

Genomic Deletion of Enhancers Uncovers Principles of Combinatorial Regulation and Cell Type-Specific Gene Expression

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

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

by
Verena Thormann
from
Konstanz
2017



The dissertation was prepared under the supervision of Dr. Sebastiaan H. Meijnsing at the Max Planck Institute for Molecular Genetics in Berlin from July 2014 to June 2017.
Berlin, June 22nd 2017

1st referee: Dr. Sebastiaan H. Meijnsing

2nd referee: Prof. Dr. Markus Wahl

Date of Disputation: 27.10.2017

Acknowledgements

First of all, I would like to express my gratitude to Sebastiaan Meijsing for his continuous support and guidance throughout my PhD. Sebastiaan had always an open door for questions and was a source of excellent advice, but at the same time gave me the freedom to develop and pursue my own ideas.

I would like to thank Martin Vingron, who gave me the opportunity to join his research team and to work together with the talented bioinformaticians that are part of his group. Here, I would like to especially mention Robert Schöpflin, whose computational analysis contributed a lot to this work.

My special gratitude goes to Maika Rothkegel and Edda Einfeldt for their great help and technical assistance in the lab. Without their support a lot of experiments would have taken many more years. In addition, I would like to thank Martin Franke and Ivana Jerkovic for sharing their knowledge and technical experience in chromosome conformation capture and chromatin immunoprecipitation.

I would like to thank the entire thesis committee and especially Prof. Markus Wahl for taking the time to review this work.

A big thank you goes to the administrative staff at the MPI, foremost to Martina Lorse, who was always ready to help with administrative and organizational issues.

Stefanie Schöne, Marina Borschiwer, Laura Glaser, Melissa Bothe, Nilofar Abdavi Azar and all present and former members of the Vingron group: Thank you for all the coffee breaks, for sharing the ups and downs of the PhD life and for all the fun we had in the last years!

I am grateful to my parents (all three of them) the biggest support in my life, to Amélie Böhm continuously joining me along the way from school to PhD and finally, I would like to thank Timo Heroth for always providing a shoulder to lean on when I needed it the most.

Zusammenfassung

Die Bindung von Transkriptionsfaktoren (TF) an genomische Enhancer-Elemente ist elementar für die Regulation von Genen. Bis heute ist es jedoch weitgehend unklar, welche Bindungsereignisse zur Regulation eines spezifischen Zielgenes beitragen, wie unterschiedliche Genexpressionsmuster zelltypspezifisch reguliert und welche minimalen Regulationsanordnungen benötigt werden, um ein Gen TF-abhängig zu regulieren. In meiner Doktorarbeit verwende ich als Modellsystem den Glukokortikoidrezeptor (GR), einen hormonaktivierbaren TF, um die molekularen Mechanismen zu untersuchen, die die funktionelle Rolle von Enhancer-Elementen beeinflussen.

Durch die genomische Zerstörung von einzelnen und mehreren GR-Bindestellen (GBS) des *GILZ* Enhancers, habe ich das regulatorische Zusammenspiel von mehreren TF-Bindestellen systematisch untersucht. Diese Mutationsanalyse zeigte, dass mehrere GBS zwar voneinander unabhängig durch GR gebunden werden, aber dennoch kooperativ als funktionelle Einheit die Genexpression beeinflussen, so lange alle beinhalteten GBS intakt sind. Durch die genomische Zerstörung einer GBS, die in zwei verschiedenen Zelltypen von GR gebunden wird, konnte ich zudem zeigen, dass zelltypspezifische Unterschiede in der dreidimensionalen (3D) Genomorganisation und Enhancer-Blocking Enhancer-Promoter Kontakte neu verbinden kann. Die Verbindung unterschiedlicher Enhancer-Promoter Kontakte ermöglicht die Expression verschiedener Transkriptvarianten eines Gens und kann dadurch zu den zelltypspezifischen Effekten von Glukokortikoiden beitragen.

Zudem habe ich den Effekt von DNA Motifsequenzen auf die Aktivität von GR untersucht, indem ich die Sequenz einer GBS des *GILZ* Enhancers in verschiedene GBS-Motifsequenzen umgewandelt habe. Während in Reporter Assays der Austausch von GBS-Motifvarianten in quantitativen Unterschieden in der Genexpression resultierte, zeigte sich im genomischen Kontext kein Effekt auf die Regulation von *GILZ*. Demzufolge, kann der genomische Kontext die regulatorische Wirkung von GBS-Motifvarianten beeinflussen, z.B. durch die Integration von regulatorischer Information von mehreren GBS. Um GBS-Motifvarianten in einem isolierten, endogenen Umfeld zu untersuchen, habe ich eine einzelne GBS an der Promoter-Region eines endogen nicht exprimierten Gens platziert. Diese Integration zeigt, dass die Präsenz einer einzelnen GBS ausreichen kann, um ein Gen GR-abhängig zu regulieren und stellt zudem ein Modellsystem dar, um die Rolle von GBS-Motifvarianten in Isolation zu testen.

Zusammenfassend kann die genomische Editierung Einblicke in die Funktionsweise der kombinatorischen Regulation durch mehrere TF-Bindestellen ermöglichen und zeigt zudem, dass eine einzige Promoter-proximale GBS ausreichen kann, um ein Gen GR-abhängig zu regulieren. Zusätzlich zeigt diese Arbeit, dass eine GBS, die in zwei verschiedenen Zelltypen gleichermaßen gebunden wird, durch Unterschiede in der 3D Genomorganisation und Enhancer-Blocking zur zelltypspezifischen Genexpression beitragen kann.

Abstract

Transcription factors (TFs) are fundamental to the regulation of genes by binding to genomic enhancer elements and orchestrating the expression of their target genes. However, it is largely unclear which TF binding event(s) contribute to the regulation of a specific gene, how cell type-specific plasticity in gene expression is achieved and what minimal circuitry is required to regulate a gene depending on the activity of a specific TF. Here, I used the glucocorticoid receptor (GR), a hormone-activated TF, as a model system to study the molecular mechanisms that determine the functional role of enhancers.

By genomically deleting, either alone or in combination, multiple GR binding sites (GBSs) located within the *GILZ* enhancer, I systematically investigated the interplay of multiple TF binding sites. This mutational analysis demonstrated that multiple GBSs are bound independently but can act cooperatively on gene expression as a single functional unit, which is only active when all of its GBSs are intact.

Furthermore, the deletion of GBSs shared between two different cell types demonstrates how cell type-specific differences in the three-dimensional (3D) genome organization and enhancer blocking can rewire enhancer-promoter contacts. This rewiring enables a GBS bound in two different cell types to direct the expression of distinct transcript variants, thereby contributing to the cell type-specific consequences of glucocorticoid signaling.

Finally, I investigated the effect of DNA motif sequence on GR activity, by exchanging the sequence of a single GBS of the *GILZ* enhancer into different GBS motif variants. Whereas in reporter assays this exchange resulted in quantitative changes in gene expression, no effect was observed upon exchange in the endogenous context. Hence, the genomic context can influence the regulatory potency of individual GBS variants, for example by integrating regulatory information from multiple GBSs. To investigate the role GBS variants in an isolated endogenous context, I integrated a GBS at the promoter region of an endogenously silenced gene, thereby activating its expression in a GR-dependent manner. This demonstrates that a single GBS can be sufficient to induce GR-mediated regulation of an associated gene and further provides a model system to investigate the effect of GBS sequence variants in isolation.

Together, genomic editing of GBSs in their endogenous genomic context enables insights into the operating principles of combinatorial gene regulation by multiple TF binding sites, but also demonstrates that a single promoter-proximal GBS is sufficient to induce GR-dependent gene expression. Furthermore, a GBS equally bound in two different cell types can contribute to the establishment of cell type-specific gene expression patterns by differences in 3D genome organization and enhancer blocking.

Contents

1	Introduction	1
1.1	Transcriptional regulation of gene expression	1
1.2	The regulatory activity of a genomic transcription factor binding site is influenced by its context	3
1.2.1	The influence of epigenetic context	4
1.2.2	Combinatorial regulation by multiple TFBSs	5
1.2.3	The influence of genome architecture	6
1.3	Genome editing using CRISPR/Cas9	9
1.4	The glucocorticoid receptor: a model system to study the transcriptional regulation of gene expression	12
1.5	Allosteric regulation of glucocorticoid receptor activity	14
1.6	Aim of this thesis	15
2	Materials and Methods	17
2.1	Materials	17
2.1.1	Chemicals	17
2.1.2	Cell lines	17
2.1.3	Antibodies	17
2.1.4	Plasmids	18
2.1.5	Media	18
2.1.6	Buffers	19
2.2	Methods	21
2.2.1	Cell culture	21
2.2.2	Polymerase chain reaction	21
2.2.3	Quantitative polymerase chain reaction	22
2.2.4	Site-directed mutagenesis	22
2.2.5	Genomic editing by CRISPR-Cas9	23
2.2.6	Quantification of hormone-induced gene expression	26
2.2.7	Activation of gene expression by dCas9-VP64 SAM	26
2.2.8	Luciferase reporter assays	27
2.2.9	Chromatin immunoprecipitation	29
2.2.10	Circularized chromosome conformation capture	32

2.2.11	siRNA knock-down of CTCF	36
2.2.12	Fluorescence <i>in situ</i> hybridization	36
2.2.13	Computational prediction of transcription factor affinity	37
3	Results	39
3.1	Part 1: Genomic deletion of glucocorticoid receptor binding sites	39
3.1.1	Genomic editing of GBSs located upstream of GR target genes . . .	39
3.1.2	Genomic deletion of GBSs influences the regulation of nearby genes	42
3.1.3	Multiple promoter-proximal GBSs cooperatively regulate <i>GILZ</i> expression	43
3.1.4	Genomic deletion of GBSs does not influence GR binding at neighboring sites	46
3.1.5	Reporter gene assays cannot recapitulate the genomic cooperativity of promoter-proximal <i>GILZ</i> GBSs	48
3.1.6	The regulatory activity of <i>GILZ</i> GBS1 differs between cell types . .	49
3.1.7	<i>GILZ</i> GBS1 regulates the expression of <i>GILZ</i> transcript variants in a cell type-specific manner	52
3.1.8	The pattern of H3K27Ac at the <i>GILZ</i> locus	55
3.1.9	The cell type-specific regulatory activity of <i>GILZ</i> GBS1 requires its endogenous genomic context	57
3.1.10	The cell type-specific regulatory activity of <i>GILZ</i> GBS1 coincides with differential CTCF binding and long-range interactions	59
3.1.11	Knock-down of CTCF does not alter the GR-dependent regulation of <i>GILZ</i> transcript variants	62
3.2	Part 2: Genomic exchange and insertion of glucocorticoid receptor binding sites	64
3.2.1	GR motif variants differentially affect reporter gene expression . . .	64
3.2.2	Genomic exchange of a low-affinity into a high-affinity GR motif variant does not affect <i>GILZ</i> expression	66
3.2.3	The introduction of a single promoter-proximal GR binding sequence is sufficient to regulate gene expression in a GR-dependent manner .	69
4	Discussion	73
4.1	Reporter assays only partially simulate endogenous enhancer activity	73
4.2	Genomic editing of TFBSs using CRISPR/Cas9	74
4.3	Not all bound GBSs contribute to the GR-dependent regulation of nearby genes	76
4.4	Multiple GBSs cooperatively regulate <i>GILZ</i> expression	78
4.5	Regulation of cell type-specific gene expression by shared TFBSs	80

4.6	The influence of genome architecture, DNA-looping and CTCF binding on enhancer activity	82
4.7	Investigating the effect of GBS motif variants at endogenous loci	86
4.8	Conclusions	89
5	Supplement	91
5.1	Supplementary figures	91
5.2	Genomic location of TFBSs	104
5.3	gRNA sequences	104
5.4	Primer sequences	106
6	List of Abbreviations, Tables and Figures	i
7	References	vii

1 Introduction

The function and identity of each cell in the human body is determined by the correct quantitative, spatial and temporal expression of specific sets of genes. The mammalian genome contains roughly 20,000 protein-coding genes. However, within the genome, protein-coding regions represent only a small proportion of the entire genetic information [1]. For decades, the vast majority of non-coding regions in the genome were considered to represent 'junk' DNA, attributing only minor relevance for about 98% of the human genome [2]. With an advance in genetic knowledge, it became increasingly clear that non-coding regions serve a fundamental regulatory function for the expression of protein-coding genes and that defects in non-coding regions can lead to misregulation of gene expression and disease [3].

Among the numerous cellular mechanisms that regulate the correct expression of protein-coding genes and the cellular abundance of their products, the first regulatory step in gene expression occurs at the level of DNA transcription into mRNA. Given that, cellular gene expression can vary dramatically during the course of cellular development and in comparison of different tissues and cell types, their pattern of expression needs to be regulated in a complex and coordinated fashion. However, up to this date the regulatory role of the vast majority of genomic loci is still undefined and the exact molecular mechanisms that contribute to the transcriptional control of cellular gene expression patterns remain poorly understood. For example, linking a putative regulatory region to changes in the expression of a specific gene still represents a challenging problem. Therefore, investigating the molecular mechanisms that define the activity of regulatory regions and shape the complex pattern of cellular gene expression programs are the primary focus of this work.

1.1 Transcriptional regulation of gene expression

Cellular gene expression is precisely controlled at the transcriptional level by the integrated action of transcription factors (TFs) and regulatory genomic regions. Typically, regulatory regions contain clusters of short sequence-encoded DNA motifs, providing a binding platform for the direct interaction with multiple TFs. Besides direct binding to DNA, TFs can also be recruited to regulatory regions without directly contacting the DNA. Namely, by binding to other DNA-bound proteins, a mechanism referred to as DNA tethering (Fig. 1.1) [4]. In the genome of eukaryotic cells regulatory DNA elements can be located proximal to the core promoter of genes but can also be found far away from the

1. Introduction

genes they regulate for example at enhancers, silencers and insulators. Silencers can repress the transcription of a gene by recruiting repressing factors or by sterical interference with activating factors [5]. Insulators restrict the regulatory range of regulatory elements and prevent promiscuous gene regulation by blocking the interaction between regulatory elements and their target promoters [5, 6].

In contrast to silencers and insulators, active enhancers have the potential to increase the transcription of an associated gene. Therefore, one of the most crucial steps in the induction of gene expression is binding of a TF to its cognate transcription factor binding site (TFBS) located within an associated enhancer region. Upon enhancer binding, TFs can promote the initiation of transcription by recruiting co-activators and the pre-initiation complex (PIC) to the core promoter (Fig. 1.1), ultimately leading to the assembly of the transcriptional machinery and the transcription of DNA into mRNA by RNA-polymerase II. Furthermore, TFs can additionally influence the process of elongation during RNA synthesis by RNA-polymerase II and the re-initiation of transcription during multiple cycles of transcription [7, 8]. Importantly, to regulate the expression of its target gene, TFs can either bind to enhancers located proximal to the core promoter or to distal enhancer elements. Distal enhancers can interact with promoter regions by looping of the flexible DNA polymer [9, 10]. Therefore, DNA-looping can bring together stretches of DNA in a three dimensional (3D) conformation that might be far away on a linear scale (Fig. 1.1).

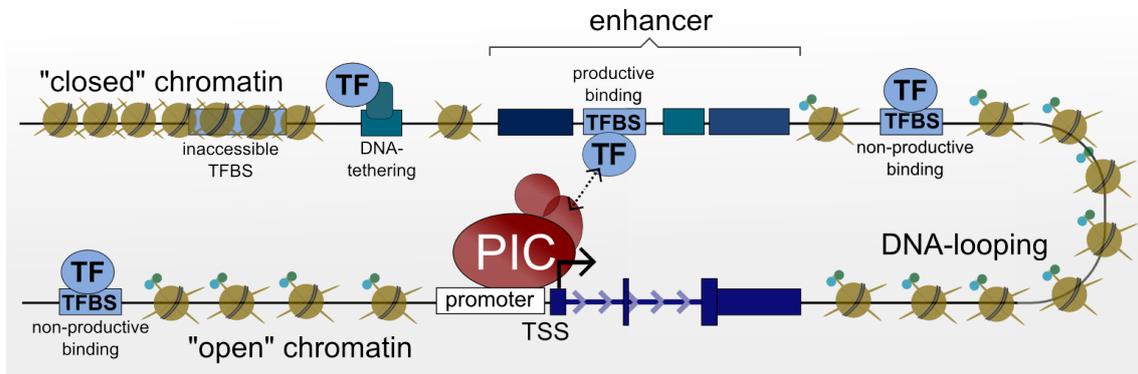


Fig. 1.1: Transcriptional regulation of gene expression. In the nucleus, the majority of bound transcription factor binding sites (TFBSs) are located within open chromatin depleted from nucleosomes. Generally, not all TFBSs bound by a given transcription factor (TF) are productive for gene regulation. TFs bound to a productive TFBS that is located at a distal enhancer can regulate gene expression by DNA-looping, bringing the transcription start site (TSS) of the target gene into close physical proximity to the bound TF. Thereby, TF-induced recruiting of co-activators and the pre-initiation complex (PIC) promotes the induction of transcription from the associated TSS.

Experimentally, genome-wide DNA binding sites of a TF of interest are detected by techniques based on chromatin immuno-precipitation (ChIP), therefore enabling the identification of the interaction of a TF and its corresponding TFBS. As a first step of the ChIP procedure, protein-DNA interactions are preserved and cross-linked by the addition

of formaldehyde. Next, the cross-linked chromatin is sheared into small fragments, e.g. by sonication and the TF of interest and its bound genomic region are co-precipitated using an antibody binding to epitopes displayed on the outer surface of the TF. Finally, these co-precipitated genomic regions bound by the TF are either identified by qPCR-based methods or next generation sequencing (NGS)(Fig. 1.2).

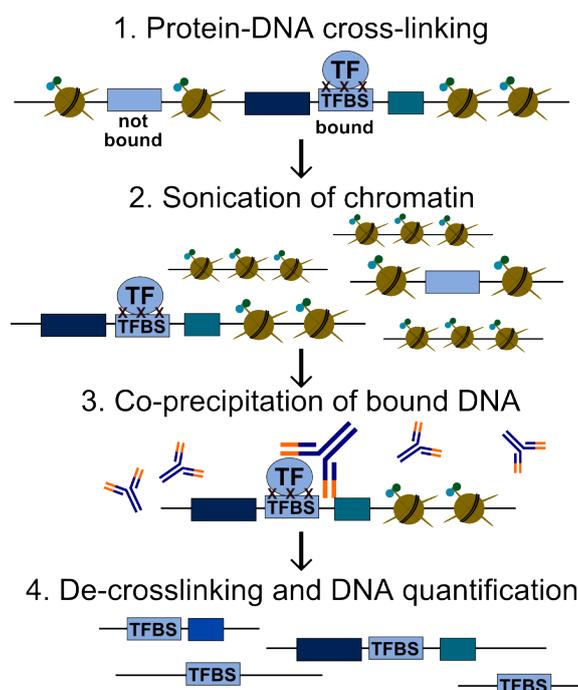


Fig. 1.2: Overview of the basic steps of the ChIP procedure. As a first step, protein-DNA interactions are cross-linked, preserving the interaction of a bound transcription factor (TF) to its cognate transcription factor binding site (TFBS). Next, the fixed chromatin is sheared by sonication into smaller fragments that can be co-precipitated using a TF-specific antibody. As a last step, the immunoprecipitated DNA fragments are de-crosslinked and quantified by qPCR or DNA-sequencing.

Together, binding of TFs to TFBSs located within enhancer regions, are a key regulatory step in the induction of gene expression. However, up to this date, only a minor proportion of all TFBSs have been biologically validated with regard to their regulatory functionality or their target gene specificity.

1.2 The regulatory activity of a genomic transcription factor binding site is influenced by its context

In the past decade, the advance in NGS-methods resulted in a wealth of available ChIP-seq data for different TFs and from a wide variety of different cell types, tissues and model organisms [1]. The analysis of these data revealed that not all potential DNA sequences

1. Introduction

predicted to be bound by a given TF are actually bound *in vivo*. In fact, only a minor proportion of all possible sequences, matching the DNA binding motif of a TF are bound in the genome of living cells.

However, even though considering the precise knowledge of TF binding as provided for example by ChIP-seq data, linking gene regulation to TF binding remains a challenging problem because of several reasons. First, given that ChIP-seq experiments usually result in several thousands of genomic peaks for an individual TF, and TF perturbations usually result in only a small number of affected genes [11, 12], the relationship between an enhancer and a specific TSS is far from definite and genomic TF binding does not necessarily imply a productive regulatory function (Fig. 1.1). Second, the analysis of TF binding patterns revealed that TF occupancy at individual TFBSs varies drastically among different cell types and depending on the cellular differentiation stage [13–16]. Nevertheless, cell type-specific patterns of TF binding can explain only some but not all observed differences in the establishment of differential gene expression patterns. Third, unlike gene-coding regions, productive TFBSs are not reliably determined by location or clear sequence features. Furthermore, given that on a linear scale enhancers can be located far away from the gene they regulate, it is impossible to explicitly define target genes solely based on proximity.

In fact, one main reason that complicates the prediction of productive TFBSs and their target genes is the influence from its endogenous genomic environment. Therefore, the next paragraph is particularly focussed on the influence from epigenetic modifications, the combinatorial regulation from multiple TFBSs and the 3D genome architecture that can shape the regulatory activity of a TFBSs.

1.2.1 The influence of epigenetic context

In the nucleus, DNA is wrapped around histone octamers, composed of four core histones (H3, H4, H2A and H2B), thereby forming nucleosomes. The formation of nucleosomes both facilitate the tight packaging of nuclear DNA but also contributes to the regulation of gene expression.

Nucleosomes are able to restrict TF binding, because many TFs are less capable to bind to DNA that is tightly wrapped around nucleosomes [17]. Therefore, the vast majority of genomic TFBSs are located in “open” nucleosome-depleted regions [18, 19] (Fig. 1.1), which can change dynamically in response to environmental signals and cellular differentiation status [20].

The connection between TF binding and nucleosome occupancy is further influenced by epigenetic modifications. Epigenetic modifications are heritable but reversible changes that occur without modification of the DNA sequence. One of the best studied epigenetic modifications are post-translational modifications (PTM) of histones, e.g. by covalent modification through phosphorylation, methylation or acetylation. A wide variety of

studies have shown that specific histone modifications are associated with distinct DNA features. For example, the acetylation of histone H3 at lysine 27 (H3K27ac) was shown to correlate with actively transcribed promoters and active enhancers [21, 22].

Besides the covalent modification of histones, the exchange of histone variants can alter nucleosome stability and their DNA binding properties [17, 23]. Furthermore, TF occupancy can also be epigenetically influenced by direct methylation of DNA at the 5' position of cytosines at CpG dinucleotides [24].

Thus, epigenetic modifications can change the regulatory activity of TFBSs on several levels, for example by altering DNA accessibility and chromatin structure. However, although the occurrence of specific chromatin marks and DNA accessibility patterns correlates with enhancer function, it is still unclear whether these represent the cause or consequence of gene regulation.

1.2.2 Combinatorial regulation by multiple TFBSs

Given that, genomic TFBSs by far outnumber the number of genes, many genes are most likely to be regulated by several enhancers. Supporting this hypothesis, many promoters were shown to interact with multiple enhancer regions [25][26]. In addition, analyzing the effect of individual TF knock-down on gene expression revealed a negative correlation to the number of TFBSs that interact with the TSS of a given gene [12]. Hence, promoters are able to integrate signals from multiple enhancers. Indeed, at the population level single nucleotide polymorphisms (SNP) within TFBSs rarely result in dramatic gene expression changes or disease phenotypes [27]; and if they do so they were reported to simultaneously arise in multiple enhancer regions [28].

Hence, cooperative regulation of gene expression by multiple TFBS and the creation of regulatory hubs might represent a common mechanism to ensure regulatory robustness, integrating signals from both remote contacts and promoter-proximal regions. However, how multiple TFBSs interact to quantitatively regulate gene expression remains largely unclear and has only been examined for a small number of loci.

Foremost, combinatorial regulation can occur at the level of DNA binding, when TF binding to one TFBS facilitates TF binding at an interacting TFBS. Such cooperative binding is most often involved with direct or indirect (via common interacting factors) protein-protein interactions, that stabilize protein-DNA contacts [29]. In this context, the concentration of a given TF in the nucleus might additionally influence the extent of cooperative interaction. Conceivably, depending on its cellular concentration the same TF can bind to different subsets of TFBSs to differentially affect their combinatorial interplay [30]. In addition, TFs can also cooperatively interact at the level of DNA binding by indirect mechanisms such as through triggering nucleosomal re-positioning by the pioneering action of another TF [19], or by inducing local bending of the DNA that may assist TF

1. Introduction

binding at a neighboring site [31].

Besides cooperation in DNA binding, TFs can also cooperatively interact with regard to their transcriptional output by at least three theoretical mechanisms (Fig. 1.3). First, TFs can act in an additive manner on transcriptional output, if their effect on gene expression is proportional to the number of bound TFBSs [32, 33]. Second, TFs can cooperatively interact with each other, if their collective effect on transcriptional output exceeds the regulatory activity of its individual parts [30]. Notably, in the most extreme case of transcriptional cooperativity the effect of each individual TFBS is required to affect transcription, thereby acting as a functional unit. Third, multiple TFBSs can be redundant, if these mutually compensate for the loss of others [34, 35].

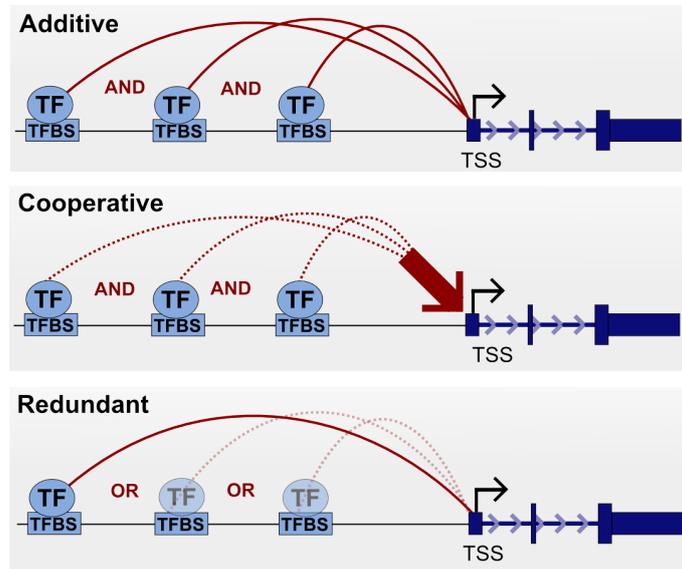


Fig. 1.3: Modes of combinatorial interaction. Transcription factors (TFs) can combinatorially regulate the expression of a specific gene by binding to multiple transcription factor binding sites (TFBSs). These can act with regard to transcriptional output in an additive, cooperative or redundant manner.

Together, combinatorial regulation by multiple TFBSs can shape gene expression depending on the activity of their individual regulatory parts or their combinatorial interplay, that may also change depending on cell type or differentiation stage [10, 36]. Thereby, the combinatorial regulation by multiple TFBSs adds another layer of complexity to the mechanisms of transcriptional gene regulation.

1.2.3 The influence of genome architecture

In the nucleus, TFBSs do not exist as an isolated linear stretch of DNA but are embedded in a sequence-specific higher order chromatin context within a tightly packed nuclear environment. In a classical view of transcriptional regulation, TFs bind to TFBSs in the

vicinity of their target genes, serving as a binding platform to recruit other co-factors and RNA polymerase II to ultimately initiate gene expression (Fig. 1.1). This classic view of gene regulation makes it reasonable to predict target genes of a specific TFBS based on its *in vivo* binding location. However, proximity-based prediction strategies exclude the fact that physical proximity between enhancers and promoters can occur over long distances by DNA-looping (Fig. 1.1). Hence, complicating the prediction of target genes for a given TFBS. Moreover, for most TFs, the majority of TFBSs as identified by ChIP-seq are located distal from TSSs, suggesting that long-range enhancer-promoter interactions do play a profound role in cellular gene regulation [37–39]. Indeed, integrating information about the genomic 3D architecture can increase the correlation between TF binding at enhancers and active transcription from an interacting promoter [26].

Experimentally, genomic long-range interactions can be investigated using microscopy-based techniques such as fluorescence *in situ* hybridization (FISH) or by chromosome conformation capture (3C)-based techniques such as 4C or Hi-C. Importantly, whereas the fluorescent labeling of DNA by FISH enables the direct visualization of a limited number of interacting loci at the single cell level, classical 3C-based techniques investigate the interaction landscapes of a population mean [40].

In 3C-based techniques enhancer-promoter interactions are preserved by fixation using formaldehyde, thereby maintaining the spatial association of enhancers and its interacting promoters (Fig. 1.4). Next, the fixed chromatin is cut with a restriction enzyme and a subsequent ligation step joins the ends of neighboring DNA fragments resulting in unique chimeric ligation products. Importantly, 3C-based techniques rely on the assumption that loci in close spatial proximity have a higher probability to become incorporated into the same DNA-hybrid molecule. Therefore, 3C-based techniques identify pairs of loci that interact with a higher frequency than expected just from random collision [41]. The main difference between currently available 3C-based techniques lies in the way the DNA-hybrid molecules are identified. Originally, the 3C-technology was limited to selected pairs of candidate interaction loci, however recent developments, such as Hi-C, allow the investigation of genome-wide interactions. In case of circularized chromosome conformation capture (4C) DNA-hybrid molecules are first amplified by inverse PCR using primers specific for the locus of interest and are subsequently identified by DNA sequencing. Therefore, 4C allows the genome-wide identification of genomic loci that interact with one specific locus of interest, referred to as viewpoint [42].

Experimental techniques such as 3C and FISH suggested that despite its apparent promiscuity enhancer activity is constrained by chromosome structure. During interphase, individual chromosomes occupy distinct regions of the cellular nucleus that only intermingle with other chromosomes at their periphery, referred to as chromosomal territories [43]. Furthermore, chromosome folding tends to cluster actively transcribed regions in the nuclear interior, whereas inactive regions are most often found in the nuclear pe-

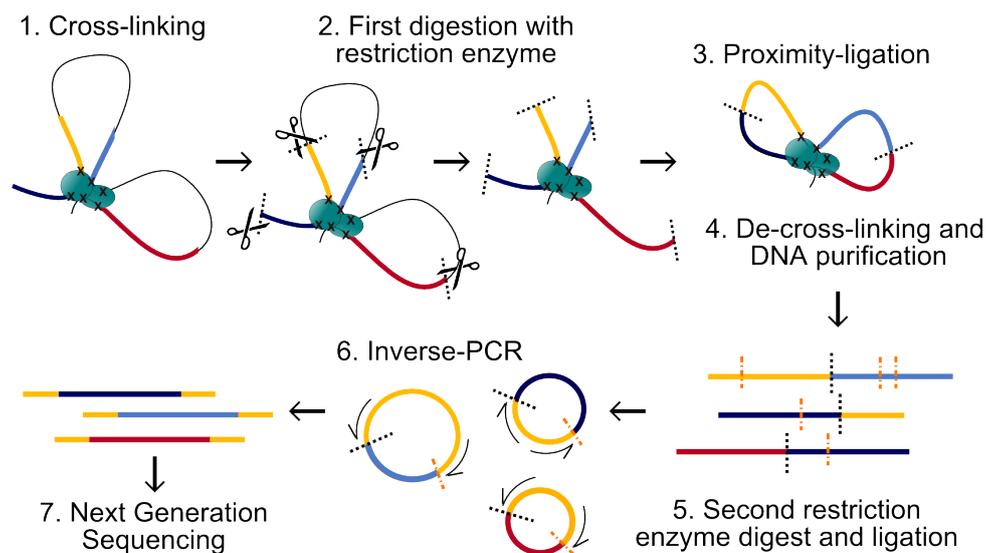


Fig. 1.4: Overview of a circularized chromosome conformation capture (4C) experiment. To preserve the spatial organization of interacting loci cells are treated with a cross-linking agent. Next, the fixed DNA is cut with a restriction enzyme (primary restriction sites are indicated with black dashed lines) followed by ligation of the cut DNA. Thereby, DNA fragments located in close physical proximity are more likely to be joined together into one DNA-hybrid molecule. Upon de-cross-linking and DNA purification, a second round of enzymatic digestion is performed using another restriction enzyme (secondary restriction sites are indicated with orange dashed lines) to reduce the size of DNA-hybrid molecules. Finally, using the generated 4C library as a template an Inverse-PCR is performed using a primer pair specific for the viewpoint of interest. DNA fragments amplified by inverse-PCR are identified by Next Generation Sequencing (NGS).

riphery, thereby contributing to the structural organization of the genome [44]. Moreover, genome-wide analysis of enhancer-promoter interactions revealed that interactions most commonly take place in regulatory units up to several megabases in size, referred to as topologically associating domain (TAD)[45]. Although the overall domain organization of individual TADs at megabase scale remains relatively stable among different species and individual cell types, a relatively small number of cell type-specific enhancer-promoter interactions can be found within TADs [25]. Supporting this picture of variant promoter-enhancer contacts, the locus control region, a distal regulatory region of the β -globin locus, interacts with a different set of globin promoters depending on the differentiation status of erythroid cells [10, 46]. However, the 3D genomic architecture can not only influence transcriptional regulation by bringing together distal regulatory elements and their target promoters but also by blocking inappropriate enhancer promoter interactions through the use of insulators.

Much remains unclear about the organization of TADs and the mechanisms of DNA-looping, but two proteins, CCCTC-binding factor (CTCF) and cohesin, were shown to contribute to both DNA-looping and TAD establishment [47]. CTCF is a highly conserved DNA binding factor that is able to homodimerize and interact with cohesin [47, 48]. Interestingly, CTCF binding is associated with both transcriptional activation and repression and it was found to block the activity of nearby enhancers, representing the main insulator protein in vertebrates [48–51]. CTCF binds to thousands of genomic sites and the majority of these sites were reported to be invariant among different cell types [52]. Especially, the establishment of TADs was found to strongly depend on the presence of CTCF at its borders. However, considering the location and number of CTCF binding sites, not all genomic CTCF-bound loci are involved in TAD boundary formation. Within TADs, regulatory DNA-loops of functional enhancer-promoter interactions often rely on CTCF binding at its bases [25]. Furthermore, the motif orientation of CTCF’s asymmetric DNA binding motif was shown to correlate with DNA-looping, as most interactions are enriched in CTCF motifs that pair in a convergent forward-reverse orientation [25, 53].

In summary, the genomic architecture structures the regulatory landscape of the genome, functionally promoting enhancer-promoter interactions by physical interaction but also constraining interactions by insulator elements. However, to which extent the genomic architecture contributes to the establishment of cell type and cell stage-specific gene expression patterns and how specificity is achieved remains unknown.

1.3 Genome editing using CRISPR/Cas9

Traditionally, reporter gene assays serve as a gold standard for testing the regulatory activity of a given DNA sequence. However, because regulatory regions are tested in a heterol-

1. Introduction

ogenous context, reporter gene assays usually lack the specific endogenous sequence context, genome architecture and epigenetic environment of the investigated TFBSs. Hence, DNA sequences that activate transcription in reporter assays are not necessarily active in its genomic context and vice versa sequences that show no regulatory effect in reporter assays may nevertheless be active for gene regulation in its genomic context. Hence, reporter gene assays only partially reflect the endogenous, *in situ* activity of a putative regulatory region.

In recent years, the rise of gene editing by programmable nucleases such as TALENs (TAL effector nucleases), ZFNs (zinc-finger nucleases) and CRISPR (clustered regularly interspersed short palindromic repeats) revolutionized the field of genetic engineering by enabling a rapid and efficient generation of targeted changes within the genome of living cells. Specifically, the type II CRISPR/Cas9 system from *Streptococcus pyogenes* has become a versatile tool for genetic engineering. The target specificity of the CRISPR/Cas9 system is determined by a guide RNA (gRNA), thereby avoiding the elaborate construction of target-specific protein modules as for TALENs and ZFNs [54–57]. Therefore, the CRISPR/Cas9 system provides a simple approach for the generation of targeted DNA changes within non-coding regulatory regions to study their regulatory activity in their endogenous genomic context.

The CRISPR gene editing system is based on an adaptive molecular defense mechanism directed against foreign DNA that is present in many bacteria and the majority of known archaea [58–60]. In the CRISPR/Cas9 system, the Cas9 endonuclease interacts with a target-specific gRNA, which guides the endonuclease to the genomic DNA target sequence based on simple nucleotide complementarity. Importantly, the specificity of the system is ensured by two mechanisms. First, the target DNA sequence must be complementary to the gRNA protospacer sequence. And second, the complementary target region must be followed by a protospacer adjacent motif (PAM), a 5'-NGG motif in the case of the CRISPR/Cas9-system from *Streptococcus pyogenes*. Upon successful recognition of the target DNA sequence, the two endonuclease domains of the Cas9 molecule introduce a double-strand break three base-pair upstream of the respective PAM sequence (Fig. 1.5). In the cell, this Cas9-introduced DNA double strand break is either repaired by the error-prone non-homologous end joining (NHEJ) pathway, which most often results in the introduction of small insertion and deletions (indels) at the site of the introduced double strand break. Alternatively, the Cas9-introduced DNA double strand break can be repaired by homology directed repair (HDR) if a cognate HDR-template is available. For example, a HDR-template can be provided using a plasmid-based DNA-template containing homology arms identical to the flanking regions of the targeted sequence (Fig. 1.5).

While the targeted mutations induced by the NHEJ repair machinery is useful for the induction of frame-shift mutations in coding regions or the disruption of TFBS motifs

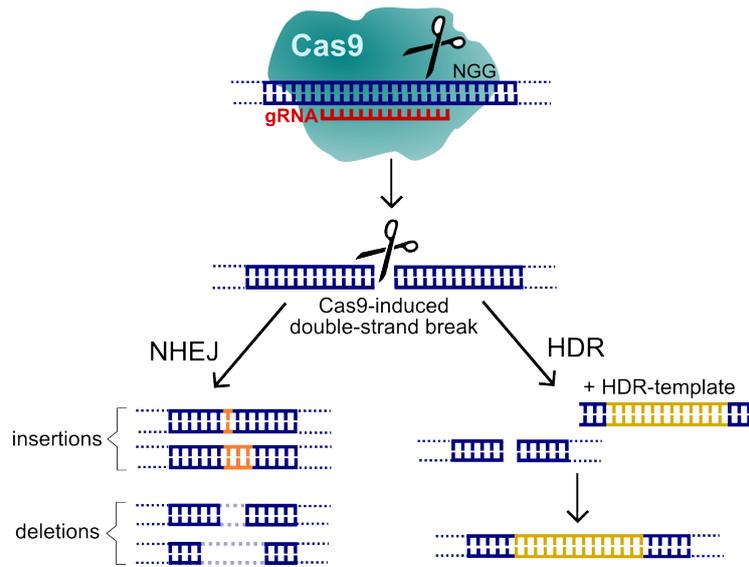


Fig. 1.5: Gene editing using the CRISPR/Cas9 system. The Cas9 nuclease is guided to its target sequence based on sequence complementarity of its associated guide RNA (gRNA). In the presence of a 5'-NGG protospacer-adjacent motif (PAM) the Cas9 nuclease introduces a double-strand break 3 bp upstream of the PAM. The Cas9-introduced DNA double-strand break is either repaired by the non-homologues end joining (NHEJ) or the homology-directed repair (HDR) pathway, in case a cognate HDR template plasmid is available. In contrast to the template-based repair by HDR, the repair of the targeted site by NHEJ is error-prone, resulting in DNA indels of various lengths.

within regulatory regions, a precise exchange of nucleotides requires the repair by HDR. However, DNA repair by NHEJ outcompetes repair by HDR, therefore the introduction of a desired change as provided on the HDR template typically occurs with a much lower frequency than genome editing by NHEJ [61, 62].

