



Published in final edited form as:

Nat Commun. ; 5: 5395. doi:10.1038/ncomms6395.

Conserved sequence-specific lincRNA-steroid receptor interactions drive transcriptional repression and direct cell fate

William H. Hudson^{1,2}, Mark R. Pickard³, Ian Mitchell S. de Vera⁴, Emily G. Kuiper¹, Mirna Mourtada-Maarabouni³, Graeme L. Conn¹, Douglas J. Kojetin⁴, Gwyn T. Williams³, and Eric A. Ortlund^{1,2}

¹Department of Biochemistry, Emory University School of Medicine, Atlanta, GA, 30322, USA.

²Discovery and Developmental Therapeutics, Winship Cancer Institute, Atlanta, GA, 30322, USA.

³Institute for Science and Technology in Medicine, Keele University, Keele, ST5 5BG, UK.

⁴Department of Molecular Therapeutics, The Scripps Research Institute, Jupiter, FL, 33458 USA.

Abstract

The majority of the eukaryotic genome is transcribed, generating a significant number of long intergenic non-coding RNAs (lincRNAs). While lincRNAs represent the most poorly understood product of transcription, recent work has shown lincRNAs fulfill important cellular functions. In addition to low sequence conservation, poor understanding of structural mechanisms driving lincRNA biology hinders systematic prediction of their function. Here, we report the molecular requirements for the recognition of steroid receptors (SRs) by the lincRNA Gas5, which regulates steroid-mediated transcriptional regulation, growth arrest, and apoptosis. We identify the functional Gas5-SR interface and generate point mutations that ablate the SR-Gas5 lincRNA interaction, altering Gas5-driven apoptosis in cancer cell lines. Further, we find that the Gas5 SR-recognition sequence is conserved among haplorhines, with its evolutionary origin as a splice acceptor site. This study demonstrates that lincRNAs can recognize protein targets in a conserved, sequence-specific manner in order to affect critical cell functions.

Introduction

An emerging phenomenon in biology is the widening gap between the availability of genomic sequences and the functional understanding of the products encoded within¹. The vast majority of functional research has focused on protein coding genes, which constitute

Correspondence should be addressed to E.A.O. (eortlun@emory.edu).

Author Contributions

W.H.H. performed binding studies, crystallographic analysis, mutagenesis, protein expression, *in vitro* transcription, and evolutionary analysis. M.R.P and G.T.W. performed apoptosis, survival, and RTPCR assays in MCF-7 and 22Rv1 cells. M.M.M. and G.T.W. performed apoptosis, survival, and RTPCR assays in suspension cell lines. I.M.d.V and D.A.K performed NMR assays on the GR-Gas5 complex and HADDOCK modeling, and E.G.K. and G.L.C. performed OH• footprinting on the GR-Gas5 complex. W.H.H. and E.A.O conceived the experiments and wrote the manuscript with input from all authors.

Conflict of Interest Statement

The authors declare there is no conflict of interest with the work presented in this article.

Accession Codes

The Gas5 GREM duplex crystal structures have been deposited to the PDB under accession numbers 4MCE and 4MCF.

only 2% of the human genome². Recently, attention has turned to the non-protein coding elements of the genome and transcriptome. While some non-coding RNA molecules (ncRNAs) such as small nucleolar and transfer RNAs have well defined function, the majority of non-ribosomal, nuclear-encoded RNA molecules present in a cell are of unknown function and/or structure³. Of these, long intergenic ncRNAs (lincRNAs), defined as intergenic ncRNAs greater than 200 nucleotides in length, remain the most poorly understood product of transcription⁴. lincRNAs have been shown to function in diverse processes such as pluripotency, development, cell cycle, immunity, transcription, and apoptosis⁵⁻⁹. Until recently, conservation of these lincRNAs was considered non-existent; however, emerging studies show higher conservation in lincRNAs than originally perceived, although well below the level of protein-coding exons^{5,10,11}. Consistent with this observation, a recent study searched for similarity between mouse and zebrafish lincRNAs and found only 9.0% of annotated mouse lincRNAs contained detectable sequence homology to whole-genome zebrafish alignments, with most of this conservation clustered in small domains separated by longer stretches of divergent sequence¹². Remarkably, mammalian lincRNA orthologs were still able to rescue developmental defects in zebrafish caused by specific lincRNA knockouts, indicating lincRNA function can be conserved despite divergent primary sequence¹². However, the mechanisms by which lincRNA functions are conserved in a rapidly changing sequence space are unclear.

The study of such mechanisms requires the identification of functional domains within lincRNAs. Unfortunately, there are currently few lincRNAs with well-characterized function, and still fewer in which functional domains have been identified and characterized. Known lincRNA functional domains include the 5' terminal region of lincRNA-p21, which interacts with hnRNP K to mediate the p53 response⁷. HOTAIR, a regulator of gene expression and chromatin state, uses a 5' domain to bind polycomb repressive complex 2, and a 3' domain to bind the LSD1/CoREST/REST complex¹³. Another lincRNA, growth arrest specific 5 (Gas5), uses a region in its 12th exon to interact with and repress the steroid hormone receptors (SRs)¹⁴. Without detailed structural information about these protein – lincRNA interactions, the basis of conservation (or lack thereof) within lincRNA functional domains is impossible to determine. By focusing on the interaction between SRs and Gas5, we sought to determine the molecular and structural requirements of a particularly important lincRNA-protein interaction.

The *GAS5* gene is a non-coding 5' TOP (terminal oligo-pyrimidine) gene consisting of twelve exons in humans and forms two mature spliced and polyadenylated isoforms in addition to numerous expressed sequence tags^{14,15}. The eleven introns of the *GAS5* gene contain ten box C/D snoRNA genes which are critical for ribosome modification and embryonic development¹⁶. Gas5 is present in all tissue types and is one of the most highly expressed non-coding RNAs in the human genome¹⁷. Gas5 has been implicated in many important cellular and physiological processes including p53 signaling⁷, embryogenesis⁶, growth arrest and cell cycle^{18,19}, and apoptosis^{14,20-22}. In addition, Gas5 is required for rapamycin action in immune cells²⁰, and Gas5 knockdown affects mRNA levels for M-Ras, prion protein (PrP), parkin, and other genes⁶. Gas5 inhibits the transcriptional activity of SRs through direct competition for DNA binding¹⁴. The SR-binding site was previously

mapped to a putative stem-loop region (bases 546-566) and described as a “mimic” of genomic SR binding sites, or glucocorticoid response elements (GREs)¹⁴. This GRE mimic (GREM) within Gas5 represents a compact lincRNA-protein interaction domain, optimal for examining the effects of altered lincRNA sequence and structure on protein-RNA interactions, and the downstream consequences on cellular processes. Here, we describe detailed studies of the mechanistic, structural, and evolutionary aspects of the SR-Gas5 interaction.

Results

The Gas5 GREM selectively binds 3-keto SRs

To measure the binding affinity between GR and Gas5, we monitored via fluorescence polarization the ability of full-length Gas5 and shorter GREM-containing constructs (Supplementary Fig. 1) to compete with fluorescently-labeled DNA activating GREs, or (+)GREs, for binding to the glucocorticoid receptor (GR) DNA binding domain (DBD; Fig. 1a). Full-length Gas5 competed for DNA binding the most strongly, with a K_i of 158 nM – within the range of affinities of GR for (+)GREs (80 nM – 1 μ M)²³. Shorter constructs (Supplementary Fig. 1) containing only the GREM were sufficient for DNA competition, although affinity was lowered approximately five-fold compared to the full-length Gas5 transcript (Fig. 1a). In addition, a fluorescently labeled Gas5 GREM construct containing only the stem bound directly to GR, indicating that only the putative double helical region within the GREM is required for GR binding (Fig. 1b), much like GR – DNA binding requires a short, double helical response element.

Previous work demonstrated that Gas5 regulates the transcriptional activity of not only GR, but also the androgen receptor (AR), progesterone receptor (PR), and mineralocorticoid receptor (MR), which share the ability to bind 3-keto steroids as well as DNA (+)GREs¹⁴. Other nuclear receptors such as the estrogen receptor (ER), PPAR, and RXR, which bind different DNA response elements, are not susceptible to Gas5 mediated repression¹⁴. To determine the relative affinity of the Gas5 GREM for all SRs, we tested the ability of recombinantly expressed DBDs from all five SRs to bind both (+)GRE DNA and the Gas5 GREM stem (Fig. 1c,d). All 3-keto SRs bound both to a DNA (+)GRE and to the Gas5 GREM, which is consistent with their transcriptional repression by Gas5¹⁴. As predicted, the estrogen receptor DBD bound only to an estrogen receptor response element (ERE), and was unable to bind the Gas5 GREM (Fig. 1c,d, Supplementary Fig. 1m,n).

Gas5 exploits sequence difference among protein targets

Given their similar structure and sequence (>50% within the DBD), it is unclear how Gas5 discriminates among SRs; for example, the ER and GR DBDs share an identical fold and insert an α -helix to read the major groove of DNA (Fig. 1e). However, key amino acid differences in this helix allow the two receptors to bind disparate response elements. Notably, one ER side chain, Glu203, makes contact with DNA when bound to an ERE, while the 3-keto SRs contain a glycine at this position (Fig. 1f). When bound to a ERE, ER Glu203 contacts a cytosine, forming an interaction required for DNA binding and subsequent gene activation²⁴. The 3-keto SRs do not contact DNA at this position, and

substitution of glycine with a bulkier amino acid side chain would introduce steric clashes with (+)GRE DNA bases (Fig. 1e).

Remarkably, mutation of this single amino acid in ER to glycine (E203G) confers ER binding not only to a DNA (+)GRE (Fig. 1g), but also to the Gas5 GREM (Fig. 1h). Likewise, the converse mutation in GR (G439E) reduced GR binding to both a DNA (+)GRE and the Gas5 GREM (Fig. 1g,h). The discrimination among evolutionary-related proteins by Gas5 demonstrates that lincRNAs are capable of sequence-specific interactions with their target proteins. Additionally, the strong correlation between DNA binding and Gas5 RNA binding suggest the presence of sequence-specific, (+)GRE-like SR contacts in the Gas5 major groove.

A widened major groove facilitates GREM-SR binding

Previous work postulated that a putative double-helical region of the GREM mediated SR-Gas5 binding¹⁴. UV melting experiments confirm that the Gas5 GREM stemloop indeed forms a secondary structure in solution (T_m 58 °C; Supplementary Fig. 2a). To determine the local RNA structure required for Gas5-SR recognition, we determined the x-ray crystal structure of the Gas5 GREM stem to a resolution of 1.9 Å. Two merohedrally twinned crystal forms were obtained in the rhombohedral space group R3, but twin refinement yielded excellent electron density and statistics (Table 1, Supplementary Table 1, and Supplementary Fig. 2b). In both crystal forms, the Gas5 GREM adopts an A-form double helix with a widened major groove. The average minor and major groove widths are greater than 15.0 Å in both crystal structures; wider than most A-form RNA helices found in crystal structures, but within the broader range of major groove widths in solution as determined by NMR²⁵. GR-bound DNA response elements of GR exhibit major groove widths approximately 1.5 – 3.0 Å wider than the free Gas5 GREM structure^{23,26}, indicating that only minimal secondary structure perturbation may be necessary for GR – Gas5 binding (Supplementary Fig. 2c,d).

