DNA-encoded signals regulate genomic binding of transcription factors

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

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

vorgelegt von

Jonas Telorac
aus Berlin

Juni 2014

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

1. Gutachter:	Dr. Sebastiaan H. Meijsing
	Max-Planck-Institut für molekulare Genetik
2. Gutachter:	Prof. Dr. Markus Wahl
	Freie Universität Berlin

Disputation am: 13.10.2014

Acknowledgements

I am very grateful to my advisor Dr. Sebastiaan H. Meijsing for always letting me follow own research ideas and being open for scientific discussion as well as his valuable advice. I also want to thank Prof. Dr. Markus Wahl for reviewing this work.

For the great atmosphere and all the fruitful discussions I want to thank everybody who accompanied me during my time at the MPI-MG, especially all my fellow lab members. I really enjoyed all those fun and controversial discussions, especially with Stephan Starick. In addition, I also want to thank Dr. Sergey Prykhozhij for all his help with the Zebrafish experiments and for a nice time during the late hour microscopy evenings, but most importantly for his patience in trying to teach me injections.

I also want to express my gratitude to Dr. Morgane Thomas-Chollier. Large parts of this work would not have been possible without her help and sophisticated bioinformatical analyses.

I also want to thank Dr. Saskia Hutten and the rest the Lamond Lab at the University of Dundee in Scotland for teaching me a lot about subnuclear structures and for a great time during my stay in Dundee.

For their commendable technical assistance I am very grateful to Edda Einfeldt and Katja Borzym.

Table of Content

1 Introduction

1.1 Transcription

Genes encoded in the genome provide the information to generate functional gene products either in the form of RNAs or of proteins. Gene expression is the process that generates gene products such as proteins from a gene. The expression level of such gene products is regulated in different stages (e.g.: transcription of the DNA-encoded gene to mRNA, subsequent mRNA processing and nuclear export to the ribosomes and finally translation of the mRNA to protein). In my thesis, I focused on mechanisms that regulate gene activity at the initial step of gene expression, to be more specific, the transcription of genes by the RNA polymerase II (PolII) enzyme. The largest subunit of PolII, Rpb1, harbors a C-terminal domain (CTD) containing a repetitive amino acid sequence of Y-S-P-T-S-P-S, with 52 repeats in mammals[1]. Post translational modifications of this CTD are important to control PolII-mediated transcription (for a review see[2]), a process that can be sub-divided into three steps: initiation, elongation and termination. Regulation of transcription is facilitated by specialized proteins, so-called transcriptional regulatory factors (TFs) that bind to specific genomic DNA sequences associated with target genes to control PolII activity especially during initiation and elongation. The set of TFs expressed largely depends on the given cell type, thereby defining which genomic TF binding sites are bound and which genes are expressed[3].

1.1.1 Initiation

Initiation of transcription starts with the recruitment of RNA PolII and other general transcription factors (gTFs) to the core promoter of genes. This is facilitated by special DNA-encoded sequence elements (for example the TATA box or the CAT box) located approximately 25-35 base pairs (bp) upstream of the transcription start site[4]. The assembly of this complex causes a local melting of the DNA double-helix[5,6] and PolII can initiate abortive rounds of transcription of the first nucleotides while remaining bound to the pre-initiation complex[7]. Subsequent initiation requires phosphorylation of serine 5 in the CTD of PolII by the general transcription factor TFIIH[8]. This phosphorylation is thought to destabilize PolII interactions with the other promoter bound proteins, thus allowing the escape of the PolII from the promoter to initiate transcription[9].

As indicated above, key players in the initiation of transcription are TFs and the multi-subunit mediator complex. This complex helps recruit PolII to the promoter to form

the pre-initiation complex[10] and plays an important role in initiation by serine 5 phosphorylation in the CTD of PolII[11] to initiate the transition from abortive transcription to elongation. TFs that regulate this process bind to binding sites encoded either directly within the promoter region of target genes or in so-called enhancer elements that can be up to several 100 kilo base pairs (kbs) up- or downstream of the promoter. Bound TFs in turn recruit co-activators or co-repressors that can influence various steps of the transcription cycle. To allow gene regulation, these proteins have to interact either directly or indirectly with PolII. In addition to its role in forming the pre-initiation complex, the interaction of these regulatory proteins and RNA PolII is bridged by the mediator complex. For example, interactions between TFs and the mediator complex are important to connect enhancers to the promoter of genes[12]. Here, the mediator complex integrates the signals from TFs to control PolII activity and recruitment to enhance or repress the transcription of associated genes. Together, TFs can employ a variety of mechanisms to influence the initiation of transcription to regulate the expression level of genes.

1.1.2 Elongation

After promoter escape, the next step in the transcription cycle is the transition to elongation. The transition to elongation starts with phosphorylation of the CTD of PolII[13]. NELF and DSIF inhibit downstream elongation and facilitates pausing of transcription after approximately 50 bp[14] to allow capping of the transcript, which is essential for transcript stability. Next, P-TEFb binding to the paused PolII results in the release of NELF and the phosphorylation of DSIF, which is thought to be important for efficient elongation (for a review see[15]). In addition, P-TEFb phosphorylates serine 2 of the PolII CTD to allow the binding of other factors that are important for efficient elongation and for RNA processing.

In addition to transcriptional initiation, elongation can be regulated by transcription factors as well. An example is given by GR, a hormone activated transcription factor, that can inhibit recruitment of P-TEFb causing the transcription to stay paused[16]. A different example is given by NfkB, a transcription factor that can bind to pTEFb to activate elongation[17].

1.1.3 Termination

Signals to initiate termination are encoded within the sequence of the RNA transcript itself (AAUAAA). Once this sequence is transcribed, it can be bound by CPSF a PolII-

associated protein. Upon binding, CPSF catalyzes the cleavage of the RNA downstream of the binding site and subsequently poly-A polymerase adds approximately 200 adenosines to the cleaved 3' end (for a review see[18]). Poly-A binding proteins assemble on the Poly-A tail to allow transcript transport to the nucleus and enhance transcript stability.

1.2 Regulation of transcription

In multicellular organisms, all cells essentially contain the same genomic information. Despite identical genomic information, cells that form these organisms vary in their form and function. To facilitate this diversity, cells have to use sub-sets of the genomic information present. Therefore, mechanisms that control which set of genes is used in each cell-type have to exist. TFs play a central role in this process by binding to genomically encoded regulatory sequences and subsequently influencing various steps in the transcription cycle to regulate the expression level of associated genes. However, because the different cells of the body essentially have the same genome, the presence of these genomically encoded binding sites alone is not sufficient to explain cell-type-specific gene expression. Another factor to take into account is whether such binding sites are actually accessible for a TF to regulate the expression of associated genes[19–21]. Accessible sites are typically located in so-called open chromatin (euchromatin) whereas inaccessible regions are located in closed chromatin (heterochromatin).

Hence, to understand how the different cell-types are generated from identical genomic information, it is necessary to understand the mechanisms that regulate accessibility of TF binding sites. Differences in accessibility of binding sites are generated during differentiation of cells from common progenitor cells. At transition points where different cell types are formed from a common progenitor cell, daughter cells start expressing diverse sets of TFs. Some of these TFs (so called pioneering factors) have the potential to locate binding sites in inaccessible chromatin and to open up these regions to allow binding of other TFs and the regulation of associated genes (for a review see[22]). This difference in the set of active TFs as well as the differences in accessibility of TF-target regions can explain how cells with identical genomes can have differential gene expression patterns. However, even for a given cell type the accessibility of binding sites is not static. To be able to respond to external stimuli and adapt to new requirements, cells are able to change the expression level of genes, for example by modifying the accessibility of TF binding sites. Such stimuli can be the presence of nutrients, changes in temperature or exposure to hormones. An example for such a cell-type specific response to external stimuli is the

response to glucocorticoids (GCs). GCs are hormones that are released into the bloodstream in response to various stresses, for example when glucose levels in the blood are low. In cells, GCs activate GR, an otherwise inactive TF (for a review see[23]). When the released GCs reach the liver, they enhance gluconeogenesis by up-regulation of corresponding genes to ensure glucose homeostasis[24]. Another example of stress resulting in the release of GCs is when a bacterial infection is present. In this case, the released GCs play a role in moderating the immunological responses of the host by decreasing the number[25] and activity[26] of lymphocytes thus resulting in immunosuppressive[27], anti-inflammatory effects. Longer exposure to GCs, for example when they are used therapeutically in various auto-immune diseases, also leads to various side effects such as osteoporosis and obesity. All these effects are caused, at least in part, by cell-type-specific changes in the expression level of genes induced by the same transcription factor, showing how cells from different tissues can show vastly different reactions to the same stimulus.

1.3 Chromatin influences accessibility of binding sites

As described above, transcription is controlled by TFs that bind to genomically encoded DNA binding sites associated with target genes. However, the presence of binding sites alone is not sufficient to explain where TFs bind in the genome. Another important factor that influences if TFs can actually bind and regulate associated genes is the accessibility of the binding sites. The DNA of eukaryotes is present as a DNA-protein complex called chromatin whose repeating unit is the nucleosome. Nucleosomes are found approximately every 200 bp throughout the genome and consist of an octamer of four core histones with 147bp of DNA wrapped around them forming a structure like “beads on a string” on a 10 nm fiber[28]. The presence of nucleosomes influences all DNA-dependent processes including DNA repair, transcription and TF binding. In addition, the 10 nm fiber of DNA can be further compacted to a solenoid structure to form a 30 nm fiber of DNA[29,30]. However, further compaction, which requires additional proteins like heterochromatin protein 1, leads to inaccessibility of binding sites encoded in these regions. Hence, these proteins can be used as markers for genomic regions of heterochromatin[31,32]. Such regions are not accessible for typical TF binding, although certain TFs (so-called pioneering factors) through yet unknown mechanisms appear to retain their ability to bind to TF binding sites in these regions (for a review see[22]).

Taken together, chromatin compaction has a major impact on TF binding to DNA and thereby on which genes are expressed, with active genes typically associated with open and

inactive genes with closed chromatin[33]. In addition, other chromatin features influence the availability of TF binding sites. These include nucleosome positioning, histone modifications[34], DNA methylation[35,36], subnuclear organization[37] and localization[38] (For a review see[39]).

1.3.1 Nucleosome positioning

TFs selectively recognize their DNA binding sites by making base-specific contacts in the minor or major groove. When DNA is wrapped around nucleosomes, the encoded binding sites can either face to the interior or the exterior of the nucleosome, depending on the exact positioning of the nucleosome. As a consequence, DNA binding sites may not be available for TF binding and accordingly, on average, accessibility of DNA wrapped around nucleosomes is decreased[40] and the position of nucleosomes is correlated to lower DNA accessibility[41].

1.3.2 Histone modifications and histone variants

The N-terminal tails of histone proteins protrude from the nucleosomes and can be subjected to several post-translational modifications (for a review see[42]). Such modifications are correlated with overall chromatin structure, with accessibility of enhancers and promoters for TF binding and with the structure at gene bodies. For example, histone acetylation of lysines reduces its positive charge and thereby reduces the strength of interaction with the negatively charged DNA thereby opening up the chromatin to generate a more accessible structure (for a review see[43]). Accordingly, histone acetylation correlates with active gene expression[44]. In contrast, H3K9me3 and H3K27me3 are marks of closed chromatin (for a review see [45]). H3K9me3 for example allows proteins associated with heterochromatin formation, like HP1, to bind[46]. Other modifications associated with genomic regions that serve different functions include H3K4me2, H3K4me3 and acetylation which localize to active promoters; H3K36me3 and H3K79me2 that map to transcribed regions and H3K4me1, H3K4me2 and H3K27ac that are marks of active enhancers. Although it is not always clear what the underlying mechanisms responsible for these correlations are (for example whether modifications found at enhancers are a cause or a consequence of TF binding), several combinations of histone modifications are highly correlated with TF binding thereby suggesting a tight connection between histone modifications and TF binding.

1.3.3 DNA methylation and CpG content

CpG methylation is another feature that can influence the binding of transcription factors in eukaryotic cells[47,48]. Methylation of genomic DNA sites can reduce genomic binding of TFs either directly by inhibition of TF-DNA interactions[49], or indirectly by recruiting histone deacetylases leading to heterochromatin formation[42,50–52]. Combined these mechanisms explain the negative correlation between CpG methylation and TF binding.

During evolution spontaneous deamination of methylated CpG di-nucleotides lead to C-to-T transitions[53]; resulting in decrease of CpG-dinucleotide frequency[47,54]. High CpG promoters of genes that are broadly expressed[55–58] are not methylated and therefore were not affected by C-to-T transition during evolution. Hence, promoters with high CpG content are usually accessible for TF binding. Likewise, pre-accessible enhancers are also rich in unmethylated CpGs[35]. These findings are supported by CpG enrichment present at enhancers bound by TFs like GR[35]; Oct4 and Stat2[35,59]. In contrast, promoters in chromatin that is opened up only after TF binding have a low CpG density[35]. Combined these results show that in addition to the DNA methylation also the relative number at CpG-dinucleotides can be used to predict the accessibility of encoded TF binding sites. Binding sites in high CpG promoters are usually more accessible, whereas the accessibility of low CpG promoters depends on the cell type.

1.3.4 Spatial genomic organization

Another important level of regulation is provided by the three-dimensional organization of the genome in the nucleus. It is becoming increasingly clear that chromosomes are organized within so-called chromosome territories[60] and that different nuclear processes like RNA splicing and transcription take place in distinct genomic regions.

Chromosome territories of gene-dense chromosomes tend to be located towards the center of the nucleus[61] with active genes being exposed to the surface of these territories reaching into the interchromatin compartment[62–64]. Within the interchromatin compartment, accumulation of RNA polymerase II foci was observed to form so-called transcription factories[65–67] and co-localization studies revealed that active genes co-localize with these transcription factories[68]. Other subnuclear structures include nucleoli, the nuclear lamina and paraspeckles. Genomic regions that interact with the nuclear lamina usually contain genes that are not actively transcribed[38,69–71] and such lamin-associated domains often accumulate in the nuclear periphery or near nucleoli[71] (see Figure 1). Another example for

the importance of spatial genomic organization is given by genomic interaction with nucleoli: active rDNA genes were shown to co-localize in nucleoli, whereas inactive rDNA genes are excluded from these structures[72]. Paraspeckles are additional subnuclear structures formed around the long non-coding RNA Neat1, whose exact role we are just beginning to understand. For example, recent findings indicate that they affect transcription by sequestration of transcriptional regulatory factors[73]. Together, these findings indicate that the subnuclear organization plays an important role in gene regulation. Although it is still largely unclear how this organization is achieved, it appears to be facilitated, at least in part, by interactions of chromosomes with subnuclear structures like the nuclear matrix[74,75], nuclear lamins[71,76,77] and nucleoli[78].

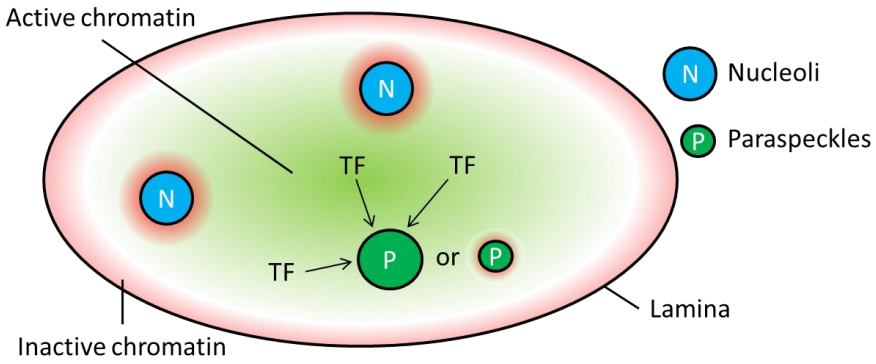


Figure 1: Correlation of nuclear organization and transcriptional activity.
The subnuclear localization is correlated with gene activity, with transcriptionally inactive chromatin (red shaded) often associated with nuclear lamina[69,70] or nucleoli[71]. In contrast, transcriptionally active chromatin (green shaded) tends to be located more to the center of the nucleus[70]. Potentially, other subnuclear structures might also have a role in control of gene expression. For example, recent reports indicate that paraspeckles affect gene expression via sponge-like mechanisms to sequestrate transcriptional repressors[73]. Another potential function for such structures is that genomic co-localization with these structures results in transcriptional repression via yet unknown mechanisms.

1.4 The nuclear receptor superfamily as a tool to study regulation of TF binding

Proper development and homeostasis requires communication between the different specialized cells in multicellular organisms. For communication, cells utilize different molecules, for example cytokines, water-soluble peptide hormones or small lipophilic steroid hormones. To initiate a response, cells have to detect these molecules. One way to detect and react to such molecules is via nuclear hormone receptors, a special class of transcription factors specific to metazoans that bind small lipophilic steroid or thyroid hormones that can enter the cells by diffusion[79]. Upon binding, their cognate receptor

becomes activated[80–82] thereby triggering a cellular response for example by changing the expression level of genes[81].

Steroid hormone receptors share a common structure, consisting of a ligand binding domain, a hinge region, a DNA binding domain and a N-terminal transcriptional activation domain (for a review see[83]). The ligand binding domain contains the activation function (AF-2 domain) which interacts with a broad range of co-regulators. The ligand binding domain is highly conserved in structure and consists of 12 α-helices and 2-3 β-strands[84,85]. Ligand binding induces conformational changes in helix 12[86], which enables the AF2 domain to interact with various co-factors[87]. Depending on which exact ligand is bound, helix 12 can adopt slightly different conformations[88] which can influence interactions with co-regulators thereby explaining how the same hormone receptor can induce ligand-specific transcriptional responses. In addition to its role in ligand binding, the ligand binding domain is important for dimerization of the nuclear receptors. DNA binding by steroid hormone receptors is mediated by the DNA binding domain, which is highly conserved and contains two zinc finger motifs. On the N-terminal end another interaction domain is located, called activation function 1 (AF-1). In contrast to AF-2, the activity of AF-1 does not depend on ligand binding. Instead, conformational changes in the AF-1 domain of the receptor[89], caused by binding to DNA-encoded binding sites induces interaction with co-regulators[90].

Binding of nuclear receptors to genomic binding sites can be facilitated by different mechanisms. Binding to the DNA can either be direct to genomically encoded binding sites that can be enhanced by co-regulators or by combinatorial binding together with other TFs that bind nearby. Alternatively, interactions with the DNA can be indirect via interactions with DNA-bound proteins that tether nuclear receptors to DNA[16,91–94]. This versatile repertoire of mechanisms controlling recruitment of nuclear receptors to DNA makes them a good model to study how recruitment to TFs to specific loci is controlled.

Another useful feature of hormone receptors is that their binding can be controlled by the presence of hormone. Using hormones as a switch allows analysis of the target locus pre and post receptor binding to study signals that initiate TF binding as well as the effects of binding on transcription.

1.5 The glucocorticoid receptor as model transcription factor

The GR is a member of the nuclear receptor superfamily and is ubiquitously expressed. In the absence of hormone, GR resides in the cytoplasm in a hormone-binding competent state which is stabilized by bound chaperones[95]. Ligand binding induces conformational changes[96] and translocation to the nucleus where it binds to specific genomic DNA sequences to regulate transcription[97,98]. Typically, GR binding is controlled by DNA accessibility[20]. At accessible sites, GR usually binds as a homodimer[99–101] to its canonical sequence motif [102] consisting of an of an inverted hexamer sequence with a three bp spacer. In addition to its primary function to bind GR, the exact sequence of the binding site also serves as an allosteric ligand adding an additional layer of information to control transcriptional outcome subsequent to TF binding[103]. Additionally, GR can bind to genomic DNA either in a combinatorial binding mechanism with other transcription factors[104,105], or indirectly via a tethering mechanism[93] (see Figure 2). Upon binding, GR interacts with a broad spectrum of co-regulators that collaborate with GR to either directly or indirectly influence the recruitment and or activity of RNA PolII to control the expression of associated target genes.

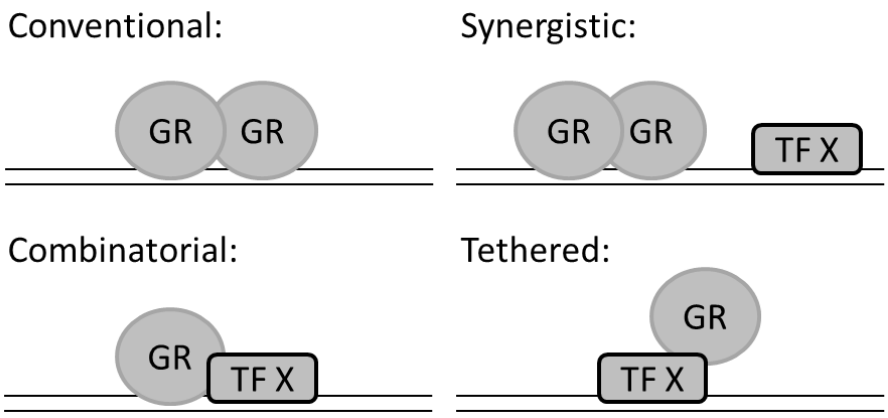


Figure 2: GR can bind to target sites in several binding modes.
Four potential GR binding modes to genomic DNA are shown. In the conventional mode; GR binds as homodimer to canonical GBSs[100,101]. The synergistic binding is a variation of the conventional binding, where genomic GR homodimer binding is enhanced by interactions with other TFs that bind the same locus[93]. In contrast to the conventional and synergistic binding modes, the presence of a canonical GBS is not essential for combinatorial and tethered genomic binding of GR binding. Combinatorial binding occurs when one half-site of the canonical GBS is combined with a binding sited for a co-regulator, thereby recruiting GR and the co-regulator to facilitate heterodimeric binding[104,105]. Tethered binding is initiated by other DNA bound TFs that recruit GR through protein-protein interactions[93].

In addition, previous studies focusing on GR activity within three cell lines (Nalm-6, U2-OS and A549) revealed that GR target genes vary depending on the cell line tested[106]. 1969 genes were found to be specifically regulated in U2-OS, 180 for Nalm-6 and 637 for

A549 cells. Only 25 genes are regulated in all three cell lines and 68 both in U2-OS and Nalm-6, 294 in U2-OS and A549 and 36 in Nalm-6 and A549 (Figure 3). This raises the question what the mechanisms are that allow GR to regulate such diverse sets of genes in different cell types.

The fact that hormone serves as an on/off switch for GR greatly facilitates the identification of genomic binding sites and of target genes by comparing RNA expression levels before and after hormone treatment. In addition, using these hormones as a switch, allows analysis of the target loci pre and post receptor binding to study signals that initiate TF binding as well as effects of subsequent binding on transcription. Also, the potential to bind genomic DNA in various modes (see Figure 2) allows using GR as a model to study how TFs are recruited to the genome. In addition, GR is a great tool for studying how activity of bound TFs is regulated. Even when bound to genomic DNA, GR activity depends largely on interaction with co-regulators, potentially amongst other mechanisms controlled by allosteric interaction with the DNA. Hence, it can be utilized to investigate how such interactions are facilitated, and how site specific interactions are controlled. Furthermore, the large difference in target genes for GR within the tested cell lines indicates that in addition to general mechanisms of GR recruitment, GR can also be used to study cell-type-specific regulation of TF binding.

Taken together, GR is a useful tool to study mechanisms that control TF binding in general, and several aspects of GR biology make it a great model to study various aspects of transcription.

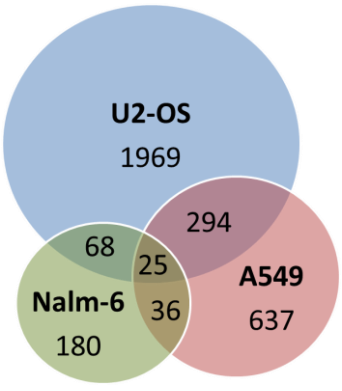


Figure 3: GR-regulated genes show little overlap between cell types (figure modified from[106]).
Depicted is the overlap of GR-regulated genes in Nalm-6, A549 and U2-OS cells as determined using microarrays (data from Samantha Cooper[106]**)**. 1969 genes are specifically regulated in U2-OS, 180 in Nalm-6 and 637 in A549. 294 genes are commonly regulated in U2-OS and A549, 68 genes in U2-OS and Nalm-6 and 36 genes commonly in Nalm-6 and A549. Only 25 genes are commonly regulated between the three cell lines.

1.6 DNA-encoded signals that specify GR binding

Typically, GR binds to GR binding sequences (GBSs) that match a loosely defined binding motif. Such sequences however are ubiquitously found in the genome and only approximately one out of 1000 potential binding sites is actually bound by GR. This indicates that the presence of a GBS alone is insufficient to specify where GR binds and that additional signals are needed to define its genomic binding pattern. As described before, DNA accessibility plays an important function in specifying which putative binding sites are in fact bound. Notably, most of the binding takes place at open chromatin as defined by DNase-I sensitivity[20]. However, this raises the question how the different cell types facilitate usage of different genomic binding sites. Potential mechanisms are provided by cell type specific transcription factors that open up cell-type-specific genomic regions for TF binding as pioneering factors[107,108]. Alternatively, the tissue-specific expression of TFs that tether GR to the DNA could serve as a mechanism to recruit GR to cell-type specific target sites and consequently result in tissue specific gene expression.

1.7 Background of my research: Analysis of ChIP-Seq data

A common approach to identify DNA sequences that recruit TFs to the genome is to search for sequences that are enriched at genomic regions of TFs binding. When applied to GR, an example of such an enriched motif is the canonical GR binding site. In addition, sequence motifs for other transcription factors are enriched that either open up chromatin to increase the accessibility of canonical GR binding sites or recruit GR to genomic locations by direct protein-protein interactions. Hence, enriched motifs serve as candidates to explain the genomic binding profile of TFs. Consequently, comparing such enriched motifs between cell types might yield clues about the mechanisms responsible for cell-type specific binding and transcriptional regulation by GR[109].

Genomic GR target regions in different tissues (pre-B (Nalm-6), lung epithelial (A549) and osteosarcoma (U2-OS) cells) were identified from ChIP-Seq data generated by Samantha Cooper[106]. Using this data, we (in cooperation with Dr. Morgane Thomas-Chollier) set out to identify selectively enriched sequence motifs at GR binding sites in these three cell types. Consistent with findings by Samantha Cooper[106], our analysis revealed that GR binding is mostly cell-type specific with little overlap in bound regions between cells. To compare enrichment of sequence motifs at ChIP-seq peaks between cell types, we scanned a window of 8 kb centered on the peak summit for the frequency of sequences that match the consensus motif for vertebrate TF binding sites from the Transfac and Jaspar databases

(for a more detailed description of the analysis performed by Dr. Morgane Thomas-Chollier see section 3.1)

As expected, this analysis shows that the canonical GR binding site sequence is enriched in all three cell lines (Figure 4 A) although the degree of enrichment appears to vary substantially between cell types with only a moderate enrichment for the Nalm-6 cell type. In addition to the GBS, other sequence motifs were enriched at bound loci in a cell-type specific manner. For example, in Nalm-6 sequence motifs resembling PU box or E-Box motifs and a motif similar to a RunX1 binding site show a much more pronounced enrichment in Nalm-6 cells than in the other two cell types examined.

Unexpectedly, we also identified sequence motifs that are depleted at GR bound sites bound across all cell-lines tested. Common to almost all of these depleted sequence motifs is that they have a high AT content and match sequences like TTTGTTT, TTAATTAA, TTAATTGAATTAA.

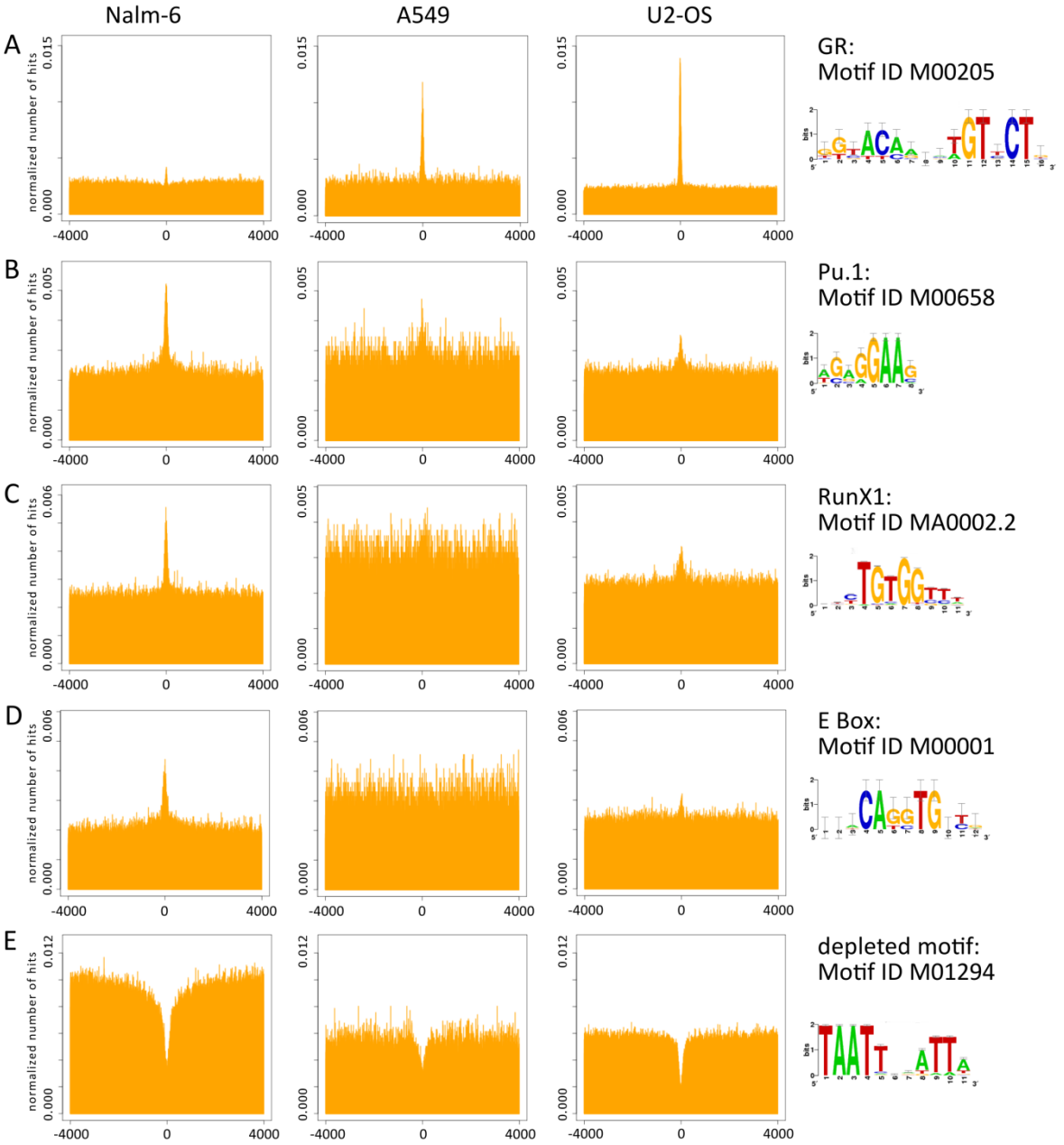


Figure 4: Cell-type specific distribution of DNA sequence motifs around GR-bound loci.
DNA sequences from ChIP-Seq peaks were aligned at the peak summit and flanking genomic DNA +/- 4000 bp was sub-divided into 50 bp bins. For each bin, the relative frequency distribution of sequence motifs was determined by scanning for alignment to these motifs (represented by position weight matrices (PWMs) from Jaspar as shown on the right). The normalized number of hits for each PWM is depicted for each bin. The analysis was done for GR ChIP-Seq data for Nalm-6 cells (left diagrams), A549 cells (middle plots) and U2-OS cells (right plots). Shown is the data for PWMs for GR (**A**), PU.1 (**B**), Runx1 (**C**), an E Box motif (**D**) and an example for a sequence motif that is depleted at GR-bound loci according to this analysis (**E**).

1.8 Nalm-6 as a model to study cell type specific signals

The enrichment of the canonical GBS in Nalm-6 is less pronounced than in U2-OS or A549. This indicates that in Nalm-6 alternative sequences are important to specify where GR binds in the genome. Good candidates for such sequences are those motifs that are

selectively enriched in Nalm-6 cells. For example, RunX1 is an ideal candidate to guide GR binding in Nalm-6 cells, because in addition to it being highly expressed in Nalm-6 [110](it is a transcription factor important for the development of cells in the hematopoietic lineage from which Nalm-6 cells are derived [111,112]) its binding site is enriched at GR bound loci in these cells. In addition to RunX1, the enrichment of the Pu-box motif indicates that members of the ETS family like PU.1[113,114] also participate in guiding GR to its genomic target sites in Nalm-6. Like RunX1, ETS transcription factors are crucial for the development of hematopoietic cells[112,115]. As a member of the ETS family, PU.1 is expressed in Nalm-6[116] and is important for development of the B-cell lineage[115,117]. Hence, PU.1 is a promising candidate to facilitate GR binding via the PU box. In addition, PU.1 has been described to control positioning of the transcriptional start site (TSS) at the TLR4 promoter in myeloid cells[118]. In combination with the observed strong co-localization of GR binding sites at TSSs specific for Nalm-6 (see Figure 5) this might indicate a general role for PU.1 in specifying GR binding sites and TSS positions.

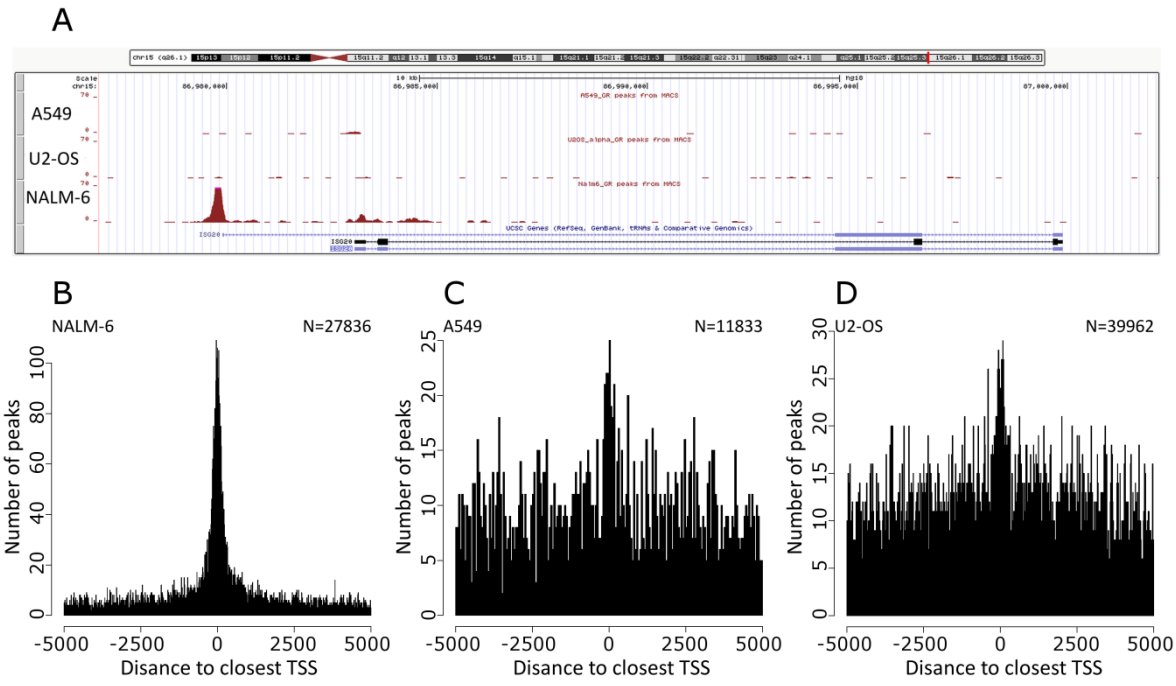


Figure 5: Genomic binding by GR is cell-type specific (Analysis by Morgane Thomas-Chollier).
(**A**) Genomic GR binding for the three cell lines tested is visualized for a genomic region (chr15:86,977,002-87,002,161, UCSC Genome Browser on Human Mar. 2006 (NCBI36/hg18) Assembly) encoding a GR regulated gene in Nalm-6, but not in U2-OS or A549 cells. Panels (**B**)-(**D**) show the specific distribution of GR binding around transcriptional start sites (TSSs) for the individual cell types as indicated above the diagram. Only binding sites located within a 5 kb window of a TSS were included in the analysis. Corresponding regions were centered on the closest TSS (position 0 on the X-axis), and then for each locus the distance of the peak center to the TSS was plotted.

1.9 Aim of this thesis

I was interested in how several mechanisms are integrated to specify which genes are regulated in a particular cell type. An essential step in the regulation of genes is the binding of TFs to regulatory sequences associated with their target genes. Thus, understanding mechanisms that specify where TFs bind in the genome is a key to understanding which genes are regulated by the TF of interest. In my thesis, I focused on the role of DNA-encoded signals (sequences) and their role in guiding TFs to the appropriate genomic locations. In theory, such signals can be divided in two groups: those that facilitate binding to certain genomic regions and those that restrict binding to other loci (see Figure 6).

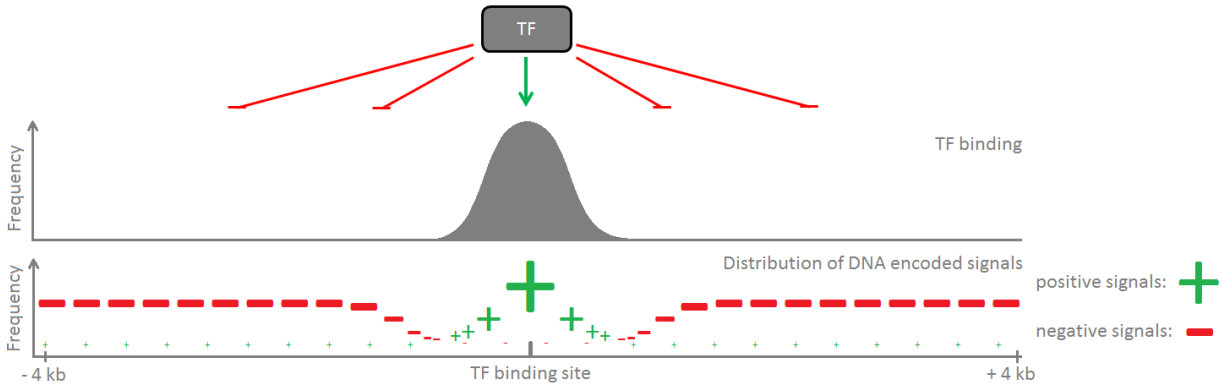


Figure 6: Illustration of potential distribution of DNA-encoded signals that regulate TF binding. Shown is a scheme of a genomic TF binding site as a ChIP-Seq peak (in grey). Likely, DNA-encoded positive signals for TF binding (e. g. the canonical transcription factor binding site or binding sites for cofactors) are enriched near the binding site (peak-summit). In contrast, signals that restrict TF binding are likely to be absent at these binding sites and instead present in non-target regions to restrict unspecific TF binding.