Besides genome editing, the CRISPR-system can also be used as an artificial transcriptional regulator for the programmable regulation of gene expression. Mutating the two nuclease domains of the Cas9-enzyme results in an enzymatically dead Cas9 (dCas9), which is no longer able to cut DNA but still can bind to a specific genomic region as dictated by the specificity of its associated gRNA [63]. This enzymatically dead Cas9 can be used to guide effector proteins such as transcriptional activators, repressors or histone modifiers to a specific locus of interest. For example, fusing the dCas9 enzyme to a potent transcriptional activator, such as the VP64 domain, enables the targeted recruitment of the transcriptional machinery to selected loci, ultimately inducing the activation of gene expression [64, 65]. To increase the potency for transcriptional activation, additional transcriptional coactivators can also be recruited by the gRNA scaffold itself. For example, in the case of the dCas9-VP64 synergistic activation mediator (SAM) system the gRNA scaffold provides two MS2 RNA aptamers that additionally recruit MS2-P65-HSF1 activation helper proteins boosting transcriptional activation [66]. Thereby, dCas9 fusion to transcriptional activators can generate a RNA-guided transcriptional activation complex

that can be used to interrogate the capability of individual regulatory sequences to activate the expression of a specific gene.

Thus, the CRISPR/Cas9-system provides a versatile toolbox to interrogate the function of regulatory regions. First of all, by the generation of targeted mutations within the genome of living cell but also by the targeted recruitment of transcriptional regulators to endogenous genomic loci.

1.4 The glucocorticoid receptor: a model system to study the transcriptional regulation of gene expression

The glucocorticoid receptor (GR), a ligand-activated member of the nuclear steroid hormone receptor family, is a constitutively expressed TF in the majority of vertebrate cells [67]. GR controls the expression of thousands of different genes and is key for the regulation of very diverse gene regulatory networks involved in metabolism, inflammation, stress response and development [68–72]. All members of the nuclear hormone receptor family, including GR, contain a common domain structure that is composed of an amino-terminal activation function domain, a central DNA binding domain, a hinge region and a carboxy-terminal ligand binding domain that includes its ligand-dependent activation function [73]. Alternative splicing and the usage of alternative TSSs can give rise to diverse GR isoforms with different biochemical activities, most importantly, the GR α , which is the predominant form of cellular GR [74].

The activity of GR as a transcriptional regulator strictly depends on the availability of glucocorticoids (GC), small steroid hormones, such as cortisol, the natural ligand of GR. However, GR can also be activated by synthetic GC derivatives, such as by its high-affinity ligand dexamethasone (dex). In the absence of hormone, inactivated GR resides as a monomer in the cytosol, stabilized by a chaperoning complex that maintains GR's high affinity state for its ligand and additionally prevents the exposure of its nuclear localization signal and DNA binding domain [75]. Upon ligand binding, GR changes its conformation, causing it to dissociate from its chaperoning complex and translocates into the nucleus. In the nucleus, GR can directly bind to genomic GR binding sites (GBSs) or interact with other DNA-bound TFs and transcriptional co-activators to either repress or activate the transcription of GC responsive genes (Fig. 1.6).

The consensus motif of canonical GBSs is composed of a hexameric pseudo-palindromic repeat that is separated by a three-base pair spacer facilitating GR binding as a dimer in a head-to-head fashion [76]. Besides binding to canonical GBSs, GR was reported to directly interact with the DNA via inverted-repeat GBSs in a head-to-tail fashion, via monomeric binding of GR half-site motifs or by binding to composite GBSs by simultaneous interaction with other DNA-bound TFs [77–79]. Importantly, although GR is ubiquitously

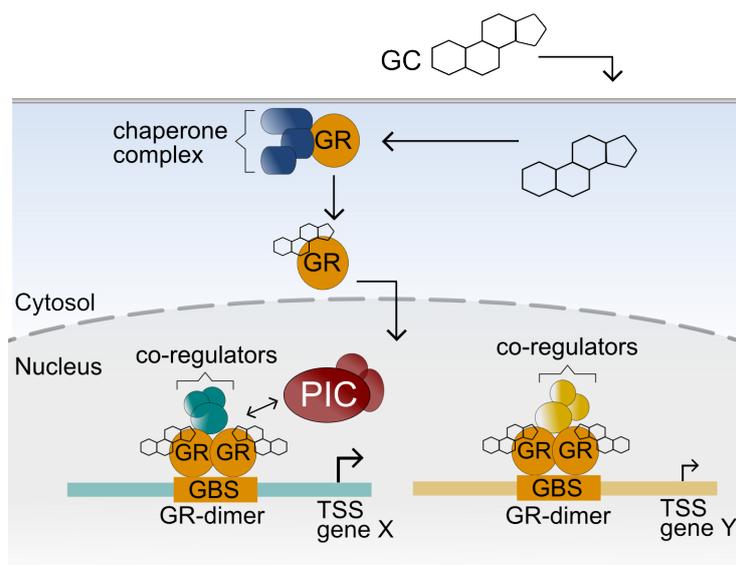


Fig. 1.6: Signaling pathway of the glucocorticoid receptor. In the cytosol glucocorticoid (GC) binding results in the dissociation of monomeric glucocorticoid receptor (GR) from its associated chaperoning complex. GC-bound GR translocates into the nucleus and binds to canonical GBSs as a dimer. Upon genomic DNA binding, activated GR recruits other co-regulators and the pre-initiation complex (PIC) to regulate the transcription of its target genes in a locus and cell type-specific manner.

expressed in distinct vertebrate cell types, GR activates the transcription of different sets of genes in a context-specific manner depending on the cell type and its physiological state [80].

There are several reasons to use GR as a model system to study the transcriptional regulation of gene expression. First, the GR is one of the best characterized metazoan TFs and its activity strictly depends on the presence of hormone. Therefore, GR induction represents an hormonal on-off system enabling a relatively simple identification of possible GR target genes. Second, because the effect of GR activity is highly cell type-specific, the GR represents an ideal system to investigate the mechanisms of cell type-specific gene regulation. Third, principles of transcriptional regulation that apply for the GR are also very likely to apply for other TFs, thereby giving insights into the general molecular mechanisms of metazoan gene regulation. Finally, GCs are a widely used clinical therapeutic due to its potent anti-inflammatory and immune-suppressive effects. However, prolonged GC treatment can cause severe side effects including metabolic disorders such as diabetes, obesity and osteoporosis [81]. Therefore, a precise understanding of the mechanisms of GR action and the induction of GR-regulated transcription is a prerequisite for the development of improved GC-mediated therapies.

Thus, due to its ligand inducibility and its cell type-specific activity the GR represents an ideal model system to study the mechanisms of cell type-specific transcriptional regulation.

1.5 Allosteric regulation of glucocorticoid receptor activity

Cellular gene expression is not an all-or-nothing event but must be quantitatively regulated depending on cell type, physiological context or in response to environmental changes. Such plasticity in transcriptional programs can only be achieved by the integration of multiple layers of regulatory information. Thus, besides regulatory information derived from the accessibility of a specific subset of TFBSs, their cooperative interaction and the presence or absence of specific co-factors, GR activity can also be changed by allosteric regulation. Such allosteric regulation is induced by interaction with allosteric effectors, resulting in conformational changes in protein structure that can ultimately alter GR activity.

Allosteric effectors of GR, e.g. covalent PTMs, identity and concentration of ligand or co-regulatory factors, can influence GR activity in several ways. For example, by changing GR's affinity for its chaperoning complex, which ultimately influences its rate of translocation into the nucleus. Furthermore, allosteric effectors can alter GR's interaction preferences with other co-regulatory factors or its DNA binding sites and eventually influence which subset of sites are occupied in the genome [82].

However, besides the aforementioned allosteric effectors, context-specific regulatory information can also arise from GR's interaction with the DNA itself. In the genome, GR binds to thousands of different genomic loci. Therefore, the plethora of DNA sequences bound by GR and their associated DNA shape was proposed to not only provide a binding platform, but also to contain locus-specific regulatory information for quantitative gene regulation [83–86]. In fact, the individual DNA sequence of canonical GBS half sites that interact with individual GR monomers does not only influence the conformation of its bound monomer, but also affects the interacting monomer bound at the other half site [84]. Similarly, the spacer sequence in between the half sites of the GR motif, can alter the DNA shape and thereby allosterically influence GR's conformation. Furthermore, the induced conformational change can alter GR's regulatory potential for transcriptional activation [83]. Additionally, the DNA sequence directly flanking the GR binding motif was suggested to alter the structure of GR's DNA binding domain, influencing transcription of both episomal and genomically integrated reporters [86].

Hence, allosteric effectors, including the exact DNA sequence bound by GR, can influence GR conformation and thereby changes its regulatory activity as a function of genomic context and cellular state. The interplay of several layers of regulatory information, including allosteric regulation, provides a complex regulatory network, enabling the cell to quantitatively fine-tune the regulation of gene expression depending on cell type or in response to environmental changes or during differentiation.

1.6 Aim of this thesis

The present work is aimed to provide insights into the molecular mechanisms and operating principles that define the regulatory activity of a bound TFBS to quantitatively contribute to cellular gene expression. To this end, I used GR as a model system and investigated the regulatory activity of selected GBSs in their natural genomic context using the CRISPR/Cas9 genome editing tool.

In the first part of my thesis, I linked selected GBSs to their target genes, by genomically deleting GBSs located upstream of two GR-regulated genes, *GILZ* and *DUSP1*. Furthermore, focusing on *GILZ*, I set out to investigate the mechanisms of combinatorial regulation by multiple GBSs, by systematically deleting GBSs individually or in combination. Next, my objective was to investigate whether TFBSs bound in different cell types share the same regulatory activity. Therefore, I edited a shared GBS that is bound by GR in two different cell types and investigated the principles of cell type-specific enhancer-promoter wiring that influence its target specificity.

In the second part of my thesis, I investigated whether the exact sequence of a GR binding motif influences the transcriptional output of *GILZ*. For this, I genomically exchanged a single GR binding motif into different sequence variants and determined its effect on endogenous *GILZ* expression. Finally, to ensure that effects from GBS motif exchange only depend on the influence from a single GBS, I artificially introduced a GBS upstream of an endogenously silenced gene. This setup provides an ideal model system for future experiments to study the regulatory effects from a single GBS in an isolated setting.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

Unless stated otherwise all chemicals and consumables were purchased from the following suppliers Calbiochem, Carl Roth, Eppendorf, Invitrogen, Merck, Sarstedt, Sigma Aldrich and TPP.

2.1.2 Cell lines

A549 (ATCC CCL-185) human lung epithelial carcinoma cell line originally derived from a 58-year-old male patient

U2OS-GR18 human bone osteosarcoma cell line stably expressing rat GR α [87] originally derived from a 15-year-old female patient

2.1.3 Antibodies

N499 polyclonal anti-GR antibody raised against the N-terminus of human GR (residues 1-499) (generated by R. M. Nissen, B. Darimont and K. R. Yamamoto)

CTCF polyclonal anti-CTCF antibody (Active Motif, Cat. No. 61311) raised against a peptide within the N-terminal region of human CTCF

H3K27Ac polyclonal anti-H3K27Ac antibody (Diagenode, Cat. No. C15410196) raised against histone H3 containing acetylated lysine 27, using a synthetic KLH-conjugated peptide

IgG control polyclonal spectrum of IgG sub-classes from rabbit serum (Diagenode, Cat. No. C15410206)

2. Materials and Methods

2.1.4 Plasmids

Tab. 2.1: Plasmids

Plasmid	Origin
hCas9	Addgene # 41815
dCas9-VP64	Addgene # 48223
MS2-p65-HSF1	Addgene # 61423
pmaxGFP	Lonza
pPur	Clontech
pGL3 <i>GILZ</i> promoter	kind gift from Wang et al. [88]
pGL3-Basic	Promega
pRL-CMV	Promega
p6R	Invitrogen

2.1.5 Media

Tab. 2.2: LB medium pH 7.5

Yeast extract	0.5 % (w/v)
Bacto-tryptone	1 % (w/v)
Sodium chloride	1 % (w/v)

Tab. 2.3: LB agar

Yeast extract	0.5 % (w/v)
Bacto-tryptone	1 % (w/v)
Sodium chloride	1 % (w/v)
Agarose	1.5 % (w/v)

Tab. 2.4: SOC medium

Yeast extract	0.5 % (w/v)
Bacto-tryptone	2 % (w/v)
NaCl	10 mM
MgCl ₂	10 mM
KCl	2.5 mM
MgSO ₄	10 mM
Glucose	20 mM

2.1.6 Buffers

Tab. 2.5: IP lysis buffer for ChIP

HEPES-KOH pH 7.4	50 mM
EDTA	1 mM
NaCl	150 mM
Glycerol	10 % (v/v)
Triton X-100	0.5 % (v/v)

Tab. 2.6: RIPA buffer for ChIP

Tris-HCl pH 8.0	10 mM
EDTA	1 mM
NaCl	150 mM
Glycerol	5 % (v/v)
Sodium deoxycholate	0.1 % (w/v)
Sodium dodecyl sulfate (SDS)	0.1 % (w/v)
Triton X-100	1 % (v/v)

Tab. 2.7: RIPA wash buffer for ChIP

Tris-HCl pH 8.0	10 mM
EDTA	1 mM
NaCl	500 mM
Glycerol	5 % (v/v)
Sodium deoxycholate	0.1 % (w/v)
Sodium dodecyl sulfate (SDS)	0.1 % (w/v)
Triton X-100	1 % (v/v)

Tab. 2.8: RIPA wash buffer for anti-CTCF ChIP

HEPES-KOH pH 7.5	50 mM
EDTA	1 mM
NP-40 alternative	1 % (v/v)
Sodium deoxycholate	0.7 % (w/v)
LiCl	500 mM

2. Materials and Methods

Tab. 2.9: LiCl wash buffer for ChIP

Tris-HCl pH 8.0	20 mM
EDTA	1 mM
LiCl	250 mM
NP-40 alternative	0.5 % (v/v)
Sodium deoxycholate	0.5 % (w/v)

Tab. 2.10: Cross-link reversal solution for ChIP

Tris-HCl pH 8.0	10 mM
EDTA	1 mM
Sodium dodecyl sulfate (SDS)	0.7 % (w/v)

prior to use: add 1 μ l proteinase K (20 mg/ml, Ambion)
per 88 μ l reversal solution

Tab. 2.11: Lysis buffer for 4C

Tris-HCl pH 7.5	50 mM
NaCl	150 mM
EDTA	5 mM
NP-40 alternative	0.5 % (v/v)
Triton X-100	1.15 % (v/v)
25x proteinase inhibitors	0.5 % (v/v)

prepare fresh before each use

Tab. 2.12: 10x ligation buffer for 4C

Tris-HCl pH 7.8	0.4 M
MgCl ₂	0.1 M
DTT	0.1 M
ATP	0.0083 M

aliquot and store at -20 °C

Tab. 2.13: 20x SSC buffer for FISH

Sodium citrate pH 7.0	0.3 M
NaCl	3 M

Tab. 2.14: qPCR master mix

Tris-HCl pH 8.3	100 mM
MgCl ₂	6 mM
Bovine serum albumin	1 mg/ml
dNTP mix	4 mM
SYBR-Green (10.000x)	0.66x
ROX reference dye	500nM

prior to use: add 10 μ l perpetual Taq polymerase (EurX)
per 990 μ l master mix

2.2 Methods

2.2.1 Cell culture

A549 and U2OS cells stably transfected with rat GR α (U2OS-GR18) [87] were cultured to confluence in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 5% fetal bovine serum (FBS) in a humidified incubator at 37 °C and 5% CO₂. For long-term storage cells were resuspended in FBS supplemented with 10% dimethyl sulfoxide (DMSO) (Serva) and frozen gradually in an isopropanol chamber until transferred to liquid nitrogen.

2.2.2 Polymerase chain reaction

For the amplification of DNA by polymerase chain reaction (PCR) a PCR reaction mix was set up on ice according to the pipetting scheme shown in Tab. 2.15, adding 2x Phusion Master Mix (NEB) as the last component.

Tab. 2.15: PCR mix

Forward primer (10 μ M)	1.25 μ l
Reverse primer (10 μ M)	1.25 μ l
Template DNA (40 ng/ μ l)	1 μ l
2x Phusion Master Mix	12.5 μ l
Sterile H ₂ O	9 μ l
<hr/>	
Total volume	25 μ l

The reaction mix was transferred to a preheated thermocycler and the PCR was conducted according to the thermocycling program shown in Tab. 2.16. For each specific primer pair both the annealing and the extension temperature of the PCR thermocycling program was optimized depending on the primer melting temperature and the size of the DNA sequence to be amplified.

Tab. 2.16: PCR thermocycling program

Step	Temperature	Duration
Initial denaturation	98 °C	30 sec
<hr/>		
Cycles 32x		
Denaturation	98 °C	10 sec
Annealing	45-72 °C	10-30 sec
Extension	72 °C	15-30 sec per kb
<hr/>		
Final extension	72 °C	10 min
Storage	4 °C	∞

2. Materials and Methods

2.2.3 Quantitative polymerase chain reaction

Quantification of DNA by qPCR was performed in a total reaction volume of 10 μ l according to the pipetting scheme shown in Tab. 2.17, containing a home-made qPCR master mix (composition shown in Tab. 2.14) and sequence-specific primer pairs (primer sequences for quantification of gene expression listed in Tab. S5.11 and for ChIP in Tab. S5.12, S5.13 and S5.14).

Tab. 2.17: qPCR mix

Template DNA	2 μ l
qPCR Master Mix	5 μ l
Primer mix (0.66 μ M)	3 μ l
<hr/>	
Total volume	10 μ l

qPCR reactions were run on a real-time PCR machine (ABI 7900 HT Applied Biosystems) according to the thermocycling program shown in Tab. 2.18. To quantify the amount of DNA during each cycle of the thermocycling program, the fluorescence of SYBR Green I dye (Invitrogen) was detected and normalized to carboxy-X-rhodamine (ROX)(Invitrogen) as an internal reference dye.

Tab. 2.18: qPCR thermocycling program

Step	Temperature	Duration	Ramp rate
Initial denaturation	95 °C	10 min	100 %
<hr/>			
Cycles 40x			
Denaturation	95 °C	15 sec	100 %
Amplification	60 °C	1 min	100 %
<hr/>			
Dissociation curve	95 °C	15 sec	100 %
	60 °C	15 sec	100 %
	95 °C	15 sec	2 %

2.2.4 Site-directed mutagenesis

To introduce specific modifications in the DNA sequence of plasmids, site-directed mutagenesis (SDM) was performed using primers inserting the desired modification based on imperfect sequence complementarity to the target region. SDM primers were specifically designed to obtain an optimal primer melting temperature of 70 °C preferably containing the mutated sites in the center of the sequence (primer sequences listed in Tab. S5.8 and S5.9). A SDM reaction mixture containing the plasmid to be modified and its corresponding SDM primer pair was prepared as shown in Tab. 2.19 and placed in a thermocycler conducting the thermocycling program shown in Tab. 2.20. Upon initial denaturation the thermocycling program was paused to add 0.5 μ l Pfu ultra polymerase (Agilent) to the reaction mix.

Tab. 2.19: SDM reaction mix

10x Pfu ultra buffer	2.5 μ l
plasmid (40 ng/ μ l)	0.5 μ l
SDM primer mix (6.25 μ M)	1 μ l
dNTP (2 mM)	2.5 μ l
Sterile H ₂ O	18 μ l
<hr/>	
Total volume	24.5 μ l

Tab. 2.20: SDM thermocycling program

Step	Temperature	Duration
Initial denaturation	95 °C	1 min
<hr/>		
Cycles 16x		
Denaturation	95 °C	30 sec
Annealing	55 °C	1 min
Extension	68 °C	4 min
<hr/>		
Final extension	68 °C	10 min
Storage	4 °C	∞

To digest the original non-mutated DNA template plasmid, 1 μ l DpnI restriction enzyme (NEB) was added and incubated for 2 h at 37 °C. 5 μ l of the digested SDM reaction mix was transformed into zymocompetent *E.coli* DH5 α (Zymo Research) according to the manufacturer’s instructions. Transformed bacteria were spread on LB agar plates containing an appropriate antibiotic. LB agar plates were incubated overnight at 37 °C. Plasmids from bacterial overnight cultures of single cell colonies were purified using the plasmid Miniprep kit (Qiagen) according to the manufacturer’s instructions. Finally, successful modification of targeted nucleotides was verified by Sanger sequencing (MWG).

2.2.5 Genomic editing by CRISPR-Cas9

Design and cloning of gRNAs

gRNAs for genome editing were designed using the Zhang lab web tool (<http://crispr.mit.edu>) [89]. gRNAs were embedded in the corresponding gRNA scaffold sequence (sequences listed in Tab. S5.3 and S5.4) and ordered as double-stranded gBlock gene fragments with 5’ phosphorylated ends (Integrated DNA Technologies). gBlocks were resuspended in H₂O to a final concentration of 10 ng/ μ l and ligated into pCR blunt vectors of the Zero Blunt PCR cloning kit (Thermo Fisher Scientific) using the pipetting scheme shown in Tab. 2.21.

Ligation reactions were incubated for 5 min at RT and transformed into zymocompetent *E. coli* DH5 α (Zymo Research). To transform zymocompetent bacteria 5 μ l of the reaction mix was added to the bacteria and incubated on ice for 5 min. Transformed cells were

2. Materials and Methods

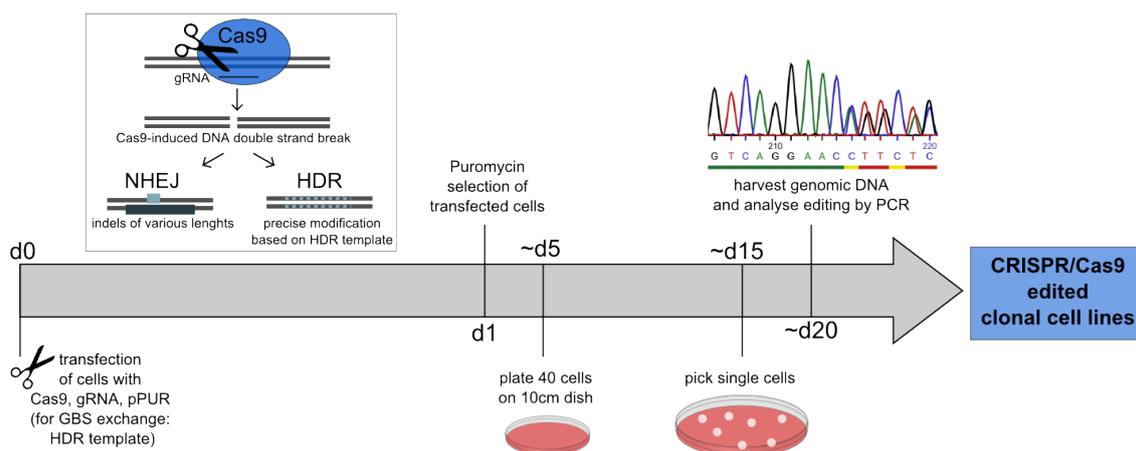


Fig. 2.1: Workflow for the generation of CRISPR/Cas9 edited clonal cell lines. Cells were transfected with a plasmid mix, encoding for the Cas9 nuclease, a locus-specific gRNA and a puromycin resistance cassette. In case of gene editing by homology-directed repair (HDR) the plasmid mix additionally contained a suitable HDR-template. Successfully transfected cells underwent gene editing by the CRISPR/Cas9 system following either the NHEJ or the HDR pathway. Single cells were seeded on culture dishes and grown until confluence, generating single-cell derived clonal cell lines. To identify successfully edited cell lines and to determine their genotype, targeted regions were PCR-amplified followed by Sanger sequencing.

Tab. 2.21: Ligation reaction

PCR blunt vector (25 ng/ μ l)	1 μ l
gBlock (10 ng/ μ l)	5 μ l
5x T4 DNA Ligase Buffer	2 μ l
T4 DNA Ligase (5 U/ μ l)	1 μ l
Sterile H ₂ O	2 μ l
Total volume	10 μl

incubated in SOC medium for 1 h at 37 °C with constant shaking at 300 rpm, allowing the bacteria to outgrow and establish antibiotic resistance against kanamycin. Upon bacterial outgrowth, bacteria were spread on LB agar plates containing 50 μ g/ml kanamycin. LB agar plates were incubated overnight at 37 °C. Plasmids from bacterial overnight cultures of single cell colonies were purified using the plasmid Miniprep kit (Qiagen) according to the manufacturer's instructions. Finally, successful integration of gRNA sequences into zeroBlunt vectors was verified by Sanger sequencing (MWG).

Generation of HDR templates

CRISPR/Cas9 gene editing by homology directed repair (HDR) allows the precise insertion of genomic modifications as provided on a HDR template. To generate HDR templates the genomic region containing the site of targeted modification including 1 kb homology arms on each site were PCR amplified as described in chapter 2.2.2 (primer sequences listed in Tab. S5.7) and cloned into a pCR blunt vector of the Zero Blunt PCR cloning

kit (Thermo Fisher Scientific) according to the manufacturer’s instructions. Nucleotide modifications of HDR templates were introduced by site directed mutagenesis (SDM) as described in chapter 2.2.4. To avoid CRISPR/Cas9 cleavage of HDR templates as well as repeated editing upon successful integration by HDR, SDM primers were designed to both insert the desired modification and to disrupt the PAM motif (SDM primer sequences listed in Tab. S5.8).

Transfection and generation of single cell-derived clonal cell lines

For gene editing by CRISPR/Cas9 cells were transiently transfected with a plasmid mix encoding the Cas9-nuclease, a puromycin resistance cassette and the respective gRNA complementary to the targeted genomic sequence (plasmids listed in Tab. 2.1). In case of generating large DNA deletions by introducing a double strand break at two genomic sites an additional gRNA was added to the plasmid mix shown in Tab. 2.22. In case of gene editing by HDR, the transfection mix additionally contained a HDR template providing the desired genomic modification (plasmid mix shown in Tab. 2.23). 1×10^6 cells were transfected with the corresponding plasmid mix by nucleofection according to the manufacturer’s instructions using either Amaxa kit T (Lonza) for the transfection of A549 cells or Amaxa kit V (Lonza) for the transfection of U2OS-GR18 cells.

Tab. 2.22: Plasmid mix for gene editing by NHEJ

hCas9	600 ng
gRNA in PCR blunt vector	600 ng
optional: second gRNA	600 ng
pPur	600 ng
<hr/>	
Total amount	2400 ng

Tab. 2.23: Plasmid mix for gene editing by HDR

hCas9	600 ng
gRNA in PCR blunt vector	600 ng
pPur	600 ng
HDR template	3000 ng
<hr/>	
Total amount	4800 ng

In case of gene editing by HDR, after transfection the cells were treated for 24 h with 10 μ M SCR7 (XcessBio Biosciences), a ligase IV inhibitor. 24 h after nucleofection, the culture medium was changed and successfully transfected cells were enriched by puromycin treatment using a concentration 10 μ g/ml of puromycin for U2OS-GR18 cells and 2 μ g/ml for A549 cells. After another 24 h, puromycin was removed and cells were grown until confluence. To expand single cells, bulk cells were counted and approximately 40 cells were transferred into 15 cm tissue culture dishes containing 20 ml of DMEM supplemented

2. Materials and Methods

with 5% fetal bovine serum (FBS). After two to three weeks, single cell-derived clones were visible by eye and transferred individually into 6-well plates using cloning cylinders.

Genotyping of CRISPR/Cas9-edited cells

To screen for successfully edited cells, genomic DNA from single cell-derived clones was purified using the DNeasy blood and tissue kit (Qiagen) according to the manufacturer's instructions. Clonal cell lines were genotyped by PCR as described in chapter 2.2.2 (primer sequences listed in Tab. S5.10). In case of gene editing by HDR at least one genotyping primer was designed to bind outside the HDR template, thereby excluding amplification of residual plasmid DNA from HDR templates.

2.2.6 Quantification of hormone-induced gene expression

Hormone induction and RNA purification

Cells were cultured to confluence in 6-well plates and treated overnight with 1 μ M dexamethasone (dex)(Alfa Aesar) or vehicle control (ethanol). The next day, RNA was extracted using the RNeasy kit (Qiagen) according to the manufacturer's instructions including the optional DNaseI digestion step.

Reverse transcription of RNA into cDNA

To generate cDNA, 500 ng of total RNA was reverse transcribed by a M-MuLV reverse transcriptase (NEB) using random nonanucleotides (NEB). For the generation of cDNA reverse transcription mix 1 (mix shown in Tab. 2.24) was incubated at 70 °C for 10 min. Subsequently 4 μ l of reverse transcription mix 2 (mix shown in Tab. 2.25) containing recombinant RNasin ribonuclease inhibitor (Promega) was added to mix 1 and incubated at 42 °C for 1 h. To thermally inactivate the reverse transcriptase for proceeding applications, the reaction mixture was finally heated for 10 min to 90 °C.

Quantification of gene expression by qPCR

First strand cDNA obtained from reverse transcription of RNA was diluted 1:25 and gene expression of candidate genes was quantified in duplicates by qPCR as described in chapter 2.2.3 using gene specific primer pairs (primer sequences listed in Tab. S5.11). The results were normalized to *RPL19*, a housekeeping gene.

2.2.7 Activation of gene expression by dCas9-VP64 SAM

gRNAs for gene activation using the dCas9-VP64 synergistic activation mediator (SAM) system [66] were designed using the Zhang lab web tool (<http://crispr.mit.edu>)[89]. gRNAs were embedded in the SAM gRNA scaffold sequence including MS2 stem loops

Tab. 2.24: Reverse transcription mix 1

Random primer 9 (0.4125 µg/µl)	2 µl
dNTP mix (2.5 mM)	4 µl
total RNA (50 ng/µl)	10 µl
<hr/>	
Total volume	16 µl

Tab. 2.25: Reverse transcription mix 2

10x M-MuLV reverse transcriptase buffer	2 µl
Recombinant RNasin ribonuclease inhibitor	0.25 µl
M-MuLV reverse transcriptase	0.125 µl
Sterile H ₂ O	1.625 µl
<hr/>	
Total volume	4 µl

(sequences listed in Tab. S5.3 and S5.6) and cloned into pCR blunt vectors of the Zero Blunt cloning kit (Invitrogen) as described in chapter 2.2.5. For gene activation by dCas9-VP64 cells were transiently transfected with a plasmid mix encoding the respective SAM-gRNA including MS2 stem loops, dCas9-VP64, the MS2-p65-HSF1 activator complex and a green fluorescent protein (GFP) expression construct (plasmids listed in Tab. 2.1) according to the pipetting scheme shown in Tab. 2.26.

Tab. 2.26: Plasmid mix for gene activation by dCas9-VP64

dCas9-VP64	600 ng
MS2-p65-HSF1	600 ng
SAM-gRNA in PCR blunt vector	600 ng
pmaxGFP	600 ng
<hr/>	
Total amount	2400 ng

Transfection was performed by nucleofection (Lonza) according to the manufacturer's instructions using either Amaxa kit T (Lonza) for the transfection of A549 cells or Amaxa kit V (Lonza) for the transfection of U2OS-GR18 cells. To enrich for successfully transfected cells, the cells were FACS sorted 24 h post-transfection (FACS Aria II SORP BD Biosciences) and RNA was subsequently isolated from GFP⁺ cells using the RNeasy kit (Qiagen) according to the manufacturer's instructions including the optional DNase I digestion step. Total RNA was transcribed into cDNA and quantified by qPCR as described in chapter 2.2.6 and 2.2.3.

2.2.8 Luciferase reporter assays

Generation of luciferase reporter constructs

The luciferase reporter construct containing *GILZ* GBS1-4 (GRCh37/hg19 ChrX: 106,961,136-106,962,152) followed by the SV40 promoter sequence and the firefly reporter gene was

2. Materials and Methods

a kind gift from Wang et al. and described previously [88]. For experiments comparing luciferase activity in both A549 and U2OS-GR18 cells, reporter constructs were generated by PCR amplification of the *GILZ* promoter region at *GILZ* transcript variant 2 (GRCh37/hg19 ChrX: 106,960,191- 106,962,152) including its endogenous promoter. For this, the region was subcloned into the pGL3-Basic reporter (Promega) containing the firefly luciferase reporter gene. For mutational analysis GBSs were mutated by SDM as described in chapter 2.2.4 (primer sequences listed in Tab. S5.9).

Transfection of U2OS-GR18 cells

To transfect U2OS-GR18 cells for luciferase assays, cells were cultured to confluency in a 48-well plate and transfected with the reporter constructs shown in Tab. 2.27 by transfection with lipofectamine (Invitrogen). In case of activation of reporter gene expression by dCas9-VP64 SAM, the plasmid mix contained all additional components required for activation (dCas9-VP64 expression plasmid, respective SAM-gRNA and an expression plasmid encoding the MS2-p65-HSF1 activator complex) (Tab. 2.28). To normalize luciferase activity for differences in cell number and transfection efficiency the plasmid mix included an internal control plasmid stably expressing *Renilla* luciferase (pRL-CMV).

Tab. 2.27: Plasmid mix for activation by GR

Firefly reporter construct (40 ng/ μ l)	1 μ l
pRL-CMV (0.4 ng/ μ l)	1 μ l
empty p6R (40 ng/ μ l)	6 μ l
<hr/>	
Total volume	8 μ l

Tab. 2.28: Plasmid mix for activation by dCas9-VP64

Firefly reporter construct (40 ng/ μ l)	0.5 μ l
pRL-CMV (0.4ng/ μ l)	0.5 μ l
dCas9-VP64 (40 ng/ μ l)	1 μ l
MS2-p65-HSF1(40 ng/ μ l)	1 μ l
SAM-gRNA in PCR blunt vector (40 ng/ μ l)	1 μ l
<hr/>	
Total volume	4 μ l

To transfect U2OS-GR18 cells the corresponding plasmid mix was mixed with 3.2 μ l Plus reagent (Invitrogen) and diluted in 50 μ l serum-free DMEM. In a separate tube 1.6 μ l lipofectamine (Invitrogen) was diluted with 50 μ l serum-free DMEM. Both mixes were vortexed and incubated for 15 min at room temperature. In the meantime U2OS-GR18 cells were washed briefly with phosphate buffered saline (PBS) and 100 μ l serum-free DMEM was added to the cells. Upon incubation, 50 μ l of lipofectamine mix was added to the plasmid mix and incubated for additional 15 min at room temperature to allow DNA complex formation. Next, 25 μ l of the DNA-lipofectamine complex was added to the cells

and mixed by gently rocking the 48-well plate. 3 h after transfection the medium was exchanged with 200 μ l DMEM supplemented with 5% fetal bovine serum (FBS).

Transfection of A549 cells

To transfect A549 cells for luciferase assays, cells were cultured to confluence in a 48-well plate and transfected with the reporter constructs shown in Tab. 2.27 by transfection with lipofectamine 2000 (Invitrogen). In case of activation of reporter gene expression by dCas9-VP64 SAM, the plasmid mix contained all additional components required for activation (dCas9-VP64 expression plasmid, respective SAM-gRNA and an expression plasmid encoding the MS2-p65-HSF1 activator complex) (Tab. 2.28). To normalize luciferase activity to differences in cell number and transfection efficiency the plasmid mix included an internal control plasmid stably expressing *Renilla* luciferase (pRL-CMV). To transfect A549 cells the corresponding plasmid mix was diluted in 100 μ l serum-free DMEM. In a separate tube 6 μ l lipofectamine 2000 (Invitrogen) was diluted with 100 μ l serum-free DMEM. Both mixes were vortexed and incubated for 5 min at room temperature. Upon incubation, 100 μ l of lipofectamine mix was added to the plasmid mix and incubated for additional 20 min at room temperature to allow DNA complex formation. Next, 50 μ l of the DNA-lipofectamine complex was added to the cells and mixed by gently rocking the 48-well plate.

Activation of luciferase expression by GR

To test the ability of GR to activate reporter gene expression, 6 h post-transfection transfected cells were treated overnight with 1 μ M dex (Alfa Aesar) or vehicle control (ethanol).

Measurement of luciferase activity

15 h after transfection reporter gene activity was measured in duplicates using the Dual Luciferase Reporter Assays Kit (Promega) in a 384-well plate reading luminometer (LUMIstar Omega BMG Labtech) according to the manufacturer's instructions. Briefly, cells were lysed by gentle shaking at room temperature for 15 min in 65 μ l passive lysis buffer (Promega). 2.5 μ l of cell lysate was subsequently transferred to a 384-well multititer plate (Greiner) and reporter gene activity was measured in the LUMIstar luminometer by sequential addition of 12.5 μ l firefly and *Renilla* luciferase substrates (LAR-II and Stop&Glo reagent, Promega).

2.2.9 Chromatin immunoprecipitation

The composition of all buffers used for chromatin immunoprecipitation (ChIP) is indicated in chapter 2.1.6.

2. Materials and Methods

For the detection of H3K27Ac or IgG control, ChIP experiments were essentially performed as described in this section for anti-GR ChIP but using 0.5 μ l of a polyclonal-H3K27Ac antibody (Diagenode, Cat. No. C15410196) or 2 μ l of a polyclonal IgG antibody (Diagenode, Cat. No. C15410206), respectively. For the detection of CTCF binding, ChIP experiments were performed in the same way but using 2 μ l of a polyclonal CTCF-antibody (Active Motif, Cat. No. 61311) and a different RIPA wash buffer for the washing steps.

Hormone induction and harvest

Cells were cultured to confluency in 10 cm dishes and treated for 90 min with 1 μ M dex (Alfa Aesar) or vehicle control (ethanol). To preserve protein-DNA interactions, the chromatin was fixed at room temperature for 3 min with 1% formaldehyde. After quenching the fixation reaction with 200 mM glycine for 10 min at 4 °C, the medium was discarded and the cells were washed twice with ice-cold 20 ml PBS, including a 5 min incubation step with PBS at 4 °C between the washing steps. Next, the cells were scraped with 10 ml ice-cold PBS into Falcon tubes and cells were pelleted by centrifuged for 5 min at 645 g at 4 °C. Upon centrifugation the supernatant was removed and pellets were snap-frozen in liquid nitrogen and stored at -80 °C until further processing.

Cell lysis and fragmentation of chromatin

To lyse the cross-linked cells 2 ml IP lysis buffer supplemented with 0.5% (v/v) proteinase inhibitor cocktail set III (EDTA-free, Merck) and 0.5 mM PMSF was added to the cells and the cell suspension was nutated for 30 min at 4 °C. Cells were pelleted by centrifugation for 5 min at 645 g at 4 °C and resuspended in 300 μ l RIPA buffer supplemented with proteinase inhibitors and PMSF as described above. Genomic DNA was fragmented in a cooled water bath using 24 cycles of sonication at high intensity (per cycle: 30 sec sonication, 30 sec pausing) using a Bioruptur (Diagenode). Upon sonication nuclei were centrifuged for 15 min at 20.000 g at 4 °C. Cellular lysates were transferred into a new tube and resuspended in 400 μ l RIPA buffer supplemented with proteinase inhibitors and PMSF as described above. For normalization 100 μ l of each sample was set aside as input control sample.

Immunoprecipitation of chromatin

Protein-bound regions were immunoprecipitated with the respective antibody listed in section 2.1.3 using concentrations indicated in section 2.2.9. For GR-ChIP 2 μ l N499 anti-GR antibody was added to the sheared chromatin and nutated overnight at 4 °C. To equilibrate protein A/G agarose beads (Santa Cruz Biotechnology), beads were incubated overnight at 4 °C in RIPA buffer supplemented with proteinase-inhibitors and PMSF as

described above. The next day, GR-bound regions were pulled down by addition of 30 μ l of a 50% bead slurry of equilibrated protein A/G agarose beads and nutation for 4 h at 4 °C. Upon incubation, antibody-bead complexes were pelleted by centrifugation for 1 min at 1000 g. To avoid unspecific binding the supernatant was removed and beads were washed four times each with 1 ml RIPA wash buffer and subsequently with 1 ml LiCl wash buffer. The beads were finally resuspended in 1 ml RIPA wash buffer and pelleted by centrifugation for 1 min at 1000 g.

