

RESEARCH

Open Access

# Glucocorticoid receptor binds half sites as a monomer and regulates specific target genes

Benjamin J Schiller<sup>1</sup>, Rajas Chodankar<sup>2</sup>, Lisa C Watson<sup>1</sup>, Michael R Stallcup<sup>2</sup> and Keith R Yamamoto<sup>1\*</sup>

## Abstract

**Background:** Glucocorticoid receptor (GR) is a hormone-activated, DNA-binding transcriptional regulatory factor that controls inflammation, metabolism, stress responses, and other physiological processes. *In vitro*, GR binds as an inverted dimer to a motif consisting of two imperfectly palindromic 6 bp half sites separated by 3 bp spacers. *In vivo*, GR employs different patterns of functional surfaces of GR to regulate different target genes. The relationships between GR genomic binding and functional surface utilization have not been defined.

**Results:** We find that A477T, a GR mutant that disrupts the dimerization interface, differs from wild-type GR $\alpha$  in binding and regulation of target genes. Genomic regions strongly occupied by A477T are enriched for a novel half site motif. *In vitro*, GR $\alpha$  binds half sites as a monomer. Through the overlap between GR $\alpha$ - and A477T-bound regions, we identify GR $\alpha$ -bound regions containing only half sites. We further identify GR target genes linked with half sites and not with the full motif.

**Conclusions:** Genomic regions bound by GR differ in underlying DNA sequence motifs and in the GR functional surfaces employed for regulation. Identification of GR binding regions that selectively utilize particular GR surfaces may discriminate sub-motifs, including the half site motif, that favor those surfaces. This approach may contribute to predictive models for GR activity and therapy.

## Background

Glucocorticoid receptor (GR, HUGO symbol NR3C1) is a DNA-binding transcriptional regulatory factor that is activated specifically by binding glucocorticoid hormones, and which regulates diverse aspects of physiology. Genome-wide measurements of gene expression in U2OS human osteosarcoma cells stably expressing GR [1] have revealed that thousands of genes are differentially expressed in response to glucocorticoid treatment (that is, regulated by GR) [2]. Moreover, U2OS lines stably expressing three mutant GR alleles, each disrupting one functional surface, have revealed that different GR target genes depend on different patterns of functional surfaces of GR for proper regulation; A477T disrupts the dimerization interface; and 30iiB (E219K/F220L/W234) and E773R disrupt the activation function 1 and 2 (AF1, AF2) domains, respectively (Figure 1A) [2]. Similarly, GR $\gamma$ , a minor isoform differing from the major

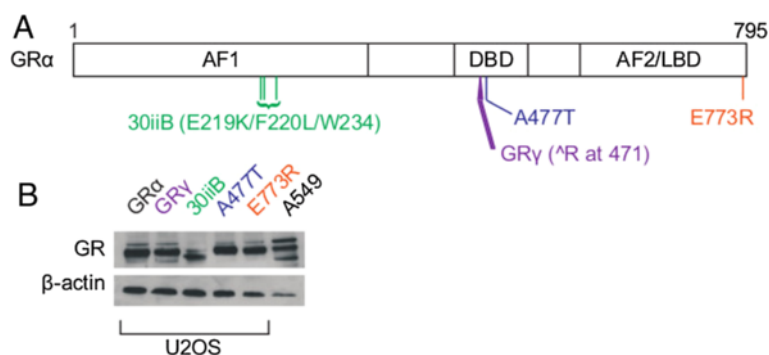
isoform GR $\alpha$  by insertion of a single arginine in the 'lever arm' region [3], differentially regulates some GR regulated genes, implicating this region as another functional surface. X-ray crystallography [3] and nuclear magnetic resonance spectroscopy (NMR) [4] of GR bound to different oligonucleotides containing high-affinity binding sites indicate that GR assumes distinct conformations at different sites. These results imply that GR also assumes different conformations *in vivo* at different glucocorticoid response elements (GREs), genomic regions that confer specific context-dependent GR responsiveness upon a nearby gene.

*In vitro*, purified GR recognizes specifically a GR binding sequence (GBS) motif composed of imperfect palindromic 6 base pair (bp) 'half sites' separated by 3 bp 'spacers', binding as an inverted dimer [5]. *In vivo*, GR occupies specific genomic GR binding regions (GBRs), identified using chromatin immunoprecipitation coupled with quantitative PCR (ChIP) or deep sequencing (ChIP-seq). The genomic locations of GBRs differ strikingly in different cell or physiologic contexts, and not all GBRs contain an identifiable GBS motif. Hence, the relationship between

\* Correspondence: yamamoto@cmp.ucsf.edu

<sup>1</sup>Department of Cellular and Molecular Pharmacology, University of California, 600 16th Street, S-222, San Francisco, CA 94158-2280, USA

Full list of author information is available at the end of the article



**Figure 1 GR alleles used in this study.** Rat GR $\alpha$ , the major isoform of GR, is a 795 amino acid protein with a modular structure typical of nuclear receptors: an N-terminal activation function 1 (AF1) domain, a DNA-binding domain (DBD), and a C-terminal ligand binding domain (LBD), which contains the activation function 2 (AF2) domain. We used mutations of GR $\alpha$  in AF1 (30iiB); AF2 (E773R); the dimerization interface of the DBD (A477T); and a minor isoform GR $\gamma$ , which bears a single amino acid (arginine) insertion at position 471 in GR $\alpha$  as a result of alternative splicing. Domain boundaries are approximate, based on functional assays. Western blots show that GR was expressed at similar levels in all cell lines, which were similar to the level of endogenous GR found in the lung adenocarcinomic human alveolar basal epithelial cell line A549.

sequences recognized *in vitro* and genomic regions occupied *in vivo* is complex. It is reassuring that GBS motifs drive GR-regulated transcription in a simple reporter context [5], and therefore reasonable to assume that GR occupancy at GBRs that contain GBSs reflects *in vivo* sequence recognition. Interestingly, different GBSs, tested in isolation *in vivo*, trigger utilization of different patterns of GR functional surfaces [3], supporting the notion that GR assumes different conformations at these sequences *in vivo*. However, rules dictating the relationship between DNA sequence, GR conformation, and utilization of distinct GR functional surfaces remain unknown.

We therefore examined whether sequence motifs, including the canonical GBS, found within GBRs might in fact be an aggregate of specific motifs, each biased toward utilization of particular functional surfaces of GR. In this study, we measured changes in gene expression and genomic occupancy upon glucocorticoid treatment of isogenic cultured human cell lines expressing alleles of GR bearing differences in particular functional surfaces. We further used statistical, computational, and biophysical methods to characterize changes in GR conformation and function, as well as the sequence motifs that are present at GREs, in relation to these functional surfaces.

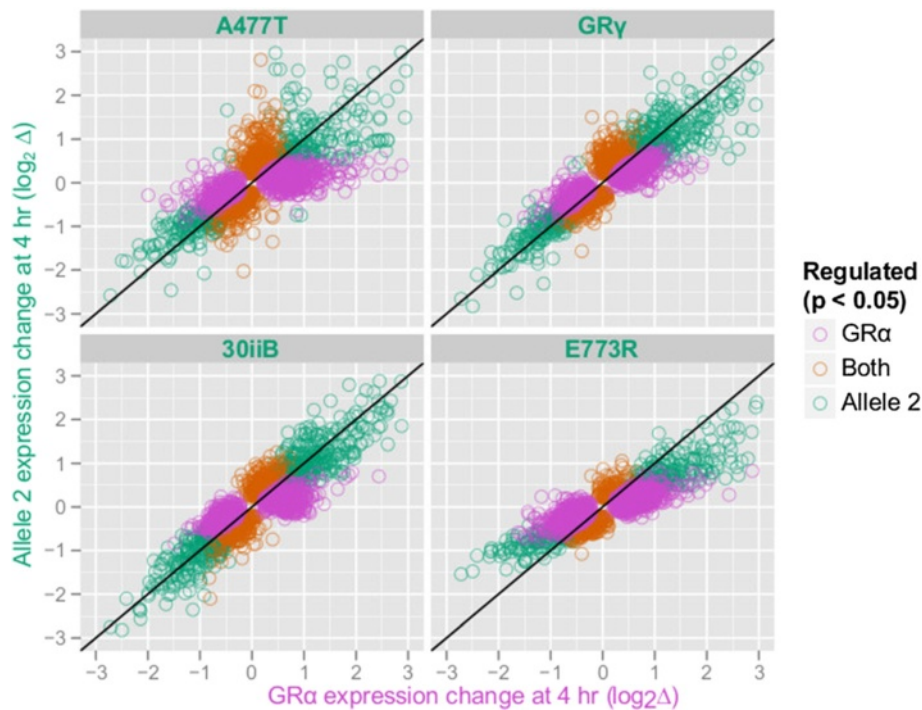
## Results

### Different GR alleles induce diverse, gene-specific transcriptional responses

To improve our understanding of the patterns of GR functional surfaces used during transcriptional regulation, we carried out a genome-wide analysis in a series of human U2OS osteosarcoma cell lines with stably integrated GR alleles [2,6]. We isolated RNA from U2OS cells expressing integrated rat GR $\alpha$ , GR $\gamma$ , GR $\alpha$  30iiB (30iiB; E219K/F220L/W234R), GR $\alpha$  A477T (A477T), or

GR $\alpha$  E773R (E773R) (Figure 1A, B) after treatment for 2, 4, or 24 h with 100 nM dexamethasone, a synthetic glucocorticoid that binds GR with high affinity, or for 4 h with ethanol only as a '0 hour' control. Each GR $\alpha$  mutation alters in a distinct way the utilization of various GR regulatory surfaces; among cells expressing the 30iiB, A477T, or E773R mutations, a panel of 10 regulated genes displayed six out of the eight possible patterns of surface utilization of three surfaces compared to GR $\alpha$ ; that is, (2 states, utilizes or does not)<sup>(3 mutations)</sup> = 8 patterns [2]. These results imply that other functional surfaces and a multitude of patterns are likely used at regulated genes. We assayed all gene expression from isolated RNA samples (Additional file 1). After 4 h of dexamethasone treatment, thousands of genes were regulated by GR [2]. Because of the statistical complexities with assessing utilization of each of the four GR surfaces at every regulated gene, we instead analyzed pairwise the regulation by GR $\alpha$  compared individually with our four GR alleles (Additional file 1). As expected, we found that regulation was affected in a highly allele-specific and gene-specific manner (Additional file 1, Figure 2).

We calculated the statistical significance of the difference in regulation between GR $\alpha$  and each allele for each gene after 4 h of dexamethasone treatment (Figure 3A to D), using limma, a software package for the analysis of gene expression microarray data using linear models. We instructed limma to estimate the regression coefficients (that is, the per-gene differences in regulation between a given allele and GR $\alpha$ ) for each such simple contrast (the difference of GR $\gamma$  and GR $\alpha$ , and so on), as well as the standard errors of those coefficients [7]. This method allows the application of a *t*-test to determine significance, which may reduce the number of identified targets but ensures a high confidence gene list by properly controlling the false discovery rate [7].



