# Mechanisms Directing Receptor-Specific Gene Regulation by the Androgen and Glucocorticoid Receptor

Inaugural-Dissertation to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry,
Pharmacy of Freie Universität Berlin

by **Marina Kulik** • 2021

The dissertation was prepared under the supervision of Dr. Sebastiaan H. Meijsing at the Max Planck Institute for Molecular Genetics in Berlin from September 2015 to February 2021.

1st Reviewer: Dr. Sebastiaan Meijsing

2nd Reviewer: Prof. Dr. Markus Wahl

Date of defense: 21.05.2021

# Selbstständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbstständig und unter Zuhilfenahme der angegebenen Literatur erstellt habe.

## Acknowledgments

First of all, I would like to express my gratitude to my supervisor Sebastiaan Meijsing for his support and guidance during my PhD. I would like to thank Martin Vingron for the opportunity to be part of his research group and for the great collaborations. Especially, I would like to thank Stefan Haas for introducing me to the world of RNA-seq and Gözde Kibar, who contributed with her bioinformatical analyses to this work. I would like to thank Sarah Kinkley for her support and for giving me the opportunity to finish my PhD in her group.

I wish to thank Stefan Prekovic, Isabel Mayayo-Peralta and Wilbert Zwart from the NKI in Amsterdam for sharing their expertise in "nuclear signaling" and the great collaboration.

I am particularly grateful to Melissa Bothe, her computational analyses and support in the lab contributed a lot to this work. My special gratitude goes to Laura Glaser for her continuous advice and for freezing my cells in the evening countless times.

I wish to thank Edda Einfeldt for her great technical support and for always having an open ear. I would like to thank Verena Thormann, Nilofar Badra-Azar, Stefanie Schöne and all the members of the Vingron group for the discussions, life saving coffee breaks and the fun outside the institute.

Finally, I would like to thank my parents, my friend Aline Behn, who accompanied me all the time since school years, and my husband Stanislav Kulik for his loving support and eternal optimism.

# Contents

# Abstract

The androgen receptor (AR) and glucocorticoid receptor (GR) are closely related transcription factors that have distinct biological functions but recognize nearly identical DNA binding sites. To provide insights into this apparent paradox, I studied the role of genomic binding, the chromatin environment and cofactor interactions in driving receptor-specific gene regulation.

To study differences of the receptors in an identical cellular context I generated a U2OS cell line stably expressing AR and compared it to the existing GR-expressing U2OS cell line. Activation of the receptors revealed the expression of receptor-specific as well as shared target genes. I found that receptor-specific gene expression can in part be driven by subtle binding differences of the receptors.

In order to investigate whether divergent DNA binding can be a consequence of different abilities to bind to inaccessible chromatin I performed ATAC-seq in both cell lines. The analysis revealed a receptor-specific ability of GR to bind to closed chromatin which in turn directs GR-specific binding and gene regulation.

Furthermore, to understand the effect of the sequence composition on differential binding, I compared the genomic loci that are selectively occupied by one of the receptors. Notably, GR showed a preference for GC-rich loci suggesting a role in local sequence composition in directing GR to its genomic binding sites and thereby contributing to GR-specific gene regulation.

Remarkably, receptor-specific gene regulation was also observed near sites that are occupied by both receptors. A genomic deletion of these sites in their endogenous context indeed resulted in a loss of regulation of the nearby gene validating that these sites are essential for transcriptional activation.

To investigate how the receptors direct differential gene regulation from shared binding sites I performed STARR-seq, an episomal reporter assay, and compared the activity of AR and GR at occupied genomic enhancers. I found receptor-specific enhancer activation at some shared binding sites suggesting that receptor-specific gene regulation can be driven by distinct abilities to activate genomic enhancers.

Finally, to explore if differential enhancer activation can be a consequence of different interactions with transcriptional cofactors, I compared the cofactors recruited by AR and GR. The analysis revealed that proteins from the mediator complex as well as enzymes harboring acetyltransferase

activity are specifically enriched among the GR interactors. This demonstrates that functional diversity at sites occupied by both receptors could be driven by differential cofactor recruitment.

Together, I found that differences in gene regulation by AR and GR can be driven by the sequence composition as well as distinct abilities to bind closed chromatin. Moreover, gene regulation can be driven by mechanisms downstream of binding, such as differential enhancer activation and cofactor recruitment. These observations might potentially apply to other paralogous transcription factors and help explain how related proteins achieve functional diversification despite similar DNA sequence preferences. Moreover, the model system I have generated provides a valuable setup to further study different mechanisms of signaling by AR and GR in the future.

# Zusammenfassung

Der Androgenrezeptor (AR) und Glukokortikoidrezeptor (GR) sind eng verwandte Transkriptionsfaktoren, die unterschiedliche biologische Funktionen haben, aber nahezu identische DNA-Bindestellen erkennen. Um dieses Paradoxon besser zu verstehen, habe ich den Einfluss der genomischen Bindung, der Chromatin-Umgebung und der Interaktionen mit Cofactoren auf die rezeptorspezifische Genregulation untersucht.

Um die Unterschiede der Rezeptoren in einem identischen zellulären Kontext zu untersuchen, erzeugte ich eine U2OS-Zelllinie, die AR stabil exprimiert und verglich sie mit der bestehenden GR-exprimierenden U2OS-Zelllinie. Die Aktivierung der Rezeptoren zeigte die Expression sowohl rezeptorspezifischer als auch gemeinsamer Zielgene. Es zeigte sich, dass die rezeptorspezifische Genexpression zum Teil aus subtilen Unterschieden in der DNA-Bindung der Rezeptoren resultiert.

Um zu untersuchen, ob Unterschiede in der DNA-Bindung eine Folge von unterschiedlichen Fähigkeiten an geschlossenes Chromatin zu binden sind, führte ich ATAC-seq in beiden Zelllinien durch. Die Analyse ergab, dass GR die Fähigkeit besitzt an geschlossenes Chromatin zu binden, was wiederum eine GR-spezifische Bindung und Genregulation auslöst.

Um zu verstehen, ob die Sequenzzusammensetzung einen Einfluss auf die differentielle Bindung hat, verglich ich außerdem die genomischen Loci, die selektiv von einem der Rezeptoren besetzt werden. GR zeigte im Vergleich zu AR eine deutliche Präferenz für GC-reiche Loci. Dies deutet darauf hin, dass die lokale Sequenzzusammensetzung einen Einfluss auf die DNA-Bindung von AR und GR hat und damit zur rezeptorspezifischen Genregulation beiträgt.

Bemerkenswerterweise wurde eine rezeptorspezifische Genregulation auch in der Nähe von Bindestellen beobachtet, die von beiden Rezeptoren gleichermaßen besetzt sind. Eine genomische Deletion dieser Bindestellen in ihrem endogenen Kontext führte tatsächlich zu einem Verlust der Regulation des nahegelegenen Gens, was bestätigt, dass die Bindestellen für die Aktivierung dieses Gens verantwortlich sind.

Um zu untersuchen, wie die Rezeptoren eine differenzielle Genregulation von gemeinsamen Bindestellen aus erreichen, habe ich mithilfe von STARR-seq die Aktivierung genomischer Enhancer durch die beiden Rezeptoren verglichen. Ich fand heraus, dass AR und GR eine

rezeptorspezifische Enhancer-Aktivierung an einigen gemeinsamen Bindestellen aufweisen. Dies deutet darauf hin, dass eine rezeptorspezifische Genregulation aus unterschiedlichen Fähigkeiten der Enhancer-Aktivierung resultieren kann.

Um zu untersuchen ob Unterschiede bei der Enhancer-Aktivierung durch AR und GR möglicherweise eine Folge unterschiedlicher Interaktionen mit Cofaktoren sind, habe ich untersucht, ob die Rezeptoren Unterschiede bei der Cofaktorrekrutierung aufweisen. Die Analyse zeigte, dass der Mediator-Komplex sowie einige Acetyltransferasen spezifisch mit GR interagieren. Dies deutet darauf hin, dass Unterschiede in der Genregulation eine Folge von unterschiedlichen Interaktionen von AR und GR mit Cofaktoren sind.

Zusammenfassend habe ich herausgefunden, dass die Sequenzzusammensetzung der DNA sowie unterschiedliche Fähigkeiten geschlossenes Chromatin zu binden einen Einfluss auf die rezeptor-spezifische Genregulation von AR und GR haben. Darüber hinaus kann eine differentielle Genregulation aus unterschiedlichen Interaktionen mit Cofaktoren resultieren. Diese Beobachtungen sind möglicherweise auch auf weitere verwandte Transkriptionsfaktoren übertragbar und können dazu beitragen, zu erklären, wie Proteine trotz ähnlicher Präferenzen für die DNA-Sequenz spezifische Funktionen erreichen. Darüber hinaus bietet das von mir erzeugte Modellsystem die Möglichkeit, verschiedene Mechanismen von AR und GR in Zukunft weiter zu untersuchen.

# 1. Introduction

## 1.1 Transcriptional Regulation

The cells of a complex and multicellular organism share the same genetic information which in turn gives rise to all the different cell types of a body. Remarkably, in humans more than 98 % of the DNA comprise non-coding sequences whereas only less than 2 % of the DNA sequences are coding for proteins. Regulatory elements, which are part of the non-coding sequences, play a role in regulating which genes are expressed in a given cell (Thurman et al. 2012). The genes expressed in muscle cells strongly diverge from the genes expressed in blood cells. Thus, the genomic information contained in each cell of an organism has to be orchestrated precisely in time and space to ensure the correct development and maintenance of the different cell types.

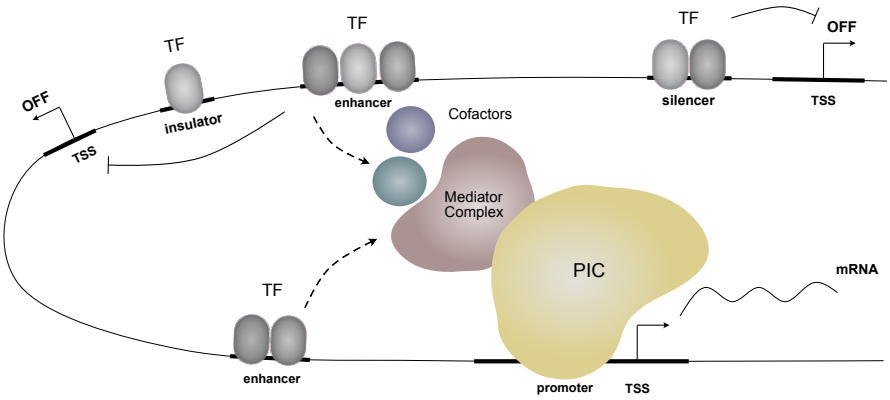



**Figure 1.1. Enhancer-mediated transcriptional regulation**. Enhancers are regulatory DNA elements that contain binding motifs for various transcription factors (TFs). Enhancer-bound TFs interact with different cofactors and induce the assembly of the preinitiation complex (PIC) leading to the transcription of a target gene from its transcriptional start site (TSS). On the contrary, binding of silencers by TFs leads to transcriptional downregulation of a target gene. Insulators are regulatory elements that form a boundary between enhancers and genes located outside the boundary. Distally located enhancers are brought into spatial proximity to their target gene by looping of the DNA, which is stabilized by different proteins such as the mediator complex. The complex patterns of cell-type specific gene regulation are orchestrated by the interplay of various enhancers that can act in a cooperative manner.

In mammalian cells, the regulatory elements acting in cis and playing an important role in gene regulation are promoters (Weingarten-Gabbay et al. 2019), enhancers (Shlyueva, Stampfel, and Stark 2014a), silencers (Doni Jayavelu et al. 2020) and insulators (Gaszner and Felsenfeld 2006a). The core promoter is a regulatory element which is typically located around 50 bp upstream and downstream of the transcriptional start site (TSS) of a gene. It comprises the binding platform for

the RNA polymerase (Pol II) as well as general transcription factors, which form the transcriptional machinery. Transcription factors (TF) are DNA binding proteins that bind to specific regulatory elements thereby regulating gene expression (Vaquerizas et al. 2009). Upon binding of regulatory elements, sequence-specific TFs interact with their specific cofactors to facilitate the recruitment of the general TFs and Pol II leading to the assembly of the preinitiation complex (PIC) at the TSS and transcriptional initiation (Hampsey 1998). After an unwiring of the DNA double helix, Pol II is being released from the TSS and the general TFs to transcribe DNA into RNA. However, after escaping the promoter, the Pol II synthesizes only a short stretch of nascent RNA and pauses upon binding of the negative elongation factor (NELF) and the sensitivity inducing factor (DSIF) (Yamaguchi, Shibata, and Handa 2013). To escape the pausing, the cyclin-dependent kinase 9 (CDK9) phosphorylates the DSIF, NELF and the C-terminus of Pol II, initiating the release of Pol II into productive transcriptional elongation (Guenther et al. 2007). Enhancers are regulatory elements that play a critical role in directing transcriptional regulation. Typically, enhancers contain clusters of transcription factor binding sites and are located either promoter proximal or at a large distance up to several hundreds of kilobases (kb) to the promoter. Enhancers recruit specific TF and enhance gene transcription. Silencers act in a similar way, but in contrast to enhancers they lead to a reduction of gene expression and are bound by repressor transcription factors. Insulators, however, are regulatory elements that represent boundaries and block signals mediated by enhancers or silencers from genes, which are located outside the boundary (Gaszner and Felsenfeld 2006b).

The complex interplay of coding DNA sequences and regulatory elements drives differential regulatory programs subsequently giving rise to distinct phenotypes. Therefore, a detailed knowledge of transcriptional regulation is not only important for the general understanding of developmental processes but also to identify disease-causing misregulations (Herz, Hu, and Shilatifard 2014).

## 1.2 Influence of the Chromatin Landscape

The three-dimensional genome organization plays an important role in specifying which gene is expressed in a given cell by directing the availability of binding sites for TFs as well as the transcription machinery. Moreover, the genome organization differs between cell types and undergoes changes during the development or upon the influence of external stimuli (Bell et al. 2011). Hence, the surrounding genome architecture and its complex interaction with various factors such as Pol II or sequence-specific TFs is an important driver of cell type-specific gene regulation.

The basic units of the DNA packaging are nucleosome core particles. In eukaryotes each 147 bp of DNA are wrapped about 1.6 times around a histone octamer (Luger et al. 1997). The histone octamer consists of two copies of the core histones H3, H4, H2A and H2B (Kornberg 1974) or variants of these core histones.
Histones can be modified posttranslationally at their histone tails by chromatin modifying enzymes thereby affecting the local chromatin structure and transcriptional state (Jenuwein and Allis 2001). Posttranslational modifications can be added, removed or recognized by so-called histone writers, erasers or readers, a protein machinery that directly modifies the chromatin state. Writers are histone modifying enzymes, such as histone acetyltransferases (HAT) or histone methyltransferases (HMT), that add posttranslational modifications to histone tails, whereas erasers are enzymes that remove posttranslational modifications. Readers are a special class of enzymes, that recognize specific posttranslational modifications. They can either comprise chromatin remodelers, such as the SWI/SNF protein complex, or enzymes regulating transcription by recruiting the transcription machinery (Bannister and Kouzarides 2011).

Histone marks associated with active chromatin are linked to opening of the chromatin to facilitate transcriptional regulation, therefore, transcriptionally active chromatin carries different posttranslational modifications compared to silent chromatin. H3K27ac and H3K4me1 are histone modifications associated with active enhancers (Creyghton et al. 2010) and H3K4me3 is associated with active promoters (Liang et al. 2004). In contrast, the histone marks H3K27me3 and H3K9me3 are found within repressed and transcriptionally silent chromatin (Ferrari et al. 2014; Richards and Elgin 2002). Particularly, H3K27ac is of special interest, because it is able to distinguish active from poised enhancers (Creyghton et al. 2010). Acetylation of histone 3 lysine 27 is catalyzed by HATs such as EP300 and CREBBP (Ogryzko et al. 1996). These enzymes are thought to play a fundamental role in gene regulation for several reasons. First, they directly interact with TFs

thereby recruiting the transcriptional machinery and regulating transcription (Q. Jin et al. 2011). Second, they catalyze H3K27ac which is an indicator of enhancer function. However, it has also been shown that H3K27ac is dispensable for enhancer function (T. Zhang et al. 2020), thus it remains unclear if this enhancer mark is simply indicative or indeed causative for enhancer activity of regulatory elements.

At the megabase-scale, chromosomes are folded into structural units called topologically associating domains (TADs) (Dixon et al. 2012; Nora et al. 2012). TADs represent regulatory hubs with increased genomic interactions of regulatory regions and their cognate gene promoters within these TADs, but only little interaction between separate TADs. TADs are formed by looping of the DNA and stabilized by the architectural proteins, such as cohesin and CTCF, a repressor transcription factor enriched at the TAD boundaries (Hansen 2020). On a larger scale of several megabases open and condensed chromatin are spatially segregated into A and B compartments, respectively (S. Wang et al. 2016) A compartments are associated with open chromatin and active transcription whereas B compartments are located in the periphery of the nucleus and associated with condensed and transcriptionally inactive chromatin (Dekker, Marti-Renom, and Mirny 2013). The last level of organization is the positioning of the chromosomes within a cell. In eukaryotes, chromosomes are not randomly packaged in the nucleus but occupy very defined nuclear territories (Cremer and Cremer 2001).

DNA accessibility is influenced by the level of DNA occupancy by nucleosomes and other architectural proteins. DNA accessibility is considered low within heterochromatin, where the DNA is densely packed by histones and other proteins, whereas active regulatory regions are depleted from histones and located within open chromatin. The majority of TFs binds to nucleosome-depleted DNA, therefore, the accessibility reflects the local regulatory state of the chromatin (Thurman et al. 2012). Moreover, chromatin accessibility is strongly cell-type specific and a major driver of cell-type-specific TF binding (Barozzi et al. 2014; John et al. 2011). However, DNA-accessibility is highly dynamic (Schones et al. 2008) and can be modulated by different proteins such as TFs or chromatin remodelers.

DNase I-seq was the first method applied to identify open chromatin in a genome-wide manner (Boyle et al. 2008; Hesselberth et al. 2009). It is based on the principle that open chromatin is less compacted and therefore relatively accessible to cleavage by the nuclease DNase I. The abundance of these so-called hypersensitive sites can be identified by sequencing of the fragment library. FAIRE-Seq (Formaldehyde-Assisted Isolation of Regulatory Elements) (Giresi et al. 2007) is an alternative method to identify hypersensitive sites in the genome that does not require the addition

of enzymes. The assay is based on the biochemical properties of the chromatin. First, the chromatin is crosslinked using formaldehyde, crosslinking is more efficient for nucleosome-bound DNA and covalently binds proteins and DNA. Second, the nucleosome-bound DNA can be separated from nucleosome-depleted DNA by phenol-chloroform extraction. The FAIRE-fragments are sheared by sonication and can be analyzed by sequencing allowing the identification of accessible chromatin genome-wide.

ATAC-seq (Buenrostro et al. 2013; Corces et al. 2017) is a further method to detect open chromatin, however, compared to DNaseI-seq and FAIRE-seq it is less laborious and requires very low amounts of input material. The method relies on the activity of the hyperactive Tn5 transposase, that simultaneously fragments open DNA and pastes sequencing adapters to the ends of the DNA fragments. Briefly, the cells are lysed to isolate the nuclei and the TN5 is added for the transposition reaction. After a DNA purification step, the fragments are amplified by PCR and can be either analyzed by sequencing or by qPCR using locus-specific primers.

## 1.3 Genomic Enhancers

Before the concept of enhancer elements was introduced, it was already proposed that proteins are involved in gene regulation (Jacob and Monod 1961) and that proteins are able to directly bind DNA (Ptashne 1967). Enhancer activity was demonstrated for the first time in 1981 for a viral DNA sequence that was able to enhance the expression of a reporter gene (Banerji, Rusconi, and Schaffner 1981; Moreau et al. 1981). Specifically, a non-coding DNA sequence from the simian virus 40 (SV40) was able to increase the expression of the gene β-globin on an episomal reporter in *cis*.

Our understanding of general enhancer function has been gained through extensive studies of a few classical enhancer examples. The functional characteristics that have been shown are that enhancers are rather small regulatory elements of only a few hundred basepairs (Andersson and Sandelin 2020; Gasperini, Tome, and Shendure 2020; Shlyueva, Stampfel, and Stark 2014). They act in a modular manner and independent of their orientation to the promoter. They can act over a large distance of hundreds kilobases or even megabases by forming DNA loops and thereby establishing physiological contacts with the gene promoter. Typically, enhancers contain whole clusters of transcription factor binding sites and the combinatorial binding of different transcription factors is able to drive complex patterns of temporal and spatial gene expression (Shlyueva, Stampfel, and Stark 2014). Moreover, enhancers are located in genomic regions depleted of nucleosomes and are therefore accessible for transcription factor binding (Tuan et al. 1985). Strikingly, DNA accessibility is different in every cell-type and also changes during developmental

processes (Bell et al. 2011). Accordingly, enhancer activity has also been shown to be highly cell-type specific. Although enhancers are located in nucleosome-free regions, they are flanked by histones marked with specific posttranslational modifications. Typically, active enhancers are accompanied by high H3K27ac (Creyghton et al. 2010) and a high signal ratio of H3K4me1 to H3K4me3 (Robertson et al. 2008).

Enhancer function is crucial for developmental processes, maintenance of cell function and a rapid response to stress and other external signals (Furlong and Levine 2018). Moreover, mutations encompassing non-coding regulatory regions often have pathogenic consequences (Visel, Rubin, and Pennacchio 2009). However, the reliable identification of enhancers specific to their target genes is rather difficult as the distance to the target promoter can strongly vary. The ZRS enhancer, for example, is around 1 Mb away from its target gene *Shh* (Symmons et al. 2016). Traditionally, enhancers were identified by scanning genomic regions in the vicinity of the promoter of a gene for conserved sequences and subsequently to test those sequences for enhancer activity in a cell culture system or with the help of transgenic mice (Forghani et al. 2001; Werner et al. 2007). This approach is rather tedious, not scalable and relies on the arbitrary decision to search for enhancers at a certain distance from the promoter. Therefore, efforts have been made with the help of large consortia to annotate enhancers based on their biochemical features like chromatin accessibility and histone modifications in a genome-wide manner (Feingold et al. 2004). However, the enhancer identification purely based on annotations has major drawbacks. First, the sequences lack the evidence of enhancer activity. The putative enhancer might not be functional. And second, the annotated enhancers cannot be directly linked to a target gene.

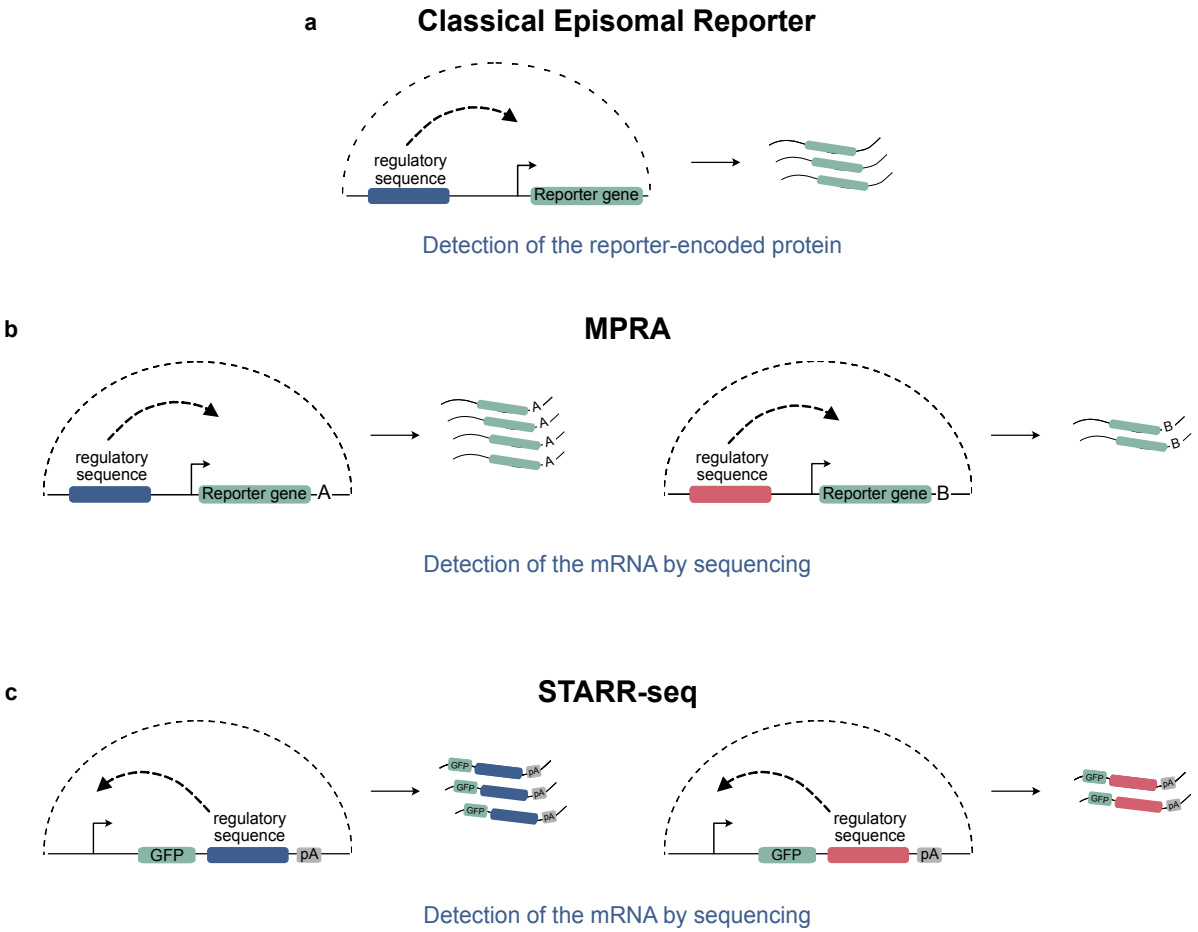



**Figure 1.2. Different types of episomal reporter assays. (a)** In a classical episomal reporter the putative enhancer is placed upstream of a reporter gene. Upon enhancer activation the reporter gene the abundance of the reporter-encoded protein can be measured by various methods. **(b)** In a massively parallel reporter assay (MPRA) many candidate sequences can be tested in parallel. A barcode is linked to each enhancer candidate and coupled to the reporter transcripts. The abundance of each barcode can be detected by sequencing. **(c)** In a self-transcribing active regulatory region sequencing (STARR-seq) assay the enhancer is placed downstream of a minimal promoter. Millions of candidates can be tested simultaneously. The enhancer activity is reflected in the abundance of the enhancer sequence among the cellular RNA and can be detected by sequencing.

Classical methods to functionally dissect enhancers are reporter assays that are based on the fact that enhancers act independent of their orientation and position on an episomal vector. Typically, the enhancer is placed upstream of a minimal promoter coupled to a reporter gene. The output can be detected as fluorescence intensity (e.g. GFP), enzymatic activity (e.g. luciferase) or the abundance of the reporter transcript. A massively parallel reporter assay (MPRA) (Patwardhan et al. 2012) is a powerful tool to directly measure the activity of thousands of candidate sequences in high-throughput. Specifically, the enhancer is placed in an episomal reporter upstream of the promoter and a reporter gene that is coupled with a barcode. The episomal reporter library is transfected into a cell line and the barcoded transcripts are isolated followed by deep sequencing. The abundance of the reporter transcript thereby reflects the strength of the enhancer sequence.

Self-transcribing active regulatory region sequencing (STARR-seq) (Arnold et al. 2013) is a unique MPRA version that allows genome-wide enhancer screens for millions of candidates in a qualitative and quantitative manner. In this assay, the enhancer sequence is placed downstream of the promoter allowing the enhancer to transcribe itself upon activation. Thus, activity of the enhancer is reflected in its abundance among the cellular RNA. With the application of STARR-seq nearly the whole Drosophila could be successfully scanned for enhancer sequences.

Episomal reporter assays have several advantages. First, they can be scaled up easily and millions of candidate sequences can be assayed in parallel. Second, the enhancer activity is examined independently of the native chromatin context, therefore, reporter assays are able to measure the intrinsic enhancer activity of regulatory elements. However, a major drawback is that reporter assays lack the native chromatin environment and the 3D structure of the DNA. Hence, enhancers identified in reporter assay are not necessarily active in their native chromatin environment and have to be functionally tested, to validate that they are indeed responsible for the regulation of the nearby genes. Genome-wide enhancer screens based on CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) (Jinek et al. 2012) can be applied to identify enhancers in their native chromatin context. CRISPR screens involve the delivery of a library of single guide RNAs that target genomic enhancers of interest. Next, a screen for changes in gene expression of one or several target genes is performed. The target genes can be labeled in different ways such as fusion with a reporter gene or labelling with fluorescent markers. This approach allows to directly link gene expression to enhancer function in the native chromatin context.

There are several ways to induce CRISPR-mediated genomic perturbations, such as delivery of the active Cas9 to induce DNA double-strand breaks (Canver et al. 2015) or, alternatively, the delivery of a dead Cas9 which is coupled to a repressor (Fulco et al. 2016) or activator domain (Simeonov et al. 2017) and induces epigenetic changes. The KRAB repressor domain tethered to a dead Cas9 enzyme can silence a target gene by inducing repressive chromatin marks (Thakore et al. 2015), also referred to as CRISPR interference or CRISPRi. Alternatively, activating domains, such as the transcriptional activators VP64 or the acetyltransferase EP300, have been fused to Cas9 to activate poised enhancers (Klann et al. 2017; Simeonov et al. 2017). However, this approach is difficult to scale up because it is limited by the screening for one or few target genes. An alternative is a CRISPR-based genome-wide screen of regulatory elements. It relies on the delivery of a guide RNA library, following by a single-cell RNA-seq analysis of the mRNA as well as the guide RNAs (Xie et al. 2017). In this way, changes in gene expression can be detected and simultaneously linked to specific guide RNAs present in the cell. However, this approach requires great multiplexing and is very cost intensive.

Especially, in genomic regions with multiple putative enhancers it still remains difficult to understand which enhancers are responsible for the regulation of nearby genes. Enhancers can be redundant and compensate for the loss of each other or multiple enhancers can act in an additive manner by stabilizing physical contacts with the promoter in a regulatory hub (Shlyueva, Stampfel, and Stark 2014).

## 1.4 Transcription Factors

A precise gene expression is crucial to establish the right cell fate decisions, for complex developmental processes and metastasis. The information for the regulation of genes is encoded in regulatory sequences that are "read" by transcription factors. The correct interpretation of those sequences by TFs is critical for all cell functions.

TFs bind in a sequence-specific manner to transcription factor binding sites to regulate gene transcription. The sequence preference of TFs can be visualized as motifs that are usually short (6-12 bp) and degenerate, meaning that they allow mismatches at certain positions. The motifs are typically represented as a sequence logo (Schneider and Stephens 1990) which is based on a position weight matrix that displays the preference at every base of the motif (Stormo and Zhao 2010).

Hence, there is a high sequence variability among the binding sites that can be bound by a specific TF and many potential binding sites can be found throughout the genome. However, only a subset of potential motifs is actually occupied by sequence-specific TFs, and therefore, additional mechanisms are needed to explain where TFs bind in the genome (Joseph et al. 2010). For instance, occupancy by TFs highly correlates with nucleosome depleted DNA (John et al. 2011). Moreover, cooperativity with other TFs plays an important role in TF binding to DNA (Jolma et al. 2015; Slattery et al. 2011; Yáñez-Cuna et al. 2012). Accordingly. enhancers typically encode clusters of different transcription factor binding sites that direct cooperative TF binding.

However, binding alone is often not sufficient to explain TF-mediated gene expression. Notably, TFs bind many thousands of sites in the genome but regulate only a couple of hundred genes (Fisher et al. 2012; X. Y. Li et al. 2008). This indicates that only a subset of the binding sites is in fact functional in terms of regulating the expression of target genes, whereas many binding sites are non-functional or redundant. Therefore, it is not obvious which sites are responsible for the regulation of which genes and mechanisms besides of binding are needed to selectively direct gene regulation by TFs. Moreover, many genes are regulated by long-range interactions with their transcription factors via looping of the DNA, that results in forming contacts between bound transcription factors and the TSS of their target gene (Splinter and de Laat 2011).

Most TFs have been identified based on their DNA binding domain (DBD) which is strongly conserved within different TF families (Weirauch and Hughes 2014). The DBD is the protein domain of TFs that directly interacts with the DNA and the binding motif. Paralog TFs belong to the same family of TFs and evolved during evolution through genomic duplications (Dehal and Boore 2005). Strikingly, given the similarity of their DBD certain paralogs may recognize identical DNA motifs (Badis et al. 2009). However, for other paralogs subtle differences between TFs in their genomic binding preference have been reported which explains the diverging functions of paralogs despite highly similar consensus motifs (Jerković et al. 2017; Sahu et al. 2013). Mechanisms that have been shown to direct divergent binding are subtle differences in the sequence preference (Crocker et al. 2015), cooperativity with other TFs (Slattery et al. 2011; Yáñez-Cuna et al. 2012) or paralog-specific pioneering activity (Bulajić et al. 2020; de Kumar et al. 2017).

Pioneer factors represent a special class of transcription factors. In contrast to conventional TFs that bind to "naked" DNA, pioneer TFs have the physical ability to bind nucleosomal DNA in condensed chromatin often resulting in opening of the chromatin and facilitating transcriptional activation of nearby genes. FoxA was the first protein discovered with pioneering activity (Cirillo L.A. et al. 2002). Purified FoxA is able to bind and locally open condensed chromatin in an ATP-independent manner (Cirillo L.A. et al. 2002). Pioneer factors either directly open the chromatin and enable binding of further TFs or they recruit chromatin remodelers to open the chromatin. Moreover, pioneer factors can recruit histone modifiers, thereby changing the epigenetic state of the chromatin (Espinosa and Emerson 2001). FoxA1 has been shown to recruit the methyltransferase MLL3, which deposits histone H3K4 monomethylation, a histone mark associated with enhancer activity (Jozwik et al. 2016). Furthermore, pioneer factors play an important role in cell differentiation, as they prime silent enhancers in closed chromatin of embryonic stem cells for their tissue-specific activation at a later stage (Zaret and Carroll 2011a).

Today, the most common method to identify transcription factor binding sites *in vitro* or *in vivo* is chromatin immunoprecipitation followed by sequencing (ChIP-seq) (Barski et al. 2007; D. S. Johnson et al. 2007). It is a powerful method to identify genome-wide binding of a DNA-associated protein of interest. In a first step, protein-DNA complexes are stabilized by crosslinking. In a second step, the protein-bound chromatin is shared by sonication. Third, the protein-DNA complexes are immunoprecipitated using an antibody specific for the protein of interest. The DNA-fragments are separated from the bound proteins and the DNA library is analyzed by high-throughput sequencing. Genome-wide protein-DNA interactions can be identified by mapping the obtained sequences back to the genome. The main advantage of this method is that it can be used

to explore genome-wide DNA binding of any protein of interest. Mapping histone modification by ChIP-seq provides valuable information about nucleosome positioning and epigenetic mechanisms. Additionally, it can be applied to define *de novo* transcription factor binding motifs. The weakness of the assay is that it strongly relies on antibody quality. Moreover, it can both capture direct protein-DNA interactions but also indirect interactions, e.g. when a TF is tethered to the DNA via another protein. Some caution in the interpretation of the results of ChIP-seq experiments is warranted given that is has a bias towards open chromatin, because shearing of open chromatin is more efficient. Moreover, identification of binding at repetitive DNA might be challenging because correct mapping of repetitive elements to the genome provides a challenge for the data analysis.

## 1.5 Role of Cofactors

In eukaryotes, most TFs are thought to regulate the expression of genes upon the interaction with cofactors (Lambert et al. 2018a). Cofactors typically lack a DNA binding domain and are recruited by DNA-bound proteins. Moreover, the expression of TFs and cofactors is highly cell-type specific and an indicator of their characteristic functions (Barozzi et al. 2014; Heinz et al. 2010).

Typically, transcriptional cofactors are characterized according to their biological functions. One class of cofactors are histone modifiers. They can induce posttranslational acetylation, methylation, phosphorylation or ubiquitylation of histones and other proteins in a reversible manner. Acetylation of histone tails may regulate transcription by loosening of the nucleosome structure allowing other proteins to bind whereas methylation of histone tails at lysines 9 and 27 is associated with a compaction of the nucleosomes (Lee and Workman 2007; Sims, Nishioka, and Reinberg 2003).

Another class of cofactors are chromatin remodelers. These are large protein complexes which actively reposition nucleosomes in an ATP-dependent manner thereby creating access to the DNA for example at promoters which allows the recruitment of the transcription machinery (Dilworth and Chambon 2001; Jenuwein and Allis 2001; Travers 1999).

Mediator is a large multiprotein complex with modular organization. It is generally required for the transcription by Pol II and is involved in various steps, such as formation of the PIC and Pol II promoter escape (Soutourina 2018). Moreover, it has been shown, that the mediator complex plays a role in DNA loopings thereby creating physical contacts of enhancers with the promoters of active genes (Kagey et al. 2010).

The classical model of transcriptional initiation requires a multistep process. First, a TF binds to its specific transcription factor binding site of an enhancer element. Second, chromatin remodelers

as well as histone modifiers are recruited thereby creating a nucleosome-free locus and modifying histone tails, respectively. And third, the Mediator complex associates with the enhancer region, which induces a looping of the DNA thereby forming a contact of the enhancer with the transcription machinery at the promoter region (Malik and Roeder 2010).

The tight interplay of TFs and their cofactors is an important factor in directing gene regulation. Thus, the knowledge of the associated cofactors is crucial for the understanding of cell-type specific gene regulation. RIME (rapid immunoprecipitation mass spectrometry of endogenous proteins) is a powerful method to identify interacting proteins of a chromatin bound TF (Mohammed et al. 2013, 2016). In a first step, the protein complexes are cross-linked using formaldehyde. Second, the protein complexes are enriched by immunoprecipitation using a specific antibody for the protein of interest. In the final step, the protein complexes are protease digested and analyzed using mass spectrometry. RIME allows the identification of endogenous protein complexes and does not require overexpression of the target protein. However, the limitations of RIME are that it requires high-affinity antibodies and cannot distinguish between direct or indirect interactions.

## 1.6 Steroid Hormone Receptors are Ligand-Inducible Transcription Factors

Steroid hormone receptors are members of the nuclear hormone receptor superfamily and comprise a class of TFs with various biological roles in metabolism, stress response and development. The receptors are ligand-inducible TFs, upon binding of their specific steroid hormones the receptors undergo conformational changes, translocate into the nucleus and bind to specific DNA sites throughout the genome leading to up- or downregulation of their target genes.

The tight regulation of nuclear activity is crucial for normal body function and misregulation of various nuclear receptor members is associated with diseases such as chronic inflammation and cancer (Tenbaum and Baniahmad 1997). Here, I will focus on 3-Ketosteroid receptors, consisting of the glucocorticoid receptor (GR), mineralocorticoid receptor (MR), progesterone receptor (PR) and androgen receptor (AR).

## 1.7 Steroid Hormone Receptors: Structure and Function

Despite their distinct biological functions, steroid receptors have a common multidomain architecture consisting of five protein domains. The amino (N)-terminal domain is the most variable among the steroid receptors and highly disordered. The N-terminus is poorly conserved between the steroid receptors and greatly varies in length and amino acid composition. Moreover, it contains the activation function 1 domain (AF-1) which can be bound by various transcriptional coregulators thereby modulating transcription and acting independent of ligand binding (Godowski et al. 1987).

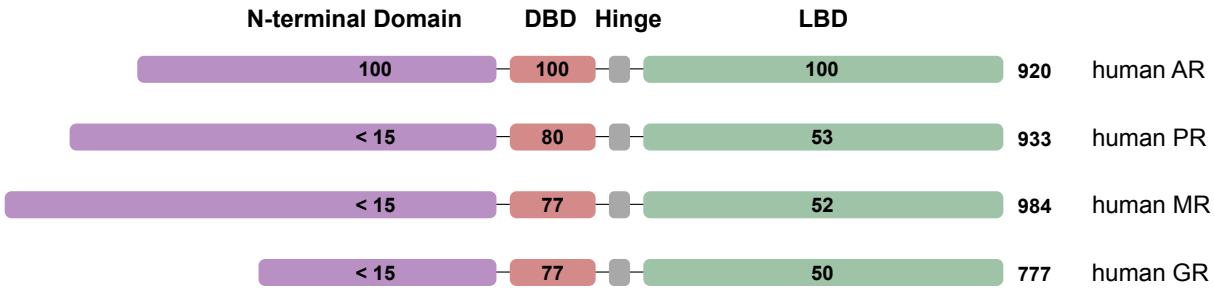



**Figure 1.3. The domain structure of human steroid hormone receptors**. Protein sequence homologies of the human androgen receptor (AR), progesterone receptor (PR), mineralocorticoid receptor (MR) and glucocorticoid receptor (GR) are represented as percent identity in primary sequence. The N-terminal domain, DNA-binding domain (DBD), hinge region and the ligand-binding domain (LBD) are shown. The figure was adapted from McEwan and Brinkmann 2000.

The DBD is the most conserved protein domain of steroid receptors and has an essential role in target gene recognition. The DBD contains two zinc finger subdomains (P and D box). The P box comprises the region that directly interacts with the DNA whereas the D box is responsible for receptor dimerization (Green and Chambon 1987; Umesono and Evans 1989). Because of the high conservation of the P box the steroid receptors, GR, MR, PR and AR all recognize virtually identical palindromic hexamers separated by a 3 bp spacer 5'-AGAACA-3', which represents the consensus motif or response element (Beato et al. 1989; Schwabe et al. 1993). The hinge region is a small linker domain connecting the DBD to the ligand-binding domain (LBD). The C-terminal LBD is crucial for receptor activation and can be directly bound by specific steroid hormones as well as coregulator proteins (Mangelsdorf et al. 1995).

Prior to hormone induction, the unliganded steroid receptors are located in the cytosol and are bound by chaperone proteins to stabilize their conformation (Pratt and Toft 1997). Upon binding of small lipophilic steroid hormones, the steroid receptors undergo conformational changes, homodimerize and translocate into the nucleus, where the receptors bind to regulatory regions

throughout the genome (Guiochon-Mantel et al. 1996). Furthermore, the steroid receptors interact with other basal transcription factors and transcriptional coregulators, thereby influencing the assembly of the transcriptional machinery and leading to transcriptional regulation of their target genes.

There are many reasons why steroid receptors have been studied extensively for decades. First, steroid receptors are highly conserved and abundant among metazoans arguing for their fundamental cellular functions. Second, the property to induce gene expression in a hormone-dependent manner makes steroid receptors a unique model system to study general mechanisms of transcriptional regulation in a controlled fashion by simply adding the specific steroid hormone. And third, misregulation of steroid receptors is associated with metabolic diseases, chronic inflammation and various cancer types. Steroid hormones are widely used therapeutically to treat a broad range of conditions such as inflammatory disorders, osteoporosis or hormone sensitive cancers (Clark and Belvisi 2012; Heidenreich et al. 2014; Vanderschueren et al. 2014). However, long-term application of steroids is often accompanied with side effects or resistance to treatment. Therefore, a detailed understanding of steroid receptor function is crucial for the development of new therapies.

## 1.8 Paralogous Transcription Factors AR and GR

Paralogous TFs have evolved during evolution upon genomic rearrangements such as DNA duplications. Some paralog TFs retained redundant activities, however, most paralogous TFs developed very specific functions (Kribelbauer et al. 2019; L. N. Singh and Hannenhalli 2008). AR and GR are close members of the nuclear receptor superfamily and are activated by androgens and glucocorticoids, respectively. Strikingly, they share a similar multidomain structure and recognize the identical consensus motif *in vitro* (Claessens, Joniau, and Helsen 2017). However, the two receptors have divergent, sometimes even opposing functions in the body. AR plays a fundamental role in the development and maintenance of the reproductive tissue. GR, on the other hand, has essential functions in metabolism and the immune system. Glucocorticoids are widely used in clinics for the treatment of inflammatory diseases (Clark and Belvisi 2012). However, elevated levels of glucocorticoids are associated with osteoporosis and muscle waste (Klein 2015) whereas androgens are associated with an increase in bone density and muscle mass (R. Singh et al. 2006; Sinha-Hikim et al. 2004).

