DNA sequence and chromatin landscape regulate genomic binding of the glucocorticoid receptor

Dissertation zur Erlangung des akademischen Grades des Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

vorgelegt von
Stephan Raphael Starick
aus Berlin

2014

Diese Arbeit wurde unter der Leitung von Dr. Sebastiaan H. Meijsing im Zeitraum von September 2010 bis September 2014 am Max-Planck-Institut für molekulare Genetik angefertigt

1. Gutachter:
Dr. Harald Seitz
Fraunhofer-Institut für Zelltherapie und Immunologie

2. Gutachter:
Prof. Dr. Rupert Mutzel
Freie Universität Berlin

Disputation am 28. April 2015

1 Danksagung

Meinem Betreuer Dr. Sebastiaan Meijsing gilt mein spezieller Dank für die Möglichkeit bei ihm meine Arbeit durchführen zu können und seinem Bestreben mich auf kurze und präzise Aussagen zu fokussieren. Auch für unsere gemeinsamen wissenschaftlichen Diskussionen und seinen Rat bin ich sehr dankbar.

Ein besonderer Dank geht an Dr. Harald Seitz, der meine Arbeit über die gesamte Zeit begleitete. Sein wertvoller Rat sowie seine Unterstützung meine Ergebnisse in einem anderen Licht zu betrachten, halfen mir sehr mein Projekt erfolgreich zu beenden und meine Ziele klar zu definieren.

Herrn Prof. Dr. Rupert Mutzel danke ich sehr herzlich für die freundliche Übernahme der Begutachtung meiner Arbeit.

Weiterer Dank geht an meiner Arbeitsgruppe, im speziellen Jonas und Marcel sowie Stephanie und Katja für die vielfältige Unterstützung, die freundliche Atmosphäre und die angeregten Diskussionen die meine Arbeit bereicherten.

Für ihre herzliche, freundschaftliche und moralische Unterstützung gebührt mein tiefer Dank Edda Einfeldt.

Auch der gesamten Bioinformatikabteilung des Max-Planck-Instituts für molekulare Genetik gilt meine Dankbarkeit, vor allem Dr. Morgane Thomas-Chollier, Jonas Ibn-Salem, sowie im speziellen Mike, Alessandro, Julia L und Yves für ihre Bereitschaft sich in die Arbeitswelt des jeweils anderen ein zudenken und die gemeinsamen Projekte erfolgreich zu beenden. Prof. Dr. Martin Vingron möchte ich für die finanzielle Unterstützung während meiner Doktorarbeit als auch für die experimentellen Anregungen danken. Ohne die sehr gute interdisziplinäre Zusammenarbeit in seine Abteilung wäre mein Projekt nicht möglich gewesen.

Nicht zuletzt möchte ich meiner Familie und Freunden für den konstanten Rückhalt danken, der mich von meinem Studium bis jetzt begleitete.

Meiner Lebenspartnerin und Ehefrau Anni gebührt der größte Dank außerhalb meines Labors. Ihre fortwährende Unterstützung und ihr liebevolles Verständnis ermutigten mich vor allem gegen Ende meiner Doktorarbeit die Wege zusammen zu gehen.

2 Abstract

Organisms have evolved cell types for different functions. Transcriptional regulation of subsets of genes by binding of transcription factors (TFs) to DNA sequences within regulatory regions is essential to manifest cellular phenotypes. Examination of genomic binding sites of the glucocorticoid receptor (GR), a hormone inducible TF, showed that only a small percentage of many possible GR binding sequences in the genome is actually GR-bound. Furthermore, only a fraction of the GR-bound regions appeared to contain the classical GR consensus sequence, indicating that this sequence is neither necessary nor sufficient to explain GR binding. This raises two questions: First, what features of the chromatin landscape discriminate bound from unbound GR binding sequences and second, what sequences, other than the classical binding sequence, can recruit GR to the genome?

Density and organization of DNA-enwound nucleosomes divides the genome's chromatin landscape into TF-accessible "open" and "closed" regions. Our GR chromatin immunoprecipitation (ChIP) showed a predominant GR-binding to "open" chromatin, with one distinctive feature: A specific depletion of GR-binding at promoter regions of the "open" chromatin universe. This depletion could be explained to some extend by a reduced frequency of GR's canonical motif-matching sequences in these regions. However, sequence composition of promoter regions explains only part of the depletion and our Hierarchical Bayes Modeling of GR-binding, using chromatin marks as model features, indicated additional candidate mechanisms. For example, we found typical promoter marks (e.g. H3K9ac) to be negatively correlated with GR-binding in the "open" chromatin universe. These acetylation marks are set by histone acetyltransferases that can also post-translationally modify GR and attenuate its interaction with DNA (Kino & Chrousos 2011) thereby providing a possible explanation for the promoter-proximal depletion observed.

As we found that a large fraction of GR-bound regions lack a canonical GR binding sequence, we asked what sequences recruit GR to individual loci? To identify these sequences at high resolution, we used ChIP-Exo, taking advantage of an exonuclease trimming of ChIP fragments to the protein:DNA cross-linking point which protects the DNA from further digestion. We exploited this signal by determining *footprint profiles* of TF binding at single base pair resolution, using ExoProfiler, a computational motif-based pipeline. Comparison of our and the few public available ChIP-Exo datasets revealed: *footprints* are protein and recognition-sequence-specific signatures of TF binding sites that allow to distinguish direct and indirect (tethering to other DNA-bound proteins) GR binding and captures information about TFs other than the one directly targeted by the antibody. We show that the absence of classical recognition sequences can be explained, in part, by direct GR binding to degenerate canonical GR binding sequences. Furthermore, my study identified a new mode of DNA binding, where GR binds as a heterodimer together with a member of the ETS or TEAD families of TFs. Finally, our studies indicate that GR can be recruited indirectly to the genome via FOX or STAT proteins. Together, our generically applicable *footprint*-based approach uncovers new structural and functional insights into the diverse ways of genomic cooperation and association of GR.

3 Zusammenfassung

Zelltypen entwickelten sich, um die unterschiedlichen Funktionen eines Organismus zu erfüllen. Die spezifische Regulation der genetischen Information ist essentiell für die Aufrechterhaltung der einzelnen Zellphänotypen und wird durch die Bindung von Transkriptionsfaktoren an DNA-Sequenzen innerhalb regulatorischer Regionen ermöglicht. Untersuchungen an dem Glucocorticoid-Rezeptor (GR), einem hormonabhängigen Transkriptionsfaktor (TF), zeigten, dass nur ein kleiner Anteil aller möglichen genomischen GR-Bindesequenzen tatsächlich von GR gebunden wird. Zudem weist nur ein Bruchteil aller GR-Bindestellen eine klassische GR-Bindesequenz auf. Dies lässt den Schluss zu, dass diese DNA-Sequenz weder notwendig, noch ausreichend ist, um zu erklären wo GR im Genom bindet. Es stellen sich daher zwei Fragen: 1.) Lassen sich Unterschiede in der Chromatinlandschaft identifizieren, welche die genomischen Bindestellen von GR erklären und 2.) welche Sequenzen, neben den klassischen GR-Bindesequenzen, bewirken eine genomische Binding des GR?

Ein Merkmal des Chromatins ist die Organisation von DNA in Nukleosomen, deren Verpackungsgrad das Genom in für TF zugängliche „offene" und „geschlossene" Bereiche teilt. Unsere GR-Chromatin-Immunpräzipitationen (ChIP) wiesen eine deutliche Präferenz der Binding von GR zu „offenem" Chromatin auf, allerdings mit einer Besonderheit: GR bindet in „offenem" Chromatin deutlich vermindert an Promoterregionen. Dies konnte teilweise durch eine verminderte Häufigkeit an Sequenzen, die dem klassischen GR-Motiv entsprechen, erklärt werden. Wir fanden jedoch auch Hinweise für zusätzliche Mechanismen. *Hierarchical Bayes Modeling* von GR-Bindestellen und Chromatinmerkmalen zeigte eine negative Korrelation von promoterspezifischen Histonmodifizierungen (z.B. H3K9ac) und der Binding von GR. Die für die Histonacetylierung verantwortlichen Histonacetyltransferasen verändern ebenfalls GR posttranslational und schwächen somit seine Interaktion mit der DNA (Kino & Chrousos 2011), was eine zusätzliche Erklärung für die von uns gefundene, verminderte promoterproximale Bindung von GR sein könnte.

Da die Sequenzanalyse von GR-Bindestellen eine große Anzahl GR gebundener Regionen aufwies, die keine klassische GR-Bindesequenz enthielten, fragten wir uns, welche zusätzlichen Sequenzen in der Lage sind, GR an spezifische Loci zu rekrutieren. Um solche Sequenzen mit hoher Auflösung zu identifizieren, verwendeten wir eine modifizierte Version des klassisches ChIPs bei der eine Exonuclease die ChIP-Fragmente bis zur Crosslink-Stelle von Protein:DNA-Interaktionstelle degradiert (ChIP-Exo). Die zuvor durch Formaldehyd-Quervernetzung fixierte Interaktionsstelle schützt die DNA vor weiterer Degradierung. Mittels unseres ExoProfilers, einem computergestützten und motivbasierten Ansatz zur Identifizierung von TF-Bindestellen, wurden die resultierenden Next-Generation-Sequencing-Signale zu sogenannten *footprint profiles* von TF-Bindestellen mit basenpaargenauer Auflösung verrechnet. Der Vergleich unserer und der wenigen frei verfügbaren ChIP-Exo-Daten machte deutlich, dass *footprints* protein- und sequenzspezifische Merkmale von TF-Bindestellen sind. Dieses Wissen ermöglichte uns, zu unterscheiden, ob GR direkt oder indirekt (über

ein zusätzliches Protein; tethering) an die DNA gebunden ist. Zudem erkannten wir, dass ChIP-Exo Informationen über Transkriptionsfaktoren liefert, die zusätzlich zu dem mit dem Antikörper erkannten TF am Genom gebunden sind. Wir konnten zeigen, dass das Fehlen einer klassischen GR-Bindesequenz teilweise mit dem Binden an degenerierte GR-Motive erklärt werden kann. Darüber hinaus konnten wir anhand meiner Experiment eine neue Art der GR-Binding identifizieren, bei der GR als heterodimer an die DNA bindet, zusammen mit Transkriptionsfaktoren von Mitgliedern der ETS- und TEAD-Proteinfamilie. Abschließend finden sich durch unsere Arbeit Hinweise darauf, dass die Rekrutierung von GR zu bestimmten genomischen Bereichen von Fox und STAT Proteinen bewirkt wird. Zusammengefasst erlangen wir mit unserem allgemein anwendbareren und auf TF-*footprints* basierenden Ansatz strukturelle, sowie funktionelle Erkenntnisse, welche die vielfältigen Interaktionen auf genomischer Ebene im Allgemeinen und des GRs im Besonderen betreffen.

4 Introduction

4.1 Gene Expression and Transcriptional Regulation

Genes hold the information to build and maintain an organism's cells. The process of transforming the information stored in the genomic DNA to a functional gene product is called gene expression. All organisms have to regulate the production, and thus the amount of gene products (proteins or functional RNAs), per cell individually in order to e.g. develop or respond to their environment. Almost all levels of gene expression can be modulated. Starting at the transcriptional initiation, RNA processing and even by changing the final gene products using post-translational modifications of the produced proteins. The study of how expression of genes is regulated is essential to understand the development and maintenance of organisms as well as the origin of cardiovascular, immune or metabolic diseases, which have all been linked (Lee & Young 2013) to changes in the early stage of gene expression: Transcription and its control by transcription factors (TFs).

To know where a TF is binding is key to understanding why a particular subset of genes is be expressed in a given cell type. My thesis therefore focuses on: the role of the DNA sequence and the role of the chromatin landscape in specifying where transcription factors bind to the genome.

Transcription is the process of copying a gene from a protein- or a non-protein coding region of the DNA into RNA (Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter 2006) and is performed by RNA-Polymerases, for example RNA polymerase II (RNAP II), a protein complex consisting of 12 subunits in humans (Sentenac 1985). If a gene is transcribed or not is initially determined by whether RNAP II is recruited to the transcriptional start site (TSS) and can furthermore be regulated by posttranslational modulations of the C-terminal domain (CTD) of Rbp1, the largest subunit of RNAP II. Additional mechanisms to regulate transcriptional rates via modulating RNAP II activity are explained later. The CTD harbors 52 repeats of the amino acids YSPTSPS in humans (Meinhart & Cramer 2004), which can be differentially phosphorylated thus allowing general transcription factors (Orphanides et al. 1996) to interact and further modulate (Schwer & Shuman 2011; de Almeida & Carmo-Fonseca 2014) the regulation of the transcript production by RNAP II. Together with general TFs, the RNAP II forms the RNAP II holoenzyme (Koleske 1994). Since eukaryotic RNAP II is not able to interact with the promoter directly (Malik & Hisatake 1991), recruitment of RNAP II to such regions requires the

assistance of specialized transcriptional regulatory factors (Butler & Kadonaga 2002). TFs are by definition proteins interacting with the DNA via a specific consensus DNA recognition sequences (Latchman 1997). The diversification of the recognition sequences associated with genes to start transcription by TFs and the cell-type-specific expression of TFs enables a fine-tuned regulation of individual genes. Thus the production of specific cohorts of TFs that bind to specific recognition sequences associated with their target genes facilitates the abovementioned task to regulate gene expression in a cell-type-specific manner (Heinz et al. 2010).

The transcriptional process can be sub-divided into four major steps that can be regulated to control the expression of genes. These steps are:

I. Pre-initiation: Five general TFs and the RNAP II form the Pre-Initiation-Complex (PIC) at the core promoter of genes (Conaway & Conaway 2013). Promoters are specific DNA sequences 25-35 base pairs (bp) upstream of the transcriptional start site (TSS) of a gene. A prominent example for a promoter is the so called TATA box (Wobbe & Struhl 1990). The assembly of the PIC is a key point for the transcriptional regulation and is facilitated directly by TFs or indirectly by interactions with the mediator complex. Furthermore, the Pre-Initiation can be blocked at this stage by RNA interference, resulting in a stop of the RNAP II assembly at promoter sequences of the DNA (Pai et al. 2014). TFs can indirectly influence the assembly of the Pre-Initiation-Complex by recruiting the multi-subunit mediator complex (Kelleher et al. 1990), of about a million daltons (Kim et al. 1994). Mediator complex is a central player in PIC assembly as it interacts with both the RNAP II, the general as well as with specialized TFs, to stabilize the transcription initiation complex (Reeves & Hahn 2003). Additionally, mediator complex is a co-activator, co-repressor and a general transcription factor all in one (Kornberg 2007; Chodankar et al. 2014), emphasizing the importance of this complex for gene regulation. Mediator complex is also known for maintaining chromatin in a hyperacetylated and therefore accessible conformation (Lorch et al. 2000).
II. Initiation: After assembly of RNAP II together with general and specialized TFs the next step in the transcriptional process is the initiation of transcription. Transcription is initiated by the release of the RNAP II, a process that can be regulated by TFs bound to enhancer

elements. Enhancers are DNA-encoded TF-binding-sites within or even several 100 - kilo base pairs (kb) up- or downstream of the promoter that control the expression of genes. TFs connect DNA elements over large distances by bending the DNA, demonstrating the complexity of transcriptional regulation (Maston et al. 2006). Additional co-activators and co-repressors can be recruited by specialized TFs to add another level of transcriptional regulation. TFs that regulate transcriptional initiation are for example the Signal Transducers and Activators of Transcription (STAT) proteins, which belong to the class of specialized TFs. STAT proteins bind site-specifically with other proteins forming enhanceosomes (Carey 1998), a protein assembly consisting of general TFs and RNAP II (Panne 2008), to increase transcription initiation (Bromberg & Jr 2000). It is also known that both activated Stat3 and the glucocorticoid receptor (GR) synergistically increase RNAP II recruitment to the enhanceosome and thus the transcriptional initiation (Lerner et al. 2003).

This GR-Stat3 interaction also results in Histone 3 acetylations at the enhanceosome, implying the recruitment of a histone acetyltransferase (Lerner et al. 2003), an important finding emphasizing the connection of transcription to the chromatin landscape, which will be introduced in more detail later.

Before RNAP II is finally released from the promoter to complete transcription, the Transcription initiation complex performs several rounds of abortive transcriptions, a process known as promoter clearance (Goldman et al. 2009). Destabilization of the RNAP II interactions with the other promoter bound proteins leads to the release of RNAP II by phosphorylation of the Serine 5 of the CTD in the Rbd1 subunit of RNAP II (Phatnani & Greenleaf 2006) by the general transcription factor TFIIH (Maxon et al. 1994). This mechanism exemplifies the importance of post-translational modification of the CTD.

III. Elongation: After the release of RNAP II by the phosphorylation of the CTD, the transcriptional process can proceed to transcriptional elongation. To stabilize the nascent RNA strand, a 5′ cap is added. After approximately 50 nucleotides have been synthesized DSIF and NELF bind to RNAP II to repress elongation (Yamaguchi et al. 1999) thereby

allowing capping of the 5′ end. The elongation is restarted by the serine 2 phosphorylation of the RNAP II-CTD by P-TEFb, resulting in the release of DSIF and NELF from RNAP II (Yamaguchi et al. 1999).

TFs can induce pausing as well as stimulation of the elongation process thereby providing another key step in the regulation of transcription. For example, the hormone activated glucocorticoid receptor (GR), a well-studied TF, can cause transcriptional pauses, by preventing p-TEFb recruitment (Gupte et al. 2013). RNAP II stalling can also be facilitated by recruitment of NELF and DSIF, as shown in recent studies showing a co-occurrence of NELF and DSIF at paused RNAP II positions (Luo et al. 2013; Quinodoz & Gobet 2014). In contrast, the transcriptional elongation can also be stimulated. For instance, the TF NF-κB is able to recruit p-TEFb, resulting in RNAP II-CTD phosphorylation (Barboric et al. 2001).

IV. Termination: Once polyadenylation (poly(A)) signals (e.g. the sequence AAUAAA) are transcribed into mRNA, two proteins transfer from the RNAP IIs CTD to the poly(A)-signal: the cleavage and polyadenylation specificity factor (CPSF) and the cleavage stimulation factor (CstF) (Lutz 2008). CPSF is cleaving the mRNA at the newly synthesized 3′ end. CstF is able to recognize DNA sequences other than the CPSF-motif thereby allowing mRNA cleavage even if the CPSF-sequence is missing (Millevoi et al. 2006). The mRNA cleavage site is associated with a poly(A) signal (Iseli et al. 2002). Together with CPSF, CstF and additional factors the ATP:polynucleotide adenylyltransferase (PNAT) is recruited and adds A-repeats of variable length to the 3′end building the poly(A) tail of the mRNA (Zhao et al. 1999). PNATs activity is further increased by the binding of PABP2, which acts also as a "molecular ruler" for the length of the poly(A) tail (Keller et al. 2000). After approx. 250 adenosines (Wahle 1991; Bienroth et al. 1993) have been added to the 3′ end of the mRNA, the interaction of PNAT and CPSF is disrupted leading to the end of polyadenylation (Viphakone et al. 2008) and the formation of pre-mRNA. Since the RNAP II is still attached to the DNA and is still synthesizing unprotected RNA, RNAP II has to be released from the DNA. Two models are describing this. First the torpedo model is hypothesizing a RNase starts to degrade the uncapped 5′ end of the still synthesized RNA until it reaches the RNAP II, resulting in the RNAP II dissociation from the DNA template (Luo & Bentley 2004). Second an

allosteric model is describing the DNA-RNAP II dissociation, at which the poly(A) transcription is inducing conformational changes in the RNAP II, reducing RNAP II's processivity and leading to a weaker interaction of RNAP II to the DNA-RNA substrate (Kuehner et al. 2011) because of a hairpin structure blocking RNAP II processivity which leads to disassembly of RNAP II (Epshtein et al. 2007).

All these four steps of transcription can be regulated by TFs to modulate the transcriptional output of genes.

Another important player in transcriptional regulation is the chromatin landscape in which the transcription takes place. For example, the promoter of a gene can be "hidden" in inaccessible chromatin thereby preventing the transcription of the gene. Furthermore, RNAP II cannot be recruited to promoter sequences in nucleosome rich regions and thus the presence and positioning of nucleosomes can influence whether a gene is expressed or not (Bai & Morozov 2010; Bargaje et al. 2012). We will have a closer look at the DNA packaging in the next chapter.

4.2 Chromatin Landscape: The natural environment of Gene Expression

Recent studies show that chromatin accessibility predetermines where in the genome specific TFs bind (John et al. 2011). It could also be revealed that in turn TFs can change the chromatin landscape and thereby determine which genes are transcribed (Sindhu et al. 2012), emphasizing the functional interplay between chromatin and TFs. In this chapter, I will focus on the various mechanisms coupling the chromatin landscape with transcription.

The chromatin landscape collectively refers to the binding and packaging of DNA around histones and other proteins, posttranslational modifications of these histones and modifications of the DNA like methylation. Chromatin plays a role in compacting the genome and thereby partitioning the DNA into accessible chromatin (called "open" or euchromatin) and "closed" regions defined as heterochromatin. This facultative additional packaging of the DNA is adding an additional level of regulation to the expression of genes. The basic repeating unit of chromatin are nucleosomes which consist of roughly 147 bp genomic DNA wrapped around an octameric set of histone core proteins (Luger et al. 1997; Richmond et al. 1984). Two nucleosomes are typically linked by nucleosome-free DNA of 10 to 80 bp length depending on the cell type (Felsenfeld & Groudine 2003). The shorter the linker length, the smaller the likelihood for a TF binding sequence being accessible for a TF

(Bai & Morozov 2010). Together, the positioning and presence of nucleosomes and thus the level of compaction of the chromatin can influence if TFs can bind and if a given gene can be expressed. Accordingly, altering the nucleosome density is used to regulate gene expression (Song et al. 2011).

In addition to correlating to chromatin accessibility, the level of transcription and the binding of TFs are linked to posttranslational modifications of histones. A large set of chemical groups can be added to the amino acids of the histone tail, to create a hypothetical "histone code" (Jenuwein & Allis 2001). This code refers to the idea that the collective presence of different combinations of histone modifications contains information that influences several processes that take place in the context of chromatin, like for example transcription (Bannister & Kouzarides 2011). Among the vast amount of histone modifications, two prominent examples are acetylation and methylation of histone tails. Acetyl groups are neutralizing the positive charge of the amino acid lysine. This diminishes the electrostatic attraction of the negatively charged DNA backbone and the positively charged lysine of a histone protein. Consequently, the nucleosome packaging loosens as a result of the decreased attraction (Roth et al. 2001; Voet & Voet 2007). Acetylation of lysine at positions 9, 14 and 27 of the histone protein H3, correlates with active transcription of nearby genes (Creyghton et al. 2010; Koch et al. 2007). Another histone modification is methylation, which can occur at lysines as well as at arginines. In contrast to acetylation, methylation does not alter the charge. The addition of one, two or even three methyl groups allows the interaction with proteins containing, for example, PHD domains, able to recognize these modifications and tether additional proteins to the methylated regions of the chromatin. These proteins can in turn alter nucleosome positions (Aasland 1995; Nair et al. 2013; Qiu et al. 2014). Mono-methylation at lysine 4, 9, 27 and 79 of histone 3 correlates with active transcription. The same is true for mono-methylation at lysine 20 of histone 4 and lysine 5 of histone H2B (Barski et al. 2007; Benevolenskaya 2007; Steger et al. 2008). Di- and tri-methylation of lysine does not result in a uniform response: Whereas the di-methylation at lysine 9 and 27 of histone 3 correlates with reduced transcriptional activity (Rosenfeld et al. 2009), the di-methylation of lysine 79 of histone 3 correlates with active transcription (Steger et al. 2008). Similarly, tri-methylations have a heterogeneous effect on transcription: tri-methylation at lysine 9 and 27 of histone 3 correlate with transcriptional repression (Barski et al. 2007), whereas tri-methylation at lysine 4 of histone 3 is a mark of active transcription (Koch et al. 2007).

These different modifications regulate chromatin structure, and by doing so they add an additional layer for transcriptional regulation. They define and restrict regions as protein-DNA interaction sites, for example enhancers, which are found to be bound by TF and have high histone 4 lysine 27 acetylation (H3K27ac) and histone 3 lysine 4 mono-methylation (H3K4me1) marks (Creyghton et al. 2010). Promoters are defined not only by their sequence, they have also enriched H4K4me3 and H3K9ac marks together with a depletion of nucleosomes (Liu et al. 2005; Koch et al. 2007). Heterochromatic regions have enriched H3K27me3 and H3K9me3 marks (Lachner et al. 2001), whereas euchromatin is generally marked by lysine acetylations of histones H2A, H2B, H3 and H4 (Wolffe & Pruss 1996; Grunstein 1997; Kurdistani et al. 2004). TSS of actively transcribed genes can be distinguished from inactive ones by H3K4me3, H3ac and H3K4me2 signals downstream from the TSS (Koch et al. 2007). In addition to defining genomic region as actively transcribed, bound TF are also associated with certain histone marks like H3K27ac (Bernstein et al. 2012).

Whether the histone modifications are caused by RNAP II elongation, TF binding or nucleosome repositioning or if these processes are consequences of these histone modifications is typically unclear, since underlying mechanisms are often not yet discovered (Henikoff & Shilatifard 2011). What remains obvious is the strong correlation of TF binding, transcription and histone modifications possibly reflecting the above mentioned histone code (Ong & Corces 2014).

Open, accessible regions in the chromatin can be determined by DNase I hypersensitivity (DHS) assays combined with next generation sequencing (Crawford et al. 2006). Therefore isolated DNA is digested with the endonuclease DNase I, whereat nucleosome dense regions (or “closed“ chromatin) are less like degraded by DNase I and yield to less signal. The digested ends were enriched, sequenced to be mapped to the genome (Song & Crawford 2010). Open regions are cell-type-specific (Waki et al. 2011) and strongly correlate with binding of TFs (John et al. 2011), therefore the “open“ chromatin also explains big parts the cell-type-specific TF binding (Xi et al. 2007). This leads to the question: How is this cell-type-specific chromatin landscape established?

The cell-type-specific chromatin landscape is established during cell differentiation in the embryogenesis by two mayor mechanisms: autonomous and conditional specification. Whereas the autonomous specification is relying on a specific set of TFs, enabling the expression of the cell specific gene products

(Whittaker 1973), the conditional specification needs external stimuli, like concentration gradients of messenger molecules (Guo et al. 2010), for activating or repressing specific gene products in order to maintain and progress the cell fate. This leads to the second part the cell-type-specific response. The former inaccessible regions in the genome have to be opened up to be addressed by TFs in order to express the genes needed for certain cell types. This rearrangement of the chromatin landscape is facilitated by pioneering factors (Zaret & Carroll 2011). One external stimulus are glucocorticoids, a class of steroid hormones that bind to the glucocorticoid receptor (GR). GR is a hormone-activated TF able to act as a pioneering factor itself (Becker et al. 1984; Richard-Foy & Hager 1987). Opening of chromatin is facilitated by the pioneering factors as they are able to recognize specific histone modifications and displaces linker histones (Cirillo et al. 2002), thus making the DNA accessible for TFs (C. S. Lee et al. 2005). Similarly, binding of specific proteins can help convert euchromatin into heterochromatin by the recruitment of, for example, Heterochromatin Protein 1 (HP1), making chromatin inaccessible for TFs and thus can also establish cell-type-specific chromatin landscapes (Wreggett et al. 1994).

These mechanisms have evolved to target subsets of genes from the shared genomic information encoded in the genome of all eukaryotic cells of an organism and enable cell-type-specific transcription. This was a key in allowing multi-cellular organisms to produce different cell types that perform specific tasks.

Since this demonstrates the tight functional connection between chromatin structure and transcriptional regulation by TF binding, one of the aims of my PhD thesis was to study chromatin features that correlate with genomic GR binding to gain mechanistic insight into their interplay. I therefore worked together with Dr. Michael I. Love, as he was analyzing the chromatin landscape of several cell lines to gain insights to the Epigenome of cells in his PhD thesis.

4.3 Transcription Factors

General TFs (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH) are involved in the formation of the pre-initiation complex at the core promoter region close to the transcriptional start sites of a gene and thereby in the regulation of many genes. They do, in general, not interact with the DNA directly. Only a few general TFs bind sequence specific (Kornberg 2007), for example the TATA-Box binding protein (Wobbe & Struhl 1990).

Specialized TFs bind to either enhancer or promoter regions, thus regulating only specific, typically small, subsets of genes by either (I) recruiting the mediator complex including RNAP II to activate transcription, (II) recruiting histone acetyltransferases, histone deacetylases or other chromatin remodelers which can either activate or repress transcription by altering nucleosome positioning or even have an intrinsic acetyltransferase activity (Narlikar et al. 2002), or (III) recruiting additional co-activator or co-repressor proteins for again activation or repression of transcription (Xu et al. 1999). The protein-DNA interaction of TFS is facilitated by DNA binding domains (DBDs) (Kummerfeld & Teichmann 2006), independently folded protein-domains able to interact with specific DNA sequences (Harrison 1991). A DBD was found in approx. 2600 (Babu et al. 2004) of the roughly 20,000 to 25,000 protein coding genes in humans (Pennisi 2012). Therefore approximately 10 % of all genes are potential TFs, making it the single largest protein family, exemplifying the importance of the transcriptional regulation. TFs are not only expressed in a cell-type-specific manner (Kim et al. 2007), their actual bound loci also differ between cell types (Choukrallah & Matthias 2014). This different binding behavior of TFs can be further expanded by a cell-type-specific set of TFs, orchestrating the cell-type-specific gene regulation (as shown by the ENCODE consortium) and building gene regulatory networks (Kim & Park 2010). There are several DBDs known, including the Helix-turn-helix (e.g. the ETS transcription factor), Zinc finger (e.g. the glucocorticoid receptor) and Leucine zipper containing DBDs (e.g. the transcription factor AP1).

The mentioned TFs are all specialized TFs, meaning they target and regulate specific genes via interaction of their DBD with certain DNA consensus sequences, also known as transcription factor binding sites (TF-BSs). Notably, not all bases of a TF-BS must have contact with the TF, which allows the bound motifs to be degenerated and creates a high number of possible TF-BSs in the genome. This leads directly to one of the underlying question of my thesis:

What defines a transcription factor binding site?

As mentioned above, the motif for a TF is degenerated and the chromatin landscape is highly flexible. Motifs are rather short, only several bases long and can therefore frequently occur by chance in the genome and accordingly most potential TF-BSs found in the genome are actually not bound (Wiench et al. 2011). A well-studied TF is the glucocorticoid receptor (GR). As mentioned earlier, GR's TF-function is

inducible by glucocorticoids, introducing an elegant way to study transcription as this allows to "switch" on the DNA binding of GR.

4.4 Glucocorticoid Receptor and the Steroid Hormone Receptor Family

The glucocorticoid receptor belongs to the family of steroid hormone receptors, which have been instrumental in advancing our understanding of various aspects of transcriptional regulation, since these types of nuclear receptors are all activated by ligands. This "switch" allows one to identify regulated target genes and GR binding sites, since GR is only binding to the DNA, once activated by its ligand.

Nuclear hormone receptors (NR) play important roles in the control of development and metabolism by regulating the expression of genes. NRs respond to steroid or thyroid hormones circulating in the bloodstream and enter cells via diffusion (Oren et al. 2004). Hormone activated nuclear receptors act as TFs by binding to their cognate sequences to change transcription of genes as a cellular response to the hormone signaling in the body of metazoans (ESCRIVA et al. 1998). NRs share a common modular structure (Figure 1) consisting of several domains (Kumar & Thompson 1999): Located at the N-terminus, one finds the ligand independent activation function 1 (AF1). AF1 changes conformation upon DNA binding of the nuclear hormone receptor (Kumar et al. 1999), this enables interaction with certain co-regulators (Kumar et al. 2001). Interaction with the DNA is mediated by the zinc-finger DNA binding domain (DBD) that is flanked by a flexible hinge region connecting the DBD and the ligand binding domain (LBD). In addition to the LBD, the C-terminus contains the activation function 2 (AF2) domain (Egea et al. 2000; Pike et al. 2000). Activity of the AF2 domain is hormone-dependent and mediates interactions with several co-repressor and co-activator proteins upon hormone activation (vom Baur et al. 1996).

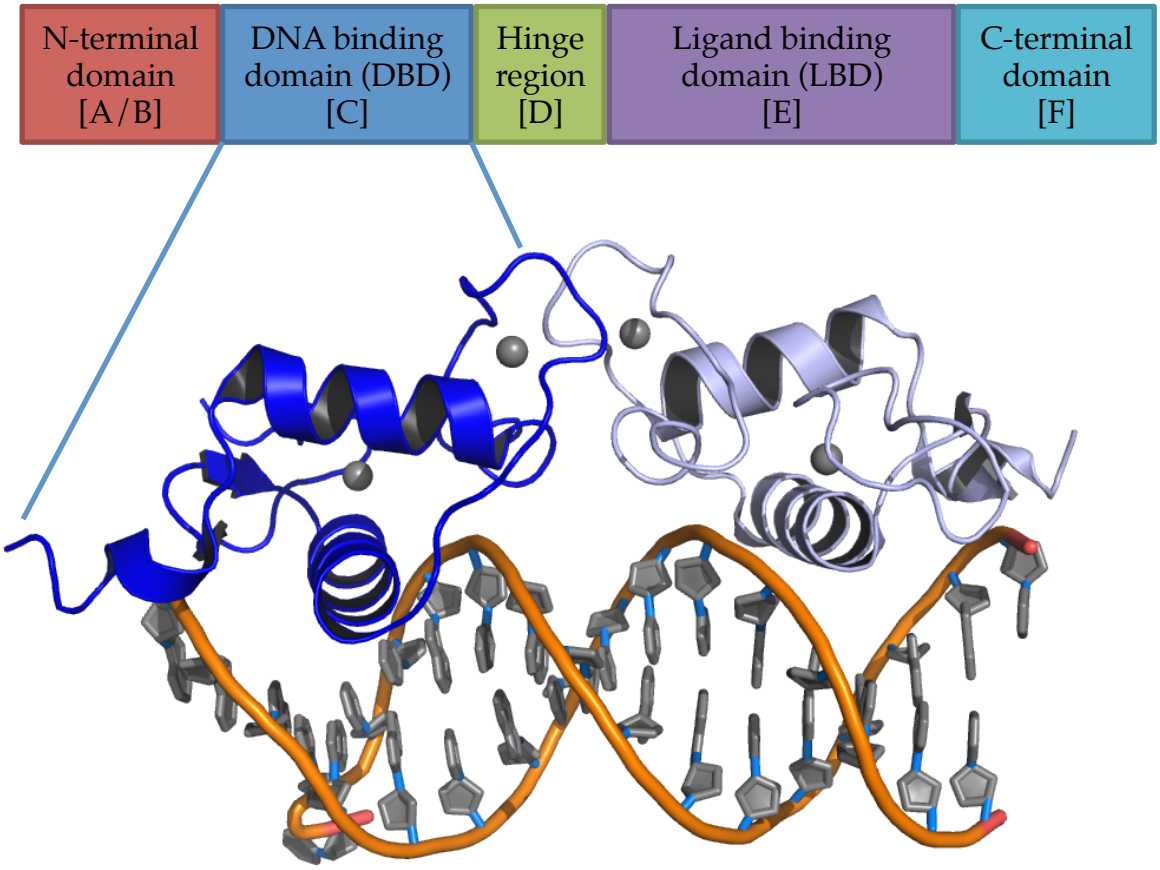

Figure 1: top: Modular structure of nuclear hormone receptors: (size of boxes does not indicate relative proportions of GR). bottom: Chrystal-structure of two GR-DBDs (blue and light blue) binding as dimer to a palindromic canonical GR sequence. Grey spheres indicate Zinc-atoms, necessary for the protein-DNA interaction by GR's zinc finger domains. Structural data visualized with PyMOL® based on PDB: 3G6U, (Meijsing et al. 2009).

Nuclear hormone receptors bind to their individual cognate sequences either directly or alternatively they can bind DNA indirectly via so-called tethered binding to other DNA-associated proteins (Gupte et al. 2013; Kassel & Herrlich 2007). The diverse mechanism underlying the recruitment of nuclear hormone receptors to the DNA together with its nature to be active only once the hormone is present making this class of TFs an ideal model to study transcriptional regulation. The NR family is divided into: Estrogen Receptor-like receptors (class A), Estrogen related receptor (class B) and 3-Ketosteroid receptors (class C). Here I will focus on glucocorticoid receptor (GR), which belongs to the class C nuclear receptors. In addition to GR, class C contains also the Mineralocorticoid Receptor, the Progesterone Receptor and the Androgen Receptor. GR is expressed throughout the body and effects the transcriptional regulation of a cell-type-specific subset of target genes including metabolic and immunologic genes (Yamamoto 1985; Gustafsson et al. 1987; Beato 1989). In the absence of hormone, GR predominantly resides in the cytosol and is stabilized by several heat shock proteins (e.g. Hsp70 and Hsp90) (Grad & Picard

2007; Pratts 1993). Hsp90 keeps the ligand binding domain of GR in a high affinity state towards its ligand (Pratt et al. 2006). GR is activated by steroid hormones called glucocorticoids (GC), for example cortisol in humans. Upon binding to GC, GR translocates to the nucleus (Yamamoto 1985; Becker et al. 1986) by exposure of two nuclear localization signals (Picard & Yamamoto 1987) and is also actively transported using the microtubule network (Czar et al. 2013; Harrell et al. 2004). After relocalization to the nucleus GR binds as a homodimer (Wrange et al. 1989) at palindromic cognate DNA sequences. The canonical GR binding sequence (GBS) consists of two inverted hexameric repeats, each interacting with a GR monomer, separated by a 3 bp spacer (Dahlman-Wright et al. 1990). Binding to the DNA changes the structure of the GR in a sequence dependent manner, adding an additional layer of regulation of the activity of GR (Meijsing et al. 2009). GBSs associated with GR target genes can either localize to the promoter regions or, more typically, are localized at great distances from the transcriptional start site of genes (Reddy et al. 2009). These interactions will be explained in chapter 4.5.

4.5 GR interactions with the genome

Activated GR can regulate gene expression in two ways: transrepression and transactivation (Rhen & Cidlowski 2005). Transrepression happens to be indirect, where an activated GR forms a complex with other transcription factors, thus preventing them to bind to DNA and hence repress the up-regulation of their target genes (Ray & Prefontaine 1994; Karin 1998). For direct transrepression, GR is able to repress the up-regulation of genes by tethering to TFs already bound to the DNA (Figure 2, bottom left) (De Bosscher et al. 2008; Langlais et al. 2012; Ratman & Berghe 2013) at so called negative Glucocorticoid Receptor Response Elements (nGREs). These nGREs contain a TF binding sequence but not a GR binding sequence (GBS). Examples for tethered binding of GR with combined transcriptional repression are interactions with AP1 (Biddie et al. 2011), NF-κB (Cato & Wade 1996), Stat3 (Langlais et al. 2012) or Stat5 (Stöcklin et al. 1996). In addition, some studies found exceptions, where activated GR binds directly to nGREs without being tethered and represses gene expression at such elements (Sakai & Helms 1988; Surjit et al. 2011).

I will focus on genomic GR-interactions, leading to activation of genes by binding of GR to the GBSs, a mechanism called transactivation. GR binds predominantly as a homodimer (Dahlman-Wright et al. 1990; Wrange et al. 1989) and GR binding studies revealed a large set of possible GR-DNA interactions (Figure 2). Besides the direct and exclusive DNA interaction of the GR homodimer (Figure 2, top), GR can perform

composite binding (Figure 2, right) with other proteins (Langlais et al. 2012). It was also shown that other proteins can be tethered to GR, resulting in transcriptional repression of GR induced genes (Figure 2, middle left) (Langlais et al. 2012). A rarely found GR interaction is the so-called GR-heterodimer at a combinatorial sequences, containing the binding sequences of a GR halfsite and a second TF (Wu & Bresnick 2007). This binding event will be analyzed in detail in chapter 7.3.6 (page 88).

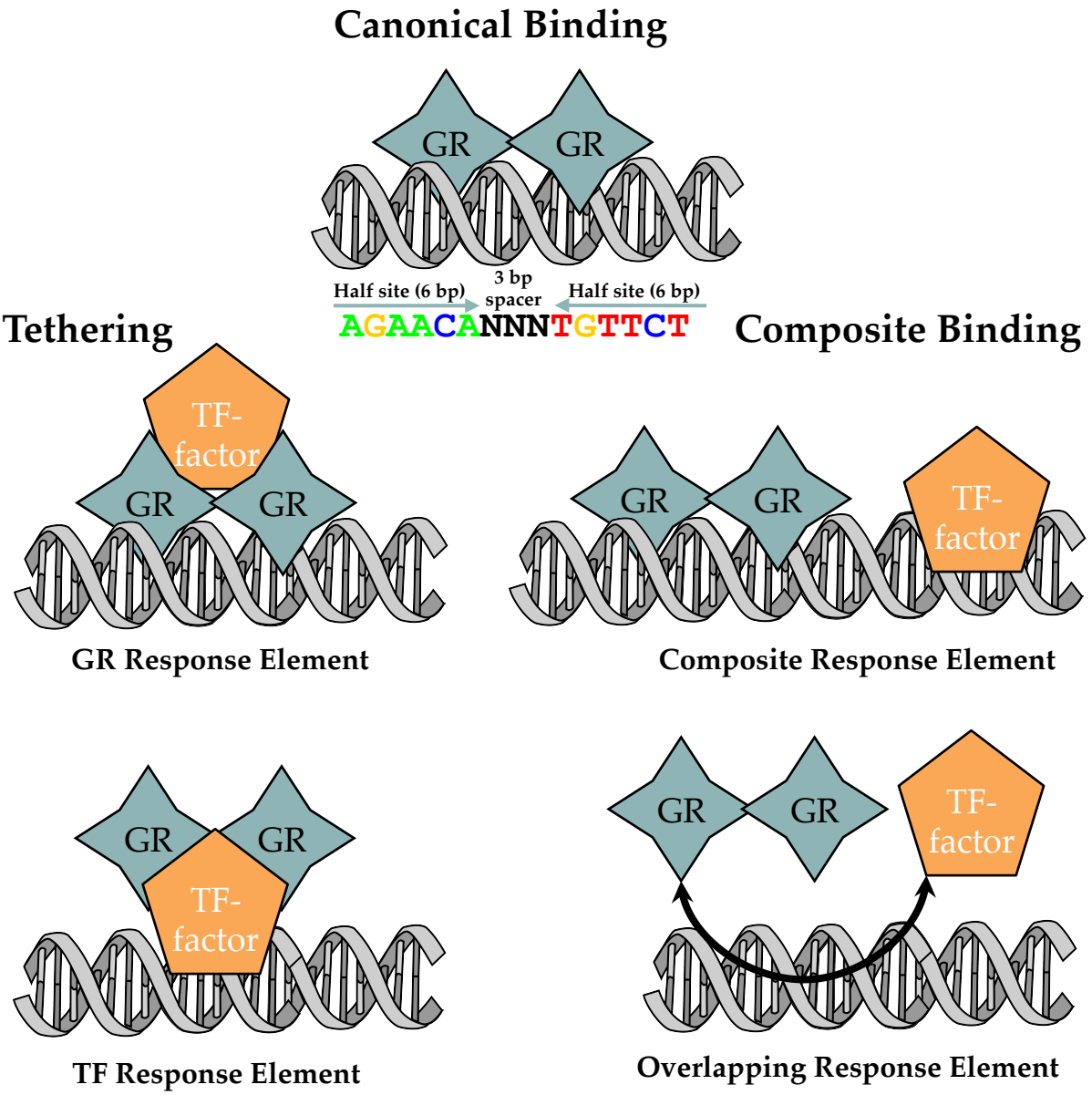

Figure 2: Overview of GR binding modes. Besides dimeric binding at canonical GR motifs (top), GR can either tether (middle left) other transcription factors (TF) or be tethered to the DNA by other TFs (lower left). In addition, GR can bind in combination with other TFs (middle right), requiring the full presence of the response elements for both TFs. Contrary, for overlapping response elements, only one TF can bind at a time resulting in a competitive binding of GR and other TFs (lower right).

The combined action of GR and GR-cofactors allows transcription of GR target genes to be fine-tuned. GR co-factors can expand the spectrum of potential GR

targeted genes, by tethering GR to new motifs or by making chromatin regions accessible for GR, as it was shown for AP1 (Biddie et al. 2011). This exemplifies again the complex network of interactions between the chromatin landscape and individual TFs and between TFs, which may have even dual roles in transcriptional regulation and chromatin remodeling, like AP1 (Milavetz 2002). As the binding of GR to the genome is influenced by the chromatin landscape in which a GBS is embedded (John et al. 2011), I will first study the role of the chromatin landscape in specifying where in the genome transcription factors bind (chapter 4.6). Second, I will focus on sequences in the genome, which are responsible for the recruitment of GR to individual loci (introduced in chapter 4.7).

4.6 What defines a TF binding site? - Role of the chromatin landscape

Previous TF binding studies already emphasizes the discrepancy between the canonical TF motif presence in the genome and the actual bound sites (Lee & Bhinge 2012). For GR only about 1 out of 1000 GR motifs present in the genome is actually GR-bound and these subsets of GR binding sites vary in a cell-type-specific manner (John et al. 2011). This phenomenon is known for several TFs (Gertz et al. 2013) and could be extended to RNAP II positioning. RNAP II was found frequently bound (approx. every 3 kb in the human genome) at functional promoters, building up promoter transcription initiation complexes (PTICs) (Venters & Pugh 2013). This large number (approx. 500,000) of PTICs differs from the actual number of protein coding genes (20,000 – 25,000; (Pennisi 2012)) and also the number of possible non-coding genes (18,400; (Pennisi 2012)) would not add up the gap. Together roughly every 3000 bases a functional RNAP II is positioned potentially ready to transcribe DNA and a matching GR motif is statistically present approx. every 1000 bp (Polman et al. 2012).

Like for RNAP II and other TF, the question arises also for GR: What differentiates GBS-like sites in the genome that are bound by GR from unbound ones? The chromatin landscape provides substantial information to guide GR and TFs in general to their actual binding site (Gao et al. 2004; Arvey et al. 2012; Kasinathan et al. 2014; Wang et al. 2014). In particular the accessibility (or "openness") of the chromatin restricts which putative GR binding sequences (GBS) are actually bound (John et al. 2011). It also determines the access of TF to regulatory regions (Han et al. 1988) that define the spatial and temporal transcriptional regulation of genes exemplified by the cell-type-specific chromatin accessibility of regions distal to a genes transcriptional start site (Waki et al. 2011). The chromatin

accessibility is driven by nucleosome density, as more nucleosomes limit TF binding in vivo (Belikov & Åstrand 2004). Even if GR in particular binds more specifically to certain regions if the DNA is organized by nucleosomes (Perlmann 1992), the presence of nucleosomes also compromises GR binding *in vitro* (Piña et al. 1990; Archer & Cordingley 1991). Thereby the chromatin accessibility limits the pool of viable GBS.

We therefore analyzed the GR-bound sites in two human cell lines with a well-defined chromatin landscape: the immortalized myelogenous leukemia cell line K562 and the primary lung fibroblast IMR90. IMR90 cells are not cancer derived and have a well-studied chromatin landscape, making them ideal cell lines to study the interplay between genomic binding of TFs and the chromatin landscape. The IMR90 cell line is part of the NIH Roadmap Epigenomics Mapping Consortium (Bernstein et al. 2010) and the ENCODE project (The ENCODE Project Consortium 2004), whereas K562 cells are studied as part of ENCODE project. Both projects created databases with histone marks and data regarding genome-wide DNase I sensitivity. The NIH Roadmap Epigenomics Mapping Consortium currently (September 2014) provides 28 different histone marks as well as DNase and mRNA sequencing data for the IMR90 cells. The Encode project also contains data regarding many histone modifications and in addition covers RNAP II-Seq, several TFs ChIP-Seq and RNA-Seq datasets for IMR90 and K562 cells. Even though these databases contain the chromatin landscape information for cells not treated with hormone, we can work with these datasets, since they represent the state of the chromatin that the GR encounters and interacts with after becoming hormone activated. With this broad range of experimental data for the actual chromatin landscape, the gene expression data and additional TFs being mapped to the genome, we set out to look how these marks correlate with GR binding.

4.7 What defines a TF binding site? - Sequences responsible for GR recruitment to individual genomic loci

The canonical GR binding motif can be represented as a position weight matrix (PWM). These PWMs can be depicted as sequence logos (Figure 3), where the height of the letter at each position reflects the conservation of this specific nucleotide to determine a functional TF-BS (e.g. GR). Not all bases of a TF binding site have direct contact with the bound transcription factor. Moreover, some individual bases that are directly contacted by TFs can diverge from the consensus preference without preventing binding of the TF. Together, this explains why the consensus recognition sequence of TFs is often loosely defined. For GR, the consensus sequence consists of 15 bases, whereof six bases represent the motif of a GR monomer.

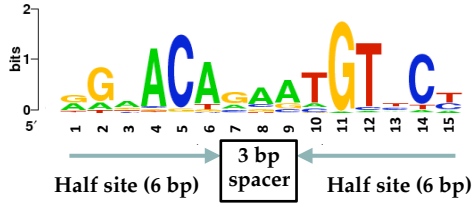

Figure 3: Position weight matrix (PWM) of the canonical GR binding motif. Motive MA0113.2 from the JASPAR database (Sandelin et al. 2004).

The PWM is derived from GR-bound regions obtained from Chromatin Immunoprecipitation (ChIP) experiments (Collas 2010). Briefly, genomic DNA and all interacting proteins are cross-linked using a chemical cross-linker, the nuclei are isolated and the DNA is fragmented. Antibodies (AB) against a specific protein are used to enrich DNA fragments cross-linked to the protein of interest. These DNA fragments are sequenced with Next Generation Sequencing (NGS) and mapped to the genome. Enriched fragments are sequenced more often leading to increased signals at these genomic positions and are called peaks. The peaks are varying in their broadness due to the fragmentation process. In general 300 – 1000 bp fragments are used for ChIP, resulting in peaks of about the same size. These bound regions contain the sequences responsible for the recruitment of the TF of interest that can be identified using computational approaches (Thomas-Chollier, Darbo, et al. 2012). However, the nature of the ChIP-protocol does not permit to conclude how the protein is interacting in detail for individual loci or if the interaction with enriched sequences is direct or indirect. Furthermore, although GBS-like sequences are highly enriched at GR-bound regions, this does not mean that they are found at all GR-bound loci. In fact, for some cell types only a small fraction of GR-bound regions appears to have a GBS-like sequence. This could either reflect the fact that other

sequences are responsible for recruitment or that GR binds to highly degenerate sequences that are found at roughly equivalent frequency at bound and unbound regions and can thus not be identified based on sequence enrichment at ChIP-Seq peaks.

My project aims to understand the GR interactions at individual loci in more detail. Since the classical ChIP-data cannot be used to obtain this information, we used a modified ChIP-protocol that can give genomic binding information at higher resolution. The ChIP-protocol was therefore changed, to reduce the fragment size for NGS to the bases bound by the protein by including an exonuclease digestion step into the ChIP protocol (ChIP-Exo) which leads to 5' digestion of DNA until the cross-linked protein protects the DNA from being degraded (Rhee & Pugh 2011). ChIP-Exo therefore synchronizes the fragments analyzed by NGS and mapped to the genome, which leads to an accumulation of reads at the exact protein-DNA protection site, indicating the position of the TF-DNA cross-link at base pair resolution. Since the protection sites can be identified for both strands individually (compare Figure 4) ChIP-Exo is generating peaks pairs, flanking the site where the TF is bound. The bases flanked by peak pairs specify at high-resolution information about the sequence bound by GR at individual loci. In addition, if the signal is weak at individual binding sites, the accumulation of reads flanking sequence motifs at many bound loci can help uncover additional modes of genomic association of TFs.

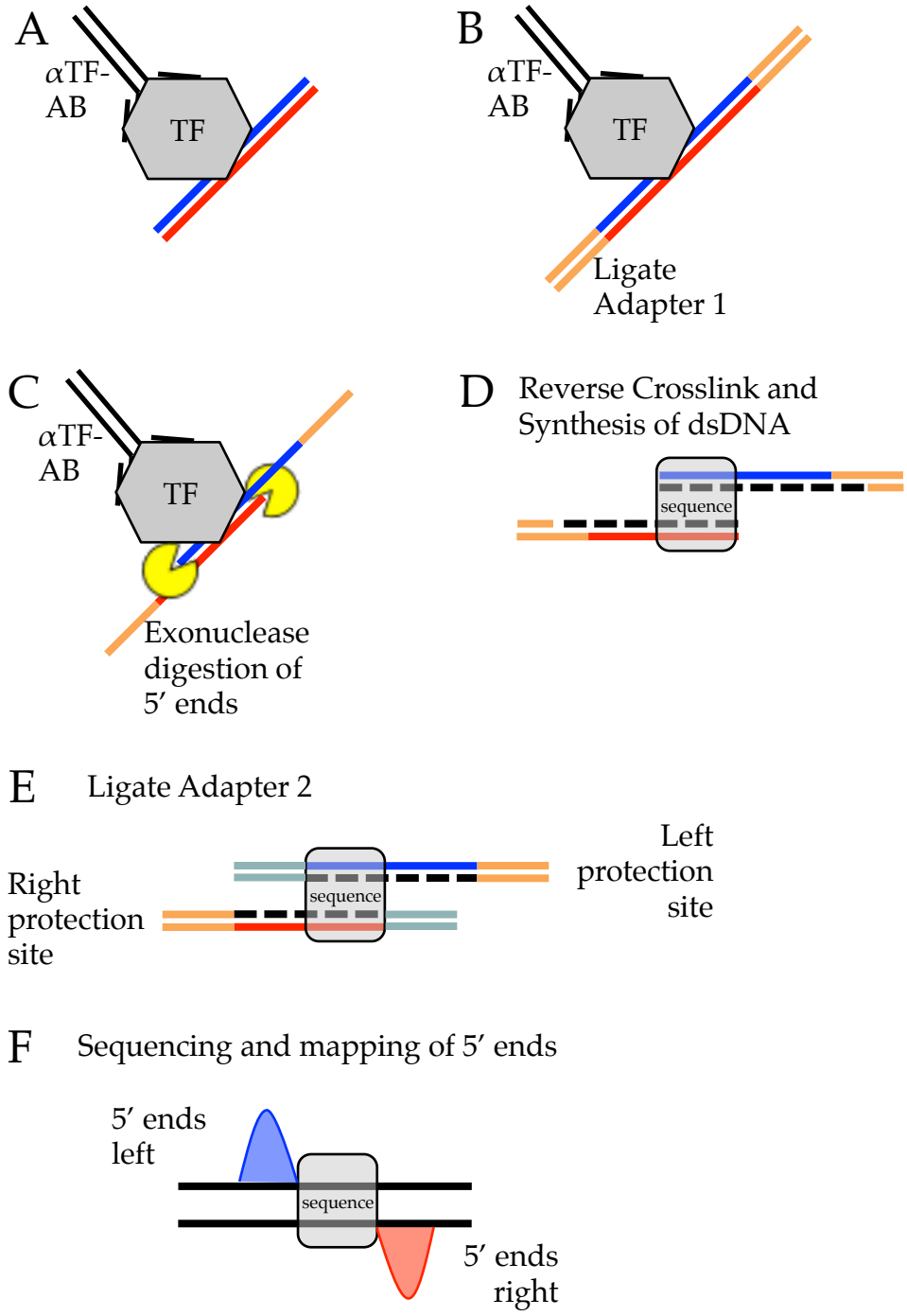

Figure 4: Schematic presentation of the process of Chromatin-Immunoprecipitation (ChIP) combined with Exonuclease treatment (ChIP-Exo) A: Starting material for the ChIP-Exo. After formaldehyde crosslinking of proteins (e.g. a Transcription Factor) to the DNA, DNA is fragmented to ~300 bp fragments by sonication and the fragments bound by the protein of interest were enriched by ChIP using a protein-specific antibody (AB). B: After blunt end polishing Next generation Sequencing (NGS) adapter 1 is ligated to both ends of the DNA fragment. C: Exonuclease digests the 5′ end of the double stranded DNA (dsDNA) until the bound protein stops digestion by sterical hindrance. D: Reverse crosslinking and separation of the two DNA strands, followed by primer hybridization to the ligated NGS adapter 1 and subsequent dsDNA synthesis from the individual strands. E: Ligation of NGS adapter 2 to the phosphorylated ends of the newly synthesized strands. F: Synchronized start points of protein-DNA protection sites (NGS adapter 2 ligation) generate accumulated reads flanking the site of protein:DNA cross-linking, generating peak pairs flanking the site where the TF is bound.

