

# Determinants of Cell- and Gene-Specific Transcriptional Regulation by the Glucocorticoid Receptor

Alex Yick-Lun So<sup>1,2</sup>, Christina Chaivorapol<sup>3,4,5</sup>, Eric C. Bolton<sup>1</sup>, Hao Li<sup>3,4,5</sup>, Keith R. Yamamoto<sup>1,2\*</sup>

**1** Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, California, United States of America, **2** Chemistry and Chemical Biology Graduate Program, University of California San Francisco, San Francisco, California, United States of America, **3** Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, California, United States of America, **4** California Institute for Quantitative Biomedical Research, University of California San Francisco, San Francisco, California, United States of America, **5** Graduate Program in Biological and Medical Informatics, University of California San Francisco, San Francisco, California, United States of America

**The glucocorticoid receptor (GR) associates with glucocorticoid response elements (GREs) and regulates selective gene transcription in a cell-specific manner. Native GREs are typically thought to be composite elements that recruit GR as well as other regulatory factors into functional complexes. We assessed whether GR occupancy is commonly a limiting determinant of GRE function as well as the extent to which core GR binding sequences and GRE architecture are conserved at functional loci. We surveyed 100-kb regions surrounding each of 548 known or potentially glucocorticoid-responsive genes in A549 human lung cells for GR-occupied GREs. We found that GR was bound in A549 cells predominately near genes responsive to glucocorticoids in those cells and not at genes regulated by GR in other cells. The GREs were positionally conserved at each responsive gene but across the set of responsive genes were distributed equally upstream and downstream of the transcription start sites, with 63% of them >10 kb from those sites. Strikingly, although the core GR binding sequences across the set of GREs varied extensively around a consensus, the precise sequence at an individual GRE was conserved across four mammalian species. Similarly, sequences flanking the core GR binding sites also varied among GREs but were conserved at individual GREs. We conclude that GR occupancy is a primary determinant of glucocorticoid responsiveness in A549 cells and that core GR binding sequences as well as GRE architecture likely harbor gene-specific regulatory information.**

Citation: So AYL, Chaivorapol C, Bolton EC, Li H, Yamamoto KR (2007) Determinants of cell- and gene-specific transcriptional regulation by the glucocorticoid receptor. *PLoS Genet* 3(6): e94. doi:10.1371/journal.pgen.0030094

## Introduction

The great challenge of metazoan transcriptional regulation is to create specialized expression pathways that accommodate and define myriad contexts, i.e., different developmental, physiological, and environmental states in distinct organs, tissues, and cell types. This is achieved by a network of transcriptional regulatory factors, which receive and integrate signaling information and transduce that information by binding close to specific target genes to modulate their expression. For example, the glucocorticoid receptor (GR) associates selectively with corticosteroid ligands produced in the adrenal gland in response to neuroendocrine cues; the GR-hormone interaction promotes GR binding to genomic glucocorticoid response elements (GREs), in turn modulating the transcription of genes that affect cell differentiation, inflammatory responses, and metabolism [1,2]. Expression profile analyses have identified glucocorticoid responsive genes in different cell types [3,4], and it is striking that there is only modest overlap in glucocorticoid-regulated gene sets between two cell types. The mechanisms by which GR selectively regulates transcription in cell-specific contexts are not well established.

An intriguing feature of GREs and other metazoan response elements is that their positions relative to their target genes are not tightly constrained [5,6]. Although certain metazoan response elements have been described that operate from long range, most searches for such regulatory sequences have nevertheless focused for technical

reasons on restricted zones just upstream of promoters, where prokaryotic and fungal elements reside. Thus, the GRE for interleukin-8 (*IL8*) is just upstream of the promoter [7], whereas the tyrosine aminotransferase GRE resides at −2.5 kb [8]. Recent, more systematic searches for response elements have revealed dramatic examples, such as an estrogen response element 144 kb upstream from the promoter of the *NRIP* gene [9], and an intragenic region 65 kb downstream from the *Fkbp5* promoter that appears to serve as an androgen response element [10]. It has been suggested that long-range regulatory mechanisms are likely to facilitate and promote regulatory evolution [11]. However, it has not been determined whether the position of a response element relative to its target gene is functionally significant.

Evidence from numerous anecdotal, gene-specific studies indicates that native response elements are typically compo-

**Editor:** Lisa Stubbs, Lawrence Livermore National Laboratory, United States of America

**Received:** January 9, 2007; **Accepted:** April 23, 2007; **Published:** June 8, 2007

**Copyright:** © 2007 So et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Abbreviations:** AR, androgen receptor; CHIP, chromatin immunoprecipitation; chip, microarray chip; CNG, conserved nongenic sequence; dex, dexamethasone; GBR, GR-binding region; GR, glucocorticoid receptor; GRE, glucocorticoid response element; *IL8*, interleukin-8; qPCR, quantitative PCR; TSS, transcription start site

\* To whom correspondence should be addressed. E-mail: yamamoto@cmp.ucsf.edu

## Author Summary

The glucocorticoid receptor (GR) regulates a myriad of physiological functions, such as cell differentiation and metabolism, achieved through modulating transcription in a cell- and gene-specific manner. However, the determinants that specify cell- and gene-specific GR transcriptional regulation are not well established. We describe three properties that contribute to this specificity: (1) GR occupancy at genomic glucocorticoid response elements (GREs) appears to be a primary determinant of glucocorticoid responsiveness; (2) the DNA sequences bound by GR vary widely around a consensus, but the precise sequences of individual GREs are highly conserved, suggesting a role for these sequences in gene-specific GR transcriptional regulation; and (3) native chromosomal GREs were generally found to be composite elements, comprised of multiple factor binding sites that were highly variable in composition, but as with the GR binding sequences, highly conserved at individual GREs. In addition, we discovered that most GREs were positioned far from their GR target genes and that they were equally distributed upstream and downstream of the target genes. These findings, which may be applicable to other regulatory factors, provide fundamental insights for understanding cell- and gene-specific transcriptional regulation.

site elements that encompass distinct sequence motifs recognized by two or more regulatory factors. In turn, the bound factors recruit non-DNA binding coregulatory factors, forming functional regulatory complexes that remodel chromatin and modify the activity of the transcription machinery. In this scheme, the structure and activity of the regulatory complex at a given response element would be specified by at least three determinants: the sequence motifs comprising the response element; the availability of those sequences for factor binding; and the availability and activity levels of regulatory factors present in the cell. For example, primary GREs, defined as those at which GR occupancy is required for glucocorticoid-responsive regulation, are a diverse family of elements that bind GR together with an array of additional factors defined by the above three determinants. Such composite response elements provide a powerful driving force for combinatorial regulation [2,12], vastly increasing the capacity of a single factor to assume multiple regulatory roles. Indeed, the mere presence of GR in a regulatory complex is not sufficient for glucocorticoid regulation [7]. It is not known, however, whether such “nonproductive” binding by GR is common, or if instead GR occupancy is a strong indicator of GRE function.

GR binds to a family of related sequences that defines a consensus motif: an imperfect palindrome of hexameric half sites separated by a three-bp spacer [13–15]. Within those 15-bp core GR binding sequences, a few positions are nearly invariant, whereas a substantial proportion can be altered with little effect on GR binding affinity [16]. However, the functional consequences of such “permitted” sequence variations are unknown. GR can mediate a range of regulatory processes within a single cell type, including activation and repression of specific genes [4,17]. These findings, together with the results of biochemical and structural studies, raise the possibility that the core GR binding sequences might themselves serve as distinct “GR ligands,” allosterically affecting GR structure to produce distinct GR functions [18,19]. Studies of other regulatory

factors have led to similar conclusions [20,21]. If different core GR binding sequences indeed produce GRE-specific (and therefore target gene-specific) regulatory activities, we could expect that the core GR binding sequence associated with a given target gene would be strongly conserved through evolution, whereas the collection of core GR binding sequences across different genes would vary substantially. Analogously, if the architecture of composite GREs, i.e., the arrangements of additional sequence motifs surrounding the core GR binding site, are also important for gene-specific regulation, we would expect flanking sequences surrounding the core GR binding site to also be evolutionarily conserved in a GRE-specific manner but not across GREs within a single genome. Neither of these notions has been examined.

In the present work, we sought to define and characterize a set of GREs in A549 human alveolar epithelial cells. Thus, we determined in A549 cells the presence of GR at specific GREs close to genes that are steroid regulated across a range of cell types. We assessed whether the GR-occupied GREs were limited mainly to genes that are GR regulated in A549 and measured within and between species the conservation of GRE sequences, architecture, and genomic positions.

## Results

### Identification of GR Binding Regions Using Chromatin Immunoprecipitation-Microarray

To assess the correlation of GR occupancy with glucocorticoid responsiveness, we examined GR binding at three classes of genes in A549 human lung carcinoma cells: first, genes regulated by GR in A549 cells; second, genes regulated by GR in U2OS human osteosarcoma cells but not in A549; third, genes regulated by GR or the androgen receptor (AR) in cells other than A549 or U2OS. The AR-responsive genes were of interest because AR is closely related to GR and shares similar DNA-binding specificity *in vitro* [14,16,22]. The first two classes of genes were identified in our lab using expression microarrays, whereas the third class was compiled from our own microarray data and from published reports of others [3,4,23,24]. Both positively and negatively regulated genes were included; together the three classes comprised 548 candidate GR target genes. By examining these genes for GR binding in A549 cells, we could determine if GR occupancy *in vivo* is restricted only at genomic sites of genes actually regulated by glucocorticoids in A549 cells; alternatively, GR might also bind at genes that are not under glucocorticoid control in A549, but are regulated by GR or AR in other cells.