**Figure 2** Glucocorticoid regulation occurs with gene-specific utilization of GR functional surfaces. U2OS cells expressing GR alleles were treated for 4 h with dexamethasone or vehicle (ethanol). Scatterplots of expression changes ( $\log_2$ -fold change) at 4 h are shown with the GR $\alpha$  response ( $\log_2$ -fold change) on the x-axis and the response for another allele on the y-axis: A477T, GR $\gamma$ , 30iiB, or E773R. Only genes with significant (adjusted  $P < 0.05$ ) changes in expression are shown, as open circles: those with changes that were significant in GR $\alpha$  are pink, significant in the other allele are green and significant in both are orange.

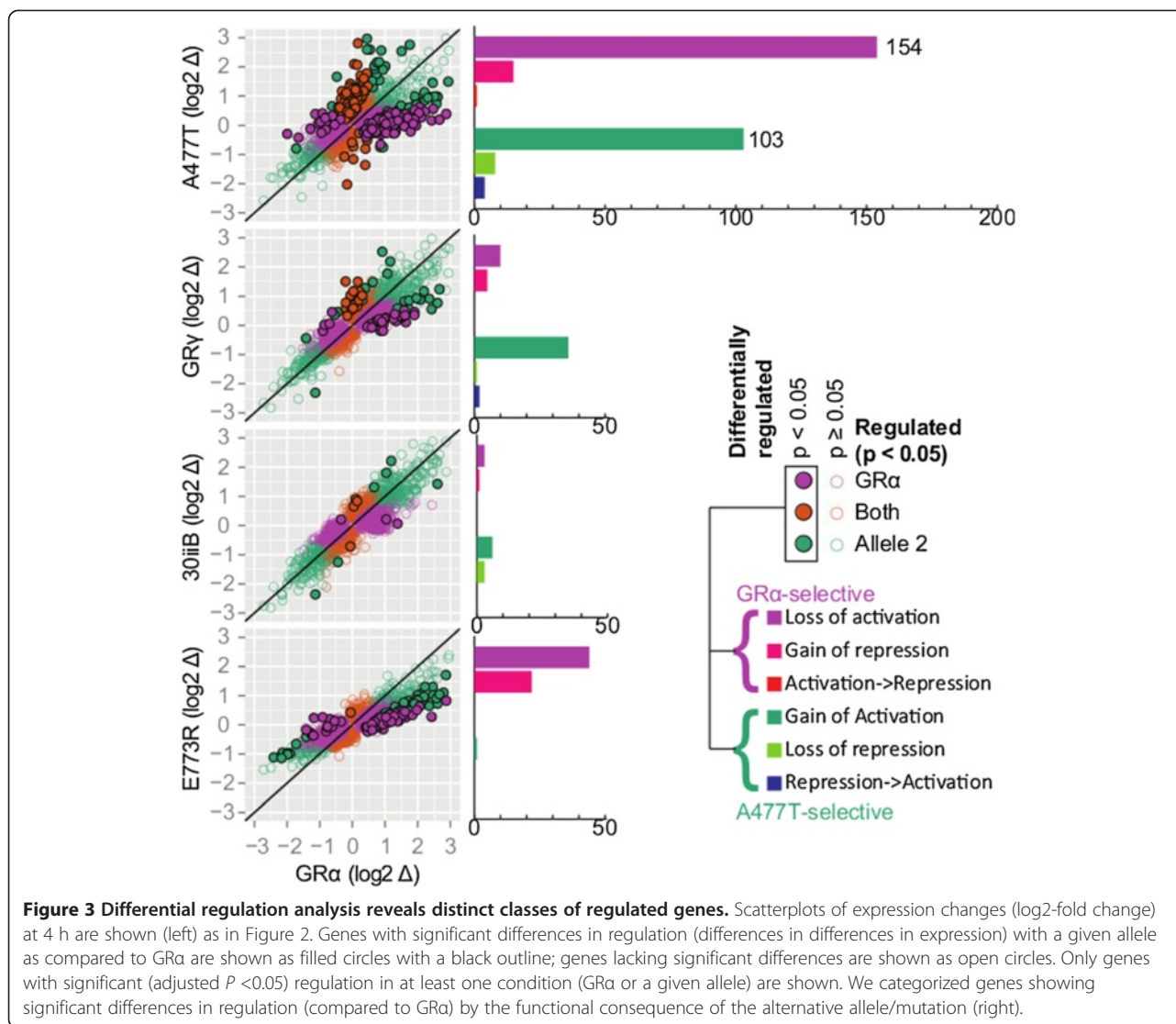
We then assigned a simple phenotypic classification to each GR regulated gene based on its valence of regulation by GR $\alpha$  (activation or repression) and the difference in regulation by that allele and by GR $\alpha$  (gain, loss, or change of valence). Some phenotypic classifications were less common than others: in particular, genes that change the valence of regulation were relatively rare. Moreover, certain phenotypes appeared to be relatively allele-restricted; for example, many genes gained activation with A477T and GR $\gamma$  (103 and 36 genes, respectively), whereas few gained activation with 30iiB or E773R (6 and 1 genes, respectively) (Figure 3). Thus, A477T and GR $\gamma$  may drive activation through one or more mechanisms unavailable to GR $\alpha$ , 30iiB, or E773R.

#### A477T and GR $\gamma$ selectively occupy GBRs near genes that gain activation

Next, we performed ChIP-seq to define GBRs genome-wide in U2OS cells expressing GR $\alpha$  or one of the three mutant alleles (30iiB, A477T, and E773R), after treatment with dexamethasone. Defining GBRs enables identification of putative primary regulated genes (those regulated by a proximal GR-occupied presumptive GRE) and allows us to distinguish genes differentially regulated due to changes in occupancy from those with altered GR activity. GR $\gamma$

has been similarly examined by others [6]. Importantly, we did paired-end sequencing of the ChIP-seq library, which improves the dynamic range of occupancy by allowing use of both ends of the double stranded molecule to determine whether the originating fragment was unique [8]. We identified GBRs (Additional file 2) using MACS2<sup>a</sup>, a software package for the analysis of ChIP-seq data, and assigned the GBRs to genes (Additional file 3) based on proximity (see Materials and Methods, 'GBR to gene assignment') [9]. We also compared occupancy between GR $\alpha$  and each mutant (Additional file 4 Additional file 5 and Additional file 6) using the 'diffstats' module of MACS2 (see Materials and Methods, 'Computational analysis of ChIP-seq data'). To avoid considering the same GBR twice, we considered all GBRs from GR $\alpha$  and those GBRs from a given mutant that did not overlap with GBRs from GR $\alpha$  (see Additional file 7, Additional file 8, Additional file 9, and Additional file 10); additionally, we analyzed only GBRs with signals of at least two reads per million total reads.

Among cells expressing GR $\alpha$  or 30iiB, we found that the locations of about 7% of GBRs differed<sup>b</sup>; and we observed the same fractional difference among cells expressing GR $\alpha$  or E773R. Similarly, the magnitude of occupancy (that is, the number of observed sequence

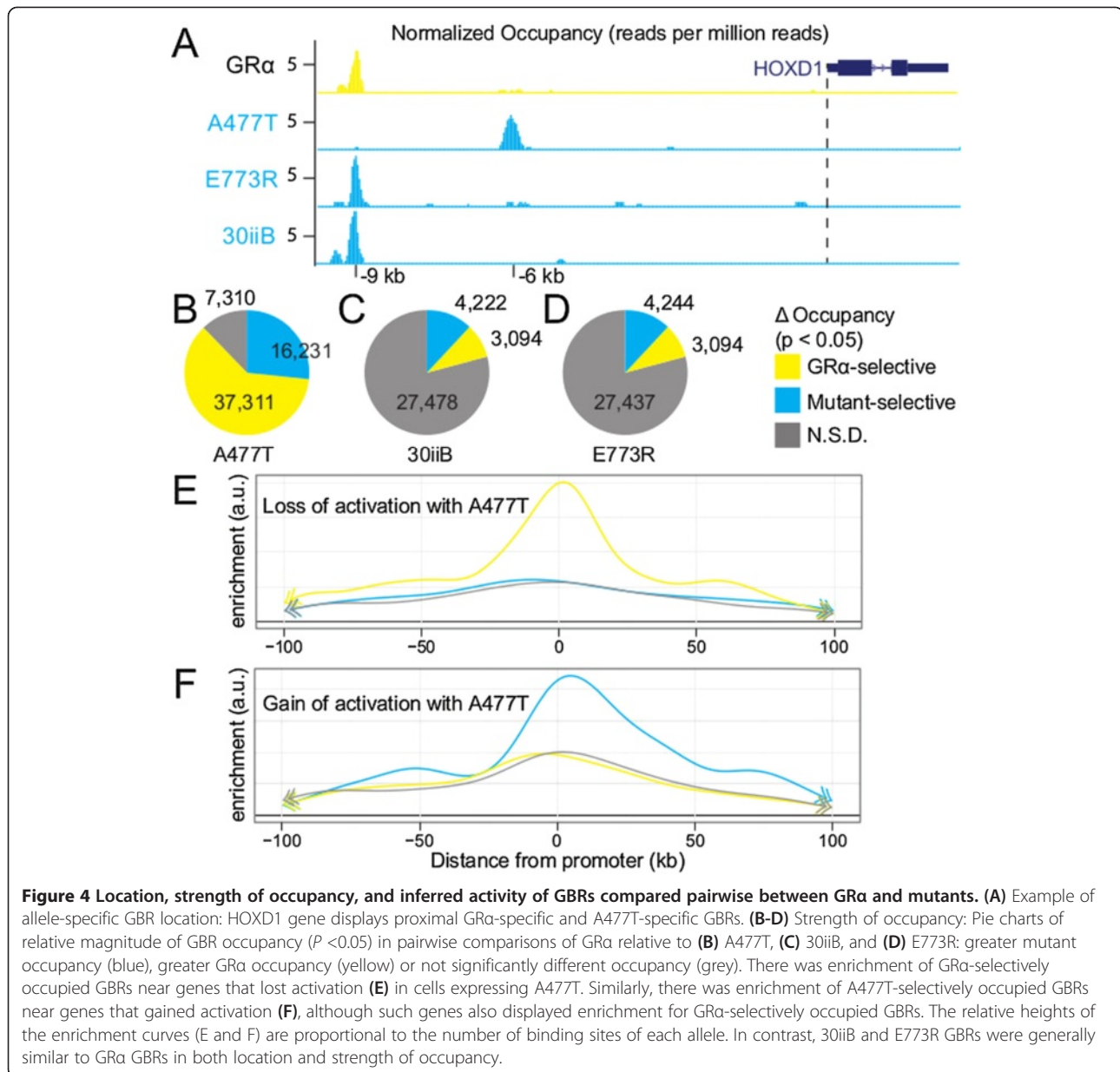


reads normalized by total reads) differed at only one-quarter of GBRs among cells expressing GR $\alpha$ , 30iiB, or E773R (Figure 4C and D). In contrast, the locations of over half of GBRs differed<sup>b</sup> among cells expressing GR $\alpha$  or A477T (for example, Figure 4A). Moreover, the magnitude of occupancy varied at most GBRs: more than half had greater occupancy by GR $\alpha$  than by A477T (GR $\alpha$ -selectively occupied) and one-quarter had greater occupancy by A477T than by GR $\alpha$  (A477T-selectively occupied) (Figure 4B).