The first part of my thesis focuses on signals that facilitate genomic binding by GR. By investigating sequence motifs that are selectively enriched at GR bound loci in Nalm-6 cells, the objective was to understand their role in facilitating Nalm-6 specific GR binding, for example in the recruitment of GR to its target regions. This might be by direct protein-protein interactions between factors binding to such sequences and GR resulting in direct DNA-bound heterodimeric complexes. Alternatively, proteins bound to such sequences might recruit GR via tethering mechanisms. Another possibility is that cell-type dependent GR binding can be explained by differences in DNA accessibility caused by cell-type specific transcription factors with pioneering activity to open up heterochromatin and thereby facilitate GR binding.

The second part of my thesis focuses on signals that restrict genomic binding by GR. The recognition sequence of GR is only loosely defined with millions of potential binding sites

encoded in the genome of which only a cell-type specific minority is actually bound. However, apart from DNA accessibility and methylation of binding sites little is known about mechanisms that restrict binding to only a subset of the genomically encoded TF binding sites. Preventing unrestricted binding to all potential sites in the genome is important as a failure to do so would likely result in insufficient binding to real target sites, because the abundance of individual TFs in the cell is very low[119]. Since DNA-encoded signals that facilitate TF binding are known, I hypothesized that there may also be DNA sequences that restrict binding. If such restricting signals exist, they should be mutually exclusive with GR binding and therefore depleted in genomic regions where GR binds. Taken together, the depleted sequence motifs from the bioinformatical ChIP-Seq analysis are ideal candidate sequence motifs (CSMs) to serve as signals to restrict genomic binding of TFs. The second part of my thesis focuses on the functional analysis of such sequences.

2 Materials

2.1 Chemicals

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

Sybr Green	Invitrogen
Tris	Carl Roth
Triton X 100	Sigma Aldrich
Tween 20	Fisher Scientific
Temed	Carl Roth
Zinc Finger Nuclease (AAVS1 locus)	Sangamo
Beta-Glycerophosphate	Sigma Aldrich

2.2 Enzymes, Proteins, DNA Kits, Plasmids

AscI	Fermentas
Asp718	Roche
BglII	Thermo Fischer
Dnase I	Qiagen
Lipofectamine	Invitrogen
Lipofectamine 2000	Invitrogen
Nucleofector II	Amaxa
Nucleofector kit T	Amaxa
Nucleofector kit V	Amaxa
Oligos	Sigma Aldrich
pcDNA3.1	Invitrogen
pGL3-Promoter	Promega
pRL (CMV)	Promega
PstI	New England Biolabs
RNase	Applichem
SAA-GFP (AAVS1 targeting vector)	Sangamo
XmaI	Thermo Fischer
Dual luciferase reporter assay system	Promega
NotI	Thermo Fischer
Zero Blunt cloning	Invitrogen
Z-Competent E.coli transformation Kit & Buffer Set	ZYMO RESEARCH
PCR purification Kit	Qiagen
Nucleospin Gel and PCR clean-up	Macherey and Nagel
mMessage mMachine Kit	Ambion
Cot1 DNA	Invitrogen
Pfu ultra polymerase	Agilent

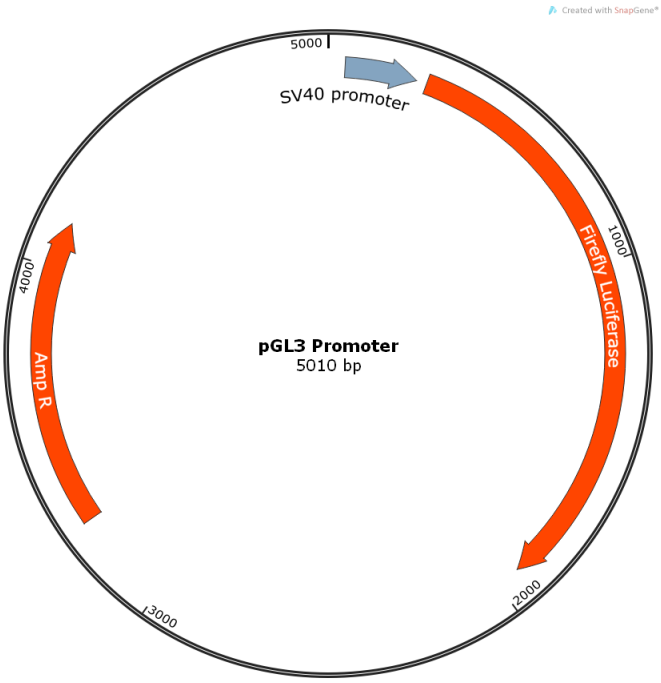
2.3 Oligonucleotides

JT117	CCGGAGAACAGGGTGTTCT
JT118	GATCAGAACACCCTGTTCT
JT123	CCGGAGAACAAAATGTTCT
JT124	GATCAGAACATTTTGTTCT
JT119	CCGGAGAACATTTTGTACG
JT120	GATCCGTACAAAATGTTCT
JT115	CCGGAGAACATTTTGTCCG
JT116	GATCCGGACAAAATGTTCT
JT125	GTACGAGGTTTGTTTG
JT126	CTAGCAAACAAACCTC
JT127	GTACGAGGTTAATTAA
JT128	CTAGTTAATTAACCTC
JT129	GTACGTTAATTCAATTAA
JT130	CTAGTTAATTGAATTAAC
JT163	CCAGGTCTCAGTACCGTGCCAGAACATTTCTCTATCGATA
JT164	CCAGGTCTCATCGACGGATCCTTATCGATTTTACC
JT177	GTACGAGAGGTTTGTTTGGCTGA
JT178	CTAGTCAGCCAAACAAACCTCTC
JT179	GTACGAGAGGTTAATTAAGCTGA
JT180	CTAGTCAGCTTAATTAACCTCTC
JT181	GTACGTTAATTCAATTAAGCTGA
JT182	CTAGTCAGCTTAATTGAATTAAC
JT183	GTACGCGAGGTAGGCTTGGCTGA
JT184	CTAGTCAGCCAAGCCTACCTCGC
JT201	GCAGATCGCAGATCAGAACA
JT202	TATGGTACCGTGCCAGAACA
JT207	CGCGCAAGCCTACCTCGGCCAAGCCTACCTCG
JT208	CGCGCGAGGTAGGCTTGGCCGAGGTAGGCTTG
JT209	CGCGCAAACAAACCTCGGCCAAACAAACCTCG
JT210	CGCGCGAGGTTTGTTTGGCCGAGGTTTGTTTG
JT211	CGCGTTAATTAACCTCGGCTTAATTAACCTCG
JT212	CGCGCGAGGTTAATTAAGCCGAGGTTAATTAA
JT213	CGCGTTAATTGAATTAAGCTTAATTGAATTAA
JT214	CGCGTTAATTCAATTAAGCTTAATTCAATTAA
JT247	GTACGAGGTATATATACG
JT248	CTAGCGTATATATACCTC
JT251	GTACGAGGAAAAAAAACG
JT252	CTAGCGTTTTTTTTCCTC
JT253	GTACGAGGTTTTAAAACG
JT254	CTAGCGTTTTAAAACCTC
JT255	GTACGAGGAAAATTTTCG
JT256	CTAGCGAAAATTTTCCTC
JT257	GTACGAGGTTTTTTTGCG
JT258	CTAGCGCAAAAAAACCTC
JT259	GTACGAGGATTTTTAGCG
JT260	CTAGCGCTAAAAATCCTC
JT261	GTACGAGGTTTTTAAGCG

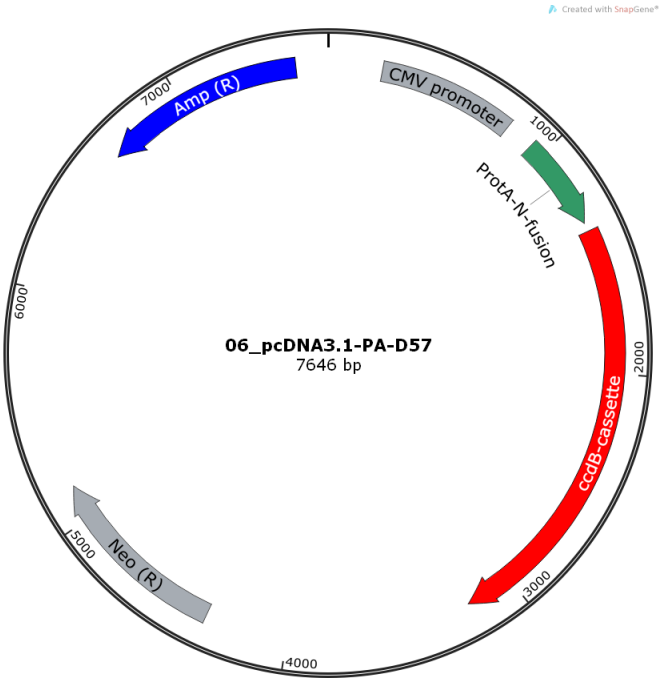
JT262	CTAGCGCTTAAAAACCTC
JT263	GTACGAGGAAAAATTGCG
JT264	CTAGCGCAATTTTTCCTC
JT-269	[BtnTg]CAAAAGATCGCTGCAGACTTGAACCGAGGTAGGCTTGCTAGCC-CGGAGAAACaaaGTTTCTACTTTGTC
JT-270	GACAAAGTAGAAACtttGTTTCTCCGGGCTAGCAAGCCTACCTCGGTTCA-AGTCTGCAGCGATCTTTTG
JT-271	[BtnTg]CAAAAGATCGCTGCAGACTTGAACAGGTTTGTTTGCGCTAGCCC-GGAGAACAaaaTGTTCTACTTTGTC
JT-272	GACAAAGTAGAACAtttTGTTCTCCGGGCTAGCGCAAACAAACCTGTTCA-AGTCTGCAGCGATCTTTTG
JT-273	[BtnTg]CAAAAGATCGCTGCAGACTTGAACAGGTTAATTAACACTAGCCC-GGAGAACAaaaTGTTCTACTTTGTC
JT-274	GACAAAGTAGAACAtttTGTTCTCCGGGCTAGTGTTAATTAACCTGTTCAA-GTCTGCAGCGATCTTTTG
JT-275	[BtnTg]CAAAAGATCGCTGCAGACTTGAACTTAATTCAATTAACTAGCCCG-GAGAACAaaaTGTTCTACTTTGTC
JT-276	GACAAAGTAGAACAtttTGTTCTCCGGGCTAGTTAATTGAATTAAGTTCAAG-TCTGCAGCGATCTTTTG
JT-277	[BtnTg]CAAAAGATCGCTGCAGACTTGAACCGAGGTAGGCTTGCTAGCCCGG
JT-278	CCGGGCTAGCAAGCCTACCTCGGTTCAAGTCTGCAGCGATCTTTTG
JT-279	[BtnTg]CAAAAGATCGCTGCAGACTTGAACAGGTTTGTTTGCGCTAGCCCGG
JT-280	CCGGGCTAGCGCAAACAAACCTGTTCAAGTCTGCAGCGATCTTTTG
JT-281	[BtnTg]CAAAAGATCGCTGCAGACTTGAACAGGTTAATTAACACTAGCCCGG
JT-282	CCGGGCTAGTGTTAATTAACCTGTTCAAGTCTGCAGCGATCTTTTG
JT-283	[BtnTg]CAAAAGATCGCTGCAGACTTGAACTTAATTCAATTAACTAGCCCGG
JT-284	CCGGGCTAGTTAATTGAATTAAGTTCAAGTCTGCAGCGATCTTTTG
JT361	GAATTCATGAGTGAGTCGAGCTCGAA
JT362	AGATCTTCACTGCTCCTCCTCCGAG
JT363	AGATCTTCACCACAATATGTACTCAGAT
efIa-paper_qfor	CTTCTCAGGCTGACTGTGC
efIa-paper_qrev	CCGCTAGCATTACCCTCC
fkbp5_qfor	CAAAAGGGGGAATGCTGTT
fkbp5_qrev	TTCTTTTCTGCCCTCTTTGC
ECFP qfor	ACGTAAACGGCCACAAGTTC
ECFP qrev	GCAGATGAACTTCAGGGTCAG
TagRFP qfor	GCTGGGAGGCCAACACCGAG
TagRFP qrev	CAGGGCCATGTCGCTTCTGC
R5	CTGGGATACCCCGAAGAGTG
LucNested	TCAAAGAGGCGAACTGTGTG
FKBP5 GBS6.1 fw	GCATGGTTTAGGGGTTCTTG
FKBP5 GBS6.1 rv	TAACCACATCAAGCGAGCTG
hRPL19 fw	ATGTATCACAGCCTGTACCTG
hRPL19 rv	TTCTTGGTCTCTTCCTCCTTG

2.4 Plasmids

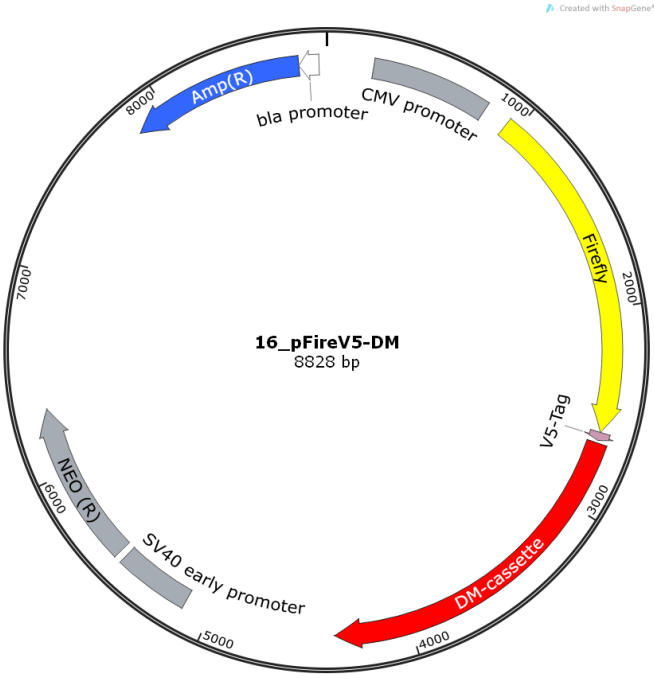
2.4.1 pGL3-Promoter



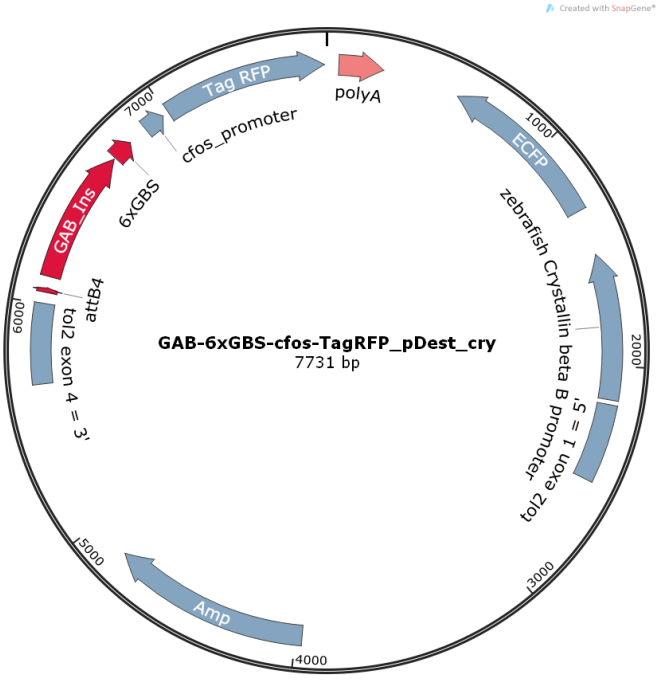
2.5 pcDNA3.1PA-D57 (kind gift by Dr. Ulrich Stelzl)



2.5.1 pFireV5-DM (kind gift by Dr. Ulrich Stelzl)



GAB-6xGBS-cFOS-TagRFP_pDest_cry



2.6 Antibodies

2.6.1 Antibodies for Lumier assay

Sheep gamma globulin	Jackson ImmunoResearch (013-000-002)
AffiniPure rabbit anti-sheep IgG Rabbit	Jackson ImmunoResearch (313-005-003)

2.6.2 Antibody for ChIP

N499 anti GR antibody, raised against a polypeptide [120] with the N-terminal amino acid sequence of the human GR (residues 1-499) (R.M. Nissen, B. Darimont, and K.R. Yamamoto, unpublished)

2.6.3 Antibodies for western blot

N499 anti GR antibody	as described above
Anti actin antibody I-19-R	Santa Cruz Biotechnology
HRP conjugate anti rabbit antibody	Invitrogen

2.6.4 Antibodies for Immunofluorescence

Primary Antibodies	***Supplier (product number)***
Rabbit anti-PSP1 antibody	homemade by University of Dundee
Rabbit anti-Lamin B1 antibody	Abcam (#ab16048)
Mouse anti-Lamin A/C antibody	Santa Cruz (#SC56140)
Rabbit anti-Coilin antibody	Proteintech (#10967-1-AP)
Mouse Anti-fibrillarin antibody	Cytoskeleton clone (#72B9)
Rabbit Anti-H4K9me^3 antibody	Abcam (#ab8898)
Rabbit Anti-SATB1 antibody	Abcam (#ab49061)

Secondary antibodies	***Supplier***
donkey anti-mouse or rabbit Alexa 488	Molecular probes
donkey anti-mouse or rabbit Alexa 594	Molecular probes

2.7 Lab Ware

Microcentrifuge	Eppendorf
Nanodrop	PeqLab
semi dry western blotter	Bio Rad
Power supply EV243	Consort
Agarose-gelchamber	homemade
Thermocycler Mastercycler gradient	Eppendorf
Thermomixer	Eppendorf
tissue culture flasks	TPP
48 Well multititer plates	TPP
12 well multititer plate	TPP
qPCR plates	Sarstedt

plastic cover for qPCR plates	Sarstedt
nitrocellulose membrane	Bio Rad
1 kb DNA ladder	Fermentas
protein ladder V	PEQLAB

2.8 Organisms

2.8.1 Bacterial strains

DH5α: *fhuA2 lac(del)U169 phoA glnV44 Φ80' lacZ(del)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*

2.8.2 Mammalian cell lines

Nalm-6: Precursor B-cells from peripheral blood of a male patient with acute lymphoblastic leukemia in relapse

U2-OS: Bone cells from a 15 year old female patient with osteosarcoma

GR-18: U2-OS cells stably expressing GR [121]

T-REx 293: Derived from human cell line 293 from embryonic kidney; modified to stably express the Tet repressor from the pcDNA6/TR plasmid (Invitrogen)

2.9 Media

2.9.1 LB Medium

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

2.9.2 SOB Medium

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

2.10 Buffers

2.11 General buffers

2.11.1 TE-Buffer

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

2.11.2 PBS buffer

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

2.12 Lumier Buffers

2.12.1 Lumier lysis buffer

50 mM HEPES; 150 mM NaCl; 1 mM EDTA; 10 % Glycerin; 1 % Triton X-100

2.12.2 TBST II

10 mM Tris-Base; 150 mM NaCl; 0.05 % Tween-20

2.12.3 Carbonate buffer

70 mM $NaHCO_3$; 30 mM Na_2CO_3

2.13 ChIP buffers

2.13.1 IP Lysis Buffer

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

2.13.2 RIPA buffer

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

2.13.3 RIPA wash buffer

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

2.13.4 LiCl wash buffer

20 mM Tris-HCl, pH 8.0; 1 mM EDTA; 250 mM LiCl; 0.5 % NP40; 0,5 % sodium deoxycholate

2.13.5 Crosslink reversal solution

10 mM Tris-HCl, pH 8.0; 1 mM EDTA; 0.7 % SDS; 0.22 mg/ml proteinase K

2.14 DNase Buffers

2.14.1 DNase I reaction buffer

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

2.15 MNase assay buffers

2.15.1 Lysis buffer

10 mM Tris pH 7.4; 10 mM NaCl; 3 mM $MgCl_2$; 0.5 % IGEPAL

2.15.2 Storage buffer

50 mM Tris pH 7.4; 40 % glycerol; 5 mM $MgCl_2$; 0.1 mM EDTA

2.15.3 Reaction buffer

50 mM Tris pH 7.4; 25 mM KCl; 2.5 mM $CaCl_2$; 5 mM $MgCl_2$; 12.5 % glycerol

2.15.4 Stopping buffer

50 mM Tris pH 7.4; 200 mM NaCl; 100 mM EDTA; 2 % SDS

2.16 Nuclear extract

2.16.1 PBSI buffer

0,5 mM PMSF; 25 mM beta-glycerophosphate; 10 mM NaF

2.16.2 Buffer A

10 mM HEPES, 1,5 mM $MgCl_2$; 10 mM KCl; 300 mM Sucrose; 0,5 % IGEPAL; 0,5 mM PMSF*; 1 mM Na_3VO_4*; 0,5 mM DTT*; 1 ‰ (v/v) Proteinase Inhibitor cocktail set III EDTA free *; 25 mM beta-glycerophosphate*; 10 mM NaF*

2.16.3 Buffer B

20 mM HEPES; 1,5 mM $MgCl_2$; 420 mM NaCl; 0,2 mM EDTA; 2,5 % Glycerol; 0,5 mM PMSF*; 1 mM Na_3VO_4*; 0,5 mM DTT*; 1 ‰ (v/v) Proteinase Inhibitor cocktail set III EDTA free*; 25 mM beta-glycerophosphate*; 10 mM NaF*

2.16.4 Buffer D

20 mM HEPES; 100 mM KCl; 0,2 mM EDTA; 8 % Glycerol; 0,5 mM PMSF*; 1 mM Na_3VO_4*; 0,5 mM DTT*; 1 ‰ (v/v) proteinase inhibitor cocktail set III EDTA free *; 25 mM beta-glycerophosphate*; 10 mM NaF*

*Add fresh directly before use

2.17 DNA pull-down buffers

2.17.1 DW buffer

20 mM Tris-HCl pH 8.0; 2 M NaCl; 0.5 mM EDTA; 0.03 % NP-40

2.17.2 Blocking buffer

20 mM HEPES, pH 7.9; 0.05 mg/ml glycogen; 0.3 M KCl; 0.02 % NP40; 0,05 mg/ml BSA; 5 mg/ml PVP

2.17.3 Buffer G

20 mM Tris-HCl pH 7.3; 10 % Glycerol; 0.1 M KCl; 0.2 mM EDTA; 10 mM K glutamate; 0,04 % NP40

2.17.4 RE buffer

100 mM NaCl; 50 mM Tris HCl; 10mM $MgCl_2$; 2 mM DTT; 2.5 % glycerol; 0.2 mM PMSF; 1 ‰ proteinase inhibitors set III EDTA free; 0.02 % NP40

2.17.5 Annealing Buffer

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

2.18 EMSA buffers

2.18.1 1x Binding buffer

20 mM Tris pH 7.5; 2 mM $MgCL_2$ 40 µl; 1 mM EDTA; 10 % Glycerol 2500 µl; 0.3 mg/ml BSA; 4 mM DTT

2.19 Immunofluorescence and FISH buffers

2.19.1 PHEM buffer

120 mM PIPES, 55 mM HEPES, 20 mM EGTA, 8 mM $MgSO_4$, pH 7

2.19.2 CSK buffer

10 mM PIPES pH 8.0; 100 mM NaCl; 300 mM Sucrose; 3 mM $MgCl_2$

2.19.3 20x SSC buffer

3 M NaCl; 0.3 M sodium citrate pH 7.0

3 Methods

3.1 ChIP-Seq analysis (as performed by Dr. Morgane Thomas-Chollier)

For each cell line, identification of genomic GR binding sites was done using "Model-based analysis of ChIP-Seq" (MACS)[122]. DNA sequence information of the identified genomic binding sites was centered on the peak summit in 8 kb windows. These regions were then subdivided into fixed bins of 50 bp. Each of these bins was scanned for the frequency of sequences that match known vertebrate TF binding sites, represented by position weight matrices (PWM) from the databases Transfac (transfac_vert_2010.1) and Jaspar (jaspar_core_vert_oct2009). The frequency of hits for each bin of all loci was then calculated as normalized number of hits. Subsequently, this frequency was plotted for each bin relative to the position of the peak summit.

3.2 Preparation and transformation of chemically competent DH5α cells

Chemically competent DH5α cells were prepared using the Z-Competent™ *E. coli* transformation Kit & Buffer Set (ZYMO RESEARCH) according to the manufacturer's protocol. For transformations, 50 µl of competent bacteria were incubated with up to 5 µl of DNA solution with a maximum of 1 ng plasmid DNA. For ligations, 5 µl of DNA ligation mix was added. Subsequently, the mixture was incubated for 5 minutes on ice. When ampicillin was used for selection, the cells were plated directly onto agar plates containing 100 µg/ml ampicillin. If kanamycin was used for selection, 250 µl of prewarmed SOB medium was added to the transformed cells and these were then incubated for one hour at 37°C prior to seeding onto an agar plate containing 20 µg/ml kanamycin.

3.3 Cloning procedures

3.3.1 Cloning of pGL3-promoter-enhancer reporters

To generate pGL3-promoter reporter plasmids (Promega) for testing of candidate genomic GR binding sites, ~500 bp target regions were amplified by PCR using mammalian genomic DNA as input. PCR primers were designed to encode Eco31I recognition sites at their 5'end that when cut with Eco31I to generate 5' overhangs compatible for ligation with the destination vector cut with XhoI. Amplicons were initially integrated into the pCR-blunt plasmid using the zero blunt cloning kit (Invitrogen) for screening for correct clones according to the manufacturer's protocol. Amplicon containing plasmids were

restriction digested using Eco31I, purified by agarose-gel extraction, and subsequently ligated into XhoI linearized pGL3-Promoter vector.

3.3.2 Cloning of pGL3-promoter-GBS-CSM reporters

The pGL3-Promoter plasmid (Promega) was linearized with the restriction enzymes XmaI and BglII. Integration of glucocorticoid receptor binding site sequences (GBSs) was facilitated by re-circularizing the plasmid using oligonucleotides that encode candidate GBSs as linkers (oligopairs used were JT117/118; JT123/124; JT119/120 or JT115/116 respectively). To test the effect of sequences on GR mediated reporter activity, the integrated GBS sequences were flanked by CSMs that were integrated using the restriction enzymes Asp718 and NotI. The linearized plasmids were then re-circularized by ligation with linker oligonucleotides encoding the candidate sequences (oligopairs used were JT125/126; JT127/128; JT129/130; JT247/JT248; JT251/JT252; JT253/JT254; JT255/JT256; JT257/JT258; JT259/JT260; JT261/JT262; JT263/JT264; JT177/JT178; JT179/JT180; JT181/JT182 or JT183/JT184 respectively).

3.3.3 Cloning of GBS-NRS luciferase reporter plasmids for stable integration

Reporters used for stable integration were designed as described previously[123]. In short, CSM luciferase reporter genes from pGL3 promoter that showed reduced activity were amplified by PCR using the primers JT163 and JT164. Amplicons were then purified by agarose gel-extraction using the NucleoSpin Gel and PCR Clean-up kit (Macherey & Nagel). Amplicons were digested with Eco31I to generate overhangs compatible with ligation into SalI and Asp718 linearized SAA-GFP plasmid[123].

3.3.4 Cloning of expression vectors for lumier assays

Gateway entry clones in pDONR221 encoding TFAP4 (gene ID: 839) Runx1 (gene ID: 861) or PU.1 (gene ID: 6688) were kind gifts from Dr. Ulrich Stelzl (Max Planck Institute for Molecular Genetics; research group Molecular Interaction Networks, Berlin), as were destination vectors to generate fusion proteins with either protein A (pcDNA3.1PA-D57, as described in[124]) of firefly luciferase (pFireV5-DM, as described in[125]) at the N-terminal end. Integration of the coding sequences into these destination vectors was facilitated by gateway cloning using the Gateway® LR Clonase™ II enzyme mix (life technologies) according to manufacturer's protocol.

3.3.5 Cloning of reporter genes for testing in Danio rerio

Reporter plasmid GAB-6xGBS-cFOS-TagRFP_pDest_cry to test for GR activity in zebrafish was generated by Dr. Sergey Prykhozhij according to Tol2Kit[126]

recommendations by multisite-gateway cloning procedure (Invitrogen). Entry clones used were p3E-TagRFP, pME-cfos_promoter and p5E-GAB-6GBS; the destination vector used was pDEST_Cry-ECFP. TagRFP coding sequence was obtained from Evrogen, the Cry:eCFP cassette for selection of transgenic fish was described before[127]. Downstream of the six encoded GBSs I integrated two copies of NRS or control sequences using the oligonucleotides JT207-JT214 using a recognition site for AscI (Fermentas). For qPCR experiments, these reporters were further modified by Dr. Sergey Prykhozhij using gateway cloning exchanging the cfos promoter against the e1b promoter.

3.3.6 Cloning of Gal4-DBD fusion plasmids

Plasmids encoding fusion proteins of NonO, SFPQ and PSPC1 and the Gal4-DNA binding domain were kind gifts of Prof. Steven A. Brown (pSCT-G4-NonO, pSCT-G4-PSPC1, pSCT-G4-SFPQ) [128]. In addition, HMGA1 isoform A and isoform HMG-R were tested. To generate expression plasmids encoding Gal4-DBD fusion proteins with these isoforms, coding sequences were amplified by PCR from cDNA, using the primer pairs JT361/362 or JT361/363 respectively. Amplicons were then cloned into pCR-Blunt for amplification and sequencing. Coding sequences were then purified by restriction digestion with BglII and EcoR1 and subsequent gel extraction. Plasmid pSG5-Gal4 (described in[129]) was linearized by restriction digestion with BglII and EcoR1. Coding sequences were then integrated into the linearized pSG5-Gal4 by ligation to generate the plasmids pSG5-Gal4-HMGA1a and pSG5-Gal4-HMG-R.

3.4 Site directed mutagenesis

Site directed mutagenesis for targeted modification of specific sequences in luciferase reporter plasmids was done in 25 µl reaction mixes containing 1x Pfu ultra buffer, 20 ng target plasmid, 0.25 µM of each primer and 0,2 mM dNTPs. After heating the samples to 95°C 0,5 µl Pfu ultra polymerase (Agilent) was added to the reaction tubes followed by cycling conditions for PCR as follows:

95°C	1 minute	
95°C	30 seconds	16 cycles
55 °C	1 minute	
68°C	4 minutes	
68°C	10 minutes	
4°C	storage	

After the PCR, 1µl of DpnI was added and the samples were incubated for 2 hours at 37°C. 5 µl of this mix was used for transformation of z-competent Dh5alpha cells. Correct clones were identified by sequencing.

Site directed mutagenesis was applied to facilitate point mutations in the luciferase reporter pGL3-P-WFS1a. To eliminate the E-box motif within the enhancer site the Primers JT133 and JT134 were used, for elimination of the RunX1 motif the primers JT99/JT100 and for elimination of the GBS the primers JT139/149.

3.5 Purification of RNA

RNA was purified from approximately 1 million cells using the RNeasy kit (Qiagen) according to the manufacturer's guidelines. Briefly, cells were disrupted in buffer RLT supplemented with 1% β-mercaptoethanol, scraped and then an equal amount of RNase free ethanol was added. Lysed cells were loaded onto the column, the membrane washed once with 350 µl buffer RW1. DNA was then degraded by on column digestion by addition of 27.3 units RNase free DNase (Qiagen) in 80 µl buffer RDD and subsequently incubated for 10 minutes at room temperature. The column was washed again with 250 µl buffer RW1, then with 350 µl buffer RPE. RNA was eluted in 40 µl RNase-free water.

3.6 Preparation of cDNA

To prepare cDNA, 500 ng RNA and 310 pmol random primer in a total volume of 16 µl with 0,625 mM dNTPs each were incubated for 10 minutes at 70°C. Then 2 µl 10 x M-MuLV reverse transcriptase buffer, 0.25 µl RNAsin, 0,125 µl M-MuLV reverse transcriptase (NEB) and 1,62 µl water were added, and the solution was incubated for one hour at 42°C, followed by 10 minutes at 90 °C. Resulting cDNA was diluted by addition of 230 µl MilliQ water prior to quantification by qPCR.

3.7 Lumier assay

3.7.1 Coating of plates with antibody

To coat plates with antibody, 100 µl of carbonate buffer with 10 µg/ml sheep gamma globulin was added to each well of a white Greiner high binding 96 well plate and plates were incubated for three hours with constant shaking at room temperature. Next, the solution was replaced with 300 µl carbonate buffer with 1 % BSA and the plates were again incubated for one hour with constant shaking at room temperature. The plates were washed three times for five minutes each with TBST II with constant shaking at room

temperature. Subsequently, to each well 100 µl carbonate buffer with 3 µg/ml rabbit anti sheep IgG was added, and the plates were incubated over night with constant shaking at 4°C. The next morning the plates were washed three times for five minutes each with constant shaking at room temperature with TBST II. Coated plates were then wrapped in foil and stored at 4°C until use.

3.7.2 Transfection and measurement

Lumier experiments were done using 293 T-REx cells. On day zero, 96 well multititer plates were inoculated with 30.000 cells per well in 100 µl DMEM with GlutaMAX/10 % FBS. On day one these cells were transfected using plasmids encoding the two candidate proteins tested for interaction, one encoded as prey as a fusion protein with protein A, the other as bait as fusion protein with firefly luciferase. For every combination of bait and prey, three wells were transfected. Briefly, the 42.5 ng of both plasmids encoding the test candidates were added to fresh reaction tubes, then filled up to a total volume of 65,5 µl with OptiMEM. To another tube 0.75 µl Lipofectamine 2000 (Invitrogen) plus 70,5 µl OptiMEM were added per combination of candidate proteins. These solutions were incubated for five minutes at room temperature. Then 65,5 µl of the Lipofectamine 2000/ OptiMEM mix was mixed with an equal volume of the expression plasmids in OptiMEM, and these transfection mixtures were incubated for another 20 minutes at room temperature. For transfection, 37.4 µl of these solutions were added to each well.

24 hours post-transfection, cells were lysed by addition of 100 µl lumier lysis buffer and incubation on ice for 30 minutes. Cell lysates were cleared by centrifugation at 4000 rpm in an Eppendorf 5810 R centrifuge with an A-4-81 rotor. Transfection efficiency was determined by testing luciferase activity for the cleared lysate. For this, 5 µl of lysate was added to 35 µl PBS in wells of a white Greiner cellstar 96well plate. 40 µl of Firefly luciferase substrate was added to each well and luciferase activity was determined immediately using a luminescence plate reader (TECAN InfiniteM200).

70 µl of the lysate were transferred to wells of IgG coated plates prepared as described above. These plates were then incubated for two hours on ice to allow protein-protein interactions to reach equilibrium. Next, the wells were washed two times with ice-cold PBS. Finally, 40 µl each of fresh PBS and firefly luciferase substrate was added to each well and luciferase activity was determined immediately in a luminescence plate reader (TECAN InfiniteM200).

3.8 Maintenance of mammalian cells

U2-OS cells and U2-OS GR18 cells (stably expressing GR[121]) were maintained in DMEM (Gibco) supplemented with 5 % FBS (Gibco). Nalm-6 cells were maintained in RPMI 1640 (Gibco) supplemented with 10 % FBS (Gibco).

3.9 Transfection of Nalm-6 cells

Nalm-6 cells were transfected with the Amaxa nucleofector (Lonza) according to the manufacturer's guideline using the nucleofector kit T. For luciferase assays, cells were transfected with 1.5 µg of pGL3-promoter reporter plasmids together with 0.5 µg of plasmid pRL (CMV) to normalize for transfection efficiency.

3.10 Transfection of U2-OS cells

To transfect U2-OS cells, 48-well multititer plates were inoculated with 20.000 cells per well in 250 µl DMEM/5 % FBS and incubated overnight. The next morning, 10 ng reporter plasmid together with the plasmids pRL (CMV) (0.1 ng), pcDNA3.1-rGR (10 ng), p6R (50 ng), 0.8 µl plus reagent (Invitrogen) and 12.5 µl serum-free DMEM (Gibco) were used per transfection.. In separate tubes, 0.8 µl Lipofectamine (Invitrogen) and 12.5 µl serum-free DMEM were mixed for each transfection. The tubes were mixed by vortexing, and then incubated for 15 minutes at room temperature. The Lipofectamine/DMEM mix was added to the DNA mix, then again mixed by vortexing and incubated for an additional 15 minutes at room temperature. The cells were washed once with PBS, then 100 µl serum free DMEM medium was added and 25 µl of the transfection mix was added per well. Three hours post transfection, medium was replaced with 250 µl DMEM/5 % FBS.

3.11 Dual luciferase assay

6 hours post transfection, the medium was replaced with DMEM/5 % FBS supplemented with 1 µM dexamethasone (from a stock of 1mM dexamethasone in ethanol) or 1‰ ethanol as vehicle control respectively. 16-18 hours after hormone treatment, luciferase activities were measured using a luminometer (LUMIstar Omega by BMG Labtech) and the dual luciferase assay kit (Promega). Briefly, cells were lysed in 50 µl passive lysis buffer, then incubated for 30 minutes at room temperature. 2.5 µl to this suspension of lysed cells was transferred to a white 384-well multititer plate (Greiner). Luciferase activities were determined with firefly- (LAR II) and renilla luciferase substrates (Stop and Glo) from the kit, using 12.5 µl each. For measurement of luciferase expression in U2-OS

and GR-18 cells the gain of the luminometer was set to 3600, when Nalm-6 was used the gain was set to 4095.

3.12 Zebrafish reporter injection

3.12.1 TOL2-transgenesis

Tol2-transgenesis was done using the Tol2Kit[126]. Briefly, 10 µg plasmid pCS2+FA-transposase from the kit was linearized with NotI. Then NotI was degraded by proteinase K and DNA purified by phenol-chloroform extraction. The DNA was precipitated and washed once with 70 % EtOH. Linearized DNA was recovered in 20 µl DEPC water and used for *in vitro* transcription using the kit mMessage mMachine by Ambion according to the manufacturer's protocol. For injection, 10-20 ng/µl reporter plasmid (as described in chapter 3.3.5) were mixed with 25 ng/µl transposase mRNA. This mixture was used for microinjection of embryos at the one-cell stage (approximately 30 pg DNA per injection) as described in[126].