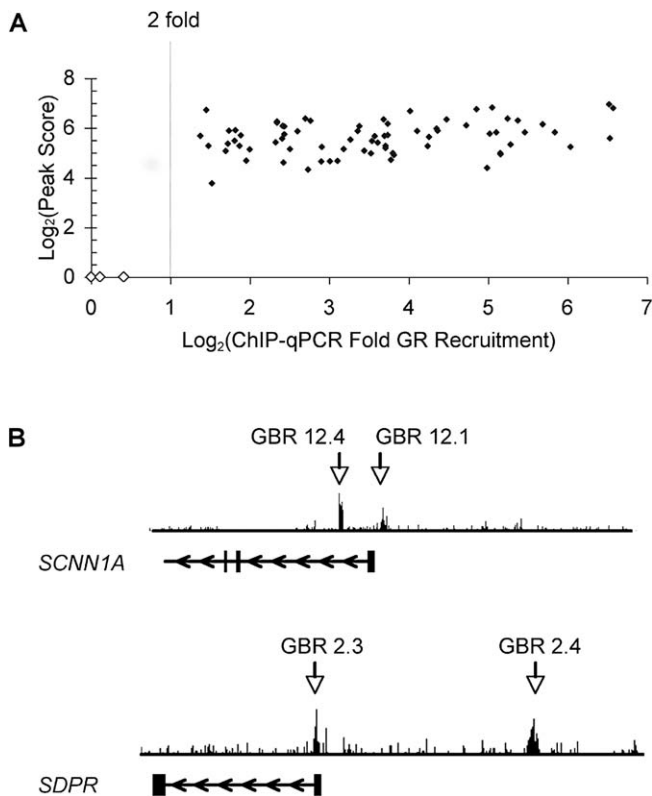
To identify GR binding regions (GBRs), we used chromatin immunoprecipitation-microarray (ChIP-chip) to interrogate 100-kb genomic segments centered on the transcription start sites (TSSs) of our set of 548 genes. This ~55-Mb sample of the genome also included or impinged upon an additional 587 genes not previously reported to be regulated by GR; thus, we assessed GR occupancy in the vicinity of more than 1,000 genes. Immunoprecipitated chromatin samples from A549 cultures treated for one hour with the synthetic glucocorticoid dexamethasone (dex) (100 nM) or ethanol were hybridized onto the ChIP-chip tiling arrays. Independent biological replicates were hybridized onto two separate arrays, and GBRs were identified using the SignalMap detection program; we detected a 3.4% false positive rate for the GBRs found in both arrays as assessed by conventional

**Table 1:** Relative Locations of GBRs

GBR	Coordinate	RefSeq	Target Gene	GBR Distance to TSS (bp)	Position of GBR	GBR	Coordinate	RefSeq	Target Gene	GBR Distance to TSS (bp)	Position of GBR
1.1	55052924	NM_014762	DHCR24	12019	Intron	10.3	70491407	NM_002727	PRG1 <sup>a</sup>	-26427	Upstream
1.2	151203962	NM_000565	IL6R <sup>a</sup>	13220	Intron	10.4	73685003	NM_019058	DDIT4 <sup>a</sup>	-18680	Upstream
1.4	166187035	NM_006996	SLC19A2 <sup>a</sup>	-136	Upstream	10.5	73678745	NM_019058	DDIT4 <sup>a</sup>	-24939	Upstream
1.5	200851096	NM_018208	ETNK2 <sup>a</sup>	1584	Intron	11.1	16991094	NG_002175	OR7E14P <sup>a</sup>	-39285	Upstream
1.6	231085797	NM_182972	IRF2BP2 <sup>a</sup>	-33791	Upstream	11.2	101682322	NM_001165	BIRC3 <sup>a</sup>	-11082	Upstream
2.1	119900074	NM_183240	TMEM37 <sup>a</sup>	-5602	Upstream	12.1	6355825	NM_001038	SCN11A <sup>a</sup>	-849	Upstream
2.2	169227864	NM_052946	NOSTRIN <sup>a</sup>	-256751	Upstream	12.2	51174824	NM_005554	KRT6A <sup>a</sup>	-1537	Upstream
2.3	192537475	NM_004657	SDPR <sup>a</sup>	12	Upstream/UTR	12.3	84732829	NM_005447	PAMC1 <sup>a</sup>	-43	Upstream
2.4	192557274	NM_004657	SDPR <sup>a</sup>	-19787	Upstream	12.4	6352704	NM_001038	SCN11A <sup>a</sup>	2273	Intron
2.5	228508484	NM_004591	CCL20 <sup>a</sup>	4409	Downstream of gene	12.5	45713321	NM_181847	AMIGO2 <sup>a</sup>	46595	Downstream of gene
3.1	42561322	NM_004624	VIPR1 <sup>a</sup>	42201	Downstream of gene	13.1	30185521	NM_001629	ALOX5A <sup>a</sup>	-22148	Upstream
3.2	100124471	NM_080927	DCBLD2	-21248	Upstream	13.2	42477337	NM_001002264	EPSTT1 <sup>a</sup>	-12960	Upstream
4.1	6870929	NM_014743	KIAA0232 <sup>a</sup>	-31610	Upstream	14.1	34942864	NM_020529	NFKBIA <sup>a</sup>	831	Intron
4.2	75330454	NM_020289	CXCL2 <sup>a</sup>	-507	Upstream	14.2	94135422	NM_001085	SERPINA3 <sup>a</sup>	-13046	Upstream
4.3	89890827	NM_014606	HERC3 <sup>a</sup>	20002	Intron	15.4	87442841	NM_007011	ABHD2 <sup>a</sup>	10412	Intron
4.4	186109853	NM_001995	ACSL1 <sup>a</sup>	12511	Intron	15.5	87461998	NM_007011	ABHD2 <sup>a</sup>	29569	Intron
4.5	74971324	NM_000584	IL8 <sup>a</sup>	14	Upstream/UTR	16.1	55198785	NM_005953	MT2A <sup>a</sup>	-1227	Upstream
5.1	142113152	NM_015071	ARHGAP26	-17325	Upstream	16.2	67307057	NM_004360	CDH1 <sup>a</sup>	-21639	Upstream
6.1	35677841	NM_004117	FKBP5 <sup>a</sup>	86852	Intron	16.3	84531505	NM_002163	IRF8	41230	Downstream of gene
6.2	35801593	NM_145028	Coorf81 <sup>a</sup>	-11245	Upstream	17.1	35896009	NM_032865	TNS4 <sup>a</sup>	15333	Intron
6.3	35802767	NM_145028	Coorf81 <sup>a</sup>	-10071	Upstream	17.2	35920025	NM_032865	TNS4 <sup>a</sup>	-8684	Upstream
6.4	134539075	NM_005627	SGK <sup>a</sup>	-1380	Upstream	17.3	35952509	NM_001838	CCR7	22741	Downstream of gene
6.5	47774006	NM_153838	GPR115 <sup>a</sup>	-242	Upstream	17.4	59513438	NM_001433	ERN1	47771	Intron
7.1	45695008	NM_000596	IGFBP1 <sup>a</sup>	-6191	Upstream	17.5	72065060	NM_134268	CYGB	-19683	Upstream
7.2	45702450	NM_000596	IGFBP1 <sup>a</sup>	1252	Intron	17.6	75652568	NM_371082	CDC40 <sup>a</sup>	27516	Intron
7.3	137096157	NM_194071	CREB3L2	47944	Intron	19.1	8326625	NM_139314	ANGPTL4 <sup>a</sup>	-8387	Upstream
8.2	82110304	NM_018440	PAG1 <sup>a</sup>	76555	Intron	19.2	12722377	NM_017682	VMD2L1 <sup>a</sup>	-2095	Upstream
9.1	72160383	NM_006007	ZAZO2 <sup>a</sup>	49063	Downstream of gene	20.1	5586238	NM_019593	KIAA1434 <sup>a</sup>	-46566	Upstream
9.2	72210285	NM_006007	ZAZO2 <sup>a</sup>	-839	Upstream	20.2	22938499	NM_001052	SSTR4 <sup>a</sup>	-25622	Upstream
9.3	121214984	NM_004099	STOM <sup>a</sup>	-2885	Upstream	20.4	25166277	NM_002862	PYGB <sup>a</sup>	-10430	Upstream
10.1	5714370	NM_024701	ASB13	34178	Downstream of gene	20.5	25206522	NM_002862	PYGB <sup>a</sup>	29816	Intron
10.2	47238469	NM_001630	ANXA8 <sup>a</sup>	21543	Downstream of gene	22.1	29143649	NM_012429	SECT14L2 <sup>a</sup>	26162	UTR
1.3	151757059	NM_003029	SHC1 <sup>a</sup>	2854	Intron	15.2	83688869	NM_006738	AKAP13 <sup>a</sup>	-36007	Upstream
5.2	175227852	NM_006650	CPLX2 <sup>a</sup>	71636	Intron	15.3	83924520	NM_006738	AKAP13 <sup>a</sup>	199645	Exon
15.1	83675562	NM_006738	AKAP13 <sup>a</sup>	-3255	Upstream	20.3	83924520	NM_006738	AKAP13 <sup>a</sup>	-96753	Upstream
8.1	37840809	NM_001002814	RAB11FIP1 <sup>a</sup>	-49313	Upstream	X.1	106791144	NM_198057	THBD <sup>a</sup>	4879	Downstream of gene
37840809	NM_025151	RAB11FIP1 <sup>a</sup>	RAB11FIP1 <sup>a</sup>	-345710	Upstream	17.1	106791144	NM_001015881	TSC22D3 <sup>a</sup>	9301	Downstream of gene
37840809	NM_001002233	RAB11FIP1 <sup>a</sup>	RAB11FIP1 <sup>a</sup>	35353	Intron	X.2	106791144	NM_004089	TSC22D3 <sup>a</sup>	34019	Intron
				9146	Intron		106767828	NM_198057	TSC22D3 <sup>a</sup>	-25288	Upstream
							106767828	NM_001015881	TSC22D3 <sup>a</sup>	-24708	Upstream
							106767828	NM_001015881	TSC22D3 <sup>a</sup>	57335	Intron
							106767828	NM_004089	TSC22D3 <sup>a</sup>	-1972	Upstream
							106767828	NM_004089	TSC22D3 <sup>a</sup>	-1392	Upstream