Comparing these results with our gene expression findings, we observed substantial enrichment of GR $\alpha$ -selectively occupied GBRs near genes that lost activation in cells expressing A477T (Figure 4E, Additional file 11, and Additional file 12), and of both A477T- and GR $\alpha$ -selectively occupied GBRs near genes that gained activation (Figure 4F, Additional file 12, and Additional file 13).

Thus, it appeared that the differentially regulated genes represent a mixture, some with altered strength of occupancy (that is, a direct relationship between binding and regulation) and others with altered GR activity (that is, an indirect relationship). The occupancy of A477T near genes that are activated by A477T is surprising in light of previous reports indicating that mutations in the dimerization interface abrogate ‘transactivation’ [10,11].

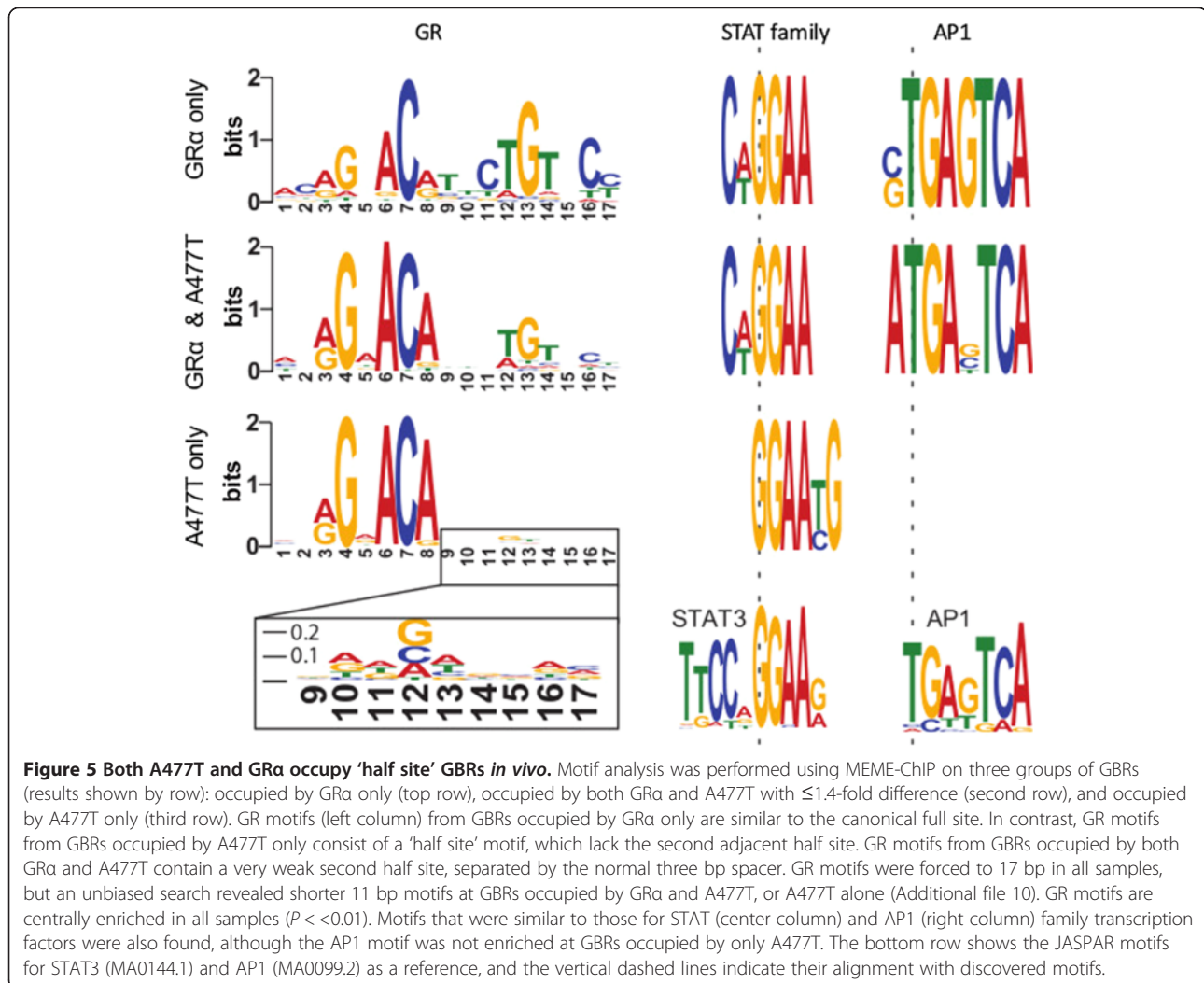
We further investigated whether the differentially occupied GBRs near differentially regulated genes were responsible for the observed regulation. As a simple measure of regulatory function, we cloned GBRs - selected on the basis of proximity to differentially regulated genes, strength of occupancy, and isolation from many other GBRs - into luciferase-based reporters and transfected them into U2OS cells expressing either GR $\alpha$  or A477T. We found GR $\alpha$ -selective activation with four GR $\alpha$ -selectively occupied



GBRs, A477T-selective activation with seven A477T-selectively occupied GBRs, and non-selective activation with five non-selectively occupied GBRs (Additional file 14). Six of 24 GBRs did not activate transcription in either cell line, but only two drove aberrant activation (opposing selectivity of occupancy and activation). These results imply that certain of the differentially occupied GBRs may indeed function to differentially regulate nearby genes. Additional experiments, using recent methods such as endogenous genome editing [12-15], will be required to assess unequivocally whether a given GBR is actually a GRE, and, if so, whether it confers differential regulation.

#### Most A477T-selectively occupied GBRs contain a half site motif

In order to understand what sequence constraints, if any, were responsible for differential occupancy by GR $\alpha$  and A477T, we performed motif analysis (Additional file 15) using MEME-ChIP [16]. Three groups of GBRs were analyzed: (1) the thousand most-occupied GBRs that were occupied by GR $\alpha$  and did not have detectable occupancy by A477T ('GR $\alpha$  only', Figure 5); (2) the thousand most-occupied GBRs that were occupied by both GR $\alpha$  and A477T with less than a 1.4-fold difference ('GR $\alpha$  & A477T', Figure 5); and (3) the thousand most-occupied GBRs that were occupied by A477T and did not have



detectable occupancy by GR $\alpha$  ('A477T only', Figure 5). We excluded from motif analysis GBRs that overlap with known repeats (see Materials and Methods, 'Computational analysis of ChIP-seq data').

*De novo* motif searches with MEME [17] revealed that GBRs occupied by only GR $\alpha$  were enriched for the canonical GR 'full site' motif, which specifies cooperative binding as a GR dimer (Figure 5). Allowing MEME to determine the optimal length for motifs in GBRs occupied by only GR $\alpha$  revealed a 17 bp GR motif (Additional file 15), which is longer than the previously reported 14 bp [18] and 15 bp motifs [19], having additional, asymmetric specificity flanking the more strongly specified half site. In contrast, GBRs occupied by only A477T were enriched for a single half site motif (Figure 5), which may provoke binding of a GR monomer to DNA. These results are consistent with the observations that A477T has an impaired dimerization interface and decreased cooperativity [4].

GBRs that were occupied by both GR $\alpha$  and A477T were enriched for an intermediate motif that consisted

of a closer-to-consensus half site and a very weak second half site, separated by the normal 3 bp spacer (Figure 5). Due to the semi-quantitative nature of ChIP-seq, some GBRs that we characterized as similarly occupied by both GR $\alpha$  and A477T may actually be differentially occupied (and conversely, some GBRs that we characterized as differentially occupied may not be). Therefore, the weak second half site seen at GBRs occupied by both GR $\alpha$  and A477T may be an experimental artifact; moreover, the lack of a second half site motif at A477T-selectively occupied GBRs implies that such GBRs indeed contain a motif distinct from that found at GR $\alpha$ -occupied GBRs. Supporting this view, length-unbiased searches with MEME for motifs at GBRs occupied by only A477T or both A477T and GR $\alpha$  revealed truncated motifs that included only the first two positions corresponding to the second half site (Additional file 15). In Figure 5, we display GR motifs found when forcing MEME to search for motifs that are exactly 17 bp in length, for ease of comparison. CentriMo, which searches for central/local enrichment of motifs in

ChIP-seq data [20], showed that GR motifs were strongly enriched at GBR signal maxima (Additional file 15), implying direct binding by GR at these GBRs.

Discriminatory motif analysis of the three different classes (GR $\alpha$ -selective, A477T-selective, and non-selective) of occupied GBRs with DREME, which exhaustively searches for all 3 to 8 bp regular expression motifs that are enriched relative to di-nucleotide shuffled sequences [21], revealed motifs similar to that of activation protein 1 (AP1) and Signal Transducers and Activators of Transcription (STAT) families of transcription factors (Figure 5 and Additional file 15). We did not detect any centrally enriched motifs that might indicate a cooperative binding partner (that is, a heterodimer) for DNA-bound GR $\alpha$  or A477T. Nonetheless, previous studies [22] indicate that there is widespread co-association among transcription factor binding sites, indicating that nearby sites may be functional.

Using DREME, we identified a STAT-like motif (CWGGAA) in 233 and 251 of 1,000 GBRs occupied by GR $\alpha$  alone, or GR $\alpha$  and A477T, respectively; a different STAT-like motif (GGAAYG) was found in 117 of 1,000 GBRs occupied by A477T alone (Table 1). AP1-like motifs, STGAGTCA and ATGABTCA, were positively identified in 35 of 1,000 GBRs occupied by GR $\alpha$  alone, and 105 of 1,000 GBRs occupied by GR $\alpha$  and A477T, respectively; AP1-like motifs were not significantly enriched at GBRs occupied by A477T alone (Table 1). Of all 23,525 GBRs found in cells expressing GR $\alpha$ , we found that 1,719 and 1,214 had STAT3 (a canonical member of the STAT family) or AP1 motifs, respectively, but lacked GR full sites, and 3,277 and 3,694 had STAT3 or AP1 motifs, respectively, and contained GR full (see Additional file 10, Additional file 12, and Table 2).

GR interacts physically and functionally with both the STAT and AP1 families of transcription factors [23]. Although GR:AP1 interaction is known to be stimulated by lipopolysaccharide (LPS) treatment, GR also interacts with AP1 under normal conditions [24]. The presence of AP1 and STAT motifs at GBRs is consistent with both 'tethering' of GR - protein:protein association of GR with DNA-bound AP1 or STAT - and with the existence of composite elements, elements that contain binding sites

**Table 1 Occurrences of STAT and AP1 motifs found using DREME.**

Found at GBRs in cells expressing	GR $\alpha$	GR $\alpha$ and A477T	A477T
STAT motif	CWGGAA	CWGGAA	GGAAYG
Of 1,000 with motif (n)	233	251	117
AP1 motif	STGAGTCA	ATGABTCA	(NA)
Of 1,000 with motif (n)	35	105	0

Found by DREME within training set of 1,000 peaks (see Figure 5 and Additional file 10).

**Table 2 Occurrences of STAT and AP1 motifs found among all GBRs.**

Motif present at GBR	With GR motif (n)	Without GR motif (n)	Ratio
STAT3	3,277	1,719	1.4
AP1	3,694	1,214	3.0

Found among all 23,525 GBRs in cells expressing GR $\alpha$ .