3.12.2 Quantification of TagRFP expression

48 hours post injection, fish were subjected to 8 hours of treatment with dexamethasone or DMSO as a vehicle control. Transgenic fish were identified by CFP expression in the eye lens. Quantification of TagRFP expression in transgenic fish was done using the confocal laser scanning microscope Zeiss LSM 700. For quantification of mRNA levels RNA was purified from transgenic Zebrafish using the RNeasy kit (Qiagen). Homogenization was done in 350-700 µl RLT buffer supplemented with 1 % β-mercaptoethanol using a ml syringes in combination with 20-22 G needles. Further processing was done according to the manufacturer's protocol. The resulting RNA was reverse transcribed to cDNA as described in chapter 3.6. The resulting cDNA was analyzed by quantitative real-time PCR (qPCR) using primers for efIa (Primers efIa-paper_qfor/rev), FKBP5 (Primers fkbp5_qfor/rev), eCFP (Primers ECFP qfor/rev) and TagRFP (Primers TagRFP qfor/rev). Data were analyzed using the ΔΔCt method using efla, a gene whose transcription is not affected by dexamethasone treatment, for normalization.

3.13 Stable transfection of GR-18 cells

To stably integrate reporter plasmids at the AAVS ("safe harbor") locus, GR-18 cells were transfected as described previously[123]. Reporter genes for stable integration at the AAVS locus were cloned into the SAA-GFP vector as described in chapter 3.3.3. . This plasmid contains a promoter-less GFP reporter preceded by a splice acceptor site and a 2A ribosome stuttering signal. Upon correct integration at the AAVS locus, a single transcript

is generated that encodes the peptide encoded by the first exon of the PPP1R12C gene and separately GFP. 10 µg of the reporter-containing SAA-GFP plasmids was co-transfected with 0,5 µg of a plasmid encoding a zinc-finger-nuclease targeting the AAVS locus[130,131]. Transfection was done using the Amaxa nucleofector Kit V (Lonza) using 1 million cells per transfection and the cells were equally distributed into 2 wells of a six well plate in 2,5 ml of DMEM/5 % FBS. 24 hours post transfection, the cell culture medium was replaced with fresh medium to wash away dead cells. 48 hours post transfection, cells were transferred to 75 cm² flasks. One week post transfection, the cells were tested for genomic integration of the reporter gene into the target locus by PCR using the primers R5 and LucNested. Three weeks post transfection transgenic GFP-positive cells were isolated by FACS. From the populations of GFP-positive cells, clonal lines were generated and each clonal line was again tested by PCR for successful genomic integration of the reporters in the AAVS locus.

3.14 Electrophoretic Mobility Shift Assay (EMSA)

Oligos were hybridized at 10 µM per oligo in water, then diluted and re-buffered to 5 nM in 2x EMSA binding buffer. Sequences for Cy5 labelled positive strand are shown below (NRS/Control sequence is underlined and black, the GBS encoding region is shown underlined and red). The negative strand was the corresponding reverse complement without fluorophore.

NRS2 coding oligo: Cy5-
TAGGTA**TTAATTCAATTAA**CTAGCCCGG**AGAACAAAATGTTCT**GATC

Control coding oligo: Cy5-
TAGGTA**CGAGGTAGGCTTG**CTAGCCCGG**AGAACAAAATGTTCT**GATC

Purified GRα DBD (amino acids 380-540[103]) was diluted in 1x EMSA binding buffer to eight different concentrations:
10 µM, 4 µM, 1,6 µM, 0,64 µM, 0,256 µM, 0,1024 µM, 0,045 µM and 0 µM.
To 6 µl of each of these protein dilutions, 3 µl of a dIdC solution at 1µg/µl and 3 µl of the 5 nM dsDNA solution were added and mixed. The samples were then incubated on ice for 30 minutes for protein-DNA interactions to reach equilibrium. Subsequently, the samples were loaded onto a running 5 % acrylamide gel in 0.5 x TBE buffer. For analysis, gels were scanned with the FLA 8000 scanner (Fujifilm) at an excitation wavelength of 640 nm and the filter 675DF20, the voltage of the photo-multiplier tube was set to 98% (980 V).

3.15 ChIP

For each condition approximately 3 - 5 million cells were used. Prior to crosslinking, the cells were treated for 90 minutes with 1 µM dexamethasone or 1 ‰ EtOH as a vehicle control. Crosslinking was done by adding formaldehyde to a final concentration of 1 % for 3 minutes at room temperature. The crosslinking reaction was then quenched by addition of glycine to a final concentration of 125 mM and subsequent incubation for 10 minutes at 4°C. Crosslinked cells were washed with PBS for 5 minutes and then scraped into 15 ml conical tubes in ice-cold IP lysis buffer (approximately 400 µl per 1 million cells) supplemented with 0.5 % proteinase inhibitor cocktail set III (EDTA-free, Merck) and 0.5 mM PMSF and nutated at 4°C for 30 minutes. Cells were pelleted by centrifugation for 5 minutes at 600 *g* and 4°C, pellets then re-suspended in RIPA buffer (300 µl per 3-5 million cells) supplemented with proteinase-inhibitors and PMSF as described above. 300 µl aliquots were transferred to fresh 1.5 ml Eppendorf reaction tubes. These aliquots were sonicated using a Diagenode bioruptor for 24 cycles, each cycle consisting of 30 seconds sonication and a 30 seconds pause to allow cooling. Sheared chromatin was either frozen at -80°C until further processing or used directly for immunoprecipitation.

For Immunoprecipitation, 400 µl fresh RIPA buffer was added to the sheared chromatin, then 70 µl of the solution was taken as an input control. 3 µl GR antibody (N499) was added to each sample of sheared chromatin and tubes were nutated overnight at 4°C. For precipitation, protein A/G coated agarose beads (Santa Cruz) were equilibrated overnight in RIPA buffer supplemented with proteinase-inhibitors and PMSF, both diluted 1:200. The next morning, 30 µl of the 50 % beads slurry per IP were added to each tube containing sheared chromatin and antibody. Next, the samples were nutated for two hours at 4°C to allow binding of antibody to the beads. The beads were then washed five times using RIPA wash buffer. If GR-18 cells were used two additional washing steps with lithium-chloride washing buffer were done. To elute the chromatin from the beads, the precipitated material was treated with proteinase K (0,2 mg in 75 µl crosslink reversal buffer) for three hours at 55°C to remove proteins and then overnight at 65°C for cross-link reversal. The next day, chromatin was purified using a Qiagen PCR purification kit, elution was done with 100 µl elution buffer. Enrichment of target loci by ChIP was tested by qPCR. A known endogenous genomic GR binding site (FKBP5; coordinates in GRCh37/hg19 assembly: chr6:35569764-35570000) in U2-OS[132] served as positive control for ChIP efficiency and was targeted using the primers FKBP5 GBS6.1 fw/rv. Five loci that

are not GR targets and can be targeted with the primers pair hRPL19 fw/rv are used as negative control, (coordinates in GRCh37/hg19 assembly: chr5:177482959-177483191; chr8:99794622-99794853; chr1:64254390-64254621; chr17:37360328-37360897; chr7:102781800-102782032 GRCh37/hg19 assembly). The reporter gene GBS is targeted using the primer pair JT201/JT202.

3.16 Preparation of Nalm-6 cells for ChIP-Exo

900 million cells were transferred to 180 ml RPMI1640/10 %FBS. Dexamethasone was added to a final concentration of 1 µM, cells then equally split into six 150 cm^2 cell culture flasks and incubated at 37°C and 5 % CO_2 for 90 minutes. Afterwards, cells were collected in one flask and crosslinking was done by addition of formaldehyde to a final concentration of 1 % and subsequent incubation for 10 minutes at room temperature. The crosslinking reaction was quenched by the addition of glycine to a final concentration of 125 mM and subsequent incubation at 4°C for 10 minutes. Cells were centrifuged at 1000 rpm for 8 minutes, and cells were then washed once with PBS. Next, cells were re-suspended in 68 ml IP lysis buffer and nutated for 30 minutes at 4°C. 11 ml lysed cells were transferred to fresh 15 ml conical tubes and centrifuged at 600 *g* for 5 minutes at 4°C. The resulting pellet was resuspended in 1350 µl RIPA buffer supplemented with proteinase inhibitors and PMSF both diluted 1:1000. Cells were sonicated in 15 ml clear polypropylene centrifuge tubes (Corning) for 20 cycles of 30 seconds sonication followed by a 30 seconds pause using a Bioruptor Plus (Diagenode) at high intensity with constant cooling to 4°C. Sheared chromatin was divided into 150 µl aliquots, and sent on dry ice to Peconic LLC for ChIP-Exo processing and sequencing.

3.17 qPCR

qPCR was performed using the power Sybr® Green PCR master mix by Applied Biosystems or a home mix consisting of 100 mM Tris pH 8.3, 6 mM $MgCL_2$, 1 mg/ml BSA, 4 mM dNTPs 0,66x SYBR-Green and 1x ROX reference dye. Each sample was measured in duplicates and at least three biological replicates were analyzed for each experiment. The total reaction volume was 10 µl, with 5 µl of the qPCR master mix and 0,2 µM of each qPCR primer. 0,5 µl ChIP DNA (or 2 µl of cDNA) or 2 µl of DNase and MNase-treated chromosomal DNA were used per reaction. qPCR was done using the ABI 7900 HT by Applied Biosystems with cycling conditions as shown in Table 1.

Table 1: Cycling conditions

Temperature	Duration	Ramp rate	
95 °C	600 seconds	100 %	Initial denaturation
95 °C	15 seconds	100 %	Amplification (40 cycles)
60 °C	60 seconds	100 %	
95 °C	15 seconds	100 %	Recording of melting curve
60 °C	15 seconds	100 %	
95 °C	15 seconds	2 %	

3.18 DNase I Hypersensitivity assay

Cells were grown in 6 well tissue culture plates to confluency, then treated with 1µM DEX or 1 ‰ EtOH for 90 minutes. Next, cells were washed with PBS and scraped into 1 ml DNaseI reaction buffer supplemented with 0.2 % NP40 alternative to lyse cell membranes. Cells were then homogenized by vortexing, incubated 5 minutes on ice and centrifuged for 5 minutes at 500 *g* and 4°C to pellet nuclei. The pellets were resuspended in 200 µl DNaseI buffer. 50 µl aliquots were then DNase I treated (or mock treated to normalize for the amount of chromatin input) by the addition of 25 µl of DNaseI buffer containing 1.5 µl DNase I (Qiagen, 2.7 u/µl) and then incubated at 37°C for 8 minutes. The reaction was stopped by addition of an equal volume of 2x stop buffer containing 200 µg/ml proteinase K and samples were incubated at 65°C for 4 hours to remove proteins. DNA was purified using the PCR purification kit by Qiagen and genomic regions of interest were analyzed by qPCR.

3.19 MNase assay

The nucleosome positioning assay was done essentially as described previously[133], with slight modifications to enable qPCR instead of Affymetrix array analysis. Cells were grown to confluence in 10 cm dishes and then treated with 1 µM dexamethasone or 1 ‰ ethanol respectively for 60 minutes. Prior to harvesting, the cells were washed once with PBS, then resuspended in 11.3 µL ice-cold lysis buffer per cm^2 and transferred to 1.5 ml Eppendorf tubes. Chromatin was collected by centrifugation at 1600 RPM for 15 min at 4°C. Resulting pellets were resuspended in 200 µL storage buffer per 10^7 cells. Chromatin samples were flash-frozen in liquid nitrogen as aliquots, and then stored at -80°C. Prior to measuring the concentration of total DNA, these aliquots were diluted 1:10 in storage buffer, then 0,4 volumes of 5 M NaCl was added to the samples to disrupt protein/DNA interactions other than those between histones and DNA. Each chromatin sample was split

in two of which one half was treated with MNase whereas the other half was untreated to normalize for the amount of input. For each sample, I mixed 50 µl MNase reaction buffer with 800 ng genomic DNA in MNase storage buffer. For the MNase treated samples, I added 50 µl reaction buffer with 1 µl MNase (NEB; ~200 kunitz units/µl) per chromatin sample. For the untreated sample for normalization, 50 µl MNase reaction buffer without MNase was added. Samples were then incubated for 10 minutes at 30°C, the reaction stopped by adding 100 µl stopping buffer supplemented with 15 µl Proteinase K. Next, the samples were incubated for 120 minutes at 60°C. To purify the MNase digested chromatin, 100 µl Phenol and 100 µl chloroform/isoamyl alcohol (24:1) were added and samples were mixed by vortexing and centrifuged for 5 minutes at room temperature and 12000 *g*. DNA in the aqueous phase was further purified by ethanol precipitation by adding 1/10 volume of sodium acetate, 2.5 volumes of 100 % EtOH and 2 µl Glycogen. Samples were mixed by vortexing and transferred to -20°C for 20 minutes, then centrifuged at 4°C for one hour at maximum speed. DNA pellets were recovered in 20 µl TE buffer (100 µl for the control samples). Mono-nucleosomal DNA from the test conditions was purified by cutting out the gel-slice corresponding to a DNA size of approximately (100-220 bp) from an 1.5 % agarose gel and DNA from the gel-slice was extracted using the NucleoSpin Gel and PCR clean-up (Macherey and Nagel) kits according to the manufacturer's protocol, except that the elution was done in 100 µl TE buffer. Finally, samples were analyzed by qPCR. To target the sequences -57 to +12 bases relative to the TSS (a region with low predicted nucleosome occupancy), I used the primers JT393/JT394; the primers JT397/JT398 to target the region +74 to +155 bases relative to the TSS (a region of high nucleosome occupancy), and the primers JT201/JT202 to target the GBS of the genomically integrated luciferase reporter gene.

3.20 CpG methylation assay

Ethanol-treated input samples of the ChIP assays as described in chapter 3.15 were used as starting material for testing the level of CpG methylation. For each assay, 3 µl input DNA was mixed with 3 µl Tango buffer (Fermentas) and 2 µl water. The samples were split equally to two fresh reaction tubes, then to one tube 1 µl of ClaI (Fermentas, 10 u/µl) was added, and to the other tube 1 µl water. The samples were then incubates for 5 hours at 37°C. Prior to qPCR analysis, to each sample 45 µl water was added. Primers FKBP5 GBS 6.1 fw/rv were used to target a genomic region without ClaI recognition site, thus enabling control for restriction specificity. The primers JT201/JT202 target the reporter gene GBS

region which includes a ClaI recognition site. CpG methylation prevents ClaI restriction, thus in case of NRS-mediated CpG methylation, the restriction of the fragment should be prevented or reduced.

3.21 Preparation of nuclear extract

Preparation of nuclear extract was done as described previously[134]. Cells were grown to confluence in 15 cm dishes and then treated for 90 minutes with 1 µM dexamethasone or with 1 ‰ ethanol as vehicle control. Then the cells were washed once with ice-cold PBS, scraped into a 15 ml falcon tube with 2ml PBSI per dish and centrifuged for 5 minutes at 550 *g* at 4°C. The supernatant was removed and the cell pellets were transferred to fresh 1.5 ml reaction tubes. The samples were centrifuged again for 30 sec at 1500 *g* and 4°C and then the supernatant was removed. The nuclei were resuspended in approximately two package volumes of Buffer A, incubated on ice for 10 minutes, then vortexed briefly. Samples were centrifuged 30 sec at 2600 *g*, supernatant was removed and the pellets were resuspended in 2/3 cell package volume of buffer B. Samples were sonicated for 5 seconds using a Bioruptor, then the supernatant was diluted 1:1 with Buffer D. The resulting nuclear extract was then flash-frozen in liquid nitrogen and stored in aliquots at -80°C.

3.22 DNA pull-down

Two sets of double stranded DNA oligonucleotides were used as baits for the DNA pull-down assays. Initially, these encoded one of the two identified NRS sequences or a control sequence flanked by a PstI restriction site on one site and a canonical GBS sequences on the other site (JT269-JT276). Alternatively, we used similar baits that did not encode the canonical GBS (JT277-JT284). For the pull-downs, the DNA baits were resuspended in annealing buffer to a concentration of 50 µM. Then, 30µl sense strand oligonucleotides with a biotin-tag on their 5'end were mixed with 40 µl of non-biotinylated antisense oligonucleotide. To hybridize these oligonucleotides, they were heated up to 90°C for 5 minutes in a thermocycler, then the temperature was gradually decreased to 65°C within 10 minutes. Samples were incubated at 65°C for 5 minutes and slowly cooled down to room temperature by switching off the thermocycler.

For pull-downs, I used 1 mg Dynal MyOne C1 streptavidin magnetic beads (Invitrogen) per 1 mg protein from the nuclear extract. Beads were first washed twice with 400 µl TE buffer supplemented with 0.01 % NP40, then twice with 750 µl DW buffer. The beads were resuspended in 400 µl DW buffer and 5 µl of the solution of the hybridized

oligonucleotide baits was added. The beads, together with the double stranded bait oligonucleotides, were incubated for 3 hours at room temperature on a rotary wheel. Beads were then washed with 400 µl TE supplemented with 0.02 % NP40, then three times with 400 µl DW buffer. Oligonucleotide-loaded beads were resuspended in to 100 µl DW buffer and stored at 4°C until use.

To prepare naked beads, 1 mg Dynal MyOne C1 streptavidin magnetic beads were washed with 400 µl TE buffer supplemented with 0.01 % NP40, then washed three times with 750 µl DW buffer, resuspended in 100 µl DW and stored at 4°C until use.

For each pull-down, I used nuclear extract containing approximately 500 µg total protein and 0.5 mg dsDNA loaded beads. Beads were resuspended in 650 µl blocking buffer containing 2.5 mM DTT, and then incubated for 1 hour at RT on a rotary wheel. Beads were then washed once with 100 µl RE Buffer containing 0.02 % NP40, then twice with 100 µl of buffer G.

In parallel, 0.5 mg naked beads were washed twice with 400 µl buffer G containing 2 mM DTT; 0,4 mM PMSF and proteinase inhibitors diluted 1:1000. Nuclear extract was centrifuged for 20 min at 15000 *g* and 4°C, and then potassium-glutamate was added to the supernatant to a final concentration of 10 mM. The solution was then diluted with one equal volume of poly dIdC (0.2 mg/ml) in buffer G supplemented with 2mM DTT, 0,4 mM PMSF and proteinase inhibitors in a dilution of 1:1000, then centrifuged for 10 minutes at 15000 *g* at 4°C. The supernatant was pre-cleared to remove proteins that bind non-specifically to beads by incubation with the naked beads for 1 hour at 4°C on a rotating wheel. The pre-cleared supernatant was transferred to the dsDNA oligo loaded beads and then incubated for 3hours at 4°C on a rotating wheel. Then the beads were washed four times with 1000 µl buffer G.

Sample preparation for mass spectrometry analysis was done in two different ways. The beads were either used directly for elution by trypsin digestion and subsequent precipitation and mass-spectrometry analysis, or protein elution was facilitated by restriction digestion directly from the beads. For restriction digestion the beads were washed again with 450 µl buffer RE resuspended in 125 µl buffer RE, then 7 µl PstI (10 *u*/µl) was added and the samples were incubated in a thermomixer at 25 °C for 1 hour with constant shaking at 1400 rpm. Then supernatant was transferred to a fresh tube and kept on ice, and the restriction digestion was repeated for a second elution step. The pooled

supernatants from the restriction digestions were then trypsin digested followed by precipitation and mass-spectrometry analysis by David Meierhofer.

Briefly, 1 ml ice-cold acetone was added to eluents, then the samples were centrifuged for 15 minutes at 20800 *g* and 4°C and the pellet was washed twice with acetone. Pellets were dried in a SC210A Speed-Vac (Savant), then 100 µl of 25 mM NH_4HCO_3 containing 10 mM DTT was added and samples were incubated for one hour at 56°C. Subsequently, 100 µl of 25 mM NH_4HCO_3 with 50 mM iodacetamide was added and samples were incubated for 30 minutes in the dark. Pellets were again washed twice with acetone and dried again in a SC210A Speed Vac (Savant). Then 100 µl of 25mM NH_4HCO_3 with 1 µl trypsin was added and samples were incubated over night at 37°C. Mass-spectrometry analysis was done as described before[135], however using the IPI database (IPI Human v.3.87).

3.23 esiRNA and DsiRNA transfection and luciferase assay

On day zero, 10,000 transgenic GR-18 cells in 250 µl medium were seeded per well of a 48-well multititer plate. On day two, these cells were transfected with esiRNA (Sigma-Aldrich) or DsiRNA (Integrated DNA Technologies).

For transfection of each well, 75 ng RNA was mixed with 12.5 µl DMEM. In a separate tube, 0.25 µl Lipofectamine 2000 (Invitrogen) was mixed with 12.5 µl DMEM. These mixes were incubated for five minutes, then combined and incubated again for an additional 20 minutes at room temperature and added to each well. 6 hours post transfection, the medium was replaced with 250 µl fresh DMEM/5% FBS. 32 hours post transfection hormone treatment was started by addition of fresh DMEM/5% FBS supplemented with either 1 µM dexamethasone or 1 ‰ ethanol as vehicle control. 16 hours post hormone treatment, the cells were lysed with 50 µl passive lysis buffer (Promega) and firefly luciferase activity was determined as described for the dual luciferase assay.

3.24 Polyacrylamide gel electrophoresis

Polyacrylamide gels (SDS-PAGE) for the analysis of proteins were prepared according to the following recipe:

	8 % resolving gel	Loading gel
1,5M Tris pH 8,8	5ml	0,63ml
30 % acrylamide	5,4ml	0,83ml
10 % SDS	200µl	50µl
TEMED	8µl	5µl
10 % APS	200µl	50µl
water	9,2ml	3,4ml

After loading of the samples, the gels were run at 85V until the dye of the marker reached the resolving gel, when the voltage was increased to 110 V.

3.25 Semi-Dry Western Blotting

Proteins from SDS-PAGE gels were transferred onto nitrocellulose membranes (0.45 µm; BIO-RAD). Briefly, two layers of gel blotting paper soaked in transfer buffer were covered with a nitrocellulose membrane, and then the PAGE gel was carefully layered on top and was covered by two additional layers of gel blotting paper soaked in transfer buffer. Proteins were transferred at 55 mA per gel. The membranes were then blocked for one hour in 5 % BSA in TBST buffer, then the primary antibody (anti actin antibody (I-19; Santa Cruz) diluted 1:1000, anti GR antibody diluted 1:4000) in 5 % BSA in TBST was added and the membranes were incubated overnight. The next day, the membranes were washed 3 times with TBST for five minutes each and washed once with 5 % BSA in TBST. Next, the secondary antibody coupled to horseradish peroxidase (HRP) (goat anti rabbit; Invitrogen) was added in a 1:400 dilution in 5 % BSA in TBST. The membranes were again incubated for one hour at room temperature and washed five times in TBST for five minutes each.

Detection of signal was done using the SuperSignal West Dura Extended Duration Substrate (Thermo Scientific) in combination with the LAS1000 camera (Fujifilm).

3.26 Immunofluorescence staining

Cells were grown on coverslips to approximately 80 % confluency. For immuno-fluorescence staining, cells were then washed once with PBS, fixed with 3.7 % paraformaldehyde in PHEM buffer for 7 minutes at room temperature, then washed again two times with PBS buffer. Cells were permeabilized with 1 % Triton-X-100 in PBS for 15

minutes at room temperature, then again washed two times with PBS. To prevent unspecific antibody interactions, cells were blocked for 10 minutes with 1 % donkey serum in PBS for 15 minutes, primary antibody in blocking buffer was added to the cells using the dilutions as indicated (Table 2) and incubated for one hour at room temperature.

Table 2: Antibodies used for immunofluorescence experiments and corresponding dilutions

Primary Antibodies	***Used Dilutions***
Rabbit anti-PSP1 antibody	1:200
Rabbit anti-lamin B1 antibody	1:300
Mouse anti-lamin A/C antibody	1:50
Rabbit anti-coilin antibody	1:400
Mouse Anti-fibrillarin antibody	1:200
Rabbit Anti-H4K9me^{3} antibody	1:500
Rabbit Anti-SATB1 antibody	1:100
Secondary antibodies	***Dilutions***
donkey anti-mouse or rabbit Alexa 488	1:1000
donkey anti-mouse or rabbit Alexa 594	1:1000

Unbound antibody was removed by washing 3 times for five minutes each with PBS containing 0.1 % Tween. Then secondary antibody was added using dilutions as indicated (Table 2) and incubated for one hour at room temperature. Cells were washed again 3 times for five minutes each in PBS containing 0.1 %Tween, DNA was stained with Hoechst at 12,5 µg/ml for 5 minutes at room temperature and slides washed again two times in PBS. Slides were mounted using Vectashield (Vector laboratories)

3.27 FISH probe labelling by nick translation

For biotin labelling of probes, the Amersham Nick Translation Kit (GE healthcare) was used according to manufacturer's protocol. Briefly, each reaction was done according to the following protocol:

First, a nucleotide mixture was prepared using the nucleotide solutions from the kit by mixing 7µl each of dATP, dCTP and dGTP with 2 µl dTTP. To 20 µl of this nucleotide mixture, 1 µl of the biotin-16-dUTP (1mM solution) was added together with 1.5 µg DNA for labeling and water to a reaction mixture total of 50 µl. Then, 10 µl of the kit's enzyme mix was added plus 0.0023 units of DNase I. The solution was then incubated for 4 hours at 15°C and the reaction was stopped by adding 2 µl of 0.5 M EDTA pH8.0 and 37 µl TE Buffer. Unincorporated nucleotides were removed using Illustra MicroSpin G-25 columns (GE healthcare) according to the manufacturer's protocol.

3.28 Fluorescence in situ hybridization (FISH)

For Fluorescence in sit hybridization (FISH), biotinylated probes were hybridized to the target loci and detection was done by tyramide signal amplification using a kit from Molecular Probes with a horseradish peroxidase-streptavidin (Kit # T20935). Briefly, cells were grown on acid-treated microscopy slides to approximately 80 % confluence. Cells were washed once with PBS, fixed with 3.7 % paraformaldehyde in CSK buffer for 5 minutes at room temperature, then washed again two times with PBS buffer. Cells were permeabilized with PBS containing 1 % Triton-X-100 for 15 minutes at room temperature, then again washed two times with PBS. Genomic DNA was dehydrated by subsequent incubation in ethanol solutions of 70 %, 90 % and 100 % respectively for five minutes each. Dehydrated genomic DNA was denatured in 70% formamide in 2x SSC buffer for 5 minutes at 80°C. The dehydration was repeated, then the slides were dried. Per slide, 130 ng labelled probe together with 20 µg Cot1 DNA were purified by ethanol precipitation, then resuspended in 15 µl hybridization solution B (Cytocell). Slides and probes were prewarmed to 37°C for 5 minutes. Then probes were added to the slides, covered with coverslips and sealed with rubber cement, heated to 80°C for 2 minutes and then incubated over night at 37°C in a humidified chamber. The next morning, coverslips were removed and slides were washed with 50 % formamide in 2x SSC buffer for five minutes at 37°C, then three times for five minutes each with 2x SCC buffer at room temperature and once with PBS. Slides were blocked with blocking buffer (1 % BSA in PBS) for 30 minutes, then the horseradish peroxidase conjugate in a 1:100 dilution in blocking buffer was added. Next, slides were incubated for 30 minutes in a humidified box, then washed 3 times with PBS for five minutes each at room temperature. Tyramide amplification buffer was prepared according to the kit's protocol and added to the slides. DNA was stained with Hoechst at 12.5 µg/ml for 5 minutes at room temperature and slides were washed again two times with PBS before mounting using Vectashield (Vector laboratories).

4 Results part 1: Sequence encoded signals that enhance genomic TF binding

4.1 Testing genomic binding sites for enhancer activity

ChIP-Seq experiments have yielded a comprehensive picture of the genomic regions were GR binds in several cell types. Interestingly, when we (in collaboration with Dr. Morgane Thomas-Chollier) compared the genomic loci bound in cell lines derived from different tissues, we found that they show little overlap. To understand the mechanisms that drive the tissue-specific binding by GR, we analyzed these binding regions to identify specifically enriched sequence motifs for the different cell types. Such motifs are candidates for serving as signals that specify where GR can bind in a particular cell type. I decided to focus on Nalm-6 cells, as the ChIP-Seq peaks for GR in this cell type show little enrichment for the canonical GR binding site sequence, indicating that other sequences play a role in guiding GR to defined genomic loci. Examples of motifs that are specifically enriched in Nalm-6 cells are: sequence motifs resembling Runx1 binding sites, PU box motifs usually bound by ETS transcription factors[136,137] and an E-box motif. TFs that associated with RunX1 binding sites and PU boxes are active in hematopoiesis[111,138], making them ideal candidates to serve as signals for cell-type specific GR binding.

To test these candidate sequences, I cloned ~500 bp genomic regions encoding the candidate sequences upstream of a luciferase reporter gene. I chose regions that were near cell-type specific GR-target genes, and for those genomic regions that recapitulated the GR-dependent transcriptional regulation, I mutated individual binding sites that matched the sequence for enriched motifs to test their function.

Following this approach, I tested 15 genomic regions, of which only one showed GR-dependent enhancer activity. This site (chr4:6,262,219-6,262,710; Human Feb. 2009 (GRCh37/hg19) Assembly) is located upstream of the gene WFS1 that is regulated by GR (Figure 7 & Figure 8). Based on its proximity to the GR target gene WFS1, I called the enhancer WFS1a. The genomic region of this reporter contains a GBS-like sequence, a sequence matching an E-Box motif and one matching a canonical RunX1 binding site.

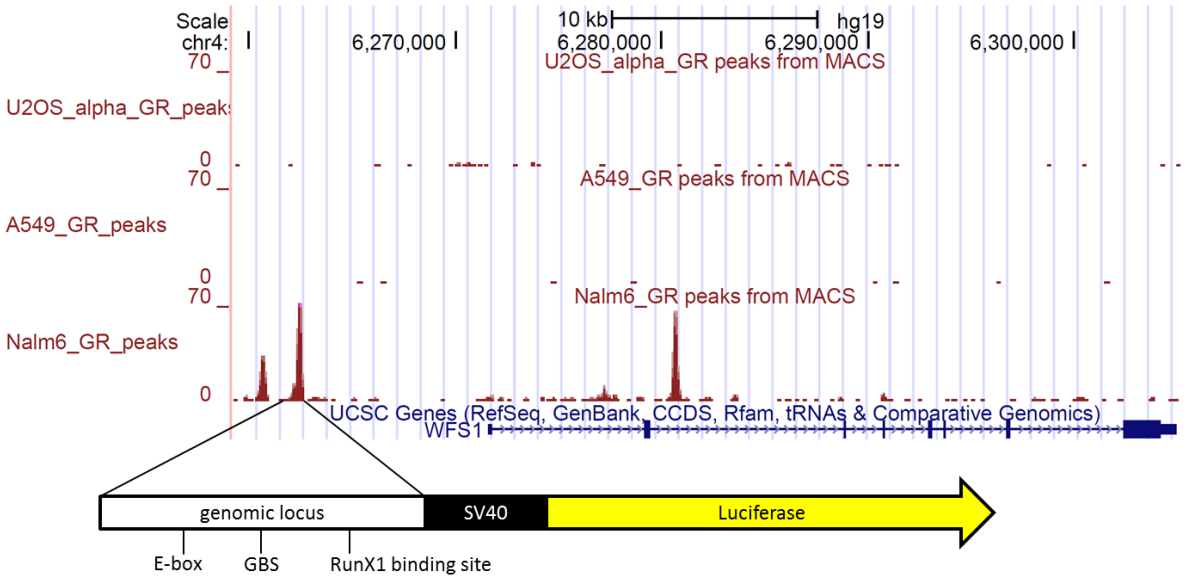


Figure 7: Design of reporter genes to test genomic GR binding regions for enhancer activity.
Nalm-6 specific genomic GR binding sites proximal to dexamethasone-regulated genes were amplified from genomic DNA by PCR. To generate luciferase reporter genes, amplicons were then cloned into the reporter plasmid pGL3-promoter, using the restriction enzyme XhoI as described in chapter 3.3.1.

4.2 Testing the influence of candidate sequences on enhancer activity

To test the role of the Nalm-6 specifically enriched Runx1 and E-box motifs in facilitating GR-activity from the WFS1a reporter, we measured the effect of mutating the sequences matching these motifs. Interestingly, the single mutation of either the RunX1 binding site or of the E-box both lead to decreased dexamethasone dependent reporter activation; from approximately 4-fold activation for the wild type reporter to a 2-fold activation for both single mutation-containing reporters. When the mutation of both sites was combined, the hormone-dependent activation of the reporter was almost completely lost (approximately 1.2-fold activation remaining). The residual activation in the absence of RunX1 and E-box motifs is comparable to the effect of mutation of the encoded canonical GR binding site which also results in a complete disruption of hormone-induced reporter activation.

These results indicate that the two sequence motifs appear to play an important role in recruiting GR to the encoded GR binding site. One possible mechanism is that they indirectly influence GR binding, for example by “opening” the chromatin to facilitate GR binding. This type of collaborative interaction between transcription factors was observed for GR[139]. Alternatively, these binding sites might recruit TFs that directly interact with GR to either tether GR to the DNA or to stabilize the interaction of GR with DNA at these loci.

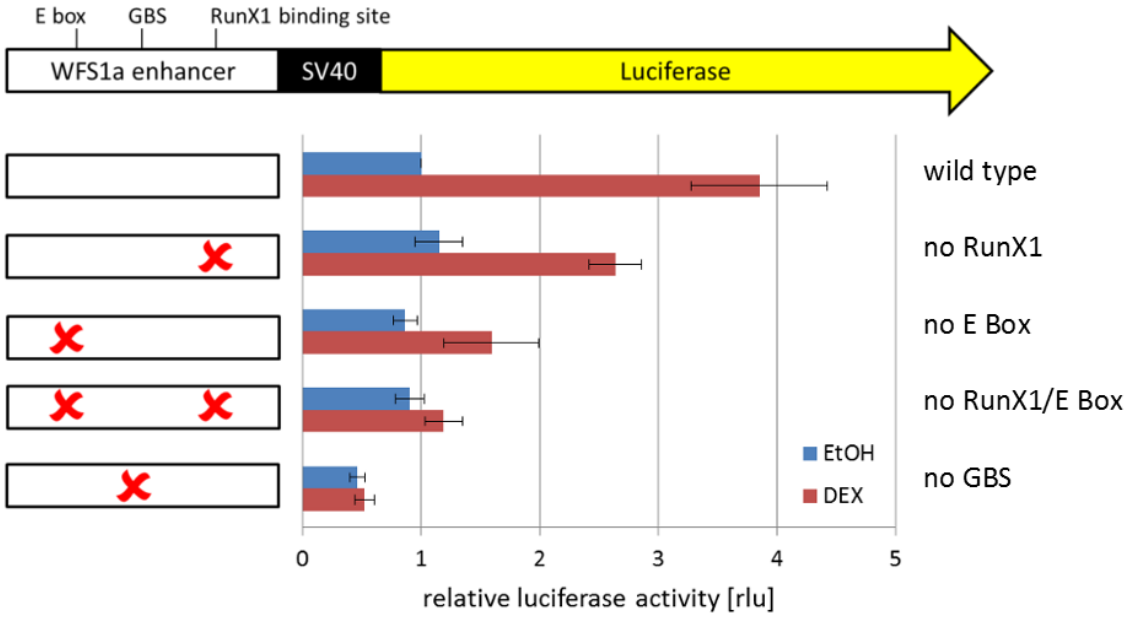


Figure 8: Runx1 and E-box motifs are important for GR-dependent WFS1a enhancer activity.
In addition to a canonical GBS, the WFS1a (see chapter 4.1) enhancer encodes motifs matching a RunX1 binding site and an E Box as shown. To test the influence of these sequence motifs, mutant versions of the WFS1a reporter were generated by site directed mutagenesis targeting each site individually, or the RunX1 binding site and the E Box motif in combination as indicated. These mutant reporters were then tested in luciferase assays and compared against the wild type reporter. The luciferase activity for all samples was normalized against the basal (EtOH) activity of the wild type enhancer reporter gene. The experiment was done in three biological replicates, and averaged normalized luciferase activity ± standard error of mean is shown.

4.3 Testing protein-protein interactions of candidate proteins and GR

The results with the WFS1a reporter indicated that GR might be recruited to genomic binding sites by protein-protein interactions with RunX1 or by interacting with E-Box-binding proteins.

Additional candidate proteins that have been shown to interact with PU-boxes[136] are ETS proteins. Experiments in mice indicate a role for most ETS proteins in hematopoiesis[138], implicating a similar role in other vertebrates. In human, 29 ETS transcription factor coding genes are found, some of which were already shown to interact with GR directly[140–142]. Since RunX1[110] and PU.1[143] are expressed in Nalm-6 cells, both are well suited candidates for protein-protein interactions with GR. In addition, TFAP4 was chosen as candidate because it is an E Box-binding protein expressed in B-cells[144] and according to the Jaspar database its binding sites has a very strong resemblance to the E-box motif that we found enriched at GR bound loci.

I used the lumier assay (luminescence-based mammalian interactome mapping) [145] to test for protein-protein interactions between GR and candidate factors. For this, one of the

candidates (bait) is fused to protein A whereas the other protein (prey) is expressed as a fusion protein with luciferase. These proteins are co-expressed in eukaryotic cells, cells are lysed and lysate is transferred to multi-titer plates that are coated with antibodies to facilitate immobilization of the bait protein. The plates are then washed to remove unspecific interactions and the retained luciferase activity is used to infer the strength of the protein-protein interaction of the bait and prey proteins.

As positive controls for the assay, I used the described interaction between the PU.1 and RunX1 proteins and between GPKOW and DHX16 [146]. As negative control, I used the proteins GASP2 and DHX16 that do not show interaction in the lumier assay (unpublished data from the Stelzl group). Candidates tested for protein-protein interaction in this experiment were GR, PU.1, RunX1 and TFAP4.

In initial experiments just using technical replicates, I tested all combinations of candidate proteins for potential interactions regardless of whether it was expressed as bait or prey, as well as the described combinations for the controls. As expected, retained luciferase activity for the positive control (GPKOW and DHX16) was high (approximately 84.150 rlu), whereas the retained luciferase activity for the negative control (GASP2 and DHX16) was low (approximately 1150 rlu), indicating that the assay works (Figure 9 A1 and A2). The expected interaction between RunX1 and PU.1 was only observed when PU.1 was used as bait and RunX1 as prey (Data not shown). Interestingly, of all potential combinations of protein-protein interactions tested, RunX1 and TFAP4 showed the strongest interaction with GR, but only when GR was used as bait. In contrast, when GR was used as prey, no interaction with any of the tested proteins was observed, indicating that the fusion of GR to the luciferase protein blocks potential protein-protein interaction by GR (data not shown). This was also seen by Sebastiaan Meijsing (unpublished results) who tested several known GR-interaction partners in the lumier assay and only observed interactions when GR was fused to protein A. The fact that fusion of both PU.1 and GR to the luciferase protein interferes with their ability to interact with other proteins suggests that the lack of interaction between these proteins observed in lumier assays likely reflects technical limitations of the assay rather than a lack of interaction between these two proteins.