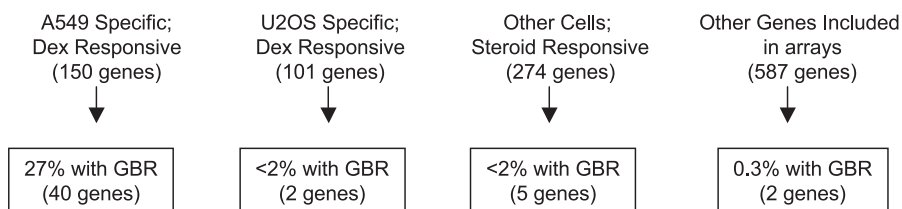
The coordinate numbers of the corresponding GBR represent the center of the GBRs obtained from UCSC Genome Browser (<http://genome.ucsc.edu>). All listed GBRs were validated as bona fide GR-occupied regions by conventional ChIP-qPCR (see Figure 1A). The GBRs were assigned to the nearest dex-responsive gene from the final list of genes that were included or impinged upon by the ChIP-chip arrays. GBRs without an associated A549 dex-responsive gene were assigned to the nearest RefSeq. Position of GBR indicates whether the GBR is: intronic ("intron"); upstream of the TSS ("upstream"); or downstream from the coding sequence of the target gene ("downstream of gene"). The longest transcript is listed first if a gene contains multiple alternative TSSs.

<sup>a</sup>Genes responsive to dex in A549 cells as shown in previous expression microarray profiles [3] or in this study (Figure S1). doi:10.1371/journal.pgen.0030094.t001



**Figure 1.** ChIP-Chip Identified Known and Novel GR Binding Regions (A) Identification of GBRs. The Log<sub>2</sub>(peak score) of GBRs obtained from the ChIP-chip arrays is plotted versus the dex-induced enrichment of GR at the corresponding GBR, which was assessed by GR ChIP-qPCR (averaged over at least three independent experiments. Note: dex-induced GR binding at the GBRs reproducibly in all the individual experiments). Solid diamonds, bona fide GBRs; open diamonds, negative control regions. (B) GBRs identified near *SCNN1A* and *SDPR* genes. Vertical bars, exons; horizontal lines, introns; arrows, direction of transcription. GBRs 12.1 and 2.3 are known promoter proximal GBRs associated with *SCNN1A* and *SDPR*, respectively. GBRs 12.4 and 2.4 are novel GBRs identified in the present work. GBR nomenclature: unique identifiers corresponding to the human chromosome number containing the GBR followed by an arbitrary integer tag. doi:10.1371/journal.pgen.0030094.g001

ChIP and quantitative PCR (qPCR) analysis. Importantly, we did not detect GR occupancy at 22 regions that showed no GR binding in the arrays (unpublished data). The ChIP-chip experiments revealed a total of 73 GBRs adjacent to 61 genes



**Figure 2.** Percentage of Genes Associated with One or More GBRs in A549 Cells A549-specific dex-responsive genes are regulated by GR in A549 cells but not U2OS cells. U2OS-specific dex-responsive genes are regulated by GR in U2OS cells but not in A549 cells. The 34 genes regulated by GR in both A549 and U2OS cells, 12 of which associated with an A549 GBR, were excluded from the analysis shown. Genes responsive to glucocorticoids or androgens in cells other than A549 and U2OS are denoted as “other cells steroid responsive.” Lastly, additional genes that were wholly or partially included in our ChIP-chip arrays due to the extensive sampling of regions around all the genes mentioned above are represented as “genes included in arrays.” doi:10.1371/journal.pgen.0030094.g002

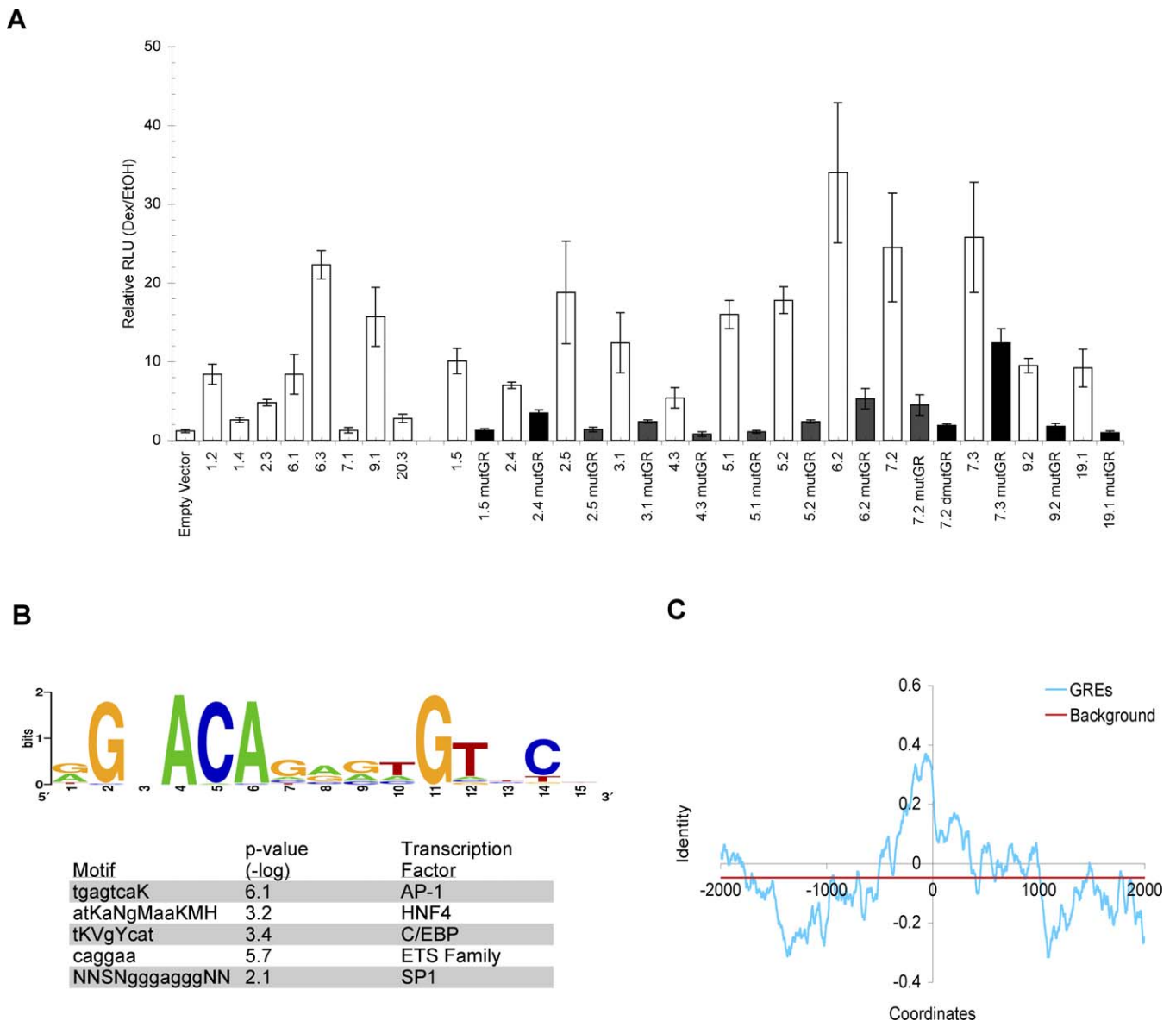
(Table 1), which were validated by GR ChIP and qPCR analysis (Figure 1A). In addition to identifying GBRs previously detected by conventional ChIP, our experiments revealed novel GBRs in regions not searched in prior studies. For example, two known promoter proximal GBRs at *SCNN1A* [25] and *SDPR* [3] were confirmed in the ChIP-chip arrays as well as newly observed GBRs +3 kb and -20 kb from the *SCNN1A* and *SDPR* TSSs, respectively (Figure 1B).

### GR Occupancy Correlates with Glucocorticoid Responsiveness

Of the 73 A549 GBRs identified in the present study, 64 (88%) were associated with genes regulated by GR in those cells (Table 1). Although the remaining nine GBRs may be nonfunctional, they may mediate responses under different biological conditions. Notably, 27% of the genes that were glucocorticoid responsive specifically in A549 but not in U2OS cells were associated with a GBR, whereas only 1.9% of the genes responsive to glucocorticoids in U2OS but not in A549 contained A549 GBRs (Figure 2). Similarly, only 1.8% of the genes that were glucocorticoid or androgen responsive in other cells and only 0.3% of the genes that were sampled by the ChIP-chip arrays but were not steroid regulatory targets were associated with A549 GBRs (Figure 2; Table S1). Thus, GR occupancy in A549 cells is generally restricted to genes that are actually regulated by glucocorticoids in those cells; specifically, GR is rarely bound in A549 cells at genes responsive to glucocorticoids in other cells. We conclude that GR occupancy is a major determinant of glucocorticoid responsiveness in A549 cells at the genes assessed in this study.