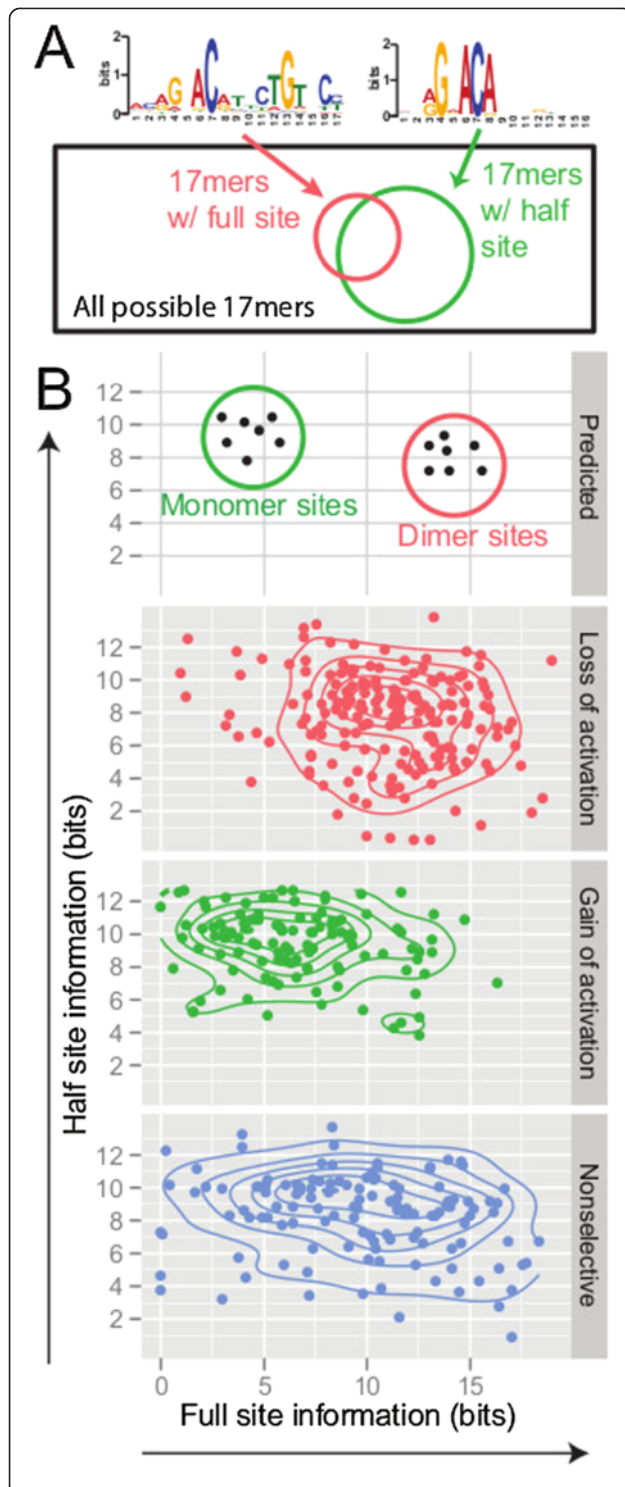
for GR, and for one or more additional transcriptional regulatory factors. The greater abundances of GBRs that contained STAT3 or AP1 motifs and also GR full sites as compared to GBRs that contained these motifs and lacked GR full sites (1.4- and 3.0-fold, respectively; Table 2) implies that composite elements may be more common than tethering elements.

#### A subset of genes regulated by GR $\alpha$ contains only half sites

Motivated by the discovery of a half site motif near genes regulated by A477T, we investigated whether any genes were regulated by GR $\alpha$  recognition of half site motifs. We identified a set of genes whose regulation at 4 h is similar (less than two-fold different) between GR $\alpha$  and A477T at the 95% confidence level<sup>c</sup>. GBR occupancy near these 'non-selective' genes (Additional file 15) did not reveal significant enrichment of GR $\alpha$ - or A477T-selectively occupied GBRs relative to genes that are selectively regulated, suggesting that GBRs near these genes are bound by both GR $\alpha$  and A477T. Consistent with this explanation, GR $\alpha$  and A477T bind specifically and with equal affinity to half sites *in vitro* [4].

We investigated whether GBRs within 20 kilo-base pairs (kb) of the transcriptional start site of non-selective, activated genes contained half sites, full sites, or both. Since the sets of sequences described by the full site motif and by the half site motif are partially overlapping (Figure 6A), we assessed the individual information, a quantitative and absolute measurement of the similarity of a sequence to a given motif [25], of the best match to each GR motif (GR $\alpha$ -derived full site motif and A477T-derived half site motif) at each GBR ( $\pm 75$  bp from the signal summit; Additional file 12 and Additional file 16). Consistent with the observed differential occupancy, genes that lost activation with A477T generally had GBRs that contained GR motifs that are good full sites (Figure 6B), which logically also constitute good half sites. Inversely, genes that gained activation with A477T had GBRs that contained poor full sites and good half sites. Non-selective genes also had GBRs that contained poor full sites and good half sites, suggesting that GBRs near non-selective genes, like those near genes that gained activation, are bound by GR $\alpha$  at half sites (Figure 6B).

Next, we investigated whether there were GR $\alpha$ -regulated genes bearing half site GBRs and apparently lacking full



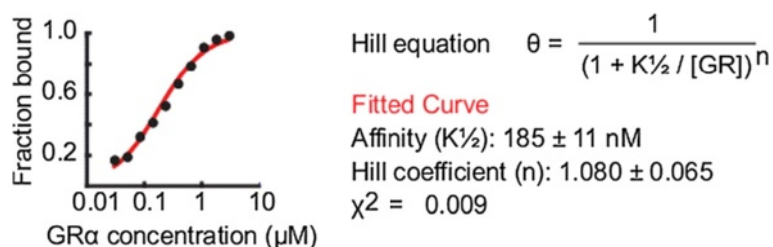
**Figure 6 GR half sites predominate at GBRs near non-selective, activated genes.** (A) More 17mers containing half sites (green circle) than full sites (red circle) are expected because the information-containing region of the motif is shorter; similarly, most full sites contain a good half site because the constraints (that is, the sequence motifs) are similar and there are two positions available in a full site for a good half site due to the palindromic nature of dimer. Nonetheless, not all full sites contain half sites, as the half site motif is more strongly constrained than the full site motif; the additional specificity may be due to the lack of cooperative binding at these sites. (B) Individual information (measured in bits) of the best full site (x-axis) and half site (y-axis) found at GBRs (circles) located within 20 kb of the transcriptional start sites of a set of genes; contour lines encompass same number of GBRs, and thus localize regions of high GBR density. Genes that lost activation with A477T (B, red) generally had peaks with good full sites (>10 bits) and good half sites (>9 bits). Genes that gained activation with A477T (B, green) generally had GBRs with poor full sites (<5 bits) and good half-sites (>10 bits). Genes that were non-selectively activated (B, blue) also generally had GBRs with poor full sites and good half sites.

site GBRs (Additional file 7 and Additional file 12). This restrictive criterion should define genes most likely to be regulated by GR $\alpha$  via half sites. We did not restrict our selection with respect to regulation by A477T, as some genes regulated by half sites might not be regulated by A477T due to impaired conformational dynamics and/or co-regulator recruitment. We defined as a full site motif any sequence with an individual information of at least 8.94 bits (the 95th percentile of randomly permuted GBR sequences), and as a half site motif any sequence with an individual information of at least 8.00 bits, with respect to the given motif. By comparison, the full site and half site motifs had 11.9 bits and 9.3 bits of information, respectively, relative to GBR background sequence (Additional file 15). We additionally required that the full site motif's individual information at candidate GBRs be less than the half site motif's individual information. In total, there were 1,441 GR $\alpha$ -regulated genes, 607 (42%) of which had associated GBRs within 20 kb of the transcriptional start site. We found that 35 of these 607 genes met our definition, and that this number declined slowly with an increasing window of association (27 within 30 kb, 21 within 40 kb, 17 within 50 kb).

#### GR $\alpha$ monomers bind half sites *in vitro*

We next investigated the *in vitro* binding of glucocorticoid receptor DNA binding domain (DBD) to an oligonucleotide containing a half site adjacent to an 'anti-motif' - that is, the least favorable binding sequence according to the half site motif - (gtacAGAACAttGTGAGGtgcac) by electrophoretic mobility shift assay [4]. GR $\alpha$  binding was fitted with the Hill equation and a Hill coefficient of  $1.080 \pm 0.065$  (mean  $\pm$  standard deviation) was determined (Figure 7A), consistent with no cooperativity. This indicates





**Figure 7 GR $\alpha$  binds a half site in distinct conformation.** Hill plot derived from an electrophoretic mobility shift assay describing the fraction of oligonucleotides containing a half site with an adjacent 'anti-motif' - that is, the least favorable binding sequence according to the half site motif -(gtacAGAACAttGTGAGGtcac) bound at increasing GR $\alpha$  concentrations [4]. The Hill coefficient was computed from the Hill Equation (right) to be  $1.080 \pm 0.065$ .

that a strong half site adjacent to an unfavorable DNA sequence is bound by GR $\alpha$  as a monomer.

## Discussion

### Half site motifs suggest GR monomers bind DNA and regulate specific genes

We set out to find whether sequence motifs bound by GR might be decomposed into multiple specific motifs that each relate to the utilization of specific functional surfaces of GR. We found that A477T, a mutation in the dimerization interface of GR leading to reduced DNA-binding cooperativity and increased dissociation rates *in vitro* [4], resulted in gains and losses in glucocorticoid-induced gene regulation and in GBR occupancy *in vivo*. Analysis of A477T-occupied GBRs revealed a hitherto unrecognized half site motif. The half site motif, coupled with a failure to detect non-GR partner motifs, suggested that A477T might bind GBRs as a monomer *in vivo*. Furthermore, A477T regulated many genes, suggesting that GR monomers are sufficient for gene regulation. The question remains, then, whether GR $\alpha$  monomers bind to DNA in a sequence-specific manner and regulate a set of target genes.

Curiously, we found that some genes were regulated by A477T and not by GR $\alpha$ , and some GBRs were occupied by A477T and not by GR $\alpha$ . We speculate that GR $\alpha$  fails to bind such GBRs because it is titrated away by cooperatively bound, high affinity full sites, composed of 'degenerate' or weak half sites; such cooperatively bound sites would be unavailable to A477T due to its reduced cooperativity. Indeed, many GR $\alpha$ -selectively occupied GBRs near GR $\alpha$ -selectively regulated genes contained full sites lacking 'consensus' or perfect half sites; rather, these full sites were composed of two degenerate half sites (Figure 6). These results suggest two orthogonal binding modes for GR $\alpha$ : (1) consensus half sites, most likely bound by monomers; and (2) degenerate full sites, most likely bound by dimers.

Alternatively, GR $\alpha$  occupancy may not be observed at A477T-specific GBRs because GR $\alpha$  dissociates from these regions more rapidly than A477T, reducing the signal

below detection limits. However, A477T dissociates from DNA more rapidly than GR $\alpha$  *in vitro* [4] and *in vivo* [26], arguing against this model. Importantly, at any given site, some fraction of binding events may involve monomers and some fraction may involve dimers. We found that a minimum of 35 genes are regulated entirely by direct binding of GR $\alpha$  monomers. While many genes may be regulated entirely by direct, cooperative binding of GR $\alpha$  as a dimer, our results suggest that others may be regulated in part by monomers and in part by dimers.