Molecular determinants and model of the GR-Gas5 interaction

To identify bases within the GREM for critical SR-Gas5 binding, we performed hydroxyl radical footprinting and used NMR to structurally map the GR-Gas5 interaction. Increasing concentrations of GR DBD protected several portions of the GREM stem from OH• hydrolysis while deprotecting bases U554-U556 within the loop (Fig. 2a-c). The pattern of protection indicates GR interaction with the major groove, much like its classical interaction with DNA. To visualize the impact of GR binding on Gas5, we obtained two-dimensional homonuclear [¹H,¹H] NOESY spectra via ¹H NMR of the Gas5 GREM stemloop both in the presence and absence of the GR DBD (Fig. 2d-f). Nucleotides within the Gas5 GREM stem showed NMR resonances with moderate or strong line broadening upon addition of GR DBD, whereas others displayed strongly shifted positions (>0.025 ppm). Broadened resonances correspond to nucleotides that are switching between two states in the intermediate exchange NMR timescale, whereas resonances that are simultaneously shifted and broadened are switching between two states in the slow-to-intermediate or intermediate-to-fast exchange regime. Only C561 and U569 have H1'-H6 resonances that are shifted but not broadened, and therefore, have two states that are in slow or fast exchange. The two

states may correspond to nucleotides situated in disparate chemical shift environments, such as (1) free vs. GR DBD-bound Gas5 RNA, (2) two Gas5 RNA conformational states in the presence of GR DBD, or (3) as a result of interactions with a neighboring Gas5 RNA or with another Gas5 RNA region within the same molecule. Both NMR and OH• footprinting demonstrated conformational changes in the four base loop upon GR binding (Fig. 2c), consistent with previous studies indicating that bulges and loops can widen the RNA major groove, presenting surfaces for protein interaction^{27,28}.

To map the surface through which GR interacts with Gas5, we collected two-dimensional [¹H, ¹⁵N]-HSQC data on the GR DBD both free and bound to the Gas5 GREM (Fig. 3a-c). Notably, chemical shift perturbations are observed in helix 1 of the DBD when bound to either DNA²⁹ or Gas5 RNA, indicating that GR uses its DNA binding interface to bind RNA (Fig. 3b-e). Additional regions of the GR DBD that contact DNA in a non-specific manner, such as helix 2 and the N-terminus, are also perturbed when GR is bound to DNA and RNA. However, no chemical shift perturbations are observed in the dimerization interface (“D-loop”) of the DBD upon RNA binding as is observed with GR DBD – (+) GRE interactions³⁰, providing no evidence for dimerization on RNA (Fig. 3c,e). These GR DBD [¹H, ¹⁵N]-HSQC data, Gas5 GREM [¹H, ¹H] NOESY spectra, and mutagenesis data were used to generate a model of the GR DBD – Gas5 GREM complex using HADDOCK³¹. In line with the hypothesis that the Gas5 GREM is a RNA mimic of DNA response elements, the top scoring models showed the GR DBD reading the Gas5 GREM by inserting helix 1 into the major groove (Fig. 4a).

A single GREM mutation abolishes the SR-Gas5 interaction

The specificity of glucocorticoid receptor and 3 keto-SR-DNA recognition is heavily dictated by the contact of arginine and lysine residues in helix 1 of the GR DBD with two guanine-cytosine (GC) pairs in the (+)GRE major groove. Surprisingly, two equivalently spaced guanine-containing pairs (G549 and G559) are present in the Gas5 RNA GREM and were identified by structure probing as participating in the GR-Gas5 interaction (Fig. 2). Mutating these two guanines to adenine, which replaces the major groove O6 carbonyl with an amine, ablates GR binding (Fig. 4b-f). Mutation of G549 to 2-aminopurine, which selectively removes a potential hydrogen bond acceptor (O6) from the major groove while leaving the minor groove unchanged, diminishes but does not abolish binding. Critically, these Gas5 RNA mutations preserve secondary structure as monitored by UV melting (Supplementary Table 2). Along with the structural data above, these experiments establish that GR indeed binds Gas5 in the major groove.

Additionally, protein recognition of RNA wobble pairs has been reported in several systems^{32,33}, and two such non-Watson-Crick interactions are present in the Gas5 GREM: U548-G563 and G549-U562. To determine if GR-GREM recognition depends on this wobble base pairing, we mutated these GU pairs to canonical Watson-Crick G-C pairs. This mutation only marginally affected GR binding (Fig. 4g), indicating that GR does not discriminate between guanine-uracil and guanine-cytosine pairs.

To examine the effect of the G549A mutation on SR-mediated transcription in cells, we transfected AR-positive 22Rv1 prostate cancer cells, with WT Gas5, Gas5 with the G549A

mutation, or empty vector (pcDNA3.1). In these cells, WT Gas5 inhibited the androgen-driven expression of several AR target genes (Fig. 4h-j). However, this effect was reversed by the G549A mutation, with expression of the AR target genes returning to vector control levels. We therefore identified a single nucleotide mutation is sufficient to ablate the interaction between SRs and Gas5 both *in vitro* and in cells.

The Gas5 G549A mutant prevents Gas5-induced apoptosis

Gas5 is downregulated in both breast and prostate cancers, and its reintroduction accelerates apoptosis in both prostate and breast cancer cell lines^{21,22,34}. However, the potential role of the Gas5-SR interaction in cancer cell proliferation and survival has not been studied. To test whether a Gas5 mutant specifically compromised in SR interaction displays aberrant function in SR-driven cancer cell lines, we tested the ability of both the WT and G549A Gas5 to drive apoptosis when expressed in 22Rv1 prostate cancer cells. Transfection of both the WT and mutant constructs increased Gas5 expression compared to the vector control (Fig. 5a), but only the WT sequence caused caspase activation in these cells, and this was accompanied by a corresponding decrease in cell viability (Fig. 5b,c). The G549A mutant, in contrast, had no effect on either parameter (Fig. 5b,c). We repeated these experiments in the PR+/ER+ breast cancer cell line MCF-7; again, the G549A mutant completely prevented the pro-apoptotic phenotype induced by transfection of WT Gas5 (Fig. 5d-f). Thus, in steroid-driven cancer cells such as 22Rv1 cells, Gas5 function appears to be directly related to the SR-Gas5 signaling axis.

The expression of Gas5 variants has differential effects in immune cells which are less dependent on SR signaling for growth¹⁸. To determine the role of SR-Gas5 interactions in apoptosis and viability in these cells, we transfected three suspension cell lines, JeKo-1, CEM-C7, and Jurkat, with the WT and G549A Gas5 constructs (Fig. 5g). Among all three cell lines, transfection of WT Gas5 induced apoptosis and a corresponding decrease in cell viability (Fig. 5h,i) as we previously reported¹⁸. Interestingly, the G549A mutation only partially reversed this phenotype (Fig. 5h,i), in contrast to the complete reversal of apoptosis induction seen in steroid-driven cancer cell lines (Fig. 5b-c, e-f). These results indicate that Gas5 has functions beyond regulating SRs, such as mediating mTOR signaling²⁰ and additional regions of the Gas5 transcript play important roles in regulating apoptosis in immune cells.

GR cannot bind its own transcribed genomic binding sites

Since 70% of the genome is transcribed, there may be many Gas5-like RNAs capable of binding and modulating the activity of SRs. In particular, there are $\sim 2.3 \times 10^6$ genomic sequences matching a GR-binding motif³⁵, and many of these lie within transcribed regions. Due to their inverted repeat sequence, transcribed (+)GREs could form a stemloop structure much like the Gas5 GREM (Fig. 6a), which led us to inquire whether a RNA GREM could readily evolve from a DNA GR binding site. However, a (+)GRE-derived stemloop differs from the Gas5 GREM in two ways. First, the loop of the (+)GRE structure would contain a 3-base loop, whereas the GREM contains a 4-base loop. Second, the (+)GRE sequence differs slightly from the bases found in the Gas5 GREM (Fig. 6a).

To test the effects of the decreased-length loop on GR binding, which would occur upon transcription of a (+)GRE, we generated the WT Gas5 GREM with a 3-base loop UUU loop rather than the 4-base UUGU loop found in WT Gas5. Although 3-base loops are often less stable than 4-base loops, a UUU loop with a closing AU pair demonstrates adequate stability³⁶. Nonetheless, the 3-base loop Gas5 GREM showed lower affinity for GR than the WT 4-loop structure, indicating that inverted repeat sequences with 3-base spacers are unlikely to be active GR-binding elements following transcription (Fig. 6b).

However, it is possible that a SR-binding RNA sequence such as the GREM could form via a single nucleotide insertion into a (+)GRE, creating a 4 base stemloop for SR binding. To eliminate effects of the loop length, we placed the consensus (+)GRE halfsite within the Gas5 GREM stem, which is sufficient for GR binding (Fig. 1b). GR could not bind this consensus sequence with either Watson-Crick or Gas5-like wobble pairs (Fig. 6c), but required additional two additional nucleotide changes to achieve GR-binding ability (Supplementary Fig. 3). These results indicate that Gas5 independently and convergently evolved a SR-binding surface, perhaps aided by the rapid sequence evolution of lincRNAs compared to coding genes. Such a mechanism, whereby a lincRNA acquires the ability to recognize protein targets could be common, and we sought to establish the evolutionary origin of the Gas5 GREM.

The Gas5 GREM evolved from a splice acceptor site

The *GAS5* gene is annotated in only three genomes in the *Ensembl* database, and the low sequence conservation of lincRNAs makes sequence alignment across multiple species challenging. Fortunately, Gas5 hosts ten snoRNAs in humans that are highly conserved throughout vertebrates. Intron 11, just upstream of the Gas5 GREM, contains the snoRNA gene *U81*, which is positionally conserved from reptiles to mammals. The length of intronic sequence from snoRNAs to the downstream 3' acceptor splice site is restricted for proper processing of box C/D snoRNAs, with lengths of 65-75 nucleotides from a box C/D snoRNA to the downstream exon ideal for efficient snoRNA excision³⁷. Since Gas5-containing snoRNAs are critical for embryonic development¹⁶, we expected the length of intronic sequence downstream of these snoRNAs to be well-conserved. Using this anticipated high conservation of snoRNA sequence and intronic length as a guide, we were able to align the Gas5 GREM sequences from over 40 vertebrates. Gas5 does not appear to be present in jawless fish. As expected, the distance from *U81* to the downstream splice acceptor site was tightly clustered across vertebrates (Fig. 7a).

The most striking feature of the *GAS5* GREM alignment is the presence of an Alu insertion in the haplorhine lineage, belonging to the AluJo family (Fig. 7b). This insertion occurred directly at the *GAS5* intron 11/exon 12 boundary, lengthening the *GAS5* exon 12 by 100 nucleotides compared to mice. In haplorhines, an Alu-contained splice acceptor site is used for Gas5 transcript processing, maintaining the strict requirement for intron length downstream of *U81*. However, this insertion demonstrates that the GREM, while exonic in humans and other haplorhines, evolved from an ancestral splice acceptor site still present in other vertebrates. Multiple previously unassigned cDNAs from GenBank support both the alignment as well as the presence of a splice site at the GREM (Supplementary Table 3).

The AluJo insertion re-positioned the ancestral splice acceptor site to be fully contained within exon 12, and it is this now-defunct splice site through which the SRs bind Gas5 (Fig. 7b).

Due to snoRNA excision restraints, the haplorhine Gas5 GREM, over 150 bases downstream of *U81*, would be under negative selection to lose the ancestral splice site features located inside exon 12 incorporated with the AluJo insertion. However, if this locus was selected for both its splice site sequence and SR binding ability before the Alu insertion, it is possible that splice-site like sequences required for SR binding would be conserved. Remarkably, when haplorhine GREM sequences are aligned, splice site-like features are readily apparent: a polypyrimidine tract and acceptor AG sequence are conserved throughout the haplorhine clade, despite a lack of splicing at this locus (Fig. 7c). The splice-site like sequence shows higher sequence identity among 14 haplorhines (93.8% vs. 84.5% for the remainder of exon 12). Additionally, relative rate analysis of each position within the ancestral splice site reveals a lower rate of nucleotide substitution within the GREM vs. the remainder of exon 12, supporting sequence conservation of the splice-site like sequences required for GR binding to the GREM (Supplementary Fig. 4). Mutation of these splice-site like features, such as the AG within the ancestral splice acceptor sequence, reduces the GREM's affinity for GR binding (Fig. 7d), suggesting that other haplorhines contain conserved, functional GREMs with both a primary sequence and secondary structure capable of binding SRs. Indeed, the GREM sequences of two other haplorhines formed secondary structures in solution (Fig. 7e) and were capable of binding the GR DBD (Fig. 7f).