### A549 GBRs Are Functional GREs

To test whether the A549 GBRs can confer glucocorticoid-directed transcriptional responses, we cloned 500-bp DNA fragments encompassing the GBRs into luciferase reporter plasmids. Of the 20 GBRs randomly selected from the 73 GBRs identified in this study, 19 were dex responsive in A549 cells as assessed by reporter analysis (Figure 3A). We define primary GREs (denoted here simply as GREs) as genomic regions that are occupied in vivo by GR and confer glucocorticoid-regulated transcription in transfected reporters. Although the reporter analyses do not prove that the identified elements are functional in their native contexts (see Discussion), they establish that the 500-bp fragments tested harbor sufficient information for GR to regulate



**Figure 3. Native Chromosomal GREs Are Composite Elements**

(A) GBRs confer glucocorticoid responsiveness. A549 cells transfected with luciferase reporter genes linked to 500 bp GBRs were treated with EtOH or 100 nM dex for 5–7 h, harvested, and measured for luciferase activity. Fold dex inductions are plotted for wildtype (white) reporters and mutant (black) reporters with singly (mutGR) or doubly mutated (dmuGR) GR binding sites; standard errors of mean over at least three independent experiments are shown. The 13 mutated GR binding sites were randomly chosen. The GREs that harbor these GR binding sites represent a range of enriched GBRs, ranging from ~6- to 40-fold dex-induced GR occupancy as assessed by ChIP-qPCR (unpublished data).

(B) Identification of enriched motifs within GBRs is shown. Top panel: Sequence logo, generated using WebLogo [58], represents all the compiled sequences resembling GR binding sites identified through computational analysis. Bottom panel shows other enriched motifs (displayed in IUPAC symbols) found in the GRE sequences. Motifs resembling AP-1, HNF4, and C/EBP binding sites were identified using BioProspector whereas motifs similar to ETS and SP1 binding sites were found with MobyDick. The *p*-values of the enriched motifs represent the random probability of these motifs occurring within the GREs.

(C) Conservation analysis of GREs. The identity of the human and mouse sequences was calculated as number of bp matches minus the number of bp deletions or insertions, divided by a 50-bp window. Shown are the average identities for each window across 50 GREs. The background level was calculated as the average of all conservation scores across the 4-kb region. The abscissa shows bp positions with 0 defined as the center of core GR binding sites for GREs.

doi:10.1371/journal.pgen.0030094.g003

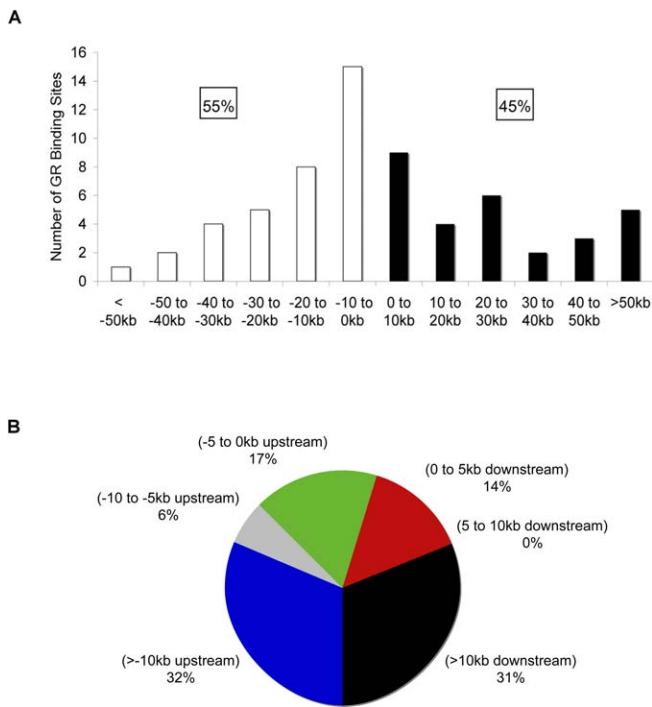
transcription. Thus, we shall refer to the GBRs henceforth as GREs.

### GREs Are Generally Distal and Evenly Distributed between Upstream and Downstream Regions

We determined the positions of the A549 GREs relative to TSSs of their respective target genes (Figure 4A). For this

analysis, the GREs were assigned to the nearest gene responsive to dex in A549 cells. Surprisingly, we found that 45% of the GREs were located downstream of the TSSs, suggesting that GR exhibits transcriptional regulation without a significant preference for regions upstream or downstream of TSSs (Table 1). Figure 4B summarizes the





**Figure 4.** Location and Position of GREs

(A) Locations of GREs relative to the TSS of target genes. The number of GREs resident in 10-kb increments relative to the TSS of the target gene are plotted. White bars and black bars represent GREs upstream and downstream of the TSS, respectively.

(B) Distribution of GREs relative to target gene transcription start site is shown. The chart presents percentage of GREs at various positions upstream and downstream of target genes. Note that only 64 of the 73 GREs detected in A549 cells were included in these analyses; the remaining nine GREs did not associate with a dex-responsive gene in these cells. The GREs were assigned to the nearest gene regulated by GR in A549 cells from the final list of genes that were included or impinged upon by the ChIP-chip arrays. Coordinates of TSSs were obtained from UCSC Genome Browser based on RefSeq. Similar results were obtained when we used TSS coordinates that were experimentally determined (DataBase of Transcriptional Start Sites) through 5' end cloning (unpublished data) [59]. The TSS of the longest transcript was used for genes that have multiple alternative TSSs. Similar results were obtained if the GREs were assigned the closest TSS of the associated dex-responsive gene: 38% of GREs were located downstream from TSS; 58% of GREs were positioned farther than 10 kb from the assigned TSSs. doi:10.1371/journal.pgen.0030094.g004

distribution of promoter proximal (within 5 kb from the TSS) and distal GREs (farther than 10 kb from the TSS). Strikingly, 63% of the GREs were distal, whereas only 31% of them were promoter proximal (Figure 4B). Mammalian response elements are commonly thought to reside upstream and proximal to their cognate promoters; thus, identification of GREs and response elements in general have mainly focused on these regions. Importantly, Figure 4B demonstrates that only a small fraction of the GREs (17%) identified in this study was positioned within these regions. These results indicate that GREs are just as likely to be located downstream of the TSSs and that the majority operate remotely from their target promoters, at least by linear DNA distance.

Our finding concerning GRE distribution is supported by two indirect analyses using nuclease sensitivity and sequence conservation. Sabo et al. found that DNase I hypersensitive sites, indicative of chromatin-bound factors, are broadly distributed with a majority located >10 kb from the nearest

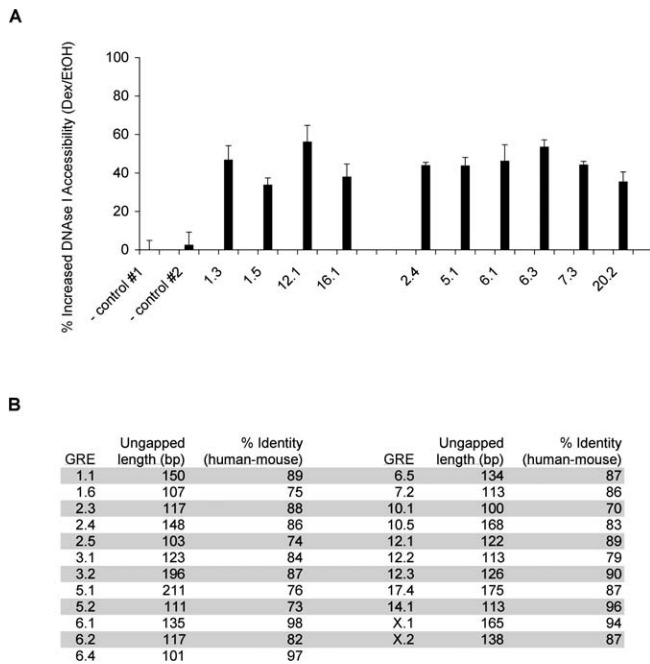
TSS [26]. Furthermore, Dermitzakis et al. showed that conserved nongenic sequences (CNGs), ungapped 100-bp fragments with at least 70% identity between human and mouse that are presumed factor-binding regions, have no significant preference for promoter proximal regions [27,28]. As expected [29], we found that GR occupancy was correlated with DNase I-hypersensitive cleavage at both promoter proximal (1.3, 1.5, 12.1, and 16.1) and distal GREs (2.4, 5.1, 6.1, 6.3, 7.3, and 20.2) (Figure 5A). In addition, by aligning the human GRE sequences with the corresponding regions in the mouse genome, we found that 23 of the GREs correspond to CNGs (Figure 5B). Moreover, GR occupancy and glucocorticoid responsiveness for several of these GREs/CNGs (6.4, 12.1, 5.1, 6.1, 6.2, 10.5, X.1, and X.2) were maintained in mouse cells (see Figure 6A, 6B). Thus, by testing the GREs identified in our study, we were able to provide direct support for the notion that DNase I hypersensitive sites and CNGs serve as regulatory elements [26,28].

### Computation and Conservation Suggest That Native GREs Are Composite Elements

Native GREs, defined as naturally evolved genomic elements that confer glucocorticoid regulation on genes in their chromosomal contexts, are likely to be “composite elements,” made up of binding sites for GR together with multiple nonreceptor regulatory factors [2]. To assess whether we could detect such complex architecture, we used computational approaches (Bioprospector and MobyDick) to survey the 500-bp GRE-containing fragments for sequences related to known regulatory factor binding sites [30–32]. The most prominent motif found, present in 68% of the GRE sequences, was a series of imperfect palindromes similar to known core GR binding sites (Figure 3B). Potentially, GR may interact with the remaining 32% of GREs through other recognition motifs or through tethering to other factors [7]. Mutagenesis of computationally predicted core GR binding sites decreased or completely abolished dex stimulation for each of 13 randomly tested sites, validating this approach for identifying functional core GR binding sequences (Figure 3A). Some GREs, such as 6.2, 7.2, and 7.3, contained multiple GR binding sites; we found that reporters mutated at only one of those sites retained residual dex inducible activity. These experiments imply that most of the core GR binding sites identified in our computational analysis are functional.