Reversal of cross-linking and DNA purification

For the reversal of protein cross-linking 80 μ l of cross-link reversal solution was added to each ChIP and input sample. Samples were incubated for 3 h at 55 °C followed by incubation at 65 °C for 6 h. Immunoprecipitated DNA was purified using the PCR purification kit (Promega) according to the manufacturer's instructions and eluted in 100 μ l elution buffer.

For quantification by qPCR the DNA was diluted 1:4 and quantified as described in 2.2.3 using region specific primer pairs (primer sequences listed in Tab. S5.12, S5.13 and S5.14). For ChIP-seq, 10 ng of DNA was used to prepare ChIP-seq libraries and was subsequently sequenced by high throughput paired end sequencing (Illumina HiSeq2500).

Computational analysis of ChIP-seq data

ChIP-seq data of wildtype A549 and U2OS-GR18 cells were previously produced by Samantha B. Cooper [90]. The computational analysis of ChIP-seq data was performed by Robert Schöpflin.

Briefly, for ChIP-seq against GR in CRISPR/Cas9-edited clonal cell lines derived from U2OS-GR18 cells, single-end short reads were mapped to the hg19 reference genome using Bowtie2 v2.1.0 (-end-to-end -very-sensitive). Reads containing mismatches were removed. However, insertions and deletions in the alignment were still allowed. Reads having a mapping quality smaller than 10 and duplicate reads were removed in addition a scaling factor of 10^6 divided by the total number of mapped reads was applied. For visualization of ChIP-seq data, coverage profiles were generated using the igv tool v2.3.55 (count -z 5 -w 25 -e 90) [91, 92] and the wigToBigWig tool [93].

For ChIP-seq against CTCF in wildtype A549 and U2OS-GR18 cells, paired-end short reads were mapped to the hg19 reference genome using Bowtie v2.1.0 (-end-to-end -sensitive -maxins 2000). Reads having a mapping quality smaller than 10 were filtered out. Duplicate reads were removed using Picard-tools v2.5.0 (<http://broadinstitute.github.io/picard/>). For visualization, coverage profiles were generated using the igv tool v2.3.55 (count -z 5 -w 25 -pairs) [91, 92] and the wigToBigWig tool [93].

2. Materials and Methods

2.2.10 Circularized chromosome conformation capture

Circularized chromosome conformation capture (4C) experiments were essentially performed as previously described by van de Werken et al. [42], using the four basepair cutters Csp61 (Thermo Fisher Scientific) and DpnII (Thermo Fisher Scientific) as a primary and secondary restriction enzyme, respectively. The composition of buffers used for 4C is described in section 2.1.6.

Hormone induction and cross-linking

Cells were cultured to confluence in 10 cm dishes and treated for 90 min with 1 μ M dex (Alfa Aesar). 5×10^6 cells were pelleted by centrifugation for 5 min at 280 g. Upon centrifugation, the supernatant was discarded and cells were resuspended in 5 ml PBS supplemented with 10% fetal bovine serum (FBS). Cells were cross-linked by addition of formaldehyde to a final concentration of 2% (v/v) and incubation for 10 min at room temperature. The cross-linking reaction was quenched by addition of 1.425 ml of 1 M glycine and transferring the samples to ice.

Cell lysis and 1st restriction digest

Cells were pelleted by centrifugation for 8 min at 400 g and 4 °C. Next, the supernatant was removed and resuspended in 5 ml ice-cold lysis buffer and incubated for 10 min on ice. Cells were again pelleted by centrifugation for 5 min at 750 g and 4 °C and resuspended in 440 μ l ddH₂O.

To prepare the cross-linked cells for the 1st restriction enzyme digest 60 μ l of 10x Csp61 restriction enzyme buffer (Thermo Fisher Scientific) was added and samples were placed at 37 °C. To remove non-cross-linked cells, 15 μ l 10% (w/v) SDS were added and incubated for 1 h with constant shaking at 900 rpm at 37 °C. Next, 75 μ l of 20% (v/v) Triton X-100 was added and the samples were again incubated for 1 h with constant shaking at 900 rpm at 37 °C. Upon incubation, a 10 μ l aliquot of the sample was set aside as "undigested control" to control for digestion efficiency on a 1% agarose gel. Finally, 600 μ l 1x Csp61 restriction enzyme buffer and 100 U Csp61 (Thermo Fisher Scientific) restriction enzyme was added to the cells and samples were incubated for 4 h with constant shaking at 900 rpm at 37 °C. Upon incubation, again 66 U Csp61 were added and incubated overnight with constant shaking at 900 rpm at 37 °C. The next day, again 66 U Csp61 were added and incubated for 4 h with constant shaking at 900 rpm at 37 °C. Upon incubation, a 10 μ l aliquot of the sample was set aside as "digested control" to control the digestion efficiency on a 1% agarose gel.

Checking the digestion efficiency

To check the digestion efficiency of the first restriction enzyme undigested and digested control samples were diluted in 90 μ l of 10 mM Tris-HCl pH 7.5. To remove RNA 2 μ l Rnase A was added to each sample and samples were incubated at 37 °C for 1 h. To de-cross-link samples 2.5 μ l Proteinase K (20 mg/ml, Ambion) were added and incubated for 1 h at 65 °C. DNA was subsequently purified by phenol-chloroform extraction and analyzed on a 1% agarose gel.

Ligation reaction

For the subsequent ligation reaction restriction enzyme Csp61 was inactivated by incubating the samples at 65 °C for 20 min. For overnight ligation the samples were transferred into 50 ml Falcon tubes and 1x ligation buffer (composition indicated in chapter 2.1.6) was added to a final volume of 7 ml. Next, 50 U ligase (Roche) was added to each sample and incubated overnight at 8 °C. The next day, a 10 μ l aliquot of each sample was set aside to determine the ligation efficiency by de-cross-linking and phenol-chloroform extraction as described for checking the digestion efficiency of control samples.

De-cross-linking and DNA purification

For chromatin de-cross-linking 15 μ l Proteinase K (20 mg/ml, Ambion) was added and incubated overnight at 65 °C. The next morning 30 μ l RNase A (10 mg/ml) was added and samples were incubated for 45 min at 37 °C. DNA was purified by phenol-chloroform extraction followed by ethanol-precipitation. For this 7 ml phenol-chloroform was added to the samples and mixed vigorously. The water phase was separated by centrifugation for 15 min at 3.300 g and transferred into a separate tube. To precipitate the DNA, 7 ml H₂O, 1.5 ml NaAc pH 5.6, 28 μ l glycogen (5 mg/ml) and 35 ml ethanol was added to the water phase and snap frozen in -80 °C. The sample was centrifuged at 8.300 g for 20 min at 4 °C. Upon centrifugation the supernatant was removed and the pellet was washed with 10 ml ice-cold 70% ethanol and centrifugation for 15 min at 3.300 g at 4 °C. The DNA pellet was air-dried and dissolved in 150 μ l 10 mM Tris-HCl pH 7.5.

2nd restriction digest

To decrease the size of DNA-ligation products, a second restriction enzyme digest was performed using DpnII (Thermo Fisher Scientific). 345 μ l 1x DpnII restriction enzyme buffer and 60 U DpnII (Thermo Fisher Scientific) was added to the cells and samples were incubated overnight with constant shaking at 900 rpm at 37 °C.

The next morning, a 10 μ l aliquot of the sample was set aside as "digested control" to control the digestion efficiency on a 1% agarose gel. To inactivate the restriction enzyme,

2. Materials and Methods

the sample was incubated at 65 °C for 20 min. Next, the DNA was purified by phenol-chloroform extraction followed by ethanol precipitation. For this 500 µl phenol-chloroform was added to the samples and mixed vigorously. The water phase was separated by centrifugation for 10 min at 3.300 g and transferred into a separate tube. To precipitate the DNA, 50 µl NaAc pH 5.6 and 1 ml ethanol was added to the water phase and snap frozen in -80 °C. The sample was centrifuged at 13.200 rpm for 20 min at 4 °C. Upon centrifugation the supernatant was removed and the pellet was washed with 1 ml ice-cold 70% ethanol and centrifugation for 5 min at 13.200 rpm at 4 °C. The DNA-pellet was air-dried and dissolved in 100 µl H₂O.

2nd ligation reaction

For ligation the DNA solution was transferred to a 50 ml Falcon tube and 1.4 ml 10x ligation buffer (composition indicated in chapter 2.1.6), 100 U T4 DNA ligase (Roche) and 12.1 ml H₂O were added to each sample and incubated overnight at 8 °C. The next day, a 10 µl aliquot of each sample was set aside as "ligation control" to determine the ligation efficiency by de-cross-linking and phenol-chloroform extraction as described for checking the digestion efficiency of control samples.

To precipitate the remaining DNA 1.4 ml NaAc pH 5.6, 7 µl glycogen (5 mg/ml) and 35 ml ethanol was added and snap frozen in -80 °C. The sample was centrifuged at 8.300 g for 45 min at 4 °C. Upon centrifugation the supernatant was removed and the pellet was washed with 15 ml ice-cold 70% ethanol and centrifugation for 5 min at 3.300 g at 4 °C. The DNA-pellet was air-dried and dissolved in 75 µl 10 mM Tris-HCl pH 7.5. Next, the DNA was purified using a PCR purification kit (Qiagen) according to the manufacturer's instructions but using at least three columns per sample. Per sample The 4C library was eluted in 50 µl 10 mM Tris-HCl pH 7.5 and pooled for subsequent inverse PCR.

Inverse PCR

4C profiles were generated in two biological replicates by inverse-PCR from 4C libraries using viewpoint-specific primers (primer sequences listed in Tab. 5.15) including adaptors for subsequent high throughput sequencing (Illumina HiSeq2500). Inverse-PCR was performed using the Expand long template PCR kit (Roche) with as many parallel PCR reactions as necessary to amplify 1.6-3.2 µg DNA according to the pipetting scheme shown in Tab.2.29 and using the thermocycling program shown in Tab. 2.30.

Upon thermocycling, the inverse PCR reactions were pooled and purified using the PCR purification kit (Qiagen) according to the manufacturer's instructions and sequenced by paired end high-throughput sequencing on Illumina HiSeq2500 machine.

Tab. 2.29: Mix for inverse PCR

10 x PCR buffer 1	5 μ l
dNTP mix (12.5 mM)	0.8 μ l
primer mix (10 μ M)	10 μ l
Template DNA	10 μ l
Expand Long Template Polymerase	0.75 μ l
Sterile H ₂ O	23.45 μ l
<hr/>	
Total volume	50 μ l

Tab. 2.30: Thermocycling program for inverse PCR

Step	Temperature	Duration
Denaturation	94 °C	2 min
<hr/>		
Cycles 29x		
Denaturation	94 °C	15 sec
Annealing	55 °C	1 min
Extension	68 °C	3 min
<hr/>		
Final extension	68 °C	7 min
Storage	4 °C	∞

Computational analysis of 4C data

The computational analysis of 4C data was performed by Robert Schöpflin. In brief, inverse primer sequences were extended to the next 3' restriction site. Next, inverse primer sequences were clipped from short reads, whereby up to three mismatches were allowed for their identification. These clipped short reads were subsequently mapped in single-end mode to the hg19 reference genome using BWA-MEM v0.7.12 [94, 95] and were sorted by their name. To obtain restriction fragments, the hg19 reference genome was virtually digested using the restriction site of the first cutter Csp61. Reads were only assigned to fragments when the first read mapped to a first cutter (Csp61) restriction site and when its mate mapped to a second cutter (DpnII) restriction site. For both mates only primary alignments with a mapping quality above 30 were considered for the generation of interaction profiles.

To avoid over representation of reads mapping to the same fragment and to exclude sequences from undigested DNA, only the first alignment was considered in case alignments from the same read pair were separated by less than nine restriction fragments. To process BAM files into interaction profiles custom Java code and the HTSJDK library v1.139 (<https://samtools.github.io/htsjdk/>) was used. To reduce noise, profiles were smoothed by averaging over a running window of five fragments. To achieve comparability between cell lines, profiles were normalized as reads per million. For this, the scaling factor was calculated from all interaction contacts on the same chromosome but excluding the viewpoint (± 5 kb on each side).

2. Materials and Methods

2.2.11 siRNA knock-down of CTCF

The Trifecta DsiRNA kit including three DsiRNAs against CTCF transcript variant 1 and 2 (IDT design ID: hs.Ri.CTCF.13.1, hs.Ri.CTCF.13.2, hs.Ri.CTCF.13.3) and two control DsiRNAs (negative control DS NC1, positive duplex control HPRT-S1 DS) were purchased from IDT. To transfect cells for siRNA knock-down, A549 cells were cultured to confluence in a 6-well plate and transfected with 10 nM of individual siRNAs by transfection with lipofectamine 2000 (Invitrogen). To transfect A549 cells, individual siRNAs were diluted in 150 μ l serum-free DMEM. In a separate tube 9 μ l lipofectamine 2000 (Invitrogen) was diluted with 150 μ l serum-free DMEM. Both mixes were vortexed and incubated for 5 min at room temperature. Upon incubation, 150 μ l of lipofectamine mix was added to the diluted siRNA and incubated for additional 20 min to allow DNA complex formation. Next, 300 μ l of the DNA-lipofectamine complex was added to the cells and mixed by gentle rocking.

48 h post-transfection the cells were treated overnight with 1 μ M dex (Alfa Aesar) or vehicle control (ethanol). The next day knock-down of mRNA levels was determined by RNA extraction using the RNeasy kit (Qiagen) according to the manufacturer's instructions including the optional DNaseI digestion step. Total RNA was transcribed into cDNA and quantified by qPCR as described in chapter 2.2.6 and 2.2.3.

2.2.12 Fluorescence *in situ* hybridization

To arrest cells in metaphase, A549 and U2OS-GR18 were cultivated for 3 h in (DMEM) supplemented with 5% FBS and 0.1 μ g/ml colcemid (Thermo Fisher Scientific). Arrested cells were transferred into a hypotonic salt solution of 0.56 % (w/v) KCl and incubated for 10 min at 37 °C. Next, the cells were fixed using fixative (75 % (v/v) methanol , 25 % (v/v) acetic acid). Fixed cells were dropped onto microscope slides and stored in 100 % ethanol at - 20 °C for at least two days.

To prepare slides for hybridization with FISH probes, the slides were rinsed in saline sodium citrate buffer (SSC) (composition described in section 2.1.6) and subsequently incubated for 10 min in a pepsin solution (1 % (v/v) HCl + 0.07 % (v/v) pepsin) at 37 °C. Next, the slides underwent a series of washing steps as described in Tab.2.31. Upon washing, the slides were air-dried until hybridization with FISH probes.

FISH-probes were purchased from Empire Genomics (RP11-81I3 FISH labeled with green 5-fluorescein) and slides were hybridized with FISH probes according to the manufacturer's instructions. In brief, 8 μ l of the supplied hybridization buffer were added to 2 μ l FISH probe. Subsequently, the probe mixture was denatured by incubation for 5 min at 73 °C, stored on ice for 2 min and re-annealed by incubation for 15 min at 37 °C. For hybridization of fixed cells, 10 μ l of the probe mixture was applied on the chromosome slides and covered by microscope cover slides. The chromosome slide was incubated for 16 h in

Tab. 2.31: FISH washing steps

Duration	Solution
5 min	PBS
5 min	PBS
3 min	50 mM MgCl ₂ in PBS
10 min	50 mM MgCl ₂ in PBS + 1% (v/v) formaldehyde
5 min	PBS
3 min	70% (v/v) ethanol
3 min	85% (v/v) ethanol
3 min	100% (v/v) ethanol

a humidified chamber containing 50% (v/v) formamide in 2x SSC as humidity control. Upon incubation, the cover slide was removed and the chromosome slide was washed by agitating in 0.4x SSC containing 0.3% (v/v) NP-40 alternative for 10 sec and subsequently incubated for another 2 min. Thereafter, the chromosome slide was transferred to 2x SSC containing 0.1% (v/v) NP-40 alternative at room temperature and incubated for 1 min. Upon hybridization, the DNA was counterstained with DAPI (Hoechst) and mounted with Antifade Mounting Medium (Vectashield). Images of metaphase spreads and intact FISH-labeled cells were captured using a fluorescence microscope (Zeiss LSM700) and quantified by visual inspection.

2.2.13 Computational prediction of transcription factor affinity

To assess the effect of indels at *DUSP1* GBS1 and *GILZ* GBS1 on the predicted affinity motif score, sTRAP (http://trap.molgen.mpg.de/cgi-bin/trap_two_seq_form.cgi) was used [96] using the matrix IDs M00955, M00192 and M00205, chordate conserved elements as a background model and Benjamini-Hochberg correction for multiple testing. To identify CTCF motif-matches at CTCF-bound ChIP peaks, regions of interest were analyzed using the Transcription Factor Affinity Prediction (TRAP) webtool (http://trap.molgen.mpg.de/cgi-bin/trap_form.cgi) [97]. All CTCF motif matches for matrix IDs M01259 and M01200 with a weight score above 4.5 were considered for analysis.

3 Results

3.1 Part 1: Genomic deletion of glucocorticoid receptor binding sites

3.1.1 Genomic editing of GBSs located upstream of GR target genes

To investigate the regulatory activity of individual TFBSs in their endogenous genomic context, the CRISPR/Cas9-system was used to destroy the GR binding motif of selected GBSs by gene editing.

For gene editing two candidate GBSs were chosen that are located upstream of two GR-regulated genes, the glucocorticoid induced leucine zipper (*GILZ* GBS1) and the dual specificity phosphatase1 (*DUSP1* GBS1), respectively. These selected GBSs appeared to be suitable candidates for several reasons. First, the selected GBSs, were located proximal to the promoter region of a GR target gene, *GILZ* or *DUSP1*, in a distance of approximately 1.5kb to the corresponding TSS (Fig. 3.1 a, b). Both *GILZ* and *DUSP1* are upregulated upon hormone addition (Fig. 3.1 c) and were shown to play an important role in the immune-suppressive and anti-inflammatory effect of glucocorticoids [98, 99]. In addition, ChIP-seq consistently showed GR-binding at the selected GBSs across diverse cell types, such as in a U2OS osteosarcoma cell line with stably integrated GR α [87] (Fig. 3.1 a, b), suggesting that GR binding at these GBSs is conserved across diverse cell types. Second, previous studies revealed that both *GILZ* GBS1 and *DUSP1* GBS1 are functionally active, as their mutation reduced the enhancer activity of reporter constructs containing the *GILZ* or *DUSP1* promoter region, respectively [88, 100]. Third, both selected GBSs contain a PAM sequence directly within their 17bp core GR binding motif (Fig. 3.1 d, e), enabling the direct editing of their GR binding motif by introducing a double-strand break 3bp upstream of the respective PAM [55].

Using the CRISPR/Cas9 system, single-cell-derived clonal lines from U2OS cells stably expressing GR α [87] were generated, containing indels of various lengths within the GR binding motif of *GILZ* GBS1 or *DUSP1* GBS1, respectively (Fig. 3.1 d, e and Fig. S5.1, S5.2). For both *GILZ* and *DUSP1* GBS1 editing efficiencies were high and the majority of analyzed cells were successfully edited at the targeted locus (*GILZ* GBS1 70% of analyzed cell, *DUSP1* 85% of analyzed cells) (Fig. S5.1 and S5.2). Notably, by using gRNAs that bind within the GR binding motif, CRISPR/Cas9-induced indels most often resulted in the disruption of nucleotide positions critical for GR binding, including the spacer sequence

3. Results

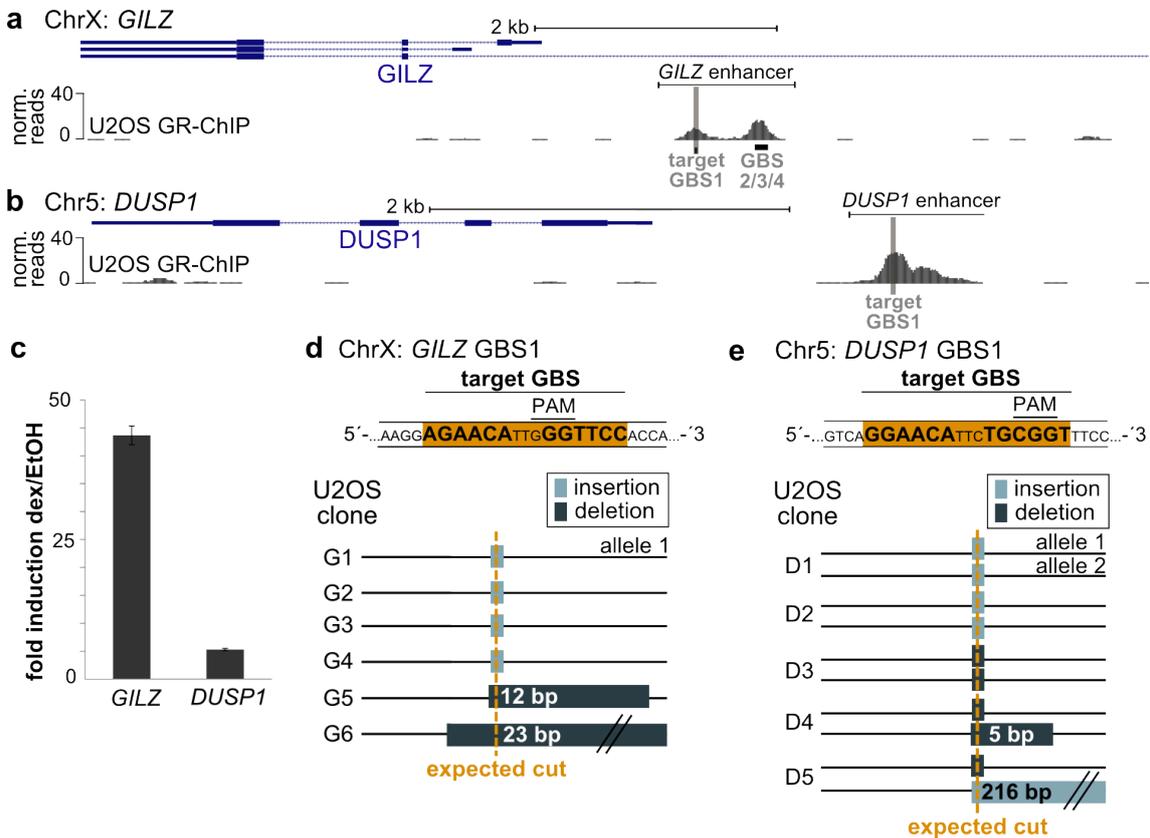


Fig. 3.1: Genomic editing of GBSs using CRISPR/Cas9 disrupts the GR binding motif. (a, b) Genome browser view showing the genomic location and GR occupancy at GBSs located upstream of the genes *GILZ* (a) and *DUSP1* (b). The individual GBSs targeted for deletion are highlighted in grey. (c) Fold induction of *GILZ* and *DUSP1* expression in U2OS cells upon overnight treatment with 1 μM dexamethasone. The graph shows the average gene expression of three independent experiments and error bars represent ± SEM. (d, e) Top: Schematics showing the target GBS sequence (highlighted in orange) and the protospacer adjacent motif (PAM) used for CRISPR/Cas9 gene editing. Bottom: Genotyping results of successfully edited clonal cell lines containing CRISPR/Cas9-induced insertions (turquoise) or deletions (dark green) at *GILZ* (d) or *DUSP1* (e) GBS1. The expected location of the Cas9-induced DNA double-strand break (3 bp upstream of the PAM) is indicated with a dashed line.

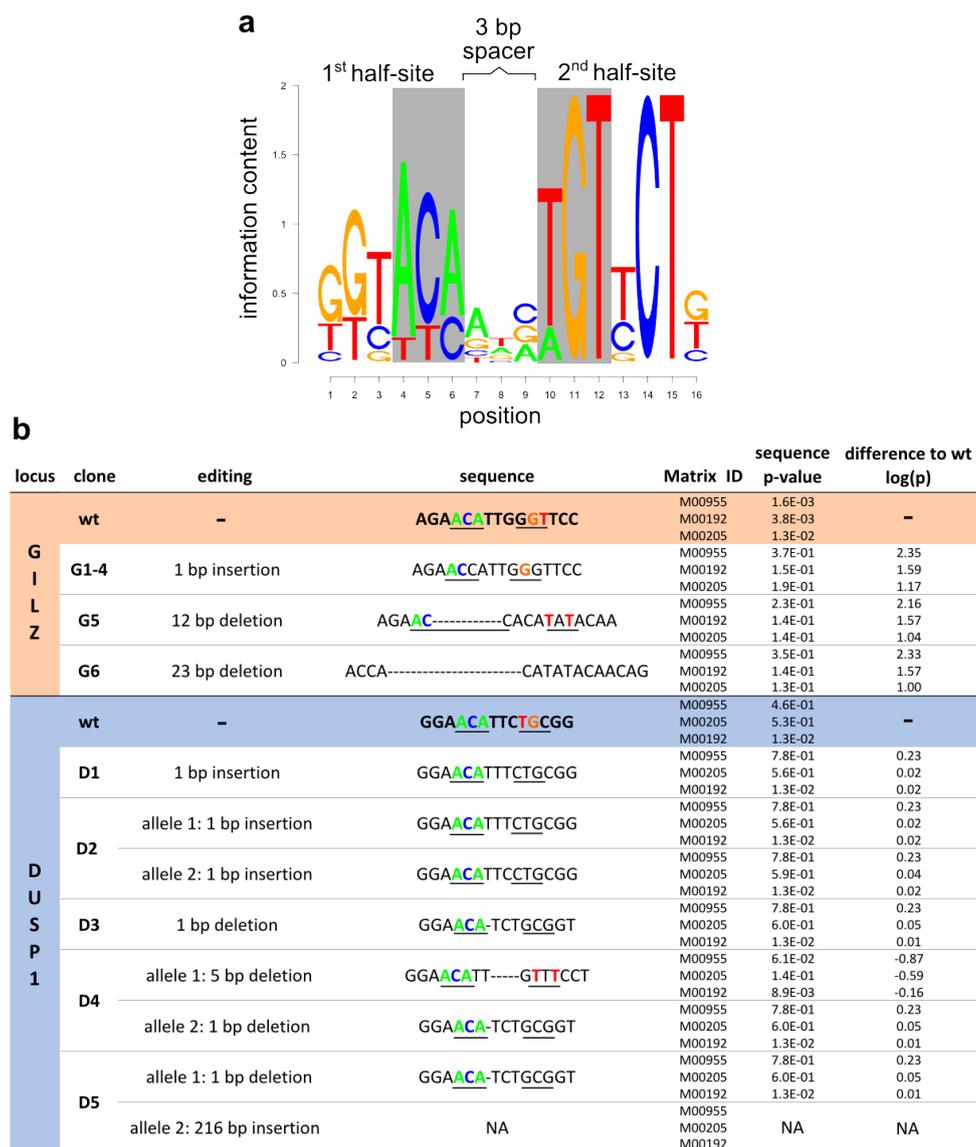


Fig. 3.2: Editing of GBSs by CRISPR/Cas9 reduces their GR binding affinity. (a) Position weight matrix of the GRE-motif (Matrix-ID: M002045), critical positions for GR binding are highlighted in grey. (b) The table summarizes the predicted difference $\log(p)$ in GR binding affinity as determined using the TRAP-tool [97] between the unedited wildtype and the CRISPR/Cas9 edited clonal cell lines for three different GR binding motifs (Matrix ID: M00955, M00192, M00205). For orientation, nucleotides that would correspond to the theoretical GBS half site motifs are underlined and nucleotides matching the GRE motif shown in (a) are colored.

between the GR half-site motif (Fig. 3.1 d, e and Fig. 3.2 a). Furthermore, a reduction in predicted GR binding affinity was confirmed for all but one generated clonal cell line by computational calculation of their GR motif score (Fig. 3.2 b), indicating that genome editing destroys the GR binding motif at these loci.

Although *GILZ* is located on chromosome X and the U2OS cell line was originally derived from a female, genotyping of *GILZ* GBS1 resulted in the detection of one edited

3. Results

allele only, whereas genotyping of *DUSP1* GBS1 resulted in the detection of up to two differentially edited alleles (Fig. 3.1 d, e). Conceivably, gene editing at *GILZ* GBS1 could result in homoallelic editing on both alleles. However, another possible explanation for the observed genotyping results could be the presence of only one allele in the U2OS cell line. To quantify the number of *GILZ* alleles fluorescence in *situ* hybridization (FISH) was performed using probes against the *GILZ* locus. In the majority of analyzed U2OS cells only one FISH signal was observed, confirming the presence of only one copy of the *GILZ* locus in U2OS cells (Fig. 3.3 a). Thus, the *GILZ* locus represents an ideal locus to study the genomic effects of targeted changes within the GR binding motif.

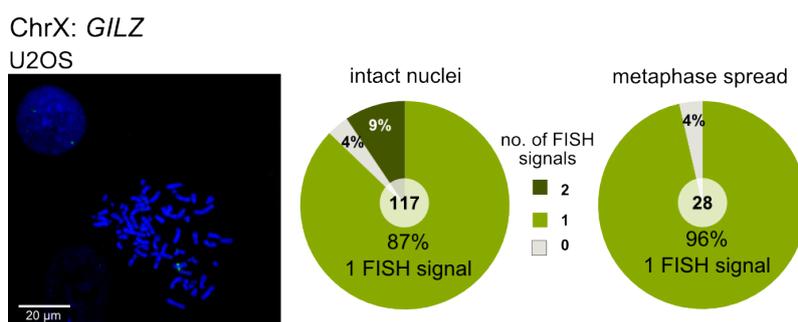


Fig. 3.3: Quantification of *GILZ* alleles in U2OS cells. (Left) Representative FISH image of U2OS cells using hybridization probes for *GILZ* binding at locus Xq22.3 (green). (Right) Pie charts summarize the number of FISH signals in all analyzed metaphase spreads and intact nuclei. The total number of analyzed cells is depicted in the inner circle of each pie chart.

Taken together, using the CRISPR/Cas9-system clonal cell lines were generated with disrupted GR binding motif located upstream of the two GR target genes *GILZ* and *DUSP1*. Indels generated by genome editing reduced the computationally predicted GR binding affinity. Hence, for further experiments I will refer to GBSs disrupted by indels as being deleted.

3.1.2 Genomic deletion of GBSs influences the regulation of nearby genes

Because genomic TF binding itself must not necessarily functionally contribute to gene expression, I next tested whether the deletion of the selected GBSs affected the GR-dependent regulation of the nearby gene.

To exclude effects from variation among clonal cell lines, gene expression was determined for at least five different clonal cell lines edited at either *GILZ* or *DUSP1* GBS1 (Fig. 3.1 d, e). The deletion of *DUSP1* GBS1 resulted in a decrease of hormone-induced *DUSP1* expression (60% of wildtype expression) (Fig. 3.4). In contrast, for *GILZ* the effect of GBS1 deletion was stronger (39% of wildtype expression). Notably, these effects were observed in comparison to both wildtype cells and CRISPR/Cas9-edited clonal cell lines that were not edited at the respective locus, thereby serving as an additional control

for the effect of GR-responsiveness of clonal cell lines (Fig. 3.4). Furthermore, both *GILZ* and *DUSP1* GBS1 deletion did neither influence basal gene expression nor the induction of other tested GR-regulated genes, as shown for *FKBP5* (Fig. 3.4).

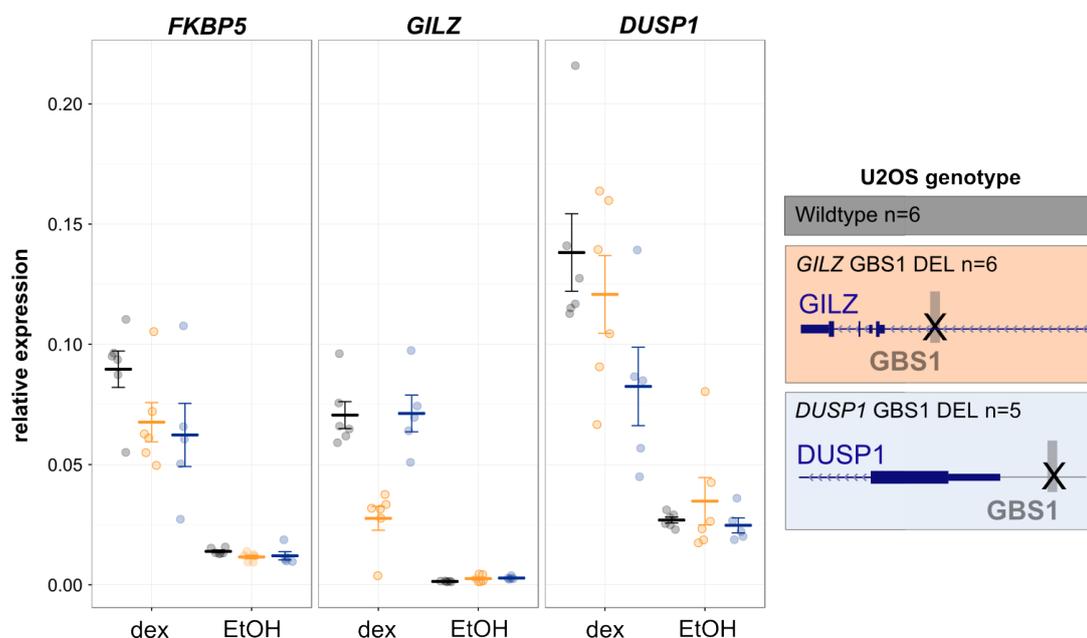


Fig. 3.4: Genomic deletion of GBSs reduces expression of the nearby gene. Relative *GILZ*, *DUSP1* and *FKBP5* expression of clonal cell lines with deleted *GILZ* GBS1 (n=6) or *DUSP1* GBS1 (n=5) as determined by qPCR after overnight treatment with 1 μ M dexamethasone (dex) or vehicle control (EtOH). A representative from two independent experiments is shown. Horizontal lines indicate the average relative gene expression of individual clonal cell lines and error bars represent \pm SEM.

Thus, both the deletion of the *GILZ* and *DUSP1* GBS1 resulted in a decrease of nearby gene expression, confirming that the deleted GBSs are functional and play a role in the GR-dependent induction of *GILZ* and *DUSP1* in U2OS cells. Notably, upon deletion of either *GILZ* or *DUSP1* GBS1, both genes are still upregulated in response to hormone treatment, indicating that additional GBSs contribute to the GR-dependent regulation of these genes.

3.1.3 Multiple promoter-proximal GBSs cooperatively regulate *GILZ* expression

The deletion of neither *GILZ* nor *DUSP1* GBS1 resulted in a complete loss of their GR-dependent expression. Thus, additional GR-bound loci might be involved in the cooperative regulation *GILZ* expression. Notably, both *GILZ* and *DUSP1* GBS1 are part of an enhancer containing multiple GBSs (Fig. 3.1 a, b). Specifically, for the *GILZ* locus,

3. Results

ChIP-seq and combined motif search identified the presence of three additional promoter-proximal GBSs (*GILZ* GBS2-4), clustering approximately 500 bp relative to GBS1 and 2 kb upstream of the TSS (Fig. 3.1 a). Furthermore, previous reporter gene assays in A549 cells revealed that mutating these three additional GBSs resulted in a reduction of the GR-dependent enhancer activity of the *GILZ* promoter [88]. Conceivably, besides *GILZ* GBS1, these previously identified promoter-proximal GBSs might additionally contribute to *GILZ* regulation in its natural genomic context.

To investigate the functional importance of these additional GBSs, clonal cell lines were generated containing a deletion of the genomic fragment containing GBS2-4 (Fig. S 5.3). Similar to what was observed upon deletion of GBS1, in comparison to wildtype cells the deletion of GBS2-4 resulted in a 50% reduction of hormone-dependent *GILZ* expression (Fig. 3.5). Furthermore, the deletion of GBS2-4 did neither affect basal *GILZ* expression nor the GR-dependent upregulation of other tested GR-regulated genes as shown by the expression of *FKBP5* (Fig. 3.5). Thus, both *GILZ* GBS1 and GBS2-4 contribute to the overall GR-dependent response of *GILZ* expression.

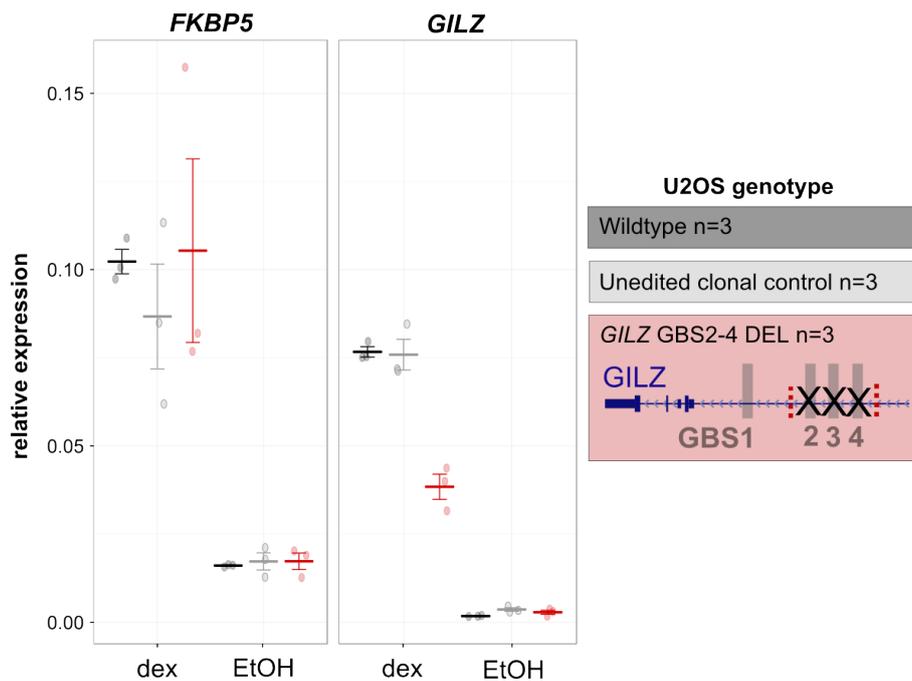


Fig. 3.5: Multiple GBSs regulate *GILZ* expression. Relative *GILZ* and *FKBP5* expression of clonal cell lines with deleted *GILZ* GBS2-4 (n=3) as determined by qPCR after overnight treatment with 1 μ M dexamethasone (dex) or vehicle control (EtOH). A representative from two independent experiments is shown. Horizontal lines indicate the average relative gene expression of individual clonal cell lines. Error bars represent \pm SEM.