Given the evolutionary origin of the human Gas5 GREM as a splice acceptor site, it is possible that alternate splicing of the *GAS5* gene may control its ability to bind and repress GR in non-haplorhines. The mouse splice isoforms with exon 11 excised or retained bound GR *in vitro* (Supplementary Fig. 5a), and the GREM derived from both the splice or unspliced Gas5 transcripts was sufficient for secondary structure formation and GR binding (Supplementary Fig. 5b,c). When we tested for GR – Gas5 binding across the vertebrate clade, eutherian mammals demonstrated the highest affinity interaction (Supplementary Fig. 5d), suggesting that the Gas5 GREM was perhaps capable of binding SRs at the time of the AluJo insertion, roughly 55 Mya³⁸. However, further study of the mouse and other mammalian Gas5 transcripts are required to definitively determine if GR regulation is a conserved property of Gas5 outside of the haplorhine clade. This is especially true given the short lifetime of unspliced Gas5 and the shorter length of exon 12 in mice compared to haplorhine Gas5 (Fig. 7b).

Gas5 is retrotransposed within multiple mammalian lineages

Finally, given the unique ability of the GREM to bind SRs, we searched for additional GREM-like sequences within the human genome which could serve as SR-repressing elements. Remarkably, we found processed copies of Gas5 in at least five locations in the human genome, including within the introns of four genes: *GABRB3*, *NAPEPLD1*, *MDN1*, and *SV2C* (Supplementary Fig. 6). At least 29 splice variants of the *GAS5* gene are known¹⁵, and *GAS5* retrotransposition events mirror this complexity. Two human copies are fully

spliced (splice variant *GAS5*-001; Supplementary Fig. 6), whereas the sequence within the *GABRB3* gene is a copy of the *GAS5*-008 splice variant (Supplementary Fig. 6). Additionally, some retrotransposed copies of Gas5 are from unreported, species-specific, or ancestral splice variants, such as a copy in gibbons that contains exons 1-4, 6, 7, 11, and 12 (Supplementary Fig. 6). Such a splice variant is not currently described in humans.

GAS5 appears to be highly susceptible to retrotransposition, with either full or partial copies of the processed Gas5 present in multiple mammalian lineages including, mice, dogs and more distal members of the mammalian clade, such as sloths (Supplementary Table 4). At least 15 independent copies appear in the haplorhine lineage with many of these copies located within transcribed loci (Supplementary Table 4). Given the limited species distribution and high sequence identity between genomic *GAS5* sequences and these copies, the retrotransposition events described here are relatively recent. It is likely that additional, ancient copies of *GAS5* are present in mammalian genomes but have diverged beyond recognition.

Discussion

Using a range of structural and biochemical approaches, we offer a mechanism to explain the potent ability of Gas5 to suppress SR-mediated transactivation. This repression is mediated through sequence-specific protein RNA-contacts within an A-form double helical structure with a widened major groove that facilitates SR binding. We show that full-length, fully spliced Gas5 binds strongly to the GR DNA binding domain, with affinity comparable to physiological DNA glucocorticoid response elements²³. In addition, GR exhibits stronger binding to full-length Gas5 than single (+)GRE half-sites or the newly-characterized negative glucocorticoid response elements^{26,39}, illustrating the strength of the GRGas5 interaction as well as the versatility of GR as a diverse nucleic acid binding protein^{23,40,41}. GR is one of a growing number of proteins that bind DNA and RNA, a class that includes ADAR, TFIIIA, and the p50 subunit of NF- κ B⁴². p50 has been crystallized in complex with its DNA response element as well as a competing RNA aptamer^{43,44}. The RNA aptamer adopts a similar conformation as NF- κ B response elements, and p50 makes nearly identical contacts with RNA as it does with DNA⁴³; we expect a similar mechanism of DNA mimicry occurs with GR and Gas5.

Here, we show that GR-Gas5 interactions are mediated through helix 1 of GR's DNA binding domain, demonstrating that GR uses the same surface to bind both DNA and Gas5 RNA. However, the dimerization loop of the GR DBD is not affected upon Gas5 binding, although more study is required to definitively determine GR's dimerization status when bound to RNA, given recent reports demonstrating the importance of the ligand binding domain in receptor dimerization⁴⁵. While the exact Gas5 GREM sequence does not appear in any human mRNAs, we have demonstrated that this sequence evolved from a splice acceptor site; therefore, GR may be able to recognize RNA transcripts similar to Gas5 at splice site sequences; GR has been previously shown to affect the alternate splicing of several genes⁴⁶⁻⁴⁸.

The potent repression of SR signaling by Gas5 is likely explained by both the strong affinity of the SR-Gas5 interaction and by the large discrepancy between Gas5 copy number and GR DNA binding sites in a cell during periods of Gas5 upregulation. In HeLa cells, Gas5 copy number increases by an order of magnitude to roughly 75,000 during serum starvation¹⁴, which is on par with or greater than the increase in Gas5 expression caused by exogenous transfection in this study (Fig. 5a,d,g). In contrast, available genomic GR binding sites are roughly fixed in number, ranging from 3,000 to 8,000 depending on cell type³⁵. This suggests that Gas5 regulates SR activity by simply titrating its own GR binding site against genomic GR binding sites. However, it is unclear whether Gas5 interferes with DNA binding-independent, SR-mediated transrepression such as “tethered” GR transrepression⁴⁹. Future studies are needed to determine if Gas5 modulates these DNA-independent effects.

The *GAS5* gene is unique in its high level of alternative splicing: at least 29 non-redundant, alternatively-spliced transcripts originate from the *GAS5* locus¹⁵, making it one of the most highly alternatively spliced genes in the genome. Although it is not clear if Gas5 splicing patterns are tissue-specific, the majority of Gas5 splice variants include the GREM, which is sufficient for SR binding. Notably, we recently showed that exon 12-containing Gas5 transcripts are downregulated in breast ductal carcinoma compared to adjacent normal tissue²¹. Nevertheless, 13 Gas5 splice variants either exclude exon 12 or shorten exon 12 such that the GREM is not included in the final transcript⁵⁰, indicating that alternative splicing may present an additional level of SR regulation by Gas5. Such control may be required for tissue-specific homeostasis and may be disrupted by SNPs that affect the splicing of GREM-containing exon 12 (rs11537772, rs111755386). An additional SNP that alters the GREM sequence (G559U) has also been described (rs186249529). Mutation of G559 in this study, even when combined with a compensatory mutation to preserve secondary structure, dramatically weakens the GR-Gas5 interaction (Fig. 4e), potentially attenuating riborepression of SR signaling. Thus, it is possible these and other Gas5 SNPs may predispose individuals to steroid-driven cancers despite apparently normal tissue levels of Gas5.

Given the paucity of identified sequence-specific lincRNA function, we used the detailed biochemical and structural data here to examine the evolutionary origin of the Gas5 GREM. The SR DBDs are highly conserved throughout vertebrates; in contrast, two recent studies with lincRNA functional domains show mixed levels of conservation: *megamind* and *cyrano* show conserved function between zebrafish and mammals¹², whereas key functional human domains of the lincRNA HOTAIR are missing in mice⁵¹. Given the comparatively faster sequence turnover within lincRNAs compared to their protein targets, we propose that Gas5 co-opted a splice acceptor site into its current SR binding role. Such a mechanism allows the primary sequence required for SR binding to be conserved due to the presence of a second, stronger pressure for conservation. The relatively loose sequence requirements of an acceptor splice site – a polypyrimidine tract followed by AG – may have allowed this locus to gradually acquire SR-binding ability in eutherian mammals (Supplementary Fig. 5d). However, it is clear that once Gas5-SR binding was established, this function was conserved, given the retention of splice-site like sequences required for SR binding at the GREM (Fig. 7c). The development of this functional domain within the translationally-

regulated 5' TOP Gas5 transcript may have provided an important link between translational down-regulation and other critical events involved in growth arrest¹⁸.

Given the large number – tens of thousands - of lincRNAs, it is likely that hundreds of lincRNAs have precise, sequence-specific functions analogous to that of Gas5. Furthermore, given their large size, it is likely that some lincRNAs contain multiple protein-recognition sites and may serve as scaffolds for protein complexes. Here, we have demonstrated that the interaction between Gas5 and steroid receptors depends on unique structural elements and precise, sequence-specific interactions to differentiate among target proteins. Subtle disruptions of these interactions have dramatic effects on cell fate. Further, we show that the lincRNA Gas5 co-opted splicing regulatory sequences into a SR-binding role, explaining functional and sequence conservation at the GREM. Together, these results demonstrate that lincRNAs can contain conserved, sequence-specific protein interaction surfaces that ensure a high level of specificity to bind to target proteins and control cell fate.

Methods

Protein expression and purification

The DBDs of the human glucocorticoid receptor (GenBank ADP91253.1; residues L381-E541), androgen receptor (UniProtKB P10275.2; residues D550-K630), progesterone receptor (UniProtKB P06401.4; residues E560-F642), mineralocorticoid receptor (GenBank AAA59571.1; residues S593-G671), and estrogen receptor (UniProtKB P03372.2; residues A175-R260) were cloned as 6X Histidine fusions into the pMCSG7 vector. Mutations were made with the QuikChange site-directed mutagenesis kit (Stratagene). The DBDs were expressed based on previous protocols^{26,52}; transformed BL-21(DE3)pLysS *E. coli* were induced with 0.3 mM IPTG for four hours at 30 °C. Cells were lysed in 1 M NaCl, 20 mM Tris-HCl pH 7.4, 25 mM imidazole, 5% glycerol via sonication, and protein was purified via affinity chromatography (HisTrap) followed by TEV protease cleavage and dialysis to 100 mM NaCl, 20 mM Tris-HCl pH 7.4, 25 mM imidazole, and 5% glycerol. The DBDs and tags were separated by affinity chromatography (HisTrap), and further purified by gel filtration chromatography. For storage, protein was concentrated to 4 mg ml⁻¹, flash frozen in liquid N₂, and stored at -80 °C. For ¹⁵N labeling, the GR DBD was expressed in M9 minimal media with ¹⁵NH₄Cl as the nitrogen source. Cultures were inoculated at an OD₆₀₀ of 0.3-0.4, and induced for 14 h with 1 mM IPTG at an OD₆₀₀ of 1.2. The GR DBD was then purified via affinity chromatography as above.

Nucleic acid synthesis and purification

Constructs for *in vitro* transcription were ligated into the puc57 vector using a forward primer with the T7 promoter sequence and a reverse primer using the HindIII restriction site. Plasmid was cut overnight with HindIII (Fermentas) and used as template for *in vitro* transcription in 50 mM Tris-HCl pH 7.5, 15 mM MgCl₂, 5 mM DTT and 2mM spermidine with 2 mM each of individual NTPs. Recombinant T7 RNA polymerase (0.1 mg ml⁻¹ per reaction) and inorganic pyrophosphatase (1 U ml⁻¹ per reaction, Fermentas) were added and the reaction was incubated for 5 hours at 37 °C. RNA was purified by anion exchange chromatography as described previously⁵³ and ethanol precipitated. Both *in vitro* transcribed

and synthetic nucleic acid duplexes (Integrated DNA Technologies) were annealed in 10 mM NaCl, 20 mM Tris-HCl pH 8.0 by heating to 85 °C and slow cooling to room temperature.