In addition, we found that motifs similar to AP-1, ETS, SP1, C/EBP, and HNF4 binding sequences were enriched in the 500-bp GRE fragments (Figure 3B). For example, motifs resembling AP-1 and C/EBP binding sites were identified in the GRE of the *IL8* gene. Importantly, the AP-1 binding site is known to be crucial for regulation of *IL8* by the AP-1 factor [33]; similarly, C/EBP $\alpha$  enhances transcription of a reporter spanning the *IL8* GRE region [34]. Thus, as with GR binding sequences, our computational analysis was capable of discovering functional nonreceptor binding sites. Detection of multiple factor binding sites within the GRE sequences is consistent with the hypothesis that native GREs are typically composite response elements that recruit heterotypic complexes for combinatorial control [2].

To estimate the extent of GRE conservation, we measured sequence identity in human and mouse across 4-kb regions centered on the core GR binding sites (see Figure 3C legend and Materials and Methods) averaged across 50-bp windows; a



**Figure 5.** Some GREs Are Conserved Nongenic Sequences That Exhibit Increased DNase I Accessibility Upon GR Occupancy

(A) Dex induces increased DNase I accessibility. Nuclei from A549 cells treated with EtOH or dex for 1 h were isolated, treated with DNase I, and harvested for DNA. The relative amount of the DNA at the corresponding region were assessed by qPCR and presented as percent cleavage, with standard error of mean averaged among at least three independent experiments. Controls #1 and #2 correspond to regions near *AMOTL2* and *CDH17* genes, respectively, which do not exhibit dex-induced GR occupancy (unpublished data).

(B) Human–mouse sequence conservation within GREs is presented. The mouse sequences aligned with 500-bp human GRE sequences were obtained from UCSC Genome Browser. The lengths of the continuous GRE sequences without gaps between human and mouse are shown, and the percent identity of these regions were calculated as number of matched base pairs divided by length of fragment.  
doi:10.1371/journal.pgen.0030094.g005

similar (albeit higher resolution) pattern was obtained with 15-bp windows (unpublished data). Strikingly, we found that flanking sequences roughly 1 kb surrounding the core GR binding sites were conserved relative to background (Figure 3C). This elevated evolutionary conservation implies that these segments are biologically functional, not only in reporter constructs (Figure 3A), but also in their native chromosomal contexts, further supporting the view that native GREs are composite elements.

### Sequence Conservation of Core GR Binding Sites and GREs

We next sought to examine in detail the extent of sequence conservation of some of the individual core GR binding sequences and GREs that we had identified in our study. We chose a subset of 12 human GREs that are occupied by GR both in another species, mouse, and in another cell type, C3H10T1/2 mesenchymal cells (Figure 6A). Consistent with the correlation between GR occupancy and glucocorticoid responsiveness (Figure 2; Table 1), we confirmed that several of these genes (*Fkbp5*, *Ddit4*, *Gilz*, *MT2A*, and *Sgk*) were indeed dex inducible in the C3H10T1/2 cells (Figure 6B). These 12 GREs resided at very different locations relative to the TSSs of their human target genes (ranging from 0.1 kb to 86 kb) (Table S2); remarkably, however, each locus was approx-

imately maintained in the mouse genome (Table S2). This finding suggests that the positions of individual GREs may be integral to their regulatory functions.

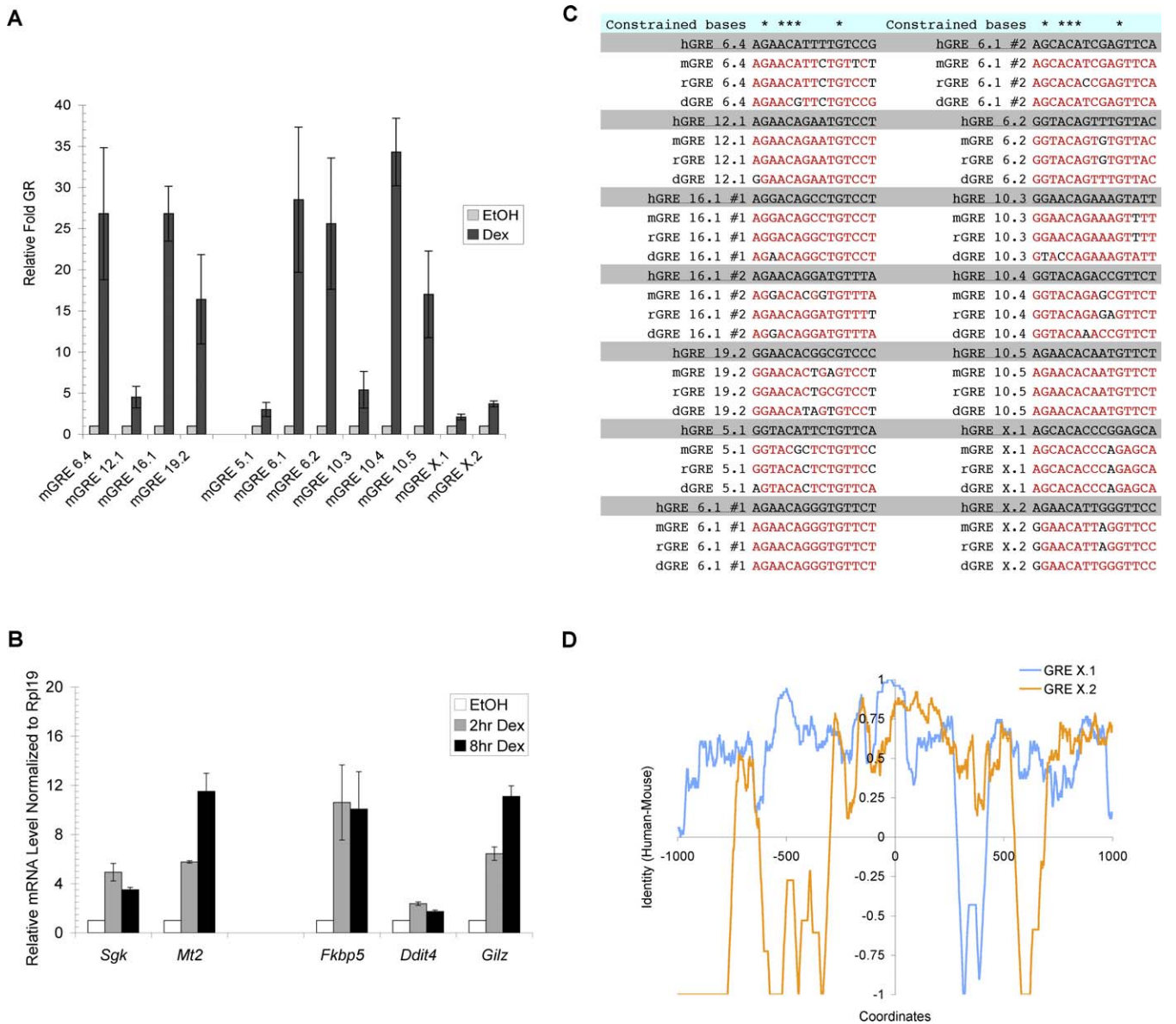
We then examined the extent of conservation of the 15-bp core GR binding sites within the GRE set defined above. As anticipated, the 12 core GR binding sites from the different human GREs differed substantially, with only five invariant positions across the 15-bp sequences (Figure 6C); for example, the binding sites of human GRE 5.1 and human GRE 10.3 match at only seven positions. In striking contrast, we found that the core GR binding site sequences within the individual GREs were highly conserved among human, mouse, dog, and rat (Figure 6C); for example, the core GR binding sequence at GRE 10.5 is identical in all four evolutionarily distant species.

Finally, we compared in human and mouse the patterns of conserved sequences flanking the core GR binding sites, which provide “architectural signatures” of individual GREs. We found that the patterns of sequence conservation differed dramatically among the different GREs (Figure 6D; Figure S3). For example, GRE X.1 contains conserved sequence elements at  $-900$ ,  $-500$ , and  $+600$ bp, whereas GRE X.2 displays no conservation at those positions (Figure 6D). Although the functional significance of the conserved regions has yet to be tested (for example, we have not ruled out incidental overlaps with conserved noncoding expressed regions), the conserved regions are likely to correspond to regulatory or structural motifs. As predicted by these findings, pair-wise calculations of sequence identity of different human GREs (using a 15-bp window centered on the core GR binding sites) demonstrated that sequences flanking the core GR binding sites varied extensively among human GREs (Figure S4). Thus, the overall family of GREs is broadly divergent in sequence and organization, but each individual GRE retains a distinctive signature of conserved sequences, suggesting that each corresponds to a composite GRE that is functionally distinct.

### Discussion

We set out to examine the organization and function of genomic elements responsible for transcriptional regulation by GR. Our study yielded five conclusions: (1) GR occupancy at a GRE is generally a limiting determinant of glucocorticoid response in A549 cells; (2) the core GR binding sequences conform to a consensus that displays substantial GRE-to-GRE variation as anticipated, but the precise binding sequences at individual GREs are highly conserved through evolution; (3) GREs appear to be evenly distributed upstream and downstream of their target genes; (4) most GREs are positioned at locations remote from the TSSs of their target TSSs; and (5) native GREs are commonly composite elements, comprised of multiple factor binding sites, and they are individually conserved in position and architecture yet very different from each other. We shall consider the implications of these conclusions in turn.

