nature genetics

Chromatin accessibility pre-determines glucocorticoid receptor binding patterns

Sam John[1], Peter J Sabo[2], Robert E Thurman[2], Myong-Hee Sung[1], Simon C Biddie[1], Thomas A Johnson[1], Gordon L Hager[1] & John A Stamatoyannopoulos[2,3]

Development, differentiation and response to environmental stimuli are characterized by sequential changes in cellular state initiated by the *de novo* binding of regulated transcriptional factors to their cognate genomic sites[1–3]. The mechanism whereby a given regulatory factor selects a limited number of *in vivo* targets from a myriad of potential genomic binding sites is undetermined. Here we show that up to 95% of *de novo* genomic binding by the glucocorticoid receptor[4], a paradigmatic ligand-activated transcription factor, is targeted to preexisting foci of accessible chromatin. Factor binding invariably potentiates chromatin accessibility. Cell-selective glucocorticoid receptor occupancy patterns appear to be comprehensively predetermined by cell-specific differences in baseline chromatin accessibility patterns, with secondary contributions from local sequence features. The results define a framework for understanding regulatory factor–genome interactions and provide a molecular basis for the tissue selectivity of steroid pharmaceuticals and other agents that intersect the living genome.

How regulatory factors interact with the chromatin landscape to effect gene regulation is one of the leading questions in genome biology. Chromatin structure is altered at *cis*-regulatory regions, resulting in hypersensitivity of the underlying DNA to nuclease attack *in vivo*[5–7]. However, how this preexisting landscape influences *de novo* binding site selection has not been determined.

Here we address this using a well-controlled model system: the endogenous glucocorticoid hormone response pathway found in most mammalian cells. The cellular actions of glucocorticoids are mediated through the glucocorticoid receptor[4], at where it selectively engages up to several thousand cognate genomic binding sites[8–10]. Glucocorticoid receptor signaling thus represents an ideal system for both qualitative and quantitative analysis of *de novo* transcription factor–genome interactions in a highly controlled fashion.

We first sought to determine the global relationship between the preexisting chromatin accessibility state of untreated cells and the pattern of glucocorticoid receptor binding following hormone induction. The glucocorticoid receptor is widely believed to function as a 'pioneer protein' that is capable of autonomous binding to genomic DNA target sites, resulting in local chromatin remodeling[11,12]. However, this concept is based largely on qualitative results from a limited set of loci[13].

To gain a genome-wide perspective, we used digital DNase I analysis[14,15] and ChIP-seq[10,16,17] to map chromatin accessibility and glucocorticoid receptor occupancy at high resolution both before and after steroid hormone (dexamethasone (Dex)) treatment in a well-studied model cell type (mouse 3134 mammary adenocarcinoma cells). Digital DNase I profiling enables quantitative delineation of chromatin accessibility, including both classical DNase I hypersensitive sites (DHSs) as well as regions of general chromatin accessibility marked by DNase I sensitivity[18] (**Supplementary Figs. 1** and **2**).

Genome-wide DNase I sensitivity and glucocorticoid receptor occupancy profiles were highly reproducible (**Supplementary Fig. 3**) and revealed a striking correspondence between the locations of glucocorticoid receptor occupancy after dexamethasone treatment and the preexisting pattern of chromatin accessibility in untreated cells (**Fig. 1** and **Supplementary Fig. 3a–c**). To quantify this phenomenon, we delineated genomic regions with high chromatin accessibility over background and identified 97,717 strongly DNase I–sensitive regions encompassing 2.1% (56.7 Mb) of the genome in untreated cells (**Supplementary Tables 1** and **2** and **Supplementary Note**), within which we localized 87,490 DHSs (0.4% of the genome at a false discovery rate (FDR) of 1%; **Supplementary Tables 1** and **3**).

Analysis of glucocorticoid receptor ChIP-seq data from hormone-treated cells revealed 8,236 sites of glucocorticoid receptor occupancy (**Supplementary Table 4**). Performing *de novo* motif discovery on the top 500 glucocorticoid receptor occupancy sites recovered a 15-bp motif that closely matched the consensus glucocorticoid receptor binding element (GRBE; **Fig. 2a**)[19,20]. More than 80% of glucocorticoid receptor occupancy sites contained some form of this GRBE consensus sequence ($P < 10^{-3}$), with 50% of these sites containing higher stringency matches ($P < 10^{-4}$).

[1]Laboratory of Receptor Biology and Gene Expression, Center for Cancer Research, National Cancer Institute, National Institutes of Health (NIH), Bethesda, Maryland, USA. [2]Department of Genome Sciences, University of Washington, Seattle, Washington, USA. [3]Department of Medicine, Division of Oncology, University of Washington, Seattle Cancer Care Alliance, Seattle, Washington, USA. Correspondence should be addressed to J.A.S. (jstam@uw.edu) or G.L.H. (hagerg@exchange.nih.gov).