Nucleic acid binding assays

Fluorescence polarization (FP) was used to detect the formation of DBD-nucleic acid complexes. Indicated amounts of DBD were added to wells containing 10 nM of 5' 6-FAM-labeled nucleic acids. For competition assays, labeled DNA was used at a concentration 1.2 times higher than the GR – (+)GRE K_d . Reactions were performed in 100 mM NaCl, 20 mM Tris-HCl pH 7.4, 5% glycerol and measured with a Biotek Synergy plate-reader at an excitation/emission wavelength of 485/528 nm. FP data were analyzed and graphed using GraphPad Prism 5 (Graphpad Software, Inc.). Error is shown as \pm SEM, and was normalized to the mean for figures showing $-\log K_d$. Sequences used in binding experiments can be found in Supplementary Fig. 1. All binding experiments were performed at least in duplicate, with three internal replicates.

UV melting assays

UV melting assays were conducted as described previously⁵⁴. Briefly, samples containing 25 μ g RNA in a solution of 10 mM Tris pH 7.4 and 20 mM NaCl were melted and UV absorption collected on a Varian Cary 400 UV/Vis spectrophotometer. Sample temperatures were measured with an in-cell temperature probe. First derivatives of the melting curves were calculated in GraphPad Prism 5.

Hydroxyl radical probing

In vitro transcribed Gas5 GREM RNA was treated with FastAP (Fermentas) and 5'-end labeled using [γ -³²P]ATP and T4 polynucleotide kinase. Labeled RNA was urea denaturing gel purified and annealed by heating at 65 °C for 10 minutes and slowly cooling to room temperature prior to use. GR was dialyzed against 20 mM Tris pH 7.5, 100 mM NaCl. In a total volume of 25 μ l, labeled RNA (50,000 cpm) was incubated without or with GR (0.2, 0.8, 3 or 12 μ M) and hydroxyl radicals produced by adding 1 μ l each of 50 mM Fe(SO₄)₂, 100 mM EDTA, 250 mM ascorbic acid and 3% (v/v) hydrogen peroxide. The reaction was allowed to proceed on ice for 5 minutes before quenching by ethanol precipitation in the presence of yeast tRNA (0.4 mg ml⁻¹) as a carrier and resolving the RNA products by 12% urea denaturing-PAGE for 1.5 hours at 55 W. Control experiments with labeled RNA and GR were performed to ensure no contaminating RNases were present in the *E. coli*-expressed protein. Band quantification was performed with ImageQuant applying the Rubber Band background subtraction. For normalization, the intensity of each band was measured and normalized to the most intense band in 0 μ M GR. The normalized intensities were compared between 0 μ M and 12 μ M GR. A nucleotide was considered strongly protected/ enhanced if the difference in normalized intensity was greater (+/-) 20%, or moderately protected if the difference was 10%.

Crystallization and structure analysis

Bipyramidal crystals of the Gas5 GREM duplex appeared in two conditions. Crystals grown in 0.05 M sodium succinate pH 5.5, 0.5 mM spermine, 0.02 M MgCl₂, and 3 M ammonium sulfate were frozen directly in liquid N₂. Data from these crystals were collected at 100K at the Emory University home source to a resolution of 2.2 Å at a wavelength of 1.54 Å. Crystals that formed in 0.1 M Citric Acid pH 4, and 3 M ammonium sulfate were cryoprotected in 0.1 M citric acid pH 4, 3 M ammonium sulfate, and 10% glycerol and flash-frozen in liquid N₂. Data from these crystals were collected at the Advanced Photon Source BM-22 at a wavelength of 1.00 Å, and data from both crystals were integrated and indexed in HKL-2000. Phases were determined using molecular replacement from an ideal RNA helix modeled in Coot⁵⁵, using PHASER in the PHENIX suite⁵⁶. Crystal twinning was detected using phenix.xtriage, and twin refinement was performed in phenix.refine⁵⁶. Molprobit indicated no incorrect sugar puckers, bad bonds, or bad angles in both crystal structures. The composite omit map utility within PHENIX was used to generate the simulated annealing composite omit map⁵⁶. Representative electron density is shown in Supplementary Fig. 7. RNA and DNA structure analysis was performed in 3DNA⁵⁷, using refined P-P distances to determine major and minor groove widths. Figures of models and crystal structures were generated in the PyMOL Molecular Graphics System (Schrödinger LLC).

NMR

NMR data were collected at 25 °C on a Bruker 700 MHz ¹H frequency equipped with a QCI cryoprobe. For RNA NMR experiments, the 33 nt Gas5 RNA (Supplementary Fig. 1) was reconstituted after ethanol precipitation in 20 mM phosphate, 50 mM KCl, 10% D₂O buffer, pH 7.4 to a final concentration of 10 mg ml⁻¹. The RNA sample was annealed by denaturing at 95 °C for 5 minutes and equilibrating to room temperature (20-23 °C) overnight. Two-dimensional ¹H-detected NOESY and TOCSY spectra were collected using 100 ms and 200 ms mixing times for Gas5 RNA and after adding equimolar concentration of GR DBD. The proton carrier frequency was set coincident with the water resonance for all experiments. For protein NMR experiments, 2D [¹H,¹⁵N]-HSQC spectra were collected for free ¹⁵N-labeled GR DBD protein or complexed with 33 nt Gas5 RNA in the same buffer as above. Chemical shift perturbations were assigned using previously published GR DBD NMR chemical shifts²⁹ and calculated using the minimum chemical shift perturbation procedure⁵⁸. Data were analyzed with NMRViewJ (OneMoon Scientific, Inc.)⁵⁹.

Modeling of the GR DBD-Gas5 RNA complex

NMR chemical shift perturbation and mutagenesis data was used as input to HADDOCK³¹ to model the GR DBD-Gas5 RNA complex. Prior to docking, the AMBER 12 package⁶⁰ was used to perform molecular dynamics on the crystal structure of the Gas5 GREM duplex in order to widen the major groove. Briefly, the tleap module was used to solvate the Gas5 GREM using a TIP3P octahedral box water model⁶¹ and neutralize the system. The ff12SB force field was employed in the calculations with the sander module used for the equilibration period and the pmemd.cuda code^{62,63} used on a graphics processor unit (GPU) for the production runs. The system was equilibrated in a multi-step process⁶⁴, starting with

10000 steps of energy minimization (5000 steps of steepest descent, followed by 5000 steps of conjugate gradient); all non-water and non-hydrogen atoms were restrained during this minimization step. The systems were then equilibrated for 200 ps in the canonical NVT ensemble⁶⁵ during which time the temperature was ramped from 10 to 295 K during the first 100 ps and the restraints on the solute atoms were retained. Particle-MESH-Ewald periodic boundary conditions and a cutoff of 1 nm for both electrostatic and Lennard-Jones interactions were employed; the SHAKE algorithm⁶⁶ was applied to constrain bonds to hydrogen atoms, and a time step of 1 fs was used. After HADDOCK docking, the lowest energy docked model from the best scoring cluster was used as the representative model. Statistics for the highest scoring cluster are presented in Supplementary Table 5.

Cell culture

Human adherent cell lines, 22Rv1 (prostate carcinoma cells) and MCF-7 (breast carcinoma cells), and suspension cell lines, Jurkat, CEM-C7 (both T-lymphoblastic leukemia cells) and JeKo-1 (Mantle cell B lymphoma cells) were routinely cultured in R-10 medium [RPMI-1640 medium supplemented with L-glutamine (2 mM), fetal bovine serum (10%) and gentamicin (50 $\mu\text{g ml}^{-1}$)] at 37 °C in a humidified incubator with 5% CO₂. CEM-C7 and Jurkat cells were derived from ref. 67, the 22Rv1 cells used are an apoptosis-sensitive clone described in ref. 68, MCF-7 cells were purchased from ATCC, and JeKo1 cells were a gift of Dr. Aneela Majid.

For experiments to determine the influence of Gas5 on androgen signaling, 22Rv1 cells were switched to $\text{prfR-10}_{\text{DCSS}}$ [phenol red-free RPMI-1640 medium supplemented with L-glutamine (2 mM), dextran charcoal-stripped fetal bovine serum (10%) and gentamicin (50 $\mu\text{g ml}^{-1}$) at 24 h prior to transfection. All experiments were carried out using cells in the logarithmic phase of growth.

Plasmid DNA transfection

WT Gas5 (NR_002578.2 bases 1-631) was cloned into the pcDNA3.1 vector and the G549A mutation made with the Quikchange site-directed mutagenesis kit (Stratagene). 22Rv1, MCF7 cells (each at 2×10^6 cells in 0.1 ml Nucleofector solution V) and JeKo-1 cells (5×10^6 cells in 0.1 ml Nucleofector solution R) were nucleofected with 2 μg plasmid, using programs X-001, E-014 and X-001, respectively. Cells were diluted to 2 ml with R-10 medium (or $\text{prfR-10}_{\text{DCSS}}$ medium, as appropriate) and cultured in 6-well plates.

Determination of apoptosis and cell survival

Adherent cell lines were trypsinized at 24 h *post*-nucleofection, samples were collected for RNA isolation, and cells were re-plated in 12-well plates (8×10^4 cells per well). For 22Rv1 and MCF-7 cells, additional samples were irradiated with ultraviolet (UV) light to induce apoptosis before re-plating: cells (10^5 per ml medium) in plastic petri dishes (lids removed) were exposed to a dose of 40 J m^{-2} UV-C light, then immediately centrifuged and resuspended in fresh medium. After culture for 24 h (22Rv1) or 48 h (MCF7), cells were trypsinized, and adherent and non-adherent cells were combined. Suspension cell lines were also collected at 24 h following transfection for total RNA isolation and for re-plating at 2×10^5 cells per well in 12-well plates. To determine viability, cells were stained with 0.1 %

nigrosin (w/v) and counted using a hemocytometer; cells that excluded the dye, were considered to be viable. To determine apoptosis, cells were stained with acridine orange (25 $\mu\text{g ml}^{-1}$), and the proportion of cells containing condensed or fragmented chromatin was scored by fluorescence microscopy (200 cells per treatment were scored). In some experiments, the proportion of cells containing activated caspase activity was determined using a commercial Caspatag assay (CaspaTag Pan-Caspase In Situ Assay Kit; Millipore, Watford, UK) and fluorescence microscopy (200 cells per treatment were scored), according to the supplied instructions.

Clonogenic assay

Long-term survival and proliferation of transfected JeKo-1, CEM-C7, and Jurkat cells were assessed by the ability of the cells to form colonies in soft agar. An equal proportion of the culture from each experimental condition was diluted in 5 ml Iscove's medium (Sigma) containing 20% heat inactivated fetal calf serum, 10% cell conditioned medium and 0.3% Noble agar (Difco) and plated in 60 mm dishes. Dishes were also overlaid with 2.5 ml Iscove's complete medium containing 10% cell conditioned medium. The number of colonies formed was counted following 2–3 weeks incubation at 37 °C in 5% CO₂.

Determination of androgen sensitivity

22Rv1 cells were trypsinized at 24 h *post*-nucleofection and cells were re-plated in 12-well plates (8×10^4 cells per well) in $\text{prfR-10}_{\text{DCSS}}$. Cells were allowed to attach to plates for 6 h, then treated with 10 nM R1881 in $\text{prfR-10}_{\text{DCSS}}$; controls received 0.05% DMSO. After 4 h, cells were collected in TRIZOL reagent (Invitrogen, Paisley, UK) for total RNA isolation.