We began by surveying more than 1,000 genes, with half of them candidates for steroid regulation, and a specific subset known to be GR-regulated in A549 cells. We found that GR occupancy of A549 GREs correlated strongly (nearly 90%) with genes that are glucocorticoid responsive in A549, suggesting that GR binding is generally a limiting determi-



**Figure 6.** Sequence of GREs as Determinants of Gene-Specific Transcriptional Regulation by GR  
 (A) Binding of GR at mouse orthologs of primary GR target genes from human A549 cells is shown. ChIP experiments were performed to monitor GR binding in EtOH and dex-treated C3H10T1/2 cells at genes shown. Immunoprecipitated DNA samples were analyzed with qPCR and normalized to a region near the mouse *Hsp70* gene. The nomenclature mGRE represents GRE sequences detected in the mouse genome.  
 (B) Genes adjacent to GREs are regulated by GR in C3H10T1/2 cells. Reverse transcribed RNA samples (cDNA) from C3H10T1/2 cells treated with EtOH or 100 nM dex were subjected to qPCR and normalized to mouse *Rpl19* transcripts.  
 (C) Core GR binding sequences are highly conserved. GR binding sequences from human (h), mouse (m), rat (r), and dog (d) are shown. Red sequences represent bases that are identical to that of human. Note that GREs 6.1 and 16.1 each contain two GR binding sites.  
 (D) Comparative sequence conservation across individual GREs. Sequence identities between human and mouse of GRE X.1 and X.2 were obtained using the same calculation as Figure 3C. The coordinates represent bp positions with 0 defined as the center of core GR binding sites.  
 doi:10.1371/journal.pgen.0030094.g006

nant for response in these cells. In a small number of cases, we observed GR occupancy close to genes that were GR-unresponsive in A549 cells, but were steroid regulated in other cells [4] (E. C. Bolton and K. R. Yamamoto, unpublished results). This implies that GR occupancy at these genes likely reflects bona fide response element binding, but that GR binding is not a limiting factor for glucocorticoid regulation of this minority class of genes in A549 cells. Collectively, our data suggest that restriction of GR occupancy in A549 cells may be responsible for much of the cell-specific GR-mediated

regulation in these cells. The mechanisms of occupancy restriction could be positive or negative mechanisms, such as accessory factors that stabilize GR binding, or chromatin packaging that precludes it. Although the strong correlation between GR occupancy and glucocorticoid responsiveness in A549 cells seems likely to hold in other cell types, it is conceivable that responsiveness may be determined differently in other cell types. Thus, it will be interesting to examine cell-specific GR regulation in other cells to complement the observations made in A549 cells. It is intriguing that



one component, GR, within such varied and complex machineries would so strongly predominate as a determinant of transcriptional regulation in A549 cells. It will be interesting to examine regulatory complexes that mediate other types of responses (e.g., heat shock and DNA damage) to assess whether response element occupancy by a single factor in each class is a dominant determinant of responsiveness.

We examined sequence conservation of a set of GREs that are occupied by GR both in human lung epithelial cells and in mouse mesenchymal stem cells. We found that the 15-bp core GR binding sequences varied greatly among the different GREs (Figure 3B), whereas the sequences of the individual binding sites were nearly fully conserved across four mammalian species (Figure 6C). Crystallographic studies demonstrate that GR makes specific contacts with only four bases of the 15-bp core binding sequence [35], yet every position, including the “spacer” between the hexameric half sites, appears to be equivalently conserved. This indicates that the binding sequences serve functions in addition to merely localizing GR to specific genomic loci and instead may carry a regulatory code that affects GR function. Leung et al. reported similarly strong evolutionary conservation of individual  $\kappa$ B binding sequences [36]. Indeed, Luecke and Yamamoto showed that GR directs distinct regulatory effects when tethered to NF $\kappa$ B at two  $\kappa$ B response elements that differ by only one base pair [7]. Thus, one interpretation of our data findings is that factor binding sites may serve as allosteric effectors [19] in which individual binding sequences convey subtle conformational differences to specify distinct factor functions. Conceivably, this hypothesis might also explain why GR predominates as a limiting determinant of responsiveness, because factors that read allosteric regulatory codes might specify the rules for assembly of GRE-specific and thus gene-specific regulatory complexes.

To characterize the architecture of GREs, we took several approaches. In unbiased computational analyses, we identified enriched sequence motifs within 500-bp segments encompassing core GR binding sites. Sequence motifs resembling binding sites for GR, AP-1, ETS, SP1, C/EBP, and HNF4 were overrepresented relative to a background of unbound GR regions, consistent with the notion that native GREs are composite elements. For most of these GREs, the role of these factors in GR transcriptional regulation remains to be tested, but it is notable that ETS-1, SP1, and HNF4 have been shown at other genes to augment glucocorticoid responses [37–39]. Moreover, Phuc Le et al. [40] described motifs resembling AP1 and C/EBP binding sites within certain mouse GREs and showed that nearly half of the GREs predicted to encompass C/EBP binding sites did indeed bind C/EBP $\beta$  [40]. These findings further the view that our computational analysis can infer factors that potentially interact with GR at GREs. Using a similar approach, Carroll et al. [9] and Laganier et al. [41] have interrogated estrogen response elements and identified FOXA1 as a factor playing an important role for both estrogen receptor binding and transcriptional activity. Thus, we anticipate that the factors that occupy the GR composite elements may interact physically, functionally, or both, thereby affecting binding as well as regulatory activity. Indeed, an averaged comparison of human and mouse sequences flanking core GR binding sites revealed that a region of approximately 1 kb was conserved above the background level (Figure 3C), suggesting that native

composite GREs are extensive and typically may contain numerous factor binding sites. Interestingly, individual GREs displayed distinctive patterns of sequence conservation extending from the core GR binding sites (Figure 6D; Figure S3). These GRE signatures likely reflect conservation of various sequence motifs at different positions within each element, producing GRE-specific (and therefore gene-specific) architecture that likely creates distinct regulatory effects.

To investigate the distribution of regulatory elements relative to their target genes, we monitored GR occupancy across 100 kb regions centered on the TSSs of glucocorticoid responsive genes. We found that GREs were evenly distributed upstream and downstream of their target genes with the majority located  $>10$  kb from their target promoters; other metazoan regulatory factors, such as estrogen receptor (ER) and STAT1, have similarly been reported to act from sites remote from their target genes [9,42–45]. In contrast to these factors, E2F1 was shown to mainly bind promoter proximal regions [42]; others have used computational approaches to infer factor binding sites close to promoters, but these have not been experimentally confirmed [46]. In parallel with our findings, Carroll et al. reported that only 4% of estrogen receptor ER binding regions was mapped within  $-800$  bp to  $+200$  bp from TSS of known genes from RefSeq [43]. Our data demonstrated that 9% of GBRs were positioned at this location. These studies together imply that steroid receptors, which include estrogen receptor and GR, in general regulate transcription from remote locations. Interestingly, we found that the positions of individual GREs were generally conserved across species (Table S2), implying that GRE position may be functionally important for target gene regulation. In any case, our findings differ dramatically from those in prokaryotes and fungi, where transcriptional regulatory elements are promoter proximal. It has been suggested that these two broad classes of regulatory mechanisms, so-called long range and short range, are mechanistically and evolutionarily related, and that long range control might facilitate regulatory evolution [11]. As predicted by that model, distal elements, far from target genes as measured by linear DNA distance, may operate in close proximity with their target promoters in 3-D space. For example, Carroll et al. detected an interaction between the *NR1P-1* promoter and its distal estrogen response element [9]. It will be interesting to determine whether response element location (i.e., promoter proximal versus distal) is somehow related to mechanism or to physiological network.

Remote response element locations can complicate assignment of cognate target genes. An extreme example is olfactory receptor gene expression, which is governed by a regulatory element that can operate on target genes located on different chromosomes [47]. In this study, we assigned the GREs to the nearest RefSeq gene responsive to dex in A549 cells. In other contexts, these GREs may be nonfunctional or may operate on genes other than those assigned in A549 cells (Table 1). Clearly, unequivocal assignment of a GRE to a given target gene will require genetic manipulations not readily accessible in mammalian cells at present. It is encouraging, however, that GR occupancy of GREs correlated strongly with glucocorticoid responsiveness of adjacent genes, supporting the view that these are bona fide direct GR targets (Figure 2; Table 1). In fact, when these genes were subjected to Gene Ontology analysis, we found that they were enriched in cell

growth and immune responses (unpublished data), two biological processes regulated by GR in A549 cells [48,49]. We found GR occupancy at genes up- and down-regulated in response to dex, consistent with GR serving either as activator or repressor in different contexts. At present, we cannot assess the significance of the finding that GR was detected at GREs adjacent to activated genes versus repressed genes at a 6:1 ratio in A549 cells; whether this difference reflects differences in GRE occupancy, epitope accessibility, cross-linking efficiency, or other variables has not been determined.

Genomic response elements orchestrate transcriptional networks to mediate cellular processes for single- and multicellular organisms. The present study advanced our understanding of the organization, evolution, and function of GREs and at the same time raised a series of interesting questions. Among the more intriguing: How is GR occupancy restricted to a small subset of potential GREs in a given cell context? What is driving the strong conservation of virtually every base pair within the core GR binding sequence at individual GREs? Addressing these and other questions raised in our study will contribute additional new insights about gene regulation by GR and by other regulatory factors.

## Materials and Methods

**Cell culture, plasmids, and reporter analysis.** A549 and C3H10T1/2 cells were grown in DMEM supplemented with 5% or 10% FBS, respectively, in a 5% carbon dioxide atmosphere. Before hormone treatment, media was replenished with DMEM containing charcoal stripped FBS, which depletes endogenous steroids. Plasmid PGL4.10 E4TATA (generously provided by Yuriy Shostak) was created by insertion of the E4TATA minimal promoter into pGL4.10 vector (Promega, <http://www.promega.com>). The 20 reporters tested (Figure 3A) represent randomly chosen GRE fragments. The QuikChange kit (Stratagene, <http://www.stratagene.com>) was used for reporter mutagenesis. The 13 core GR binding sites that were mutated in the reporters (Figure 3A) were also randomly chosen based on success of mutagenesis. GBR-containing DNA fragments (500 bp) were amplified by PCR and subcloned into pGL4.10 E4TATA using KpnI and XhoI sites (see Table S3 for primer sequences). A549 cells were grown in a 48-well plate and cotransfected with 19 ng of the reporter constructs, 10 ng pRL Luc (Promega), and 38 ng pCDNA3 hGR (human GR expression vector) using Lipofectamine 2000 (Invitrogen, <http://www.invitrogen.com>). After overnight transfection, cells were treated with hormone, harvested, and luciferase activity was measured as described for the dual luciferase reporter system (Promega) using a Tecan Ultra Evolution plate reader (Tecan, <http://www.tecan.com>).