Retained luciferase activity for the combination of pA-GR with DHX16-Luc (approximately 1250 rlu) was similar to the negative control indicating, that unspecific protein interaction of pA-GR in this assay is low (Figure 9 A2). Next, I tested the

interaction of RunX1 and TFAP4 as a prey with GR as a bait in three biological replicates consisting of three technical replicates each. In these experiments I used pA-GASP/DHX16-Luc as negative control. The retained luciferase activities for the tested combinations of GR with either TFAP4-Luc or RunX1-luc were approximately 5 or 16 times higher than observed for the negative control (Figure 9 B). This strongly suggests that both, TFAP4 and RunX1, might also interact with GR in vivo.

Together, these experiments indicate a dense interaction network between TFAP4, PU.1 and GR, consistent with the idea that protein-protein interactions between these proteins and GR can either tether GR to the DNA or might play a role in stabilizing the direct interaction of GR with DNA.

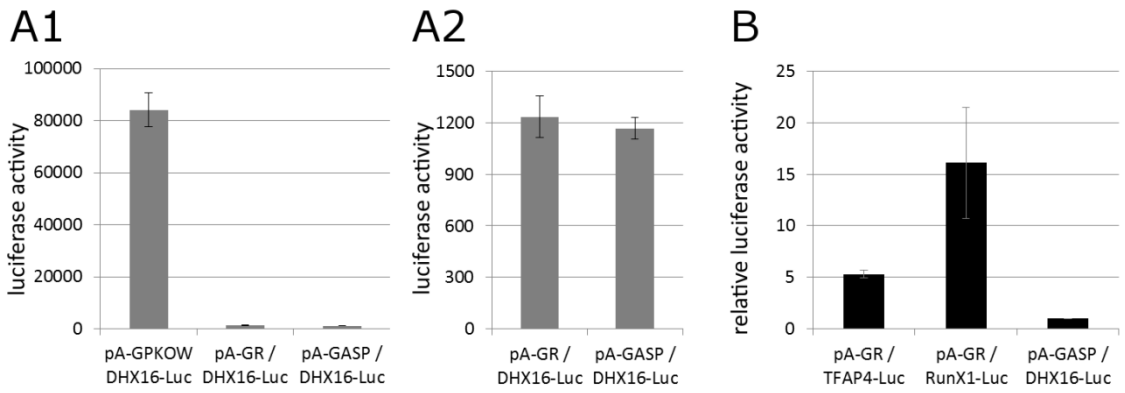


Figure 9: Protein-protein interactions between candidate transcription factors from ChIP-Seq Analysis.
GR, PU.1, RunX1 and TFAP4 were tested for protein-protein interactions in LUMIER assays, as described in chapter 3.7. (**A**) The interaction of DHX16-luc with pA-GPKOW served as positive control for a protein-protein interaction, whereas the interaction of DHX16-luc with pA-GASP served as negative control. To test for the level of unspecific interaction of GR, the combination of pA-GR and DHX16-luc, two proteins not expected to interact, was tested in parallel. These combinations of bait and prey proteins were tested for interaction in three technical replicates. (**B**) To test for interaction of GR with candidate proteins, these were tested in three biological repeats with three technical replicates each. The retained luciferase activities within the technical replicates were averaged for each biological replicate, and then normalized to the negative control. These normalized values were then averaged over all three biological replicates and plotted ± standard error of the mean for each tested combination.

4.4 Analysis of candidate sequence motif protein interactions by ChIP-Exo

Recruitment of GR to genomic targets can be facilitated by direct GR-DNA interactions. Alternatively GR can indirectly interact with DNA via interactions with other transcription factors that bind to specific DNA sequences (tethered binding). If other transcription factors recruit GR to the chromatin, such factors should be bound to these loci with GR simultaneously. Arguing for an important role for tethered binding by GR in Nalm-6 cells is the fact that there is a relatively weak enrichment of the canonical GBS in Nalm-6 (compared to the enrichment seen in U2-OS and A549 cells).

Sequences enriched in GR-bound sites are good candidates for facilitating tethered binding by GR, especially motifs that can be bound by TF directly interacting with GR like Pu.1 and Runx1. The enrichment of a motif however, does not reveal the underlying mechanism. For instance, both tethered binding and a role for these motifs in making specific genomic loci accessible for binding by GR can be the cause for the enriched binding observed. In an attempt to gain further insight into the role of enriched motifs in guiding GR to defined loci I turned to ChIP-Exo. ChIP-Exo is a modified version of the ChIP protocol and can give genome-wide information about the exact position of binding sites for a transcription factor of interest at single base pair resolution[147].

Briefly, in ChIP-Exo immunoprecipitated fragments of genomic DNA from ChIP are subjected to lambda exonuclease degradation. The exonuclease degrades the DNA fragments in the 5' to 3' direction. Complete degradation is prevented by cross-linked transcription factors that spatially block the exonuclease from further processing. This generates DNA fragments that are partially synchronized at their 5' ends at transcription factor binding sites. This synchronization results in enrichment of sequencing start sites during deep sequencing. The sequence bound by the ChIPed transcription factor will thus be flanked by an accumulation of reads on the positive strand from one side and on the negative strand from the other side (Figure 10 A).

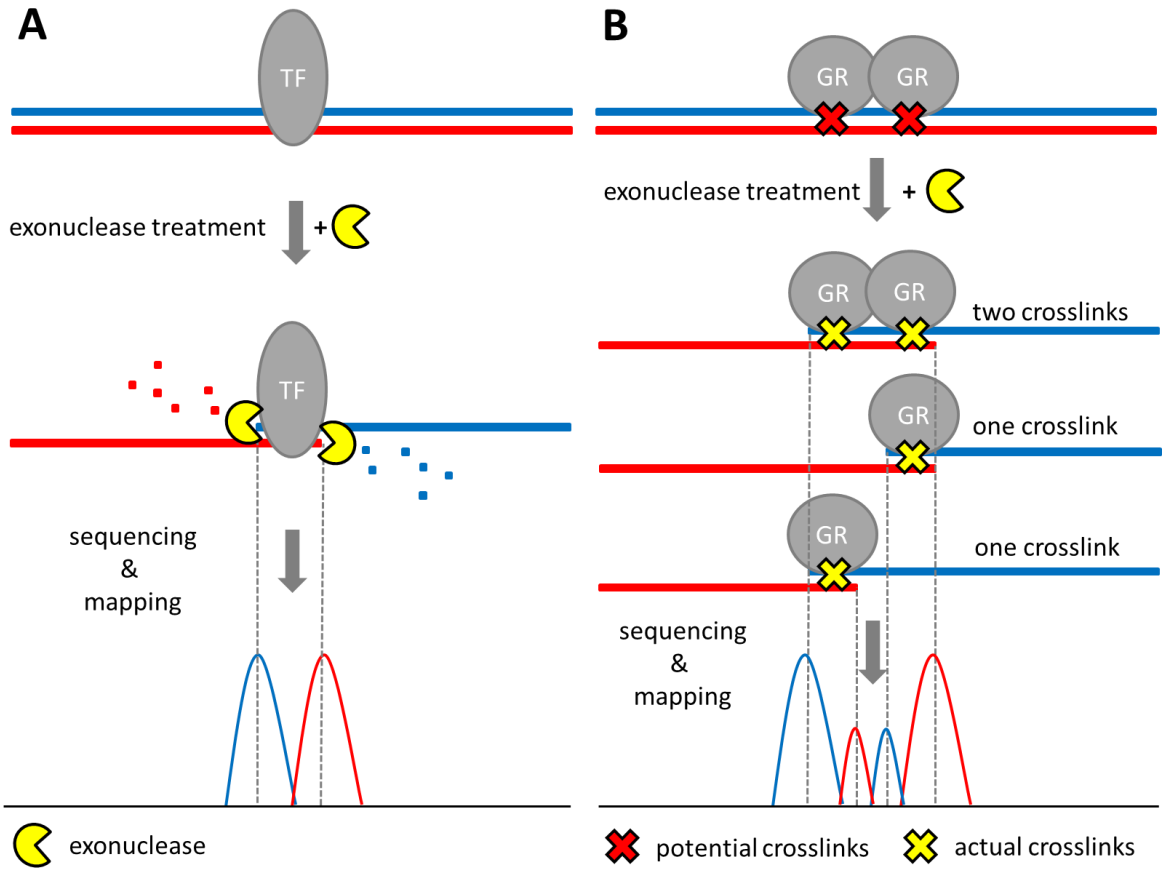


Figure 10: Model explaining the structure of footprints from ChIP-Exo data.
Immunoprecipitated genomic fragments with covalently bound TFs are treated with lambda exonuclease to degrade the 5´ ends of the fragments. Bound transcription factors form a barrier for exonuclease activity, thereby synchronizing the 5' ends adjacent to the cognate binding sites. Sequencing of the resulting fragments from the 5' ends and subsequent mapping of the sequencing start sites for both strands generates "footprints" at TF binding sites (**A**). The footprint for TFs that bind as homodimers (such as GR) generate footprints that differ from those that bind as monomers (**B**) For example, due to low crosslink efficiencies often only one of the two GR proteins is covalently attached to the DNA fragment. This results in two potential positions for both DNA strands where exonuclease activity is blocked from further processing, resulting in a special footprint pattern as observed for the canonical GBS (as shown in Figure 11 A).

Chromatin samples were prepared as described in Methods and send to Peconic LLC for chromatin immunoprecipitation using a GR antibody, Exo-processing and deep

sequencing. Bioinformatical analysis of the ChIP-Exo data was done in collaboration with Jonas Ibn-Salem and Dr. Morgane Thomas-Chollier. Briefly, genomic GR binding sites identified in ChIP-Seq were scanned for candidate sequence motifs from JASPAR using the RSA tool matrix-scan (quick version) with a p-value cutoff of 0.0001, and then centered on these sites. Next, the start sites from ChIP-Exo reads were plotted as read counts relative to the candidate motifs in a window +/- 30 bp around the motif (Figure 11).

The canonical GBS motif (Jaspar matrix MA0113.2) served as a positive control for a sequence that we know can be bound by GR and thus was expected to produce a footprint. The GBS is a palindromic sequence consisting of two inverted repeats separated by a 3bp spacer to which GR binds as a homo-dimer. Consistent with the literature[102] where DNase-I-footprinting has revealed an approximately 30 bp footprint for sites bound by GR, we found that the GBS is flanked by peaks on the plus and minus strand that are separated by approximately 30bp (Figure 11 A). In addition, the footprint shows some smaller additional peaks that can be explained by the fact that one of the GR monomers cross-links to the DNA (Figure 10 B). This could either be due to the fact that GR dimerization occurs on the DNA and thus initially one monomer interacts with DNA. Alternatively, the intrinsically inefficient formaldehyde cross-linking can result in only on half of the dimer covalently being attached to the chromatin. Together, the symmetry of the footprint is consistent with dimeric binding by GR to its canonical binding site indicating that the ChIP-Exo method is indeed capable of identifying sequences that recruit GR to the chromatin. It also shows that the low level of GBSs in GR-bound regions cannot be explained by an inability of GR to bind to such sites in Nalm-6 cells.

If the sequence motifs for PU.1 and RunX1 binding sites would indeed tether GR to the DNA, we would expect that these sequences should also leave a footprint in the ChIP-Exo experiment. To test this, the genomic regions bound by GR were scanned for the JASPAR motif IDs MA0080.3 and MA0002.2 using the same settings as for the GBS (MA0113.2). Unlike the GR, PU.1 and RunX1 bind their cognate binding sites as monomers and are thus expected to leave a footprint distinct from that seen for GR. Consistent with tethered binding by GR via Pu.1, we found a distinct footprint for the PU.1 binding motif (Figure 11 B) showing an enrichment of reads at position -11 relative to the motif center for the forward strand and at the position 13-14 relative to the motif center for the reverse strand. During the ChIP-procedure genomic fragments crosslinked to a GR protein are purified using a specific antibody. Thus, the footprint for the PU.1 motif which looks distinct from

the one seen for the GBS motif implies that the PU.1 sites are occupied by a transcription factor other than GR consistent with the idea that PU.1 might tether GR to the DNA.

In contrast to the GBS and the PU Box, the RunX1 binding sites do not show a footprint (Figure 11 C). This indicates that this site is unlikely to tether GR to the DNA, or that the crosslinking conditions that were optimized for GR do not efficiently crosslink proteins that bind this sequence.

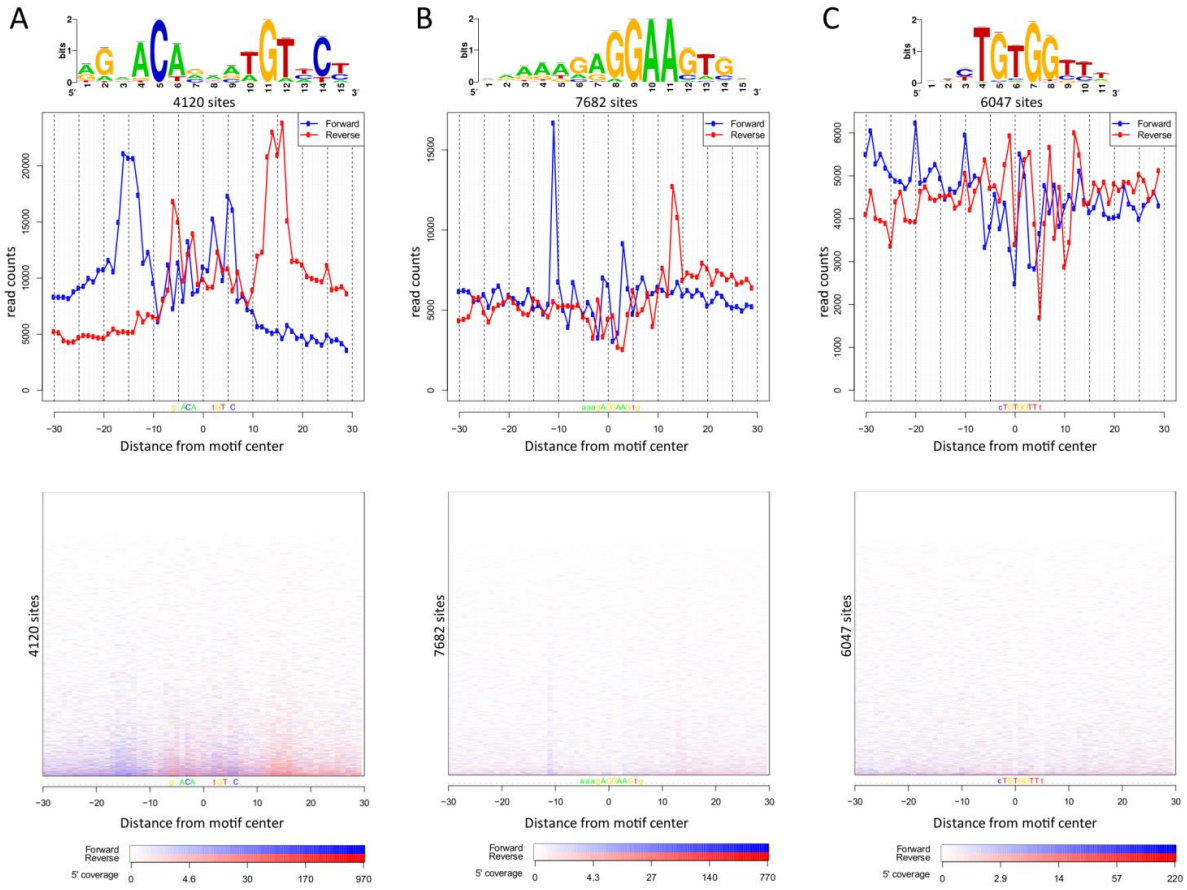


Figure 11: Different footprints at PU box motifs and GBSs indicate that the mode of GR binding differs for these motifs (Analysis performed by Jonas Ibn Salem).
GR-bound loci identified from ChIP-Seq data were scanned for PWMs (Jaspar) representing a canonical GBS (motif ID MA0113.2) (**A**); a PU Box (MA0080.3) (**B**) or a RunX1 binding sequence (MA0002.2) (**C**). The total number of sites matching these motifs is shown beneath each PWM. Aligned sequencing start sites from the ChIP-Exo experiment were plotted relative to motif center in 60 bp windows sites centered on identified sites. Upper panel: For each position the total number of sequencing start sites from the ChIP-Exo experiment was plotted for the forward and the reverse strand. Lower panel: Each genomic locus is represented individually as one row, the total number of sequencing start sites at each position for these loci is depicted by color intensity, the color indicating to which strand this start site matches.

5 Discussion part 1

To better understand how genes are regulated, we need further knowledge of signals that regulate accessibility of genomic regions to influence the binding of transcription factors. To date, prediction of genomic binding sites of transcription factors relies largely on knowledge of the genome-wide chromatin status for the cell line of interest. TF binding sites can be predicted by scanning accessible regions for cognate canonical binding sites. However, sites where TFs bind to DNA in alternative ways, such as tethering to other factors, are difficult to predict. For individual TFs of interest, it is therefore important to identify the factors that it can tether to. Using GR as a model transcription factor we (in cooperation with Dr. Morgane Thomas-Chollier) have shown, that its binding pattern in Nalm-6 is very different from the binding in U2-OS or A549 cells. For example, in Nalm-6 cells GR binding sites are more likely to be in close proximity to the transcriptional start site (TSS) of target genes than in the other cell lines. Another striking difference observed was that the canonical GBS is found less frequently at GR-bound loci in Nalm-6 cells than in U2-OS or in A549 cells. Focusing on Nalm-6, my aim was to understand how this special genomic binding pattern of GR is established.

One potential explanation for the decreased presence of canonical GBSs in Nalm-6-specific GR-bound loci is that in this cell line GR can also bind to more degenerated GBSs. Alternatively, this suggests that, in contrast to U2-OS and A549 cells, signals other than the GBS play an important role in regulating the genomic GR binding pattern in Nalm-6. One explanation is that sequence motifs other than the canonical GBS play a role in recruiting GR to its target loci in Nalm-6. Potentially these might encode binding sites for other proteins that serve as tethering platforms for GR. Enriched sequence motifs included sequences matching a RunX1 binding site, a PU-box and an E-box. RunX1 has been shown before to mediate indirect DNA binding (tethering) of the estrogen receptor[94]. Hence, RunX1 might also play a role in tethering other nuclear hormone receptors like GR, a nuclear hormone receptor closely related to the estrogen receptor.

Since GR binds predominantly to open chromatin[20], another mechanistic model to explain the GR binding pattern in Nalm-6 is that these sequence motifs recruit pioneering factors to open up chromatin to subsequently allow GR binding. Again, RunX1 is an interesting candidate for such mechanisms, as it has been described before to serve as a pioneering factor[148]. Also PU.1, another candidate protein that binds PU box sequence motifs, also has been described to serve as a pioneering factor[22,149].

The results from the luciferase assays suggest that GR binding may not only be facilitated by RunX1 opening up regions to make them available for GR binding as transiently transfected reporters likely do not restrict GR binding. Hence, if the function of the RunX1 site was to open up the region, its sequence might not be important within transiently transfected reporters. However, mutation in the RunX1 binding site within the WFS1a enhancer resulted in a decreased enhancer activity upon hormone activation. This indicates that the encoded RunX1 binding site is important for enhancer function by mechanisms other than its pioneering activity to increase accessibility. Mutation of the E-box motif had similar effects on enhancer activity as those observed for the RunX1 binding site. In addition, simultaneous mutation of both sites even completely inactivated enhancer activity, suggesting that both these sequence motifs might recruit GR via tethering by associated proteins. However, the observation that deletion of the canonical GBS also completely inactivated enhancer activity argues that the GR-dependent regulation at the WFS1a locus requires a complex interplay between the GBS, E-box and RunX1 binding motif. Possibly, proteins that bind to the E-box and RunX1 sequences tether GR to this genomic locus which in turn increases the local GR concentration which subsequently increases the likelihood of GR to bind to the cognate canonical GBS.

Candidate proteins for binding to the enriched sequence motifs are RunX1, PU.1 and TFAP4. To test the role of these proteins in facilitating GR-dependent transcriptional regulation, I set out to reduce their levels using RNAi knockdown. Unfortunately however, I was neither able to introduce an efficient knockdown using the psiRNA system (Invivogen) nor by esiRNA (Sigma Aldrich) (data not shown) and therefore was unable to study their role in GR-dependent transcriptional regulation in Nalm-6 cells.

According to the results from the lumier assay, GR interacts with two out of three candidate proteins (TFAP4 as a candidate to bind the E-box motif, and RunX1). Whether GR interacts with PU.1 is unclear due to technical difficulties and thus theoretically, all candidate proteins might help to recruit GR to the genome. However, even if these proteins have the potential to form protein-protein interactions, this does not allow conclusions if such an interaction facilitates recruitment to the locus, or if it tethers GR to the locus. One requirement for these proteins to be involved in tethering of GR is, that they bind genomic target sites simultaneously with GR. Using ChIP-Exo where GR was the target of the immunoprecipitation I found a striking footprint profile for the PU-box motif which looked different in shape from the profile for conventional GR-bound GBSs. This indicated that

another protein is bound to such DNA sequences, probably PU.1, and that this protein tethers GR to the DNA. In contrast, the other enriched motifs did not yield a footprint profile. This could either reflect the fact that they are not capable of tethering GR to the DNA or alternatively it could reflect technical limitations of the ChIP-Exo assay.

Another important observation from the ChIP-Exo data is related to the possibility that GR might be able to bind to degenerated GBS sequences in Nalm-6 which could explain the apparent absence of GBS-like sequences in GR-bound regions. As shown by the clear footprint in Figure 11, GR is able to bind to the canonical GBS. However, surprisingly, more degenerated GBSs within GR-bound regions in Nalm-6 failed to produce a footprint, whereas they did show a footprint in U2-OS cells. This indicates that the virtual absence of GBSs in GR-bound regions in Nalm-6 cells cannot be explained by binding to more degenerate sequences, consistent with the idea that other factors such as PU.1 might tether GR to the DNA.

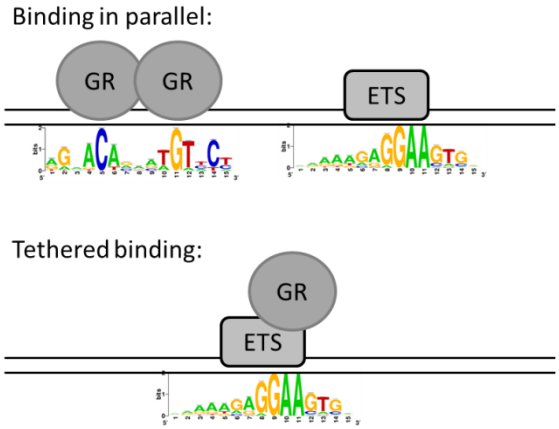


Figure 12: Potential DNA binding modes implicated from ChIP-Exo data.
Results from ChIP-Exo (see Figure 11) indicate that GR binds to genomic DNA either directly together with an ETS transcription factor (likely PU.1) in a synergistic binding mode (upper panel), or that ETS transcription factors tether GR to genomic DNA independent of an encoded canonical GBS (lower panel).

5.1 Future directions

Apart from the reporter with the active enhancer, none of the 12 additional GR binding regions I tested showed GR-dependent enhancer activity. One potential reason might be the SV40 minimal promoter used in these reporters. In the past, inconsistent findings were described for the activity of the SV40 promoter in Nalm-6 cells. Some studies describe that this promoter is not well suited to drive expression in Nalm-6 cells[150], whereas they were successfully employed to drive expression in other studies[151]. Hence, the lack of activity for most of the reporters tested might be due to the inactivity of the SV40 promoter used.

Thus, future studies for instance to study the rule of the PU.1 binding site might benefit from re-testing these reporters using an alternative promoter.

Although my studies suggest that tethered binding of GR by PU.1, Runx1 and TFAP4 might play a crucial role in recruiting GR to genomic binding sites in Nalm-6 cells, additional experiments are needed to show this convincingly. For example, co-immuno-precipitation experiments between GR and PU.1 could determine if these two proteins interact, a prerequisite for tethered binding. Similarly, DNA-pull down assays using the PU.1 binding sequence as bait and subsequent probing for GR enrichment could provide further evidence for PU.1-dependent tethered binding of GR. In addition, since I was unable to knock-down the expression of candidate proteins using RNAi, I could try to study their role in recruiting GR to the genome by knocking them out using genome editing techniques, like the CRISPR/Cas9 system [152].

5.2 Summary

My results indicate that the observation that canonical GBSs are less enriched in Nalm-6 specific GR-bound loci is not a result of GR binding to more degenerate GBSs. This suggests that GR binding at loci without canonical GBSs is facilitated by other mechanisms, likely tethered binding through other proteins. Accordingly, GR binds to several candidate proteins that occupy such sequences, a prerequisite for tethered binding. Finally, ChIP-Exo data showed a very clear footprint for the PU.1 sequence motif that looked distinct from the footprint left by DNA-bound GR. This is consistent with another protein, likely PU.1, binding that indirectly tethers GR to the DNA. In summary, although incomplete, these studies suggest that a broader spectrum of proteins than previously though is capable of tethering GR to the DNA. The tissue-specific expression of such tethering factors, like PU.1 which is mostly expressed in hematopoetic cells, could also explain the distinct genomic binding profile of GR observed in different tissues

6 Results part 2: Sequence encoded signals that restrict genomic TF binding

6.1 Candidate signals to restrict TF binding are depleted in ChIP-Seq peaks

Analysis of GR ChIP-Seq data by Dr. Morgane Thomas-Chollier has shown that in addition to the enrichment of binding sites for GR and co-factors at the center of bound genomic loci, other sequence motifs are depleted in these regions as shown in Figure 4. One explanation for the depletion of sequence motifs is that the presence of binding sites for GR is mutually exclusive with the presence of other sequence motifs. To test this, the dataset of genomic binding sites was subdivided into two sets, those with a canonical GBS within a 300 bp window surrounding the peak summits and those without. If the observed depletion for sequence motifs is caused by enrichment of the canonical GBS, this depletion should be lost or be weaker in the subset of peaks without a canonical GBS. What we found however, was that the level of sequence motif depletion for these two subsets of binding sites is similar (Figure 13 A), with a tendency to be even stronger for the subset of binding sites without an encoded GBS. Another possible explanation for the observed depletion is that they might be caused by an artifact from the experimental procedure, as AT-rich sequences tend to be depleted during the ChIP-Seq procedure[153]. If this were true, any AT-rich sequence would be expected to be depleted which is not what we observed. Moreover, with approximately 40 % GC content the base composition of the ChIPed material was not significantly different from composition of the genome arguing against such a bias being the reason for the depletion observed in our data.

If the depleted motifs indeed play a role in preventing GR binding to nearby GBSs, one would expect that the depletion we observed should be specific for bound GBSs and not (or to a lesser degree) be observed at GBS-like sequences that are not bound by GR. To test this, we generated two sets of unbound genomic loci (roughly containing as many genomic regions as for the bound set examined). The first set consisted of randomly chosen genomic regions with GBS-like sequences that are not bound by GR in any of three cell lines examined (A549, U2OS and Nalm-6) (random GBS). The second control set was generated from GBSs that are encoded two to six kb away from the centers of ChIP-Seq peaks (unbound GBSs). The rationale for this control set was that we expected that such regions are likely to have a similar chromatin state as the real datasets. For both control sets, 8 kb windows were centered on the assigned GBS, with the canonical motif being randomly assigned in a 300 bp central bin to mimic the spread of a real dataset.

When these two control sets were scanned for the frequency of the depleted sequence motifs a very small depletion was detectable at the center of the control sets. This depletion however was much smaller than that observed for the real dataset (Figure 13 A). One possible explanation for this small depletion is that the control sets likely include GBSs bound by GR in other cell lines that were not included in our analysis. Another explanation for the small depletion observed is that the depleted motif cannot be present at the exact position where the GBS is located. If this is indeed the reason, this further underscores that the depletion observed for the real dataset (which is much larger) cannot simply be caused by being mutually exclusive with the presence of a GBS.

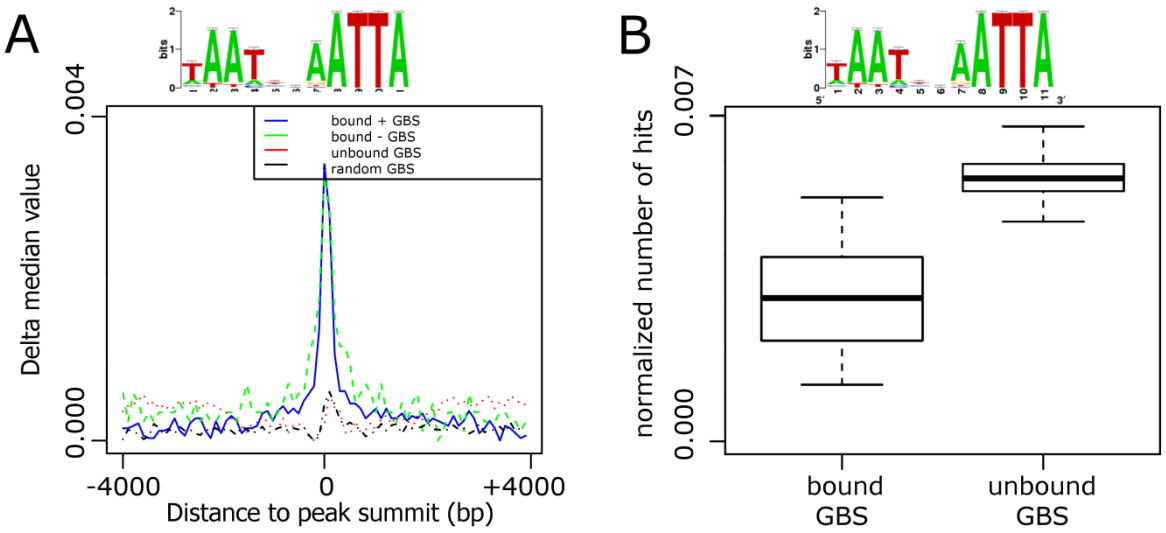


Figure 13: Depletion of AT-rich candidate sequences is specific for bound GBSs (analysis by Dr. Morgane Thomas-Chollier).
GR-bound genomic regions encoding a canonical GBS (bound + GBS) or not encoding a GBS (bound - GBS) within a 300 bp window around the ChIP-Seq Peaks (bound GBS) were compared to random GBSs at least 2 kb away from the closest ChIP-seq peak center (unbound GBS) or random umbound GBSs anywhere in the genome. For each set, the difference in mean frequency of hits is shown relative to the bin with the highest value for each set. High values indicate a strong depletion at the corresponding bin (**A**). For bound and unbound GBSs, the frequency of hits for the tested PWM for the central bin is shown as normalized number of hits in (**B**).

6.2 Conservation of depleted sequence motifs

If the depleted AT-rich sequences indeed play a functional role in restricting transcription factor binding to specific parts of the genome, their function and sequence might be conserved across species. We therefore studied the conservation of four depleted sequence motifs of interest (together with Stefanie Schöne; Figure 14 A-D). As control, we included two sequences with similar AT content (Figure 14 E-F). For the analysis we scanned the genome for the CSMs, and then generated 10 subsets of 500 sites encoding these sequences. For each subset, we calculated the median PhyloP score for each base in a

window of 100 bp centered on the motif. For each sequence tested, the average of the median values for each position is plotted in Figure 14 A-B and D-F.

The sequence of CSM 3 only had 389 exact matches in the genome (excluding repetitive sequences). Hence, for this sequence motif we modified the analysis and plotted the average and mean PhyloP score distributions for a window of 50 bp centered on the motif (Figure 14 C).

The PhyloP score for the regions surrounding all sequences tested was similar, indicating a similar evolutionary conservation of the genomic regions. For CSM 1, we observed increased PhyloP scores of up to 0.25 for the two stretches of three thymidines, which drops for the central guanosine to a negative value. Similarly, CSM 3 also showed increased PhyloP scores only for some of the bases of the tested sequence motif, and overall, the score distribution for this sequence motif shows more fluctuation than observed for the other motifs analyzed. Likely, this is caused by the fact that less genomic regions contribute to the analysis done for CSM 3 than for the other sequences tested, and that no averaging of multiple median values could be done for this sequence motif. CSM 2 shows an increased value for the PhyloP score (0.2 – 0.25) for every base of the sequence, indicating that its complete sequence is more conserved than its surrounding region. CSM 4 shows an increase in the PhyloP score only for the five central thymidines, but not for the two flanking adenosines.

Compared to the depleted sequences analyzed (CSMs 1-4), the control sequences show lower median PhyloP scores. Furthermore, when an increase is observed (control sequence 1) it does not span the whole sequence in contrast to the block of conservation observed for CSM 2. Further, the pattern of the PhyloP score distribution for sequences of CSM 1 and control 1 are similar, with the amplitude being smaller for control sequence 1. The results for these sequences indicate that stretches of three or more adenosines or thymidines in a row tend show increased PhyloP scores. This would also explain the conservation score distribution for the CSM 4, for which an increased conservation was detected only within the central thymidines, not for the single flanking adenosines.

Together, the conservation analysis showed a higher degree of conservation for the CSMs than for their surrounding and the control sequences. Furthermore, for CSM 2 the conservation covers the complete sequence. Although alternative explanations are possible,

the observed conservation is consistent with a functional role of these depleted sequences that is conserved across species.

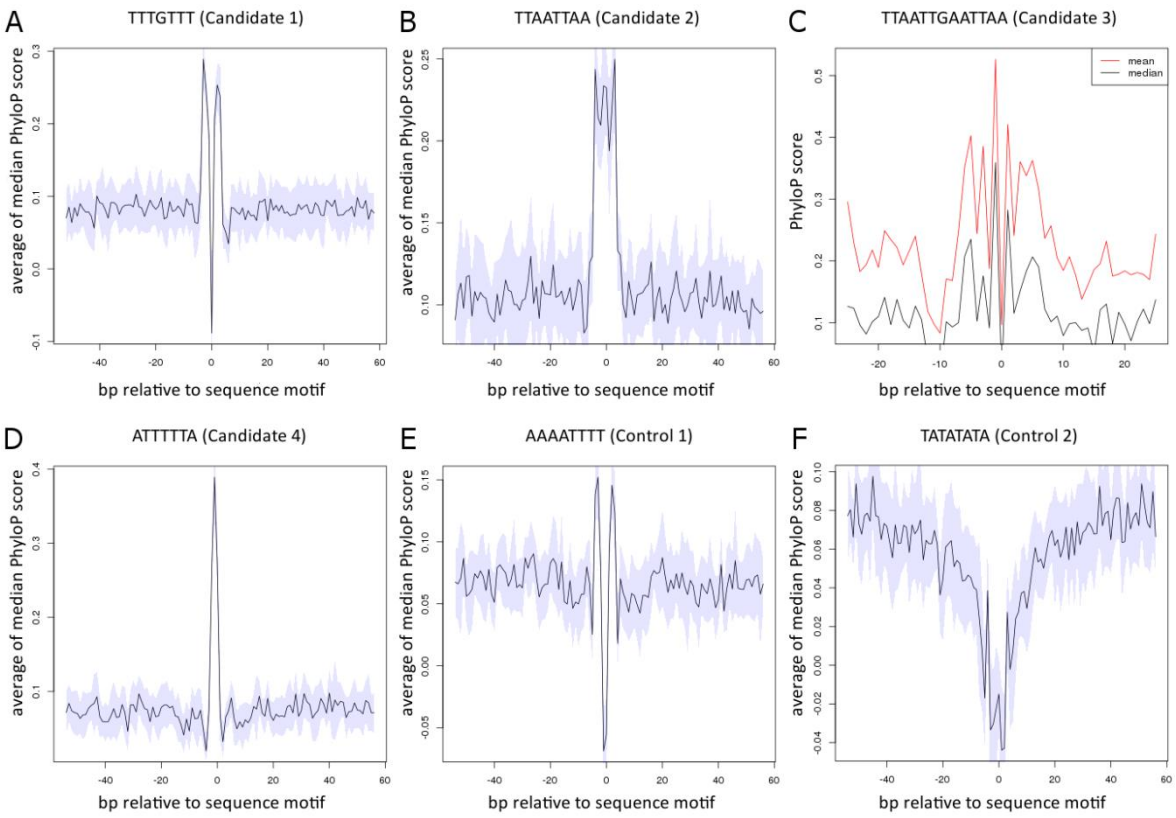


Figure 14: Genomic DNA sequence conservation of AT-rich sequences (Analysis by Stefanie Schöne). (**A**),(**B**),(**D**)-(**F**) Genomic DNA was scanned for candidate or control sequences as indicated. From all matching genomic sites, 10 subpopulations of 500 sites in total were generated. These sites were centered on the corresponding motifs in 100 bp windows and the median of the average PhyloP score of the 10 subpopulations is shown by the black line. The grey area indicates the standard deviation. (**C**) The sequence of candidate 3 only has 389 exact matches encoded in the genome and therefore it was not possible to generate 10 subpopulations of 500 individual matches. Hence, for candidate 3 the average and mean PhyloP score for each position is shown for all genomic sites.

6.3 Candidate sequences have the potential to decrease GC induced transcription

To determine if these AT-rich CSMs are indeed functional and not simply caused by an experimental bias, I wanted to test if their presence interferes with transcription factor binding. As a proxy for reduced GR binding, we first determined if the presence of a depleted sequence influences the ability of GR to activate transcription. For these experiments, I generated luciferase reporters that encode one of four different GR binding site sequences (FKBP5, PAL CGT and SGK as described[103]) upstream of a minimal promoter driving the expression of a luciferase reporter gene. These GBSs were flanked by sequences encoding the depleted AT-rich CSMs as shown in Figure 15 (CSM #1: TTTGTTT; CSM #2: TTAATTAA; CSM #3: TTAATTCAATTAA).