Real time PCR analysis of gene expression

Total RNA was isolated using TRIZOL reagent (Invitrogen, Paisley, UK), treated with RQ1 RNase-free DNase (Promega, Southampton, UK) then reverse transcribed using random hexamer priming and SuperScript II Reverse Transcriptase (Invitrogen, Paisley, UK), according to the supplied protocols. For determination of Gas5 levels, real time PCR was conducted using SensiFast Probe Hi-ROX kit (Bioline, London, UK) and Taq Man Gene Expression Assays (Applied Biosystems, Warrington, UK; assay codes Hs03464472_m1 for GAS5 and Hs99999901_m1 for 18S). For determination of expression levels of androgen-responsive genes, real time PCR was conducted using SensiFast SYBR Hi-ROX kit (Bioline, London, UK) according to the supplied instructions and primer sets specific for: KLK2 (forward: 5'-ATGTGTGCTAGAGCTTACTC-3'; reverse: 5'-AAGTGGACCCCCAGAATCAC-3'); KLK3 (forward: 5'-CCAAGTTCATGCTGTGTGCT-3'; reverse: 5'-CCCATGACGTGATACCTTGA-3'); FKBP5 (forward: 5'-CGCAGGATATACGCCAACAT-3'; reverse: 5'-GAAGTCTTCTTGCCATTGC-3'); and ACTB (forward: 5'-GTTTGAGACCTTCAACACCC-3'; reverse: 5'-ATGTCACGCACGATTTCCC-3'). Assays contained 10 ng sample cDNA in a final volume of 25 μl and were run on an ABI Prism Sequence Detection System model 7000. For adherent cell lines, standard curves, comprising 0.2 – 60 ng cDNA (prepared from parental cells), were included with each assay to allow relative quantitation; standard curves of threshold cycle (C_T) value versus log input

standard cDNA were constructed by linear regression, and the equation of the line was used to calculate input amounts of samples from their respective C_T values. For suspension cell lines, quantitation of Gas5 was determined using the comparative C_T method, using parental cells as calibrators. Gas5 levels were expressed relative to 18S rRNA, and levels of androgen-responsive genes were expressed relative to ACTB.

Evolutionary Analysis

Genomic sequences of the *GAS5* locus were downloaded from the *Ensembl* database⁵⁰. For species where Gas5 was not annotated, the exon 12 sequence was determined using splice prediction downstream of SNORD81⁶⁹. Sequences were aligned in Geneious (Biomatters), and evolutionary rate analysis was conducted with the haplorhine sequences in MEGA 5 with the Tamura-Nei model, using all sites and 3 gamma rate categories⁷⁰.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Jo Linpinsel for assistance with the UV melting experiments, Christine Dunham for the T7 polymerase expression construct, Keith Yamamoto and Lisa Watson for the NMR restraints for the GR DBD and GR DBD – Gha complex, and Naoya Kenmochi for helpful discussion regarding fish Gas5. W.H.H. was supported by the Emory-NIH Graduate Training in Pharmacological Studies Grant 5T32GM008602-14, an American Heart Association predoctoral fellowship (13PRE16920012), and E.A.O. and W.H.H. were supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under award number RO1DK095750. M.R.P, M.M-M, and G.T.W are grateful to the UK charities Breast Cancer Campaign, Leukemia and Lymphoma Research, and Prostate Cancer UK for financial support. This research was also supported by startup funds to E.A.O from Emory University and startup funds to D.J.K from The Scripps Research Institute. X-ray data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-BM beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

References

1. Galperin MY, Koonin EV. From complete genome sequence to ‘complete’ understanding? Trends Biotechnol. 2010; 28:398–406. [PubMed: 20647113]
2. Consortium IHGS. Initial sequencing and analysis of the human genome. Nature. 2001; 409:860–921. [PubMed: 11237011]
3. Kapranov P, et al. The majority of total nuclear-encoded non-ribosomal RNA in a human cell is ‘dark matter’ un-annotated RNA. BMC Biol. 2010; 8:149–149. [PubMed: 21176148]
4. Khalil AM, Rinn JL. RNA-protein interactions in human health and disease. Semin. Cell Dev. Biol. 2011
5. Guttman M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009; 458:223–7. [PubMed: 19182780]
6. Guttman M, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 2011
7. Huarte M, et al. A Large Intergenic Noncoding RNA Induced by p53 Mediates Global Gene Repression in the p53 Response. Cell. 2010; 142:409–419. [PubMed: 20673990]
8. Loewer S, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat. Genet. 2010; 42:1113–1117. [PubMed: 21057500]

9. Sharma S, et al. Dephosphorylation of the nuclear factor of activated T cells (NFAT) transcription factor is regulated by an RNA-protein scaffold complex. *Proc. Natl. Acad. Sci. U.S.A.* 2011; 108:11381–11386. [PubMed: 21709260]
10. Guttman M, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* 2010; 28:503–510. [PubMed: 20436462]
11. Ponjavic J, Ponting CP, Lunter G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* 2007; 17:556–65. [PubMed: 17387145]
12. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution. *Cell.* 2011; 147:1537–1550. [PubMed: 22196729]
13. Tsai M-C, et al. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science.* 2010; 392:689–693. [PubMed: 20616235]
14. Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* 2010; 3:ra8–ra8. [PubMed: 20124551]
15. Frankish A, Mudge JM, Thomas M, Harrow J. The importance of identifying alternative splicing in vertebrate genome annotation. *Database.* 2012; 2012
16. Higa-Nakamine S, et al. Loss of ribosomal RNA modification causes developmental defects in zebrafish. *Nucleic Acids Res.* 2011; 40:391–8. [PubMed: 21908402]
17. Gibb EA, et al. Human Cancer Long Non-Coding RNA Transcriptomes. *PLoS ONE.* 2011; 6:e25915–e25915. [PubMed: 21991387]
18. Mourtada-Maarabouni M, Hedge VL, Kirkham L, Farzaneh F, Williams GT. Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (GAS5). *J. Cell Sci.* 2010; 123:1181–1181.
19. Schneider C, King RM, Philipson L. Genes specifically expressed at growth arrest of mammalian cells. *Cell.* 1988; 54:787–93. [PubMed: 3409319]
20. Mourtada-Maarabouni M, Hasan AM, Farzaneh F, Williams GT. Inhibition of Human T-Cell Proliferation by Mammalian Target of Rapamycin (mTOR) Antagonists Requires Noncoding RNA Growth-Arrest-Specific Transcript 5 (GAS5). *Mol. Pharmacol.* 2010; 78:19–28. [PubMed: 20421347]
21. Mourtada-Maarabouni M, Pickard MR, Hedge VL, Farzaneh F, Williams GT. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene.* 2009; 28:195–208. [PubMed: 18836484]
22. Pickard MR, Mourtada-Maarabouni M, Williams GT. Long non-coding RNA GAS5 regulates apoptosis in prostate cancer cell lines. *Biochim. Biophys. Acta.* 2013; 1832:1613–1623. [PubMed: 23676682]
23. Meijssing SH, et al. DNA Binding Site Sequence Directs Glucocorticoid Receptor Structure and Activity. *Science.* 2009; 324:407–410. [PubMed: 19372434]
24. Schwabe JWR, Chapman L, Finch JT, Rhodes D. The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell.* 1993; 75:567–578. [PubMed: 8221895]
25. Tolbert BS, et al. Major groove width variations in RNA structures determined by NMR and impact of ¹³C residual chemical shift anisotropy and ¹H-¹³C residual dipolar coupling on refinement. *J. Biomol. NMR.* 2010; 47:205–19. [PubMed: 20549304]
26. Hudson WH, Youn C, Ortlund EA. The structural basis of direct glucocorticoid-mediated transrepression. *Nat. Struct. Mol. Biol.* 2013; 20:53–58. [PubMed: 23222642]
27. Battiste JL, et al. Helix-RNA Major Groove Recognition in an HIV-1 Rev Peptide-RRE RNA Complex. *Science.* 1996; 273:1547–1551. [PubMed: 8703216]
28. Weeks K, Crothers D. Major groove accessibility of RNA. *Science.* 1993; 261:1574–1577. [PubMed: 7690496]
29. Watson LC, et al. The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. *Nat. Struct. Mol. Biol.* 2013; 20:876–83. [PubMed: 23728292]

30. Luisi BF, et al. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature*. 1991; 352:497–505. [PubMed: 1865905]
31. de Vries SJ, van Dijk M, Bonvin AMJJ. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* 2010; 5:883–897. [PubMed: 20431534]
32. Strobel S, Cech T. Minor groove recognition of the conserved G.U pair at the Tetrahymena ribozyme reaction site. *Science*. 1995; 267:675–679. [PubMed: 7839142]
33. Magali F, Schimmel P. Subtle atomic group discrimination in the RNA minor groove. *Proc. Natl. Acad. Sci. U.S.A.* 1997; 94:11291–11294. [PubMed: 9326602]
34. Romanuik TL, et al. LNCaP Atlas: gene expression associated with in vivo progression to castration-recurrent prostate cancer. *BMC Med. Genomics*. 2010; 3:43. [PubMed: 20868494]
35. John S, et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* 2011; 43:264–8. [PubMed: 21258342]
36. Shu Z, Bevilacqua PC. Isolation and Characterization of Thermodynamically Stable and Unstable RNA Hairpins from a Triloop Combinatorial Library. *Biochemistry*. 1999; 38:15369–15379. [PubMed: 10563823]
37. Hirose T, Steitz JA. Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* 2001; 98:12914–9. [PubMed: 11606788]
38. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nat. Rev. Genet.* 2002; 3:370–379. [PubMed: 11988762]
39. Surjit M, et al. Widespread negative response elements mediate direct repression by agonist-liganded glucocorticoid receptor. *Cell*. 2011; 145:224–41. [PubMed: 21496643]
40. Ali M, Vedeckis W. The glucocorticoid receptor protein binds to transfer RNA. *Science*. 1987; 235:467–470. [PubMed: 3798121]
41. Ishmael FT, et al. The human glucocorticoid receptor as an RNA-binding protein: global analysis of glucocorticoid receptor-associated transcripts and identification of a target RNA motif. *J. Immunol.* 2011; 186:1189–98. [PubMed: 21148795]
42. Hudson WH, Ortlund EA. The structure, function and evolution of proteins that bind DNA and RNA. *Nat. Rev. Mol. Cell. Biol.* In press; doi: 10.1038/nrm3884.
43. Huang D-B, et al. Crystal structure of NF- κ B (p50)₂ complexed to a high-affinity RNA aptamer. *Proc. Natl. Acad. Sci. U.S.A.* 2003; 100:9268–9273. [PubMed: 12886018]
44. Müller CW, Rey FA, Sodeoka M, Verdine GL, Harrison SC. Structure of the NF- κ B p50 homodimer bound to DNA. *Nature*. 1995; 373:311–317. [PubMed: 7830764]
45. Presman DM, et al. Live cell imaging unveils multiple domain requirements for in vivo dimerization of the glucocorticoid receptor. *PLoS Biol.* 2014; 12:e1001813. [PubMed: 24642507]
46. Cote GJ, Gagel RF. Dexamethasone differentially affects the levels of calcitonin and calcitonin gene-related peptide mRNAs expressed in a human medullary thyroid carcinoma cell line. *J. Biol. Chem.* 1986; 261:15524–8. [PubMed: 3491070]
47. Kosaki A, Webster NJ. Effect of dexamethasone on the alternative splicing of the insulin receptor mRNA and insulin action in HepG2 hepatoma cells. *J. Biol. Chem.* 1993; 268:21990–21996. [PubMed: 8408055]
48. Lai G-J, McCobb DP. Opposing actions of adrenal androgens and glucocorticoids on alternative splicing of Slo potassium channels in bovine chromaffin cells. *Proc. Natl. Acad. Sci. U.S.A.* 2002; 99:7722–7727. [PubMed: 12032350]
49. Webster JC, Cidlowski JA. Mechanisms of Glucocorticoid-receptor-mediated Repression of Gene Expression. *Trends Endocrinol. Metab.* 1999; 10:396–402. [PubMed: 10542396]
50. Flicek P, et al. Ensembl 2011. *Nucleic Acids Res.* 2011; 39:D800–D806. [PubMed: 21045057]
51. Schorderet P, Duboule D. Structural and functional differences in the long non-coding RNA hotair in mouse and human. *PLoS Genet.* 2011; 7:e1002071. [PubMed: 21637793]
52. Hudson WH, Youn C, Ortlund EA. Crystal structure of the mineralocorticoid receptor DNA binding domain in complex with DNA. *PLoS ONE*. 2014; 9:e107000. [PubMed: 25188500]
53. Easton LE, Shibata Y, Lukavsky PJ. Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA*. 2010