For an in-depth analysis of the data, I relied on collaborations with computational biologist that, for example, developed a tool called ChIP-Exo-profiler (Dr. Morgane Thomas-Chollier and Jonas Ibn-Salem). This tool analyzes ChIP-Exo signals around any sequence of interest, for example *de novo* identified motifs enriched at ChIP-Seq peaks or reference motifs from e.g. the JASPAR database (Mathelier et al. 2014). As output, ExoProfiler summarizes ChIP-Exo signal around motifs of interest with strand specificity as a heatmap of the coverage at individual sites and a footprint of the accumulated signals for all bound sequences combined (footprint profile) for each motif.

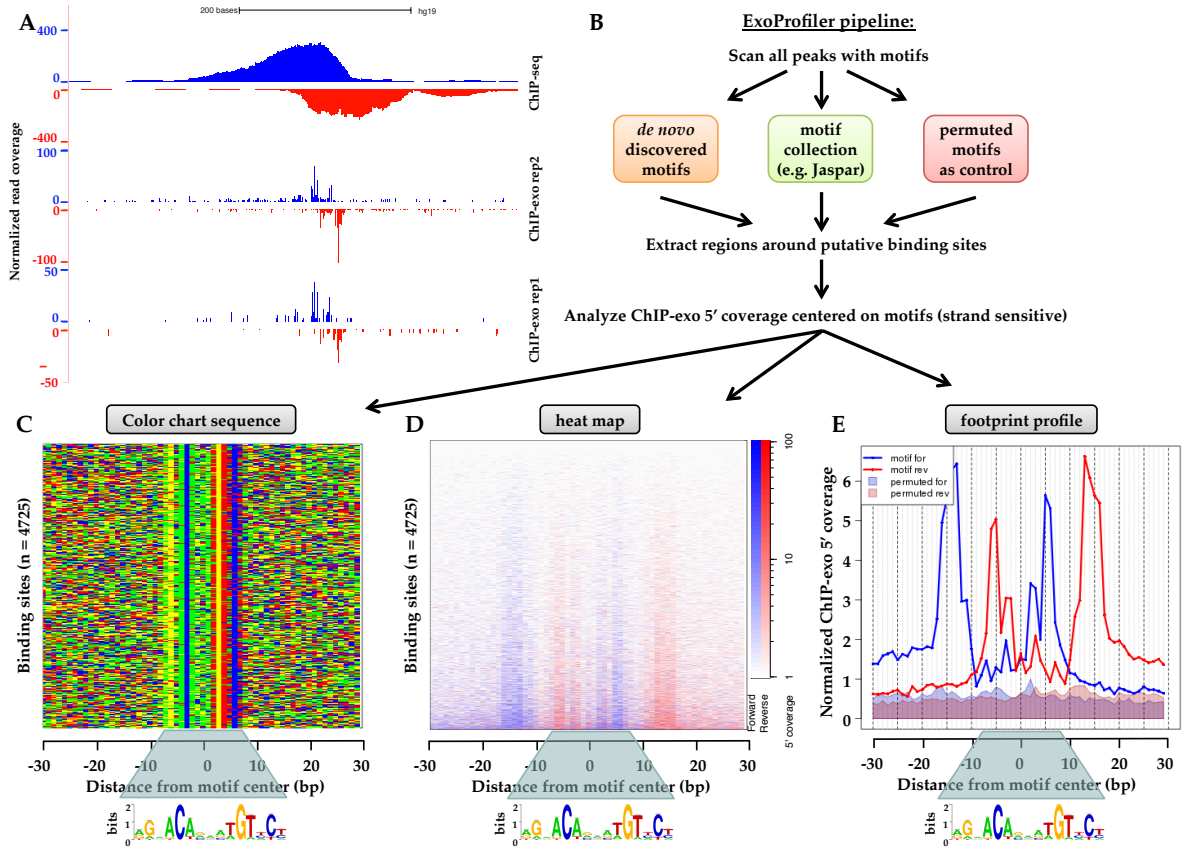

Figure 5: Pipeline for Footprint generation from ChIP-Exo data. A: comparing ChIP-Seq and ChIP-Exo signals at a given loci revealed individual protein binding rather (ChIP-Exo) than enriched genomic region (ChIP). B: Exo-Profiler-Pipeline starting with called peaks from ChIP-Seq and scanning them for either known motif from databases or *de novo* motif discoveries. Next, ChIP-Exo coverage around these motifs is extracted and the coverage is compared with the pattern observed for permutated motifs as control. As output, ExoProfiler produces: a Color chart sequence (C), heat map (D) or footprint profile (E).

For motifs identified in ChIP-Seq data, we will generate the according footprint profiles from ChIP-Exo data with our ChIP-Exo pipeline. These footprints reflect the actual protein occupancy accumulated over several individual loci. The sharp peaks could be obtained by the synchronized mapping of 5′ ends of the ChIP-Exo reads.

Depending on the strength of the signal at individual sites, we can either assign the sequence bound at individual loci when the signal is strong. Alternatively, when the signal at individual loci is weak, one can analyze the footprint profiles that results from multiple binding sites which allows one to determine if such sequence are directly bound, however in this case without getting information about binding at individual loci.

Sequences, with convincing footprint profiles, will be analyzed in detail to answer the questions: 1. Are such sequences directly or indirectly, through tethering, bound by GR? 2. Is GR is able to induce transcription when binding to these sequences?

Together, our combined approach has the potential to uncover GR binding modes apart from the classical dimeric binding to canonical motifs. Due to the increased resolution if ChIP-Exo, we reasoned that we might be able to differentiate between dimeric GR binding to canonical, but degenerative motifs and novel GR binding modes like the rarely identified GR heterodimerization (Kumar & Thompson 1999; Wu & Bresnick 2007). The identification is useful to understand the diversified GR interactions, as the might uncover novel mechanisms for the cell-type-specific and hormone induced transcriptional regulation by GR.

5 Materials

5.1 Chemicals

Acrylamide	Carl Roth
Agarose	Biozym
APS	Merck
BSA	Rockland
Calcium chloride	Calbiochem
Chloroform	Merck
Dexamethasone	Alfa Aesar
Dimethyl sulfoxide (DMSO)	Serva
Disodium hydrogen phosphate	Merck
Dithiothreitol (DTT)	Carl Roth
DMEM	Gibco
Dynal MyOne C1 streptavidin magnetic beads	Invitrogen
EDTA	Carl Roth
EMEM	Gibco
Ethanol	Merck
FBS	Gibco
Formaldehyde	Sigma Aldrich
GlutaMAX	Gibco
Glycerol	Merck
Glycine	Merck
Glycogen	Thermo Fischer
HEPES	Carl Roth
IGEPAL	Sigma Aldrich
IMDM	Gibco
Lithium Chloride	Merck
Magnesium chloride	Merck
NP40 alternative	Calbiochem
PBS	Gibco
Phenol	Carl Roth
PMSF	Carl Roth
Poly(dI-dC)	Sigma Aldrich
Polyvinylpyrrolidone (PVP)	Sigma Aldrich
Potassium chloride	Carl Roth
Potassium dihydrogen phosphate	Merck
Potassium Glutamate	Sigma Aldrich
Protein A/G beads	Santa Cruz
Proteinase inhibitor cocktail set III EDTA free	Calbiochem
Proteinase K	Ambion
ROX reference dye	Invitrogen
SDS	Carl Roth
Sodium carbonate	Merck
Sodium chloride	Carl Roth
Sodium deoxycholate	Carl Roth
Sodium fluoride	Sigma Aldrich
Sodium hydrogen carbonate	Merck
Sodium orthovanadate	Sigma Aldrich
Spermidine	Sigma Aldrich

Spermine	Fluka
Sucrose	Carl Roth
SYBR Green	Invitrogen
Temed	Carl Roth
Tris	Carl Roth
Triton X 100	Sigma Aldrich
Tween 20	Fisher Scientific
Beta-Glycerophosphate	Sigma Aldrich

5.2 Enzymes, Proteins, DNA Kits

1 kb DNA ladder	Fermentas
BglII	Thermo Fischer
Cfr9*I*	Fermentas
DNase I	Qiagen
DNeasy Blood & Tissue Kit	Qiagen
dsiRNA	Integrated DNA Technologies
Dual luciferase reporter assay system	Promega
Lipofectamine	Invitrogen
Lipofectamine 2000	Invitrogen
Nucleospin Gel and PCR clean-up	Macherey and Nagel
Oligos	Sigma Aldrich
p6R	Invitrogen
pcDNA3.1-rGR	Promega
PCR purification Kit	Qiagen
Pfu ultra polymerase	Agilent
pGL3-Promoter	Promega
Power SYBR® PCR master mix	Applied Biosystems
pRL (CMV)	Promega
RNase-A	Applichem
RNeasy Mini Kit	Qiagen
XmaI	Thermo Fischer
Z-Competent E.coli transformation Kit & Buffer Set	ZYMO RESEARCH
Zero Blunt cloning kit	Invitrogen

5.3 Antibodies

Anti glucocorticoid receptor antibody (N499): Raised against a polypeptide with the N-terminal amino acid sequence (Luecke & Yamamoto 2005) of the rat GR (residues 1-499) (R.M. Nissen, B. Darimont, and K.R. Yamamoto, unpublished).

5.4 Labware

1 kb DNA ladder	Fermentas
6 well multititer plate	TPP
24 well multititer plates	TPP
48 well multititer plates	TPP
384 multititer plate (white)	Corning
Agarose-gel chamber	homemade
Bioruptor Plus	Diagenode
Microcentrifuge	Eppendorf
NanoDrop	PeqLab
plastic cover for qPCR plates	Sarstedt
Power supply EV243	Consort
Qubit® 2.0 Fluorometer	Life Technologies
qPCR plates	Sarstedt
Thermocycler Mastercycler gradient	Eppendorf
Thermomixer	Eppendorf
tissue culture flasks	TPP

5.5 Oligonucleotides

Table 1: Oligos from Sigma Aldrich used for cloning

#	Name	Sequence
C067	C067_hs+TEAD_2_T2_F	CCCGGG TCCAAATGCTGCAGCTTTAA
C068	C068_hs+TEAD_2_T2_R	AGATCT GAAAGTTCTCTTAAATAGCA
C124	C124_3xFoxq1_F	CCGGGAAATAAACAAACGCGAAATAAACAAACGCGAAATAAACAAAA
C125	C125_3xFoxq1_R	GATCTTTTGTTTATTTCGCGTTTGTTTATTTCGCGTTTGTTTATTTC
C127	C127_Fox_1_F	CCCGGGCACAGTCTATTCCAAGACTC
C128	C128_Fox_1_R	AGATCTCAGTAGATTTAACTATTGTT
C135	C135_Fox_5_F	CCCGGGAGGCTTATTTTTCTTCATGTTG
C136	C136_Fox_5_R	AGATCTATGCTACTTCCTAACTTACC
C137	C137_Fox_6_F	CCCGGGAAATGATTTTAGGTCTTAGT
C138	C138_Fox_6_R	AGATCTATCAGACTTCAGTTGTGAGA
C143	C143_Fox_9_F	CCCGGGCTTTCTCTCTCTTTCTTTCT
C144	C144_Fox_9_R	AGATCTTAGTGGTTTAGTCATTTGAG
D059	D059_3xcombiTEAD_F	CCGGGAAAGAACATTCCAGCGAAAGAACATTCCAGCGAAAGAACATTCCAA
D060	D060_3xcombiTEAD_R	AGATCTTGGAATGTTCTTTCGCTGGAATGTTCTTTCGCTGGAATGTTCTTTC
D103	D103_3xHYBRID-newfs-noT_F	CCGGGAAAGAACAAACCAGCGAAAGAACAAACCAGCGAAAGAACAAACCAA
D104	D104_3xHYBRID-newfs-noT_R	AGATCTTGGTTTGTTCTTTCGCTGGTTTGTTCTTTCGCTGGTTTGTTCTTTC
D129	D129_3xC_noTEAD/GR-2_F	CCGGGAAAGAAAATTCCAGCGAAAGAAAATTCCAGCGAAAGAAAATTCCAA
D130	D130_3xC_noTEAD/GR-2_R	GATCTTGGAATTTTCTTTCGCTGGAATTTTCTTTCGCTGGAATTTTCTTTC
D133	D133_Fox1-noGR_F	GTAGTAAAGTCTCAGAAGTACAATTCCAGGACAGAC
D134	D134_Fox1-noGR_R	GTCTGTCCTGGAATTGTACTTCTGAGACTTTACTAC
D135	D135_3xC_noG-nf_F	CCGGGAAAAAACATTCCAGCGAAAAAACATTCCAGCGAAAAAACATTCCAA
D136	D136_3xC_noG-nf_R	GATCTTGGAATGTTTTTTCGCTGGAATGTTTTTTCGCTGGAATGTTTTTTC
D137	D137-HYBRID-1_F	CCCGGGTACTAAGTACTTGGCACATAG
D138	D138-HYBRID-1_R	AGATCTCTTTAATGGCCTCTTTCCC
D141	D141-HYBRID-3_F	CCCGGGTTACTTTAAATTCTTTCTGTATTTG
D142	D142-HYBRID-3_R	AGATCTAAGTTGAATTTGTTGCAAAGCTA
D147	D147-HYBRID-6_F	CCCGGGAGAGACAGACTGAGCAAACA
D148	D148-HYBRID-6_R	AGATCTTATAGGCGTGAGACCCCA

#	Name	Sequence
D151	D151-HYBRID-8_F	CCCGGGTTTCTATGCATTTACGTAATAAAC
D152	D152-HYBRID-8_R	AGATCTCCGGCACCTCTCGGG

Table 2: Oligos from Sigma Aldrich used for EMSA

#	Name	Sequence
D055	D055_combi-EMSA_F	Cy5-GATCTCGAAAAGAACATTCCAGTACCTAT
D056	D056_combi-EMSA_R	ATAGGTACTGGAATGTTCTTTTCGAGATC
D057	D057_COMBI-noTTCC-EMSA_F	Cy5-GATCTCGAAAAGAACAAACCAGTACCTAT
D058	D058_COMBI-noTTCC-EMSA_R	ATAGGTACTGGTTTGTTCTTTTCGAGATC
D115	D115_FOX-EMSA_F	Cy5-GATCTCGAAATAAACAAAATAGTACCTAT
D116	D116_FOX-EMSA_R	ATAGGTACTATTTTGTTTATTTCGAGATC
A005	A005_GBS (pal)-EMSA_F	Cy5-TCGAAGAACAAAATGTTCTTCGA
A006	A006_GBS (pal-EMSA)_R	TCGAAGAACATTTTGTTCTTCGA
A007	A007_random-EMSA_F	Cy5-TCGATACCAAAATATTTGAGTAC
A008	A008_random-EMSA_R	GTACTCAAATATTTTGGTATCGA

Table 3: Oligos from Sigma Aldrich used for Pulldown

#	Name	Sequence
D105	D105_Biotin-1xC-nf_F	[BtnTg]CAAAAGATCGCTGCAGACTTGAACAAAGAACATTCCAACTTTGTC
D106	D106_Biotin-1xC-nf_R	GACAAAGTTGGAATGTTCTTTGTTCAAGTCTGCAGCGATCTTTTG
D107	D107_Biotin-1xC-no-T-nf_F	[BtnTg]CAAAAGATCGCTGCAGACTTGAACAAAGAACAAACCAACTTTGTC
D108	D108_Biotin-1xC-no-T-nf_R	GACAAAGTTGGTTTGTTCTTTGTTCAAGTCTGCAGCGATCTTTTG
D109	D109_Biotin-3xC-nf_F	[BtnTg]CAAAAGATCGCTGCAGACTTGAAC AAAGAACATTCCAGCGAAAGAACATTCCAGCGAAAGAACATTCCAACTTTGTC
D110	D110_Biotin-3xC-nf_R	GACAAAGTTGGAATGTTCTTTCGCTGGAATGTTCTTTCGCTGGAATGTTCTTTGTT CAAGTCTGCAGCGATCTTTTG
D111	D111_Biotin-3xC-noT-nf_F	[BtnTg]CAAAAGATCGCTGCAGACTTGAAC AAAGAACAAACCAGCGAAAGAACAAACCAGCGAAAGAACAAACCAACTTTGTC
D112	D112_Biotin-3xC-noT-nf_R	GACAAAGTTGGTTTGTTCTTTCGCTGGTTTGTTCTTTCGCTGGTTTGTTCTTTGTT CAAGTCTGCAGCGATCTTTTG
D113	D113_Biotin-3xScrambled_F	[BtnTg]CAAAAGATCGCTGCAGACTTGAAC AAATACAATGCAAGCGAAATACAATGCAAGCGAAATACAATGCAAACTTTGTC
D114	D114_Biotin-3xScrambled_R	GACAAAGTTTGCATTGTATTTCGCTTGCATTGTATTTCGCTTGCATTGTATTTGTT CAAGTCTGCAGCGATCTTTTG

Table 4: Oligos from Sigma Aldrich used for qPCR-cDNA

#	Name	Sequence
A001	FK506-BP5-F (qPCR-cDNA)	GAGCTTCGAAAAGGCCAAAG
A002	FK506-BP5-R (qPCR-cDNA)	GCAGCCTGCTCCAATTTTTC
B011	B011 FKBP5_3F(TSS)	TGAAGGGTTAGCGGAGCAC
B012	B012 FKBP5_3R(TSS)	CTTGGCACCTTCATCAGTAGTC
B015	B015 IGFBP1_F (qPCR-cDNA)	TTTTGAACACTCAGCTCCTAGC
B016	B016 IGFBP1_R (qPCR-cDNA)	GCACCTTATAAAGGGCACAGG
B045	B045_GILZ_(qPCR-cDNA)_F	GGAGTTTGTGACATACGAGGTG
B046	B046_ GILZ _(qPCR-cDNA)_R	AGAACGAACCCAAAGCCAAG
B047	B047_ GILZ -B_(qPCR-cDNA)_F	CCATGGACATCTTCAACAGC
B048	B048_GILZ-B_(qPCR-cDNA)_R	TTGGCTCAATCTCTCCCATC
D031	D031_hTEAD1-fw	TCCACCAAAGTTTGCTCCTT
D032	D032_hTEAD1-rev	GCCATTCTCAAACCTTGCAT
D033	D033_hTEAD2-fw	TTTTGGTCTGGAGGATCTGG
D034	D034_hTEAD2-rev	ATGGGGGAGTCAGTGACAAG
D035	D035_hTEAD3-fw	TCATCCACAAGCTGAAGCAC

#	Name	Sequence
D036	D036_hTEAD3-rev	AGCAATGACAAGCAGGGTCT
D037	D037_hTEAD4-fw	TCATCCACAAGCTCAAGCAC
D038	D038_hTEAD4-rev	AATGCACAGCAAGGTCTCCT
D075	D075_ETS1_F	TGCAGGTGCCTTAATGAAGC
D076	D076_ETS1_R	TCACACACACACCTTTTGCC
D077	D077_ETS2_F	AAAGTGGCCAAGAAGCAGTG
D078	D078_ETS2_R	AATTAGCTGTGCCGTTGCTG
D079	D079_RelA_F	CTGCATCCACAGTTTCCAGAAC
D080	D080_RelA_R	TTGTTGTTGGTCTGGATGCG
D081	D081_STAT3_F	ACAGATTGCCTGCATTGGAG
D082	D082_STAT3_R	TTTTGCTGCAACTCCTCCAG
D083	D083_STAT5B_F	TGGCCAGCATTTTCCCATTG
D084	D084_STAT5B_R	TGGGTGGCCTTAATGTTCTCC
D085	D085_ELK1_F	TGGCCAAGGAAGAATCACAC
D086	D086_ELK1_R	TTGGCAGACAAAGGAATGGC

Table 5: Oligos from Sigma Aldrich used for qPCR-ChIP

#	Name	Sequence
A003	A003_FKBP5_R (qPCR-ChIP)	GCATGGTTTAGGGGTTCTTG
A004	A004_FKBP5_F (qPCR-ChIP)	TAACCACATCAAGCGAGCTG
B025	B025 gilz1816_R (qPCR-ChIP)	TTCCTAAGTGAGGCCACCTG
B026	B026 gilz1953_F (qPCR-ChIP)	ATGTGGTGGAACCCAATGTT
B027	B027 gilz23_R (qPCR-ChIP)	AGTGAATGTTCTTGATGACCC
B028	B028 gilz23_F (qPCR-ChIP)	CTCACTTTAGGTGGGAGAC
B071	B071_ZBTB16_(qPCR ChIP)_F	CTCCTTGAGGGAAAGAACACAC
B072	B072_ZBTB16_(qPCR ChIP)_R	ACAGACGCAGGGCATTTTAC
B073	B073_FKBP5_(-DNase)_F	TCAGCTCTGAAAAGCTGCAC
B074	B074_FKBP5_(-DNase)_R	TTTTCCTGGGTTGAGGACAG
B075	B075_FKBP5_(+DNase)1_F	CCTGTGGTTGCTTTTTGACC
B076	B076_FKBP5_(+DNase)1_R	TCATGCTGGTTGGGATTAGC
B077	B077_FKBP5_(+DNase)2_F	TAACCACATCAAGCGAGCTG
B078	B078_FKBP5_(+DNase)2_R	AACAGGGTGTTCTGTGCTCTTC
B079	B079_TSC22D3_(qPCR ChIP +DNase)_F	AGAACATTGGGTTCCACCAC
B080	B080_TSC22D3_(qPCR ChIP +DNase)_R	TGTGGAGCACTGATTCATGG
B080	B080_TSC22D3_(+DNase)_R	TGTGGAGCACTGATTCATGG
B081	B081_IGFB1_(-DNase)_F	ACGTCCTGGATACAGTATGTGC
B082	B082_IGFB1_(-DNase)_R	TCATGTTCTTAGGGGGCAAC
B083	B083_IL8_(+DNase)_F	AAGAAACCACCGGAAGGAAC
B084	B084_IL8_(+DNase)_R	AGCTGCAGAAATCAGGAAGG
B087	B087_ZBTB16_(+DNase)_F	CCGAACTTTTTCCCAGAGTG
B088	B088_ZBTB16_(+DNase)_R	TGGGTACAGTGTGTTCTTGGAG
C083	C083_IGFBP1-sp_F	ACGTCCTGGATACAGTATGTGC
C084	C084_IGFBP1-sp_R	GGCAATGAATGGAAGTGAAGC
B033	B033_hRPL19_F(qPCR ChIP+cDNA)	ATGTATCACAGCCTGTACCTG
B034	B034_hRPL19_R(qPCR ChIP+cDNA)	TTCTTGGTCTCTTCCTCCTTG

Table 6: Oligos from Sigma Aldrich used for qPCR of DNase I hypersensitivity

#	Name	Sequence	Remarks
B033	B033_hRPL19_F(qPCR ChIP+cDNA)	ATGTATCACAGCCTGTACCTG	
B034	B034_hRPL19_R(qPCR ChIP+cDNA)	TTCTTGGTCTCTTCCTCCTTG	
B071	B071_ZBTB16_(ChIP-DNase)_F	CTCCTTGAGGGAAAGAACACAC	GR-BS in ZBTB16 gene between chr11:114,049,950-114,050,411 (main GR peak and no DNase I peak)
B072	B072_ZBTB16_(ChIP-DNase)_R	ACAGACGCAGGGCATTTTAC	
B073	B073_FKBP5_(-DNase)_F	TCAGCTCTGAAAAGCTGCAC	GR-BS in FKBP5 gene between chr6:35,558,432-35,558,772 (lower GR peak but no DNase I peak)
B074	B074_FKBP5_(-DNase)_R	TTTTCCTGGGTTGAGGACAG	
B075	B075_FKBP5_(+DNase)1_F	CCTGTGGTTGCTTTTTGACC	GR-BS in FKBP5 gene between chr6:35,569,392-35,569,652 (lower GR and main DNase I peak)
B076	B076_FKBP5_(+DNase)1_R	TCATGCTGGTTGGGATTAGC	
B085	B085_OR1A1_(+DNase)_F	TTTCTCCTTGCCAACCTCTC	OR1A1 gene between chr17:3,118,915-3,119,842 target at Pos 3119096 (no GR peak and no Histone activating marks)
B086	B086_OR1A1_(+DNase)_R	ATGGTTGGCCAGCATCTTAG	
B087	B087_ZBTB16_(+DNase)_F	CCGAACTTTTTCCCAGAGTG	GR-BS in ZBTB16 gene between chr11:114,033,544-114,033,865 (main GR peak and DNase I peak)
B088	B088_ZBTB16_(+DNase)_R	TGGGTACAGTGTGTTCTTGGAG	

Table 7: Oligos from Sigma Aldrich used for SDM

#	Name	Sequence
C096	C096_68-TEAD-TTC-CCC_F	GGTTTGGCTGTGAGGAATGCCATGTTCTCTCTTGCCCTCC
C097	C097_68-TEAD-TTC-CCC_F	GGAGGGCAAGAGAGAACATGGCATTCCTCACAGCCAAACC
D019	D019_Fox1-noFox_F	CCAGGACAGACACATAGGCTAGTACAGGTTATTTCTGG
D020	D020_Fox1-noFox_R	CCAGAAATAACCTGTACTAGCCTATGTGTCTGTCCTGG
D021	D021_Fox5-noFox_F	CTGAGAGGTATGCCAAGCCTATTTTTATTATGTAGC
D022	D022_Fox5-noFox_R	GCTACATAATAAAAATAGGCTTGGCATACCTCTCAG
D023	D023_Fox6-noFox_F	GTACTGCCAAAATAGGCTAGCTGAGAGTGAGAACGG
D024	D024_Fox6-noFox_R	CCGTTCTCACTCTCAGCTAGCCTATTTTGGCAGTAC
D025	D025_Fox9-noFox_F	CTATAACTATCGTAGCCTATTTTGATGCTCAAAATG
D026	D026_Fox9-noFox_R	CATTTTGAGCATCAAAATAGGCTACGATAGTTATAG
D123	D123_Fox5-noGR_F	GTAGCTCACATCAGAGAATCCTATACGTCCTTCATAAC
D124	D124_Fox5-noGR_R	GTTATGAAGGACGTATAGGATTCTCTGATGTGAGCTAC
D125	D125_Fox6-noGR_F	CTGAAAGTTAGCAAATAAGGAAATAAAGCCAAAATAAAC
D126	D126_Fox6-noGR_R	GTTTATTTTGGCTTTATTTCCTTATTTGCTAACTTTCAG
D127	D127_Fox9-noGR_F	GTTTATTTTGATGCTAAAAATATTCCAGGACAATTTTTAC
D128	D128_Fox9-noGR_R	GTAAAAATTGTCCTGGAATATTTTTAGCATCAAAATAAAC
E001	E001-D137-noTEAD_F	GGCTTTTCATCCCAGGAACATGGCATTCTAACTGCTCAG
E002	E002-D137-noTEAD_R	CTGAGCAGTTAGAATGCCATGTTCCTGGGATGAAAAGCC
E003	E003-D141-noTEAD_F	GTTCACAGCTGGGAACAGAACATGGCTCTGTGATCCCCTTG
E004	E004-D141-noTEAD_R	CAAGGGGATCACAGAGCCATGTTCTGTTCCCAGCTGTGAAC
E007	E007-D147-noTEAD_F	GTCTGGAGACGGAGAGAACATGGCAGGTAGAGGGAACAG
E008	E008-D147-noTEAD_R	CTGTTCCCTCTACCTGCCATGTTCTCTCCGTCTCCAGAC
E009	E009-D151-noTEAD_F	CTTTTTAACTGGGATAGAACATGGCACAGTAGGGACATAAC
E010	E010-D151-noTEAD_R	GTTATGTCCCTACTGTGCCATGTTCTATCCCAGTTAAAAAG

Table 8: DsiRNA from IDT

#	Name	Gene	Sequence
1	HSC.RNAI.N004496.12.1	FoxA1	GGAGGAGAGAUAAGUUAUAGGGAGC
2	HSC.RNAI.N004496.12.4	FoxA1	GGCAUACGAACAGGCACUGCAAUAC
3	HSC.RNAI.N001451.12.1	FoxF1	GCAAACAGGAGACCAACAAT
4	HSC.RNAI.N001451.12.2	FoxF1	GCAGGACACGAGAAATT
5	HSC.RNAI.N005250.12.1	FoxL1	ACAGCAAACACAACAAATA
6	HSC.RNAI.N005250.12.2	FoxL1	GGCAGAGAAGGAAACACAG
7	HSC.RNAI.N001143820.12.1	ETS1	AGCAUAGAGAGCUACGAUAGUUGTG
8	HSC.RNAI.N001143820.12.2	ETS1	GCAAAGGAAACUAAGGAAGGAGGTT
9	HSC.RNAI.N005239.12.1	ETS2	GCCUAACAGUUAUGGAAACUACAGT
10	HSC.RNAI.N005239.12.7	ETS2	CCAUGUCUUUCAAGGAUUACAUCCA
11	HSC.RNAI.N001145138.12.1	RELA	GGAGUACCCUGAGGCUAUAACUCGC
12	HSC.RNAI.N001145138.12.2	RELA	AGCACAGAUACCACCAAGACCCACC
13	HSC.RNAI.N003150.12.1	STAT3	CCGUCAACAAAUUAAGAAACUGGA
14	HSC.RNAI.N003150.12.4	STAT3	AGAGUCAAGAUUGGGCAUAUGCGGC
15	HSC.RNAI.N012448.12.1	STAT5B	GCUGUUUGAAUAAGAGAAGGACAAA
16	HSC.RNAI.N012448.12.3	STAT5B	GCAUGGGACUCAGUAGAUCUUGATA
17	HSC.RNAI.N001114123.12.1	ELK1	AGGAGAAACAUAGUUCAACUGAAAG
18	HSC.RNAI.N001114123.12.5	ELK1	CCUUCUAUCAGCGUGGAUGGCCUCT
19	HSC.RNAI.N021961.12.1	TEAD1	ACUAUAGCCUUAAGAAAUGCUUGGT
20	HSC.RNAI.N021961.12.3	TEAD1	ACCAGAGAAAUAUAUGAUGAACAGT
21	HSC.RNAI.N003598.12.1	TEAD2	CCGGCAGAUCUACGACAAAUUCCCT
22	HSC.RNAI.N003598.12.2	TEAD2	ACGAGAUAUUUAUAAGUGGGUGCTA
23	HSC.RNAI.N003214.12.1	TEAD3	GGUCCUCACUGUUUGCAUAUCGCTC
24	HSC.RNAI.N003214.12.4	TEAD3	GCACGCUAUAUUAAACUGAGGACGG
25	HSC.RNAI.N003213.12.1	TEAD4	GCCGUGGACAUCCGCCAAAUCUATG
26	HSC.RNAI.N003213.12.2	TEAD4	CUAUGAGAAUGGACACUACUCUUAC

5.6 Organisms

5.6.1 Bacterial strains

DH5 α : fhuA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17

5.6.2 Mammalian cell lines

U2OS: Human bone cells from a 15-year-old female patient with osteosarcoma; cultivated in DMEM + 5 % FBS

GR18: U2OS cells stably expressing GR (Rogatsky 1997) ; cultivated in DMEM + 5 % FBS

IMR90: Human primary lung fibroblast from a 16 weeks old caucasian female embryo; cultivated in EMEM + 10 % FBS, ATCC® CCL-186™

K562: Human immortalized chronic myelogenous leukemia (CML) line from a 53 year old female CML patient; cultivated in IMDM + 10 % FBS

5.7 Media

LB Medium

5 g/l yeast extract
10 g/l bacto-tryptone
10 g/l sodium chloride

SOB Medium

5 g yeast extract
20 g/l bacto-tryptone
10 mM sodium chloride
10 mM $MgCl_2$
0.5 mM KCl
10 mM $MgSO_4$

5.8 Buffers

5.8.1 General buffers

TE-Buffer

10 mM Tris-HCl pH 8.0; 1 mM EDTA; ddH_2O

PBS buffer

137 mM NaCl, 2,7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4 pH7.4

5.8.2 ChIP buffers

IP Lysis Buffer

50 mM HEPES-KOH, pH 7.4; 1 mM EDTA; 150 mM NaCl; 10 % glycerol; 0.5 % Triton X-100

LiCl buffer

20mM Tris, pH 8.0; 1mM EDTA; 250mM LiCl; 0.5 % NP-40; 0.5 % sodium deoxycholate

RIPA buffer

10 mM Tris-HCl, pH 8.0; 1 mM EDTA; 150 mM NaCl; 5 % glycerol; 0.1 % sodium deoxycholate; 0.1 % SDS; 1 % Triton X-100

RIPA-500 mM NaCl

10 mM Tris-HCl, pH 8.0; 1 mM EDTA; 500 mM NaCl; 5 % glycerol; 0.1 % sodium deoxycholate; 0.1 % SDS; 1 % Triton X-100

Xlink reversal solution

10 mM Tris-HCl, pH 8.0; 1 mM EDTA; 0.7 % SDS; 220 µg/ml proteinase K (freshly added)

5.8.3 DNase buffers

DNase I reaction buffer

20 mM HEPES pH 7.4, 0.5 mM $CaCl_2$, 5 % glycerol, 3 mM $MgCl_2$, 0.2 mM spermine, 0.2 mM spermidine; 0.2 % NP40

DNase I reaction buffer (without NP40)

20 mM HEPES pH 7.4, 0.5 mM $CaCl_2$, 5 % glycerol, 3 mM $MgCl_2$, 0.2 mM spermine, 0.2 mM spermidine

DNase I Stop buffer

50 mM Tris pH 8.0; 200 mM NaCl; 12.5 mM EDTA; 1 % SDS; 200 µg/ml Proteinase K (freshly added)

5.8.4 EMSA buffers

1x Binding buffer

20 mM Tris pH 7.5; 2 mM $MgCl_2$; 1 mM EDTA; 10 % Glycerol; 0.3 mg/ml BSA; 4 mM DTT

5.8.5 Nuclear extract

Buffer A

10 mM HEPES, 1.5 mM $MgCl_2$; 10 mM KCl; 300 mM Sucrose; 0.5 % IGEPAL; 0.5 mM PMSF*; 1 mM Na_3VO_4*; 0.5 mM DTT*; 1 ‰ (v/v) Proteinase Inhibitor cocktail set III EDTA free *; 25 mM beta-glycerophosphate*; 10 mM NaF*

Buffer B

20 mM HEPES; 1.5 mM $MgCl_2$; 420 mM NaCl; 0.2 mM EDTA; 2.5 % Glycerol; 0.5 mM PMSF*; 1 mM Na_3VO_4*; 0.5 mM DTT*; 1 ‰ (v/v) Proteinase Inhibitor cocktail set III EDTA free*; 25 mM beta-glycerophosphate*; 10 mM NaF*

Buffer D

20 mM HEPES; 100 mM KCl; 0.2 mM EDTA; 8 % Glycerol; 0.5 mM PMSF*; 1 mM Na_3VO_4*; 0.5 mM DTT*; 1 ‰ (v/v) proteinase inhibitor cocktail set III EDTA free *; 25 mM beta-glycerophosphate*; 10 mM NaF*

*: Add fresh directly before use

PBSI buffer

0.5 mM PMSF; 25 mM beta-glycerophosphate; 10 mM NaF

5.8.6 DNA pull-down buffers

Annealing Buffer

20 mM Tris-HCl pH 8.0; 10 mM $MgCl_2$; 100 mM KCl

Blocking buffer

20 mM HEPES, pH 7.9; 0.05 mg/ml glycogen; 0.3 M KCl; 0.02 % IGEPAL; 0.05 mg/ml BSA; 5 mg/ml PVP

DW buffer

20 mM Tris-HCl pH 8.0; 2 M NaCl; 0.5 mM EDTA; 0.03 % IGEPAL

Buffer G

20 mM Tris-HCl pH 7.3; 10 % Glycerol; 0.1 M KCl; 0.2 mM EDTA; 10 mM Monopotassium glutamate; 0,04 % IGEPAL

RE buffer

100 mM NaCl; 50 mM Tris HCl; 10 mM $MgCl_2$; 2 mM DTT; 2.5 % glycerol; 0.2 mM PMSF; 1 ‰ proteinase inhibitors set III EDTA free; 0.02 % IGEPAL

6 Methods

6.1 Maintenance of human cells

U2OS cells and U2OS GR18 cells (stably expressing GR (Rogatsky 1997)) were maintained in DMEM (Gibco) supplemented with 5 % FBS (Gibco). IMR90 cells were maintained in EMEM (Gibco) supplemented with 10 % FBS (Gibco). K562 cells were maintained in IMDM (Gibco) supplemented with 10 % FBS (Gibco). All cell were cultured at 37 °C and 5 % CO_2.

6.2 RNA purification

RNA of approximately 1 million cells was purified using the RNeasy kit (Qiagen) according to the manufacturer's protocol. Cells were lysed in RLT buffer supplemented with 2 % β-mercaptoethanol and adherent cells were scraped of surface. An equal amount of RNase-free pure ethanol was added and the lysate was passed 5 times through a 0.8 gauge needle fitted to a syringe before they were loaded onto columns. Membrane was washed once with 350 µl buffer RW1, before membrane was incubated with 27.3 units of RNase free DNase (Qiagen) in 80 µl RDD buffer for 15 minutes at room temperature. The column was washed again with 350 µl RW1 buffer and then twice with 500 µl RPE buffer. RNA was eluted with 50 µl RNase-free water. RNA concentrations were measured using a NanoDrop2000 spectrophotometer.

6.3 cDNA preparation

500 ng RNA, 775 pmol random primer (NEB) and 310 µM dNTP mix in a total volume of 16 µl were incubated for 5 minutes at 70 °C. After cooling down the mix to 42 °C, 4 µl reverse transcription mix (2 µl 10xM-MuLV reverse transcriptase buffer, 0.25 µl RNasin (Promega), 0.125 µl M-MuLV reverse transcriptase (NEB) and 1.625 µl) water were added and subsequently incubated for one hour at 42 °C, followed by 10 minutes at 90 °C.

Transcript levels were quantified by qPCR using the following mRNA targeting primers: FKBP5: A001/A002 and B011/B012; IGFBP1: B015/B016; GILZ: B045/B046; GILZ B: B047/B048 and RPL19 as non GR-regulated gene: B033/B034.

Transcript levels for knocked down genes were quantified by qPCR using the mRNA targeting primers listed in Table 4.

6.4 ChIP Chromatin Immunoprecipitation (ChIP)

6.4.1 Cell growth and sonication

Adherent cells (IMR90) were seeded at 10,000 cells per cm^2 and grown at 5 % CO_2 / 37 °C until 80-90 % confluence. Non-Adherent cells (K562) were grown to 500,000 cells / ml. Media was changed the day before cells were collected. Cells were treated for 1 hour with 1 μM dexamethasone (dissolved in EtOH) or an equal volume of EtOH as vehicle control and a final concentration of 1 % formaldehyde was used to crosslink cells for 3 min at 22 °C while shaking. Formaldehyde was quenched with glycine at a final concentration of 125 mM for 10 min at 4 °C. Non-Adherent cells were centrifuged at 100 rcf. Cross-linked cells were rinsed with PBS and washed again with PBS (1 ml PBS per 1 million cells) for 5 minutes at 4 °C. Adherent cells were scraped into 15 ml conical tubes with IP lysis buffer (2 ml per 1 million cells) supplemented with 0.5 % proteinase inhibitor cocktail set II (EDTA free) and PMSF 0.5 mM. Nuclei were nutated for 30 min at 4 °C, pelleted by centrifugation at 4 °C for 5 min at 600 rcf and afterwards resuspended in 300 μl ice-cold RIPA buffer with protease inhibitors (1:200). 1.5 ml TPX microtubes were used to sonicate cells. K562 cells were sonicated for 10 min, whereas IMR90 cells were sonicated for 50 min divided into 30 s on and 30 s off intervals using a Bioruptor Plus (Diagenode) at high intensity with constant cooling to 4 °C. Samples were centrifuged at 4 °C at 20,000 rcf for 15 min and supernatant transferred to a fresh tube.

6.4.2 Immunoprecipitation

25 μl Invitrogen protein G Dynabeads® were used per ChIP (+ 10 % extra) and pelleted with a magnet to remove the liquid. Beads were washed twice (without magnet) in 1000 μl RIPA and resuspended at original bead concentration (without 10 % extra) in RIPA with 1X Proteinase Inhibitor Cocktail (PIC) (Proteinase inhibitors + PMSF 1:200 and 0.5 $\mu g/\mu l$ BSA). 2 μl GR antibody per ChIP was added to beads and mixture was subsequently nutated for 1 h at 4 °C in 0.5 ml tube. Beads were spun briefly and RIPA was added to bead mixture to 500 μl total volume and transferred to a 1.5 ml tube. Liquid was removed while beads were pelleted with magnet followed by two resuspending washes with 1000 μl RIPA. Beads were resuspended in RIPA with 4X Proteinase Inhibitor Cocktail (Proteinase inhibitor + PMSF 1:50 and 2 $\mu g/\mu l$ BSA) at original bead concentration. 25 μl of loaded beads were transferred to each tube of an 8 well strip and 75 μl of supernatant from centrifuged chromatin were added to each ChIP tube. 7.5 μl supernatant was taken as input. Bead-chromatin mixture was nutated for 4 hours at 4 °C on a rotating wheel.

6.4.3 *Washes and crosslink reversal*

Tubes were spun briefly and 100 µl RIPA-500 mM NaCl was added to the pelleted beads that were immobilized with a magnet. Liquid was removed and beads were washed 4 times with 200 µl RIPA-500 mM NaCl and 4 times with 200 µl LiCl with alternating resuspending and non-resuspending washes. Beads were resuspended in 200 µl RIPA-500 mM NaCl and transferred to a new tube. Liquid from last wash was removed and beads were resuspended in 11 µl RIPA-500 mM NaCl and 89 µl of X-link reversal solution. In parallel, 89 µl X-link reversal solution and 1 µl 5 M NaCl were added to input samples. To all samples, 1 µl proteinase K was added followed by a 2-step incubation: 3 h at 55 °C and 6 h at 65 °C.

6.4.4 *Sample purification*

To extract the immunoprecipitated DNA, the aqueous phase without beads was transferred to 1.5 ml tubes and mixed with 1/10 Vol. 3 M NaAc and 4 µl Glycogen (Fermentas 20 µg/µl) by vortexing. 3 volumes of pure ethanol were added followed by a 1 h centrifugation at 4 °C with 20,000 rcf. Pellets were washed in 500 µl 80 % EtOH keeping the pellet intact and spun again at 20,000 rcf for 5 min at 4 °C. 80 % EtOH was removed completely with a pipette avoiding the pellet. Pellets were air dried for about 5 min and should retain a moist appearance. Finally, pellets were resuspended in 20 µl MilliQ water.

6.5 ChIP verification by enrichment of target loci using qPCR

ChIP enrichment was validated using primers A003/A004 to target GR binding at coordinates chr6: 35569764-35570000 of the FKBP5 gene in GRCh37/hg19 assembly, a known endogenous genomic GR binding site in several human cell lines (Thomas-Chollier et al. 2013). Since GR does not bind at this locus in K562 cells, for this cell line, I used a GBS present in an intron of the ZBTB16 gene at position chr11: 114049950-11405041 (GRCh37/hg19 assembly) targeted with the primer pair B071/B072, known for GR binding from IMR90 ChIP-Seq (own data). GR binding was controlled additionally using primers B025/B026 (chrX: 106961461-106961597), B027/B028 (chrX:106962005-106962152) and B079/B080 (chrX: 106961576-106961650), which target GBS in the TSC22D3 (also named GILZ) gene (GRCh37/hg19 assembly). Since GR binding is also known in the IGFBP1, I also used the region chr7: 45929095-45929181 (GRCh37/hg19 assembly) as positive control for the ChIP experiments.

To control for unspecific binding in the ChIP experiments, primers B033/B034 were used, which target five different loci known to be unbound by GR. Coordinates in GRCh37/hg19 assembly are:

chr5:177482959-177483191 (RPL19 gene);
chr8:99794622-99794853;
chr1:64254390-64254621 (STK3 gene);
chr17:37360328-37360897 (ROR1 gene);
chr7:102781800-102782032 (NAPEPLD gene).

6.6 Preparation of cells for ChIP-Exo

Cells were seeded in 15 cm dishes (10,000 cells per cm^2) and grown in 20 mL medium at 5 % CO_2 / 37 °C to 80-90 % confluence. Media was changed the day before cells were collected. One 15 cm plate was used to count cells on the day of the experiment. Cells were treated with 1 μM dexamethasone or Ethanol (as vehicle control) for 1 hour. Non-adherent cells were collected in one flask. Cells were cross-linked for 3 minutes at RT by adding formaldehyde to a final concentration of 1 %. Crosslinking reaction was quenched with glycine at a final concentration of 125 mM and subsequent incubation at 4 °C for 10 minutes. Non-Adherent cells were centrifuged at 100 rcf. Cells were rinsed with PBS and washed again for 5 min with 10 ml PBS at 4 °C. 10 ml IP lysis buffer with protease inhibitors (1:200) was added. Adherent cells were scraped from the dishes, and collected in IP lysis buffer, transferred to 15 ml Falcon tubes followed by nutation at 4 °C for 30 min. Nuclei were pelleted by centrifugation at 4 °C for 5 min at 600 rcf and ~15 million cell equivalents were resuspended in 150 μl ice-cold RIPA buffer with protease inhibitors (1:200). 1.5 ml TPX microtubes were used to sonicate cells in a Bioruptor Plus (Diagenode) at high intensity with constant cooling to 4 °C. K562 cells were sonicated for 30 cycles, whereas IMR90 cells were sonicated for 50 cycles. Each cycle is divided into 30 s sonication and 30 s pause. Samples were centrifuged at 4 °C at 20,000 rcf for 15 min. 1 million cell equivalents were checked on gels for proper fragmentation (desired fragment size 200-400 bp). Suitability of samples for ChIP-Exo was assessed by conduction a conventional ChIP (6.4) and validated using qPCR prior to sending the samples (on dry ice) together with GR-AB (at 4 °C) to Peconic LLC for further ChIP-Exo processing and sequencing (Rhee & Pugh 2011).

6.7 qPCR

ChIP performance, transcriptional regulation and DNase I hypersensitivity experiments were analyzed by qPCR using an ABI 7900 HT cycler (Applied Biosystems). Samples for each experiment were measured as duplicates of at least three biological replicates.

5 µl Power Sybr® PCR master mix (Applied Biosystems) [or a home mix consisting of 100 mM Tris pH 8.3, 6 mM MgCL2, 1 mg/ml BSA, 4 mM dNTPs 0,66x SYBR-Green and 1x ROX reference dye], 0.2 µM of each qPCR primer and 2 µl DNA (ChIP or cDNA in concentration of about 1 ng/ml) in a total reaction volume of 10 µl were analyzed using cycling conditions shown in Table 9.

Table 9: cycling conditions for qPCR

	Temperature	Time	Ramp rate
Initial denaturation	95 °C	600 s	100 %
Amplification (40 cycles)	95 °C 55 °C	15 s 60 s	100 % 100 %
Recording of melting curve	68 °C 68 °C 4 °C	15 s 15 s 15 s	100 % 100 % 2 %

6.8 Generation and transformation of chemically competent DH5 alpha cells

The Z-Competent™ *E.coli* transformation Kit & Buffer Set (ZYMO RESEARCH) was used to generate chemically competent DH5α cells according to the manufacturer's protocol. 50 µl of competent bacteria were transformed with up to 1 ng plasmid in up to 5 µl DNA solution. Ligation mixtures were incubated for 5 min on ice prior to further processing. For ampicillin selection, cells were plated to LB/agar containing 100 µg/ml ampicillin. If selected with kanamycin, 250 µl SOC medium (37 °C) was added followed by a 1 h incubation at 37 °C while shaking before seeding cells to LB/agar plates containing 20 µg/ml kanamycin.

6.9 Cloning procedures

6.9.1 Cloning of pGL3-promoter - GR-reporter vectors

The pGL3-Promoter vector (Promega) was linearized with the restriction enzymes Cfr9*I* and Bgl*II* (both Thermo Fisher) and afterwards SAP (Fermentas) treated for 1 hour at 37 °C. Linear pGL3 was gel-checked and linearized vector backbone was purified using the NucleoSpin Gel - and PCR Clean-up kits; Macherey & Nagel.

Reporter plasmids were generated by cloning candidate sequences into the pGL3-promoter plasmid (Promega). Candidate sequences were either approx. 400 bp genomic regions of interest or isolated motifs.

For genomic regions of interest, genomic DNA was purified from $3*10^6$ IMR90-cells, ProteinaseK and RNase A treated using the Qiagen Blood and Tissue Kit. PCR primers used to amplify genomic DNA were introducing Cfr9I und BglII recognition sites to the ends of the PCR products (Phusion® PCR Master Mix, used as recommended by supplier). Amplicons were gel purified, blunt ligated into TOPO-Vectors using the zero blunt cloning kit (Invitrogen) and screened for insertions according to the manufacturer's protocol. To extract amplicons, clones were cultured in Kanamycin medium and plasmids were extracted using the QIAprep Spin Miniprep Kit. Amplicons were restriction digested out of 4 µg plasmid using 2 µl of Cfr9*I* (Fermentas) and Bgl*II* (Fermentas) each in 30 µl reactions and desired inserts were subsequently purified from agarose-gels. Restriction-digested amplicons were ligated into linearized pGL3-Promoter vector and finally sequenced to assure they contained the desired insert.

Used oligos, details in Table 1:

Combi sites : C067/C068; D137/D138; D141/D142; D147/D148; D151/D152

Fox sites: C127/C128; C135/C136; C137/C138; C143/C144;

Non-genomic sequences containing isolated sequences matching the motif of interest in either one or three copies were ordered as oligos. 5 µM of each corresponding oligo was mixed in 10x Ligase buffer (Fermentas) containing 125 µM ATP and 20 units of polynucleotide kinase (Fermentas) for 4 hours at 37 °C to phosphorylate their 5'ends. Phosphorylated oligos were hybridized by boiling in water bath and subsequently cooled to RT overnight. The resulting hybridized oligos have 4 bp overhangs as shown in Figure 6 compatible to ligate into Cfr9*I* and Bgl*II* linearized pGL3-Promoter vector (Promega).

CCGGG-insert-A

C-insert-TCTAG

Figure 6: Hybridized oligos containing candidate sequences flanked by 4 bp overhangs compatible to ligate into Cfr9*I* and Bgl*II* restricted pGL3 promoter (Promega).

Hybridized oligos were ligated into linearized pGL3-Promoter vector (Promega) using T4 DNA Ligase (Fermentas) according to the manufacturer's protocol. Re-circularized pGL3-promoter - GR-reporter vectors were subsequently transformed into chemically competent DH5α cells and plasmids from resulting clones were sequenced to isolated plasmids with the desired insert.

Used oligos for isolated motif-matching sequences (sequence details in Table 1):

C124/C125; D059/D060; D103/D104; D129/D130; D135/D136;

6.10 Site directed mutagenesis

Specific changes in sequences of luciferase reporter plasmids were introduced by site-directed mutagenesis (SDM) done in 25 µl reaction mixes containing 1x Pfu ultra buffer, 20 ng target plasmid, 0.25 µM of each primer and 200 µM dNTP mix. Samples were heated up to 95 °C before 0.5 µl Pfu ultra polymerase (Agilent) were added, followed by cycling conditions for the SDM as follows:

Temperature	Time	
95 °C	60 s	
95 °C	30 s	} 18 cycles
55 °C	60 s	
68 °C	4 min	
68 °C	10 min	
4 °C	hold	

Resulting PCR mixes were digested by adding 10 units of Dpn*I* to the SDM-solution and incubating at 37 °C for 2 hours. 5 µl SDM solution was used to transform Z-competent Dh5alpha cells. Correct clones were identified by sequencing.

Used oligos (sequence details in Table 7)

No GR in Fox genomic regions:

D123/D124; D125/D126; D127/D128; D133/D134

No Fox in Fox genomic regions

D019/D020; D021/D022; D023/D024; D025/D026

No TT in combi-motif regions:

C096/C097; E001/E002; E003/E004; E007/E008; E009/E010

6.11 Transfection of IMR90 cells

Per assay, each reporter was tested in duplicate both with and without hormone. Therefore, 4 wells of a 24 well cell culture multiwell plate per reporter were inoculated with 50,000 cells in 500 µl EMEM/10 % FBS and incubated overnight. The next day, each of the four wells per reporter was transfected with transfection mix. The transfection mix consisted of two separate premixes. The first contains 720 ng reporter plasmid and 8 ng pRL (CMV) in 50 µl OPTI-MEM® per well. The second premix is prepared in a separate tube containing 2 µl Lipofectamine 2000 (Invitrogen) and 50 µl OPTI-MEM® per well. Both premixes were mixed gently by pipetting and incubated for 5 minutes at room temperature. The two premixes were combined by pipetting and incubated for 20 minutes at room temperature. 100 µl transfection mix was added per well. Medium was carefully replaced with 500 µl EMEM/10 % FBS six hours post transfection containing either 1 µM dexamethasone or pure ethanol as vehicle control.