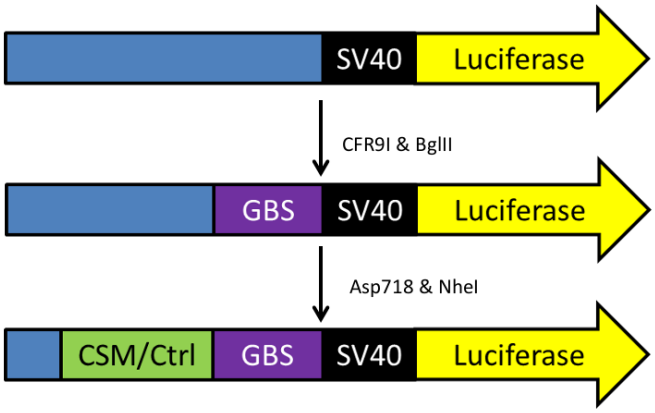


Figure 15: Generation of NRS reporters based on the plasmid pGL3-Promoter.
pGL3-Promoter was linearized with CFR9I and BglII, and the short fragment between the restriction sites replaced by a DNA oligonucleotide-linker that encoded one out of six different GBSs tested to produce the reporters pGL3-Promoter-'GBS'. These GBS-encoding reporters were linearized with Asp718 and NheI, and the short fragment between the restriction sites replaced by DNA linkers encoding the CSMs for testing.

Next, these luciferase reporters were tested in U2OS cells. I first determined the activity of the GBS reporters in the absence of CSMs. As expected, for all reporters a hormone-dependent activation was observed with varrying degrees of activation depending on the sequence of the GBS (Figure 16). Next, I examined the effect of flanking the GBSs by one of the CSMs. With CSM #1, no reduction of hormone-dependent activation was observed. In contrast, when the GBSs were flanked by CSM #2 or #3 a significant decrease of reporter gene activation by approximately 50 % was observed. This effect was seen for all four GBSs examined *(*Figure 16 C*)*. Notably, the reduced reporter gene appeared to be specific for the hormone-dependent activation as the basal activity of the reporters did not change when either CSM #2 or #3 were present (Figure 16 A).

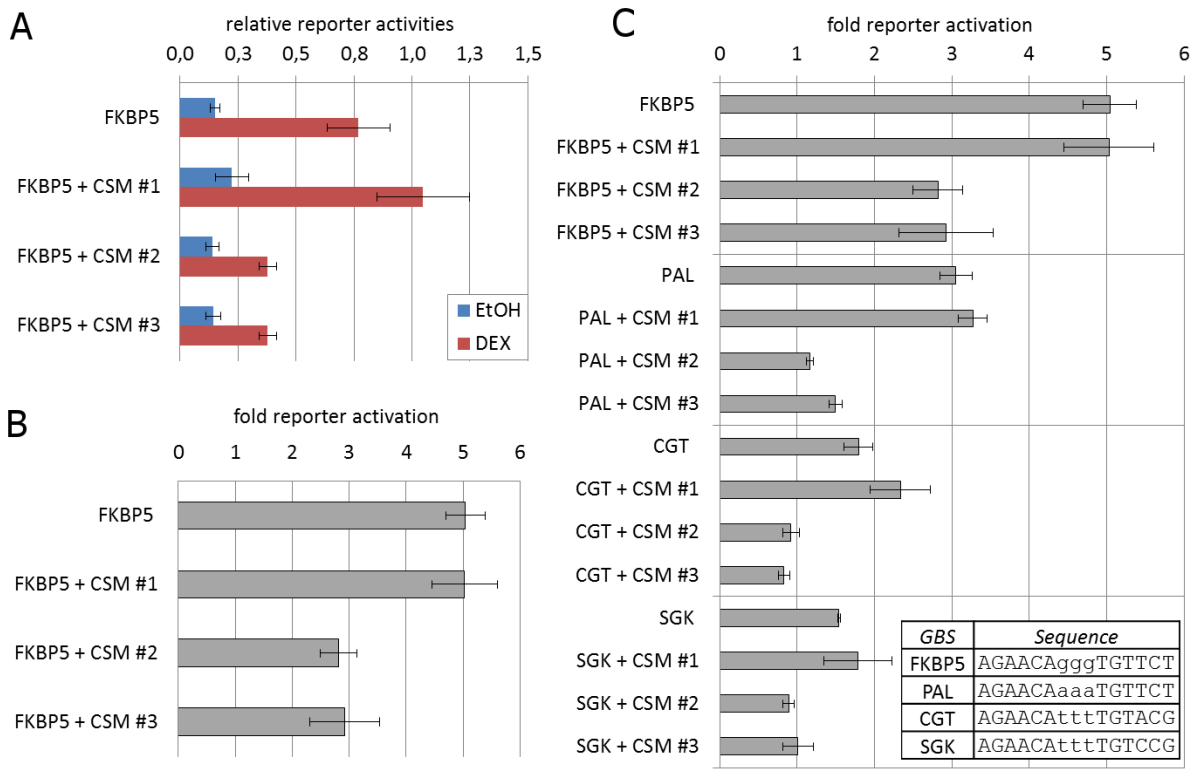


Figure 16: Influence of CSMs on reporter gene activation.
Reporter plasmids encoding one of four different GBS sequences (as described before [103] and shown (**C**)) alone or in combination with a CSM were tested in luciferase assays using U2-OS cells. For the FKBP5 reporters, the normalized relative reporter activity for ethanol and dexamethasone-treated cells is shown in (**A**). The ratio of relative reporter activities for dexamethasone and ethanol treated cells is shown as fold reporter activation in (**B**). The experiment was done also with reporters encoding other GBSs as shown in (**C**). For these, only the fold reporter gene activation upon dexamethasone treatment is shown. Experiments were done in three independent biological replicates using U2-OS cells, averages ± standard error are shown.

Because CSM #1 did not reduce reporter gene activation I decided to use CSM #1 as additional control for further experiments. Based on the observed effects for CSM #2 and CSM #3, I called these two *Negative Regulatory Sequences* (NRS).

CSM #1:	TTTGTTT	= Control
CSM #2:	TTAATTAA	= NRS 1
CSM #3:	TTAATTCAATTAA	= NRS 2

6.4 Influence of random AT rich sequences and the relative position to the GBS

The observed effects for the NRSs are consistent with the hypothesis that some of these depleted sequences negatively influence DNA-binding by GR. A shared feature of these NRSs is that they are AT-rich raising the question whether a specific sequence motif is responsible for the effects observed for any AT-rich sequence. To test this, we flanked the

GBS with the highest activity (FKBP5 sequence as shown in Figure 16) with various AT-rich sequences to study their effect.

6.4.1 Design of additional controls for testing of depleted sequence motifs

To rule out effects of subtle changes in the length of the inserted sequences, I modified the control reporter to produce a luciferase reporter with an insert of the same length as the reporter GBS-NRS2, thus differing only in the bases encoding the NRS. This reporter was used as reporter GBS-Ctrl #1 for further experiments (see table of *Figure 17)*. CSM #1 that did not reduce reporter activity (TTTGTTT) will from now on be referred to as control #2. The reporters GBS-Ctrl #1 and GBS-Ctrl #2 plus the reporters GBS-NRS 1 and GBS-NRS 2 were now tested in luciferase assays in three independent biological repeats (Figure 17 A). As observed before (chapter 6.3), dexamethasone-induced reporter gene transcription using GBS-Ctrl #2 is even stronger than the induced transcription for other control sequences. In contrast, activation is reduced for NRS 1 and NRS2 to approximately 50% of the induction observed with GBS-Ctrl #1. Together, these experiments indicate that the effects observed in the initial experiments were not caused by differences in reporter length for the reporter GBS-control but instead caused by the NRS sequences.

6.4.2 Testing of other AT rich candidate sequences

To test if the effects for the NRSs were specific for the sequences tested or a typical feature of AT-rich sequences, I tested the influence of 7 additional A/T rich sequences plus another depleted CSM identified in the ChIP-Seq data analysis ((ATTTTTA), introduced as CSM #4 in chapter 6.2) on reporter gene activity (sequences tested shown in Figure 17 C). Interestingly, the additional AT rich sequences as well as the additionally tested CSM ATTTTTA showed little to no reduction in reporter activity. Of all tested CSMs, only the reporter genes encoding NRS 1 and NRS 2 showed a significant reduction in reporter gene activation compared to GBS-Ctrl #1 according to a two-tailed t-test assuming equal variance ($p < 0.05$). This clearly shows that the effects of NRS1 and NRS2 are not simply a consequence of their AT-richness, but that only specific sequence motifs are able to interfere with GR-dependent transcriptional activation of the reporter gene.

6.4.3 Testing the influence of the relative position of NRSs to the GBS

One possible explanation for the effects observed for the NRSs is that they might recruit and bind proteins that in turn prevent GR from interacting with the encoded GBS by spatial hindrance. If this were the case, one would expect that the effect might be lost when the

NRS is moved to the opposite face of the DNA by increasing the spacing by 5 bp, approximately half a helical-DNA turn. To test this, I modified the reporter genes as described above by incorporating five additional nucleotides between the GBS and the NRS/control sequence. Similar to the observations with the initial reporters, again we saw a marked reduction of the GBS-5bp-NRS reporter activity in the presence of both NRS1 and NRS2 with NRS2 having the strongest effect (Figure 17 B).

Taken together, these results indicate that NRSs do not mediate their effects by simple sterical hindrance, as addition of half a helical turn did not affect the function of NRS 2 and only partially decreased the effects of NRS 1. In addition (see chapter 6.8), the effects of NRSs were also seen when tested in zebrafish using reporters with multiple GBSs and a different spacing between these individual GBSs and the NRSs.

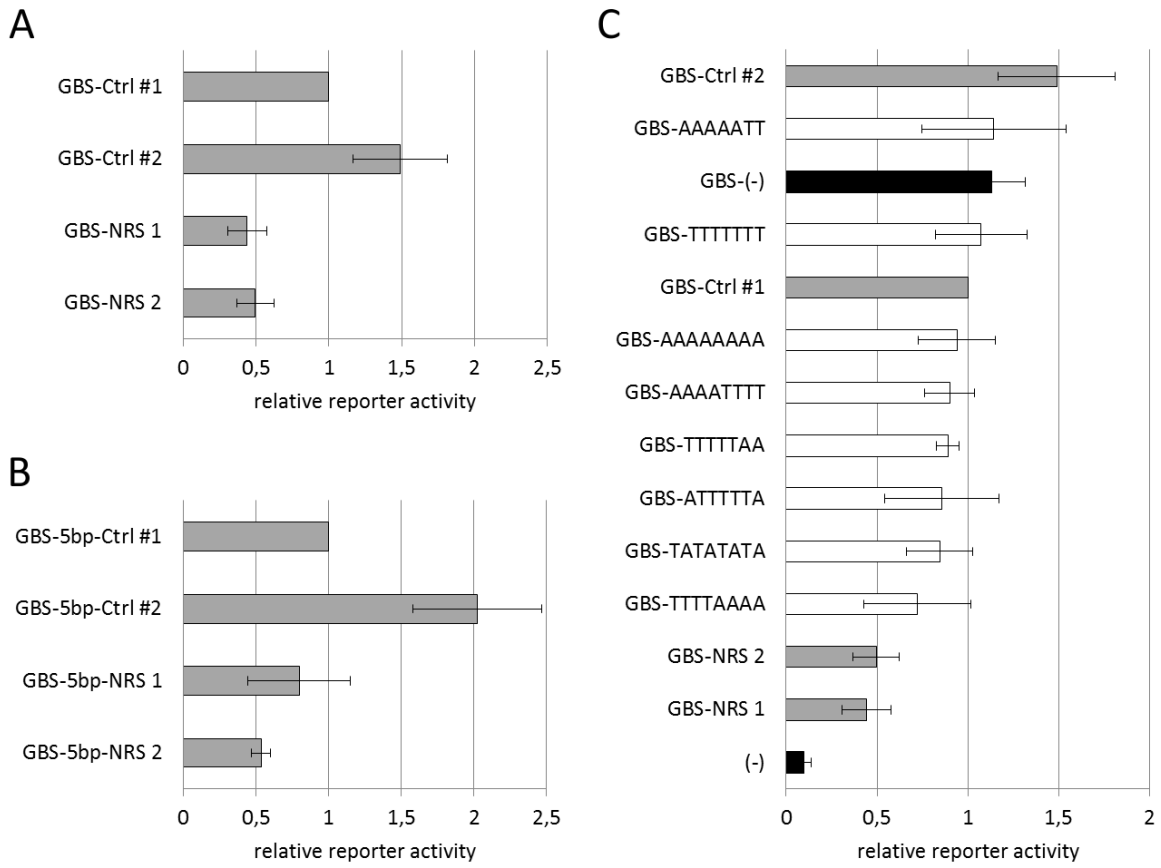


Figure 17: NRS activity is sequence-specific and independent of its exact position relative to the GBS. The reporters encoding either the control or one of the NRSs (as used in chapter 6.3) were tested in luciferase assays in parallel with a new control reporter gene (GBS-Ctrl #1) where a stretch corresponding to the AT-rich NRS/control sequence was replaced by a random sequence (**A**). Additional luciferase reporters with a 5 bp shift were tested in luciferase assays, results for fold activation are shown in (**B**). Modified versions of the NRS 2 reporter to encode stretches of other AT-rich sequences next to the GBS were tested in luciferase assays; the sequences tested are indicated in the labels of the diagram. (-) corresponds to the reporter gene plasmid without GBS (empty pGL3-Promoter) for which no induction was expected (**C**)**.** All experiments were done in triplicates and fold activation upon dexamethasone treatment is shown normalized to the activation of the corresponding Ctrl #1 reporter ± standard error of the mean.

6.5 Effect of NRS sequences on other transcription factors

The CSMs tested in this thesis were derived from ChIP-Seq data for the GR. Hence it may be that their activity is restricted to GR or to nuclear hormone receptors in general. Alternatively, the NRSs might be able to interfere with transcription factor function in general. To assess the effects of NRS on transcription factors other than GR, I tested their influence on MyoD-driven transcriptional initiation. For these experiments, I generated luciferase reporters encoding either an NRS or control sequences directly adjacent to three MyoD binding sites in the plasmid DLO-Luc as shown in the pictogram in Figure 18 A. These reporters were then used for luciferase assays. U2OS cells show no to very little MyoD expression[154,155], a muscle-specific transcription factor. Hence, the transfection scheme was modified. Instead of an expression construct for GR, either an empty vector (to assess the basal activity of the reporter) or an expression construct for MyoD was co-transfected and cells were not hormone treated as the MyoD TF is constitutively active. As observed before (chapter 6.4), the transcriptional activity in the presence of control sequence #2 was higher than in the presence of control sequence #1 (Figure 18 A). In contrast to the results with GR, no effect on MyoD driven transcriptional activation was observed for NRS 1 whereas NRS 2 very effectively inhibited MyoD-driven transcriptional activation (Figure 18 A).

The experiments on the effects of NRSs on MyoD-driven transcription were done using another luciferase reporter gene encoding plasmid (DLO-luc) than in previous experiments testing the effects on GR driven transcription (pGL3-Promoter). Thus, it may be that the lack of effect for NRS 1 when tested with MyoD is reporter specific or alternatively that this NRS has TF-specific effects. To distinguish between these two scenarios, we tested the effects of NRSs on GR-driven activity in the DLO-luc reporter. For these experiments, I cloned two GBSs upstream of the promoter next to one of the control or NRS sequences as shown in the pictogram in Figure 18 B. Basal activity for all four of these reporter genes was similar. Similar to the observation with the other reporters, both NRSs reduce the hormone-induced transcriptional activation to approximately 50% of the level observed for the reporter encoding the control sequences. This indicates that NRS1 selectively interferes with GR-driven transcription whereas NRS2 appears to interfere with TF-activity in general.

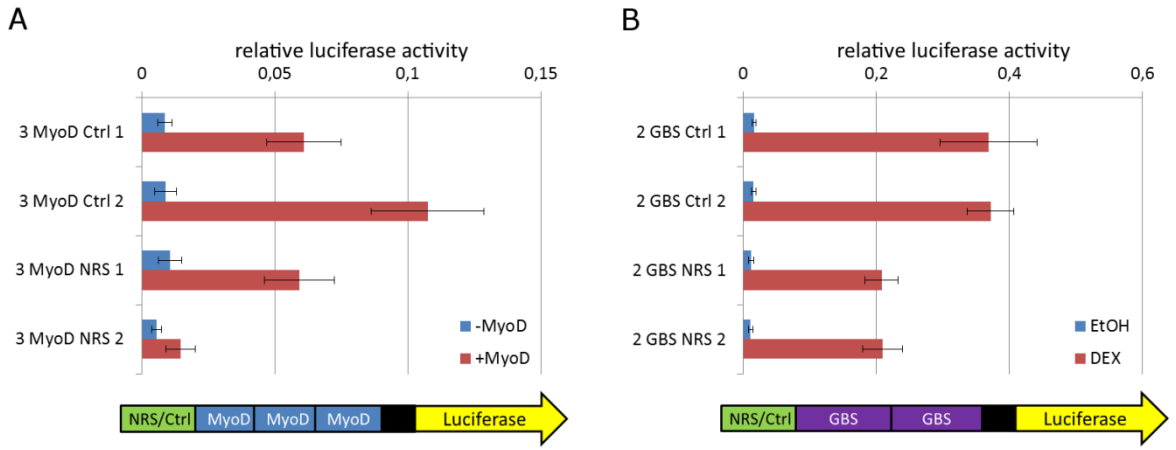
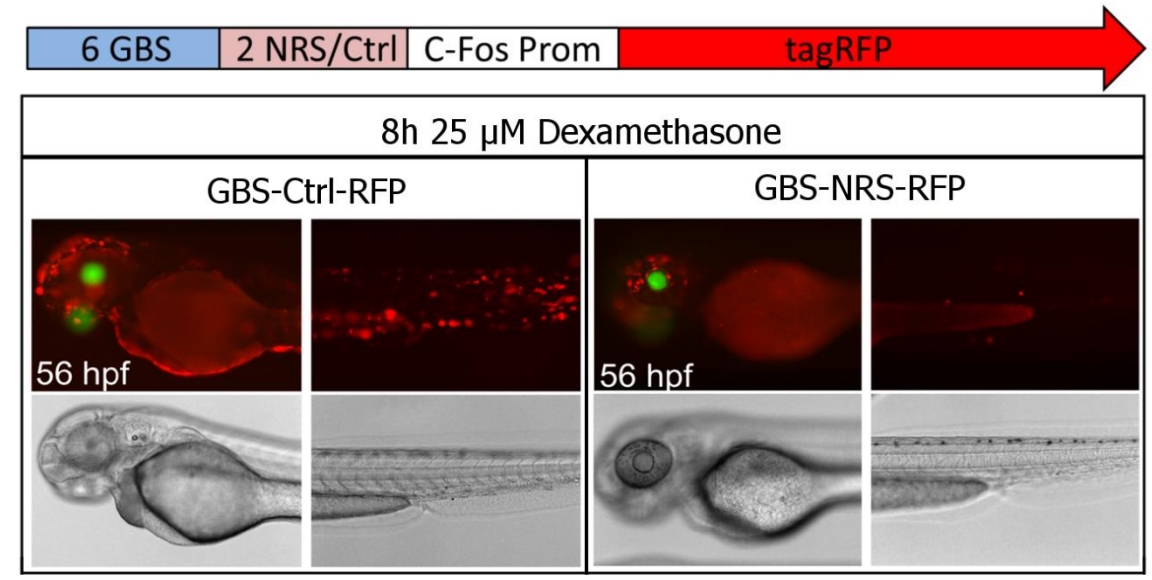
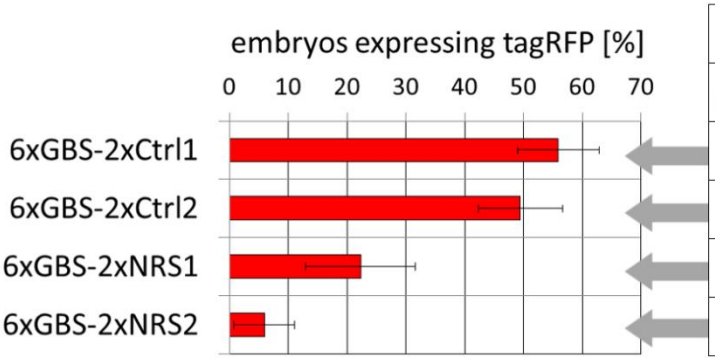


Figure 18: The effects of NRSs are not restricted to GR.
Reporter genes encoding 3 MyoD binding sites adjacent to one of the NRSs or control sequences were tested in luciferase assays using U2-OS cells. These were either co-transfected with empty or MyoD-coding expression vectors as indicated. For all conditions, the normalized reporter gene activity in either the presence of absence of MyoD is shown (**A**). To test if the effect of NRSs is recapitulated in the DLO Reporter plasmid, I tested luciferase reporters, encoding two GBSs (TAT) adjacent to a candidate or control sequence, as shown (**B**). Experiments were done in three biological replicates, shown are averages ± standard error.

6.6 Effect of NRSs are conserved across species

To test if the effects detected in mammalian cells are conserved across species and to test for potential cell-type specificity, I decided to test the effects of NRSs in *Danio rerio* (zebrafish). Zebrafish have functional glucocorticoid receptor signaling[156] and have diverged from humans some 450 million years ago[157]. I generated reporter genes similar to the luciferase reporter genes I used before to test the effects of the NRSs. For these experiments however, I modified a reporter gene that was designed by Dr. Sergey Prykhozhij (GAB-6xGBS-cFOS-TagRFP_pDest_cry). This reporter encodes a red fluorescence protein (RFP) driven by a weak c-Fos promoter. Upstream of this promoter six GBSs are encoded driving RFP expression upon hormone treatment to activate GR. The reporter is designed for random genomic integration using the Tol2 transposase system[126]. Adjacent to the GBSs, I inserted either two control or two NRS sequences. Together with Dr. Sergey Prykhozhij, I then tested these reporters by injecting them into fertilized eggs at the one-cell stage. 48 hours post injection the embryos were treated for 8 hours with dexamethasone or DMSO as vehicle control. Similar to the observations made with mammalian cells, the presence of NRS sequences efficiently reduced dexamethasone-induced transcription of the reporter gene, indicating that the effects of NRSs are conserved across species (Figure 19, Figure 20). Furthermore, the effect was seen throughout the animal and appeared not to be restricted to specific cells or tissues, arguing that the effects of NRSs are not cell-type specific. Visual quantification of dexamethasone-induced RFP expression of transgenic fish showed that 55.9 % of transgenic fish with the control #1 reporter showed expression of the RFP reporter (N_{total} = 92; $N_{positive}$ = 51.5); 49.5 % of transgenic fish with the control #2 reporter (N_{total} = 109; $N_{positive}$ = 54); 19.0 % of transgenic fish with the NRS 1 reporter (N_{total} = 92; $N_{positive}$ = 17.5) and 4.3 % of transgenic fish with the NRS 2 reporter (N_{total} = 93; $N_{positive}$ = 4).





	# of transgenic Fish				TagRFP positive (total)
	Exp 1	Exp 2	Exp 3	total	
6xGBS-2xCtrl1	20	26	46	92	51,5
6xGBS-2xCtrl2	32	38	39	109	54
6xGBS-2xNRS1	15	42	35	92	17,5
6xGBS-2xNRS2	16	36	41	93	4

Figure 19: NRS activity is conserved in zebrafish and active throughout the animal.
RFP reporter plasmids were randomly integrated into the zebrafish genome by Tol2 transgenesis, and dexamethasone-induced RFP expression was evaluated by microscopic analysis as described in chapter 3.12. Representative fish for RFP activity in dexamethasone-treated fish for control and NRS reporters are shown in the upper panel. For all four reporter plasmids the injections were done three times. The number of hormone-treated transgenic fish for each reporter and experiment is shown in the table (bottom right) together with the total number of transgenic fish from all three experiments and the corresponding number of RFP-positive fish according to visual quantification. The diagram (bottom left) represents the average [%] of transgenic embryos with RFP expression for each reporter tested, averages ± standard deviation are shown.

Visual quantification is subjective and therefore we also quantified hormone-induced activation of the TagRFP mRNA. In previous experiments, Dr. Sergey Prykhozhij found that the basal expression of TagRFP is lower when an e1B promoter is used instead of the c-Fos promoter (unpublished data). To circumvent potential problems of high basal activity, we swapped the c-Fos promoter in the reporter plasmids against an e1B promoter. Other than that, the reporters used were identical to those used before. The modified reporters recapitulated the effects observed with the c-Fos driven reporters (data not shown).

Next, we quantified the TagRFP mRNA levels for DMSO versus dexamethasone-treated fish using quantitative real time PCR (qPCR) for all reporter constructs in three

independent biological replicates. To determine if the effects observed were specific for the reporters investigated, we also examined the endogenous GR-target gene FKBP5 which was expected to be regulated similarly regardless of which reporter was integrated. Consistent with expectation, we observed a robust and reproducible hormone-dependent activation of the FKBP5 gene that was comparable for all fish injected with the control reporter and those injected with the NRS reporter (Figure 20). Interestingly, the level of transcriptional activation showed high variability between the experiments, with values from 10 to more than 100-fold increases in the expression level of FKBP5 for dexamethasone-treated fish. We also examined changes in eCFP expression (the co-injected marker gene to identify transgenic fish) upon hormone treatment and as expected eCFP levels were not affected by treatment with glucocorticoids. In contrast and paralleling what we saw visually, the expression of TagRFP increased 4-fold for the control reporter whereas this induction was almost completely lost in presence of NRS 2 (1.3-fold) (Figure 20). This significant loss of TagRFP induction (p<0.01, two tailed t-test assuming equal variance) in the presence of NRS2 validates our visual quantification and shows that the effects of NRSs are conserved across species and appears to occur regardless of the tissue examined.

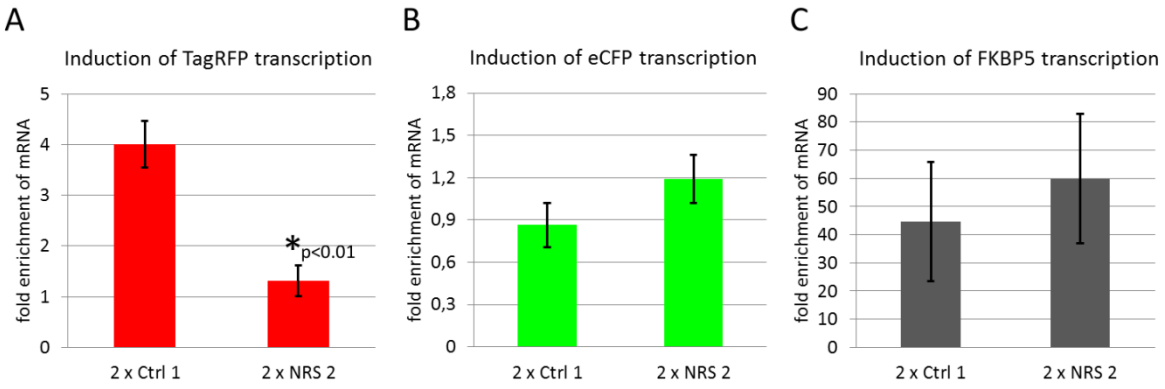


Figure 20: NRSs prevent hormone-induced transcription in zebrafish.
cDNA prepared from transgenic fish (NRS 2 or Control 1 reporter genes in combination with e1B promoter) after dexamethasone/DMSO treatment was used for qPCR analysis. Shown are the ratios of dexamethasone versus DMSO treated fish for mRNA of TagRFP (**A**), eCFP as a control for transgenesis efficiency (**B**) and FKBP5 as a positive control for dexamethasone-inducible transcription (**C**). The results are from three independent repeats with 40-60 embryos per sample; averages ± standard error are shown.

6.7 Reporter analysis in reporter cell lines after stable integration

The results with mammalian cell culture and zebrafish showed that NRSs effectively interfere with GR's ability to activate transcription suggesting that NRSs interfere with the ability of GR to bind to DNA. To test the hypothesis directly, I generated cell lines with genomically integrated reporters similar to those used for the transient luciferase reporter

assays. In addition to testing for transcriptional effects such genomically integrated reporters also allow examination of the effects of NRSs on genomic GR binding using ChIP experiments, an assay which in our hands is problematic with transient reporters. To assure that potential differences in TF binding are caused only by the NRS examined and not by the genomic integration site, I used zinc finger nucleases to direct targeted integration into the AAVS locus as described before for other GR reporters[123]. This way, I generated four cell lines encoding dexamethasone inducible luciferase reporter genes that encoded either one of the two control sequences or one of the two NRSs. The design of the donor construct is such that upon correct integration a GFP-fusion protein is generated driven by promoter of the PPP1 gene present at this locus. To isolate populations of transgenic cells, GFP positive cells were sorted and correct integration was examined by a diagnostic PCR (approximately 90% of GFP positive cells have the reporter integrated at the correct locus). From each of these cell lines, I also generated clonal cell lines to test if the effects observed for populations are recapitulated in clonal cell lines to exclude that the observed effects are caused by outliers.

I first examined the effects of NRSs on GR-activity in the genomic context by quantifying the luciferase activity for populations of transgenic cells. As observed for the transiently transfected reporter genes, the basal activity of all reporters was similar (Figure 21 B). Furthermore, whereas a robust activation was observed when the GBSs where flanked by the control sequences (38-fold for Ctrl #1, 62-fold for Ctrl #2) this activation was almost completely lost in the presence of either NRS 1 or NRS 2 (≈75% reduction for NRS 1, ≈ 95% reduction for NRS 2 when compared to Ctrl #1). We found the similar results when we examined clonal lines although here the basal reporter gene activity was below or around the detection limit. Hence, for these clonal lines the reporter gene activation cannot be compared by their fold activation, but instead was analyzed by comparison of measured absolute luciferase activity. Together, these findings show that the NRSs have even stronger effects on GR-dependent transcriptional activation when assayed in genomically integrated reporters.

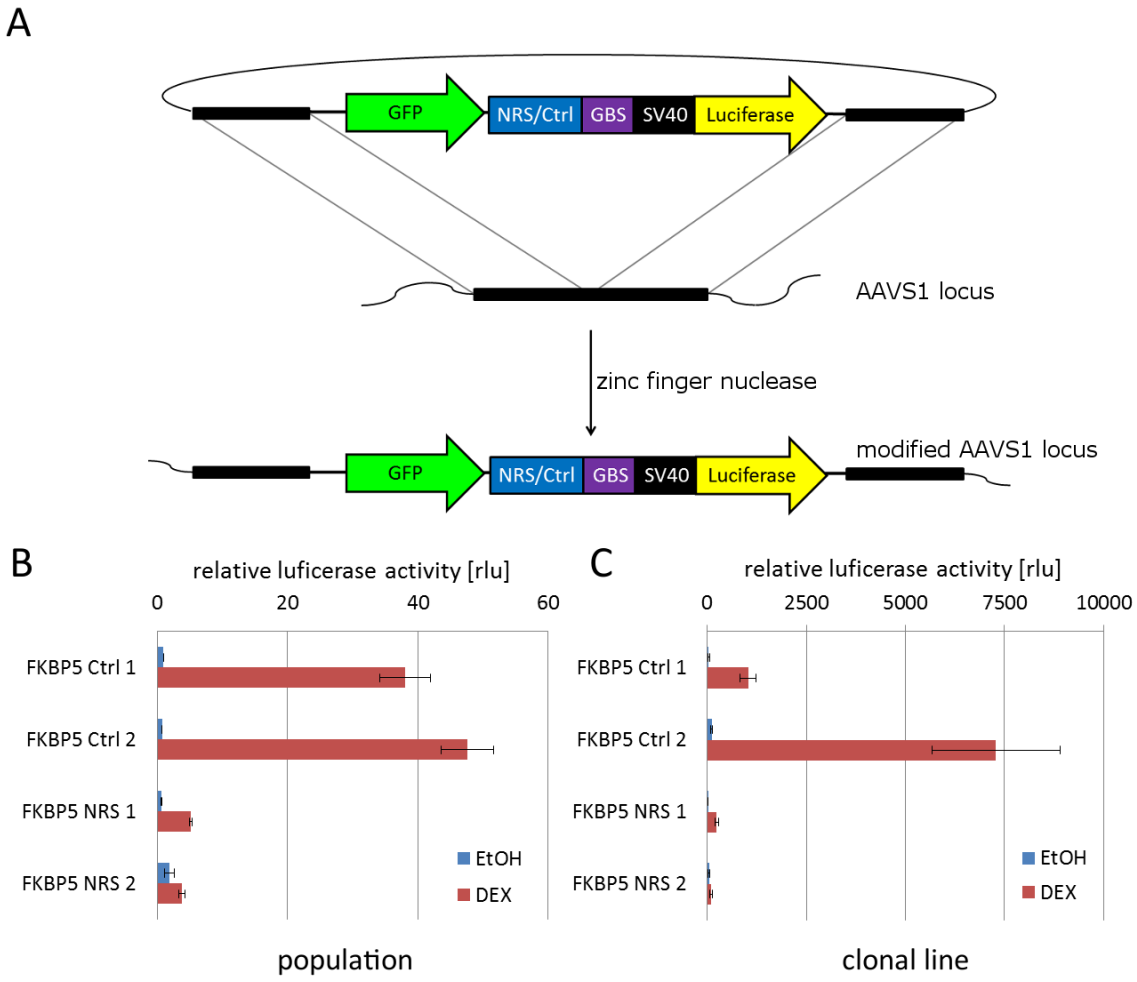


Figure 21: Influence of NRSs on genomically integrated reporters.
(**A**): (Figure modified from [123]). Luciferase reporter genes were cloned next to a GFP expression cassette. Both reporter genes are flanked by DNA sequences homologous to the AAVS1 locus. Plasmids encoding these constructs were co-transfected with a plasmid that encodes for a zinc-finger nuclease that targets the AAVS1 locus. Reporter genes were then integrated into this locus by homologous recombination. Four reporter cell lines were generated according to this scheme to test the influence of the two NRSs in a genomic context.
(**B**) The populations of reporter cell lines were treated overnight with 1µM dexamethasone or 1 ‰ ethanol as vehicle control and firefly luciferase expression was measured the next morning. Luciferase activities are shown relative to basal luciferase activity of control #1 cell line. (**C**) Clonal lines were generated from the populations of reporter cell lines. Luciferase activities were determined as in (**B**), however, luciferase activities are shown as relative luciferase activities without normalization. Averages from three biological replicates ± standard error of mean are shown.

6.8 Influence of NRSs on transcription factor binding

Next, I tested if the reduced activation of reporter gene transcription caused by NRSs is a consequence of reduced genomic GR binding at the nearby GBS. For this, I performed ChIP experiments for the populations of the reporter cell lines with an antibody against GR and quantified binding to the GBS by qPCR analysis. I did three independent biological replicates for all reporter cell lines, using input samples as controls. Calculation of the fold

enrichment was done by generating the ratio of the enrichment over the input for the Dexamethasone as well as ethanol treated samples.

The known endogenous GR target site FKBP5 showed a strong enrichment (minimum of 44-fold on average) upon hormone treatment in all four reporter cell lines, indicating that the ChIP efficiency for all these cell lines was comparable (Figure 22 A). As expected, little to no hormone-dependent increase in binding (maximum 1.6-fold) was observed for the negative control region RPL19 (Figure 22 C). Together, these findings indicated that the conditions used for the ChIP yielded a high sensitivity with reasonable specificity.

Next, I examined if NRSs affect GR binding to the GBS driving the luciferase reporter gene. For the reporters with control sequences flanking the GBS I clearly found increased GR-binding upon treatment with dexamethasone (on average more than 8-fold for control #1; 12-fold for control #2). In contrast, when the GBS was flanked by either one of the NRSs, GR-binding to the GBS was significantly decreased (on average less than 2-fold for both NRSs, a > 75 % reduction) (Figure 22 B).

Taken together, these results show that NRS sequences mediate their effects by preventing genomic binding of GR to nearby GBSs.

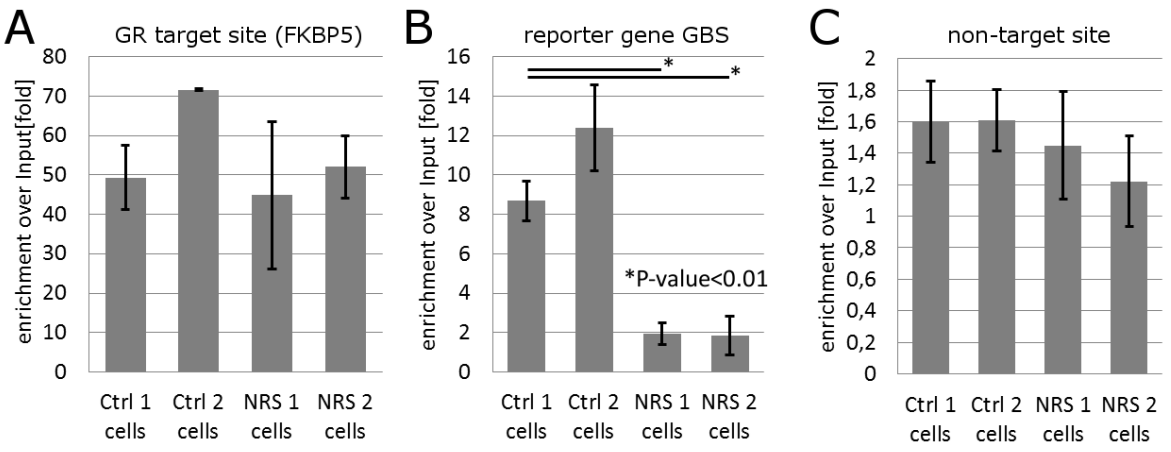


Figure 22: NRSs reduce GR binding to the reporter gene GBS.
The clones of GR-18 reporter cell lines were used to analyze GR binding by ChIP experiments. ChIP experiments were done as described in chapter 3.15 and analyzed by qPCR. A known genomic GR binding site (FKBP5) served as positive control (**A**), whereas a genomic region that is not bound by GR served as negative control (**C**). The results for the GR binding site encoded in the reporter gene are shown in (**B**). Shown are the mean hormone-dependent enrichments over input ± standard error of mean calculated for each corresponding locus, results are from three biological replicates.

6.9 Reduced GR binding is not mediated by conformational changes of the DNA

One potential explanation for the effects of NRSs on TF binding is that they induce a conformational change in the DNA, thereby preventing TF binding to their binding sites. One possible mechanism for this could be that AT-rich sequences can influence the minor groove width when present as so-called A tracts[158]. To test if NRSs mediate their effects by influencing the DNA conformation, I used dsDNA oligonucleotides encoding NRS2 or Control #1 adjacent to a GBS and tested how their presence influenced GR binding to the adjacent GBS using electrophoretic mobility shift assays (EMSAs). For these assays, I used the GRα DNA binding domain as described previously[103] and found that GR bound the GBS coding dsDNA oligonucleotide when flanked by the control sequence. The binding in the presence of the NRS was indistinguishable from that seen in the presence of the control sequence showing that at least in vitro (data not shown) the effects appear not to be a consequence of NRS-induced changes in DNA conformation that interfere with GR binding.