**ChIP and array analysis.** ChIP assays were performed as described [7] with the following modifications. The chromatin samples were extracted once with phenol-chloroform and purified using a Qiaquick column as recommended by the manufacturer (Qiagen, <http://www1.qiagen.com>). The ligation-mediated PCR (LMPCR) process was adapted from Oberley et al. [50]. We used 3.5–20 ng of amplicon for real-time qPCR analysis, and data were normalized to *Hsp70* (see Table S4 for primer sequence). Human and mouse DNA sequences were retrieved from University of California Santa Cruz (UCSC) Genome Browser (<http://genome.ucsc.edu>) NCBI Build 35, and qPCR primers were designed using Primer3 [51]. For the array, ~50 kb upstream and downstream regions were tiled with isothermal 50 mer oligos (spaced on average of every 54 bp apart) relative to the TSSs of the investigated target genes. Where 100-kb regions overlapped, the surrounding genomic region was tiled further bidirectionally. ChIP samples from (final concentration, 0.01% ethanol) or dex-treated A549 cells were labeled with Cy3 or Cy5, hybridized onto the arrays, and relative signal intensities were measured by NimbleGen (<http://www.nimblegen.com>). SignalMap was utilized to find peak enrichments with both window threshold detection (500-bp peak window size, 25% of Peak Threshold) and second derivative peak detection (500-bp peak window size, 20 bp smooth step, 25% peak threshold) (NimbleGen).

**RNA isolation, reverse transcription, and real-time qPCR.** The RNA isolation, reverse transcription, and qPCR steps were performed as previously described [4]. Primers for cDNA amplification are displayed in Table S4.

**DNase I accessibility assay.** The experiments were adapted from previous described protocol [52] with the following modifications. Briefly, nuclei from A549 cells treated with vehicle or dex for 1 h were treated with 6.25–200 units/ml of DNase I (Qiagen) for 5 min at room temperature. The reaction was stopped and treated with Proteinase K for 1 h at 65 °C. The DNA samples were extracted once with 1:1 phenol-chloroform and further purified using MiniPrep columns (Qiagen). The samples were subjected to qPCR analysis to determine the relative amount of cleaved product (see Table S4 for primer sequences), which was converted to percent DNase I cleavage.

**Computational analysis.** For computational analysis of enriched motifs, all repeat-masked DNA sequences were downloaded from the UCSC genome browser (NCBI Human Build 35). BioProspector analysis was initially performed using nucleotide widths (w) 14 and 16 on GREs to identify GR binding sites and the top motifs were masked to identify other motifs [32]. For MobyDick analysis, both the human and the human/mouse aligned sequences were used as inputs to identify enriched motifs [30,31]. Similar motifs were clustered using CAST [53–55]. All *p*-values for enrichment were Bonferroni corrected to identify putative factor binding sites [55]. The top BioProspector w14 position weight matrix (PWM) was used to score GREs for putative GR binding sites with a false positive rate of less than 10%. This upper bound was calculated from randomly sampling unbound GR regions (Figure S2).

We built position weight matrices (PWMs) of those motifs with *p*-values less than 0.05, which were used to measure similarity to known binding sites in TRANSFAC [56]. We measured the distance between the PWMs and those representing binding sites for known regulatory factors using relative entropy (Kullback-Liebler divergence) with a cutoff of less than 6.0 to associate motifs with putative regulatory factor binding sites. The known binding site matrices were obtained from TRANSFAC professional version 9.3.

The human–mouse conservation score was calculated as described [9] using a 50-mer window for 50 sequences containing a putative GR binding site based on our computational and experimental analysis (Figure S2 and Figure 3A). The conservation score was calculated as number of bp matches minus the number of bp deletions or insertions divided by the bp window size. We centered each alignment based on the highest scoring putative GR binding site in human and expanded equally on each side of the binding site to a total length of 4 kb. The background level was calculated by taking the average of all conservation scores across the 4-kb region. The human(hg)/mouse(mm) genome alignments were downloaded from Vista (<http://pipeline.lbl.gov/cgi-bin/gateway2>).

## Supporting Information

**Figure S1.** Identification of Novel GR Targets in A549 Cells by ChIP-chip

ChIP-chip analysis revealed GR occupancy at genes not previously recognized as GR targets in A549 cells. Analysis using qPCR demonstrated that 28 of these genes were dex responsive after four or eight hours of treatment. Induction or repression is represented as fold changes in positive or negative values, respectively, averaged over at least three independent experiments. The data shown represent largest fold change obtained from four or eight hours of hormone treatment.

Found at doi:10.1371/journal.pgen.0030094.sg001 (109 KB PDF).

**Figure S2.** Number of GREs with Putative GR Binding Sites

All 73 GRE sequences were scored for putative GR binding sites using a predicted position weight matrix representative of the GR binding site. Percent of sequences predicted to contain a GR binding site with varying score cutoffs is plotted as red squares. The false positive rate (blue triangles) was calculated by randomly sampling unbound sequences at varying score cutoffs.

Found at doi:10.1371/journal.pgen.0030094.sg002 (126 KB PDF).

**Figure S3.** Sequence Conservation “Signatures” Are Distinct for Each GRE

Identity scores were determined for human–mouse aligned sequences and are plotted as in Figure 6D; for clarity, data are presented as pairwise comparisons. In (A–E) comparisons of conservation of the specified GREs are represented.

Found at doi:10.1371/journal.pgen.0030094.sg003 (498 KB PDF).

#### Figure S4. GREs Vary in Sequences

(A) A sequence comparison of human GRE 10.5 with human GRE 6.1 is shown.

(B) A sequence comparison of human GRE 6.4 and X.2 is shown. The sequences were pair-wise aligned using ClustalW [57] and similarities were calculated as in Figure 3C using a 15-bp window. Coordinate 0 represents the center of the core GR binding sites. The red line represents the background level, which was calculated by taking the average of all identity scores.

Found at doi:10.1371/journal.pgen.0030094.sg004 (142 KB PDF).

#### Table S1. Dex Responsiveness of Steroid Targets from Other Cells in A549 Cells

Quantification of relative mRNA levels by qPCR of a subset of the 587 genes (denoted as ChIP-chip Spanned) included in the ChIP-chip arrays (Figure 2) showed that they were not dex responsive (less than 1.6-fold change) in A549 cells after 4 or 8 h of treatment; U2OS source genes are responsive to dex in U2OS but not in A549 cells; Other cells source genes are potentially steroid responsive in other cells but not in A549 cells. Analysis with qPCR confirms that a majority of these genes were indeed not responsive to dex in A549 cells after 4 or 8 h of treatment. Values shown are fold changes comparing dex and ethanol treatment averaged over at least two independent experiments. Bold letter genes are those that are dex responsive in A549 cells.

Found at doi:10.1371/journal.pgen.0030094.st001 (23 KB XLS).

#### Table S2. Distances of GREs Relative to Adjacent Gene Are Conserved in the Mouse Genome

The distances were calculated based on coordinates of the mouse aligned GREs (mGREs) and TSSs of the GR-regulated mouse homolog genes obtained from UCSC Genome Browser. The TSS of the longest transcript was used for this calculation when a gene has multiple variants. Bold letters represent the distance of the GREs relative to the adjacent gene in mouse.

#### References

1. Jobe AH, Ikegami M (2000) Lung development and function in preterm infants in the surfactant treatment era. *Annu Rev Physiol* 62: 825–846.
2. Yamamoto KR, Darimont BD, Wagner RL, Iniguez-Lluhi JA (1998) Building transcriptional regulatory complexes: Signals and surfaces. *Cold Spring Harb Symp Quant Biol* 63: 587–598.
3. Wang JC, Derynck MK, Nonaka DF, Khodabakhsh DB, Haqq C, et al. (2004) Chromatin immunoprecipitation (ChIP) scanning identifies primary glucocorticoid receptor target genes. *Proc Natl Acad Sci U S A* 101: 15603–15608.
4. Rogatsky I, Wang JC, Derynck MK, Nonaka DF, Khodabakhsh DB, et al. (2003) Target-specific utilization of transcriptional regulatory surfaces by the glucocorticoid receptor. *Proc Natl Acad Sci U S A* 100: 13845–13850.
5. Chandler VL, Maler BA, Yamamoto KR (1983) DNA sequences bound specifically by glucocorticoid receptor in vitro render a heterologous promoter hormone responsive in vivo. *Cell* 33: 489–499.
6. Bulger M, Groudine M (1999) Looping versus linking: Toward a model for long-distance gene activation. *Genes Dev* 13: 2465–2477.
7. Luecke HF, Yamamoto KR (2005) The glucocorticoid receptor blocks P-TEFb recruitment by NFkappaB to effect promoter-specific transcriptional repression. *Genes Dev* 19: 1116–1127.
8. Jantzen HM, Strahle U, Gloss B, Stewart F, Schmid W, et al. (1987) Cooperativity of glucocorticoid response elements located far upstream of the tyrosine aminotransferase gene. *Cell* 49: 29–38.
9. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, et al. (2005) Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 122: 33–43.
10. Magee JA, Chang LW, Stormo GD, Milbrandt J (2006) Direct, androgen receptor-mediated regulation of the FKBP5 gene via a distal enhancer element. *Endocrinology* 147: 590–598.
11. Yamamoto KR (1989) A conceptual view of transcriptional regulation. *American Zoologist* 29: 537–547.
12. Yamamoto KR (1995) Multilayered control of intracellular receptor function. *Harvey Lect* 91: 1–19.
13. Beato M, Chalepakis G, Schauer M, Slater EP (1989) DNA regulatory elements for steroid hormones. *J Steroid Biochem* 32: 737–747.
14. Nordeen SK, Suh BJ, Kuhnel B, Hutchison CD (1990) Structural determinants of a glucocorticoid receptor recognition element. *Mol Endocrinol* 4: 1866–1873.
15. Strahle U, Klock G, Schutz G (1987) 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* 84: 7871–7875.