Additionally, we found that GBRs were enriched within 20 kb from the promoters of regulated genes (relative to the whole genome), suggesting that promoter-proximal GBRs were responsible for many observed changes in gene expression. Similarly, we found that GR $\alpha$ - and A477T-selectively occupied GBRs were enriched near the promoters of genes that lost and gained activation, respectively, with A477T, consistent with these GBRs being responsible for gene regulation. Nonetheless, more than half of GBRs were located distal to genes (more than 50 kb from the transcriptional start site), and only 42% of GR $\alpha$ -regulated genes were associated with promoter-proximal GBRs ( $\pm 20$  kb). Further exploration is needed to understand the function (if any) of distal GBRs.

### Monomeric GR $\alpha$ is functional in regulation

The ability of a transcription factor to function in gene regulation alternatively as a monomer or a dimer has been previously observed for SRY (sex determining region Y)-box 10 (SOX10) [27]. There are several lines of evidence suggesting that GR $\alpha$ , too, binds DNA alternatively as a monomer or a dimer. GR $\alpha$  and A477T bind specifically and with equal affinity to half sites *in vitro* [4], with a Hill coefficient of approximately 1, consistent with binding as a monomer to half sites. Additionally, single-molecule imaging of GR binding *in vivo* indicates that A477T dissociates from DNA 10-fold faster than GR $\alpha$  [26], consistent with the observation that A477T has reduced cooperativity when binding full sites *in vitro*

[4]. These results collectively suggest that A477T binds as a monomer *in vivo*. Furthermore, deep sequencing of DNase I-treated chromatin, in combination with ChIP-seq, reveals that hormone-induced binding of androgen receptor (AR, HUGO symbol NR3C4), a related nuclear receptor, was associated with DNase footprints the size of AR monomers at two-thirds of binding regions [28]. Given the significant homology between AR and GR, these results suggest that GR $\alpha$  is also capable of binding as a monomer *in vivo*.

DNase footprints measured in aggregate, and even direct imaging of individual binding events, are insufficient to demonstrate that certain GBR binding events promote regulation of target genes. Joint measurements of gene expression and GBR occupancy are the minimum necessary data to infer function of GBRs. Our finding that at least 35 regulated genes contain only GBRs with half sites within 20 kb of their transcriptional start site provides the best evidence to date that GR $\alpha$  binds as a functional monomer *in vivo*, resulting in regulation of target genes.

#### Implications of GR monomers for physiology

If GR and other nuclear receptors bind DNA alternatively as dimers or as monomers, then these two receptor species may represent distinct drug targets [29]. Glucocorticoids, like other steroid hormones, regulate diverse aspects of physiology, making them potent but blunt tools for medicine. Significant interest in 'selective' or 'dissociated' glucocorticoid receptor agonists [30,31] that separate the beneficial and adverse clinical effects of glucocorticoid treatment has been driven by the success of similar drugs targeting the estrogen receptor, especially in treating breast cancer and osteoporosis [32]. Drugs targeting nuclear receptors commonly bind to a pocket in the ligand binding domain, normally bound by an endogenous hormone (cortisol in the case of human GR). Alternatively targeting the dimerization interface or DNA-binding interface to separate dimer and monomer activity may help produce improved selective modulators for GR and other receptors.

Encouragingly, drugs modulating AR conformation and activity, subsequent to hormone binding, have been previously found [33]. Our data make it difficult to infer the potential effects of such selective modulators on GR activity. However, observations of mice harboring the equivalent mutation to A477T indicate that inhibiting GR dimerization may prevent glucocorticoid-induced decreases in skin and bone collagen synthesis, as well as decreases in insulin sensitivity, while maintaining or even exaggerating glucocorticoid-induced increases in triglyceride synthesis [29]. Moreover, such an inhibitor may cause an increase in lipogenesis, possibly due to the regulation (by monomers) of target genes not normally regulated by GR [29]. A selective modulator that inhibits

the regulatory function of GR dimers may have different effects from one that inhibits their formation; both would inhibit regulation of dimerization-dependent genes, but the latter may also drive regulation of new target genes, similar to those observed in the dimerization interface mutant.

#### Evolution and structure of hormone-regulated gene networks

Evidence for existence of functional AR [28] and GR monomers bound to consensus half sites suggests that similar monomeric forms may exist for all nuclear receptor subfamily 3 group C transcription factors (GR, mineralocorticoid receptor, progesterone receptor, and AR), and possibly more distantly related nuclear receptors. These results may have important implications for how hormone-regulated gene networks evolved. Binding of receptor monomers to consensus half sites may present an evolutionary path for establishing robust response elements wherein one perfect half site appears, at random or by insertion or rearrangement, and allows for regulation by monomers, while also promoting selection of a second site with the proper spacing for cooperative binding by dimers; the full site would then be resistant to mutations that might have perturbed binding by monomers. Alternatively, the half site may have preferred function, and may therefore remain stable. Analysis of mutations equivalent to A477T in other nuclear receptors and cell lines may help clarify the functional and evolutionary role of half sites.

Additionally, other GR functional surfaces, including but not limited to AF1, AF2, and the lever arm, may be associated with more narrowly defined motifs that may either be subsets of the full site GBS motif, or may differ from it. Such motifs may be detected by analyzing differentially occupied GBRs near regulated genes selectively affected by perturbations to a functional surface. In cases where binding is not perturbed, discriminatory motif analysis between GBRs proximal to differentially regulated genes and GBRs proximal to non-selectively regulated genes may yield similar answers.

Moreover, structural methods such as X-ray crystallography, electron microscopy, and NMR may define directly the conformational changes induced in GR upon binding to these motifs. These approaches may confirm that GR binding to a suspected 'sub-motif' induces changes in the functional surface being studied and drives allosteric changes in remote functional surfaces.

#### Conclusions

GREs differ in their precise sequence motifs, in their modes of GR binding (monomer or dimer; direct or indirect), and in the functional GR surfaces required for binding or regulation. Structural changes induced in GR

upon binding different 'sub-motifs' (refinements of the canonical motif) may be important determinants of GR activity at particular response elements. We have shown that GR occupies *in vivo* many genomic regions containing half sites, and that these regions are likely responsible for regulation of a subset of target genes. Moreover, GR specifically binds to half sites *in vitro*, without apparent cooperativity, and thus as a monomer.

Half sites are easily overlooked among those sequences that conform to the known GR full site motif, but they have distinct functional properties compared to degenerate full sites, which may be predicted to bind with similar affinity. Identification of GREs with other sub-motifs, the conformations those GREs induce, and the functional surfaces they utilize will be important steps in defining predictive models of GR activity and in developing therapeutic strategies to modulate GR activity.

## Materials and methods

### Cell culture

U2OS cells stably transfected with rat GR $\alpha$ , GR $\alpha$  30iiB (30iiB; E219K/F220L/W234R), GR $\alpha$  A477T (A477T), GR $\gamma$ , or GR $\alpha$  E773R (E773R) were grown as previously described [2]. Parent U2OS cells do not manifest any endogenous response to glucocorticoid treatment. DNA sequencing analysis was used to confirm the presence of point mutations in all cell lines. A549 cells were obtained from American Type Culture Collection and maintained in DMEM supplemented with 10% volume/volume (v/v) fetal bovine serum.

### Immunoblotting

Whole cell extracts were collected using 20 mM Tris-Cl pH 8.0, 150 mM NaCl, 1% NonidetP-40 (NP-40), 1% sodium deoxycholate, 0.1% sodium dodecyl sulphate (SDS), 2 mM ethylenediaminetetraacetic acid (EDTA). Immunoblot analysis was performed as previously described [34]. Antibodies to GR (N499) and  $\beta$ -actin (Sigma-Aldrich) were used.

### Gene expression measurements using microarrays

Cells were plated in 6-well plates using DMEM supplemented with 5% v/v fetal bovine serum. Dexamethasone (Sigma) at a final concentration of 100 nM was added to the medium for 2, 4, or 24 h; for the '0 h' time point, cells were treated with vehicle (ethanol) only for 4 h. Cells were lysed and total RNA was isolated using QIAshredder and RNeasy mini columns (Qiagen). The quality of RNA samples was evaluated by A260/A280 ratio which was at least 1.9 and the integrity was analyzed using the Bioanalyzer 2100 (Agilent) with the Experion RNA StdSens analysis kit (Biorad). For each experimental condition 2  $\mu$ g of high quality total RNA was submitted to the University of Southern California Epigenome Center. Four biological replicates of each experimental

sample from independent experiments that were collected on different days were randomly placed on Illumina Human Ref8 beadchips and processed following standard Illumina procedures.

### Genome build

All coordinates reported are for the UCSC human genome build hg19 (GRCh37).

### Computational analysis of gene expression data

Bead-level data outputted from Illumina's BeadScan software were read using the beadarray package [35] from Bioconductor [36] with the R programming language [37]. BASH [38] was used to correct for compact and diffuse spatial artifacts on the arrays. Bead-level data were summarized using beadarray, filtering out beads that were more than three mean absolute deviations from the median. Probe annotations were retrieved using the illumina-Humanv3.db package from Bioconductor; poor quality probes and probes not corresponding to known HUGO gene symbols were removed from the data before further processing. Bead summary data were log<sub>2</sub> transformed and quantile normalized. We then used the non-parametric empirical Bayes method described by [39] to correct for chip-to-chip batch effects. The raw data and batch-corrected, normalized data have been deposited in NCBI's Gene Expression Omnibus [40] and are accessible through GEO Series accession number GSE45407 [41].

The limma package [7] from Bioconductor was used to determine differential gene expression/regulation: the lmFit function was used to determine relative expression at each condition (time point, expressed allele); the contrasts.fit function was used twice, first to determine fold-change at 2, 4, and 24 h relative to the '0 h' time point (4 h ethanol treatment), and second to determine the difference in fold-change between each expressed allele and GR $\alpha$ ; the eBayes function was used to compute statistics in each case. An adjusted [42] *P* value cutoff of *P* < 0.05 was used to determine which genes are differentially expressed. We found there was insufficient statistical power to identify many differentially regulated genes at 2 h. This is likely due in part to the limitations of measuring steady state total RNA levels rather than newly generated transcripts, which was not feasible at the time of this study. The ggplot2 package [43] was used to visualize the results. For a more detailed procedure see (Additional file 17).

### Chromatin immunoprecipitation (ChIP)

U2OS cells stably expressing one of GR $\alpha$ , A477T, 30iiB, or E773R were grown in 3  $\times$  75 cm<sup>2</sup> dishes and treated with 100 nM dexamethasone for 90 min. The following procedure is adapted from [19]. We added formaldehyde to a final concentration of 1% for 3 min at room temperature

and then quenched by adding 2.5 M glycine to a final concentration of 125 mM glycine. Cells were incubated at 4°C for 10 min and then washed with PBS for 5 min. We added 5 mL ice-cold IP lysis buffer (50 mM HEPES-KOH pH 7.4, 1 mM EDTA, 150 mM NaCl, 10% v/v glycerol, 0.5% v/v Triton X-100) with 1× protease inhibitor cocktail added (PIC; Roche Complete) to each flask to lyse cells and scraped cells into 50 mL conical tubes. Lysed cells were nutated for 30 min and pelleted by centrifugation at 4°C for 5 min at 600 times gravity. After removing the supernatant, nuclei were resuspended in 900 µL RIPA buffer (10 mM Tris pH 8.0, 1 mM EDTA, 150 mM NaCl, 5% v/v glycerol, 0.1% sodium deoxycholate, 0.1% SDS, 1% v/v Triton X-100) with 1× PIC added and frozen at -80°C.