6.12 Transfection of U2OS cells

Each reporter was tested in duplicates both with and without hormone. Therefore, 4 wells of a 48 well cell culture multiwell plate per reporter were inoculated with 30,000 cells in 250 µl DMEM/5 % FBS and incubated overnight. The next day, each of the four wells per reporter was transfected with transfection mix. The transfection mix consists of two separate premixes. The first contains 10 ng reporter plasmid, 0.1 ng pRL (CMV), 10 ng pcDNA3.1-rGR, 50 ng p6R and 0.8 µl plus reagent (Invitrogen) in 12.5 µl serum free DMEM (Gibco) per well. The second premix was prepared in a separate tube containing 0.8 µl Lipofectamine (Invitrogen) and 12.5 µl serum free DMEM per well. Both premixes were vortexed and incubated for 15 minutes at room temperature. The two premixes were combined by pipetting and incubated for an additional 15 minutes at room temperature. Cells were carefully washed once with 0.5 ml PBS and 100 µl serum free DMEM medium was added per well. 25 µl transfection mix was added per well. Three hours post transfection, medium was carefully replaced with 250 µl DMEM/5 % FBS. Six hours post transfection, medium was replaced again with 250 µl DMEM/5 % FBS containing either 1µM dexamethasone or EtOH as vehicle control.

6.13 dsiRNA and pGL3-promoter - GR-reporter co-transfection

6.13.1 Co-transfection in IMR90 cells

20,000 IMR90 cells were seeded in 200 µl EMEM/10 % FBS per well of a 48 well cell culture multiwell plate. After an overnight incubation, cells of each well

were transfected with Lipofectamine 2000 (Invitrogen) and 25 nM dsiRNA (IDT). 0.25 µl Lipofectamine 2000 (Invitrogen) and 1,5 µl dsiRNA (2.5 µM) were premixed separately with 12.5 µl SF-EMEM and incubated for 5 minutes at room temperature. The two premixes were combined and incubated for an additional 20 minutes at room temperature. 75µl media was removed from each well before 25 µl transfection mix was added. Six hours post transfection, medium was replaced twice with 250 µl EMEM/10 % FBS. 24 hours post dsiRNA transfection, cells of each well were transfected with 1 µl Lipofectamine 2000 (Invitrogen) as well as 360 ng reporter and 4 ng pRL (CMV) plasmid. Both plasmids and Lipofectamine 2000 (Invitrogen) were premixed separately with 25 µl OptiMEM® (Invitrogen) and incubated for 5 minutes at room temperature. The two premixes were combined and incubated for an additional 20 minutes at room temperature. 50 µl transfection mix was added to the 250 µl media of each well. Six hours post transfection, medium was replaced with 250 µl EMEM/10 % FBS containing either 1µM dexamethasone or EtOH as vehicle control. Luciferase activities were measured as described in 6.14 Dual luciferase assay.

6.13.2 Co-transfection in U2OS cells

20,000 U2OS cells were seeded in 200 µl DMEM/5 % FBS per well of a 48 well cell culture multiwell plate. After an overnight incubation, cells of each well were transfected with Lipofectamine 2000 (Invitrogen) and 25 nm dsiRNA (IDT). 0.25 µl Lipofectamine 2000 (Invitrogen) and 1,5 µl dsiRNA (2.5 µM) were premixed separately with 12.5 µl SF-DMEM and incubated for 5 minutes at room temperature. The two premixes were combined and incubated for additional 20 minutes at room temperature. 75 µl media was removed from each well before 25 µl transfection mix was added. Six hours post transfection, medium was replaced twice with 250 µl DMEM/5 % FBS containing. 24 hours post dsiRNA transfection, cells of each well were transfected with 0,4 µl Lipofectamine (Invitrogen) as well as 10 ng reporter, 0,1 ng pRL (CMV), 10 ng pcDNA3.1-rGR and 50 ng p6R plasmid. Plasmids were premixed with 0,8 µl plus reagent in 12,5 µl serum free DMEM (Gibco), whereas Lipofectamine (Invitrogen) was premixed separately with 12,5 µl serum free DMEM (Gibco) and incubated for 15 minutes at room temperature. The two premixes were combined by vortexing and incubated for an additional 15 minutes at room temperature. Cells were washed with 500 µl PBS (37 °C) and 100 µl serum free DMEM (Gibco) was added to the cells. The 25 µl transfection mix was added to the SF-media of each well. Three hours post transfection, medium was replaced with

250 µl DMEM/5 % FBS. Six hours post transfection, medium was replaced again with 250 µl DMEM/5 % FBS containing either 1 µM dexamethasone (DEX) or EtOH as vehicle control. Luciferase activities were measured as described in 6.14 Dual luciferase assay.

6.14 Dual luciferase assay

Luciferase activities were measured 16-18 hours after hormone treatment using a luminometer (LUMIstar Omega by BMG Labtech) and the dual luciferase assay kit (Promega). Cells were lysed in 65 µl (U2OS cells) or 100 µl (IMR90 cells) passive lysis buffer for 30 minutes at room temperature while shaking. Lysed cells (2.5 µl) were transferred to a white 384 well microplate (Greiner). 12.5 µl of LARII- as well as Stop and Glo - substrate were used to determine Firefly- and Renilla-luciferase activities respectively in each well. Renilla activities were used to normalize firefly activities for transfection efficiencies. Luciferase expression in U2OS cells required a gain of the luminometer of 3600, whereas luciferase activity in IMR90 cells was determined at a gain of 4095.

6.15 Electrophoretic Mobility Shift Assay (EMSA)

EMSAs were performed to determine DNA binding affinity using the GR-DBD (amino acids 380-540 (Meijsing et al. 2009)). The forward oligos for were ordered with an AmC6F label at their 5'ends to allow Cy5-labelling. Oligos were hybridized at 50 µM per Oligo by boiling in a water bath, cooled to room temperature overnight and subsequently Cy5 labeled using the Cy5 Post-Labeling Reactive Dye Pack (Amersham) according to manufacturer's technical bulletin. Sequences for Cy5 labeled positive strands are shown below. Candidate motifs are marked in yellow and bases marked in red indicate changes to diverge the recognition sequence away for the consensus recognition sequence motif.

Combi	(D055/D056)	Cy5-GATC tcga AAAGAACATTCCA gtac CTAT
Combi-noT	(D057/D058)	Cy5-GATC tcga AAAGAACAAACCA gtac CTAT
FOX	(D115/D116)	Cy5-GATC tcga AATAAACAAAATA gtac CTAT
PAL		Cy5-TCGA AGAACAAAATGTTCT TCGA
random		Cy5-TCGA TACCAAAATATTTGA GTAC

Cy5-labeled DNA was re-buffered to MilliQ water using a Bio-Rad Bio-spin 6 column and concentration was measured using a NanoDrop2000 spectrophotometer to set oligo concentrations to 10 µM in 1x EMSA binding buffer (without poly(dI-dC)) as stock solution. Cy5-labeled DNA is further diluted to 20 nM with 2x EMSA

binding buffer and mixed in a 1:1 ratio with poly(dI-dC) solution (0.2 µg/µl). 1x EMSA binding buffer was used to set purified GR DBD to: 20 µM; 8 µM; 3.2 µM; 1.28 µM; 0.512 µM; 0.2048 µM; 0.09 µM and 0 µM. Each protein dilution was mixed in a 1:1 ratio (9 µl each) with 10 nM premixed Cy5-labeled DNA solution and directly incubated in the dark at room temperature for 30 minutes to allow protein:DNA interactions to reach equilibrium. A native 6 % polyacrylamide gel in 0.5 x a TBE buffer was prepared and prerun at 250 V for 30 min. Samples (18 µl) were subsequently loaded to the running gel. Gel electrophoresis was stopped after 15 minutes and gels were scanned for analysis with the FLA 5100 scanner (Fujifilm) at an excitation wavelength of 640 nm and corresponding filter, while voltage of photo-multiplier was set to 800 V.

6.16 DNase I Hypersensitivity assay

Cells were grown in 10 cm tissue plates to 80-90 % confluence. Media was renewed the day before hormone treatment with 1 µM DEX or 1 ‰ EtOH for 1 hour. Cells were washed with PBS and scraped into 6 ml ice-cold DNase I buffer/0.2 % NP40 alternative to lyse cell membranes and split into 1 ml samples in 1.5 ml reaction tubes. Cells were homogenized by 10 s vortexing, incubated 5 minutes on ice and centrifuged for 5 minutes at 500 rcf at 4 °C to pellet nuclei. Pellets were equilibrated to room temperature after resuspending in 200 µl DNase I buffer (without NP40 alternative). Samples were split and 100 µl were treated with 2.5 units DNase I (Qiagen) for 5 minutes at 37 °C. Untreated samples were used to normalize for chromatin input. An equal volume of 2x stop buffer containing 200 µg/ml proteinase K was added to each sample to stop the DNase I digestion and degrade proteins by incubation at 65 °C for 4 hours. DNA was purified using the PCR purification kit (Qiagen), set to 25 ng/µl to analyze genomic regions of interest by qPCR using the following primers (details in Table 6):

B071/B072; B087/B088; B073/B074; B075/B076; B077/B078; B079/B080; B081/B082; B083/B084

6.17 Preparation of nuclear extract

Preparation of nuclear extract was done as described previously (Mittler et al. 2009). Cells were grown to 80-90 % confluence in 15 cm dishes. Medium was replaced the day before the cells were treated for 1 hour with 1 µM dexamethasone or with 1 ‰ EtOH as vehicle control. Buffers A, B and D were freshly mixed with 1:500 proteinase inhibitor and 1 µM dexamethasone. Cells were washed once with ice

cold PBS and cells were scraped with 2 ml PBSI per dish into a 15 ml falcon tube. Supernatant was removed after 5 minutes centrifugation at 550 rcf at 4 °C and soft cell pellets were transferred to fresh 1.5 ml reaction tubes. Cells were centrifuged at 1500 rcf for 30 seconds for removal of remaining supernatant and resuspended afterwards in two package volumes of buffer A. Cells were incubated for 10 minutes on ice. Isolated nuclei were subsequently mixed by gentle vortexing and centrifuged for 30 s at 2600 rcf. After removal of supernatant, cells were resuspended in 2/3 of cell package volumes of buffer B, sonicated for 7 s using a Bioruptor Plus (Diagenode) at high intensity with constant cooling to 4 °C. Nuclear extract was centrifuged for 5 minutes at 10,000 rcf at 4 °C and supernatant was finally diluted in 1:1 ratio with buffer D. Aliquots were stored at -80 °C

6.18 DNA Pulldown

For DNA pulldown assays, the following biotinylated DNAs coupled to streptavidin beads were used (Table 10): recognition sequence of interest underlined, modified bases to diverge the sequence away from the consensus motif are highlighted in red. Further details are listed in Table 3

Table 10: Sequence for DNA-Pulldown of nuclear extracts to identify proteins associated to the combi motif.

#	Sequence
D105_ 1xC	CAAAAGATCGCTGCAGACTTGAAC AAAGAACATTCCA ACTTTGTC
D107_ 1xC-no-TT	CAAAAGATCGCTGCAGACTTGAAC AAAGAACAAACCA ACTTTGTC
D109_ 3xC	CAAAAGATCGCTGCAGACTTGAACAAAGAACATTCCAGCGAAAGAACATTCCAGCGAAAGAACATTCCAACTTTGTC
D111_ 3xC-noTT	CAAAAGATCGCTGCAGACTTGAACAAAGAACAAACCAGCGAAAGAACAAACCAGCGAAAGAACAAACCAACTTTGTC
D113_ 3x Scrambled	CAAAAGATCGCTGCAGACTTGAACAAATACAATGCAAGCGAAATACAATGCAAGCGAAATACAATGCAAACTTTGTC

Corresponding oligos were resuspended in annealing buffer at 50 μM and hybridized by heating up a mixture of 30 μl of 5'-biotinylated sense strand with 40 μl non-biotinylated antisense-strand oligos to 90 °C for 5 minutes followed by gradually decreasing temperature to 65 °C within 10 minutes. After a 5 minutes pause at 65 °C, oligos were slowly cooled down to RT by letting samples stand in the switched of thermocycler until they had cooled down to room temperature. Prior to loading 1 mg Dynal MyOne C1 streptavidin magnetic beads (Invitrogen) with 250 pmol hybridized oligos, 100 μl of beads were washed twice with 400 μl TE buffer (0.01 % IGEPAL), twice with 750 μl DW buffer and finally resuspended in 400 μl DW buffer. 250 pmol (11.7 μl) hybridized oligos were coupled to 1 mg Dynal MyOne C1

streptavidin magnetic beads during a three hour incubation on a rotary wheel at RT. Uncoupled oligos were removed by washing the loaded beads once with 400 µl TE buffer (0.02 % IGEPAL), and three times with 400 µl DW buffer. Loaded beads with coupled oligos were resuspended at 200 pmol/mg in 100 µl DW buffer. Naked beads were prepared by washing 1 mg Dynal MyOne C1 streptavidin magnetic beads with 400 µl TE buffer (0.01 % IGEPAL), three times with 750 µl buffer DW and resuspended in 100 µl DW. Naked and loaded beads were stored at 4 °C.

Nuclear extract containing 1 mg total protein and 1 mg dsDNA loaded beads were used for each pulldown. Loaded beads were prepared for pulldown assays by adding 1.3 ml blocking buffer (2.5 mM DTT) followed by a 1-hour incubation on rotary wheel at RT. Beads were washed once with 1.3 ml RE buffer (0.02 % IGEPAL) and twice with 2.67 ml buffer G. As control, 1 mg naked beads were prepared for pulldown by washing them twice with 2.67 ml buffer G (2 mM DTT; 0.4 mM PMSF and proteinase inhibitors diluted 1:500. Nuclear extract was centrifuged at 15,000 rcf for 20 minutes at 4 °C and potassium-glutamate was added to the cleared supernatant to a final concentration of 10 mM. Next, nuclear extract was diluted 1:1 with poly(dI-dC) (0.2 mg/ml) in buffer G (2 mM DTT, 0.4 mM PMSF and proteinase inhibitors dilution of 1:5000) and centrifuged for 10 minutes at 15,000 rcf at 4 °C. Proteins binding non-specifically to beads were first removed from supernatant by adding 1 mg washed naked beads at a final concentration of 1.5 mg/ml. The resulting mixture was incubated for one hour on rotating wheel at 4 °C. The cleared supernatant was then incubated with the loaded beads for three hours at 4 °C on a rotary wheel. Beads were washed with 1.8 ml buffer G (resuspending), followed by two non-resuspending washes with 1.8 ml buffer G and a final wash with 1.8 ml buffer G without IGEPAL.

Beads were now ready for elution by trypsin digestion, subsequent precipitation and mass-spectrometry analysis (which were done by David Meierhofer), which were done as follows:

1 ml ice-cold acetone was added to samples followed by a 15 minutes centrifugation at 20,800 rcf at 4 °C. Pellets were washed twice with acetone, dried in a SC210A Speed-Vac (Savant) and resuspended in 100 µl of 25 mM NH_4HCO_3 (10 mM DTT). Samples were incubated for one hour at 56 °C before 100 µl of 25 mM NH_4HCO_3 with 50 mM Iodacetamide were added followed by a 30 minutes incubation in the dark. Afterwards, pellets were washed twice with acetone and

dried again in a SC210A Speed Vac (Savant). Samples were combined with 100 µl of 25 mM NH_4HCO_3 and 1 µl trypsin and incubated overnight at 37 °C. Mass spectrometry analysis was done as described before (Meierhofer et al. 2013), and analyzed using the IPI database (IPI Human v.3.87) and the MaxQuant quantitative proteomics software package. Each sample was background corrected. Proteins were only considered as verified, if they were identified by at least two unique peptide hits.

The following computational analyses were cited from our submitted paper Starick et al. 2015, if not stated otherwise.

6.19 Microarray analysis of transcriptional regulation in IMR90 cells

IMR90 cells were analyzed for their transcriptional response to hormone treatment using the HumanHT-12 v3 Microarray (Illumina). Total RNA of vehicle control or 4-hour hormone (dexamethasone, 1 µM) treated cells was purified. Biotinylated cDNA was synthesized from 500 ng RNA for 3 biological replicates for each condition using the TotalPrep RNA amplification Kit (Ambion). Handling of bead arrays was kindly performed by Aydah Sabah. Samples were hybridized to bead arrays according to Illuminas technical bulletin. After washing and staining, the BeadArrays were scanned using the Illumina BeadStation 500 Bead array. I was analyzing the data for regulated genes using Illumina Genomestudio software. Data was extracted and further analyzed by Dr. Annalisa Marsico as follows: Resulting images were quantified and text files containing raw values were analyzed. Data preprocessing, differential expression analysis and gene annotation were done in R, using available Bioconductor packages (www.bioconductor.com). First, the signals were background corrected with the normexp meth- od16 (Limma package), and an offset of 1 was added to the intensities before normalization and log transformation to ensure that all intensity values are positive. After background correction, the data were normalized between arrays using the quantile method. Differentially expressed probes among different conditions were identified by means of the linear model implemented in the Limma package. In addition, the empirical Bayes method was used to construct moderated t- statistics and adjust for multiplicity of the tests. The Benjamini and Hochberg's method was used to control the false discovery rate. The biomaRt annotation package was used to assign the corresponding gene accessions (Ensembl IDs, Entrez Gene Ids) to each Agilent probe ID. Each gene was then assigned to the median expression value from all the corresponding probes. Genes

with a false discovery rate less than 0.1 and a fold change higher than 2 were considered differentially expressed.

6.20 Hierarchical Bayes Modeling

Dr. Michael Love constructed a Hierarchical Bayes Model (Gelman et al. 2003) to correlate binding of GR (measured by ChIP-Seq read counts) with chromatin features and motif score in DNase hypersensitivity sites (DHS) across experiments and across cell types. The log of read counts for various chromatin features (plus a pseudocount of 1) and the motif score over the annotated DHS of a cell type are arranged as columns of a matrix. For further details I'm referring to Dr. Michael Loves PhD thesis "Statistical analysis of high-throughput sequencing count data ".

6.21 ChIP-Seq: Read Mapping and Peak Calling

The bioinformatic analysis was performed by Morgane Thomas-Chollier, by mapping the reads using bowtie v.1 (Langmead et al. 2009), removal of duplicate reads and subsequent peak calling with MACS v1.4 (Zhang et al. 2008) at an false discovery threshold of 0.01. This results in a high confident set of genome wide GR binding sites for IMR90 and K562 cells.

6.22 *De novo* motif discovery

The following description of the *de novo* motif discovery is cited from the "Transcription factor binding profiles in ChIP-Exo data" master thesis of Jonas Ibn-Salem, with whom I collaborated for the computational part of the ChIP-Exo project.

To identify enriched sequence recognition motifs of the targeted TF and potential cofactors in ChIP-Seq peak sequences, the peak-motifs tool (Thomas-Chollier, Darbo, et al. 2012) from the Regulatory Sequence Analysis Tools (RSAT) suite (Thomas-Chollier et al. 2008; Thomas-Chollier et al. 2011; van Helden 2003) was applied using the four complementary algorithms oligo-analysis, dyad-analysis, position-analysis, and local-word-analysis. With the algorithm oligo-analysis, all possible DNA words of length six or seven nucleotides are tested for over-representation in the peak sequences (van Helden et al. 1998). The expected number of word occurrences is calculated using a Markov model (option -markov auto was used). In addition, the program dyadanalysis detects over-represented dyads, which are pairs of short oligonucleotides (monads) of length 3 bp spaced by a region of fixed width (between 0 and 20 bp) but variable content. The expected frequency of each dyad is estimated as the product of its monads probabilities from the Markov model (van Helden, Rios, et al. 2000). The algorithm position-analysis calculates the

positional distribution of oligonucleotides in the set of ChIP-Seq sequences and detects words that show a positional bias in the peak regions and thus significantly discards from a homogeneous distribution (van Helden, del Olmo, et al. 2000). Finally, the program local-word-analysis tests the over-representation of each possible oligonucleotide word or dyad in positional windows in the test peak sequence sets. To take the positional biased enrichment with respect to the peak summit into account, we also analyzed only the region ±30 bp around the ChIP-seq peak summit separately.

The over-represented words or dyads are then assembled to sets of overlapping words (assemblies) using the program pattern-assembly. This assembly is then used by the program matrix-from-pattern as seed to perform a matrix-based scanning of the peak sequences to collect the most likely instances of the motif and merge them to the final position-specific scoring matrices (PSSM) that represent the recognition motif. For each of the four algorithms, the five most significantly overrepresented motifs were reported and compared to known motifs of TFs in the JASPAR (Sandelin et al. 2004; Mathelier et al. 2014) and TransFac (Matys 2003) database.

6.23 Fraction of ChIP-Seq peaks with GBS

Dr. Morgane Thomas-Chollier performed the ChIP-Seq analysis to find GR binding regions with and without the canonical GR binding motif as follows: GR-ChIP-Seq peak sequences (+/-50 bp around the peak summit) were scanned with the JASPAR motif MA0113.2 (Mathelier et al. 2014) for GR, using the program RSAT *matrix-scan* (Thomas-Chollier et al. 2008; Thomas-Chollier et al. 2011). The background model trained for each cell line on the corresponding peak sequences is a Markov chain of order 1, which accounts for the CpG depletion of vertebrate genomes. To determine which sequence segments are considered as match, we set the threshold on the p-value associated to the weight score. This threshold ranged from 10^{-6} (very stringent) to 10^{-1} (very loose). As control sequences, the coordinates of GR peaks from all cell lines were randomly shifted into the regions flanking the actual peaks. The flanking regions were defined as 2 kb on each side of the peak after extending the peaks by 200 bp on both sides. This was achieved with slop, flank and shuffle from the BEDTools suite (Quinlan & Hall 2010). As above, the background model was trained on this dataset.

6.24 ChIP-Exo-Seq: Read Mapping and Peak Calling

Morgane Thomas-Chollier, Jonas Ibn-Salem and Alejandra Medina-Rivera performed this part of the project. Peconic LLC delivers the ChIP-Exo datasets of my prepared samples (compare chapter 6.6, page 43) as mapped reads for GR-bound regions of IMR90 and K562 cells in BAM files and a separate file for the called peaks. Peconic LLC was using GeneTrack (Albert et al. 2008) to separately call peaks for the positive and negative strand. The two peaks were defined as "peak pair" using the genomic coordinates of peak summits as start and end points. To create BED files as input for the ChIP-Exo analysis pipeline and Exo-profiler tool, we compared different peak callers. Therefore raw peaks were generated by Alejandra Medina Rivera by running additional peak calling programs with default parameters on the ChIP-Exo dataset. The peak callers used were:

MACE (http://chipexo.sourceforge.net ; April 2014),

MACS (Zhang et al. 2008),

SWEMBL (http://www.ebi.ac.uk/~swilder/SWEMBL/ ; April 2014) and

peakzilla (Bardet et al. 2013).

6.25 ExoProfiler pipeline

To analyze the local 5' coverage distribution centered on TF-BSs, we (Dr. Morgane Thomas-Chollier and Jonas Ibn-Salem) developed a program called ExoProfiler, implemented in Python. It takes as input the mapped reads from a ChIP-Exo experiment (BAM format) and TF-BS coordinates for a motif of interest (BED format). Thanks to the python package HTSeq (Simon Anders, Posted August 19 2014), ExoProfiler is computationally efficient, processing a typical dataset in a few minutes on a common desktop computer.

TF-BS predictions (matrixScan WS): Before running ExoProfiler, a TF-BS coordinates BED file must first be obtained from a motif-scanning program. Here, we used RSAT *matrix-scan* (Thomas-Chollier et al. 2008; Thomas-Chollier et al. 2011), as described above, with a stringent threshold set on the weight score p-value 10^{-4}. For palindromic motifs reported at the same position on both strands, the match associated to the lowest p-value was retained. The motifs used for scanning (using *matrix-scan*) were obtained from a collection of reference motifs (JASPAR November 2013 ((Mathelier et al. 2014), vertebrates only, 205 motifs) and from *de novo* motifs discovered on ChIP-Seq peaks sequences with RSAT *peak-motifs* (Thomas-Chollier,

Darbo, et al. 2012; Thomas-Chollier, Herrmann, et al. 2012) (default parameters, using the four algorithms, 5 motifs per algorithm), both on the complete peak length or on ±30 bp around the peak summit to better benefit from the two algorithms based on positional bias. The results shown are for ± 30 bp around the peak summit, except for a few cases explicitly indicated in the figure legends. As control, each motif had its columns randomly permuted ten times independently using RSAT *convert-matrix* (Thomas-Chollier et al., 2011), which maintain the statistical properties of the original matrix, but not its biological significance. RSAT *compare-matrices* (Thomas-Chollier et al. 2011) was finally run to ensure that the permuted matrices are distinct (-lth Ncor 0.99), and not too similar to the original matrix (-lth Ncor 0.4).

For the *in silico* mutated GBS consensus analysis, this TF-BS prediction step was replaced by a pattern-matching approach using RSAT *dna-pattern* (Thomas-Chollier et al. 2011). The patterns were expressed with IUPAC code; the "mutation" is achieved replacing a chosen letter (e.g. A) by "not this letter" (e.g. B coding for C or G or T).

ChIP-Exo 5′ coverage (ExoProfiler): For each TF-BS from the BED file, ExoProfiler defines a short region (e.g. +/-30 bp) centered on this TF-BS. Within this region, the ChIP-Exo coverage is analyzed as follows: First, ExoProfiler discards regions not covered by at least 5 ChIP-Exo reads to limit unnecessary computations. Next, ExoProfiler reduces all mapped ChIP-Exo reads to their 5′-most base. For the ChIP-Exo footprint, only this most 5′ position of the reads is informative as it marks the boundary of protection from lambda exonuclease digestion provided by cross-linked proteins. The program is fully strand-sensitive to ensure that forward and reverse read coverage is calculated properly with respect to the motif orientation. If the TF-BS is located on the reverse strand, reads on the direct strand are counted as reverse and reads on the reverse strand as forward, and all counts are adjusted to the correct distance from the motif center.

Plotting footprint profiles (ExoPlotter):

The pipeline outputs 4 plots of the short regions centered on motifs, with a companion R script:

(1) A color chart representation, which mainly serves to control that the motifs are correctly aligned and that the regions are not shifted by one base pair, a relatively common error when working with genomic coordinate files.

(2) A heatmap of the 5′ coverage combining the forward (blue) and reverse (red) strand. The color intensities are log transformed after a pseudocount of 1 is added to all 5′ coverage counts.

(3) A similar heatmap, ordered after a hierarchical clustering of the ChIP-Exo 5′coverage at individual short regions. The distance between individual sites is calculated as follows: After adding a pseudo-count of 1, each 5'coverage count *c* is log normalized by *log(c)/log(c_max))*, where *c_max* is the maximal count for forward or reverse 5' coverage. For each individual site, the coverage count signal is than smoothed along the genomic positions using the 'smooth' function in R with default parameters. The Euclidean distances on the log-normalized and smoothed count vectors is used for hierarchical clustering.

(4) A footprint profile, summing the coverage at each position for all regions, for the forward (blue) and reverse (red) strand. The raw sum is plotted unless the user chooses to add the permuted motif control. In this case, the values are normalized by dividing the counts by the number of motifs matches in the assay and in each permutation. A p-value, determining the significance of the enrichment of ChIP-Exo reads around the motif, is calculated using a Wilcoxon rank-sum test. It tests if the total coverage on the short region is significantly higher than on the short regions extracted when using permutated motifs.

For all these plots, there is no shift in the position of the reads. Optionally, the program calculates the consensus sequence of all regions aligned by the motif midpoint, which necessitates as additional input the reference genome in FASTA format.

6.26 Structural alignment

Structural alignments of protein and DNA complexes were obtained as follows and performed by Dr. Marcel Jurk, at this time Postdoc in our group. A structural model of a DNA hybrid sequence was generated using 3DDart (van Dijk & Bonvin 2009). The hybrid sequence always consisted of the GR half site AGA ACA and the binding motif of the alignment partner. The latter was derived from the corresponding PDB file and comprised a sequence matching the JASPER consensus sequence. For instance, a hybrid sequence for the GR-Stat3-DNA complex consisted of the 5′-TG CAT TTC CC-3′ motif of the Stat3 structure (PDB entry 1BG1) and the 5′-AGA ACA CCC TGT TCT-3′ for GR (PBD entry 3G6U). An overview of all hybrid sequences used is given in Table 11. GR and potential interaction partner binding

motifs were aligned using the CE-align algorithm (Jia et al. 2004) to the 3D-DART DNA model of the hybrid sequence. Only the GR-Stat3-DNA complex was energy minimized using XPLOR-NIH version 2.30 (Schwieters et al. 2003) to analyze the potential interface. A complete list of structures used for alignment is provided in Table 11.

Table 11: Structures used for alignment of DNA-Protein interactions of GR and possible co-factors.

Protein	PDB	sequence of bound DNA[$]	hybrid sequence[$]
GR	3G6U	A AGA ACA CCC TGT TCT	-
STAT1	1BF5	TGC ATT TAC GGG AAA CTG	A AGA ACA TTC CAG GAA
TEAD1	2HZD	[#]	A AGA ACA TTC CTC TGC
ETS1	1K79	CAC ATT TCC GGC ACT	A AGA ACA TTC CGG CAC T
ELK1	1DUX	TGA CCG GAA GTG T	A AGA ACA TTC CGG TCA
STAT3	1BG1	TG CAT TTC CCG TAA ATC T	TG CAT TTC CC AGA ACA CCC TGT TCT

[$]all sequences listed are in 5'->3' orientation

[#]not bound to DNA - alignment via Mos1 (PDB 3HOS)

7 Results

7.1 Characterization of GR binding and GR-dependent gene regulation in IMR90 & K562 cells

A prerequisite for correlating GR binding to chromatin state, as catalogued for IMR90 and K562 cell lines, is GRs functionality in these cell lines. I therefore examined the transcriptional response of three different genes known to be activated by dexamethasone (Dex), a synthetic ligand for GR, in other cell lines (Cvoro et al. 2011; Dougherty et al. 2012; Pereira et al. 2014). For IMR90 cells, I tested different hormone treatment times and concentrations and found a hormone-dependent up-regulation of all three genes examined regardless of treatment time or hormone concentration (Figure 7, left). Glucocorticoid dependent gene response in K562 cells is exemplarily shown for the *FKBP5* gene (Figure 7, right).

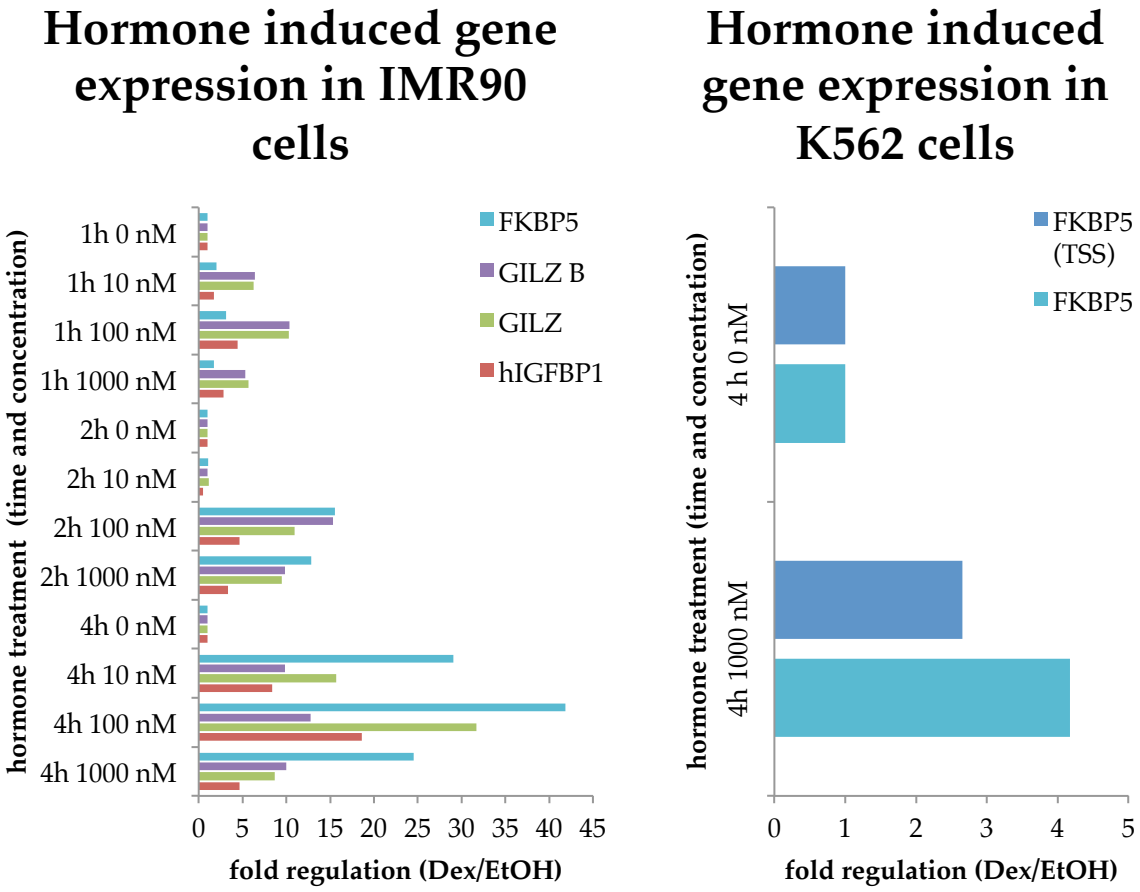

Figure 7: Time and concentration-dependent transcriptional response to hormone treatment of IMR90 and K562 cells. Shown are fold inductions for 3 genes (IMR90 cells) and two loci (K562 cells) analyzed with qPCR of mRNA levels for hormone over vehicle treated cells. Results were normalized to gene expression values of *RPL19*, a gene that does not respond to hormone treatment (compare Microarray data, Figure 2).

Since a 4 hour treatment yielded the strongest up-regulation of mRNA levels, this time point was used to analyze transcriptional response to dexamethasone globally in IMR90 cells using microarrays. Although longer treatments typically result in more pronounced expression changes, we chose a relatively short treatment time to increase our chances of finding primary target genes.

Next, we examined the genome-wide transcriptional response to hormone treatment in IMR90 cells using HumanHT-12 v3 Microarrays (Illumina) that have 48,804 probes covering 27,455 annotated human genes and additional long non-coding RNAs. For 32,813 of 48,804 probes a signal was detected. The analysis showed a significant regulation for 3856 probes (Figure 8), revealing a differential response to dexamethasone of approx. 8 % of all tested probes.

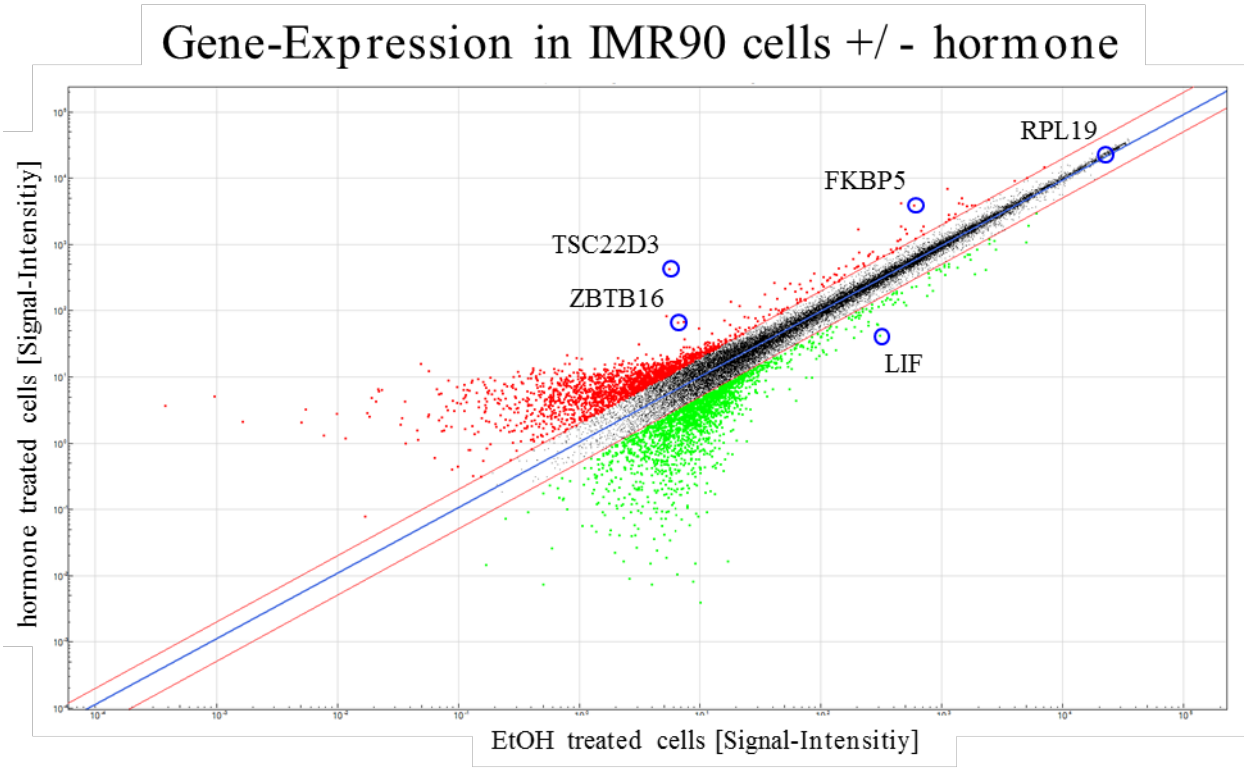

Figure 8: Logarithmic plot of probe intensities for different genes from a Microarray (Illumina). Differential transcript levels for IMR90 cells to 4h 1 μM hormone (Dex plus, Y-axis) versus vehicle (Dex minus, x-axis) treated IMR90 cells. Genes with an at least 2 fold increase in intensity upon hormone treatment are marked in red, genes with a decrease greater than two-fold in green. Genes in between the red lines were considered not to be regulated by GR. Microarrays were done in quadruplicates. Plot shows the average from all four experiments. Plot was generated using the Genomestudio software (Illumina).

1873 probes were up and 1983 down regulated, when a 2-fold threshold for intensity values was applied. To identify genes with significant changes in expression upon hormone treatment, 4 biological replicates were analyzed by Dr. Mike Love, using the beadarray package (Dunning et al., 2007). Target genes

included genes known from other studies indicating that the genes we identified likely reflect true GR target genes, like *FKBP5* (Scharf et al. 2011), *TSC22D3* (Ayroldi et al. 2007) and *ZBTB16* (Wasim et al. 2010). Also, the gene used for normalization in qPCR assay, *RPL19* was identified and indeed does not respond to hormone treatment, consistent with our qPCR results.

Together, the analysis of transcriptional responses to the synthetic glucocorticoid dexamethasone shows that the glucocorticoid receptor is functional in IMR90 and K562 cells.

The well-characterized epigenome of IMR90 and K562 making these cells a good model to study the correlation between GR binding and the chromatin landscape. We therefore set out to identify where in the genome GR binds using Chromatin Immunoprecipitation (ChIP). The optimal ChIP-conditions for both cells were optimized independently using different lengths of sonication, to shear chromatin into 100-400 bp fragments, and different cross-link times. This is essential for ChIP, as smaller fragments lead to the identification of bound regions at higher resolution. I found that for both IMR90 and K562 cells the chromatin was sheared to the desired range after 50 min sonication and cross-linking for 3 minutes.

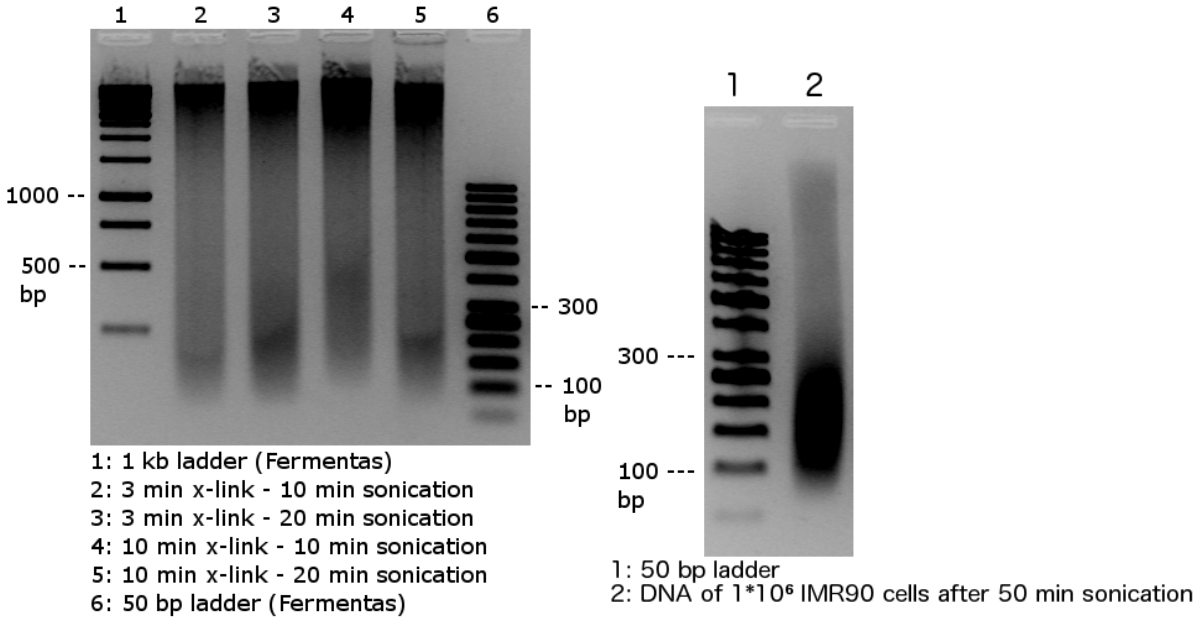

Figure 9: SYBR-Green stained agarose gel. Left: Sheared DNA from 1x10^6 cells for different crosslink and sonication times. Most of the genomic DNA is not in the right range of approximately 300 bp. Right: DNA of 1x10^6 cells cross-linked for 3 minutes and sonicated for 50 minutes shows the desired size range and these conditions were used for ChIP-Seq experiments.

To find optimal conditions for enriching GR-bound loci via Chromatin Immunoprecipitation (ChIP), I compared agarose beads and magnetic Dynabeads®.

Genomic binding of GR upon hormone treatment was analyzed after 1 h in IMR90 cells. Although both types of beads show a hormone-dependent enrichment of genomic loci bound by GR (Figure 10), the enrichment when using magnetic beads was consistently higher. These single experiments represent the GR-binding site enrichment in both IMR90 and K562 cells that was reconfirmed for the pooled ChIP-Sequencing (ChIP-Seq) samples (Figure 12, page 64).

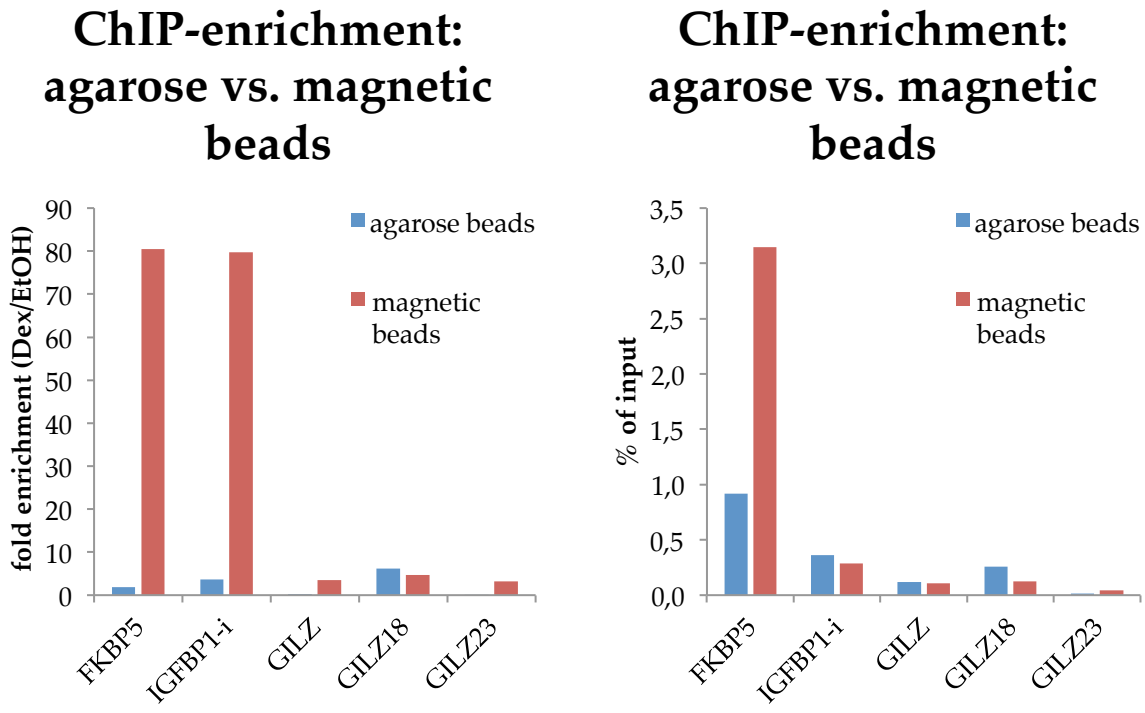

Figure 10: Comparison of Agarose and magnetic (DYNABEADS®) beads for ChIP. Enriched GR-bound regions (GBR) were either normalized to the control HSP70 locus (left) or shown as % of input DNA (right) after 1-hour hormone treatment of IMR90 cells.

In addition, the amount of non-specific binding quantified by qPCR in vehicle control was lower using the Dynabeads® (Figure 11), explaining the superior enrichment of GR-bound regions (GBR) when using Dynabeads®. Therefore, these beads were used for subsequent experiments in IMR90 as well as K562 cells.

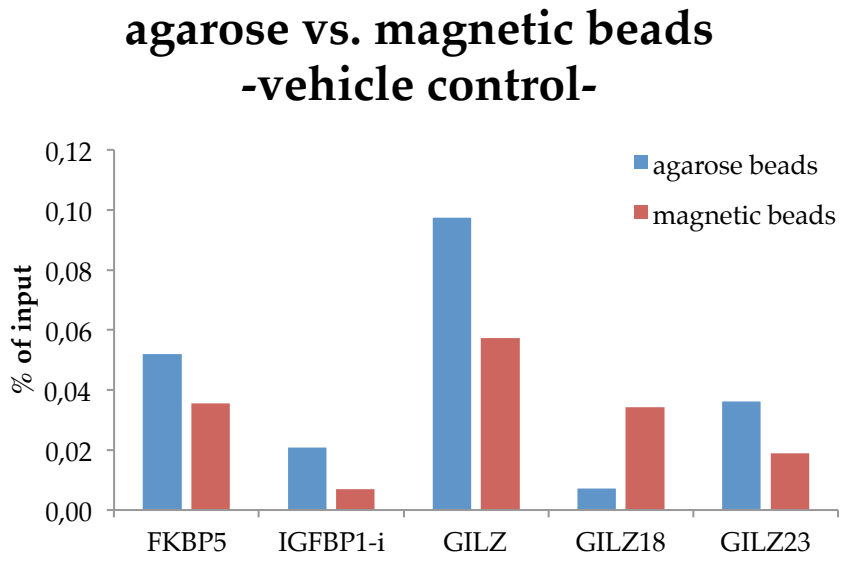

Figure 11: ChIP-comparison of agarose and magnetic (DYNABEADS®) beads for unspecific binding in vehicle (EtOH) control treated IMR90 cells. The amounts of precipitated material are shown as % of input DNA. Magnetic beads have less unspecific binding, as the percent of input DNA is mostly lower compared to agarose beads.

After optimizing the ChIP-procedure, we set out to sequence the bound GR regions (GBRs). Multiple GR-ChIPs were performed in parallel for 1-hour hormone-treated cells. The precipitated DNA was pooled for IMR90 and K562 cells independently and tested for GBR enrichment at specific loci (Figure 12). Afterwards the samples were handed over to the sequencing facility for library preparation and subsequent next generation sequencing.

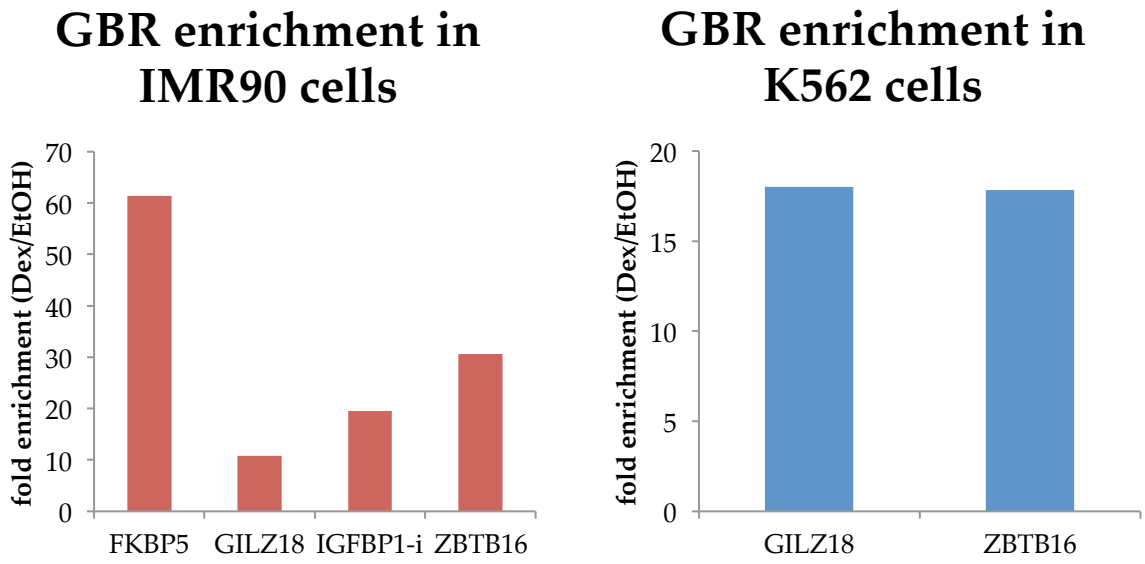

Figure 12: Comparison of enriched GR-bound regions (GBRs) shown as fold (Hormone (Dex) over vehicle control (EtOH)) for IMR90 (left) and K562 (right) cells of pooled ChIP-Seq samples after 1 hour hormone (dexamethasone) treatment.

In collaboration (Dr. Morgane Thomas-Chollier) the sequenced reads were mapped to the genome and bound regions (peaks) were called using MACS (Zhang et al. 2008). This analysis resulted in the identification of 47.630 bound loci in IMR90 and 6.329 in K562 cells (Table 12). Subsequent motif searches in the GR-bound regions (GBR) in IMR90, K562 and U2OS cells showed a striking difference in the fraction of peaks with an apparent GR binding sequence (GBS) (see chapter 7.2, page 67). Previous ChIP of GR in U2OS cells (osteosarcoma) revealed 41.402 GBRs (Thomas-Chollier et al. 2013), which were downloaded to compare to our findings in IMR90 and K562 cells.

Table 12: ChIP-Seq data for GR

Cell-line	Total reads	Mapped reads	Peaks	Peak size [bp], mean (SD)
IMR90	28262570	20061690 (70.98 %)	47630	387.8 (±165.2)
K562	32716747	20090476 (61.41 %)	6329	423.1 (±192.3)

Furthermore, consistent with findings by others (Reddy et al. 2009; Yu et al. 2010; John et al. 2011; Pan et al. 2011), we found that the identified GR-bound loci were different between the two cell lines, exemplarily shown for GBRs on a locus of chromosome 6 (Figure 13, blue box). Another observation we made is that out of the many potential GR binding sequences in the genome (Figure 13, marked by green ticks) only a small subset is actually bound, here exemplified for a region containing the *FKBP5* gene on chromosome 6, leading to the question: what is different for the regions where GBSs are bound when compared to regions with bona fide GBSs that are not bound?

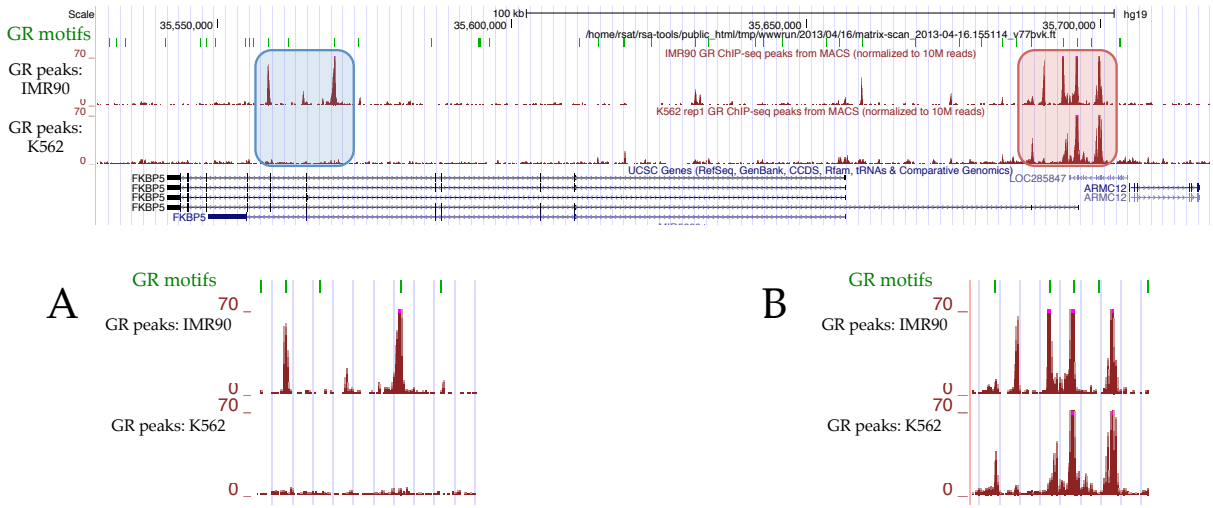

Figure 13: top: UCSC-genome browser screenshot of canonical GR motifs found in the genome (green ticks) and actual GR binding shown as ChIP-Seq peaks for IMR90 and K562 cells in GRCh37/hg19 assembly. Zoomed blue (A) and red (B) regions illustrate the cell-type-specific binding pattern of GR. It also shows the relatively rare GR binding events compared with the frequent occurrence of the canonical GR motif. The missing GR binding in the *FKBP5* gene in K562 cells explains the necessity of additional controls for ChIP enrichment as used in Figure 12.

Together, we found that GR signaling is functional in K562 and IMR90 cells. The ChIP-Seq data we generated in these cell lines was used to study the role of chromatin in specifying where GR binds to the genome (results presented in chapter 7.2). In chapter 7.3 (page 77) this data is used to study the role of DNA sequences specifying where in the genome GR binds.