Found at doi:10.1371/journal.pgen.0030094.st002 (21 KB XLS).

#### Table S3. Primers Used for Cloning and Mutating GRE Reporters

Capitalized letters represent the restriction digestion sites used for cloning the constructs into pGL4.10 E4TATA.

Found at doi:10.1371/journal.pgen.0030094.st003 (25 KB XLS).

#### Table S4. Primers Used for qPCR Analysis

FO primer and RE primer represent the forward and reverse primer, respectively, for the corresponding amplified genomic regions or cDNA sequences of the indicated genes.

Found at doi:10.1371/journal.pgen.0030094.st004 (39 KB XLS).

## Acknowledgments

We thank all members of the Yamamoto laboratory and V. Chubukov, K. Kechris, and B. Tuch of the Li laboratory for extensive discussion involving this project. We give special thanks to K. M. Shokat, G. J. Narlikar, M. Pufall, S. Taubert, and S. H. Meijnsing for helpful critiques of the manuscript.

**Author contributions.** AYLs, CC, ECB, HL, and KRY conceived and designed the experiments. AYLs, CC, and ECB performed the experiments. This manuscript was written by AYLs, CC, and KRY with guidance from ECB and HL.

**Funding.** AYLs received support from the California Breast Cancer Research Program of the University of California, Grant Number 10GB-0104; CC is recipient of a predoctoral fellowship from the National Science Foundation; ECB was recipient of a National Research Service Award postdoctoral fellowship F32DK65402 from the National Institutes of Health (NIH). Research support was from NIH (CA20535 to KRY and GM070808 to HL) and a Packard Fellowship in Science and Engineering to HL.

**Competing interests.** The authors have declared that no competing interests exist.

16. La Baer J, Yamamoto KR (1994) Analysis of the DNA-binding affinity, sequence specificity and context dependence of the glucocorticoid receptor zinc finger region. *J Mol Biol* 239: 664–688.
17. Rogatsky I, Luecke HF, Leitman DC, Yamamoto KR (2002) Alternate surfaces of transcriptional coregulator GRIP1 function in different glucocorticoid receptor activation and repression contexts. *Proc Natl Acad Sci U S A* 99: 16701–16706.
18. Lefstin JA, Thomas JR, Yamamoto KR (1994) Influence of a steroid receptor DNA-binding domain on transcriptional regulatory functions. *Genes Dev* 8: 2842–2856.
19. Lefstin JA, Yamamoto KR (1998) Allosteric effects of DNA on transcriptional regulators. *Nature* 392: 885–888.
20. Geserick C, Meyer HA, Haendler B (2005) The role of DNA response elements as allosteric modulators of steroid receptor function. *Mol Cell Endocrinol* 236: 1–7.
21. Scully KM, Jacobson EM, Jepsen K, Lunyak V, Viadiu H, et al. (2000) Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science* 290: 1127–1131.
22. Roche PJ, Hoare SA, Parker MG (1992) A consensus DNA-binding site for the androgen receptor. *Mol Endocrinol* 6: 2229–2235.
23. DePrimo SE, Diehn M, Nelson JB, Reiter RE, Matese J, et al. (2002) Transcriptional programs activated by exposure of human prostate cancer cells to androgen. *Genome Biol* 3: RESEARCH0032.
24. Nelson PS, Clegg N, Arnold H, Ferguson C, Bonham M, et al. (2002) The program of androgen-responsive genes in neoplastic prostate epithelium. *Proc Natl Acad Sci U S A* 99: 11890–11895.
25. Sayegh R, Auerbach SD, Li X, Loftus RW, Husted RF, et al. (1999) Glucocorticoid induction of epithelial sodium channel expression in lung and renal epithelia occurs via trans-activation of a hormone response element in the 5'-flanking region of the human epithelial sodium channel alpha subunit gene. *J Biol Chem* 274: 12431–12437.
26. Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, et al. (2006) Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods* 3: 511–518.
27. Dermitzakis ET, Reymond A, Lyle R, Scamuffa N, Ucla C, et al. (2002) Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature* 420: 578–582.
28. Dermitzakis ET, Kirkness E, Schwarz S, Birney E, Reymond A, et al. (2004) Comparison of human Chromosome 21 conserved nongenic sequences (CNGs) with the mouse and dog genomes shows that their selective constraint is independent of their genetic environment. *Genome Res* 14: 852–859.

29. Zaret KS, Yamamoto KR (1984) Reversible and persistent changes in chromatin structure accompany activation of a glucocorticoid-dependent enhancer element. *Cell* 38: 29–38.
30. Bussemaker HJ, Li H, Siggia ED (2000) Building a dictionary for genomes: Identification of presumptive regulatory sites by statistical analysis. *Proc Natl Acad Sci U S A* 97: 10096–10100.
31. Bussemaker HJ, Li H, Siggia ED (2000) Regulatory element detection using a probabilistic segmentation model. *Proc Int Conf Intell Syst Mol Biol* 8: 67–74.
32. Liu X, Brutlag DL, Liu JS (2001) BioProspector: Discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput*: 127–138.
33. Hoffmann E, Dittrich-Breiholz O, Holtmann H, Kracht M (2002) Multiple control of interleukin-8 gene expression. *J Leukoc Biol* 72: 847–855.
34. Freund A, Jolivel V, Durand S, Kersual N, Chalbos D, et al. (2004) Mechanisms underlying differential expression of interleukin-8 in breast cancer cells. *Oncogene* 23: 6105–6114.
35. Luisi BF, Xu WX, Otwinowski Z, Freedman LP, Yamamoto KR, et al. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352: 497–505.
36. Leung TH, Hoffmann A, Baltimore D (2004) One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers. *Cell* 118: 453–464.
37. Geng CD, Vedeckis WV (2004) Steroid-responsive sequences in the human glucocorticoid receptor gene 1A promoter. *Mol Endocrinol* 18: 912–924.
38. Marinovic AC, Zheng B, Mitch WE, Price SR (2002) Ubiquitin (UbC) expression in muscle cells is increased by glucocorticoids through a mechanism involving Sp1 and MEK1. *J Biol Chem* 277: 16673–16681.
39. Wang JC, Stromstedt PE, Sugiyama T, Granner DK (1999) The phosphoenolpyruvate carboxykinase gene glucocorticoid response unit: Identification of the functional domains of accessory factors HNF3 beta (hepatic nuclear factor-3 beta) and HNF4 and the necessity of proper alignment of their cognate binding sites. *Mol Endocrinol* 13: 604–618.
40. Phuc Le P, Friedman JR, Schug J, Brestelli JE, Parker JB, et al. (2005) Glucocorticoid receptor-dependent gene regulatory networks. *PLoS Genet* 1: e16. doi:10.1371/journal.pgen.0010016.
41. Laganieri J, Deblois G, Lefebvre C, Bataille AR, Robert F, et al. (2005) From the cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proc Natl Acad Sci U S A* 102: 11651–11656.
42. Bieda M, Xu X, Singer MA, Green R, Farnham PJ (2006) Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. *Genome Res* 16: 595–605.
43. Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, et al. (2006) Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* 38: 1289–1297.
44. Hartman SE, Bertone P, Nath AK, Royce TE, Gerstein M, et al. (2005) Global changes in STAT target selection and transcription regulation upon interferon treatments. *Genes Dev* 19: 2953–2968.
45. Martone R, Euskirchen G, Bertone P, Hartman S, Royce TE, et al. (2003) Distribution of NF-kappaB-binding sites across human Chromosome 22. *Proc Natl Acad Sci U S A* 100: 12247–12252.
46. Bajic VB, Tan SL, Christoffels A, Schonbach C, Lipovich L, et al. (2006) Mice and men: Their promoter properties. *PLoS Genet* 2: e54. doi:10.1371/journal.pgen.0020054
47. Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, et al. (2006) Interchromosomal interactions and olfactory receptor choice. *Cell* 126: 403–413.
48. Wang JC, Shah N, Pantoja C, Meijsing SH, Ho JD, et al. (2006) Novel arylpyrazole compounds selectively modulate glucocorticoid receptor regulatory activity. *Genes Dev* 20: 689–699.
49. Rosen J, Miner JN (2005) The search for safer glucocorticoid receptor ligands. *Endocr Rev* 26: 452–464.
50. Oberley MJ, Tsao J, Yau P, Farnham PJ (2004) High-throughput screening of chromatin immunoprecipitates using CpG-island microarrays. *Methods Enzymol* 376: 315–334.
51. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132: 365–386.
52. Rao S, Procko E, Shannon MF (2001) Chromatin remodeling, measured by a novel real-time polymerase chain reaction assay, across the proximal promoter region of the IL-2 gene. *J Immunol* 167: 4494–4503.
53. Ben-Dor A, Shamir R, Yakhini Z (1999) Clustering gene expression patterns. *J Comput Biol* 6: 281–297.
54. Li H, Rhodius V, Gross C, Siggia ED (2002) Identification of the binding sites of regulatory proteins in bacterial genomes. *Proc Natl Acad Sci U S A* 99: 11772–11777.
55. Patil CK, Li H, Walter P (2004) Gcn4p and novel upstream activating sequences regulate targets of the unfolded protein response. *PLoS Biol* 2: e246. doi:10.1371/journal.pbio.0020246
56. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, et al. (2003) TRANSFAC: Transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 31: 374–378.
57. Higgins DG, Thompson JD, Gibson TJ (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol* 266: 383–402.
58. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Res* 14: 1188–1190.
59. Suzuki Y, Yamashita R, Sugano S, Nakai K (2004) DBTSS, DataBase of transcriptional start sites: Progress report 2004. *Nucleic Acids Res* 32: D78–D81.