The remaining ChIP procedure was performed at 4°C. Protein G Dynabeads (Invitrogen) were prepared for IP as follows: beads were pelleted (using a magnet); liquid was removed; beads were washed twice with RIPA, then resuspended in RIPA with 1× PIC and 0.5 mg/mL BSA at the original bead concentration; 12 µg N499 α-GR antibody were added per 100 µL beads, and the beads were nutated for 2 h.

Nuclei were thawed in ice water. Chromatin from nuclei were fragmented by sonication at 4°C in 300 µL aliquots in 1.5 mL microcentrifuge tubes using the Diagenode Biorupter for 8 cycles of 7 min (0.5 min on / 0.5 min off). Samples were spun at 14,000 revolutions per minute (rpm) and for 15 min to pellet insoluble molecules, and the supernatant was transferred into new microcentrifuge tubes. We saved 10% of each sample to isolate later as input DNA (GRα and A477T only). Protein G Dynabeads with antibody bound were pelleted, washed twice with RIPA, and resuspended at the original bead concentration in RIPA with 4× PIC and 2 mg/mL BSA. Sonicated chromatin (900 µL) from each sample was incubated with 100 µL of antibody-bound beads for 16 h.

After pulldown, beads were washed four times with RIPA with 500 mM NaCl, then four times with LiCl buffer (20 mM Tris, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate), to decrease non-specific binding. To reverse crosslinks, beads were then resuspended in 11 µL RIPA with 500 mM NaCl plus 89 µL Rev-Xlink buffer (0.7% SDS in TE pH 8.0 with 0.2 mg/mL Proteinase K). A total of 10 µL of each input DNA sample was similarly treated by adding 1 µL 5 M NaCl and 89 µL Rev-Xlink buffer. Samples were incubated in a PCR block at 55°C for 3 h, then 16°C for 16 h, then column purified using the Clean & Concentrator-5 kit (Zymogen) at a DNA binding buffer:sample ratio of 5:1.

#### Library preparation for ChIP sequencing (ChIP-seq)

Fragments were end-repaired for 30 min at 20°C in 25 µL reactions with 1× T4 DNA Ligase Buffer w/ 10 mM

ATP (NEB), 100 µM dNTPs (Invitrogen), 3.75 U T4 DNA Polymerase (NEB), 1.25 U DNA Polymerase I, Large (Klenow) Fragment (NEB), and 12.5 U T4 Polynucleotide Kinase. Samples were column purified using the Clean & Concentrator-5 kit (Zymogen) at a DNA binding buffer: sample ratio of 2:1 following this and subsequent enzymatic reactions. dATP was added to the 3' end of molecules by incubating for 30 min at 37°C in 25 µL reactions with Buffer 2 (NEB), 200 µM dATP (Invitrogen), and 7.5 U Klenow Fragment (3' → 5' exo-). Sample concentrations were measured using PicoGreen (Invitrogen), and then ligated to sequencing adapters at a 2:1 (adapter:sample) molar ratio in 20 µL reactions with 1X T4 DNA ligase buffer w/ 10 mM ATP (NEB) with 6% PEG-8000 and 200 U T4 DNA Ligase (NEB). Libraries were amplified by PCR for 17 cycles using the standard Illumina paired-end primers, and purified by PAGE as described in [44]. Sample concentrations and size distribution were measured using the High Sensitivity DNA kit for the Bioanalyzer 2100 (Agilent) prior to sequencing.

#### Adapters and barcodes for ChIP-seq

In-house barcodes were used for ChIP samples, which were added to the 3' ends of the sequencing adapters and determined using the first four bases of the read. Barcodes are TCAT (BC1), GACG (BC2), AGTC (BC3), and CTGA (BC4). The ChIPs for GRα and A477T were each ligated to all barcodes, pooled, and sequenced in their own lanes. Samples for 30iiB and E773R were ligated to BC1-2 and BC3-4, respectively, and pooled. Input samples were ligated to indexed TruSeq adapters using indices 5 and 7 for GRα and A477T inputs, respectively.

#### ChIP-seq procedure

ChIP samples were sequenced at the University of California San Francisco Center for Advanced Technology using the Genome Analyzer II (Illumina) with 2× 75 bp paired end reads as per the manufacturer's instructions. Input DNA samples were sequenced using the HiSeq 2000 (Illumina) with 2× 100 bp paired end reads. Raw sequence reads were deposited in the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) [45] and are accessible through SRA study accession number SRP020242.

#### Computational analysis of ChIP-seq data

We automated the analysis procedure using the Python programming language in order to make our results easily reproducible and to facilitate similar analysis by others. The scripts used - preprocess\_reads.py, align2.py, call\_peaks.py, and analyze.py - are described here and are available in the software package seriesoftubes on the Python Package Index (PyPI). seriesoftubes depends in part on Biopython [46]. We used preprocess\_reads.py (seriesoftubes)

to separate the sequencing results by barcode/index, saving only reads with valid barcodes/indices (on both ends in a pair for internal barcodes), and merged barcodes/indices as applicable to consolidate sequence data for each sample. We then used align2.py (seriesoftubes) to align sequences to the genome using bowtie2 [47], allowing a maximum insert size of 600 bp (based on our observation of an average library size of <150 bp) and saved in BAM format using samtools [48]. Note: for internal barcodes, the first five bases contain the barcode plus an additional 'T'; these positions were ignored when running bowtie2.

We used call\_peaks.py (seriesoftubes), which relies on the MACS2<sup>a</sup> callpeak module, to find significantly occupied peak regions ('peaks') and summit positions using an adjusted *P* value cutoff of *P* < 0.01 and to generate pileup tracks (signal per million reads). Additionally, the MACS2 diffstats module was used to compute differential occupancy statistics and find (mutant versus GR $\alpha$ ) differentially occupied peaks at using an adjusted *P* value cutoff of *P* < 0.05. For convenience, we have created a track hub formatted for the UCSC genome browser [49] containing paired-end read alignments, signal pileups, peak region locations, and peak summit locations [50].

#### GBR to gene assignment

Peaks were assigned to genes using GREAT [9], with a basal domain of  $\pm 20$  kb of a gene's primary transcriptional start site (where assignment is guaranteed) and an extended domain of  $\pm 100$  kb (where assignment is made only if it does not overlap with another gene's basal domain); if a peak was assigned to multiple genes, it was ignored in subsequent analyses.

#### Motif analysis

MEME-ChIP [16] was used for motif discovery as described in the text. For all motif analyses, we used only peak regions ( $\pm 75$  bp from the peak summit) that do not overlap with repeats, found by using bedtools [51] to intersect peaks regions with the 'rmsk' table from the RepeatMasker [52] track for hg19, extracted using the Table Browser [53]. Peak sequences ( $\pm 75$  bp from the peak summit) were then extracted from the UCSC hg19 build of the human genome using the twobitreader package (available on PyPI). Occurrences of motifs were found using analyze.py (seriesoftubes), which relies on MOODS [54] for efficiency. JASPAR [55] motifs MA0144.1 and MA0099.2 were used to find STAT3 and AP1 sites, respectively. Peak sequences were randomly permuted and the 95th percentile of individual information for the best motif in each permuted peak sequence was used as a cutoff for determining presence of GR full site, STAT3, or AP1 motifs (8.94, 8.06, and 9.49 bits, respectively). Data were integrated for further analysis using the R

programming language (Additional file 10) and visualized with ggplot2 [43].

#### Luciferase transcriptional reporters

Selected GR binding regions were cloned into the firefly luciferase reporter vector pGL4.10-E4TATA<sup>d</sup> using conventional methods (that is, PCR and restriction enzymes). A table of reporter names, primers, and genomic coordinates is provided in Additional file 18. We used the *Renilla* luciferase reporter vector pGL4.70-MLE2<sup>e</sup> as a normalization control, and the mammalian expression vector p6R as 'filler' DNA to reach the recommended concentration of DNA for transfections.

U2OS cells stably expressing either GR $\alpha$  or A477T were grown in three 10 cm dishes each, split, and seeded into two 96-well plates at a final density of about  $5 \times 10^4$  cells/well. The next day, cells were transfected with 3.33 ng firefly reporter, 150 pg *Renilla* reporter, and 20 ng p6R per well using PLUS Reagent and Lipofectamine (Invitrogen) according to manufacturer instructions. After 16 h, cells were treated with 100 nM dexamethasone or vehicle only (ethanol) for 3 h, and then luciferase expression was assayed using the Dual-Luciferase Reporter Assay System (Promega) according to manufacturer's instructions.

#### Endnotes

<sup>a</sup>MACS2 will be described in more detail in a forthcoming publication. Briefly, peak calling uses a Poisson test as in the first version of MACS [56] and differential occupancies are determined using a log-likelihood ratio test, and transformed to *P* values using Wilks' theorem [57]. Pre-release versions of MACS2 are available online at [58].

<sup>b</sup>GBRs with different locations among two samples were determined as those GBRs that were detectable in one sample and not the other (by overlap) and that did not significantly differ in occupancy among the two samples. This quantification of GBRs with differing locations is only an estimate, as some GBRs that differ in occupancy may appear to have similar occupancy. This effect results from variation in the observed numbers at a given GBR, primarily due to Poisson counting error.

<sup>c</sup>To determine whether fold-changes in gene expression between two conditions were less than two-fold different at a significance level of 95%, we constructed 97.5% confidence intervals for the fold-change in gene expression under each condition and required that they do not overlap at all. The 97.5% confidence intervals were computed as the fold-change plus or minus 2.447 - the *t*-statistic for six degrees of freedom (eight total sample measurements minus two degrees for determining their respective means) - divided by the square root of two (because subtracting two groups of four yields an equivalent *n* of 2), and

multiplied by the posterior standard deviation for a given gene as determined by limma.

<sup>d</sup>pGL4.10-E4TATA was constructed by ligating the E4TATA minimal promoter 5'-tttttagtcc tatatatac cgctctgtac ttggcccttt tacactgtg to the promoter-less vector pGL4.10 at BglII and HindIII restriction sites. The E4TATA promoter was selected because it does not respond to dexamethasone.

<sup>e</sup>pGL4.70-MLE2 was constructed by ligating the MLE2 promoter (engineered ML promoter with stronger induction ability) 5'-gaaggggggc tataaaact cgctctggcg cgctctcct cactctctc cgatcgct. The MLE2 promoter was chosen because it does not respond to dexamethasone and drives strong expression.