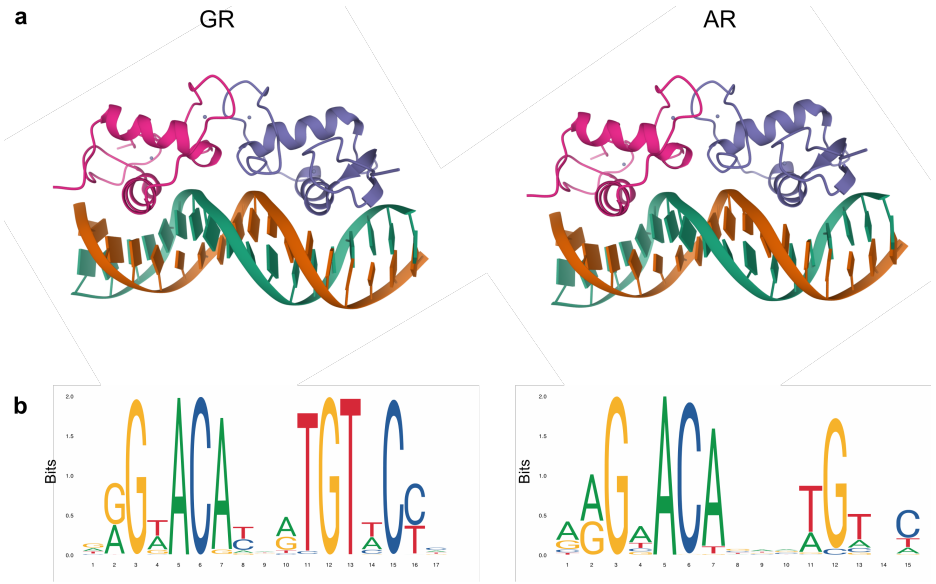



**Figure 1.4. Crystal structures of the GR and AR DBD in a complex with the DNA binding sites.** **(a)** The crystal structures of rat GR DBD dimer bound to a palindromic repeat (Luisi et al. 1991) and rat AR DBD dimer bound to a direct repeat binding site (Shaffer et al. 2004) are shown. Both receptors show a head-to-head conformation of the receptor dimers. The crystal structures were deposited at the Protein Data Bank archive (PDB) (Berman, Henrick, and Nakamura 2003) **(b)** Sequence logo representing the human GR (Matrix ID MA0113.3) and human AR (Matrix ID MA0007.2) binding motifs from the JASPAR CORE database (Khan et al. 2018).

It is intriguing that both AR and GR have structurally conserved DBSs and virtually the same DNA motif preference yet direct distinct regulatory programs (Fig 1.4). It has been shown that the specificity of transcription factors is directed on different levels of regulation such as small variations in DNA binding, differential interactions with cofactors and the chromatin environment. Partially, the specificity of AR and GR can be explained by their tissue-specific expression. Diverse tissue-specific expression has been previously shown to mediate differential gene regulation of paralogous TFs (L. N. Singh and Hannenhalli 2008b). AR is crucial for the normal function of the prostate tissue, where GR is not expressed under normal conditions. However, GR has been shown to be upregulated in castration-resistant prostate cancer and consequently becomes co-expressed with AR (Arora et al. 2013).

Differences in gene regulation can also be caused by small variations in DNA binding. Although, it was previously demonstrated that both receptors recognize the identical consensus motif *in vitro*, ChIP experiments have shown that both receptors have overlapping as well as divergent binding sites when AR and GR are coexpressed in the same cell line (Arora et al. 2013; Sahu et al. 2013). Moreover, it was shown that AR is able to bind a motif representing the direct repeat of 5'-AGAACA-3' unlike the classical palindromic consensus which consists of an inverted repeat (Schoenmakers et al. 2000; Zhou, Corden, and Brown 1997). Strikingly, this direct repeat is not recognized by GR. A comparison of the crystal structures of the two receptors has revealed that

small differences in the dimerization domain and an additional hydrogen bond likely increase the interaction strength of the AR DBD with the DNA (Shaffer et al. 2004).

Of note, binding to differential bindings sites can direct receptor specific gene regulation. However, binding fails to explain receptor-specific gene regulation near genes with overlapping binding sites. Interestingly, when AR and GR are coexpressed some genes show an overlapping binding but none-the-less are regulated by only one receptor (Arora et al. 2013; Sahu et al. 2013). But how can differential gene expression be explained when both receptors show overlapping binding?

Besides binding, the recruitment of cofactors and transcriptional coregulators is another driver of receptor-specific gene expression (Wiench, Miranda, and Hager 2011). There is little to no sequence conservation of the NTD between AR and GR (Lavery and McEwan 2005). Nevertheless, the NTD has been found to be critical for steroid receptor-specific gene regulation. Specifically, steroid receptors "communicate" via the NTD with the transcription machinery as well as various cofactors thereby regulation transcription. Due to their highly variable NTD, receptor-specific recruitment of cofactors might provide an explanation for the differential gene regulation by AR and GR observed downstream of genomic binding.

Furthermore, differential interaction with chromatin remodelers allows to create a permissive chromatin environment and further binding of TFs and cofactors enabling transcriptional activation. The interaction with FoxA1, a pioneer factor, has been shown to differentially direct AR and GR binding in a cell type specific manner. FoxA1 interacts specifically with AR in LNCaP cells but GR in the VCaP prostate cancer cell line (Sahu et al. 2013).

Finally, the chromatin environment plays a crucial role in directing receptor-specific gene expression. The binding specificity of steroid receptors is determined by the chromatin accessibility, which is highly cell-type specific (John et al. 2011). However, different abilities to access enhancers within nucleosomal regions and to induce nucleosomal rearrangements might provide a mechanism for receptor-specific gene regulation. It has been demonstrated repeatedly that GR has pioneering potential and is able to bind inaccessible sites (T. A. Johnson et al. 2018; Wu et al. 2015). But also, in case of AR, it was proposed that it is able to initiate opening of the chromatin (He et al. 2010).

Despite decades of studies and a wealth of knowledge regarding their working mechanism, many questions still remain. For example, it is still unclear how receptor-specific binding is achieved for

AR and GR when both receptors are coexpressed in the same cell line, although both receptors appear to recognize identical DNA motifs. Moreover, cofactor recruitment seems to be highly context-specific and therefore an important driver of receptor-specific gene regulation. A better understanding of gene regulation by AR and GR is particularly relevant for understanding of the general rules of transcriptional regulation by related TFs but also for the clinical application of steroids. Prostate cancer is classically treated with antiandrogens, unfortunately, resistance to hormone therapy is a common side effect of the treatment. GR has been shown to substitute for AR in castration-resistant prostate cancer by regulating androgen-specific pathways (Arora et al. 2013; Isikbay et al. 2014). Therefore, a detailed understanding of differential gene regulation by AR and GR could lead to new therapeutic approaches.

## 1.9 Aim of the Thesis

The aim of this work is to provide insights into mechanisms that are responsible for functional diversification of two paralogous transcription factors, AR and GR, that have unique functions despite their nearly identical DNA binding preferences. Therefore, I used an osteosarcoma cell line either expressing AR or GR as model system where both receptors can be studied in an identical cellular background. Specifically, I have examined the roles of genomic binding, the chromatin environment and cofactor interactions in directing receptor-specific gene regulation by AR and GR.

In the first part, my objective was to examine if selective gene regulation by AR or GR can be explained by differences in their genomic binding patterns. Using bioinformatical approaches the genes upregulated by either AR or GR were intersected with receptor-specific binding. In addition, I have used CRISPR/Cas9-mediated genome-editing to disrupt receptor-specific binding sites and functionally test whether they contribute to selective regulation of the nearby gene by one of the receptors.

Further, I wanted to investigate whether AR and GR display different properties to interact with the chromatin environment. Therefore, I examined the chromatin accessibility as well as changes of various histone modifications in response to AR and GR signaling. Moreover, I investigated whether AR and GR show different abilities to interact with genomic enhancers.

In the next part of the thesis, I wanted to test whether AR and GR show different preferences to their binding sequences which might contribute to differential gene regulation by the two receptors.

Lastly, I set out to investigate if differential cofactors interactions play a role directing receptor-specific gene regulation by AR and GR.

# 2. Materials and Methods

## 2.1 General Materials

If not stated differently all general chemicals and lab-ware were purchased from the following suppliers: Calbiochem, Carl Roth, Eppendorf, Invitrogen, Merck, Sarstedt, Sigma Aldrich and TPP.

### 2.1.1 Cell Lines

**U2OS** Human osteosarcoma cell line derived from a 15-years old female

**GR18** Cell line was derived from U2OS cells and stably expresses the GR isoform GR$\alpha$ from rat. (Rogatsky, Trowbridge, and Garabedian 1997)

### 2.1.2 DNA Plasmids

**AR vector** Vector containing human AR gene (Farla et al. 2004) *

**STARR-Seq vector** Human STARR-seq screening vector containing sgGFP (Arnold et al. 2013)

**PX459** Vector containing Cas9 from S. pyogenes and cloning backbone for sgRNA (Ran et al. 2013)

* Following changes were performed by Christine Helsen. EGFP and 510bp of the *hAR* were cut out using *NheI* and *BspTI*. ATG and a *NheI* cutting side were added at 5' end of *hAR* by PCR with primers containing a *BspTI* overhang. Insert with 510 bp of the *hAR* was ligated using *NheI* and *BspTI* sites.

### 2.1.3 Antibodies

**Table 2.1 Antibodies**

| Antibody | Description | Supplier |
|---|---|---|
| Actin | I-19 rabbit polyclonal anti-Actin | Santa Cruz Biotechnology, Cat. No. sc-1616 |
| AR | PG-21 rabbit polyclonal anti-AR | Merck, Cat. No. 06-680 |
| HRP-goat anti rabbit | Secondary antibody | Invitrogen, Cat. No. 656120 |
| N499 | Polyclonal rabbit anti-GR, raised against the N-terminus of human GR (residues 1- 499) | Produced by R. M. Nissen, B. Darimont and K. R. Yamamoto |
| H3K27Ac | Polyclonal rabbit anti-H3K27Ac | Diagenode, Cat. No. C15410196 |
| H3K4me1 | Polyclonal rabbit anti-H3K4me1 | Diagenode, Cat. No. C15410194 |
| H3K4me3 | Polyclonal rabbit anti-H3K4me3 | Diagenode, Cat. No. C15410003 |
| H3K9me3 | Polyclonal rabbit anti-H3K9me3 | Diagenode, Cat. No. C15410193 |
| H3K27me3 | Polyclonal rabbit anti-H3K27me3 | Diagenode, Cat. No. C1541095 |
| H3K36me3 | Polyclonal rabbit anti-H3K36me3 | Diagenode, Cat. No. C15410058 |
| H3K79me2 | Polyclonal rabbit anti-H3K79me2 | Active Motif, Cat. No. 39143 |
| Med1 (Anti-CRSP1/TRAP220) | Polyclonal rabbit anti-Med1 | Bethyl Laboratories, Cat. No A300-793A |
| P300 | Monoclonal mouse antibody raised against the C-terminus of human P300 | Diagenode, Cat. No. C15200211 |
| GR | Monoclonal rabbit anti-GR | Cell Signalling Technology, Cat. No. 12041 |
| IgG | Normal mouse IgG | Santa Cruz Biotechnology, Cat. No. sc-2027 |

## 2.2 Bacteria Transformation, Bacterial Cell Culture and Plasmid Isolation

Bacteria from the strain DH5 Alpha were made chemically competent using the Mix & Go kit (Zymo Research) and stored at -80°C. The competent cells were thawed on ice, mixed with 1-5 µl of plasmid DNA and incubated for 5 min on ice. The bacteria were resuspended in 250 µl SOC medium and incubated for 1 h at 37 °C. Then 100 µl were plated on LB-Agar plates containing antibiotics. Bacteria with plasmids containing an ampicillin resistance cassette were plated directly on LB-Agar plates without SOC medium.

Bacterial liquid cultures were grown in LB medium containing antibiotics. The liquid cultures were grown overnight shaking at 37 °C and 190 rpm. Plasmid preparations were carried out using the Mini, Midi and Maxi Plasmid Kit (Qiagen) or the NucleoSpin Plasmid Kit (Macherey-Nagel). For long-term storage of the bacteria, glycerol stocks were prepared. 1 ml of bacterial liquid culture was mixed with a 60 % glycerol solution for a final glycerol concentration of 30 %. The bacteria were snap frozen and stored at -80 °C.

**Table 2.2 SOC Medium**

| | |
|---|---|
| Bactotryptone | 20 g/l |
| Yeast extract | 5 g/l |
| Magnesium choride | 10 mM |
| Sodium chloride | 10 mM/l |
| Potassium chloride | 0,5 mM |
| Magnesium sulfate | 10 mM |

**Table 2.3 LB Agar**

| | |
|---|---|
| Agarose | 15 g/l |
| Sodium chloride | 10 g/l |
| Bactotryptone | 10 g/l |
| Yeast extract | 5 g/l |

**Table 2.4 LB Medium**

| | |
|---|---|
| Sodium chloride | 10 g/l |
| Bactotryptone | 10 g/l |
| Yeast extract | 5 g/l |

## 2.3 Cell Culture

U2OS cells were grown in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 5 % fetal bovine serum (FBS, Gibco) at 37 °C and 5 % $CO_2$. To store the cells, the cells were resuspended in cryopreservation medium, FBS containing 10 % dimethyl sulfoxide (DMSO, Serva), transferred to cryovials and frozen in an isopropanol freezing container at -80 °C overnight. For long-term storage, the samples were moved to a liquid nitrogen storage container.

### 2.3.1 Transfection of U2OS Cells with Lipofectamine for Stable AR Integration

**Table 2.5 DNA Plus Mix**

| | |
|---|---|
| Plus reagent (Invitrogen) | 1,6 µl |
| AR plasmid (30 ng/µl) | 1 µl |
| Serum-free DMEM (Gibco) | 25 µl |

**Table 2.6 Lipofectamine Mix**

| | |
|---|---|
| Lipofectamine (Invitrogen) | 0,8 µl |
| Serum-free DMEM (Gibco) | 25 µl |

The AR vector containing the human *AR* gene was a kind gift from Frank Claessens and Christine Helsen. A day before transfection, 40.000 cells of the U2OS cell line were seeded per well of a 24-well plate and incubated overnight. Next, the DNA Plus Mix and Lipofectamine Mix were prepared and incubated for 15 minutes at room temperature. Then both solutions were mixed, vortexed and incubated a second time for 15 minutes at room temperature. Meanwhile, the medium was aspirated from the cells, cells were washed with 500 µl PBS and 200 µl serum-free DMEM was added per well. 50 µl of the transfection mix was added to the cells per well and incubated for three hours at 37 °C and 5 % $CO_2$. The medium was changed to DMEM with 5 % FBS and a second time after another three hours.

Next day, the cells were washed once with 500 µl PBS and 250 µl Trypsin were added. The trypsin was neutralized with 250 µl DMEM with 5 % FBS, the cells were carefully resuspended by pipetting up and down and 100 µl of the cell suspension was transferred to a 15 cm petri dish. Stably transfected cells were then selected by adding 320 µl of G418 (800 µg/µl) per 20 ml medium until we obtained single-cell derived colonies. Then, single cell-derived clonal cells were grown and

tested for AR expression, where one cell line stably expressing AR was found. Successful generation of an AR-U2OS cell line was tested by qPCR and Western blot

### 2.3.2 Transfection of U2OS Cells with Amaxa Nucleofector

U2OS cells were transfected with the Amaxa Cell Line Nucleofector Kit V (Lonza) according to the manufacturer's manual. Transfection using the Amaxa Kit led to more than 90 % of the living cells being transfected. For one transfection, 1 million cells were transfected with 2 μg DNA and Nucleofector 2b Device was used. After transfection, samples were transferred to two wells of 6-well plate.

## 2.4 Polymerase Chain Reaction (PCR)

DNA amplification was carried out by Polymerase Chain Reaction (PCR) using a 2x Phusion Master Mix from NEB. The PCR reaction was made according to the following table.

**Table 2.7 PCR Reaction Mix**

| | |
|---|---|
| Template DNA (30ng/μl) | 1 μl |
| Forward primer (10 μM) | 1 μl |
| Reverse primer (10 μM) | 1 μl |
| Phusion Master Mix 2x | 10 μl |
| $H_2O$ | 7 μl |
| Total volume | 20 μl |

The PCR reaction was carried out in a thermocycler according to the following cycling program. The annealing temperature was chosen according to the melting temperature (Tm) of the used primer pair and the elongation time was set to 30 sec per kb of the desired PCR product.

**Table 2.8 PCR Thermocycling Program**

| | Temperature | Duration |
|---|---|---|
| Initial Denaturation | 98 °C | 30 s |
| Cycles 35x | | |
| Denaturation | 98 °C | 10 sec |
| Annealing | 55-70 °C | 10-30 sec |
| Elongation | 72 °C | 30 sec per kb |
| Final Elongation | 72 °C | 10 min |
| Storage | 4 °C | ∞ |

## 2.5 DNA Sanger Sequencing

Sanger sequencing of plasmid DNA, genomic DNA or specific PCR products was conducted by Eurofins. The samples were mixed with the corresponding primer and sent to Eurofins.

## 2.6 mRNA Isolation and cDNA Synthesis

Purification of RNA was performed with the RNAeasy Mini Kit (Qiagen) according to the manufacturer's manual. Instead of beta-Mercaptoethanol, 40 µl DTT (1 M) were added to 1 ml RLT lysis buffer to inactivate intracellular RNAses released upon the lysis step. RNA Elution was completed in 50 µl RNAse-free $H_2O$. cDNA synthesis was carried out with the PrimeScript One Step Kit (Takara) according to the manufacturer's manual. For cells transfected with a STARR-Seq reporter, specific primers MB108_STARR and MB108_RPL19 were used for targeted reverse transcription (Tab. S12). For all other applications, cDNA synthesis was performed with the Random Primer and Oligo dT primer provided by the PrimeScript Kit. 325 ng RNA were used per reaction. The cDNA was diluted to 5 ng/µl with $H_2O$ for subsequent qPCR analysis.

## 2.7 Quantitative Real-Time PCR

DNA quantification by quantitative real-time PCR (qPCR) was performed to analyze ChIP experiments, ATAC experiments and cDNA. The reaction was set up with 2 µl of template DNA, 5 µl home-made qPCR Master Mix and 3 µl primer mix (0,66 µM forward and reverse primer) in a 10 µl reaction volume. All samples were quantified in duplicates. The qPCR reactions were run on an ABI 7900 HT machine (Applied Biosystems) or the QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystems).

The primer sequences for quantification of gene expression are listed in Tab. S1 and for ChIP quantification in Tab. S2-S4.

**Table 2.9 qPCR Master Mix**

| | |
|---|---|
| Tris pH 8.3 | 100 mM |
| MgCl2 | 6 mM |
| BSA | 1 mg/ml |
| dNTPs | 4 mM |
| SYBR-Green (Invitrogen) | 0.66x |
| ROX reference dye (Invitrogen) | 1x |
| Perpetual Taq (EURx) | 0.2 U/µl |

**Table 2.10 qPCR Thermocycling Program**

| | Temperature | Duration | Ramp rate |
|---|---|---|---|
| Initial Denaturation | 95 °C | 10 min | 100 % |
| **Cycles 40x** | | | |
| Denaturation | 95 °C | 15 sec | 100 % |
| Annealing/Elongation | 60 °C | 60 sec | 100 % |
| **Dissociation step** | | | |
| Denaturation | 95 °C | 15 sec | 100 % |
| Annealing/Elongation | 60 °C | 15 sec | 100 % |
| Denaturation | 95 °C | 15 sec | 2 % |

## 2.8 Polyacrylamide Gel Electrophoresis and Western Blotting

The protein sample was mixed with 6x sample buffer and heated for 5 min at 95 °C. 10 µl of the samples or 5 µl PageRuler Plus prestained protein ladder (10 to 250 kDa; ThermoFisher) were loaded on a NuPage Gradient 4-12% Bis-Tris Mini Gel (Invitrogen). The gel was run in MES SDS Running Buffer 20x (Invitrogen) diluted to 1x with $H_2O$ for 40 min at 150 V.

**Table 2.11 Transfer Buffer**

| | |
|---|---|
| Trizmabase | 3,03 g |
| Glycine | 14,4 g |
| H2O | 500 ml |
| Methanol | 140 ml |
| H2O | ad 1000 ml |

**Table 2.12 TBST Buffer**

| | |
|---|---|
| Tris | 50 mM |
| NaCl | 500 mM |
| Tween 20 | 0,05 % |

For the protein transfer, a PVDF membrane (Millipore) was activated by soaking it in 100 % methanol and incubation for 2 min in transfer buffer. The protein sandwich was assembled using the PVDF membrane, the protein gel and Whatman blotting paper and transferred with transfer buffer using a BioRad blotting chamber and run at 150 V for 1,5 h at 4°C. A magnetic stir and radiator block were added. The membrane was blocked with a 5 % BSA-TBST solution for 1h on a rotating shaker at room temperature. The membrane was washed with 3x TBTS and 1x with 5%

BSA-TBST solution. The primary antibody (Actin or AR) was diluted 1:1000 in BSA-TBST and the membrane incubated over night at 4°C. Next, the membrane was washed 3x with TBTS and 1x with 5% BSA-TBST solution. The secondary antibody (HRP-goat anti rabbit) was diluted 1:4000 in BSA-TBST and the membrane was incubated for 1h at room temperature. The membrane was then washed 5x with TBST for 5 min. SuperSignal West Dura Extended Duration Substrate (ThermoFisher) was used for signal detection using a LAS1000 Fujifilm camera.

## 2.9 Hormone Binding Assay

The hormone binding assay was performed to quantify the cellular androgen receptor or glucocorticoid receptor content in U2OS-AR cells and U2OS-GR cells respectively. The assay was performed as described (Meijsing et al. 2007).

25.000 U2OS-GR or U2OS-AR cells per well were seeded on a 24-well-plate. To deplete the medium for steroid hormones, cell culture medium was replaced with DMEM containing 5 % charcoal-stripped FBS one day before the assay. The cells were washed three times with hormone-free medium for 5 min. Next, U2OS-GR cells were treated with 100 nM [$^3$H-]Dex (81 Ci/mmol, PerkinElmer) in the presence or absence of a 10 μM (100 x) excess of unlabeled Dex for 45 min. U2OS-AR were treated with 100 nM [$^3$H]-R1881 (81,4 Ci/mmol, PerkinElmer) in the presence or absence of a 10 μM (100 x) excess of unlabeled R1881 for 45 min. The cells were then washed five times with cold PBS and replaced by 250 μl cold ethanol to extract the steroids. To measure the radioactivity, 250 μl ethanol cell mix was added to 5 ml scintillation cocktail (Rotiszint Eco Plus, Carl Roth) and measured in a liquid scintillation counter (Wallac 1409, model 1409-001). Additionally, 1 μl of each labeled hormone was measured in 5 ml scintillation cocktail to determine the counts per nM steroid hormone. To determine the specific binding of AR or GR, nonspecifically bound radioactivity was subtracted from the total bound radioactivity. The experiments were performed in triplicates for both cell lines.

## 2.10 Calculation to Determine the Bound Receptor Fraction

The binding affinity is the measure for the ability of a molecule to interact with its ligand. Once the ligand (*L*) associates with its receptor (*R*) it forms a non-covalent complex ($R \cdot L$).

$$R \cdot L \leftrightarrows R + L \qquad \text{E.1}$$

The dissociation constant ($K_d$) represents the tendency of the complex to dissociate into its components. Once association and dissociation reach an equilibrium, the dissociation constant can be determined with the following equation.

$$K_d = {[R][L]}/{[R \cdot L]} \qquad \text{E.2}$$

To determine the $K_d$, usually the concentration of a ligand is being varied with a fixed receptor number and the fraction of bound receptor (*Y*) is determined. $R_{total}$ represents the total receptor amount.

$$Y = {[R \cdot L]}/{[R]_{total}} \qquad \text{E.3}$$

Usually the ligand is present at much higher concentrations compared to the receptor. In this case, the following equation can be applied. It assumes that the receptor concentration is negligible compared to the ligand concentration and ranges the fraction of bound receptor from zero to one.

$$Y = {[L]}/{(K_d + [L])} \qquad \text{E.4}$$

The $K_d$ values were previously determined experimentally (Cleutjens et al. 1997) and were used for an approximate calculation of the receptor occupancy at a given hormone concentration. The determined $K_d$ for R1881 binding to AR is around 1,1 nM and 3,1 nM for Dex binding to GR The following equation shows an example calculation for the fraction of bound GR in the presence of 1 nM Dex.

$$Y = {1\ nM}/{(3{,}1\ nM + 1\ nM)} = 0{,}24 \qquad \text{E.5}$$

A fraction of 0,24 bound receptor equals around 89,542 molecules in the case of GR.

**Table 2.13 Bound GR fraction at different hormone concentrations**

| Dex concentration | Bound receptor fraction | Total bound receptor molecules |
|---|---|---|
| 100 nM | 1 | 373.090 |
| 5 nM | 0,61 | 227.585 |
| 1 nM | 0,24 | 89.542 |
| 0,5 nM | 0,14 | 52.233 |
| 0,1 nM | 0,03 | 11.193 |
| 0,01 nM | 0,003 | 1.119 |

**Table 2.14 Bound AR fraction at different hormone concentrations**

| R1881 concentration | Bound receptor fraction | Total bound receptor molecules |
|---|---|---|
| 100 nM | 1 | 120.151 |
| 5 nM | 0,82 | 98.524 |

## 2.11 Chromatin Immunoprecipitation

ChIP assays for GR using the N499 antibody were performed as described (Meijsing et al. 2009). ChIPs targeting the androgen receptor were done with 2 µl anti-rabbit polyclonal antibody PG-21 (Anti-AR; Sigma Aldrich; 06-680). ChIP experiments for qPCR analysis targeting H3K27ac were performed using 0,5 µl of the Diagenode polyclonal rabbit antibody (H3K27Ac; C15410196). ChIP experiments targeting Med1 and P300 were carried out with 5 µl antibody.

### Hormone induction and harvest

Five million GR cells were grown on a 10 cm petri dish and induced with 1 µM Dex or 0,1 % EtOH as vehicle control for 1,5 h. AR cells were treated with 5 nM R1881 or 0,1 % DMSO as a vehicle. After hormone induction, cells were cross-linked with a 37 % formaldehyde solution and a final concentration of 1 % at RT for 3 minutes (10 minutes for the Med1 and P300 ChIP). The reaction was quenched with 2,5 M glycine to a final concentration of 200 mM. Next, the cells were incubated at 4 °C for 10 minutes, washed with 20 mL ice-cold PBS, scraped off and transferred into a falcon tube. The cells were pelleted by centrifugation at 645 g 4 °C for 5 minutes to be snap frozen and stored at -80 °C.

### Lysis and chromatin fragmentation

Samples were thawed on ice and 2 mL ice-cold IP lysis buffer was added (50 mM HEPES-KOH pH 7,4, 1mM EDTA, 150 mM NaCl, 10% Glycerol, 0,5 % Triton X-100, 1:200 Protease inhibitor, 1:200 PMSF). The suspension was nutated at 4 °C for 30 minutes. The nuclei were pelleted by centrifugation 645 g 4 °C for 5 minutes. Next, cells were resuspended in 300 µl ice-cold RIPA buffer (10 mM Tris-HCL, pH 8, 1 mM EDTA, 150 mM NaCl, 5 % Glycerol, 0,1 % DOC, 0,1 % SDS, 1 % Triton-X100, 1:200 Protease inhibitor, 1:200 PMSF) and sonicated using a Diagenode Bioruptur using 24 cycles (30 sec on, 30 sec off) at 4 °C (27 cycles for the Med1 and P300 ChIP). Samples were centrifuged at max speed for 15 minutes at 4 °C. The lysate was resuspended in 400 µl RIPA buffer and 10 % of the sample are set aside as the input control.

### Immunoprecipitation and reversal of the cross-linking

Next, 2 µl of N499 antibody was added and incubated rotating overnight at 4 °C. 30 µl of a 50% agarose bead slurry (Santa Cruz protein A/G agarose beads) were washed 2x 750 µl RIPA buffer and nutated overnight at 4 °C. RIPA buffer was removed to obtain a 50 % bead slurry and 30 µl were added to each sample and incubated for 4 h at 4 °C on a rotation wheel. The beads were pelleted by centrifugation and washed 5x with 1 ml RIPA wash buffer containing 500 mM NaCl. Each ChIP sample was then supplemented with Proteinase K solution (TE buffer pH 8, 0,7 % SDS, 200 µg/ml proteinase K) and incubated for at 55 °C for 3 h and 65 °C overnight. The next day, 1 µl RnaseA (2 µg/ml) was added and incubated at 25 °C for 1 h. The ChIP samples were purified using the PCR Purification Kit (Promega, Wizard SVGel and PCR Clean-Up system) and eluted using 100 µl RNAse-free water.

### ChIP-Seq for Histone Marks

ChIP-Seq experiments targeting histone marks in U2OS-AR and U2OS-GR18 cells were performed by Alisa Fuchs. The following antibodies were used (1µg per ChIP reaction): anti-H3K4me1 (Diagenode C15410194), anti-H3K4me3 (Diagenode C15410003), anti-H3K9me3 (Diagenode C15410193), anti-H3K27me3 (Diagenode C1541095), anti-H3K27ac (Diagenode C15410196), anti-H3K36me3 (Diagenode C15410058) and anti-H3K79me2 (Active Motif 39143).

Three million cells were grown in a petri dish for each condition. Hormone induction and harvest of the cells was performed as described above. The cell pellets were gently passed through a 25-

gauge needle and resuspended in 600 µl RIPA Plus (TE Buffer with 0,1%SDS; 1:200 Complete protease inhibitor [Roche], 1:200 PMSF) and incubated for 20 minutes on ice. Each 200 µl were then transferred to a 1,5 ml Bioruptur PicoTube and sheared for 30 cycles. Next, the samples were centrifuged at max. speed for 15 minutes at 4°C and the chromatin pooled from each condition. For the immunoprecipitation 200 µl sa,ple were taken for each IP and 20 µl were kept for the input fraction. The IP was performed using the IP-Star automated system (Diagenode). Next, the sample and input fraction was adjusted with 100 µl ChIP elution buffer, 4 µl 5M NaCl were added and incubated at 65 °C overnight. Then, 4 µl RNAseA 5 mg/ml were added and incubated 1 h at 55°C. 3 µl Proteinase K were then added and the samples were incubated for 4 h at 55 °C. Lastly, the DNA was purified with the Zymo Clean and Concentrator Kit and eluted in 20 µl elution buffer.

Sequencing libraries were prepared using the NEBNext Ultra DNA Library Prep kit (NEB E7370) according to manufacturer's instructions and submitted for sequencing using the HiSeq2500 sequencer. Paired-end Illumina sequencing was carried out with read-length of 50 bp.

## 2.12 Omni-ATAC-Seq

Omni ATAC was performed as described (Corces et al. 2017). First, 250.000 U2OS-GR or U2OS-AR cells were seeded in a 6-well plate per well and grown overnight. GR cells were induced with 1 µM Dex or 0,1 % EtOH as vehicle control for 1,5 h. AR cells were treated with 5 nM R1881 or 0,1 % DMSO as a vehicle.

**Table 2.15 Resuspension Buffer (RSB)**

| | Final Concentration | Volume |
|---|---|---|
| Tris-HCl, pH 7.4 (1M) | 10 mM | 500 µl |
| NaCl (5M) | 10 mM | 100 µl |
| $MgCl_2$ (1M) | 3 mM | 150 µl |
| Tween-20 (10%) | 0.1 % | 500 µl |
| $H_2O$ | | 48,75 ml |
| Total volume | | 50 ml |

**Table 2.16 Resuspension Buffer Plus (RSB)**

| | Final Concentration | Volume |
|---|---|---|
| Igepal CA-630 (10%) | 0.1 % | 2.5 µl |
| Digitonin (1%) | 0.01 % | 2.5 µl |
| RSB | | 245 µl |
| Total volume | | 250 µl |

**Table 2.17 Transposase Mix**

| | Final Concentration | Volume |
|---|---|---|
| TD Nextera buffer 2x | 1x | 25.0 µl |
| TDE1 (Nextera Tn5 Transposase) | 100 nM | 2.5 µl |
| Digitonin (1%) | 0.01% | 0.5 µl |
| Tween-20 (10%) | 0.1% | 0.5 µl |
| PBS | | 16.5 µl |
| $H_2O$ | | 5.0 µl |
| Total Volume | | 50µl |

To prepare the nuclei, the cells were washed with PBS, trypsinized with 400 µl trypsin and 1600 µl DMEM were added. 100.000 viable cells were counted and pelleted by centrifugation at 500 g for 5 minutes at 4 °C. The supernatant was carefully aspirated. 50 µl ice-cold Resuspension Buffer Plus were added and pipetted up and down three times. The samples were incubated on ice for 3 minutes. Next, the samples were washed with 1 ml Resuspension Buffer and the tube inverted three times. Then the nuclei were pelleted by centrifugation at 500 g for 10 minutes at 4 °C. All supernatant was carefully aspirated.

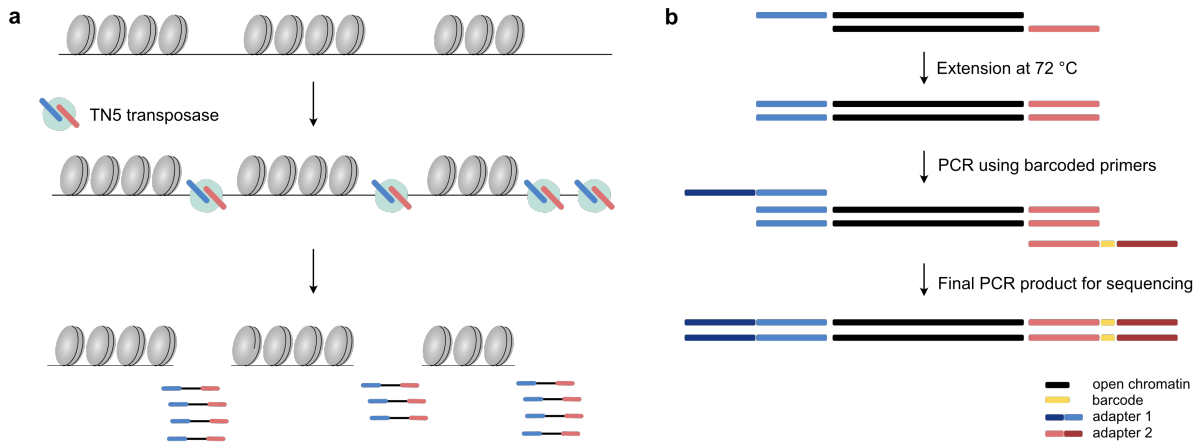



**Figure 2.1. Detection of open DNA using ATAC-seq**. (a) Nextera adapters are inserted by the TN5 transposase into open chromatin regions thereby fragmenting the DNA (b) In the first step of the PCR, both ends of the DNA fragments are completed. In the second step, barcoded primers containing sequencing adapters are added. The final DNA fragments are amplified in the last PCR step. Adapted from (Buenrostro et al. 2015).

The TN5 transposase is pre-loaded with Nextera adapter sequences. During the transposition reaction, the adapters are inserted into the DNA resulting in the fragmentation of the DNA. For the transposition reaction, the pellet was resuspended in 50 µl Transposition mix by pipetting up and down six times and incubated in a thermomixer at 37 °C for 30 minutes at 1000 rpm. The reaction was immediately stopped with 2,5 µl 10 % SDS. The DNA purification was carried out with the Zymo Clean and Concentrator Kit (Ref. D4014) according to manufacturer's protocol and the samples were eluted with 22 µl Elution buffer.

To amplify the transposed fragments, 20 µl of the sample were used for the following amplification PCR reaction. To establish the number of cycles was a critical step for the ATAC experiment and had to be optimized for our cell line. The primers used for the ATAC enrichment PCR are listed in Tab. S5. The universal primer was combined with one barcoded primer.

**Table 2.18 PCR Mix**

| | |
|---|---|
| Transposed DNA | 20.0 µl |
| Custom Nextera Primer 1 (Mb032, 25 µM) | 2.5 µl |
| Custom Nextera Primer2 (25 µM) | 2.5 µl |
| NEBNext High-Fidelity 2x PCR Master Mix | 25.0 µl |
| Total volume | 50 µl |

**Table 2.19 PCR Thermocycling Program**

| | Temperature | Duration |
|---|---|---|
| Annealing/Elongation | 72°C | 5 min |
| Denaturation | 98°C | 30 sec |
| Cycles **11x** | | |
| Denaturation | 98°C | 10 sec |
| Annealing | 63°C | 30 sec |
| Final Elongation | 72°C | 1 min |

The DNA purification of the ATAC libraries was performed using AMPure XP beads (Agencourt) according to the manufacturer's protocol. A two-sided size selection was performed to remove very large and very small fragments. In the first step, beads were added with a ratio of 0,7 to remove fragments larger than 500 bp and the supernatant was retained. In the second step, the AMPure beads were added to the supernatant with a ratio of 1,8 to remove fragments smaller than 100 bp. The DNA fragments were eluted in 20 µl $H_2O$. The concentration was measured using a Qubit dsDNA HS Assay Kit (Thermofisher) according to the manufacturer's protocol. The fragment range of the ATAC libraries was determined with the help of the Bioanalyzer system, an automated DNA electrophoresis, and the High Sensitivity DNA Kit (Agilent).

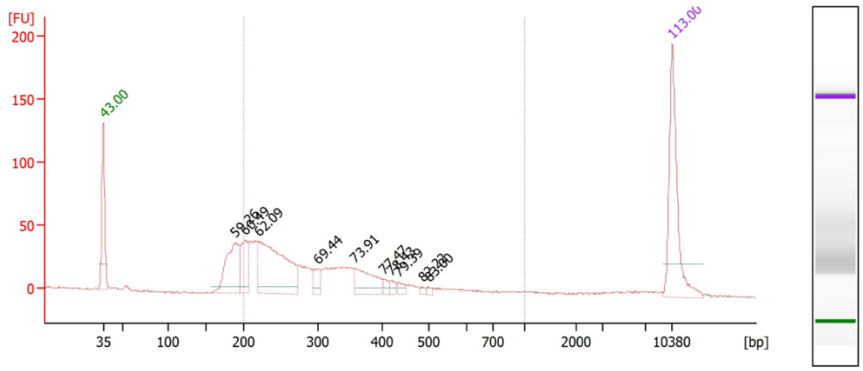



**Figure 2.2. Example electropherogram showing ATAC fragment distribution in AR cells**. High Sensitivity DNA Assay shows the ATAC library from 100.000 U2OS-AR cells after a two-sided size selection. The fragment sizes range from 189 bp to 449 bp.

## 2.13 CRISPR/Cas9 Genomic Editing

CRISPR/Cas9 genomic editing was carried out to induce genomic deletions in U2OS-AR cells. Therefore, single guide RNA's (sgRNA) were designed using the CRISPOR online design tool (Haeussler et al. 2016). The two sgRNA's were placed left and right from the genomic locus to be deleted. The sgRNA's were targeting a 20 nucleotide sequence followed by the PAM sequence "NGG". The first nucleotide of the sgRNA's was a "C" which is required for sgRNA expression from the U6 promoter. The sgRNA's were cloned into the PX459 plasmid. The pSpCas9(BB)-2A-Puro (PX459) V2.0 was a gift from Feng Zhang (Addgene plasmid #62988). The plasmid contains the Cas9 from S. pyogenes with a puromycin selection cassette and cloning backbone for a sgRNA.

### Plasmid Cloning

In a first step, the sgRNA's (oligos are listed in Tab. S6) were phosphorylated and annealed. The phosphorylation mix was set up incubated in the thermocycler at 37 °C for 30 min followed by 95 °C for 5 min and then ramped down to 25 °C at 5 °C per minute. Next, the oligos were diluted 250-fold with $ddH^2O$.

**Table 2.20 Phosphorylation Mix**

| | |
|---|---|
| Oligo 1 (100 µM) | 1 µl |
| Oligo 2 (100 µM) | 1 µl |
| 10x T4 Ligation Buffer (NEB) | 1 µl |
| $ddH^2O$ | 6,5 µl |
| T4 PNK (NEB) | 0,5 µl |

**Table 2.21 Digestion-Ligation Mix**

| | |
|---|---|
| PX459 backbone vector (50n g/µl) | 2 µl |
| Phosphorylated and annealed oligos (1:250 dilution) | 2 µl |
| 10x Tango buffer (or FastDigest Buffer) | 2 µl |
| DTT (10 mM) | 2 µl |
| ATP (10 mM) | 2 µl |
| FastDigest BbsI (Thermo Fisher Fermentas) | 1 µl |
| T7 DNA ligase | 0,5 µl |
| $ddH^2O$ | 8,5 µl |

**Table 2.22 Digestion-Ligation Thermocycling Program**

| | Temperature | Duration |
|---|---|---|
| Cycles 6x | 37 °C | 5 min |
| | 23 °C | 15 sec |

To insert the oligos into the PX459 vector, the digestion-ligation reaction was set up and incubated in a thermocycler. Finally, 5 µl of the ligated plasmid containing the sgRNA insert were used for transformation into Zymo-competent cells and plated on LB Ampicillin agar plates. After overnight incubation at 37 °C, bacterial colonies were picked and grown in liquid LB culture containing ampicillin at 37 °C overnight. After plasmid purification, sgRNA insertion was verified by Sanger sequencing.

## Transfection, Isolation and Genotyping of Single Cell-derived Colonies

Transfection of the U2OS-AR cells was performed with the Amaxa V Kit (Lonza) according to the manufacturer's manual. The cells were transfected with a mix of the two PX459 plasmids containing the sgRNAs. One million U2OS-AR cells were transfected with 1,2 µg of the PX459-AQP3_214 and 1,2 µg of the PX459-AQP3_216 plasmid and plated into two wells of 6-well plate. The next day the medium was replaced by fresh medium containing puromycin (10 µg/ml) to select the cells that contain the transfected plasmids. The day after, the medium was replaced by fresh medium without puromycin. Next day the 50-80 cells were transferred in a 150 mm petri dish to grow single cell-derived colonies. Once the colonies were visible, they were picked with the help of a cloning cylinder. First, cells were washer 2x with PBS, then a cylinder was pressed into sterile silicone grease and set on top of each cell colony. Then 20 µl of trypsin were added. When all colonies were covered the plate was incubated for 5 min at 37 °C in the incubator and the colonies were transferred in a 24-well microtiter plate with fresh medium until grown for confluency.

To genotype the clonal cell lines, genomic DNA was isolated with a Blood and Tissue Kit (Qiagen) according to the manufacturer's instructions. Genotyping was carried out with a diagnostic PCR followed by Sanger sequencing. The PCR primers were designed outside of the breakpoints to show the newly formed junction of the clones containing a genomic deletion (primers listed in Tab.S7). To identify WT cells or heterozygous clones containing WT alleles, primers were designed to amplify the breakpoints left and right of the targeted locus (primers listed in Tab.S7). The PCR products were analyzed by gel electrophoresis and Sanger sequencing.

## 2.14 STARR-Seq with FAIRE-Library

### FAIRE library and Input Plasmid Library

The FAIRE library and input plasmid library were generated by Stefanie Schöne and Alisa Fuchs. To perform a genome-wide screening for enhancers, a FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements) library was generated and integrated into the STARR-seq (Self-Transcribing Active Regulatory Region Sequencing) (Arnold et al. 2013) screening vector. FAIRE library was generated according to the protocol (Simon et al. 2013). In brief, 5 million U2OS-GR cells were treated with 1 μM dexamethasone for 1,5 h. Then the cells were cross-linked with 1 % formaldehyde for 3 min. The reaction was stopped with 125 mM glycine and incubated for 10 min at 4 °C. The cells were washed with PBS for 10 min at 4 °C and frozen at – 80°C. Then the cells lysed and the chromatin was fragmented by sonication with the Bioruptor Pico. Separation of open-chromatin DNA from the nucleosome-bound fraction was done by phenol-chloroform extraction. After DNA isolation, the cross-link was reversed and the DNA purified. The FAIRE library was generated by adding Illumina adapters (NEB #E7335) to the DNA fragments with the NEBNext Ultra DNA Library kit (NEB #E7370) according to manufacturer's manual, only the last PCR step was modified.

To integrate the FAIRE library into the STARR-seq vector, In-Fusion cloning sites were added by PCR to the adapter ligated DNA according to the original STARR-seq protocol(Arnold et al. 2013). Six PCR reactions containing 2 μl adapter ligated DNA as template and primers with a 15 nt extension for In-Fusion cloning were performed with the NEBNext Q5 Hot Start HiFi PCR Master Mix (PCR primers are listed in Tab S8).