54. Wahid AM, Coventry VK, Conn GL. The PKR-binding domain of adenovirus VA RNAI exists as a mixture of two functionally non-equivalent structures. *Nucleic Acids Res.* 2009; 37:5830–5837. [PubMed: 19635730]
55. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr.* 2004; D60:2126–2132.
56. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* 2010;213–221.
57. Lu X-J, Olson WK. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* 2008; 3:1213–1227. [PubMed: 18600227]
58. Williamson MP. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Mag. Res. Sp.* 2013; 73:1–16.
59. Johnson, B. Using NMRView to Visualize and Analyze the NMR Spectra of Macromolecules.. In: Downing, AK., editor. *Protein NMR Techniques*. Vol. 278. Humana Press; 2004. p. 313-352.
60. Case DA, et al. The Amber biomolecular simulation programs. *J. Comput. Chem.* 2005; 26:1668–88. [PubMed: 16200636]
61. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J. Biol. Chem.* 1983; 79:926–935.
62. Gotz AW, et al. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* 2012; 8:1542–1555. [PubMed: 22582031]
63. Le Grand S, Götz AW, Walker RC. SPFP: Speed without compromise, A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* 2013; 184:374–380.
64. Shang Y, Simmerling C. Molecular dynamics applied in drug discovery: the case of HIV-1 protease. *Methods Mol. Biol.* 2012; 819:527–49. [PubMed: 22183556]
65. Allen, MP.; Tildesley, DJ. *Computer simulation of liquids*. Oxford University Press; Oxford, New York: 1987.
66. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 1977; 23:327–341.
67. Williams GT, Critchlow MR, Hedge VL, O'Hare KB. Molecular failure of apoptosis: inappropriate cell survival and mutagenesis? *Toxicol. Lett.* 1998; 102-103:485–9. [PubMed: 10022300]
68. Pickard MR, Darling D, Farzaneh F, Williams GT. Preparation and characterization of prostate cell lines for functional cloning studies to identify regulators of apoptosis. *J. Androl.* 2009; 30:248–58. [PubMed: 19023142]
69. Reese M, Eeckman F, Kulp D, Haussler D. Improved Splice Site Detection in Genie. *J. Comput. Biol.* 1997; 4:311–23. [PubMed: 9278062]
70. Tamura K, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 2011; 28:2731–9. [PubMed: 21546353]

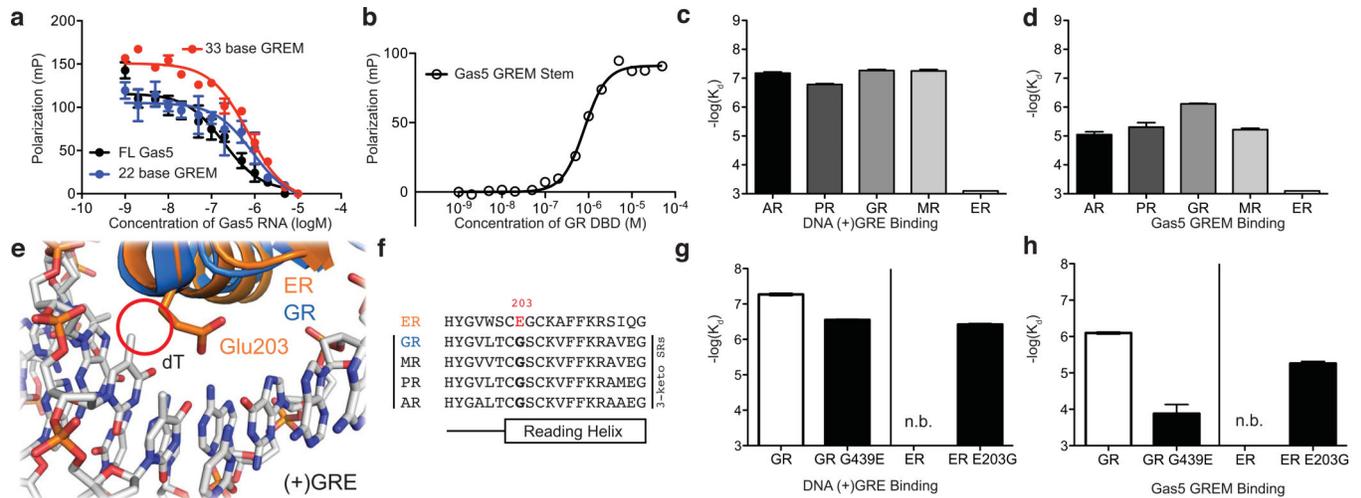


Figure 1. Gas5 selectively binds the 3-keto steroid receptors

(a) Full length Gas5 competes strongly for DNA binding to the glucocorticoid receptor (GR; $K_i = 158$ nM, fit to a one-site binding competition model). Shorter Gas5 constructs (Supplementary Fig. 1) containing only the GREM stemloop are sufficient for GR binding (K_i for the 23 and 33 base constructs are 825 and 654 nM, respectively). Error is shown as \pm s.e.m., and was normalized to the mean for figures showing $-\log K_d$. (b) A double helical region containing only the GREM stem is sufficient for direct GR binding ($K_d = 801$ nM, fit to a one-site specific binding model). (c) The 3-keto steroid receptors but not the estrogen receptor are able to bind to an activating DNA glucocorticoid response element ((+)GRE). (d) Similarly, the GREM stem construct shown in (b) binds the 3-keto steroid receptors, but not the estrogen receptor. However, the estrogen receptor is capable of binding its own response element (Supplementary Fig. 1m). (e) GR and ER recognize their cognate DNA response elements via different sequence-specific contacts. Notably, ER contains a glutamic acid at position 203 (Gly439 in GR), which excludes binding to an activating GR response element, or (+)GRE (f). (g) Mutation of this residue (Glu203) to the homologous GR residue (E203G) restores ER binding to a GRE (right), and the converse mutation in GR (left, G439E) reduces affinity for a GRE. n.b., no binding. (h) Remarkably, the ER E203G mutation confers binding to Gas5 (right), demonstrating that Gas5 exploits differences in GR/ER nucleic acid specificity to selectively target the non-ER steroid receptors. Binding experiments were performed in duplicate (full-length Gas5, panel a) or triplicate (all others), with three internal replicates each.

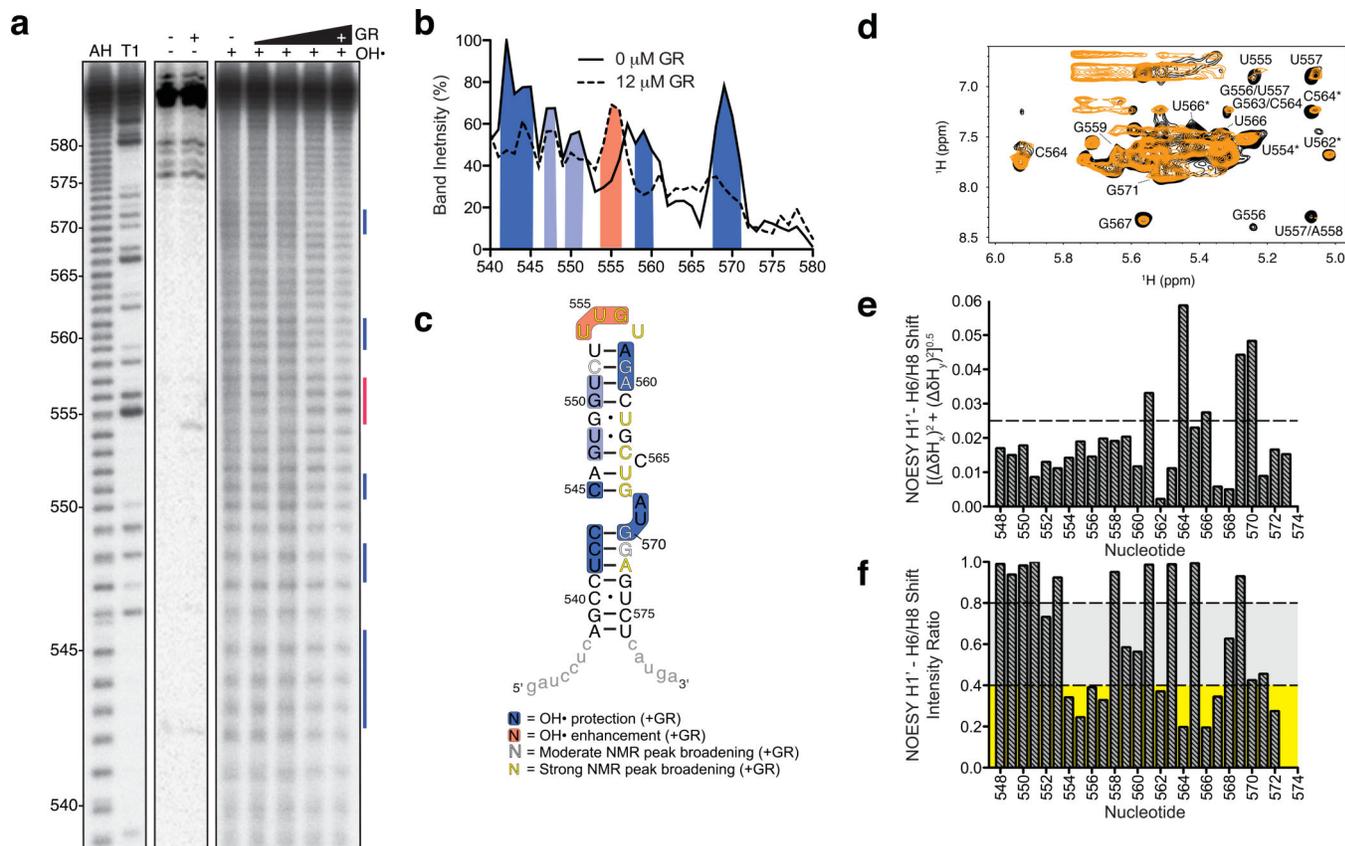


Figure 2. Base specific-interactions mediate the glucocorticoid receptor – Gas5 interaction

(a) GR protects the Gas5 GREM from OH• hydrolysis. Increasing amounts of GR DBD were added to *in vitro* transcribed Gas5 GREM RNA and subjected to OH• hydrolysis, followed by denaturing polyacrylamide gel electrophoresis; a representative gel is shown. GR-mediated protection from OH• hydrolysis is quantified in (b). Light and dark blue indicate moderate and strong protection (> +/- 10 and 20%, respectively) with GR DBD, and red indicates deprotection in the presence of GR DBD. (c) Results from NMR and OH• footprinting are mapped onto the Gas5 GREM. (d) Zoomed in overlay of free Gas5 RNA (black) and 1:1 Gas5 RNA:GR DNA binding domain (DBD; orange) 2D [^1H , ^1H]-NOESY spectra with 200 ms mixing time. Broadened resonances in the presence of GR DBD are labeled. Intra-nucleotide H1'-H6/H8 resonances are labeled according to nucleotide type and position number. H5-H6 peaks are labeled in the same way, followed by an asterisk (*). Peak labels separated by a slash (/) are for inter-nucleotide H1'-H6/H8 NOESY crosspeaks. (e) NMR crosspeak shifting calculated from NOESY spectra for free and GR DBD-bound Gas5 GREM stem reveals strongly shifted resonances (changes >0.025 ppm; horizontal dashed line). (f) NMR crosspeak line broadening determined from NOESY spectra as calculated from the crosspeak intensity ratio for free and GR-DBD bound Gas5 GREM stem (“moderately broadened” resonances, >0.4, gray; “strongly broadened” resonances, >0.8, yellow).