7.2 What defines a GR binding site? – Role of the chromatin landscape

7.2.1 Chromatin state at GR binding sites

We and others (Reddy et al. 2009; Yu et al. 2010; Pan et al. 2011; Gertz et al. 2013) found that only a subset of all potential GR binding sites (GBS) found in the genome are actually bound by GR and that the set of bound sites varies among different cell types. Since GR-bound loci are mainly predetermined by the accessibility of the chromatin (John et al. 2011), which is also cell-type-specific, we studied the role of various chromatin features in determining where TFs bind. Our analysis of GR binding sites in IMR90 cells showed that approx. 15 % (44.772) of all DNase I hypersensitive sites (DHS or "open" chromatin) are bound by GR and these 44.772 GR-bound DHS are almost all (~96 %) of GR's binding sites (Figure 14) in IMR90 cells. This emphasizes two things: First, the wide distribution of GR binding sites in the genome and thus diversity of potential GR-regulated genes and second, the strong correlation between accessible chromatin and GR binding, as it was also found by other studies, where 95 % of all GR binding sites were found in DHS in AtT-20 cells (John et al. 2011).

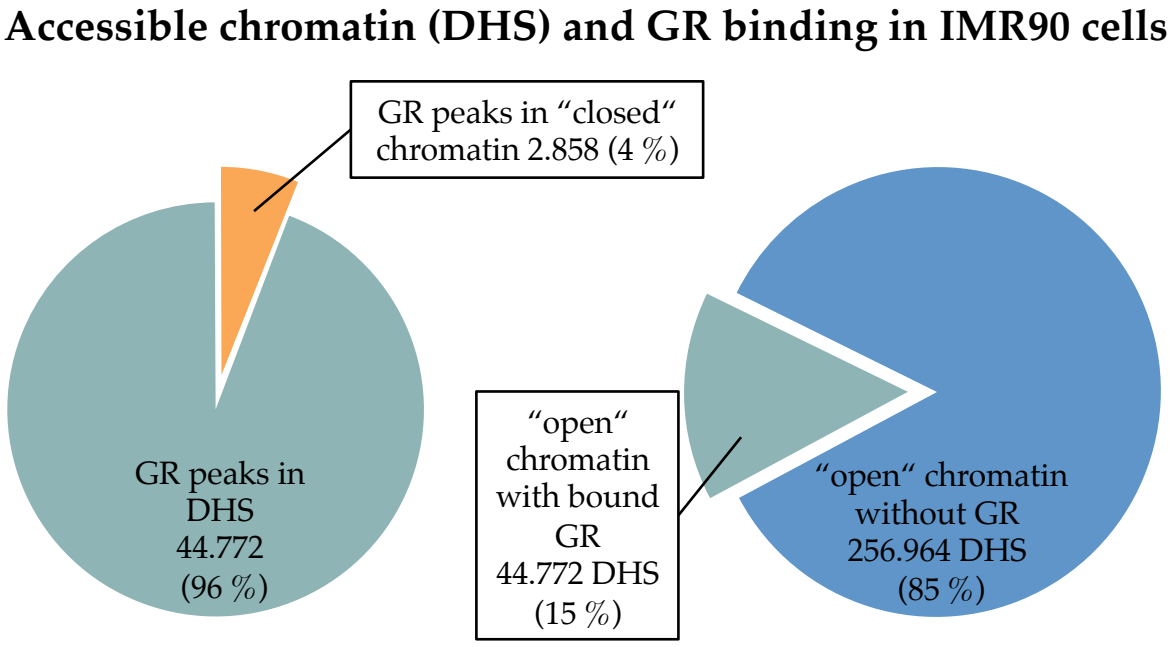

Figure 14: Intersect between the distribution of GR binding regions (GR peaks in ChIP) and DNase I hypersensitive sites (DHS or "open" chromatin) in IMR90 cells. Left: The majority (44.772) of all GR binding is detected in predefined "open" chromatin in IMR90 cells. Only a small amount (2.858) of GR peaks were not found in predefined DHS, emphasizing the importance of DNA accessibility for GR binding as was also found by others (John et al. 2011). Right: Roughly 15 % of all accessible chromatin is bound by GR. Analysis done in collaboration with Dr. Michael I. Love.

GR was believed to act as a pioneering factor itself to determine its own binding sites into the chromatin landscape of different cell lines (Richard-Foy & Hager 1987;

Becker et al. 1984). This was redefined after comparing GR-bound regions (GBRs) globally pre- and post hormone treatment, showing GR was binding in 95 % of the cases to a set of predefined accessible chromatin (John et al. 2011). We could confirm this finding by comparing GBRs to non-hormone treated DHS in IMR90 cells, in particular hormone induced GR binding to specific regions pre-defined as "open" (DHS) or "closed" (no DHS) prior to hormone treatment. Upon hormone treatment we indeed found GR-bound to "closed" chromatin regions (Figure 15, orange boxes) and subsequently tested whether these GBRs are at false negative DHS, or if GR can indeed act as a pioneering factor, rearranging nucleosome density and consequently DNase I sensitivity.

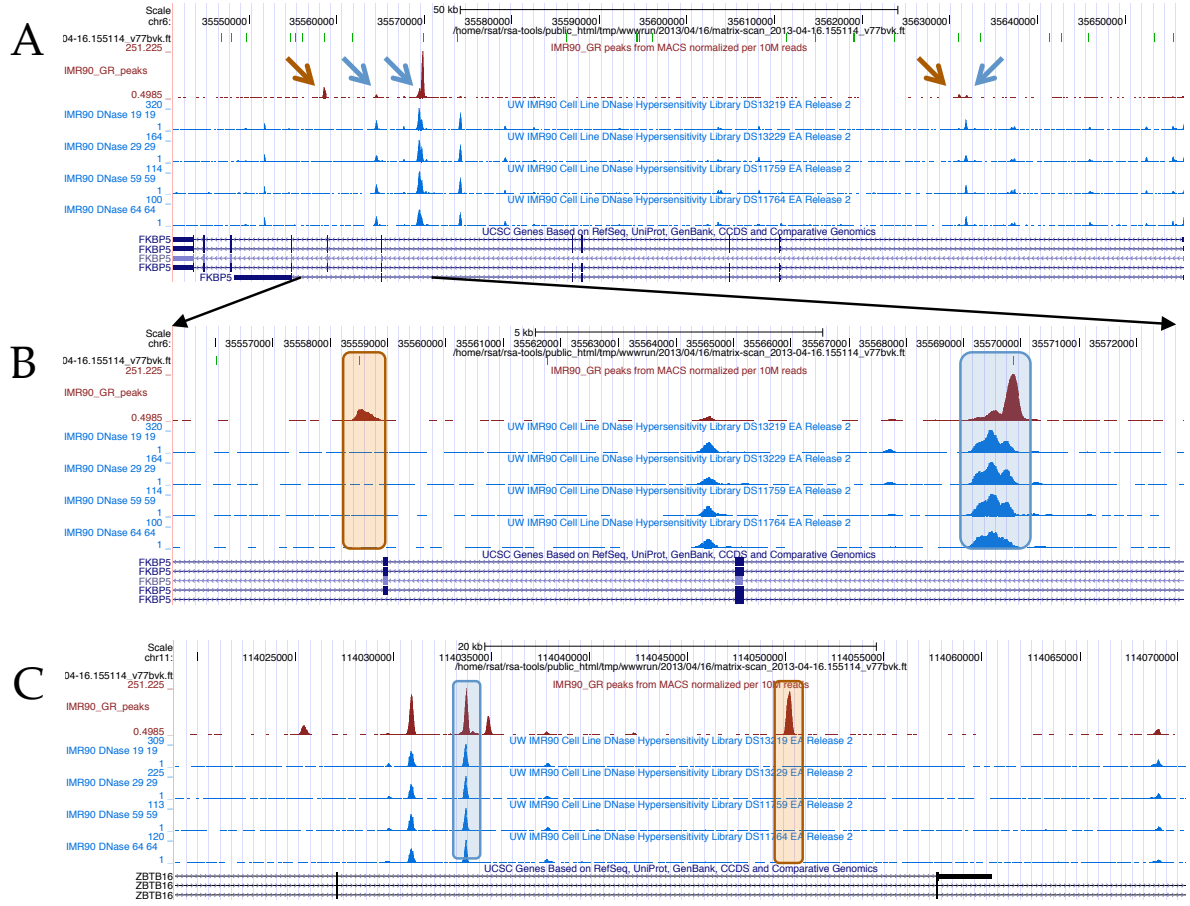

Figure 15: IMR90 GR-ChIP-Seq peaks, DNase I hypersensitivity (GEOs: GSM530665, GSM530666, GSM468792, GSM468801) and presence of canonical GR motif shown as UCSC Genome Browser screenshot (GRCh37/hg19) assembly. A: GR binds to five different loci within the FKBP5 gene. Two GR peaks (orange arrows) were found in "closed" chromatin and contain a canonical GR motif (defined in RSAT). B + C: Comparing GR-bound regions with (blue box) and without (orange box) a predefined DNase I hypersensitivity site in the FKBP5 (B) and ZBTB16 (C) gene. These four regions (boxes) were compared for DNase I sensitivity (Figure 16).

We compared DNase I sensitivity in regions of the GR-regulated genes FKBP5 and ZBTB16 at positions with and without pre-existing DNase I hypersensitivity (Figure 15, B and C). Chromatin changes in IMR90 cells were quantified using qPCR, by

comparing the amount of DNase I resistant chromatin at GR-bound loci in regions defined as "open" (Figure 15, blue boxes) or "closed" (Figure 15, orange boxes) pre and post 1 hour hormone treatment. The DNase I sensitivity serves as a proxy for the accessibility of DNA: The more DNA is degraded by DNase I, the more accessible the specific DNA loci is to e.g. TFs. Our DHS assay shows that regardless of whether a site was "open" or "closed" before hormone treatment, GR is binding to both predefined regions (as shown in Figure 15 B and C) in the FKBP5 and ZBTB16 gene and all tested loci became more accessible or are "opened" by GR binding (Figure 16, compare red and blue bars per loci). This shows upon hormone induced stimulation GR is able to bind to predefined "closed" regions and invalidates the speculation that predefined "closed" regions could be false negative DHS.

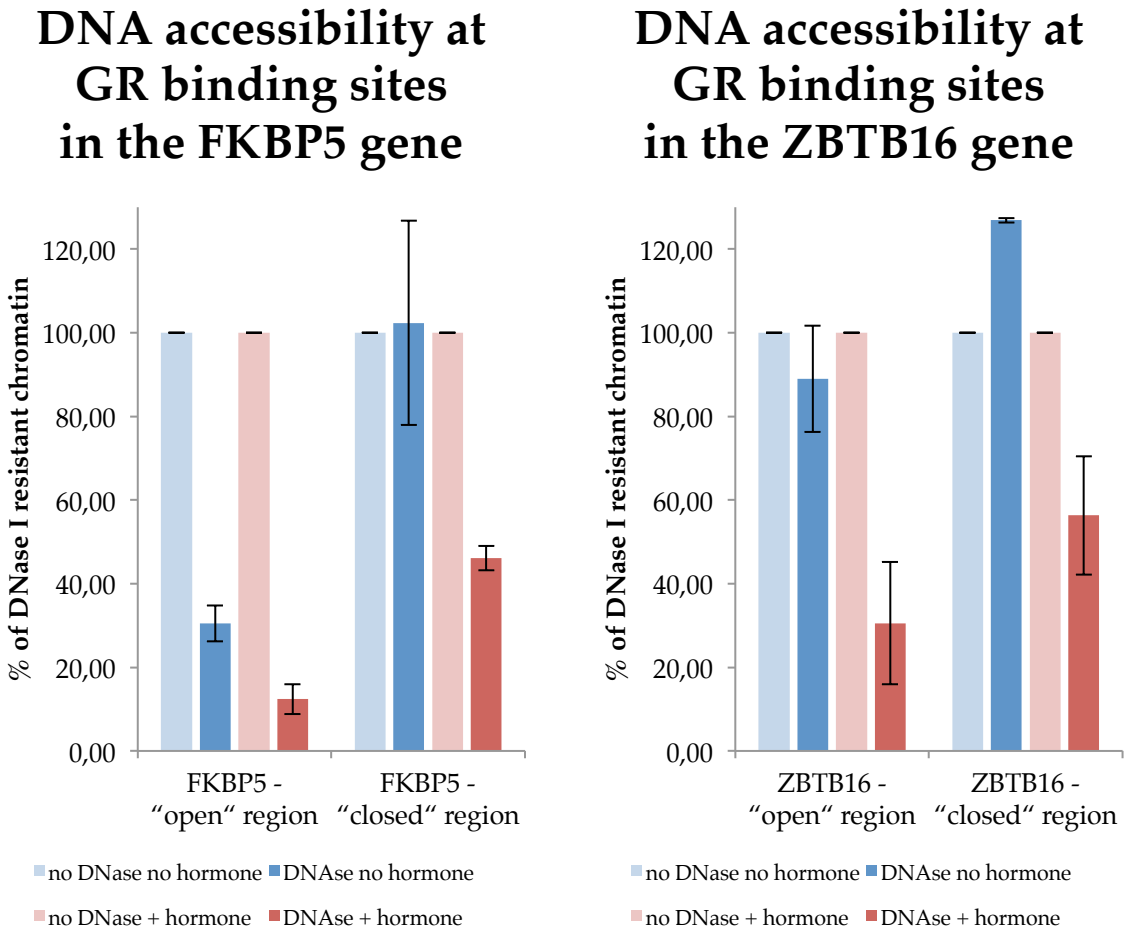

Figure 16: Comparison of DNase I sensitivity at two different loci for the FKBP5 (left) and ZBTB16 (right) genes before (blue bars) and after (red bars) 1 hour hormone-treatment (1 μM dexamethasone, IMR90 cells). Shown is the percentage of DNA left after DNase I treatment (dark colors) compared to non DNase I treated (light colors) purified chromatin. The "closed" region is less prone to DNase I degradation before hormone treatment, compared to predefined "open" region in the FKBP5 gene (blue bars). Whereas hormone treatment leads to DNA degradation of both predefined "open" and "closed" regions for the FKBP5 and the ZBTB16 gene. Values were normalized to a non-GR-bound region in predefined "open" chromatin (OR1A1 gene). Experiment was done in two biological replicates; Error bars indicate standard error of mean.

We show here, as also found by others (John et al. 2011), a hormone induced chromatin remodeling of the chromatin landscape resulting in increased DNase I accessibility or "openness" at specific loci. However, our assay does not permit to draw a conclusion whether GR is altering the DNA accessibility independently or if other hormone induced processes, like cofactors, are involved.

Next, we will have a closer look at chromatin features and the bound motifs of GR-bound regions in "open" and "closed" chromatin. The presented analyses were done in cooperation with Dr. Michael I. Love and complement my sequence-based analysis of GR binding sites (chapter 7.3, starting page 77).

7.2.2 *Chromatin features at GR binding sites*

The finding of GR binding to predefined "open", but also to not predefined "open", so called "closed" chromatin (compare Figure 15), raises the question what is different for these two GR binding environments in terms of for example motif score? Interestingly, no single motif was found to be specifically enriched for GR-bound regions (GBRs) in "closed" chromatin, which would indicate a specific cofactor enabling GR to open up the predefined "closed" chromatin. On the other side, if GR is binding to "closed" chromatin regions independently of a cofactor, are bound motifs matching to the canonical GR motif preferred, since they could enable a stronger GR-DNA interaction with higher affinity towards the canonical GR motif? To study if the canonical motif score is more important in guiding GR to chromatin regions predefined as not DNase I hypersensitive (or "closed") than for predefined "open" regions, we compared motif scores in GR peaks in predefined "open" and "closed" chromatin (Figure 17). Since my former experiments (Figure 16, page 69) showed GR binding in "closed" chromatin alters the chromatin structure, by making it accessible to DNase I degradation, we thought of the possibility of GR acting as a pioneering factor himself. If GR would bind independently of cofactors, we assume binding happens at constrained versions of the canonical GR motif and this could coincide with higher motif scores in GR binding sites in "closed" regions.

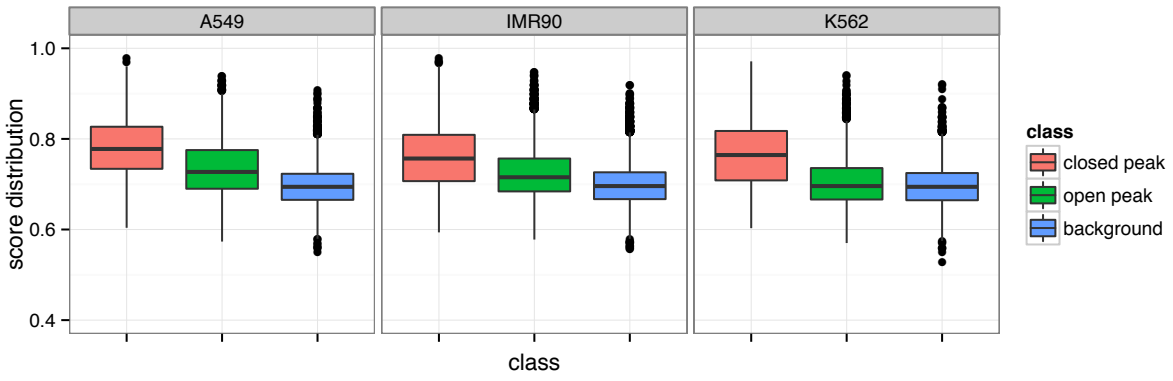

Figure 17: Distribution of top scoring motifs for various regions: GR peaks not overlapping any DHS ("closed" peaks), GR peaks overlapping DHS, ("open" peaks), and regions which are 2 kb randomly upstream or downstream from a peak (background). In all three tested cell lines the motif scores of "closed" peaks are higher compared to "open" peaks. For A549 and IMR90 cells "open "and "closed" peaks have higher motif scores compared to background, only in K562 cell motifs score of "open" peaks are close to background. Analysis done in collaboration with Dr. Michael I. Love and is cited from his PhD-thesis: "Statistical analysis of high-throughput sequencing count data". Black line indicates median average surrounded by the first and third quartile (boxed areas).

The motif scores in "closed peaks" are indeed higher compared to GR binding sites in predefined "open" chromatin (Figure 17) in all three tested cell lines, meaning

the GR tends to bind sequences better matching the canonical consensus motif, if the site is embedded in "closed" chromatin, as it was also found by others (Gertz et al. 2013). This supports the well known theory of GR acting as a pioneering factor itself (Becker et al. 1984; Richard-Foy & Hager 1987) for the small fraction (4 % in IMR90 cells and 5 % in AtT-20 cells (John et al. 2011)) of GR peaks found to be present in "closed" chromatin.

To study what chromatin features discriminate GR-bound sites in "open" and "closed" chromatin would not reveal particular GR binding site predictors, it would rather result in chromatin features correlating with DNase accessibility in general. We therefore focused our study to discriminating features of GR-bound and unbound regions in "open" chromatin, as GR-bound regions in predefined "open" chromatin accounts for the largest part of GR interactions with the genome (compare Figure 14 and (John et al. 2011)). Dr. Michael I. Love implemented Hierarchical Bayes modeling for multiple cell lines. This can uncover positive or negative correlations of a given chromatin feature with GR binding within the accessible chromatin universe. To develop and implement such an algorithm was one of the goals for Dr. Love's PhD-thesis "Statistical analysis of high-throughput sequencing count data". Therefore we could both work with my data to find answers to the emerging question of GR binding.

We (Dr. Michael I. Love) therefore analyzed different chromatin features, as well as the GR motif and how they correlate with GR binding (Figure 18). The 12 analyzed chromatin features, as well as the GR motif score and the input control for GR-bound sites in DHS are mostly consistent across cell types and in general reflect known properties of distal regulatory elements and enhancers.

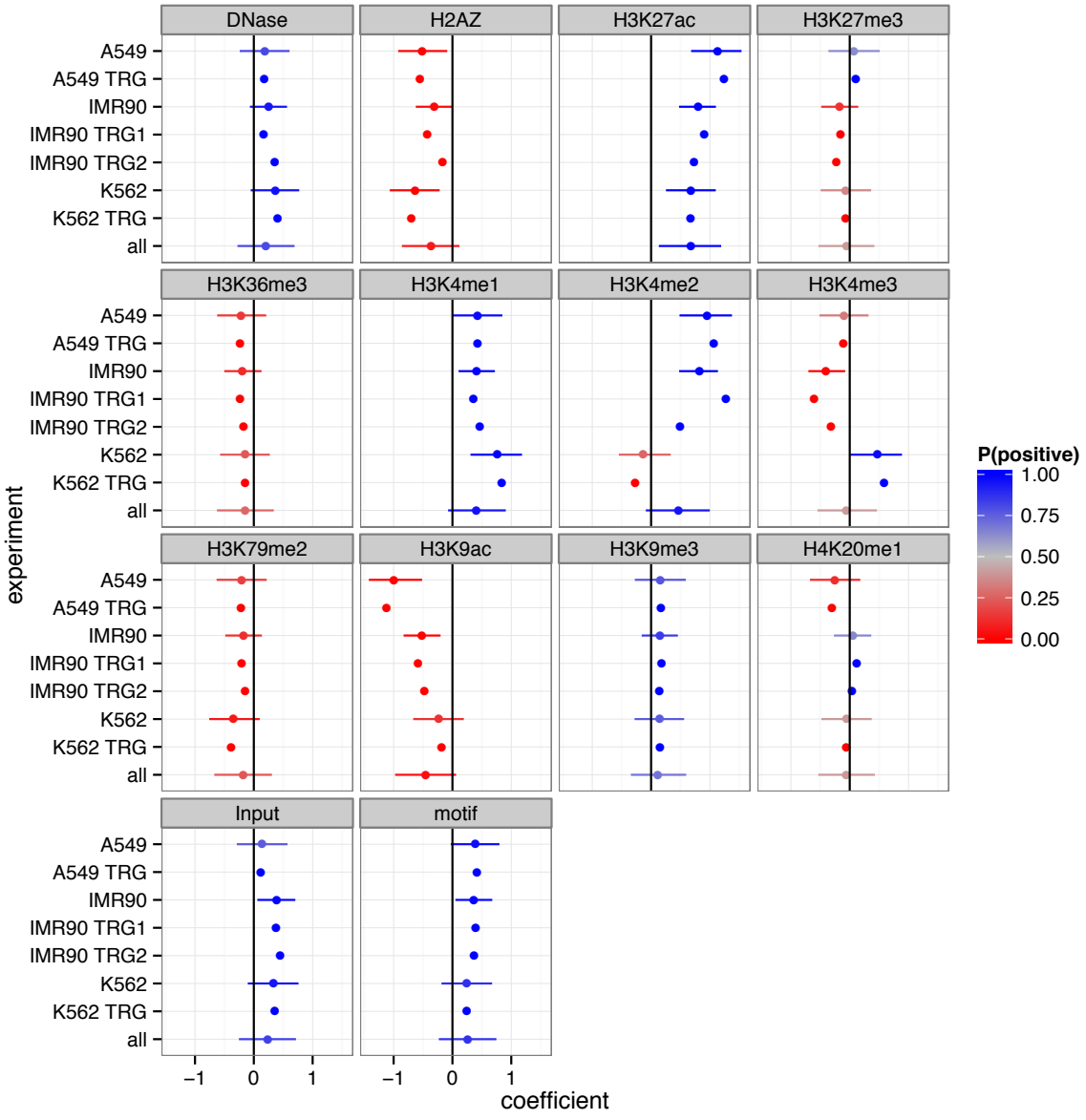

Figure 18: Hierarchical Bayes modeling to compare GR-bound and unbound DNA hypersensitive sites for 12 different chromatin features, GR motif and ChIP-input (as control). Datasets marked TRG (Transcriptional Regulation Group), indicate our own ChIP-Seq data. In all tested cell lines DHS together with H3K4me1 (enhancer mark), H3K27ac (an active enhancer mark) and H3K9me3 (constitutive repression), together with the input control and the GR motif itself could be identified as positive marks for GR binding. Negatively correlating marks are H2A.Z (nucleosome stabilizing), H3K36me3 (increased nucleosome density), H3K79me2 (associated with DNA replication) and H3K9ac (marks promoter regions). Analysis done in collaboration with Dr. Michael I. Love and the plot is from his PhD-thesis: "Statistical analysis of high-throughput sequencing count data".

Consistently positive features correlating with GR binding at DHS are for example H3K27ac, indicating active enhancers, and H3K4me1, which is used to identify enhancer regions in general (Creyghton et al. 2010). H3K9me3, associated with constitutive gene repression (Hublitz et al. 2009) also positively correlates with GR binding in "open" chromatin. Furthermore, the ChIP-Seq input control, DNase I sensitivity and the GR motif itself have also a positive correlation with GR binding in

"open" chromatin. Negative correlations were found with following histone marks. H2A.Z is associated with stronger nucleosome-DNA interactions (Kumar & Wigge 2010). H3K36me3 is associated with increased nucleosome occupancy in particular at exons (Schwartz et al. 2009) H3K79me2 is linked to cell cycle dependent DNA replication (Fu et al. 2013), whereas H3K9ac is associated to promoters from actively transcribed genes (Koch et al. 2007).Hierarchical Bayes modeling identified also cell-type-specific features correlating with GR binding sites. Those are typical promoters marks, e.g. H3K4me3 is a histone modification typically found at high-CpG promoters (HCP) (Mikkelsen et al. 2007), and if found at low-CpG promoters (LCP), H3K4me3 marks active transcription (Karlić et al. 2010).

Together, we found H3K4me3 and H3K9ac to be negatively correlated with GR binding in IMR90 and A549 cells, whereas K562 cells shows a positive correlation between GR binding and the H3K4me3 mark. To test if the negative correlation with marks typically found at promoters reflects decreased binding of GR to such regions we analyzed GR binding in A549, IMR90 and K562 cells at "open" chromatin regions (DHS) for bound or unbound GR binding site, focusing on different genomic regions, namely: promoters, exons, introns and distal regions (Figure 19).

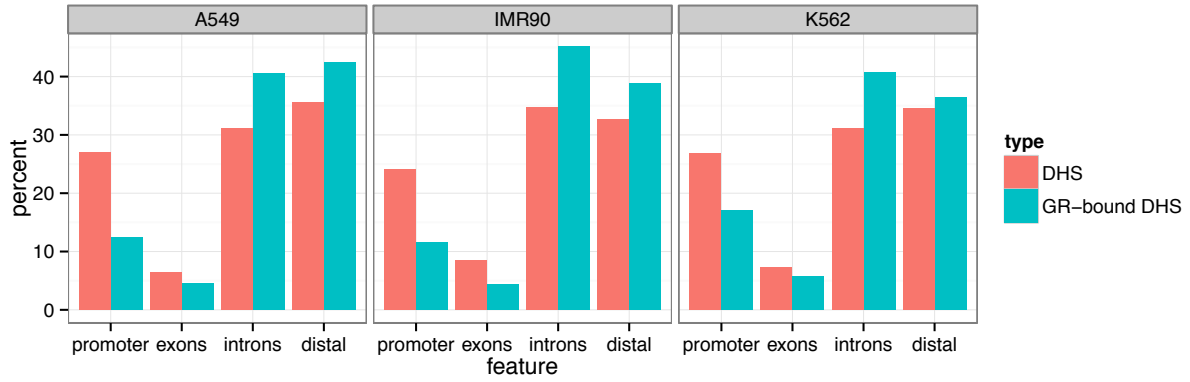

Figure 19: Proportion of GR binding (based on my ChIP-Seq data) in DNase I hypersensitive sites (DHS) sorted by the genomic location: promoter proximal (± 2.5 kb from TSS), exonic, intronic and promoter distal regions. All three analyzed cell lines were showing a relative depletion of GR binding in promoter proximal regions. GR tends to bind to exonic, intronic and promoter distal regions instead. Analysis done in collaboration with by Dr. Michael I. Love and cited from his PhD-thesis: "Statistical analysis of high-throughput sequencing count data".

We first mapped DHSs to the aforementioned genomic regions (Figure 19, red bars) and found, for example, that about 25 % of DHSs map to promoter regions for all three cell lines examined. Interestingly, when we compared this to the percentage of GR-bound DHSs that map to promoters (Figure 19, green bars) this number was much lower. This depletion is more pronounced in A549 (55 % less promoter binding) and IMR90 (~52 % less promoter binding), than in K562 cells (37 % less

promoter binding). Thus the negative correlation of H3K4me3 to GR binding in A549 and IMR90 cells (Figure 18) could be a cause of depleted GR binding to promoter regions.

For all three cell lines (A549, IMR90 and K562; K562 and IMR90 data was generated by me), we found a relative depletion of GR binding to "open" promoter regions when compared to the distribution seen for DHSs in general. One straightforward explanation for this could be that the sequence composition at promoters is different and specifically that fewer sequences matching the GR binding motif are present in this region. This hypothesis was tested, by analyzing the sequence composition of promoter-proximal and promoter-distal regions in DHS. In addition, we compared the distribution for promoter-proximal and promoter-distal regions that were actually bound by GR (Figure 20).

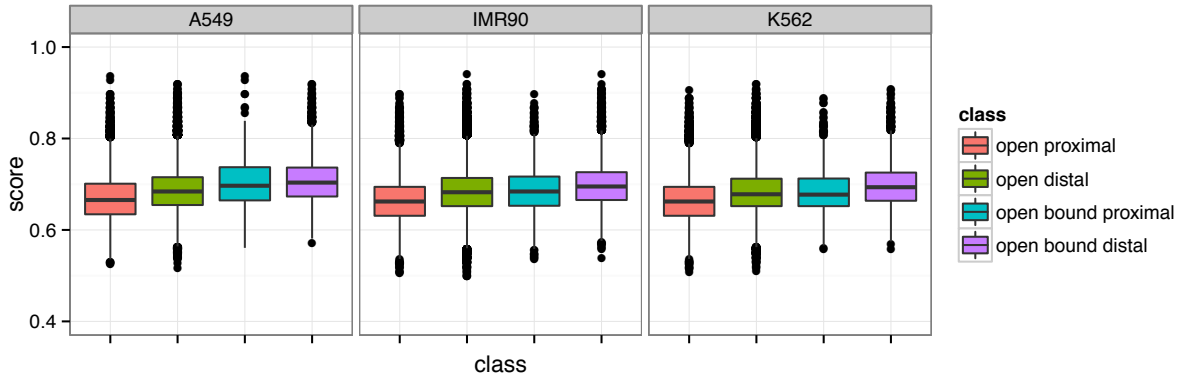

Figure 20: Distribution of motif scores for various groups: DNase I hypersensitive sites (DHS) which are proximal or distal to promoters, and DHSs that are bound by GR and proximal or distal to promoters. Bound DHS have a slightly higher score distributions compared to the universe of all DHS, and distal DHS have slightly higher score distributions than promoter-proximal DHS. Distributions of "closed" and "open" binding sites are also compared (Figure 17). Analysis done in collaboration with Dr. Michael I. Love and cited from his PhD-thesis: "Statistical analysis of high-throughput sequencing count data". Black line indicates median average surrounded by the first and third quartile (boxed areas).

We indeed found a slight decrease in the motif score distribution (higher scores indicate a closer match to the canonical GR motif) for "open" promoter-proximal regions (Figure 20, red boxes) compared to "open" distal ones (green boxes) in all three examined cell lines. Also, the GR-bound promoter-proximal regions (Figure 20, cyan boxes) were showing slightly decreased motif scores compared to promoter-distal regions bound by GR (Figure 20, violet boxes) in "open" chromatin. The lower motifs scores found at "open" and GR-bound proximal positions partly explains the reduced GR binding in promoter proximal regions, as GR is binding more likely to regions containing its canonical binding motif. This is also consistent with the observation, that GR-bound regions have slightly increased GR motif scores,

compared to GR unbound “open“ chromatin in both distal and promoter proximal regions consistent.

Due to the close distributions of motif scores for GR-bound and unbound proximal and distal promoters we cannot answer the question if the depleted presence of GR in promoter proximal regions can be fully explained by the lack of canonical GR sequences. This justified our more sophisticated approach that Dr. Michael I. Love performed, to test whether the depletion in promoter proximal GR binding in “open“ chromatin can be fully explained by sequence composition or if additional mechanisms might be involved. He therefore controlled for the motif scores, within GR peaks in DHS, by assigning a variable to each DHS, whether or not it overlaps with a GR peak. Second he modeled this binary variable by logistic regression, by comparing two models: The first model considers solely the motif score explaining the depleted GR binding. The second model involves the motif score plus the information about whether the DHS was proximal to a promoter. The result is a percentage where both models were compared for their deviance, meaning to what percent is model one in accordance with model two. If promoter proximity explained nothing beyond the motif score alone, the percentage would be 0, whereas 100 % mean promoter proximity and motif score are equally contributing to explain the depleted GR binding in promoter proximal regions. Values between 0 % and 100 % indicate a contribution of the promoter proximity, but the motif score is still more important. For K562 cells we found the percentage of deviance was 36 %, whereas A549 (45 %) and IMR90 cells (52 %) had higher values.

For the three analyzed cell types, Dr. Michael I. Love found indications that mechanisms other than the depletion of GBS-like motifs are needed to fully explain the observed promoter-proximal depletion of GR binding.

7.3 What defines a GR binding site? - Sequences responsible for GR recruitment to individual genomic loci

7.3.1 Comparing GR binding in different cell lines

To study the role of the classical GR binding sequence (GBS) in guiding GR to the genome we analyzed GR-ChIP-Seq data from different cell lines to determine the frequency of GR binding sites (peaks in ChIP-Seq) with and without the canonical motif (done in collaboration with Jonas Ibn-Salem and Dr. Morgane Thomas-Chollier). For GR binding in IMR90 and K562 cells, we used my own ChIP-Sequencing (ChIP-Seq) data, whereas data for GR binding in U2OS cells and A549 cells came from previous studies ((Thomas-Chollier et al. 2013), unpublished results Yamamoto lab). We found that of all tested cell lines, U2OS cells have a higher frequency of GR binding sites containing the canonical motif (Figure 21, p-value cutoff for motif score < 10^{-3}).

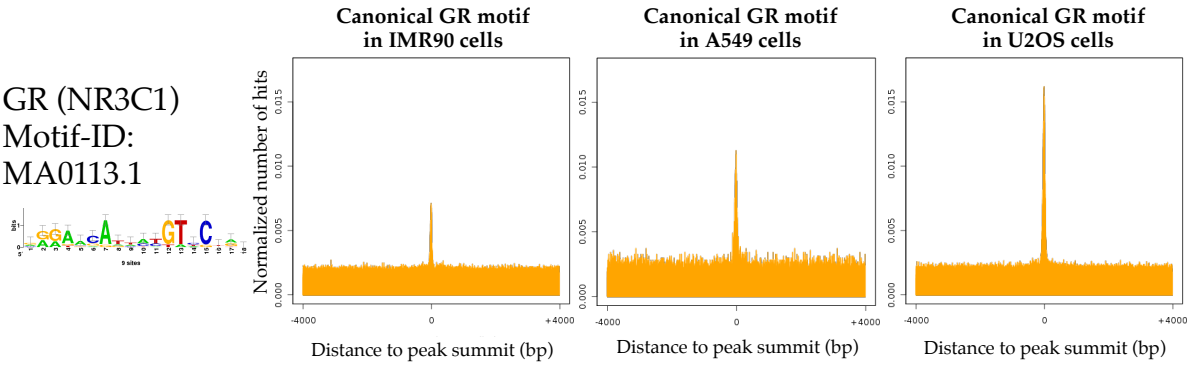

Figure 21: Motif distribution at GR-bound loci. Canonical GR motif was differently enriched for different cell lines. DNA sequences +/- 4000 base pairs (bp) centered on the sequenced GR-ChIP peak summits were aligned and sub-divided into 50 bp bins. Alignment scores for matching motifs of the PWM of GR (from TRANSFAC, at a P-value of 10^{-3} used for cut-off) in every bin were assigned to calculate the relative frequency distribution of GR (MA0113.1, JASPAR, (Sandelin et al. 2004)) motif around GR-ChIP-Seq peaks. The orange plot shows the normalized number of hits for the GR-PWM for each 50 bp bin. A549 GR-ChIP-Seq was performed by the Yamamoto lab. Analysis was done in collaboration with by Dr. Morgane Thomas-Chollier.

To exclude the possibility, that the reduced usage of the canonical motif at GR binding sites in IMR90 cells depends on the assigned p-value, the binding sites usage in three cell lines was compared at multiple p-value cut-offs (Figure 22). 85 % of GR-bound regions in U2OS cells contain a GR binding sequence (GBS) at a low stringency cut-off (P<10^{-3}), compared to 54-55 % of peaks in K562 and IMR90 cells. K562 and IMR90 cells have therefore only slightly more canonical motifs in their peaks than control regions, which are not bound by GR (47 %).

At higher stringency (P < 10^{-4}), the enrichment of GBS sequences for bound regions is more obvious while at the same time a much smaller fraction of peaks

contains a GBS (46 % U2OS; 20 % IMR90; 13 % K562; 7 % for unbound control regions). Notably, the absence of apparent GBSs for a substantial fraction of genomic binding sites occupied by GR has also been reported by others (John et al. 2011; Siersbæk et al. 2011). We found the reduced binding of GR to its canonical motif was independent to the given p-value and therefore considered as a global effect in IMR90 cells and K562 cells (Figure 22).

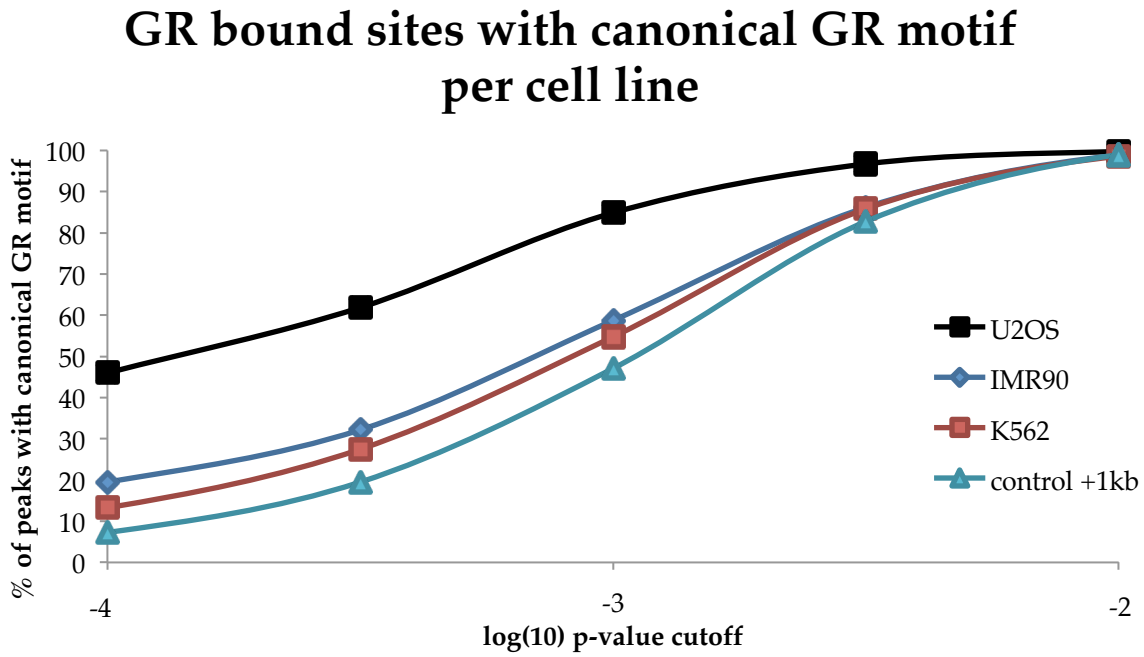

Figure 22: Frequency of GR-bound sites (ChIP-Seq peaks) containing the canonical motif in different cell lines. A higher fraction of GR-ChIP-Seq peaks in U2OS cells contains a canonical motif in U2OS cell when compared to IMR90 and K562 cells. This trend is independent of the p-value cut-off. Analysis was done in collaboration with Jonas Ibn-Salem.

The apparent absence of the conventional GR recognition sequence in IMR90 cells raises the question: What sequences are responsible for the recruitment to individual loci? One potential explanation for this low fraction of peaks comprising a high-stringency motif match is that GR binds to highly degenerate sequences. Alternatively, other sequences present at GR-bound regions may recruit GR directly, or indirectly to the DNA. We find several of those candidate sequences enriched at GR-bound sites in IMR90, but not in U2OS cells (exemplarily shown in Figure 23). Since ChIP-Sequencing (ChIP-Seq) lacks the resolution to discriminate between these two mentioned scenarios, we turned into ChIP-Exonuclease-Sequencing (ChIP-Exo) (Rhee & Pugh 2011) to study the genomic interactions of GR at higher resolution with the aim to identify the sequences responsible for recruitment at individual loci.

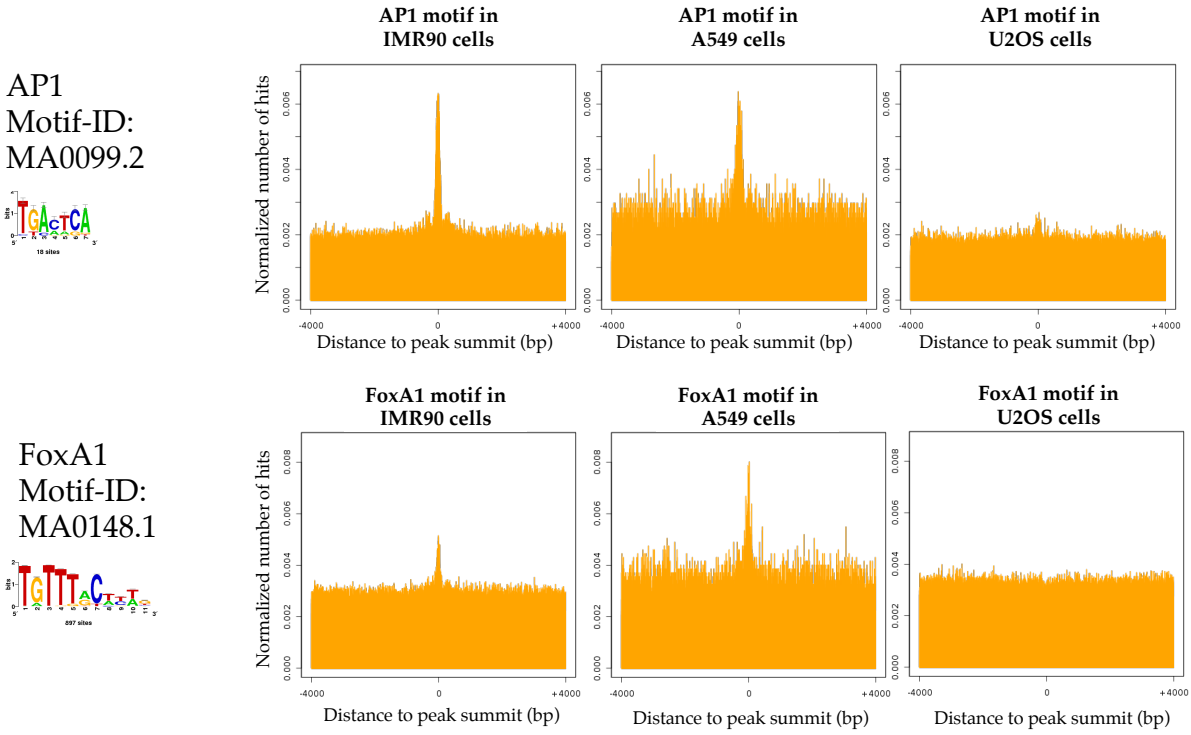

Figure 23: Motif distribution at GR-bound loci for non-GR motifs. AP1 and FoxA1 motifs were found to be enriched at GR peaks for different cell lines. DNA sequences +/- 4000 base pairs (bp) centered on the GR-ChIP-Seq peak summits were aligned and sub-divided into 50 bp bins. Alignment scores for matching motifs of the PWM of AP1 or FoxA1 (from TRANSFAC, at a P-value of 10-3 used for cut-off) in every bin were assigned to calculate the relative frequency distribution of AP1 and FoxA1 (JASPAR, (Sandelin et al. 2004)) motif around GR-ChIP-Seq peaks. The orange plot shows the normalized number of hits for the GR-PWM relative to the peak center for each bin. A549 GR-ChIP-Seq was performed by the Yamamoto lab. Analysis was done in collaboration with Dr. Morgane Thomas-Chollier.

7.3.2 GR-DNA interactions in detail: ChIP-Exo

After preparing IMR90 and K562 cells by cross-linking and sonication, I handed the chromatin over to PECONIC for further ChIP-Exo processing using our GR antibody. We got back the data, including the GR binding regions as reads, which were mapped to the genome and further processed by Dr. Morgane Thomas-Chollier. A comparison of ChIP-Seq and ChIP-Exo signal at a GR-bound region illustrates the increased resolution (Figure 24) of ChIP-Exo performed in IMR90, K562 and U2OS cells. U2OS cells were prepared for ChIP-Exo by Dr. Marcel Jurk.

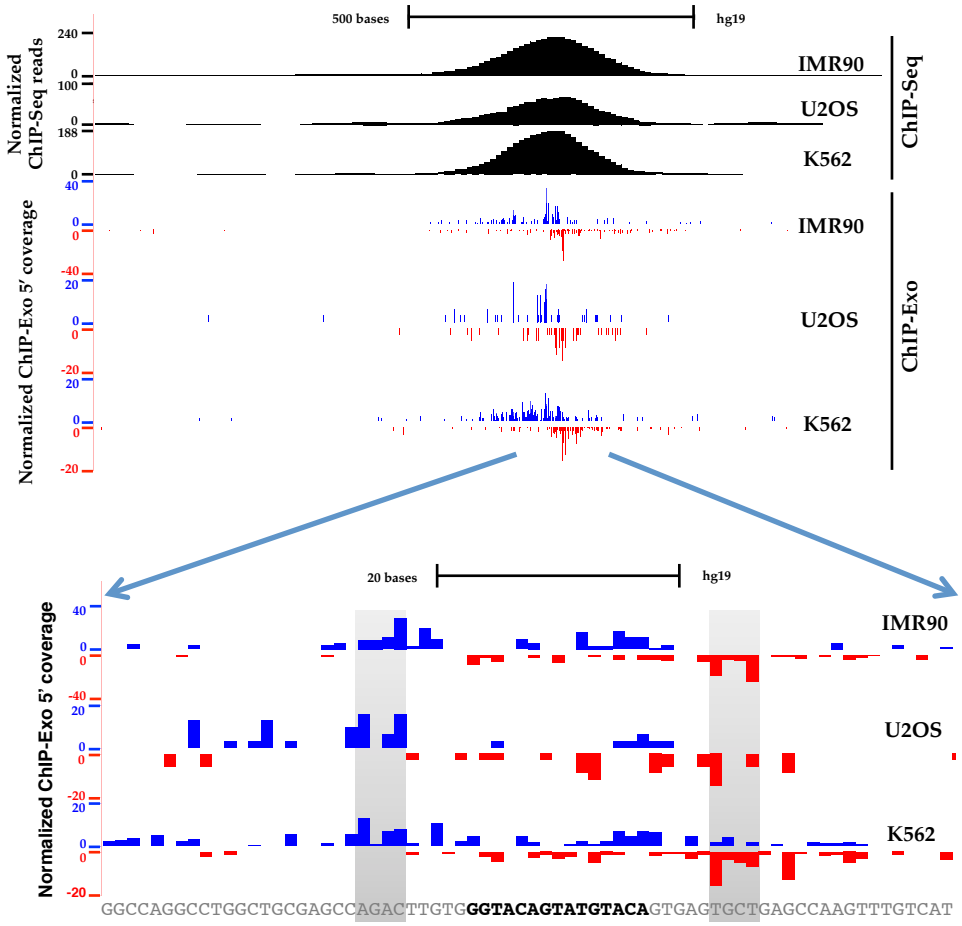

Figure 24: UCSC genome browser screenshot at the GR-bound region at the *ZBTB16* locus (chr11:114,030,896-114,030,965 in hg 19 assembly) exemplifying increased resolution of ChIP-Exo (bottom lanes) compared to ChIP-Seq (top lanes). The coverage at a sequence resembling the canonical GR motif is comparable in IMR90, U2OS and K562 cells. U2OS cells were prepared for ChIP-Exo by Dr. Marcel Jurk. Illustration also used in our submitted publication Starick et al. 2015.

We (Dr. Morgane Thomas-Chollier) compared GR-bound regions identified with ChIP and ChIP-Exo. Analysis of ChIP-Seq data to identify GR binding regions in IMR90 (primary fetal lung fibroblast) and K562 cells (erythromyeloblastoid leukemia cell line) resulted in the identification of 47.630 bound loci in IMR90 and 6.329 in K562 cells (Table 12), whereas ChIP-Exo identified 38.706 GR-bound loci in IMR90 cells and 27.631 sites in K562 cells respectively (Table 13).

Table 13: ChIP-Exo data for GR

Cell-line	Total reads	Mapped reads (unique)	Peaks from Peconic	Peak size [bp], mean (SD)
IMR90	73099097	42787006 (58.5 %)	38715	35.37 (± 21.37)
K562	152287529	137220232 (90.1 %)	27631	40.57 (± 20.88)

A direct comparison of GR-ChIP-Seq and ChIP-Exo peaks showed decrease in the average peak size from ChIP-Seq (IMR90: 387.8 ± 165.2; K562: 423.1 ± 192.3 bp) to approx. 10 % in ChIP-Exo peaks (IMR90: 35.37 ± 21.37 bp; K562: 40.57 ± 20.88 bp), suggesting that the intention of increased resolution for ChIP experiment was achieved with ChIP-Exo. This increase in resolution was used to determine the exact binding sites of the immunoprecipitated protein by the development of a computational method called ExoProfiler (Jonas Ibn-Salem and Dr. Morgane Thomas-Chollier) to further study protein binding at specific DNA motifs.

7.3.3 ExoProfiler pipeline and GR binding in IMR90 cells

GR binding to canonical and non-canonical motifs could lead to different ChIP-Exo signals, since any additional protein, would also be cross-linked like GR himself. These cross-linked proteins could shift the 5′ starts of the NGS-reads, as they function as an additional protection for the DNA degradation by the Exonuclease. Since we found non-GR sequences to be enriched even in our ChIP-Seq data (compare Figure 23, page 79), we set our to analyze and identify sequences possibly involved in GR recruitment to the DNA as well as the canonical GR binding, by developing (Jonas Ibn-Salem and Dr. Morgane Thomas Chollier) a computational pipeline named ExoProfiler, which analyzes the ChIP-Exo signal around sites matching a given motif (Figure 5, page 28; see Methods section for a complete description). Our objective was to identify sequences responsible for recruiting GR to specific genomic loci, rather than obtaining a comprehensive genome-wide picture of its binding. We therefore focused on ChIP-Exo reads mapping to GR-bound regions which were also identified by ChIP-Seq, as they most likely reflect the protection from real binding events. In a nutshell, our ExoProfiler scans bound regions with motifs of interest first, to identify putative TF-BSs with a high-scoring motif match. Notably, only the most 5′ base of forward and reverse NGS-reads were counted as signal and summed up to generate footprints relative to the centered TF-BS as it marks the boundary of protection from lambda exonuclease digestion provided by cross-linked proteins. The ExoProfiler produces several plots (Figure 5, page 45), including a footprint profile displaying the total sum of counts over all sites (Figure 25). The resulting ChIP-Exo footprint profile is plotted along with a profile obtained with permuted motifs and the ChIP-Exo coverage is compared between these two conditions. We prioritized motifs showing distinct peak pairs complementary to significant coverage enrichment.

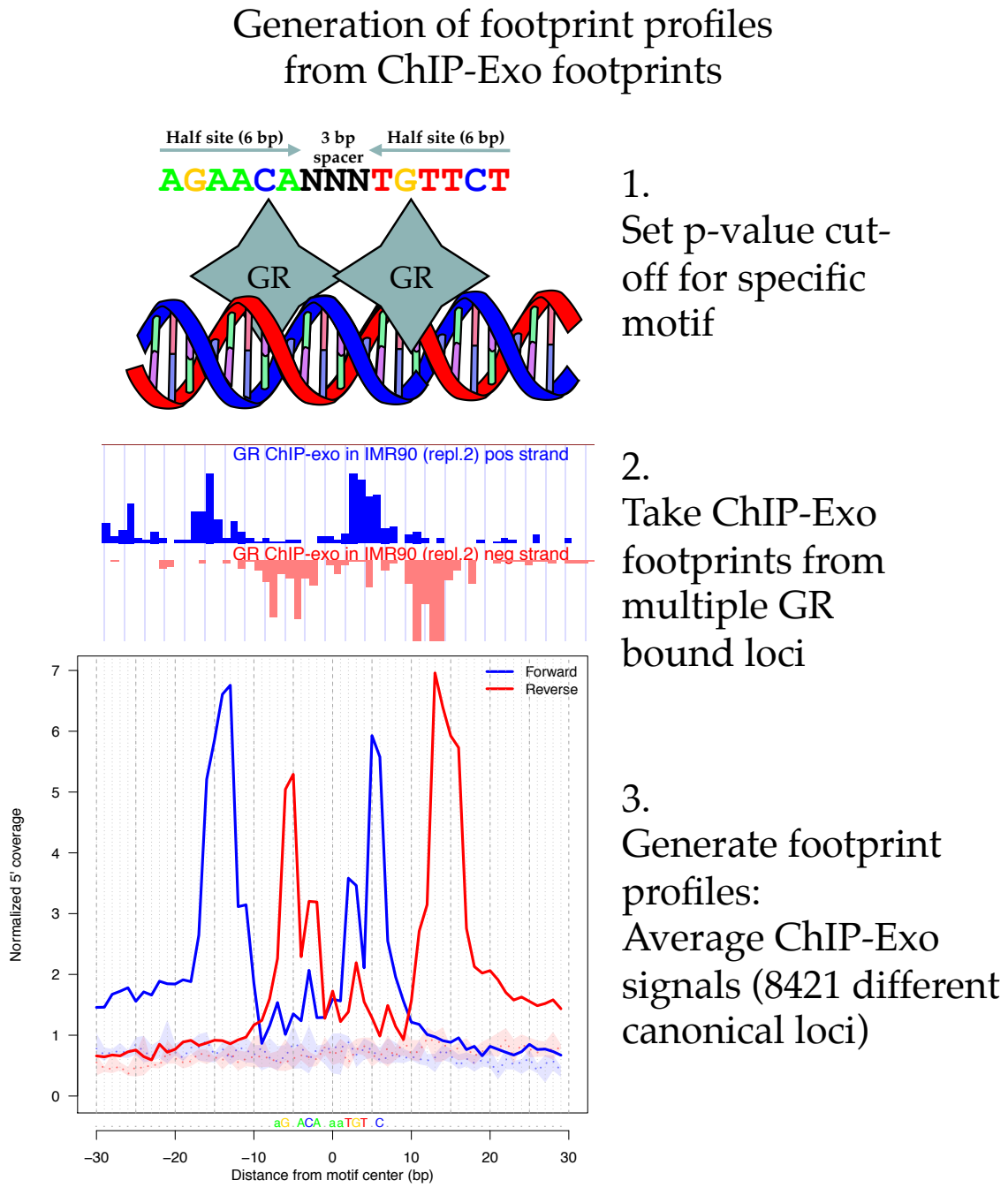

Figure 25: Footprint profiles of GR binding at the canonical motif were build from the average ChIP-Exo signal from 8421 individual footprints of GR-bound loci in IMR90 cells using the ChIP-Exo profiler. The footprint profile is recapitulating the 5′coverage for all short regions after aligning the reads to the motif. As control, this plot displays the 5′ coverage for regions matching permutated motifs. The 10 permutations are summarized by the median (dotted line) and the interquartile range (shaded area). Plot was generated with our ExoProfiler.