**Table 2.23 PCR Thermocycling Program**

| | Temperature | Duration |
|---|---|---|
| Initial Denaturation | 98 °C | 45 s |
| Cycles 10x | | |
| Denaturation | 98 °C | 10 sec |
| Annealing | 65 °C | 75 sec |
| Elongation | 72 °C | 30 sec |
| Final Elongation | 72 °C | 10 min |
| Storage | 4 °C | ∞ |

The STARR-seq vector was a gift from Alexander Stark (Addgene plasmid #71509) The STARR-seq vector was linearized for subsequent cloning by AgeI-SalI digestion. After the PCR amplification of the FAIRE library and addition of In-Fusion cloning sites, the library was cloned into the linearized STARR-seq vector using the In-Fusion HD kit (Takara) according to the manufacturer's instructions. Next, a total of 20 transformations of the cloned FAIRE-STARR vector into MegaX DH10B T1R electrocompetent cells (ThermoFisher Scientific) were performed. Five transformed *E.coli* batches were pooled and grown in 0,5 l ampicillin-containing LB medium. The bacteria were harvested when the medium reached an Optical Density (OD) of 1.0. The input plasmid library was isolated with the Plasmid Plus Mega Kit (Qiagen).

## Input Library Transfection

U2OS-GR and U2OS-AR cells were transfected with the input plasmid library using the Amaxa Nucleofector kit V (Lonza) according to the manufacturer's instructions, except that for each electroporation 5 μg of DNA was used per 5 million cells. For each treatment (hormone or vehicle) 20 million cells were transfected and 3 biological replicates were generated.
It was shown that type I interferon (IFN-I) response to plasmid DNA transfection is able to compromise activation of reporter assays (Muerdter et al. 2018). To avoid the activation of a type I interferon (IFN-I), interferon inhibitors were added to the complete growth medium before resuspending the electroporated cells. PKR (Sigma) and TBK1/IKK (Sigma) were added directly after resuspension in complete growth medium according to the protocol (Muerdter et al. 2018), except that a final concentration of 0,5 μM per inhibitor was used. Due to low viability upon inhibitor treatment, 10 million electroporated cells were seeded on a 150 mm petri-dish. The cells were hormone treated for 14 h overnight. U2OS-GR cells were either treated with 1 μM Dex or 0,1 % EtOH as a vehicle control for 1,5 h. U2OS-AR cells were treated with 5 nM R1881 or 0,1 % DMSO as a vehicle. The cells were harvested by trypsinization, snap frozen and stored at –80°C.

## Preparation of the FAIRE-STARR-seq Library

Total RNA was isolated using the RNAeasy Midi Kit (Qiagen). One column was used per treatment. Elution was done 2x for each construct in 150 μl RNAse-free $H_2O$. The isolation of the poly(A) mRNA fraction was carried out using the Dynabeads mRNA Purification Kit (ThermoFisher Scientific) according the manufacturer's protocol. For each 500 ng RNA, 330 ng beads were used. The mRNA was eluted in 90 μl RNAse-free $H_2O$. mRNA was treated with

TURBO DNase (Ambion) for 30 min at 37°C and then purified with Agencourt RNAClean XP beads (Beckman Coulter) according to the manufacturer's instructions. The cDNA synthesis was carried out using SuperScript III (ThermoFisher Scientific) to reverse transcribe the mRNA and add unique molecular identifiers (UMIs) in the same reaction (UMI barcode primer for reverse transcription is listed in Tab S9). Per reaction, 500 ng mRNA were used.

The reactions were pooled and treated with RNAse (10 mg/ml) for 1 h at 37 °C. The cDNA containing UMIs was purified using Agencourt RNAClean XP beads (Beckman Coulter) according to the manufacturer's instructions and eluted in 50 µl RNAse-free $H_2O$. The cDNA concentration was measured with the Qubit ssDNA Assay Kit (Thermo Fisher Scientific). Next, a PCR reaction was set up with Kapa Hifi hotstart ready mix (Roche) to add Illumina HiSeq platform-compatible primers (PCR Primers with Illumina Adaptors are listed in Tab S10).

**Table 2.24 PCR Mix for aDNA amplification and Illumina Adaptors**

| | |
|---|---|
| cDNA (10 ng/µl) | 2,5 µl |
| Primer Forward + Reverse (10 µM) | 1,5 µl |
| Kapa Hifi hotstart ready mix | 25 µl |
| $H_2O$ | 11 µl |
| **Total Volume** | **50 µl** |

**Table 2.25 PCR Thermocycling Program**

| | Temperature | Duration |
|---|---|---|
| Initial Denaturation | 98 °C | 45 s |
| Cycles 17x | | |
| Denaturation | 98 °C | 15 sec |
| Annealing | 65 °C | 30 sec |
| Elongation | 72 °C | 30 sec |
| Storage | 4 °C | ∞ |

The PCR samples were pooled and purified using AmPure XP beads (Beckman Coulter) with a ratio of 1 (beads to PCR reaction) and eluted with 25 µl $H_2O$. Finally, the FAIRE-STARR cDNA libraries were submitted for paired-end Illumina sequencing with a read length of 50 nt and a sequencing depth of 100 million reads.

## 2.15 STARR-Seq Analysis for Individual Enhancers

### Preparation of the STARR-seq Vector containing Enhancer Insert

Putative enhancer fragments were ordered for integration into the STARR-seq vector and designed as previously described (Schöne et al. 2018). The fragments contained In-Fusion cloning sites for direct integration into the STARR-Seq vector as well as adapter sequences for Illumina sequencing containing P5 and P7 sequences. The enhancer fragments were ordered as a gBlock (IDT) or GeneStrand (Eurofins). The sequences of the enhancer inserts are listed in Tab. S14.

5' **TAGAGCATGCACCGG**ACACTCTTTCCCTACACGACGCTCT--INSERT--
AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC**TCGACGAATTCGGCC** 3'

**Figure 2.3. Enhancer Design for cloning into the STARR-seq vector**. Sequences contained P5 (colored in red) and P7 (colored in blue) Illumina adaptors and In-Fusion cloning sites for integration into the STARR-Seq vector (in bold). Sequence design was made by Stefanie Schöne.

The STARR-seq vector was linearized for subsequent cloning by AgeI-SalI digestion. Integration of the enhancer sequence into the STARR-Seq vector was carried out with the In-Fusion HD Cloning Kit (Takara Clontech) according to the manufacturer's manual. A 10 µl reaction was set up with 100 ng linearized STARR-seq vector, 20 ng insert DNA and 2 µl In-Fusion HD Enzyme Premix and incubated at 50 °C for 15 min. 3.8 µl of the reaction were used for transformation of 50ul DH5α chemically competent cells. The *E. coli* were plated on ampicillin-containing LB-Agar plates and incubated overnight at 37 °C. 3-5 colonies of each construct were picked for overnight liquid cultures. The plasmids were isolated using the QIAprep Spin Miniprep Kit (Qiagen) and sent for Sanger Sequencing (Eurofins) to check for the correct insertion (sequencing primer listed in Tab. S11).

### Transfection and Harvest

U2OS-GR and U2OS-AR cells were transfected with STARR-seq constructs by Nucleofection using the Amaxa V Kit (Lonza) according to the manufacturer's manual. 1 million cells were transfected with 2 µg DNA. PKR (Sigma) and TBK1/IKK (Sigma) were added to the complete growth medium before resuspension of the cells at a final concentration of 0,5 µM per inhibitor. 1 million transfected cells were split in two wells of a six-well plate. U2OS-GR cells were either treated with 1 µM Dex or 0,1 % EtOH whereas U2OS-AR cells were treated with 5 nM R1881 or

0,1 % DMSO for 14 h overnight. The next day, the RNA was purified using the RNeasy Kit (Qiagen).

### Reverse Transcription and qPCR Analysis of STARR-seq Activation

The RNA was reverse transcribed with the PrimeScript Kit (Takara) using specific primers for the housekeeping gene RPL19 and the GFP sequence on the STARR-seq vector (primers listed in Tab. S.12). STARR-seq activation was analyzed by qPCR using specific primers for the housekeeping gene RPL19 and the GFP sequence of the STARR-seq vector (primer sequences Tab. S13).

## 2.16 Rapid immunoprecipitation mass spectrometry of endogenous proteins

The rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) experiments were performed in a collaboration with Stefan Prekovic and Isabel Mayayo Peralta from the The Netherlands Cancer Institute (NKI). After a hormone deprivation step, U2OS-GR cells were incubated with 1 μM Dex and AR-U2OS cells with 5 nM R1881 for 4h. RIME experiments were done as previously described (Mohammed et al. 2016). The antibodies anti-GR (12041, Cell Signalling Technology), anti-AR (06-680, Merck), and anti-rabbit IgG (sc-2027, Santa Cruz Biotechnology) were used. The tryptic digestion of the proteins bound to the beads was done as previously described (Stelloo et al. 2018). LC-MS/MS analysis of digests was carried out on an Orbitrap Fusion Tribrid mass spectrometer and a Proxeon nLC1000 system (Thermo Scientific) using the previously described settings, except that the samples were eluted from the analytical column in a 90-min linear gradient.

Data analysis of the raw data was performed by Proteome Discoverer (PD) (version 2.3.0.523, Thermo Scientific) with standard settings. The MS/MS data was scanned in the Swissprot database (release 2018-06) using Mascot (version 2.6.1, Matrix Science, UK) with Homo sapiens as taxonomy filter (20,381 entries). For the fragment ion masses the maximum allowed precursor mass tolerance was 50 ppm and 0,6 Da. Trypsin was chosen as cleavage specificity tolerating two missed cleavages. Furthermore, oxidation (M) and deamidation (NQ) were used as variable modifications whereas Carbamidomethylation (C) was set as a fixed modification. For the identification of peptide and protein, false discovery rates were set to 1 % and Mascot peptide ion score > 20 or Sequest HT XCorr > 1 was set as additional filter. Proteome Discoverer output file that contained the abundances was loaded into Perseus (version 1.6.1.3). Next, LFQ intensities

were Log2-transformed and the proteins were filtered for at least 66 % valid values. The values missing were replaced by assumption based on the standard settings of Perseus, in example a normal distribution with a width of 0,3 and a downshift of 1,8. A t-test was applied for determination of the differentially enriched proteins.

## 2.17 Computational Analysis

### 2.17.1 RNA-Seq

RNAseq analysis Sequencing libraries were made using the TruSeq RNA library Prep Kit (Illumina, San Diego, CA). Poly (A) RNA was extracted by using oligod(T) beads (Illumina) followed by reverse transcription. RNA-seq experiment was performed in three biological replicates of each condition (hormone or vehicle treatment).

RNA-seq data analysis was done in collaboration with Stefan Haas. Paired end 50 bp reads from Illumina sequencing (roughly 40 million per sample) were mapped against the hg19 reference genome using the Spliced Transcripts Alignment to a Reference (STAR) software (Dobin et al. 2013) (options: alignIntronMin 20, alignIntronMax 500000, chimSegmentMin 10, outFilterMismatchNoverLmax 0.05, outFilterMatchNmin 10, outFilterScoreMinOverLread 0, outFilterMatchNminOverLread 0, outFilterMismatchNmax 10, outFilterMultimapNmax 5). The differential gene expression of hormone and vehicle-treated samples (n=3) was identified with the tool DESeq2 (PMID 25516281). Default parameters were used except betaPrior=FALSE.

### 2.17.2 RNA-Seq Gene-Expression Categories of AR and GR (Venn Diagram)

This data analysis was performed by Gözde Kibar to identify AR-specific, GR-specific and shared gene clusters. Differentially expressed genes between hormone and vehicle treatment condition were obtained for GR (4h hormone against vehicle) and AR cells (24h hormone against vehicle) with the *LfcShrink* function of the DESeq2 tool (version 1.24.0) (Love, Huber, and Anders 2014). Genes with a p-value < 0,05 and log2(fold change) > 1.5 were considered as significant. In total 777 GR-specific and 364 AR-specific upregulated genes were found. 187 upregulated genes were shared between AR and GR.
Moreover, 227 GR-specific, 14 AR-specific and 3 shared downregulated genes were identified.

### 2.17.3 Heatmaps for Top 50 AR, GR, Shared and Non-Regulated Genes

First, the unnormalized gene counts of the genes from the differential gene categories were obtained (for AR-specific, GR-specific genes, shared genes adjusted p-value < 0.05 and log2(fold change) > 1.5 and for non-regulated genes adjusted p-value < 0.5 and 0.5> log2(fold change) > 0). Second, the unnormalized gene counts were transformed into log normalized gene counts for the heatmap visualizations using *DESeq2* function *normTransform*.
To identify AR-specific, GR-specific and shared upregulated genes after hormone induction in both cell lines, first, the genes were sorted by their log fold change. And second, the top 50 AR-specific, GR-specific and shared target genes between AR and GR with the highest fold change were plotted in heatmaps using the pheatmap and ggplot2 packages. Third, RNA-seq heatmaps of 50 randomly selected genes were plotted for the non-regulated gene category.

### 2.17.4 ChIP-Seq

The ChIP-seq data analysis was performed in collaboration with Melissa Bothe. The ChIP-seq data of the U2OS-GR replicate 1 was from a previous study (SRP020242) (Schiller et al. 2014). ChIP-seq data of the U2OS-GR replicate 2 was produced by Verena Thormann.

The reads were mapped using the tool Bowtie v2.1.0 (Langmead and Salzberg 2012) to the hg19 reference genome (options: --very-sensitive). Reads of GR replicate 1 were mapped using Bowtie2 to hg19 (options: --very-sensitive -X 600 --trim5 5). With the tool SAMtools v1.10 (H. Li et al. 2009) the reads with a mapping quality < 10 were removed by filtering. Duplicate reads were removed using Picard tools v.2.17.0 (http://broadinstitute.github.io/picard/; MarkDuplicates).
The BigWig tracks used for data visualization in the IGV genome browser were made with deepTools v3.4.1 (Ramírez et al. 2014) and bamCoverage (options: --normalizeUsing RPKM --binSize 20 --smoothLength 60).
Peak calling of the AR and GR ChIP-seq peaks over the input was done with the tool MACS2 v2.1.2 (PMID: 18798982) and a q-value cut-off of 0.05. The overlapping peaks were extracted with BEDtools intersect v2.27.1 (-u) (Quinlan and Hall 2010). Finally, the ENCODE blacklisted regions for hg19 (Dunham et al. 2012)as well as regions within unassigned contigs (chrUn) and mitochondrial genes (chrM) were removed.

Receptor-specific and shared binding peaks were identified with BEDtools intersect. The GR-specific peaks at loci with low accessibility were received by sorting ATAC-seq signal of GR cells (vehicle treated cells) from high to low at all GR peaks (+/- 250 of peak center) with the tool computeMatrix (deepTools) followed by extraction of the lowest 17.125. The peaks were called on the ATAC-seq data in U2OS-GR (vehicle treated cells) cells as well as ATAC-seq data in U2OS-AR (vehicle treated cells) cells with the tool MACS2 v2.1.2 (Y. Zhang et al. 2008) (options: --broad –broad-cutoff 0.05). The peaks obtained were removed from the GR-specific peaks with a low ATAC signal to obtain a list with only inaccessible regions.

### 2.17.5 Omni-ATAC-Seq

The data analysis was carried out by Melissa Bothe. ATAC-seq data was processed as previously described (Thormann et al. 2019). Additionally, reads with a mapping quality <10 were removed with SAMtools v1.10 (H. Li et al. 2009).
The BigWig tracks used for data visualization in the IGV genome browser were made with the tool bamCoverage (--normalizeUsing RPKM; --binSize 20; --smoothLength 60) from deepTools v3.4.1 (Ramírez et al. 2014).

### 2.17.6 STARR-Seq

The data analysis was carried out by Melissa Bothe. The reads were mapped with Bowtie2 v2.1.0 (Langmead and Salzberg 2012) (options: --very-sensitive) to the human hg19 reference genome. According to their UMIs and genomic coordinates, the reads were deduplicated with the help of the UMI-tools v1.0.0 dedup function (Smith, Heger, and Sudbery 2017). To filter reads with a low mapping quality < 10 SAMtools v1.10 (H. Li et al. 2009) was applied. For further downstream analyses, the BAM files of the STARR-seq replicates were merged with SAMtools merge.
The BigWig tracks of the merged BAM files used for data visualization in the IGV genome browser were made using bamCoverage (--normalizeUsing RPKM; --binSize 20; --smoothLength 60) from deepTools v3.4.1 (Ramírez et al. 2014)

### 2.17.7 Heatmaps and Profile Plots at AR and GR Peaks

The heatmaps were generated in collaboration with Melissa Bothe. The heatmaps as well as their corresponding profile plots (+/- 2 kb around the peak center) of the peak categories AR-specific,

GR-specific and shared were generated with the tool computeMatrix (options: reference-point) and plotHeatmap from deepTools v3.4.1 (Ramírez et al. 2014), with the BigWig files used as input. The BigWig files have previously been generated with deepTools bamCoverage (--normalizeUsing RPKM).

### 2.17.8 Intersecting Genomic Binding and Gene Regulation

The data analysis was carried by Gözde Kibar. In a first step, the gene categories were defined. Receptor-specific genes: adjusted p-value < 0.05 and log2(fold change) > 1.5 for either AR or GR. Shared genes: adjusted p-value < 0.05 and log2(fold change) > 1.5 for both AR and GR cells. Non-regulated: adjusted p-value < 0.5 and 0.5 > log2(fold change) > 0. Second, ChIP-seq peaks were assigned to the genes: A 60-kb genomic window centered on TSSs was scanned for each gene for overlap with each peak file (AR-specific peaks, GR-specific peaks, shared peaks between AR and GR cells) using bedtools window. With the SeqMiner package (Zhan and Liu 2015) the TSS annotations of genes (hg19) were extracted from the NCBI RefSeq database. Each peak was only assigned to one gene categorie. In case a peak was assigned to more than one gene category, the peak was then assigned to the nearest gene with the tool ChIPpeakAnno in R (Zhu et al. 2010). Genes without a peak in 60 bp window were indicated as "no peaks". The binding events assigned to each gene category are represented in stacked bar graph.
The same analysis was used to link GR binding peaks with a low accessibility ('pioneering peaks') and without a low accessibility to the different gene categories (AR-specific, GR-specific and shared).
A statistical test was carried out with the Fisher Exact test on 2x2 contingency tables to compare the number of genes in each gene category that have specific binding events to the number of non-regulated genes with the corresponding binding.

### 2.17.9 Shared FAIRE-STARR and H3K27ac Peaks near Different Gene Categories

The data analysis was performed in collaboration with Melissa Bothe. Shared FAIRE-STARR-seq peaks (window +/- 250 bp) or shared H3K27ac ChIP-seq signal for shared ChIP-seq peaks from AR and GR cells were plotted for four different gene categories (AR-specific, GR-specific, shared and non-regulated genes) with the help of deepTools v3.4.1 (Ramírez et al. 2014) computeMatrix

(options: reference-point) and plotProfile. BigWigs previously generated files were used as input with bamCoverage (--normalizeUsing RPKM).

### 2.17.10 Motif Analysis and GC Content of AR and GR-specific Sites

The motif analysis was performed in collaboration with Melissa Bothe. The tool AME (McLeay and Bailey 2010) of the MEME suite v5.1.1 (Bailey et al. 2009) was used to do a motif enrichment analysis at AR and GR binding peaks. AR-specifc peaks as well as GR-specific peaks +/- 250pb around the peak summit were used as input. For GR, 6593 randomly sampled peaks or 6593 peaks with the highest chromatin accessibility (sorted ATAC-seq signal in vehicle condition) were used as input. The ChIP-seq peak sequences were then scanned for JASPAR 2018 Vertebrate Clustering motifs (Khan et al. 2018) including the DR3 (direct repeat) motif. Shuffled input sequences were set as control sequences. Motif hits with an E value $<10^{-30}$ for either AR or GR were included for the heatmap visualization. The 7 motif logos are shown additionally.

The profile plot showing the GC content for U2OS cells were generated at AR-specific and GR-specific peaks and at AR- and GR-specific peaks with a high accessibility. The high accessibility regions were received by sorting the ATAC-seq signal (vehicle condition) at all AR- or GR-specific peaks +/- 250 bp around the peak summit from high to low and obtaining the top 3296 peaks. In a next step, the human reference genome hg19 was binned into 50 bp bins using the makewindows function of BEDtools v2.27.1 (Quinlan and Hall 2010). The GC content was obtained for each bin with BEDtools nuc and followed by a conversion of the bedgraph file into bigWig format using bedGraphToBigWig (Kent et al. 2010). The deepTools v3.4.1 (Ramírez et al. 2014) commands computeMatrix (reference-point) and plotProfile were used to generate GC content profiles +/- 5 kb around peak summit.

### 2.17.11 AR and GR Binding Sites in the AQP3 Enhancer

The *AQP3* enhancer sequence [hg19: chr9:33437258-33437811] was scanned for AR (JASPAR ID MA0007.1-3) and GR (JASPAR ID MA0113.1-3) motifs using the JASPAR CORE database (Sandelin et al. 2004) and the relative profile score threshold 80%. Six putative AR and four GR sites were identified. The top three JASPAR hits are shown.

# 3. Results

## 3.1. Generation and Characterization of a Cell Line with Stable AR Expression

Already in the eighties it has been shown that GR and AR can recognize and bind the same consensus motif in the DNA (Beato et al. 1989). But despite of their similar binding preference, both receptors have different biological roles in the body. GR is important for gluconeogenesis as well as the immune system and has anti-inflammatory effects. AR on the other hand is, for example, crucial for the development and maintenance of the reproductive tissue. Both receptors even have contradictory functions. High cortisone levels lead to an accumulation of fat deposits and depression whereas high testosterone dosage is correlated with muscle growth and aggressive behavior. Although AR and GR identify similar DNA sequences, they activate different regulatory programs. To address this paradox, I set out to generate model cell lines that either express the human AR or GR in an identical cellular environment. With this model system, I was able to individually study AR and GR actions to understand receptor-specific properties of the androgen and glucocorticoid receptor.

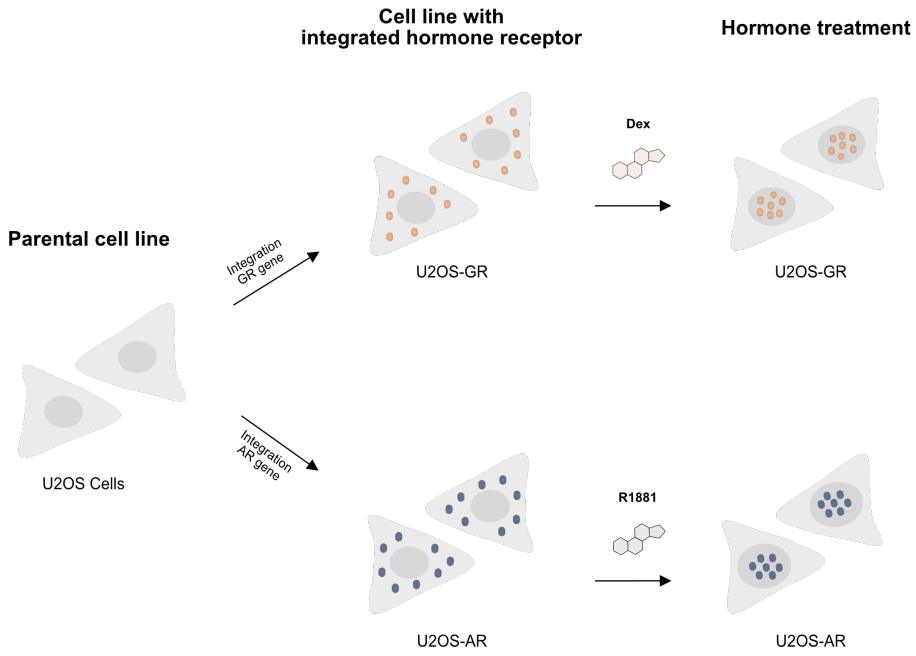



**Figure 3.1. Experimental Design.** Parental cell line U2OS was transfected with an GR plasmid for stable expression of GR in a previous study (Rogatsky, Trowbridge, and Garabedian 1997). U2OS-AR cells were generated using the same procedure. GR cells were activated with the synthetic steroid hormone Dexamethasone (Dex) whereas AR cells were activated using the synthetic steroid R1881.

U2OS-GR is an osteosarcoma cell line with a stable expression of the GRα isoform from rat. Many studies on GR were carried out successfully using the U2OS-GR cells (Rogatsky, Trowbridge, and Garabedian 1997) and the cells were already available in our lab. U2OS-GR cells are a valuable model system to study GR and transcriptional regulation for multiple reasons. First, U2OS-GR cells show strong hormone inducible gene expression. By simply adding a steroid, GR-dependent gene regulation can be studied in presence and absence of the hormone. Second, U2OS-GR cells have shown robust inducible reporter expression in various reporter assays in the past (Schöne et al. 2018; Telorac et al. 2016; Thormann et al. 2018). Therefore, we decided to generate an analogous model to study AR regulation which is directly comparable with GR.

The U2OS-AR cell line was generated by transfection of the parental U2OS cell line with a plasmid containing the human androgen receptor followed by random integration of the receptor into the genome (Fig. 3.1). Upon induction with its specific steroid, the hormone receptor can be activated. The receptor translocates into the nucleus and directly binds to DNA resulting in gene expression (Guiochon-Mantel et al. 1996; Mangelsdorf et al. 1995). AR was activated with R1881, a synthetic steroid hormone, whereas GR was induced with the synthetic steroid hormone Dexamethasone (Dex).

### 3.1.1. mRNA and Protein Expression of AR

After a transfection of U2OS cells with the AR plasmid, the cells were selected for a stable integration of AR using antibiotics. Then, single cell-derived clonal cells were grown and tested for AR expression, where one cell line stably expressing AR was found. Successful generation of an AR-U2OS cell line was tested by qPCR and Western blot. As expected, *AR* mRNA was detected in the U2OS-AR cells only whereas the parental cell line U2OS as well as U2OS-GR cells showed *AR* expression levels that were barely detectable (Fig 3.2a). In addition, Western blot analysis confirmed AR protein expression in the U2OS-AR cells indicative of successful integration of the receptor whereas the parental cell line U2OS showed no detectable AR protein expression (Fig.3.2b)

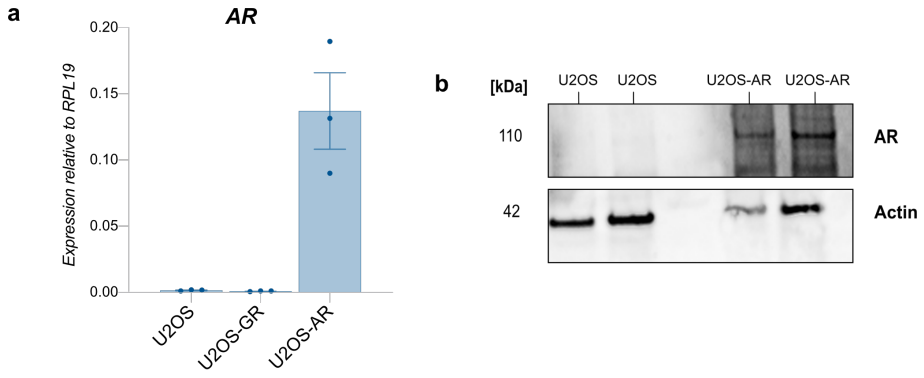
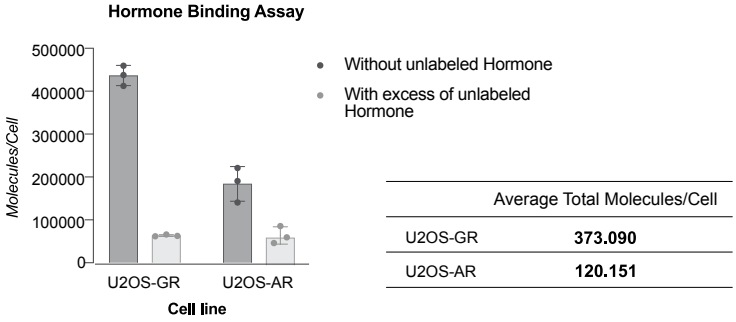



**Figure 3.2. Characterization of the AR cell line. (a)** Relative expression of *AR* mRNA in U2OS, the U2OS-GR cells and U2OS-AR cells. The average gene expression of three independent replicates ±SEM is shown. **(b)** Protein expression of AR and actin as internal control in U2OS cells and U2OS-AR cells was assayed by Western blotting. Only U2OS-AR cells showed a band indicative of AR protein expression.

### 3.1.2. Receptor Quantification by Liquid Scintillation Counting

A hormone binding assay was performed to further characterize the AR and GR model system and quantify the number of receptors expressed per cell. A comparable receptor expression level is crucial to identify differences in regulation of the two receptors that are a consequence of the individual receptor properties and not a consequence of their expression level.

To quantify receptor levels, the GR and AR cells were incubated with the same concentration of radioactively labeled hormone. To take unspecific binding of the hormone into account, a second experiment was performed where the cells were treated with radioactively labeled hormone in the presence of an excess of unlabeled hormone. Besides of the specific binding to steroid receptors, steroid hormones can unspecifically bind to lipoproteins in the cell (Leszczynski and Schafer 1990). By adding an excess of unlabeled hormone, it competes out the labeled hormone at its specific binding domain of the hormone receptor. However, unspecific hormone binding cannot be competed out and is therefore still detected in the presence of an excess of unlabeled hormone. The total number of molecules per cell was calculated by subtracting the number of molecules per cell with unlabeled hormone from the number of molecules per cell without addition of unlabeled hormone (Fig 3.3). Using this approach, I found that U2OS-GR cells expressed, on average, a total of 373.090 GR molecules per cell whereas U2OS-AR cells expressed, on average, a total of 120.151 AR molecules per cell. The expression levels of the receptors differ between AR and GR cells, however, the expression is still in a similar range. Nevertheless, differences in receptor levels might have an influence on differential actions of AR and GR in the cells.




| | Average Total Molecules/Cell |
|---|---|
| U2OS-GR | **373.090** |
| U2OS-AR | **120.151** |

**Specific Binding**

*Total [Molecules/Cell] = [Molecules/Cell (without unlabeled Hormone)] -*

*[Molecules/Cell (with excess of unlabeled Hormone)]*

**Figure 3.3. Receptor quantification by liquid scintillation counting**. Receptor concentration of GR and AR was assayed by liquid scintillation counting in U2OS-GR and U2OS-AR cells, respectively. U2OS-GR cells were treated with 100 nM [$^3$H]-Dex in the presence or absence of a 10 μM excess of unlabeled Dex. U2OS-AR cells were treated 100 nM [$^3$H]-R1881 in the presence or absence of a 10 μM excess of unlabeled R1881. The average of three independent replicates ±SEM is shown. The number of total molecules per cell was calculated by subtracting the average number of molecules per cell with excess of unlabeled hormone from the average number of molecules per cell without excess of unlabeled hormone.

## 3.2. Comparing Gene Expression and Genomic Binding of AR and GR Cells

### 3.2.1. RNA-seq Analysis

To identify the genes specifically regulated either by GR or AR in U2OS cells, I carried out a RNA-seq analysis in both cell lines. The experiment was performed in absence and presence of hormone. In this way, I was able to identify the basal expression level of the transcriptome in both cell lines as well as the target genes specifically regulated by either GR or AR. Before the RNA-seq analysis, I have initially performed an qPCR analysis for the gene *FKBP5* in the AR and GR cells with different hormone treatment times. *FKBP5* is known target gene of GR (Vermeer et al. 2003) and AR (Amler et al. 2000) and was therefore suitable to test the transcriptional response followed by receptor activation and to establish the hormone treatment conditions in both cell lines. In the AR cells, robust hormone-dependent gene expression was observed after 24 h, whereas GR cells showed a regulation after 4 h and an even stronger effect after 24 h (Fig.3.4). Consistent with this observation, it was already previously shown, that both receptors have different gene activation kinetics and that activation of a reporter by AR was slower compared to GR (Nenseth et al. 2014).

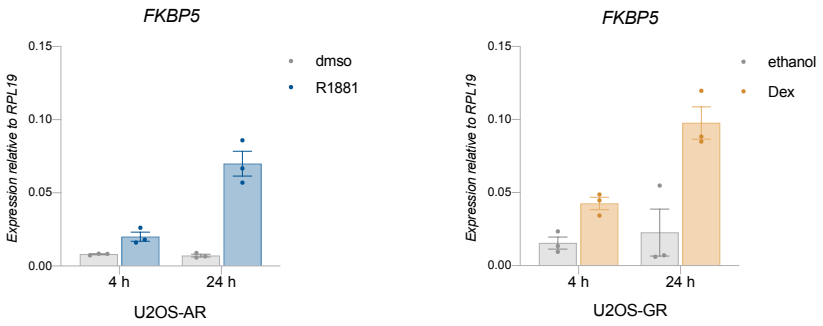



**Figure 3.4. *FKBP5* is regulated by AR and GR**. Relative mRNA expression of *FKBP5* is shown in AR and GR cells. AR cells were treated with dmso as vehicle control or 5 nM R1881 and GR cells treated with ethanol as vehicle control or 1 μM Dex. Both cell lines were induced for 4 h and 24 h. Average gene expression of three independent replicates ±SEM is shown.

#### 3.2.1.1. Venn Diagram Showing Receptor-Specific and Shared Gene Regulation

I have performed the RNA-seq analysis for GR cells using 4 h and 24 h hormone treatment. However, after 24 h a very large number of genes was regulated. To avoid unspecific secondary effects of a long GR activation, I decided to use the GR RNA-seq data with a 4 h hormone induction only. Because of the slower activation kinetics of AR, the AR cells were still treated for 24 h with hormone and vehicle to ensure a robust signal. The RNA-seq data analysis was done in collaboration with Stefan Haas and Gözde Kibar. Genes with differential expression upon hormone-treatment were identified from three biological replicates using the tool DeSeq2 (Version 1.24.0) (Love, Huber, and Anders 2014).

In total, 590 genes were found to be significantly upregulated upon GR induction only and 177 genes upon AR induction (Fig.3.5). Furthermore, 187 genes were significantly up-regulated by both receptors. Next, we looked at down-regulation and found that 227 genes were down-regulated by GR whereas only 14 genes were significantly down-regulated by AR. Only three genes were down-regulated by both receptors.

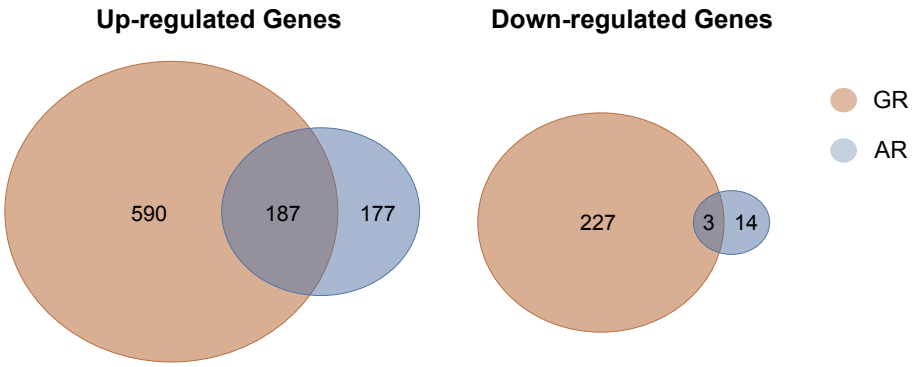



**Figure 3.5. Gene expression analysis of GR and AR cells.** Up and down-regulated differentially expressed genes are shown in GR cells (orange) and AR cells (blue). GR cells were treated with 1 μM Dex and ethanol as vehicle for 4 h, AR cells were treated with 5 nM R1881 and DMSO as vehicle for 24 h. Differential gene expression analysis was performed using the tool DeSeq2 on three biological replicates ($p < 0.05$, abs(log2fc) > 1.5). Data analysis was performed in collaboration with Stefan Haas and Gözde Kibar.

Together, both receptors have their specific target genes but also genes that are regulated by both receptors. In general, GR regulates more genes compared to AR in U2OS cells, which is true for down-regulation and up-regulation of gene expression. Only a very small number of genes were significantly down-regulated in the AR cells. In this study, we decided to exclusively focus on up-regulated genes since the mechanisms responsible for down-regulation are poorly understood and cannot be directly linked to binding of the transcription factors as down-regulation can be a consequence of secondary effects of binding (Langlais et al. 2012).

#### 3.2.1.2. Top 50 AR-Specific, GR-Specific and Shared Genes

The gene expression analysis described above showed that both AR and GR have shared as well as receptor-specific target genes. Importantly, since these categories of regulated genes were based on cut-offs for fold change and p-value, we set out to determine if the differential genes are indeed only regulated by one of the two receptors in a quantitative manner. Therefore, the top 50 AR-specific, GR-specific and shared genes were visualized in a heatmap. The heatmaps were generated by Gözde Kibar. In addition, a heatmap with genes is shown that are expressed in both cell lines but not regulated by either AR or GR.

The heatmap visualization demonstrates that the majority of the top 50 receptor-specific genes are regulated upon hormone induction specifically by only AR or GR (Fig 3.6 a and b). The top 50 shared genes are regulated by both AR and GR (Fig 3.6 c). However, the non-regulated genes show no significant change in expression upon hormone treatment (Fig 3.6 d).

Together, the analysis and visualization of the RNA-seq data revealed that both AR and GR specifically regulate an individual set as well as an overlapping set of genes.

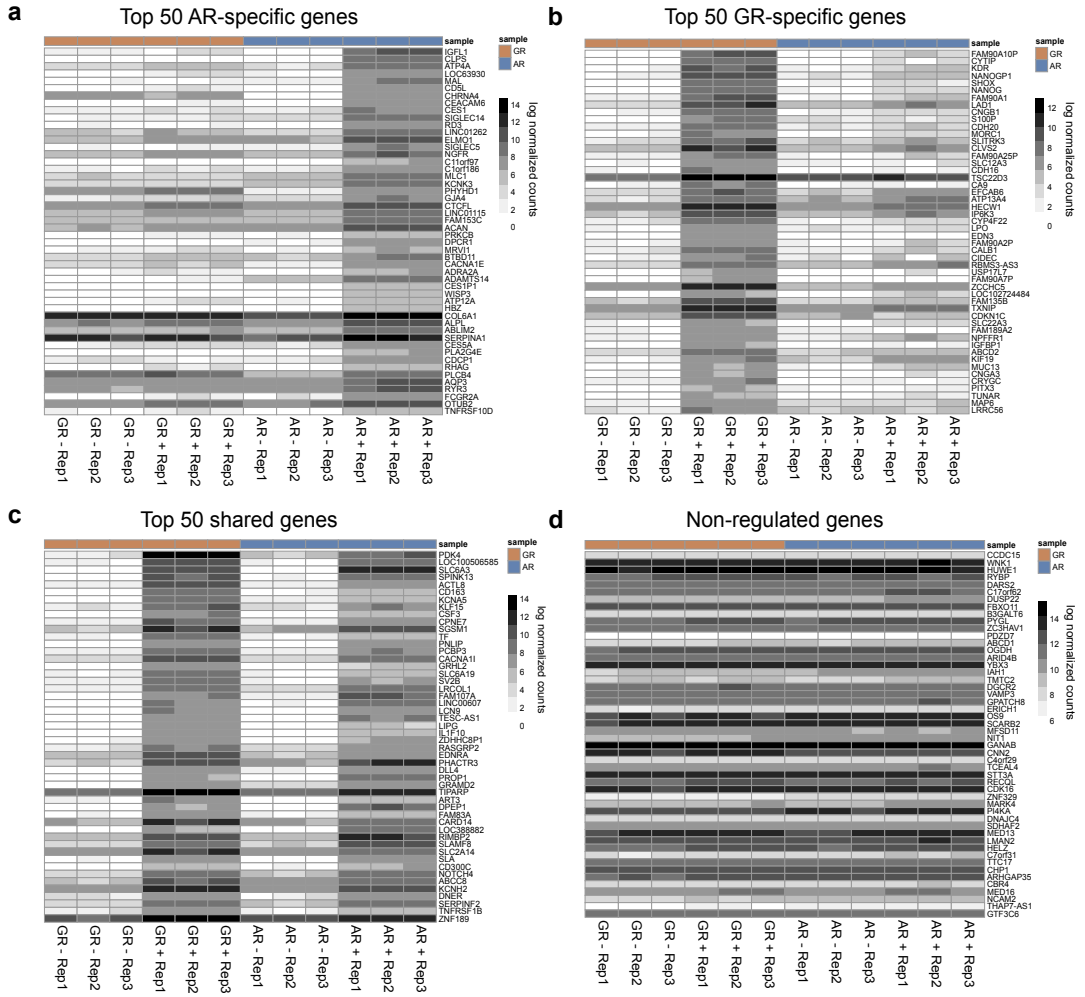



**Figure 3.6. Top 50 genes differentially regulated by GR, AR, shared and non-regulated genes**. Heatmaps representing normalized gene expression of the top 50 differentially regulated genes specific for AR **(a)**, GR **(b)**, both receptors **(c)** and genes not regulated by either AR or GR and randomly selected **(d)**. Columns represent the individual replicates and rows are representing the genes. Three replicates are shown for each receptor either treated with hormone or a vehicle control. Data analysis was done by Gözde Kibar.

### 3.2.1.3. qPCR Analysis of Receptor-Specific Gene Regulation

The RNA-seq analysis of AR and GR revealed that both receptors have an overlapping set of differentially expressed genes but also receptor-specific genes that are differentially expressed upon hormone treatment. To experimentally validate the RNA-seq results, I have performed a qPCR analysis for two genes of each gene category: AR-specific, GR-specific and shared, that were identified by RNA-seq. Therefore, AR and GR expressing cells were treated for 4h and 24h with

its specific hormone prior to RNA isolation, reverse transcription and qPCR. Moreover, by looking at two time-points I was able to get an idea if the genes are indeed regulated differentially by AR and GR or if the regulation of these genes is simply following different kinetics.

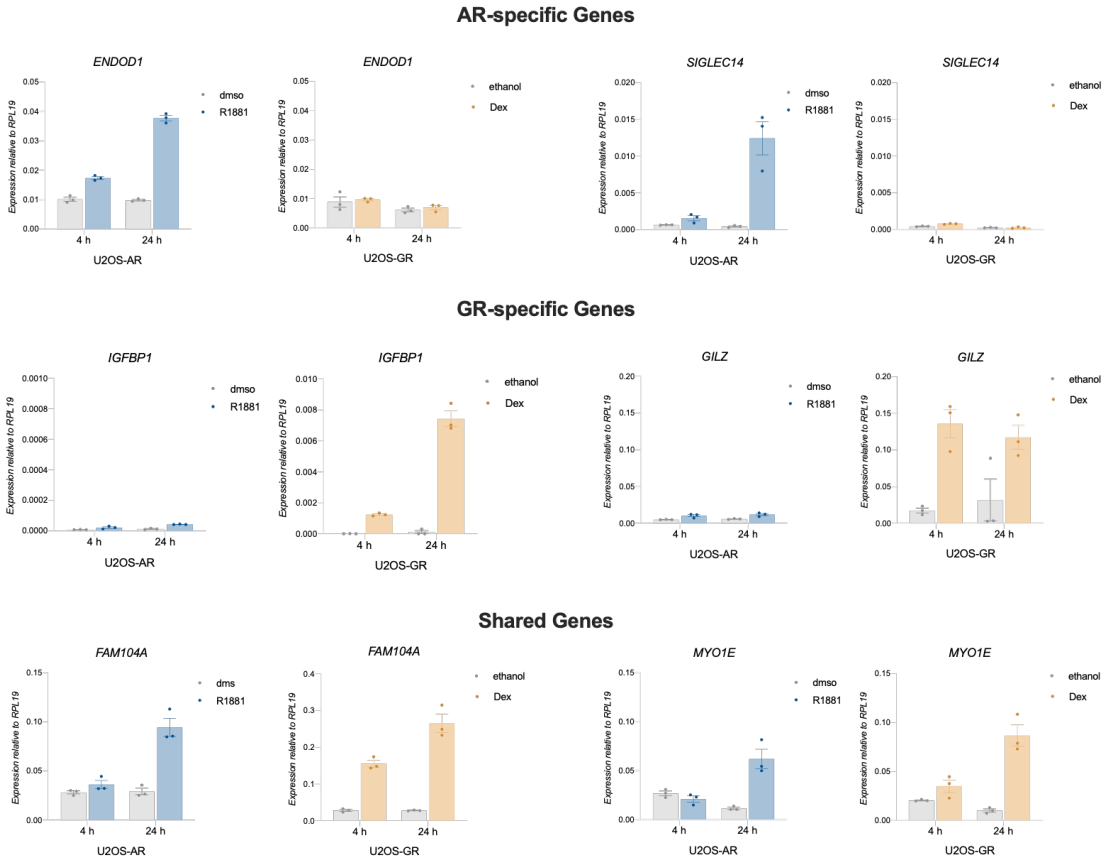



**Figure 3.7. Gene Expression of AR-specific, GR-specific and shared genes**.
Relative mRNA expression of two AR-specific, GR-specific and shared genes is shown in AR and GR cells. AR cells were treated with dmso as vehicle control or 5 nM R1881 and GR cells treated with ethanol as vehicle control or 1 µM Dex. Both cell lines were induced for 4 h and 24 h. Average gene expression of three independent replicates ±SEM is shown.