In principal, multiple TFBSs such as *GILZ* GBS1 and GBS2-4, can combinatorially regulate gene expression by at least two possible mechanisms. First, an independent

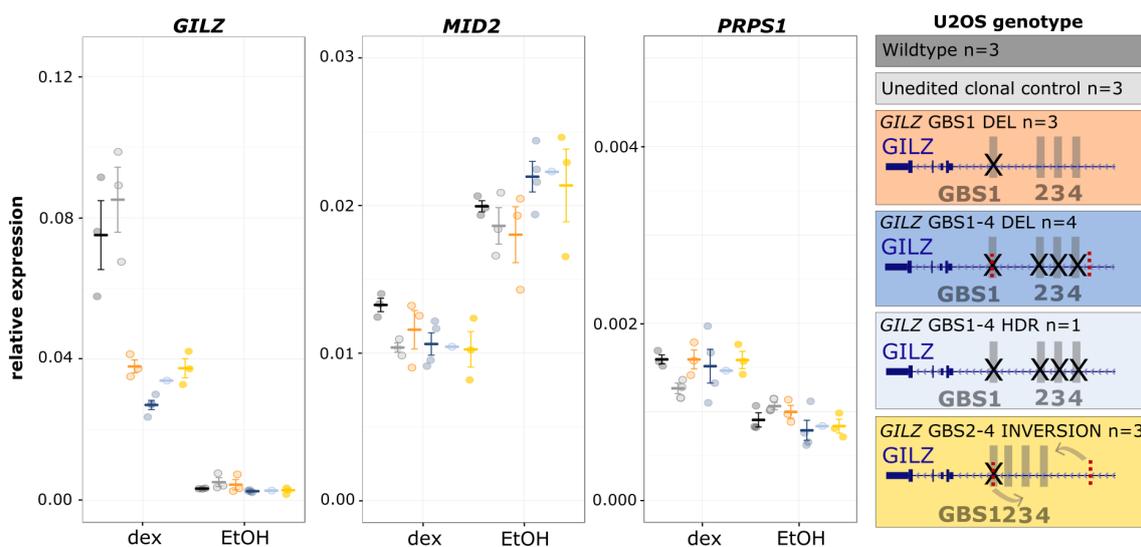


Fig. 3.6: *GILZ* GBS1-4 cooperatively regulate *GILZ* expression as a functional unit. Relative *GILZ*, *MID2* and *PRPS1* expression of clonal cell lines with deleted *GILZ* GBS1 (n=3), GBS1-4 (n=3) or inverted *GILZ* GBS2-4 (n=3) as determined by qPCR after overnight treatment with 1 μ M dexamethasone (dex) or vehicle control (EtOH). A representative from two independent experiments is shown. Horizontal lines indicate the average relative gene expression of individual clonal cell lines. Error bars represent \pm SEM.

enhancer model, in which individual GBSs contribute to the overall GR response in an additive manner. And second, a cooperative enhancer model in which the presence of all GBSs is required to jointly establish GR-dependent induction.

To test whether GBS1 and GB2-4 regulate *GILZ* in an additive or a cooperative manner, clonal cell lines were generated containing combinatorial deletions of both *GILZ* GBS1 and GBS2-4, following two different strategies. First, clonal deletion cell lines were generated using two gRNAs: one cutting directly within *GILZ* GBS1 and another one downstream of GBS4 (*GILZ* GBS1-4 DEL) (Fig. S 5.4), thereby genomically removing a \sim 600 bp DNA segment containing all four GBSs. Second, because cutting removes a large DNA fragment, a clonal cell line was generated using a HDR template introducing point mutations in each of the four GBSs at positions critical for GR binding (*GILZ* GBS1-4 HDR) (Fig. S 5.4).

Upon hormone treatment, both *GILZ* GBS1-4 deletion cell lines, containing either a \sim 600 bp deletion (*GILZ* GBS1-4 DEL) or individual mutations within GBS1-4 (*GILZ* GBS1-4 HDR), showed a similar reduction in the level of *GILZ* expression (Fig. 3.6), indicating that the removal of the DNA fragment in between GBS1 and GBS4 did not further affect *GILZ* expression. Surprisingly, the deletion of GBS1-4 (*GILZ* GBS1-4 DEL) did not result in a greater loss of GR-dependent *GILZ* induction than the single deletion of GBS1 (*GILZ* GBS1 DEL)(Fig. 3.6), suggesting that GBS1-4 cooperatively regulate *GILZ* as a single functional unit in which the presence of each GBS is required for maximum induction of gene expression.

3. Results

For some clonal cell lines, gene editing by CRISPR/Cas9 using two gRNAs resulted in an inversion of the targeted genomic fragment (Fig. S 5.5). To assess whether the genomic position of GBS2-4 is crucial for its activity in regulating *GILZ* expression, the effect of inverting GBS2-4 and placing them in a closer distance relative to the TSS was determined. Notably, this set-up simultaneously destroyed GBS1. Upon hormone treatment, the inversion of GBS2-4 (*GILZ* GBS2-4 INVERSION) did result in a similar level of *GILZ* expression in comparison to clonal cell lines with deleted GBS1 (Fig. 3.6). Thus, situating GBS2-4 into closer proximity relative to the TSS of *GILZ* was not able to rescue the effect of GBS1 deletion, suggesting that an intact regulatory unit of GBS1 and GBS2-4 is required to confer maximum GR-dependent *GILZ* expression.

In the genome *GILZ* is directly flanked by two other GR-regulated genes, *MID2* and *PRPS1*, whose TSSs are located in a distance of 108 and 90 kb relative to GBS1, respectively. Whereas *PRPS1* is upregulated by GR upon hormone addition, *MID2* is downregulated by GR (Fig. 3.6). To investigate whether *GILZ* GBS1-4 are additionally involved in the GR-dependent regulation of these genes, the effect of GBS1-4 deletion on the expression of *MID2* and *PRPS1* was determined. Although in all cell lines analyzed both *MID2* and *PRPS1* were regulated by GR, clonal cell lines with deleted GBS1 or GBS1-4 showed similar expression levels of both genes in comparison to both wildtype or an unedited clonal control cell line, suggesting that GBS1-4 are not involved in the regulation of *MID2* and *PRPS1*.

In summary, these findings indicate that in their genomic context, both *GILZ* GBS1 and GBS2-4 are required to fully establish the GR-dependent induction of *GILZ*. Furthermore, *GILZ* GBS1 and *GILZ* GBS2-4 cooperatively regulate *GILZ* expression as a single functional unit. Notably, upon deletion of four promoter-proximal GBSs *GILZ* is still upregulated by GR upon hormone addition, indicating that additional distal GBSs contribute to the regulation of *GILZ* expression.

3.1.4 Genomic deletion of GBSs does not influence GR binding at neighboring sites

TFs are able to mutually interact by protein-protein interactions, creating a regulatory hub to cooperatively regulate gene expression. In this context, direct and indirect protein-protein interactions between clustering TFs or its associated co-factors were reported to mutually stabilize TF binding and occupancy at individual TFBSs [101, 102].

To examine whether the deletion of individual GBSs influences GR occupancy at nearby loci, I compared the GR ChIP-seq peaks of clonal cell lines in which either the *GILZ* or *DUSP1* GBS1 was deleted to a control cell line that was unedited at the respective locus. Both the deletion of either *GILZ* GBS1 or *DUSP1* GBS1 resulted in a reduction of GR binding at the deleted GBSs (Fig. 3.7 a, b). However, upon deletion GR binding was

not completely abolished, most probably resulting from the limited resolution of ChIP-seq experiments between nearby GR-peaks. Although GR binding at the deleted GBSs was diminished in both *GILZ* and *DUSP1* deletion clones (Fig. 3.7 a, b), GR binding at GBSs nearby remained unaffected (Fig. 3.7). Specifically, the deletion of *GILZ* GBS1 did not alter GR binding at *GILZ* GBS2-4 (Fig. 3.7 a). Thus, these findings indicate that the cooperative interaction among GBS1-4 does not result from cooperativity at the level of DNA binding.

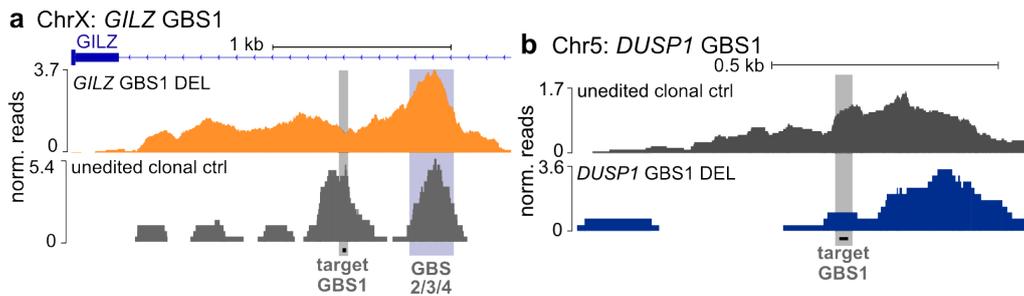


Fig. 3.7: Genomic deletion diminishes GR binding at deleted GBSs. (a, b) ChIP-seq tracks showing GR binding in a representative clonal cell line with deleted *GILZ* GBS1 (clone G2) or *DUSP1* GBS1 (clone D2) at the *GILZ* (a) and *DUSP1* (b) promoter region. The deleted GBSs are highlighted in grey. The location of *GILZ* GBS2-4 is highlighted in blue.

To determine, whether the deletion of *GILZ* or *DUSP1* GBS1 affects GR binding at more distal loci, I compared GR binding at GBSs surrounding the *GILZ* or *DUSP1* locus. Similar to what was found for nearby GBSs, GR binding at more distal loci was neither affected by *GILZ* nor *DUSP1* GBS1 deletion. Besides ChIP-seq experiments with one representative deletion clone, the maintenance of GR binding upon GBS deletion was additionally confirmed by ChIP-qPCR in three individual clonal cell lines for a subset of 5 neighboring *GILZ* and 3 neighboring *DUSP1* ChIP-peaks (Fig. 3.8 a-d).

Together, as determined by ChIP the genomic deletion of *GILZ* and *DUSP1* GBS1 led only to a small reduction of GR binding at the respective GBSs themselves. Furthermore, GR occupancy at interacting GBSs was maintained upon deletion, indicating that GR binding at the deleted GBSs did not affect the recruitment of GR to adjacent loci. Thus, although *GILZ* GBS1-4 cooperatively interact to regulate *GILZ* expression, these GBS are bound individually by GR, indicating that their cooperativity is not a result of cooperative DNA binding.

3. Results

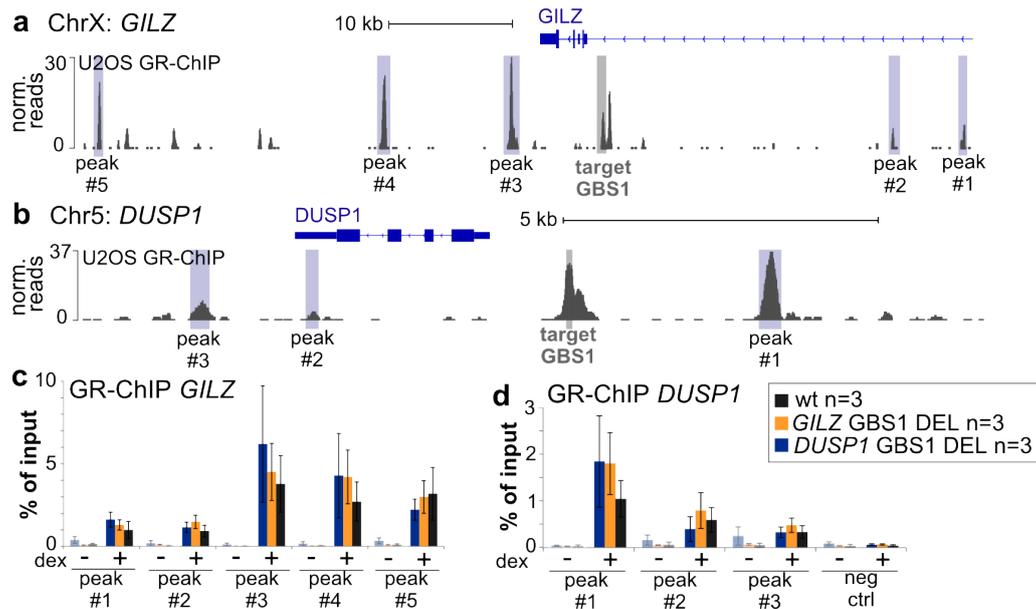


Fig. 3.8: Genomic deletion remains GR binding at neighboring GBSs unaffected. (a, b) ChIP-seq tracks showing GR binding in U2OS cells at the genomic region surrounding *GILZ* (a) and *DUSP1* GBS1 (b). The main neighboring GR-ChIP peaks selected for quantitative analysis by ChIP-qPCR are highlighted in blue. (c, d) ChIP-qPCR analysis of GR binding at selected loci surrounding the *GILZ* (c) or *DUSP1* (d) locus in three representative clonal cell lines with deleted *GILZ* or *DUSP1* GBS1. Error bars represent \pm SEM.

3.1.5 Reporter gene assays cannot recapitulate the genomic cooperativity of promoter-proximal *GILZ* GBSs

Reporter gene assays are a standard tool for studying the enhancer activity of *cis*-regulatory elements. These assays are performed by transient transfection of episomal reporter gene constructs containing the DNA segment to be tested in a non-native, heterologous context.

To compare the effect of genomic GBS deletion to classical reporter gene assays, the enhancer activity of the *GILZ* promoter region encompassing GBS1-4 was investigated in luciferase reporter constructs. For this, constructs were generated containing individual and combinatorial deletions of GBS1-4, mutated at positions most crucial for GR binding. Thereby, this setup mimics the genomic situation in the generated cell lines with deleted GBSs.

As expected, upon hormone treatment reporter constructs containing the wildtype *GILZ* promoter region highly upregulated reporter gene expression. Similar to the genomic deletion, the deletion of GBS2-4 reduced the induction of reporter gene expression by 51% in comparison to the wildtype construct. In contrast, the deletion of GBS1, similar to the combinatorial deletion of GBS1-4 completely abolished GR-dependent induction of reporter gene expression (Fig. 3.9).

Hence, in reporter gene assays, similar to its genomic context both GBS1 and GBS2-4

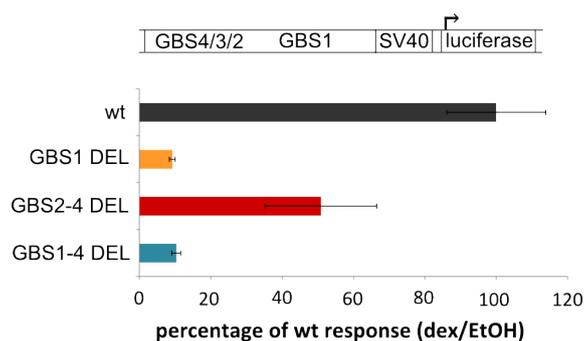


Fig. 3.9: In an episomal context enhancer activity of the *GILZ* promoter depends on the presence of GBS1. Hormone-induced luciferase activity of reporter constructs containing the *GILZ* promoter region encompassing GBS1-4 in comparison to reporter constructs with individual and combinatorial mutations of the indicated GBSs in U2OS cells. The graph shows the average of three independent experiments. Error bars represent \pm SEM.

are required for the full regulatory activity of the *GILZ* promoter region. However, in an episomal setting the overall activity of the *GILZ* enhancer strictly depends on the presence of GBS1 but not of GBS2-4. Thus, reporter gene assays do not detect *GILZ* GBS1-4 as a functional regulatory unit, in which the presence of both GBS1 and GBS2-4 are strictly required for GR-dependent regulatory activity.

3.1.6 The regulatory activity of *GILZ* GBS1 differs between cell types

The regulation of gene expression by GR in response to GCs is highly cell type-specific [103]. Besides binding to only a cell type-specific subset of all possible genomic GBSs [13, 104], an additional mechanism contributing to the cell type-specific effect of GR activity might result from differences in the individual regulatory activity of GR-bound GBSs. As previously demonstrated, both *GILZ* and *DUSP1* GBS1 are bound by GR in both A549 and U2OS cells [88, 100] (Fig. 3.10 a, b). Furthermore, in both cell lines *DUSP1* and *GILZ* are regulated by GR and induced upon hormone addition (Fig. 3.4 and Fig. 3.12). Notably, the origin of these two cell lines differs, as the U2OS cell line is derived from human osteosarcoma cells, while the A549 cell line is derived from lung adenocarcinoma cells.

To test whether the functional significance of *GILZ* and *DUSP1* GBS1 on gene expression is conserved among different cell types, CRISPR/Cas9-induced deletions of *GILZ* and *DUSP1* GBS1 were generated in A549 cells. Exactly as for GBS deletion experiments in U2OS cells, the same gRNAs were used to target *GILZ* or *DUSP1* GBS1 directly within their binding motif. For *GILZ* GBS1 the efficiency of successful editing was comparable between U2OS (70% of analyzed clones) and A549 cells (67% of analyzed clones) (Fig. S 5.1 and S5.6). However, for *DUSP1* GBS1 the fraction of successfully edited cells was considerably lower in A549 (44% of analyzed clones) than in U2OS cells (85% of analyzed clones) (Fig. S5.2 and S5.7). Notably, in contrast to U2OS cells, in A549 cells genotyping

3. Results

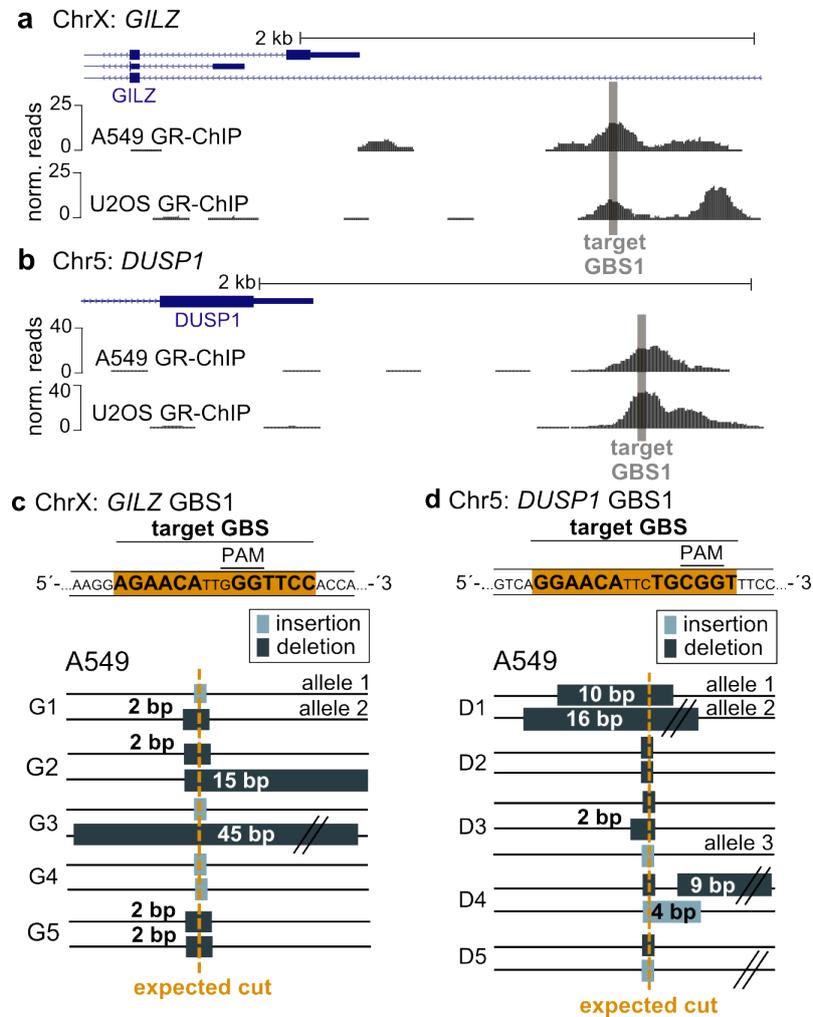


Fig. 3.10: Genomic editing of *GILZ* and *DUSP1* GBS1 in A549 cells. (a, b) Genome browser view showing the genomic location and GR occupancy at *GILZ* (a) and *DUSP1* (b) GBS1 (highlighted in grey) in U2OS and A549 cells. (c, d) Top: Schematics showing the target GBS sequence (highlighted in orange) and the PAM used for CRISPR/Cas9 gene editing. Bottom: Genotyping results of successfully edited clonal cell lines containing CRISPR/Cas9-induced insertions (turquoise) or deletions (dark green) at *GILZ* (c) or *DUSP1* (d) GBS1. The expected location of the Cas9-induced DNA double-strand break (3 bp upstream of the PAM) are indicated.

of *GILZ* GBS1 revealed the presence of two edited alleles, whereas genotyping of *DUSP1* GBS1 resulted in the detection of up to three differentially edited alleles (Fig. 3.10 c, d and Fig. S5.6, S5.7). In concordance with these findings, the presence of two copies of the *GILZ* locus in A549 cells was subsequently confirmed by FISH (Fig. 3.11).

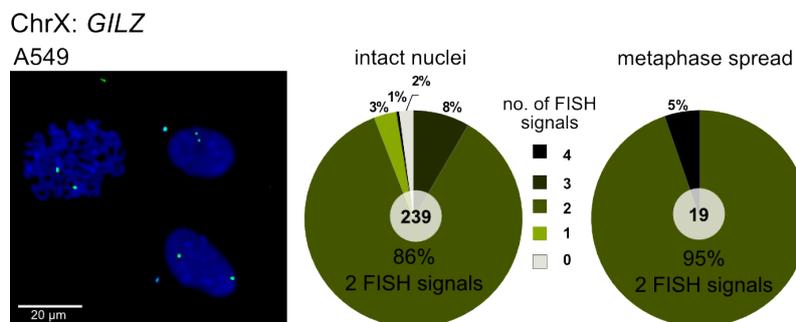


Fig. 3.11: Quantification of *GILZ* alleles in A549 cells. (Left) Representative FISH image of A549 cells using hybridization probes for *GILZ* binding at locus Xq22.3 (green). (Right) Pie charts summarize signals of metaphase spreads or intact nuclei. The total number of analyzed cells is depicted in the inner circle of each pie chart.

Similar to U2OS cells, CRISPR/Cas9-induced gene editing resulted in indels of various lengths, efficiently disrupting the GR binding motif (Fig. 3.10 c, d and Fig. S5.6, 5.7). Thus, just as for the U2OS cell line, at least five clonal A549 cell lines were generated containing disrupted GBSs located upstream of the *GILZ* or *DUSP1* gene, respectively.

To test whether in A549 cells the deletion of these GBSs resulted in an effect on the GR-dependent expression of *GILZ* or *DUSP1*, their level of gene expression in clonal cell lines with deleted GBSs was determined upon hormone induction. Importantly, just as for U2OS cells, in A549 cells both *GILZ* and *DUSP1* GBS1 deletion did neither affect basal gene expression nor the induction of other tested GR-regulated genes, as shown by the expression of *FKBP5* (Fig. 3.12). Comparable to U2OS cells, the deletion of *DUSP1* GBS1 resulted in a small decrease of hormone-induced *DUSP1* expression (73% of wild-type expression). Surprisingly, in contrast to U2OS cells, the deletion of *GILZ* GBS1 in A549 cells showed no detectable effect on the induction of *GILZ* expression (Fig. 3.12).

Thus, in contrast to U2OS cells, in A549 cells only the deletion of *DUSP1* GBS1 but not of *GILZ* GBS1 detectably influenced nearby gene expression.

3. Results

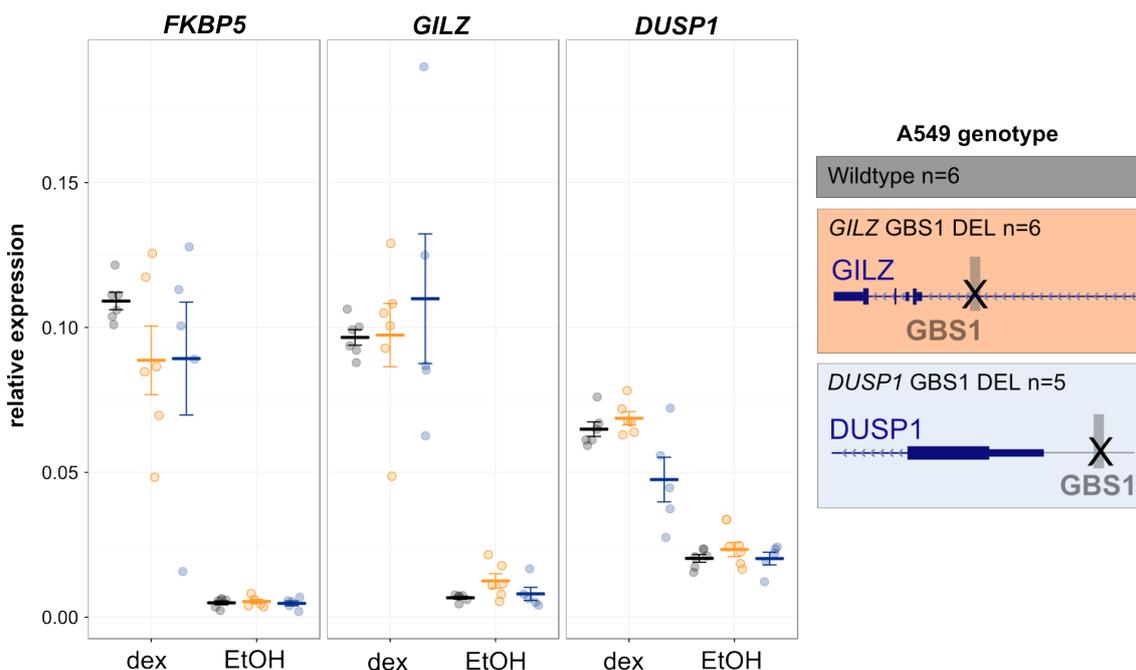


Fig. 3.12: Effect of genomic GBS deletion in A549 cells. Relative *GILZ*, *DUSP1* and *FKBP5* expression of clonal cell lines with deleted *GILZ* GBS1 (n=6) or *DUSP1* GBS1 (n=5) as determined by qPCR after overnight treatment with 1 μ M dexamethasone (dex) or vehicle control (EtOH). A representative from two independent experiments is shown. Horizontal lines indicate the average relative gene expression of individual clonal cell lines. Error bars represent \pm SEM.

3.1.7 *GILZ* GBS1 regulates the expression of *GILZ* transcript variants in a cell type-specific manner

Enhancers can regulate gene expression from distal promoters by DNA-looping [9, 10]. Previously published HiC contact matrices indicate that the *GILZ* locus is part of a TAD, encompassing numerous genes including three *GILZ* transcript variants transcribed from alternative promoters (Fig. 3.13). Notably, the TSSs of transcript variants 2 and 3 are located 1.5 - 2.5 kb upstream of GBS1, whereas the TSS of transcript variant 1 is located in a distance of more than 57 kb from GBS1 (Fig. 3.14 a).

Importantly, until now the effect on *GILZ* expression was detected using a qPCR primer pair binding in an exon shared by the majority of annotated *GILZ* transcript variants (*GILZ* standard primer, Fig. 3.14 a and S 5.11). Given that *GILZ* is composed of at least three transcript variants, the deletion of GBS1 could possibly differentially affect the expression of individual transcript variants.

To test the effect of GBS1 deletion on individual *GILZ* transcript variants, three different transcript variant-specific primer pairs were designed, each binding in the first variant-specific exon (Fig. 3.14 a). In concordance with previous results, in U2OS cells the deletion of GBS1 resulted in a diminished fold induction of *GILZ* transcript variant 2 and 3 in comparison to both a clonal control cell line that is unedited at the *GILZ* locus

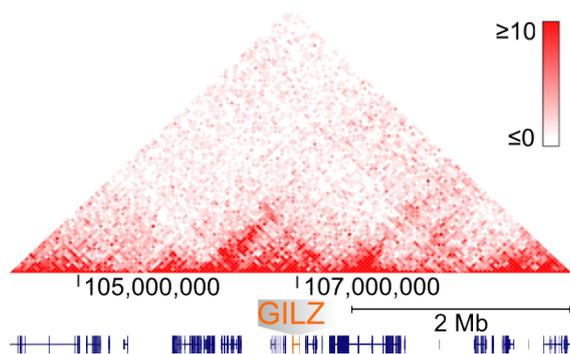


Fig. 3.13: *GILZ* transcript variants are located within the same TAD. HiC contact matrix in A549 cells for a 5 Mb region encompassing the *GILZ* locus at chromosome X (data from [25]).

and to wildtype cells. However, the expression of transcript variant 1 was not induced by GR in U2OS cells (Fig. 3.14 b). Conversely, in A549 cells, the expression of transcript variant 1 was highly induced upon hormone induction. In A549 cells the deletion of *GILZ* GBS1 resulted in a clearly diminished GR-dependent induction of transcript variant 1 (39% of wildtype expression). Surprisingly, although both *GILZ* transcript variant 2 and 3 were induced by GR in A549 cells, the deletion of *GILZ* GBS1 showed no effect on their GR-dependent expression (Fig. 3.14 c). Remarkably, although according to the NCBI RNA reference sequences database [105] the last exon is shared by all of the three transcript variants, no effect on gene expression was observed using the *GILZ* standard primer pair. Hence, additional *GILZ* isoforms must exist that are not detected with the standard primer pair.

In summary, these results indicate that the endogenous regulatory activity of *GILZ* GBS1 differs among different cell types as shown for U2OS and A549 cells. In U2OS cells *GILZ* GBS1 regulates the GR-dependent expression of transcript variant 2 and 3. Whereas, in A549 cells the same GBS regulates the expression of transcript variant 1, whose TSS is located more than 57 kb apart from GBS1. Hence, by deleting GBS1 in A549 cells, the long-range promoter-enhancer interaction between *GILZ* GBS1 and the TSS of *GILZ* transcript variant 1 was identified. Notably, although not regulated by GBS1 in A549 cells, the expression of transcript variant 2 and 3 is nonetheless regulated in a GR-dependent manner.

3. Results

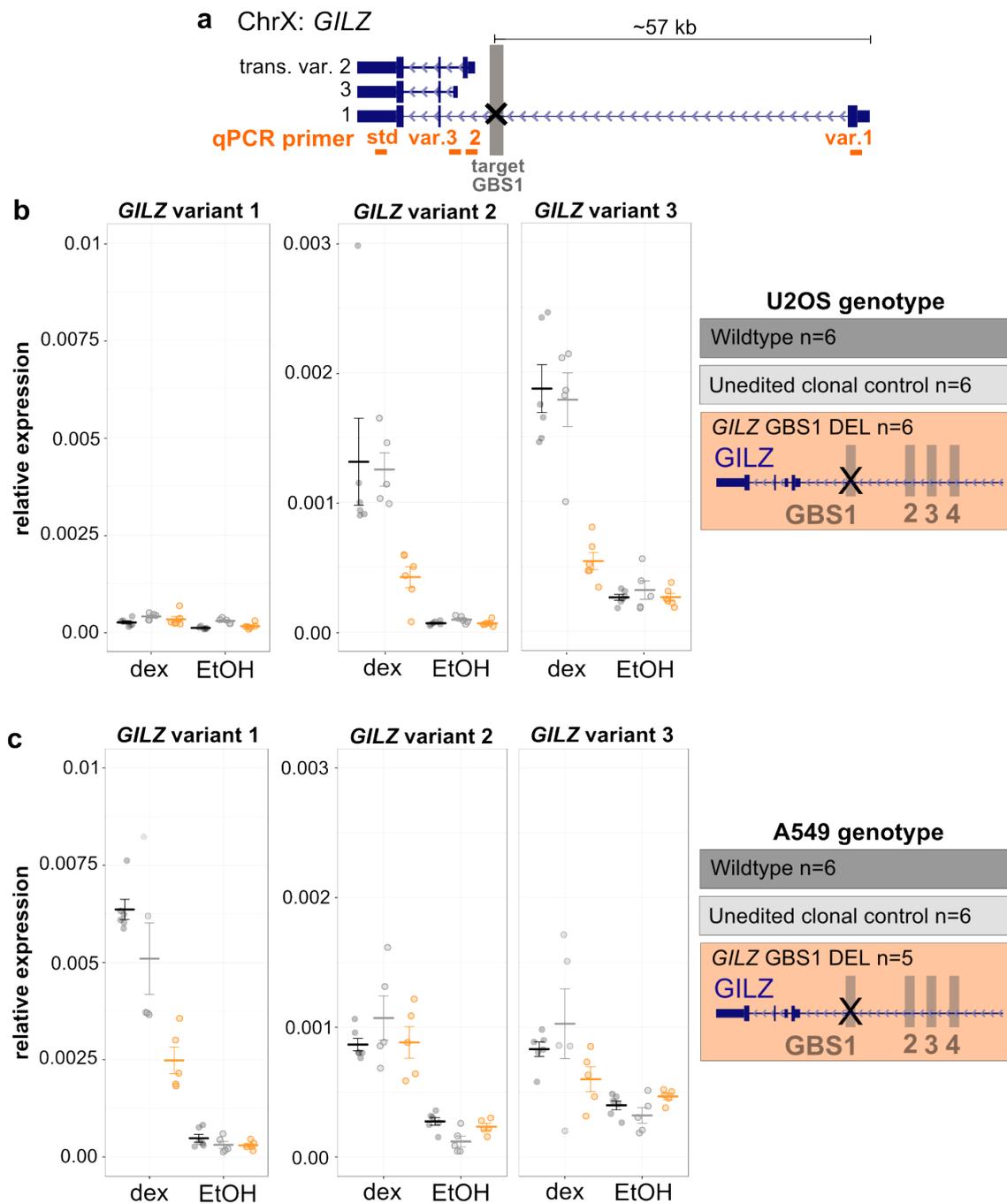


Fig. 3.14: In A549 cells *GILZ* GBS1 is unable to activate the expression of nearby *GILZ* transcript variants. (a) Schematics of *GILZ* transcript variants (not to scale) showing the location of *GILZ* GBS1 and individual primers used for their detection in qPCR. (b, c) Relative mRNA expression of *GILZ* transcript variants in A549 (n=5) (b) or U2OS (n=6) (c) clonal cell lines with deleted *GILZ* GBS1 as determined by qPCR after overnight treatment with 1 μ M dexamethasone (dex) or vehicle control (EtOH). A representative from two independent experiments is shown. Horizontal lines indicate the average relative gene expression of individual clonal cell lines. Error bars represent \pm SEM.

3.1.8 The pattern of H3K27Ac at the *GILZ* locus

Regulatory elements are associated with the occurrence of specific histone modifications at their surrounding chromatin. For example, a genome-wide analysis of enhancer activity and histone modifications revealed that the pattern of H3K27Ac is tightly associated with active enhancer and promoters [21, 22].

To test whether the cell type-specific regulation of *GILZ* transcript variants coincides with differences in the acetylation state of H3K27, the level of H3K27Ac at the *GILZ* locus was determined by ChIP and compared between A549 and U2OS cells. As expected, in both A549 and U2OS cells, *GILZ* GBS1 and the promoter regions of nearby transcript variants 2 and 3 showed an elevated level of H3K27Ac, mirroring their regulatory and transcriptional activity (Fig. 3.15 a). While the level of H3K27Ac at the promoter region of transcript variants 2 and 3 was comparable between A549 and U2OS cells, the level of H3K27Ac at GBS1 was higher in U2OS than in A549 cells (Fig. 3.15 a). Furthermore, although the level of H3K27Ac at the TSS1 was relatively low in both cell lines, A549 cells showed a markedly higher level of H3K27Ac at TSS1 than U2OS cells. Thus, these results match to preceding observations, demonstrating the absence of transcription from the TSS of transcript variant 1 in U2OS cells (Fig. 3.15 a). As a control, a ChIP against histone H3 resulted in a comparable level of precipitation between cell lines (Fig. 3.15 b), excluding differences in nucleosome positioning or ChIP efficiency to account for the observed cell type-specific differences in the pattern of H3K27Ac.

Thus, the observed cell type-specific difference in transcriptional activity of *GILZ* transcript variants is associated with a difference in the level of H3K27Ac at the *GILZ* locus. Notably, GBS1 is functionally relevant for GR-dependent *GILZ* expression in both cell lines. However, U2OS cells show a considerably higher level of active enhancer marks than A549 cells, potentially indicating that H3K27Ac marks are correlated to nearby gene expression rather than long-range regulation.

3. Results

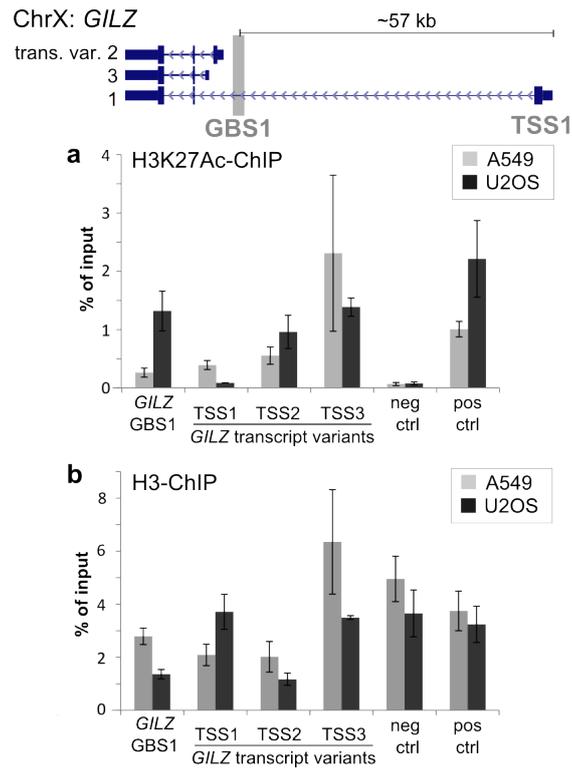


Fig. 3.15: The pattern of H3K27Ac at the *GILZ* locus. (a) Level of H3K27Ac at *GILZ* GBS1 and the promoter region of individual *GILZ* transcript variants as determined by ChIP-qPCR in U2OS or A549 cells treated with 1 μ M dexamethasone. (b) Histone H3 occupancy as determined by ChIP-qPCR in U2OS or A549 cells treated with 1 μ M dexamethasone. (a, b) The graphs show the average of three independent experiments. Error bars represent \pm SEM.

3.1.9 The cell type-specific regulatory activity of *GILZ* GBS1 requires its endogenous genomic context

To investigate the mechanisms that are responsible for the cell type-specific effect of *GILZ* GBS1 deletion, I next tested whether the observed cell type-specific differences in the regulatory activity of GBS1 could be recapitulated in an episomal context using luciferase reporter assays.

For this, the enhancer activity of GBS1 was compared between A549 and U2OS cells in an episomal context using two different strategies. First, in the presence of hormone by mutating GBS1 in reporter constructs containing the endogenous *GILZ* promoter region of transcript variant 2. And second in the absence of hormone, using the same reporter construct by targeted recruitment of an inactivated Cas9 nuclease fused to activation modules (dCas9-VP64 SAM [66]) to GBS1. As expected, in both cell lines wildtype reporter constructs containing the *GILZ* promoter showed a robust hormone-dependent induction of reporter gene expression. Interestingly, the mutation of GBS1 strongly reduced the hormone-induced expression of reporter genes and resulted in a comparable reduction in both cell types, accounting for only 11% and 12% of wildtype response in A549 and U2OS cells, respectively (Fig. 3.16 a). Similarly, guiding the dCas9-VP64 SAM complex to GBS1 was able to induce reporter gene expression in both A549 and U2OS cells to a comparable level (Fig. 3.16 b).

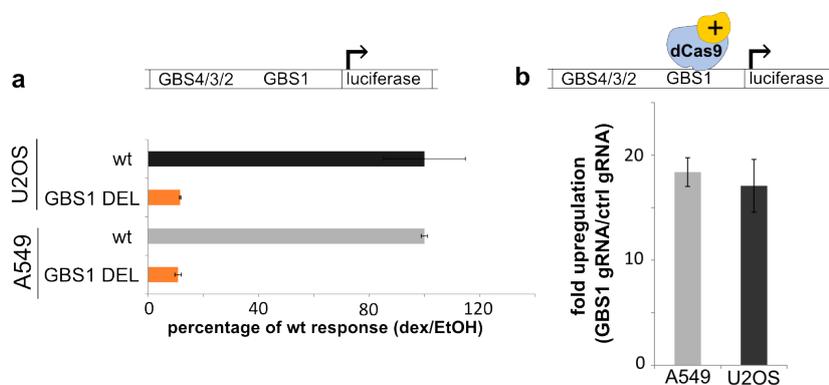


Fig. 3.16: In an episomal context GBS1 is required for enhancer activity of the *GILZ* promoter region irrespective of cell type. Luciferase activity of reporter constructs encompassing *GILZ* GBS1-4 including its endogenous promoter in U2OS and A549 cells. **(a)** Comparison of the effect of GBS1 deletion on hormone-induced reporter gene expression in A549 and U2OS cells. **(b)** Comparison of the induction of reporter gene expression by targeted recruitment of the activator complex, dCas9-VP64 SAM, to GBS1 in A549 and U2OS cells. **(a, b)** The graphs show the average of three independent experiments. Error bars represent \pm SEM.