6.10 NRSs have no effect on chromosomal accessibility

GR binding is clearly correlated with accessible chromatin, with the majority of binding (>90%) occurring at DNase I hypersensitive sites[20]. Thus, one explanation for the effects observed for NRSs could be that they exert their effects by making nearby regions inaccessible. Therefore, I tested how the presence of NRSs influenced the DNA accessibility in DNase-I hypersensitivity assays using the transgenic cell lines with integrated reporters. To establish the assay, I first examined an “open” (FKBP5) and a “closed” (IGFBP1) genomic region from our parental U2OS cell line (Meijsing lab, unpublished observation). Consistent with these unpublished observations, qPCR analysis showed that DNase I treatment degraded more than 90 % of the “open” locus whereas for the “closed” locus only up to 22 % was degraded (Figure 22). This confirmed that this locus is relatively protected from DNase I degradation. These results show that our assay can indeed discriminate between “open” and “closed genomic loci.

When I examined the integrated reporters, I found that the locus of the reporter gene GBSs was degraded on average by 80 % independent of the presence of an NRS sequence (Figure 22). This shows that this genomic locus is relatively open and that NRSs do not appear to prevent GR binding by changing chromatin accessibility.

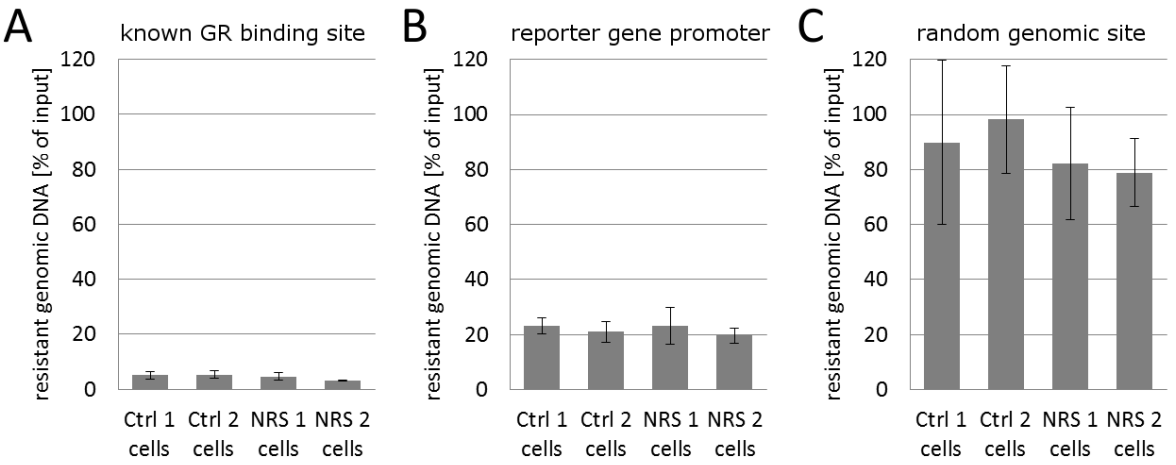


Figure 23: NRSs do not affect DNase I accessibility.
The clones of GR-18 reporter cell lines were used to analyze the influence of NRSs on DNA accessibility in DNase I hypersensitivity assays as described in chapter 3.18. The fraction of resistant genomic DNA for each locus was calculated from qPCR experiments by comparing DNase I treated versus mock treated genomic DNA. Low values indicate high accessibility, whereas higher values indicate low accessibility. In (**A**) the result from qPCR analysis for a known GR binding site that is accessible to DNase I is shown. In (**C**) the result for a DNase I inaccessible genomic region is shown. DNase accessibility to the reporter gene GBS is not affected by the presence of NRSs (**B**), as the corresponding genomic loci all show a very similar DNase I sensitivity. Averages from three biological replicates ± SEM are shown.

6.11 NRSs do not change nucleosome occupancy

Another possible explanation for the effects of NRSs could be that they influence DNA availability by changing the presence and/or positioning of nucleosomes. In fact, sequences with a high AT content have been implicated to play an important role in nucleosome positioning[159]. Nucleosome positioning is important for regulation of gene activity [41]. To analyze potential effects on nucleosome positioning caused by the AT rich NRS sequences, I conducted a MNase-based approach to detect the presence of nucleosomes at a given genomic target region[133].

Previous findings described that nucleosomes flanking the transcriptional start site of active genes follow a certain distribution, with a well positioned +1 nucleosome directly downstream of the TSS, and a nucleosome depleted region upstream[160]. Therefore, as control for the specificity of the assay, I analyzed the nucleosome positioning around the transcriptional start site of the housekeeping gene GAPDH transcript variant 1 (NCBI Reference Sequence: NM_002046.4). According to these findings this gene should have a well positioned +1 nucleosome directly downstream and a nucleosome-free region immediately upstream of its transcriptional start site (TSS). The level of nucleosome occupancy for each locus was determined by comparing the quantity present in the MNase-treated sample against that of the untreated chromosomal DNA input.

Each cell line was tested in two different conditions; either after treatment with 1 µM dexamethasone or when mock-treated with ethanol for one hour. As predicted, the region upstream of the GAPDH TSS is efficiently degraded by MNase (Figure 24), in agreement with the fact that this region is nucleosome-free. In contrast, the region downstream of the GAPDH TSS was relatively resistant to MNase in agreement with having a well-positioned +1 nucleosome. These experiments showed that the assay is indeed capable of discriminating between nucleosome-occupied and nucleosome-free regions. With the assay working, I next examined the nucleosome occupancy at the GBS of the integrated luciferase reporter genes. For all integrated reporter cell lines, I found a comparable MNase sensitivity that resembled the sensitivity seen for the +1 nucleosome of the GAPDH gene indicating that there is a nucleosome present at the GBS. However, the level of sensitivity did not change in the presence of an NRS, indicating that NRSs do not prevent transcription factor binding by increased recruitment of nucleosomes.

Although not reaching statistical significance (according to a t-test), the results for dexamethasone treated cells indicated that hormone treatment resulted in a slight increase in sensitivity to MNase (Figure 24), specifically for control reporters. This suggests that the activation by GR is accompanied by a displacement of the nucleosome, a phenomenon that was been observed for studies with the MMTV promoter[161]. In the presence of an NRS however, the lack of GR binding and transcriptional activation does not result in a displacement of the nucleosome present. Together, these data indicate that the effects of NRSs do not appear to be a consequence of increased nucleosome recruitment. Given the resolution of the assay however, I cannot rule out that slight changes in the positioning of the nucleosome might influence whether the GBS is available for GR binding or not

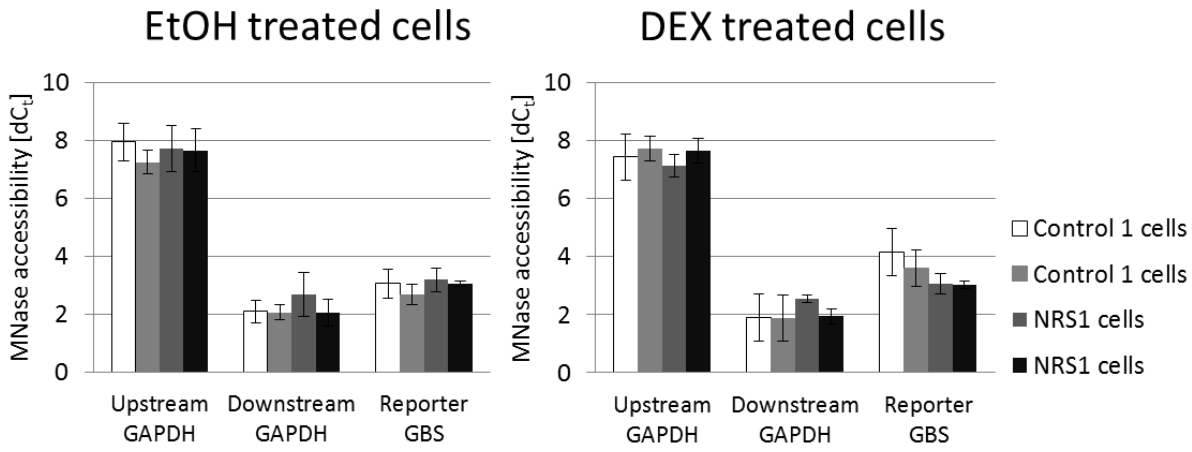


Figure 24: NRSs do not influence nucleosome occupancy.
The clones of GR-18 reporter cell lines were used to analyze the influence of NRSs on nucleosome occupancy using MNase assays as described in chapter 3.19. For each locus tested, the MNase accessibility was determined by qPCR, the dC_t values of both conditions served as indirect measure of nucleosome occupancy. Low nucleosome density is indicated by high MNase accessibility, whereas higher nucleosome occupancy is indicated by lower accessibility. The upstream GAPDH region served as control for a region with low expected nucleosome density, and the downstream GAPDH region as control for a region with high nucleosome density. Experiments were done in three biological replicates, averages ± standard error are shown.

6.12 NRSs do not change CpG methylation levels

NRSs might introduce local CpG methylation. However, that alone is not sufficient to induce heterochromatin formation. As pointed out before however, CpG methylation within TF binding sites can directly prevent TF binding[49]. A similar observation was made before with GR binding, where CpG methylations within the GBS at positions that form contacts to GR affected its binding[35]. Likely, GR-DNA binding is not only facilitated by interactions with the stretch encoding the binding site, but in addition it is enhanced by unspecific interactions of helix 3 with DNA 3-6 bp outside of its canonical binding site[103]. Similar observations were made for the progesterone receptor, indicating that such interactions with DNA outside of the canonical binding site might increase the binding affinity. Taken together, these results indicate that CpG methylation both within and outside of the canonical binding site prevents GR binding by inhibition of interactions with helix 3 of GR.

I tested for a potential NRS-mediated CpG methylation with the CpG methylation assay as described above. The control region that does not encode a binding site for the CpG methylation sensitive restriction enzyme ClaI shows that the DNA is not cut unspecifically, as after restriction 100 % (+/- 10 %) of initial DNA are recovered (Figure 25 B). For all

cell lines tested, 18 % (+/- 5%) of initial DNA was recovered after restriction digestion (Figure 25 A). This shows that the ClaI restriction digestion of the reporter gene GBS locus is independent of the presence of NRSs and that NRSs do not increase CpG methylation. However, although this indicates that NRSs do not mediate their effects by inducing increased methylation levels globally, they do not rule out that NRSs induce methylation of specific residues.

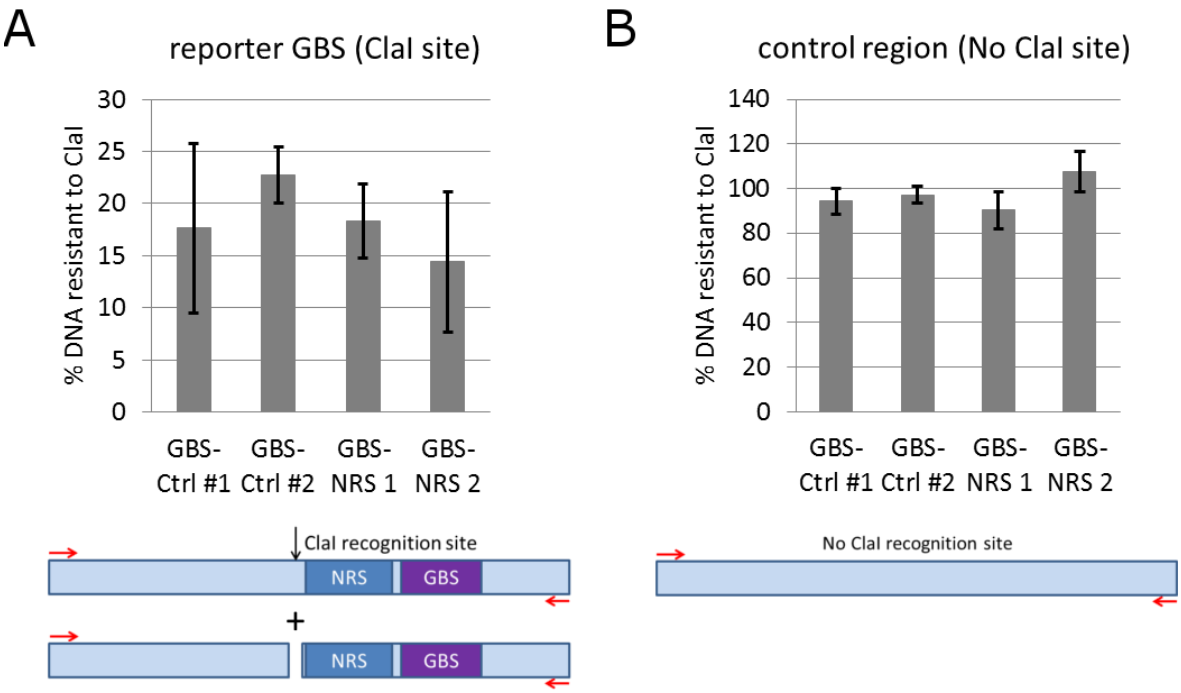


Figure 25: NRSs do not influence CpG methylation.
The clones of GR-18 reporter cell lines were used to analyze the influence of NRSs on CpG methylation as described in chapter 3.20. The tested reporter gene GBS region (**A**) encodes a ClaI recognition site; the control region does not (**B**). The level of degradation for both loci was determined by qPCR comparing the mock-treated to the ClaI-treated samples. If NRSs induce CpG methylation, the ClaI recognition site within the reporter gene GBS region should be more resistant to ClaI digestion. Averages from three biological replicates ± standard error of mean are shown.

6.13 Identification of Proteins that might mediate the effect of the NRSs

To identify the mechanism that prevents the binding of TFs to genomic regions surrounding NRSs, I hypothesized that they encode binding sites for proteins that in turn interfere with DNA-binding by GR. To identify such proteins, I performed DNA pull-down assays using the NRSs as baits to identify interacting proteins from a nuclear extract by subsequent mass-spectrometry analysis[134]. The double strand DNA oligonucleotides used as bait encoded a PstI restriction enzyme recognition site, one of the NRS/control sequences and a GBS. Furthermore, one of the two complementary oligonucleotides was labeled with a biotin-tag on the 5' end for immobilization of the baits on streptavidin-coated magnetic beads. A schematic drawing for the design of these DNA baits is shown in

Figure 26. Apart from the stretch encoding the control/candidate sequence the baits were identical.

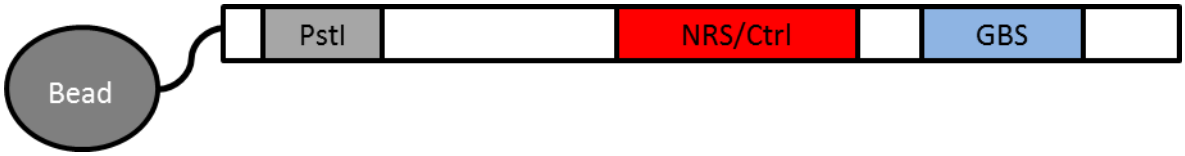


Figure 26: Schematic drawing representing the initial design of the DNA baits for the DNA pull-downs. The baits used for the DNA pull-down assays consist of two hybridized antisense oligonucleotides. To one of the oligonucleotides a biotin tag is attached for immobilization on streptavidin-coated beads. Next to this tag, a PstI recognition site is encoded for elution of the DNA baits after the washing. In the center of the bait, either one of the two NRSs or one of the two controls sequences is encoded next to a canonical GBS.

As described in chapter 6.9, EMSA experiments showed that the presence of the depleted sequences did not prevent GR binding to nearby GBSs. However, these experiments were done in the absence of other nuclear proteins. To test if NRSs can influence GR binding when nuclear proteins are present, I performed a DNA pull-down and compared the amount of GR precipitated between baits with and without an encoded NRS. The resulting eluents were split and half of the sample was tested for the amount and overall composition of proteins by PAGE with subsequent silver staining. The other half was used to compare the amount of GR precipitated by western blot analysis as described in chapter 3.25. The silver-stained gels showed that the total amount of protein and the composition of proteins between the two samples is similar (Figure 27). Unexpectedly, the western blot analysis showed that actually more GR was bound I the presence of NRS 2 when compared to the control sequence (data not shown). Staining for actin as loading control showed that this effect was not generated by loading of different amounts of protein from the eluents. In an additional DNA-pull down experiment, I tested GR binding to all four different baits by western blot analysis. Here, I found comparable binding of GR to all baits except for reduced binding in the presence of control sequence 1 (Figure 27), as observed in the initial experiment.

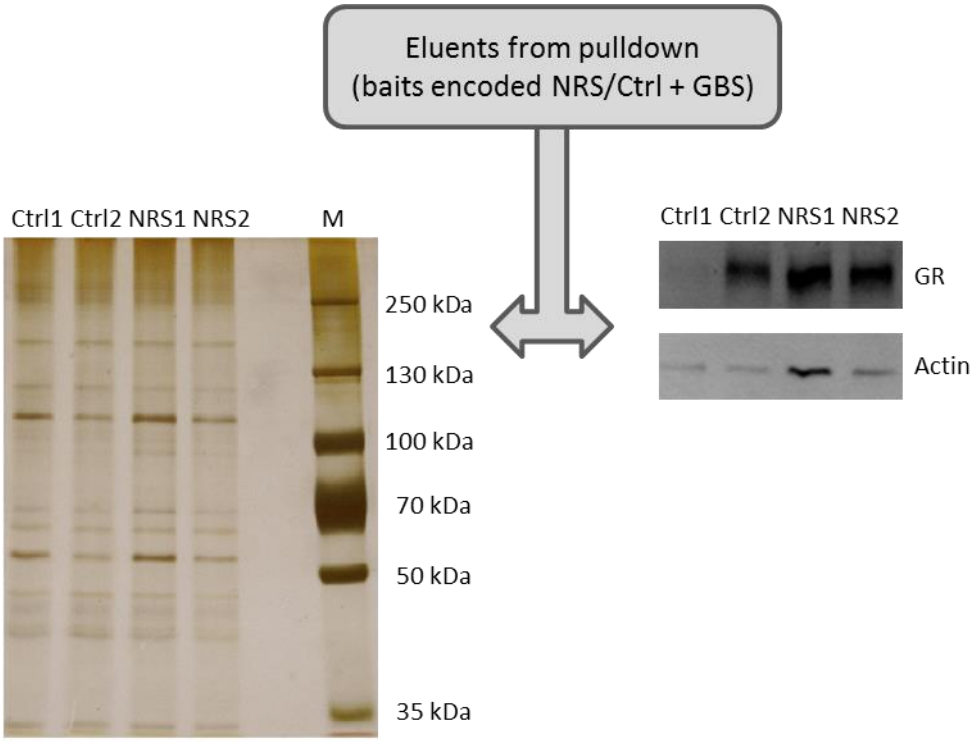


Figure 27: NRSs do not affect GR binding in DNA pull-downs.
DNA pulldowns were done using baits encoding a GBS in addition to one of the two NRSs or control sequences. Eluents from the pulldowns were split, and one part was analyzed by SDS-PAGE followed by silver staining (left), the other part was analyzed by western blot staining for GR and actin as loading control (right).

Together the results from the initial pull-down experiments showed that the presence of other proteins in the nuclear extract are not able to prevent GR binding in to the encoded GBS *in vitro* suggesting that the effects of NRSs require for instance the endogenous chromatin environment or the endogenous organization of the nucleus.

6.14 Identification of NRS-binding proteins by DNA pull-down

Because in the pull-down assays the effects of NRSs on TF binding was not recapitulated, I decided to modify the DNA baits by removing the GBS and therefore only having the NRS sequence present (Figure 28). With these modified reporters, I performed three independent DNA pull-down experiments. In the first experiment the proteins were not eluted from the beads by restriction digestion, but instead were trypsin digested directly on the beads. For the second and third experiment, I eluted the protein-bound DNA baits by PstI restriction digestion prior to sample preparation for mass spectrometry as described in chapter 0. Mass spectrometry analysis identified a total number of 363 different proteins in the eluents of the first experiment; elution by restriction digestion prior to mass spectrometry increased this number to 654 for the second and 794 different proteins for the third experiment.

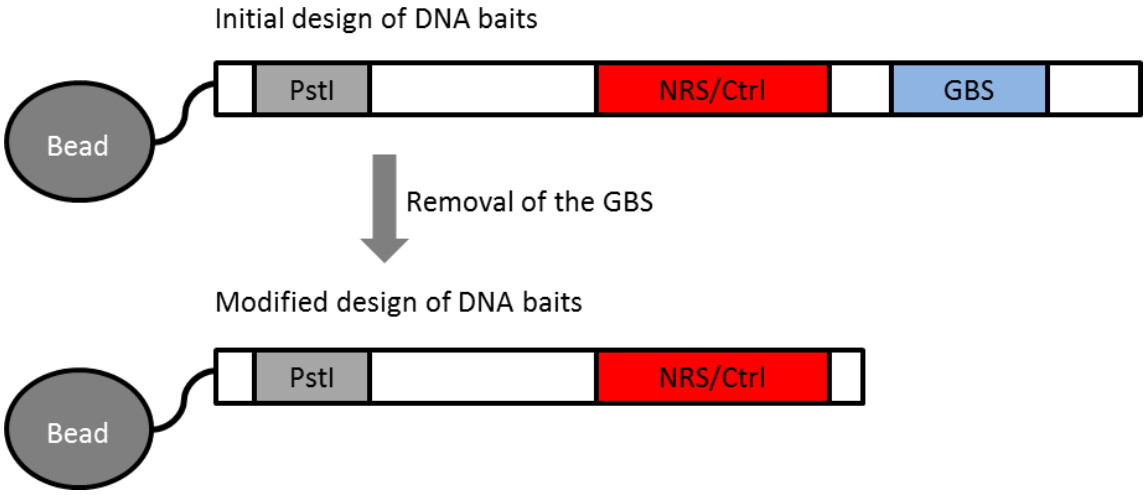


Figure 28: Modification of baits to reduce unspecific protein interactions.
To reduce unspecific interactions of proteins other than those that interact with the NRS, the encoded GBS was removed. The resulting baits were 23 bp shorter than the initial bait; otherwise the sequence of the baits was not modified.

To identify NRS-binding proteins, I compared the abundance of proteins in the eluents from DNA pull-downs using NRS coding baits against the abundance in eluents from pull-downs using the control sequences. Therefore, the intensity scores from label-free quantification for each detected protein were compared between NRSs and control sequences. For analysis I generated the sum of intensity scores for the control and the test conditions, and then added a pseudo-count. Then, I formed the ratio of intensity scores for the control and NRS-conditions and only considered those with at least two-fold enrichment for the NRS-condition. Many of the detected proteins were found only in one or two of the three samples, with intensity scores of zero for the other samples. To filter for real interaction partners, only proteins that were identified in at least two of the three experiments were further considered. The 11 proteins that were identified to be more than two-fold enriched in all three experiments and hence considered top candidates were the following:

- HNRNPD
- HOXA5
- KARS
- NONO
- NPM1
- PSPC1
- RBM14
- RPL7A
- SFPQ
- TARDBP
- TOMM20

6.14.1 Candidate proteins of the DBHS family

Interestingly, the proteins SFPQ, NonO and PSPC1 that selectively bound to NRS encoding bait belong to the DBHS protein family and have been implicated in RNA processing[162,163]. PSPC1 accumulates at nucleoli when transcription is inactive[164]. PSPC1 and the other DBHS family proteins are core component of a subnuclear structure called paraspeckle[165]. This suggests that paraspeckles might play a role in in mediating the effects of NRSs. Consistent with this idea, another paraspeckle component, RBM14[165], was also identified as an NRS-interacting protein. Supporting the hypothesis that paraspeckles are important for NRS activity, two of the DBHS proteins, namely SFPQ and NonO, appear to modulate the transcriptional activity of AR by impeding the interaction with its response element[166]. Likely this effect is caused by the potential of SFPQ and NonO to interact with the DBD of nuclear hormone receptors, with SFPQ being active as transcriptional repressor (probably through interaction with Sin3a)[167]. In contrast DHBS family members can also act as coactivators by interacting with the androgen receptor-AF1[168] in a complex containing NonO, SFPQ and PSPC1. In addition, SFPQ can act as a co-repressor for the progesterone receptor, another steroid hormone receptor related to GR, by preventing the progesterone receptor DBD from binding to its response element and by enhancing degradation of the progesterone receptor[169]. This makes these proteins ideal candidates to mediate NRS activity.

6.14.2 Candidate proteins identified only in two out of three experiments

In addition to the proteins that were identified in all three experiments, I also considered those that were identified only twice for further analysis.

For example, because the NRS sequences resemble DNA-binding sites for homeodomain proteins[170], I expected such proteins to be enriched in pull-downs using the NRS encoding baits. Accordingly, I identified three homeodomain proteins (HMBOX1, Oct1 and HoxA5), to selectively bind to the NRS-encoding baits in two of the three experiments. HMBOX1 binds to the sequence CTAGTTAA[170], the exact PWM describing its binding site preference (motif) is depicted in Figure 29 and Figure 32 A. The motif indicates that the two central positions of the motif (GT) are essential for HMBOX1 binding. This sequence is encoded completely within the sequence of the oligonucleotide bait encoding NRS 2 whereas for NRS 1 the first five positions of the sequence are encoded, including the two essential central GT bases. Conversely, the baits that were used as negative controls lack the essential central bases explaining that HMBOX binding to the control

sequences is weaker (see Figure 29). Hence, HMBOX1 identification from mass spectrometry indicates the specificity of the identified proteins and was used as an additional candidate protein. HMBOX1 is a homeodomain containing transcription factor that can act as a transcriptional repressor in experiments with transiently transfected reporter genes[171]. In addition, it is expressed in a broad range of cell types[172] making it a promising candidate to mediate the effects of NRSs.

Ctrl #1	CTAGCAAGCCTACCTCG
Ctrl #2	CTAGCGCAAACAAACCT
NRS 1	CTAGTGTTAATTAACCT
NRS 2	CTAGTTAATTGAATTAA

Figure 29: Alignment of bait nucleotide sequences with the HMBOX1 motif.
Shown is the overlap of the sequences of the four different baits used for the DNA pull-downs (table) and the binding site sequence for HMBOX1 as a PWM[170]. According to the PWM, the central positions of the motif consisting of GT are most important for HMBOX1 binding. Only baits encoding NRS 1 and NRS2 align at this position, with NRS 2 showing perfect alignment for the whole sequence motif.

Similarly to HMBOX1, Oct1 is a broadly expressed transcription factor and was hence also tested for its role in NRS mediated repression of TF binding. However, knockdown of Oct1 did not affect NRS activity, indicating that it is not necessary for its activity. HoxA5 is usually expressed only during embryogenesis and in adult kidneys, with HeLa being an exception to express HoxA5 (according to the databases PaxDb version 3.0, MaxQB Version 3.9.4 and MOPED version 2.5). Therefore, it is likely that it was identified in the DNA pull-down only because HeLa cells were used to generate the nuclear extract. Because NRSs are active in U2-OS cells, where HoxA5 is not expressed (according to the databases PaxDb version 3.0, MaxQB Version 3.9.4 and MOPED version 2.5) it is likely not important for NRS activity and will not be tested for a potential role in mediating NRS activity.

Additional candidate proteins that were identified twice in the Mass-Spec data are HMGA1 isoforms A1 and Isoform HMG-R. HMGA1 proteins bind to DNA through a structure called AT-hook peptide motif that recognizes the structure of AT-rich DNA rather than exact sequence motifs[173]. Each HMG protein has three AT-hook peptide motifs that bind to the minor grove of B-DNA[174] and they were described to serve as architectural transcription factors[175]. In contrast to a potential repressive role in mediating the NRS activity, HMGA1 proteins play a role in the formation of multi-protein complexes in

enhancer regions[176], indicative of an important role in the activation of transcription factor binding. Consistent with these findings, HMGA1 was described to enhance the binding of the estrogen receptor to its response element and thereby to activate transcription[177]. Interestingly, despite its association with transcriptional activation, HMGA1 is found mainly in heterochromatin[178], but little is known of its role in heterochromatin. Taken together, these reports indicate that HMGA1 is a reasonable candidate to mediate the effects of NRSs.

Also Lamin B1 was identified to interact with the NRS coding baits. As described above, genomic interactions with the nuclear lamina have been described to correlate to transcriptional activity of corresponding genes[38,69,70,76]. Hence, another potential mechanism for NRS activity that will be investigated is interaction with nuclear lamins.

6.15 Testing of candidates identified in DNA pull-down experiments

6.15.1 Testing of HMGA1 and DBHS proteins by RNAi

To test for a possible functional role of the candidate proteins interacting with the NRS sequences, I decided to test the effect of reducing their levels using esiRNAs, pools of short RNase-III digested double strand RNA targeting the coding sequence of interest[179]. For all candidates tested, transfection of cells with the esiRNAs resulted in reduced RNA levels by >-67%.

The individual knock-down of SFPQ, NonO or HMGA1, however, did not result in enhanced reporter gene activation for the integrated reporter cell lines harboring the NRSs (data not shown). However, DBHS proteins were reported to have redundant roles[128]. Therefore, knockdown of individual DBHS proteins might be insufficient to observe effects on NRS activity. I thus modified the experimental design and transfected esiRNA targeting all three human DBHS proteins or both SFPQ and NonO. When targeting all three DBHS proteins at once, the efficiency of knockdown for NonO mRNA was approximately 60%, < 20% for SFPQ and no decrease in mRNA for PSPC1 was observed. Hence I decided to only target SFPQ and NonO by esiRNA transfection in which enhanced the knockdown efficiency for SFPQ to 60%. In contrast to the effect observed for individual knockdowns, simultaneous knockdown of SFPQ and NonO differentially affected the reporter cell lines. For the controls, either no effect was observed (Ctrl 2) or a slightly enhanced activation by approximately 1.2-fold was observed (Ctrl 1). For the lines with NRS reporters these effects are more pronounced with an increase in activity of

approximately 1.6-fold for both NRS1 and NRS2 consistent with a potential role for DBHS proteins in mediating the repressive effects of NRSs.

In summary, these results point to a potential role for the DBHS proteins in mediating NRS activity, although the effects are only weak. This could be due to the relatively low efficiency of the knockdowns that could perhaps be improved by further optimization of the knockdown of DBHS proteins, ideally targeting all 3 DBHS protein simultaneously.

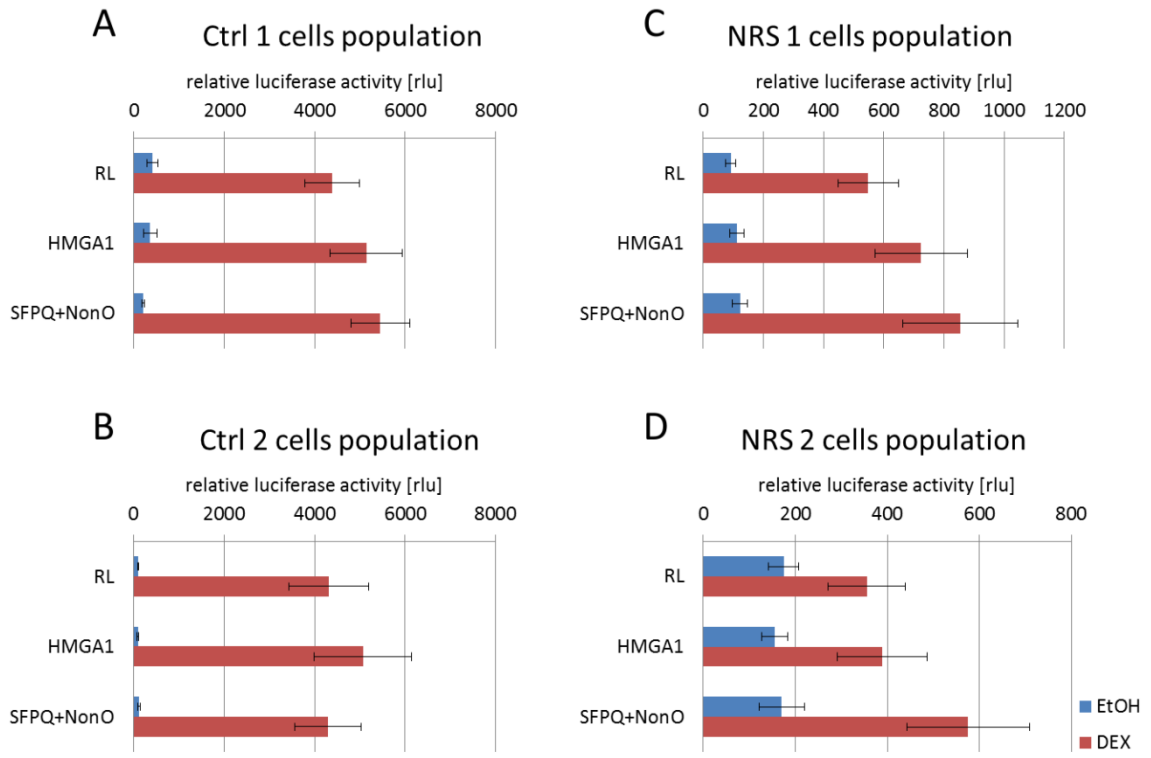


Figure 30: Effects of knockdowns of DBHS proteins on NRS activity.
Candidate proteins were tested in knockdown experiments in the reporter cell lines encoding either one of the two controls ((**A**) and (**B**)) or one of the two NRSs ((**C**) and (**D**)). As control, a non-target esiRNA (RL) was transfected to account for effects caused by the transfection. qPCR analysis of knockdown experiments indicated knockdown efficiencies of approximately 60 % or higher. 36 hours post transfection, cells were treated with dexamethasone or EtOH overnight, then the effects of knockdown on reporter gene activity was tested by comparing luciferase activity. Average relative luciferase activities ± SEM are shown.

6.15.2 Testing HMGA1 and DHBS-proteins in rescue experiments

Given the modest effects seen when knocking-down candidate proteins and potential issues with knock-down efficiency, I decided to test the effect of specifically recruiting candidate proteins to GBS-driven reporters. The expectation being that if these factors are indeed responsible for the effects of NRSs, their targeted recruitment should result in reduced GR-dependent transcriptional activity. The targeted recruitment was done by flanking the GBS with binding sites for Gal4. Subsequent recruitment of candidate proteins to the Gal4 binding sites of the reporter was facilitated by expressing them as fusion proteins with the

DBD of Gal4. To control for recruitment-independent effects of overexpression of candidate proteins, the effect of the fusion proteins was also tested using a modified reporter lacking the Gal4 binding sites flanking the GBS. Luciferase assays were done as described, with slight modification to co-transfect the fusion protein coding plasmids: per well I transfected 0.1 ng pRL (CMV), 10 ng pcDNA3.1-rGR and 45 ng p6R plus 30 ng of plasmid encoding the Gal4-DBD fusion candidate proteins. Each reporter was tested in three independent biological replicates, testing all reporters in parallel. As control, I tested the influence of the Gal4-DBD on reporter gene activation alone by co-transfection of the plasmid pSG5-Gal4.

Co-transfection of the Gal4-DBD fusion proteins with either HMGA1 or HMG-R isoform did not result in changes in either basal or hormone-induced activation of the $(gal4)_2$-$(GBS)_2$-luc reporter when compared to the activity seen when only the Gal4-DBD was co-transfected (Figure 31). Similarly, co-transfection of these fusion proteins with the control reporter gene without the encoded Gal4 binding sites also showed comparable values for the dexamethasone induced as well as for the basal luciferase activities. In contrast, the SFPQ- and the NonO-Gal4-DBD fusion proteins reduced reporter activation upon dexamethasone treatment when the test-reporter was co-transfected. This effect was not observed for the control reporter, indicating that the effects depend on the targeted recruitment of SFPQ and NonO to the GBS region of the reporter. In addition, a similar, albeit weaker, effect was seen for PSPC1. This supports previous findings that these proteins can interact with the DBD of nuclear hormone receptors and thereby interfere with their ability to interact with DNA[167]. Taken together, these results support the hypothesis that DBHS proteins participate in mediating the effects of the NRSs.

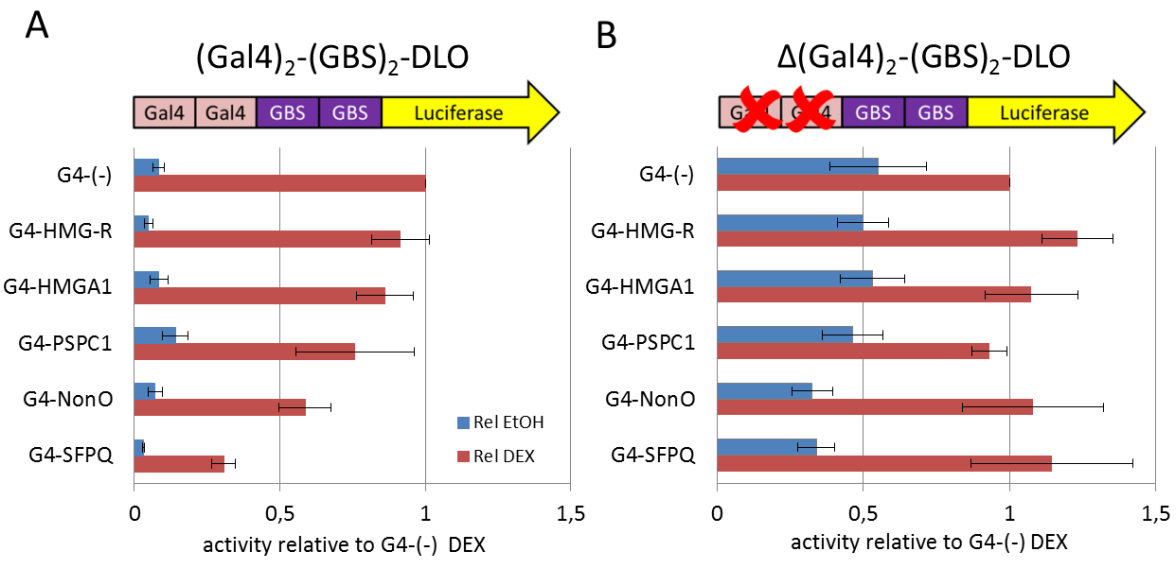
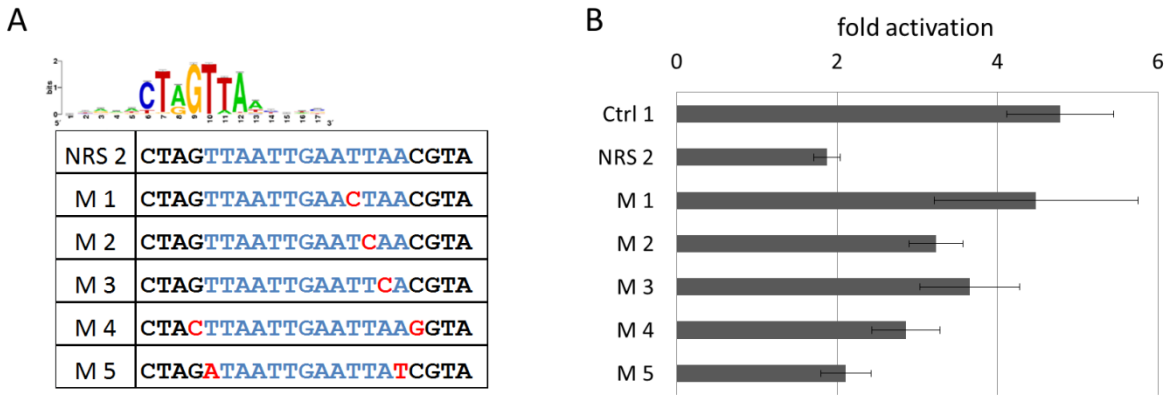


Figure 31: Targeted recruitment of DBHS proteins near GBSs interfered with GR-dependent activation.
Vectors encoding candidate NRS interaction partners as fusion proteins with the Gal4 DNA binding domain were co-transfected with reporter genes encoding two GBSs next to two Gal4 binding sequences as test reporter (**A**) or with reporter genes only encoding two GBSs as control (**B**). The influence of the overexpressed fusion proteins on reporter gene activation was determined in luciferase assays for ethanol and dexamethasone treated cells. For both reporters used, the level of reporter gene activity for each condition was normalized to the activity of the reporter in dexamethasone treated cells expressing the Gal4 DBD only. Averages of three biological replicates ± SEM are shown.