Consistently with the RNA-seq results, *ENDOD1* and *SIGLEC14* are differentially regulated by AR, whereas *IGFBP1* and *GILZ* show GR-specific regulation (Fig 3.7). The genes *FAM104A* and *MYO1E* are regulated by both receptors. Moreover, different dynamics of gene regulation could again be observed in AR and GR cells. Genes regulated by AR, show only a weak induction after 4h of hormone and a strong and robust signal after 24h. On the other hand, GR-regulated genes show a robust expression already after 4h of hormone induction.

Together, I was able to confirm receptor-mediated regulation of four differentially expressed genes identified by RNA-seq. Moreover, the analysis of these genes confirmed that AR and GR show

different dynamics of regulation. In GR cells, a robust gene expression could be observed after 4h of induction, whereas AR cells showed only weak gene expression after 4h of hormone treatment and a strong signal after 24h.

#### 3.2.1.4. Influence of the Dex Concentration on Gene Regulation by GR

The receptor quantification has shown that the U2OS-AR cells are expressing around 120.000 AR molecules whereas U2OS-GR cells express around 370.000 GR receptors. The expression level of a transcription factor can be an important driver of its cell-type specific function (L. N. Singh and Hannenhalli 2008). Differential gene expression driven by steroid receptors is often a consequence of their cell-type specific expression. AR is crucial for the function of the male reproductive organs, furthermore, it is highly expressed in androgen-sensitive prostate cancer cells LNCaP (Horoszewicz et al. 1983) where GR is not expressed (Schuurmans et al. 1988). For these cells, receptor levels indeed explain the fact, that AR is an important transcriptional regulator.

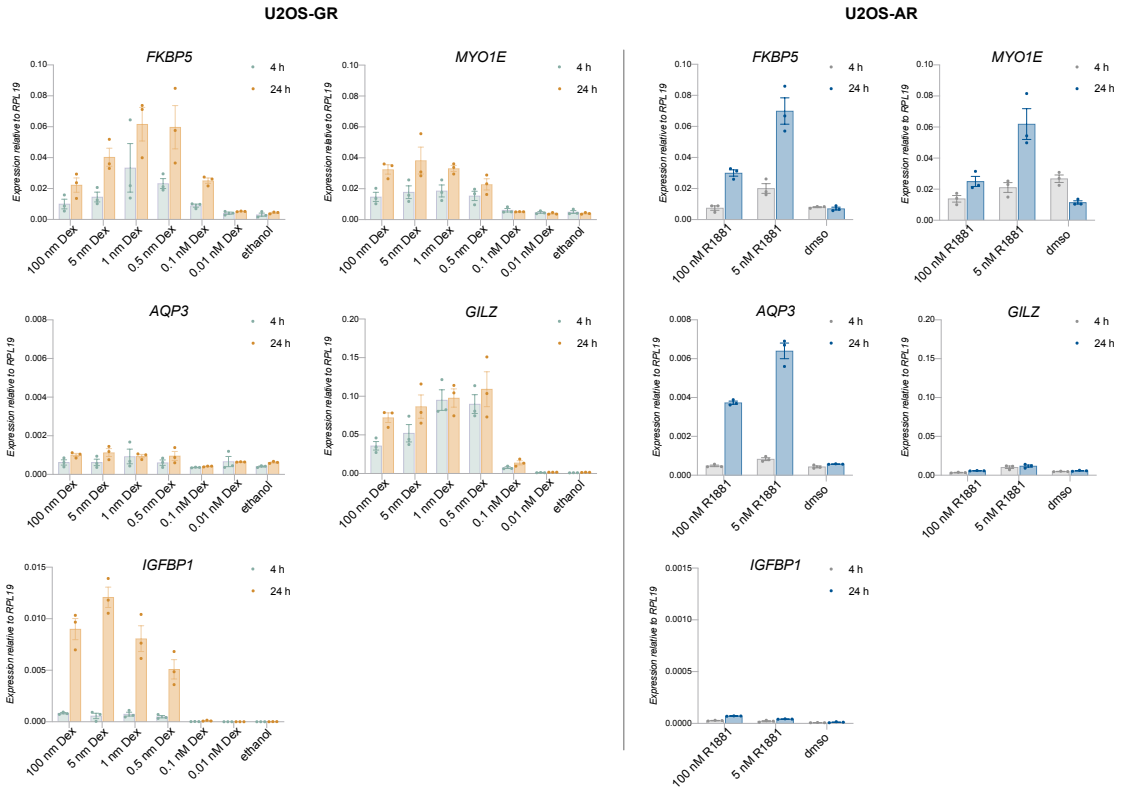



**Figure 3.8. qPCR analysis of GR and AR target genes using different hormone concentrations.** Relative mRNA expression of two shared (*FKBP5, MYO1E*), one AR-specific (*AQP3*) and two GR-specific genes (*GILZ, IGFBP1*) is shown in AR and GR cells. GR cells were treated with ethanol as vehicle control or with different Dex concentrations. AR cells were treated with dmso as vehicle control, 100 nM and 5 nM R1881. Both cell lines were induced for 4 h and 24 h. Average gene expression of three independent replicates ±SEM is shown.

To test if the receptor-specific effects observed for GR are simply a result of the higher receptor number in the cell, I have performed qPCR analyses for various GR target genes with reduced Dex concentrations. At lower hormone concentrations, only a subset of GR molecules in the cell are hormone-bound thus mimicking the lower level of AR that is present in the U2OS-AR cell line. I have used previously published $K_d$ values (Cleutjens et al. 1997) in order to perform an approximate calculation of the bound receptor fraction at a given hormone concentration. The dissociation constant $K_d$ represents the tendency of the receptor-ligand complex to dissociate into its components. 50 % of receptor molecules are bound at a hormone concentration that equals the $K_d$. Given previously reported $K_d$ values, 1 nM Dex would bind about 24 % of the GR receptors which corresponds to around 89.000 molecules, a level comparable to AR expression in the U2OS-AR cell line. The qPCR analysis showed that the GR target genes still show saturated expression levels when treated with 1 nM Dex and even 0,5 nM Dex. Upon induction with 0,1 nM Dex, which corresponds to 11.193 GR molecules, the expression levels of all target genes drastically dropped (Fig. 3.8).

These results indicate that GR-specific gene activation is not simply a result of the high expression levels but rather of the receptor-specific properties.

5 nM R1881 is a nearly saturating hormone concentration for AR and is estimated to activate around 98.000 molecules. To see if AR-specific gene expression is compromised by not using saturating R1881 concentration, I performed qPCR analyses for AR and GR-specific target genes using a saturating 100 nM R1881 concentration. I found that adding a higher hormone concentration did not further enhance gene activation by AR and also did not induce activation of genes, which were not regulated by AR before (*GILZ, IGFBP1*). This is again indicating that a lack of gene regulation is not simply a consequence of lower levels of hormone-occupied AR receptors.

### 3.2.2. ChIP-seq Analysis Reveals Distinct Groups of Binding Sites

Both AR and GR have a highly similar DNA binding domain and they bind to the same consensus motif *in vitro*. In order to understand if differential gene expression is a consequence of receptor specific binding, previous studies have compared the cistromes of AR and GR using ChIP-seq (Arora et al. 2013; Sahu et al. 2013). The analyses have been carried out in prostate cancer cells, where GR is not expressed under normal conditions. Although these datasets revealed that both receptor have individual binding sites as well as shared ones, in many cases differential gene expression could not be explained by binding.

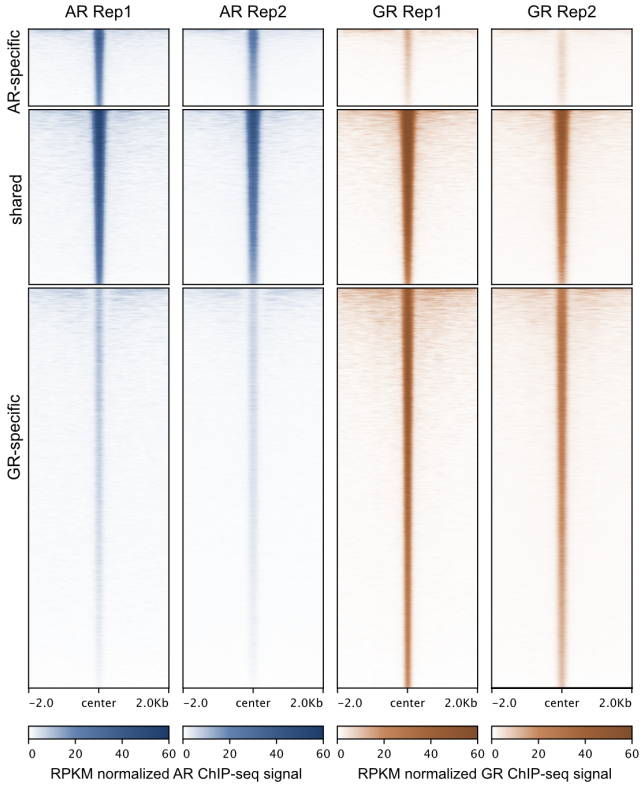



**Figure 3.9. ChIP-seq analysis shows three categories of receptor binding sites**. AR-specific sites comprise peaks called in AR cells only, GR-specific sites comprise peaks called in GR cells and shared sites comprise peaks called in both cell lines. The AR cells were treated with 5 nM R1881 for 4 h and GR with 1 µM Dex for 1,5 h. Two biological ChIP-seq replicates are shown. One GR ChIP-seq experiment was downloaded (SRP020242) (Schiller et al. 2014), the second ChIP-seq experiment was performed by Verena Thormann (Thormann et al. 2018). Data analysis was done by Melissa Bothe.

To understand if receptor-specific regulation in U2OS cells is directed by differential binding of AR and GR, I performed ChIP-seq to determine the genomic binding sites of both receptors *in vivo*. In brief, the transcription factors are linked covalently to their trancription factor binding sites using formaldehyde cross-linking. Next, the chromatin is sheared and the DNA bound transcription factor of interest is immunoprecipitated with its specific antibody. The DNA fragments are de-crosslinked and purified followed by deep sequencing and computational analysis.

One GR ChIP-seq replicate was downloaded (SRP020242) (Schiller et al. 2014), the second GR ChIP-seq experiment was carried out by Verena Thormann (Thormann et al. 2018). The data analysis of the ChIP experiments was carried out by Melissa Bothe.
The peaks identified by ChIP-seq were clustered in the categories AR-specific, shared and GR-specific peaks for the two cell lines (Fig 3.9). AR-specific peaks were only called for AR, shared peaks were called in both cell lines and GR-specific peaks were called for GR only. Similar to the gene expression, the ChIP-seq analysis revealed that both receptors bind individual sites but also share a set of binding sites in the genome.

### 3.2.3. Intersecting Genomic Binding and Gene Regulation

Our data showed that both receptors have specific binding peaks and also regulate their specific genes, therefore, we wanted to test whether receptor-specific binding indeed contributes to differential gene expression. The data analysis was performed by Gözde Kibar.

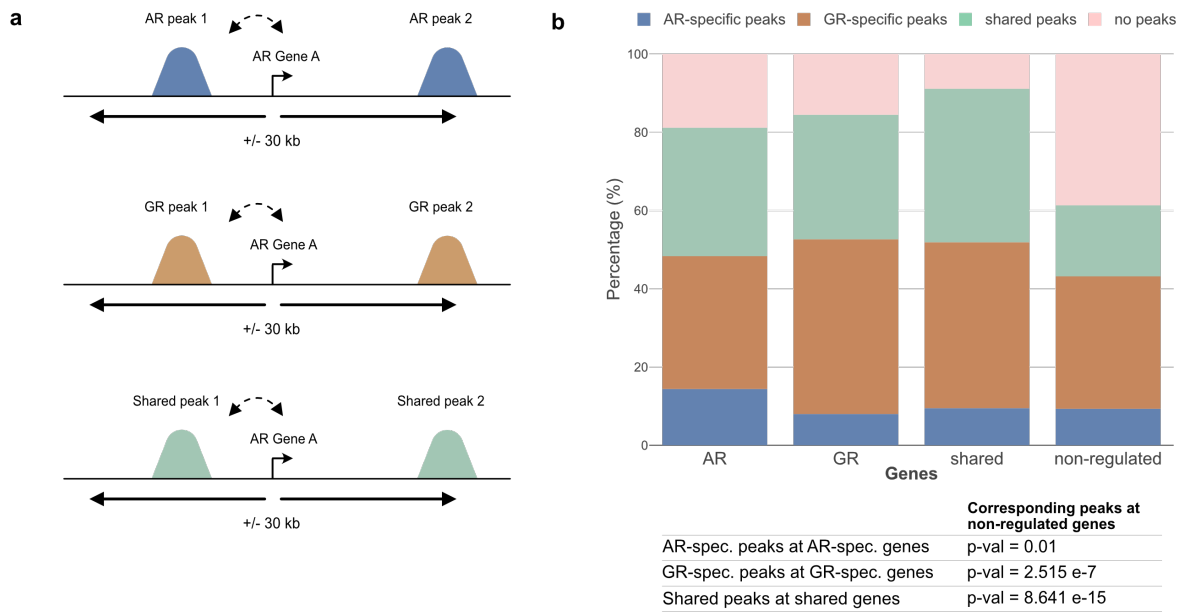



**Figure 3.10. Intersecting receptor-specific binding and differential gene regulation genome-wide**. **(a)** Cartoon showing how ChIP peaks of each category were assigned to AR regulated genes. A region of +/- 30 kb around each TSS was scanned for ChIP peaks of each category (AR, GR, shared, no peak). Analysis was done in the same way for GR-specific, shared and non-regulated genes. **(b)** The stacked bar graphs represent the percentage of each peak category (AR, GR, shared, no peak) assigned to each category of differentially regulated genes (AR, GR, shared, non-regulated). p-values were calculated using a Fisher's exact test. The data analysis was done by Gözde Kibar.

In this analysis, the ChIP peaks of each category were assigned to each category of regulated genes (receptor-specific, shared, non-regulated) (Fig 3.10 a). Therefore, a region around each TSS of each gene of each gene category (+/- 30 kb) was scanned for the presence of an AR-specific peak, GR-specific peak, shared peak or no peaks. As expected, the non-regulated genes have the largest fraction of genes with no receptor-specific gene in the vicinity (Fig 3.10 b). Consistent with receptor-specific gene regulation being driven by receptor-specific binding, we find that AR-specific genes are enriched for AR-specific peaks. Similarly, GR-specific genes are mostly enriched for GR-specific peaks and shared genes are significantly enriched for shared peaks.

This data suggests, that when intersecting genome-binding with gene expression, indeed, receptor-specific peaks are enriched in the vicinity of receptor-specific target genes. Therefore, receptor-specific regulation can in part be explained by binding. However, most of the receptor-specific genes show no receptor-specific binding. This finding indicates that mechanisms downstream of binding play an import role in conferring receptor-specific gene regulation.

### 3.2.4. Examples of AR-Specific, GR-Specific and Shared Genes

The genome-wide RNA-seq analysis revealed that both AR and GR regulate specific sets of genes but also show an overlap in gene regulation. Below some examples of genes where binding explains the pattern of gene regulation observed.

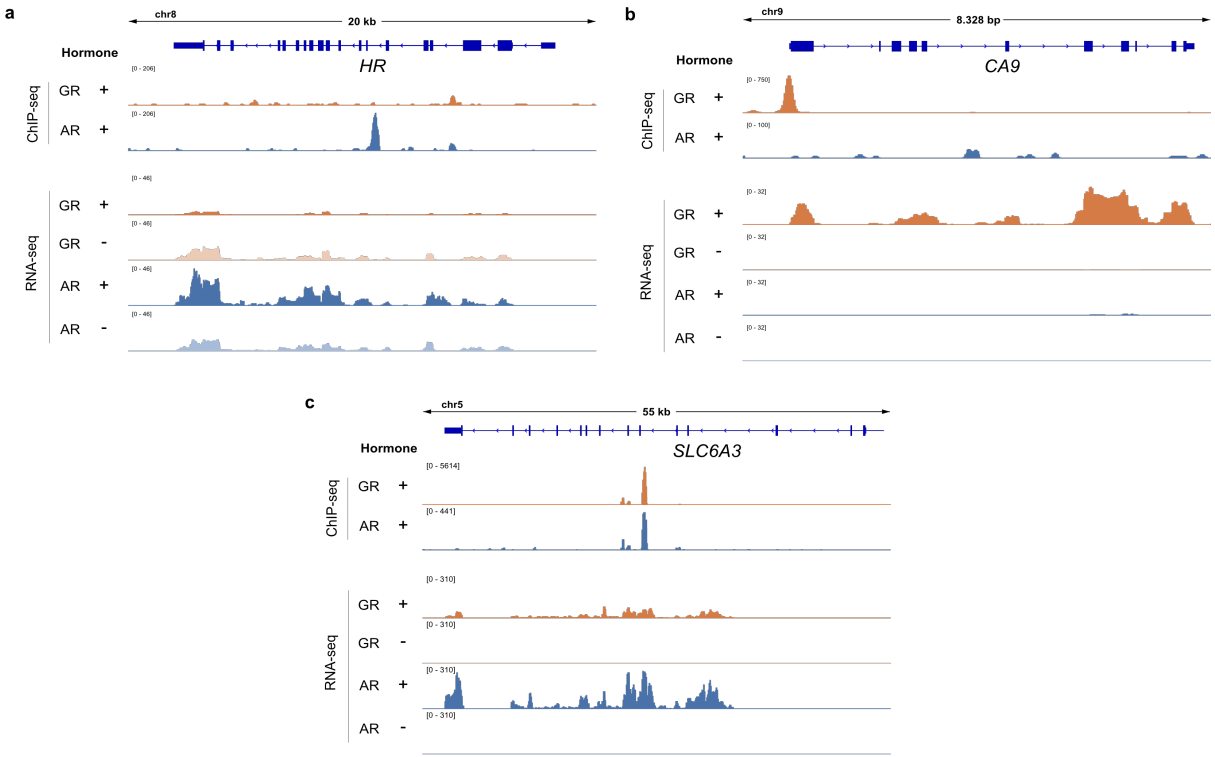



**Figure 3.11. Representation of transcript levels and receptor binding for shared, AR and GR-specific genes**. Genome-browser screenshot of the genomic locus of AR-specific *HR* **(a)**, GR-specific *CA9* **(b)** and shared *SLC6A3* **(c)** showing RefSeq gene, ChIP-seq and RNA-seq tracks of GR and AR. For the ChIP analysis GR cells were treated with ethanol as vehicle control or 1 μM Dex for 1,5 h and AR cells were treated with dmso as vehicle control or 5 nM R1881 for 4h. One representative ChIP-seq track is shown from two biological replicates. For the RNA-seq analysis GR cells were treated with ethanol as vehicle control or 1 μM Dex for 4h and AR cells were treated with dmso as vehicle control or 5 nM R1881 for 24h. RNA-seq tracks are the merged signal from three biological replicates. The BigWig tracks for data visualization were generated by Melissa Bothe.

The following genes are examples where gene expression can indeed be explained by receptor-specific binding. The gene *CA9* contains a GR binding peak in the promoter region (Fig 3.11 a). Moreover, *CA9* is specifically regulated by GR in a hormone-dependent manner. AR, however is not binding or regulating *CA9*. The gene *HR*, on the other hand, is specifically bound and regulated by AR only (Fig 3.11 b). Both examples support that regulation of certain genes can be explained by selective transcription factor binding. In contrast, shared binding at *SLC6A3* coincides with

shared regulation (Fig 3.11 c). Both AR and GR have a strong binding peak in an intronic region of *SLC6A3* and induce hormone-dependent gene expression of this gene.
Together these examples exemplify that for a subset of genes receptor-specific gene regulation can indeed be explained by binding of the corresponding hormone receptor.

### 3.2.5. ChIP-qPCR Validation of Receptor-Specific Binding Peaks

The ChIP-seq analysis of both receptors revealed that AR and GR have shared binding sites but each receptor also binds to receptor-specific sites in the genome.
To confirm these findings experimentally, I performed ChIP-qPCR analyses for AR and GR with primers for either GR- or AR-specific binding sites (Fig 3.12). Additionally, I treated GR cells with different Dex concentrations in the range of 1 μM to 0,5 nM to test if GR-specific binding is a result of the higher receptor number compared to AR or a result of GR-specific properties (Fig 3.13). A Dex concentration of 1 μM is saturating and activates all GR molecules present in the cell (see calculation for hormone-bound GR fraction in Tab 2.13). Incubating GR cells with 1 nM Dex yields comparable cellular levels of hormone-bound GR and AR (see calculation for hormone-bound AR fraction in Tab 2.14).

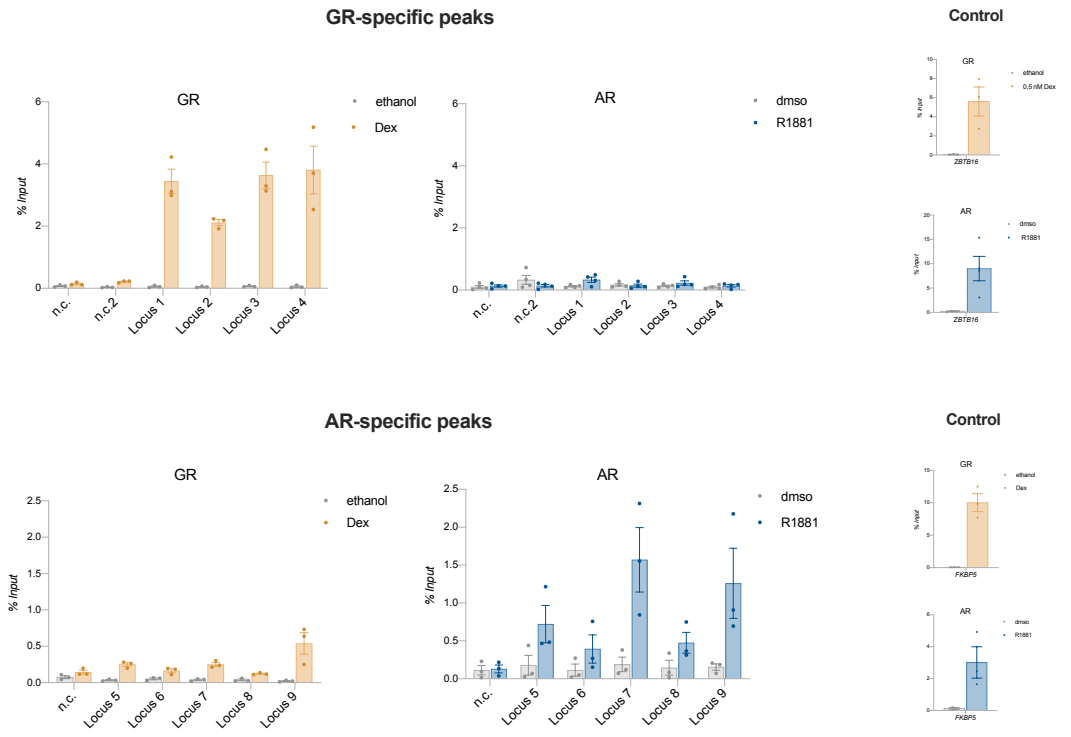



**Figure 3.12. Receptor-specific binding sites in GR and AR cells**. ChIP-qPCR of GR- and AR-specific peaks in GR and AR cells. GR cells were treated with 0,5 nM, 1 nM, 1 μM Dex or ethanol as a vehicle control for 1,5h. AR cells were treated with 5 nM R1881 or dmso for 4h. Negative control regions (n.c.) are neither occupied by AR or GR. The error bars represent ±SEM.

Consistent with the ChIP-seq data, the genomic loci 1-4 (GR-specific peaks) were all enriched in the GR ChIP but no enrichment was detected in the AR ChIP. Furthermore, five AR-specific peaks were validated by ChIP-qPCR in both cell lines (Fig b). All peaks showed a significant enrichment in the AR ChIP. However, a weak GR binding could also be detected, but to a lesser extend compared to AR.

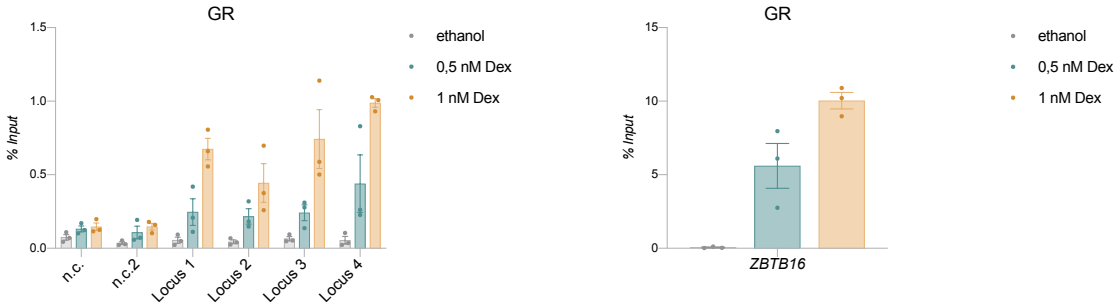



**Figure 3.13. GR-specific binding remains with reduced hormone**. ChIP-qPCR analysis of GR-specific peaks (Locus 1-4) and a positive control region (*ZBTB16*) in U2OS-GR cells. GR cells were treated with 0,5 nM, 1 nM Dex and ethanol as a vehicle control for 1,5h. Negative control regions (n.c.) are neither occupied by AR or GR. The error bars represent ±SEM.

With the ChIP-qPCR analysis, I was able to confirm four GR-specific binding peaks that show an enrichment for GR only. At lower Dex concentrations, I found decreased GR enrichment. Nevertheless, an enrichment was still detectable even at a Dex concentration of 0,5 nM. Together, these results indicate that GR-specific binding we have detected is not a result of the high receptor number but is rather due to GR-specific molecular properties.

## 3.3. Deletion of DNA Binding Sites of Receptor-Specific Genes

### 3.3.1. Differential Regulation of GILZ and Dissection of the Locus

The stacked bar graphs (Fig 3.10 b) indicate that for some genes receptor-specific regulation can be explained by receptor-specific binding. However, this is not true for the majority of genes that are regulated in a receptor-specific manner but harbor binding sites that are shared. To better understand what drives receptor-specific regulation and if it can always be explained by binding, I looked into receptor specific regulation of individual genes with shared binding sites.

The Glucocorticoid-Induced Leucine Zipper (*GILZ*) is a known target gene of GR and plays a role in the anti-inflammatory response upon GR signaling (Berrebi et al. 2003). I have examined the binding profile of both GR and AR in the *GILZ* locus by analyzing ChIP-seq data of the two receptors and found that both receptors share their binding peaks in this locus (Fig 3.14 a). Next, I analyzed by qPCR the gene expression of *GILZ* in GR and AR cells upon hormone induction (Fig 3.14 b). Interestingly, *GILZ* is selectively and very robustly regulated by GR and no regulation was observed upon AR activation. Although there is no differential binding of GR and AR at this locus, *GILZ* is only differentially regulated by GR. This observation indicates that mechanisms downstream of binding seem to be crucial for the receptor-specific regulation of *GILZ*.

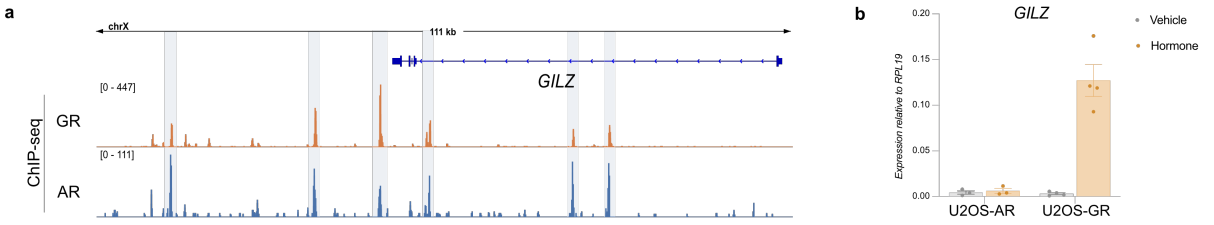



**Figure 3.14. Representation of the *GILZ* locus and transcriptional regulation of *GILZ* by GR. (a)** Genome-browser screenshot of the *GILZ* locus showing the *GILZ* RefSeq gene and ChIP-seq tracks of GR and AR. The ChIP peaks bound by both receptors are highlighted in grey. One representative track is shown from two biological replicates. **(b)** Relative expression of *GILZ* in U2OS-AR cells treated with dmso as vehicle control or 5 nM R1881 and U2OS-GR cells treated with ethanol as vehicle control or 1 µM Dex for 24h. The average gene expression of three biological replicates ±SEM is shown. The BigWig tracks for data visualization were generated by Melissa Bothe.

In a previous study it was shown that the GR peaks upstream and downstream of the promoter are functional and crucial for the regulation of *GILZ* by GR (Thormann et al. 2018). Genomic deletions using CRISPR/Cas9 were carried out to confirm that the binding peaks at the *GILZ* locus are indeed functional and responsible for the gene regulation by GR. The deletion experiments were performed by Verena Thormann and Laura Glaser. Binding peaks upstream of the TSS of *GILZ* as well as the first binding site downstream of the TSS within an intronic region were deleted in the GR cells (Fig 3.15a). The single-cell derived deletion clones were genotyped by Sanger sequencing in the previously published study. First, the clones were tested for the expression of *FKBP5* by qPCR to confirm that GR is still functional and the CRISPR experiment did not interfere with receptor function in general (Fig. 3.15 b). *FKBP5* levels were slightly reduced in the clonal lines compared to the wild-type cells, nevertheless the receptor was functional given that its expression increases in response to hormone-treatment. Next, the cells were tested for their *GILZ* expression (Fig 3.15 c). Indeed, the clonal cells carrying the deletion did not show *GILZ* regulation in response to GR activation. This experiment confirms, that the removed binding peaks are responsible for the regulation of *GILZ* by GR. However, even though each of the peaks in these regions responsible for GR-dependent regulation are also occupied by AR, AR fails to regulate this gene indicating that events that occur downstream of binding are responsible for the receptor-specific regulation of the *GILZ* gene observed.

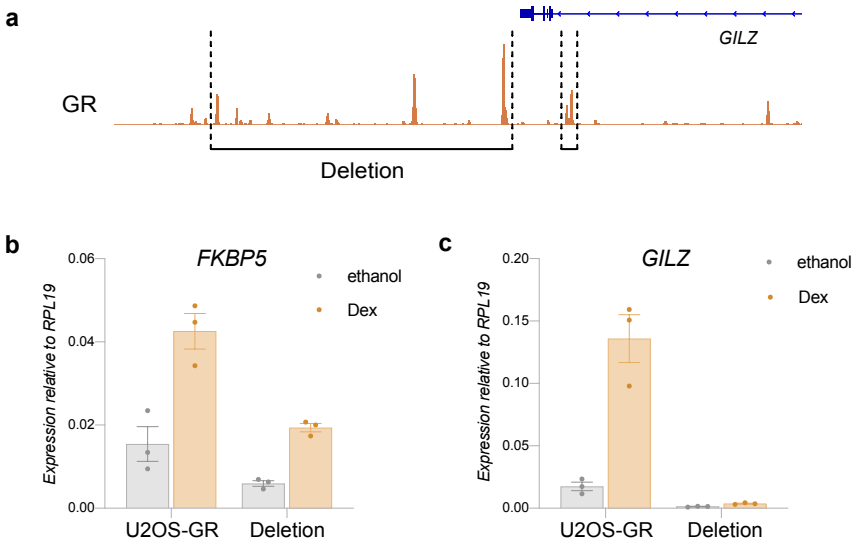



**Figure 3.15. Genomic deletion of GR binding sites leads to a loss of *GILZ* regulation in the GR cells. (a)** Genome-browser screenshot of the *GILZ* locus showing the *GILZ* gene and CRISPR/Cas9 deleted regions in the GR ChIP-seq tracks. **(b)** Relative expression of *FKBP5* in U2OS-GR wt cells and GR clones with the genomic deletion treated with ethanol or 1 μM Dex for 4h **(c)** Relative expression of *GILZ* in U2OS-GR wt cells and GR clones with the genomic deletion treated with ethanol or 1 μM Dex. The genome-edited clonal lines were previously generated by Verena Thormann and Laura Glaser (Thormann et al. 2018).

### 3.3.2. Differential Regulation of AQP3 and Dissecting the AQP3 Locus

Next, I have looked into the RNA-seq and ChIP-seq data to find genes that are specifically regulated by AR, but harbor nearby genomic binding sites which are occupied by both AR and GR. Aquaporin 3 is membrane protein that plays a role in water and glycerol transport (Sasaki, Ishibashi, and Marumo 1998) and is expressed in various cancer types (Pinho, Matias, and Gaspar 2019). Downstream of the *AQP3* gene, I have found two sites that are both strongly bound by both AR and GR (Fig 3.16 a). Yet, consistent with the RNA-seq data, qPCR analysis showed that *AQP3* is regulated only by AR (Fig 3.16 b). So, although GR equally binds the locus it is not able to induce *AQP3* expression.

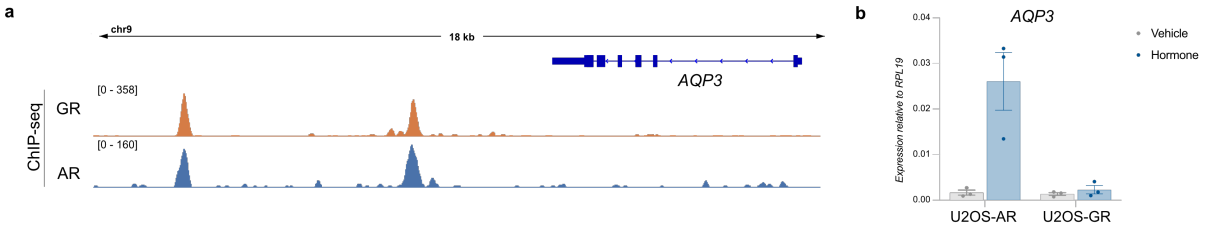



**Figure 3.16. Representation of the *AQP3* genomic locus and differential regulation of *AQP3* by AR.** **(a)** Genome-browser screenshot of the *AQP3 locus* showing the *AQP3* RefSeq gene and ChIP-seq tracks of GR and AR. One representative ChIP track is shown from two biological replicates. **(b)** Relative expression of *AQP3* in U2OS-AR cells treated with dmso as vehicle control or 5 nM R1881 and U2OS-GR cells treated with ethanol as vehicle control or 1 μM Dex for 24h. The average gene expression of three biological replicates ±SEM is shown. The BigWig tracks for data visualization were generated by Melissa Bothe.

In a next step, I wanted to test if the two binding peaks near the *AQP3* gene are indeed required for its regulation by AR. Therefore, I deleted the region comprising the two binding peaks using CRISPR/Cas9 genome editing in the AR cell line (Fig 3.18 a). First, I designed two single guide RNAs targeting genomic regions spanning the peaks. Second, U2OS-AR cells were transfected with plasmids containing the Cas9 enzyme, Puromycin resistance and a single guide RNA. Third, single cell-derived clones were grown and deletions of the region was tested by a diagnostic PCR and Sanger sequencing (Fig 3.17).

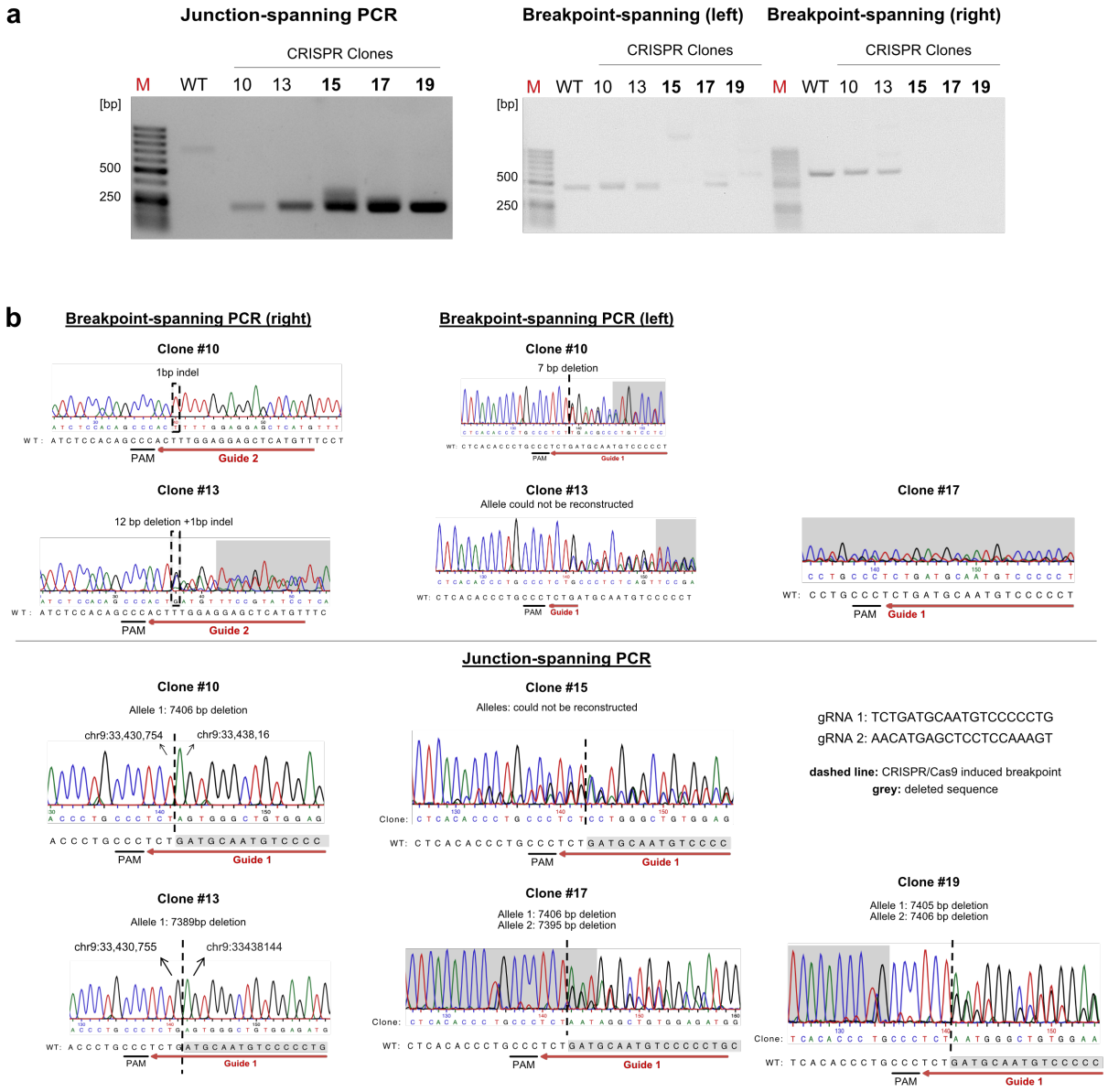



**Figure 3.17. Genotyping the genome-edited *AQP3* AR clones.** **(a)** The CRISPR targeted *AQP3* locus was PCR amplified from genomic DNA following size-separation of the PCR products by electrophoresis on a 1% agarose gel. For the junction PCR, primers are placed outside the breakpoints to detect CRISPR clones carrying the desired genomic deletion at the *AQP3* locus. Primers spanning the breakpoints detect WT alleles. The expected amplicon size of the left breakpoint is 476 bp and 686 bp for the right breakpoint for WT samples. Amplicon of left breakpoint of clone #17 indicates a partial deletion of the second allele. M marks the lane with the DNA size marker GeneRuler 50bp. **(b)** Sequencing of the breakpoint PCR amplicon validated the presence of a WT allele of the clones #10 and #13. Sequencing of the junction PCR amplicons evidenced a genomic deletion of the CRISPR clones #10, #13, #15, #17, and #19 at the *AQP3* locus.

I obtained two single-cell derived clones with a heterozygous deletion (clone 10 and 13) and three clones carrying a homozygous deletion (clone 15, 17 and 19). The heterozygous clones showed a band in the PCR using primers, that were spanning the CRISPR/Cas9-induced breakpoint whereas homozygous clones did not show a band (Fig 3.17 a). Clone 17 on the other hand showed a band for left breakpoint only. However, in the junction-spanning PCR there are two edited alleles of clone 17, this is why I assigned it as homozygous deletion. The deletions were confirmed using primers spanning the created junction. The heterozygous clones contained one allele with a deletion whereas the homozygous clones had two, or more edited alleles (Fig 3.17 b).

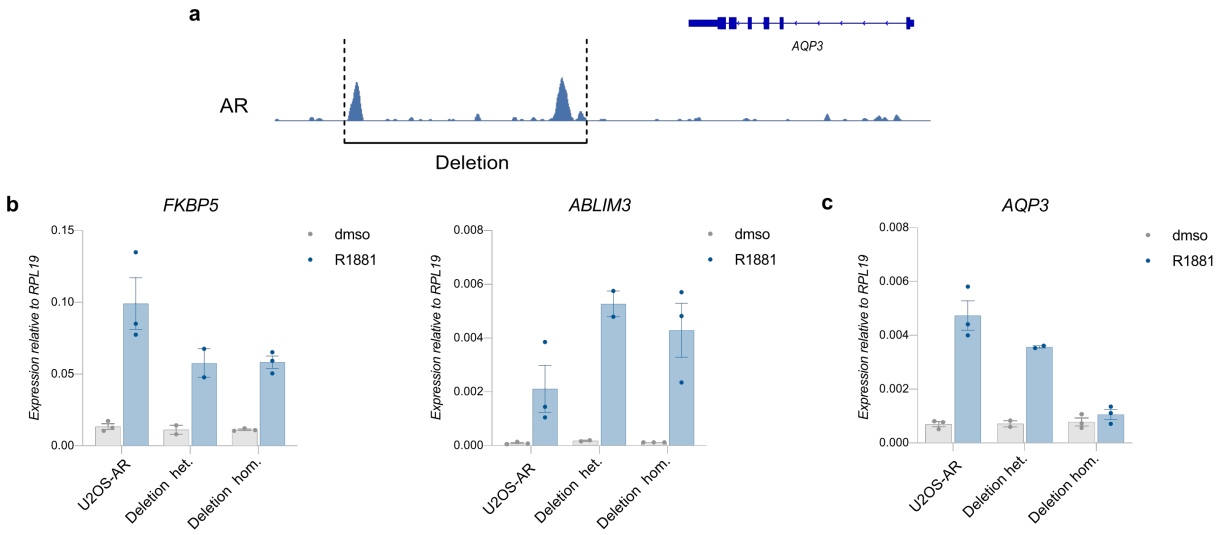



**Figure 3.18. Genomic deletion of nearby AR binding sites leads to a loss of *AQP3* gene regulation in the AR cells. (a)** Genome-browser screenshot of the *AQP3* locus showing AR ChIP-seq tracks at the *AQP3* gene and the region that was deleted using CRISPR/Cas9. **(b)** Relative expression of *FKBP5 and ABLIM3* in wildtype AR and clones with a deletion in the *AQP3* locus. The average gene expression of three biological replicates ±SEM is shown. **(c)** Relative expression of *AQP3* in AR wt cells, AR clones carrying a heterozygous deletion and AR clones carrying a homozygous deletion treated with dmso or 5 nM R1881 for 24 h. The average gene expression of three biological replicates ±SEM is shown. ChIP-seq data analysis was done by Melissa Bothe.

To confirm that AR is still functional in the clonal cell lines, I first performed a qPCR analysis for the genes *FKBP5* and *ABLIM3* in the clonal lines (Fig. 3.18 b). Both genes are inducible in a hormone-dependent manner indicating that AR regulation is still intact. The gene expression analysis of the *AQP3* gene demonstrated that the clones with a heterozygous deletion show a reduced *AQP3* activation by AR (Fig 3.18 c). However, the clones with the homozygous deletion did not show any induction of *AQP3* expression by AR indicating that AR directs the regulation of *AQP3* by binding at the locus upstream of the gene. Thus, similar to what we observe for *GILZ*, these experiments indicate that events downstream of receptor binding direct receptor-specific regulation given that even though GR and AR bind these genomic regions near the *AQP3* gene, this binding only results in regulation of this gene for AR.