Hence, both the deletion of GBS1 and the targeted recruitment of transcriptional activators resulted in a comparable effect on reporter expression in both cell types. Therefore, the cell type-specific regulatory activity of *GILZ* GBS1 cannot be recapitulated using reporter

3. Results

constructs, suggesting that the cell type-specific regulatory activity of GBS1 requires its endogenous genomic context. Moreover, these findings indicate that the differential regulatory activity of GBS1 does not result from the requirement of cell type-specific factors, such as the expression of cell type-specific cofactors or differences in endogenous promoter activity.

Conceivably, another possible explanation for the differential effect of GBS1 deletion in A549 and U2OS cells could be a redundancy in enhancer usage that masks the effect of GBS1 deletion on nearby *GILZ* transcript variants. To assess the genomic regulatory activity of GBS1 in isolation, the inactivated Cas9-enzyme fused to VP64 activation modules that can activate gene expression with the help of the synergistic activation mediator (dCas9-VP64 SAM) complex [66] was guided to GBS1. This strategy allowed to study the individual regulatory capacity of GBS1 in isolation by the targeted recruitment of artificial transcriptional activators to GBS1 in the absence of hormone treatment and concomitant influence from additional genomic GBSs.

In U2OS cells the recruitment of transcriptional activator complexes to GBS1 increased the expression of *GILZ* transcript variant 2 by 18-fold in comparison to cells expressing no or a control gRNA (Fig. 3.17 a). In addition, the recruitment of transcriptional activators to GBS1 increased the expression of transcript variant 3 in average by 4-fold above basal levels. Thus, consistent with our previous observations obtained from genomic deletion experiments, transcriptional regulators recruited to GBS1 were able to induce the expression of nearby *GILZ* transcript variants 2 and 3 in U2OS cells (Fig. 3.17 a).

In contrast, in A549 cells activators recruited to GBS1 resulted in a much weaker induction of transcription of nearby *GILZ* transcript variants (Fig. 3.17 a), indicating that in the genomic context these are shielded from transcriptional regulation by *GILZ* GBS1. Furthermore, neither in A549 nor in U2OS cells an upregulation of *GILZ* transcript variant 1 was observed upon recruitment of transcriptional activators to GBS1 (Fig. 3.17 a), suggesting that the long-range interaction between *GILZ* GBS1 and the TSS of transcript variant 1 specifically requires the presence of bound GR or hormone treatment or may be sterically blocked by the presence of the dCas9-complex. However, in both cell lines the direct recruitment of transcriptional activators to the promoter region of transcript variant 1 was able to activate its expression to a comparable level (Fig. 3.17 b), indicating that differences in transfection efficiency and dCas9-VP64 SAM activity between the A549 and U2OS cells are probably not responsible for the observed differences. Thus, although transcript variants 2 and 3 are located in close distance to GBS1, in their genomic context these are isolated from the regulatory activity of GBS1 in A549 cells.

Collectively, these findings indicate that in A549 cells the differential role of GBS1 on *GILZ* expression does not result from a redundancy of regulatory elements or the

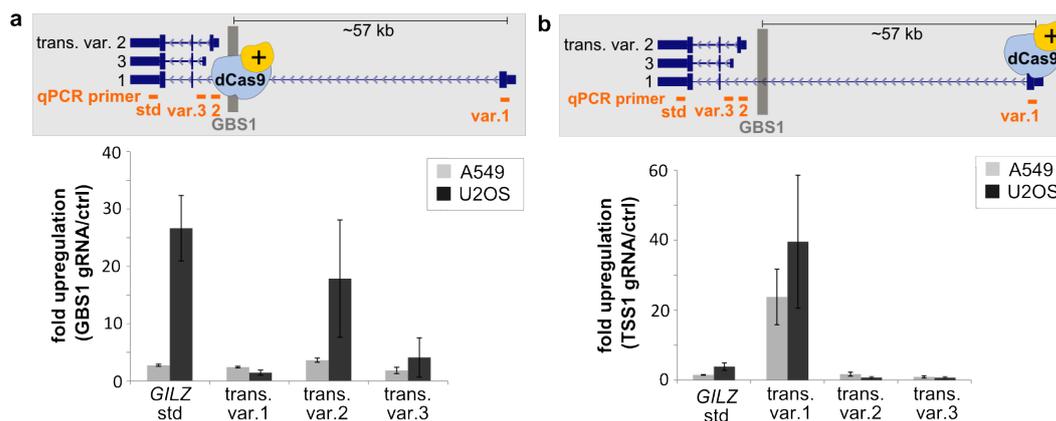


Fig. 3.17: In its genomic context *GILZ* GBS1 is unable to activate the expression of nearby transcript variants in A549 cells. Fold upregulation of *GILZ* transcript variants in A549 or U2OS cells transfected with the activator complex dCAS9-VP64 SAM and a corresponding gRNA binding either at *GILZ* GBS1 (a) or *GILZ* TSS1 (b). The graph shows the average fold upregulation upon targeted recruitment of dCAS9-VP64 SAM to *GILZ* GBS1 or TSS1 relative to untransfected cells and gRNAs targeting other loci in three independent experiments. Error bars represent \pm SEM.

requirement of cell type-specific factors. Furthermore, in an episomal context the cell type-specific regulatory activity of GBS1 could not be recapitulated, suggesting that the endogenous genomic context of *GILZ* GBS1 contributes to its cell type-specific regulatory activity. Specifically for the A549 cell line, the recruitment of a potent artificial activator to GBS1 was also not able to upregulate the expression of nearby transcript variants. Thus, these findings indicate that in its natural context an enhancer blocking mechanism prevents *GILZ* GBS1 from regulating nearby transcript variants.

3.1.10 The cell type-specific regulatory activity of *GILZ* GBS1 coincides with differential CTCF binding and long-range interactions

Enhancers can regulate the expression of distantly located genes by DNA-looping, thereby bringing remote TFBSs and interacting TSSs into close spatial proximity [9, 10].

To test whether differences in the chromatin interaction landscape at the *GILZ* locus are responsible for the cell type-specific regulatory activity of GBS1, circular chromosome conformation capture (4C) experiments were performed in both U2OS and A549 cells, using the genomic region at GBS1 as a viewpoint. Notably, the 4C-profile of both A549 and U2OS cells was highly similar, suggesting that the majority of genomic interactions with GBS1 are not cell type-specific (Fig. 3.18 a). However, in comparison to U2OS cells, A549 cells revealed an increased relative interaction frequency between *GILZ* GBS1 and the region surrounding the TSS of *GILZ* transcript variant 1 (TSS1) (Fig. 3.18 a).

Thus, consistent with previous results from GBS deletion experiments, these experiments indicate that in A549 cells GBS1 and the promoter region of transcript variant 1 are

3. Results

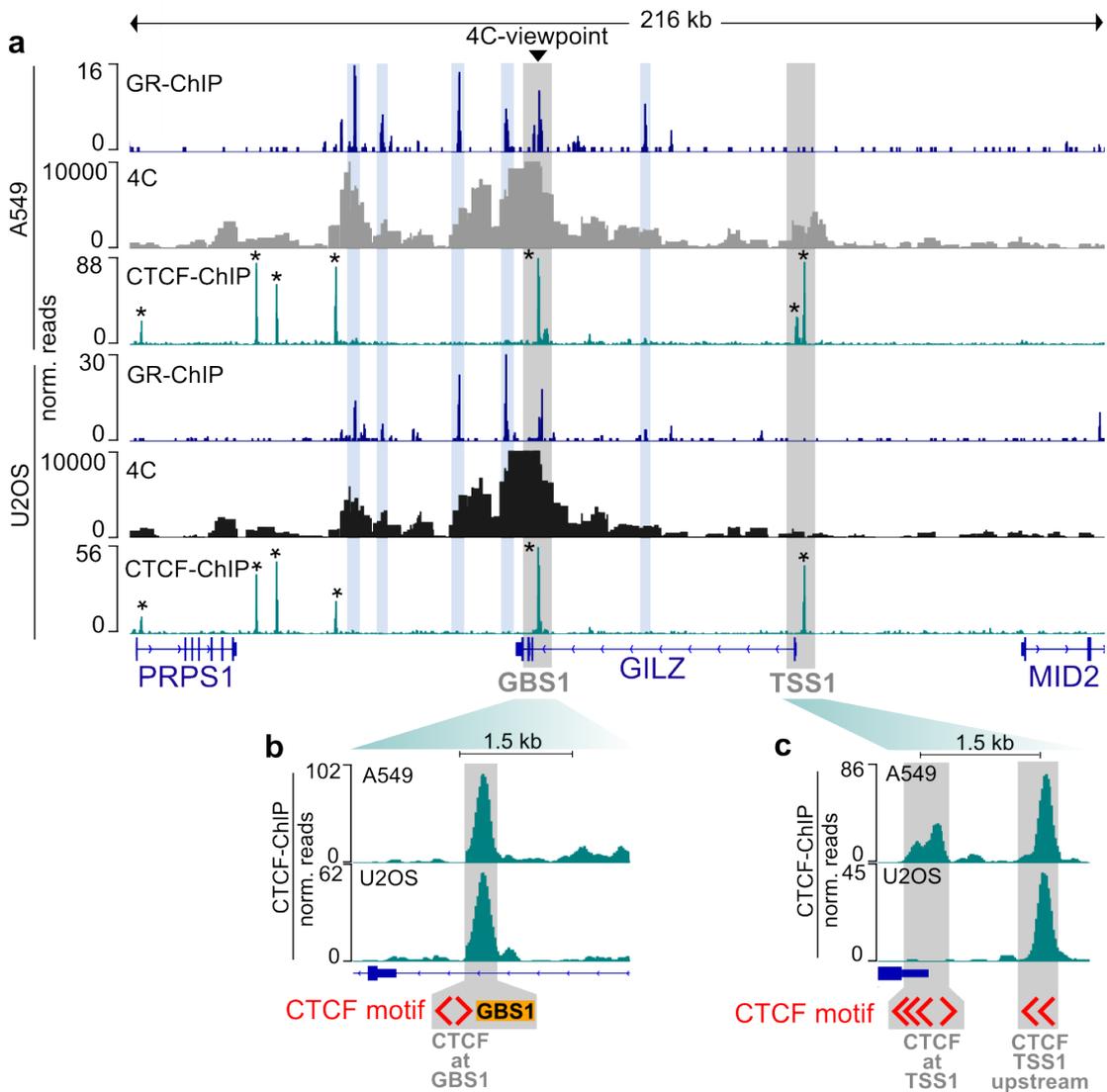


Fig. 3.18: A549 cells but not U2OS cells show increased interactions and cell type-specific CTCF-binding at *GILZ* TSS1. (a) 4C contact profile for a 216 kb region surrounding the *GILZ* locus using GBS1 as a viewpoint. The 4C contact profile shows a representative from two independent experiments and cells were treated with 1 μ M dexamethasone prior to fixation. ChIP-seq tracks below show GR and CTCF binding in U2OS and A549 cells treated with 1 μ M dexamethasone. The genomic region at GBS1 and TSS1 is highlighted in grey, GR-bound loci are highlighted in blue and CTCF-bound loci are marked with an asterisk. (b, c) Zoom-in showing CTCF binding at *GILZ* GBS1 (b) and TSS1 (c) including the individual motif orientation of CTCF motifs located within CTCF-bound peaks.

functionally connected by a long-range DNA-looping interaction, which is not established in U2OS cells. Furthermore, in both A549 and U2OS cells, several genomic regions with a high relative interaction frequency to GBS1 that simultaneously overlap with GR-bound loci (Fig. 3.18 a). Conceivably, these GR-bound regions possibly represent additional interacting GBSs that might cooperatively regulate *GILZ* expression, thereby explaining why the deletion of GBS1-4 resulted in a partial loss of GR-dependent *GILZ* regulation only.

The CCCTC-binding factor (CTCF) plays a key role in the establishment of chromatin organization and DNA-looping contacts [47]. Furthermore, not only CTCF binding but also the orientation of the asymmetric CTCF DNA binding motif was shown to influence DNA-looping interactions and chromosome topology[25, 53, 106].

To investigate whether the cell type-specific long-range interaction between *GILZ* GBS1 and TSS1 correlates with differences in CTCF binding, CTCF binding was determined in both A549 and U2OS cells by ChIP-seq. In both A549 and U2OS cells, several CTCF-bound ChIP peaks were detected in a region of 216 kb encompassing the *GILZ* locus (Fig. 3.18 a, indicated by asterisks). However, in this window only six out of seven CTCF peaks that were bound in A549 cells were also detected in U2OS cells. Specifically, for both A549 and U2OS cells CTCF binding nearby GBS1 was detected (Fig. 3.18 b). Whereas, at TSS1, representing the other end of the putative looping interaction, CTCF binding was only observed in A549 cells (Fig. 3.18 c) and completely absent in U2OS cells. In contrast, upstream of TSS1 a CTCF binding peak was detected that is equally bound in both cell lines (Fig. 3.18 c). Notably, cell type-specific mutations at the CTCF-bound genomic region at *GILZ* TSS1 that could explain differences in CTCF occupancy, were excluded by Sanger-sequencing (Fig. S 5.9).

To investigate the potential role of CTCF motif orientation in mediating the chromatin interaction landscape at the *GILZ* locus, I computationally screened for the presence and orientation of CTCF-motifs at the CTCF-bound peaks located at the *GILZ* locus. Interestingly, several CTCF motifs were found in the vicinity of both *GILZ* GBS1 and TSS1 in both reverse and forward orientations (Fig. S 5.8). In addition, a tandem reverse CTCF binding motif was identified in a distance of 1.5 kb upstream of the TSS of *GILZ* transcript variant 1 (CTCF TSS1 upstream) (Fig. S 5.8).

To test whether CTCF binding at the *GILZ* locus is influenced by the activity of GR or hormone treatment, CTCF binding was assayed in the presence or absence of hormone. In both cell lines, the level of CTCF binding at *GILZ* GBS1, TSS1 and the CTCF-bound region located upstream of TSS1 remained unchanged upon hormone addition, indicating that hormone treatment does not influence genomic CTCF binding at the *GILZ* locus (Fig. 3.19).

3. Results

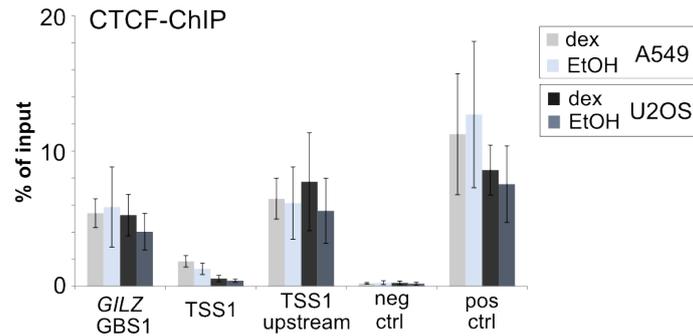


Fig. 3.19: CTCF binding at the *GILZ* locus is not affected by hormone treatment. ChIP-qPCR of CTCF-binding at the *GILZ* locus in U2OS or A549 cells upon treatment with 1 μ M dexamethasone or vehicle control (EtOH). The graph shows the average of three independent experiments. Error bars represent \pm SEM.

In summary, these findings indicate that in A549 cells *GILZ* GBS1 and TSS1 establish a cell type-specific long-range enhancer-promoter interaction to regulate the expression of *GILZ* transcript variant 1 in a GR-dependent manner. Furthermore, the establishment of this interaction coincides with the cell type-specific binding of CTCF to convergent CTCF motifs located at the base of the DNA-loop, thereby possibly facilitating the looping interaction.

3.1.11 Knock-down of CTCF does not alter the GR-dependent regulation of *GILZ* transcript variants

To explore whether CTCF binding at the *GILZ* locus participates in the cell type-specific regulatory activity of GBS1 and the regulation of *GILZ* transcript variants, CTCF was knocked down in A549 cells by siRNAs. In comparison to cells transfected with control siRNAs the knock-down of CTCF resulted in a clear reduction of CTCF mRNA abundance by more than 50% (Fig. 3.20), confirming that siRNAs efficiently targeted CTCF. Notably, CTCF knock-down by siRNAs did not result in a complete loss of CTCF expression, as CTCF mRNA transcripts were still detectable (Fig. 3.20).

Theoretically, if CTCF binding at the bases of the looping interaction between GBS1 and the promoter region of transcript variant 1 is strictly required for its functionality, a knock-down of CTCF should result in a reduction of GR-dependent regulation of *GILZ* transcript variant 1. However, CTCF knock-down did not result in a detectable change of the hormone-dependent regulation of *GILZ* when detected with the standard qPCR primer pair or with primers that target transcript variant 1. Furthermore, if a knock-down of CTCF would result in a loss of its insulator function, a more pronounced induction of transcript variants 2 and 3 could be expected, because these are no longer shielded from the regulatory activity of GBS1. However, in comparison to cells transfected with control siRNAs, CTCF depletion did not result in a detectable induction of transcript variants 2 and 3. Instead, expression of transcript variant 2 was slightly decreased upon

CTCF knock-down. Notably, the expression of *FKBP5*, also regulated by GR, remained unchanged upon CTCF knock-down (Fig. 3.20).

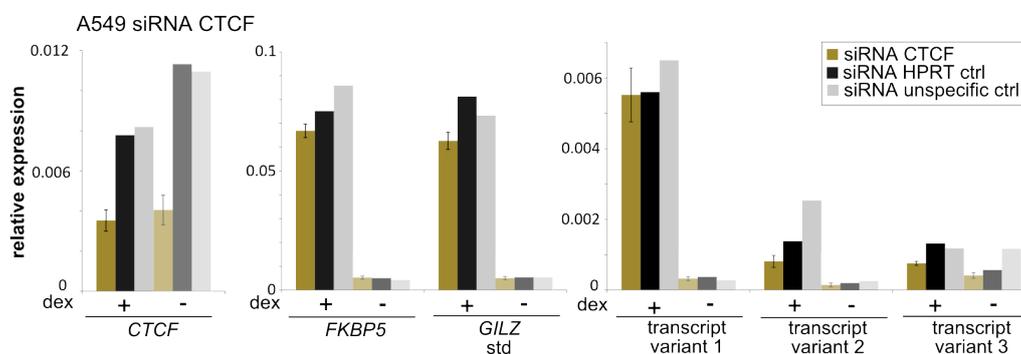


Fig. 3.20: CTCF depletion by siRNAs does not affect the regulation of *GILZ* transcript variants. Relative expression of *GILZ* transcript variants, *CTCF* and *FKBP5* as determined by qPCR after overnight treatment with 1 μ M dexamethasone or vehicle control in A549 cells transfected with siRNAs targeting CTCF or control loci. Error bars represent \pm SEM from three different siRNAs targeting CTCF.

Thus, CTCF depletion by siRNA knock-down did not remarkably affect the GR-dependent regulation of *GILZ* transcript variants. These findings could either indicate that the insulation of GBS1 from nearby transcript variants and its long-range interaction with TSS1 does not depend CTCF binding. Alternatively, given that CTCF knock-down did not completely abolish cellular CTCF expression, the CTCF-knock down efficiency might be insufficient to observe regulatory effects.

3.2 Part 2: Genomic exchange and insertion of glucocorticoid receptor binding sites

3.2.1 GR motif variants differentially affect reporter gene expression

Represented by its consensus sequence, GR preferentially binds to palindromic hexameric repeats separated by a three basepair spacer. However, in the genome GR is able to directly bind to thousands of genomic DNA sequences differing by up to several basepairs from its consensus sequence. Previous studies suggested that the exact sequence of the GR binding motif is able to act as an allosteric effector, thereby modulating the magnitude of GR-dependent gene regulation [83, 84, 86]. Having determined that *GILZ* GBS1 exhibits a profound effect on nearby *GILZ* expression, I wondered whether the exchange of *GILZ* GBS1 into other variants of the GR binding motif is associated with changes in the GR-dependent regulation of gene expression.

Therefore, different GR motif variants situated in the *GILZ* promoter context were tested for differences in their enhancer activity in luciferase assays. To this end, reporter constructs were generated containing the *GILZ* promoter region including either the sequence of the wildtype *GILZ* GBS1 or one of seven GR motif variants (Fig. 3.21 a). These variants were either originally derived from the sequence of genomic GBSs located near GR target genes (e.g. FKBP5, FKBP5-2, SGK, CGT) or have a synthetic origin (e.g. consensus, PAL, TAT). Notably, all of the tested variants were previously shown to activate reporter gene expression in a GR-dependent manner [83]. Furthermore, as previously demonstrated by others, these selected GR motif variants highly differed in their binding affinities for GR [83]. For example, the dissociation constant of PAL was approximately ten times lower than that of *GILZ* GBS1.

Although the basal level of reporter gene expression for individual GR motif variants remained unchanged, their ability of hormone-dependent induction highly differed between constructs, ranging from 9% of wt response for the TAT sequence to 95% for the FKBP5-2 sequence (Fig. 3.21 b). Notably, for all motif variants the exchange at GBS1 resulted in a reduction of hormone-induced expression in comparison to the wildtype sequence, indicating that the sequence of *GILZ* GBS1 is most suitable for gene activation in the *GILZ* promoter context. Importantly, of all functional GR motif variants tested the FKBP5-2 sequence and the PAL sequence showed the most profound differential effect on reporter gene expression. For example, the induction of reporter gene expression of FKBP5-2 was comparable to the wildtype response. In contrast, regardless of its markedly higher affinity for GR, GBS exchange into PAL induced reporter expression to only 32% of wildtype response (Fig. 3.21 b).

Thus, depending on the sequence the exchange of *GILZ* GBS1 into other GR motif variants differentially affected their GR-dependent enhancer activity in reporter assays, suggesting that the exact sequence of the GR motif is able to modulate the capability of

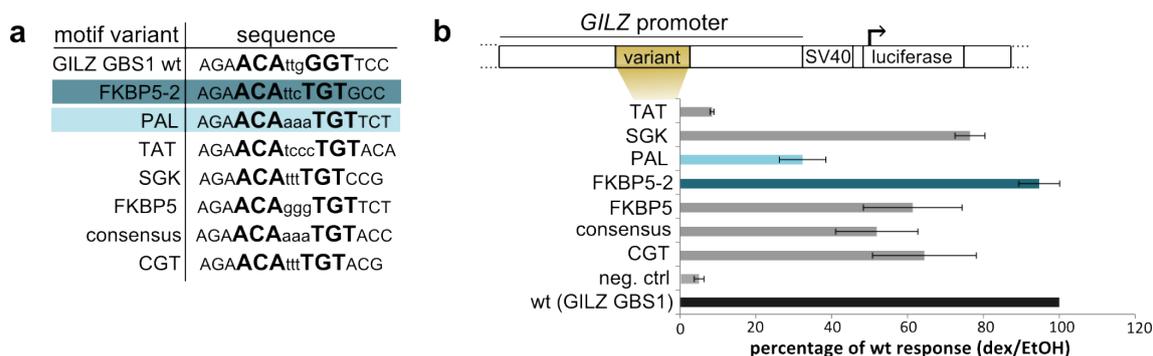


Fig. 3.21: GR motif variants differentially affect reporter gene expression. (a) DNA sequence of eight selected GBS variants, which are either synthetic sequences or derived from genomic GR-bound loci and named based on the closest gene (FKBP5: FK506-binding protein 5, PAL: synthetic palindromic consensus sequence containing AAA spacer, TAT: tyrosine aminotransferase, SGK: serum/glucocorticoid-regulated kinase, consensus: synthetic palindromic consensus sequence 1, CGT: ceramide UDP-galactosyltransferase). (b) Top: Schematics of the luciferase reporter construct containing the *GILZ* promoter region. Individual GBS variants were cloned into the *GILZ* promoter context in exchange of the wildtype *GILZ* GBS1 sequence. Bottom: The graphs show the average percentage of wildtype fold response for the indicated constructs in three independent experiments. Error bars represent \pm SEM. (a, b) The GBS variants selected for genomic exchange at *GILZ* GBS1 are highlighted in turquoise.

GR to regulate gene expression.

3. Results

3.2.2 Genomic exchange of a low-affinity into a high-affinity GR motif variant does not affect *GILZ* expression

To test whether the exchange of GBS1 into another GR motif variant not only differentially affects reporter gene expression but also endogenous GR regulation, GR motif variants were genomically exchanged at *GILZ* GBS1. First, to choose candidate GR binding motifs for genomic exchange, GR motif variants with a strong effect on reporter gene expression were selected. Because the exchange into PAL and FKBP5-2 resulted in the strongest differences in reporter gene expression (Fig. 3.21 b), these two GR motif variants were selected for subsequent genomic exchange.

To this end, clonal cell lines were generated by CRISPR/Cas9 providing a HDR template containing the desired GR motif variant flanked by homology arms for the *GILZ* promoter region. To increase the probability of successful homoallelic exchange by HDR, gene editing was performed in U2OS cells, since both previous GBS deletion experiments (Fig. 3.1 d) and FISH (Fig. 3.3 a) consistently revealed that the U2OS cell line contains only one allele of the *GILZ* locus. The number of generated clonal cell lines containing the desired GR motif variant, was generally lower than indels generated by NHEJ (FKBP5-2 HDR efficiency: 10% of all analyzed clones, PAL HDR efficiency: 16% of all analyzed clones)(Fig. S5.10 and S5.11).

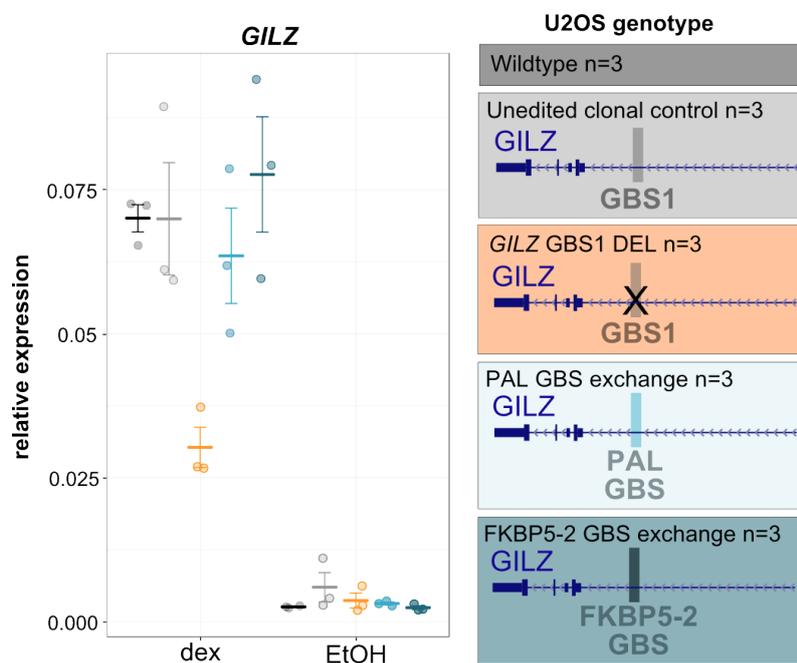


Fig. 3.22: Genomic exchange of *GILZ* GBS1 into other GR motif variants does not affect *GILZ* expression. Relative *GILZ* expression of clonal cell lines containing the FKBP5-2 (n=3) or the PAL (n=3) sequence instead of *GILZ* GBS1 as determined by qPCR after overnight treatment with 1 μ M dexamethasone (dex) or vehicle control (EtOH). A representative from two individual experiments is shown. Horizontal lines indicate the average relative gene expression of independent clonal cell lines. Error bars represent \pm SEM.

Contrary to what was found in reporter gene assays, *GILZ* expression upon hormone treatment was comparable between wildtype and clonal cell lines with GR motif variants, suggesting that the genomic exchange into PAL or FKBP5-2 did not affect endogenous *GILZ* expression (Fig. 3.22). In contrast, previously generated cell lines with deleted GBS1 showed a clear reduction in *GILZ* upregulation (Fig. 3.22). Importantly, for both clonal cell lines with exchanged GR motif variant, neither basal *GILZ* expression nor the induction of other GR-regulated genes, as shown for the expression of *FKBP5* (Fig. S5.12) was changed.

Given that, exchanging *GILZ* GBS1 into PAL results in the exchange of a low-affinity GBS into a high affinity GBS [83], the effect of GBS exchange might only be prevalent under low concentrations of hormone. Thus, I next tested the effect of GR motif exchange on gene expression at low hormone concentrations. Importantly, at a concentration of 10 nM dexamethasone, the level of *GILZ* expression was comparable to preceding results using 1 μ M (Fig. 3.22), indicating that at a concentration of 10 nM the response already reached its plateau. Furthermore, for the two dex concentrations tested, *GILZ* expression increased with increasing amounts of hormone. However, similar to what was observed under high hormone concentrations, the exchange of GBS1 into GR motif variants did not result in a detectable change of *GILZ* expression, suggesting that endogenous gene expression is not affected from the exchange of a low affinity into a high affinity GBS.

Together, the genomic exchange of the regulatory active *GILZ* GBS1 into different GR motif variants does not detectably affect endogenous *GILZ* expression. Thus, allosteric regulation of GR induced by GBS sequence variants does not influence transcriptional output at this particular genomic locus.

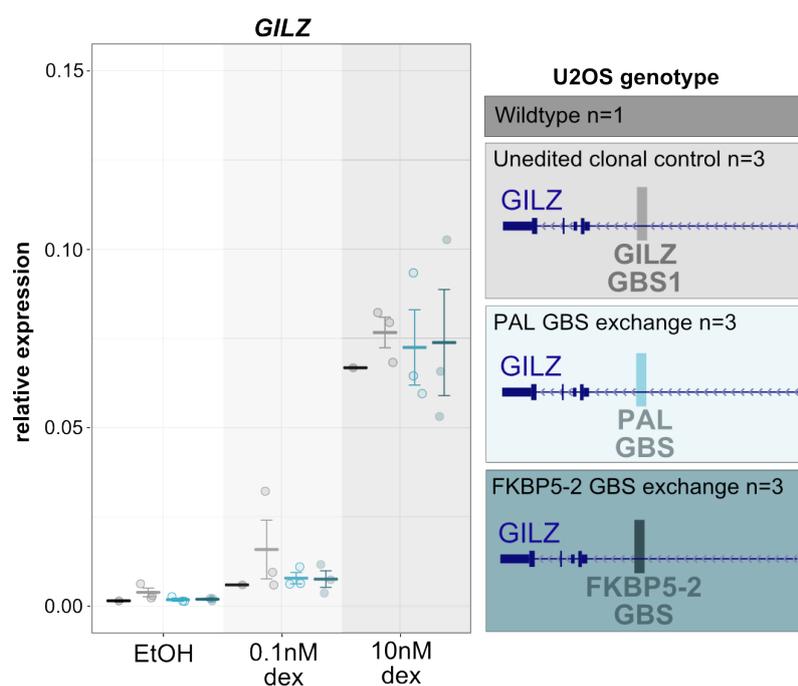


Fig. 3.23: Genomic exchange of GR motif variants does not affect *GILZ* expression at low concentrations of hormone. Relative *GILZ* expression of clonal cell lines containing the FKBP5-2 (n=3) or the PAL (n=3) sequence variant instead of *GILZ* GBS1. Gene expression was determined by qPCR after overnight treatment with 0.1 nM or 10 nM dexamethasone (dex) or vehicle control (EtOH). Horizontal lines indicate the average relative gene expression of individual clonal cell lines. Error bars represent \pm SEM.

3.2.3 The introduction of a single promoter-proximal GR binding sequence is sufficient to regulate gene expression in a GR-dependent manner

Based on observations made in GBS deletion experiments, the genomic expression of *GILZ* is under the control of a regulatory hub composed of multiple promoter-proximal and distal GBSs. Therefore, the regulatory role of GR motif variants present at only one of these GBSs might potentially be masked by the regulatory activity of the remaining interacting GBSs.

To circumvent this problem, and to generate a model to study the regulatory role of GR motif variants in an isolated context, a single GR binding sequence was placed proximal to the promoter of a gene endogenously not expressed. As a target locus, the interleukin 1 receptor type 2 (*IL1R2*) gene was chosen because of two main reasons.

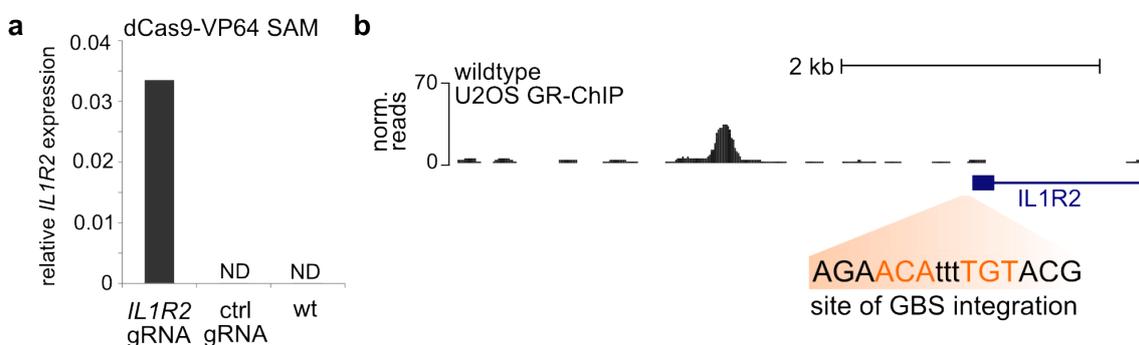


Fig. 3.24: Genomic location of the site for GR binding sequence integration. (a) Relative *IL1R2* mRNA expression as detected by qPCR upon transfection of U2OS cells with the dCas9-VP64 SAM system using either a gRNA targeting the *IL1R2* promoter, a control region located at another chromosome or in untransfected wildtype cells. (b) Genome browser view showing the genomic location of the locus for GR binding sequence insertion at the promoter region in a 8 bp distance to the TSS of *IL1R2*. The GR ChIP-seq track above shows GR occupancy in a 6 kb window including the site of integration.

First, while *IL1R2* is highly expressed in B- and T-cells, the gene is completely silenced in U2OS cells (Fig. 3.24 a). Hence, introducing a GBS directly at the promoter of *IL1R2* in U2OS cells, could therefore result in a qualitative (silenced vs. expressed) rather than quantitative change in gene expression, which are harder to detect when a gene is already expressed.

Second, gRNA sequences that efficiently target the promoter region of *IL1R2* were already published by others [66] and also upregulated *IL1R2* expression using the dCas9-VP64 SAM system. Hence, these gRNAs bound to its target sequence in U2OS cells and could therefore be used to target the *IL1R2* locus for genomic integration. Given that in reporter gene assays, the CGT sequence (Fig. 3.21 a) consistently resulted in a high induction of gene expression either when placed in front of a minimal promoter [83] or when tested in the *GILZ* context (Fig. 3.21 b), the CGT sequence was chosen for genomic integration. Notably, in unedited U2OS cells a GR binding peak is located approximately

3. Results

2 kb upstream of the chosen site for GBS integration (Fig. 3.24 b), indicating that this nearby GBS is bound by GR but its binding is not sufficient to induce *IL1R2* expression in a GR-dependent manner.

Using CRISPR/Cas9 single-cell derived clonal cell line were generated by providing a HDR template containing homology arms for *IL1R2*. Notably, successful integration by HDR was relatively low (4% of all analyzed clones) resulting in only one clonal cell line in which the GR binding sequence was successfully inserted. Furthermore, for the generated clonal cell line the selected GR binding sequence was introduced at only one allele, whereas the second allele was edited by NHEJ resulting in a 6 bp deletion (Fig. S5.13). Interestingly, in the generated clonal cell line expression of *IL1R2* was considerably up-regulated upon addition of hormone, resulting in a 13-fold induction of gene expression (Fig. 3.25). Notably, in the absence of hormone, expression of *IL1R2* was slightly elevated in comparison to wildtype cells. But given that *IL1R2* expression levels were generally low, this effect could also result from difficulties in the quantification of lowly expressed genes by qPCR. Specifically for wildtype cells and in the absence of hormone for clonal cell lines containing a GBS at *IL1R2*, dissociation curves of PCR products indicated that the qPCR amplification was not specific, arising as artifacts from qPCR primers. In contrast, in wildtype cells both in the presence and in the absence of hormone *IL1R2* expression was not detectable, demonstrating that the induction of gene expression is specific for the presence of a promoter-proximal GR binding sequence (Fig. 3.25).

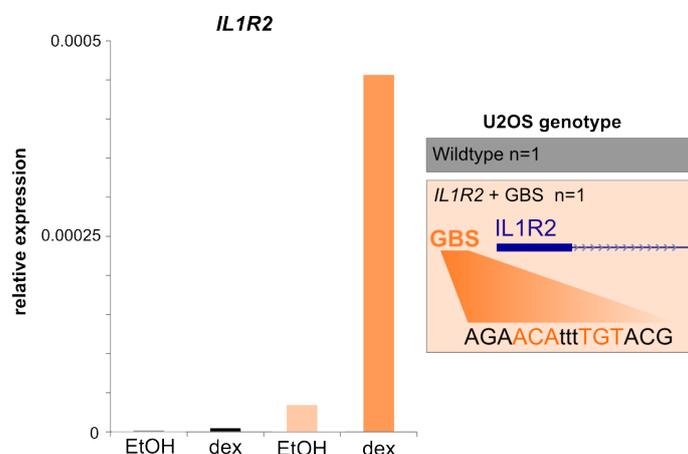


Fig. 3.25: *IL1R2* becomes regulated by GR upon introduction of a single promoter-proximal GR binding sequence. Relative *IL1R2* expression in U2OS wildtype cells and a clonal cell line containing an artificially introduced GR binding sequence located upstream of *IL1R2* as determined by qPCR after overnight treatment with 1 μ M dexamethasone (dex) or vehicle control (EtOH).

Together these data indicate that, first, GR is able to bind to the artificially introduced GR binding sequences and second that its binding induces *IL1R2* expression. Thus, although *IL1R2* is not expressed in the U2OS cell line, the insertion of a single GR binding

sequence in its promoter region is sufficient to induce gene expression in a GR-dependent manner. Given that in the generated cell line, the expression of *IL1R2* depends on the regulatory activity of the introduced GR binding motif, it represents an ideal model to study the effects of GR motif variants in future experiments.

4 Discussion

4.1 Reporter assays only partially simulate endogenous enhancer activity

Understanding the mechanisms that regulate the complex patterns of gene expression requires a detailed knowledge of the location and activity of the regulatory regions that are responsible for the expression of a specific target gene. For decades, reporter gene assays served as a gold standard to investigate the activity of putative regulatory elements. In these assays, the region of interest is typically studied under the control of a minimal promoter located on transiently transfected reporter plasmids. However, my studies and those of others indicate that reporter assays are not able to fully mimic the endogenous activity of regulatory regions [107, 108]. Exemplified by the loci studied in this work, reporter assays lack several layers of regulatory information naturally provided in the genomic context. Most importantly, this includes regulatory information from the surrounding chromatin environment, from the chromosomal 3D architecture and from the integration of combinatorial signals from multiple regulatory regions and factors.

