S. 8255–8265. DOI: 10.1093/nar/gks595.
- Shukla, Sanjeev; Kavak, Ersen; Gregory, Melissa; Imashimizu, Masahiko; Shutinoski, Bojan; Kashlev, Mikhail et al. (2011): CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. In: *Nature* 479 (7371), S. 74–79. DOI: 10.1038/nature10442.
- Sleutels, Frank; Soochit, Widia; Bartkuhn, Marek; Heath, Helen; Dienstbach, Sven; Bergmaier, Philipp et al. (2012): The male germ cell gene regulator CTCFL is functionally different from CTCF and binds CTCF-like consensus sites in a nucleosome composition-dependent manner. In: *Epigenetics & chromatin* 5 (1), S. 8. DOI: 10.1186/1756-8935-5-8.
- Soltanian, Sara; Dehghani, Hesam; Matin, Maryam M.; Bahrami, Ahmad Reza (2014): Expression analysis of BORIS during pluripotent, differentiated, cancerous, and non-cancerous cell states. In: *Acta biochimica et biophysica Sinica* 46 (8), S. 647–658. DOI: 10.1093/abbs/gmu045.
- Staines, W. A.; Craig, J.; Reuhl, K.; McBurney, M. W. (1996): Retinoic acid treated P19 embryonal carcinoma cells differentiate into oligodendrocytes capable of myelination. In: *Neuroscience* 71 (3), S. 845–853.
- Suzuki, H.; Osaki, K.; Sano, K.; Alam, AHMK; Nakamura, Y.; Ishigaki, Y. et al. (2011): Comprehensive Analysis of Alternative Splicing and Functionality in Neuronal Differentiation of P19 Cells. In: *PloS one* 6 (2). DOI: 10.1371/journal.pone.0016880.
- Suzuki, Teruhiko; Kosaka-Suzuki, Natsuki; Pack, Svetlana; Shin, Dong-Mi; Yoon, Jeongheon; Abdullaev, Ziedulla et al. (2010): Expression of a testis-specific form of Gal3st1 (CST), a gene essential for spermatogenesis, is regulated by the CTCF paralogous gene BORIS. In: *Molecular and cellular biology* 30 (10), S. 2473–2484. DOI: 10.1128/MCB.01093-09.
- Takai, Daiya; Jones, Peter A. (2002): Comprehensive analysis of CpG islands in human chromosomes 21 and 22. In: *Proceedings of the National Academy of Sciences of the United States of America* 99 (6), S. 3740–3745. DOI: 10.1073/pnas.052410099.

Tan, Minjia; Luo, Hao; Lee, Sangkyu; Jin, Fulai; Yang, Jeong Soo; Montellier, Emilie et al. (2011): Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. In: *Cell* 146 (6), S. 1016–1028. DOI: 10.1016/j.cell.2011.08.008.

The ENCODE Project Consortium (2004): The ENCODE (ENCyclopedia Of DNA Elements) Project. In: *Science* 306 (5696), S. 636–640. DOI: 10.1126/science.1105136.

The ENCODE Project Consortium (2007): Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. In: *Nature* 447 (7146), S. 799–816. DOI: 10.1038/nature05874.

The ENCODE Project Consortium (2012): An integrated encyclopedia of DNA elements in the human genome. In: *Nature* 489 (7414), S. 57–74. DOI: 10.1038/nature11247.

Thurman, Robert E.; Rynes, Eric; Humbert, Richard; Vierstra, Jeff; Maurano, Matthew T.; Haugen, Eric et al. (2012): The accessible chromatin landscape of the human genome. In: *Nature* 489 (7414), S. 75–82. DOI: 10.1038/nature11232.

van de Nobelen, Suzanne; Rosa-Garrido, Manuel; Leers, Joerg; Heath, Helen; Soochit, Widia; Joosen, Linda et al. (2010): CTCF regulates the local epigenetic state of ribosomal DNA repeats. In: *Epigenetics & chromatin* 3 (1), S. 19. DOI: 10.1186/1756-8935-3-19.

Vandesompele, J.; Preter, K. de; Pattyn, F.; Poppe, B.; van Roy, N.; Paepe, A. de; Speleman, F. (2002): Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. In: *Genome biology* 3 (7), S. research0034.1–11.

Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G. et al. (2001): The sequence of the human genome. In: *Science (New York, N.Y.)* 291 (5507), S. 1304–1351. DOI: 10.1126/science.1058040.

Wan, Le-Ben; Pan, Hua; Hannenhalli, Sridhar; Cheng, Yong; Ma, Jun; Fedoriw, Andrew et al. (2008): Maternal depletion of CTCF reveals multiple functions during oocyte and preimplantation embryo development. In: *Development (Cambridge, England)* 135 (16), S. 2729–2738. DOI: 10.1242/dev.024539.

Wang, Hao; Maurano, Matthew T.; Qu, Hongzhu; Varley, Katherine E.; Gertz, Jason; Pauli, Florencia et al. (2012): Widespread plasticity in CTCF occupancy linked to DNA methylation. In: *Genome research* 22 (9), S. 1680–1688. DOI: 10.1101/gr.136101.111.