## 3.4. Differential Binding of AR and GR to Low-Accessibility Sites

### 3.4.1. ATAC-Seq Analysis

Chromatin accessibility varies across different cell types and subsequently also does transcription factor binding and gene regulation. The majority is of transcription factors binds to accessible and nucleosome depleted chromatin (John et al. 2011). This in turn results in a further opening of the chromatin by nucleosome rearrangements and the expression of specific target genes. Thus, chromatin accessibility is an important driver of TFs binding followed by gene activation.

I performed ATAC-seq in both cell lines, to answer the question if chromatin accessibility plays a role in directing AR and GR to their genomic binding sites, and subsequently, plays a role in receptor-specific gene regulation. Moreover, I wanted to compare the consequences of AR and GR binding on the chromatin accessibility. ATAC-seq is a method to detect nucleosome-depleted "accessible" regions genome-wide. By including vehicle-treated cells, I was able to detect changes in chromatin accessibility upon hormone activation. Additionally, I was able to determine the basal ATAC-seq signal levels and the chromatin state at the receptor binding sites prior to binding. The ATAC-seq experiments in GR cells and data analysis were performed by Melissa Bothe.

In brief, the cells were lysed to prepare the nuclei in a first step. In a second step, the Tn5 transposase fragmented the DNA by inserting Nextera adapters in nucleosome depleted regions. The DNA fragments were isolated and amplified by PCR and sequencing adapters were added. After a purification step to get rid of adapters, the libraries were sent for sequencing.
The ATAC-seq results of AR and GR are represented as a heatmap with the ATAC signal centered on the summit (Fig 3.19 a). Profile plots show the centered average of the ATAC-signal and allow for an easier direct quantitative comparison (Fig 3.19 b). The ATAC signal was clustered in an AR-specific, shared and GR-specific group based on the ChIP-seq data (Fig. 3.9). Treatment of the U2OS-AR cell line with R1881 to activate AR results in a significant increase in ATAC signal indicative of chromatin opening at AR-specific and shared binding sites. For GR-specific sites, only little increase is detected upon AR activation by addition of its hormone consistent with a lack (or lower levels) of AR binding at these sites. GR on the other hand shows a strong induction at GR-specific sites and shared binding sites whereas no change is observed at AR-specific binding sites. Interestingly, when comparing the basal ATAC levels of the different groups of binding sites without hormone treatment, we found that the GR-specific sites show the lowest ATAC signal in both cell lines (Fig. 3.19 b).

This finding suggests, that AR and GR have different binding preferences regarding the chromatin state. GR-specific sites are enriched in inaccessible chromatin whereas AR preferentially binds to accessible sites in the genome.

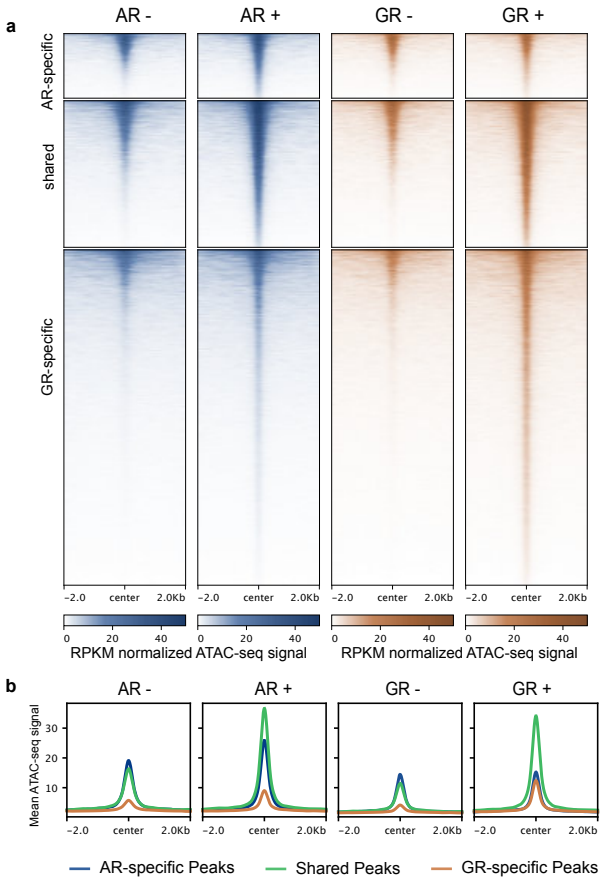



**Figure 3.19. GR-specific sites are preferentially located in inaccessible chromatin. (a)** ATAC-seq summit-centered heatmap of ATAC-seq signal in AR and GR cells with vehicle (-) and hormone treatment (+) conditions. The AR cells were treated with 5 nM R1881 for 4 h and GR with 1 µM Dex for 1,5 h. The ATAC-seq peaks are clustered into AR-specific, shared and GR-specific binding sites. The merged ATAC signal from two biological replicates is shown. **(b)** Profile plot showing the average ATAC-signal in AR and GR at AR-specific, shared and GR-specific sites. Data analysis was done by Melissa Bothe.

### 3.4.2. Opening Genomic Sites with Low-Accessibility

The ATAC-seq analysis revealed that a large fraction of GR-specific peaks is located in inaccessible chromatin. To validate this finding, I performed a qPCR analysis with additional ATAC libraries from AR and GR cells. For this analysis I chose primers for peaks that are specifically bound by GR and show a low basal ATAC signal. GR-binding was previously confirmed for all for sites (Fig. 3.12) whereas no binding was observed for AR.

The ATAC analysis showed that each of the loci open up upon GR induction, which is reflected in an increased ATAC signal (Fig. 3.20 a). The low basal ATAC signal of these loci confirms that those sites were initially closed prior to receptor binding. Treatment of the GR cells with a lower Dex concentration still resulted in a chromatin opening. This indicates that opening of these sites is not simply a result of the higher number of GR molecules but rather might be due to a GR-ability to bind to sites embedded in closed chromatin. In contrast, hormone activation of AR showed no increase in ATAC signal at these loci 1-4 (Fig. 3.20 b) consistent with a lack of binding of AR at these loci. The binding region in the *ZBTB16* locus was used as a control to show, that the experiments worked for AR as this region is bound by AR and opens upon addition of AR's ligand R1881.

Together, I was able to validate by ATAC qPCR that GR is endowed with a receptor-specific ability to interact with inaccessible chromatin and that such sites are "opened" upon GR induction whereas AR cannot bind or induce opening of these sites. The GR activation with a lower Dex concentration indicate that binding and opening of closed DNA is a GR-specific feature and not a simple result of the higher number of GR receptors compared to AR in the cell lines examined.

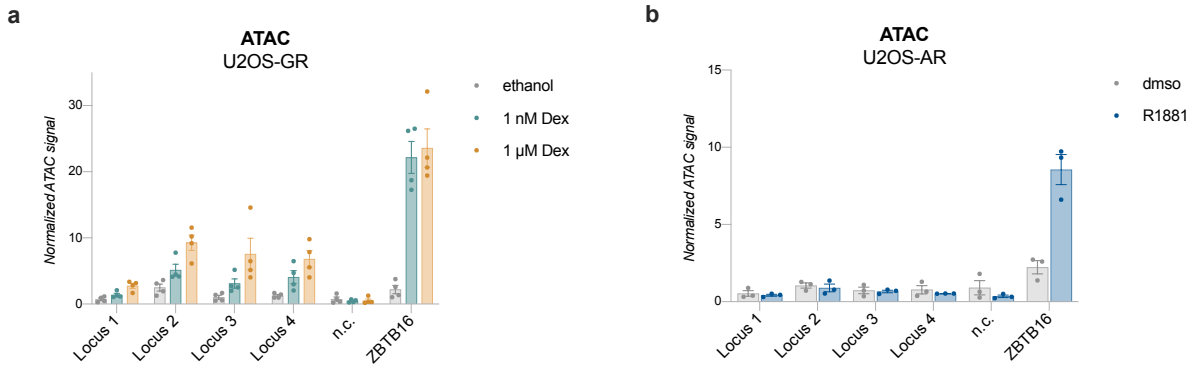



**Figure 3.20. Opening of inaccessible sites by GR**. Bar graphs showing ATAC-qPCR results of GR-specific peaks in GR **(a)** and AR cells **(b)**. GR cells were treated with 0,5 nM, 1 nM, 1 µM Dex and ethanol for 1,5h. AR cells were treated with 5 nM R1881 or dmso for 4h. The negative control region (n.c.) is neither occupied by AR or GR. The error bars represent ±SEM.

### 3.4.3. GR-Specific Peaks Associated with Closed Chromatin

Specific histone modifications have been correlated with open as well condensed chromatin (Jenuwein and Allis 2001). Acetylation is generally linked to active transcription and open chromatin. For example, H3K27ac is a widely used mark to identify active enhancers whereas H3K27me3 acts in an antagonistic manner and is associated with transcriptional silencing. H3K9me3 is a further histone mark associated with silenced transcription and condensed chromatin.

Our ATAC-seq results suggested that various GR-specific sites are located in inaccessible chromatin. To further substantiate this observation, we performed ChIP-seq analyses in both cell lines for H3K27me3 and H3K9me3, two marks that correlate with closed chromatin. The ChIP-seq experiments were done in collaboration with Alisa Fuchs and the data analysis by Melissa Bothe. These ChIP-seq results are represented as profile plots showing the average ChIP signal centered on the summit. The ChIP-seq peaks are again clustered in AR-specific, shared and GR-specific binding sites. The ChIPs were performed with AR and GR under hormone and vehicle treatment conditions. Both histone marks H3K27me3 (Fig. 3.21 a) and H3K9me3 (Fig. 3.21 b) show the highest basal signal at GR-specific sites followed by the shared and AR-specific sites. This indicates that GR-specific sites are more enriched in marks associated with condensed chromatin compared to the shared and AR-specific binding sites. Upon hormone induction of GR, we can observe a dip of the ChIP signal at the peak summit of all three peak clusters indicating that these regions become less condensed upon GR activation. AR induction, however, did not show significant changes of the H3K27me3 or H3K9me3 ChIP-seq signal at the cluster of GR-specific peaks consistent with a lack of binding of AR.

These results further indicate that, in contrast to AR, GR is able to bind to inaccessible chromatin enriched for H3K27me3 or H3K9me3. Furthermore, after hormone activation GR induces a dip in ChIP-seq signal that is likely associated with nucleosome rearrangement and opening of the chromatin.

Interestingly, although AR and GR have a shared sequence preference, the two receptors show differential genomic binding in U2OS cells. One explanation might lie in their different abilities to interact with inaccessible chromatin with GR being able to bind genomic sites that are not accessible for AR.

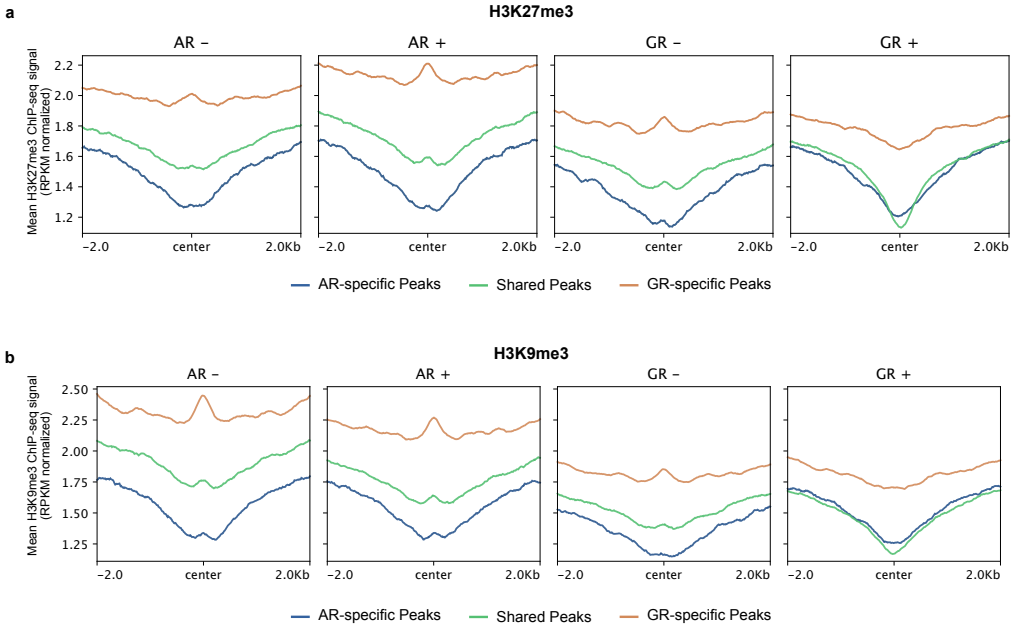


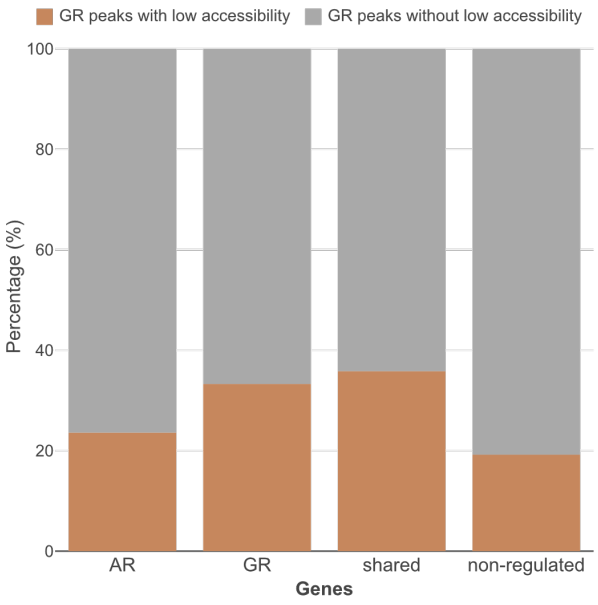

**Figure 3.21. GR-specific sites are enriched for histone marks associated with closed chromatin.** Profile plots showing the average H3K27me3 **(a)** and H3K9me3 **(b)** ChIP-seq signal centered on the peak summit in AR and GR cells. Peaks are clustered in AR-specific, shared and GR-specific sites. The AR cells were treated with 5 nM R1881 or vehicle for 4 h and GR with 1 μM Dex or vehicle for 1,5 h. The ChIP-seq experiment was performed in collaboration with Alisa Fuchs and Data analysis was done by Melissa Bothe.

### 3.4.4. GR-Regulated Genes are Enriched for GR Peaks in Inaccessible Chromatin

The previous analyses have shown that a set of GR peaks is located in closed chromatin with a low ATAC signal and enriched for chromatin marks that are specific for condensed chromatin. In the following analysis we wanted to test whether GR binding in inaccessible chromatin is contributing to GR-specific gene regulation (Fig. 3.22). This computational analysis was performed by Gözde Kibar.

A region around each TSS of each gene (+/- 30 kb) was scanned for the presence of an GR peak with or without low accessibility. Strikingly, when comparing to the non-regulated genes GR-specific genes are indeed enriched for GR peaks with a low accessibility.

This data suggests that GR-binding at inaccessible chromatin seems to be able to drive GR-specific gene regulation.




| | Low access. GR peaks at non-regulated genes |
|---|---|
| Low access. GR peaks at AR-spec genes | p-val = 0.1072e-10 |
| Low access. GR peaks at GR-spec genes | p-val = 4.059e-10 |
| Low access. GR peaks at shared genes | p-val = 1.13e-10 |

**Figure 3.22. Intersecting GR binding at inaccessible sites and differential gene regulation genome-wide**. The stacked bar graphs represent the percentage of GR peaks with low accessibility (orange) and without low accessibility (grey) assigned to each category of differentially regulated genes (AR, GR, shared, non-regulated). p-values were calculated using a Fisher's exact test. The data analysis was done by Gözde Kibar.

## 3.5. Differential H3K27 Acetylation at Genomic Loci Occupied by GR and AR

H3K27ac is a histone modification that is associated with active enhancers as well as transcriptional activity (Creyghton et al. 2010). This histone mark can be exclusively found in open chromatin and is widely used to study enhancer activity and gene regulation. To better understand how H3K27ac is associated with receptor-specific binding and activity, ChIP-seq experiments for H3K27ac were done in the AR and GR cells.

The ChIP-seq experiments were carried out in collaboration with Alisa Fuchs and the data analysis was done by Melissa Bothe. The ChIP-seq results for both cell lines are represented as a heatmap with the ChIP-seq signal centered on the summit of ChIP-seq peaks for AR and GR respectively (Fig. 3.23 a). Profile plots are showing the centered average ChIP-signal for a direct quantitative comparison (Fig. 3.23 b), however, due to the very subtle changes upon AR activation a different scale of the y-axis was chosen for the profile plots. Again, the ChIP-seq peaks were clustered in the groups AR-specific, shared and GR-specific. GR activation induces very strong acetylation in the shared and GR-specific binding sites whereas acetylation at AR-specific sites is large unchanged. Notably, GR-specific binding sites show the lowest basal H3K27ac levels in both cell lines consistent with GR-specific loci mapping to regions of low chromatin accessibility.

Remarkably, there is only very little H3K27ac induced by AR activation in all three binding categories. Only the shared binding sites show a clear increase in H3K27ac upon AR induction but to a lesser extent compared to GR. This finding is surprising because H3K27ac is postulated to be correlated with enhancer activity and transcription (Creyghton et al. 2010). Thus, although AR induces active transcription of many genes, this activation does not appear to require an increase in H3K27ac levels.

Together, the ChIP-seq experiments have shown that GR is able to induce strong H3K27ac whereas AR is only capable to induce very little changes. This result indicates that transcriptional activity can be uncoupled from an increase in H3K27ac. Moreover, the low basal acetylation levels at GR-specific sites correlate well with our findings that many of those sites are inaccessible.

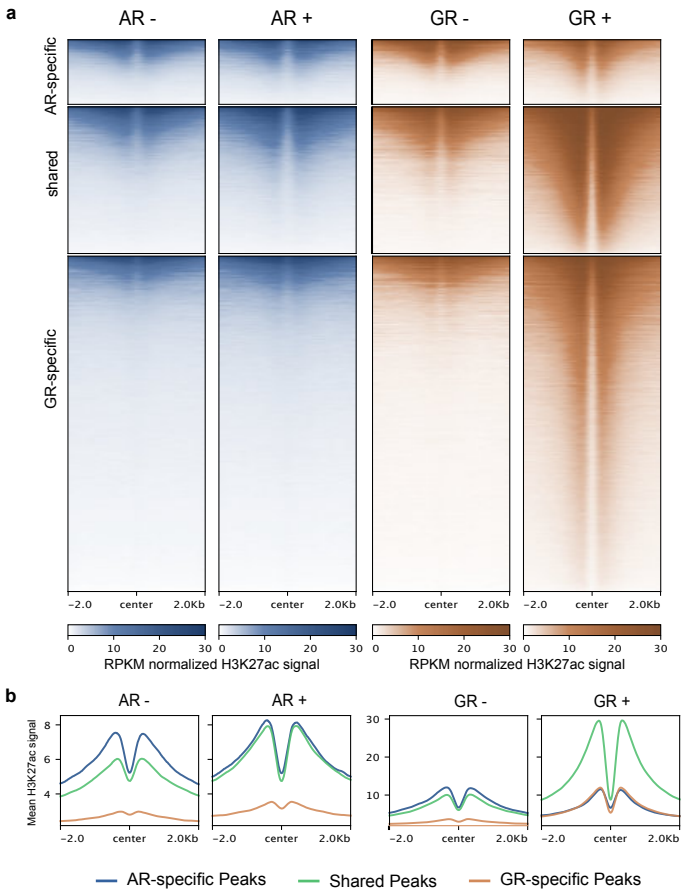



**Figure 3.23. GR induces a stronger increase in H3K27ac signal compared to AR**. **(a)** Heatmaps showing H3K27ac ChIP-seq at genomic sites bound by AR only, shared by both receptors and bound by GR only in AR and GR cells. U2OS-GR cells were treated with ethanol as vehicle control or 1 μM Dex for 1,5 h and U2OS-AR cells were treated with dmso as vehicle control or 5 nM R1881 for 4 h. **(b)** Profile plot showing the average H3K27ac ChIP-seq signal in AR and GR at AR-specific, shared and GR-specific sites. The ChIP-seq experiments were done in collaboration with Alisa Fuchs and the data analysis was performed by Melissa Bothe.

### 3.5.1. H3K27ac Levels at Genes Regulated by AR and GR

The H3K27ac ChIP-seq analysis in both cell lines suggested that AR induces only little change in H3K27ac levels when compared to GR. To understand in more detail how well H3K27ac can be connected with gene regulation in our model system I analyzed the H3K27ac levels at shared binding sites near genes that regulated by both receptors.

First, I confirmed the regulation of the four genes *TMEM63C, FAM105A, FAM104A* and *MYO1E* in GR (Fig. 3.24 a) and AR cells (Fig. 3.24 b). I found that *TMEM63C, FAM105A, FAM104A* are regulated stronger by GR, whereas *MYO1E* showed a similar regulation level. However, the four genes are inducible by both receptors. Second, I performed H3K27ac ChIP experiments and analyzed the H3K27ac enrichment at receptor binding sites near the four genes

regulated by AR (Fig. 3.24 c) and GR (Fig. 3.24 d). Prior to the ChIP experiment, the cells were treated with hormone for 4 h and 24 h. GR induces a significant increase in H3K27ac after 4 h and an even stronger signal after 24 h at all four regions. Moreover, the H3K27ac signal correlates with the gene expression level in the GR cells. The gene *FAM104A* shows the strongest regulation as well as the highest H3K27ac. Interestingly, AR cells did not show an increased H3K27ac after 4 h and only a small increase in signal after 24 h.

Thus, although these four genes are significantly regulated by both receptors, a strong increase of H3K27ac could be detected only in the GR cell line. AR induction, however, did not result in a significant change in H3K27ac levels. This finding suggest that gene regulation can be uncoupled from the increase in H3K27ac as can be seen in AR cells where regulation of genes occurs in the absence of marked changes in H3K27ac levels.

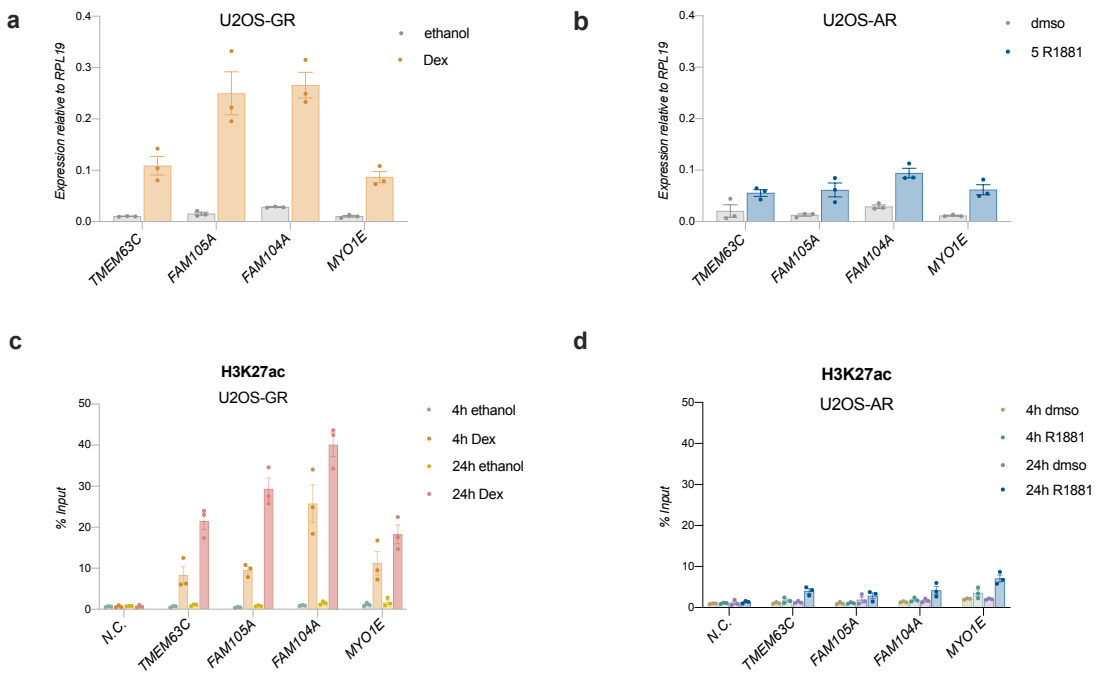



**Figure 3.24. Genes regulated by GR and AR show little increase in H3K27 acetylation in AR cells**. qPCR analysis of genes regulated by both receptors and ChIP-qPCR of peaks within these genes bound by both receptors. Relative expression of the genes *TMEM63C, FAM105A, FAM104A* and *MYO1E* in **(a)** GR cells treated with ethanol or 1 µM Dex and **(b)** AR cells treated with dmso or 5 nM R1881 for 24h. The average gene expression of three biological replicates ±SEM is shown. ChIP-qPCR analysis for H3K27ac within genomic regions regulated and bound by both receptors. **(c)** H3K27ac ChIP-qPCR analysis in GR cells treated with ethanol as vehicle control or Dex for 1,5 h. and **(d)** AR cells treated with dmso or 5 nM R1881 for 4 h. Error bars indicate ±SEM (n=3). Negative control regions (n.c.) are neither occupied by AR or GR.

### 3.5.2. H3K27ac Levels at the AQP3 Locus

H3K27ac ChIP-seq and qPCR analyses of selected targets showed that AR induces only little H3K27ac in comparison to GR. However, compared to the receptor-specific binding sites, the category of the shared binding sites showed the strongest H3K27ac signal increase in AR cells.
*AQP3* is a gene specifically regulated by AR. However, GR and AR both share the identical two binding peaks upstream of the *AQP3* TSS. In order to see if the AR-specific *AQP3* expression is accompanied by an increase in H3K27ac levels, I generated a genome-browser screenshot of the *AQP3* locus showing ChIP-seq data of both receptors as well their corresponding H3K27ac ChIP-seq signal (Fig. 3.25). Although *AQP3* is not regulated by GR, there is a strong increase in H3K27ac at both binding sites. Interestingly, AR induces almost no H3K27ac signal at these two binding sites although *AQP3* is significantly regulated by AR.
This finding indicates, that an increase in H3K27ac is not necessarily connected to gene regulation and enhancer activity can possibly be decoupled from an increase in H3K27ac levels.

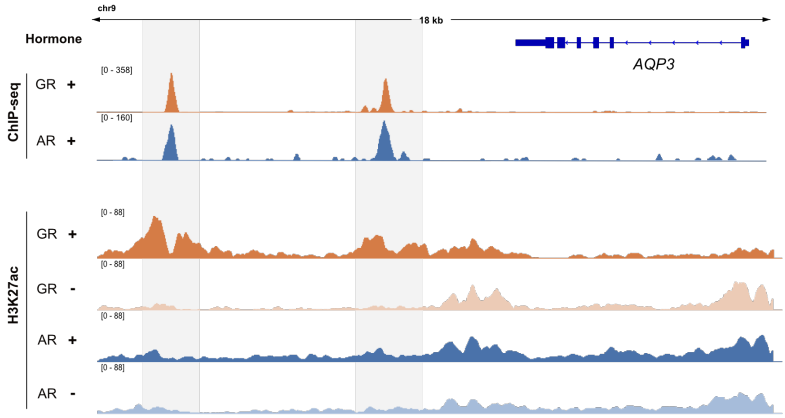



**Figure 3.25. *AQP3* locus shows stronger H3K27ac by GR compared to AR**. Genome-browser screenshot of the *AQP3* locus showing the *AQP3* gene, GR and AR ChIP-seq and H3K27ac ChIP-seq tracks of GR and AR cells. U2OS-GR cells were treated with ethanol as a vehicle control or 1 µM Dex whereas U2OS-AR cells were treated with dmso as a vehicle control or 5 nM R1881. The Peaks bound by both receptors are highlighted in grey.

## 3.6. Comparing Intrinsic Enhancer Activities of AR and GR

### 3.6.1. Establishing STARR-seq for AR Cells

STARR-seq is a massively-parallel reporter assay for the identification of enhancers in a genome-wide and quantitative manner (Arnold et al. 2013). Different sources of DNA can be used as an insert for the STARR-seq reporter and subsequently tested for their enhancer activity. The DNA insert is placed downstream of a minimal promoter and a GFP-tag (Fig. 3.26 a). Upon activation of the reporter the enhancer sequence directs its own transcription that can be detected as mRNA output. Thus, the mRNA output directly correlates with the activity of the enhancer. That is why the STARR-seq assay not only identifies active enhancers but also provides information about the strength of the enhancer activity.

We performed a STARR-seq assay to find out if GR and AR have different preferences in enhancer usage and if enhancer activity can explain why the two receptors are able to differentially regulate genes with nearby binding sites that are bound by both receptors. The STARR-seq assay can be applied to either measure the enhancer activity of a single sequence analyzed by qPCR (Fig. 3.26 b) or to study a whole sequence library when analyzed by Next Generation Sequencing (Fig. 3.26 c).

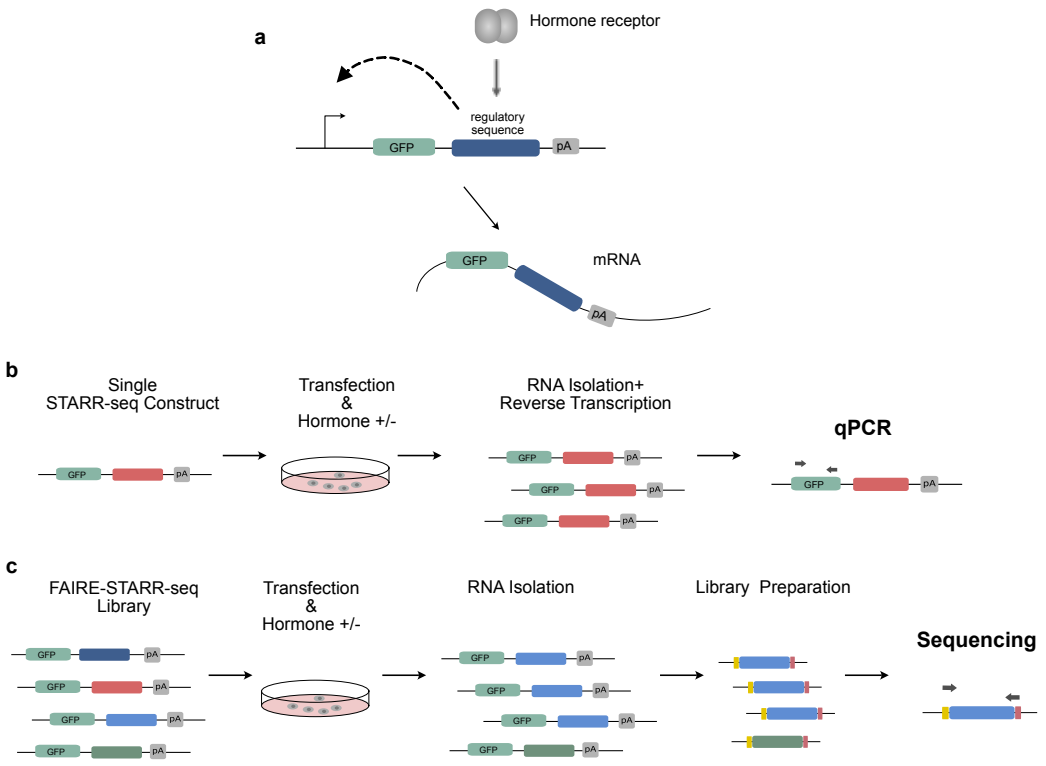



**Figure 3.26. Experimental STARR-seq procedure for testing individual enhancers or a STARR-seq library**. **(a)** The STARR-seq vector contains a minimal promoter followed by GFP, a potential regulatory sequence and a polyadenylation site (pA). Upon activation of a hormone receptor, the enhancer drives its own mRNA expression. **(b)** A single STARR-seq vector containing one individual DNA insert was transfected into AR or GR cells and treated with hormone or vehicle followed by RNA isolation and reverse transcription. The mRNA output was measured by qPCR with primers targeting GFP. **(c)** STARR-seq library was generated by inserting FAIRE DNA fragments (FAIRE ref) comprising genomic and accessible DNA into the STARR-seq vector. The library was transfected into AR or GR cells and treated with hormone or vehicle followed by RNA isolation and library preparation. The library was sequenced with primers targeting the adapters up- and downstream of the DNA insert. The STARR-seq library was generated by Alisa Fuchs and Stefanie Schöne.

The STARR-seq assay was previously established by Stefanie Schöne for the GR cell line (Schöne et al. 2018). In a first approach, I transfected AR and GR cells separately with two different STARR-seq constructs to optimize the STARR-seq assay for AR. One STARR-seq construct contained the hormone-inducible and previously identified enhancer *IP6K3* and the other construct contained a scrambled GR binding site (GBS) as negative control. I identified the *IP6K3* enhancer which is strongly inducible by AR and GR, by scanning preliminary STARR-seq results from AR and GR cells. Then, I introduced this sequence in a STARR-seq reporter and tested it for activation in both AR and GR expressing cells. The *IP6K3* reporter was indeed inducible by both receptors, and therefore, I decided to use it as a positive control for further improvements of the assay. The cells were either treated with hormone or vehicle control overnight. Furthermore, the cells were treated with the interferon inhibitors PKR and TBK1/IKK. A previous study has shown that

transfection with foreign DNA can cause an intracellular interferon response that in turn can interfere with reporter assays leading to false results (Muerdter et al. 2018). Moreover, it was shown that AR interacts with the interferon-induced RNase L and that interferons can antagonize AR signaling (Bettoun et al. 2005).

Without the addition of interferon inhibitors, AR showed only little activation of the *IP6K3* enhancer comparable to the activation of the scrambled GBS (Fig. 3.27 a). After adding interferon inhibitors, *IP6K3* expression was more robustly increased. In contrast, GR showed strong activation of *IP6K3* and no activation of the scrambled control without the addition of interferon inhibitors (Fig. 3.27 b). Addition of the inhibitors led to slight a reduction of the STARR-seq signal for GR possibly due to a lower cell viability. However, the signal was still strong and robust so I decided to perform the STARR-seq experiments I the presence of interferon inhibitors.

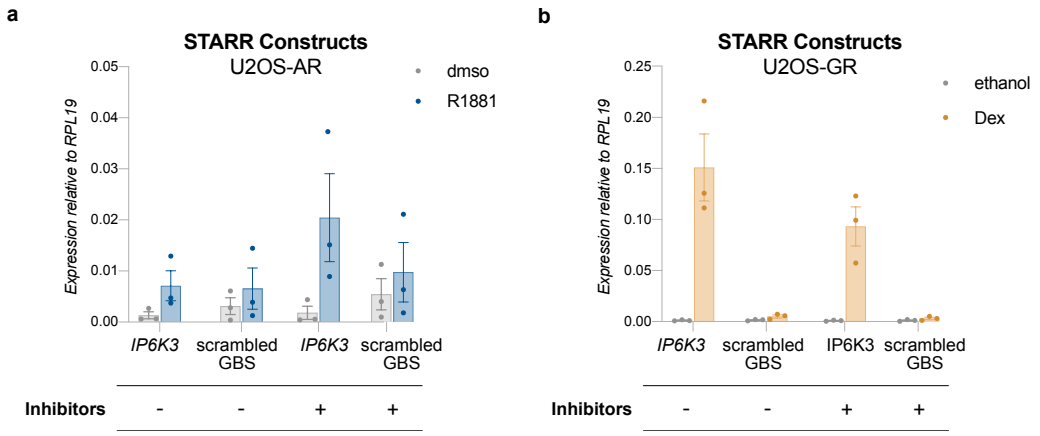



**Figure 3.27. Stable STARR-seq activation of selected constructs with addition of interferon inhibitors in AR cells**. Two STARR reporter constructs were tested for activation in **(a)** AR cells or **(b)** GR cells. *IP6K3* is a receptor-inducible enhancer whereas the scrambled GBS represents a mutated GR binding site that serves as negative control. Both STARR constructs were transfected in the presence and absence of the interferon inhibitors PKR and TBK1/IKK. Relative expression of the constructs is shown in U2OS-AR cells treated with dmso as vehicle control or 5 nM R1881 and U2OS-GR cells treated with ethanol as vehicle control or 1 µM Dex overnight. The average gene expression of three biological replicates ±SEM is shown.

### 3.6.2. Genome-wide Enhancer Activity

STARR-seq is a powerful method to detect active enhancers in a genome-wide and quantitative manner. I performed this method in the two cell lines expressing either AR or GR to directly compare their ability to activate enhancers and to answer the question if receptor-specific activities can be explained by differential enhancer activation by AR and GR.

Therefore, a STARR-seq library was generated that comprised DNA fragments from open chromatin. The FAIRE-STARR-seq library was generated by Alisa Fuchs and Stefanie Schöne. The open DNA fragments that where later integrated into the STARR-seq reporter, were generated in a FAIRE experiment (Giresi et al. 2007; Simon et al. 2013). In brief, GR cells were treated with hormone and crosslinked with formaldehyde. After the DNA extraction, the chromatin was sonicated and the open DNA fragments were separated from the nucleosome-bound DNA in a phenol-extraction step. In a last step, the open DNA fragments where cloned into the empty STARR-seq vector. The two cell lines were transfected with the FAIRE-STARR-seq library in the presence of interferon inhibitors and treated with hormone overnight. During the library preparation process, unique molecular identifier (UMI) were added to each fragment. During the data analysis, the UMIs were used to remove artifacts caused by PCR amplification.

The data analysis was carried out by Melissa Bothe. The STARR-seq signal was clustered according to the AR-specific, shared and GR-specific binding sites (Fig. 3.28 a). Overall, the STARR-seq signal intensity driven by GR is stronger compared to AR. However, the pattern of activation across all three binding categories is surprisingly similar. The enhancers of the shared sites show the strongest activity in both cell lines. Interestingly, the androgen receptor is activating enhancers within the AR- as well as the GR-specific cluster. In the cellular context, the GR-specific sites are not occupied by AR. But taken out of chromatin context AR becomes able to activate the enhancers present in those sites. GR, on the other hand, is a strong activator of enhancers in the shared and GR-specific sites. In contrast to AR, GR shows only weak activation of enhancers that are located at AR-specific binding sites (Fig 3.28 b) suggesting that the receptor-specific binding is driven by sequence rather than chromatin context.

Taken together, the genome-wide enhancer screening showed that GR is a stronger enhancer activator when comparing the signal intensities. Both receptors are able to activate the enhancers of the shared binding sites. Moreover, AR and GR both show activation of the enhancers located at their receptor-specific sites. Surprisingly, AR is also able to activate enhancers of GR-specific sites arguing that the chromatin context is indeed preventing AR from binding at these sites and that the lack of binding observed is not due to the sequence composition of these regions.

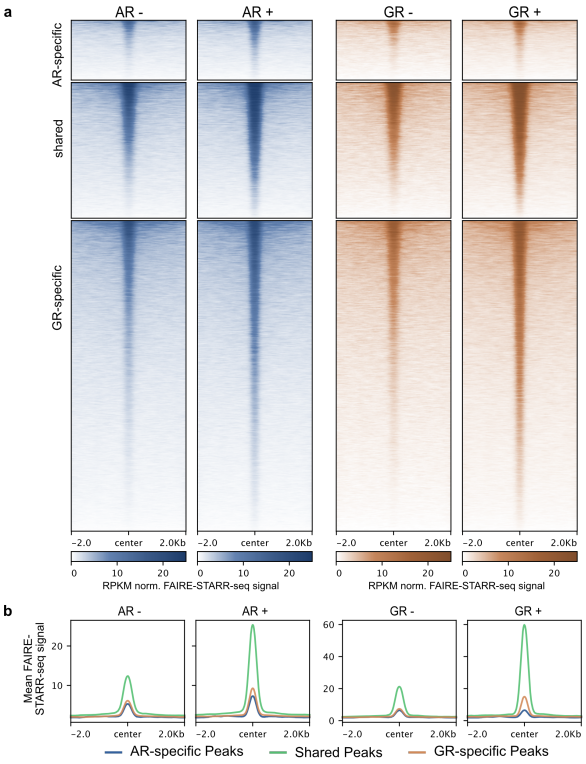



**Figure 3.28. Genome-wide STARR-seq activity of AR and GR cells**. **(a)** The Heatmap is representing the STARR-seq activity of AR (orange) and GR cells (blue) at AR-specific, shared and GR-specific binding sites. Both cell lines were transfected with the FAIRE-STARR library in presence of the interferon inhibitors PKR and TBK1/IKK. AR cells treated with dmso as vehicle control or 5 nM R1881 and GR cells with ethanol as vehicle control or 1 μM Dex overnight. The average of three independent replicates is shown. **(b)** Profile plot showing the average STARR-seq signal centered on the peak summit in AR and GR at AR-specific, shared and GR-specific sites. The FAIRE-STARR library was generated by Alisa Fuchs and Stefanie Schöne. Data analysis was performed by Melissa Bothe.

### 3.6.3. Gene Regulation Coincides with Differential Activity at Shared Peaks

The STARR-seq analysis revealed, that AR and GR both show comparable activation of enhancers that are located at shared genomic sites. Nevertheless, shared binding can also drive receptor-specific and differential gene regulation, as seen for *GILZ* and *AQP3*. Both genes have shared binding sites in their loci but show hormone-inducible regulation by only one of the two receptors.

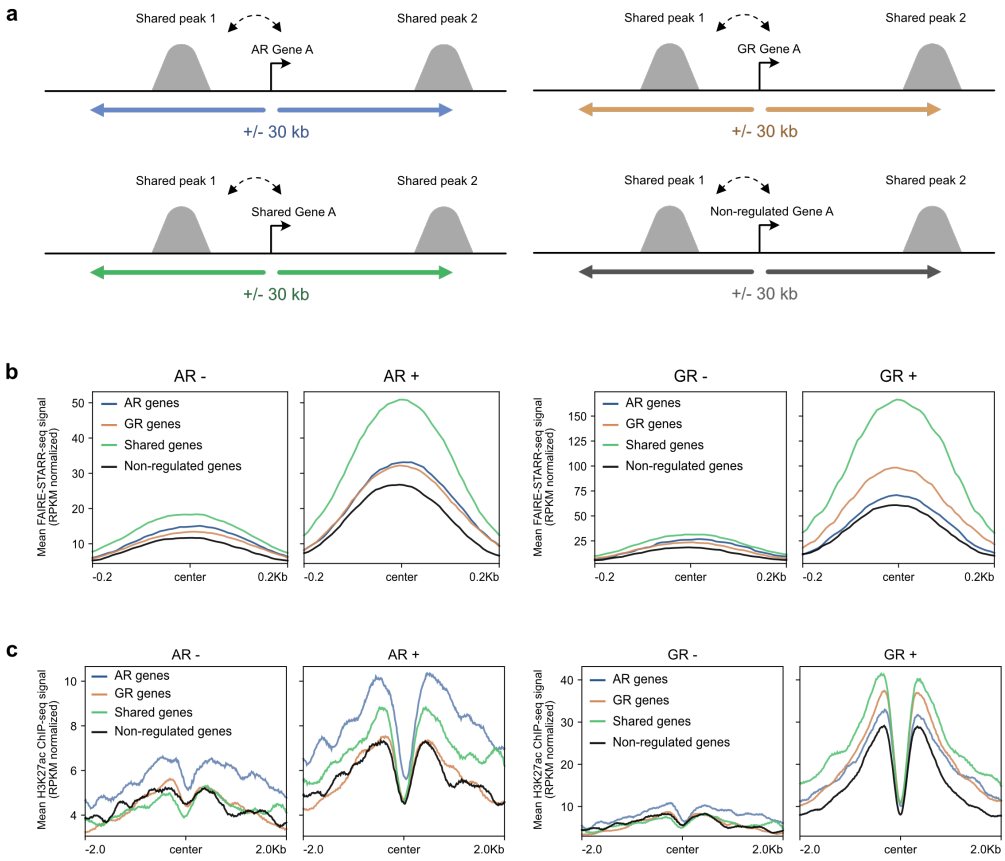



**Figure 3.29. Enhancer activity correlates with receptor-specific gene regulation. (a)** Cartoon showing how shared STARR-seq peaks were intersected with receptor specific gene expression. The TSS of each gene of each gene category (AR, GR, shared, non-regulated genes) was scanned for shared STARR peaks at a distance of +/- 30 kb in both cell lines. **(b)** Profile plot representing mean FAIRE-STARR signal (RPKM normalized) at shared sites (+/- 250 bp around peak summit) in the vicinity of each gene category. AR-U2OS cells were treated with vehicle (AR -) or R1881 (AR +) and GR-U2OS cells were treated with vehicle (GR -) or Dex (GR +) overnight. **(c)** Profile plot representing mean H3K27ac signal (RPKM normalized) at shared sites (+/- 2 kb around peak summit) in the vicinity of each gene category. Data analysis was done by Melissa Bothe.