7.3.4 Insights into canonical GR binding from ChIP-Exo signals in multiple cell lines

The aforementioned analysis was performed in U2OS, K562 and IMR90 cells, revealing similar footprint profiles s for the canonical GR motif in all tested cell lines (Figure 26). Since we failed to produce a clear footprint profile for the canonical GR motif, in ChIP-Exo data for CTCF of HeLa cells (Rhee & Pugh 2011) (CTCF ChIP-Seq

data (Ohlsson et al. 2010)) our footprint profiles obtained with the ExoProfiler appear to be specific for the immunoprecipitated factor.

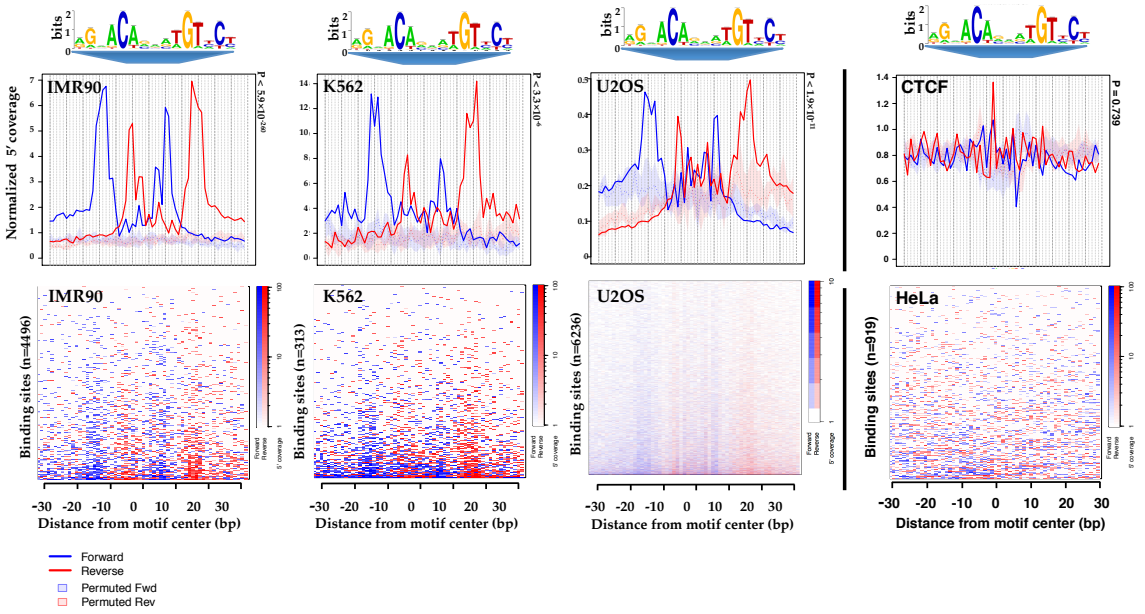

Figure 26: GBS ChIP-Exo footprints (top) in different cell lines matching the canonical GR motif (JASPAR MA0113.2) and ChIP-Exo coverage heatmaps (bottom). Even if the total number of matching sites differs between cell lines, the generated ChIP-Exo footprints are strikingly similar. On the very right side the GBS footprint for CTCF ChIP-Exo data from HeLa cells is shown (Ohlsson et al. 2010; Rhee & Pugh 2011), no distinct footprint was observed, showing that the observed footprint is specific for GR and not a general pattern observed in ChIP-Exo data. The 10 permutations are summarized by the median (dotted line) and the interquartile range (shaded area).

GR protects a region of approximately 30 bp, which is comparable to the footprint obtained by DNase I footprinting (Payvar et al. 1983). For all cell lines, the footprint profiles for GBSs were showing a very similar shape of peaks (Figure 26). The peaks of the footprint profile can be divided into two smaller inner peak pairs (Figure 27, peak pairs 1+2 and 3+4), representing the two GR monomers, and a broader outer peak pair at forward and reverse strand for the GR homodimer (Figure 27, peaks labeled with 1 and 4). This broader peak pair covers the complete 15 bp core homodimeric GBS, whereas the two independently cross-linked GR monomers can explain the inner peak pairs protecting only one GR halfsite. Our footprint is compatible with dimeric binding of GR, with each GR monomer having an independent crosslink point (Figure 27, black crosses), which are both visible due to inefficient formaldehyde crosslinking or could reflect GR dimerizing to the DNA by binding of one monomer prior to the other (Dahlman-Wright et al. 1990). Even tough the ExoProfiler is generating the footprint profiles from multiples single ChIP-Exo signals to further increase the base pare resolution, the inner and outer peak pairs are also visible at single loci (Figure 27, C) and also in the heatmaps (Figure 26, bottom).

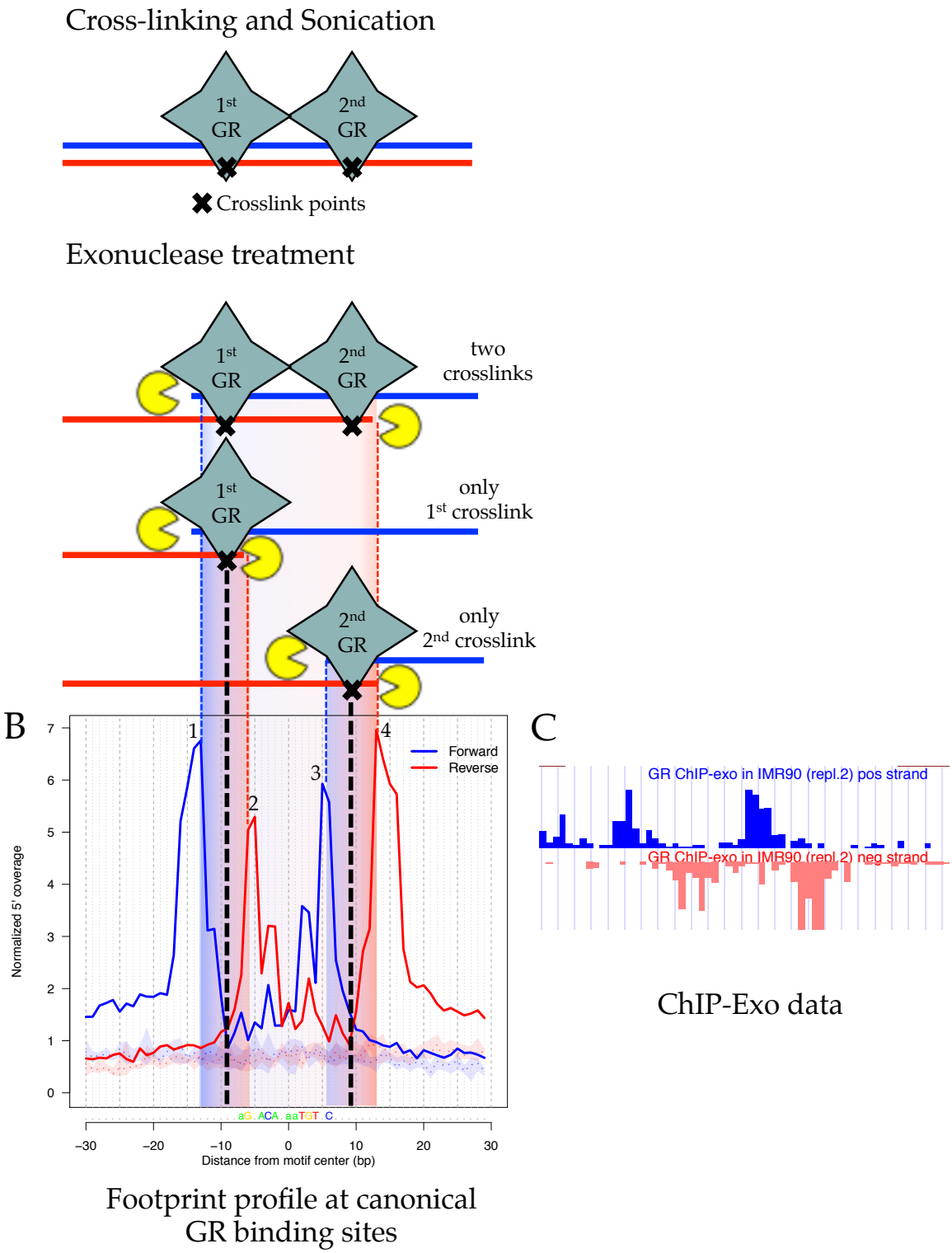

Figure 27: A: Explanation of GR-dimer signal in ChIP-Exo-data. B: Each monomer yields an inner pair of peaks (peaks 1+2 and 3+4) from each DNA-strand, because of inefficient formaldehyde crosslinking. Two cross-linked GR monomers can explain the broader outer peak pair (peaks labeled with 1 and 4). Footprint profile was generated using ExoProfiler. The 10 permutations are summarized by the median (dotted line) and the interquartile range (shaded area). C: zoomed in screenshot from ChIP-Exo data in the UCSC genome browser (chr10: 47391495-47391561; hg 19 assembly), showing two peak pairs of each GR dimer at a single location (compare Figure 24).

Crosslinking is a time-dependent chemical reaction connecting the primary amino groups of proteins with nearby nitrogen atoms of either DNA-backbone or

other proteins with a $–CH_2$-linkage. I used short incubation times (3 minutes). For these conditions, the crosslinking-reaction is far from 100 % efficient at each locus in the millions of cells used per experiment. Each monomer had its own individual signal in ChIP-Exo, because even if only one monomer was cross-linked at a specific locus in one cell, it will be summed up with reads of the second monomer being cross-linked at the same locus in a different cell (Figure 27). Cross-linked GR-dimers lead to signal at the outer positions, therefore explaining the higher read count at the outer borders (Figure 27, peaks 1 and 4), because the signal is summed up with monomeric binding events at these sites.

Taken together, the low cross-linking efficiency at individual loci is reflected in the footprint profiles, which summarizes the protection offered by populations of cells with either one or both of the GR monomers cross-linked to the DNA. In addition, the footprint profiles were hypothesized to be influenced by the DNA:protein cross-link point itself and not by the native folding of the protein as suggested by the initial ChIP-Exo study (Rhee & Pugh 2011). Exonuclease I digestions stops approximately 5-6 bp upstream of the cross-linking point, which protects the 5' ends from being degraded. A closer look at the footprint profile (Figure 27 B) obtained at canonical GR motifs revealed a compatibility to the two main cross-link sites (black crosses) of each GR monomer centered in the two smaller colored boxes between peak pairs 1,2 and 3,4. Earlier examinations of the GR binding domain structure by our group (Meijsing et al. 2009) identified several potential DNA:protein cross-linking sites mapping to the hypothesized cross-linking point, in particular the contacts of R510 and K514 of the C-terminal helix 3 (Figure 28).

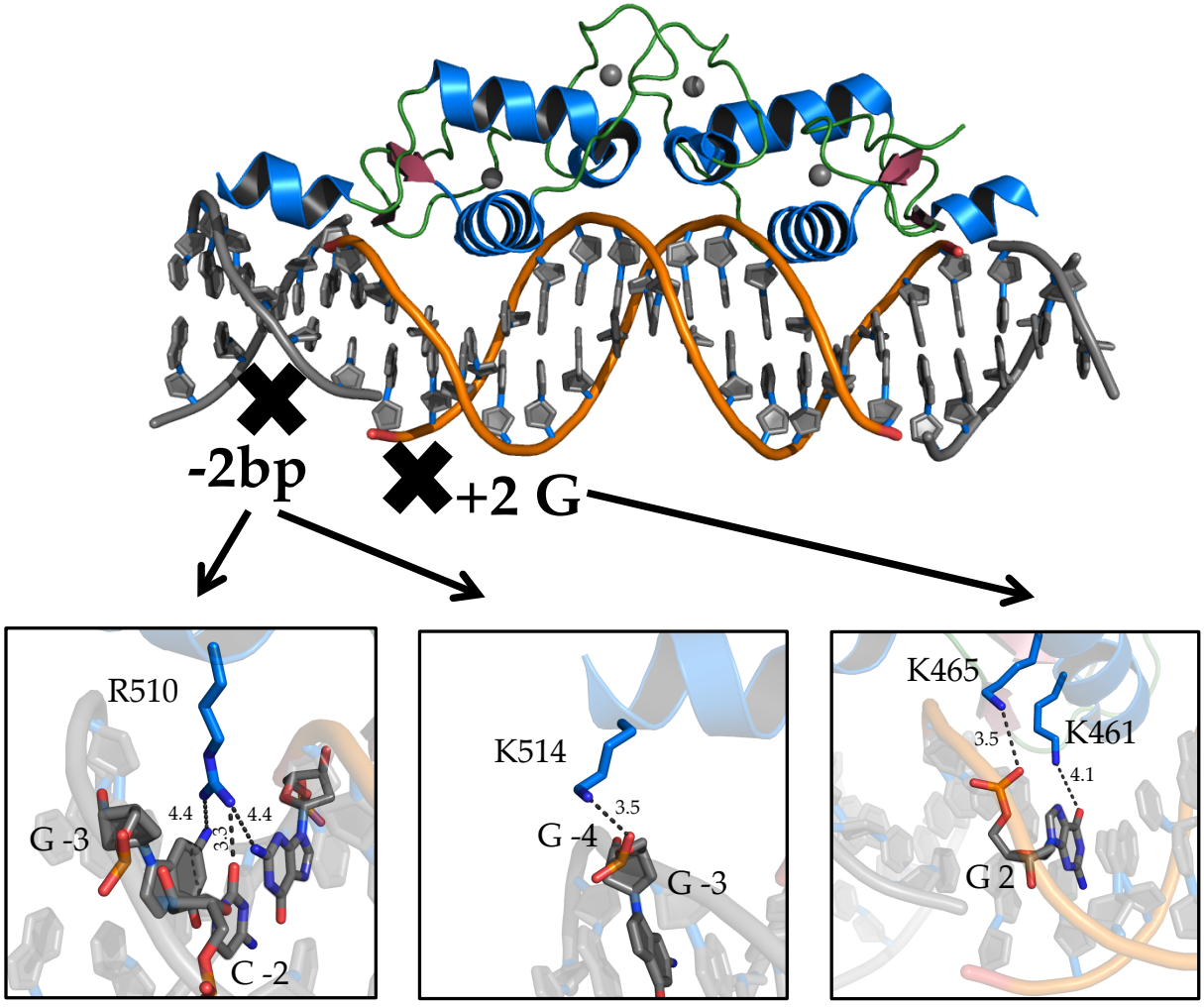

Figure 28: Contacts mapping to the hypothesized GR:DNA cross-linking region based on the crystal structure of the DNA binding domain (DBD) of GR (PDB 3G6U) (Meijsing et al. 2009). Figure produced in collaboration with Dr. Marcel Jurk.

Together, the footprint profile (Figure 27 B) suggested that GR binds as a dimer to genomic GBSs *in vivo*, which is consistent with structural data (Figure 28) and indicates that cross-linking occurs, at least in part, by DNA contacts made by helix 3 outside the 15 bp GBS.

7.3.5 Footprints at degenerate GR binding sites

Since we found the actual amount of canonical motifs at GR binding sites varies among different cell lines (Figure 22, page 78), we set out to test if this could be explained by GR binding to highly degenerate sequences for cell lines with an apparently low fraction of GBS matches. By loosening the motif-scanning threshold for ChIP-Seq peaks, the fraction of motif-matching sequences in both the peak and control regions increases, but shows comparable percentages for bound and control regions and is thus not informative. In contrast, the distinct footprints of binding of

the ChIP-Exo method provided us with the opportunity to test, if degenerate sequences are bound by GR.

As mentioned earlier in this chapter (compare also Figure 22, page 78), fewer GBS-like sequences are present at stringent p-values especially for K562 compared to U2OS and IMR90 cells, suggesting that in K562 cells GR is either recruited by other sequences or that GR preferably associates with more degenerate GBS sequences. To test the latter possibility, we (analysis done by Dr. Morgane Thomas-Chollier and Jones Ibn-Salem) analyzed the motifs binned by p-value and with fixed consensus sequences with varying numbers of constrained bases for all cell lines (Figure 29).

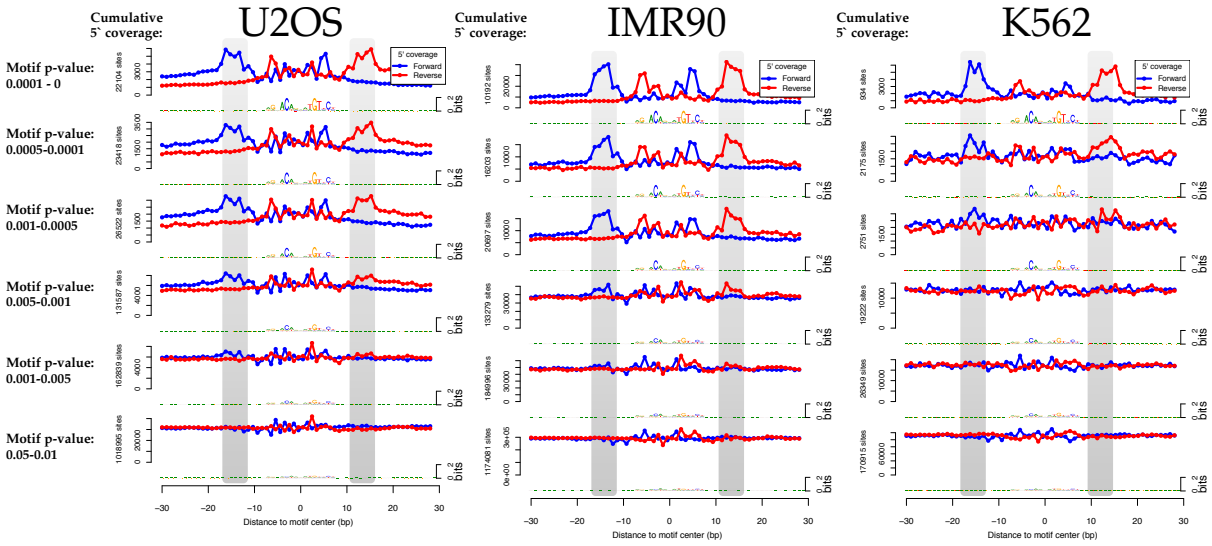

Figure 29: GR-ChIP-Exo footprint profiles for sequences matching the GBS consensus motif (JASPAR MA0113.2) at different motif p-value thresholds in U2OS, IMR90 and K562 cells. Grey bars represent the specific "outer" peak pairs. For p-values 0.0005 – 0.0001 one can still recognize the characteristic double peak pairs representing the dimeric GR binding for all three tested cell lines. For less stringent p-values (0.001 – 0.0005) these shapes were lost in the K562 cell line, indicating other non GBS-like sequences might be responsible for GR recruitment. Underlying plots were generated in collaboration with Jonas Ibn-Salem.

For the first analysis of the GBS footprint profile in IMR90 cells (Figure 27 B, page 84), only regions matching the GR motif at a stringent threshold (p-value ≤ 10^{-4}) were included. Here, we divided the motif matches ChIP-Seq peaks for U2OS, IMR90 and K562 cells into 6 subsets of increasing p-values before applying ExoProfiler. GR motif matches at stringent thresholds (p-value ≤ 0.0001) yielded footprints resembling the GBS footprint in both the shape and position of peaks for all three tested cell lines (Figure 29, first row). At a less stringent cut-off (p-value 0.005 - 0.001) at which most control (83 %) and ChIP-Seq peaks harbor a motif match (IMR90: 86 %; U2OS: 97 %; K562: 86 %; Figure 22, page 78) still yielded a clear GBS-like footprint indicating GR binding to such sequences (Figure 29, fourth row). As indicated by our

motif score analysis, the percentage of bound regions harboring a high scoring GBS motif-match varies among cell types (Figure 22, page 78). This could be recapitulated when we analyzed the motifs by increasing the p-values. Only in U2OS and IMR90 cells, the ExoProfiler generated striking footprints at increased p-values (0.001-0.0005, Figure 29, third row). ExoProfiler failed to produce a comparable footprint in K562 cells at the same p-value (0.001-0.0005) setting. Thus, the smaller fraction of peaks with an apparent GBS for the K562 cell line (Figure 22, page 78) cannot be explained by an inability of GR to bind to GBS-like sequences. Contrary to our expectation and in contrast to U2OS and IMR90 cells, more degenerate GBS sequences failed to produce a footprint profile in K562 cells (Figure 29). Furthermore, when a footprint profile was observed, fewer genomic regions contributed to these profiles (Figure 26, bottom). This analysis indicates that the low percentage of peaks with a high-scoring GBS motif match in K562 cells is not a consequence of GR binding to highly degenerate sequences.

Our binned p-value motif analyses indicates that GR can bind to degenerate GBS-like sequences in IMR90 and U2OS cells, whereas binding in K562 cells is restricted to sequences that closely match the consensus.

7.3.6 ExoProfiler identifies profiles for non-GBS motifs in GR-ChIP-Exo data

An alternative explanation for the absence of GBS sequences at GR-bound regions is that other sequences might be responsible for recruiting GR to the DNA, either directly or indirectly. Therefore, we assayed non-GBS motifs resulting from the *de novo* motif discovery in GR-ChIP-Seq peaks and from the JASPAR database (Sandelin et al. 2004) and could identify several footprint profiles with significant ChIP-Exo coverage and enrichment. Since ChIP was performed with a GR antibody, the found footprints could, for example, reflect the binding of proteins that interact with GR and tether it to the DNA. The additional proteins other than GR can be co-immunoprecipitated, if these proteins were either tethering GR to the DNA (Figure 30 A), bind in composition with GR (Figure 30 B) or represent binding and cross-linking events for proteins near bound GR that do not necessarily have physical connection to GR-DNA interactions. Also GR might bind to non-canonical sequences directly when part of a combinatorial motif, allowing GR to interact as a monomer with another protein (Figure 30 C). These three binding modes of GR would, in principal, lead to different footprint profiles (Figure 30 D) at the individual sequences (Figure 30 E). The interpretation and functional analysis of several individual footprint profiles and their role in recruiting GR to the genome is discussed below.

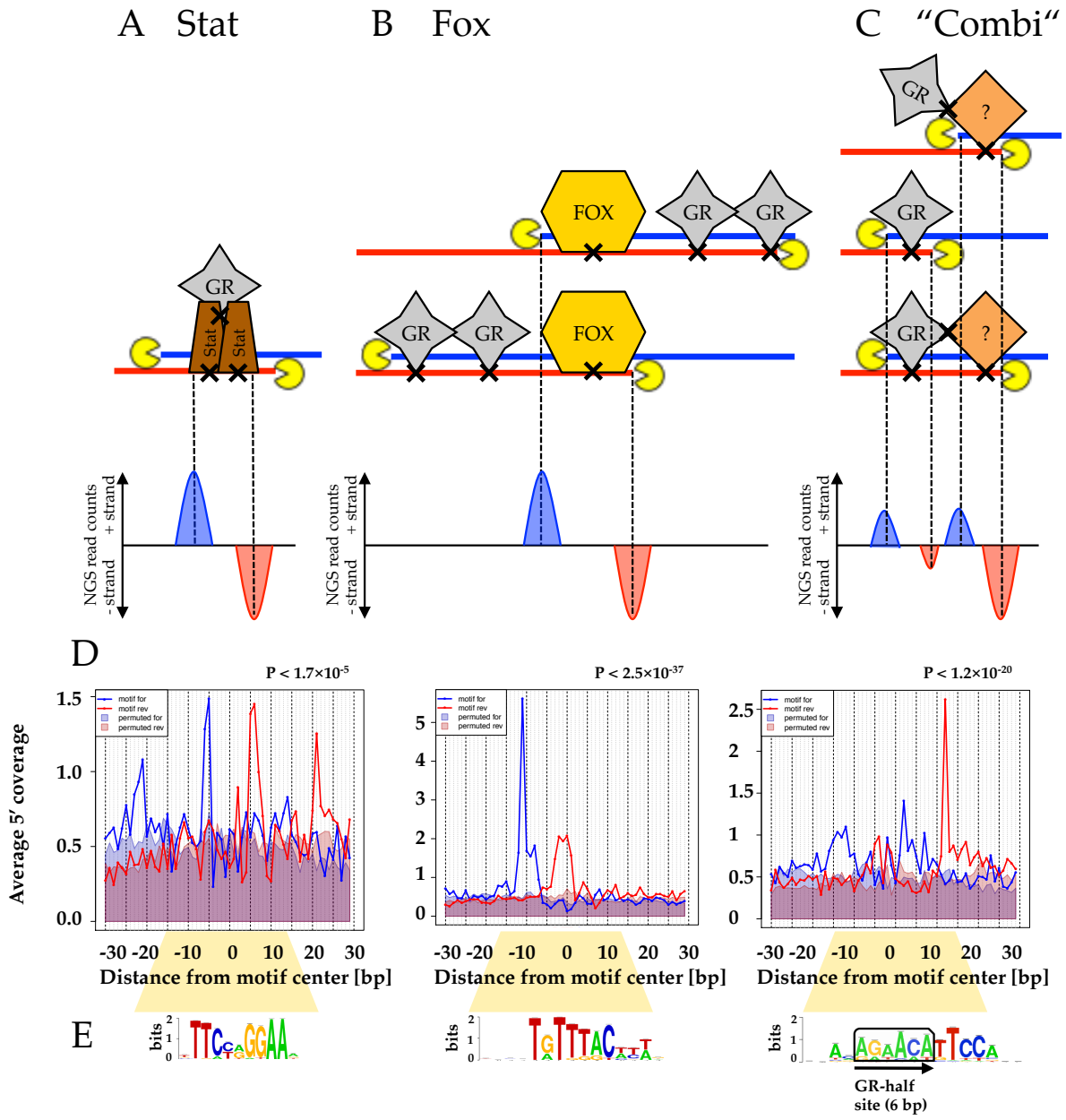

Figure 30: Explanation of footprints at motifs other than the GBS in GR-ChIP-Exo data. X marks the crosslink point of protein-protein or protein-DNA interactions A: GR is tethered to the DNA by another protein (e.g. Stat), B: GR and other proteins are both bound at the same time (e.g. Fox) or C: GR binds in combination with another protein allowing GR to bind non-GR motifs (e.g. GR binding at the *de novo* "combi" motif). D: Resulting footprint profiles (including permutated motifs as background; shaded area) for the analyzed DNA sequences (E).

Among the *de novo* identified non-canonical motifs in IMR90-ChIP-Seq analysis, approx. 10 % of all GR-bound sites showed a combinatorial motif, consisting of a half-glucocorticoid binding motif site and a TTCCA containing part (Figure 31). The combinatorial motif resembles the motif for TEAD/TEF transcription factors (Wasserman & Fickett 1998). Several other studies have also found enrichment of this motif at GR-bound regions (Biddie et al. 2011; Polman et al. 2012) but its role in

recruiting GR to the genome has not been investigated yet. The TEAD motif resembles a GR-halfsite but has a TTCCA part instead of a second GR binding site separated by a 3 bp spacer. The TTCCA part is incompatible with the canonical GR binding motif, since its last two bases (CA) are at the constrained positions of the canonical GR motif (GT) of the canonical motif (Figure 31).

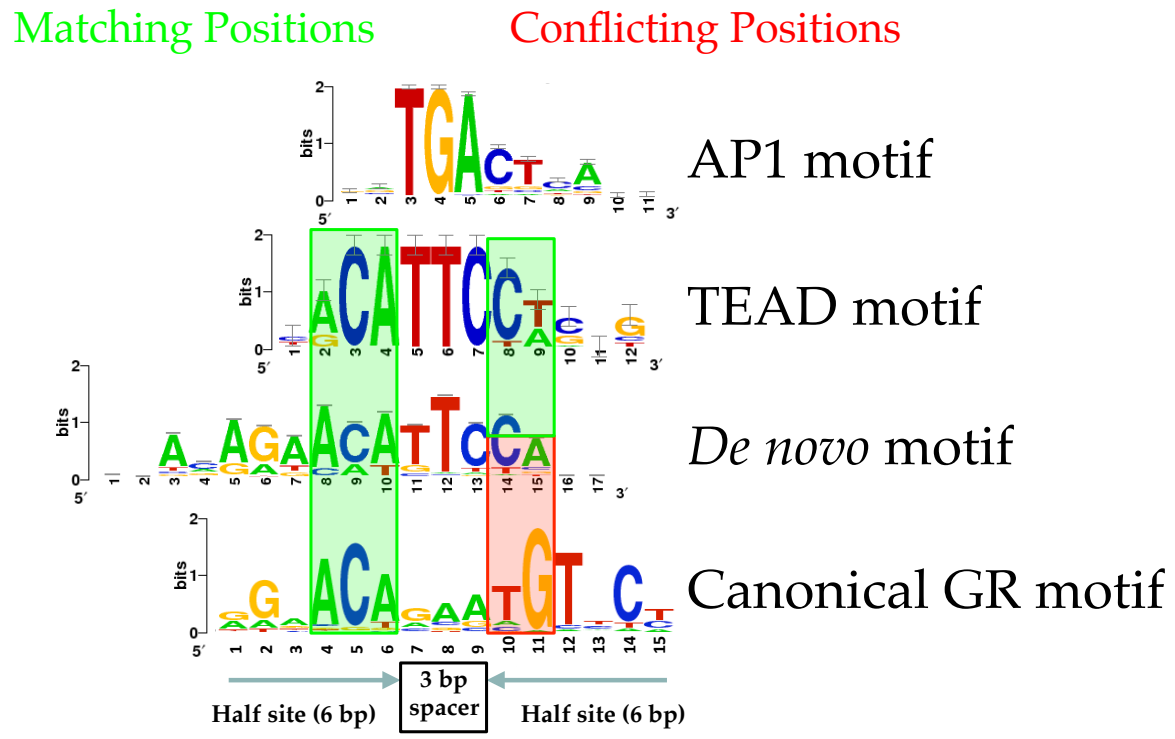

Figure 31: Comparison of the PWM of DNA sequences. Three bases in the first and two bases in the second GR-halfsite are constrained in the canonical GR binding motif (boxes). Whereas the three constrained positions in the first halfsite were matching (green), the two bases in the second halfsite of the canonical motif were not compatible (red box) with the positions in the *de novo* motif. *De novo* motif was found to be enriched in motif analysis (Dr. Morgane Thomas-Chollier) of GR binding sites in IMR90 cells. The PWM of a known GR partner (AP1) also shows no compatibility to the *de novo* motif.

At individual loci the footprints are weak and the profile only becomes obvious when the signal from several bound loci is summarized. Comparing footprints at "combi" and canonical motifs showed a striking overlap of the first peak pairs regarding shape, position and relative intensities, both resemble the first GR monomer. However, ChIP-Exo signals at palindromic canonical motif look like mirror images due to the dimeric binding of GR. The second part with the TTCCA motif however yielded a different footprint at the "combi" motif, not looking like the first part when both footprints were aligned (Figure 32): peak height and width for second peak pair were strongly increased. Together, the ChIP-Exo footprint for the "combi" motif indicated that a different protein complex binds these sequences when

compared to the regular GBS, possibly only monomeric binding of GR-DBD to the "combi" motif together with another partnering protein.

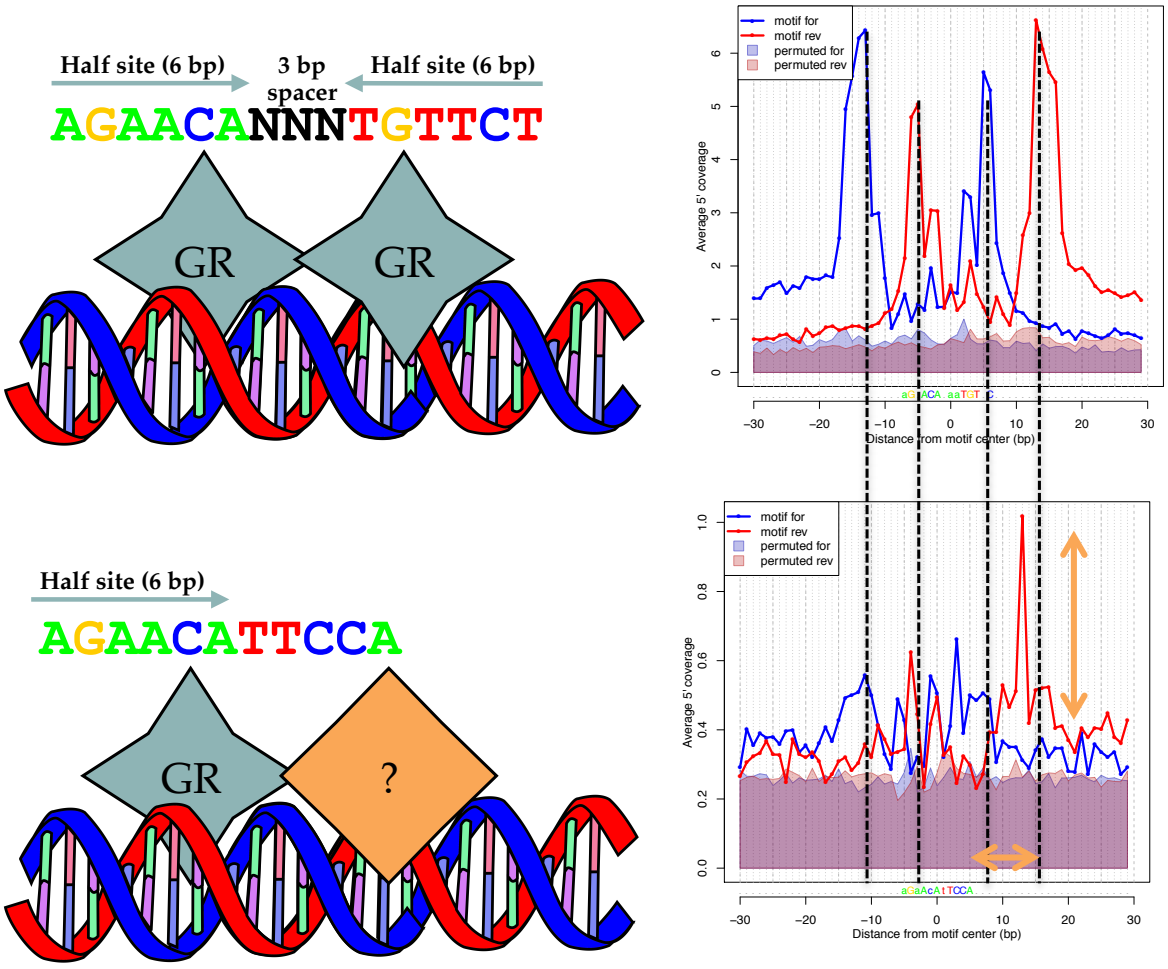

Figure 32: Comparison of footprint profiles for the canonical (top) and the "combi" (bottom) motif based on ChIP-Exo data, revealed a striking similarity for the first half of both motifs (concerning shape, position and relative intensities) resembling the first GR monomer, but different for the second GR monomer. Footprint profiles were generated using ExoProfiler. Shaded area shows signal for permutated sequences.

Since the second GR-halfsite is missing in the *de novo* "combi" motif, I hypothesize a combinatorial GR binding with another factor, other than the previously described and not compatible motif of AP1 (Wu & Bresnick 2007) (Figure 31, top). This GR interaction is clearly distinguishable from dimeric GR binding (Figure 33, top) and tethering, where no GR motif would be found at all. Also composite or overlapping binding of GR and an additional TF (Figure 33, middle) are not applicable, since this interaction would need a full canonical GR site. My hypothesis of a GR monomer partnering with a protein other than GR (Figure 33, bottom), could explain the observed differences in the footprint profiles (Figure 32).

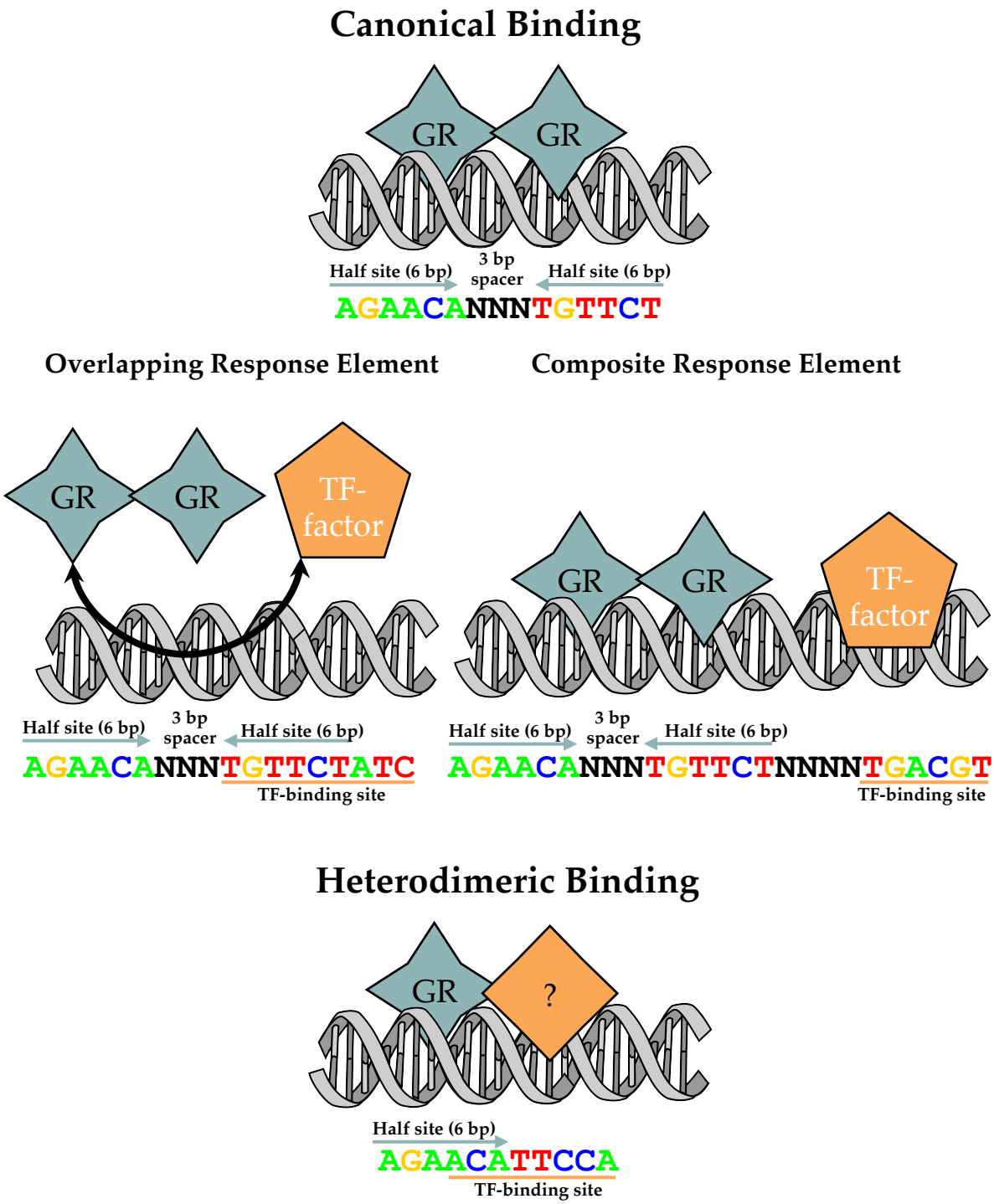

Figure 33: Comparing composite and hypothesized heterodimeric GR binding. Besides dimeric binding at canonical GR motifs (top), GR can bind in composition with other TF (center right), which requires the non-overlapping presence of both full response elements. If the response elements are overlapping (center left) the GR dimer has to compete for binding with the second TF, since monomeric GR binding is less likely even at overlapping binding sites. The enriched sequence found in *de novo* motif search, does not have a full GR binding sequence. GR binding to this sequence (as verified with αGR-ChIP) could be explained by the heterodimerization (bottom) of one GR monomer with another protein, matching to the GR-incompatible TTCCA part.

To study GR interactions at the "combi" sites, I performed several functional analyses. First I studied the binding behavior of the GR-DNA binding domain (GR-DBD) using electrophoretic mobility shift assays (EMSAs).

7.3.7 Functional analysis of the combi motif

To examine whether GR alone is able to bind as a classical dimer, or if GR is binding in another multimeric conformation, I performed EMSAs. Therefore, Cy-5 labeled DNA containing either the *de novo* "combi" motif or a "combi" motif with mutated TTCC, were tested along with a canonical binding site and a randomized sequence as positive and negative controls respectively.

The EMSA assays showed that the DBD of GR binds as a monomer to the "combi" sequence regardless of whether the additional TTCC flanked the GR half-site (Figure 34 A) or not (Figure 34 B). Dimeric interaction of GR-DBD was only detected with its canonical motif (Figure 34 A and B, lanes: pal; Figure 34 C), whereas a random sequence showed nearly no interaction with GR (Figure 34 D). Since GR was not able to interact with the "combi" motif as a dimer or even multimer, this further suggests that GR interacts together with another protein at "combi" sites. This interaction partner needs to interact with the TTCCA part or even more bases of the found "combi" motif (compare Figure 31). The found *de novo* motif is from now on called combi motif to further distinguish it from composite binding (Figure 33, middle)

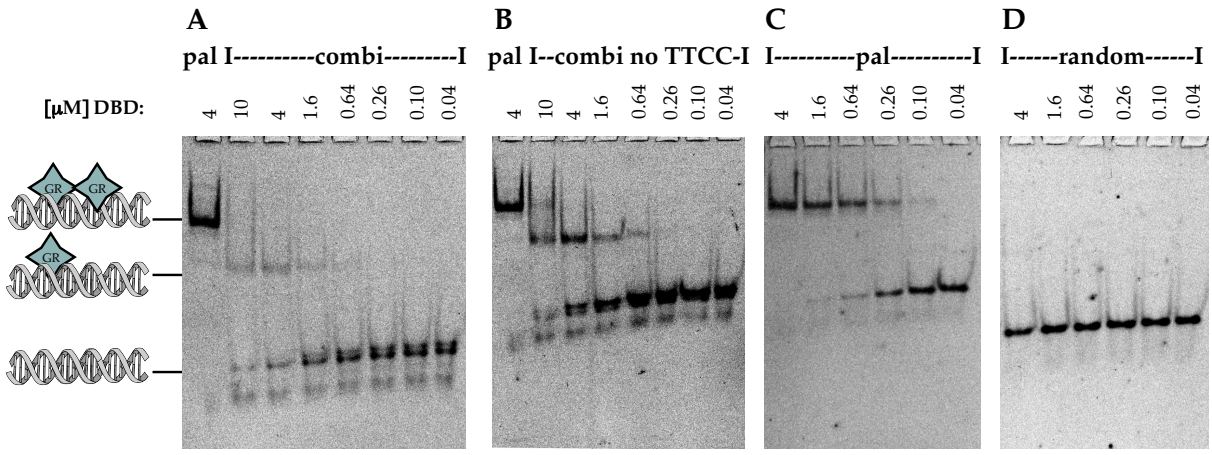

Figure 34: EMSA of Cy5-labeled DNA and different concentrations of GR-DBD (amino acids 380-540 (Meijsing et al. 2009)). GR-bound as a monomer to the "combi" motif (A) or a mutated version lacking the TTCC part (B). Only monomeric binding was detected compared to lanes "pal" containing 4 µM DNA with canonical motif as control. Lanes are labeled with the amount of GR-DBD (4 µM; 1.6 µM; 0.64 µM; 0.256 µM; 0.1 µM and 0 µM) that is titrated to the Cy5-labeled DNA: "combi" motif (A), "combi-noTTCC" motif (B), canonical motif (C) or a random motif (D). Lanes labeled "pal" contain the Cy5-DNA with the canonical GR motif like in panel C.

Next, we studied the transcriptional functionality of the combi motif *in vivo*, by inserting three isolated copies (Figure 35, B) of the combi motif as well as the motif within its genomic context (Figure 35, C) upstream of a minimal promoter driving the expression of a luciferase reporter gene.

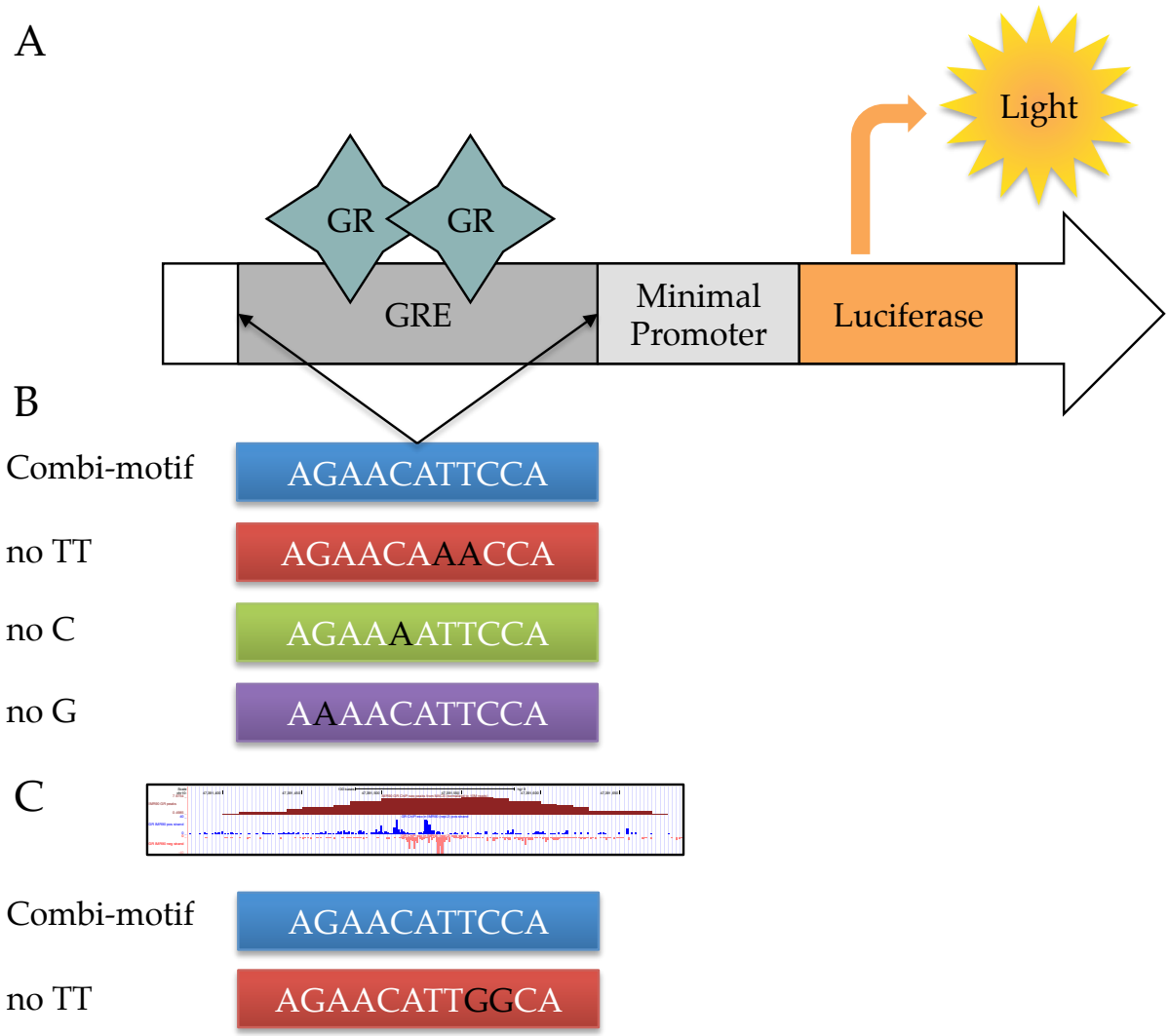

Figure 35: (A) Composition of pGL3-promoter - GR-luciferase reporter constructs. The GR binding region (GBR) containing GR-response element (GRE) activates the downstream SV40 minimal promoter to activate transcription of the Luciferase gene once GR is bound to the GRE. Luminescence of the Luciferase protein is measured using a Luminometer. (B) Three different combi motif-mutants were tested containing mutations either in the TTCC part (no TT), at a position effecting GR and additional factor (no C) or at the GR-halfsite alone (no G). (C) Combi motif and no TT mutant within genomic content.

If GR is able to bind to the glucocorticoid response element (GRE) upon hormone activation, Luciferase expression would be expected to increase, thus allowing the detection of transcriptional activity upon hormone treatment.

Since the combi motif was found in multiple cell lines, motif and mutants were tested in both IMR90 and U2OS cells.

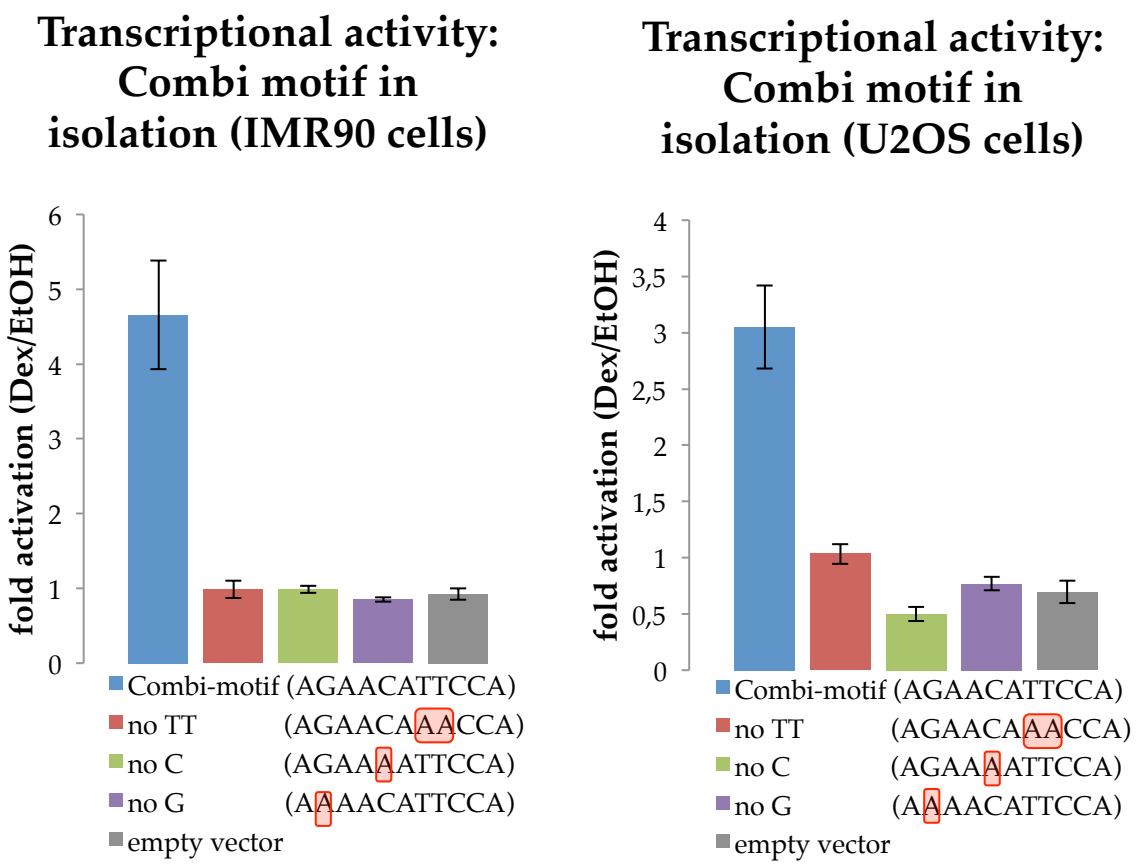

Figure 36: Isolated Combi motif was able to activate transcription in IMR90 (left) as well as U2OS (right) cells using the hormone dependent Luciferase reporter system (Figure 35). Whereas motif mutations were not able to induce transcription, since hormone dependent fold activation is equal to empty vector control. Experiments were done in three biological replicates; Error bars indicate standard error of mean.

Three copies of the combi motif were likewise able to induce transcription upon hormone treatment in IMR90 and U2OS cells. The GR-dependent transcriptional regulation of the reporters was lost if important positions of the GR-halfsite were mutated (no G and no C). Mutation of the unknown GR-interaction partner in the TTCCA-part (no TT) was showing a similar loss of activity, indicating that GR half-site alone was insufficient to drive GR-dependent transcriptional activation of the reporter. Since both cell lines were showing similar results, all further tests were performed in U2OS cells.

To determine the transcriptional regulation of the combi motif in the genomic context, I cloned five genomic loci (400 bp) containing the combi motif activity into pGL-reporter vector. The loci were chosen among all combi sites after scoring for p-value (10^{-3}) and high read-counts in ChIP-Seq data. The influence of the motif itself is determined by SDM of the combi motif, by changing the C part into TGGC (compare

Figure 35 C). Leaving the GR-halfsite intact allows for testing the possible transcription regulation driven by a GR monomer.

Combi-Motif: AGACAATTCC
Combi-Motif-noTT: AGAACATGGC
Canonical motif: AGAACAxxxTGTTCT

The transcriptional response of cells treated with hormone and transfected with the pGL3-promoter - GR-reporter constructs was compared to basal transcriptional activity in vehicle treated cells, since GR is only able to interact with DNA when hormone-bound. All five genomic loci containing the combi motif were less transcriptional active, if the TTCC part of the AGAACATTCC was mutated to TGGC (Figure 37).