## Additional files

**Additional file 1: Genome-wide expression analysis.** Compressed folder (zip archive) containing the following tab-delimited text files with probe-wise results of gene expression analysis using limma, including statistics and gene-level annotations: Expression\_Levels\_log2.txt, expression levels at all conditions; Fold-changes\_vs\_0hr\_log2.txt, relative gene expression in all cell lines after 2, 4, and 24 h of dexamethasone treatment versus 0 hours; Fold-difference\_vs\_GRa\_2hr\_log2.txt, difference in regulation (changes in gene expression) after 2 h of dexamethasone treatment in cells expressing A477T, 30iib, E773R, or GR $\gamma$  versus regulation in cells expressing GR $\alpha$ ; 'Fold-difference\_vs\_GRa\_4hr\_log2.txt' and 'Fold-difference\_vs\_GRa\_24hr\_log2.txt', same as 'Fold-difference\_vs\_GRa\_2hr\_log2.txt' but after 4 and 24 h of dexamethasone treatment, respectively. Low quality and non-genic probes have been removed.

**Additional file 2: Genome-wide description of GR binding regions (GBRs).** gzip-compressed tarball (tar.gz file) containing tab-delimited text files and BED files (UCSC format), created with MACS2, describing significantly enriched GBRs from ChIP-seq data for GR $\alpha$  (GR-WT), A477T (GR-Dim), 30iib (GR-AF1), E773R (GR-AF2). Each allele is organized into a folder that contains three files (named as described above): [NAME]\_peaks.bed and [NAME]\_summits.bed files contain peak/GBR locations and (signal peak) summit locations as BED files, respectively, and [NAME]\_peaks.xls files describing statistics about each peak. Chromosomal coordinates in .xls files start with 1 and intervals contain both the start and end position.

**Additional file 3: Assignments of GBRs to genes (GREAT output).** Gzip-compressed tarball (tar.gz file) containing tab-delimited text files with assignments of peaks to genes and vice versa for each GR allele. Assignments were made with GREAT (see Methods, 'Analysis of ChIP-seq data'). Alleles and GBRs are named as in Additional file 2.

**Additional file 4: MACS2 analysis of differential occupancy between GR $\alpha$  and A477T.** Tab-delimited text files describe differentially occupied regions genome-wide or at peaks (GBRs) in each separate sample. Three files created with MACS2 are provided: GR-WT\_vs\_GR-Dim\_diffpeaks.xls, which lists all differentially occupied peaks; GR-WT\_vs\_GR-Dim\_diffpeaks\_by\_peaks1.xls and GR-WT\_vs\_GR-Dim\_diffpeaks\_by\_peaks2.xls, which list all peaks originally found with GR-Dim/A477T (annotated as sample 1) or GR-WT/GR $\alpha$  (annotated as sample 2), respectively, and statistics about their differential occupancy. Chromosomal coordinates in .xls files start with 1 and intervals contain both the start and end position.

**Additional file 5: MACS2 analysis of differential occupancy between GR $\alpha$  and 30iib.** Tab-delimited text files describe differentially occupied regions genome-wide or at peaks (GBRs) in each separate sample. Three files created with MACS2 are provided: GR-WT\_vs\_GR-AF1\_diffpeaks.xls, which lists all differentially occupied peaks; GR-WT\_vs\_GR-AF1\_diffpeaks\_by\_peaks1.xls and GR-WT\_vs\_GR-AF1\_diffpeaks\_by\_peaks2.xls, which list all peaks originally found with GR-AF1/30iib (annotated as sample 1) or GR-WT/GR $\alpha$  (annotated as sample 2), respectively, and statistics about

their differential occupancy. Chromosomal coordinates in .xls files start with 1 and intervals contain both the start and end position.

**Additional file 6: MACS2 analysis of differential occupancy between GR $\alpha$  and E773R.** Tab-delimited text files describe differentially occupied regions genome-wide or at peaks (GBRs) in each separate sample. Three files created with MACS2 are provided: GR-WT\_vs\_GR-AF2\_diffpeaks.xls, which lists all differentially occupied peaks; GR-WT\_vs\_GR-AF2\_diffpeaks\_by\_peaks1.xls and GR-WT\_vs\_GR-AF2\_diffpeaks\_by\_peaks2.xls, which list all peaks originally found with GR-AF2/E773R (annotated as sample 1) or GR-WT/GR $\alpha$  (annotated as sample 2), respectively, and statistics about their differential occupancy. Chromosomal coordinates in .xls files start with 1 and intervals contain both the start and end position.

**Additional file 7: Pairwise comparisons of GR $\alpha$  GBRs and A477T GBRs.** The comparison includes information about full site and half motifs, and contains all GBRs from GR $\alpha$  and those GBRs from A477T that do not overlap with GBRs from GR $\alpha$ . Applicable information is merged from Additional files 3, 4, 12, and 15. Alleles are named as in Additional file 2. See Additional file 17 for details about the construction of these files.

**Additional file 8: Pairwise comparisons of GR $\alpha$  GBRs and 30iib GBRs.** The comparison contains all GBRs from GR $\alpha$  and those GBRs from 30iib that do not overlap with GBRs from GR $\alpha$ . Applicable information is merged from Additional files 3 and 5. Alleles are named as in Additional file 2. See Additional file 17 for details about the construction of these files.

**Additional file 9: Pairwise comparisons of GR $\alpha$  GBRs and E773R GBRs.** The comparison contains all GBRs from GR $\alpha$  and those GBRs from E773R that do not overlap with GBRs from GR $\alpha$ . Applicable information is merged from Additional files 3 and 6. Alleles are named as in Additional file 2. See Additional file 17 for details about the construction of these files.

**Additional file 10: R script for GBR analysis (text file).** Assembly of Additional files 7 and 8 and combination of gene expression data (Additional file 1) with occupancy data. See Materials and Methods, 'Computational analysis of ChIP-seq data' for details. Execution requires R <<http://www.r-project.org/>>.

**Additional file 11: GBR positions and strengths of occupancy at genes that lost activation with A477T.** Genes are aligned at their transcription start sites with coding regions in the plus direction and promoter regions in the minus direction. GBRs are indicated as circles; size indicates the magnitude of occupancy; color indicates differential occupancy: blue indicates more binding by A477T and yellow indicates more binding by GR $\alpha$ . A yellow box from -20 kb to +20 kb indicates where there is significant enrichment of GR $\alpha$ -selective GBRs.

**Additional file 12: Motif information at GBRs occupied by GR $\alpha$  or A477T.** Gzip-compressed tarballs containing tab-delimited text files with information about the best match to the full site or half site motifs at GBRs occupied by GR $\alpha$  or A477T. Additionally, there are files containing STAT3 (MA0144.1) and AP1 (MA0099.2) for GBRs occupied by GR $\alpha$ . Alleles are named as in Additional file 2.

**Additional file 13: GBR positions and strengths of occupancy at genes that gained activation with A477T.** Blue box from -20 kb to +20 kb indicates region of significant enrichment of GR $\alpha$ -selectively occupied GBRs. See Additional file 7 legend for further description of features.

**Additional file 14: GR $\alpha$  and A477T selectively activate transcription with transcriptional reporters containing selectively occupied GBRs.** (A) Selected GBRs (approximately 500 bp, centered on the signal maximum) were cloned into a luciferase-based transcriptional reporter and transfected into U2OS cells expressing either GR $\alpha$  or A477T. (B) GBRs were selected on the basis of proximity to activated genes, strength of occupancy, and isolation from many other GBRs: GR $\alpha$ -selectively occupied GBRs proximal to GR $\alpha$ -selectively activated genes; A477T-selectively occupied GBRs proximal to A477T-selectively activated GBRs genes (with the exception of HOXD1\_GRE1, which showed GR $\alpha$ -selective occupancy); and non-selectively occupied GBRs proximal to non-selectively activated genes. GBRs drove GR $\alpha$ -selective activation with several GR $\alpha$ -selectively occupied GBRs, A477T-selective activation with

several A477T-selectively occupied GBRs, and non-selective activation with several non-selectively occupied GBRs. Some GBRs did not activate transcription in either cell line (for example, KCNA5\_GRE1-3), but only two (HOXD1\_GRE1 and CHST15\_GRE1) drove aberrant activation (opposing selectivity of occupancy and activation).

**Additional file 15: Motif analysis of selectively and nonselectively occupied GBRs.** Gzip-compressed tarball (.tar.gz file) containing folders with motif results (MEME-ChIP) for three groups of GBRs, with and without enforcement of a 17 bp motif in MEME (six folders total). 'GRa\_only' and 'GRa\_only.17 bp': the thousand most-occupied GBRs that were occupied by GRa and did not have detectable occupancy by A477T. 'GRa\_and\_A477T' and 'GRa\_and\_A477T.17 bp': the thousand most-GRa-occupied GBRs that were occupied by both GRa and A477T with less than a 1.4-fold difference. 'A477T\_only' and 'A477T\_only.17 bp': The thousand most-occupied GBRs that were occupied by A477T and did not have detectable occupancy by GRa. Each folder contains a file named 'index.html', which can be opened in a web browser to view the results. See Figure 5 for a summary.

**Additional file 16: GBR positions and strength of occupancy at non-selectively activated<sup>c</sup> genes.** See Additional file 7 legend for a description of features.

**Additional file 17: R script for analysis of gene expression microarray data (text file).** See Materials and Methods, 'Computational analysis of gene expression data' for details. Execution requires R, available at <<http://www.r-project.org/>>.

**Additional file 18: Table of primers used for reporters.**

#### Abbreviations

AF1: Activation function 1; AF2: Activation function 2; AP1: Activation protein 1; AR: Androgen receptor; bp: Base pairs; ChIP: Chromatin immunoprecipitation; ChIP-seq: Chromatin immunoprecipitation sequencing; DNA: Deoxyribonucleic acid; GBR: Glucocorticoid receptor binding region; GBS: Glucocorticoid receptor binding sequence; GR: Glucocorticoid receptor; GRE: Glucocorticoid receptor response element; kb: Kilo-base pairs; NMR: Nuclear magnetic resonance spectroscopy; STAT: Signal Transducers and Activators of Transcription.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

BJS participated in the conception and design of the study, measurements of occupancy by GR, and reporter assays; carried out all computational and statistical analyses, prepared figures, and drafted the manuscript. RC participated in the conception and design of the study, confirmed the presence of GR mutations in all stable cell lines, carried out microarray and PCR gene expression measurements, and validated occupancy by GR at genomic regions used with the reporter assay. LCW fitted electrophoretic mobility shift assay data, and participated in measurements of occupancy by GR and in reporter assays. MRS participated in the conception and design and coordination of the study. KRY participated in the conception and design and coordination of the study and helped to draft the manuscript. All authors read and approved the final manuscript.

#### Acknowledgements

Samantha Cooper was involved in the design of the study and optimizing chromatin immunoprecipitation sequencing protocols. Sheng-Hong Chen provided the luciferase reporter vectors pGL4.10-E4TATA and pGL4.70-MLE2. Clement Chu and Jessica Lund at the UCSF Center for Advanced Technology aided in the acquisition of sequencing data. Processing of RNA samples and gene expression microarrays were performed by the University of Southern California Epigenome Center. Joel Mefford suggested the statistical test used to identify non-selective, activated genes. Funding for this study was provided by grants to KRY (R01 CA020535) and to MRS (R01 DK043093) from the National Institute of Health and also by Cancer Center Support Grant P30CA014089 from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

#### Author details

<sup>1</sup>Department of Cellular and Molecular Pharmacology, University of California, 600 16th Street, S-222, San Francisco, CA 94158-2280, USA. <sup>2</sup>Department of Biochemistry and Molecular Biology, University of Southern California, 1441 Eastlake Avenue, NOR 6316, Los Angeles, CA 90089-9176, USA.