To address the question if receptor-specific regulation correlates with enhancer activity, we linked gene-expression to the enhancer activity of shared binding sites. The data analysis was performed by Melissa Bothe. In brief, each gene of a differential gene category (AR-specific, GR-specific, shared and non-regulated) was assigned to a shared STARR-seq peak in a region of +/- 30 kb (Fig. 3.29 a). When two STARR-seq peaks were found in this region only the closest peak was assigned to the gene.

The analysis revealed that AR shows the strongest enhancer activity near shared genes followed by AR-specific and GR-specific genes (Fig. 3.29 b). Non-regulated genes, however, show the weakest enhancer activity. Similarly, GR shows the strongest enhancer activity near shared genes followed by GR-specific, AR-specific genes and non-regulated genes. This data indicates that receptor-specific enhancer activity of the shared binding sites indeed correlates with receptor-specific gene regulation.

H3K27ac is known to strongly correlate with enhancer activity in a cell type-specific manner. In order to see if differential gene regulation at the shared sides is reflected in the H3K27ac signal, we plotted the H3K27ac signal for the same shared sites that were assigned to each gene category as described above (Fig. 3.29 c). In this case, AR showed the strongest H3K27ac signal at AR-specific binding sites followed by shared sites. GR-specific and non-regulated genes, however, showed the weakest H3K27ac induction. Similarly, GR showed the strongest signal near shared and GR-specific genes, whereas the signal was weaker at AR-specific and non-regulated sites.
Consistently with the analysis of shared STARR-seq peaks, we could observe a correlation of receptor-specific gene regulation and the H3K27ac signal at shared binding sites. Together, these results suggest that receptor-specific enhancer activity at shared binding sites contributes to the receptor-specific regulation of nearby genes.

### 3.6.4. Dissecting the Enhancer Landscapes of *GILZ* and *AQP3*

As described above, I found that *GILZ* is regulated by GR only but, interestingly, it is equally bound by GR and AR. Thus, in this case binding alone cannot explain the differential regulation of *GILZ,* so there has to be a mechanism downstream of binding that drives *GILZ* expression. To get a better understanding of the *GILZ* regulation, I looked into the enhancer landscape of this gene. The genome browser screenshot shows ChIP-seq data of GR and AR at the *GILZ* locus as well as the STARR-seq signal representing the enhancer activity (Fig. 3.30). The STARR-seq tracks are shown for both cell lines with or without hormone treatment. The binding peaks of GR and AR are highlighted in grey. The hormone inducible and therefore receptor-specific STARR-seq peaks are shown in blue. Remarkably, the same enhancers can be found in both cell lines and are induced upon hormone treatment for both receptors leading to the conclusion that differential enhancer usage is not responsible for the specific regulation of *GILZ* by GR. However, although the inducible enhancers are shared the overall enhancer activation is stronger upon GR signaling.

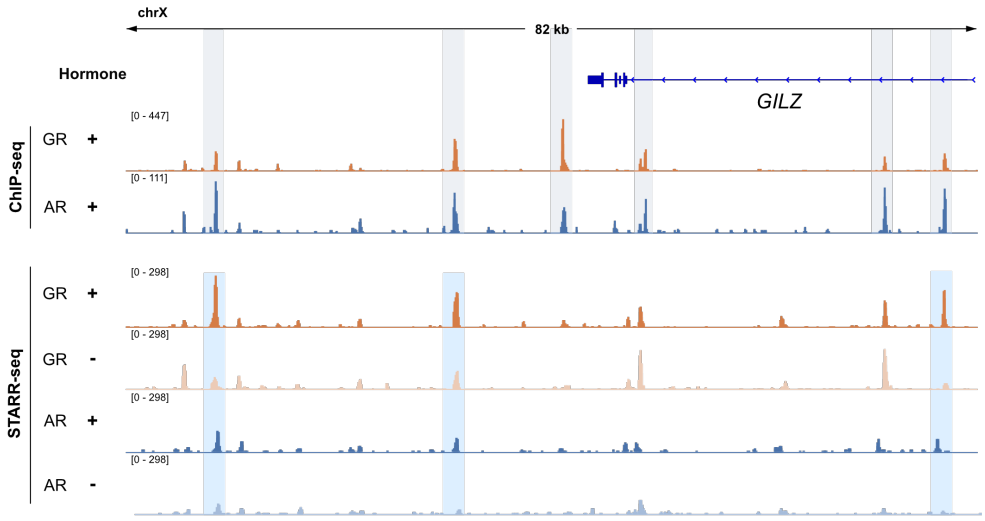



**Figure 3.30. Enhancer landscape of the *GILZ* locus**. Genome-browser screenshot of the *GILZ* locus showing the *GILZ* gene, ChIP-seq and STARR-seq tracks of GR and AR. The ChIP peaks bound by both receptors are highlighted in grey. One representative ChIP track is shown from two independent replicates. The STARR-seq tracks represent enhancer activity in U2OS-AR cells treated with dmso as vehicle control or 5 nM R1881 and U2OS-GR cells treated with ethanol as vehicle control or 1 μM Dex overnight. Interferon inhibitors were added in both cell lines. The hormone-inducible STARR-seq enhancers are highlighted in blue. STARR-seq tracks represent the merge from three biological replicates.

Similarly, as seen for GILZ, the *AQP3* locus is bound by both receptors but only AR regulates *AQP3* expression. The two binding peaks of AR and GR in the *AQP3* locus are upstream of the gene. Here, a genome-browser screenshot is shown for the *AQP3* locus showing the ChIP-seq tracks of both receptors as well as the STARR-seq data (Fig. 3.31 a). The more distal peak from the TSS shows hormone-inducible enhancer activity for both receptors. Interestingly, the more proximal peak is only induced by AR activation. Thus, this selective enhancer usage might in fact be responsible for the receptor-specific regulation of *AQP3* by AR as only AR is able to activate this enhancer.

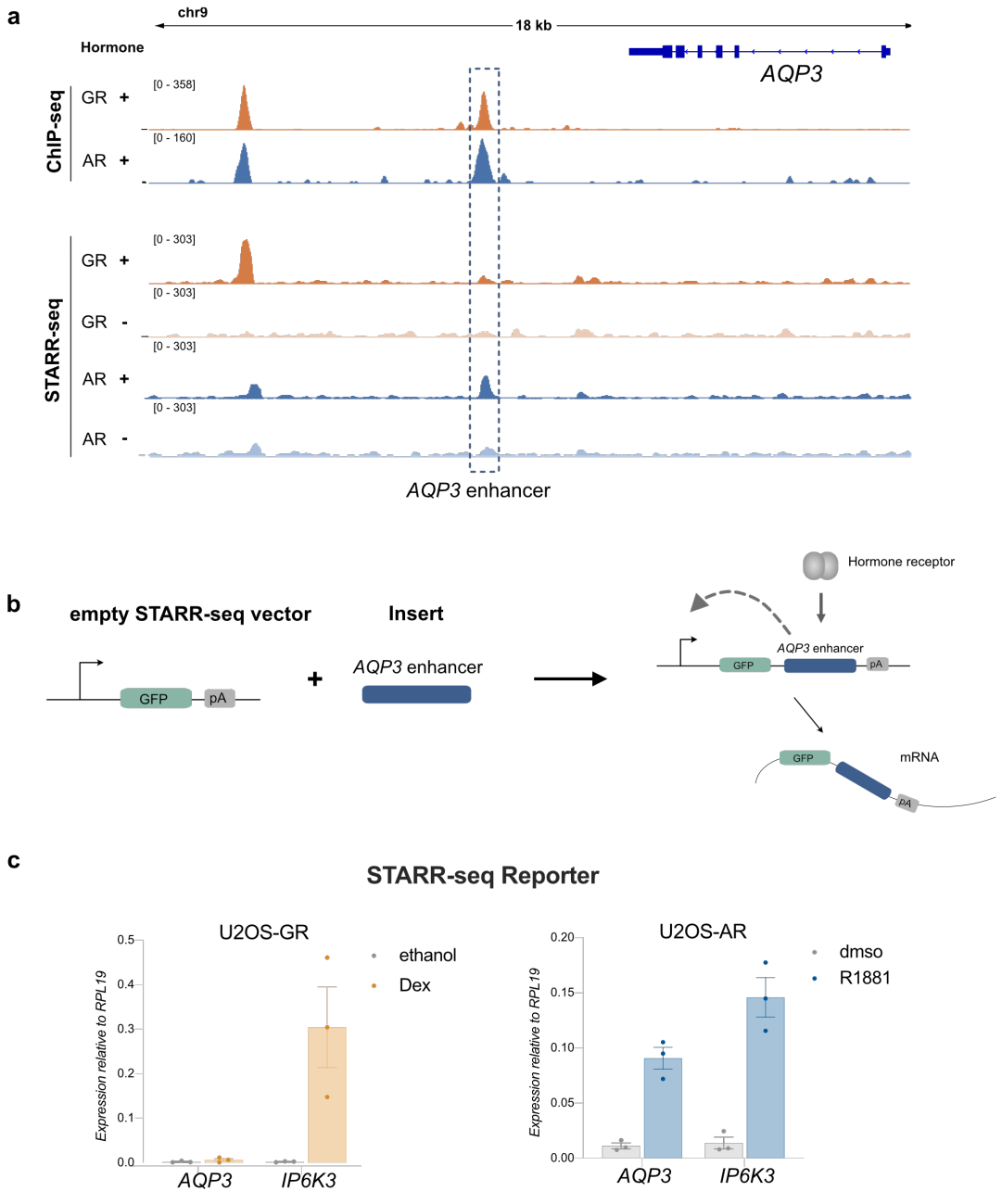



**Figure 3.31. Enhancer landscape of the *AQP3* locus shows an AR-specific enhancer.** **(a)** Genome-browser screenshot of the *AQP3* locus showing the *AQP3* RefSeq gene, ChIP-seq and STARR-seq tracks of GR and AR. One representative ChIP track from two independent replicates is shown. The STARR-seq tracks represent enhancer activity in U2OS-AR cells treated with dmso as vehicle control or 5 nM R1881 and U2OS-GR cells treated with ethanol as vehicle control or 1 μM Dex overnight. Interferon inhibitors were added in both cell lines. The *AQP3* STARR-seq enhancer selectively regulated by AR is shown in the dashed box. STARR-seq tracks represent the merge from three biological replicates. **(b)** *AQP3* enhancer DNA fragment was inserted into the STARR-seq vector and can be tested for its enhancer activity in the cell. **(c)** GR and AR cells were transfected with one individual STARR-seq reporter containing the *AQP3* enhancer sequence or the inducible enhancer *IP6K3*. Relative expression of the constructs was measured by qPCR in U2OS-AR cells treated with dmso as vehicle control or 5 nM R1881 and U2OS-GR cells treated with ethanol as vehicle control or 1 μM Dex overnight. Interferon inhibitors were added in both cell lines. The average gene expression of three biological replicates ±SEM is shown.

To confirm that the *AQP3* enhancer is indeed specifically activated by AR, the sequence was cloned into an empty STARR-seq vector (Fig. 3.31 b). The construct was then transfected into AR as well as GR cells. The cells were subsequently treated with their specific hormone or a vehicle control. The enhancer output was measured by qPCR. Confirming what we observed in our STARR-seq screen, GR was not able to activate the *AQP3*-STARR-seq reporter even when the *AQP3* enhancer was taken out of its chromatin context (Fig. 3.31 c). In contrast, AR was able to activate the *AQP3* reporter thereby confirming that the enhancer is receptor-specific. Interestingly, although the ChIP-seq analysis has shown that GR is able to bind the *AQP3* enhancer region, GR is not able activate transcription from this enhancer.

**Table 3.1 AR and GR motifs in the *AQP3* enhancer**

| | Matrix ID | Name | Score | Rel. Score | Position | Strand | Predicted Sequence |
|---|---|---|---|---|---|---|---|
| 1 | MA0007.2 | AR | 11.548 | 0.897677752161 | 206 - 220 | + | AGGAACATGC**AGA**GA |
| 2 | MA0007.2 | AR | 11.2647 | 0.893866602854 | 258 - 272 | + | AGGTACACAA**AGA**AA |
| 3 | MA0007.2 | AR | 10.8739 | 0.888609954453 | 182 - 196 | + | AAGTACACTG**AGA**GA |
| 4 | MA0113.2 | GR | 9.57914 | 0.867795861151 | 174 - 188 | + | TGTACTCAA**AGT**ACA |
| 5 | MA0113.2 | GR | 8.68432 | 0.856656193582 | 238 - 252 | + | AGTCCAGGA**TGT**TCA |
| 6 | MA0113.2 | GR | 7.02742 | 0.836029346487 | 207 - 221 | - | ATCTCTGCA**TGT**TCC |
| 7 | MA0113.1 | GR | 10.0465 | 0.830057193494 | 236 - 253 | - | GTGAACATCC**TGG**ACTGT |
| 8 | MA0007.2 | AR | 5.99922 | 0.823036069688 | 161 - 175 | + | GAGTACATGG**AGG**TG |
| 9 | MA0007.2 | AR | 5.39579 | 0.814918782213 | 237 - 251 | + | CAGTCCAGGA**TGT**TC |
| 10 | MA0007.2 | AR | 4.74408 | 0.806152061195 | 239 - 253 | - | GTGAACATCC**TGG**AC |

To understand in more detail what makes the *AQP3* enhancer AR-specific, I scanned the enhancer sequence for AR and GR motifs from the JASPAR CORE database (Sandelin et al. 2004). The analysis revealed six AR and four potential GR motifs (Tab. 3.1). When looking closer at the motifs, it is striking that the top three AR motif matches contain a 'AGA' sequence in the second receptor binding half-site. This sequence diverges from the consensus motif for both AR and GR. Our ChIP-seq data suggest that GR is able to bind the *AQP3* enhancer region. Because of the low resolution of several hundred base pairs it is not possible to determine which exact motifs are bound by GR. Nevertheless, I wanted to find out if the top three motifs containing the 'AGA' sequence are preventing GR from activating the AQP3 enhancer. Therefore, I first disrupted the top three AR binding motif matches by introducing point mutations and inserted the mutated enhancer into an empty STARR-seq vector (Fig. 3.32 a). By introducing mutations, I was able to determine if the top three motifs are responsible for the activation of the enhancer by the androgen receptor. Second, I replaced the 'AGA' sequence by 'TGT' in the second half site of all the three motifs and thereby converted them to a more consensus-like binding motif. Then I inserted the transformed *AQP3* enhancer into an empty STARR-seq vector. In this way I was able to test if GR

can activate the *AQP3* enhancer when the top binding motifs are more consensus-like and if sequence specificity is responsible for differential *AQP3* enhancer activity of the two receptors.

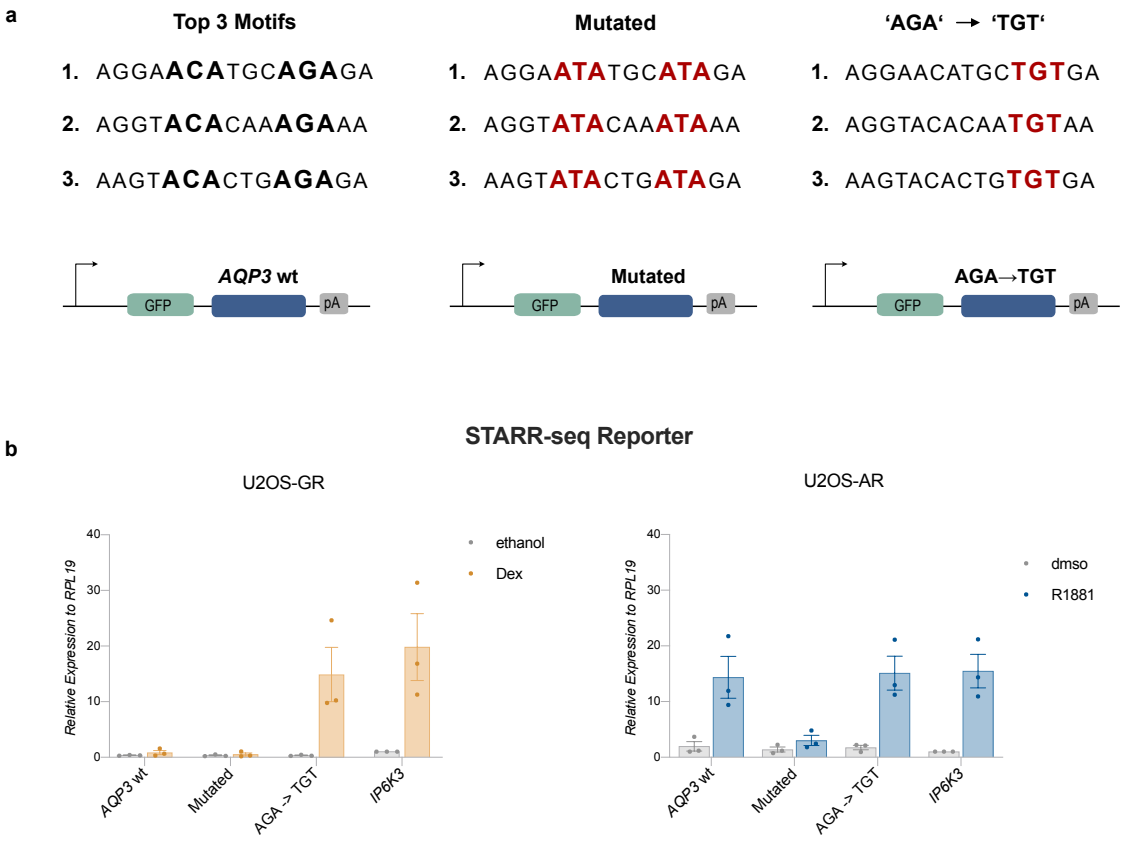



**Figure 3.32 GR is able to activate the AQP3 enhancer containing a 'TGT'. (a)** The top 3 AR motif-matches of the *AQP3* enhancer were mutated to disrupt the motif (Mutated) or converted from 'AGA' to a 'TGT' sequence (AGA -> TGT) **(b)** GR and AR cells were transfected with one individual STARR-seq reporter containing the *AQP3* enhancer sequence, the mutated *AQP3* enhancer, the converted *AQP3* enhancer or the inducible enhancer *IP6K3*. Relative expression of the constructs was measured by qPCR in U2OS-AR cells treated with dmso as vehicle control or 5 nM R1881 and U2OS-GR cells treated with ethanol as vehicle control or 1 µM Dex overnight. Interferon inhibitors were added in both cell lines. The average gene expression of three biological replicates ±SEM is shown.

I transfected both cell lines with the *AQP3* wildtype enhancer, the mutated *AQP3* enhancer with the disrupted binding motifs, the transformed *AQP3* enhancer containing a 'TGT' in the second half site of the motif and an inducible enhancer as a positive control (Fig. 3.32 b). As I have shown before, GR not able to activate the wild type *AQP3* enhancer upon hormone induction. The same is true for the mutated *AQP3* enhancer. Interestingly, GR is able to robustly activate the transformed *AQP3* enhancer containing a 'TGT' sequence. On the other hand, AR is able to activate the wild type *AQP3* enhancer. Upon mutation of the three "AGA" motifs, AR loses the ability to activate it. This indicates that the identified top three binding sites are indeed responsible

for the selective activation by AR. After transforming these motifs into more consensus-like motifs AR is still able to activate this STARR-seq reporter.
First, this finding is indicating that the sequence composition of *AQP3* is responsible for the differential enhancer activation by the two receptors. And second, that in contrast to GR, AR is able to activate enhancers containing a direct repeat-like motif.

## 3.7. Motif Enrichment Analysis

Previous studies have shown that different DNA binding preferences of related transcription factors are responsible for differential binding resulting in differential gene expression (Shen et al. 2018). To better understand the impact of motif preference of AR and GR, an analysis of motif enrichment (AME) (McLeay and Bailey 2010) of the ChIP-seq data of the two receptors was performed by Melissa Bothe. We performed a search using the clustered JASPAR core vertebrate motif collection which is an open-source database for non-redundant eukaryotic transcription factor binding profiles (Khan et al. 2018). The profiles were derived from DNA binding sites that were shown to be bound by TFs experimentally. In addition, we added a direct repeat consensus motif (DR3). The classical consensus motif comprises a palindromic hexamer separated by three nucleotides. The direct repeat motif, however, represents a direct repeat of the classical hexamer separated by three nucleotides and was included in the analysis for two reasons. First, the direct repeat consensus motif has been previously described as being AR-specific and second, I found that AR is able to specifically activate the *AQP3* enhancer which required direct-repeat like motifs (Fig. 3.32).

A heatmap was generated showing the top enriched motifs in both cell lines (Fig. 3.33 a). As expected, given their similar DNA sequence preference, several of the motifs are shared between both receptors. Interestingly, the first and most enriched motif at AR-specific peaks is AP-1. At GR-specific binding sites, however, AP-1 shows no enrichment. Notably, as expected the DR3 motif shows a stronger enrichment in AR compared to GR. On the other hand, the classical AR/GR consensus motif on the other hand is stronger enriched at GR-specific peaks.
The top 7 motif hits are also represented with their logos (Fig. 3.33 b). I have previously shown, that in contrast to AR, many GR-specific sites are located in closed chromatin. The condensation state of the chromatin correlates with sequence and consequently also does the motif composition (J. Wang et al. 2012). Thus, to remove the potential bias that might result from differential condensation states of the chromatin an additional motif analysis was performed for GR using receptor-specific peaks located in accessible chromatin. The additional analysis showed similar

motif hits for GR. One exception was that AP-1 was enriched in open sites bound by GR. The reason we find AP-1 being enriched at GR bound sites only in open chromatin might be that AP-1 is a pioneer factor, which is able to open chromatin and maintain the opening and is therefore associated with open chromatin (Biddie et al. 2011).

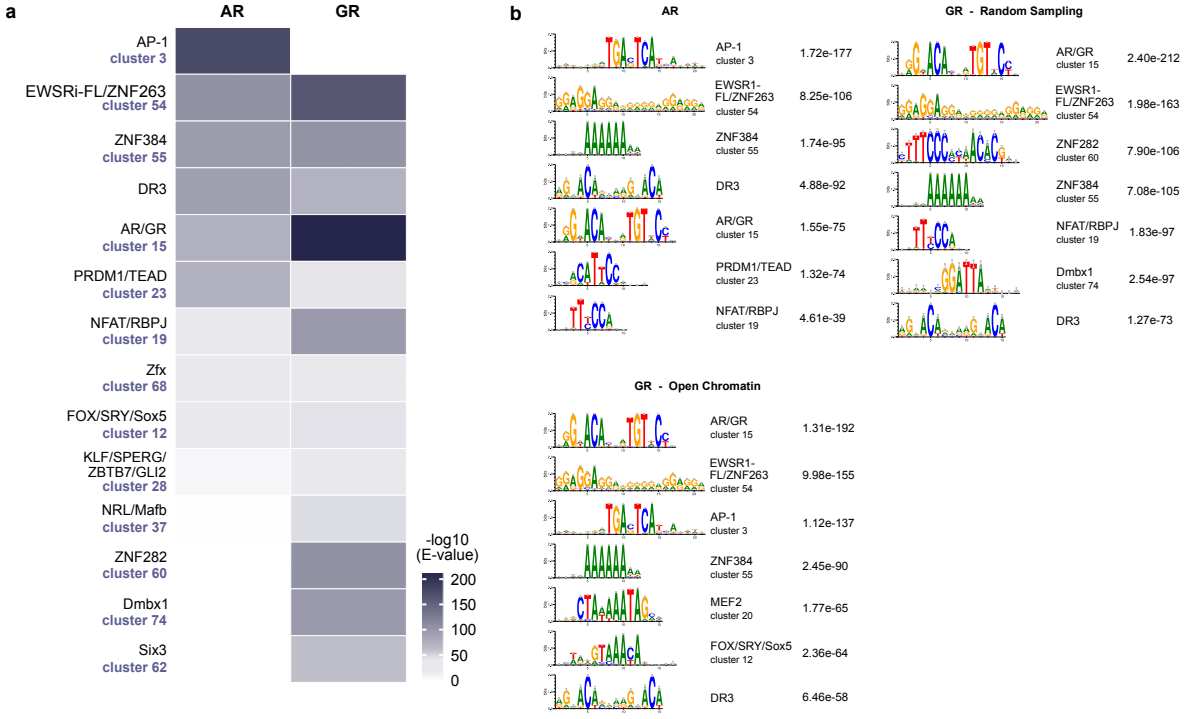



**Figure 3.33. Comparing motif enrichment at AR and GR binding sites. (a)** The heatmap represents motif clusters at all AR and GR-specific peaks (+/- 250 pb around peak summit). Motifs are shown with an E value $<10^{-30}$. Shuffled input sequences represent the background for the motif analysis. **(b)** Top 7 enriched differential motif clusters and corresponding E value at all AR-specific peaks, GR-specific peaks randomly sampled or GR-specific peaks at genomic sites with a high chromatin accessibility (+/- 250 pb around the peak summit). Shuffled input sequences represent the background for the motif analysis. Data analysis was done by Melissa Bothe.

It was reported in a previous study that sites directly flanking the consensus core motif are slightly enriched for poly(A) nucleotides in AR compared to GR (L. Zhang et al. 2018). We therefore performed a screening of AR-specific and GR-specific binding sites identified by ChIP-seq for their GC content.

The data analysis was performed in cooperation with Melissa Bothe. In this analysis, the following genomic binding sites were scanned for their GC content: AR-specific, GR-specific as well as AR- and GR-specific sites located in open chromatin (Fig. 3.34). We individually scanned open chromatin sites because nucleosome-depleted sites are known to be GC-rich whereas nucleosome-bound sites are more enriched for AT nucleotides (Jabbari, Chakraborty, and Wiehe 2019). Strikingly, GR shows a stronger signal for GC-rich DNA compared to AR. This difference

becomes more prominent when looking at the GR-specific sites in open chromatin as the GC content of GR-specific regions is even higher in accessible chromatin. AR, however, binds fragments with a lower GC content compared to GR. Interestingly, the GC-content of the GR-specific fragments is higher in GR cells not only at the core motif and regions directly flanking the core but also throughout a large region of at least 5 kb.

This finding indicates that that the nucleotide composition of the DNA has an impact on recruiting AR and GR to their specific binding sites with GR showing a preference for GC-rich regions.

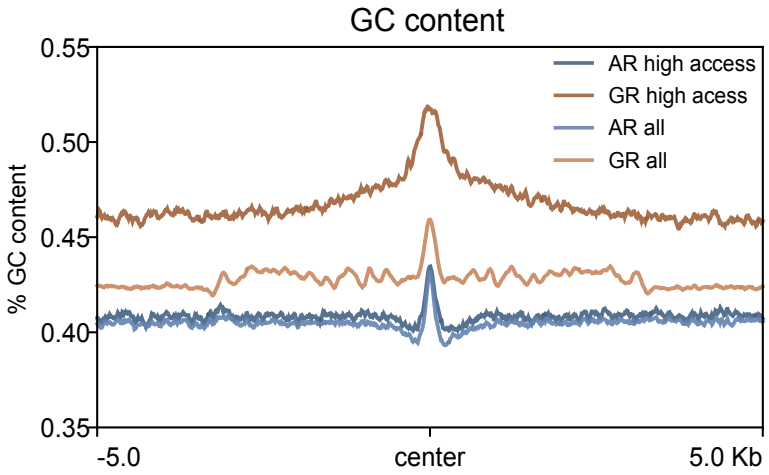



**Figure 3.34. GC content at AR-specific and GR-specific binding sites.** The profile plot shows the mean GC content at all AR-specific peaks (AR all), all GR-specific peaks (GR all) and for peaks in genomic regions of accessible chromatin (AR high access or GR high access). Data analysis was done by Melissa Bothe.

## 3.8. Comparing Differential Cofactor Interactions of GR and AR

### 3.8.1. RIME Analysis

Biological processes require coordinated actions of individual proteins to achieve specific functions. Therefore, the identification of interacting proteins is crucial for a detailed understanding of biological processes such as transcriptional regulation. In the past, many cofactors have been identified to specifically interact with GR and AR. Interactions with different cofactors are highly cell type-specific and depend on the variety of cofactors expressed in the cell (Goi, Little, and Xie 2013). However, in our model system I compared AR and GR in the same cellular background, therefore differential regulation likely cannot be explained by the availability of cofactors. Many genes differentially regulated by GR and AR in U2OS cells show an overlapping binding profile for the two receptors. Therefore, mechanisms downstream of binding seem to play a role in driving receptor-specific gene regulation. The differential recruitment of cofactors by AR

and GR might thus provide an explanation for receptor-specific gene activation driven by events downstream of binding.

Rapid immunoprecipitation mass spectrometry of endogenous protein (RIME) (Mohammed et al. 2016) is a method to endogenously identify interacting proteins. This method was performed by Stefan Prekovic and Isabel Mayayo Peralta from the group of Wilbert Zwart at the NKI in the Netherlands to specifically detect interacting proteins of both AR in GR in U2OS cells.
In brief, both cell lines were treated with their specific hormone followed by formaldehyde cross-linking. In this way, the chromatin bound receptor is fixed and protein complexes are stabilized. After a sonication step, the protein complexes were immunoprecipitated with antibodies either specific for AR or GR. Next, the protein complexes were stringently washed and digested into peptides followed by a liquid chromatography with tandem mass spectrometry (LC-MS-MS) analysis. The untargeted label-free quantification (LFQ) method was used to detect the relative quantity of proteins in four biological replicates (Bantscheff et al. 2007). The relative amount of proteins is indicated as LFQ value

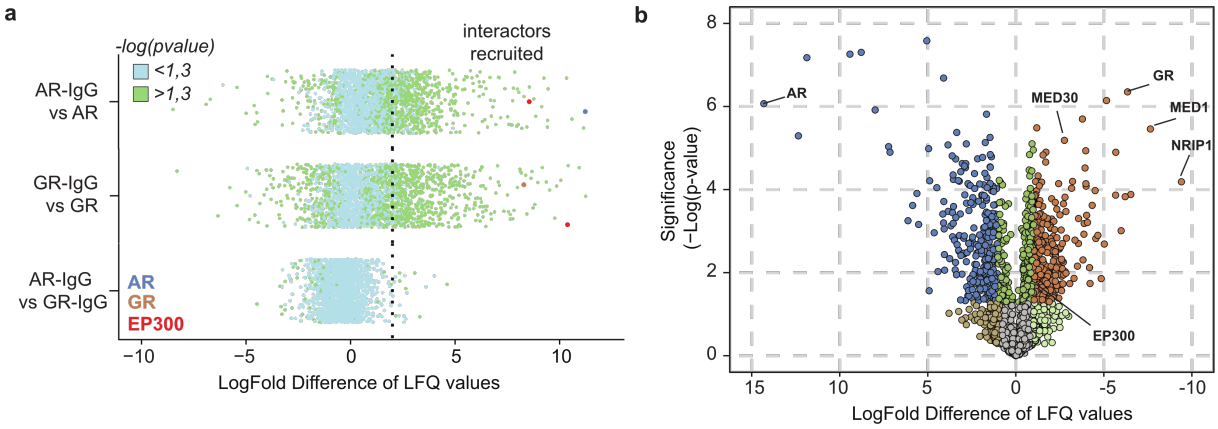



**Figure 3.35. Specific interactors detected by RIME for AR and GR. (a)** Scatterplot showing an enrichment over IgG for AR and GR. Proteins are considered to interact when they are ≥ 2 LFQ enriched over IgG (dotted line) and are regarded significant when-log(padj)>1,3 (green). n=4. **(b)** The cofactors differentially enriched for AR (blue) and GR (orange) are represented in the volcano plot (n=4). The RIME experiments as well as data analysis were performed by Stefan Prekovic and Isabel Mayayo Peralta.

Initial quality control of the data analysis showed that both AR and GR show recruitment of specific interactors that are enriched over the IgG control (Fig 3.35 a). The proteins AR, GR and EP300 are highlighted. As expected, AR and its known interaction partner EP300 (Fu et al. 2000, 2003) are found among recruited proteins in the AR cells similarly GR are and EP300 are present

in GR cells. Differential analysis of recruited proteins by AR and GR revealed that both receptors have shared interactors (e.g. EP300) but also show a significant recruitment of receptor-specific proteins (Fig 3.35 b) with AR-specific interactors on the left and GR-specific cofactors on the right. These initial results show, that with the help of RIME we were able to specifically enrich for interacting proteins of AR and GR. Further, these datasets can provide us with potential candidates that are differentially recruited by AR and GR and therefore could play a role receptor-specific gene regulation.

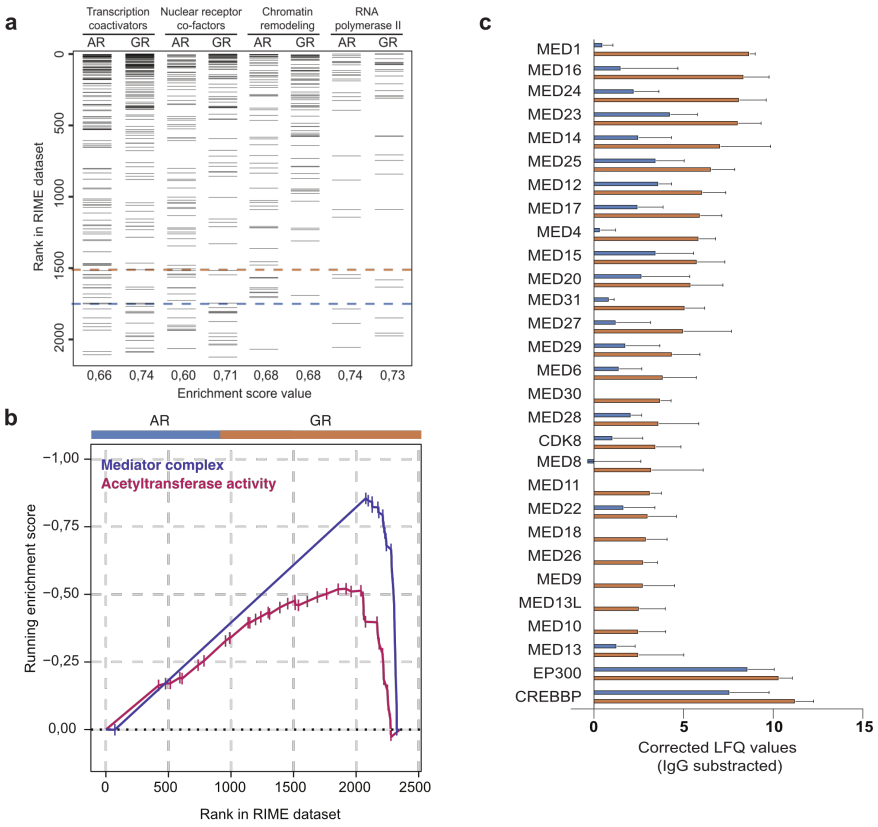



**Figure 3.36. GR specifically interacts with the mediator complex and histone acetyltransferases. (a)** GSEA (Geneset Enrichment Analysis) for transcription coactivator activity (M19071), nuclear receptor binding (M26456), chromatin remodeling (M19139), and RNA polymerase transcription factor binding (M19139) in the AR and GR RIME data. **(b)** Geneset enrichment of Acetyltransferase activity (M18150) and mediator complex (M17759) in the AR and GR RIME experiments. **(c)** LFQ enrichment over the IgG control for various mediator subunits and EP300 and CREBBP in the AR and GR RIME datasets. The RIME experiments were performed by Stefan Prekovic and Isabel Mayayo Peralta.

As a sanity check for the RIME data quality, an enrichment analysis was performed for the genesets transcription coactivators, nuclear receptor cofactors, chromatin remodeling and RNA polymerase II (Fig. 3.36 a). As expected, both AR and GR show an enrichment of coregulators in all four groups. However, a further geneset enrichment analysis for GR revealed an enrichment of the mediator complex as well as a geneset containing proteins involved in histone acetylation. A direct

comparison of the enrichment of these two protein categories between AR and GR has shown that both Mediator and histone acetyltransferases are more enriched in GR compared to AR (Fig. 3.36 b). We then individually plotted the data for various members of the mediator complex and, interestingly, most of them are not enriched for AR or to a lesser extend compared to GR (Fig. 3.36 c). In contrast, the two acetyltransferases EP300 and CREBBP, also known as CBP or CREB (cyclic AMP response element binding protein binding protein), are recruited similarly by both receptors AR and GR. Previously, I have demonstrated that GR has a stronger potential to induce H3K27ac compared to AR (Fig.3.23). EP300 and CREBBP are both histone acetyltransferases that can deposit the H3K27ac mark (Ogryzko et al. 1996). Therefore, differential recruitment of these factors could provide an explanation for the dramatic difference in H3K27ac signal between AR and GR cells. However, given the comparable recruitment of EP300 and CREBBP it seems unlikely that differential recruitment of these proteins by AR and GR explains the dramatic difference in induced H3K27 acetylation upon hormone treatment for these receptors.

### 3.8.2. Recruitment of MED1 and EP300 to the Chromatin

The mediator complex is important for transcriptional regulation and is able to connect transcription factors bound to regulatory regions to gene promotors which is achieved by looping of the DNA (Kagey et al. 2010). This in turn leads to a recruitment of the preinitiation complex and the Pol II followed by transcriptional activation (Soutourina 2018). The RIME experiments showed that GR is selectively interacting with the mediator complex whereas AR shows no or only little interactions.

To validate the finding, that GR and AR show striking differences in interacting with multiple mediator subunits, I performed ChIP-qPCR analyses using antibodies specific for Med1, one of the mediator subunits, in both cell lines (Fig. 3.37 a). In this way, I could test whether Med1 recruitment to genomic sites bound by AR and GR is differential in the two cell lines. The genomic regions I have chosen are occupied by both AR and GR and near genes regulated by one receptor or by both AR and GR. *ZBTB16, GILZ, IGFBP1* and *USP18* are genes regulated by GR. *FKBP5* is regulated by both receptors and *ABLIM3* is specifically regulated by AR. The enrichment of Med1 was measured by qPCR with primers targeting several regions near genes that are bound by both AR and GR.

Consistent with the results from the RIME experiments, I have found a strong and hormone inducible enrichment of Med1 at various receptor binding sites in the GR cells. In contrast, AR cells showed no enrichment of Med1 at most sites except maybe for low MED1 levels at the loci *FKBP5* and *ABLIM3*.

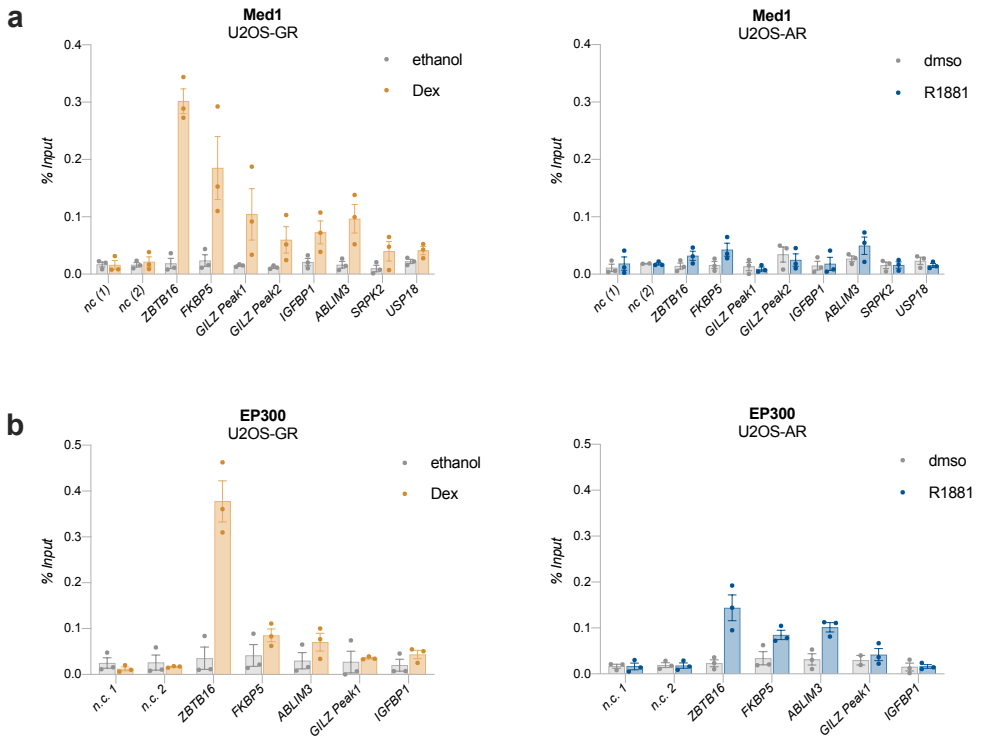



**Figure 3.37. GR specifically recruits with Med1 to GR-bound regions**. ChIP-qPCR analysis for Med1 and EP300 within genomic regions bound by both AR and GR. **(a)** Med1 ChIP-qPCR analysis in GR cells treated with ethanol as vehicle control or Dex for 1,5 h. and AR cells treated with dmso or 5 nM R1881 for 4 h. The error bars indicate ±SEM (n=3) **(b)** EP300 ChIP-qPCR analysis in GR cells treated with ethanol as vehicle control or Dex for 1,5 h. and AR cells treated with dmso or 5 nM R1881 for 4 h. Negative control regions (n.c.) are neither occupied by AR or GR. The error bars indicate ±SEM (n=3).

Histone acetylation of specific lysine residues in the tails of histones, in particular H3 K27, by acetyltransferases is associated with transcriptional activation. Acetylation of lysine residues on histone proteins leads to a decompaction of condensed chromatin (G. Li and Reinberg 2011; Robinson et al. 2008) and makes it accessible for the transcription machinery. It has been shown in mammals, that inhibition of EP300 together with CREBBP results in the global loss of H3K27ac (Visel et al. 2009; Weinert et al. 2018). Thus, differential interaction of GR and AR with EP300 and CREBBP might explain receptor-specific gene activation as well the stronger H3K27ac signal induced by GR compared to AR in U2OS cells.

To test this hypothesis, I performed ChIP-qPCR analyses using antibodies specific for EP300 in both cell lines and tested whether EP300 is specifically recruited to genomic sites by GR or AR (Fig. 3.37 b).

EP300 was clearly enriched upon hormone induction at several GR binding sites in U2OS-GR. The strongest enrichment was observed in the *ZBTB16* locus. In U2OS-AR, however, an

enrichment of EP300 could be detected at the *ZBTB16*, *FKBP5* and *ABLIM3* locus after hormone induction. Consistent with the results from RIME, GR and AR are both able to specifically recruit EP300 to their specific receptor-binding sites. Yet, H3K27ac levels increase for GR and not for AR. This finding indicates, that perhaps histone acetyltransferases other than EP300 are differentially recruited by GR and play a role in inducing the increase in H3K27ac upon GR signaling.

# 4. Discussion

## 4.1 Binding as a Driver of Receptor-Specific Gene Expression

AR and GR are very close members of the steroid receptor family of TFs but have divergent biological functions. The androgen receptor is, for example, crucial for the development of sex organs whereas the glucocorticoid receptor is important for the immune response of the body as well as the glucose metabolism. These divergent roles are also reflected in their clinical use. Androgens are applied for the treatment of muscle waste or osteoporosis (Vanderschueren et al. 2014) whereas glucocorticoids are widely used to treat various inflammatory or allergic diseases.
Despite their different physiological functions, both receptors share strong structural similarities. Most strikingly, AR and GR share a conserved DBD and the amino acids forming direct contacts with the DNA are identical (Beato et al. 1989). Consistently, *in vitro* experiments have shown that both receptors recognize the same consensus motif consisting of two palindromic hexamers separated by a 3 bp spacer (Mangelsdorf et al. 1995). Despite these similarities, previous genome-wide binding studies in prostate cancer cells have revealed, that both AR and GR have not only overlapping but also receptor-specific binding sites (Arora et al. 2013; Sahu et al. 2013). Moreover, it was previously shown that paralogous TF exhibit subtle differences in genomic binding (Shen et al. 2018). Consistent with these studies, when expressed in U2OS cells, both AR and GR show an overlapping set of binding sites as well as GR- and AR-specific sites (Fig. 3.9).