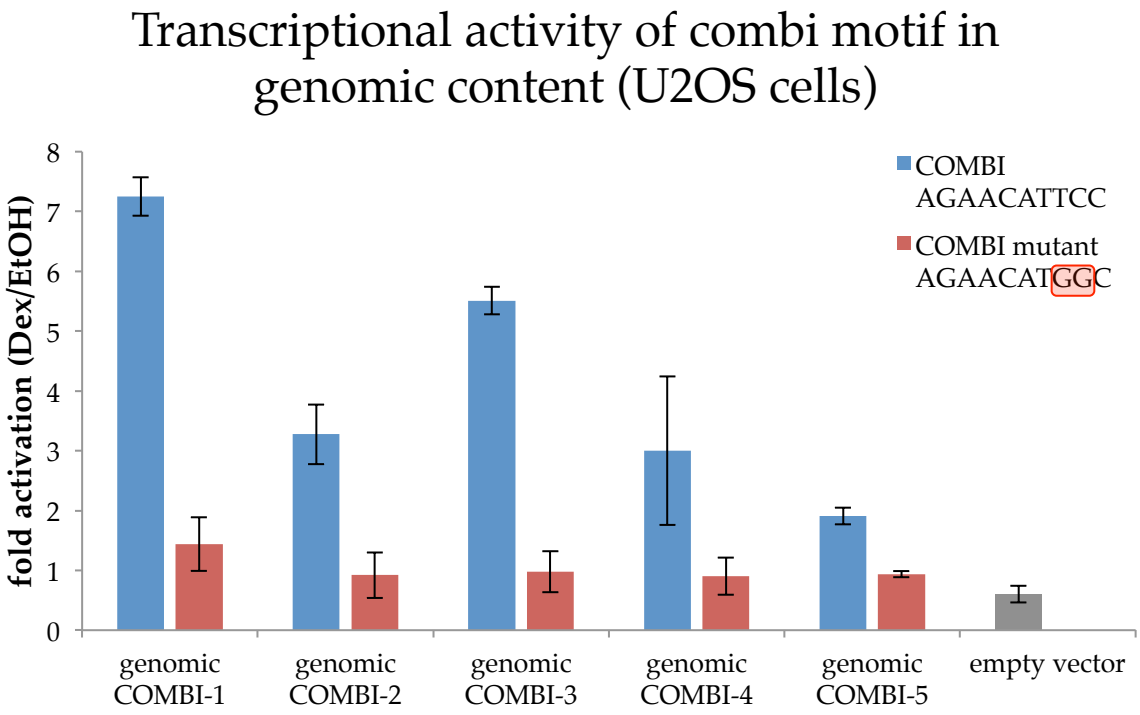

Figure 37: Genomic regions containing the GR-Combi motif were compared to mutants missing the combi motif. Transcriptional activity of genomic regions missing the combi binding site is diminished in four out of five tested regions. Experiment was done in three biological replicates; Error bars indicate standard error of mean. Genomic regions in hg19 assembly: Combi-1 (chr6:108975209+108975621); Combi-2 (chr1:27325909+27326321); Combi-3 (chr1:61647749+61648161); Combi-4 (chr20:49039019+49039431); Combi-5 (chr3:140995189+140995601);

Mutations of the TTCC sequence of the combi motif in isolation (Figure 36) and in genomic content (Figure 37) resulted in a loss of activity, indicating that GR half-sites alone were insufficient to drive GR-dependent transcriptional activation. This indicates a GR partner, which is binding in combination as a possible heterodimer.

We therefore searched for the TTCC motif in the JASPAR database (Sandelin et al. 2004) to identify candidate TFs that could act as interaction partners of GR by binding to this sequence.

7.3.7.1 Possible GR binding partners for the combi motif

The TTCC part can be found in binding motifs of several other proteins, which are therefore candidates for the hypothesized GR-heterodimer. Members of the ETS-family, a family of transcription factors defined by the shared DNA binding domain (G. M. Lee et al. 2005), can bind to the TTCC part of the combinatorial sequence and are known to interact with GR (Verger & Duterque-Coquillaud 2002). Additionally, the TEAD family (Anbanandam et al. 2006; Pobbati & Hong 2013), all sharing the TAE-domain, can also bind to the ACATTCC part. Four different TEAD proteins are known, none of them is identified as a GR cofactor so far. Together the TTCC part is also part of the binding motif of the ELK, ETS1, ETS2, RelA, Stat3, Stat5B and the four TEAD proteins (Sandelin et al. 2004) (Figure 38). The found proteins containing the TTCC motif as consensus sequence were structurally aligned (Dr. Marcel Jurk) with GR protein at the combi site and showed that monomers of all cofactor candidates allow combinatorial binding with a GR monomer (Figure 38).

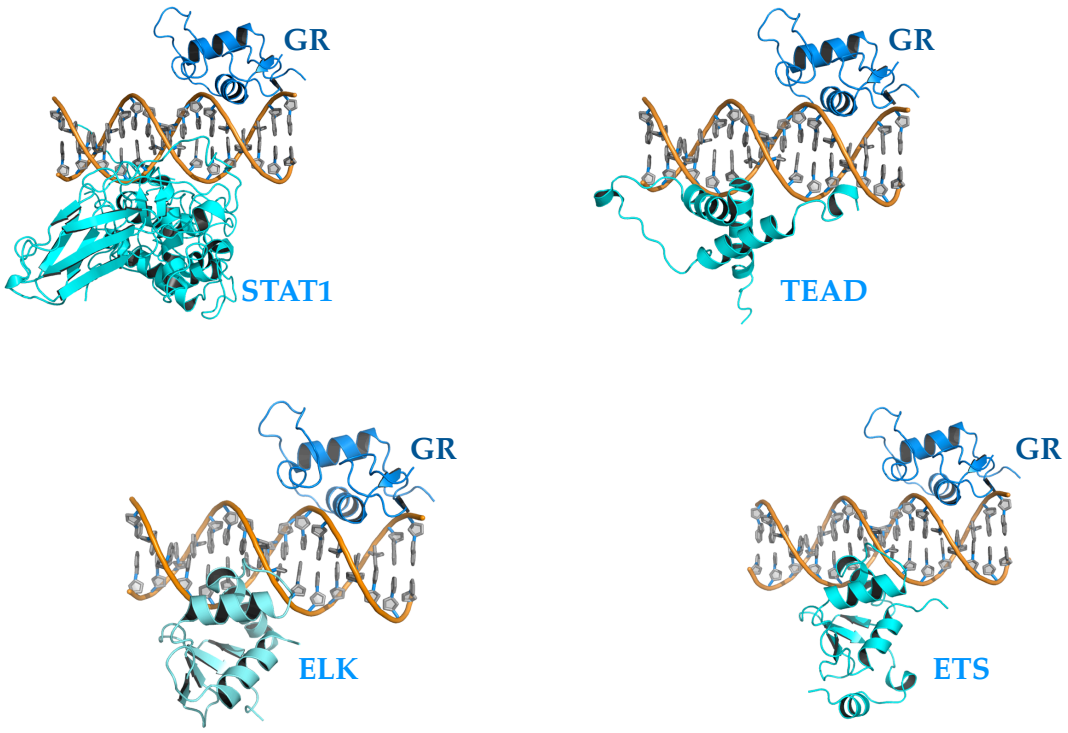

Figure 38: Structural alignment of GR monomer with possible co-factors at the combi motif (Figure 31). According to the structural alignment, all cofactors are compatible to interact at the combi sites without any structural constrains. Structural alignments was done in collaboration with Dr. Marcel Jurk using PyMOL®.

Since the structural alignments indicates no spatial constrictions of GR and all examined possible heterodimerization partners at combi sites, I searched for the possible GR interaction partner using a global approach, named DNA-Pulldown using the combi sequence as bait to identify interacting proteins.

7.3.7.2 DNA affinity chromatography and quantitative proteomics to identify GR interaction partners

The TTCC part of the combinatorial motif found to be enriched in GR-ChIP- and GR-ChIP-Exo-Seq and data of IMR90 cells has similarity to the consensus sequence of several DNA binding proteins according to the JASPAR database. Whole transcription factor families were found to match with the TTCC part of the combinatorial sequence. We therefore used an affinity purification assay called DNA-Pulldown to search globally for possible GR heterodimerization partners (Mittler et al. 2009). Therefore, biotinylated doublestranded DNA oligos were used as baits for proteins of a nuclear extract. Baits were containing either the combi motif, a mutated version of the combi motif (COMBI no TTCC) or a scrambled motif (negative control). The oligo-loaded beads were incubated with nuclear extract (including 1μM dexamethasone) to allow proteins to interact with the DNA. Non-specific interactions of beads, baits and proteins were removed by extensive washing. Protein-DNA complexes were eluted from beads by trypsin digestion, subsequently precipitated and analyzed by orbitrap mass-spectrometry (MS, in collaboration with David Meierhofer, MPI for Molecular Genetics, Berlin). Max-Quant was used to analyze the MS-spectra for the three different samples (1:combinatorial motif 2: combinatorial motif noTTCC and 3: scrambled motif). Proteins were further analyzed only if they were identified by at least 2 or more unique peptides in the label free quantification.

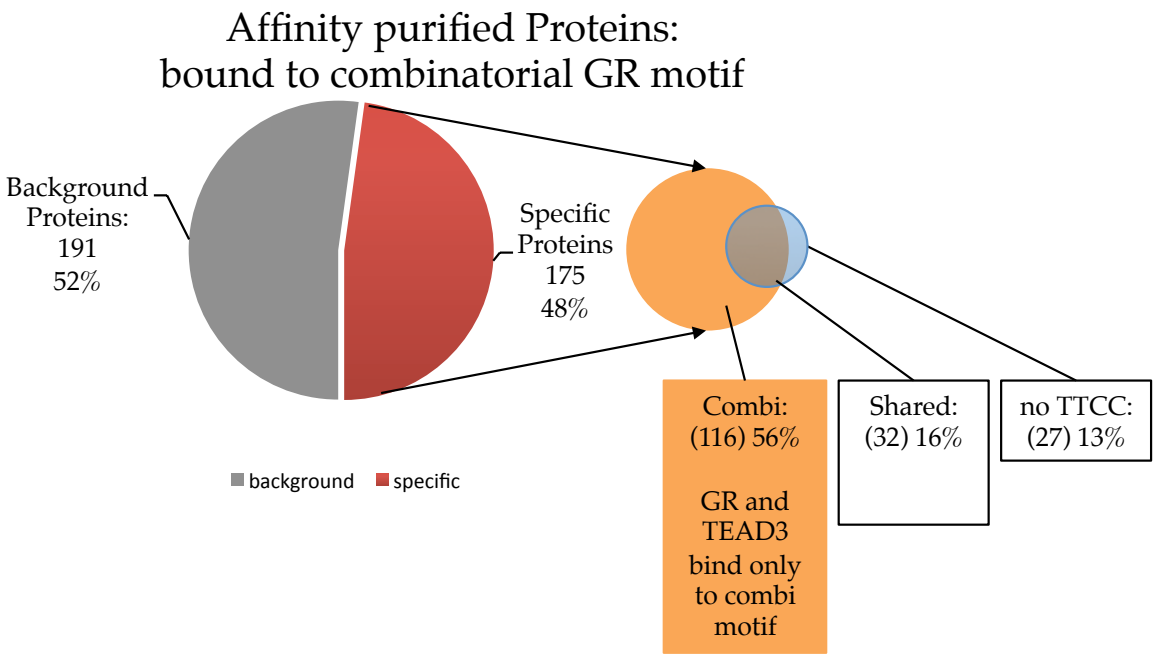

Figure 39: Proteins pulled down from nuclear extract and identified by Mass-Spec. TEAD3 and GR were only bound to COMBI motif and not in the no TTCC motif containing control. 543 proteins were identified in total; roughly half of them were bound on scrambled sequence too and therefore classified as background.

All proteins (191 hits) identified by MS in the scrambled negative control were considered as background binding and therefore removed from further analysis (Figure 39). 116 proteins were found to interact with the combi motif, whereas 27 proteins interacted with the control combi no TTCC motif respectively. 32 proteins were found to interact with both motif variants. GR (NR3C1) and was identified only at combi motif sample (Table 14), indicating a preferred binding if the motif for the candidate heterodimerization partner is present. Among the unique proteins identified by affinity purification for the combi motif, we identified TEAD3 as the only protein from the predefined motif matching proteins, which could be identified using the global combined approach of DNA-Pulldown and Mass-Spec.

Table 14: Top 20 proteins uniquely bound to combi (orange), combi-noTTCC (grey) motif or shared among both (verified by Mass-Spectrometry). TEAD3 as candidate factor for the combi motif was found only in combi bound proteins.

rank	Proteins bound to combi motif	Intensity: combi over combi-noTTCC	Protein bound to both motifs	Intensity combi over noTTCC	Proteins bound to combi-noTTCC-motif	Intensity: combi-noTTCC over combi
1	MSH6	527280001	KIF20B	4,44	EXOSC3	1056400001
2	PURB	405400001	HIST2H3A; HIST3H3;H3F3A; HIST1H3A; H3F3C; HIST2H3PS2	3,10	EHMT2	535510001
3	EIF2S2	403640001	CDC5L	2,34	WTAP	495460001
4	ZC2HC1A	392150001	CAPZA1	1,77	EXOSC4	490170001
5	AHNAK	386860001	C19orf47	1,72	EXOSC10	331100001
6	rGR	342480001	SAFB	1,53	CIRBP	242390001
7	BAZ2A	289920001	SF3B2	1,47	KIAA1429	236350001
8	BCAS2	281950001	DNTTIP1	1,16	CHTOP	162550001
9	TOP2B	255540001	NOP58	1,05	EXOSC6	156100001
10	DDX24	253360001	HSPA5	1,02	THRAP3	130010001
11	HDAC1	233530001	KMT2A	1,02	IGF2BP3; IGF2BP1; IGF2BP2	117220001
12	DDX54	219090001	RFC2	1,02	PLEKHA5	117190001
13	TEAD3	188710001	NOC3L	0,92	KIFC1	90939001
14	RBBP4	186080001	ORC1	0,72	ZC3H13	86662001
15	RFC1	184820001	CORO1C	0,70	MSL1	79154001
16	SMARCA1	180900001	PHF3	0,65	PPP1CB	76017001
17	FEN1	177700001	RPS17L;RPS17	0,60	DRG1	71337001
18	SPTAN1	173740001	DKC1	0,59	RUNX1	63346001
19	RPL32	167240001	SART1	0,57	FLII	63247001
20	CAPZB	166280001	GNL3	0,55	PSRC1	57536001

The other top 15 proteins identified by MS in the DNA pull-down for the combi motif were either chromatin remodelers (BAZ2A; HDAC1; RBBP4; SMARCA1), parts of DNA repair systems (MSH6, TOP2B, DDX24, DDX54, RFC1, FEN1), belong to the translational machinery (EIF2S2, BCAS2). Other proteins belong to cell cycle associated proteins (ANHAK), were part of scleroprotein or related class (SPTAN1, CAPZB) or were of ribosomal origin. The only DNA-binding proteins associated with transcription were NR3C1 itself, PurB and TEAD3. PurB is implicated in transcriptional regulation and DNA replication, thus binding preferentially to purine rich single stranded DNA (Gupta et al. 2003). The consensus sequence of PurB does not match to the one present in the combi motif. Of all proteins identified, TEAD3 is

the only one whose consensus sequence matching to the TTCC part of the combi motif.

To further analyze the hypothesized combinatorial binding of GR and factor X, I set out to knockdown the candidates (ELK, ETS1, ETS2, RelA, Stat3, Stat5B and the four TEAD proteins) independently to investigate a possible impact to the GR-dependent transcriptional regulation of the combi reporter. These proteins were chosen because they showed up in the MS analysis (TEAD), or because they bind to the TTCC part of the combi motif and have been shown to interact with GR either physically or functionally (ELK, ETS1, ETS2, RelA, Stat3, Stat5B).

7.3.7.3 Cofactor analysis

To identify possible cofactors of GR at the combi motif, cofactor candidates were knocked down individually using dsiRNA. Each knockdown was done with two different dsiRNA, targeting another part of the particular mRNA. Only the dsiRNA with the strongest knockdown was used for further tests. Expression levels were normalized to the mRNA-expression-levels of scrambled dsiRNA treated and set to 100 %. Scrambled dsiRNA does not targeting any mRNA. For all mRNAs I obtained knock down efficiencies of at least 50 % or more (Mean: 30.6 ± 2.7 %) (Figure 40 and Figure 41).

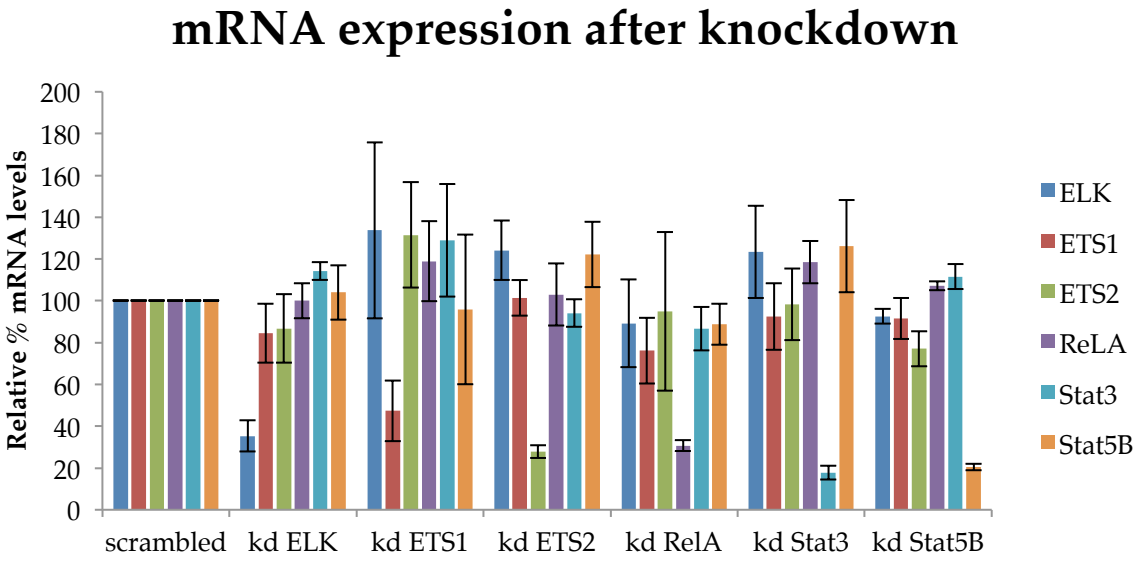

Figure 40: Expression levels of mRNA analyzed by qPCR and compared to control dsiRNA (scrambled). Reduced mRNA levels Stat5B, Stat3, ETS1, ETS2 and ELK correspond to individual knockdown of candidate genes. Experiment was done in three biological replicates; Error bars indicate standard error of mean.

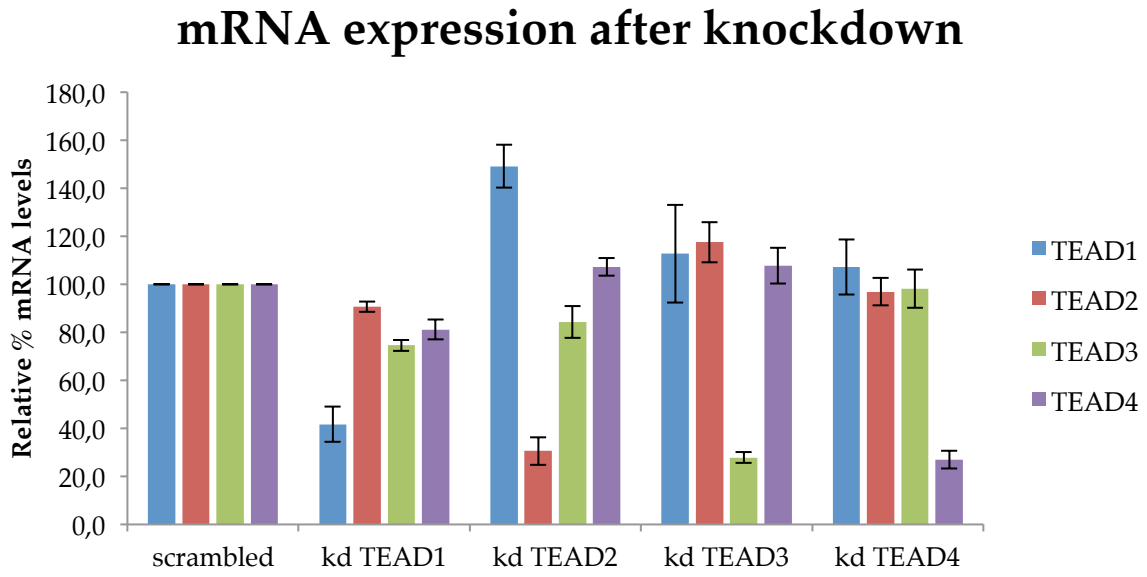

Figure 41: Expression levels of mRNA analyzed by qPCR and compared to control dsiRNA (scrambled). Knockdown of TEAD variants leads to reduced mRNA levels for each TEAD variant individually. Experiments were done in three biological replicates; Error bars indicate standard error of mean.

Next, I tested the effect of knockdown of candidate factors on the GR-dependent activation of the combi luciferase reporter.

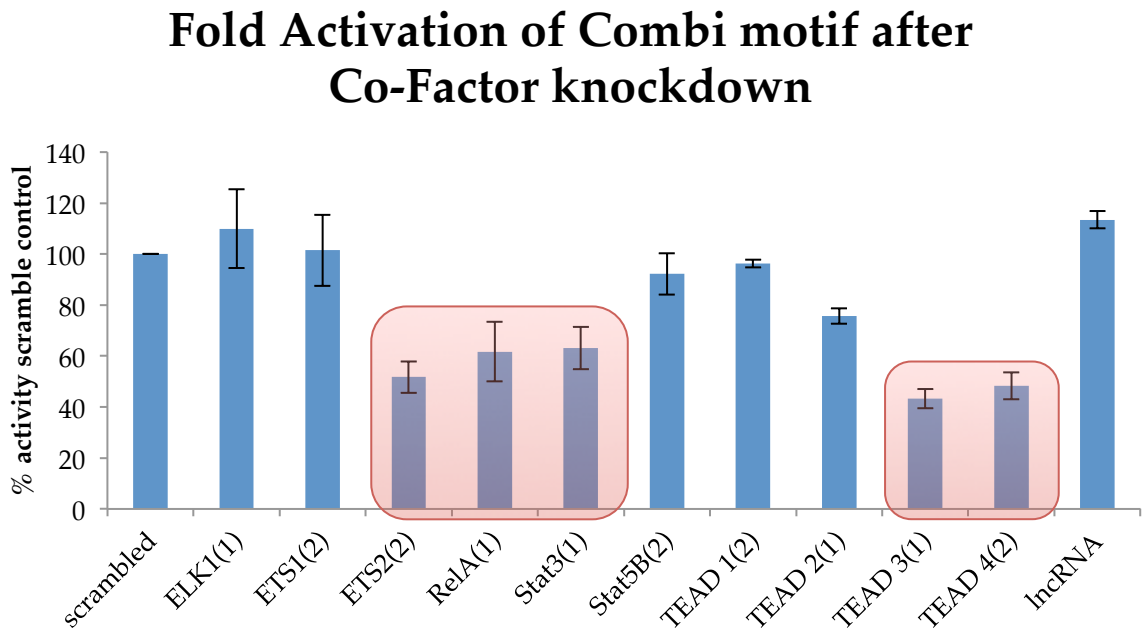

Figure 42: Fold activation of the combi motif after knockdown of several candidate co-factors. Cells were treated with hormone and fold activation of combi motif of scrambled dsiRNA treated U2OS cells was set to 100 %. Knockdown of ETS2, RelA and Stat3 as well as two of four TEAD variants reduced the fold activation, thus indicating a possible interaction of GR with the particular factor at the combi motif. Experiments were done in three biological replicates; Error bars indicate standard error of mean.

To test if the effects observed were specific to the combi motif, in parallel we tested the effect of their knockdown on a conventional GBS reporter. Here, we only tested genes where the knockdown had an effect on the combi motif: ETS2, RelA, Stat3 as well as TEAD3 and TEAD4. If the knockdown only affects the transcriptional regulation of the combi motif and not the hormone induced regulation of GR at the GBS (CGT) motif, this would indicate a specific interaction of GR and factor X for the combi motif.

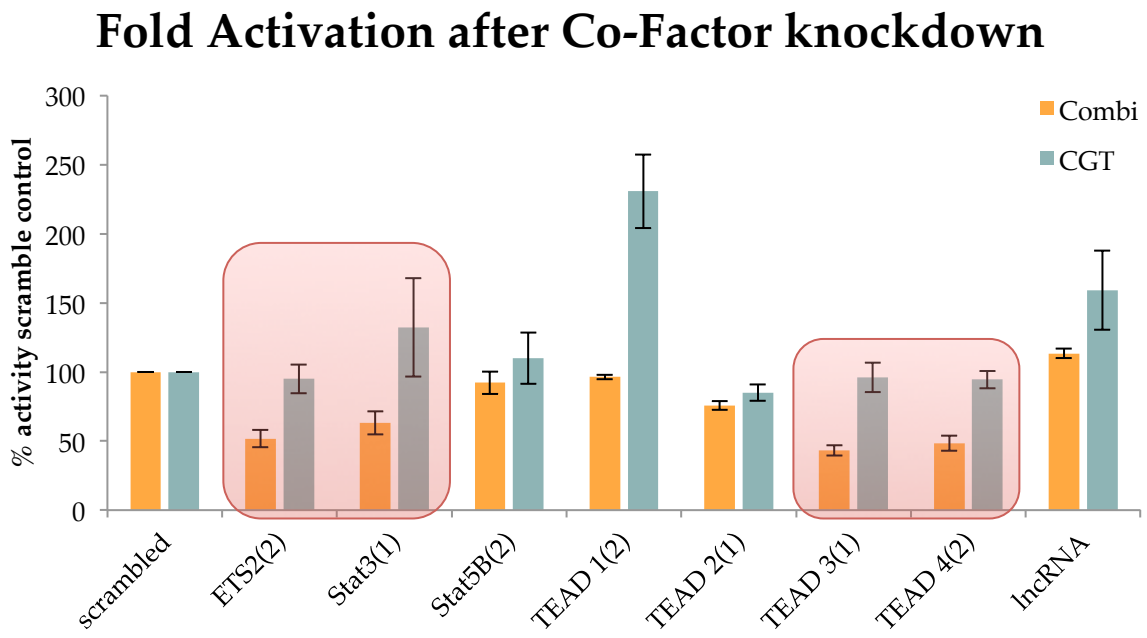

Figure 43: Fold activation of the combi motif compared to canonical GR motif (CGT) after knockdown of several candidate co-factors. Cells were treated with hormone and fold activation of combi motif of scrambled dsiRNA treated U2OS cells was set to 100 % for each motif. ETS2, Stat3 together with TEAD3 and TEAD4 were reducing the fold activation of the combi motif containing pGL3-reporter, but not the fold activation of the CGT reporter. These four factors were showing independent effects on regulation, indicating a specific interaction of GR with the particular factor. Experiments were done in three biological replicates; Error bars indicate standard error of mean.

I found that knockdown of ETS2, Stat3, TEAD3 and TEAD4 were diminishing the fold regulation of the combi reporter in hormone-treated U2OS cells, but not the fold activation of the CGT reporter vector, containing the canonical GR motif (Figure 43). This indicates a specific effect of the knocked-down candidates, corroborating these four factors as possible candidates for partnering with GR at the *de novo* identified combi motif.

Together with (I) the data of the EMSA assays, indicating no multi- or monomeric GR interaction at combi site, (II) the structural alignment, showing no structural constrains for GR monomer and another protein binding to the same location at the combi motif, (III) the Mass-Spec data showing TEAD3 as only

candidate binding to the combi motif when used as bait in a DNA pulldown assay, the knockdown of TEAD3 indicates a functional relation of GR and TEAD3 at combi site to regulate transcription. However, also knockdown of TEAD4, Stat3 and ETS2 were diminishing the transcriptional activity of the combi reporter, indicating a possible larger protein complex responsible for the transcriptional regulation or the fact that GR can hetereodimerize with several alternative partner proteins at this motif.

7.3.8 Fox footprint profile

Of all the enriched motifs at GR-ChIP-Seq regions in IMR90 cells, only a few showed a clear footprint profile. Among them, we found a sequence bound by members of the forkhead box (Fox) family of transcription factors (Figure 44 A), that was enriched in GR-ChIP-Seq data (Figure 44 B) and delivered a very defined footprint in GR-ChIP-Exo data (Figure 44 D). It was also obvious that not all sequences were likewise enriched in all cell types. The FoxA1 motif for example was found to be enriched in IMR90 but not in U2OS cells (Figure 44 C), consistent with the cell-type-specific protein:DNA interactions of TFs.

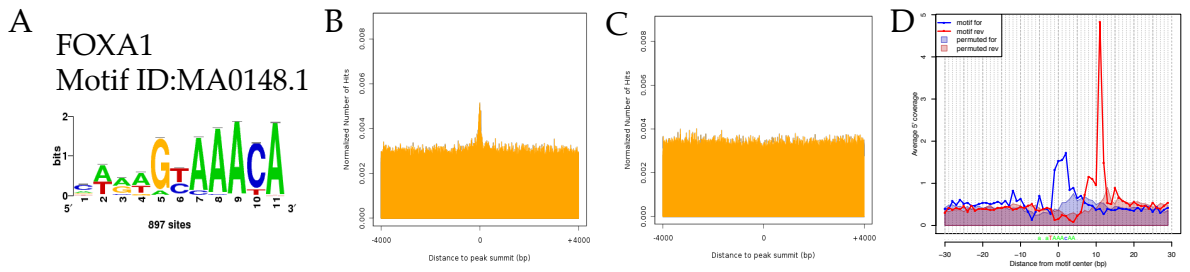

Figure 44: A: FoxA1 motif. B and C: FoxA1 motif distribution at GR-bound loci in IMR90 (B) and U2OS cells (C). DNA sequences from +/- 4000 base pairs (bp) from ChIP-Seq peak summits were aligned and sub-divided into 50 bp bins. Alignment scores for matching motifs of the PWM of FoxA1 (from Jaspar) in every bin were assigned to calculate the relative frequency distribution of FoxA1 motif around GR-ChIP-Seq peaks. The orange plot shows the normalized number of hits for FoxA1-PWM relative to the to the peak center on position 0. D: Fox footprint found in GR-ChIP-Exo data for the Fox motif. Shaded area show the signal for a permutated Fox binding sequence. Analysis was done in collaboration with Dr. Morgane Thomas-Chollier and Jonas Ibn-Salem.

Our Fox footprint looked distinct from the one observed for footprints generated at canonical GR binding sites (Figure 45 A). Fox footprint profiles have a well-defined peak 8 bp upstream of the motif on the forward strand and a broader complementary peak on the reverse strand. The well-defined peak could be a result of one of the three putative Fox-crosslink points, which is more prominently cross-linked than the others (Figure 45 B). This crosslink point yields therefore in higher peaks (more reads) in ChIP-Exo.

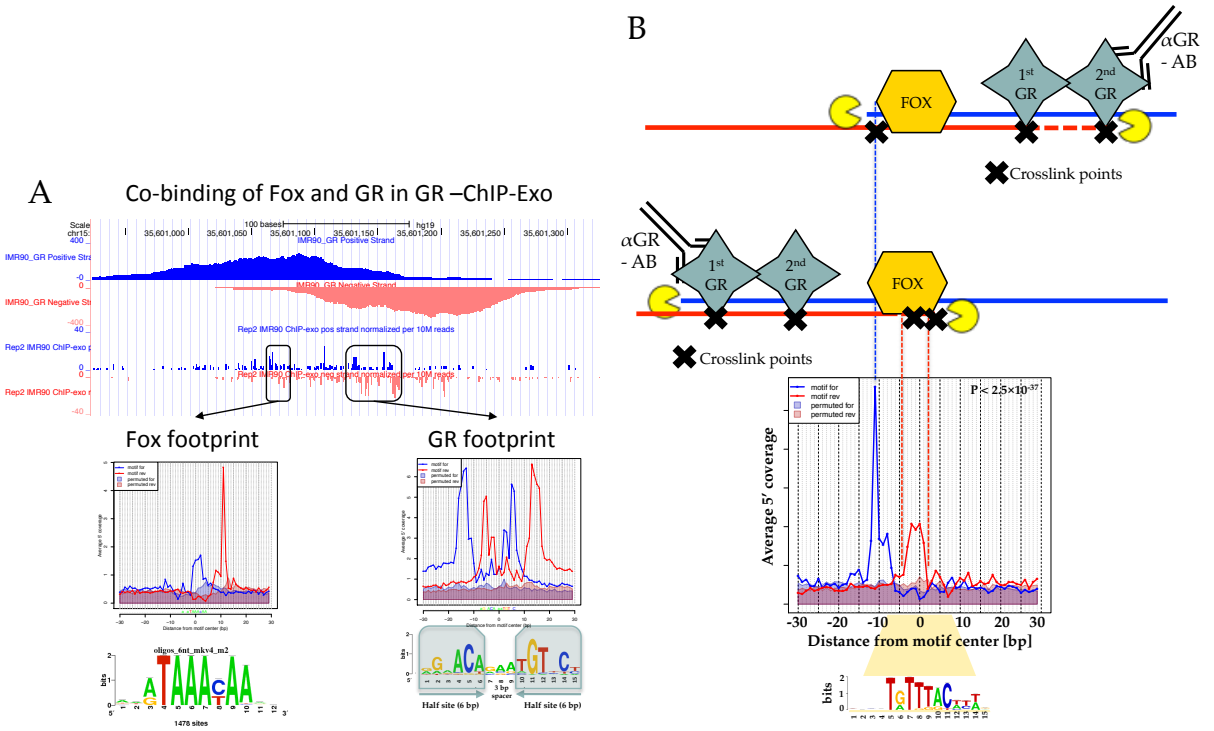

Figure 45: A: ChIP-Exo analysis revealed footprints at Fox motifs close to GR footprints. B: Schematic explanation for Fox signals in GR-ChIP-Exo. GR can be bound at both sides of the Fox protein, depending on the orientation the merged signals accumulate for both crosslink variants. When one of the three Fox-crosslink points is cross-linked more efficiently than the others, this crosslink yields a higher and sharper peak (more reads) in ChIP-Exo.

The different footprint suggested that a protein other than GR binds to these Fox sites despite the fact that the ChIP was done with an antibody specific to GR. This could happen when the associated protein tethers GR to the DNA, or when it is efficiently cross-linked to the DNA near sites where GR binds and consequently co-precipitates during the ChIP procedure. This is corroborated by our electrophoretic mobility shift assay (EMSA) of the Fox motif, showing no increased affinity of GR DBD for the Fox sequence when compared to a randomized control sequences (Figure 46). GR binds only non-specifically to DNA as a monomer at high GR-concentrations (shown for the 10 μM GR-DBD samples for both sequences in Figure 46). GR therefore appears not to be able to interact directly with the Fox motif enriched at GR-bound regions.

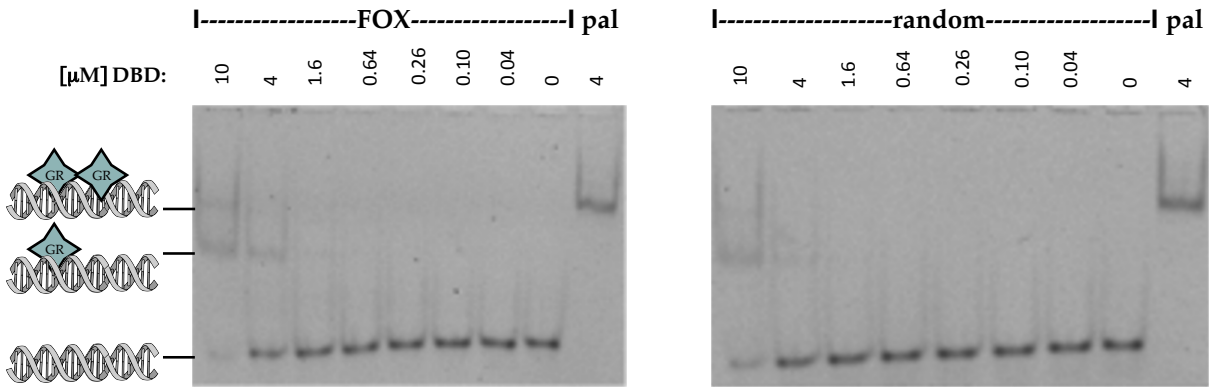

Figure 46: EMSA of Cy5-labeled DNA and different concentrations of GR-DBD (amino acids 380-540 (Meijsing et al. 2009)). GR does not bind the Fox motif. Shown is binding data for the isolated FOX sequence (left) or a random sequence (right). Lane "pal" contains labeled DNA with a canonical GBS sequence, where GR binds as a dimer leading to a shift of DNA by GR binding in the EMSA.

The most obvious candidate proteins to bind are members of the Fox family of transcription factors that share a similar DNA binding domain. To determine if Fox could be the protein causing our Fox footprint, we compared "our" footprint with the footprint for ChIP-Exo data targeting FoxA1 (Serandour et al. 2013). Therefore, we reprocessed the FoxA1 and the ER ChIP-Exo data using our ExoProfiler pipeline to generate footprints for a direct comparison. A similar footprint would indicate, that a Fox protein is indeed responsible for the footprint observed in our ChIP-Exo data for GR.

We found (Figure 47) a footprint profile at the FoxA1 motif that looks very similar to the specific FoxA1 footprint profile from our ChIP-Exo data for GR. Thus, our observed Fox footprint profile might reflect binding of a member of the Fox family of TFs, since they share the distinct sharp peak on the forward strand 8 bp upstream of the motif and the broader peak on the opposite strand, which was also observed in the original study (Serandour et al. 2013). The striking similarity of the footprint suggests that the same, or a structurally related, protein is responsible for the observed footprints. Again, we were not able to generate footprint profiles for the Fox motif in the CTCF ChIP-Exo data (Rhee & Pugh 2011), showing that this footprint profile is specific for the protein immunoprecipitated and not an artifact of the ChIP-Exo procedure.

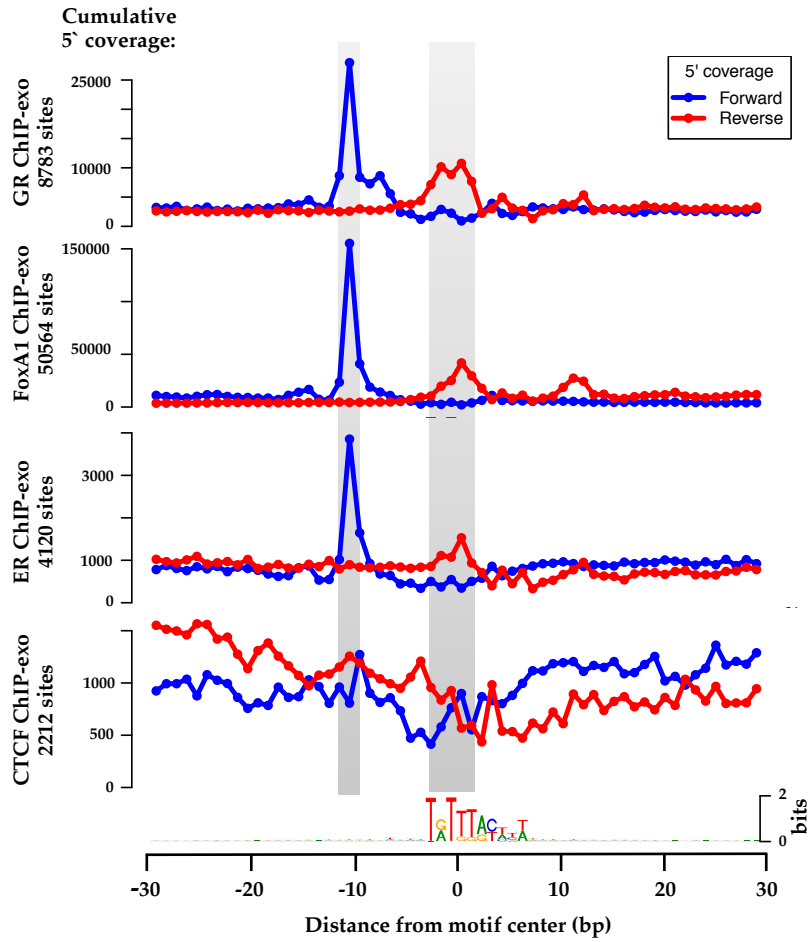

Figure 47: Footprints generated with the ChIP-Exo pipeline (ExoProfiler) for the Fox motif from our GR-ChIP-Exo data (top) and generated from FoxA1 (second) and ER-ChIP-Exo (third) (Serandour et al. 2013). All three footprints showed a striking similarity concerning the distinct sharp peak on the forward strand 8 bp upstream of the motif and the broader peak on the opposite strand. No footprint profile was detected for Fox motifs in the CTCF ChIP-Exo data (bottom) (Rhee & Pugh 2011). Analysis was done in collaboration with Jonas Ibn-Salem.

In conclusion, although we haven't directly identified the protein responsible for the Fox footprint in our GR study, the striking resemblance to the FoxA1 footprint argues that binding of a member of the Fox family of TFs is responsible for the Fox footprint observed in our study.

Fox motifs are bound by proteins identified in ChIP-Exo and these motifs are also enriched in our ChIP-Seq analysis even though the ChIP was performed with an antibody specific to GR. The question arises if GR alone is sufficient to induce transcription at genomic sequences with GR and Fox motifs in combination, or if Fox is needed to enable GR to drive transcription. However, I first studied the Fox motif in isolation to test if GR was able to regulate transcription when only Fox binding sites are present (Figure 48). Here I found no clearly discernable regulation indicating that GR cannot be active from Fox sites in isolation.

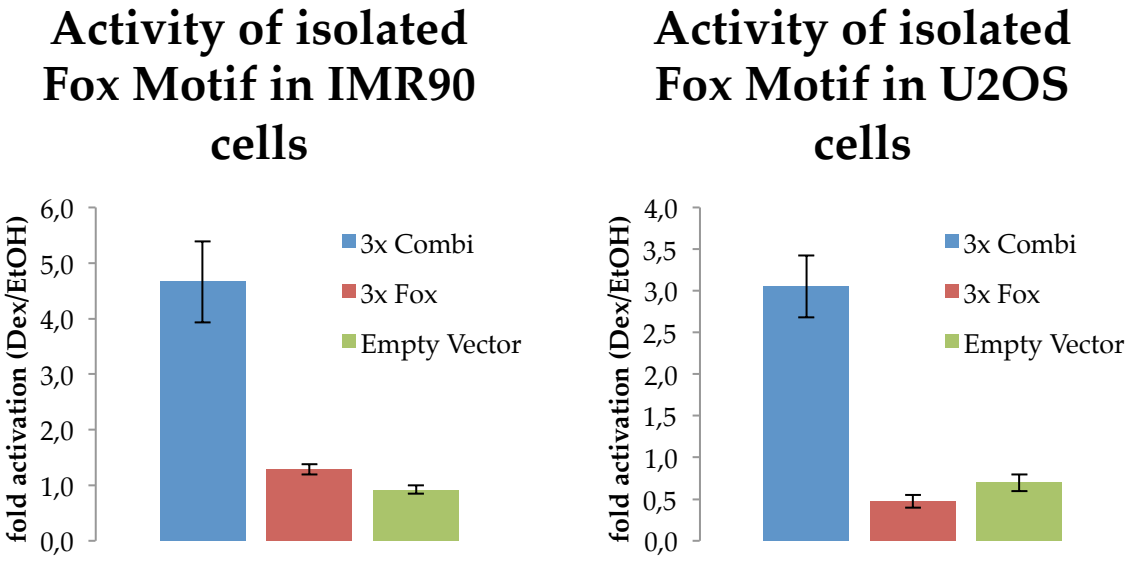

Figure 48: Isolated Fox motif in luciferase reporters is not able to induce transcription in IMR90 (left) and U2OS (right) cells, since hormone dependent fold action is equal to empty vector. Experiments were done in three biological replicates; Error bars indicate standard error of mean.

To further test the functional relation between GR and Fox, several GR-bound genomic loci near GR-regulated genes containing Fox and GR motifs in combination were cloned into reporters and tested for transcriptional activation of a luciferase reporter gene. From 6 reporters containing a Fox and a GR motif, 4 showed a reproducible GR-dependent activation (Figure 50). For all active reporters, mutation of the Fox sequence at key positions TAAACA to TAGGCT resulted in reduced GR-dependent regulation. In addition, these reporters contained GBS-like sequences, which were mutated individually for all reporters as indicated in Figure 49.

```
Canonical: AGAACAxxxTGTTCT
 Fox1-GBS:  CAGACGtacTGTTCC to CAGAAGtacAATTCC

Canonical: AGAACAxxxTGTTCT
 Fox5-GBS: AGAGCAtccTGTACT to AGAGAAtccTATACG

Canonical: AGAACAxxxTGTTCT
 Fox6-GBS: AGATAAggaAGTACT to AAATAAggaAATAAA

Canonical: AGAACAxxxTGTTCT
 Fox9-GBS: TGCTCAaaaTGTTCT to TGCTAAaaaTATTCC
```

Figure 49: Mutations (underlined) in genomic Fox sites to destroy GR binding sites.

Similar to the observation for the Fox motif, mutating the GR motif also resulted in reduced activation of all four combinatorial regions tested (Figure 50). Diminished activity caused by mutating the Fox motifs indicates that Fox motifs are

indeed involved in the recruitment of co-regulatory factors of GR (Sahu et al. 2011; Jitrapakdee 2012) and play an important role for GR to drive transcription.

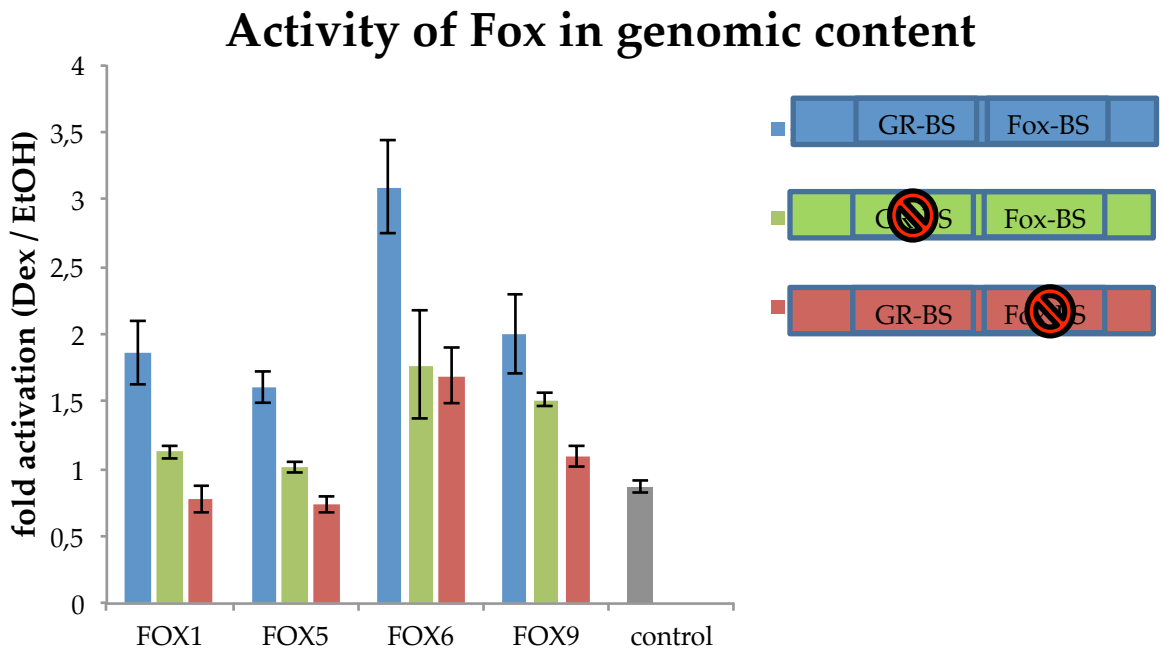

Figure 50: Genomic regions containing Fox and GR motifs were compared to mutants missing the GR-BS or the Fox-BS respectively. Transcriptional activity of genomic regions missing the Fox (red bars) or GR (green bars) binding site is diminished in all tested regions, compared to the genomic regions (blue bars). Experiment was done in three biological replicates; Error bars indicate standard error of mean.

These studies clearly indicate a functional connection between GR-dependent transcriptional activation and the presence of a Fox sequence, which was found to be enriched at GR-ChIP-Seq peaks. The Fox family is known to pave the way for hormone receptor binding potentially by opening the DNA in vicinity to GBS (Carroll et al. 2005).

In this composite binding of GR and Fox, the Fox footprint could be a consequence of immunoprecipitation of genomic regions where both GR and Fox were bound and cross-linked to their consensus sequence, rather than Fox tethering GR to the DNA. Since tethered GR binding has been linked to transcriptional repression (Ratman & Berghe 2013), we analyzed the transcriptional regulation of genes close to ChIP-Seq peaks harboring a Fox motif, but missing the GR motif. Here we found that these genes were indeed, on average, down regulated (Figure 51) whereas genes near peaks with only a GBS or with both a GBS and a Fox motif match were, on average up, regulated.

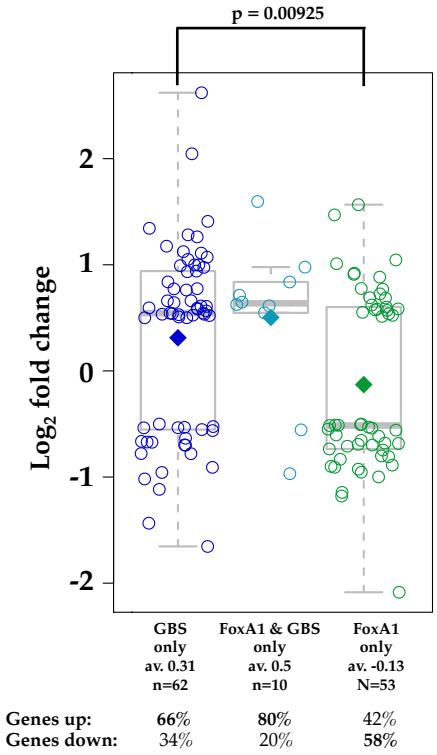

Figure 51: Boxplot of log fold change for differentially expressed genes upon hormone treatment for 4 hours with 1 μM dexamethasone (log fold change ≤ -0.5 or ≥ 0.5). Genes with ChIP-Seq peaks in the region ± 20 kb around the TSS with only a FoxA1 motif (p<0.0001) are marked in green, genes with peaks with only a GBS (p<0.0001) motif in dark blue, genes with peaks with sequences matching both motifs in turquoise. Center lines show the medians; diamonds show the mean; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles. Analysis was done in collaboration with Dr. Morgane Thomas-Chollier

Together, our data suggests that Fox proteins are co-bound at genomic regions of GR binding and can either tether GR to such regions, or play a role in facilitating GR binding to GBSs nearby.

7.3.9 Stat Footprint profile

A non-canonical footprint identified by ExoProfiler was for the recognition motifs for various members of the signal transducer and activator of transcription (STAT) family of transcription factors. The footprint profile for the Stat1 motif (MA 0137.3, Figure 52 B) looks similar to the footprint profiles found for Stat3, Stat5 and Stat6 motifs. Likewise to GR, STAT proteins bind as dimers to STAT binding elements (SBEs) containing a STAT motif. The STAT binding motif consist of an inverted repeat of 4 bp half-sites separated by a 1 bp spacer (Langlais et al. 2008). SBEs are enriched at GR-ChIP-Seq peaks in IMR90 cells (Figure 52 A), and a large fraction of these peaks appears to have an SBE-matching sequence but lack a GBS (Figure 52 C).

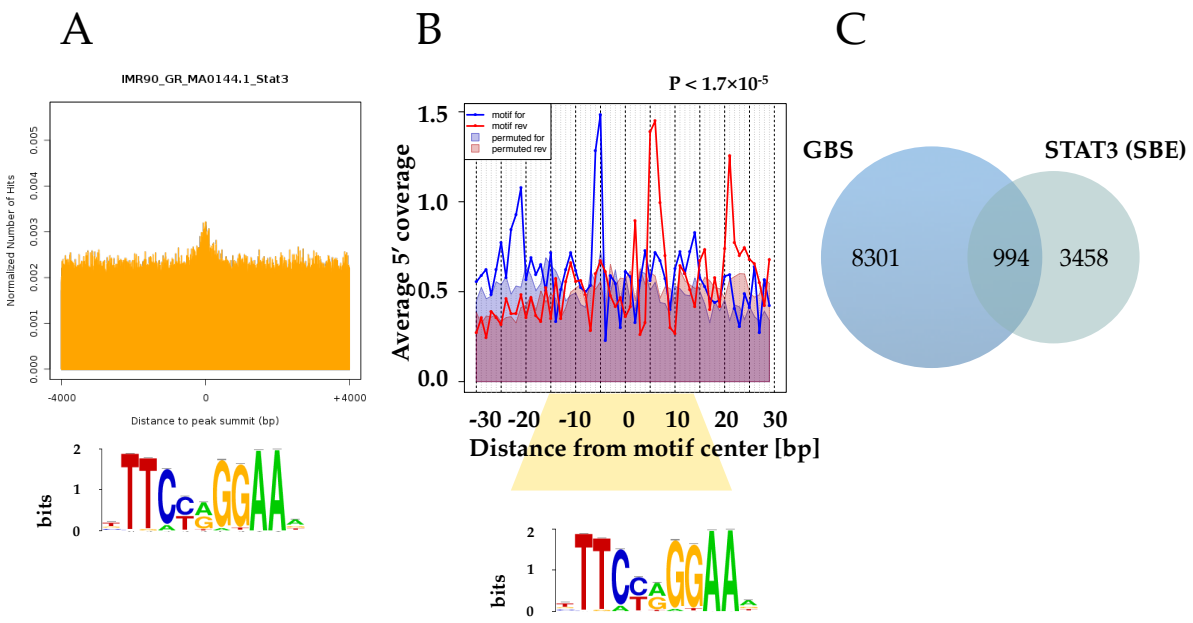

Figure 52: (A) The Stat3 motif was enriched in motif distributions at GR-bound loci. DNA sequences +/- 4000 base pairs (bp) centered on the ChIP-Seq peak summits were aligned and sub-divided into 50 bp bins. Alignment scores for matching motifs of the PWM of Stat3 (from TRANSFAC) in every bin were assigned to calculate the relative frequency distribution of Stat3 motif around GR-ChIP-Seq peaks. The orange plot shows the normalized number of hits for the GR-PWM relative to the peak center on position 0. (B) ExoProfiler result of ChIP-Exo footprint profile at Stat3 sequences. (C) Binding sites of Stat3 and GR in whole ChIP-Seq peaks at p-values for each motif of 0.0005. Analyses were done in collaboration with Dr. Morgane Thomas-Chollier.

The symmetry of the STAT footprint profile (Figure 52, B) is consistent with dimeric STAT binding and the fact that the footprint profile looks distinct from the one observed for GBSs indicates that a protein other than GR is bound. Subsequently, we analyzed the regulation of genes with nearby ChIP-Seq peaks containing STAT binding sites (SBE) in IMR90 cells. Genes associated with GR-ChIP-Seq peaks containing an SBE but lacking a GBS were, on average, transcriptionally repressed by GR, whereas genes associated with ChIP-Seq peaks containing GBSs but lacking SBEs were, on average, activated (Figure 53). These findings corroborate the recent findings (Langlais et al. 2012) of GR-dependent transcriptional repression if Stat3 is tethering GR to the DNA. Our group (Dr. Marcel Jurk) is currently working on Stat3 binding profiles. These ChIP experiments could show a co-occurrence of GR and Stat3 (hormone independent) at regions containing an SBE consistent with Stat3-dependent tethering of GR.

Together, these findings indicate that ChIP-Exo data can uncover footprints reflecting tethered GR binding.