For example, in reporter assays GBS2-4 contributed only little to the overall enhancer activity of the *GILZ* reporter construct (Fig. 3.9). In contrast, genome editing in the endogenous genomic context revealed that *GILZ* GBS1-4 cooperatively interact as a single functional unit that is only active when both GBS1 and GBS2-4 are intact. Hence, reporter assays are not able to fully reflect the endogenous cooperative interplay of GBS1-4. Indeed, due to its plasmid-based nature, conventional reporter assays lack several layers of regulatory information. This includes regulatory information from the surrounding chromatin environment, e.g. the presence of specific histone modifications. However, besides these differences, reporter assays may additionally be influenced by position-dependent effects. Therefore, the regulatory activity of GBS2-4 that is located in a greater distance to the minimal promoter than GBS1 might be generally under-estimated in reporter assays. Thus, another possible explanation for the diminished activity of GBS2-4 in reporter assays could be the position of the GBSs with regard to the minimal promoter of the reporter construct.

Similarly, whereas in reporter constructs the exchange of GBS motif variants resulted in differences in reporter activities, these effects were absent upon exchange in the genomic context (Fig. 3.21 b and Fig. 3.22). Due to limitations in vector size, reporter assays lack

the presence of distal possibly cooperatively interacting TFBSs that are present in the genomic context. Specifically for the *GILZ* locus, mutational analysis showed that in its genomic context *GILZ* is regulated by multiple promoter-proximal and distal enhancers. Therefore, sequence-specific differences in regulatory activity could be more prevalent in reporter assays because these lack the genomic 3D context in which the regulatory region is embedded.

Furthermore, in reporter assays the *GILZ* promoter region was similarly active in both A549 and U2OS cells (Fig. 3.16 a, b). In contrast, in the genomic context GBS1 regulated nearby gene expression in a cell type-specific manner. Hence, reporter assays are not able to reflect the cell type-specific activity of *GILZ* GBS1. Indeed, 4C experiments indicated that the establishment of cell type-specific looping interactions contribute to the cell type-specific regulatory activity of GBS1 (Fig. 3.18 a). However, because reporter constructs are tested in a heterologous context, long-range looping interactions cannot be captured in luciferase reporter assays. Therefore, the inability of reporter assays to reflect the cell type-specific activity of GBS1 most likely results from a lack of genomic 3D architecture and long-range DNA-looping interactions.

Supporting these profound differences, a previous study systematically compared the activity of genomically integrated reporter constructs to their episomal activity [108]. Of note, and in contrast to what was tested in this work, the study tested the activity of enhancers outside their native context upon genomic integration of entire reporter constructs including its minimal promoter. Nevertheless, the activity between genomic and episomal reporter constructs highly differed, suggesting that there are substantial differences between these assays.

Thus, although reporter assays can provide an important indication of the regulatory activity of candidate regulatory regions, reporter assays cannot deterministically reflect their endogenous activity and function. Consequently, to investigate the endogenous functional relevance of TFBSs, requires a general shift from reporter assays to an analysis in the endogenous genomic context.

4.2 Genomic editing of TFBSs using CRISPR/Cas9

Today, the recent advance in genome editing tools provides the opportunity to interrogate the function of non-coding regulatory elements directly in their endogenous genomic environment in a highly efficient and flexible manner.

In this work, I investigated the relationship between TF binding and the regulation of gene expression by editing several non-coding regulatory regions using the CRISPR/Cas9 system. Although observed editing efficiencies were generally high (Fig. S5.1, S5.2, S5.6 and S5.7), these differed depending on both locus and cell type. This is in concordance

with other studies, reporting that the cleavage activity of a gRNA depends on the gRNA sequence and its associated secondary structure [109–111]. Furthermore, locus-specific factors influencing DNA accessibility that may differ among cell types, were shown to additionally affect the activity of gRNAs [111].

In this work, non-coding regulatory regions were edited by three different strategies. First, individual GBSs were edited by NHEJ-mediated DNA repair, introducing small indels within the GR binding motif. Importantly, in contrast to most other studies this experimental setup allows for a fine-scaled investigation of individual TFBSs. Second, multiple genomic GBSs were simultaneously deleted by removing larger DNA fragments containing both the GBSs and the surrounding DNA context. And third, GBSs were edited using a HDR template introducing point mutations at positions most critical for GR binding. Notably, as previously described by others [61, 112], the majority of generated clonal cell lines contained small indels at the site of the CRISPR/Cas9-induced DNA double strand break (Fig. 3.1 d, e and Fig. 3.10 c, d), indicating that DNA-repair by NHEJ was most prevalent in the studied cell types. In contrast, gene editing efficiencies for the introduction of a precise modification by HDR were generally low, occurring in only a small fraction of edited cells (Fig. S 5.11, S5.12, S5.13).

Of note, especially in case of the fine-scaled editing of TFBSs by small indels, TFs might theoretically still be able to bind to the edited site, even though their sequence differs from the original binding motif. However, a functional contribution of the GBSs edited in this work is unlikely for several reasons.

First, both small one basepair modifications as well as bigger genomic deletions disturbed the functionality of the edited GBSs to the same extent. For example, both the complete deletion of *GILZ* GBS1-4 by removing a DNA fragment of approximately 600 bp in size and the modification of individual GR binding motifs at positions critical for GR binding resulted in a comparable reduction of GR-induced *GILZ* expression (Fig. 3.6). Therefore, although residual GR binding at the edited motifs cannot be completely excluded, the edited GBSs were impaired in their functional contribution to gene expression.

Second, previous studies [88, 100], in concordance with this study, showed that the mutation of individual GBSs within reporter constructs containing either the *GILZ* or *DUSP1* promoter region was sufficient to abrogate their enhancer activity (Fig. 3.9). Hence, small modifications within the GR binding motif also likely affect transcriptional regulation by GR in its genomic context.

Third, represented by its consensus motif GR can bind as a dimer to hexameric imperfect palindromic repeats separated by a three basepair spacer, in which each GR monomer binds a repeat motif (Fig. 3.2 a)[78]. Thus, targeting the spacer sequence in between these repeats by CRISPR/Cas9 generates indels within the spacer sequence. These indels shifts the distance between the GR half-site binding motifs, most probably impeding GR binding as a dimer. Indeed, editing of GBSs by NHEJ reduced the computationally

predicted GR binding affinity for all except of one generated clonal cell line (Fig. 3.2 b). Notably, target gene expression in this particular clonal cell line was similarly reduced as in all other generated clonal cell lines edited at this locus. Thus, indels introduced within the GR binding motif most probably diminish its affinity for GR. Accordingly, in ChIP experiments GR binding was reduced at the edited GBSs but not completely abolished (Fig. 3.7). Of note, it cannot be excluded that GR is still able to bind either as a monomer or dimer to the edited GBS motifs. However, the residual GR binding observed in ChIP experiments could also result from the relatively low resolution of conventional ChIP-seq. For example, both *GILZ* and *DUSP1* GBS1 are located in a relatively broad GR-bound ChIP peak, probably encompassing several direct or indirect GBSs (Fig. 3.1 a, b). Conceivably, in ChIP experiments the effect of editing a single GBS on GR binding might be masked by the presence of nearby GR-bound GBSs. Furthermore, the lack of a complete abolishment of GR binding at the edited GBSs could also reflect its interaction with other GR-bound loci in a regulatory hub. Thus, GR binding might emerge as an artifact from interacting GBSs by covalent linkage during the fixation step of the ChIP procedure.

Together, these findings indicate that small indels located within the GR binding motif can effectively impair the functionality of GR binding. Therefore, genomic editing of TFBSs by CRISPR/Cas9 provides the opportunity to study the endogenous regulatory function of non-coding regions and their contribution to gene expression in their natural genomic context.

4.3 Not all bound GBSs contribute to the GR-dependent regulation of nearby genes

In recent years, ChIP experiments provided a wealth of information regarding the location of genomic binding sites for a plethora of TFs [1]. In fact, genomic TF binding and the regulation of gene expression are statistically clearly connected [113, 114]. However, ChIP-seq experiments usually result in several thousands of peaks for an individual TF, whereas TF perturbations affect only a small number of genes [11, 12]. This indicates that only a subset of all bound TFBSs are actually productive for gene expression and raises the question how TFBSs are specifically linked to the regulation of their target genes. By dissecting multiple GBSs at the *GILZ* locus, I found two possible mechanisms that can prevent a bound GBS from regulating the expression of a nearby gene.

First, exemplified by the cell type-specific activity of *GILZ* GBS1 (Fig. 3.14), enhancer blocking and changes in the genomic 3D architecture can redirect the activity of a GBS to other genomic loci. Thus, cell type-specific enhancer blocking and rewiring of enhancer-promoter contacts can prevent a generally active enhancer from acting on a nearby gene. In fact, given that enhancers are generally able to act over long distances,

enhancer blocking can provide a means to ensure enhancer-promoter specificity. Thereby, enhancer blocking can help to structure and organize the genomic regulatory landscape and restrict inappropriate enhancer-promoter interactions. Mechanistically, many recent studies and experimental evidence obtained in this work indicate that CTCF binding and the formation of closed loop domains are involved in the establishment of enhancer blocking [115, 116](discussed in more detail in section 4.6).

Second, for some enhancers simultaneous GR binding at multiple GBSs might be required to cooperatively induce gene expression. For example, the *GILZ* enhancer containing GBS1-4 regulates *GILZ* expression cooperatively as a single functional unit, but only if both GBS1 and GBS2-4 are functionally intact. Hence, the absence of the correct array of cooperative interactions could explain why some TFBSs might lack the ability to contribute to gene expression. In line with these findings, a recent study reported a low general correlation of genomic GR binding and regulatory activity, as the majority of GR-bound regions showed only minor activity in reporter assays [117].

Although, not all GBSs that were edited in this study contributed to the GR-dependent expression of the nearby gene, all selected GBSs were generally productive and contributed to GR-dependent gene expression. Nevertheless, genomic deletion of TFBSs was repeatedly reported to result in no or only marginal effects on gene expression [118–120]. Importantly, in the endogenous cellular context, TF binding patterns can change dynamically depending on cell type or during differentiation [121]. Furthermore, additional features such as the pattern of histone modifications, the availability of specific co-factors, the status of DNA-methylation or chromatin accessibility can also change dynamically or depending on cell type [21, 122]. Therefore, these features might represent an important mechanism to shape and finetune the individual regulatory activity of a bound TFBS. In fact, several enhancers whose deletion did not result in an apparent effect under standard conditions were reported to affect gene expression only during a specific cell-stage, in a specific cell type or environmental condition [118, 123–127]. Hence, the absence of a detectable effect on gene expression for a given TFBSs does not generally imply its insignificance. Instead of this qualitative classification, the regulatory activity of individual TFBSs depends on its biological context.

In summary, both enhancer blocking and combinatorial binding at several TFBSs can explain why a subset of TFBSs are not able to functionally contribute to nearby gene expression. Thus, although there is a well-established statistical connection of promoter-proximal TF binding and gene regulation, simple proximity cannot be used to deterministically predict the regulatory activity of a bound TFBS.

4.4 Multiple GBSs cooperatively regulate *GILZ* expression

Given that in the genome the number of enhancers by far outnumbers the number of promoters [128], an important question remains how multiple enhancers cooperatively interact to regulate gene expression.

Here, I could show that within the promoter-proximal *GILZ* enhancer, multiple GBSs are required to achieve the correct quantitative expression of *GILZ* (Fig. 3.6). Supporting this finding, reporter gene assays presented here and previously performed by others [88] additionally confirmed, that the enhancer activity of the *GILZ* promoter depends on multiple GBSs (Fig. 3.9). Furthermore, individual and combinatorial deletion of the GBSs located within the enhancer revealed that in its genomic context GBS1-4 do not act in an additive manner. Instead, the GBSs regulate *GILZ* expression as a single functional unit. The disruption of this functional unit, for example by deletion of one of its parts, leads to a complete loss of its entire regulatory activity (Fig. 4.1). Therefore, the contribution of its individual parts is only meaningful for gene expression, when the functional unit exists in its entirety. Hence, the individual GBSs within the *GILZ* enhancer regulate *GILZ* expression in a cooperative manner as a single functional unit.

Theoretically, the requirement for multiple bound TFBSs for an enhancer to be active could ensure that not all genomic sites actively contribute to gene expression. Given that, for many TFs, including GR, ChIP-seq experiments result in several thousands of bound genomic TFBSs [90, 103], such an assembly could prevent inappropriate regulatory activity by ensuring that a subset of bound genomic sites does not actively contribute to gene expression. Furthermore, rendering a subset of TFBSs in an unproductive but preset state could also allow to rapidly respond to a changing environment and to dynamically fine-scale the contribution of individual TFBSs by expanding or limiting the binding of interacting factors.

However, the regulation of gene expression using a functional unit of multiple TFBSs most probably does not represent a general operating principle of TFBSs. As it is shown here for the *IL1R2* locus, the single insertion of a promoter-proximal GBS is sufficient to induce the GR-dependent regulation of an endogenously silenced gene. In support of this finding, both the presence of a single GBS in reporter assays and its genomic integration in front of reporter genes was reported to be sufficient for GR-dependent regulation [86, 117]. Furthermore, given the fact that a number of studies provide evidence for enhancer redundancy [34, 124], cooperativity [118] and additivity [32, 36], most probably different modes of interaction among multiple TFBSs exist and might dynamically change during development or in response to environmental stimuli.

Notably, up to this date most studies investigated the mechanisms of combinatorial regulation by multiple TFBSs either in the context of reporter assays or within the genomic context of super-enhancers. For example, a recent study showed that in its genomic

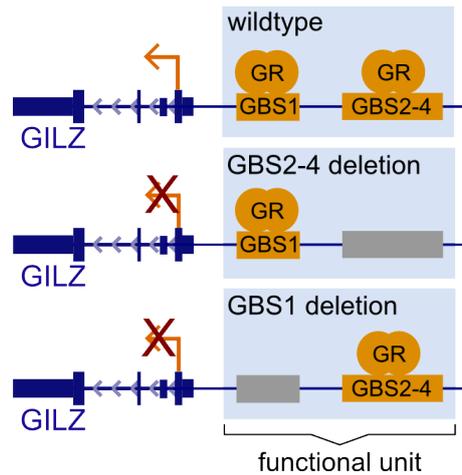


Fig. 4.1: GBS1 and GBS2-4 cooperatively regulate *GILZ* expression as a single functional unit. Binding of GR at either GBS1 or GBS2-4 in isolation, as mimicked by CRISPR/Cas9 mediated deletion, is not able to contribute to the induction of *GILZ* expression. In contrast, GR binding at both GBS1 and GBS2-4 is required for their cooperative regulation and maximum GR-dependent induction of gene expression as observed in wildtype cells.

context multiple enhancers of the *Wap* super-enhancer act independently but provide distinct functional contributions to the overall enhancer activity [36]. Similarly, the genetic dissection of the α -globin super-enhancer showed that its individual enhancers act in an additive fashion [32].

Theoretically, the cooperative regulation of gene expression by multiple TFBSs may be associated with direct protein-protein interactions of the interacting TFs and mutual stabilization by cooperative binding in a regulatory hub. Being part of such a regulatory hub, the disruption of one TFBS might therefore destabilize the binding of an interacting TF. In line with this hypothesis, directly bound GBSs were proposed to interact with remote AP-1 binding sites that could stabilize GR binding by protein-protein interactions [117]. Moreover, for individual loci the genomic deletion of TFBSs was shown to destabilize nearby TF binding [118]. In contrast to these findings, this work provides experimental evidence that cooperativity among multiple TFBSs does not necessarily occur on the level of DNA binding. In fact, as detected by ChIP, the deletion of *GILZ* and *DUSP1* GBSs did not affect GR binding at interacting or neighboring sites (Fig. 3.7 and Fig. 3.8).

Since the data presented here suggest that cooperatively interacting GBSs are bound independently, their cooperativity in transcriptional regulation must be achieved at a later stage than DNA binding. As already shown for other TFs, TFs might cooperatively regulate gene expression by cooperative recruitment of other co-factors [129] or by the cooperative recruitment of components of the basal transcriptional machinery [130, 131]. Thereby, cooperative recruitment may ultimately achieve a higher local concentration of regulatory factors. Conceivably, in the case of the cooperative interaction of GBS1-4 the presence of all GBSs may be required to reach a threshold level of regulatory factors to

ultimately initiate transcription by RNA Pol II.

Importantly, upon the deletion of promoter-proximal GBSs, both *DUSP1* and *GILZ* were still regulated by GR (Fig. 3.4 and Fig. 3.12), suggesting that additional distal GBSs regulate their expression. Therefore, both genes are most probably not only regulated by multiple TFBSs but also by multiple enhancer regions. These findings are in line with other published data, indicating that cooperative regulation of gene expression by multiple regulatory regions represents a common phenomenon [26, 123]. Ultimately, the combinatorial regulation by multiple distal and promoter-proximal enhancers may provide a means to ensure regulatory robustness not only against environmental perturbations, but also against genetic variation, simultaneously providing a playground for evolutionary diversification [124]. In fact, the 4C profile using *GILZ* GBS1 as a viewpoint revealed several other genomic loci that contain GR-bound GBSs showing a high relative interaction frequency with the promoter region of *GILZ* transcript variant 2 and 3 (Fig. 3.18 a). Whether these contribute to the GR-dependent regulation of *GILZ* and in which manner these multiple enhancer regions interact to establish the robust induction of *GILZ* expression could be determined by genomically deleting candidate GBSs in future experiments. Furthermore, many GR-regulated genes can additionally perceive regulatory information from other signaling pathways [132]. However, in this work the combinatorial interaction by multiple TFBSs was analyzed regarding the interaction of homo-typic TFBSs for the same TF. Conceivably, combinatorial regulation of GR-regulated gene expression is additionally influenced by the interaction with other TFs, such as AP-1 or NFkB, thereby integrating several regulatory signals to contribute to the cellular plasticity of gene expression.

Together, combinatorial regulation by multiple distal and promoter-proximal enhancers most probably represents a common mechanism to ensure regulatory robustness and plasticity for gene expression. Within a single enhancer multiple TFBSs can cooperatively regulate target gene expression as a single functional unit, requiring the presence of multiple GBSs to be active. Theoretically this arrangement can simultaneously ensure that only a subset of TFBSs establish appropriate regulatory activity. Furthermore, this work shows that cooperative regulation by multiple TFBSs must not necessarily occur at the level of DNA binding. Instead, cooperativity could also be established downstream of DNA binding, for example by the cooperative recruitment of additional regulatory factors.

4.5 Regulation of cell type-specific gene expression by shared TFBSs

The consequences of GR activity highly depend on the cell type examined and can range from the induction of anti-inflammatory responses as for example in hematopoietic cells

to the induction of gluconeogenesis in hepatic cells [72, 133]. Similarly, for the two cell lines examined in this thesis, the A549 and U2OS cell line, the number and identity of GR-regulated genes highly varies depending on cell type [82].

Reflecting the observed differences in gene expression, the pattern of genomic GR binding also varies drastically among different cell types [103, 104]. Thus, one plausible explanation for the cell type-specific effects of GR activity could be based on distinct GR binding patterns in different cell types. Indeed, for the estrogen receptor cell type-specific gene expression changes were shown to be associated with the cell type-specific occupancy of nearby estrogen receptor binding sites [13].

Conceivably, besides the occupancy of different TFBSs, a TFBS that is bound in different cell types may also contribute to the establishment of cell type-specific gene expression patterns by providing distinct regulatory activities. In this study I could show that a GBS that is equally bound in two different cell types, exhibits distinct regulatory properties depending on its genomic context. The data presented here suggest that the cell type-specific activity of *GILZ* GBS1 is achieved by rewiring of its interaction landscape, resulting in the differential expression of distinct transcript variants. Such a concept of differential enhancer wiring was also found for other loci. For example, the locus control region, a distal regulatory region for the β -globin locus, is able to dynamically interact with the promoters of different globin genes depending on the differentiation status of the cell. This enables the cell to sequentially express the correct globin gene during erythroid differentiation [10]. However, in non-hematopoietic cells the locus control region is inactive [46]. In contrast, in this work a GBS that is bound in two different cell types interacts with different TSSs of the same gene, thereby contributing to the establishment of cell type-specific gene expression programs.

Functionally, the expression of different *GILZ* transcript variants may serve different cellular functions depending on the needs of the cell. For example, the long transcript variant 1, referred to as L-*GILZ*, but not its short transcript variants were shown to interact with p53 involved in the p53-mediated stress response [134]. Hence, in different cell types distinct *GILZ* transcript variants possibly fulfill different functional roles. Notably, in A549 cells the quantification of *GILZ* transcript variant 1 by qPCR revealed a loss of GR-dependent upregulation upon GBS1 deletion, whereas a primer binding to an exon, shared by all transcript variants according to the NCBI RNA reference sequences database [105] showed no effect (Fig. 3.12 and 3.14 c). Thus, most probably additional truncated *GILZ* transcript variants exist in the A549 cell line. Indeed, other gene annotation databases, such as ENSEMBL [135], indicated the presence of such truncated *GILZ* isoforms.

Together, distinctive cellular functions are established and maintained by transcriptionally regulating the abundance of gene products with different functions in a cell type-

specific manner. Besides differences in TF binding, plasticity in gene expression is achieved by TFBSs that are bound in different cell types yet regulate the expression of cell type-specific transcript variants.

4.6 The influence of genome architecture, DNA-looping and CTCF binding on enhancer activity

Although the DNA is tightly packed in the nucleus, its 3D architecture is highly organized. In fact, this structural organization enables the cell to regulate gene expression by the establishment of long-range enhancer-promoter interactions. Supporting this tight relationship between genome organization and the regulation of gene expression, the integration of information regarding the 3D landscape from HiC data improves the correlation of TF binding to its target genes [26].

At the *GILZ* locus, a combination of 4C experiments and artificial activation using the dCas9-VP64 SAM system indicated substantial differences in the genome topology of A549 and U2OS cells (Fig. 3.17 a and Fig.3.18 a), suggesting that these are able to explain observed differences in the GR-mediated regulation by GBS1.

Notably, regarding experiments using the dCas9-VP64 SAM system the observed cell type-specific differences in gene activation could theoretically also result from differences in the activity of gRNAs or the expression of the components of the dCas9-VP64 SAM system itself. In fact, in A549 cells both the deletion of GBS1 and 4C experiments indicate the establishment of a functional enhancer-promoter interaction between *GILZ* GBS1 and the promoter region of transcript variant 1. However, this long-range interaction was not observed using the artificial activator dCas9-VP64 SAM (Fig. 3.17 a). Of note, experiments using the dCas9-VP64 SAM system were performed in the absence of hormone. Thus, for distal regulatory regions both GR activity and the simultaneous binding of multiple GBSs might be required to robustly induce the transcription of transcript variant 1. Therefore, artificial activation using a single gRNA might not be sufficient to reach a necessary threshold level for gene induction from a distal locus. However, regarding the upregulation of the nearby *GILZ* transcript variants, differences in gRNA binding at GBS1 between both cell types are unlikely to explain the observed cell type-specific differences for two main reasons. First of all, *GILZ* GBS1 is similarly bound by GR in both cell types, suggesting that the genomic region is also equally accessible for gRNA binding. Additionally supporting this hypothesis, Cas9-editing efficiencies using the same gRNA targeting *GILZ* GBS1 were generally high and comparable between both cell types (Fig. S5.1, S5.2, S5.6 and S5.7). Second, a gRNA placed directly at the promoter region of transcript variant 1 was able to upregulate its expression in both cell types (Fig. 3.17 b). Therefore, differences in the activity of gRNAs or dCas9-VP64 SAM are unlikely to explain the observed cell type-specific differences in activation of nearby *GILZ* transcript

variants. Instead, and in concordance with 4C experiments, the regulatory activity of GBS1 is most probably influenced by cell type-specific differences in genome architecture.

Thus, the endogenous genomic environment and particularly the 3D environment in which a TFBS is embedded can crucially shape its regulatory activity, thereby contributing to the establishment of cell type-specific gene expression programs. Notably, overall TAD structures were reported to remain remarkably stable in different cell types and in response to different stimuli [26, 136]. However, individual interactions within TADs are able to change during the course of differentiation and in response to signaling pathways [26, 137]. This indicates that, similar to what was found here for *GILZ* GBS1, differential enhancer-promoter interactions within TADs can contribute to the establishment of distinct gene regulation programs.

Furthermore, at the *GILZ* locus, differences in CTCF binding correlated with differences in the regulatory activity of GBS1, supporting the pivotal role of CTCF in the formation of cell type-specific long-range interactions. Based on these findings, I propose a model in which differential CTCF binding is a crucial mediator in defining the cell type-specific regulatory activity of *GILZ* GBS1. This model proposes that in A549 cells CTCF binding at the promoter region of *GILZ* transcript variant 1 and GBS1 enables the formation of a DNA-looping interaction (Fig. 4.2 top). Thereby, the long-range DNA-looping interaction established between GBS1 and the distal promoter region of transcript variant 1 facilitates its GR-dependent induction. Furthermore, this looping interaction places the GBS1 into the same regulatory unit as the TSS1 resulting in the isolation of transcript variant 2 and 3 from its regulatory activity (Fig. 4.2 top). In contrast, in U2OS cells CTCF binding at TSS1 is absent. Therefore, the GBS1 is contained within the same regulatory unit as the promoter region of transcript variant 2 and 3, regulating their GR-dependent transcription (Fig. 4.2 bottom). Thus, CTCF binding at *GILZ* TSS1 and the resulting long-range DNA-looping interaction in A549 cells follows two functions. On the one hand it enables the GBS1-regulated transcription of *GILZ* transcript variant 1. And on the other hand, it isolates the nearby transcript variants 2 and 3 from regulation by GBS1. In contrast, in U2OS cells the absence of CTCF binding and DNA-looping interaction between GBS1 and TSS1 facilitates the expression of the nearby transcript variants 2 and 3, resulting in the establishment of distinct cell type-specific gene expression patterns.

Importantly, as already suggested by others [138–140], this model is consistent with CTCF acting as both a promoter of long-range interactions and an insulator of regulatory elements. Thereby, the establishment of closed looping interactions can shield enhancer regions from inappropriate interactions with promoters and simultaneously redirect its activity. However, genomic deletion of CTCF binding sites will be required to ultimately verify the proposed model and to answer the question whether differences in CTCF binding are the cause for the cell type-specific expression of *GILZ*. In fact, differences in CTCF binding could represent a general means to influence the regulatory activity of individual

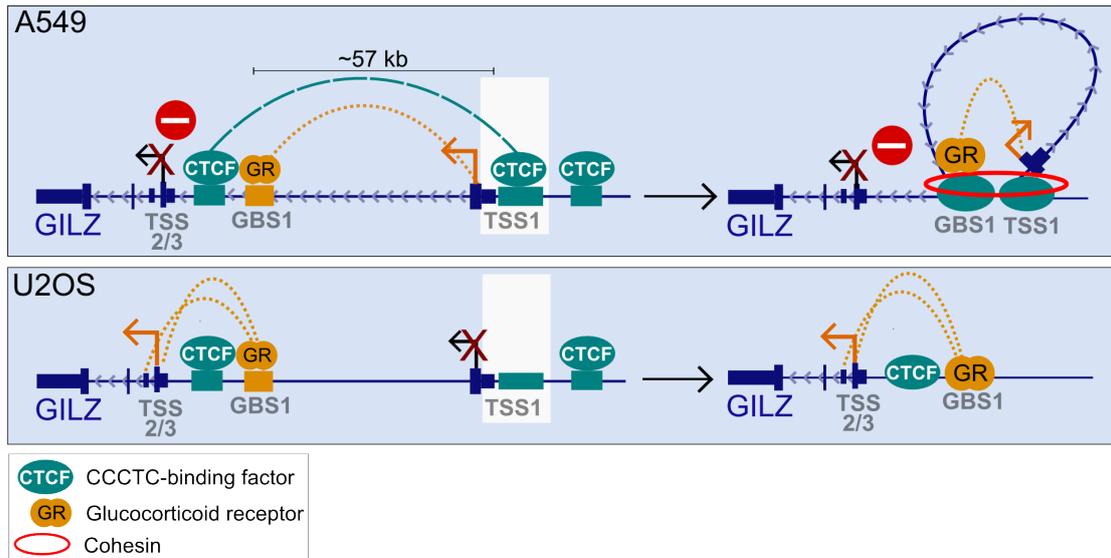


Fig. 4.2: Model for the cell type-specific regulatory activity of *GILZ* GBS1 mediated by differential CTCF binding and DNA-looping. In A549 cells CTCF binds to a CTCF binding site located at the promoter region of *GILZ* transcript variant 1 (TSS1), enabling the interaction with CTCF bound near GBS1. Therefore, CTCF and cohesin are able to establish a long-range DNA-looping interaction between GBS1 and TSS1, facilitating the GBS1-dependent upregulation of *GILZ* transcript variant 1. Furthermore, the establishment of the regulatory loop between GBS1 and TSS1 results in the isolation of *GILZ* transcript variant 2 and 3 (TSS2/3) from the regulatory activity of GBS1. In contrast in U2OS cells, CTCF does not bind to the CTCF motif located at TSS1. Thus, CTCF bound near GBS1 is not able to establish a DNA-looping interaction with the promoter of transcript variant 1. Therefore, GBS1 does not interact with the promoter of transcript variant 1 but instead interacts with the promoters of nearby transcript variants 2 and 3 to regulate their expression in a GR-dependent manner.

TFBSs. Although, a major proportion of genomic CTCF binding sites were reported to be conserved among different cell types, 30-60% of binding sites show a cell type-specific binding pattern [141]. Thus, a subset of CTCF binding sites could indeed be involved in the regulation of cell type-specific gene expression. In support of this, genome-wide analysis showed that enhancer regions are specifically enriched for CTCF binding [142, 143]. Furthermore, recent insights indicate that the motif orientation of paired CTCF binding sites located at the bases of putative loops are indicative for the strength of DNA looping interactions. Specifically, motifs in a convergent forward-reverse orientation appeared to be most stable [25, 47, 53, 106]. Notably, a recent analysis showed that convergent orientated CTCF motifs are mainly associated with TAD borders, whereas tandem sites form weaker interactions that are associated with loops within TAD structures [144]. Of note, at the *GILZ* locus several CTCF binding motifs were identified at both ends of the putative DNA-looping interactions between *GILZ* GBS1 and the promoter region of transcript variant 1 (Fig. S5.8). Therefore, at the *GILZ* locus CTCF could preferentially interact in a forward-reverse orientation. However, whether the establishment of long-range interactions strictly requires such a configuration and how multiple CTCF binding sites are involved remains to be determined.

A critical question remains whether long-range interactions consequently require CTCF to establish a functional enhancer-promoter contacts. Given that the GR-dependent up-regulation of *GILZ* transcript variants was not completely abolished upon GBS1-4 deletion (Fig. 3.6), additional distal GBSs probably maintain their GR-dependent expression. However, no GBSs contained in a window of 200 kb surrounding GBS1 overlaps with a nearby CTCF-binding peak (Fig. 3.18 a). Other studies already reported the existence of long-range interactions that were established independent of CTCF [25, 106, 144], suggesting that other factors might be involved in their establishment. Supporting this hypothesis, a knock-down of CTCF did not result in changes of the expression of *GILZ* transcript variants (Fig. 3.20). Given that residual CTCF expression was still detectable upon siRNA knock-down, genomic deletion of CTCF motifs will be ultimately required to investigate the role of CTCF binding in the cell type-specific regulation of *GILZ* transcript variants. In contrast, for specific genes CTCF depletion performed in other studies reduced intra-domain contacts and led to changes in the expression of some genes. However, a depletion of CTCF did not result in global changes of gene expression or a general break-down of TADs [47]. However, the targeted degradation of CTCF using an auxin-inducible degron system resulted in more substantial defects in TAD establishment and further resulted in changes of gene expression for a subset of genes [145]. Because CTCF-mediated long-range interactions are stabilized by and co-localize with cohesin, another open question remains whether differences in cohesin binding also contribute to the cell type-specific expression of *GILZ* transcript variants. Furthermore, given the sheer number of genomic CTCF binding sites, likely not all CTCF sites are involved in the formation of

4. Discussion

long-range interactions and closed loop domains. Therefore, additional mechanisms must exist that shape the role of CTCF binding sites in the genome.

At the *GILZ* locus cell type-specific DNA mutations within CTCF motifs were not responsible for the observed differences in CTCF binding. Hence, the question arises what is the restrictive factor that directs cell type-specific CTCF binding in A549 and U2OS cells. Given that the consensus CTCF motif contains CpGs, DNA-methylation was suggested to be involved in the regulation of genomic CTCF binding [146]. Supporting this hypothesis, targeted methylation of CTCF sites by dCas9-Dnmt3A resulted in a loss of CTCF binding and changes in gene expression at interacting loci [147]. However, because DNA methylation is not a deterministic predictor of CTCF binding [146], additional mechanisms must exist that regulate its binding.

Of note, CTCF is also implicated in the regulation of alternative RNA-splicing by RNA polymerase II pausing [148], suggesting an alternative but not necessarily mutually exclusive explanation for the correlation between cell type-specific CTCF binding and the expression of *GILZ* transcript variants. Theoretically, CTCF-induced alternative splicing of exons at the *GILZ* locus might result in the cell type-specific generation of mature mRNAs. Importantly, the NCBI RNA reference sequences database [105] only provides an indication for the expression of *GILZ* transcript variants in the cell types studied here. Hence, CTCF-binding could also correlate with differences in RNA splicing rather than in the induction of transcription from alternative TSSs.

In summary, the regulatory activity of a TFBS can be rewired in a cell type-specific manner by differences in its genomic architecture and the establishment of specific long-range interactions. For the *GILZ* GBS1 studied here, cell type-specific differences in its regulatory activity coincided with differences in CTCF binding. Importantly, CTCF was reported to possess a dual function as both an insulator and a mediator of long-range interactions. Thereby, differences in CTCF binding could be responsible for the establishment of cell type-specific long-range interactions and simultaneously shield enhancers from interacting with specific promoters.

4.7 Investigating the effect of GBS motif variants at endogenous loci

Regulatory information enabling the correct quantitative expression of genes is contained at several levels of input. One of these inputs represents the DNA sequence itself, where TFs bind to its cognate TFBSs based on the recognition of specific DNA sequence motifs. Conceivably, quantitative differences in gene expression could be based on differences in GR affinity for its cognate GR binding motif. However, *in vitro* binding affinities of different GBS motif variants and their ability for transcriptional activation of gene

expression in transient and genomically integrated reporters do not correlate [83, 86, 127]. In fact, structural studies suggested that the DNA sequence of a GBS can alter the shape of the DNA, thereby allosterically regulating GR activity [83, 86]. However, whether fine-tuning of gene expression by GBS motif variants is also relevant at endogenous loci is still unclear.

Indeed, embedding different GBS motif variants in the context of the *GILZ* promoter differentially affected reporter gene expression in luciferase reporter assays. Of note, here GBS motif variants were tested in the context of the *GILZ* promoter, whereas in previous studies motif variants were directly placed in front of a minimal promoter [83, 86]. In concordance to previous studies [83], the exchange of GBS motif variants within reporter constructs resulted in distinct transcriptional activities. However, the relative activity of individual motifs differed in comparison to previous studies. For example, when placed directly upstream of a minimal promoter, reporter gene expression induced by the CGT sequence was considerably higher than induction by the FKBP5 sequence [83]. However, this pronounced difference was not observed upon integration in the context of *GILZ* reporter constructs (Fig. 3.21 b). Hence, the sequence context in which a GBS motif variant is contained highly influences its regulatory effect on reporter gene expression. Supporting these results, the wildtype GBS sequence, which is endogenously contained at the *GILZ* locus showed the highest transcriptional activity in reporter assays (Fig. 3.21 b). This suggests that the wildtype *GILZ* GBS1 sequence is best suited for transcriptional induction within the *GILZ* promoter context. Generally, similar to previous reports [83], the affinity of the exchanged GBS motif variants as determined by EMSA did not correlate to their transcriptional activity. For example, in reporter assays the PAL sequence, having a 10-fold higher affinity for GR than the *GILZ* GBS1 sequence showed a lower activity than the low affinity *GILZ* GBS1. Therefore, differences in GR binding affinity are most probably not responsible for the observed differences in transcriptional activity.

In contrast to reporter assays, the genomic exchange of two selected GBS motif variants, PAL and FKBP5-2, which showed the greatest differences in reporter assays, displayed no effect on endogenous *GILZ* expression (Fig. 3.22). Thus, this either indicates that GBS motif variants do not play a role in endogenous gene regulation or that their effect in isolation might be masked by the contribution of other regulatory regions. Arguing for the latter possibility, the combinatorial deletion of *GILZ* GBSs showed that *GILZ* is regulated by several GBSs, namely GBS1-4 and at least one additional distal GBS. Of note, GBS2-4 were also contained in reporter constructs used for luciferase reporter assays. However, mutational analysis of reporter constructs already indicated that its individual contribution on transcriptional activation does not reflect its endogenous interplay (Fig. 3.9). Thus, because in the genomic context several GBSs regulate the expression of *GILZ*, the quantitative effect of exchanging GBS1 into different GBS motif variants might be masked by the regulatory influence from other GBSs. In fact, previous experiments show-

4. Discussion

ing an allosteric effect of GBS motif variants on GR activity, were mainly performed in settings using one single GBS [83, 86]. Thus, fine-tuning of gene expression by GBS motif variants may be relevant for gene expression depending on the context in which the GBS is active.

Therefore, I choose the endogenously silenced *IL1R2* gene as a model locus to investigate GBS motif variants in an isolated setting by placing a single GBS directly at its promoter region. Furthermore, the genomic deletion of promoter-proximal *GILZ* GBSs suggested that multiple GBSs are required to regulate *GILZ* expression. Hence, this set-up also allowed me to investigate whether a cluster of multiple GBSs is a general requirement for GR-dependent gene expression at endogenous loci. Surprisingly, the insertion of a single GBS at the promoter region of *IL1R2* induced gene expression in a GR-dependent manner (Fig. 3.25), suggesting that its insertion is sufficient to regulate gene expression by GR. Hence, for future experiments the *IL1R2* locus represents an ideal model system to study the effect of GBS motif variants in an endogenous but isolated setting in which gene expression depends on the presence of a single GBS.

Although its mechanism of action is not clear, *IL1R2* is implicated in the regulation of inflammation, reflecting its role as a GR target gene [149]. Specifically, although being completely silenced in U2OS cells, *IL1R2* expression is upregulated in a GR-dependent manner in other cell types [150]. Conceivably, for *IL1R2* the introduction of a single promoter-proximal GBS was sufficient to induce gene expression because the *IL1R2* gene is generally responsive for regulation by GR. Reflecting its regulation in other cell types, a GR ChIP-peak is located in a distance of only 2 kb from the TSS of *IL1R2* (Fig. 3.24 b). Despite its location, this GR-bound peak is not sufficient to induce the GR-dependent expression of *IL1R2* in U2OS cells. However, it remains to be determined whether other proximal and distal GBSs, including GBSs of the aforementioned GR ChIP-peak, are able to combinatorially regulate *IL1R2* expression upon induction by the artificially introduced GBS. And furthermore, whether silenced genes can be generally activated by introduction of a single promoter-proximal GBS. Indeed, some enhancers were described to create compatible enhancer-promoter pairs with only certain promoters [151–153], thereby providing an additional means to restrict enhancer activity to the appropriate target gene. In future experiments, a single GBS could be introduced at the promoter region of multiple GR-regulated and unregulated genes to test whether a single promoter-proximal GBS is generally sufficient to induce GR-dependent gene regulation.

Collectively, at the *GILZ* locus the role of GBS motif variants in endogenous gene expression is complicated due to the integration of regulatory information from multiple regulatory active GBSs. In contrast, introducing a single GBSs at the *IL1R2* locus was sufficient to induce gene expression. Therefore, in future experiments this setting enables the investigation of GBS motif variants and their effect on endogenous gene expression.

Furthermore, the introduction of GBSs at the promoter-region of additional genes could help to uncover the general requirements and mechanisms required for GR-dependent regulation.