To better understand if receptor-specific binding is indeed contributing to receptor-specific gene regulation, ChIP-seq data was intersected with the genes upregulated by either AR, GR or both receptors (Fig. 3.10). The focus was set on upregulated genes only for several reasons. First, very few genes were repressed upon AR induction, therefore, a meaningful bioinformatical analysis was impossible. Second, gene repression by GR is important for the inhibition of inflammatory pathways, because GR interferes with the expression of cytokines and chemokines driven by the factor NFKappa B (de Bosscher, vanden Berghe, and Haegeman 2006). This mechanism, called transrepression, does not necessarily require direct receptor binding to DNA but might be a result of protein-protein interactions (Pascual and Glass 2006). Third, attempts to identify repressive GR binding sites did not have much success in the past, because repressed genes were depleted for GR binding sites (Langlais et al. 2012; Reddy et al. 2009). And lastly, gene repression by GR is possibly a consequence of cofactor squelching. Thereby, different cofactors are recruited to the receptor

binding sites and in turn those factors are absent in other loci leading to a reduced gene expression (Schmidt et al. 2016).

As expected, AR and GR-specific binding sites were enriched near AR and GR-regulated genes, respectively. This indicates, that genomic binding, at least in part, contributes to the functional specificity of steroid receptors. For example, the gene *HR* contains a binding peak of AR in the intronic region and is specifically regulated by AR. The gene *CA9*, on the other hand, has a promoter-proximal GR peak and is regulated by GR only. These examples serve to illustrate that receptor-specific peaks are enriched near receptor-specific target genes. Moreover, the examples of *HR* and *CA9* indicate that receptor-specific binding is likely driving the regulation of these receptor-specific target genes.
Furthermore, shared binding likely explains the regulation of a shared set of genes by both receptors as genes regulated by AR and GR are enriched for shared binding sites compared to non-regulated genes (Fig. 3.10). This is for example true for the gene *SLC6A3*, which is regulated by both AR and GR and contains a shared binding peak of both receptors (Fig 3.11.).

Strikingly, this study shows, that shared binding sites are also able to direct differential gene regulation by AR and GR. A case in point is *GILZ*, a known target gene of GR. It contains several GR binding sites downstream of its promoter and within intronic regions (Fig. 3.14). Interestingly, AR also occupies each of these binding sites upon hormone induction. Yet, in contrast to GR, AR is not able to drive the expression of *GILZ* despite their identical binding patterns. In a previous study it has been shown that the GR binding peaks upstream of the promoter, as well as the promoter proximal peaks, are crucial for the regulation by GR (Thormann et al. 2018). Upon deletion of these receptor binding sites, GR loses the ability to upregulate *GILZ* expression (Fig 3.15) demonstrating that these binding sites, that are occupied by both AR and GR, are indeed responsible for the GR-dependent regulation observed. *AQP3*, on the other hand, is a gene specifically regulated by AR, however, binding of AR and GR is identical in this locus (Fig. 3.16). Similarly, to what I observed for *GILZ*, when deleting the receptor binding sites in the *AQP3* locus AR is no longer able to drive gene expression of this gene (3.17). This result demonstrates that these binding sites contribute to receptor-specific regulation of *AQP3* by AR only.

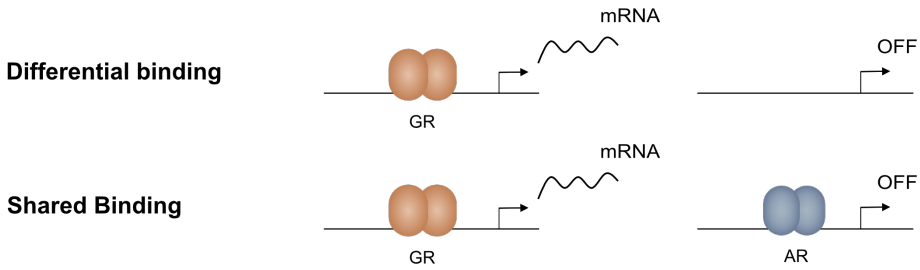



**Figure 4.1. Model explaining receptor-specific gene regulation by binding.** Differential gene regulation by AR and GR can be directed by differential binding but also by shared binding of the two receptors. Downstream mechanisms likely play a role in driving selective regulation from sites occupied by both receptors.

Together, these findings suggest that differential binding is indeed an important driver of differential gene expression by AR and GR, which is in line with many other studies indicating that diversification of genomic binding of related TFs is critical to fulfill their unique biological functions (Datta et al. 2018; Jerković et al. 2017; Kribelbauer et al. 2019; Sahu et al. 2013; L. N. Singh and Hannenhalli 2008b). However, it is still mostly unclear how AR and GR can differentially regulate genes when they bind to the same sites. One possibility is that the receptors differentially interact with the surrounding chromatin and that one receptor is able to stabilize enhancer-promoter contacts whereas the other is not. Moreover, the differential ability of AR and GR to activate genomic enhancers, for example as shown in this study for the *AQP3* enhancer, might lead to differential transcriptional outputs. Another explanation is differential interaction with transcription factors and transcriptional coactivators that ultimately lead to Pol II recruitment and transcriptional activation. The different mechanism that might direct receptor-specific gene regulation downstream of binding will be discussed in detail in the following paragraphs.

## 4.2 Comparing Motif and Sequence Preference of AR and GR

*In vitro* experiments have shown in previous studies that AR and GR recognize a virtually identical consensus motif. This motif is an inverted repeat of the hexamer 5'-AGAACA-3' which is separated by a three-nucleotide spacer (Mangelsdorf et al. 1995). Another study has demonstrated that even in cellular context AR and GR have similar consensus motifs (Sahu et al. 2013). However, one major difference in binding is that AR also recognizes a motif distinct from the classical consensus, which resembles a direct repeat of the first half-site (Haelens et al. 2003). It has been reported, that unlike AR, GR is not able to recognize direct repeats (Denayer et al. 2010).

To better understand if diverging motif preferences play a role in directing differential gene regulation by the two receptors, a motif enrichment analysis based on our ChIP-seq data was

performed by Melissa Bothe. Interestingly, GR showed a stronger preference for the classical AR/GR consensus motif compared to AR. AR, on the other hand, showed a higher motif preference for the direct repeat (DR3) (Fig. 3.33). Nevertheless, the direct repeat was also present among the GR-specific motifs in this analysis. These results are in line with a previous study arguing that AR binding is more relaxed in terms of sequence stringency compared to GR and that AR achieves specificity through binding to sites stronger diverging from the consensus (Sahu et al. 2014). One potential explanation for the recognition of the direct repeat by AR are subtle structural differences in the second zinc finger of the DBD (Shaffer et al. 2004). Chimeric AR, where this second zinc finger was replaced by the second zinc finger from GR, looses the ability to recognize direct repeats and has a stronger preference for the classical consensus motif (Schauwaers et al. 2007). Hence, relaxed motif stringency and recognition of direct repeat-like motifs are potential drivers of AR-specific DNA binding and possibly gene regulation.

Additionally, this study has shown, that GR and AR have divergent preferences for the local nucleotide composition. A previous study has shown that AR prefers poly(A)-rich sequences directly flanking the consensus sequence (L. Zhang et al. 2018). In accordance with this study, our data demonstrates, that GR has a clear preference for GC-rich DNA sequences compared to AR (Fig. 3.34). But unlike the previous study, our data has shown that this preference was reaching beyond the nucleotides flanking the core and could be observed over a large region of 5 kb. This finding is in line with studies, which have demonstrated that the sequence context, flanking the core motif but also sequences reaching beyond the direct flanks, has an influence on TF binding (Dror et al. 2015; Levo et al. 2015). Moreover, different GC sequence contents in the surrounding of the core binding site have been shown to directly impact TF binding (Dror et al. 2015).
Together this data suggests, that the local sequence composition plays a role in directing receptor-specific binding of AR and GR despite their similar consensus motifs.

In the past, studies have mainly focused of the core binding motif to understand genome-wide TF binding. However, genomic TF binding sites are embedded in a sequence environment which is specific for each binding site, therefore, unique features of the DNA that extend beyond the consensus motif likely play a role in directing TF binding. Accordingly, the nucleotides directly flanking the core motif have been repeatedly shown to strongly influence TF binding (Levo et al. 2015) as well as gene regulation (Rajkumar, Dénervaud, and Maerkl 2013). The sequence composition of the flanking sites has a direct impact on the shape of the DNA and can contribute to TF binding by shape recognition (Rohs et al. 2009).

Further, it has also been demonstrated that distal DNA regions widely extending beyond the flanking sites differ between distinct TFs (Afek et al. 2014; Dror et al. 2015; White et al. 2013). However, it is unlikely that the distal nucleotide composition plays a role in directing TF-DNA interactions by DNA sequence or shape readout but rather contributes to non-specific interactions of TFs with the DNA. One plausible explanation is, that TF have an intrinsic preference for the sequence composition of the surrounding chromatin environment which helps to guide the TFs to their cognate DNA binding sites.

## 4.3 Receptor Expression Level as a Potential Driver of Differential Gene Regulation

Cell-type specific expression of related TFs can be a driver of TF specificity in gene regulation (Singh and Hannenhalli 2008). To a large degree, receptor-specific actions of AR and GR can be explained by their tissue-specific expression (Claessens, Joniau, and Helsen 2017). However, when both receptors are coexpressed in a cell it becomes trickier to understand how the receptors achieve transcriptional specificity.

I have generated a model system to study receptor-specific actions in the same cellular background, however, the receptor quantification revealed that the number of GR receptor expressed per cell is three times higher compared to AR (Fig 3.3). To rule out that the receptor-specific actions observed in this study are simply a result of different receptor numbers in the cell, I performed a set of control experiments using a lower Dex concentration, thereby activating only subset of GR molecules. qPCR analyses for several GR target genes have shown, that hormone-dependent activation by GR does not change when GR was induced with 1 nM Dex, a hormone concentration that was below saturation and was expected to activate a comparable number of GR molecules to AR molecules (Fig. 3.8). Similarly, binding of GR-specific sites (Fig. 3.12) and subsequent opening of those loci (Fig. 3.20) could be still observed, when U2OS-GR cells were treated with only 1 nM Dex, a hormone level which is predicted to lead to around 24 % of the GR molecules per cell being hormone-bound. Bound receptor fractions were calculated based on previously reported $K_D$ values from multiple cell lines and, therefore, represent an estimate rather than exact numbers.

These results suggest, that the differences in receptor specificity identified in this experimental setup are in fact a consequence of receptor-specific properties of AR and GR and can not be explained by the different receptor numbers expressed per cell.

## 4.4 Differential Interaction of AR and GR with the Chromatin Environment

The local chromatin environment is highly cell-type specific and an important driver for TF binding. It has been previously reported that the majority of TF binding sites are located within accessible chromatin (John et al. 2011). However, a special group of TFs, so-called pioneer factors, is able to interact with relatively inaccessible DNA and access nucleosomal enhancers (Zaret and Carroll 2011b). Strikingly, GR has been repeatedly reported to have "pioneering" properties, being able to access nucleosomal DNA (Archer et al. 1991; Richard-Foy and Hager 1987) and induce an opening of the local chromatin structure (Hoffman et al. 2018; T. A. Johnson et al. 2018).

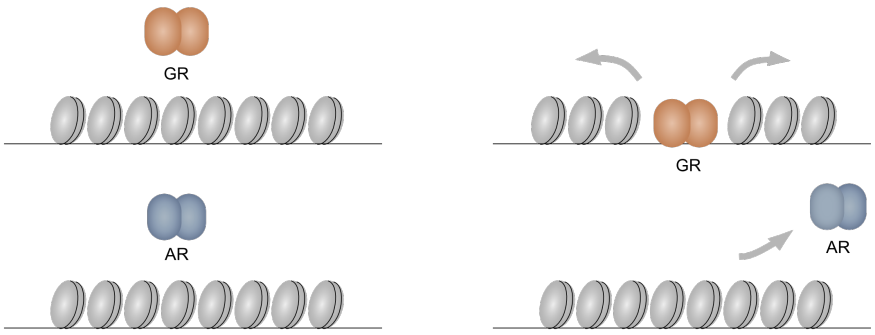



**Figure 4.2. Model explaining receptor-specific gene regulation by the ability to interact with inaccessible chromatin.** The property of GR to bind inaccessible chromatin can contribute to receptor-specific binding and gene regulation.

To study a potential link between selective AR and GR binding and chromatin accessibility, ATAC-seq was performed in both cell lines before and after hormone treatment (Fig 3.19). AR was able to induce opening of AR-specific and shared binding regions. In general, AR-specific sites were found in regions that were already accessible prior to hormone treatment. GR, however, induced opening of GR-specific and shared binding sites. Interestingly, a large fraction of GR-specific sites that were opened by GR showed only a low ATAC-seq signal prior to hormone treatment, indicating that those sites are located in inaccessible chromatin. Accordingly, the repressive histone marks H3K27me3 and H3K9me3 showed a higher signal at GR-specific sites when compared to AR-specific and shared binding sites, suggesting that these sites lie within more condensed chromatin (Fig 3.21). These results indicate, that GR has "pioneering" activity and is able to bind and open sites within condensed chromatin, whereas AR shows only binding in accessible regions. One explanation for the differences in the binding ability to closed chromatin might be differential

protein-protein interactions with cofactors and chromatin remodelers, allowing the receptor to induce nucleosomal rearrangements that support robust GR binding.
Our data clearly supports the previous findings, that GR is able to interact with nucleosomal DNA and has pioneering properties, whereas AR binding was only observed in accessible chromatin. It has been demonstrated before that nucleosomal rearrangements at regulatory elements correlate well with transcriptional activation (He et al. 2010; Schones et al. 2008). Thus, the ability to interact with chromatinized enhancers might indeed be a strong driver of receptor-specific gene regulation by GR.

To test, whether GR binding in closed chromatin is in fact contributing to receptor-specific gene regulation, GR ChIP-seq peaks in regions with a low chromatin accessibility were intersected with AR-specific, GR-specific, shared or non-regulated genes (Fig. 3.22). Indeed, genes regulated in a GR-specific manner were significantly enriched for GR-peaks with a low chromatin accessibility ("pioneering peaks") compared to non-regulated genes. This observation indicates, that the intrinsic pioneering ability of GR, which was not observed for AR, is a potential driver of differential gene regulation. This could provide an explanation how receptors bind to distinct genomic regions although they recognize the same binding motif which in turn can direct different regulatory programs in the same cell.

H3K27ac is a posttranslational histone modification that correlates with active enhancers and is therefore widely used to identify or predict active enhancers (Creyghton et al. 2010; Rada-Iglesias et al. 2011; Zentner, Tesar, and Scacheri 2011). Differential H3K27ac signal could potentially explain receptor-specific gene regulation by AR and GR because of the strong prediction power of H3K27ac on enhancer activity.
Surprisingly, AR induced only little H3K27ac across all three cluster of binding sites. GR, however induced strong H3K27ac compared to AR, with the highest signal intensities at shared followed by GR-specific binding sites. AR is able to specifically induce gene expression, however, the accompanied increase of H3K27ac signal is very subtle. For example, although AR strongly activates *AQP3* expression, it induces only weak H3K27ac in this region (Fig. 3.25). GR binding, however, induced a strong increase in H3K27ac signal in the *AQP3* locus, although it is not able to activate this gene.

Nevertheless, overall our data indicates that H3K27ac induction can not be uncoupled from transcriptional activation as, in general, H3K27ac levels correlated well with activation for both AR and GR (Fig 3.29). Moreover, H3K27ac levels showed more robust changes for enhancers near

regulated genes than for their counterparts that are AR occupied, yet nearby genes are not regulated by AR. Similarly, GR induced the strongest H3K27ac changes in the vicinity of GR-regulated genes compared to regions that are bound, but show no GR regulation of nearby genes. Overall, AR induced a much weaker H3K27sc signal compared to GR, however, the H3K27ac signal still correlated with receptor-specific gene regulation. This data demonstrates, that gene regulation is indeed accompanied by an increase in H3K27ac, but an H3K27ac induction alone is not able to predict gene expression.

One possible reason for differential gene expression downstream of binding is that AR and GR exhibit distinct interactions with the 3D-chromatin architecture. In humans, millions of chromatin loops are preestablished in the genome (F. Jin et al. 2013). During differentiation processes in the cell, the higher order chromatin organization is restructured by the formation of new chromatin loops to connect target gene promoters to their distal regulatory elements (Dixon et al. 2015; Phanstiel et al. 2017). This suggests that, the contact frequency of regulatory elements at loop anchors is an important and fundamental driver of transcriptional activation.

A previous study has demonstrated, that GR binding increased interaction frequencies of GR target genes with their corresponding distal regulatory enhancers (D'Ippolito et al. 2018). Interestingly, after hormone treatment of GR the authors did not observe a formation of new chromatin contacts but rather an increase in the contact frequency of preexisting chromatin loops. One potential explanation for the differential gene regulation from shared binding sites is that the two receptors have distinct abilities to enhance the frequency of chromatin contacts. Chromosome conformation capture techniques (3C) allow us to identify contact frequencies of genetic elements which are in close proximity in the 3D chromatin space but can be far away on linear distance (Dekker et al. 2002). Hi-C is a 3C-based method that has been applied to examine changes in chromatin interactions in a genome-wide manner (Lieberman-Aiden et al. 2009; Schoenfelder et al. 2018). First, the cells are crosslinked with formaldehyde, lysed and the chromatin is digested using restriction enzymes. Next, the chromatin is isolated and proximity ligated under diluted conditions. In this way, the ligated fragments represent physical interactions of genomic fragments in the three-dimensional space. The chimeric fragments can be analyzed using next generation sequencing. However, Hi-C requires deep sequencing to gain a decent resolution and is very cost-intensive. To examine contacts between enhancers and promoters at a distance lower than 1 Mb, capture-based methods can be applied, to enrich for specific regulatory elements such as promoters (Schoenfelder et al. 2018).

By performing Hi-C in AR and GR expressing cells before and after the addition of hormone, one could study the hormone-induced changes of chromatin interactions. Because the formation of chromatin contacts has a mechanistic role in driving gene expression, differences between AR and GR in forming or stabilizing long-range contacts might provide an explanation for receptor-specific gene expression where binding is shared. In this scenario, a receptor-specific ability to form or stabilize long-range contacts would allow just one of the receptors to regulate the nearby gene.

## 4.5 Interaction of AR and GR with Genomic Enhancers

One of the biggest challenges in understanding gene regulation by steroid receptors is the difficulty to precisely assign enhancers to their cognate genes. One reason is, that the majority of binding sites are located at a great distance to the gene promoter, and second, that the number of bound genomic sites highly exceeds the number of regulated genes (John et al. 2011; Reddy et al. 2009). These observations suggest that chromatin looping and the presence of enhancer-promoter contacts is a central mechanism that accommodates gene regulation by steroid receptors (Hakim et al. 2009). Moreover, promoters receive their regulatory information from multiple enhancers which are placed within a regulatory hub where several enhancers contribute to the transcriptional activation of a gene (Vockley et al. 2016).
The first step of gene activation by steroid receptors is the binding of regulatory elements. Steroid receptors do not act alone but interact with other bound TFs to activate transcription (Biddie et al. 2011; Zhao et al. 2016). In a next step, transcriptional cofactors are recruited to the bound enhancers followed by the assembly of transcriptional machinery and transcriptional initiation. Therefore, differential cooperativity with other TFs as well as recruited cofactors are likely to be critical for the transcriptional output specific for AR and GR.

To test whether AR and GR show differences in their abilities to activate genomic enhancers, which in turn might explain receptor-specific gene regulation, I performed STARR-seq, a method to determine the activity of enhancer sequences in a genome-wide manner. Overall, GR showed stronger enhancer activation when compared to AR. Higher activity levels were also found for genomic sequences from the GR-specific and shared cluster of binding sites (Fig 3.28.). Strikingly, AR showed activation of the AR-specific, shared as well as GR-specific sequence cluster. Notably, the GR-specific sequences are not bound by AR in their native chromatin environment, however, when taken out of the chromatin context and placed in an episomal reporter, AR becomes able to activate these sequences. Many of the GR-specific binding sites are located in inaccessible chromatin. In our study, AR showed no pioneering activity and is not able interact with genomic

sites, when they are chromatinized. However, the STARR-seq assay has demonstrated, that in principle AR is able to interact with GR-specific sequences when they are not embedded in their native chromatin context. GR, on the other hand, did not activate the AR-specific cluster, indicating that GR is not able interact with these sites because of their sequence composition.

This study has demonstrated that shared binding sites of AR and GR are none-the-less able to drive differential gene regulation. To understand, if differences in enhancer activation could explain receptor-specific gene regulation, the STARR-seq signal of the shared binding cluster was intersected with each different gene category (AR-specific, GR-specific and shared) in both cell lines. Interestingly, enhancer activity indeed strongly correlated with receptor-specific gene regulation for both AR and GR, supporting the fact, that shared enhancers are able to drive receptor-specific gene regulation by differences in activation once they are bound.

As an example of an AR-specific gene with shared nearby binding sites, we studied the gene *AQP3*. Typically, it is unclear which enhancers are responsible for the regulation of a given gene given that an experimental validation of candidate enhancers is typically lacking. To test if the shared enhancers near the receptor-specific target genes *AQP3* and *GILZ* are indeed involved in regulation of the nearby gene, we delete such binding sites using the CRISPR/Cas9-system. The deletions led to loss of receptor-specific gene activation of these genes supporting the regulatory role of the corresponding enhancers. Interestingly, the STARR-seq data revealed that one of the two binding sites near the *AQP3* gene shows enhancer activity upon AR induction but not GR (Fig. 3.31). Accordingly, when I inserted the *AQP3* enhancer in a STARR-seq reporter and transfected AR- and GR-expressing cells with this reporter construct, only AR was able to induce activation. Strikingly, the enhancer sequence contained three motifs resembling a direct repeat of the consensus motif 5'-AGAACA-3'. When I changed the sequence of the second halfsite of the direct repeats into a more consensus-like palindromic motif (AGA→TGT), GR became able to activate the reporter (Fig. 3.32). It is striking that although the ChIP-seq data showed that GR is able to bind in the *AQP3* enhancer region, it is not able to activate the enhancer. Our motif analysis of all GR ChIP-Seq peaks identified a direct repeat motif (DR3) as being enriched for GR binding, suggesting that GR is indeed able to bind these sequences. However, GR is not able to activate enhancers containing direct-like motifs. This result indicates, that direct repeat-like sequences are likely playing a role in driving AR-specific regulation, because unlike GR, AR is able to activate enhancers containing direct repeat motifs.

There a several possible explanations as to why GR appears to be able to bind but not able to activate AR-specific enhancers containing a DR3 motifs. First, due to structural differences in the dimerization domain of AR and GR, AR harbors an additional hydrogen bond stabilizing its contact with the DNA (Shaffer et al. 2004). This increased stability of the DBD-DNA complex of AR might allow mismatches in the second half site of the consensus sequence leading to a recognition of enhancers with DR3 motifs. GR binding to these motifs is possibly too weak to induce enhancer activation but was still captured by ChIP due to efficient formaldehyde crosslinking. Alternatively, GR might be binding nearby genomic binding sites, which can not be distinguished by ChIP-seq due to the low resolution.

A second potential explanation is that binding of the DR3 motif induces conformational changes of AR allowing it to activate transcription when bound whereas GR remains in an inactive conformation. In this scenario, the DNA elements serve as allosteric ligands that that can induce distinct conformations for the steroid receptors. Upon binding of the DNA, the receptors might undergo structural changes which could in turn alter their affinity for transcriptional coregulators. Allosteric signaling is well studied for the DBD of GR. More specifically, the DNA serves as an allosteric ligand for GR and induces structural changes of the receptor in a sequence-dependent manner that are transmitted to the LBD and result in the recruitment of cofactors and distinct activities (Meijsing et al. 2009; Watson et al. 2013). Similarly, the DR3 motif could serve as an allosteric ligand for AR resulting in cofactor binding and transcriptional activation whereas it fails to induce an activation-competent state in GR.

One advantage of episomal reporter assays is that one can measure the enhancer activity independent of the chromatin context to identify the intrinsic and sequence-specific enhancer potential of regulatory elements. At the same time, this is also a major drawback of reporter assays. The regulatory elements are tested outside their native chromatin environment, and therefore, once identified, enhancers need to be functionally validated. One possibility to test enhancers in their native environment is to apply the CRISPR/Cas9 system to induce genomic perturbations, such as point mutations or deletions, and to assay the phenotypic consequences.

## 4.6 Influence of Differential Cofactor Interactions

Numerous studies have focused on differences in genome-wide TF binding as explanation for TFs regulating distinct sets of genes. However, there are many layers of complexity between TF binding and the resulting phenotype. For example, binding of one TF is probably not enough to induce gene expression but requires the interaction with other TFs as well as cofactors, which contribute to gene activation in several ways. Chromatin remodelers induce nucleosome rearrangements thereby generating a binding interface for further TFs or mediate post-translational modifications on histones. Other cofactors such as mediator or EP300/CREBBP are involved in the recruitment and the assembly of the transcriptional machinery, but also the release of the Pol II into productive elongation requires specialized cofactors.
When both AR and GR show identical binding patterns near differentially regulated genes, binding fails to explain receptor-specific gene regulation. Therefore, mechanisms downstream of genomic binding, such as cofactor interaction, likely play a role in directing differential gene expression from shared binding sites.

A multitude of cofactors that are crucial for the function of steroid hormone receptors has been identified in the past (Rosenfeld, Lunyak, and Glass 2006). Differential interaction of various cofactors with AR and GR could affect gene regulation on several levels. First, differential recruitment of chromatin remodelers could allow one receptor to induce a shifting of the nucleosomes and create a permissive state for gene activation. Second, differences in the recruitment of the mediator complex could stabilize enhancer promoter contacts and lead to the PIC assembly in one case but not the other. And lastly, differential interaction with EP300 could lead to Pol II recruitment in one case but not the other (Boija et al. 2017). Therefore, differential recruitment of cofactors might explain receptor-specific gene activation by AR and GR from shared binding sites.

The DBD is the most conserved protein domain of steroid receptors, whereas the N-terminal domain shows only little conservation and is the most variable domain. Distinct cofactor interactions of the two receptors might be a result of structural differences of the N-terminus between AR and GR, which contains the activation function 1 domain and mediates recruitment of specific coactivators (Wärnmark et al. 2003). Furthermore, it has been shown that the recruitment of cofactors by GR is gene-specific (W. Chen, Rogatsky, and Garabedian 2006; Sacta et al. 2018; Stallcup and Poulard 2020). Therefore, differential cofactor interaction of AR and GR possibly drive receptor-specific gene activation from sites that are bound by both receptors.

To address this question, the interactomes of both AR and GR in U2OS cells were examined using RIME in collaboration with Stefan Prekovic and Isabel Mayayo Peralta from the Wilbert Zwart lab. For both receptors, shared but also receptor-specific interactors could be identified (Fig 3.35). Moreover, a geneset enrichment analysis identified that the mediator complex as well as enzymes harboring acetyltransferase activity are enriched among the interactors of GR compared to AR (Fig 3.36). Strikingly, almost every component of the mediator complex examined was enriched for GR compared to AR (Fig. 3.37). This finding is consistent with a previous study in another cell type, which has shown a strong interaction of the mediator complex with GR but only little interaction with AR in human cells (Lempiäinen et al. 2017). In addition, ChIP-qPCR analyses using antibodies for the mediator complex subunit Med1 in both cell lines has shown strong enrichment of Med1 at shared binding sites in GR cells but only little Med1 enrichment in the AR expressing cell line, suggesting that GR is able to recruit Med1 to its binding sites in a hormone-dependent manner whereas AR shows almost no Med1 recruitment (Fig 3.37). Thus, selective recruitment of the mediator complex could explain GR-specific regulation at genes that require recruitment of this complex for activation.

Both EP300 and CREBBP play a central role in catalyzing H3K27ac (Bannister and Kouzarides 1996; Ogryzko et al. 1996) and our data has demonstrated that GR induces strong H3K27ac signal whereas AR induces only subtle H3K27ac changes. Surprisingly, despite an enrichment of acetyltransferases among the GR interactors, the acetyltransferases EP300 and CREBBP similarly interacted with both receptors AR and GR. Consistent with the RIME data, ChIP-qPCR analysis with antibodies specific for EP300 has confirmed, that both AR and GR are able to recruit EP300 to their shared binding sites (Fig 3.37). Because no differences in interaction with EP300 or CREBBP could be observed, other histone acetyltransferases might be specifically recruited by GR and be responsible for the GR-specific induction of H3K27ac. Although CREBBP shows comparable interaction frequencies with both receptors based on RIME, it is possible that this cofactor is still differentially recruited to their DNA binding sites. Finally, different levels of H3K27ac could be driven by differences in the activity of the recruited histone acetyltransferases.

Further, H3K27ac has been show to distinguish active from poised enhancers (Creyghton et al. 2010) and induction of this histone mark correlates well with enhancer activation followed by transcriptional activation. Thus, differential H3K27ac induced by AR and GR might explain receptor-specific gene activation resulting from receptor-specific interactions with acetyltransferases. Moreover, the more robust H3K27ac changes might be a reason why GR, in

general, activated more genes with higher expression levels and induced stronger enhancer signals compared to AR.

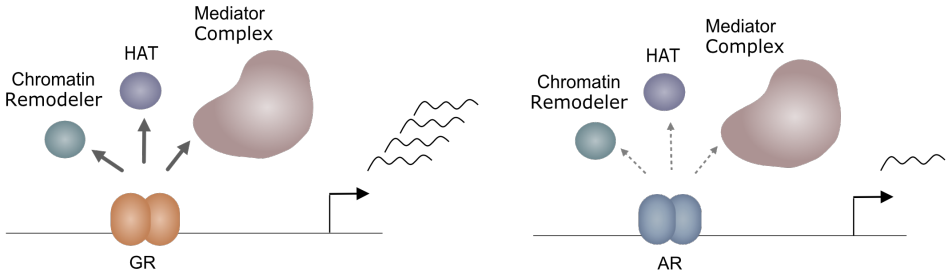



**Figure 4.2. Model explaining receptor-specific gene regulation by different cofactor interactions.** Differential cofactor interaction of AR and GR can drive receptor-specific gene activation from sites that are occupied by both receptors.

Together, these results suggest that differential interaction with the mediator complex might contribute to hormone-dependent gene regulation by GR from binding sites shared by AR and GR. Moreover, enriched interactions of GR with acetyltransferases could contribute to the stronger H3K27ac induction by GR compared to AR and are a potential driver of GR-specific gene expression. However, EP300 and CREBBP did not demonstrate differential recruitment to the chromatin. Therefore, other candidate acetyltransferases have to be examined using ChIP-qPCR, and subsequently, they can be functionally validated with the help of RNA interference-based knock-down-experiments or CRISPR/Cas-9-mediated genomic perturbations.

Of note, the recruitment of CREBBP to DNA does not necessarily provide a functional link to gene regulation. Interestingly, one study has found that the majority of CREBBP binding events did not activate the nearby genes arguing that a ChIP binding analysis alone is a poor predictor to identify the genes controlled by this factor and to characterize its biological role (Kasper et al. 2014). Thus, similar recruitment of CREBBP by both AR and GR could still lead to differential regulation of nearby genes, where additional cofactors might be crucial for receptor-specific gene regulation.

## 4.7 Conclusion

With this study, I have addressed the long-standing question how the closely related TFs AR and GR direct distinct regulatory programs although they recognize nearly identical DNA binding sites. In eukaryotes most TFs belong to few large TF families that tend to bind similar DNA motifs, yet, most TFs developed their specific regulatory functions (Lambert et al. 2018b).
With various layers of data ranging from the transcriptome and cistrome of the receptors to chromatin accessibility, enhancer activation and interactome profiling, I could identify unique functions of the receptors potentially contributing to differences in gene regulation. I have discovered that AR and GR are able to drive differential gene expression through differential but also shared binding sites. Moreover, the receptors show distinct properties to interact with the chromatin with GR being able to bind inaccessible chromatin sites whereas AR does not. Further, I have identified different preferences for the local sequence composition. Finally, I have found differences in the recruitment of cofactors that might direct functional diversification between AR and GR.

Together, I was able to gain valuable insights into the unique features of AR and GR signaling in an identical cellular background, which might give an explanation for the divergent biological functions of the receptors. Moreover, I provided evidence that divergent binding as well as mechanisms downstream of binding play a central role in receptor-specific gene regulation. The mechanisms I have identified in this work might also apply to other TF paralogs explaining functional diversification despite similar binding preferences.

.

# Abbreviations

| | |
|---|---|
| AR | Androgen receptor |
| ATAC | Assay for transposase-accessible chromatin using sequencing |
| bp | Base pair |
| Cas9 | CRISPR-associated protein 9 |
| cDNA | Complementary deoxyribonucleic acid |
| ChIP | Chromatin immunoprecipitation |
| CRISPR | Clustered regularly interspaced short palindromic repeats |
| DBD | DNA-binding domain |
| Dex | Dexamethasone |
| DNA | Deoxyribonucleic acid |
| FAIRE | Formaldehyde-assisted isolation of regulatory elements |
| GR | Glucocorticoid receptor |
| HAT | Histone acetyltransferase |
| HMT | Histone methyltransferase |
| Indel | Insertion and deletion |
| LBD | Ligand-binding domain |
| LFQ | Label-free quantification |
| MPRA | Massively parallel reporter assay |
| mRNA | Messenger ribonucleic acid |
| NTD | N-terminal domain |
| PAM | Protospacer adjacent motif |
| PCR | Polymerase chain reaction |
| PIC | Pre-initiation complex |
| Pol II | RNA polymerase II |
| qPCR | Quantitative Polymerase Chain Reaction |
| R1881 | Metribolone |
| RIME | Rapid immunoprecipitation mass spectrometry of endogenous proteins |
| RNA | Ribonucleic acid |
| SEM | Standard error of mean |
| sgRNA | Single guide RNA |
| STARR-seq | Self-transcribing active regulatory region sequencing |
| TAD | Topologically associating domain |

# Primer Sequences

**Table S1: qPCR Primer for cDNA**

| | |
|---|---|
| *RPL19*_Fwd | ATGTATCACAGCCTGTACCTG |
| *RPL19*_Rev | TTCTTGGTCTCTTCCTCCTTG |
| *AR*_Fwd | ATCCCAGTCCCACTTGTGTC |
| *AR*_Rev | CCACAGATCAGGCAGGTCTT |
| *ENDOD1*_Fwd | TCCTTCCAGGAACGGTGGTA |
| *ENDOD1*_Rev | TCAGGGACTGCCACTTTGTC |
| *SIGLEC14*_Fwd | TGGAGGTGACAGCCCTGATA |
| *SIGLEC14*_Rev | GAATGTGAGAGGTGGTCCCG |
| *GFP*_Fwd | GGCCAGCTGTTGGGGTGTC |
| *GFP*_Rev | TTGGGACAACTCCAGTGAAGA |
| *GILZ*_Fwd | CCATGGACATCTTCAACAGC |
| *GILZ*_Rev | TTGGCTCAATCTCTCCCATC |
| *AQP3*_Fwd | GCAGCCTGTCCATCTGTG |
| *AQP3*_Rev | ACCCTACTTCCCAAAAGCC |
| *FKBP5*_Fwd | TGAAGGGTTAGCGGAGCAC |
| *FKBP5*_Rev | CTTGGCACCTTCATCAGTAGTC |
| *ABLIM3*_Fwd | TATTAGTCCACGCGCCTTCA |
| *ABLIM3*_Rev | TGCTGATAAGGAATGCTAGTGT |
| *IGFBP1*_Fwd | TCACAGCAGACAGTGTGAGAC |
| *IGFBP1*_Fwd | AGACCCAGGGATCCTCTTC |
| *SIGLEC14*_Fwd | TGGAGGTGACAGCCCTGATA |
| *SIGLEC14*_Rev | GAATGTGAGAGGTGGTCCCG |
| *TMEM63C*_Fwd | GTCTGGGGTCACTCTTCTGC |
| *TMEM63C*_Rev | AGAGCAACCCAAAAGGCACA |
| *FAM105A*_Fwd | TGTGGAGGCAGAGGTTGATTT |
| *FAM105A* _Rev | AGCTCCTCATAAGCCTTCCTC |
| *FAM104A*_Fwd | GTAGTGTCAGCGGGCATCTT |
| *FAM104A*_Rev | GGCGGCGCAATAGAGAAGTA |
| *MYO1E*_Fwd | AAGACCGTCCGGAACAACAA |
| *MYO1E* _Rev | GCCCTCGATGAGCTGGTAAA |

**Table S2: qPCR Primer for ChIP and ATAC**

| | |
|---|---|
| n.c1_Fwd | AATGGCAGCCCCTAGTCATTC |
| n.c1_Rev | AACTGGGAGTGATACTGGTTCC |
| n.c2_Fwd | TGCATGACGCAGACCTTTCT |
| n.c2_Rev | ATGAGAACCACATGGGCCAG |
| locus1_Fwd | CCTTTTTCAATTTGGGTGGTT |
| locus1_Rev | GATGTCCATTTTCACCACGA |
| locus2_Fwd | ATGCCACTCCCTTCTCCATT |
| locus2_Rev | CACAGGTCTCGGCTAACAGA |
| locus3_Fwd | CCCATTGGTGCCAGTACTGA |
| locus3_Rev | AGAGGTCCGAGGTTTGAGAG |
| locus4_Fwd | CCTCTTAGGTTGGTGCAGATT |
| locus4_Rev | TACCTACCCAGTTCCAGAGC |
| locus5_Fwd | TGGGATCTGCTGACAAGTGT |
| locus5_Rev | CCTGTCTGCCTCCTCAAGAA |

| | |
|---|---|
| locus6_Fwd | GACCGCACTTCTCAGTGTCA |
| locus6_Rev | GGACATCACAAACACCAGCA |
| locus7_Fwd | ATGCAAAAGCCCCTACACAG |
| locus7_Rev | AGACAGCCAGAGCGTAGAGC |
| locus8_Fwd | AGAGGTGAACGAGGTGGATG |
| locus8_Rev | AGGGACTTGGGAGGTCTGTT |
| locus9_Fwd | CTTCTCTGCCCAGGTGCTAT |
| locus9_Rev | CTGCCACTGAAGGAGACACA |
| *ZBTB16*_Fwd | CTCCTTGAGGGAAAGAACACAC |
| *ZBTB16*_Rev | ACAGACGCAGGGCATTTTAC |
| *FKBP5*_Fwd | GCATGGTTTAGGGGTTCTTG |
| *FKBP5*_Rev | TAACCACATCAAGCGAGCTG |

**Table S3: qPCR Primer for Med1 and EP300 ChIP**

| | |
|---|---|
| *GILZ* Peak 1_Fwd | ACTGCCTCTTTTTCTAAGGGC |
| *GILZ* Peak 1_Rev | TCTCTCATCTCATCCTCATGGA |
| *GILZ* Peak 2_Fwd | CTCAGCAGCTTTTCTTCGTG |
| *GILZ* Peak 2_Rev | AACCAAGGAATTGGGTCACA |
| *IGFBP1*_Fwd | CCAGGAGGTGTTTGGAATGT |
| *IGFBP1*_Rev | TCATGTTCTTAGGGGGCAAC |
| *ABLIM3*_Fwd | GAGGTTTGATTCCCATTCCA |
| *ABLIM3*_Rev | CCTGGAGTGGAACACTGTGA |
| *SRPK2*_Fwd | GACATCACACCTCGTCTC |
| *SRPK2*_Rev | GGATGTGCTCTTCATGTC |
| *USP18*_Fwd | TGCTGGCAGAACAAGATGTC |
| *USP18*_Rev | AAGGAACCAATGTTGCTTGG |

**Table S4: qPCR Primer for H3K27ac ChIP**

| | |
|---|---|
| *TMEM63C*_Fwd | CAAAGGGAAACCGAAGCATA |
| *TMEM63C* _Rev | CAGAGTGAAAGGCTGGGAAA |
| *FAM105A*_Fwd | AAGGGGAGGAGGTGAGAGAA |
| *FAM105A* _Rev | AAGCTGGGACAGTTGGTCAC |
| *FAM104A*_Fwd | GGAGTGGCTCAACAACTGAAT |
| *FAM104A* _Rev | AAAATCCGTGCATTGGTCTC |
| *MYO1E*_Fwd | ACCTACAGCCATGGGTTTCA |
| *MYO1E* _Rev | TGGGCTTATCATCATCTGCA |

## Oligomers for Omni-ATAC

**Table S5: Primer for ATAC enrichment PCR ***

| Name | Description | Sequence 5' to 3' |
|---|---|---|
| MB001 | Universal Primer_fwd | AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTC |
| MB002 | Barcode Primer1_rev | CAAGCAGAAGACGGCATACGAGAT**GGATGTTCT** GTCTCGTGGGCTCGG |
| MB003 | Barcode Primer2_rev | CAAGCAGAAGACGGCATACGAGAT**CTTATCCAG** GTCTCGTGGGCTCGG |
| MB004 | Barcode Primer3_rev | CAAGCAGAAGACGGCATACGAGAT**GTAAGTCAC** GTCTCGTGGGCTCGG |
| MB005 | Barcode Primer4_rev | CAAGCAGAAGACGGCATACGAGAT**TTCAGTGAG** GTCTCGTGGGCTCGG |
| MB006 | Barcode Primer5_rev | CAAGCAGAAGACGGCATACGAGAT**CTCGTAATG** GTCTCGTGGGCTCGG |
| MB007 | Barcode Primer6_rev | CAAGCAGAAGACGGCATACGAGAT**CATGTCTCA** GTCTCGTGGGCTCGG |

* The primer sequences anneal to the Nextera adapters pasted by the TN5 transposase. The barcoded primers contain a P7 Nextera barcode (in bold). The primer sequences were adapted from (Wang et al., 2013)

## Oligomers for CRISPR

**Table S6: Single guide RNA Oligos**

| Oligo | Sequence |
|---|---|
| *AQP3*_214_up_fwd | CACCGCAGGGGGACATTGCATCAGA |
| *AQP3*_214_up_rev | AAACTCTGATGCAATGTCCCCCTGC |
| *AQP3*_266_down_fwd | CACCGAACATGAGCTCCTCCAAAGT |
| *AQP3*_266_down_rev | AAACACTTTGGAGGAGCTCATGTTC |

**Table S7: Genotyping PCR Primer for AQP3 clonal lines**

| | |
|---|---|
| MB183_Junction_F | ACACATTTTCCCACCCCTCT |
| MB184_Junction_R | GGGAGAATTGCATCCCCTAT |
| MB185_Breakpoint_left_R | AAGCTCATCCATCACCAACC |
| MB187_Breakpoint_right_F2 | TCCAGGATGTTCACTCAGAGG |

## Oligomers for STARR-seq

**Table S8: PCR Primers for FAIRE (Arnold et al, 2013)**

| | |
|---|---|
| In_Fusion_fwd | TAGAGCATGCACCGGACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| In_Fusion_rev | GGCCGAATTCGTCGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT |

**Table S9: UMI Barcode Primer for Reverse Transcription**

| | |
|---|---|
| RT primer | 5'-CAAGCAGAAGACGGCATACGAGATnnnnnnnnGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCT -3' |

**Table S10: PCR Primers with Illumina Adaptors**

| | | |
|---|---|---|
| Forward Primer | CAAGCAGAAGACGGCATACGA | P7 index |
| Reverse Index 1 | AATGATACGGCGACCACCGAGATCTACAC -**TATCCTCT**-ACACTCTTTCCCTACACGACGCTC | |
| Reverse Index 2 | AATGATACGGCGACCACCGAGATCTACAC -**AGAGTAGA**-ACACTCTTTCCCTACACGACGCTC | |

**Table S11: Sanger Sequencing Primer**

| | |
|---|---|
| STARR-seq_Fwd | GGTCCTTCTTGAGTTTGTAAC |

**Table S12: cDNA Synthesis Primer for STARR-seq Analysis**

| | |
|---|---|
| MB108_*RPL19* | GAGGCCAGTATGTACAGACAAAGTGG |
| MB108_STARR | CAAACTCATCAATGTATCTTATCATG |

**Table S13: qPCR Primer for STARR-seq Analysis**

| | |
|---|---|
| MB109_STARR_Fwd | TTGGGACAACTCCAGTGAAGA |
| MB109_STARR_Rev | GGCCAGCTGTTGGGGTGTC |
| MB110_*RPL19*_Fwd | ATGTATCACAGCCTGTACCTG |
| MB110_*RPL19*_Rev | TTCTTGGTCTCTTCCTCCTTG |

**Table S14: Enhancer Fragments for STARR-seq ***

| | |
|---|---|
| *IP6K3* (p.c.) | **TAGAGCATGCACCGGACACTCTTTCCCTACACGACGCTCTTCCGATCT**TGGGATAATCTGCC CAAGGTCACATGGTCGGTGTGGGGTGGAGCTCTGGGAGGGCCATGCCAGACGTGGGGATG GAGGAAAGGACAGGCTGTTCCAGAAGGTCTGGTGTTGTCACGCTGTTTACCCCAAGTTTGCAT TAGGGACATTCCTGCTGTGTCCCCTGTCATGTTTGTTCTGGTCATTCTGTTCCATAGGAGAGAA CAGAGACGCTGTGATTCCCTCCCCTCAGGGAGGGTCGTCAGCGCTGAGGGCTGGGAGCCCAG GCTGCAAGGAGATGGTGTTTACATTTCCAGGCCTGTGTCTTGGGGAAGGAGGATTGGAGTGC TTTGTTCTAGGGAGGAGAACTGACCAGATCGGAAGAG**CACACGTCTGAACTCCAGTCACTCG ACGAATTCGGCC** |
| Scrambled GBS (n.c.) | **TAGAGCATGCACCGGACACTCTTTCCCTACACGACGCTCTTCCGATCT**CAGCGAAAGAAACT CCGTTGCCCGTCGCT**CACACGTCTGAACTCCAGTCACTCGACGAATTCGGCC** |
| *AQP3* (WT) | **TAGAGCATGCACCGGACACTCTTTCCCTACACGACGCTCTTCCGATCT**TGGAGTAGTCCCTC TTCTCCCCTTCCCACCCCACCCAACTCTACCCCACCCACGTTTCCCCAAGCCTAGAAGTGCCCCA CCTAGGTTTTCAGAAAAACAAAATCAAGAGGAAAGAGGAAGGAGGGAGCTCATTACCAGGAA TTAAGGAGGAGGCCTGCCCTAGGGGAACACAAGTGAGACTTGGCTGGCAGGCAGAGATATG CACTGAGTACATGGAGGTGTACTCAAAGTACACTGAGAGAGGGGCTCAGAGGAACATGCAGA GATGTGATCTGAGGGTCACAGTCCAGGATGTTCACTCAGAGGTACACAAAGAAAGACCCAGA CACAAGGCTGGGCACTGTGGTTCATGCCTGTAATCCCAGCACTTTGGGTGGCCAAGGCAGGT GGATCACCTGAGGTCAGGAGTTCGACACCAGCCTGGGCAACACGGTGAAACCCTGTCTCTACT AAAAATACAAAAATTAGCTGGGCGTGGTGGCATGTGCTTGTAATCCCAGCTACTCGGGAGGC TGAGGCAGGAGAATTGCTTCAACCCAGGAGGCAGAGGTTGC**CACACGTCTGAACTCCAGTCA CTCGACGAATTCGGCC** |
| *AQP3*-Mut | **TAGAGCATGCACCGGACACTCTTTCCCTACACGACGCTCTTCCGATCT**TGGAGTAGTCCCTC TTCTCCCCTTCCCACCCCACCCAACTCTACCCCACCCACGTTTCCCCAAGCCTAGAAGTGCCCCA CCTAGGTTTTCAGAAAAACAAAATCAAGAGGAAAGAGGAAGGAGGGAGCTCATTACCAGGAA TTAAGGAGGAGGCCTGCCCTAGGGGAACACAAGTGAGACTTGGCTGGCAGGCAGAGATATG CACTGAGTACATGGAGGTGTACTCAAAGTATACTGATAGAGGGGCTCAGAGGAATATGCATA GATGTGATCTGAGGGTCACAGTCCAGGATGTTCACTCAGAGGTATACAATATAAGACCCAGAC ACAAGGCTGGGCACTGTGGTTCATGCCTGTAATCCCAGCACTTTGGGTGGCCAAGGCAGGTG GATCACCTGAGGTCAGGAGTTCGACACCAGCCTGGGCAACACGGTGAAACCCTGTCTCTACTA AAAATACAAAAATTAGCTGGGGCCACCACGATGTGCTTGTAATCCCAGCTACTCGGGAGGCTG AGGCAGGAGAATTGCTTCAACCCAGGAGGCAGAGGTTGC**CACACGTCTGAACTCCAGTCACT CGACGAATTCGGCC** |
| *AQP3*-Trans | **TAGAGCATGCACCGGACACTCTTTCCCTACACGACGCTCTTCCGATCT**TGGAGTAGTCCCTC TTCTCCCCTTCCCACCCCACCCAACTCTACCCCACCCACGTTTCCCCAAGCCTAGAAGTGCCCCA CCTAGGTTTTCAGAAAAACAAAATCAAGAGGAAAGAGGAAGGAGGGAGCTCATTACCAGGAA TTAAGGAGGAGGCCTGCCCTAGGGGAACACAAGTGAGACTTGGCTGGCAGGCAGAGATATG CACTGAGTACATGGAGGTGTACTCAAAGTACACTGTGTGAGGGGCTCAGAGGAACATGCTGT GATGTGATCTGAGGGTCACAGTCCAGGATGTTCACTCAGAGGTACACAATGTAAGACCCAGAC ACAAGGCTGGGCACTGTGGTTCATGCCTGTAATCCCAGCACTTTGGGTGGCCAAGGCAGGTG GATCACCTGAGGTCAGGAGTTCGACACCAGCCTGGGCAACACGGTGAAACCCTGTCTCTACTA AAAATACAAAAATTAGCTGGGGCCACCACGATGTGCTTGTAATCCCAGCTACTCGGGAGGCTG AGGCAGGAGAATTGCTTCAACCCAGGAGGCAGAGGTTGC**CACACGTCTGAACTCCAGTCACT CGACGAATTCGGCC** |

* The genomic fragments were either ordered as Gblock (IDT) of GeneStrand (Eurofins)

# References


Afek, Ariel et al. 2014. "Protein-DNA Binding in the Absence of Specific Base-Pair Recognition." *Proceedings of the National Academy of Sciences of the United States of America*.