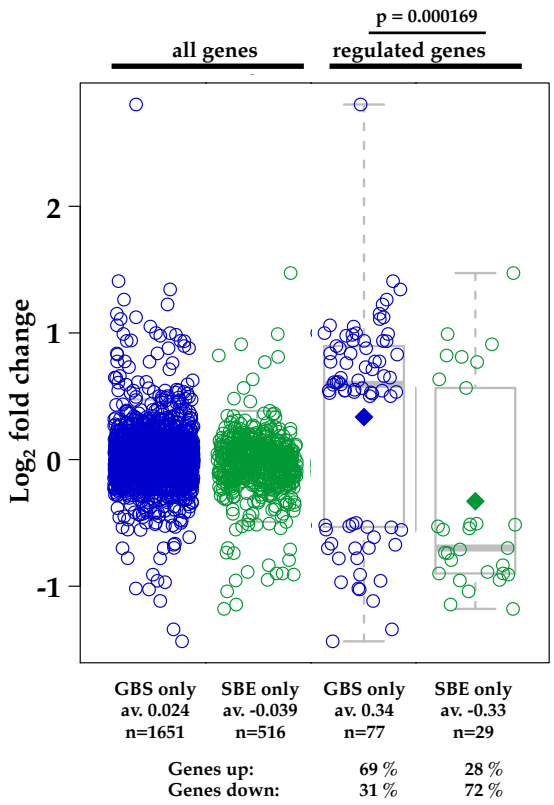

Figure 53: Boxplot of log fold change for (left) all genes upon treatment for 4 hours with 1 μM dexamethasone and (right) for genes that are differentially expressed with a log fold change ≤ -0.5 or ≥ 0.5. Center lines show the medians; diamonds show the mean; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles. Genes with ChIP-Seq peaks in windows ± 20 kb around the TSS, harboring a specific motif are indicated by circles (Genes with peaks with Stat3 (p<0.0001) but not GBS are marked in green, GBS (p<0.0001) but not Stat3 in blue). Analysis was done in collaboration with Dr. Morgane Thomas-Chollier

8 Discussion

To address the diverse tasks in the human body, transcription factors (TFs) define the expression level of specific parts of the genome holding the information to respond properly, by binding to specific DNA-sequences or motifs. This DNA sequence recognized by a TF does not provide enough information to explain where in the genome a TF binds, as only a small percentage of many possible TF binding sequences in the genome is actually TF-bound (Myers et al. 2011). The glucocorticoid receptor (GR) is studied to understand transcriptional regulation and controls a wide array of physiological processes including proliferation, development, inflammation and metabolic homeostasis (Sapolsky et al. 2000). One mechanism to fulfill these tasks is GR's ability to act as a hormone-inducible transcription factor, making GR an ideal candidate to study transcriptional regulation. GR and other TFs are guided to specify genomic regions by the chromatin landscape (Burd & Archer 2013) and their interaction with other proteins (Langlais et al. 2012) thereby regulate the transcription of genes.

First, we studied the influence of the chromatin landscape in specifying where in the genome GR binds, by focusing on histone modifications and DNA accessibility. Although an increasing number of histone modifications have been identified (Suganuma & Workman 2011; Kimura 2013), only some could be functional understood (Petty & Pillus 2013). The diverse histone modifications are thought to provide a histone code, whereby combinations of histone modifications have specific functions (Jenuwein & Allis 2001). We therefore analyzed histone modifications, which have been linked to specific genomic regions (e.g. promoter regions) or could be linked to transcriptional activity, to search for a correlation to GR binding, beginning on the next page.

Second, we complemented the chromatin aspect of GR binding by a detailed analysis of the bound GR sequences. As mentioned above GR and other TFs interact with other proteins to bind genomic regions. This coordinated recruitment of GR and its cofactors to regulatory elements is reflected by the typical co-enrichment of multiple sequence motifs at genomic GR binding sites defined by ChIP-Seq. However, how and if these enriched motifs participate in genomic binding of GR remains elusive. We will answer the functional relevance of some enriched motifs by analyzing GR binding sites in base pair resolution in the second part (page 120).

8.1 What defines a TF binding site? – Role of the chromatin landscape

To analyze GR binding and the influence of the chromatin landscape we (in collaboration with Dr. Michael I. Love) worked on IMR90 cells, a cell line with a well-studied chromatin landscape. In concordance with other studies, we found a predominant binding of GR to chromatin, which is "open" and accessible (based on DNase I hypersensitivity site (DHS) sequencing) already before hormone treatment (John et al. 2011). To test whether the minority of GR-bound sequences located in "closed" chromatin are not false negative DHS, I examined the DNase I sensitivity at an "open" and a "closed" locus in the FKBP5 gene, as it is a known target gene for the steroid receptor family (Sinars et al. 2003) as well as for the ZBT16 gene. In this experiment I also tested the influence of hormone treatment on the chromatin accessibility as determined by DNase I sensitivity for all four loci pre and post hormone treatment. For the GR binding sites in "closed" chromatin, we could show a hormone-induced gain of accessibility as indicated by an increase of DNase I sensitivity at both regions of the FKBP5 and the ZBTB16 gene, thereby showing that these regions were indeed DHSs only after hormone treatment and thereby not false negative DHS regions. The hormone dependent chromatin remodeling, which was also shown by others (John et al. 2011), also indicates GR can indeed function as a pioneering factor as GR is able to interact with the DNA within "closed" chromatin.

This raises the question: what distinguishes GR-bound loci in "open" and "closed" chromatin? In agreement with other studies (John et al. 2011), we found that GR binding in "closed" regions correlates positively with binding to sequences with high motif scores arguing that GR's pioneering function depends on a tighter protein:DNA interaction. Further, our data indicates that hormone treated GR-bound regions appear more sensitive to DNase I degradation in predefined "open" chromatin, than GBS in "closed" chromatin. This could indicate a spatial limitation of GR-induced DNA accessibility, especially since qPCR primers for GBS span a region of 81 bp (FKBP5) to 130 bp (ZBTB16) at "closed" chromatin regions. The GR motif itself however is only 15 bp long and the gained accessibility could therefore be limited to a smaller region. Since no single motif could be specifically assigned to predefined "closed" GR binding sites, we conclude that GR is using its intrinsic function to act as a pioneering factor (Becker et al. 1984; Richard-Foy & Hager 1987) for opening up the chromatin at such loci and not with the help of a specific co-factor as indicated by other studies. These studies found C/EBPb, a cell-type-specific co-regulator of GR, is responsible for the maintenance of DNA accessibility on the one

hand, but it could also be shown that the chromatin remodeling of *de novo* sites bound by GR requires C/EBPb presence for non-tethered GR recruitment (Grøntved et al. 2013). This dual role for chromatin remodeling by co-regulators indicates the necessity of further studies for cell-type-specific TF interactions as started with the Encode project.

We further examined GR binding to "open" chromatin regions and besides the GR motif itself we set out to identify chromatin marks that distinguish GR-bound from unbound regions within the "open" chromatin universe. The discrimination of chromatin marks in "open" and "closed" GR-bound regions would lead to marks distinguishing "open" and "closed" chromatin, rather than identify GR-specific chromatin marks. We therefore analyzed GR binding to "open" chromatin genome-wide using Hierarchical Bayes modeling in multiple cell lines to correlate GR binding patterns with chromatin and sequence features in sites of "open" chromatin, so called DNase I hypersensitive sites (DHS). This approach identified positively and negatively correlated chromatin marks.

Positive features for GR binding in IMR90 cells, besides the GR binding motif itself, were identified as DNA accessibility, H3K27ac, H3K4me1, H3K9me3. H3K27ac and H3K4me1 were used to identify enhancer regions associated with increased GR binding, whereas H3K9me3 was found at cell-type-specific enhancers (Zhu et al. 2012), explaining also a positive correlation to GR binding. Negative correlated features of GR binding in IMR90 cells were H2A.Z, H3K36me3, H3K4me3, H3K79me2 and H3K9ac. The four last negatively correlated features were found at actively transcribed regions in general, in particular the histone modifications H3K36me3 (actively transcribed gene bodies, (Guenther et al. 2007)), H3K4me3 (active transcription, (Karlić et al. 2010)), H3K79me2 (activation of transcription, (Steger et al. 2008)) and H3K9ac (actively transcribed promoters, (Koch et al. 2007)).

Consistently positive features correlating with GR binding at DHS are for example H3K27ac, indicating active enhancers, and H3K4me1, used to identify enhancer regions in general (Creyghton et al. 2010). Both could be linked to the findings that the majority of GR binding sites occur outside classically defined promoters (also corroborated by the negative correlation of promoter marks, as mentioned above), indicating that enhancers may be the predominant docking sites for GR recruitment. H3K9me3, is also associated with constitutive gene repression (Hublitz et al. 2009) and positively correlates with GR binding in "open" chromatin.

Together with the aforementioned negative correlation of GR binding with regions marked as actively transcribed in general, this could be two sides of a mechanism to enable GR to target genes cell-type-specifically: keeping GR away from general transcribed genes (negative marks: H3K36me3, H3K79me2 and H3K9ac) and attracting GR to constitutive repressed, but cell-type-specifically active enhancer regions (positive marks: H3K9me3). Furthermore, the ChIP-Seq input control, DNase I sensitivity and the GR motif itself have also a positive correlation with GR binding in "open" chromatin. This shows the preferred binding of GR to its cognate sequence in "open" chromatin, corroborated by our finding that GR binds even in the universe of "open" chromatin to the most "open" sites. It also emphasizes the necessity to control for biases (e.g. input).

As mentioned above, negatively correlated features (H3K36me3, H3K79me2 and H3K9ac) are associated with active transcription of genes in general, it is therefore most likely to find these marks negatively associated with GR binding to enable GR to target genes in a cell-type-specific way. H3K36me3 could also be associated with the prevention of cryptic initiation of transcription, by a possible interaction with TFs (Carrozza et al. 2005). Cryptic initiation leads to non-coding RNA transcribed from an intragenic and intergenic region and occurs at nucleosome depleted regions (Colin et al. 2011). Keeping GR away from these sites could prevent a burst of non-coding cryptic RNAs upon hormone induction of GR. A second way to prevent GR binding to regions of general transcription is indicated by negatively correlated marks of the chromatin landscape associated to nucleosome positioning effects, like increased nucleosome density (H3K36me3 (Schwartz et al. 2009)) and stability (H2A.Z, (Kumar & Wigge 2010)). Nucleosome dense regions might limit the accessibility of cognate sequences thereby preventing TF binding. GR's negative correlated binding at H2A.Z-containing nucleosomes could reflect this limited access of TFs in general and is corroborated by the fact that only if H2A.Z is incorporated in the +1 nucleosome of a transcriptional start site (TSS) it was linked to increased transcription and the recruitment of general TFs (Bargaje et al. 2012; Weber et al. 2014). Therefore, a general attraction of GR to H2A.Z-containing nucleosomes would be in contrast to GR's function to target genes in a cell-type-specific way. Additionally, the negative correlation of H2A.Z could be linked to the enrichment of H2A.Z in promoter regions of a gene and simply reflect our aforementioned finding of promoter-proximal depleted GR binding. Also the negative correlated H3K4me3

and H3K9ac marks typically found at promoters (Koch et al. 2007; Karlić et al. 2010) could be linked to promoter-proximal depleted GR binding.

We could indeed show a diminished GR presence at promoter-proximal regions in A549, IMR90 and K562 cells, when compared to the distribution found for all DHS. Part of this depletion of GR binding to promoter regions can be explained by a disfavor of GR binding based on the sequence composition at promoters (Figure 54 A, page 118). This is a major finding as it shows the direct interaction of GR to promoter regions in "open" chromatin is hampered to a large degree on the level of the DNA-sequence itself. Since not all depletion of GR binding could be explained by the sequence composition, we see indications that also other mechanisms are involved in the reduced promoter proximal binding of GR. The histone marks have to be set by Histone acetyltransferase (HATs) and methyltransferases (HMTs), if these histone-modifying enzymes are localized at the chromatin it has been shown that they also post-translational modify GR (Kovacs et al. 2005; Minucci & Pelicci 2006). GR is sensitive to e.g. acetylations and loses its affinity to NF-κB upon acetylation (Ito et al. 2006). Furthermore, GR-acetylation was shown to attenuate GR binding and consequently represses GR-induced genes (Nader & Hardt 2009; Kino & Chrousos 2011; Zelin et al. 2012). This demonstrates the connection of TF binding and the chromatin landscape, as they can mutually influence each other. Moreover, the connection between histone marks and GR binding might be indirect and simply reflect the fact that enzymes exist that modifying both GR and histone tails. In this scenario, the actual mechanisms responsible for reduced GR binding is that GR is modified by the same enzyme that deposits the histone mark and thus the correlation with the histone mark simply reflects the presence of the enzyme.

Together, these findings indicate that GR binding is to a large degree predetermined by DNA accessibility in IMR90 cells (consistent with other studies (John et al. 2011)). The enriched histone marks found at GR-bound regions in "open" chromatin were correlated to enhancer regions are also found by others (Burd & Archer 2013) and in combination with the negatively correlation of promoter marks (e.g. H3K9ac) support the findings of reduced GR binding to promoter regions in "open" chromatin. This could only be found since we focused on determining features of GR binding sites in "open" chromatin only (Figure 54 B). If GR binding is analyzed regardless of the chromatin state one can detect enriched GR binding to promoters. All these intriguing results were found by our unique approach of looking at the DHS universe only and could contribute to cell-type-specific

transcriptional regulation. Promoter regions are typically in "open" chromatin across several cell types, whereas enhancer regions are usually located in cell-type-specific "open" chromatin (Thurman et al. 2012). GR binding to promoter regions would therefore not lead to cell-type-specific gene regulation, but our finding that GR binding is associated with histone marks that are correlated with distal enhancer regions does follow this line of thinking.

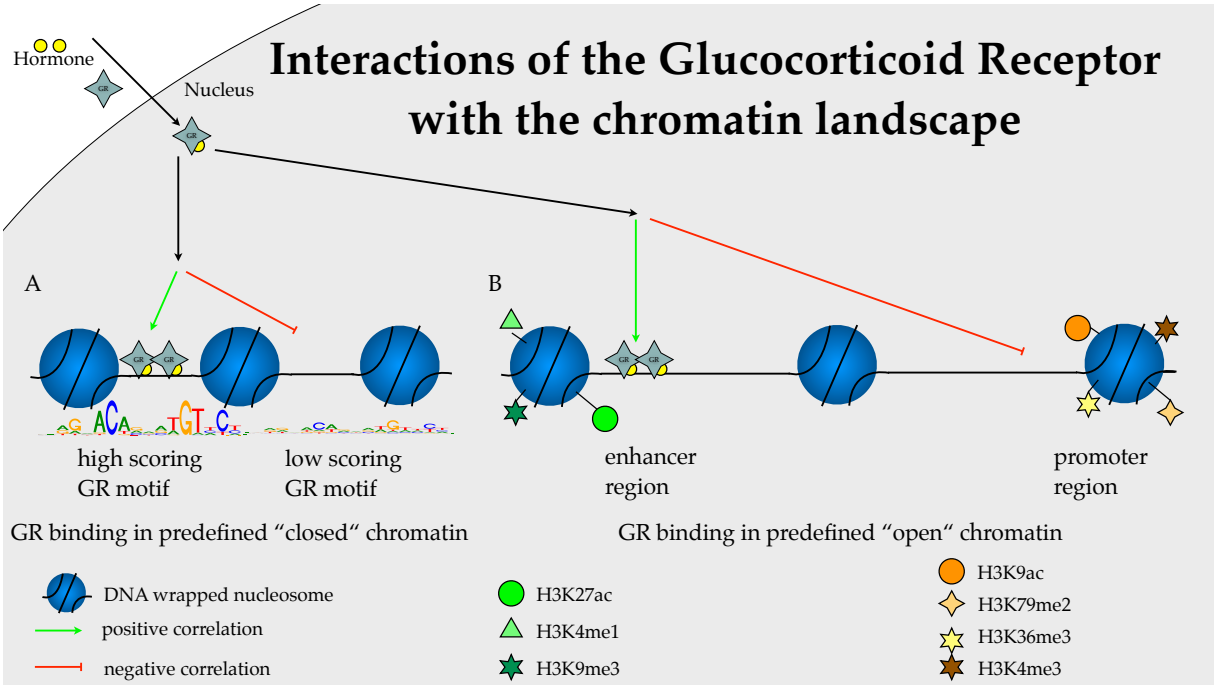

Figure 54: Overview of possible GR interactions with the genome influenced by the chromatin landscape. Our findings indicate GR interactions in predefined "closed" as well as "open" chromatin in IMR90 cells. A: Activated GR binds in predefined "closed" chromatin region more likely to DNA sequences matching more to the canonical GR binding sequence (high scoring motifs). B: In "open" chromatin we found GR binding to be positively (green symbols) correlated to histone marks found to be present at cell-type-specific enhancer marks (e.g H3K4me1), whereas histone marks linked to promoter regions, that are shared among multiple cell lines (e.g H3K4me3), are found to be negatively (red and orange symbols) correlated with GR binding in our Hierarchical Bayes modeling.

The answer to the emerging question of the role of the chromatin landscape in defining a TF binding site is: a combination of several histone marks are guiding GR to its loci, but rather not as an independent code. The histone marks within the chromatin landscape work more as a framework (Gusmão et al. 2014), as indicated by the negative (H3K4me3, indication of promoter) and positive (H3K4me1, indication of enhancers) signals. Together with the accessibility (DNase I sensitivity) and the TF motif itself the chromatin landscape provides essential information to specify TF binding sites in a cell-type-specific matter and to define genetic programs that are regulated in response to stimuli like glucocorticoids. However, the chromatin landscape has to be flexible in terms of accessibility to allow TFs to

recognize their binding motifs. This flexibility is provided by pioneering factors that can adjust the spectrum of binding sites available for binding by other factors.

Further experimental approaches could determine whether solely histone modifications or other proteins associated to histone modifications guide TFs, in particular GR, to its binding sites, or whether they are merely correlative.

The use of GAL4 fusion proteins could for example show, if H3K4me3 deposited at a nucleosome close to a GR binding site leads to reduced GR-dependent transcriptional activity, as indicated by our preliminary results. These experiments are based on the recruitment of enzymes depositing the modification of interest and are fused to the GAL4 DNA-binding-domain to interact with the GAL4 binding site, inserted upstream to a GR binding site. GR itself drives a minimal promoter linked to a luciferase reporter gene, allowing the quantitative measurement of GRs regulatory activity. If the modification of interest interferes with transcriptional regulation by GR, the reporter gene activity would be reduced.

The aforementioned preliminary results suggest a reduced GR-dependent transcription of the reporter, caused by H3K4 trimethylation (deposited by several WDR5 and SET-protein containing complexes, (Robert et al. 2014)), H3K9 acetylation (deposited by GCN5) and the presence of H2A.Z-containing nucleosomes. Since these features of the chromatin landscape were found to be negatively correlated with GR binding they argue for a potential causative connection between these histone modifications with GR binding and regulatory activity. The recently developed CRISPR/CAS system (Jinek et al. 2012) offers an elegant way to test this hypothesis directly in the genomic setting without the need for an artificial reporter gene. Specifically, the CRISPR/CAS system can be used to recruit enzymes to deposit histone marks of interest directly to the nucleosomes at GR binding sites. If GR binding is subsequently reduced, one has a direct read-out of GRs response to histone modifications and test for a causative connection to histone marks.

8.2 What defines a transcription factor binding site? – Sequences responsible for GR recruitment to individual genomic loci

8.2.1 GR binding at canonical and non-canonical motifs

To orchestrate transcription, glucocorticoid receptor (GR) interacts with several other proteins to bind to regulatory DNA-elements (Stöcklin et al. 1996; Zhang et al. 1997; Wakui et al. 1997; Chang et al. 1998; Boruk et al. 1998; Almlöf et al. 1998; Lerner et al. 2003). Concordantly, we detected enrichment of the known GR consensus recognition sequence as well as motifs for several co-regulating factors in our GR-chromatin immunoprecipitations (ChIP) in several cell lines.

For example, we identified Fox and AP1 motifs besides the canonical GR motif (compare Figure 21 and Figure 23, pages 77 and 79).

Fox proteins are well-known co-binding factors of the steroid receptor family (Tuteja & Kaestner 2007). Besides their function as independent TFs, Fox proteins are pioneering factors making the chromatin landscape accessible to other DNA binding proteins (Zaret & Carroll 2011). This pioneering function might explain the presence of the Fox motif in vicinity of GR binding sites, and the resulting enrichment in our GR-ChIP-Seq analysis.

Activating protein 1 (AP1) is a heterodimeric complex consisting of members of the Jun and Fos protein families, (Hess et al. 2004) which binds to response DNA elements, thus enabling transcriptional regulation. GR is able to interact with AP1 (Ray & Prefontaine 1994) using its DNA binding domain (DBD). The GR-DBD was shown to be responsible for transactivation as well as transrepression (Heck et al. 1994; Liden et al. 1997; Tao et al. 2001), explaining the importance of the GR-DBD for our further experiments (e.g. EMSA and structural alignments). AP1's function as transcription factor is repressed by direct protein-protein interaction with hormone-activated GR monomer (Yang-Yen et al. 1990). Besides the well-established direct interaction of GR and AP1, which results for example in transrepression (Lucibello et al. 1990; Teurich & Angel 1995), the presence of AP1 motifs close to GR binding sites could be linked to the pioneering function of AP1, maintaining genomic GR binding sites accessible (Biddie et al. 2011). This link could be the basis for the correlation between AP1 sequences, "open" chromatin and GR binding (John et al. 2011).

If and how these and other co-enriched motifs are involved in the GR recruitment to the DNA cannot be answered without additional experiments. Due to their relatively short lengths, DNA binding motifs of e.g. TFs are often found by

chance and it is therefore unclear to draw conclusions regarding the reason for enrichment at an individual locus. Here we analyzed GR binding in several cell lines and in agreement with other studies we found a cell-type-specific binding pattern of GR, with an additional specificity: GR binding to its canonical motif is less frequent in certain cell lines (Figure 22, page 78).

A possible explanation for the large amount of GR-bound loci containing non-canonical motifs could be that GR binds degenerated motifs. Another likely explanation is, that GR interacts with other proteins (Figure 55), which recruit GR to additional loci besides the ones holding only a canonical GR consensus sequence (Figure 55 B and C), or a canonical GR consensus sequence in composition to a second TF binding sequence (Figure 55 D). These additional loci could be due to combinatorial binding (Figure 55 E), exhibiting part of a GR-DNA interaction and an additional motif, belonging to the partnering protein. Also tethered binding (Figure 55 E) would lead to enriched GR binding sites in GR-ChIP-experiments and reveals the binding sequence of the protein is tethered to.

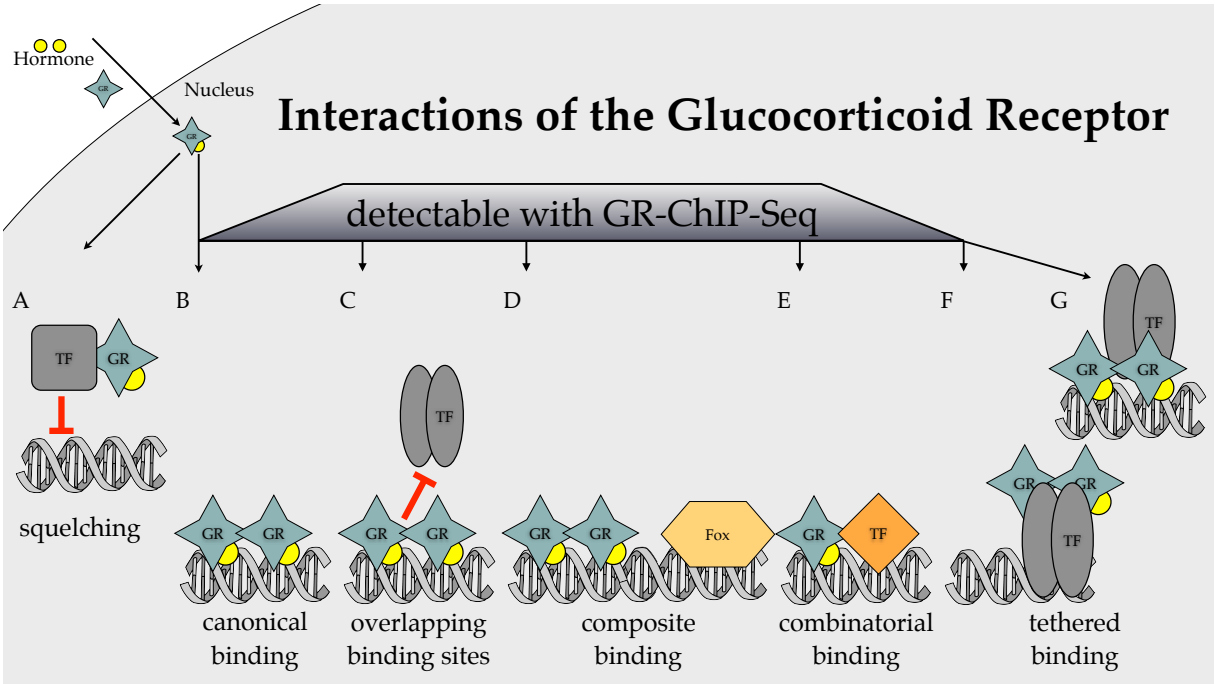

Figure 55: Overview of GR interactions. Hormone activated GR interacts with other proteins to prevent them from (A) binding their DNA consensus sequence, a process called transrepression (e.g. GR and AP1 or NF-κB). Besides transrepression, GR is able to bind to DNA at canonical GR binding sequences (GBS) (B), and GBSs overlapping to recognition sequences of other TFs (C). These additional TF binding sequences can also by in composition to a GBS (D). Earlier studies (Wu & Bresnick 2007) found rare binding events of a GR dimerizing with another TF (e.g. AP1) to bind non-canonical GR binding sequences (E). GR and can also be tethered to other TFs (F), or other TFs are tethered to GR (G). GR binding events B till F are all leading to ChIP-Seq signals, but if and how these non-canonical GBS found at composite (D), combinatorial (E) or tethered (F) GR binding events are involved in GR recruitment can not fully answered, since theses short DNA sequences are frequently found by chance.

Simple motif-finding approaches in classical ChIP-Seq cannot distinguish between GR binding to degenerative motifs (including combinatorial / overlapping) or if GR is tethered to the DNA at other motifs. GR would be bound in all cases, thus resulting in ChIP-Seq peaks. To discriminate between these variants, we worked in a multidisciplinary approach to better understand the mechanisms behind genomic GR binding by combining diverse experimental techniques with computational biology. We employed a technique allowing one to identify protein-DNA interactions with base pair resolution: ChIP-Exo (Rhee & Pugh 2011). ChIP-Exo trims the sequenced fragments for NGS from around 300-400 bp down to around 30-40 bp of sequence that is protected from digestion by the cross-linked protein.

In contrast to other studies, we only found a partial overlap of ChIP-Seq and ChIP-Exo peaks and vice-versa. The higher sensitivity of the ChIP-Exo procedure (Rhee & Pugh 2011) could be one reason for this discrepancy of peaks, or it could be a consequence of the missing controls in ChIP-Exo. An alternative explanation for the partial overlap of peaks in ChIP-Seq and ChIP-Exo data are the different peak calling algorithms which are needed to identify peaks in ChIP-Seq and peak pairs in ChIP-Exo data. Input or mock IgG were used to control for biases in classical ChIP-Seq and non-specific peaks were subsequently filtered out of the data-set, like the on position chr17:22020400-22021224 at hg19 assembly, which is found in the ChIP-Exo as well as in the input control of ChIP-Seq (Figure 56).

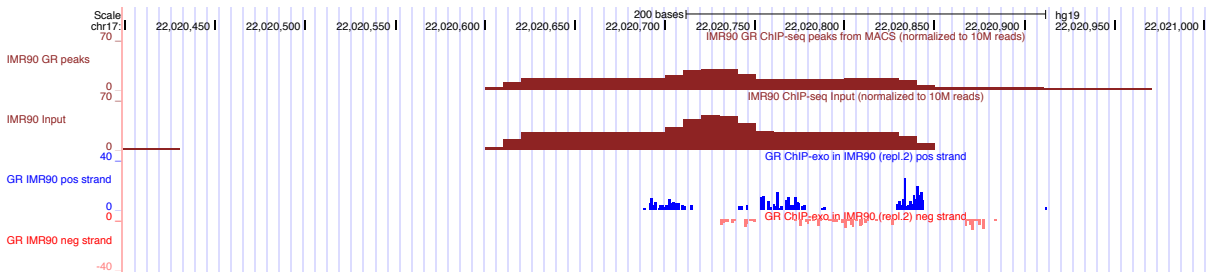

Figure 56: Genome browser screenshot showing a signal in GR-ChIP-Seq and ChIP-Exo, which has to be filtered out because of its appearance in input control.

Our priority was to identify which sequences are responsible for binding at individual loci and we thus opted to focus our analysis on GR-bound regions identified by both methods as these are most likely reflecting real binding events. We (in collaboration with Morgane Thomas Chollier and Jonas Ibn-Salem) developed a motif-based approach of the 5' ChIP-Exo coverage to merge the reliable peak calling from classical ChIP-Seq with the variable and complex ChIP-Exo footprints. The resulting footprints and associated plots are not novel (Rhee & Pugh 2011), but the

systematical test of hundreds of motifs enriched in classical ChIP-Seq is expanding the analysis beyond just analyzing the known consensus sequence of the TF of interest.

This approach uncovered a striking footprint for the canonical GR binding motif for several cell lines (compare Figure 26, page 83). Our ChIP-Exo data combined with the ExoProfiler was yielding *in vivo* structural insights, for example since the canonical GR-footprint profile consists of two mirror symmetric peak pairs, each for one GR monomer. This indicates the dimeric GR binding to inverted repeats separated by a 3 bp spacer *in vivo*, meaning that ChIP-Exo data gives structural insights, which can be found with our ExoProfiler. Since direct GR interaction leads to GR-specific mirror symmetric footprint profiles, tethered GR binding should be distinguishable. We therefore analyzed additional motifs to uncover possible GR interactions on the DNA to find an answer to the question: Are motifs enriched in classical GR-ChIP-Seq peaks indeed involved in GR recruitment or is GR directly bound to these or other sequences?

As our analysis was aimed at understanding GR binding at non-canonical DNA sequences, we analyzed footprint profiles in the cell line with the lowest fraction of GR-ChIP-Seq peaks with a high-scoring motif match, K562 (Figure 22, page 78). Since no clear footprint profile in K562 cells for degenerate GBS-like sequences was found, this indicates that other non GBS-like sequences might be responsible for directing GR to the chromatin in K562 cells. For IMR90 and U2OS cells the question if the apparent lack of canonical motifs means that GR is recruited to more degenerative motifs could at least in part be answered with ChIP-Exo and the ExoProfiler. In both cells we could indeed show that GR binds to GBS-like sequences, as motifs with higher p-values still lead to the distinct mirror-image-like footprint profile of GR binding. In contrast, GR binding in K562 cells is limited to GR motif-matching sequences with low p-values, as the ExoProfiler did not show the striking footprint profile for direct GR binding for GR motif-matching sequences with higher p-values.

Second, we analyzed enriched motifs (apart from GR) in our GR-ChIP-Exo data starting with the Fox binding motif, which was also enriched in our GR-ChIP-Seq peaks and provides information regarding sequences responsible for recruiting TFs to specific loci. To further understand the role of for example Fox motifs enriched in GR-ChIP-Seq peaks, I tested the role of these sequences in mediating GR-dependent transcriptional regulation.

8.2.2 Fox footprints at GR-bound regions

The Fox family is known to pave the way for hormone receptor binding, in particular for the estrogen receptor (ER), by opening the DNA at ER response elements (Carroll et al. 2005), which could explain the co-enriched motifs of the Fox family members in ChIP-Seq data of GR. Our analysis of isolated Fox motifs showed no functional transcriptional regulation. Only for genomic regions containing both a GR and a Fox sequence motif we detected GR-dependent transcriptional regulation, which was mostly lost once the Fox or the GR motif was mutated, indicating a tight functional coupling of Fox and GR. The connection between steroid hormone receptors and Fox proteins is perhaps best characterized for the estrogen receptor (ER). Almost all genomic ER-binding events require the presence of FoxA1. Furthermore, if FoxA1 is introduced into cells expressing low levels of FoxA1, a large fraction of previously unbound ER binding sites become bound (Hurtado et al. 2010). Part of this might be explained by FoxA1's role in establishing accessible chromatin (Cirillo et al. 2002), but in contrast to the ER, GR binding is not simply lost upon FoxA1 depletion. GR binding seems to be reprogramed upon FoxA1 depletion with GR relocating to many other genomic loci (Sahu et al. 2013). This indicates a possible tethering mechanism of GR by Fox proteins, which has also been suggested for FoxA1/FoxA2 and the androgen receptor (Yu et al. 2005), a structurally related hormone receptor having an almost identical DBD as GR. Taken together with our findings, we found indications for a dual role of Fox proteins in GR-dependent transcriptional regulation.

First, a pioneering function is indicated by the mutual dependence of the GR and the Fox motifs to activate transcription using luciferase reporter assays. This pioneering mechanism of Fox could indeed be working in plasmid structures of our reporter assay, since plasmids do acquire nucleoprotein structure after transfection into mammalian cells, by complexing with histones *in vitro* (Jeong et al. 1991; Gracey et al. 2010)

The second finding is an indication for Fox as a factor that tethers GR to the genome. We find footprint profiles at Fox binding sequences in our GR-ChIP-Exo data which are clearly distinguishable from the footprint profiles found at canonical GR sequences. These different footprints might reflect GR being tethered to the DNA by Fox proteins, which was also indicated in other studies (Yu et al. 2010). This is also supported by the fact that the footprint profile at Fox motifs is very different from the ones at canonical GR binding sites, indicting that GR is not directly interacting with

the DNA at these binding sites. It could also be shown, that GR and Fox motifs are in close vicinity to each other, thus resulting in co-enrichment of both motifs in ChIP-Seq and could lead to signals in ChIP-Exo (Figure 30, page 89). The interplay of Fox proteins and GR is fairly complex requiring further studies to understand the exact nature of this interaction.

Together, the analysis of the ChIP-Exo footprint profiles for Fox and GR to their canonical sequences argues that these structures reflect the *in vivo* binding of these factors, since the ChIP-Exo signals found at canonical GR binding were very specific for GR and in addition were distinct from footprint profiles at other, e.g. Fox, motifs. This shows that ChIP-Exo footprint profiles are TF-specific signals building the basis to elaborate new testable hypotheses to uncover binding mechanisms.

8.2.3 Stat footprints at GR binding sites

In agreement with recent studies showing a tight functional connection between GR and Stat proteins (Langlais et al. 2012), our ChIP-Exo approach found footprint profiles for Stat sequence motifs at ChIP-Seq peaks where no GBS was found. The specific footprint profile for directly bound GR at canonical GBS allows us to discriminate between direct and indirect GR binding. Like the footprint profiles at Fox binding sites, the lack of double peak structure of footprint profiles at Stat binding sites indicates that GR is not directly bound to these sequences. Instead, these footprint profiles might reflect Stat3-dependent tethering of GR to the DNA. GR-Stat3 tethering was shown to be transcriptional regulative, as Stat3 induced genes were repressed, once GR was hormone activated (Langlais et al. 2012). This down regulation of genes with GR-ChIP-Seq peaks containing only a Stat motif could be recapitulated in IMR90 cells and further support the idea that this footprint profile reflects tethered GR binding via Stat proteins.

8.2.4 The role of AP1 at GR binding sites

For the AP1 binding motif we did not find a discernible footprint profile. This could indicate that the well studied functional connection between GR and AP1 (Biddie et al. 2011) might depend on AP1's role in providing access to the chromatin, rather then a tethering mechanism of GR.

8.2.5 *New insights into GR binding: Combi footprints at GR binding sites*

The accumulated reads at multiple loci sharing the motif consisting of a GR halfsite and additional bases (AGAACATTCCA) resulted in a footprint profile, which looks similar to GR binding for the AGAACA part (one GR monomer), but does not resemble this binding profile for the second part (TTCCA) (Figure 57). Our motif analysis and footprint profiles showed only a partial overlap of the *de novo* combi and the canonical GR motif. Whereas the footprint profile at the first halfsite of the combi motif is matching to the footprint profile of the first GR monomer at canonical GR binding sites (Figure 57, peaks 1 and 2), the ChIP-Exo signal in the footprint profile of the second part is incompatible to a second GR monomer (Figure 57, peaks 3 and 4). We therefore hypothesize that GR binds as a monomer and heterodimerizes, or co-binds, with protein other than GR, as speculated earlier (Kumar & Thompson 1999).

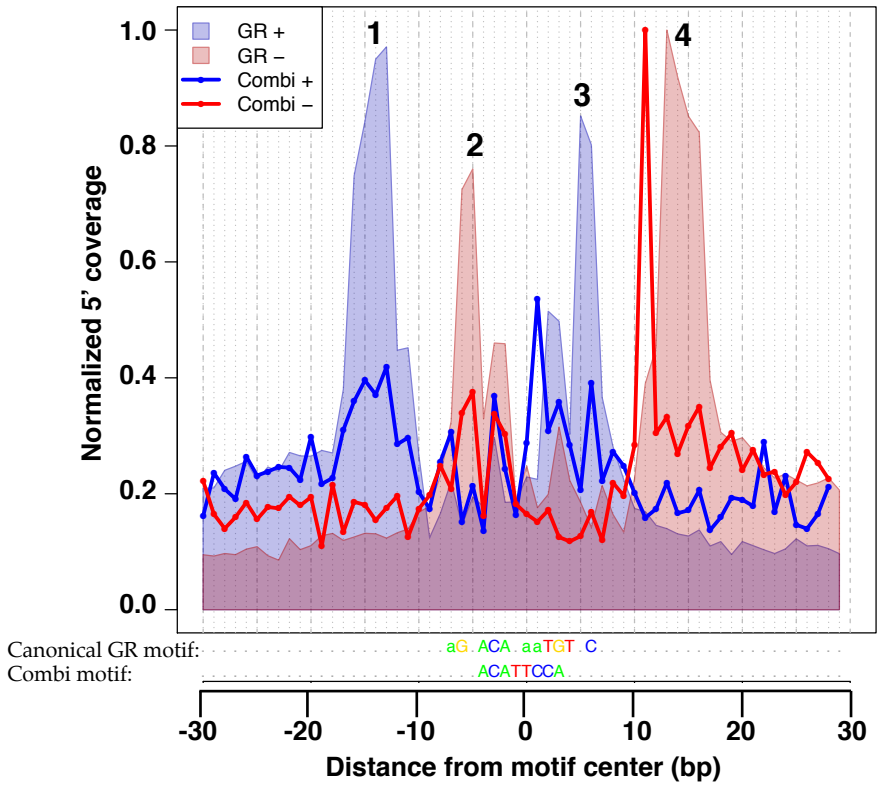

Figure 57: Overlay of the two footprint profiles obtained at canonical GR motifs (area) and at the combi motif (lines). For the canonical GR motif each monomer yields an inner pair of peaks (peaks 1+2 and 3+4) from each DNA-strand, because of imperfect formaldehyde crosslinking. The combi motif is matching to the first peak pair (1+2), but the profile obtained for the second part of the combi motif does not overlap with the one of the canonical motif (peaks 3+4). This indicates that GR interactions to combi motifs is based on heterodimeric binding of GR. Plot was generated in collaboration with Jonas Ibn-Salem.

Consistent with the speculated monomeric GR binding, a recent study suggests that GR is indeed able to bind as a monomer *in vivo* (Schiller et al. 2014). This is similar to our EMSA results with isolated GR-DBD, indicating that GR binds to the isolated AGAACATTCCA motif as a monomer. At these isolated motifs as well as for luciferase reporters harboring the motif in its genomic sequence context, GR is able to induce transcription, which is lost once the TTCC or the GR-monomer part is mutated. The lost transcriptional activity due to the mutations in the TTCC part of the combi motif is indicating a second protein, other than GR, that might bind to this motif.

Possible heterodimerization partners of GR are DNA-binding proteins, whose consensus motif matches the TTCCA part. A JASPAR database search revealed a whole set of possible candidates partnering the GR at the combi motif: the ELK protein, the ETS family, RelA, the STAT family and the four TEAD proteins. Our structural alignment of all mentioned candidates with a GR monomer showed no structural hindrance for both a GR monomer and the candidate to be present simultaneously at DNA sequences containing the combi motif.

The dsiRNA-based knockdown experiments indicate a functional connection of GR and ETS2, RelA, Stat3 as well as TEAD3 and TEAD4. The hormone induced transcriptional activity of a luciferase reporter containing the combi motif was specifically reduced upon knockdown of the aforementioned candidates, compared to the unaffected control vector containing the canonical GBSs. These five proteins are therefore candidates for the hypothesized GR heterodimerization partner.

The ETS superfamily, is defined by their shared helix-turn-helix DNA binding domain (Petersen et al. 1995; Jonsen & Petersen 1996) and is sub-divided in 12 families. ETS1 and ETS2 proteins bind to sequence motifs containing TTCCA and are both known GR interaction partners (Espinás et al. 1994; Mullick et al. 2001) and therefore corroborate the idea of a combined interaction with GR. ETS factors have been shown to act as transcriptional repressors and activators (Sharrocks 2001).

RelA also known as the p65 subunit is part of the nuclear factor, called NF-κB complex that regulates transcription (Chen & Greene 2004). NF-κB is a known target for the transrepression by GR (Ray & Prefontaine 1994). GR-NF-κB interaction leads to hampered NF-κB-DNA binding, resulting in a repression of inflammatory genes like IL6 by glucocorticoids (Ray & Prefontaine 1994).

The known GR interaction partner Stat3 (Zhang et al. 1997; Lerner et al. 2003) is also a reasonable candidate for the hypothesized GR-heterodimer. Previous studies revealed a wide range of GR-Stat3 interactions (Langlais et al. 2012), each resulting in a different transcriptional outcome: If GR is tethered to Stat3 (cluster 4), Stat3 induced genes are repressed, once Stat3 is tethered to GR (cluster 6), one can see a synergistic effect of GR induced genes. If GR binds in composition with Stat3 (cluster 5), genes were up regulated, but only if both factors are bound to their individual consensus motifs. They show an overlapping motif, where a GR- and a STAT-binding site were separated by 8 bp. However, the motif identified by the Langlais group contains a full GR binding motif, and is therefore different to the combi motif identified with our ExoProfiler, consisting only of a GR halfsite.

TEAD is a group of four proteins sharing the a TEA DNA binding domain (Bürglin 1991) and consisting of TEAD1 (TEA domain factor 1 or SV40 transcriptional enhancer factor 1 (TEF-1)), TEAD2, TEAD3 (also known as TEAD5) and TEAD4. All are derived from different genes and bind to the same DNA sequence (Kaneko & DePamphilis 1998). The four proteins have an ~80 % similarity concerning their amino acid sequence and additionally bind the same transcriptional co-activators (Vassilev et al. 2001).

Our pulldown assay showed that from all the above-mentioned candidates matching the combi motif, TEAD3 is specifically pulled down at sequences containing the combi motif, compared to scrambled and mutated motif. We did not detect the other candidate proteins, however this does not mean that they are not active at the combi motif as their absence might reflect technical limitations of the pull-down assay. Besides GR (NR3C1) itself and TEAD3, PurB was also enriched. PurB is a sequence-specific DNA-binding protein linked to both replication and transcriptional processes. Since PurB binds to purine rich elements (guanine and adenine) it is likely that the TTCCA (especially the reverse TGGAA) of the combi motif is attracting PurB to bind in the pulldown assay.

As we found indications that GR is able to bind DNA by partnering with other proteins at the combi motif by a rarely studied mechanism for GR-DNA interactions, called combinatorial binding (Wu & Bresnick 2007), I will have a closer look to protein:protein interaction with a focus on GR based interactions.

One distinguishes general or unspecific and specific protein-protein interactions. General protein-protein interactions are found for example in

chaperones, which help to maintain the structure of other proteins especially under stress conditions like heat or changed pH (Ellis & van der Vies 1991; Grad & Picard 2007; Tapley et al. 2010). One example for unspecific protein-protein interactions is HdeA, a pH-regulated chaperone, which resides in a dimeric form in the periplasm of *E. coli* (Richard & Foster 2003). Upon low pH, HdeA dissociates into two monomers (Gajiwala & Burley 2000), which start to interact with the wide range of periplasmatic proteins to protect them from losing structure upon exposure to low pH. These unspecific interactions are caused by HdeA's exposed hydrophobic core, which is attracted to unusual hydrophobic sites of miss- and unfolded proteins (Xu et al. 1997; Hong et al. 2005).

Protein-protein interactions can also be specialized and involve very specific protein domains (Lock and key model, (Koshland 1958)), for example the GR-specific transcriptional repression of the pro-inflammatory protein IL6, at which GRs DNA binding domain (GR-DBD) is interacting with p65 to repress IL-6 expression (Ray & Prefontaine 1994). Other examples for specialized protein-protein interactions are the heat shock proteins (HSPs) Hsp90 and Hsp70, which interact with GRs ligand binding domain (GR-LBD). Both HSPs prevent the aggregation of non-hormone activated GR and maintain GR's high affinity state to its activating ligand cortisol (Dittmar & Pratt 1997). GR-LBD also serves as co-regulator binding site (Pfaff & Fletterick 2010). A third specialized GR interaction is the well understood homodimerization (Kumar & Thompson 1999) of two GR monomers at their dimer interface, involving the second zinc finger of the GR-DBD, the so-called D-box (Kumar & Thompson 1999). Interactions with c-Jun, a part of the TF AP1, are also performed using GR's DBD (Touray et al. 1991).

Summarizing the protein:protein interaction of GR: GR alone serves as a hub for several GR interacting proteins. These interactions are conducted via specialized domains of the GR, each interacting with a very specific set of other proteins. In addition, heterodimeric interaction of DNA-bound GR with other transcription factors was speculated to be present and expected to be found at genomic sequences containing only a GR halfsite (Kumar & Thompson 1999). One example for a GR heterodimer was found with AP1 at a genomic sequence consisting of such a speculated GR halfsite and additional bases coding for an AP1 binding site (Wu & Bresnick 2007). Our *de novo* identified combi motif consists of a GR halfsite, but the additional bases are not compatible to AP1 consensus motif, so far the only known GR heterodimerization partner. Instead our sequence is compatible with proteins

binding to the motif ACATTCCA (Figure 58). It is possible that several proteins, which share this specific DNA binding domain to bind to the ACATTCCA motif, also share a protein-protein interaction site, which enables them to interact with GR. These two requirements would ensure GR-specific interactions at the combi motif and accounts for multiple proteins that can bind at the combi motif together with GR, as we found multiple candidates in our knockdown experiments. The transcriptional activity of the combi motif is not only lost due to mutations of the sequence itself. This activity is specifically lost upon knockdown of candidate proteins, which share the TTCC consensus sequences present in the second part of the combi motif (AGAACATTCCA). Namely ETS2, TEAD3 and TEAD4 were confirmed in this dsiRNA-based approach to have the strongest effects, in congruence with our findings from the pulldown assays identifying TEAD3 as possible GR binding partner at the combi motif (Figure 58).

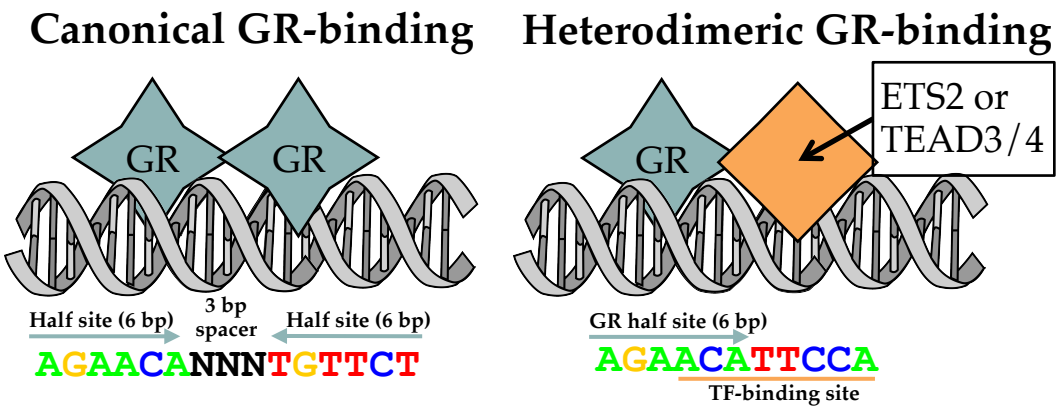

Figure 58: Comparison of homodimeric GR binding (left) at canonical GR binding sequence (GBS) and heterodimeric binding (right) of GR and partnering protein at GBSs containing a combi motif. Both binding mechanisms lead to transcriptional activation.

As stated above, the results of our knockdown experiments indicate that multiple proteins are able to function as GR partner. This could be explained by earlier findings, which show GRs ability to interact via its LBD with different proteins (e.g. Hsp90 or a second GR, (Trebble et al. 2013)). If the partnering protein has to fulfill the two functions, first binding to the ACATTCCA motif and second, to be able to interact with GR, multiple candidate proteins might have structural similarities. Of course these interactions could be indirect as well, as GR-partnering proteins could make the binding sites accessible for GR, like Fox, whose binding sites are enriched at GR-bound regions and might potentiate chromatin accessibility (Cirillo et al. 2002) at GR binding sites. In addition, the partnering protein could be

non-ubiquitously expressed to enable cell-type-specific GR binding and transcriptional regulation of motif associated target genes.

To investigate a potential role in cell-type-specific regulation, my knockdown experiments should be expanded to multiple cell lines. Unfortunately however, I could not perform these experiments in IMR90 cells, as dsiRNA treatment of IMR90 cells lead to loss of transcriptional activation of all tested luciferase reporters, even with the nonspecific control dsiRNA.

We identified a new motif that can also be bound by GR and can thus explain GR-peaks without an obvious GBS. Expanding the search for other combinatorial motifs to additional cell lines might uncover additional proteins, which heterodimerize with GR. This newly identified mechanism of GR-dependent transcriptional regulation could at some point help to minimize adverse effects of glucocorticoids, when applied as cancer treatment or for immunosuppression (Harousseau et al. 2006). Even after decades of intensive research on GR binding, the part of GR heterodimerizing with other proteins to interact with combinatorial binding sites is not well understood, as indicated by the low numbers of found examples (e.g. AP1, (Wu & Bresnick 2007)). The number of adverse effects on the contrary is fairly big, based on the nonselective application of glucocorticoid drugs with tissues affected throughout the body. Among the adverse effects are neural implications, like memory and attention deficits (Keenan et al. 1996) as well as immunodeficiency (Klein et al. 2001). If one could target cell-type-specific GR-cofactors, or at least have a deeper understanding of the transcriptional regulatory network of GR, adverse effects could be explained and ultimately possibly minimized.

8.2.6 GR interactions revealed by footprint profiles

Our combined approach to distinguish GR binding to degenerative motifs from GR binding to non-canonical motifs was facilitated by the base pair resolution of ChIP-Exo, which was also be shown by other studies (Rhee & Pugh 2012; Mendenhall & Bernstein 2012). The novelty of our approach is: 1. We expanded our analysis to motifs other than the known consensus motif. 2. We show that ChIP-Exo footprints are protein-specific signatures. 3. Based on the peak shape, peak number and relative position of peak pairs we can distinguish between different modes of GR binding, as exemplified for a combinatorial motif in our ChIP-Exo data. 4. We could furthermore see protein-specific footprint profiles for degenerative GBS-like motifs indicating that GR binds to such sequences even though they are found at roughly equal frequencies in bound and unbound regions based on ChIP-Seq.

With this approach we found indications that GR is able to bind DNA by partnering with other proteins at a *de novo* identified DNA sequence, called combi motif, where both proteins could interact with each other and are both bound to the DNA. The identified *de novo* motif consists of a GR halfsite and an additional TTCCA sequence. Our initial findings showed sequences containing this TTCCA part and other sequences besides the canonical GR sequence to be enriched in our GR-ChIP-Seq analysis (Figure 59).

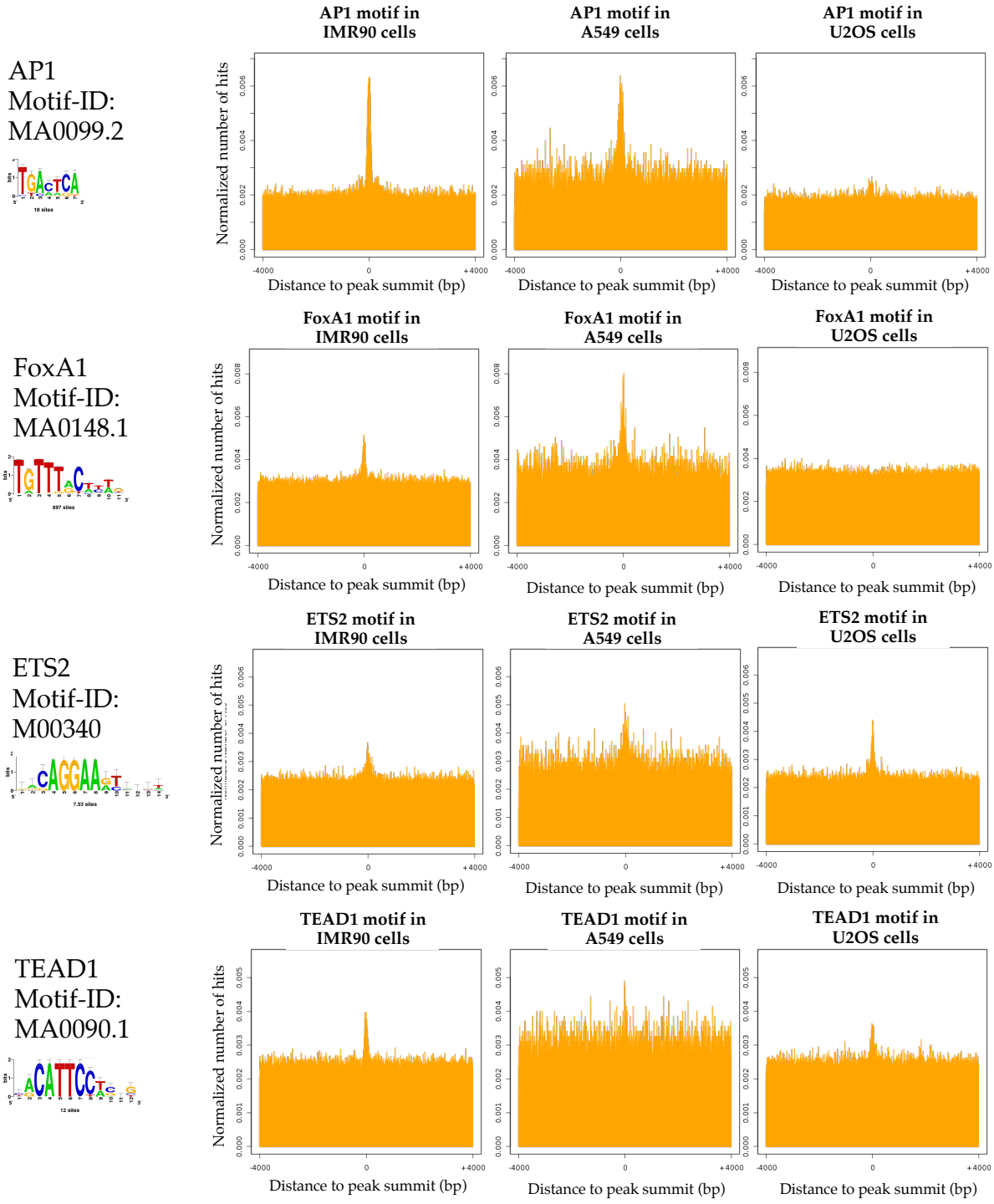

Figure 59: Motif distributions at GR-bound loci for non-GR motifs. Different Motifs are enriched at GR peaks for different cell lines. DNA sequences +/- 4000 base pairs (bp) centered on the GR-ChIP-Seq peak summits were aligned and sub-divided into 50 bp bins. Alignment scores for matching motifs of the PWM of AP1, FoxA1, ETS2 or TEAD1 (from TRANSFAC, at a P-value of 10^-3 used for cut-off) in every bin were assigned to calculate the relative frequency distribution of AP1, FoxA1, ETS2 or TEAD1 (JASPAR, (Sandelin et al. 2004)) motifs around GR-ChIP-Seq peaks. The orange plot shows the normalized number of hits for the appropriate PWM relative to the peak center on position 0. A549 GR-ChIP-Seq was performed by the Yamamoto lab. Analysis was done in collaboration with Dr. Morgane Thomas-Chollier.