4.8 Conclusions

The sheer number of genomic TFBSs complicates the discrimination of apparently non-productive TFBSs and productive TFBSs that regulate the expression of a specific gene. Furthermore, the different possible modes of interactions among multiple combinatorially interacting TFBSs and the influence of a dynamically changing chromatin context and environment renders the identification of productive TFBSs and their target genes even more complex. By functionally analyzing TFBSs in their genomic context, this project sheds light on the mechanisms of combinatorial regulation by multiple TFBSs. Furthermore, this work highlights the importance of the 3D architecture for the establishment of cell type-specific gene expression programs.

Up to this date, only few TFBSs have been comprehensively investigated in their genomic context. Importantly, neither linear proximity nor reporter assays can deterministically predict the genomic role of a putative TFBSs. Therefore, further systematic analysis in the endogenous genomic context is required to refine our understanding of the dynamic regulatory activity of genomic TFBSs and the mechanisms that shape their individual role in transcriptional regulation.

5 Supplement

5.1 Supplementary figures

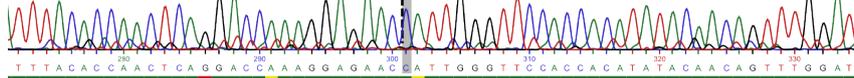
GILZ GBS1 deletion

gRNA GCAGGACCAAAGGAGAACATT

editing efficiency: 70%

U2OS clone G1

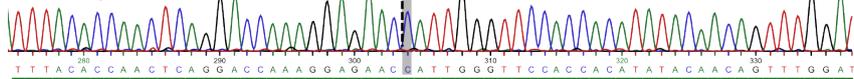
allele 1: 1bp insertion



wt ACACCAACTCAGGACCAAAGGAGAACATTGGGTTCCACCACATATACAACAGTTTGGAT
gRNA

U2OS clone G2

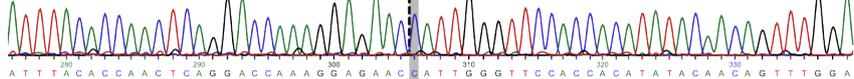
allele 1: 1bp insertion



wt ACACCAACTCAGGACCAAAGGAGAACATTGGGTTCCACCACATATACAACAGTTTGGAT
gRNA

U2OS clone G3

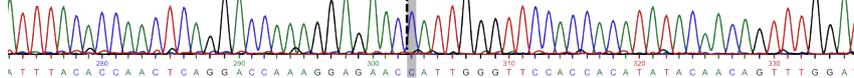
allele 1: 1bp insertion



wt ACACCAACTCAGGACCAAAGGAGAACATTGGGTTCCACCACATATACAACAGTTTGGAT
gRNA

U2OS clone G4

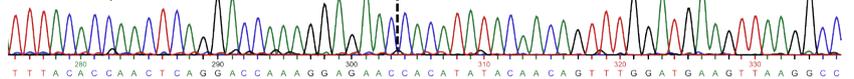
allele 1: 1bp insertion



wt ACACCAACTCAGGACCAAAGGAGAACATTGGGTTCCACCACATATACAACAGTTTGGAT
gRNA

U2OS clone G5

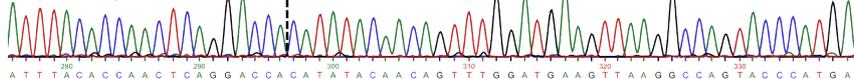
allele 1: 12bp deletion



wt ACACCAACTCAGGACCAAAGGAGAACAATTGGGTTCCACCACATATACAACAGTTTGGAT
gRNA

U2OS clone G6

allele 1: 23 bp deletion



wt ACACCAACTCAGGACCAAAGGAGAACAATTGGGTTCCACCACATATACAACAGTTTGGAT
gRNA

Fig. 5.1: Genotyping results of U2OS *GILZ* GBS1 deletion clones. Genotyping results of successfully CRISPR/Cas9-edited clonal U2OS cells. The respective gRNAs used for CRISPR/Cas9-editing, the corresponding expected location of the Cas9-induced DNA double-strand break and editing efficiency are indicated.

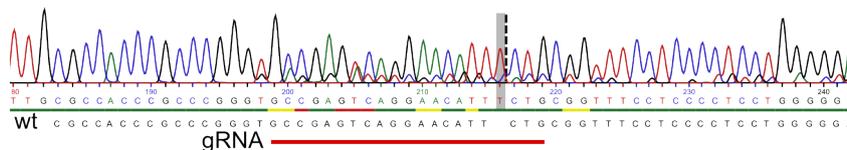
DUSP1 GBS1 deletion

gRNA GCCGAGTCAGGAACATTCTG

editing efficiency 85%

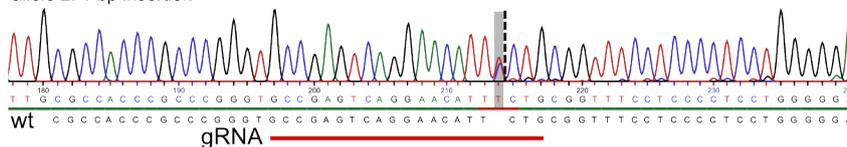
U2OS clone D1

allele 1: 1 bp insertion

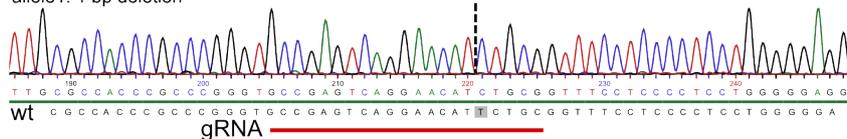
**U2OS clone D2**

allele 1: 1 bp insertion

allele 2: 1 bp insertion

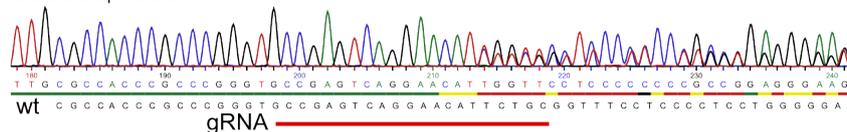
**U2OS clone D3**

allele 1: 1 bp deletion

**U2OS clone D4**

allele 1: 1 bp deletion

allele 2: 5 bp deletion

**U2OS clone D5**

allele 1: 1 bp deletion

allele 2: 216 bp deletion

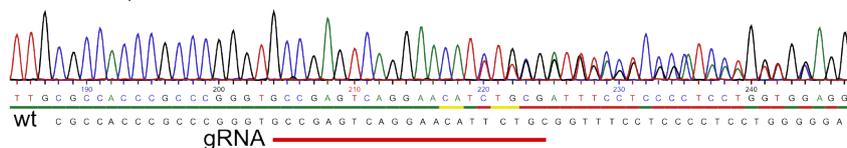


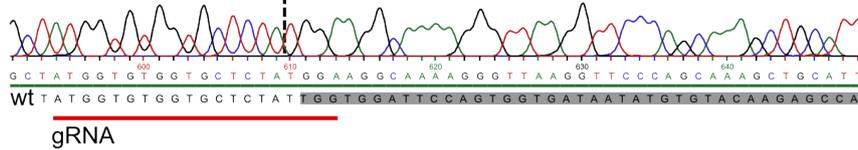
Fig. 5.2: Genotyping results of U2OS *DUSP1* GBS1 deletion clones. Genotyping results of successfully CRISPR/Cas9-edited clonal U2OS cells. The respective gRNAs used for CRISPR/Cas9-editing, the corresponding expected location of the Cas9-induced DNA double-strand break and editing efficiency are indicated.

GILZ GBS2-4 deletion

gRNA1 GATGGTGTGGTGCTCTATTGG
 or
 gRNA2 GTAAACCTGCTGCACTAGCCC
 GCCAGGTGGTATGGGAAGGGA

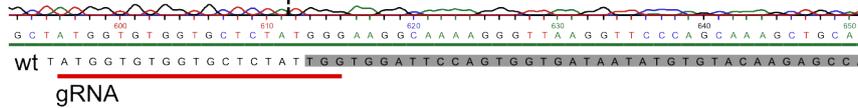
U2OS clone 1

allele 1: 601 bp deletion



U2OS clone 2

allele 1: 601 bp deletion



U2OS clone 3

allele 1: 512 bp deletion

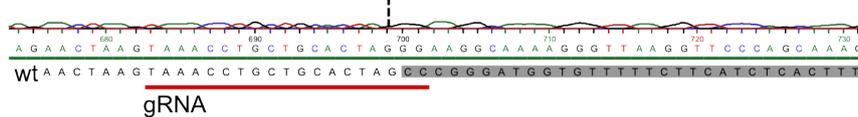


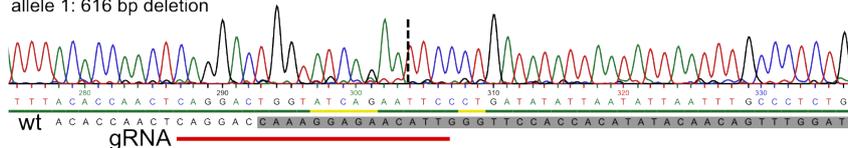
Fig. 5.3: Genotyping results of U2OS *GILZ* GBS2-4 deletion clones. Genotyping results of successfully CRISPR/Cas9-edited clonal U2OS cells. The respective gRNAs used for CRISPR/Cas9-editing and the corresponding expected location of the Cas9-induced DNA double-strand break are indicated.

GILZ GBS1-4 deletion

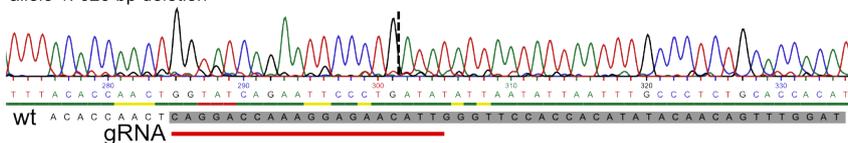
gRNA1 GCAGGACCAAAGGAGAACATT
gRNA2 GTCTTAAATCAGGAGCTTAAC

U2OS CUT1-4 clone 1

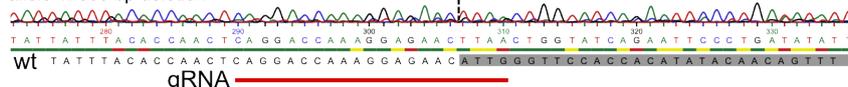
allele 1: 616 bp deletion

**U2OS CUT1-4 clone 2**

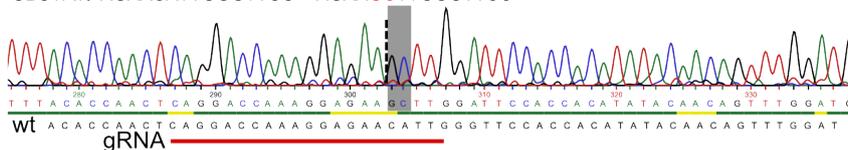
allele 1: 623 bp deletion

**U2OS CUT1-4 clone 3**

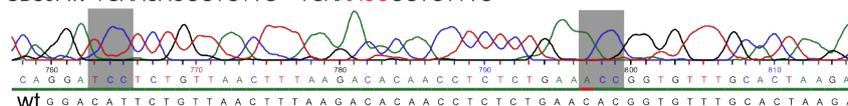
allele 1: 600 bp deletion

**U2OS HR1-4**

GBS1: wt AGAACATTGGGTTCC > AGAAGCTTGGGTTCC



GBS2: wt AGGACATTCTGTAA > AGGATCCTCTGTAA
GBS3: wt TGAACACGGTGTG > TGAACCGGTGTG



GBS4: wt AGAACAAAGTGCAGG > AGATCTAAGTGCAGG

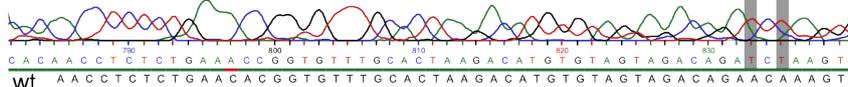


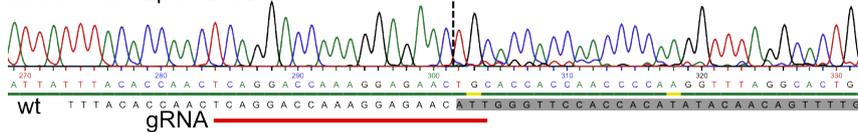
Fig. 5.4: Genotyping results of U2OS *GILZ* GBS1-4 deletion clones. Genotyping results of successfully CRISPR/Cas9-edited clonal U2OS cells. The respective gRNAs used for CRISPR/Cas9-editing and the corresponding expected location of the Cas9-induced DNA double-strand break are indicated.

GILZ GBS1-4 inversion

gRNA1 GCAGGACCAAAGGAGAACATT
 gRNA2 GTCTTAAATCAGGAGCTTAAC

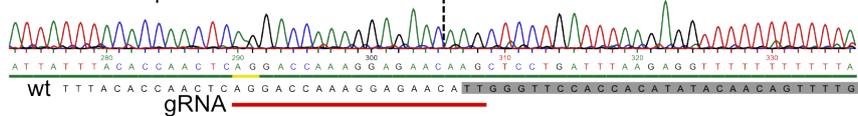
U2OS INVERSION GBS1-4 clone 1

allele 1: 601 bp inversion



U2OS INVERSION GBS1-4 clone 2

allele 1: 600 bp inversion



U2OS INVERSION GBS1-4 clone 3

allele 1: 600 bp inversion

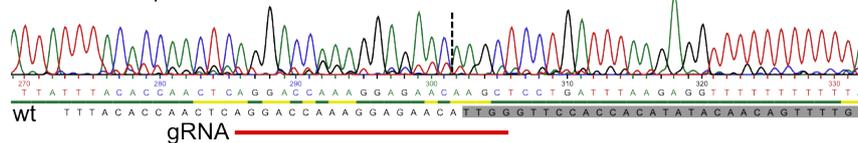


Fig. 5.5: Genotyping results of U2OS *GILZ* GBS2-4 inversion clones. Genotyping results of successfully CRISPR/Cas9-edited clonal U2OS cells. The respective gRNAs used for CRISPR/Cas9-editing and the corresponding expected location of the Cas9-induced DNA double-strand break are indicated.

GILZ GBS1 deletion

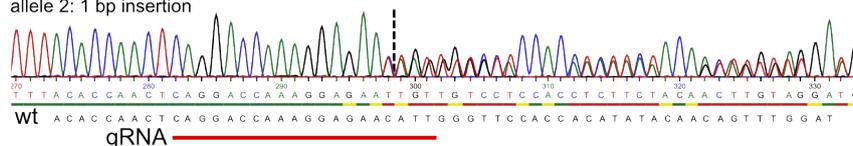
gRNA GCAGGACCAAAGGAGAACATT

editing efficiency: 67%

A549 clone G1

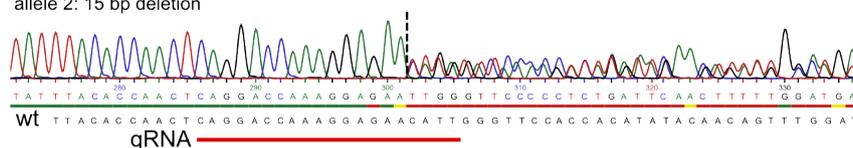
allele 1: 2 bp deletion

allele 2: 1 bp insertion

**A549 clone G2**

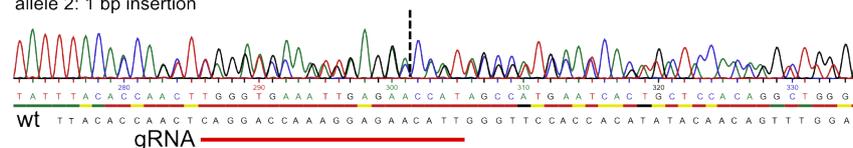
allele 1: 2 bp deletion

allele 2: 15 bp deletion

**A549 clone G3**

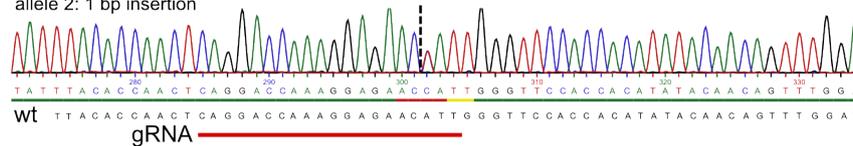
allele 1: 45 bp deletion

allele 2: 1 bp insertion

**A549 clone G4**

allele 1: 1 bp insertion

allele 2: 1 bp insertion

**A549 clone G5**

allele 1: 2 bp deletion

allele 2: 2 bp deletion

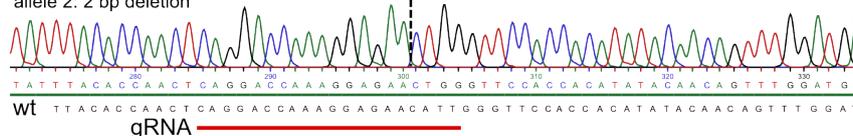


Fig. 5.6: Genotyping results of A549 *GILZ* GBS1 deletion clones. Genotyping results of successfully CRISPR/Cas9-edited clonal A549 cells. The respective gRNAs used for CRISPR/Cas9-editing, the corresponding expected location of the Cas9-induced DNA double-strand break and editing efficiency are indicated.

DUSP1 GBS1 deletion

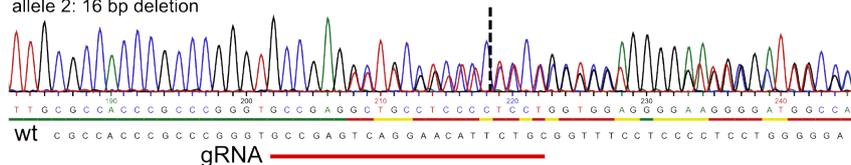
gRNA GCCGAGTCAGGAACATTCTG

editing efficiency: 44%

A549 clone D1

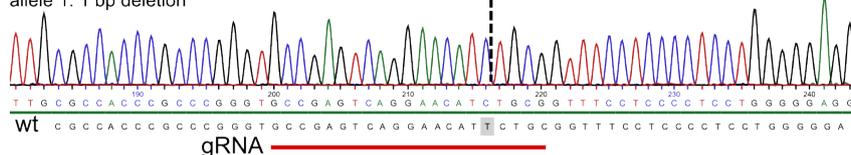
allele 1: 10 bp deletion

allele 2: 16 bp deletion



A549 clone D2

allele 1: 1 bp deletion

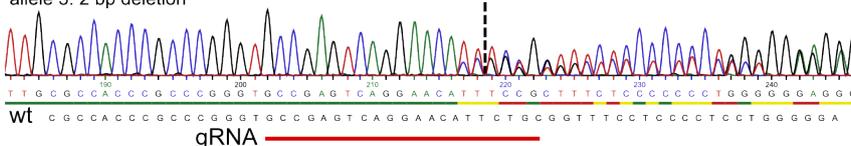


A549 clone D3

allele 1: 1 bp insertion

allele 2: 1 bp deletion

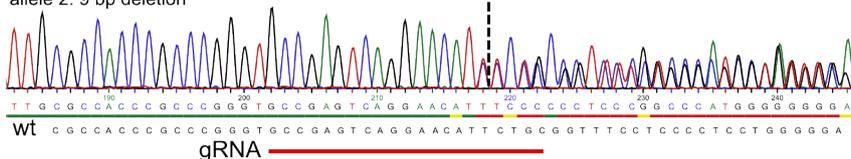
allele 3: 2 bp deletion



A549 clone D4

allele 1: 1 bp insertion

allele 2: 9 bp deletion



A549 clone D5

allele 1: 1 bp deletion

allele 2: 1 bp insertion

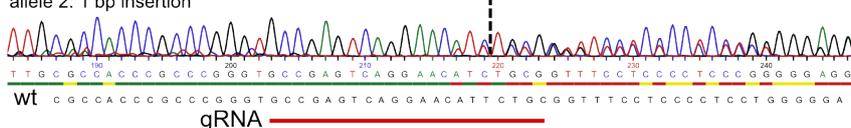


Fig. 5.7: Genotyping results of A549 *DUSP1* GBS1 deletion clones. Genotyping results of successfully CRISPR/Cas9-edited clonal A549 cells. The respective gRNAs used for CRISPR/Cas9-editing, the corresponding expected location of the Cas9-induced DNA double-strand break and editing efficiency are indicated.

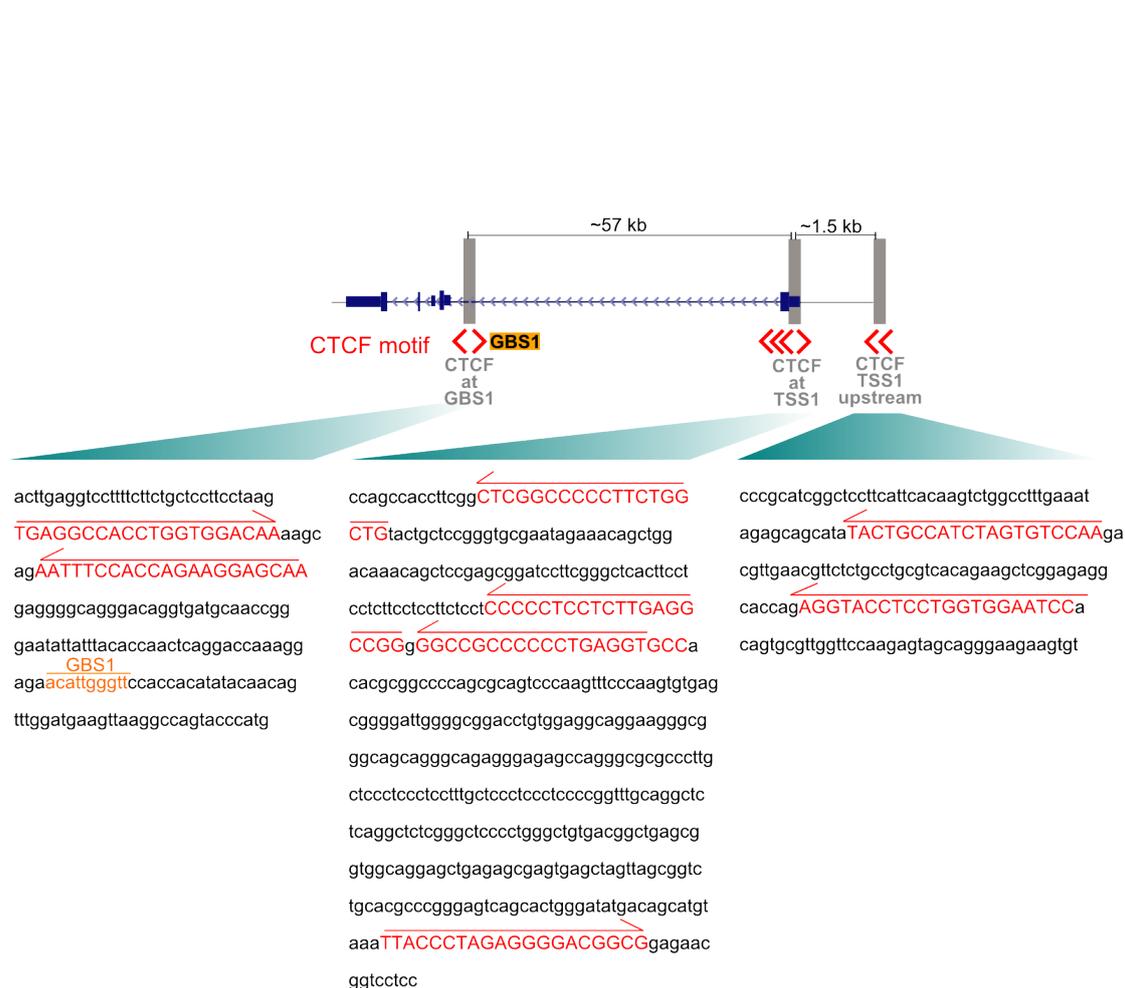


Fig. 5.8: CTCF motif orientation at CTCF-bound regions of the *GILZ* locus. CTCF motif matches of the CTCF-bound regions at *GILZ* GBS1 (left), TSS1 (middle) and the region 2 kb upstream of TSS1 (right). Individual motifs (Matrix ID: M01200) are highlighted in red and arrows indicate the orientation of individual motifs. The location of *GILZ* GBS1 is highlighted in orange.

Sequence alignment CTCF motifs at *GILZ* TSS1

```

U2OS 176 CCTAAAGCCAGCCACCTTCGGCTCGGCCCTTCTGGCTGTACTGCTCCGGGTGCGAATA ;
A549 175 CCTAAAGCCAGCCACCTTCGGCTCGGCCCTTCTGGCTGTACTGCTCCGGGTGCGAATA ;

U2OS 236 GAAACAGCTGGACAAACAGCTCCGAGCGGATCCTTCGGGCTCACTTCCTCCTTCTCTCC 2
A549 235 GAAACAGCTGGACAAACAGCTCCGAGCGGATCCTTCGGGCTCACTTCCTCCTTCTCTCC 2

U2OS 296 TTCTCCTCCCCCTCCTTTGAGGCCGGGGGCCGCCCTGAGGTGCCACACGCGGCCCC
A549 295 TTCTCCTCCCCCTCCTTTGAGGCCGGGGGCCGCCCTGAGGTGCCACACGCGGCCCC

U2OS 356 AGCGCAGTCCCAAGTTTCCCAAGTGTGAGCGGGGATTGGGGCGGACCTGTGGAGGCAGGA
A549 355 AGCGCAGTCCCAAGTTTCCCAAGTGTGAGCGGGGATTGGGGCGGACCTGTGGAGGCAGGA

U2OS 416 AGGGCGGGCAGCAGGGCAGAGGAGAGCCAGGGCGCGCCCTTGCTCCCTCCCTCTTTGC
A549 415 AGGGCGGGCAGCAGGGCAGAGGAGAGCCAGGGCGCGCCCTTGCTCCCTCCCTCTTTGC

U2OS 476 TCCCTCCCTCCCCGGTTTGACAGGCTCTCAGGCTCTCGGGCTCCCCTGGGCTGTGACGGCT
A549 475 TCCCTCCCTCCCCGGTTTGACAGGCTCTCAGGCTCTCGGGCTCCCCTGGGCTGTGACGGCT

U2OS 536 GAGCGGTGGCAGGAGCTGAGAGCGAGTGAGCTAGTTAGCGGTCTGCACGCCCGGGAGTCA
A549 535 GAGCGGTGGCAGGAGCTGAGAGCGAGTGAGCTAGTTAGCGGTCTGCACGCCCGGGAGTCA

U2OS 596 GCACTGGGATATGACAGCATGTAATACCCCTAGAGGGGACGGCGGAGAACGGTC 650
A549 595 GCACTGGGATATGACAGCATGTAATACCCCTAGAGGGGACGGCGGAGAACGGTC 650

```

Fig. 5.9: Genomic sequence at *GILZ* TSS1 in U2OS and A549 cells. DNA sequence alignment of the genomic region covering the identified CTCF motifs at *GILZ* TSS1 comparing the sequence of U2OS and A549 cells.

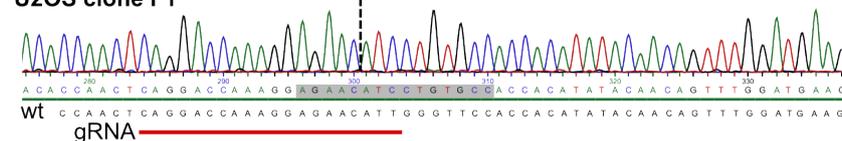
FKBP5-2 exchange at *GILZ* GBS1

gRNA GCAGGACCAAAGGAGAACATT

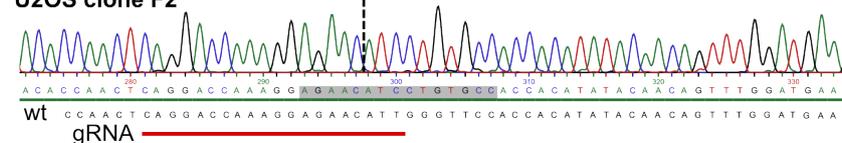
FKBP5-2 sequence: AGAACAtccTGTGCC

editing efficiency: 10%

U2OS clone F1



U2OS clone F2



U2OS clone F3

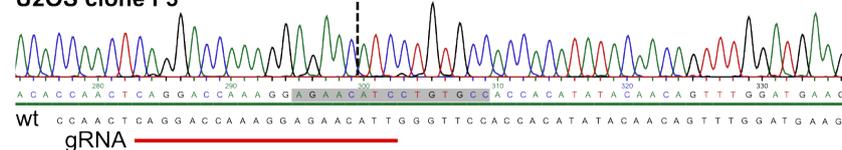


Fig. 5.10: Genotyping results of U2OS cell lines with *GILZ* GBS1 exchanged into the FKBP5-2 sequence. Genotyping results of successfully CRISPR/Cas9-edited clonal U2OS cells containing the FKBP5-2 GBS sequence instead of *GILZ* GBS1. The respective gRNAs used for CRISPR/Cas9-editing, the corresponding expected location of the Cas9-induced DNA double-strand break and editing efficiencies are indicated.

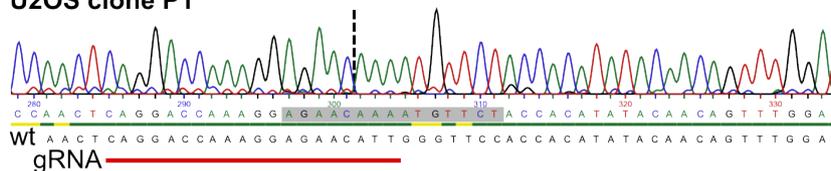
PAL exchange at *GILZ* GBS1

gRNA GCAGGACCAAAGGAGAACATT

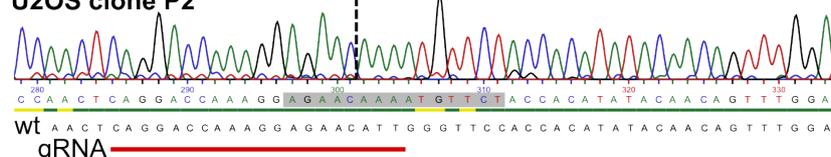
PAL sequence: AGAACAAAATGTTCT

editing efficiency: 16%

U2OS clone P1



U2OS clone P2



U2OS clone P3

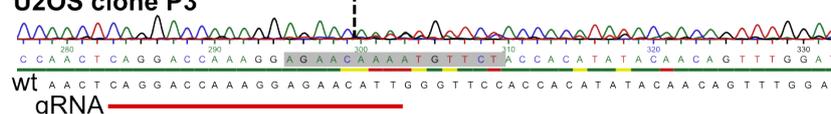


Fig. 5.11: Genotyping results of U2OS cell lines with *GILZ* GBS1 exchanged into the PAL sequence. Genotyping results of successfully CRISPR/Cas9-edited clonal U2OS cells containing the PAL GBS sequence instead of *GILZ* GBS1. The respective gRNAs used for CRISPR/Cas9-editing, the corresponding expected location of the Cas9-induced DNA double-strand break and editing efficiencies are indicated.

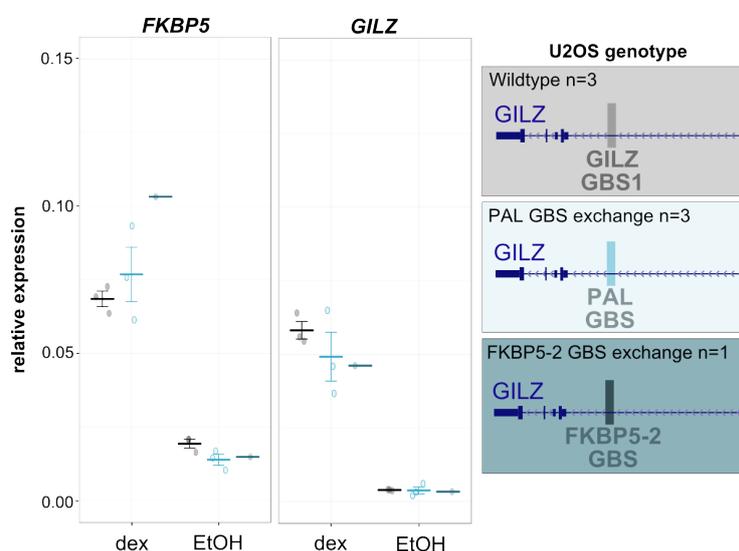
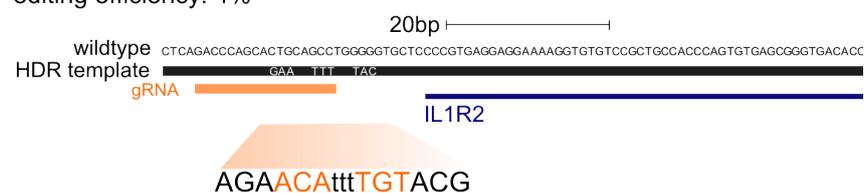


Fig. 5.12: Genomic exchange of *GILZ* GBS1 into other GBS sequence variants does not affect *GILZ* or *FKBP5* expression. Relative *GILZ* and *FKBP5* expression of clonal cell lines containing the FKBP5-2 GBS (n=3) or the PAL GBS (n=3) sequence instead of *GILZ* GBS1 as determined by qPCR after overnight treatment with 1 μ M dexamethasone or vehicle control. A representative from two individual experiments is shown. Horizontal lines indicate the relative average gene expression of individual clonal cell lines. Error bars represent \pm SEM.

IL1R2 GBS intro

gRNA GACCCAGCACTGCAGCC

editing efficiency: 4%



U2OS clone A

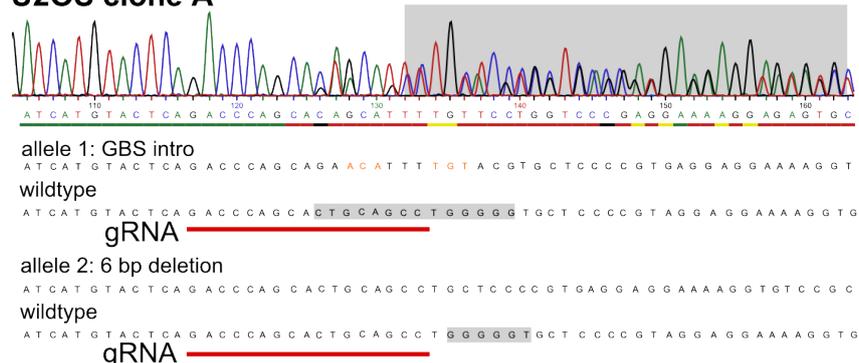


Fig. 5.13: Genotyping results of U2OS cell lines containing a GBS motif upstream of the *IL1R2* gene. Top: Schematics of the *IL1R2* locus showing the location of GBS integration by sequence alignment with the HDR template introducing the CGT GBS sequence. **Bottom:** Genotyping results of successfully CRISPR/Cas9-edited clonal U2OS cells containing the CGT GBS sequence introduced by HDR upstream of the *IL1R2* gene. The respective gRNAs used for CRISPR/Cas9-editing are indicated.

5.2 Genomic location of transcription factor binding sites

Tab. 5.1: Genomic location of target GBSs

target GBS	Location (GRCh37/hg19)
GILZ GBS1	ChrX:106,961,573-106,961,593
GILZ GBS2	ChrX:106,962,033-106,962,053
GILZ GBS3	ChrX:106,962,067-106,962,087
GILZ GBS4	ChrX:106,962,104-106,962,124
DUSP1 GBS1	Chr5:172,199,533-172,199,553

Tab. 5.2: Genomic location of CTCF peaks

CTCF peak	Location (GRCh37/hg19)
CTCF at GBS1	ChrX:106,961,246-106,961,729
CTCF at TSS1	ChrX:107,018,612-107,019,322
CTCF upstream TSS1	ChrX:107,020,386-107,020,778

5.3 gRNA sequences

Tab. 5.3: gRNAs for gene editing

gRNA	Usage	Sequence 5' to 3'
GILZ GBS1	GBS deletion by NHEJ GBS exchange by HDR	CAGGACCAAAGGAGAACATT
GILZ GBS2 a	GBS deletion using two gRNAs	ATGGTGTGGTGCTCTATTGG
GILZ GBS2 b	GBS deletion using two gRNAs	TAAACCTGCTGCACTAGCCC
GILZ GBS4	GBS deletion using two gRNAs	CCAGGTGGTATGGGAAGGGA
DUSP1 GBS1	GBS deletion by NHEJ	GCCGAGTCAGGAACATTCTG
IL1R2 (from [66])	GBS introduction by HDR	GACCCAGCACTGCAGCCTGG

Tab. 5.4: gRNA backbone sequence

gRNA backbone sequence 5' to 3'

TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGT
 CGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGG
 CTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAAT
 ACGTGACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAA
 AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGGCTTTATATAT
 CTTGTGAAAAGGACGAAACACCCGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAAT
 AGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
 TTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATT

Structural features are color-coded (gray: U6 promoter, green: gRNA sequence w/o PAM, blue: gRNA scaffold, red: termination signal) [56].

Tab. 5.5: gRNAs for gene activation by dCas9-VP64 SAM

gRNA	Sequence 5' to 3'
GILZ GBS1	CAGGACCAAAGGAGAACATT
GILZ TSS1	GAGGGAGCAAGGGCGCGCCC
DUSP1 GBS1	GCCGAGTCAGGAACATTCTG
IL1R2 (from [66])	GACCCAGCACTGCAGCCTGG

Tab. 5.6: gRNA backbone sequence dCas9-VP64 SAM

gRNA backbone sequence 5' to 3'

TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGT
 CGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGG
 CTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAAT
 ACGTGACGTAGAAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAA
 ATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTTCGATTTCTTGGCTTTATATATC
 TTGTGAAAAGGACGAAACACCCGNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGGCCAA
 CATGAGGATCACCCATGTCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATC
 AACTTGGCCAACATGAGGATCACCCATGTCTGCAGGGCCAAGTGGCACCGAGTCGGTGC
 TTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATT

Structural features are color-coded (gray: U6 promoter, green: gRNA sequence w/o PAM, blue: gRNA scaffold including tetraloop and two MS2 loops, red: termination signal)[66].

5.4 Primer sequences

All primers were purchased as dried oligonucleotides from Sigma Aldrich.