Amler, Lukas C. et al. 2000. "Dysregulated Expression of Androgen-Responsive and Nonresponsive Genes in the Androgen-Independent Prostate Cancer Xenograft Model CWR22-R." *Cancer Research*.

Andersson, Robin, and Albin Sandelin. 2020. "Determinants of Enhancer and Promoter Activities of Regulatory Elements." *Nature Reviews Genetics*.

Archer, T K, M G Cordingley, R G Wolford, and G L Hager. 1991. "Transcription Factor Access Is Mediated by Accurately Positioned Nucleosomes on the Mouse Mammary Tumor Virus Promoter." *Molecular and Cellular Biology*.

Arnold, Cosmas D. et al. 2013. "Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-Seq." *Science*.

Arora, Vivek K. et al. 2013. "Glucocorticoid Receptor Confers Resistance to Antiandrogens by Bypassing Androgen Receptor Blockade." *Cell*.

Badis, Gwenael et al. 2009. "Diversity and Complexity in DNA Recognition by Transcription Factors." *Science*.

Bailey, Timothy L. et al. 2009. "MEME Suite: Tools for Motif Discovery and Searching." *Nucleic Acids Research*.

Banerji, Julian, Sandro Rusconi, and Walter Schaffner. 1981. "Expression of a β-Globin Gene Is Enhanced by Remote SV40 DNA Sequences." *Cell*.

Bannister, Andrew J., and Tony Kouzarides. 1996. "The CBP Co-Activator Is a Histone Acetyltransferase." *Nature*.

Bannister, Andrew J., and Tony Kouzarides 2011. "Regulation of Chromatin by Histone Modifications." *Cell Research*.

Bantscheff, Marcus et al. 2007. "Quantitative Mass Spectrometry in Proteomics: A Critical Review." *Analytical and Bioanalytical Chemistry*.

Barozzi, Iros et al. 2014. "Coregulation of Transcription Factor Binding and Nucleosome Occupancy through DNA Features of Mammalian Enhancers." *Molecular Cell*.

Barski, Artem et al. 2007. "High-Resolution Profiling of Histone Methylations in the Human Genome." *Cell*.

Beato, Miguel, Georges Chalepakis, Michael Schauer, and Emily P. Slater. 1989. "DNA Regulatory Elements for Steroid Hormones." *Journal of Steroid Biochemistry*.

Bell, Oliver, Vijay K. Tiwari, Nicolas H. Thomä, and Dirk Schübeler. 2011. "Determinants and Dynamics of Genome Accessibility." *Nature Reviews Genetics*.

Berman, Helen, Kim Henrick, and Haruki Nakamura. 2003. "Announcing the Worldwide Protein Data Bank." *Nature Structural Biology*.

Berrebi, Dominique et al. 2003. "Synthesis of Glucocorticoid-Induced Leucine Zipper (GILZ) by Macrophages: An Anti-Inflammatory and Immunosuppressive Mechanism Shared by Glucocorticoids and IL-10." *Blood*.

Bettoun, David J. et al. 2005. "Interaction between the Androgen Receptor and RNase L Mediates a Cross-Talk between the Interferon and Androgen Signaling Pathways." *Journal of Biological Chemistry*.

Biddie, Simon C. et al. 2011. "Transcription Factor AP1 Potentiates Chromatin Accessibility and Glucocorticoid Receptor Binding." *Molecular Cell*.

Boija, Ann et al. 2017. "CBP Regulates Recruitment and Release of Promoter-Proximal RNA Polymerase II." *Molecular Cell*.

de Bosscher, K., W. vanden Berghe, and G. Haegeman. 2006. "Cross-Talk between Nuclear Receptors and Nuclear Factor KB." *Oncogene*.

Boyle, Alan P. et al. 2008. "High-Resolution Mapping and Characterization of Open Chromatin across the Genome." *Cell*.

Buenrostro, Jason D. et al. 2013. "Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position." *Nature Methods*.

Buenrostro, Jason D. et al. 2015. "ATAC-Seq: A Method for Assaying Chromatin Accessibility Genome-Wide." In *Current Protocols in Molecular Biology*, Hoboken, NJ, USA: John Wiley & Sons, Inc.

Bulajić, Milica et al. 2020. "Differential Abilities to Engage Inaccessible Chromatin Diversify Vertebrate HOX Binding Patterns." *Development*.

Canver, Matthew C. et al. 2015. "BCL11A Enhancer Dissection by Cas9-Mediated in Situ Saturating Mutagenesis." *Nature*.

Chen, Charlie D. et al. 2004. "Molecular Determinants of Resistance to Antiandrogen Therapy." *Nature Medicine*.

Chen, Weiwei, Inez Rogatsky, and Michael J. Garabedian. 2006. "MED14 and MED1 Differentially Regulate Target-Specific Gene Activation by the Glucocorticoid Receptor." *Molecular Endocrinology*.

Cirillo L.A., Lin F R Cuesta I Friedman D Jarnik M Zaret K S et al. 2002. "Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4." *Molecular cell*.

Claessens, Frank, Steven Joniau, and Christine Helsen. 2017. "Comparing the Rules of Engagement of Androgen and Glucocorticoid Receptors." *Cellular and Molecular Life Sciences*.

Clark, Andrew R., and Maria G. Belvisi. 2012. "Maps and Legends: The Quest for Dissociated Ligands of the Glucocorticoid Receptor." *Pharmacology and Therapeutics*.

Cleutjens, C. B.J.M. et al. 1997. "Both Androgen Receptor and Glucocorticoid Receptor Are Able to Induce Prostate-Specific Antigen Expression, but Differ in Their Growth-Stimulating Properties of LNCaP Cells." *Endocrinology*.

Corces, M. Ryan et al. 2017. "An Improved ATAC-Seq Protocol Reduces Background and Enables Interrogation of Frozen Tissues." *Nature Methods*.

Cremer, T., and C. Cremer. 2001. "Chromosome Territories, Nuclear Architecture and Gene Regulation in Mammalian Cells." *Nature Reviews Genetics*.

Creyghton, Menno 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*.

Crocker, Justin et al. 2015. "Low Affinity Binding Site Clusters Confer HOX Specificity and Regulatory Robustness." *Cell*.

Datta, Rhea R. et al. 2018. "A Feed-Forward Relay Integrates the Regulatory Activities of Bicoid and Orthodenticle via Sequential Binding to Suboptimal Sites." *Genes and Development*.

Dehal, Paramvir, and Jeffrey L. Boore. 2005. "Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate." *PLoS Biology*.

Dekker, Job, Marc A. Marti-Renom, and Leonid A. Mirny. 2013. "Exploring the Three-Dimensional Organization of Genomes: Interpreting Chromatin Interaction Data." *Nature Reviews Genetics*.

Dekker, Job, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. 2002. "Capturing Chromosome Conformation." *Science*.

Denayer, Sarah et al. 2010. "The Rules of DNA Recognition by the Androgen Receptor." *Molecular Endocrinology*.

"Dexamethasone in Hospitalized Patients with Covid-19 — Preliminary Report." 2020. *New England Journal of Medicine*.

Dilworth, F. J., and P. Chambon. 2001. "Nuclear Receptors Coordinate the Activities of Chromatin Remodeling Complexes and Coactivators to Facilitate Initiation of Transcription." *Oncogene*.

D'Ippolito, Anthony M. et al. 2018. "Pre-Established Chromatin Interactions Mediate the Genomic Response to Glucocorticoids." *Cell Systems*.

Dixon, Jesse R. et al. 2012. "Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions." *Nature*.

Dixon, Jesse R. et al. 2015. "Chromatin Architecture Reorganization during Stem Cell Differentiation." *Nature*.

Dobin, Alexander et al. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics*.

Doni Jayavelu, Naresh, Ajay Jajodia, Arpit Mishra, and R. David Hawkins. 2020. "Candidate Silencer Elements for the Human and Mouse Genomes." *Nature Communications*.

Driscoll, M. C., C. S. Dobkin, and B. P. Alter. 1989. "Γδβ-Thalassemia Due to a de Novo Mutation Deleting the 5' β-Globin Gene Activation-Region Hypersensitive Sites." *Proceedings of the National Academy of Sciences of the United States of America*.

Dror, Iris et al. 2015. "A Widespread Role of the Motif Environment in Transcription Factor Binding across Diverse Protein Families." *Genome Research*.

Dunham, Ian et al. 2012. "An Integrated Encyclopedia of DNA Elements in the Human Genome." *Nature*.

Espinosa, Joaquin M., and Beverly M. Emerson. 2001. "Transcriptional Regulation by P53 through Intrinsic DNA/Chromatin Binding and Site-Directed Cofactor Recruitment." *Molecular Cell*.

Farla, Pascal et al. 2004. "The Androgen Receptor Ligand-Binding Domain Stabilizes DNA Binding in Living Cells." *Journal of Structural Biology*.

Feingold, E. A. et al. 2004. "The ENCODE (ENCyclopedia of DNA Elements) Project." *Science*.

Ferrari, Karin J. et al. 2014. "Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity." *Molecular Cell*.

Fisher, William W. et al. 2012. "DNA Regions Bound at Low Occupancy by Transcription Factors Do Not Drive Patterned Reporter Gene Expression in Drosophila." *Proceedings of the National Academy of Sciences of the United States of America*.

Forghani, Reza et al. 2001. "A Distal Upstream Enhancer from the Myelin Basic Protein Gene Regulates Expression in Myelin-Forming Schwann Cells." *Journal of Neuroscience*.

Forrester, William C. et al. 1987. "Evidence for a Locus Activation Region: The Formation of Developmentally Stable Hypersensitive Sites in Globin-Expressing Hybrids." *Nucleic Acids Research*.

Fu, Maofu et al. 2000. "P300 and P300/CAMP-Response Element-Binding Protein-Associated Factor Acetylate the Androgen Receptor at Sites Governing Hormone-Dependent Transactivation." *Journal of Biological Chemistry*.

Fu, Maofu et al. 2003. "Acetylation of Androgen Receptor Enhances Coactivator Binding and Promotes Prostate Cancer Cell Growth." *Molecular and Cellular Biology*.

Fulco, Charles P. et al. 2016. "Systematic Mapping of Functional Enhancer-Promoter Connections with CRISPR Interference." *Science*.

Furlong, Eileen E.M., and Michael Levine. 2018. "Developmental Enhancers and Chromosome Topology." *Science*.

Gasperini, Molly, Jacob M. Tome, and Jay Shendure. 2020. "Towards a Comprehensive Catalogue of Validated and Target-Linked Human Enhancers." *Nature Reviews Genetics*.

Gaszner, Miklos, and Gary Felsenfeld. 2006. "Insulators: Exploiting Transcriptional and Epigenetic Mechanisms." *Nature Reviews Genetics*.

Giresi, Paul G. et al. 2007. "FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) Isolates Active Regulatory Elements from Human Chromatin." *Genome Research*.

Godowski, Paul J., Sandro Rusconi, Roger Miesfeld, and Keith R. Yamamoto. 1987. "Glucocorticoid Receptor Mutants That Are Constitutive Activators of Transcriptional Enhancement." *Nature*.

Goi, Chin Lui, Peter Little, and Chao Xie. 2013. "Cell-Type and Transcription Factor Specific Enrichment of Transcriptional Cofactor Motifs in ENCODE ChIP-Seq Data." *BMC Genomics*.

Green, Stephen, and Pierre Chambon. 1987. "Oestradiol Induction of a Glucocorticoid-Responsive Gene by a Chimaeric Receptor." *Nature*.

Grosveld, Frank, Greet Blom van Assendelft, David R. Greaves, and George Kollias. 1987. "Position-Independent, High-Level Expression of the Human β-Globin Gene in Transgenic Mice." *Cell*.

Guenther, Matthew G. et al. 2007. "A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells." *Cell*.

Guiochon-Mantel, Anne, Karine Delabre, Pierre Lescop, and Edwin Milgrom. 1996. "Intracellular Traffic of Steroid Hormone Receptors." In *Journal of Steroid Biochemistry and Molecular Biology*,

Haelens, Annemie et al. 2003. "DNA Recognition by the Androgen Receptor: Evidence for an Alternative DNA-Dependent Dimerization, and an Active Role of Sequences Flanking the Response Element on Transactivation." *Biochemical Journal*.

Haeussler, Maximilian et al. 2016. "Evaluation of Off-Target and on-Target Scoring Algorithms and Integration into the Guide RNA Selection Tool CRISPOR." *Genome Biology*.

Hakim, Ofir et al. 2009. "Glucocorticoid Receptor Activation of the Ciz1-Lcn2 Locus by Long Range Interactions." *Journal of Biological Chemistry*.

Hampsey, Michael. 1998. "Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery." *Microbiology and Molecular Biology Reviews*.

Hansen, Anders S. 2020. "CTCF as a Boundary Factor for Cohesin-Mediated Loop Extrusion: Evidence for a Multi-Step Mechanism." *Nucleus*.

He, Housheng Hansen et al. 2010. "Nucleosome Dynamics Define Transcriptional Enhancers." *Nature Genetics*.

Heidenreich, Axel et al. 2014. "EAU Guidelines on Prostate Cancer. Part II: Treatment of Advanced, Relapsing, and Castration-Resistant Prostate Cancer." *European Urology*.

Heinz, Sven et al. 2010. "Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities." *Molecular Cell*.

Herz, Hans Martin, Deqing Hu, and Ali Shilatifard. 2014. "Enhancer Malfunction in Cancer." *Molecular Cell*.

Hesselberth, Jay R. et al. 2009. "Global Mapping of Protein-DNA Interactions in Vivo by Digital Genomic Footprinting." *Nature Methods*.

Hoffman, Jackson A., Kevin W. Trotter, James M. Ward, and Trevor K. Archer. 2018. "BRG1 Governs Glucocorticoid Receptor Interactions with Chromatin and Pioneer Factors across the Genome." *eLife*.

Horoszewicz, Julius S. et al. 1983. "LNCaP Model of Human Prostatic Carcinoma." *Cancer Research*.

Ikuta, T., and Yuet Wai Kan. 1991. "In Vivo Protein-DNA Interactions at the β-Globin Gene Locus." *Proceedings of the National Academy of Sciences of the United States of America*.

Isikbay, Masis et al. 2014. "Glucocorticoid Receptor Activity Contributes to Resistance to Androgen-Targeted Therapy in Prostate Cancer." *Hormones and Cancer*.

Jabbari, Kamel, Maharshi Chakraborty, and Thomas Wiehe. 2019. "DNA Sequence-Dependent Chromatin Architecture and Nuclear Hubs Formation." *Scientific Reports*.

Jacob, François, and Jacques Monod. 1961. "Genetic Regulatory Mechanisms in the Synthesis of Proteins." *Journal of Molecular Biology*.

Jenuwein, T., and C. D. Allis. 2001. "Translating the Histone Code." *Science*.

Jerković, Ivana et al. 2017. "Genome-Wide Binding of Posterior HOXA/D Transcription Factors Reveals Subgrouping and Association with CTCF." *PLoS Genetics*.

Jin, Fulai et al. 2013. "A High-Resolution Map of the Three-Dimensional Chromatin Interactome in Human Cells." *Nature*.

Jin, Qihuang et al. 2011. "Distinct Roles of GCN5/PCAF-Mediated H3K9ac and CBP/P300-Mediated H3K18/27ac in Nuclear Receptor Transactivation." *EMBO Journal*.

Jinek, Martin et al. 2012. "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity." *Science*.

John, Sam et al. 2011. "Chromatin Accessibility Pre-Determines Glucocorticoid Receptor Binding Patterns." *Nature genetics*.

Johnson, David S., Ali Mortazavi, Richard M. Myers, and Barbara Wold. 2007. "Genome-Wide Mapping of in Vivo Protein-DNA Interactions." *Science*.

Johnson, Thomas A et al. 2018. "Conventional and Pioneer Modes of Glucocorticoid Receptor Interaction with Enhancer Chromatin in Vivo." *Nucleic Acids Research*.

Jolma, Arttu et al. 2015. "DNA-Dependent Formation of Transcription Factor Pairs Alters Their Binding Specificity." *Nature*.

Joseph, Roy et al. 2010. "Integrative Model of Genomic Factors for Determining Binding Site Selection by Estrogen Receptor-α." *Molecular Systems Biology*.

Jozwik, Kamila M. et al. 2016. "FOXA1 Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3." *Cell Reports*.

Kagey, Michael H. et al. 2010. "Mediator and Cohesin Connect Gene Expression and Chromatin Architecture." *Nature*.

Kasper, Lawryn H. et al. 2014. "Genome-Wide and Single-Cell Analyses Reveal a Context Dependent Relationship between CBP Recruitment and Gene Expression." *Nucleic Acids Research*.

Kent, W. J. et al. 2010. "BigWig and BigBed: Enabling Browsing of Large Distributed Datasets." *Bioinformatics*.

Khan, Aziz et al. 2018. "JASPAR 2018: Update of the Open-Access Database of Transcription Factor Binding Profiles and Its Web Framework." *Nucleic Acids Research*.

Klann, Tyler S. et al. 2017. "CRISPR-Cas9 Epigenome Editing Enables High-Throughput Screening for Functional Regulatory Elements in the Human Genome." *Nature Biotechnology*.

Klein, Gordon L. 2015. "The Effect of Glucocorticoids on Bone and Muscle." *Osteoporosis and Sarcopenia*.

Kornberg, Roger D. 1974. "Chromatin Structure: A Repeating Unit of Histones and DNA." *Science*.

Kribelbauer, Judith F., Chaitanya Rastogi, Harmen J. Bussemaker, and Richard S. Mann. 2019. "Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes." *Annual Review of Cell and Developmental Biology*.

de Kumar, Bony D. et al. 2017. "Dynamic Regulation of Nanog and Stem Cell-Signaling Pathways by Hoxa1 during Early Neuro-Ectodermal Differentiation of ES Cells." *Proceedings of the National Academy of Sciences of the United States of America*.

Lambert, Samuel A. et al. 2018a. "The Human Transcription Factors." *Cell*.

Langlais, David, Catherine Couture, Aurélio Balsalobre, and Jacques Drouin. 2012. "The Stat3/GR Interaction Code: Predictive Value of Direct/Indirect DNA Recruitment for Transcription Outcome." *Molecular Cell*.

Langmead, Ben, and Steven L. Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." *Nature Methods*.

Lavery, Derek N., and Iain J. McEwan. 2005. "Structure and Function of Steroid Receptor AF1 Transactivation Domains: Induction of Active Conformations." *Biochemical Journal*.

Lee, Kenneth K., and Jerry L. Workman. 2007. "Histone Acetyltransferase Complexes: One Size Doesn't Fit All." *Nature Reviews Molecular Cell Biology*.

Lempiäinen, Joanna K. et al. 2017. "Agonist-Specific Protein Interactomes of Glucocorticoid and Androgen Receptor as Revealed by Proximity Mapping." *Molecular and Cellular Proteomics*.

Leszczynski, D. E., and R. M. Schafer. 1990. "Nonspecific and Metabolic Interactions between Steroid Hormones and Human Plasma Lipoproteins." *Lipids*.

Levo, Michal et al. 2015. "Unraveling Determinants of Transcription Factor Binding Outside the Core Binding Site." *Genome Research*.

Li, Guohong, and Danny Reinberg. 2011. "Chromatin Higher-Order Structures and Gene Regulation." *Current Opinion in Genetics and Development*.

Li, Heng et al. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics*.

Li, Qiliang et al. 1991. "Primary Structure of the Goat β-Globin Locus Control Region." *Genomics*.

Li, Xiao Yong et al. 2008. "Transcription Factors Bind Thousands of Active and Inactive Regions in the Drosophila Blastoderm." *PLoS Biology*.

Liang, Gangning et al. 2004. "Distinct Localization of Histone H3 Acetylation and H3-K4 Methylation to the Transcription Start Sites in the Human Genome." *Proceedings of the National Academy of Sciences of the United States of America*.

Lieberman-Aiden, Erez et al. 2009. "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome." *Science*.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology*.

Luger, Karolin et al. 1997. "Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution." *Nature*.

Luisi, B. F. et al. 1991. "Crystallographic Analysis of the Interaction of the Glucocorticoid Receptor with DNA." *Nature*.

Malik, Sohail, and Robert G. Roeder. 2010. "The Metazoan Mediator Co-Activator Complex as an Integrative Hub for Transcriptional Regulation." *Nature Reviews Genetics*.

Mangelsdorf, David J. et al. 1995. "The Nuclear Receptor Superfamily: The Second Decade." *Cell*.

Margot, Jean B., G. William Demers, and Ross C. Hardison. 1989. "Complete Nucleotide Sequence of the Rabbit β-like Globin Gene Cluster. Analysis of Intergenic Sequences and Comparison with the Human β-like Globin Gene Cluster." *Journal of Molecular Biology*.

McEwan, Iain J, and Albert O Brinkmann. 2000. Endotext *Androgen Physiology: Receptor and Metabolic Disorders*.

McLeay, Robert C., and Timothy L. Bailey. 2010. "Motif Enrichment Analysis: A Unified Framework and an Evaluation on ChIP Data." *BMC Bioinformatics*.

Meijsing, Sebastiaan H et al. 2007. "The Ligand Binding Domain Controls Glucocorticoid Receptor Dynamics Independent of Ligand Release." *Molecular and cellular biology*.

Meijsing, Sebastiaan H. et al. 2009. "DNA Binding Site Sequence Directs Glucocorticoid Receptor Structure and Activity." *Science*.

Mohammed, Hisham et al. 2013. "Endogenous Purification Reveals GREB1 as a Key Estrogen Receptor Regulatory Factor." *Cell Reports*.

Mohammed, Hisham et al. 2016. "Rapid Immunoprecipitation Mass Spectrometry of Endogenous Proteins (RIME) for Analysis of Chromatin Complexes." *Nature Protocols*.

Moon, A. M., and T. J. Ley. 1990. "Conservation of the Primary Structure, Organization, and Function of the Human and Mouse β-Globin Locus-Activating Regions." *Proceedings of the National Academy of Sciences of the United States of America*.

Moreau, P. et al. 1981. "The SV40 72 Base Repair Repeat Has a Striking Effect on Gene Expression Both in SV40 and Other Chimeric Recombinants." *Nucleic Acids Research*.

Muerdter, Felix et al. 2018. "Resolving Systematic Errors in Widely Used Enhancer Activity Assays in Human Cells." *Nature Methods*.

Nenseth, Hatice Z. et al. 2014. "Distinctly Different Dynamics and Kinetics of Two Steroid Receptors at the Same Response Elements in Living Cells." *PLoS ONE*.

Nora, Elphège P. et al. 2012. "Spatial Partitioning of the Regulatory Landscape of the X-Inactivation Centre." *Nature*.

Ogryzko, Vasily v. et al. 1996. "The Transcriptional Coactivators P300 and CBP Are Histone Acetyltransferases." *Cell*.

Pascual, Gabriel, and Christopher K Glass. 2006. "Nuclear Receptors versus Inflammation: Mechanisms of Transrepression." *Trends in Endocrinology & Metabolism*.

Patwardhan, Rupali P. et al. 2012. "Massively Parallel Functional Dissection of Mammalian Enhancers in Vivo." *Nature Biotechnology*.

Phanstiel, Douglas H. et al. 2017. "Static and Dynamic DNA Loops Form AP-1-Bound Activation Hubs during Macrophage Development." *Molecular Cell*.

Pinho, Jacinta Oliveira, Mariana Matias, and Maria Manuela Gaspar. 2019. "Emergent Nanotechnological Strategies for Systemic Chemotherapy against Melanoma." *Nanomaterials*.

Pratt, William B., and David O. Toft. 1997. "Steroid Receptor Interactions with Heat Shock Protein and Immunophilin Chaperones*." *Endocrine Reviews*.

Ptashne, Mark. 1967. "Specific Binding of the λ Phage Repressor to λ DNA." *Nature*.

Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics*.

Rada-Iglesias, Alvaro et al. 2011. "A Unique Chromatin Signature Uncovers Early Developmental Enhancers in Humans." *Nature*.

Rajkumar, Arun S., Nicolas Dénervaud, and Sebastian J. Maerkl. 2013. "Mapping the Fine Structure of a Eukaryotic Promoter Input-Output Function." *Nature Genetics*.

Ramírez, Fidel et al. 2014. "DeepTools: A Flexible Platform for Exploring Deep-Sequencing Data." *Nucleic Acids Research*.

Ran, F. Ann et al. 2013. "Genome Engineering Using the CRISPR-Cas9 System." *Nature Protocols*.

Reddy, Timothy E. et al. 2009. "Genomic Determination of the Glucocorticoid Response Reveals Unexpected Mechanisms of Gene Regulation." *Genome Research*.

Richard-Foy, H., and G. L. Hager. 1987. "Sequence-Specific Positioning of Nucleosomes over the Steroid-Inducible MMTV Promoter." *The EMBO journal*.

Richards, Eric J., and Sarah C.R. Elgin. 2002. "Epigenetic Codes for Heterochromatin Formation and Silencing: Rounding up the Usual Suspects." *Cell*.

Robertson, A. Gordon et al. 2008. "Genome-Wide Relationship between Histone H3 Lysine 4 Mono- and Tri-Methylation and Transcription Factor Binding." *Genome Research*.

Robinson, Philip J.J. et al. 2008. "30 Nm Chromatin Fibre Decompaction Requires Both H4-K16 Acetylation and Linker Histone Eviction." *Journal of Molecular Biology*.

Rogatsky, I, J M Trowbridge, and M J Garabedian. 1997. "Glucocorticoid Receptor-Mediated Cell Cycle Arrest Is Achieved through Distinct Cell-Specific Transcriptional Regulatory Mechanisms." *Molecular and Cellular Biology*.

Rohs, Remo et al. 2009. "The Role of DNA Shape in Protein-DNA Recognition." *Nature*.

Rosenfeld, Michael G., Victoria v. Lunyak, and Christopher K. Glass. 2006. "Sensors and Signals: A Coactivator/Corepressor/Epigenetic Code for Integrating Signal-Dependent Programs of Transcriptional Response." *Genes and Development*.

Sacta, Maria A. et al. 2018. "Gene-Specific Mechanisms Direct Glucocorticoid-Receptor-Driven Repression of Inflammatory Response Genes in Macrophages." *eLife*.

Sahu, Biswajyoti et al. 2013. "FoxA1 Specifies Unique Androgen and Glucocorticoid Receptor Binding Events in Prostate Cancer Cells." *Cancer research*.

Sahu, Biswajyoti et al. 2014. "Androgen Receptor Uses Relaxed Response Element Stringency for Selective Chromatin Binding and Transcriptional Regulation in Vivo." *Nucleic acids research*.

Sandelin, Albin et al. 2004. "JASPAR: An Open-Access Database for Eukaryotic Transcription Factor Binding Profiles." *Nucleic Acids Research*.

Sasaki, S., K. Ishibashi, and F. Marumo. 1998. "Aquaporin-2 and -3: Representatives of Two Subgroups of the Aquaporin Family Colocalized in the Kidney Collecting Duct." *Annual Review of Physiology*.

Schauwaers, K. et al. 2007. "Loss of Androgen Receptor Binding to Selective Androgen Response Elements Causes a Reproductive Phenotype in a Knockin Mouse Model." *Proceedings of the National Academy of Sciences*.

Schiller, Benjamin J. et al. 2014. "Glucocorticoid Receptor Binds Half Sites as a Monomer and Regulates Specific Target Genes." *Genome Biology*.

Schmidt, Søren Fisker, Bjørk Ditlev Larsen, Anne Loft, and Susanne Mandrup. 2016. "Cofactor Squelching: Artifact or Fact?" *BioEssays*.

Schneider, Thomas D., and R. Michael Stephens. 1990. "Sequence Logos: A New Way to Display Consensus Sequences." *Nucleic Acids Research*.

Schoenfelder, Stefan et al. 2018. "Promoter Capture Hi-C: High-Resolution, Genome-Wide Profiling of Promoter Interactions." *Journal of Visualized Experiments*.

Schoenmakers, Erik et al. 2000. "Differences in DNA Binding Characteristics of the Androgen and Glucocorticoid Receptors Can Determine Hormone-Specific Responses." *Journal of Biological Chemistry*.

Schöne, Stefanie et al. 2018. "Synthetic STARR-Seq Reveals How DNA Shape and Sequence Modulate Transcriptional Output and Noise." *PLoS Genetics*.

Schones, Dustin E. et al. 2008. "Dynamic Regulation of Nucleosome Positioning in the Human Genome." *Cell*.

Schuurmans, Alex L.G. et al. 1988. "Regulation of Growth and Epidermal Growth Factor Receptor Levels of LNCaP Prostate Tumor Cells by Different Steroids." *International Journal of Cancer*.

Schwabe, John W.R., Lynda Chapman, John T. Finch, and Daniela Rhodes. 1993. "The Crystal Structure of the Estrogen Receptor DNA-Binding Domain Bound to DNA: How Receptors Discriminate between Their Response Elements." *Cell*.

Shaffer, Paul L. et al. 2004. "Structural Basis of Androgen Receptor Binding to Selective Androgen Response Elements." *Proceedings of the National Academy of Sciences of the United States of America*.

Shen, Ning et al. 2018. "Divergence in DNA Specificity among Paralogous Transcription Factors Contributes to Their Differential In Vivo Binding." *Cell Systems* 6(4): 470-483.e8.

Shlyueva, Daria, Gerald Stampfel, and Alexander Stark. 2014a. "Transcriptional Enhancers: From Properties to Genome-Wide Predictions." *Nature Reviews Genetics*.

Simeonov, Dimitre R. et al. 2017. "Discovery of Stimulation-Responsive Immune Enhancers with CRISPR Activation." *Nature*.

Simon, Jeremy M., Paul G. Giresi, Ian J. Davis, and Jason D. Lieb. 2013. "A Detailed Protocol for Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE)." *Current Protocols in Molecular Biology*.

Sims, Robert J., Kenichi Nishioka, and Danny Reinberg. 2003. "Histone Lysine Methylation: A Signature for Chromatin Function." *Trends in Genetics*.

Singh, Larry N., and Sridhar Hannenhalli. 2008. "Functional Diversification of Paralogous Transcription Factors via Divergence in DNA Binding Site Motif and in Expression." *PLoS*.

Singh, Rajan et al. 2006. "Testosterone Inhibits Adipogenic Differentiation in 3T3-L1 Cells: Nuclear Translocation of Androgen Receptor Complex with β-Catenin and T-Cell Factor 4 May Bypass Canonical Wnt Signaling to down-Regulate Adipogenic Transcription Factors." *Endocrinology*.

Sinha-Hikim, Indrani et al. 2004. "Androgen Receptor in Human Skeletal Muscle and Cultured Muscle Satellite Cells: Up-Regulation by Androgen Treatment." *Journal of Clinical Endocrinology and Metabolism*.

Slattery, Matthew et al. 2011. "Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins." *Cell*.

Smith, Tom, Andreas Heger, and Ian Sudbery. 2017. "UMI-Tools: Modeling Sequencing Errors in Unique Molecular Identifiers to Improve Quantification Accuracy." *Genome Research*.

Soutourina, Julie. 2018. "Transcription Regulation by the Mediator Complex." *Nature Reviews Molecular Cell Biology*.

Splinter, Erik, and Wouter de Laat. 2011. "The Complex Transcription Regulatory Landscape of Our Genome: Control in Three Dimensions." *EMBO Journal*.

Stallcup, Michael R., and Coralie Poulard. 2020. "Gene-Specific Actions of Transcriptional Coregulators Facilitate Physiological Plasticity: Evidence for a Physiological Coregulator Code." *Trends in Biochemical Sciences*.

Stelloo, S. et al. 2018. "Endogenous Androgen Receptor Proteomic Profiling Reveals Genomic Subcomplex Involved in Prostate Tumorigenesis." *Oncogene*.

Stormo, Gary D., and Yue Zhao. 2010. "Determining the Specificity of Protein-DNA Interactions." *Nature Reviews Genetics*.

Symmons, Orsolya et al. 2016. "The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances." *Developmental Cell*.

Telorac, Jonas et al. 2016. "Identification and Characterization of DNA Sequences That Prevent Glucocorticoid Receptor Binding to Nearby Response Elements." *Nucleic Acids Research*.

Tenbaum, S., and A. Baniahmad. 1997. "Nuclear Receptors: Structure, Function and Involvement in Disease." *International Journal of Biochemistry and Cell Biology*.

Thakore, Pratiksha I. et al. 2015. "Highly Specific Epigenome Editing by CRISPR-Cas9 Repressors for Silencing of Distal Regulatory Elements." *Nature Methods*.

Thormann, Verena et al. 2018. "Genomic Dissection of Enhancers Uncovers Principles of Combinatorial Regulation and Cell Type-Specific Wiring of Enhancer-Promoter Contacts." *Nucleic Acids Research*.

Thormann, Verena et al. 2019. "Expanding the Repertoire of Glucocorticoid Receptor Target Genes by Engineering Genomic Response Elements." *Life Science Alliance*.

Thurman, Robert E. et al. 2012. "The Accessible Chromatin Landscape of the Human Genome." *Nature*.

Travers, Andrew. 1999. "An Engine for Nucleosome Remodeling." *Cell*.

Tuan, D., W. Solomon, Q. Li, and I. M. London. 1985. "The 'β-like-Globin' Gene Domain in Human Erythroid Cells." *Proceedings of the National Academy of Sciences of the United States of America*.

Umesono, Kazuhiko, and Ronald M. Evans. 1989. "Determinants of Target Gene Specificity for Steroid/Thyroid Hormone Receptors." *Cell*.

Vanderschueren, Dirk et al. 2014. "Sex Steroid Actions in Male Bone." *Endocrine Reviews*.

Vaquerizas, Juan M., Sarah K. Kummerfeld, Sarah A. Teichmann, and Nicholas M. Luscombe. 2009. "A Census of Human Transcription Factors: Function, Expression and Evolution." *Nature Reviews Genetics*.

Vermeer, Harry et al. 2003. "Glucocorticoid-Induced Increase in Lymphocytic FKBP51 Messenger Ribonucleic Acid Expression: A Potential Marker for Glucocorticoid Sensitivity, Potency, and Bioavailability." *Journal of Clinical Endocrinology and Metabolism*.

Visel, Axel et al. 2009. "ChIP-Seq Accurately Predicts Tissue-Specific Activity of Enhancers." *Nature*.

Visel, Axel, Edward M. Rubin, and Len A. Pennacchio. 2009. "Genomic Views of Distant-Acting Enhancers." *Nature*.

Vockley, Christopher M. et al. 2016. "Direct GR Binding Sites Potentiate Clusters of TF Binding across the Human Genome." *Cell*.

Wang, Jie et al. 2012. "Sequence Features and Chromatin Structure around the Genomic Regions Bound by 119 Human Transcription Factors." *Genome Research*.

Wang, Siyuan et al. 2016. "Spatial Organization of Chromatin Domains and Compartments in Single Chromosomes." *Science*.

Wärnmark, Anette, Eckardt Treuter, Anthony P.H. Wright, and Jan Åke Gustafsson. 2003. "Activation Functions 1 and 2 of Nuclear Receptors: Molecular Strategies for Transcriptional Activation." *Molecular Endocrinology*.

Watson, Lisa C. et al. 2013. "The Glucocorticoid Receptor Dimer Interface Allosterically Transmits Sequence-Specific DNA Signals." *Nature Structural and Molecular Biology*.

Weinert, Brian T. et al. 2018. "Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/P300 Acetylome." *Cell*.

Weingarten-Gabbay, Shira et al. 2019. "Systematic Interrogation of Human Promoters." *Genome Research*.

Weirauch, Matthew T., and T. R. Hughes. 2014. "A Catalogue of Eukaryotic Transcription Factor Types, Their Evolutionary Origin, and Species Distribution." *Sub-Cellular Biochemistry*.

Werner, Torsten et al. 2007. "Multiple Conserved Regulatory Elements with Overlapping Functions Determine Sox10 Expression in Mouse Embryogenesis." *Nucleic Acids Research*.

White, Michael A., Connie A. Myers, Joseph C. Corbo, and Barak A. Cohen. 2013. "Massively Parallel in Vivo Enhancer Assay Reveals That Highly Local Features Determine the Cis-Regulatory Function of ChIP-Seq Peaks." *Proceedings of the National Academy of Sciences of the United States of America*.

Wiench, Malgorzata, Tina B. Miranda, and Gordon L. Hager. 2011. "Control of Nuclear Receptor Function by Local Chromatin Structure." *FEBS Journal*.

Wu, Jennifer N. et al. 2015. "Functionally Distinct Patterns of Nucleosome Remodeling at Enhancers in Glucocorticoid-Treated Acute Lymphoblastic Leukemia." *Epigenetics and Chromatin*.

Xie, Shiqi et al. 2017. "Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells." *Molecular Cell*.

Yamaguchi, Yuki, Hirotaka Shibata, and Hiroshi Handa. 2013. "Transcription Elongation Factors DSIF and NELF: Promoter-Proximal Pausing and Beyond." *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*.

Yáñez-Cuna, J. Omar et al. 2012. "Uncovering Cis-Regulatory Sequence Requirements for Context-Specific Transcription Factor Binding." *Genome Research*.

Zaret, Kenneth S., and Jason S. Carroll. 2011. "Pioneer Transcription Factors: Establishing Competence for Gene Expression." *Genes and Development*.

Zentner, Gabriel E., Paul J. Tesar, and Peter C. Scacheri. 2011. "Epigenetic Signatures Distinguish Multiple Classes of Enhancers with Distinct Cellular Functions." *Genome Research*.

Zhan, Xiaowei, and Dajiang J. Liu. 2015. "SEQMINER: An R-Package to Facilitate the Functional Interpretation of Sequence-Based Associations." *Genetic Epidemiology*.

Zhang, Liyang et al. 2018. "SelexGLM Differentiates Androgen and Glucocorticoid Receptor DNA-Binding Preference over an Extended Binding Site." *Genome research*.

Zhang, Yong et al. 2008. "Model-Based Analysis of ChIP-Seq (MACS)." *Genome Biology*.

Zhao, J. C. et al. 2016. "FOXA1 Acts Upstream of GATA2 and AR in Hormonal Regulation of Gene Expression." *Oncogene*.

Zhou, Zhifeng, Jeffry L. Corden, and Terry R. Brown. 1997. "Identification and Characterization of a Novel Androgen Response Element Composed of a Direct Repeat." *Journal of Biological Chemistry*.

Zhu, Lihua J. et al. 2010. "ChIPpeakAnno: A Bioconductor Package to Annotate ChIP-Seq and ChIP-Chip Data." *BMC Bioinformatics*.