6.15.3 Testing of HMBOX1by mutagenesis of binding site

To test if HMBOX1 binding is important for mediating the effects of the NRSs, I introduced point mutations in the luciferase reporter plasmid pGL3-Promoter-FKBP5-NRS2 in the sequence encoding the NRS. Next, I tested the effect of five different mutations on NRS activity, with two of them selectively perturbing the HMBOX1 binding site sequence as shown in Figure 32 A. If HMBOX1 was important for NRS activity, I would expect that only mutations that affect the HMBOX1 binding site sequence should affect NRS activity. As observed before, the presence of wild type NRS2 resulted in a decrease of reporter activity for dexamethasone-treated cells by approximately 50% (Figure 32 B). This effect was reduced to approximately 20 % for all mutants (data not shown). For mutants 4 and 5, which selectively mutate the most constrained bases in the HMBOX1 motif, the basal activity of the reporter is slightly increased. Therefore, when looking at the fold activation, the effect of these two mutations is the weakest whereas for the M1 mutation that leaves the HMBOX1 sequence the effect of the NRS is almost completely lost. Together, these data indicate that although HMBOX1 can bind to the NRSs they are unlikely to be responsible for mediating its effect on GR-dependent transcriptional activation.



NRS 2	CTAGTTAATTGAATTAACGTA
M 1	CTAGTTAATTGAACTAACGTA
M 2	CTAGTTAATTGAATCAACGTA
M 3	CTAGTTAATTGAATTCACGTA
M 4	CTACTTAATTGAATTAAGGTA
M 5	CTAGATAATTGAATTATCGTA

Figure 32: Effect of NRS2 is not mediated by the HMBOX1 binding sequence.
To test if the HMBOX1 interaction is important for the effects of NRS 2, five mutant variants of the NRS reporter were tested, with mutations of mutant 4 and 5 specifically targeting the HMBOX1 binding site sequence (shown as PWM as described before[170]) (**A**). These mutant reporters were tested in luciferase assays, as controls I chose pGL3-P-FKBP5-Ctrl#1 with the NRS replaced with a random sequence as described in 6.4 and the wild-type pGL3-P-FKBP5-NRS2 reporter (**B**). Average fold activation upon dexamethasone treatment from three biological replicates ± SEM are shown.

6.15.4 Testing for influence of interaction with the nuclear lamina

Previously studies showed that genomic regions associated with the nuclear lamina are typically lowly expressed[71,76]. Such genomic regions often have a high AT content, similar to the NRSs[76]. In addition, lamin-associated domains are enriched in binding sites for Pou2f1[180], another protein that was enriched twice in the NRS pull-down assays (specifically enriched for NRS 2). To test if the NRS may serve as genomic anchoring points and thereby mediate the activity of NRSs, I tested the influence of knock-down of the individual lamin proteins lamin B1, B2 and A/C by transfection of dicer-substrate RNAs (DsiRNAs; Integrated DNA Technologies) targeting these isoforms. For all three lamin isoforms, I tested two different DsiRNAs. The mRNA knock-down efficiencies for all tested DsiRNAs were approximately 80% or higher. No effects on NRS-reporter activity were observed for lamin B1 and for lamin A/C (data not shown). Interestingly however, knockdown of lamin B2 resulted in increased reporter activation to approximately 120% of initial activity for both NRS1 and NRS2 coding reporter cell lines. In contrast, for both control reporter cell lines, activation was reduced to approximately 60-70% of initial activity.

Taken together, these results indicate that lamin B2 appears to act as a coactivator that is required for full GBS-driven activity by facilitating appropriate genomic 3D organization. Potentially, NRSs serve as an interaction site for lamin B2 to anchor such sequences to the nuclear lamina and thereby affect TF binding. Consistent with this idea the repressive effects of the NRSs are reduced when lamin B2 levels are decreased. Hence, the observed

effects could either be a direct effect NRS-lamin interaction; or alternatively perturbation of the subnuclear genomic organization might indirectly affect NRS activity.

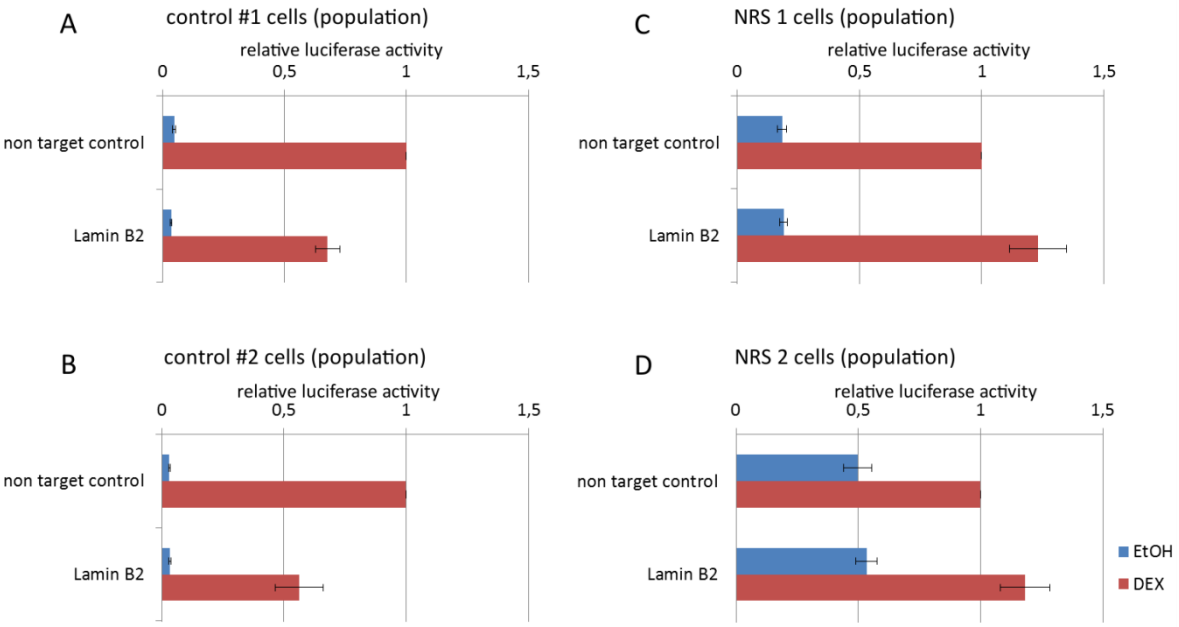


Figure 33: Lamin B2 knockdown partially reverses the NRS effect.
Lamin B2 was tested in knockdown experiments in the reporter cell lines encoding one of the two controls ((**A**) & (**B**)) or one of the two NRSs ((**C**) & (**D**)). As control, a non-target DsiRNA was transfected. qPCR analysis of knockdown experiments indicated knockdown efficiencies of approximately 80 %. The effect of knockdown on NRS activity was tested by measuring luciferase activity. 36 hours post transfection, cells were treated with dexamethasone or EtOH overnight, then the effects of knockdown on luciferase reporter gene activity was determined. Averages of three independent experiments ± standard error are shown.

6.15.5 Testing for influence of interaction with the nuclear matrix

Our experiments with DHBS proteins and lamins indicated a role of the nuclear architecture in NRS activity. Therefore, another potential mechanism by which NRSs might restrict TF binding is by interaction with proteins that organize nuclear organization.

Apart from paraspeckles and the nuclear lamina, another structure that might be important for gene regulation is the nuclear matrix. Interaction of the genomic DNA with this matrix is facilitated via special matrix attachment regions that were described to have a high AT content[181]. Hence, proteins associated with the nuclear lamina that bind genomic matrix attachment regions are further candidates to mediate NRS activity. Cux1, SATB1 and SATB2 are such proteins[182–184] that might be important for specific interactions with the nuclear matrix[185]. Also, SATB1 and Cux1 have been shown to directly interact and their binding was associated with transcriptional repression[186], and at least for SATB1 binding to nuclear matrix is important for repressive effects[182].

Similar to previous experiments, I tested the influence of these nuclear matrix associated proteins in knockdown experiments using the four GR-18 reporter cell lines. In these experiments, I transfected the reporter cell lines with esiRNA targeting Cux1, Satb1 or SATB2. Knockdown efficiency for Cux1 was approximately 70 %, for SATB1 and SATB2 40 % (data not shown). SATB2 knockdown did not show any effects on reporter gene activation. For all reporter cell lines, SATB1 knockdown partially reduced GR-dependent reporter gene activation. Cux1 knockdown in contrast, enhanced hormone induced reporter gene activity for all reporter cell lines to a similar degree, independent of the reporter gene encoding an NRS or a control sequence (data not shown). Taken together, these results indicate that SATB1 and Cux1 knockdown efficiencies were sufficient to affect reporter gene transcription. However, since the effects observed were not specific for NRS or control coding reporter genes these results do not indicate a role for the nuclear matrix in mediating NRS activity.

7 Discussion part 2

7.1 How is binding to non-target regions restricted in the genome

Sequences similar to the consensus GBS are ubiquitously found in the genome, raising the question why only a subset of these potential binding sites is bound. Over the last years, research focused mainly on understanding the mechanisms, more specifically sequence elements, which enhance TF binding to target loci. An example is given by in part 1 of this thesis, where I investigated the role of DNA sequence motifs (other than the consensus GBS) enriched at GR target sites in recruiting GR to specific genomic loci. Another potential mechanism to direct binding to specific loci only is by restricting access to non-target sites. This could be, for instance, through genomically encoded sequences that restrict binding either within a given cell type, or restrict binding in general. The positive signals for TF binding were identified by their enrichment at GR target loci. Therefore, I argued that opposing signals that restrict TF binding should be depleted at GR target sites, for the reason of being mutually exclusive. Following this logic, we (in co-operation with Morgane Thomas-Chollier) identified several candidate sequence motifs (CSMs) that are depleted at sites where GR binds in the genome. However, apart from a potential role in restricting binding of TFs, such depletions can have various origins. For example, such motifs might recruit proteins that do not prevent TF binding directly but overwrite signals from TFs bound nearby. In such a scenario, the presence of CSMs would be mutually exclusive only with TF activity, but not with TF binding. As the total number of TF in a cell is limited[119] such effects could result in a loss of nearby (non-functional) TF binding sites over the course of evolution, to avoid excessive TF binding without transcriptional effects, thus leading to the observed depletion. Another potential origin for the observed depletion of CSMs is that they are an artifact from the analysis or the experimental procedure[153].

Functional sequence motifs, such as binding sites for TFs, usually show increased conservation, as it is seen for the GBS[187]. Hence, if functional, the CSMs would likely show increased conservation. We (in co-operation with Stefanie Schöne) analyzed the PhyloP[188] score (a measure of nucleotide substitution rates relative to expected neutral drift) to test for a conservation of the CSMs (see Figure 14). Indicating conservation and consistent with a functional role, we found an increased PhyloP score across the whole CSM 2 (later called NRS 1) sequence motif. Similar to CSM 2, CSM 3 (later called NRS 2) also showed an increased PhyloP score when compared to its surrounding sequence, the

only exception being the central base, suggesting that this position is not important for NRS activity. However, because of its large sequence length, only 389 exact matching sites for CSM 3 were identified in the genome, which could explain the spikier PhyloP score profile distribution as shown in Figure 14. Interestingly, the PhyloP score distribution of the other sequence motifs tested (and which did not interfere with GR binding in our assays) indicated that longer stretches of adenosines or thymidines show increased conservation in general, whereas the transition between these nucleotides show decreased conservation. However, although such sequences are depleted at GR binding sites, their function does not appear to involve interfering with GR binding as their presence near GBSs did not interfere with GR function in the luciferase reporter studies.

7.2 Candidate sequences are functional

Using luciferase reporters, I found that two of four depleted sequences tested (hence called NRS 1 and NRS 2) were functional and indeed reduced GR-dependent activation of the reporter gene (Figure 16 and Figure 17). Importantly, the effects of NRSs on TF-mediated activation were highly specific for the sequence motifs I tested and not a simple consequence of having a high AT content (see Figure 17). The finding that NRSs also affect MyoD-driven transcription (see Figure 18) indicates that these might serve as a general signal to restrict TF binding to certain genomic regions. The identification of the other two non-functional CSMs might originate from artifacts or reasons other than those postulated above. Alternatively, or our reporters might not recapitulate the endogenous chromatin environment. Another striking observation was that stable genomic integration greatly enhanced the observed effects (up to 95 % reduction compared to approximately 50 % reduction with transiently transfected reporter genes). This might be caused by depletion in the pool of repressor proteins caused by the high number of NRSs introduced by transfection. Another reasonable explanation is that NRSs need the chromatin environment to be fully active.

7.3 NRSs restrict genomic binding of TFs

I tested the effect of NRSs on GR binding using ChIP experiments. Indeed the presence of NRS1 and NRS2 specifically decreased binding of GR to the nearby GBS by approximately 75 %, arguing that decreased reporter gene activation is a consequence of decreased TF binding. In summary, the ChIP experiments with our integrated NRS-reporters showed that NRSs interfere with TF binding and thus could explain the depletion of NRSs at genomic GR binding sites.

7.4 Effects of NRSs appear independent of changes in chromatin accessibility

Given the striking effects of NRSs on GR binding, I wanted to understand the mechanisms by which NRSs restrict TF binding. Several mechanisms are known that influence the availability of genomic regions for TF binding including DNA compaction and nucleosome positioning. For example, studies have shown that 90% of GR binding occurs at genomic regions that are "open" as assessed by DNase I hypersensitivity assays [20], arguing that GR binding is mostly restricted to open chromatin. The preferred binding to open chromatin and the finding that genomic integration of our NRS reporters enhanced the effects of NRSs suggested a role for chromatin. So, one straightforward explanation for the effects observed would be that NRSs serve to render regions of chromatin inaccessible. Contrary to our expectations however (Figure 23), the presence of an NRS did not influence chromatin accessibility. Similarly, other chromatin features tested, like nucleosome occupancy and CpG methylation, were not affected in the presence of NRSs (Figure 24 and Figure 25). This indicates that mechanisms other than simple chromatin compaction are responsible for the effect of NRSs.

7.5 Candidate proteins that interact with NRSs

To gain insight into the mechanism(s) by which NRSs interfere with TF binding, I set out to identify interacting proteins using NRSs as baits in DNA pull-down assays[134]. In the original design the DNA pull-down was used in combination with stable isotope labelling of amino acids in cell culture (SILAC) to test how binding of proteins is affected by point mutations in their cognate binding sites[134]. The change in affinity from these mutations did not necessarily completely impair protein-DNA interactions, hence SILAC was necessary for a quantitative assessment of the observed differences between the test and control condition. However, in contrast to the baits with only single point mutations in target sites as used in the original protocol, I designed the baits to either encode one NRS, or completely replaced the NRS with another sequence. Hence, I expected the difference in precipitation efficiency for NRS-interacting proteins between NRS and control-encoding baits to be larger. For these reasons, I decided to omit SILAC and instead normalize the results by label-free quantification using the MaxQuant software (performed by David Meierhofer). Interestingly, the sequences of both NRSs consist mainly of adenosines and thymidines in repetitive dimers of these bases. Because of the sequence similarity of these

two sequence motifs and their similar effects in our reporter assays, it is likely that both motifs serve as recognition sites for the same protein.

In the three DNA pull-downs I performed, 107 enriched proteins were identified at least twice in eluents from NRS-coding baits, and 11 of these were identified in all three experiments. One of these proteins (TOMM20) is usually found in mitochondria and hence likely to be a contamination of the nuclear extract with cytosolic proteins. From the remaining 10 proteins four are core components of paraspeckles (SFPQ, NONO, PSPC1 and RBM14)[189] and at least one which accumulates in Nucleoli (NPM1)[71,190]. The enriched binding of proteins that accumulate in distinct structures implicates that somehow interaction with subnuclear structures might play a role in mediating the effects of NRSs. Interestingly, in U2-OS cells only a small fraction of cells (~1 %) forms paraspeckles with a similar size as observed for with other cells HeLa, MCF7, Hek293, DFSF1 and 1787htert according to literature[165]. In agreement with previous findings[128], I have shown by immunofluorescence-staining that paraspeckles in U2-OS cells accumulate in small granules spread all over the nucleus (Figure 36). A role in NRS activity for the paraspeckle-associated proteins as indicated by the DNA pull-downs was supported by the experiments where these proteins were artificially recruited to the promoter region of reporter genes after fusion to a Gal4-DBD (Figure 31). Similarly, knockdown experiments targeting SFPQ and NONO in parallel partially reduced the effects of the NRSs, although these effects were not statistically significant (Figure 30). A potential explanation for the small effects on NRS activity might be that the knockdown efficiency for SFPQ and NONO was insufficient and that the amount of remaining protein after knockdown remained sufficient to mediate the effects of NRS sequences. A systematic analysis of the human proteome indicates that paraspeckle proteins are expressed at medium to high levels in all tissues and cell types examined (www.proteinatlas.org). Hence knockdown efficiencies of approximately 70 % on the mRNA level might be inefficient to reduce the protein concentration to a degree that is sufficient to prevent NRS-binding and activity.

7.6 Two hypotheses for mechanisms that mediate the effects of NRSs

7.6.1 Working Hypothesis 1

Hypothesis 1 is that NRSs serve as anchoring sites for subnuclear translocation of stretches of DNA to genomic regions that are not available for TF binding (see Figure 34). Similar to this hypothesis, the Polycomb 2 protein regulates the expression of growth-control genes by its facultative interaction with the lncRNAs TUG1 or Neat2. This interaction depends

on its methylation-state[191], thereby controlling the re-location either to polycomb bodies (genes repressed) or to interchromatin granules (genes expressed). In contrast to the mechanism observed for the Polycomb 2 protein, according to working hypothesis 1 paraspeckle-associated proteins (DBHS proteins and RBM14) bind to NRSs and relocate such bound loci to paraspeckle structures. Paraspeckles are structures that are formed by accumulation of proteins on clusters of lncRNA Neat1. According to this hypothesis, these paraspeckle regions of the nucleus interfere with TF binding and could thereby explain the effects of NRSs. In addition, interaction of NRSs with other subnuclear structures like the nuclear lamina might also contribute to the effects observed for the NRSs.

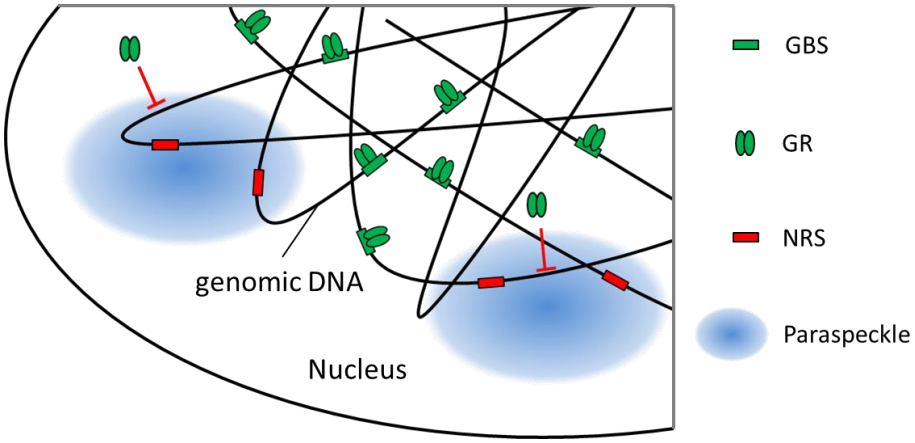


Figure 34: Illustration of working hypothesis 1.
NRSs encoded in genomic DNA serve as anchoring points for proteins that translocate genomic DNA to subnuclear structures like the paraspeckles. These structures then prevent binding of TFs to genomic DNA nearby (red spheres) by unknown mechanisms.

7.6.2 Working Hypothesis 2

Working hypothesis 2 is that the NRS-associated proteins directly interfere with TF binding. Consistent with this idea, binding of DBHS proteins to promoters of target genes has been described to have repressive effects on transcription[73,128]. For the type II steroid receptor subfamily, the interaction of SFPQ with Sin3A was proposed to repress transcription in a mechanism that involves recruitment of HDACs to the DBDs of these receptors[167]. For type I receptor subfamily members, such as GR, such an interaction with HDACs might also be present to restrict genomic binding of these receptors. The recruited HDACs might induce changes in translational modifications (de-acetylation) within GR's DBD or of other proteins that subsequently prevent binding to genomic binding sites[192]. Alternatively, the effects of HDACs might be a consequence of de-acetylation of histones and related chromatin remodeling. Another possible explanation for the repressive effects of DBHS proteins is that they directly interact with the DBD of GR and thereby prevent

that this domain can interact with DNA. Arguing for this possibility, studies by others have shown that PSPQ interacts with the DBDs of AR and PR which are virtually identical to the DBD of GR. Accordingly, another study has shown that GR interacts with PSPQ[193] Furthermore, EMSA experiments showed that the presence of PSPQ interfered with interaction of the DBD of PR with DNA[169]. Together, these findings indicate that the effects of NRSs might be a consequence of their ability to recruit DBHS proteins, thereby increasing their local concentration, which allows them to interact with the DBD of GR to prevent the interaction of this domain with DNA.

The formation of paraspeckles in the nucleus is triggered by Neat1 expression, with DBHS proteins accumulating on this lncRNA simultaneously with its transcription[194,195]. Increased Neat1 transcription results in excessive binding of DBHS proteins with the result that these accumulate in growing Paraspeckles[73,196]. The concentration of DBHS protein in the nuclei is not affected by changes in Neat 1 expression. Consequently, the enrichment of these proteins in paraspeckles upon increased Neat1 expression leads to their dissociation from bound promoters with subsequent loss of repression of corresponding genes[73]. Similarly, the repressive effects of paraspeckle components on GR binding to loci with NRSs might be modulated by changes in NEAT1 expression levels. This could modulate the genomic binding pattern and consequently the genes that are regulated by GR.

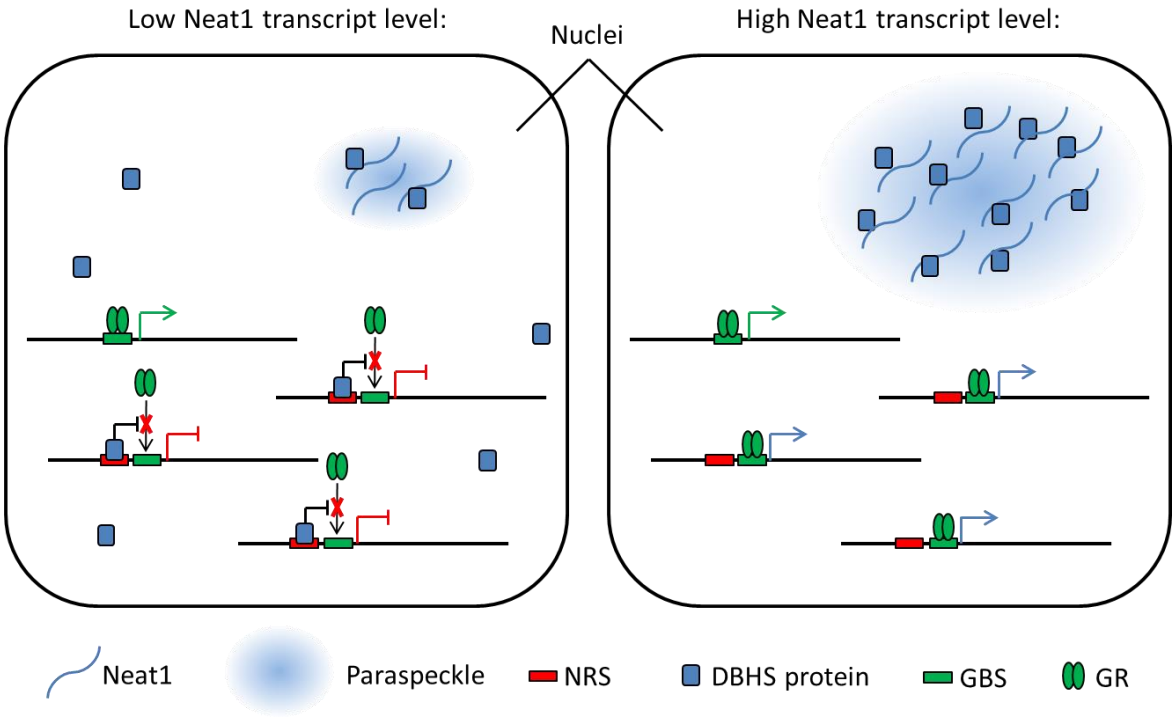


Figure 35: Illustration of working hypothesis 2.
When the Neat1 transcript is expressed at low levels (left panel) only a few DBHS proteins assemble in paraspeckles, the rest remaining either unbound or bound to NRSs, where they prevent binding of GR to GBSs nearby and subsequently transcriptional activation. When Neat1 levels are high, (right panel) DBHS proteins are recruited to the growing paraspeckles. Because the total number of DBHS proteins is not affected[120], the reduced availability of DBHS proteins results in a loss of NRS activity. As a result, GR can now bind GBSs near NRSs to regulate an additional subset of genes (represented by blue arrow) in addition to its regular set of genes (represented by green arrow).

7.7 Future directions

7.7.1 Testing of working hypothesis 1

I plan to test for NRS-mediated changes in subnuclear localization in co-localization studies of genomic regions encoding these NRSs with candidate structures such as paraspeckles, nucleoli or the nuclear lamina with three different approaches as described below.

7.7.1.1 Combine FISH with immunostaining

If NRS serve as anchoring points for proteins that translocate genomic regions, I might be able to show a co-localization of genomic regions encoding NRSs with the subnuclear structures that are formed by the candidate proteins. This will be tested using a combination of DNA-FISH and immunostainings of candidate proteins.

In initial experiments, I used the four reporter cell lines generated in previous experiments described above that have genomically integrated reporters in isogenic settings[123]. In two of the four cell lines these encode one of the two identified NRS, whereas in the other cell

lines either a random sequence or permutated AT-rich sequence is encoded as control. By localizing these reporter genes by FISH in combination with immunostainings for candidate subnuclear structures (as shown in Figure 36) I plan to test for potential effects on subnuclear repositioning caused by the NRSs.

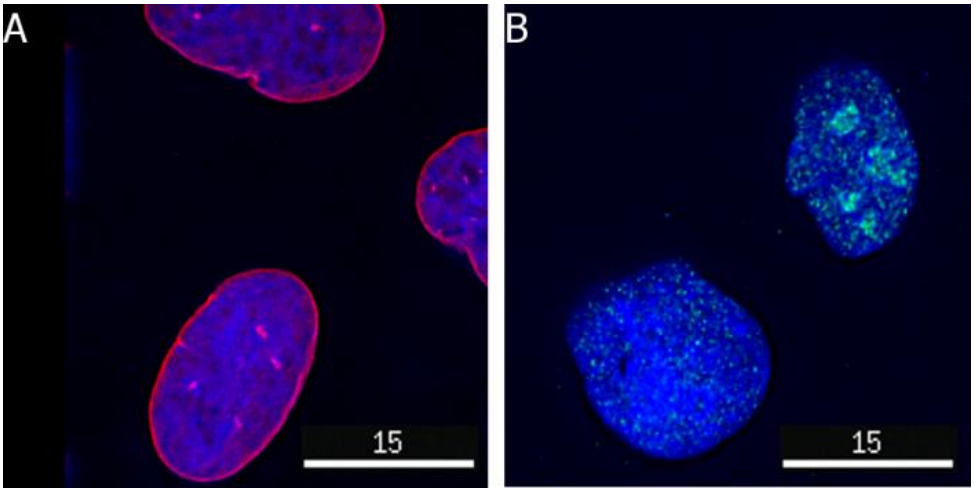


Figure 36: Immunostaining of candidate structures.
U2-OS were used for immunofluorescence staining of the nuclear lamina (**A**) or paraspeckles (**B**). Images were acquired with the a DeltaVision Core Restoration microscope mounted on an Olympus IX71 stand, using a ×60 oil immersion objective lens by Olympus with a numerical aperture of 1.42. A 1×1 bin with a section spacing of 0.2 µm was applied and images taken with a 12-bit Coolsnap HQ camera (Roper, USA) with the gain set to 4. Image processing was performed with SoftWorx (Applied Precision) and GIMP2. For better visualization, images were processed individually and therefore staining intensities cannot be compared between images.

Similar experiments focusing on the association of subnuclear structures with specific gene loci have been done successfully in the past[78,197]. Importantly, the isogenic setting of the reporter genes assures that the effects observed can be ascribed to the encoded NRS. The subnuclear positioning of NRSs will be compared to that of control sequences. I will initially focus on co-localization with paraspeckles, as these are main candidates for co-localization. A specific co-localization of NRSs with these structures would indicate a novel role for paraspeckles in modulating GR binding by influencing subnuclear positioning. In case I don't observe co-localization of NRSs with the paraspeckles, I will expand my analysis to test additional candidate proteins from my list that show a non-uniform subnuclear distribution. For proteins that show co-localization to the NRSs, I plan to do ChIP-Seq experiments to identify the genomic regions they occupy, with the ultimate goal to better understand their role in the global organization of the genome.

7.7.1.2 Analyzing the distribution of transiently transfected NRS coding reporters

In addition to the combination of FISH with immunostaining, I plan to investigate how NRSs influence the distribution of transiently transfected reporter-genes within the nuclei

of cells. To overcome limitations of efficient and reproducible fluorophore labelling of plasmid reporters, I will use MIDGE (minimalistic immunogenically defined gene expression) vectors designed by MOLOGEN AG. These MIDGE vectors, in contrast to regular vectors, are linear and consist only of the promoter region and coding sequence of the reporter gene. Exonuclease degradation of MIDGE vectors is prevented by hairpin structures at both ends. These hairpin structures will be used to incorporate fluorophores that can be tracked in the nucleus by confocal microscopy to study the influence of NRSs on subnuclear positioning. This experimental setup will also allow direct comparison of the distribution of control coding vectors against those encoding NRSs in the same nuclei by using different fluorophores for both constructs. A NRS-dependent change in localization would indicate that part of the effects of such sequences might be due to subnuclear positioning.

7.7.1.3 Testing for position relative to the nuclear matrix (DAM-ID)

If NRSs interact with subnuclear structures, they might also interact with the nuclear lamina which plays a key role in nuclear organization[198,199]. Supporting the idea that the AT-rich NRSs translocate DNA to the nuclear periphery, reports on DNA-lamin interactions indicate that lamin associated domains (LADs) have a high AT content[76]. Moreover, the lamin B2 knockdown experiments (chapter 6.15.4) indicated that NRSs might also interact with the nuclear lamina. Hence, in addition to the experiments described above that rely on microscopy, I also plan to investigate another possible mechanism for NRS activity: Their location relative to the nuclear periphery which is correlated to the gene activity[200]. However, so far no distinct sequence motif has been described that is necessary for the DNA-lamin interaction. Hence, if NRSs are responsible for such interactions, my research might provide a better understanding of the mechanisms by which the nuclear lamina interacts with genomic DNA. Genomic interaction with the nuclear lamina can be studied using a tool that has been described before[71]. Briefly, by expressing a DAM-LaminB1 fusion protein, genomic DNA that interacts with the nuclear lamina becomes methylated[71]. Using methylation-sensitive restriction enzymes, such interacting regions can be identified and the degree of interaction quantified. Here, I plan to investigate if the reporter genes show different levels of lamin interaction depending on the presence of NRSs to study the role of lamin-association in mediating the effects of NRSs.

7.7.2 Testing of working hypothesis 2

According to this hypothesis, NRS activity is mediated directly by DBHS proteins, for example by their interaction with the co-repressor Sin3A, as described before for SFPQ[167,201]. According to[167], treatment with sodium butyrate (an inhibitor of HDACs[202,203]) reduced the effects caused by SFPQ and reactivated nuclear receptor mediated transcription. Therefore, I plan to test hypothesis 2 either by treatment of the reporter cell lines with sodium butyrate or by knock-down of Sin3A. If DBHS interactions with Sin3A and subsequently with HDACs were important for NRS activity, such treatment should reactivate dexamethasone induced reporter gene transcription for the NRS-coding reporter cell lines. In contrast, for the control-coding reporter cell lines no effect would be expected.

An alternative approach to test hypothesis 2 is to enhance sequestration of DBHS proteins by increasing the Neat1 transcript abundance. The small granule structure of paraspeckles in the U2-OS cells I used in our studies indicates a low expression level of Neat1. Hence, I will use poly I:C to induce overexpression of the Neat1 lncRNA in the reporter cell lines, as described before[73]. Neat1 levels will be monitored by RT-qPCR or immunostainings for paraspeckles, and effects on NRS-activity will be determined by assaying how increased Neat1 levels influences the activity of the luciferase reporter gene in my reporter cell lines. If treatment with of poly I:C blunts the effects of NRSs, I will assay how this influences the genome-wide binding pattern of GR by ChIP-Seq.

7.8 Biological relevance of NRS action: Role of paraspeckles in regulating context-dependent GR binding

The lncRNA Neat1 is essential for paraspeckle formation[204] and protein recruitment to paraspeckles is regulated by the level of Neat1 lncRNA, which is one of the most-highly expressed long non-coding RNAs in the nucleus[73]. Neat1 expression is regulated via the TLR3-p38 pathway[73], which plays a fundamental role in the recognition of pathogens like viruses. Accordingly, activation of this pathway results in a marked increase in NEAT1 levels[73,205]. Increased transcription of Neat1 results in DBHS protein sequestration, with similar effects likely also for other paraspeckle-associated proteins like RBM14. This raises the intriguing possibility that GR binding sites that are normally unavailable for binding due to the repressive effects of DBHS proteins, might become available for binding during viral infections. Consequently, during viral infection, GR might be able to bind to a different sub-set of binding sites and thereby regulate a viral-infection-specific

subset of target genes. Of particular note in this context, glucocorticoids are stress hormones that play an essential role in the protection against cytokine-mediated lethality during viral infection.

Interestingly, for all cell lines tested, ChIP-Seq data indicates that GR binds to genomic DNA directly upstream of the Neat1 coding region (see Figure 37). This suggests that GR might also play a role in the regulation of the Neat1 transcript, and consequently could influence the sequestration of inhibitory proteins, such as the DBHS proteins. Depending on the direction of regulation (activation versus repression), this could either reduce or increase the availability of the inhibitory proteins. Such a mechanism would add another layer of complexity to how GR regulates the expression of genes. For instance via changes in the sequestration of the repressor proteins and thereby indirectly the expression of target genes of these repressors without the need for GR directly bind to regulatory sequences near such genes. Another possible consequence of GR-dependent regulation of Neat1 could be that it changes the genomic binding profile of GR after long-term exposure to glucocorticoids. For example, in the case of Neat1 activation, this could result in de-repression of NRSs and consequently allow GR to bind to additional GBSs to regulate associated genes.

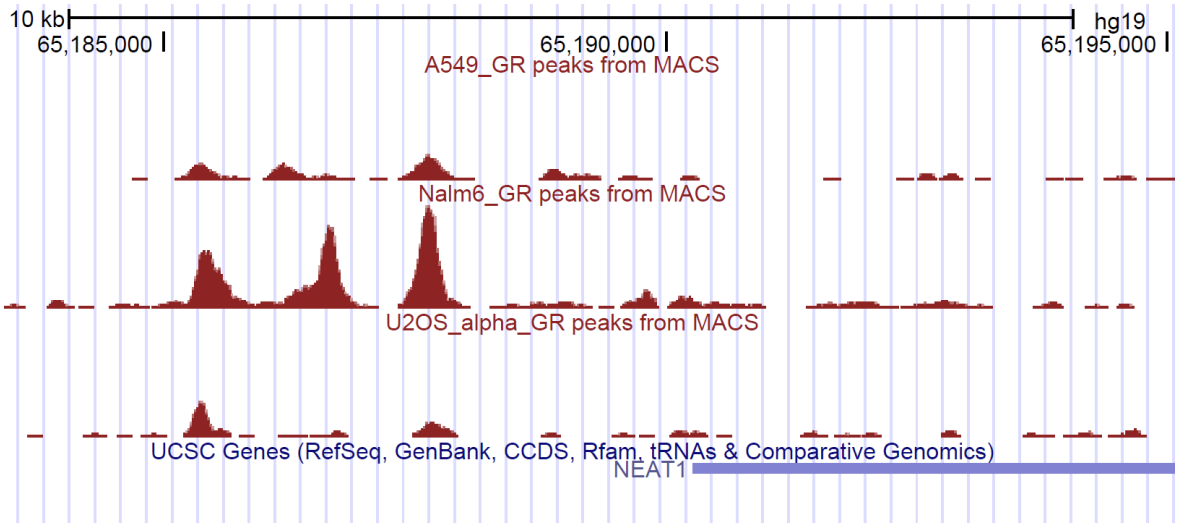


Figure 37: GR binding upstream the TSS of the lncRNA Neat1 is shared between multiple cell lines. Shown is GR ChIP-Seq data for A549, Nalm-6 and U2-OS in the UCSC genome browser in the region of the TSS of lncRNA Neat1. Peaks in this region are found in ChIP-Seq for all three cell lines showing that GR binding to the region upstream of the Neat1 TSS is conserved across cell types.