Weth, Oliver; Paprotka, Christine; Günther, Katharina; Schulte, Astrid; Baierl, Manuel; Leers, Joerg et al. (2014): CTCF induces histone variant incorporation, erases the H3K27me3 histone mark and opens chromatin. In: *Nucleic acids research* 42 (19), S. 11941–11951. DOI: 10.1093/nar/gku937.

Wilkinson, K. D. (2000): Ubiquitination and deubiquitination: targeting of proteins for degradation by the proteasome. In: *Seminars in cell & developmental biology* 11 (3), S. 141–148. DOI: 10.1006/scdb.2000.0164.

Woloszynska-Read, Anna; Zhang, Wa; Yu, Jihnhee; Link, Petra A.; Mhawech-Fauceglia, Paulette; Collamat, Golda et al. (2011): Coordinated cancer germline antigen promoter and global DNA hypomethylation in ovarian cancer: association with the BORIS/CTCF expression ratio and advanced stage. In: *Clinical cancer research : an official journal of the American Association for Cancer Research* 17 (8), S. 2170–2180. DOI: 10.1158/1078-0432.CCR-10-2315.

Xiao, Tiaojiang; Wallace, Julie; Felsenfeld, Gary (2011): Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. In: *Molecular and cellular biology* 31 (11), S. 2174–2183. DOI: 10.1128/MCB.05093-11.

Yoon, Se-Lyun; Roh, Yun-Gil; Lee, Sang-Ho; Kim, Si-Hoon; Kim, Min Chan; Kim, Su Jin; Leem, Sun-Hee (2011): Analysis of promoter methylation and polymorphic minisatellites of BORIS and lack of association with gastric cancer. In: *DNA and cell biology* 30 (9), S. 691–698. DOI: 10.1089/dna.2011.1248.



## Danksagung

An dieser Stelle möchte ich mich bei allen Personen bedanken, die mich während dieser Arbeit unterstützt und auf dem etwas längeren Weg begleitet haben. Sind doch einige gewesen! ;-)

Mein besonderer Dank gilt Herrn Prof. Dr. Rainer Renkawitz, für die Annahme als Doktorand, die gute Betreuung und nicht zuletzt Finanzierung des Projektes. Danke auch an die weiteren Professoren, Prof. Dr. Lienhard Schmitz, Prof. Dr. Reinhard Dammann und Prof. Dr. Adriaan Dorresteijn die sich als Gutachter bzw. Prüfer der vorliegenden Arbeit zur Verfügung gestellt haben.

Auch bedanken möchte ich mich bei allen Mitgliedern, über die Jahre, des Labors 329, in dem eigentlich immer eine sehr kollegiale und gute Arbeitsatmosphäre herrschte. Besonders an den „harten Kern“: Laborleiter Ölle, für unzählige Diskussion über Fußball, Kinder und Finanzen. Die Montag Morgen werden mir sehr fehlen! Sven, ehemals Betreuer meiner Diplomarbeit, dann Mitstreiter im Sachen CTCFL, es war mir eine Ehre. Und natürlich Jörg, Meister der Protein Biologie, schwedischen Hard-Rocks und des kreativen Chaos am Arbeitsplatz. Ich werde viele schöne Erinnerungen wie z.B. den Halbmarathon und die jährlichen eisigen Frankfurt-Bremen Spiele mitnehmen. Natürlich danke auch an Priyanka, CP, Flo, Sinah, Anne, Katharina S., Nadine, Manuel und Artem. Ihr wart das Salz in der Suppe des Laboralltages ☺

Danke an Marek, ohne den viele Teile dieser Arbeit nicht möglich gewesen wären. Der beste Bioinformatiker und Basketball-Kenner bestimmt von mindestens Mittelhessen!

Danke allen Mitgliedern des restlichen Instituts die in Seminaren und auch im Privaten zu guten Kollegen und teilweise Freunden wurden: Martin (auch wenn er Köln Fan ist), Steffen (obwohl er Schalke und Karneval Fan ist), Katharina G., Dorte, Theresa, Antje, Tanja.

Ein großer Dank auch an alle administrativen und technischen Mitarbeiter des Institutes die immer für einen reibungslosen Ablauf sorgten: Leni, Nadine, Anna, Angelika, Sonja, Andreas, Bashir, Israil und Christa.

Abschließend möchte ich mich bei meiner Familie, die im Laufe der Doktorarbeit stetig wuchs, und vor allem bei meinen Eltern Ottilie & Peter bedanken für andauernde Unterstützung, Vertrauen und Liebe. Bei meinem Bruder Gabriel mit Nana, Ottilie (klein), Herman und Harald. Schade, dass Norwegen so weit weg ist!

---

Zuletzt bleibt noch eine ehemalige Arbeitskollegin bei der ich mich besonderes bedanken muss und will, da sie inzwischen mit mir als Partnerin durchs Leben geht und deren Familie mich auch sehr gut aufgenommen hat. Danke für alles, Steffi. Seit einer tollen Konferenz in New York bist du an meiner Seite, hast uns unsere tolle Tochter Leonie, der diese Arbeit gewidmet ist, geschenkt und auf das Abenteuer Familie mit mir eingelassen. Danke für die andauernde Unterstützung auch wenn die unweigerliche Frustration beim Schreiben der Arbeit hin- und wieder einmal auftrat!

## Eigenständigkeitserklärung

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten.

---

Gießen, 19.04.2016

---

Philipp Bergmaier