Tab. 5.7: Primer HDR templates

Primer		Sequence 5' to 3'
GILZ GBS1-4	fwd	GCTGCAGAACGAACCCAAAAG
	rev	CCTCCAAACCCCTACCAAC
IL1R2 promoter	fwd	TCCTAATTGAGGCAGGTCAGC
	rev	CCTGAGCCTCCACCAGAAATG

Tab. 5.8: Primer for SDM of HDR templates

Primer		Sequence 5' to 3'
SDM GILZ GBS1	fwd	ACTCAGGACCAAAGGAGAAGCTTGGGTTCCACC ACATATAACAACAG
	rev	CTGTTGTATATGTGGTGGAAACCCAAGCTTCTCC TTTGGTCCTGAGT
SDM GILZ GBS2	fwd	TGGGAGACAATAATGATCTCAGGATCCTCTGTT AACTTTAAGACACAACCTCT
	rev	AGAGGTTGTGTCTTAAAGTTAACAGAGGATCCT GAGATCATTATTGTCTCCCA
SDM GILZ GBS3/4	fwd	TCTGAAACCGGTGTTTGCACCTAAGACATGTGTA GTAGACAGATCTAAGTGCAGGCT
	rev	AGCCTGCACTTAGATCTGTCTACTACACATGTC TTAGTGCAAACACCGGTTTCAGA
SDM GILZ GBS1 into FKBP5-2	fwd	CAGGACCAAAGGAGAACATCCTGTGCCACCACA TATAC
	rev	GTATATGTGGTGGCACAGGATGTTCTCCTTTGG TCCTG
SDM GILZ GBS1 into PAL	fwd	CAGGACCAAAGGAGAACAAAATGTTCTACCACA TATAC
	rev	GTATATGTGGTAGAACATTTTGTTCCTTTGG TCCTG
SDM IL1R2 GBS intro	fwd	TACTCAGACCCAGCAGAACATTTTGTACGTGCT CCCCGTGAG
	rev	CTCACGGGGAGCACGTACAAAATGTTCTGCTGG GTCTGAGTA

Tab. 5.9: Primer for SDM of luciferase reporter constructs

Primer		Sequence 5' to 3'
SDM GILZ GBS1	fwd	ACTCAGGACCAAAGGAGAAGCTTGGGTTC ACCACATATACAACAG
	rev	CTGTTGTATATGTGGTGGAAACCAAGCTTC TCCTTGGTCCTGAGT
SDM GILZ GBS2	fwd	TGGGAGACAATAATGATCTCAGGATTCTCT GTTAACTTTAAGACACAACCTCT
	rev	AGAGGTTGTGTCTTAAAGTTAACAGAGAAT CCTGAGATCATTATTGTCTCCCA
SDM GILZ GBS3/4	fwd	TCTGAAACCGGTGTTTGCACCTAAGACATGT GTAGTAGACAGATCTAAGTGCAGGCT
	rev	AGCCTGCACTTAGATCTGTCTACTACACATG TCTTAGTGCAAACACCGGTTTCAGA
SDM GILZ GBS1 into FKBP5-2	fwd	CAGGACCAAAGGAGAACATCCTGTGCCACC ACATATAC
	rev	GTATATGTGGTGGCACAGGATGTTCTCCTT TGGTCCTG
SDM GILZ GBS1 into FKBP5	fwd	CAGGACCAAAGGAGAACAGGGTGTCTACC ACATATAC
	rev	GTATATGTGGTAGAACACCCTGTTCTCCTT TGGTCCTG
SDM GILZ GBS1 into PAL	fwd	CAGGACCAAAGGAGAACAAAATGTTCTACC ACATATAC
	rev	GTATATGTGGTAGAACATTTTGTCTCCTT TGGTCCTG
SDM GILZ GBS1 into CGT	fwd	CAGGACCAAAGGAGAACATTTTGTACGACC ACATATAC
	rev	GTATATGTGGTCGTACAAAATGTTCTCCTT TGGTCCTG
SDM GILZ GBS1 into TAT	fwd	CAGGACCAAAGGAGAACATCCCTGTACAAC CACATATAC
	rev	GTATATGTGGTTGTACAGGGATGTTCTCCT TTGGTCCTG
SDM GILZ GBS1 into CONSENSUS	fwd	CAGGACCAAAGGAGAACAAAATGTACCACC ACATATAC
	rev	GTATATGTGGTGGTACATTTTGTCTCCTT TGGTCCTG
SDM GILZ GBS1 into SGK	fwd	CAGGACCAAAGGAGAACATTTTGTCCGACC ACATATAC
	rev	GTATATGTGGTCGGACAAAATGTTCTCCTT TGGTCCTG

Tab. 5.10: Primer for genotyping of edited cells

Primer	Sequence 5' to 3'	
GILZ GBS	fwd	GGAAAACACCTGCCCTGTGA
	rev	CGGGAGGAAATCAAGGCCTT
GILZ GBS1-4	fwd	GGAAAACACCTGCCCTGTGA
	rev	GTCTGAGTCTGGGCTGAACC
DUSP1 GBS1	fwd	CAACCCTCGCTCCCTGTC
	rev	CCTCTTTGCTGTCCTCGACC
IL1R2 promoter	fwd	CTTGGGTGTCTGTTGGGTCT
	rev	CTGGAGAAATCCGGAGAGTCG

Tab. 5.11: qPCR primer for the quantification of gene expression

Primer	Sequence 5' to 3'	
GILZ standard	fwd	CCATGGACATCTTCAACAGC
	rev	TTGGCTCAATCTCTCCCATC
GILZ transcript variant 1	fwd	TACAGTGAGCAACTTTCGGC
	rev	GTTGATCAGGTAGCAGGGGT
GILZ transcript variant 2	fwd	TGGAGTTTGTGACATACGAGG
	rev	AGAACGAACCCAAAGCCAAG
GILZ transcript variant 3	fwd	AATTCCTAGCTAGCTTCAGAGC
	rev	GGCCTGTTCGATCTTGTTGT
DUSP1	fwd	CTGCCTTGATCAACGTCTCA
	rev	GTCTGCCTTGTGGTTGTCCT
FKBP5	fwd	TGAAGGGTTAGCGGAGCAC
	rev	CTTGGCACCTTCATCAGTAGTC
IL1R2 [66]	fwd	CAGGTGAGCAGCAACAAGG
	rev	TGCTCCTGACAACTTCCAGA
RPL19	fwd	ATGTATCACAGCCTGTACCTG
	rev	TTCTTGGTCTCTTCCTCCTTG

Tab. 5.12: qPCR primer for anti-GR ChIP

Primer	Sequence 5' to 3'	
GILZ GBS1	fwd	GTGAGGCCACCTGGTGG
	rev	TATATGTGGTGGAAACCCAATG
GILZ neighboring peak 1	fwd	CTTGCTCTGACAGGGAACAA
	rev	AGATCCCAGAAGAATTGGCAG
GILZ neighboring peak 2	fwd	GATGGAGATAGGAAAAGGGGAG
	rev	GGAGTACTGCCAAGTGCTTTAT
GILZ neighboring peak 3	fwd	ACTGCCTCTTTTCTAAGGGC
	rev	TCTCTCATCTCATCCTCATGGA
GILZ neighboring peak 4	fwd	AACTCAGCAGCTTTTCTTCGT
	rev	AACCAAGGAATTGGGTACAT
GILZ neighboring peak 5	fwd	TCAACGTCCAGACATAGCAAG
	rev	ATAGCTGGGAAATGGTAGCAG
DUSP1 neighboring peak 1	fwd	ATCTTTACAAACAGATCTCCATGC
	rev	TCACACAATGCTGACTACGG
DUSP1 neighboring peak 2	fwd	AAACCGGATCACACACTGAG
	rev	TAAC TTCACCCGAGTTCCTCT
DUSP1 neighboring peak 3	fwd	TGTCGCTGGTACATTTCCAC
	rev	CAGCTGGGTTTCCGATTACA
negative control GR	fwd	AATGGCAGCCCCTAGTCATTC
	rev	AACTGGGAGTGATACTGGTTCC

Tab. 5.13: qPCR primer for anti-CTCF ChIP

Primer	Sequence 5' to 3'	
GILZ GBS1	fwd	GTGAGGCCACCTGGTGG
	rev	TATATGTGGTGGAAACCCAATG
GILZ TSS1	fwd	GAGTGAGCTAGTTAGCGGTC
	rev	CCGTCCCCTCTAGGGTAATTT
GILZ TSS1 upstream	fwd	GGCCTTTGAAATAGAGCAGC
	rev	TACTCTTGGAACCAACGCAC
negative control CTCF	fwd	AATGGCAGCCCCTAGTCATTC
	rev	AACTGGGAGTGATACTGGTTCC
positive control CTCF	fwd	GTGATCGGTCCAGTGCATAG
	rev	CTGGCATGTCATGGTAGAGC

Tab. 5.14: qPCR primer for anti-H3K27Ac ChIP

Primer	Sequence 5' to 3'	
GILZ GBS1	fwd	GTGAGGCCACCTGGTGG
	rev	TATATGTGGTGGAAACCCAATG
GILZ TSS1	fwd	GAGTGAGCTAGTTAGCGGTC
	rev	CCGTCCCCTCTAGGGTAATTT
GILZ TSS1 upstream	fwd	GGCCTTTGAAATAGAGCAGC
	rev	TACTCTTGGAACCAACGCAC
negative control H3K27Ac	fwd	AATGGCAGCCCCTAGTCATTC
	rev	AACTGGGAGTGATACTGGTTCC
positive control H3K27Ac	fwd	AGGAATATTTGCTGACACTTCCA
	rev	ACAGCACCTACCATATAGGCTT

Tab. 5.15: Primer for inverse PCR

Viewpoint	Sequence 5' to 3'	
GILZ GBS1 adjacent	fwd	CTACACGACGCTCTTCCGATCTAATG TTCAGGTGTGGGAGTAC
	rev	CAGACGTGTGCTCTTCCGATCTGCCCTCTGCCTCTTGTTAGG

6 List of Abbreviations

3C	Chromosome conformation capture
3D	Three-dimensional
4C	Circularized chromosome conformation capture
Cas9	CRISPR-associated protein 9
ChIP	Chromatin immunoprecipitation
CRISPR	Clustered regularly interspaced short palindromic repeats
CTCF	CCCTC- binding factor
dCas9	enzymatically dead CRISPR-associated protein 9
Dex	Dexamethasone
DNA	Deoxyribonucleic acid
DUSP1	Dual specificity protein phosphatase 1
EMSA	Electrophoretic mobility shift assay
EtOH	Ethanol
FISH	Fluorescence <i>in situ</i> hybridization
GBS	Glucocorticoid receptor binding site
GC	Glucocorticoids
GILZ	Glucocorticoid induced leucine zipper
GR	Glucocorticoid receptor
gRNA	guide RNA
H3K27Ac	Acetylation of histone H3 at lysine 27
HDR	Homology-directed repair
indels	insertions and deletions
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
SAM	Synergistic activation mediator
SDM	Site-directed mutagenesis
SNP	Single nucleotide polymorphism
TAD	Topologically associating domain
TALEN	TAL effector nucleases
TF	Transcription factor
TFBS	Transcription factor binding site
TSS	Transcription start site
ZFN	Zinc finger nucleases

List of Figures

1.1	Transcriptional regulation of gene expression.	2
1.2	Overview of the basic steps of the ChIP procedure.	3
1.3	Modes of combinatorial interaction.	6
1.4	Overview of a circularized chromosome conformation capture (4C) experiment.	8
1.5	Gene editing using the CRISPR/Cas9 system.	11
1.6	Signaling pathway of the glucocorticoid receptor.	13
2.1	Workflow for the generation of CRISPR/Cas9 edited clonal cell lines.	24
3.1	Genomic editing of GBSs using CRISPR/Cas9 disrupts the GR binding motif.	40
3.2	Editing of GBSs by CRISPR/Cas9 reduces their GR binding affinity.	41
3.3	Quantification of <i>GILZ</i> alleles in U2OS cells.	42
3.4	Genomic deletion of GBSs reduces expression of the nearby gene.	43
3.5	Multiple GBSs regulate <i>GILZ</i> expression.	44
3.6	<i>GILZ</i> GBS1-4 cooperatively regulate <i>GILZ</i> expression as a functional unit.	45
3.7	Genomic deletion diminishes GR binding at deleted GBSs.	47
3.8	Genomic deletion remains GR binding at neighboring GBSs unaffected.	48
3.9	In an episomal context enhancer activity of the <i>GILZ</i> promoter depends on the presence of GBS1.	49
3.10	Genomic editing of <i>GILZ</i> and <i>DUSP1</i> GBS1 in A549 cells.	50
3.11	Quantification of <i>GILZ</i> alleles in A549 cells.	51
3.12	Effect of genomic GBS deletion in A549 cells.	52
3.13	<i>GILZ</i> transcript variants are located within the same TAD.	53
3.14	In A549 cells <i>GILZ</i> GBS1 is unable to activate the expression of nearby <i>GILZ</i> transcript variants.	54
3.15	The pattern of H3K27Ac at the <i>GILZ</i> locus.	56
3.16	In an episomal context GBS1 is required for enhancer activity of the <i>GILZ</i> promoter region irrespective of cell type.	57
3.17	In its genomic context <i>GILZ</i> GBS1 is unable to activate the expression of nearby transcript variants in A549 cells.	59

List of Figures

3.18	A549 cells but not U2OS cells show increased interactions and cell type-specific CTCF-binding at <i>GILZ</i> TSS1.	60
3.19	CTCF binding at the <i>GILZ</i> locus is not affected by hormone treatment. . .	62
3.20	CTCF depletion by siRNAs does not affect the regulation of <i>GILZ</i> transcript variants.	63
3.21	GR motif variants differentially affect reporter gene expression.	65
3.22	Genomic exchange of <i>GILZ</i> GBS1 into other GR motif variants does not affect <i>GILZ</i> expression.	66
3.23	Genomic exchange of GR motif variants does not affect <i>GILZ</i> expression at low concentrations of hormone.	68
3.24	Genomic location of the site for GR binding sequence integration.	69
3.25	<i>IL1R2</i> becomes regulated by GR upon introduction of a single promoter-proximal GR binding sequence.	70
4.1	GBS1 and GBS2-4 cooperatively regulate <i>GILZ</i> expression as a single functional unit.	79
4.2	Model for the cell type-specific regulatory activity of <i>GILZ</i> GBS1 mediated by differential CTCF binding and DNA-looping.	84
5.1	Genotyping results of U2OS <i>GILZ</i> GBS1 deletion clones.	92
5.2	Genotyping results of U2OS <i>DUSP1</i> GBS1 deletion clones.	93
5.3	Genotyping results of U2OS <i>GILZ</i> GBS2-4 deletion clones.	94
5.4	Genotyping results of U2OS <i>GILZ</i> GBS1-4 deletion clones.	95
5.5	Genotyping results of U2OS <i>GILZ</i> GBS2-4 inversion clones.	96
5.6	Genotyping results of A549 <i>GILZ</i> GBS1 deletion clones.	97
5.7	Genotyping results of A549 <i>DUSP1</i> GBS1 deletion clones.	98
5.8	CTCF motif orientation at CTCF-bound regions of the <i>GILZ</i> locus.	99
5.9	Genomic sequence at <i>GILZ</i> TSS1 in U2OS and A549 cells.	100
5.10	Genotyping results of U2OS cell lines with <i>GILZ</i> GBS1 exchanged into the FKBP5-2 sequence.	101
5.11	Genotyping results of U2OS cell lines with <i>GILZ</i> GBS1 exchanged into the PAL sequence.	102
5.12	Genomic exchange of <i>GILZ</i> GBS1 into other GBS sequence variants does not affect <i>GILZ</i> or <i>FKBP5</i> expression.	102
5.13	Genotyping results of U2OS cell lines containing a GBS motif upstream of the <i>IL1R2</i> gene.	103

List of Tables

2.1	Plasmids	18
2.2	LB medium	18
2.3	LB agar	18
2.4	SOC medium	18
2.5	IP lysis buffer for ChIP	19
2.6	RIPA buffer for ChIP	19
2.7	RIPA wash buffer for ChIP	19
2.8	RIPA wash buffer for anti-CTCF ChIP	19
2.9	LiCl wash buffer for ChIP	20
2.10	Cross-link reversal solution for ChIP	20
2.11	Lysis buffer for 4C	20
2.12	10x ligation buffer for 4C	20
2.13	SSC buffer for FISH	20
2.14	qPCR master mix	20
2.15	PCR mix	21
2.16	PCR thermocycling program	21
2.17	qPCR mix	22
2.18	qPCR thermocycling program	22
2.19	SDM reaction mix	23
2.20	SDM thermocycling program	23
2.21	Ligation reaction for Zero Blunt cloning	24
2.22	Plasmid mix for gene editing by NHEJ	25
2.23	Plasmid mix for gene editing by HDR	25
2.24	Reverse transcription mix 1	27
2.25	Reverse transcription mix 2	27
2.26	Plasmid mix for gene activation by dCas9-VP64	27
2.27	Luciferase assay activation by GR plasmid mix	28
2.28	Luciferase assay activation by dCas9-VP64 plasmid mix	28
2.29	Mix for inverse PCR	35
2.30	Thermocycling program for inverse PCR	35
2.31	FISH washing steps	37

List of Tables

5.1	Genomic location of target GBSs	104
5.2	Genomic location of CTCF peaks	104
5.3	gRNA sequences for gene editing by CRISPR/Cas9	104
5.4	gRNA backbone sequence for gene editing by CRISPR/Cas9	105
5.5	gRNA sequences for gene activation by dCas9-VP64 SAM	105
5.6	gRNA backbone sequence for gene activation by dCas9-VP64 SAM	105
5.7	Primer sequences for amplification of HDR templates from genomic DNA .	106
5.8	Primer sequences for SDM of HDR templates	106
5.9	Primer sequences for SDM of luciferase reporter constructs	107
5.10	Primer sequences for genotyping of CRISPR/Cas9-edited cells	108
5.11	qPCR primer sequences for the quantification of gene expression	108
5.12	ChIP qPCR primer sequences anti-GR	109
5.13	ChIP qPCR primer sequences anti-CTCF	109
5.14	ChIP qPCR primer sequences anti-H3K27Ac	109
5.15	Primer sequences for inverse PCR	110

7 References

1. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74. ISSN: 0028-0836 (2012).
2. Gregory, T. R. Synergy between sequence and size in large-scale genomics. *Nat Rev Genet* **6**, 699–708. ISSN: 1471-0056 (Print) 1471-0056 (2005).
3. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–5. ISSN: 0036-8075 (2012).
4. Ratman, D. *et al.* How glucocorticoid receptors modulate the activity of other transcription factors: a scope beyond tethering. *Mol Cell Endocrinol* **380**, 41–54. ISSN: 0303-7207 (2013).
5. Riethoven, J. J. Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. *Methods Mol Biol* **674**, 33–42. ISSN: 1064-3745 (2010).
6. Bushey, A. M., Dorman, E. R. & Corces, V. G. Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Mol Cell* **32**, 1–9. ISSN: 1097-2765 (2008).
7. Yankulov, K., Blau, J., Purton, T., Roberts, S. & Bentley, D. L. Transcriptional elongation by RNA polymerase II is stimulated by transactivators. *Cell* **77**, 749–59. ISSN: 0092-8674 (Print) 0092-8674 (1994).
8. Yudkovsky, N., Ranish, J. A. & Hahn, S. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**, 225–9. ISSN: 0028-0836 (Print) 0028-0836 (2000).
9. Amano, T. *et al.* Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell* **16**, 47–57. ISSN: 1534-5807 (2009).
10. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**, 1453–65. ISSN: 1097-2765 (Print) 1097-2765 (2002).
11. Cusanovich, D. A., Pavlovic, B., Pritchard, J. K. & Gilad, Y. The functional consequences of variation in transcription factor binding. *PLoS Genet* **10**, e1004226. ISSN: 1553-7390 (2014).
12. Gitter, A. *et al.* Backup in gene regulatory networks explains differences between binding and knockout results. *Mol Syst Biol* **5**, 276. ISSN: 1744-4292 (2009).

7. References

13. Gertz, J. *et al.* Distinct properties of cell-type-specific and shared transcription factor binding sites. *Mol Cell* **52**, 25–36. ISSN: 1097-2765 (2013).
14. Morikawa, M. *et al.* ChIP-seq reveals cell type-specific binding patterns of BMP-specific Smads and a novel binding motif. *Nucleic Acids Res* **39**, 8712–27. ISSN: 0305-1048 (Print) (2011).
15. Kvon, E. Z. *et al.* Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**, 91–5. ISSN: 0028-0836 (2014).
16. Sandmann, T. *et al.* A temporal map of transcription factor activity: mef2 directly regulates target genes at all stages of muscle development. *Dev Cell* **10**, 797–807. ISSN: 1534-5807 (Print) 1534-5807 (2006).
17. Talbert, P. B. & Henikoff, S. Histone variants—ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* **11**, 264–75. ISSN: 1471-0072 (2010).
18. He, H. H. *et al.* Differential DNase I hypersensitivity reveals factor-dependent chromatin dynamics. *Genome Res* **22**, 1015–25. ISSN: 1088-9051 (2012).
19. Biddie, S. C. *et al.* Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol Cell* **43**, 145–55. ISSN: 1097-2765 (2011).
20. West, J. A. *et al.* Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. *Nat Commun* **5**, 4719. ISSN: 2041-1723 (2014).
21. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931–6. ISSN: 0027-8424 (2010).
22. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311–8. ISSN: 1061-4036 (Print) 1061-4036 (2007).
23. Weber, C. M. & Henikoff, S. Histone variants: dynamic punctuation in transcription. *Genes Dev* **28**, 672–82. ISSN: 0890-9369 (2014).
24. Li, E. & Zhang, Y. DNA methylation in mammals. *Cold Spring Harb Perspect Biol* **6**, a019133. ISSN: 1943-0264 (2014).
25. Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–80. ISSN: 0092-8674 (2014).
26. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–4. ISSN: 0028-0836 (2013).
27. Kilpinen, H. *et al.* Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* **342**, 744–7. ISSN: 0036-8075 (2013).

-
28. Corradin, O. *et al.* Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res* **24**, 1–13. ISSN: 1088-9051 (2014).
 29. Meyer, M. B., Benkusky, N. A. & Pike, J. W. Selective Distal Enhancer Control of the Mmp13 Gene Identified through Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) Genomic Deletions. *J Biol Chem* **290**, 11093–107. ISSN: 0021-9258 (2015).
 30. Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, 613–26. ISSN: 1471-0056 (2012).
 31. Panne, D., Maniatis, T. & Harrison, S. C. An atomic model of the interferon-beta enhanceosome. *Cell* **129**, 1111–23. ISSN: 0092-8674 (Print) 0092-8674 (2007).
 32. Hay, D. *et al.* Genetic dissection of the alpha-globin super-enhancer in vivo. *Nat Genet* **48**, 895–903. ISSN: 1061-4036 (Print) 1061-4036 (2016).
 33. Giorgetti, L. *et al.* Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. *Mol Cell* **37**, 418–28. ISSN: 1097-2765 (2010).
 34. Cannavo, E. *et al.* Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr Biol* **26**, 38–51. ISSN: 0960-9822 (2016).
 35. Barolo, S. Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays* **34**, 135–41. ISSN: 0265-9247 (2012).
 36. Shin, H. Y. *et al.* Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat Genet* **48**, 904–11. ISSN: 1061-4036 (2016).
 37. So, A. Y., Chaivorapol, C., Bolton, E. C., Li, H. & Yamamoto, K. R. Determinants of cell- and gene-specific transcriptional regulation by the glucocorticoid receptor. *PLoS Genet* **3**, e94. ISSN: 1553-7390 (2007).
 38. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* **4**, 651–7. ISSN: 1548-7091 (Print) 1548-7091 (2007).
 39. MacQuarrie, K. L., Fong, A. P., Morse, R. H. & Tapscott, S. J. Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet* **27**, 141–8. ISSN: 0168-9525 (Print) 0168-9525 (2011).
 40. Ramani, V., Shendure, J. & Duan, Z. Understanding Spatial Genome Organization: Methods and Insights. *Genomics Proteomics Bioinformatics* **14**, 7–20. ISSN: 1672-0229 (Print) (2016).

7. References

41. Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet* **14**, 390–403. ISSN: 1471-0056 (2013).
42. Van de Werken, H. J. *et al.* 4C technology: protocols and data analysis. *Methods Enzymol* **513**, 89–112. ISSN: 0076-6879 (2012).
43. Cremer, T. & Cremer, M. Chromosome Territories. *Cold Spring Harb Perspect Biol* **2**. doi:10.1101/cshperspect.a003889 (2010).
44. Kind, J. & van Steensel, B. Genome-nuclear lamina interactions and gene regulation. *Curr Opin Cell Biol* **22**, 320–5. ISSN: 0955-0674 (2010).
45. Dixon, J. R. *et al.* Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* **485**, 376–80. ISSN: 0028-0836 (Print) (2012).
46. Palstra, R. J. *et al.* The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* **35**, 190–4. ISSN: 1061-4036 (Print) 1061-4036 (2003).
47. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences* **111**, 996–1001 (2014).
48. Phillips, J. E. & Corces, V. G. CTCF: Master Weaver of the Genome. *Cell* **137**, 1194–211. ISSN: 0092-8674 (Print) (2009).
49. Liu, M. *et al.* Genomic discovery of potent chromatin insulators for human gene therapy. *Nat Biotechnol* **33**, 198–203. ISSN: 1087-0156 (2015).
50. Bell, A. C., West, A. G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**, 387–96. ISSN: 0092-8674 (Print) 0092-8674 (1999).
51. Lobanenko, V. V. *et al.* A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken *c-myc* gene. *Oncogene* **5**, 1743–53. ISSN: 0950-9232 (Print) 0950-9232 (1990).
52. Lee, B. K. *et al.* Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells. *Genome Res* **22**, 9–24. ISSN: 1088-9051 (Print) (2012).
53. De Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* **60**, 676–84. ISSN: 1097-2765 (2015).
54. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–21. ISSN: 0036-8075 (2012).

-
55. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–6. ISSN: 0036-8075 (2013).
 56. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–23. ISSN: 0036-8075 (Print) (2013).
 57. Cho, S. W., Kim, S., Kim, J. M. & Kim, J. S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**, 230–2. ISSN: 1087-0156 (2013).
 58. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–8. ISSN: 0028-0836 (2012).
 59. Fineran, P. C. & Charpentier, E. Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* **434**, 202–9. ISSN: 0042-6822 (2012).
 60. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–70. ISSN: 0036-8075 (2010).
 61. Miyaoka, Y. *et al.* Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci Rep* **6**, 23549. ISSN: 2045-2322 (2016).
 62. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)* **7**, 1765–71. ISSN: 1568-7864 (Print) 1568-7856 (2008).
 63. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–83. ISSN: 0092-8674 (Print) (2013).
 64. Mali, P. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. **31**, 833–8. ISSN: 1087-0156 (Print) (2013).
 65. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat Meth* **10**, 977–979. ISSN: 1548-7091 (2013).
 66. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–8. ISSN: 0028-0836 (2015).
 67. Pujols, L. *et al.* Expression of glucocorticoid receptor alpha- and beta-isoforms in human cells and tissues. *Am J Physiol Cell Physiol* **283**, C1324–31. ISSN: 0363-6143 (Print) 0363-6143 (2002).
 68. Kuo, T., Harris, C. A. & Wang, J. C. Metabolic functions of glucocorticoid receptor in skeletal muscle. *Mol Cell Endocrinol* **380**, 79–88. ISSN: 0303-7207 (2013).

7. References

69. Wang, J. C., Gray, N. E., Kuo, T. & Harris, C. A. Regulation of triglyceride metabolism by glucocorticoid receptor. *Cell Biosci* **2**, 19 (2012).
70. Phuc Le, P. Glucocorticoid Receptor-Dependent Gene Regulatory Networks. **1**. ISSN: 1553-7390 (Print). doi:10.1371/journal.pgen.0010016 (2005).
71. Cohen, S. *et al.* Chronic stress, glucocorticoid receptor resistance, inflammation, and disease risk. *Proc Natl Acad Sci U S A* **109**, 5995–9. ISSN: 0027-8424 (Print) (2012).
72. Baschant, U. & Tuckermann, J. The role of the glucocorticoid receptor in inflammation and immunity. *J Steroid Biochem Mol Biol* **120**, 69–75. ISSN: 0960-0760 (2010).
73. Carlstedt-Duke, J. *et al.* Domain structure of the glucocorticoid receptor protein. *Proc Natl Acad Sci U S A* **84**, 4437–40. ISSN: 0027-8424 (Print) 0027-8424 (1987).
74. Lu, N. Z. & Cidlowski, J. A. Glucocorticoid receptor isoforms generate transcription specificity. *Trends Cell Biol* **16**, 301–7. ISSN: 0962-8924 (Print) 0962-8924 (2006).
75. Kirschke, E., Goswami, D., Southworth, D., Griffin, P. R. & Agard, D. A. Glucocorticoid receptor function regulated by coordinated action of the Hsp90 and Hsp70 chaperone cycles. *Cell* **157**, 1685–97. ISSN: 0092-8674 (2014).
76. Luisi, B. F. *et al.* Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, 497–505. ISSN: 0028-0836 (Print) 0028-0836 (1991).
77. Hudson, W. H., Youn, C. & Ortlund, E. A. The structural basis of direct glucocorticoid-mediated transrepression. *Nat Struct Mol Biol* **20**, 53–8. ISSN: 1545-9985 (2013).
78. Starick, S. R. *et al.* ChIP-exo signal associated with DNA-binding motifs provides insight into the genomic binding of the glucocorticoid receptor and cooperating transcription factors. *Genome Res* **25**, 825–35. ISSN: 1088-9051 (2015).
79. Schiller, B. J., Chodankar, R., Watson, L. C., Stallcup, M. R. & Yamamoto, K. R. Glucocorticoid receptor binds half sites as a monomer and regulates specific target genes. *Genome Biol* **15**, 418. ISSN: 1474-7596 (2014).
80. Sacta, M. A., Chinenov, Y. & Rogatsky, I. Glucocorticoid Signaling: An Update from a Genomic Perspective. *Annu Rev Physiol* **78**, 155–80. ISSN: 0066-4278 (2016).
81. Vegiopoulos, A. & Herzig, S. Glucocorticoids, metabolism and metabolic diseases. *Mol Cell Endocrinol* **275**, 43–61. ISSN: 0303-7207 (Print) 0303-7207 (2007).
82. Weikum, E. R., Knuesel, M. T., Ortlund, E. A. & Yamamoto, K. R. Glucocorticoid receptor control of transcription: precision and plasticity via allostery. *Nat Rev Mol Cell Biol* **18**, 159–174. ISSN: 1471-0072 (2017).
83. Meijsing, S. H. *et al.* DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* **324**, 407–10. ISSN: 0036-8075 (2009).

-
84. Watson, L. C. *et al.* The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. *Nat Struct Mol Biol* **20**, 876–83. ISSN: 1545-9993 (Print) (2013).
 85. Lefstin, J. A. & Yamamoto, K. R. Allosteric effects of DNA on transcriptional regulators. *Nature* **392**, 885–8. ISSN: 0028-0836 (Print) 0028-0836 (1998).
 86. Schone, S. *et al.* Sequences flanking the core-binding site modulate glucocorticoid receptor structure and activity. *Nat Commun* **7**, 12621. ISSN: 2041-1723 (2016).
 87. Rogatsky, I., Trowbridge, J. M. & Garabedian, M. J. Glucocorticoid receptor-mediated cell cycle arrest is achieved through distinct cell-specific transcriptional regulatory mechanisms. *Mol Cell Biol* **17**, 3181–93. ISSN: 0270-7306 (Print) 0270-7306 (1997).
 88. Wang, J. C. *et al.* Chromatin immunoprecipitation (ChIP) scanning identifies primary glucocorticoid receptor target genes. *Proc Natl Acad Sci U S A* **101**, 15603–8. ISSN: 0027-8424 (Print) 0027-8424 (2004).
 89. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827–32. ISSN: 1087-0156 (2013).
 90. Love, M. I. *et al.* Role of the chromatin landscape and sequence in determining cell type-specific genomic glucocorticoid receptor binding and gene regulation. *Nucleic Acids Res.* ISSN: 0305-1048. doi:10.1093/nar/gkw1163 (2016).
 91. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24–6. ISSN: 1087-0156 (2011).
 92. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178–92. ISSN: 1467-5463 (2013).
 93. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–7. ISSN: 1367-4803 (2010).
 94. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–60. ISSN: 1367-4803 (2009).
 95. Heng, L. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Genomics* **1303**. doi:arXiv:1303.3997 (2013).
 96. Manke, T., Heinig, M. & Vingron, M. Quantifying the effect of sequence variation on regulatory interactions. *Hum Mutat* **31**, 477–83. ISSN: 1059-7794 (2010).
 97. Thomas-Chollier, M. *et al.* Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. *Nat Protoc* **6**, 1860–9. ISSN: 1750-2799 (2011).

7. References

98. Owens, D. M. & Keyse, S. M. Differential regulation of MAP kinase signalling by dual-specificity protein phosphatases. *Oncogene* **26**, 3203–13. ISSN: 0950-9232 (Print) 0950-9232 (2007).
99. Ayroldi, E. & Riccardi, C. Glucocorticoid-induced leucine zipper (GILZ): a new important mediator of glucocorticoid action. *Faseb j* **23**, 3649–58. ISSN: 0892-6638 (2009).
100. Shipp, L. E. *et al.* Transcriptional regulation of human dual specificity protein phosphatase 1 (DUSP1) gene by glucocorticoids. *PLoS One* **5**, e13754. ISSN: 1932-6203 (2010).
101. Mirny, L. A. Nucleosome-mediated cooperativity between transcription factors. *Proc Natl Acad Sci U S A* **107**, 22534–9. ISSN: 0027-8424 (2010).
102. Lin, Y. S., Carey, M., Ptashne, M. & Green, M. R. How different eukaryotic transcriptional activators can cooperate promiscuously. *Nature* **345**, 359–61. ISSN: 0028-0836 (Print) 0028-0836 (1990).
103. John, S. *et al.* Kinetic complexity of the global response to glucocorticoid receptor action. *Endocrinology* **150**, 1766–74. ISSN: 0013-7227 (2009).
104. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* **43**, 264–8. ISSN: 1061-4036 (2011).
105. Pruitt, K. D. *et al.* RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res* **42**, D756–63. ISSN: 0305-1048 (2014).
106. Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900–10. ISSN: 0092-8674 (2015).
107. Smith, C. L. & Hager, G. L. Transcriptional regulation of mammalian genes in vivo. A tale of two templates. *J Biol Chem* **272**, 27493–6.
108. Inoue, F. *et al.* A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res* **27**, 38–52.
109. Xu, H. *et al.* Sequence determinants of improved CRISPR sgRNA design. *Genome Res* **25**, 1147–57. ISSN: 1088-9051 (2015).
110. Chari, R., Mali, P., Moosburner, M. & Church, G. M. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat Methods* **12**, 823–6. ISSN: 1548-7091 (2015).
111. Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E. & Schier, A. F. Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat Commun* **7**, 11750. ISSN: 2041-1723 (2016).
112. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281–308. ISSN: 1754-2189 (Print) (2013).

-
113. Cheng, C. *et al.* Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome Res* **22**, 1658–67. ISSN: 1088-9051 (2012).
 114. Ouyang, Z., Zhou, Q. & Wong, W. H. ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. *Proc Natl Acad Sci U S A* **106**, 21521–6. ISSN: 0027-8424 (2009).
 115. Gaszner, M. & Felsenfeld, G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* **7**, 703–13. ISSN: 1471-0056 (Print) 1471-0056 (2006).
 116. Ali, T., Renkawitz, R. & Bartkuhn, M. Insulators and domains of gene expression. *Curr Opin Genet Dev* **37**, 17–26. ISSN: 0959-437x (2016).
 117. Vockley, C. M. *et al.* Direct GR Binding Sites Potentiate Clusters of TF Binding across the Human Genome. *Cell* **166**, 1269–1281.e19. ISSN: 0092-8674 (2016).
 118. Huang, J. *et al.* Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Dev Cell* **36**, 9–23. ISSN: 1534-5807 (2016).
 119. Korkmaz, G. *et al.* Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol* **34**, 192–8. ISSN: 1087-0156 (2016).
 120. Sanjana, N. E. *et al.* High-resolution interrogation of functional elements in the noncoding genome. *Science* **353**, 1545–1549. ISSN: 0036-8075 (2016).
 121. Tsankov, A. M. *et al.* Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**, 344–9. ISSN: 0028-0836 (2015).
 122. Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477–81. ISSN: 0028-0836 (Print) (2013).
 123. Perry, M. W., Boettiger, A. N. & Levine, M. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proc Natl Acad Sci U S A* **108**, 13570–5. ISSN: 0027-8424 (2011).
 124. Frankel, N. *et al.* Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490–3. ISSN: 0028-0836 (2010).
 125. O’Meara, M. M. *et al.* Cis-regulatory mutations in the *Caenorhabditis elegans* homeobox gene locus *cog-1* affect neuronal development. *Genetics* **181**, 1679–86. ISSN: 0016-6731 (Print) 0016-6731 (2009).
 126. Dunipace, L., Saunders, A., Ashe, H. L. & Stathopoulos, A. Autoregulatory feedback controls sequential action of cis-regulatory modules at the brinker locus. *Dev Cell* **26**, 536–43. ISSN: 1534-5807 (2013).
 127. Crocker, J. *et al.* Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* **160**, 191–203. ISSN: 0092-8674 (2015).

7. References

128. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–61. ISSN: 0028-0836 (2014).
129. Merika, M., Williams, A. J., Chen, G., Collins, T. & Thanos, D. Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. *Mol Cell* **1**, 277–87. ISSN: 1097-2765 (Print) 1097-2765 (1998).
130. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569–77. ISSN: 0028-0836 (Print) 0028-0836 (1997).
131. Tjian, R. & Maniatis, T. Transcriptional activation: a complex puzzle with few easy pieces. *Cell* **77**, 5–8. ISSN: 0092-8674 (Print) 0092-8674 (1994).
132. Rao, N. A. S. *et al.* Coactivation of GR and NFkB alters the repertoire of their binding sites and target genes. *Genome Res* **21**, 1404–16. ISSN: 1088-9051 (Print) (2011).
133. Rui, L. Energy Metabolism in the Liver. *Compr Physiol* **4**, 177–97.
134. Ayroldi, E. *et al.* L-GILZ binds p53 and MDM2 and suppresses tumor growth through p53 activation in human cancer cells. *Cell Death Differ* **22**, 118–30. ISSN: 1350-9047 (2015).
135. Hubbard, T. *et al.* The Ensembl genome database project. *Nucleic Acids Res* **30**, 38–41.
136. Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100. ISSN: 0028-0836 (2014).
137. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336. ISSN: 0028-0836 (2015).
138. Geyer, P. K. & Corces, V. G. DNA position-specific repression of transcription by a Drosophila zinc finger protein. *Genes Dev* **6**, 1865–73. ISSN: 0890-9369 (Print) 0890-9369 (1992).
139. Muravyova, E. *et al.* Loss of insulator activity by paired Su(Hw) chromatin insulators. *Science* **291**, 495–8. ISSN: 0036-8075 (Print) 0036-8075 (2001).
140. Udvardy, A., Maine, E. & Schedl, P. The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *J Mol Biol* **185**, 341–58. ISSN: 0022-2836 (Print) 0022-2836 (1985).
141. Ong, C. T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* **15**, 234–46. ISSN: 1471-0056 (2014).
142. Song, L. *et al.* Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res* **21**, 1757–67. ISSN: 1088-9051 (Print) (2011).

-
143. DeMare, L. E. *et al.* The genomic landscape of cohesin-associated chromatin interactions. *Genome Res* **23**, 1224–34. ISSN: 1088-9051 (2013).
 144. Tang, Z. *et al.* CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**, 1611–27. ISSN: 0092-8674 (2015).
 145. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930–944.e22. ISSN: 0092-8674 (2017).
 146. Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res* **22**, 1680–8. ISSN: 1088-9051 (2012).
 147. Liu, X. S. *et al.* Editing DNA Methylation in the Mammalian Genome. *Cell* **167**, 233–247.e17. ISSN: 0092-8674 (2016).
 148. Shukla, S. *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–9. ISSN: 0028-0836 (2011).
 149. Peters, V. A., Joesting, J. J. & Freund, G. G. IL-1 receptor 2 (IL-1R2) and its role in immune regulation. *Brain Behav Immun* **32**, 1–8. ISSN: 0889-1591 (Print) (2013).
 150. Vambutas, A. *et al.* Alternate splicing of interleukin-1 receptor type II (IL1R2) in vitro correlates with clinical glucocorticoid responsiveness in patients with AIED. *PLoS One* **4**, e5293. ISSN: 1932-6203 (2009).
 151. Zabidi, M. A. *et al.* Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556–9. ISSN: 0028-0836 (2015).
 152. Butler, J. E. & Kadonaga, J. T. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev* **15**, 2515–9. ISSN: 0890-9369 (Print) 0890-9369 (2001).
 153. Merli, C., Bergstrom, D. E., Cygan, J. A. & Blackman, R. K. Promoter specificity mediates the independent regulation of neighboring genes. *Genes Dev* **10**, 1260–70. ISSN: 0890-9369 (Print) 0890-9369 (1996).