7.9 Conclusion: DNA encoded signals partition the genome in accessible and inaccessible regions

Binding of proteins to the genome is controlled by a variety of mechanisms that act together. Hence, to understand how genome-wide binding is controlled, we first have to identify the mechanisms that control TF binding. Eventually, only by integration of information on all these levels we will be able to understand how the binding pattern for TFs is established. A detailed understanding of these mechanisms might also help to determine how perturbations affect expression, thus allowing progress in personalized medicine. Ultimately, the information necessary to regulate the mechanisms of transcriptional regulation is encoded on the DNA. Hence, in my thesis I focused on potential sequence signals that partition the genome into sites that are accessible for TF binding and those that are not.

To date, ChIP-Seq data has yielded a lot of insight regarding genomic transcription factor binding sites and their role in facilitating transcription factor binding to specific genomic loci. Similarly, in part I of my thesis I have shown that sequence analysis of TF-bound genomic loci has the potential to reveal signals that promote transcription factor binding. In contrast, in part II I focused on opposing sequence signals that restrict binding of TFs. Such sequences can be identified from ChIP-Seq data and we (in cooperation with Morgane Thomas-Chollier) identified two motifs that prevent genomic binding of GR which I called NRSs (negative regulatory sequences). Experiments with MyoD indicate that NRSs likely affect binding of a broad range of TFs and studies with zebrafish indicate that the effects are conserved across species and appear to occur regardless of tissue examined. Interestingly, although many studies have shown that chromatin accessibility is a key driver in specifying which genomic loci can be bound by a TF in a particular cell type, the NRSs I identified here appear to mediate their effects through alternative mechanisms possibly involving subnuclear organization.

In summary, I have shown that sequence analysis of TF bound genomic loci has the potential to reveal signals that enhance transcription factor binding as well as signals that restrict TF binding to other loci. Interestingly, the negative signals I identified do not seem to influence chromatin accessibility arguing that a variety of mechanisms exist to restrict were TFs can bind in the genome.

In der vorliegenden Arbeit wurden DNA-kodierte Signale analysiert, welche die Genexpression durch Modulation der Bindung von Transkriptionsfaktoren (TF) steuern. Potentielle Signale in Form von DNA Sequenzmotiven wurden durch die Analyse von ChIP-Seq Daten unter der Annahme identifiziert, dass entsprechende Signale an genomischen Bindungsstellen angereichert (TF-Bindung induzierend) oder abgereichert (TF-Bindung verhindernd) sind. Diese Arbeit beschäftigt sich mit dem Einfluss solcher Sequenzen auf die genomische Bindung des Glukokortikoid Rezeptors (GR), einem durch Hormone gesteuerten TF.

Der erste Teil der vorliegenden Arbeit befasst sich mit Signalen, welche die genomische Bindung von GR begünstigen. In diesem Teil werden Signale in einer Zelllinie näher untersucht, in der die konventionelle GR-Bindungssequenz an gebundenen Regionen nur selten vorhanden ist. Dies wirft die Frage auf, welche Sequenzen für die Rekrutierung des GR an die Zielregionen verantwortlich sind. Die hier gezeigten Ergebnisse deuten an, dass die Rekrutierung von GR durch Interaktion mit anderen DNA-gebundenen Proteinen stattfindet. Solche Interaktionen könnten entweder GR indirekt über ein anderes Protein an die DNA binden, oder sie könnten die Konzentration von GR lokal erhöhen, und somit auch die Wahrscheinlichkeit einer direkten Interaktion mit der DNA.

Der zweite Teil dieser Arbeit befasst sich mit der Identifizierung und Charakterisierung von Sequenz-kodierten Signalen, welche die genomische Binding von GR einschränken. Interessanterweise scheinen die Mechanismen, welche die Bindung verhindern, nicht die Zugänglichkeit der DNA zu beschränken. Vielmehr scheint es, dass durch Interaktion mit speziellen subnuklearen Strukturen, den sogenannten Paraspecklen, die Bindung verhindert wird. Die Sequenzmotive könnten einerseits als Ankerpunkte dienen, um entsprechende genomische Abschnitte zu verlagern und so genomische Bindung von GR beeinträchtigen. Andererseits könnten die Sequenzmotive Proteine rekrutieren, welche mit GR interagieren und so dessen Bindung an entsprechende genomische Bereiche verhindern. Indem Paraspeckles die Bindung beschränkenden Proteine wie ein Schwamm absorbieren, könnten solche Bereiche als Reaktion auf äußere Reize wieder für die Bindung zugänglich gemacht werden.

Zusammengefasst zeigt diese Arbeit neuartige Mechanismen auf, welche TF-Bindung induzieren oder beschränken, und trägt zum besseren Verständnis genomischer Information zur Steuerung der Genaktivität bei.

Abstract

In the present thesis, I analyzed DNA-encoded signals that regulate gene expression by modulating transcription factor (TF) binding. Candidate signals in the form of DNA sequence motifs were identified by analysis of ChIP-Seq data, on the assumption that corresponding signals are enriched (enhancing TF binding) or depleted (restricting TF binding) at genomic binding sites. The influence of such signals on the genomic binding of the glucocorticoid receptor (GR), a hormone-activated TF, was the topic of my thesis.

One part focused on signals that enhance genomic binding of GR. Specifically, I studied signals in a cell line where the conventional binding site of GR is rarely found at GR-bound regions raising the question: what are the sequences responsible for GR recruitment in this cell line? The results presented here indicate that recruitment of GR is likely facilitated via interactions with other DNA-bound proteins. Such interactions might either indirectly tether GR to the DNA, or might elevate the local concentration of GR and thereby increase the likelihood of it to bind to DNA directly.

Another part of my thesis focused on the identification and functional characterization of sequence-encoded signals that restrict genomic GR binding. I identified such sequences and found that they indeed interfere with GR binding to the genome. Interestingly, my studies indicate that the mechanisms employed by these sequences to restrict GR binding do not involve changes in DNA accessibility. Rather, my results suggest that interactions with proteins found at specific subnuclear structures, called paraspeckles, are important to restrict GR binding. One possibility is that these sequence motifs serve as anchoring points for subnuclear re-localization of corresponding genomic loci and thereby influence GR binding. Another possibility is that the proteins recruited by these sequence motifs prevent GR interactions with the genome. As a response to external stimuli such genomic regions can be reactivated for TF binding by sponge-like activities of paraspeckles to sequestrate such proteins.

Together, my studies uncovered novel mechanisms that allow or restrict where TFs bind in the genome and thereby increase our understanding of how the information in the genome is decoded to produce the gene-products needed by cells.

10 Bibliography

1. Chapman, R. D., Conrad, M. & Eick, D. Role of the Mammalian RNA Polymerase II C-Terminal Domain (CTD) Nonconsensus Repeats in CTD Stability and Cell Proliferation. *Mol. Cell. Biol.* **25,** 1665–7674 (2005).

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

3. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38,** 576–89 (2010).

4. Wobbe, C. R. & Struhl, K. Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. *Mol. Cell. Biol.* **10,** 3859–67 (1990).

5. Kim, T., Ebright, R. H. & Reinberg, D. Mechanism of ATP-Dependent Promoter Melting by Transcription Factor IIH. *Science* **288,** 1418–1421 (2000).

6. Pant, G. & Greenblatt, J. Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region upstream of the initiation site. *J. Biol. Chem.* **269,** 30101–30104 (1994).

7. Young, B. A., Gruber, T. M. & Gross, C. A. Views of transcription initiation. *Cell* **109,** 417–420 (2002).

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

9. Søgaard, T. M. M. & Svejstrup, J. Q. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J. Biol. Chem.* **282,** 14113–20 (2007).

10. Cantin, G. T., Stevens, J. L. & Berk, A. J. Activation domain-mediator interactions promote transcription preinitiation complex assembly on promoter DNA. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 12003–8 (2003).

11. Esnault, C. *et al.* Mediator-dependent recruitment of TFIIH modules in preinitiation complex. *Mol. Cell* **31,** 337–46 (2008).

12. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467,** 430–5 (2010).

13. Søgaard, T. M. M. & Svejstrup, J. Q. Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. *J. Biol. Chem.* **282,** 14113–20 (2007).

14. Yamaguchi, Y. *et al.* NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97,** 41–51 (1999).

15. Peterlin, B. M. & Price, D. H. Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23,** 297–305 (2006).

16. Gupte, R., Muse, G. W., Chinenov, Y., Adelman, K. & Rogatsky, I. Glucocorticoid receptor represses proinflammatory genes at distinct steps of the transcription cycle. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 14616–21 (2013).

17. Barboric, M., Nissen, R. M., Kanazawa, S., Jabrane-ferrat, N. & Peterlin, B. M. NF-kB Binds P-TEFb to Stimulate Transcriptional Elongation by RNA Polymerase II. *Mol. Cell* **8,** 327–337 (2001).

18. Zhao, J., Hyman, L. & Moore, C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* **63,** 405–45 (1999).

19. Guertin, M. J. & Lis, J. T. Chromatin landscape dictates HSF binding to target DNA elements. *PLoS Genet.* **6,** e1001114 (2010).

20. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* **43,** 264–268 (2011).

21. Pique-regi, R. *et al.* Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* **21,** 447–455 (2011).

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

23. Nagy, L. & Schwabe, J. W. R. Mechanism of the nuclear receptor molecular switch. *Trends Biochem. Sci.* **29,** 317–24 (2004).

24. Ray, P. D., Foster, D. O., Henry, A. & Chem, J. B. Mode of Action of Glucocorticoids : I . Stimulation of Gluconeogenesis independant of synthesis de novo of enzymes. *J. Biol. Chem.* **239,** 3396–3400 (1964).

25. Leung, D. Y. M. & Bloom, J. W. Update on glucocorticoid action and resistance. *J. Allergy Clin. Immunol.* **111,** 3–22 (2003).

26. Lee, S. W. *et al.* Glucocorticoids selectively inhibit the transcription of the interleukin 1beta gene and decrease the stability of interleukin 1beta mRNA. *Proc. Natl. Acad. Sci.* **85,** 1204–1208 (1988).

27. Herrlich, P. Cross-talk between glucocorticoid receptor and AP-1. *Oncogene* **20,** 2465–75 (2001).

28. Olins, a L. & Olins, D. E. Spheroid chromatin units (v bodies). *Science* **183,** 330–2 (1974).

29. Gall, J. Chromosome fibers from an interphase nucleus. *Science* **139,** 120–121 (1963).

30. Gall, J. G. Chromosome fibers studied by a spreading technique. *Chromosoma* **20,** 221–233 (1966).

31. James, T. C. & Elgin, S. C. Identification of a Nonhistone Chromosomal Protein associated with heterochromatin in Drosophila melanogaster and Its Gene. *Mol. Biochem. Parasitol.* **6,** 3862–3872 (1986).

32. Wongtawan, T., Taylor, J. E., Lawson, K. a, Wilmut, I. & Pennings, S. Histone H4K20me3 and HP1α are late heterochromatin markers in development, but present in undifferentiated embryonic stem cells. *J. Cell Sci.* **124,** 1878–90 (2011).

33. Pajoro, A. *et al.* Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol.* **15,** R41 (2014).

34. Karlić, R., Chung, H.-R., Lasserre, J., Vlahovicek, K. & Vingron, M. Histone modification levels are predictive for gene expression. *Proc. Natl. Acad. Sci.* **107,** 2926–31 (2010).

35. Wiench, M. *et al.* DNA methylation status predicts cell type-specific enhancer activity. *EMBO J.* **30,** 3028–3039 (2011).

36. Elango, N. & Yi, S. V. DNA methylation and structural and functional bimodality of vertebrate promoters. *Mol. Biol. Evol.* **25,** 1602–8 (2008).

37. Galande, S., Purbey, P. K., Notani, D. & Kumar, P. P. The third dimension of gene regulation: organization of dynamic chromatin loopscape by SATB1. *Curr. Opin. Genet. Dev.* **17,** 408–14 (2007).

38. Bickmore, W. a & van Steensel, B. Genome architecture: domain organization of interphase chromosomes. *Cell* **152,** 1270–84 (2013).

39. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12,** 7–18 (2011).

40. Svaren, J. & Hörz, W. Transcription facors vs nuleosomes: regulaion of the PHO5 promoter in yeast. *TIBS* **22,** 93–97 (1997).

41. Bai, L. & Morozov, A. V. Gene regulation by nucleosome positioning. *Trends Genet.* **26,** 476–83 (2010).

42. Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* **3,** 662–73 (2002).

43. Eberharter, A. & Becker, P. B. Histone acetylation: a switch between repressive and permissive chromatin. *EMBO Rep.* **3,** 224–9 (2002).

44. Allfrey, G., Faulkner, R. & Mirsky, A. E. Acetylation and Methylation of Histones and their possible Role in the Regulation of RNA Synthesis. *Proc. Natl. Acad. Sci.* **51,** 786–794 (1964).

45. Hublitz, P., Albert, M. & Peters, A. H. F. M. Mechanisms of transcriptional repression by histone lysine methylation. *Int. J. Dev. Biol.* **53,** 335–54 (2009).

46. Bannister, a J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410,** 120–4 (2001).

47. Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8,** 1499–1504 (1980).

48. Gruenbaum, Y., Stein, R., Cedar, H. & Razin, a. Methylation of CpG sequences in eukaryotic DNA. *FEBS Lett.* **124,** 67–71 (1981).

49. Watt, F. & Molloy, P. L. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.* **2,** 1136–1143 (1988).

50. Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **19,** 187–91 (1998).

51. Bird, a. Methylation talk between histones and DNA. *Science* **294,** 2113–5 (2001).

52. Yusa, K., Takeda, J. & Horie, K. Enhancement of Sleeping Beauty Transposition by CpG Methylation: Possible Role of Heterochromatin Formation. *Mol. Cell. Biol.* **24,** 4004–4018 (2004).

53. Coulondre, C., Miller, J. H., Farabaugh, P. J. & Gilbert, W. Molecular basis of base substitution hotspots in Escherichia coli. *Nature* **274,** 775–780 (1978).

54. Shen, J., Ill, W. M. R. & Jones, P. A. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.* **22,** 972–976 (1994).

55. Shen, L. *et al.* Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet.* **3,** 2023–36 (2007).

56. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9,** 465–76 (2008).

57. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39,** 457–66 (2007).

58. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 1412–7 (2006).

59. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133,** 1106–17 (2008).

60. Cremer, T. *et al.* Rabl's Model of the Interphase Chromosome Arrangement Tested in Chinese Hamster Cells by Premature Chromosome Condensation and Laser-UV-Microbeam Experiments. *Hum. Genet.* **60,** 46–56 (1982).

61. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* **30,** 90–8 (2012).

62. Chambeyron, S. & Bickmore, W. a. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* **18,** 1119–30 (2004).

63. Volpi, E. V *et al.* Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J. Cell Sci.* **113,** 1565–76 (2000).

64. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2,** 292–301 (2001).

65. Wansink, D. G. *et al.* Fluorescent Labeling of Nascent RNA Reveals Transcription by RNA Polymerase II in Domains Scattered Throughout the Nucleus. *J. Cell Biol.* **122,** 283–293 (1993).

66. Jackson, D. a, Hassan, a B., Errington, R. J. & Cook, P. R. Visualization of focal sites of transcription within human nuclei. *EMBO J.* **12,** 1059–65 (1993).

67. Fraser, P. & Bickmore, W. Nuclear organization of the genome and the potential for gene regulation. *Nature* **447,** 413–7 (2007).

68. Osborne, C. S. *et al.* Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* **36,** 1065–71 (2004).

69. Akhtar, W. *et al.* Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* **154,** 914–27 (2013).

70. Boyle, S. *et al.* The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum. Mol. Genet.* **10,** 211–9 (2001).

71. Kind, J. *et al.* Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153,** 178–92 (2013).

72. Pontvianne, F. *et al.* Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states. *Genes Dev.* **27,** 1545–50 (2013).

73. Imamura, K. *et al.* Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. *Mol. Cell* **53,** 393–406 (2014).

74. Kumar, P. P. *et al.* Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat. Cell Biol.* **9,** 45–56 (2007).

75. Ma, H., Siegel, a J. & Berezney, R. Association of chromosome territories with the nuclear matrix. Disruption of human chromosome territories correlates with the release of a subset of nuclear matrix proteins. *J. Cell Biol.* **146,** 531–42 (1999).

76. Meuleman, W. *et al.* Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* **23,** 270–280 (2013).

77. Ma, H., Siegel, a J. & Berezney, R. Association of chromosome territories with the nuclear matrix. Disruption of human chromosome territories correlates with the release of a subset of nuclear matrix proteins. *J. Cell Biol.* **146,** 531–42 (1999).

78. Hutten, S. *et al.* An intranucleolar body associated with rDNA. *Chromosoma* **120,** 481–99 (2011).

79. Oren, I., Fleishman, S. J., Kessel, A. & Ben-Tal, N. Free diffusion of steroid hormones across biomembranes: a simplex search with implicit solvent model calculations. *Biophys. J.* **87,** 768–79 (2004).

80. Escriva, H. *et al.* Ligand binding was acquired during evolution of nuclear receptors. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 6803–8 (1997).

81. Evans, R. M. The steroid and thyroid hormone receptor superfamily. *Science* **240,** 889–95 (1988).

82. Mangelsdorf, D. J. *et al.* The nuclear receptor superfamily: the second decade. *Cell* **83,** 835–9 (1995).

83. Kumar, R. & Thompson, E. B. The structure of the nuclear hormone receptors. *Steroids* **64,** 310–319 (1999).

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

85. Pike, A. C. ., Brzozowski, A. M. & Hubbard, R. E. A structural biologist's view of the oestrogen receptor. *J. Steroid Biochem. Mol. Biol.* **74,** 261–268 (2000).

86. Danielian, P. S., White, R., Lees, J. A. & Parker, M. G. Identification of a conserved region required for hormone dependent transcriptional activation by steroid hormone receptors. *EMBO* **11,** 1025–1033 (1992).

87. Feng, W. *et al.* Hormone-Dependent Coactivator Binding to a Hydrophobic Cleft on Nuclear Receptors. *Science* **280,** 1747–1749 (1998).

88. Schoch, G. a *et al.* Molecular switch in the glucocorticoid receptor: active and passive antagonist conformations. *J. Mol. Biol.* **395,** 568–77 (2010).

89. Kumar, R. *et al.* Interdomain Signaling in a Two-domain Fragment of the Human Glucocorticoid Receptor. *J. Biol. Chem.* **274,** 24737–24741 (1999).

90. Kumar, R., Lee, J. C., Bolen, D. W. & Thompson, E. B. The conformation of the glucocorticoid receptor AF1/tau1 domain induced by osmolyte binds co-regulatory proteins. *J. Biol. Chem.* **276,** 18146–52 (2001).

91. Chandran, U. R., Warren, B. S., Baumann, C. T., Hager, G. L. & Defranco, D. B. The Glucocorticoid Receptor Is Tethered to DNA-bound Oct-1 at the Mouse Gonadotropin-releasing Hormone Distal Negative Glucocorticoid Response Element. *J. Biol. Chem.* **274,** 2372–2378 (1999).

92. Schoneveld, O. J. L. M., Gaemers, I. C. & Lamers, W. H. Mechanisms of glucocorticoid signalling. *Biochim. Biophys. Acta* **1680,** 114–28 (2004).

93. Langlais, D., Couture, C., Balsalobre, A. & Drouin, J. The Stat3/GR interaction code: predictive value of direct/indirect DNA recruitment for transcription outcome. *Mol. Cell* **47,** 38–49 (2012).

94. Stender, J. D. *et al.* Genome-wide analysis of estrogen receptor alpha DNA binding and tethering mechanisms identifies Runx1 as a novel tethering factor in receptor-mediated transcriptional activation. *Mol. Cell. Biol.* **30,** 3943–55 (2010).

95. Pratts, W. B. The Role of Heat Shock Proteins in Regulating the Function,Folding and Trafficking of the Glucocorticoid Receptor. *J. Biol. Chem.* **268,** 21455–21458 (1993).

96. Bledsoe, R. K. *et al.* Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* **110,** 93–105 (2002).

97. Yamamoto, K. R. Steroid Receptor regulated Transcription of specific Genes and Gene Networks. *Annu. Rev. Genet.* **19,** 209–52 (1985).

98. Becker, P. B., Gloss, B., Schmid, W., Strähle, U. & Schutz, G. In vivo protein-DNA interactions in a glucocorticoid response element require the presence of the hormone. *Nature* **234,** 686–688 (1986).

99. Dahlman-Wrights, K., Siltala-Roos, H., Carlstedt-Duke, J. & Gustafson, J.-A. Protein-Protein Glucocorticoid Interactions Facilitate DNA Binding by the Receptor DNA-binding Domain. *J. Biol. Chem.* **265,** 14030–14035 (1990).

100. Weatherill, P. J. & Bell, P. a. Characterization of the molybdate-stabilized glucocorticoid receptor from rat thymus. *Biochem. J.* **206,** 633–40 (1982).

101. Wrange, O., Eriksson, P. & Perlmann, T. The purified activated glucocorticoid receptor is a homodimer. *J. Biol. Chem.* **264,** 5253–9 (1989).

102. Beato, M., Chalepakis, G., Schaupr, M. & Slater, E. P. DNA regulatory elements for steroid hormones. *J. Steroid Biochem.* **32,** 737–748 (1989).

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

104. Wu, J. & Bresnick, E. H. Glucocorticoid and growth factor synergism requirement for Notch4 chromatin domain activation. *Mol. Cell. Biol.* **27,** 2411–22 (2007).

105. Ou, X. M., Storring, J. M., Kushwaha, N. & Albert, P. R. Heterodimerization of mineralocorticoid and glucocorticoid receptors at a novel negative response element of the 5-HT1A receptor gene. *J. Biol. Chem.* **276,** 14299–307 (2001).

106. Cooper, S. Glucocorticoid receptor DNA occupancy and transcriptional regulation across cell types. *Proquest, Umi Diss. Publ.* (2011).

107. Biddie, S. C. *et al.* Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol. Cell* **43,** 145–55 (2011).

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

109. McLeay, R. C. & Bailey, T. L. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11,** 165 (2010).

110. Thathia, S. H. *et al.* Epigenetic inactivation of TWIST2 in acute lymphoblastic leukemia modulates proliferation, cell survival and chemosensitivity. *Haematologica* **97,** 371–8 (2012).

111. Niebuhr, B. *et al.* Runx1 is essential at two stages of early murine B-cell development. *Blood* **122,** 413–23 (2013).

112. Lutterbach, B. & Hiebert, S. W. Role of the transcription factor AML-1 in acute leukemia and hematopoietic differentiation. *Gene* **245,** 223–35 (2000).

113. Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C. & Maki, R. A. The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene. *Cell* **61,** 113–24 (1990).

114. Moreau-Gachelin, F. Spi-1/PU.1: an oncogene of the Ets family. *Biochim. Biophys. Acta* **1198,** 149–63 (1994).

115. Fisher, R. C. & Scott, E. W. Role of PU.1 in hematopoiesis. *Stem Cells* **16,** 25–37 (1998).

116. Himmelmann, A. *et al.* PU.1/Pip and basic helix loop helix zipper transcription factors interact with binding sites in the CD20 promoter to help confer lineage- and stage-specific expression of CD20 in B lymphocytes. *Blood* **90,** 3984–95 (1997).

117. Sokalski, K. M. *et al.* Deletion of genes encoding PU.1 and Spi-B in B cells impairs differentiation and induces pre-B cell acute lymphoblastic leukemia. *Blood* **118,** 2801–2808 (2011).

118. Lichtinger, M., Ingram, R., Hornef, M., Bonifer, C. & Rehli, M. Transcription factor PU.1 controls transcription start site positioning and alternative TLR4 promoter usage. *J. Biol. Chem.* **282,** 26874–83 (2007).

119. Simicevic, J. *et al.* Absolute quantification of transcription factors during cellular differentiation using multiplexed targeted proteomics. *Nat. Methods* **10,** 570–6 (2013).

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

121. 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 (1997).

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

123. Dekelver, R. *et al.* Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res.* **20,** 1133–1142 (2010).

124. Weimann, M. *et al.* A Y2H-seq approach defines the human protein methyltransferase interactome. *Nat. Methods* **10,** 339–42 (2013).

125. Palidwor, G. a *et al.* Detection of alpha-rod protein repeats using a neural network and application to huntingtin. *PLoS Comput. Biol.* **5,** e1000304 (2009).

126. Kwan, K. M. *et al.* The Tol2kit: a multisite gateway-based construction kit for Tol2 transposon transgenesis constructs. *Dev. Dyn.* **236,** 3088–99 (2007).

127. Donà, E. *et al.* Directional tissue migration through a self-generated chemokine gradient. *Nature* **503,** 285–9 (2013).

128. Kowalska, E. *et al.* Distinct roles of DBHS family members in the circadian transcriptional feedback loop. *Mol. Cell. Biol.* **32,** 4585–94 (2012).

129. Hong, S. H., David, G., Wong, C. W., Dejean, A. & Privalsky, M. L. SMRT corepressor interacts with PLZF and with the PML-retinoic acid receptor alpha (RARalpha) and PLZF-RARalpha oncoproteins associated with acute promyelocytic leukemia. *Proc. Natl. Acad. Sci.* **94,** 9028–33 (1997).

130. Miller, J. C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25,** 778–85 (2007).

131. Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* **27,** 851–7 (2009).

132. Thomas-Chollier, M. *et al.* A naturally occurring insertion of a single amino acid rewires transcriptional regulation by glucocorticoid receptor isoforms. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17826–31 (2013).

133. Pham, C. D., Sims, H. I., Archer, T. K. & Schnitzler, G. R. Multiple Distinct Stimuli Increase Measured Nucleosome Occupancy around Human Promoters. *PLoS One* **6,** e23490 (2011).

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

135. Meierhofer, D. *et al.* Protein sets define disease states and predict in vivo effects of drug treatment. *Mol. Cell. Proteomics* **12,** 1965–79 (2013).

136. Karim, F. D. *et al.* The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. *Genes Dev.* **4,** 1451–1453 (1990).

137. Sharrocks, A. D., Brown, A. L., Ling, Y. & Yates, P. R. The ETS-domain transcription factor family. *Int. J. Biochem. Cell Biol.* **29,** 1371–87 (1997).

138. Bartel, F. rank O., Higuchi, T. & Spyropoulos, D. D. Mouse models in the study of the Ets family of transcription factors. *Oncogene* **19,** 6443–6454 (2000).

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

140. Espinás, M. L., Roux, J., Ghysdael, J., Pictet, R. & Grange, T. Participation of Ets transcription factors in the glucocorticoid response of the rat tyrosine aminotransferase gene. *Mol. Cell. Biol.* **14,** 4116–25 (1994).

141. Aurrekoetxea-hernández, K. & Buetti, E. Synergistic Action of GA-Binding Protein and Glucocorticoid Receptor in Transcription from the Mouse Mammary Tumor Virus Promoter. *J. Virol.* **74,** 4988–4998 (2000).

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

143. Himmelmann, A. *et al.* PU.1/Pip and basic helix loop helix zipper transcription factors interact with binding sites in the CD20 promoter to help confer lineage- and stage-specific expression of CD20 in B lymphocytes. *Blood* **90,** 3984–95 (1997).

144. Tsujimoto, K. *et al.* Regulation of the expression of caspase-9 by the transcription factor activator protein-4 in glucocorticoid-induced apoptosis. *J. Biol. Chem.* **280,** 27638–44 (2005).

145. Barrios-Rodiles, M. *et al.* High-throughput mapping of a dynamic signaling network in mammalian cells. *Science* **307,** 1621–5 (2005).

146. Hegele, A. *et al.* Dynamic protein-protein interaction wiring of the human spliceosome. *Mol. Cell* **45,** 567–80 (2012).

147. Rhee, H. S. & Pugh, B. F. Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147,** 1408–19 (2011).

148. Lichtinger, M. *et al.* RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J.* **31,** 4318–33 (2012).

149. Krysinska, H. *et al.* A two-step, PU.1-dependent mechanism for developmentally regulated chromatin remodeling and transcription of the c-fms gene. *Mol. Cell. Biol.* **27,** 878–87 (2007).

150. Adachi, N., Nishijima, H. & Shibahara, K. Gene targeting using the human Nalm-6 pre-B cell line. *Biosci. Trends* **2,** 169–80 (2008).

151. Chen, H. & Boxer, L. M. Pi 1 binding sites are negative regulators of bcl-2 expression in pre-B cells. *Mol. Cell. Biol.* **15,** 3840–3847 (1995).

152. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–826 (2013).

153. Hillier, L. W. *et al.* Whole-genome sequencing and variant discovery in C . elegans. *Nat. Methods* **5,** 183–188 (2008).

154. Yuan, W., Condorelli, G., Caruso, M., Felsani, A. & Giordano, A. Human p300 Protein Is a Coactivator for the Transcription Factor MyoD. *J. Biol. Chem.* **271,** 9009–9013 (1996).

155. Maclellan, W. R., Xiao, G., Abdellatif, M. & Schneider, M. D. A Novel Rb- and p300-Binding Protein Inhibits Transactivation by MyoD. *Mol. Cell. Biol.* **20,** 8903–8915 (2000).

156. Schaaf, M. J. M., Chatzopoulou, A. & Spaink, H. P. The zebrafish as a model system for glucocorticoid receptor research. *Comp. Biochem. Physiol.* **153,** 75–82 (2009).

157. Liu, T. X. *et al.* Evolutionary conservation of zebrafish linkage group 14 with frequently deleted regions of human chromosome 5 in myeloid malignancies. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 6136–41 (2002).

158. Rohs, R. *et al.* The role of DNA shape in protein-DNA recognition. *Nature* **461,** 1248–53 (2009).

159. Trifonov, E. N. Nucleosome positioning by sequence, state of the art and apparent finale. *J. Biomol. Struct. Dyn.* **27,** 741–6 (2010).

160. He, H. H. *et al.* Nucleosome dynamics define transcriptional enhancers. *Nat. Genet.* **42,** 343–7 (2010).

161. Belikov, S., Gelius, B. & Wrange, O. Hormone-induced nucleosome positioning in the MMTV promoter is reversible. *EMBO J.* **20,** 2802–2811 (2001).

162. Patton, J. G., Porro, E. B., Galceran, J., Tempst, P. & Nadal-Ginard, B. Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes Dev.* **7,** 393–406 (1993).

163. Peng, R. *et al.* PSF and p54nrb bind a conserved stem in U5 snRNA . *RNA* **8,** 1334–1347 (2002).

164. Andersen, J. S. *et al.* Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **12,** 1–11 (2002).

165. Fox, A. H. *et al.* Paraspeckles: a novel nuclear domain. *Curr. Biol.* **12,** 13–25 (2002).

166. Dong, X., Sweet, J., Challis, J. R. G., Brown, T. & Lye, S. J. Transcriptional activity of androgen receptor is modulated by two RNA splicing factors, PSF and p54nrb. *Mol. Cell. Biol.* **27,** 4863–75 (2007).

167. Mathur, M., Tucker, P. W. & Samuels, H. H. PSF Is a Novel Corepressor That Mediates Its Effect through Sin3A and the DNA Binding Domain of Nuclear Hormone Receptors. *Mol. Biochem. Parasitol.* **21,** 2298–2311 (2001).

168. Ishitani, K. *et al.* P54Nrb Acts As a Transcriptional Coactivator for Activation Function 1 of the Human Androgen Receptor. *Biochem. Biophys. Res. Commun.* **306,** 660–665 (2003).

169. Dong, X., Shylnova, O., John, R. G. & Lye, S. J. Identification and Characterization of the Protein-associated Splicing Factor as a Negative Co-regulator of the Progesterone Receptor. *J. Biol. Chem.* **280,** 13329–13340 (2005).

170. Berger, M. F. *et al.* Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133,** 1266–76 (2008).

171. Chen, S. *et al.* Isolation and functional analysis of human HMBOX1, a homeobox containing protein with transcriptional repressor activity. *Cytogenet. Genome Res.* **114,** 131–6 (2006).

172. Dai, J. *et al.* Recombinant expression of a novel human transcriptional repressor HMBOX1 and preparation of anti-HMBOX1 monoclonal antibody. *Cell. Mol. Immunol.* **6,** 261–8 (2009).

173. Reeves, R. & Nissen, M. S. The A/T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. *EMBO J.* **265,** 8573–8582 (1990).

174. Reeves, R. Structure and Function of the HMGI (Y) Family of Architectural Transcription Factors. *Environ. Health Perspect.* **108,** 803–809 (2000).

175. Reeves, R. Nuclear functions of the HMG proteins. *Biochim. Biophys. Acta* **1799,** 3–14 (2010).

176. Thanos, D. & Maniatis, T. Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell* **83,** 1091–100 (1995).

177. Massaad-Massade, L. *et al.* HMGA1 enhances the transcriptional activity and binding of the estrogen receptor to its responsive element. *Biochemistry* **41,** 2760–8 (2002).

178. Harrer, M., Lührs, H., Bustin, M., Scheer, U. & Hock, R. Dynamic interaction of HMGA1a proteins with chromatin. *J. Cell Sci.* **117,** 3459–71 (2004).

179. Yang, D. *et al.* Short RNA duplexes produced by hydrolysis with Escherichia coli RNase III mediate effective RNA interference in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 9942–7 (2002).

180. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453,** 948–51 (2008).

181. Mirkovitch, J., Mirault, M. E. & Laemmli, U. K. Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell* **39,** 223–32 (1984).

182. Seo, J., Lozano, M. M. & Dudley, J. P. Nuclear matrix binding regulates SATB1-mediated transcriptional repression. *J. Biol. Chem.* **280,** 24600–9 (2005).

183. Chattopadhyay, S., Whitehurst, C. E. & Chen, J. A Nuclear Matrix Attachment Region Upstream of the T Cell Receptor β Gene Enhancer Binds Cux / CDP and SATB1 and Modulates Enhancer-dependent Reporter Gene Expression but Not Endogenous Gene Expression. *J. Biol. Chem.* **273,** 29838–29846 (1998).

184. Britanova, O., Akopov, S., Lukyanov, S., Gruss, P. & Tarabykin, V. Novel transcription factor Satb2 interacts with matrix attachment region DNA elements in a tissue-specific manner and demonstrates cell-type-dependent expression in the developing mouse CNS. *Eur. J. Neurosci.* **21,** 658–68 (2005).

185. Sansregret, L. & Nepveu, A. The multiple roles of CUX1: insights from mouse models and cell-based assays. *Gene* **412,** 84–94 (2008).

186. Liu, J., Barnett, a, Neufeld, E. J. & Dudley, J. P. Homeoproteins CDP and SATB1 interact: potential for tissue-specific regulation. *Mol. Cell. Biol.* **19,** 4918–26 (1999).

187. So, A. Y.-L., 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 (2007).

188. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20,** 110–121 (2010).

189. Fox, A. H. & Lamond, A. I. Paraspeckles. *Cold Spring Harb. Perspect. Biol.* **2,** a000687 (2010).

190. Frehlick, L. J., Eirín-López, J. M. & Ausió, J. New insights into the nucleophosmin/nucleoplasmin family of nuclear chaperones. *Bioessays* **29,** 49–59 (2007).

191. Yang, L. *et al.* ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* **147,** 773–88 (2011).

192. Popov, V. M. *et al.* The functional significance of nuclear receptor acetylation. *Steroids* **72,** 221–230 (2007).

193. Tyson-Capper, a J., Shiells, E. a & Robson, S. C. Interplay between polypyrimidine tract binding protein-associated splicing factor and human myometrial progesterone receptors. *J. Mol. Endocrinol.* **43,** 29–41 (2009).

194. Bond, C. S. & Fox, A. H. Paraspeckles: nuclear bodies built on long noncoding RNA. *J. Cell Biol.* **186,** 637–44 (2009).

195. Mao, Y. S., Sunwoo, H., Zhang, B. & Spector, D. L. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* **13,** 95–101 (2011).

196. Hirose, T. *et al.* NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol. Biol. Cell* **25,** 169–83 (2014).

197. Koningsbruggen, S. Van *et al.* High-Resolution Whole-Genome Sequencing Reveals That Specific Chromatin Domains from Most Human Chromosomes Associate with Nucleoli. *Mol. Biol. Cell* **21,** 3735–3748 (2010).

198. Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38,** 603–13 (2010).

199. Gruenbaum, Y., Wilson, K. L., Harel, A., Goldberg, M. & Cohen, M. Review: nuclear lamins--structural proteins with fundamental functions. *J. Struct. Biol.* **129,** 313–23 (2000).

200. Boyle, S., Rodesch, M. J., Halvensleben, H. a, Jeddeloh, J. a & Bickmore, W. a. Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. *Chromosom. Res.* **19,** 901–909 (2011).

201. Duong, H. a, Robles, M. S., Knutti, D. & Weitz, C. J. A molecular mechanism for circadian clock negative feedback. *Science* **332,** 1436–9 (2011).

202. Stanley, F. & Samuels, H. H. n-Butyrate effects thyroid hormone stimulation of prolactin production and mRNA levels in GH1 cells. *J. Biol. Chem.* **259,** 9768–75 (1984).

203. Cousens, L. S., Gallwitz, D. & Alberts, B. M. Different Accessibilities in Chromatin to Histone Acetylase. *J. Biol. Chem.* **254,** 1716–1723 (1979).

204. Clemson, C. M. *et al.* An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell* **33,** 717–26 (2009).

205. Kawai, T. & Akira, S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat. Immunol.* **11,** 373–84 (2010).

bp	base pair
ChIP	chromatin immunoprecipitation
ChIP-Seq	ChIP sequencing
CSM	candidate sequence motif
CTD	C-terminal domain
Ctrl	control
DBD	DNA binding domain
DsiRNA	dicer-substrate RNA
EMSA	electrophoretic mobility shift assay
esiRNA	endoribonuclease-prepared siRNA
EtOH	ethanol
FISH	fluorescence in situ hybridization
GBS	GR binding site
GC	glucocorticoid
GR	glucocorticoid receptor
GRE	glucocorticoid receptor response element
gTF	general transcription factor
IP	immunoprecipitation
kb	kilo base pair
LAD	lamin associated domain
LUMIER	luminescence-based mammalian interactome mapping
NRS	negative regulatory sequence
PolII	polymerase II
RFP	red fluorescence protein
SILAC	stable isotope labeling by/with amino acids in cell culture
TF	transcription factor
TSS	transcriptional start site