Received: 19 May 2014 Accepted: 17 July 2014

Published: 31 July 2014

#### References

1. Rogatsky I, Trowbridge JM, Garabedian MJ: **Glucocorticoid receptor-mediated cell cycle arrest is achieved through distinct cell-specific transcriptional regulatory mechanisms.** *Molecular and Cellular Biology* 1997, **17**:3181–3193.
2. Rogatsky I, Wang JC, Derynck MK, Nonaka DF, Khodabakhsh DB, Haqq CM, Darimont BD, Garabedian JM, Yamamoto KR: **Target-specific utilization of transcriptional regulatory surfaces by the glucocorticoid receptor.** *Proceedings of the National Academy of Sciences* 2003, **100**:13845–13850.
3. Meijsing SH, Pufall MA, So AY, Bates DL, Chen L, Yamamoto KR: **DNA binding site sequence directs glucocorticoid receptor structure and activity.** *Science* 2009, **324**:407–410.
4. Watson LC, Kuchenbecker KM, Schiller BJ, Gross JD, Pufall MA, Yamamoto KR: **The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals.** *Nat Struct Mol Biol* 2013, **20**:878–883.
5. Strähle U, Klock G, Schütz G: **A DNA sequence of 15 base pairs is sufficient to mediate both glucocorticoid and progesterone induction of gene expression.** *Proc Natl Acad Sci U S A* 1987, **84**:7871–7875.
6. Thomas-Cholliera M, Watson LC, Cooper SC, Pufall MA, Liud JS, Borzym K, Vingron M, Yamamoto KR, Meijsing SH: **A naturally occurring insertion of a single amino acid rewires transcriptional regulation by glucocorticoid receptor isoforms.** *Proc Natl Acad Sci U S A* 2013, **110**:17826–17832.
7. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3**:3.
8. Chen Y, Negre N, Li Q, Mieczkowska JO, Slattery M, Liu T, Zhang Y, Kim TK, He HH, Zieba J, Ruan Y, Bickel PJ, Myers RM, Wold BJ, White KP, Lieb JD, Liu XS: **Systematic evaluation of factors influencing ChIP-seq fidelity.** *Nat Methods* 2012, **9**:609–614.
9. McLean CY, Bristol D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G: **GREAT improves functional interpretation of cis-regulatory regions.** *Nature Biotechnol* 2010, **28**:495–501.
10. Heck S, Kullmann M, Gast A, Ponta H, Rahmsdorf HJ, Herrlich P, Cato AC: **A distinct modulating domain in glucocorticoid receptor monomers in the repression of activity of the transcription factor AP-1.** *EMBO J* 1994, **13**:4087–4095.
11. Reichardt HM, Kaestner KH, Tuckermann J, Kretz O, Wessely O, Bock R, Gass P, Schmid W, Herrlich P, Angel P, Schütz G: **DNA binding of the glucocorticoid receptor is not essential for survival.** *Cell* 1998, **93**:531–541.
12. Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC: **Highly efficient endogenous human gene correction using designed zinc-finger nucleases.** *Nature* 2005, **435**:646–651.
13. Porteus MH, Baltimore D: **Chimeric nucleases stimulate gene targeting in human cells.** *Science* 2003, **300**:763.
14. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F: **Multiplex genome engineering using CRISPR/Cas systems.** *Science* 2013, **339**:819–823.
15. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM: **RNA-guided human genome engineering via Cas9.** *Science* 2013, **339**:823–826.
16. Machanic P, Bailey TL: **MEME-ChIP: motif analysis of large DNA datasets.** *Bioinformatics* 2011, **27**:1696–1697.
17. Bailey TL, Elkan C: **Fitting a mixture model by expectation maximization to discover motifs in bipolymers.** San Diego, CA: Department of Computer Science and Engineering, University of California; 1994.
18. So A, Chaiworapal C, Bolton E, Li H, Yamamoto K: **Determinants of cell-and gene-specific transcriptional regulation by the glucocorticoid receptor.** *PLoS Genet* 2007, **3**:e94.
19. Cooper S: **Glucocorticoid receptor DNA occupancy and transcriptional regulation across cell types.** In *PhD thesis*. San Francisco, CA: University of California; 2010.

20. Bailey TL, Machanick P: **Inferring direct DNA binding from ChIP-seq.** *Nucleic Acids Res* 2012, **40**:e128.
21. Bailey TL: **DREME: motif discovery in transcription factor ChIP-seq data.** *Bioinformatics* 2011, **27**:1653–1659.
22. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addleman N, Bhardwaj N, Boyle AP, Cayting P, Charos A, Chen DZ, Cheng Y, Clarke D, Eastman C, Euskirchen G, Fritze S, Fu Y, Gertz J, Grubert F, Harmanci A, Jain P, Kasowski M, *et al*: **Architecture of the human regulatory network derived from ENCODE data.** *Nature* 2012, **489**:91–100.
23. Necela BM, Cidlowski JA: **Mechanisms of glucocorticoid receptor action in noninflammatory and inflammatory cells.** *Proc Am Thorac Soc* 2004, **1**:239–246.
24. Rogatsky I, Luecke HF, Leitman DC, Yamamoto KR: **Alternate surfaces of transcriptional coregulator GRIP1 function in different glucocorticoid receptor activation and repression contexts.** *Proc Natl Acad Sci U S A* 2002, **99**:16701–16706.
25. Schneider TD: **Information content of individual genetic sequences.** *J Theor Biol* 1997, **189**:427–441.
26. Gebhardt JCM, Suter DM, Roy R, Zhao ZW, Chapman AR, Basu S, Maniatis T, Xie XS: **Single-molecule imaging of transcription factor binding to DNA in live mammalian cells.** *Nat Methods* 2013, **10**:421–426.
27. Peirano RI, Wegner M: **The glial transcription factor Sox10 binds to DNA both as monomer and dimer with different functional consequences.** *Nucleic Acids Res* 2000, **28**:3047–3055.
28. Tewari AK, Yardimci G, Shibata Y, Sheffield NC, Song L, Taylor BS, Georgiev SG, Coetzee GA, Ohler U, Furey TS, Crawford GE, Febbo PG: **Chromatin accessibility reveals insights into androgen receptor activation and transcriptional specificity.** *Genome Biol* 2012, **13**:R88.
29. Roohk DJ, Mascharak S, Khambatta C, Leung H, Hellerstein M, Harris C: **Dexamethasone-mediated changes in adipose triacylglycerol metabolism are exaggerated, not diminished, in the absence of a functional GR dimerization domain.** *Endocrinology* 2013, **154**:1528–1539.
30. Wang JC, Shah N, Pantoja C, Meijising SH, Ho JD, Scanlan TS, Yamamoto KR: **Novel arylpyrazole compounds selectively modulate glucocorticoid receptor regulatory activity.** *Genes Dev* 2006, **20**:689–699.
31. Zhang JZ, Cavet ME, VanDerMeid KR, Salvador-Silva M, López FJ, Ward KW: **BOL-303242-X, a novel selective glucocorticoid receptor agonist, with full anti-inflammatory properties in human ocular cells.** *Mol Vis* 2009, **15**:2606.
32. Wood AJ, Riggs BL, Hartmann LC: **Selective estrogen-receptor modulators – mechanisms of action and application to clinical practice.** *New Engl J Med* 2003, **348**:618–629.
33. Jones JO, Diamond MI: **A cellular conformation-based screen for androgen receptor inhibitors.** *ACS Chem Biol* 2008, **3**:412–418.
34. Won Jeong K, Chodankar R, Purcell DJ, Bittencourt D, Stallcup MR: **Gene-specific patterns of coregulator requirements by estrogen receptor- $\alpha$  breast cancer cells.** *Mol Endocrinol* 2012, **26**:955–966.
35. Dunning MJ, Smith ML, Ritchie ME: **Tavaré S: beadarray: R classes and methods for Illumina bead-based data.** *Bioinformatics* 2007, **23**:2183–2184.
36. Gentleman RC, Carey VJ, Bates DM: **Bioconductor: Open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5**:R80.
37. R Core Team: **R: A Language and Environment for Statistical Computing.** Vienna, Austria: R Foundation for Statistical Computing; 2012. [<http://www.R-project.org/>]
38. Cairns JM, Dunning MJ, Ritchie ME, Russell R, Lynch AG: **BASH: a tool for managing BeadArray spatial artefacts.** *Bioinformatics* 2008, **24**:2921–2922.
39. Johnson WE, Li C, Rabinovic A: **Adjusting batch effects in microarray expression data using empirical Bayes methods.** *Biostatistics* 2006, **8**:118–127.
40. Edgar R: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic Acids Res* 2002, **30**:207–210.
41. Schiller BJ, Chondankar R, Watson LC, Stallcup MR, Yamamoto KR: **Gene expression of hormone-treated U2OS cells expressing GR alleles.** [<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45407>]
42. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J Roy Stat Soc B Stat Meth* 1995, **57**:289–300.
43. Wickham H: **ggplot2: elegant graphics for data analysis.** New York: Springer; 2009.
44. Chory J, Pollard JD: **Current Protocols in Molecular Biology.** Chichester: Wiley Blackwell (John Wiley & Sons); 2001.
45. **Sequence Read Archive.** [<http://www.ncbi.nlm.nih.gov/sra>]
46. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL: **Biopython: freely available Python tools for computational molecular biology and bioinformatics.** *Bioinformatics* 2009, **25**:1422–1423.
47. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat Meth* 2012, **9**:357–359.
48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPP: **The sequence alignment/map format and SAMtools.** *Bioinformatics* 2009, **25**:2078–2079.
49. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human genome browser at UCSC.** *Genome Res* 2002, **12**:996–1006.
50. **GBR occupancy of hormone-treated U2OS cells expressing GR alleles.** [[http://kryweb.ucsf.edu/data/Schiller\\_et\\_al\\_2014/hub.txt](http://kryweb.ucsf.edu/data/Schiller_et_al_2014/hub.txt)]
51. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, **26**:841–842.
52. Smit A, Hubley R, Green P: **RepeatMasker Open-3.0b.** Institute for Systems Biology: Seattle, WA; 1996.
53. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ: **The UCSC Table Browser data retrieval tool.** *Nucleic Acids Res* 2004, **32**:D493–D496.
54. Korhonen J, Martinmaki P, Pizzi C, Rastas P, Ukkonen E: **MOODS: fast search for position weight matrix matches in DNA sequences.** *Bioinformatics* 2009, **25**:3181–3182.
55. Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B: **JASPAR: an open-access database for eukaryotic transcription factor binding profiles.** *Nucleic Acids Res* 2004, **32**:D91–D94.
56. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based analysis of ChIP-Seq (MACS).** *Genome Biol* 2008, **9**:R137.
57. Wilks SS: **The large-sample distribution of the likelihood ratio for testing composite hypotheses.** *Ann Math Stat* 1938, **9**:60–62.
58. **MACS – Model-based Analysis of ChIP-Seq.** [<https://github.com/taoliu/MACS/>]

doi:10.1186/s13059-014-0418-y

**Cite this article as:** Schiller *et al*: Glucocorticoid receptor binds half sites as a monomer and regulates specific target genes. *Genome Biology* 2014 **15**:418.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