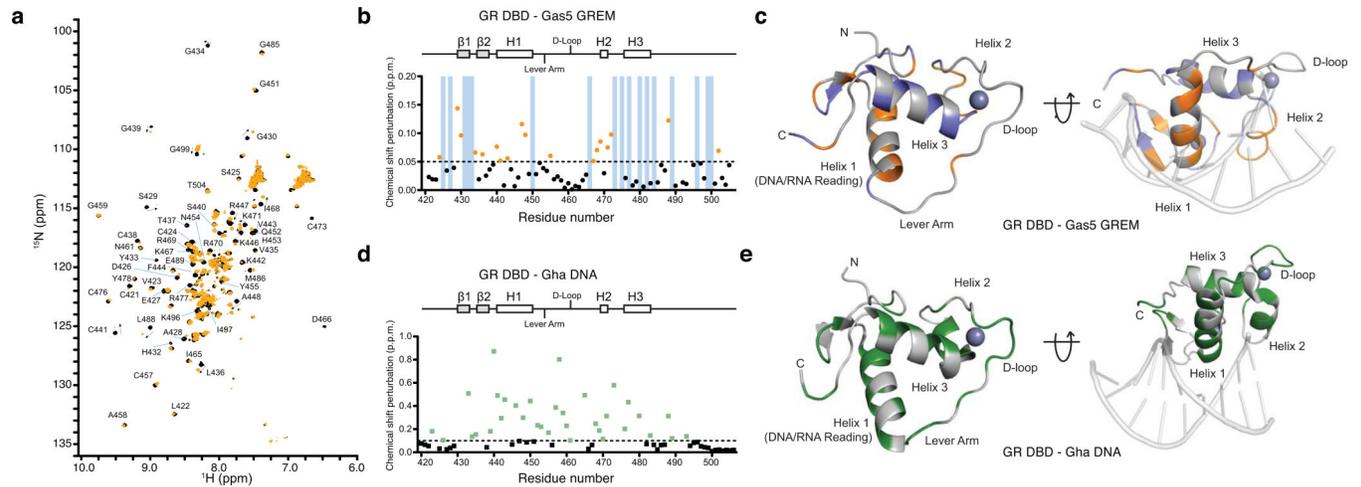


Figure 3. NMR reveals that Gas5 interacts directly with residues that comprise GR's DNA-binding interface

(a) 2D [^1H , ^{15}N]-HSQC spectra of GR DBD free (black) and bound to the Gas5 GREM (orange). (b) Chemical shift perturbations induced in GR DBD upon binding Gas5. Perturbations greater than 0.05 ppm are colored orange; blue vertical lines represent resonances that broaden upon binding Gas5 and are no longer observed. (c) GR chemical shift perturbations upon GREM binding mapped onto the HADDOCK model of GR DBD bound to the Gas5 GREM. (d) Chemical shift perturbations greater than 0.1 ppm (green) induced in GR DBD upon binding Gha DNA (Watson *et al.*, 2013) plotted against GR DBD primary and secondary structure. (e) GR chemical shifts upon Gha DNA binding mapped onto the GR DBD crystal structure (PDB 4HN5).

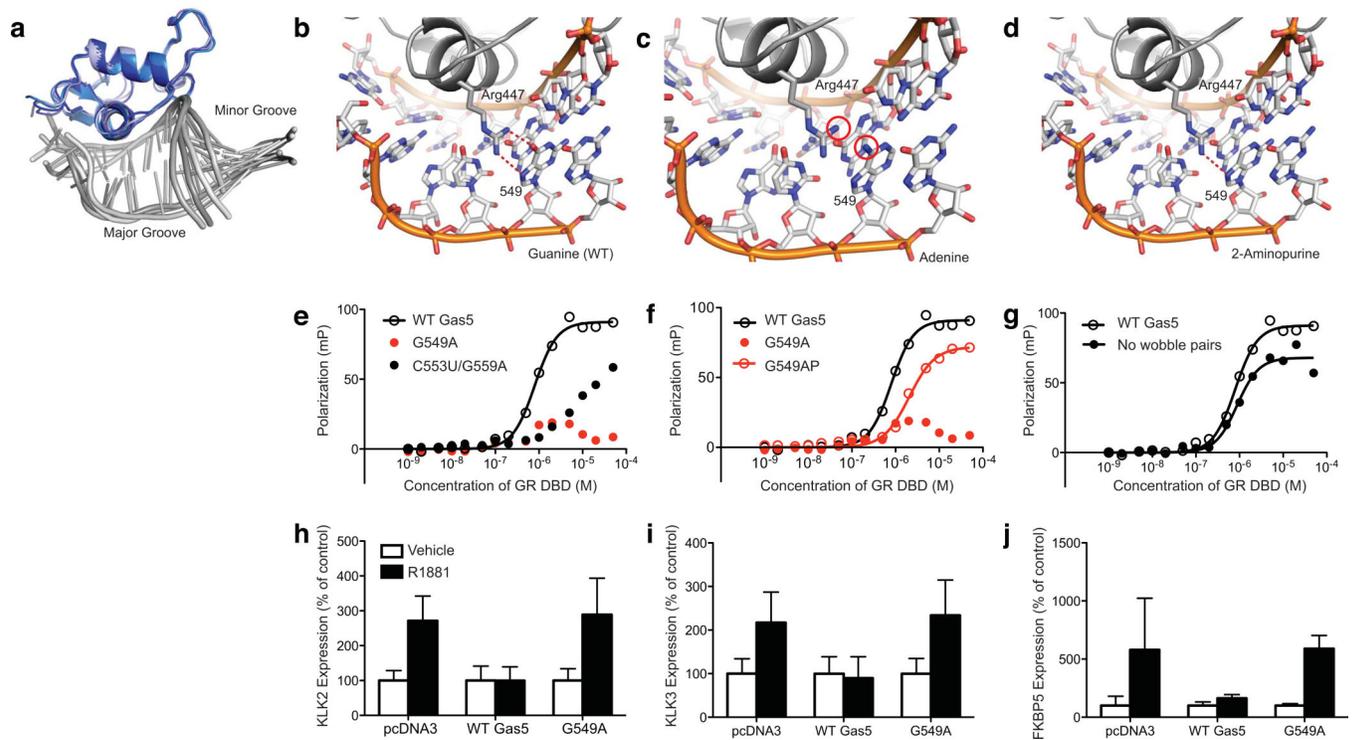


Figure 4. A single nucleotide mutation within Gas5 eliminates its ability to bind steroid receptors and repress steroid-driven transcription

(a) Superposition of the highest-scoring cluster of GR – Gas5 GREM models generated by HADDOCK. GR binds in the major groove to the A-form helix of the Gas5 GREM. (b) Guanine 549 presents a carbonyl moiety toward the major groove, which GR may recognize with a basic amino acid such as Arg447 (modeled here by HADDOCK) or Lys442. (c) Mutation of guanine 549 to adenine introduces an amine in the major groove, which would clash with the basic major-groove reading amino acids of GR. However, mutation of this guanine to 2-aminopurine (d) removes this carbonyl moiety without affecting the minor groove interaction surface. (e) Mutation of the two guanine bases, including position 549, identified by structure probing (Fig. 2) to adenine ablates binding of GR to the Gas5 GREM stem. Binding (panels e-g) data are fit to a one-site specific binding model. (f) The G549A mutation ablates GR binding to the Gas5 GREM stem. However, mutation of G549 to 2-aminopurine reduces GR – Gas5 GREM stem binding only 2.5-fold, indicating that GR contacts the major groove of Gas5 (see panels b-d). (g) Two G-U wobble pairs (including G549) are present in the GREM; however, Watson-Crick G-C pairs are also compatible with GR binding. (h-j) The G549A mutation, which ablates SR-Gas5 binding *in vitro*, also compromises Gas5-mediated repression of steroid-driven transcriptional activation in 22Rv1 cells. Three androgen receptor-responsive genes were measured and normalized to *ACTB* via RT-PCR in 22Rv1 prostate cancer cells with and without the AR agonist R1881. In all panels, error is shown as \pm s.e.m, and binding experiments and RT-PCR were both performed in triplicate.

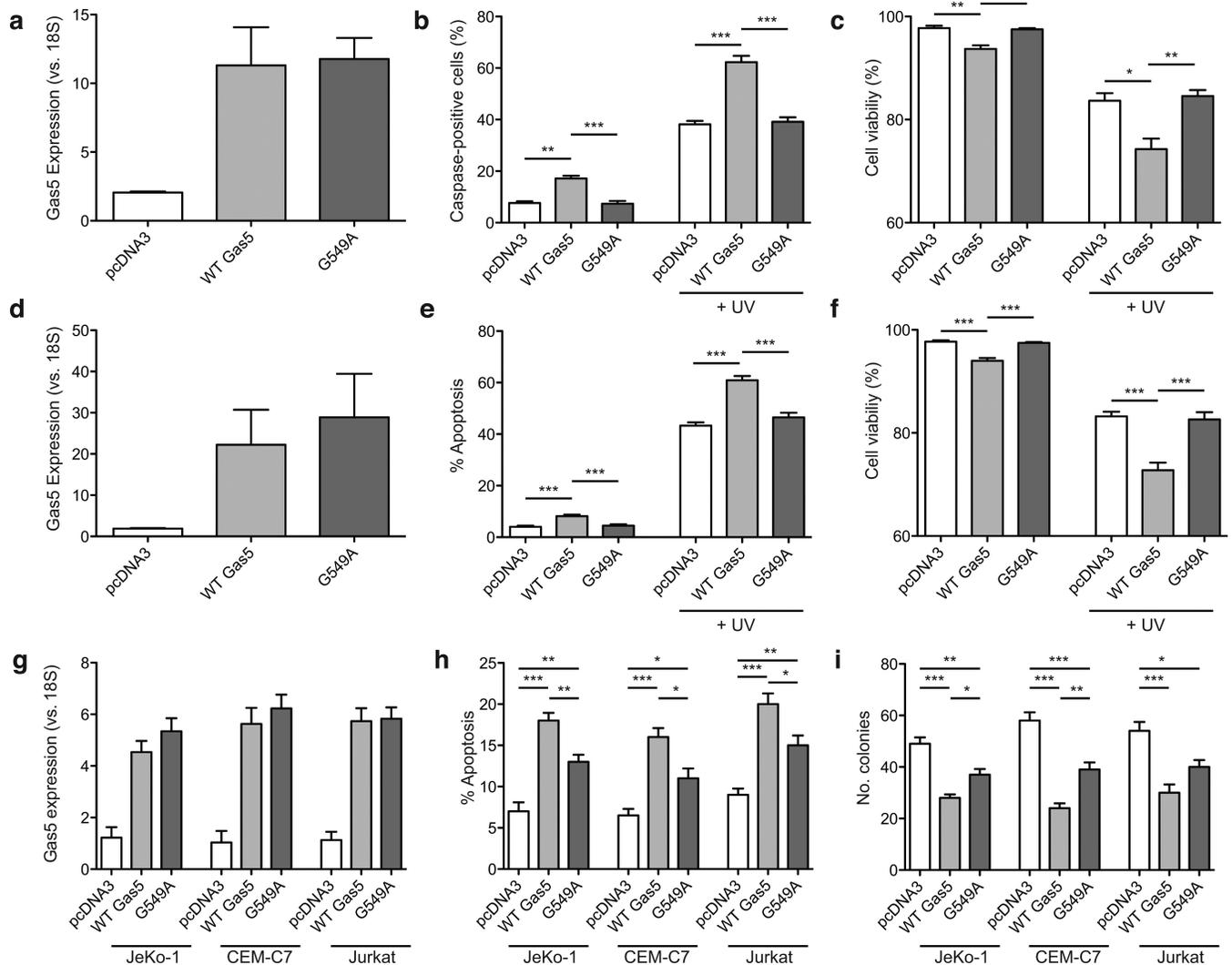


Figure 5. Selective ablation of the steroid receptor-Gas5 interaction has dramatic effects on cell fate