If GR or another protein is actually bound to these additional sequences could not be answered. As stated above, the ChIP-Exo footprint profiles allowed this

discrimination between different GR binding modes (compare Figure 60 A and C) and if GR is actually bound with other proteins to non-canonical sequences (compare Figure 60 B and D). Interestingly, and up until now not shown by others, we found that proteins other than the immunoprecipitated factor can lead to ChIP-Exo signals. These non-GR footprints additional to GR's dimeric one (Figure 60 A) are present at, for example, composite binding sites. Composite binding sites are harboring full binding sites for a protein and GR (Figure 60 B), which can be bound at both sides of the e.g. Fox protein. Depending on the orientation the merged signals accumulate for both crosslink variants leading to a footprint profile. This explains the enriched Fox sequences initially found in our GR-ChIP-Seq analysis.

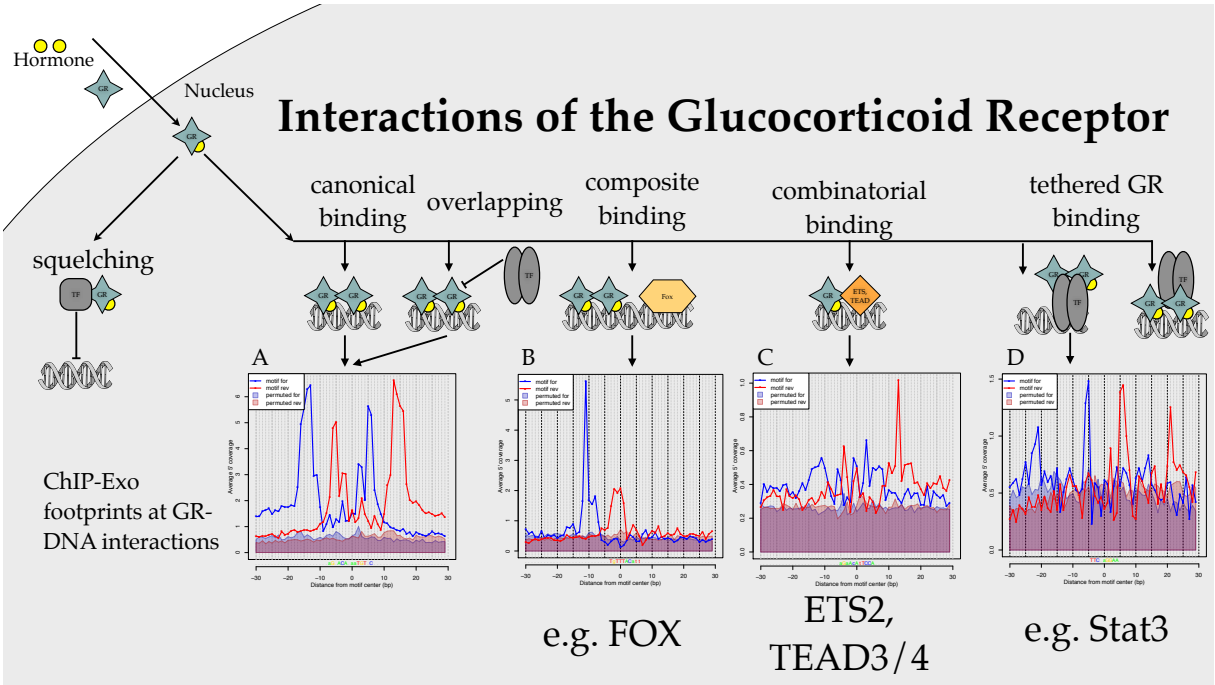

Figure 60: Overview of possible GR interactions with the genome including detailed information regarding the bound GR-bound sequences. ChIP-Exo was used to generate footprint profiles of homodimeric GR binding to its canonical motif (A), were each GR monomer leads to a peak pair. We were also able to identify footprint profiles at composite binding sites of GR and e.g. Fox proteins (B). Depending on the orientation of both proteins, the merged signals accumulate for both crosslink variants leading to a footprint profile and reveal the bound sequence of the non-GR protein. The novel finding is the rare binding event of a GR monomer heterodimerizing with a second protein at combinatorial binding sites (3). Our footprint profile indicates two different binding partners at corresponding motifs if compared to the profile found at canonical GR binding sites (1). GR could also be tethered via e.g. Stat3 (4), leading to different footprint profiles and revealing the bound sequence of the GR tethering protein. Footprint profiles were generated using our ExoProfiler.

When GR is bound at combinatorial sequences consisting of a GR halfsite and an additional motif, GR might heterodimerize with a second protein to bind there, which leads to specifically assignable ChIP-Exo signals (Figure 60 C). Similarly, for tethering events: GR is bound to the DNA via another protein (Figure 60 D), thus

leading to footprint profiles at sequences bound by other proteins e.g. Stat3. This exemplifies the broad protein interactions of GR (Figure 60), which all lead to specific and distinguishable footprint profiles.

Together, besides the known and well-studied mechanisms of GR-tethering (e.g. with Stat3) and composite binding (e.g. with AP1) (Kassel & Herrlich 2007), we found indications for a less common mechanism for GR-DNA interactions, called combinatorial binding (Wu & Bresnick 2007), meaning both proteins could interact with each other and are both bound to the DNA. Only the combination of computational and wet-lab approaches, allowed us to conclude that the GR is able to bind DNA by partnering with other proteins at the combi motif.

An improvement to our approach would be the direct generation of footprints at single loci. So far we were successfully generating footprint profiles by merging ChIP-Exo signals of multiple loci, like for the combi and STAT motifs, to get a convincing signal. Perhaps this could be resolved by deeper sequencing of ChIP-Exo material. Alternatively, better cross-linking reagents and conditions might improve the coverage and resolution at individual loci. This would allow studying single binding events to determine if protein interactions are indeed based on two or more proteins at a specific binding site. This is of clear interest as the revolution in genomic editing by the CRISPR/CAS system (Hale et al. 2012) allows one to study the role of such sequences at individual loci *in vivo*, for example to determine if the footprint profiles changes when the sequence at a genomic locus is deleted or changed.

Our ExoProfiler can be applied to ChIP-Exo data for any TF, which might help uncover how they bind DNA as they also often lack a clear consensus sequence at regions that are enriched in ChIP-Seq data (Worsley Hunt et al. 2014), explaining their binding and leading to the very same question: Is the TF of interest tethered to the DNA or binding to degenerative motifs? Similar to the approach described in my thesis, for motifs that are enriched at GR-bound regions, detectable footprint profiles can give clues about their role in recruiting the TF to the genome. For potential partners, I envision that in the future, a collection of ChIP-Exo footprint profiles for many TFs might facilitate the identification of candidate proteins responsible for such profiles based on similarities in the profile, as we show here for the Fox motif.

In summary our wet-lab and computational analysis to address the emerging question what defines a TF binding site and what sequences are responsible for GR recruitment in particular, yielded new insights regarding GR binding. The absence of

classical recognition sequences in GR-bound regions can be explained by direct GR binding to a broader spectrum of sequences than previously known: as classical GR-homodimer, as heterodimer together with a member of the ETS or TEAD families of TFs and by indirect recruitment via FOX or STAT proteins. This was only possible by our unique interpretation of footprint profiles generated from high-resolution ChIP-Exo data: footprint profiles are protein and recognition-sequence-specific signatures of TF binding sites that allow the discrimination of direct and indirect (tethering to other DNA-bound proteins) GR binding. Thereby ChIP-Exo footprints reveal which specific sequences of regions enriched by classical ChIP are actually protein bound and if this protein is the TF itself since ChIP-Exo captures information about TFs other than the one targeted by the antibody.

Together, our generically applicable footprint-based approach uncovers new structural and functional insights into the diverse ways of genomic cooperation and association of TFs.

9 Bibliography

Aasland, R., 1995. The PHD finger: Implications for chromatin-mediated transcriptional regulation. *Trends in Biochemical Sciences*, 20(2), pp.56–59.

Albert, I. et al., 2008. GeneTrack--a genomic data processing and visualization framework. *Bioinformatics (Oxford, England)*, 24(10), pp.1305–6.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P., 2006. The Molecular Biology of the Cell, 4th Edition.

De Almeida, S.F. & Carmo-Fonseca, M., 2014. Reciprocal regulatory links between cotranscriptional splicing and chromatin. *Seminars in cell & developmental biology*, 32, pp.2–10.

Almlöf, T. et al., 1998. Role of important hydrophobic amino acids in the interaction between the glucocorticoid receptor tau 1-core activation domain and target factors. *Biochemistry*, 37(26), pp.9586–94.

Anbanandam, A. et al., 2006. Insights into transcription enhancer factor 1 (TEF-1) activity from the solution structure of the TEA domain. *Proceedings of the National Academy of Sciences of the United States of America*, 103(46), pp.17225–30.

Archer, T. & Cordingley, M., 1991. Transcription Factor Access Is Mediated by Accurately Positioned Nucleosomes on the Mouse Mammary Tumor Virus Promoter. *Molecular and cellular …*, 11(2), pp.688–698.

Arvey, A. et al., 2012. Sequence and chromatin determinants of cell-type-specific transcription factor binding. *Genome research*, 22(9), pp.1723–34.

Ayroldi, E. et al., 2007. GILZ mediates the antiproliferative activity of glucocorticoids by negative regulation of Ras signaling. , 117(6), pp.1605–1615.

Babu, M.M. et al., 2004. Structure and evolution of transcriptional regulatory networks. *Current opinion in structural biology*, 14(3), pp.283–91.

Bai, L. & Morozov, A. V, 2010. Gene regulation by nucleosome positioning. *Trends in genetics : TIG*, 26(11), pp.476–83.

Bannister, A.J. & Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell research*, 21(3), pp.381–95.

Barboric, M. et al., 2001. NF-kB Binds P-TEFb to Stimulate Transcriptional Elongation by RNA Polymerase II. *Molecular Cell*, 8, pp.327–337.

Bardet, A.F. et al., 2013. Identification of transcription factor binding sites from ChIP-seq data at high resolution. *Bioinformatics (Oxford, England)*, 29(21), pp.2705–13.

Bargaje, R. et al., 2012. Proximity of H2A.Z containing nucleosome to the transcription start site influences gene expression levels in the mammalian liver and brain. *Nucleic acids research*, 40(18), pp.8965–78.

Barski, A. et al., 2007. High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4), pp.823–37.

Vom Baur, E. et al., 1996. Differential ligand-dependent interactions between the AF-2 activating domain of nuclear receptors and the putative transcriptional intermediary factors mSUG1 and TIF1. *The EMBO journal*, 15, pp.110–124.

Beato, M., 1989. Gene regulation by steroid hormones. *Cell*, 56(3), pp.335–44.

Becker, P., Renkawitz, R. & Schütz, G., 1984. Tissue-specific DNaseI hypersensitive sites in the 5'-flanking sequences of the tryptophan oxygenase and the tyrosine aminotransferase genes. *The EMBO journal*, 3(9), pp.2015–20.

Becker, P.B. et al., 1986. In vivo protein-DNA interactions in a glucocorticoid response element require the presence of the hormone. *Nature*, 234(18).

Belikov, S. & Åstrand, C., 2004. Chromatin-Mediated Restriction of Nuclear Factor 1 / CTF Binding in a Repressed and Hormone-Activated Promoter In Vivo. *Molecular and cellular biology*, 24(7), pp.3036–3047.

Benevolenskaya, E. V, 2007. Histone H3K4 demethylases are essential in development and differentiation. *Biochemistry and cell biology*, 85(4), pp.435–43.

Bernstein, B.E. et al., 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), pp.57–74.

Bernstein, B.E. et al., 2010. The NIH Roadmap Epigenomics Mapping Consortium. *Nature biotechnology*, 28(10), pp.1045–8.

Biddie, S.C. et al., 2011. Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Molecular cell*, 43(1), pp.145–55.

Bienroth, S., Keller, W. & Wahle, E., 1993. Assembly of a processive polyadenylation complex. *The EMBO journal*, 12(2), pp.585–594.

Boruk, M., Savory, J.G. & Haché, R.J., 1998. AF-2-dependent potentiation of CCAAT enhancer binding protein beta-mediated transcriptional activation by glucocorticoid receptor. *Molecular endocrinology (Baltimore, Md.)*, 12(11), pp.1749–63.

De Bosscher, K. et al., 2008. Selective transrepression versus transactivation mechanisms by glucocorticoid receptor modulators in stress and immune systems. *European journal of pharmacology*, 583(2-3), pp.290–302.

Bromberg, J. & Jr, J.D., 2000. The role of STATs in transcriptional control and their impact on cellular function. *Oncogene*, pp.2468–2473.

Burd, C. & Archer, T., 2013. Chromatin architecture defines the glucocorticoid response. *Molecular and cellular endocrinology*, 380(1-2), pp.25–31.

Bürglin, T.R., 1991. The TEA domain: A novel, highly conserved DNA-binding motif. *Cell*, 66, pp.11–12.

Butler, J.E.F. & Kadonaga, J.T., 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes & development*, 16(20), pp.2583–92.

Carey, M., 1998. The Enhanceosome and Transcriptional Synergy. *Cell*, 92(1), pp.5–8.

Carroll, J.S. et al., 2005. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*, 122(1), pp.33–43.

Carrozza, M.J. et al., 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*, 123(4), pp.581–92.

Cato, A.C. & Wade, E., 1996. Molecular mechanisms of anti-inflammatory action of glucocorticoids. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 18(5), pp.371–8.

Chang, C., Chen, Y. & Lee, S., 1998. Coactivator TIF1b Interacts with Transcription Factor C/EBPb and Glucocorticoid Receptor To Induce a1-Acid Glycoprotein Gene Expression. *Molecular and …*, 18(10), pp.5880–5887.

Chen, L.-F. & Greene, W.C., 2004. Shaping the nuclear action of NF-kappaB. *Nature reviews. Molecular cell biology*, 5, pp.392–401.

Chodankar, R. et al., 2014. Hic-5 is a transcription coregulator that acts before and/or after glucocorticoid receptor genome occupancy in a gene-selective manner. *Proceedings of the National Academy of Sciences of the United States of America*, 111(11), pp.4007–12.

Choukrallah, M.A. & Matthias, P., 2014. The Interplay between Chromatin and Transcription Factor Networks during B Cell Development: Who Pulls the Trigger First? *Frontiers in immunology*, 5(April), p.156.

Cirillo, L.A. et al., 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular cell*, 9(2), pp.279–89.

Colin, J., Libri, D. & Porrua, O., 2011. Cryptic transcription and early termination in the control of gene expression. *Genetics research international*, 2011, p.653494.

Collas, P., 2010. The current state of chromatin immunoprecipitation. *Molecular biotechnology*, 45(1), pp.87–100.

Conaway, R. & Conaway, J., 2013. The Mediator Complex and Transcription Elongation. *Biochimica et Biophysica Acta (BBA)-Gene …*, 1829(1), pp.69–75.

Crawford, G.E. et al., 2006. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome research*, 16(1), pp.123–31.

Creyghton, M.P. et al., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, 107(50), pp.21931–6.

Cvoro, A. et al., 2011. Cross talk between glucocorticoid and estrogen receptors occurs at a subset of proinflammatory genes. *Journal of immunology (Baltimore, Md. : 1950)*, 186(7), pp.4354–60.

Czar, M.J. et al., 2013. The hsp56 immunophilin component of untransformed steroid receptor complexes is localized both to microtubules in the cytoplasm and to the same nonrandom regions within the nucleus as the steroid receptor. *Molecular Endocrinology*, 8(12).

Dahlman-Wright, K. et al., 1990. Protein-Protein Glucocorticoid Interactions Facilitate DNA Binding by the Receptor DNA-binding Domain. *The Journal of biological chemistry*, 265(23), pp.14030–14035.

Van Dijk, M. & Bonvin, A.M.J.J., 2009. 3D-DART: a DNA structure modelling server. *Nucleic acids research*, 37(Web Server issue), pp.W235–9.

Dittmar, K.D. & Pratt, W.B., 1997. Folding of the Glucocorticoid Receptor by the Reconstituted hsp90-based Chaperone Machinery. *Journal of Biological Chemistry*, 272(20), pp.13047–13054.

Dougherty, E.J. et al., 2012. Deducing the temporal order of cofactor function in ligand-regulated gene transcription: theory and experimental verification. *PloS one*, 7(1), p.e30225.

Egea, P.F., Klaholz, B.P. & Moras, D., 2000. Ligand-protein interactions in nuclear receptors of hormones. *FEBS letters*, 476, pp.62–67.

Ellis, R.J. & van der Vies, S.M., 1991. Molecular chaperones. *Annual review of biochemistry*, 60, pp.321–47.

Epshtein, V. et al., 2007. An allosteric path to transcription termination. *Molecular cell*, 28(6), pp.991–1001.

ESCRIVA, H. et al., 1998. Evolution and Diversification of the Nuclear Receptor Superfamilya. *Annals of the New York Academy of Sciences*, 839(1 TRENDS IN COM), pp.143–146.

Espinás, M.L. et al., 1994. Participation of Ets transcription factors in the glucocorticoid response of the rat tyrosine aminotransferase gene. *Molecular and cellular biology*, 14(6), pp.4116–25.

Felsenfeld, G. & Groudine, M., 2003. Controlling the double helix. *Nature*, 421(6921), pp.448–53.

Fu, H. et al., 2013. Methylation of histone H3 on lysine 79 associates with a group of replication origins and helps limit DNA replication once per cell cycle. *PLoS genetics*, 9(6), p.e1003542.

Gajiwala, K.S. & Burley, S.K., 2000. HDEA, a periplasmic protein that supports acid resistance in pathogenic enteric bacteria. *Journal of molecular biology*, 295(3), pp.605–12.

Gao, F., Foat, B.C. & Bussemaker, H.J., 2004. Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. *BMC bioinformatics*, 5, p.31.

Gelman, A. et al., 2003. *Bayesian Data Analysis, Second Edition (Texts in Statistical Science)*,

Gertz, J. et al., 2013. Distinct Properties of Cell-Type-Specific and Shared Transcription Factor Binding Sites. *Molecular Cell*, pp.1–12.

Goldman, S., Ebright, R. & Nickels, B., 2009. Direct detection of abortive RNA transcripts in vivo. *Science*, 324(5929), pp.927–928.

Gracey, L.E. et al., 2010. An in vitro -identified high-affinity nucleosome-positioning signal is capable of transiently positioning a nucleosome in vivo. *Epigenetics & Chromatin*, 3(13), pp.1–7.

Grad, I. & Picard, D., 2007. The glucocorticoid responses are shaped by molecular chaperones. *Molecular and cellular endocrinology*, 275(1-2), pp.2–12.

Grøntved, L. et al., 2013. C/EBP maintains chromatin accessibility in liver and facilitates glucocorticoid receptor recruitment to steroid response elements. *The EMBO journal*, 32(11), pp.1568–83.

Grunstein, M., 1997. Histone acetylation in chromatin structure and transcription. *Nature*, 389, pp.349–352.

Guenther, M., Levine, S. & Boyer, L., 2007. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, 130(1), pp.77–88.

Guo, G. et al., 2010. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Developmental cell*, 18(4), pp.675–85.

Gupta, M. et al., 2003. Single-stranded DNA-binding Proteins PUR-alpha and PUR-beta Bind to a Purine-rich Negative Regulatory Element of the alpha-Myosin Heavy Chain Gene and Control Transcriptional and Translational Regulation of the Gene Expression. *The Journal of biological chemistry*, 278(45), pp.44935–48.

Gupte, R. et al., 2013. Glucocorticoid receptor represses proinflammatory genes at distinct steps of the transcription cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 110(36), pp.14616–21.

Gusmão, E.G. et al., 2014. Detection of Active Transcription Factor Binding Sites with the Combination of DNase Hypersensitivity and Histone Modifications. *Bioinformatics (Oxford, England)*.

Gustafsson, J.A. et al., 1987. Biochemistry, molecular biology, and physiology of the glucocorticoid receptor. *Endocrine reviews*, 8(2), pp.185–234.

Hale, C. et al., 2012. Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Molecular cell*, 45(3), pp.292–302.

Han, M. et al., 1988. Depletion of histone H4 and nucleosomes activates the PH05 gene in Saccharomyces cerevisiae. *The EMBO journal*, 7(7), pp.2221–2228.

Harousseau, J.-L. et al., 2006. Bortezomib plus dexamethasone as induction treatment prior to autologous stem cell transplantation in patients with newly diagnosed multiple myeloma: results of an IFM phase II study. *Haematologica*, 91(11), pp.1498–505.

Harrell, J.M. et al., 2004. Evidence for glucocorticoid receptor transport on microtubules by dynein. *The Journal of biological chemistry*, 279(52), pp.54647–54.

Harrison, S., 1991. A structural taxonomy of DNA binding domains. *Nature*, 353, pp.715–719.

Heck, S. et al., 1994. A distinct modulating domain in glucocorticoid receptor monomers in the repression of activity of the transcription factor AP-1. *The EMBO journal*, 13(17), pp.4087–95.

Heinz, S. et al., 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*, 38(4), pp.576–589.

Van Helden, J., 2003. Regulatory Sequence Analysis Tools. *Nucleic Acids Research*, 31(13), pp.3593–3596.

Van Helden, J., André, B. & Collado-Vides, J., 1998. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *Journal of molecular biology*, 281(5), pp.827–42.

Van Helden, J., del Olmo, M. & Pérez-Ortín, J.E., 2000. Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. *Nucleic acids research*, 28(4), pp.1000–10.

Van Helden, J., Rios, a F. & Collado-Vides, J., 2000. Discovering regulatory elements in non-coding sequences by analysis of spaced dyads. *Nucleic acids research*, 28(8), pp.1808–18.

Henikoff, S. & Shilatifard, A., 2011. Histone modification: cause or cog? *Trends in genetics : TIG*, 27(10), pp.389–96.

Hess, J., Angel, P. & Schorpp-Kistner, M., 2004. AP-1 subunits: quarrel and harmony among siblings. *Journal of cell science*, 117(Pt 25), pp.5965–73.

Hong, W. et al., 2005. Periplasmic protein HdeA exhibits chaperone-like activity exclusively within stomach pH range by transforming into disordered conformation. *The Journal of biological chemistry*, 280(29), pp.27029–34.

Hublitz, P., Albert, M. & Peters, A.H.F.M., 2009. Mechanisms of transcriptional repression by histone lysine methylation. *The International journal of developmental biology*, 53(2-3), pp.335–54.

Hurtado, A. et al., 2010. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nature genetics*, 43(1), pp.27–33.

Iseli, C. et al., 2002. Long-range heterogeneity at the 3′ ends of human mRNAs. *Genome research*, 12(7), pp.1068–74.

Ito, K. et al., 2006. Histone deacetylase 2-mediated deacetylation of the glucocorticoid receptor enables NF-kappaB suppression. *The Journal of experimental medicine*, 203(1), pp.7–13.

Jenuwein, T. & Allis, C.D., 2001. Translating the histone code. *Science (New York, N.Y.)*, 293(5532), pp.1074–80.

Jeong, S., Lauderdale, J.D. & Stein, A., 1991. Chromatin assembly on plasmid DNA in vitro. *Journal of Molecular Biology*, 222(4), pp.1131–1147.

Jia, Y. et al., 2004. A new scoring function and associated statistical significance for structure alignment by CE. *Journal of computational biology : a journal of computational molecular cell biology*, 11(5), pp.787–99.

Jinek, M. et al., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, 337(6096), pp.816–21.

Jitrapakdee, S., 2012. Transcription factors and coactivators controlling nutrient and hormonal regulation of hepatic gluconeogenesis. *The international journal of biochemistry & cell biology*, 44(1), pp.33–45.

John, S. et al., 2011. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature genetics*, 43(3), pp.264–268.

Jonsen, M. & Petersen, J., 1996. Characterization of the cooperative function of inhibitory sequences in Ets-1. *Molecular and cellular biology*, 16(5), pp.2065–2073.

Kaneko, K.J. & DePamphilis, M.L., 1998. Regulation of gene expression at the beginning of mammalian development and the TEAD family of transcription factors. *Developmental genetics*, 22(1), pp.43–55.

Karin, M., 1998. New twists in gene regulation by glucocorticoid receptor: is DNA binding dispensable? *Cell*, 93(4), pp.487–90.

Karlić, R. et al., 2010. Histone modification levels are predictive for gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 107(7), pp.2926–31.

Kasinathan, S., Orsi, G. & Zentner, G., 2014. High-resolution mapping of transcription factor binding sites on native chromatin. *Nature …*, 11(2), pp.203–209.

Kassel, O. & Herrlich, P., 2007. Crosstalk between the glucocorticoid receptor and other transcription factors: molecular aspects. *Molecular and cellular endocrinology*, 275(1-2), pp.13–29.

Keenan, P.A. et al., 1996. The effect on memory of chronic prednisone treatment in patients with systemic disease. *Neurology*, 47(6), pp.1396–402.

Kelleher, R.J., Flanagan, P.M. & Kornberg, R.D., 1990. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell*, 61(7), pp.1209–1215.

Keller, R.W. et al., 2000. The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail. *Journal of molecular biology*, 297(3), pp.569–83.

Kim, T.-M. & Park, P.J., 2010. Advances in analysis of transcriptional regulatory networks. *Wiley interdisciplinary reviews. Systems biology and medicine*.

Kim, T.H. et al., 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, 128(6), pp.1231–45.

Kim, Y.J. et al., 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, 77(4), pp.599–608.

Kimura, H., 2013. Histone modifications for human epigenome analysis. *Journal of human genetics*, 58(7), pp.439–45.

Kino, T. & Chrousos, G., 2011. Acetylation-mediated Epigenetic Regulation of Glucocorticoid Receptor Activity: Circadian Rhythm-associated Alterations of Glucocorticoid Actions in Target Tissues. *Molecular and cellular endocrinology*, 336, pp.23–30.

Klein, N.C., Go, C.H. & Cunha, B.A., 2001. Infections associated with steroid use. *Infectious disease clinics of North America*, 15(2), pp.423–32, viii.

Koch, C., Andrews, R. & Flicek, P., 2007. The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Research*, pp.691–707.

Koleske, A.J.& R.A.Y., 1994. An RNA polymerase II holoenzyme responsive to activators. *Nature*, 368, pp.466–469.

Kornberg, R.D., 2007. The molecular basis of eukaryotic transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 104(32), pp.12955–61.

Koshland, D.E., 1958. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences*, 44(2), pp.98–104.

Kovacs, J.J. et al., 2005. HDAC6 regulates Hsp90 acetylation and chaperone-dependent activation of glucocorticoid receptor. *Molecular cell*, 18(5), pp.601–7.

Kuehner, J.N., Pearson, E.L. & Moore, C., 2011. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature reviews. Molecular cell biology*, 12(5), pp.283–94.

Kumar, R. et al., 1999. Interdomain Signaling in a Two-domain Fragment of the Human Glucocorticoid Receptor. *Journal of Biological Chemistry*, 274(35), pp.24737–24741.

Kumar, R. et al., 2001. The conformation of the glucocorticoid receptor af1/tau1 domain induced by osmolyte binds co-regulatory proteins. *The Journal of biological chemistry*, 276(21), pp.18146–52.

Kumar, R. & Thompson, E.B., 1999. The structure of the nuclear hormone receptors. *Steroids*, 64(5), pp.310–9.

Kumar, S.V. & Wigge, P. a, 2010. H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. *Cell*, 140(1), pp.136–47.

Kummerfeld, S.K. & Teichmann, S. a, 2006. DBD: a transcription factor prediction database. *Nucleic acids research*, 34(Database issue), pp.D74–81.

Kurdistani, S.K., Tavazoie, S. & Grunstein, M., 2004. Mapping global histone acetylation patterns to gene expression. *Cell*, 117, pp.721–733.

Lachner, M. et al., 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, 410, pp.116–120.

Langlais, D. et al., 2008. Regulatory network analyses reveal genome-wide potentiation of LIF signaling by glucocorticoids and define an innate cell defense response. *PLoS genetics*, 4(10), p.e1000224.

Langlais, D. et al., 2012. The Stat3/GR interaction code: predictive value of direct/indirect DNA recruitment for transcription outcome. *Molecular cell*, 47(1), pp.38–49.

Langmead, B. et al., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3), p.R25.

Latchman, S., 1997. Transcription Factors: An Overview. *The International Journal of Biochemistry & Cell Biology*, 2725(97), pp.1305–1312.

Lee, B. & Bhinge, A., 2012. Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells. *Genome research*, pp.9–24.

Lee, C.S. et al., 2005. The initiation of liver development is dependent on Foxa transcription factors. *Nature*, 435(7044), pp.944–7.

Lee, G.M. et al., 2005. The structural and dynamic basis of Ets-1 DNA binding autoinhibition. *The Journal of biological chemistry*, 280(8), pp.7088–99.

Lee, T.I. & Young, R. a, 2013. Transcriptional regulation and its misregulation in disease. *Cell*, 152(6), pp.1237–51.

Lerner, L. et al., 2003. STAT3-dependent enhanceosome assembly and disassembly: synergy with GR for full transcriptional increase of the alpha 2-macroglobulin gene. *Genes & development*, 17(20), pp.2564–77.

Liden, J. et al., 1997. A New Function for the C-terminal Zinc Finger of the Glucocorticoid Receptor: REPRESSION OF RelA TRANSACTIVATION. *Journal of Biological Chemistry*, 272(34), pp.21467–21472.

Liu, C.L. et al., 2005. Single-Nucleosome Mapping of Histone Modifications in S. cerevisiae. *PLoS Biology*, 3(10), p.e328.

Lorch, Y. et al., 2000. Mediator-nucleosome interaction. *Molecular Cell*, 6(1), pp.197–201.

Lucibello, F., Slater, E. & Jooss, K., 1990. Mutual transrepression of Fos and the glucocorticoid receptor: involvement of a functional domain in Fos which is absent in FosB. *The EMBO …*, 9(9), pp.2827–2834.

Luecke, H.F. & Yamamoto, K.R., 2005. The glucocorticoid receptor blocks P-TEFb recruitment by NFkappaB to effect promoter-specific transcriptional repression. *Genes & development*, 19(9), pp.1116–27.

Luger, K. et al., 1997. Crystal structure of the nucleosome core particle at 2 . 8 A resolution. *Nature*, 389, pp.251–260.

Luo, M. et al., 2013. A conserved protein motif is required for full modulatory activity of negative elongation factor subunits NELF-A and NELF-B in modifying glucocorticoid receptor-regulated gene induction properties. *The Journal of biological chemistry*, 288(47), pp.34055–72.

Luo, W. & Bentley, D., 2004. A Ribonucleolytic Rat Torpedoes RNA Polymerase II. *Cell*, 119, pp.911–914.

Lutz, C.S., 2008. Alternative polyadenylation: a twist on mRNA 3′ end formation. *ACS chemical biology*, 3(10), pp.609–17.

Malik, S. & Hisatake, K., 1991. Sequence of general transcription factor TFIIB and relationships to other initiation factors. *Proceedings of the National Academy of Sciences*, 88(November), pp.9553–9557.

Maston, G.A., Evans, S.K. & Green, M.R., 2006. Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics*, Vol. 7.

Mathelier, A. et al., 2014. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic acids research*, 42(Database issue), pp.D142–7.

Matys, V., 2003. TRANSFAC(R): transcriptional regulation, from patterns to profiles. *Nucleic Acids Research*, 31(1), pp.374–378.

Maxon, M.E., Goodrich, J. a & Tjian, R., 1994. Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. *Genes & Development*, 8(5), pp.515–524.

Meierhofer, D. et al., 2013. Protein sets define disease states and predict in vivo effects of drug treatment. *Molecular & cellular proteomics : MCP*, 12(7), pp.1965–79.

Meijsing, S.H. et al., 2009. DNA Binding Site Sequence directs GR Structure and Activity. *Science*, 324(April), pp.407–410.

Meinhart, A. & Cramer, P., 2004. Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. *Nature*, 430(6996), pp.223–6.

Mendenhall, E.M. & Bernstein, B.E., 2012. DNA-protein interactions in high definition. *Genome biology*, 13(1), p.139.

Mikkelsen, T. et al., 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448(7153), pp.553–560.

Milavetz, B.I., 2002. SP1 and AP-1 elements direct chromatin remodeling in SV40 chromosomes during the first 6 hours of infection. *Virology*, 294(1), pp.170–9.

Millevoi, S. et al., 2006. An interaction between U2AF 65 and CF I(m) links the splicing and 3′ end processing machineries. *The EMBO journal*, 25(20), pp.4854–64.

Minucci, S. & Pelicci, P.G., 2006. Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nature reviews. Cancer*, 6(1), pp.38–51.

Mittler, G., Butter, F. & Mann, M., 2009. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome research*, 19(2), pp.284–93.

Mullick, J. et al., 2001. Physical interaction and functional synergy between glucocorticoid receptor and Ets2 proteins for transcription activation of the rat cytochrome P-450c27 promoter. *The Journal of biological chemistry*, 276(21), pp.18007–17.

Myers, R.M. et al., 2011. A user's guide to the Encyclopedia of DNA elements (ENCODE). *PLoS Biology*, 9.

Nader, K. & Hardt, O., 2009. A single standard for memory: the case for reconsolidation. *Nature reviews. Neuroscience*, 10(3), pp.224–34.

Nair, S., Li, D. & Kumar, R., 2013. A Core Chromatin Remodeling Factor Instructs Global Chromatin Signaling Through Multivalent Reading of Nucleosome Codes. *Molecular cell*, 49(4), pp.704–718.

Narlikar, G., Fan, H. & Kingston, R., 2002. Cooperation between Complexes that Regulate Chromatin Structure and Transcription. *Cell*, 108, pp.475–487.

Ohlsson, R., Bartkuhn, M. & Renkawitz, R., 2010. CTCF shapes chromatin by multiple mechanisms: the impact of 20 years of CTCF research on understanding the workings of chromatin. *Chromosoma*, 119(4), pp.351–60.

Ong, C.-T. & Corces, V.G., 2014. CTCF: an architectural protein bridging genome topology and function. *Nature reviews. Genetics*, 15(4), pp.234–46.

Oren, I. et al., 2004. Free diffusion of steroid hormones across biomembranes: a simplex search with implicit solvent model calculations. *Biophysical journal*, 87(2), pp.768–79.

Orphanides, G., Lagrange, T. & Reinberg, D., 1996. The general transcription factors of RNA polymerase II. *Genes & Development*, 10(21), pp.2657–2683.

Pai, D.A. et al., 2014. RNAs nonspecifically inhibit RNA polymerase II by preventing binding to the DNA template. *RNA*, 20(5), pp.644–655.

Pan, D., Kocherginsky, M. & Conzen, S.D., 2011. Activation of the glucocorticoid receptor is associated with poor prognosis in estrogen receptor-negative breast cancer. *Cancer Res*, 71, pp.6360–6370.

Panne, D., 2008. The enhanceosome. *Current opinion in structural biology*, 18(2), pp.236–42.

Payvar, F. et al., 1983. Sequence-specific binding of glucocorticoid receptor to MTV DNA at sites within and upstream of the transcribed region. *Cell*, 35(2), pp.381–392.

Pennisi, E., 2012. ENCODE Project Writes Eulogy For Junk DNA. *Science (New York, N.Y.)*, 337(6099), pp.1159, 1161.

Pereira, M.J. et al., 2014. FKBP5 expression in human adipose tissue increases following dexamethasone exposure and is associated with insulin resistance. *Metabolism: clinical and experimental*, pp.0–10.

Perlmann, T., 1992. Glucocorticoid receptor DNA-binding specificity is increased by the organization of DNA in nucleosomes. *Proceedings of the National Academy of Sciences*, 89(9), pp.3884–3888.

Petersen, J.M. et al., 1995. Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an alpha helix. *Science (New York, N.Y.)*, 269(5232), pp.1866–9.

Petty, E. & Pillus, L., 2013. Balancing chromatin remodeling and histone modifications in transcription. *Trends in genetics : TIG*, 29(11), pp.621–9.

Pfaff, S.J. & Fletterick, R.J., 2010. Hormone binding and co-regulator binding to the glucocorticoid receptor are allosterically coupled. *The Journal of biological chemistry*, 285(20), pp.15256–67.

Phatnani, H.P. & Greenleaf, A.L., 2006. Phosphorylation and functions of the RNA polymerase II CTD. *Genes & development*, 20(21), pp.2922–36.

Picard, D. & Yamamoto, K.R., 1987. Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. *The EMBO journal*, 6(11), pp.3333–3340.

Pike, A.C.., Brzozowski, A.M. & Hubbard, R.E., 2000. A structural biologist's view of the oestrogen receptor. *The Journal of Steroid Biochemistry and Molecular Biology*, 74(5), pp.261–268.

Piña, B., Brüggemeier, U. & Beato, M., 1990. Nucleosome positioning modulates accessibility of regulatory proteins to the mouse mammary tumor virus promoter. *Cell*, 60(5), pp.719–731.

Pobbati, A. & Hong, W., 2013. Emerging roles of TEAD transcription factors and its coactivators in cancers. *Cancer biology & therapy*, (May), pp.1–9.

Polman, J.A.E. et al., 2012. A genome-wide signature of glucocorticoid receptor binding in neuronal PC12 cells. *BMC neuroscience*, 13, p.118.

Pratt, W.B. et al., 2006. Chaperoning of glucocorticoid receptors. *Handbook of experimental pharmacology*, (172), pp.111–38.

Pratts, W.B., 1993. The Role of Heat Shock Proteins in Regulating the Function,Folding and Trafficking of the Glucocorticoid Receptor. *The Journal of Biological Chemistry*, 268(29), pp.21455–21458.

Qiu, Z. et al., 2014. Functional Interactions Between NURF and CTCF Regulate Gene Expression. *Molecular and cellular biology*.

Quinlan, A.R. & Hall, I.M., 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, pp.841–842.

Quinodoz, M. & Gobet, C., 2014. Characteristic bimodal profiles of RNA polymerase II at thousands of active mammalian promoters. *Genome biology*.

Ratman, D. & Berghe, W. Vanden, 2013. How glucocorticoid receptors modulate the activity of other transcription factors : A scope beyond tethering. *Molecular and Cellular Endocrinology*, 380(1-2), pp.41–54.

Ray, a & Prefontaine, K.E., 1994. Physical association and functional antagonism between the p65 subunit of transcription factor NF-kappa B and the glucocorticoid receptor. *Proceedings of the National Academy of Sciences of the United States of America*, 91(2), pp.752–6.

Reddy, T.E. et al., 2009. Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome research*, 19(12), pp.2163–71.

Reeves, W. & Hahn, S., 2003. Activator-independent functions of the yeast mediator sin4 complex in preinitiation complex formation and transcription reinitiation. *Molecular and cellular biology*, 23(1), pp.349–358.

Rhee, H. & Pugh, B., 2011. Comprehensive Genome-wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution. *Cell*, 147(6), pp.1408–1419.

Rhee, H.S. & Pugh, B.F., 2012. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*, 483(7389), pp.295–301.

Rhen, T. & Cidlowski, J.A., 2005. Antiinflammatory action of glucocorticoids--new mechanisms for old drugs. *The New England journal of medicine*, 353, pp.1711–1723.

Richard, H.T. & Foster, J.W., 2003. Acid resistance in Escherichia coli. *Advances in applied microbiology*, 52, pp.167–86.

Richard-Foy, H. & Hager, G., 1987. Sequence-specific positioning of nucleosomes over the steroid- inducible inducible MMTV promoter. *The EMBO journal*, 6(8), pp.2321–2328.

Richmond, T.J. et al., 1984. Structure of the nucleosome core particle at 7 A resolution. *Nature*, 311(5986), pp.532–7.

Robert, V.J. et al., 2014. The SET-2/SET1 Histone H3K4 Methyltransferase Maintains Pluripotency in the Caenorhabditis elegans Germline. *Cell reports*, 9(2), pp.443–50.

Rogatsky, I., 1997. Glucocorticoid receptor-mediated cell cycle arrest is achieved through distinct cell-specific transcriptional regulatory mechanisms. *Molecular and cellular biology*, 17(6), pp.3181–93.

Rosenfeld, J. a et al., 2009. Determination of enriched histone modifications in non-genic portions of the human genome. *BMC genomics*, 10, p.143.

Roth, S.Y., Denu, J.M. & Allis, C.D., 2001. Histone acetyltransferases. *Annual review of biochemistry*, 70, pp.81–120.

Sahu, B. et al., 2011. Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. *The EMBO journal*, 30(19), pp.3962–76.

Sahu, B. et al., 2013. FoxA1 specifies unique androgen and glucocorticoid receptor binding events in prostate cancer cells. *Cancer research*, 73(5), pp.1570–80.

Sakai, D. & Helms, S., 1988. Hormone-mediated repression: a negative glucocorticoid response element from the bovine prolactin gene. *Genes & …*, (1985), pp.1144–1154.

Sandelin, A. et al., 2004. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic acids research*, 32(Database issue), pp.D91–4.

Sapolsky, R.M., Romero, L.M. & Munck, A.U., 2000. How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocrine reviews*, 21(1), pp.55–89.

Scharf, S.H. et al., 2011. Expression and regulation of the Fkbp5 gene in the adult mouse brain. *PloS one*, 6(2), p.e16883.

Schiller, B.J. et al., 2014. Glucocorticoid receptor binds half sites as a monomer and regulates specific target genes. *Genome Biology*, 15(8), p.418.

Schwartz, S., Meshorer, E. & Ast, G., 2009. Chromatin organization marks exon-intron structure. *Nature structural & molecular biology*, 16(9), pp.990–5.

Schwer, B. & Shuman, S., 2011. Deciphering the RNA polymerase II CTD code in fission yeast. *Molecular cell*, 43(2), pp.311–8.

Schwieters, C.D. et al., 2003. The Xplor-NIH NMR molecular structure determination package. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 160(1), pp.65–73.

Sentenac, A., 1985. Eukaryotic RNA polymerases. *CRC critical reviews in biochemistry*, 18(1), pp.31–90.

Serandour, A. a et al., 2013. Development of an Illumina-based ChIP-exonuclease method provides insight into FoxA1-DNA binding properties. *Genome biology*, 14(12), p.R147.

Sharrocks, A.D., 2001. The ETS-domain transcription factor family. *Nature reviews. Molecular cell biology*, 2(11), pp.827–37.

Siersbæk, R. et al., 2011. Extensive chromatin remodelling and establishment of transcription factor "hotspots" during early adipogenesis. *The EMBO journal*, 30(8), pp.1459–72.

Sinars, C.R. et al., 2003. Structure of the large FK506-binding protein FKBP51, an Hsp90-binding protein and a component of steroid receptor complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 100(3), pp.868–73.

Sindhu, C., Samavarchi-Tehrani, P. & Meissner, A., 2012. Transcription factor-mediated epigenetic reprogramming. *The Journal of biological chemistry*, 287(37), pp.30922–31.

Song, L. & Crawford, G.E., 2010. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harbor protocols*, 2010(2), p.pdb.prot5384.

Song, N. et al., 2011. Immunohistochemical Analysis of Histone H3 Modifications in Germ Cells during Mouse Spermatogenesis. *Acta histochemica et cytochemica*, 44(4), pp.183–90.

Steger, D.J. et al., 2008. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Molecular and cellular biology*, 28(8), pp.2825–39.

Stöcklin, E. et al., 1996. Functional interactions between Stat5 and the glucocorticoid receptor. *Nature*, 383(6602), pp.726–8.

Suganuma, T. & Workman, J.L., 2011. Signals and combinatorial functions of histone modifications. *Annual review of biochemistry*, 80, pp.473–99.

Surjit, M. et al., 2011. Widespread negative response elements mediate direct repression by agonist- liganded glucocorticoid receptor. *Cell*, 145(2), pp.224–41.

Tao, Y., Williams-Skipp, C. & Scheinman, R.I., 2001. Mapping of glucocorticoid receptor DNA binding domain surfaces contributing to transrepression of NF-kappa B and induction of apoptosis. *The Journal of biological chemistry*, 276(4), pp.2329–32.

Tapley, T.L. et al., 2010. Protein refolding by pH-triggered chaperone binding and release. *Proceedings of the National Academy of Sciences of the United States of America*, 107(3), pp.1071–6.

Teurich, S. & Angel, P., 1995. The glucocorticoid receptor synergizes with Jun homodimers to activate AP-1-regulated promoters lacking GR binding sites. *Chemical senses*, 20(2), pp.251–5.

The ENCODE Project Consortium, 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science (New York, N.Y.)*, 306(5696), pp.636–40.

Thomas-Chollier, M., Darbo, E., et al., 2012. A complete workflow for the analysis of full-size ChIP-seq (and similar) data sets using peak-motifs. *Nature protocols*, 7(8), pp.1551–68.

Thomas-Chollier, M. et al., 2013. A naturally occurring insertion of a single amino acid rewires transcriptional regulation by glucocorticoid receptor isoforms. *Proceedings of the National Academy of Sciences of the United States of America*, 110(44), pp.17826–31.

Thomas-Chollier, M. et al., 2011. RSAT 2011: regulatory sequence analysis tools. *Nucleic acids research*, 39(Web Server issue), pp.W86–91.

Thomas-Chollier, M., Herrmann, C., et al., 2012. RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic acids research*, 40(4), p.e31.

Thomas-Chollier, M. et al., 2008. RSAT: regulatory sequence analysis tools. *Nucleic acids research*, 36(Web Server issue), pp.W119–27.

Thurman, R.E. et al., 2012. The accessible chromatin landscape of the human genome. *Nature*, 489(7414), pp.75–82.

Touray, M. et al., 1991. Characterisation of functional inhibition of the glucocorticoid receptor by Fos/Jun. *Oncogene*, 6(7), pp.1227–34.

Trebble, P.J. et al., 2013. A ligand-specific kinetic switch regulates glucocorticoid receptor trafficking and function. *Journal of cell science*, 126(Pt 14), pp.3159–69.

Tuteja, G. & Kaestner, K.H., 2007. SnapShot: forkhead transcription factors I. *Cell*, 130(6), p.1160.

Vassilev, a et al., 2001. TEAD/TEF transcription factors utilize the activation domain of YAP65, a Src/Yes-associated protein localized in the cytoplasm. *Genes & development*, 15(10), pp.1229–41.

Venters, B.J. & Pugh, B.F., 2013. Genomic organization of human transcription initiation complexes. *Nature*, 502(7469), pp.53–8.

Verger, A. & Duterque-Coquillaud, M., 2002. When Ets transcription factors meet their partners. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 24(4), pp.362–70.

Viphakone, N., Voisinet-Hakil, F. & Minvielle-Sebastia, L., 2008. Molecular dissection of mRNA poly(A) tail length control in yeast. *Nucleic acids research*, 36(7), pp.2418–33.

Voet, J.G. & Voet, D., 2007. Biochemistry. *Biochemistry and molecular biology education : a bimonthly publication of the International Union of Biochemistry and Molecular Biology*, 35(1), p.1.

Wahle, E., 1991. A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell*, 66(4), pp.759–68.

Waki, H. et al., 2011. Global mapping of cell type-specific open chromatin by FAIRE-seq reveals the regulatory role of the NFI family in adipocyte differentiation. *PLoS genetics*, 7(10), p.e1002311.

Wakui, H. et al., 1997. Interaction of the Ligand-activated Glucocorticoid Receptor with the 14-3-3 Protein. *Journal of Biological Chemistry*, 272(13), pp.8153–8156.

Wang, Y., Li, X. & Hu, H., 2014. H3K4me2 reliably defines transcription factor binding regions in different cells. *Genomics*, 103(2-3), pp.222–8.

Wasim, M. et al., 2010. PLZF/ZBTB16, a glucocorticoid response gene in acute lymphoblastic leukemia, interferes with glucocorticoid-induced apoptosis. *The Journal of steroid biochemistry and molecular biology*, 120(4-5), pp.218–27.

Wasserman, W.W. & Fickett, J.W., 1998. Identification of regulatory regions which confer muscle-specific gene expression. *Journal of molecular biology*, 278(1), pp.167–81.

Weber, C.M., Ramachandran, S. & Henikoff, S., 2014. Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Molecular cell*, 53(5), pp.819–30.

Whittaker, J.R., 1973. Segregation during ascidian embryogenesis of egg cytoplasmic information for tissue-specific enzyme development. *Proceedings of the National Academy of Sciences of the United States of America*, 70(7), pp.2096–100.

Wiench, M., Miranda, T.B. & Hager, G.L., 2011. Control of nuclear receptor function by local chromatin structure. *The FEBS journal*, 278(13), pp.2211–30.

Wobbe, C.R. & Struhl, K., 1990. Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. *Molecular and cellular biology*, 10(8), pp.3859–67.

Wolffe, a P. & Pruss, D., 1996. Targeting chromatin disruption: Transcription regulators that acetylate histones. *Cell*, 84(6), pp.817–9.

Worsley Hunt, R. et al., 2014. Improving analysis of transcription factor binding sites within ChIP-Seq data based on topological motif enrichment. *BMC genomics*, 15(1), p.472.

Wrange, O., Eriksson, P. & Perlmann, T., 1989. The purified activated glucocorticoid receptor is a homodimer. *The Journal of biological chemistry*, 264(9), pp.5253–9.

Wreggett, K.A. et al., 1994. A mammalian homologue of Drosophila heterochromatin protein 1 (HP1) is a component of constitutive heterochromatin. *Cytogenetics and cell genetics*, 66(2), pp.99–103.

Wu, J. & Bresnick, E.H., 2007. Glucocorticoid and growth factor synergism requirement for Notch4 chromatin domain activation. *Molecular and cellular biology*, 27(6), pp.2411–22.

Xi, H. et al., 2007. Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS genetics*, 3(8), p.e136.

Xu, L., Glass, C.K. & Rosenfeld, M.G., 1999. Coactivator and corepressor complexes in nuclear receptor function. *Current opinion in genetics & development*, 9(2), pp.140–7.

Xu, Z., Horwich, A.L. & Sigler, P.B., 1997. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature*, 388(6644), pp.741–50.

Yamaguchi, Y. et al., 1999. NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress RNA Polymerase II Elongation. *Cell*, 97(1), pp.41–51.

Yamamoto, K.R., 1985. Steroid Receptor regulated Transcription of specific Genes and Gene Networks. *Annual Review of Genetics*, 19, pp.209–52.

Yang-Yen, H.-F. et al., 1990. Transcriptional interference between c-Jun and the glucocorticoid receptor: Mutual inhibition of DNA binding due to direct protein-protein interaction. *Cell*, 62(6), pp.1205–1215.

Yu, C.Y. et al., 2010. Genome-wide analysis of glucocorticoid receptor binding regions in adipocytes reveal gene network involved in triglyceride homeostasis. *PLoS ONE*, 5.

Yu, X. et al., 2005. Foxa1 and Foxa2 interact with the androgen receptor to regulate prostate and epididymal genes differentially. *Annals of the New York Academy of Sciences*, 1061, pp.77–93.

Zaret, K.S. & Carroll, J.S., 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes & development*, 25(21), pp.2227–41.

Zelin, E. et al., 2012. The p23 molecular chaperone and GCN5 acetylase jointly modulate protein-DNA dynamics and open chromatin status. *Molecular cell*, 48(3), pp.459–470.

Zhang, Y. et al., 2008. Model-based analysis of ChIP-Seq (MACS). *Genome biology*, 9(9), p.R137.

Zhang, Z. et al., 1997. STAT3 Acts as a Co-activator of Glucocorticoid Receptor Signaling. *Journal of Biological Chemistry*, 272(49), pp.30607–30610.

Zhao, J., Hyman, L. & Moore, C., 1999. Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiology and molecular biology reviews : MMBR*, 63(2), pp.405–45.

Zhu, Y., van Essen, D. & Saccani, S., 2012. Cell-Type-Specific Control of Enhancer Activity by H3K9 Trimethylation. *Molecular Cell*, 46, pp.408–423.

DNA helix in Figure 2, Figure 32, Figure 33, Figure 58, Figure 55 and Figure 60 is from Wikipedia (DNA simple2.svg). The author (Forluvof) released it into public domain.

10 Abbreviations

AB	Antibody
AF1	Activation Function 1
AF2	Activation Function 2
BtnTg	Biotin Tag
ChIP	Chromatin Immunoprecipitation
ChIP-Exo	ChIP-Exonuclease
ChIP-Seq	ChIP-Sequencing
CPSF	cleavage and polyadenylation specificity factor
CstF	cleavage stimulation factor
CTD	C-terminal domain
DBD	DNA binding domain
ddH_2O	aqua bidestillata
Dex	dexamethasone
DHS	DNase I hypersensitive site
DNA	Deoxyribonucleic acid
GBR	GR-bound region
GBS	GR binding sequence
GC	Glucocorticoid
GR	Glucocorticoid Receptor
GRE	Glucocorticoid Receptor Response Element
HAT	Histone acetyltransferase
HMT	Histone methyltransferase
kb	kilo base pair
LBD	ligand binding domain
nGRE	negative Glucocorticoid Receptor Response Element
NGS	Next Generation Sequencing
NR	Nuclear hormone receptors
PIC	Pre-Initiation-Complex
PNAT	ATP:polynucleotide adenylyltransferase
PTIC	promoter transcription initiation complex
RNA	Ribonucleic acid
RNAP II	RNA polymerase II
TF	transcription factor
TF-BS	TF binding sites
TSS	transcriptional start site
αGR-ChIP	anti-GR-ChIP

Hiermit versichere ich, Stephan Raphael Starick, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer, als die der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Schriften entnommen sind, sind als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form von mir noch nicht als Prüfungsarbeit eingereicht worden.

Berlin, den 16. Dezember 2014