(a) Quantification of Gas5 levels normalized to 18S RNA in 22Rv1 cells after transfection with WT Gas5, G549A Gas5, and pcDNA3 constructs. (b) WT Gas5 induces caspase-dependent cell death in 22Rv1 cells, but this effect is completely reversed in the G549A mutant. Similar effects were seen for cell viability (panel c). (d) MCF-7 cells were transfected with identical constructs, and levels quantified via RT-PCR and normalized to 18S RNA. The G549A mutation completely ablated the pro-apoptotic (e) effects and reduction in cell viability (f) caused by transfection of WT Gas5. (g) Likewise, three suspension cell lines were transfected with pcDNA3, WT Gas5, or the G549A mutant and expression were quantified as above. In these cell lines, the G549A mutation only partially reverses the apoptotic effects (h) and colony-forming effects (i) seen by transfection of WT Gas5. For all panels, experiments were performed in triplicate, and error is shown as \pm s.e.m. and was normalized to the mean for figures showing $-\log K_d$. One, two, and three asterisks indicate $p < 0.05$, 0.01 , and 0.001 , respectively by one-way ANOVA with Tukey's post-hoc test.

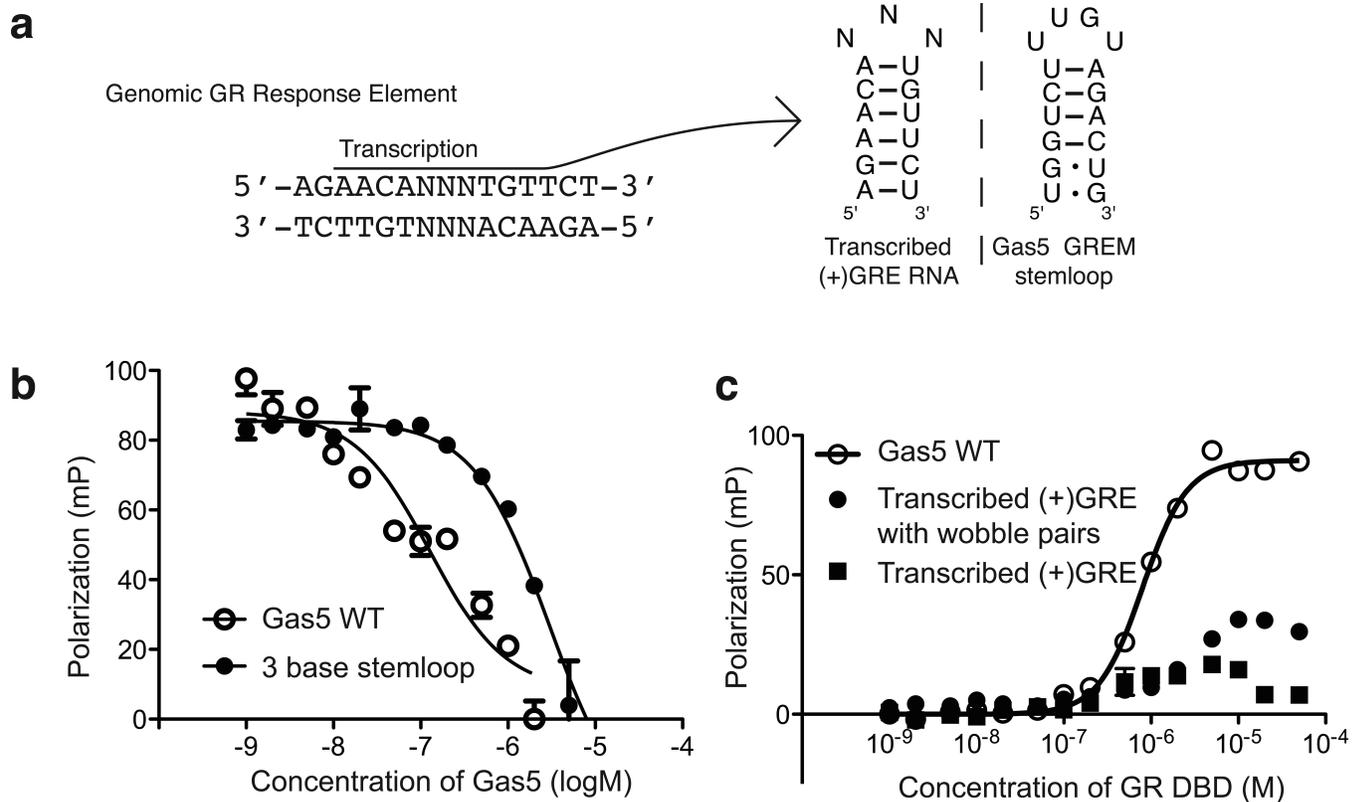


Figure 6. The glucocorticoid receptor will not bind its transcribed genomic response elements (a) Within genomic DNA, GR binds (+)GREs, which are inverted repeats separated by three nucleotides. Such inverted repeat sequences would form a stemloop with similar sequence to the Gas5 GREM when transcribed (on right). However, this sequence is incapable of binding GR due to both secondary structure constraints caused by a 3 base stemloop (b) as well as a different primary sequence (c). (b) Competition for a fluorescently-labeled (+)GRE DNA construct was performed with a GREM-containing 51 nt construct of WT or the same construct with a 3-base stemloop (G556). The affinity of the WT construct for GR (K_i) was 125 nM, similar to that of full-length Gas5 (Fig. 1a). (c) A fluorescently-labeled transcribed (+)GRE sequence with Gas5 GREM stem flanking sequences was tested for binding to the GR DBD. Binding (panels b,c) data are fit to a one-site competition or specific binding model, with error shown as \pm s.e.m. Binding experiments were performed in duplicate (panel b) or triplicate (c), with three internal replicates each.

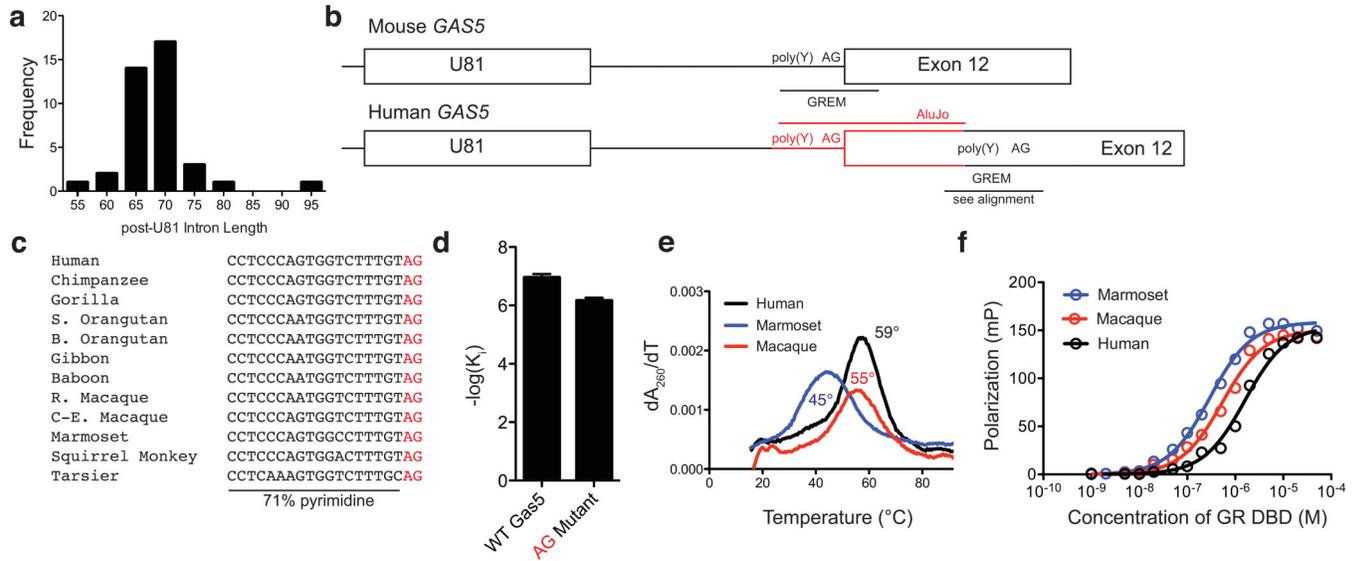


Figure 7. The Gas5 GREM evolved from a splice acceptor site

(a) The downstream intron length from the snoRNA U81 is highly conserved due to excision constraints, illustrating the importance of sequence and positional conservation of the splice acceptor site at this locus. (b) The Gas5 GREM was interrupted by an AluJo insertion (red) in haplorhines, moving the ancestral splice acceptor site into exon 12, where splicing no longer occurs. Poly(Y) indicates the position of a polypyrimidine tract. (c) However, despite the lack of splicing at the haplorhine GREM, the haplorhine GREM maintains a splice-site like sequence with a polypyrimidine tract followed by an AG dinucleotide, which is required for high-affinity steroid receptor–Gas5 binding (mutated to GA in panel d). Other haplorhine species contain Gas5 GREM sequences that exhibit similar secondary structure (e) and GR-binding affinity to the human Gas5 GREM (f). Error is shown as \pm s.e.m., and was normalized to the mean for figures showing $-\log K_d$. Binding data (panel f) were fit to a one site, specific binding model. Binding experiments were performed in duplicate (panel d) or triplicate (panel f), with three internal replicates each.

Table 1

Data collection and refinement statistics for Gas5 GREM x-ray crystal structures.

	Gas5 GREM (4 Duplexes)	Gas5 GREM (2 Duplexes)
Data collection		
Space group	R3	R3
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	43.3, 43.3, 304.2	43.5, 43.5, 173.2
α , β , γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Resolution (Å)	1.9 (1.97 – 1.90) *	2.2 Å (2.24-2.20) *
<i>R</i> _{merge}	5.2 (42.3)	6.9 (25.6)
<i>I</i> / σI	22.5 (2.8)	28.5 (5.5)
Completeness (%)	99.7 (100.0)	99.8 (82.1)
Redundancy	2.9 (2.8)	3.5 (2.5)
Refinement		
Resolution (Å)	1.9	2.2
No. reflections	16794	6140
<i>R</i> _{work} / <i>R</i> _{free}	19.8 / 26.3	17.7 / 25.2
No. atoms		
RNA	1649	857
Water	114	85
<i>B</i> -factors		
RNA	31.6	34.6
Water	24.4	22.2
R.m.s. deviations		
Bond lengths (Å)	0.0053	0.0065
Bond angles (°)	0.77	0.77
Coordinate error (Å)	0.26	0.40

* Data for each structure were collected from a single crystal. Values in parentheses are for highest-resolution shell. Maximum-likelihood coordinate error is shown.