Received 30 August 2010; accepted 29 December 2010; published online 23 January 2011; doi:10.1038/ng.759

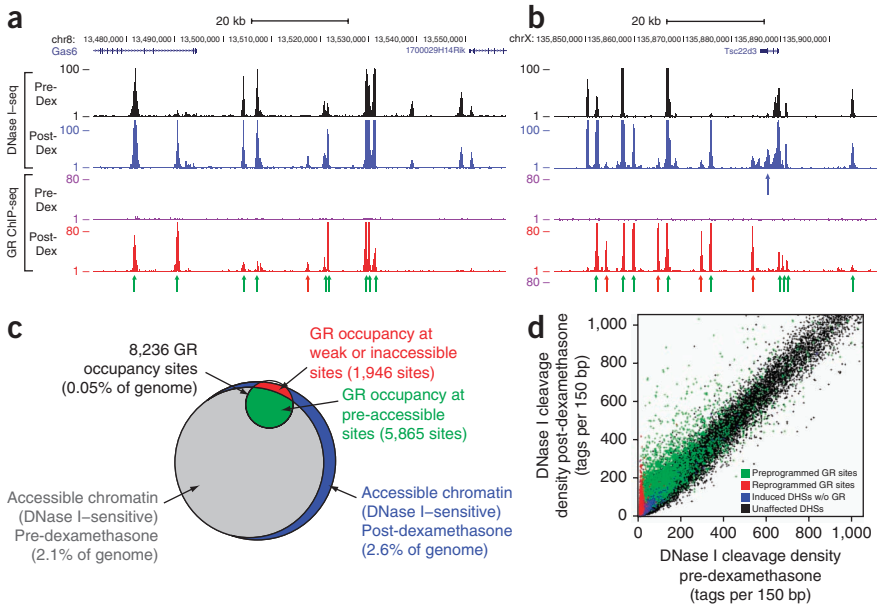

Figure 1 Dominant effect of chromatin accessibility on glucocorticoid receptor occupancy patterns. (**a**,**b**) Examples of DNase I sensitivity and glucocorticoid receptor occupancy patterns in relation to dexamethasone exposure (see **Supplementary Fig. 3a–c** for additional examples). Each data track shows tag density (150-bp sliding window) from either DNase I–seq or glucocorticoid receptor ChIP-seq, normalized to allow comparison across different samples (Online Methods). Green arrows mark sites of post-hormone glucocorticoid receptor occupancy in pre-existing DNase I–sensitive chromatin ('preprogrammed' sites). Red arrows mark glucocorticoid receptor occupancy sites in pre-hormone inaccessible chromatin that result in post-hormone chromatin remodeling ('reprogrammed' sites). Blue arrows mark hormone-induced DHSs not directly associated with glucocorticoid receptor occupancy (**Supplementary Fig. 4c**). (**c**) Venn diagram summarizing global glucocorticoid receptor occupancy compared to the chromatin accessibility landscape (~25 million read depth) in mammary cells (for legibility, the glucocorticoid receptor circle is shown at ×5 scale). Most glucocorticoid receptor occupancy occurs within pre-hormone accessible chromatin. A small fraction of generally weak glucocorticoid receptor peaks (5.2% of total peaks) are not associated with reprogrammed or preprogrammed chromatin. (**d**) DNase I sensitivity (tag density) pre-hormone (horizontal axis) compared to post-hormone (vertical axis) treatment. Colors match those used in **c**. Black, pre-hormone accessible regions with no post-hormone glucocorticoid receptor occupancy; blue, DNase I–sensitive regions induced post-hormone without glucocorticoid receptor occupancy (**Supplementary Fig. 4c**); green; pre-hormone DNase I–sensitive regions occupied by glucocorticoid receptor post-hormone ('preprogrammed' sites); red, pre-hormone inaccessible chromatin remodeled by glucocorticoid receptor occupancy ('reprogrammed' sites), resulting in marked alteration in DNase I sensitivity (**Supplementary Fig. 4a,b**). GR, glucocorticoid receptor.

The great majority of glucocorticoid receptor occupancy sites (71%, 5,865 sites) were targeted to the 2.1% of the genome defined by preexisting (that is, pre-hormone or 'baseline') strongly DNase I–sensitive regions ($P < 10^{-300}$). An additional ~9% of binding localized to weakly DNase I–sensitive regions, with 80% of glucocorticoid receptor binding occurring within 4.9% of the genome (**Supplementary Fig. 3d**). However, this estimate represents a lower limit. For example, increasing the sequencing depth of the pre-hormone DNase I–seq sample approximately eightfold increased the proportion of glucocorticoid receptor sites falling within pre-hormone accessible chromatin from 71% to 88.3% ($P < 10^{-300}$; **Supplementary Note** and **Supplementary Fig. 3d**). In hormone-treated cells, 95% of glucocorticoid receptor occupancy sites (and >99% after deep sequencing) localized to accessible chromatin ($P < 10^{-300}$). Additionally, we observed DHSs unique to hormone-treated cells that were not directly associated with glucocorticoid receptor binding (**Fig. 1a**, blue arrows and **Fig. 1d**, blue crescent). Most of these DHSs derived from sites of very weak pre-hormone chromatin accessibility that were potentiated following hormone treatment (**Supplementary Fig. 4**) and may thus represent indirect or 'network' effects of glucocorticoid receptor action.

Taken together, these results indicate that preexisting patterns of chromatin accessibility exert a dominant, global effect on *de novo* regulatory factor localization, and that factor occupancy is almost invariably associated with local chromatin remodeling.

Despite the fact that average pre-hormone chromatin accessibility at promoter regions was high, we observed 93% of glucocorticoid receptor occupancy sites >2.5 kb distal to the nearest transcriptional start site (compared to 61% of all DHSs; **Supplementary Fig. 5**). Glucocorticoid receptor sites were also highly clustered along the genome (**Supplementary Fig. 6**). However, we found no clear relationship between glucocorticoid receptor occupancy patterns and transcriptional activation of nearby genes (**Supplementary Table 5** and **Supplementary Fig. 7**), raising the possibility that the glucocorticoid receptor acts through long-range mechanisms, or that many glucocorticoid receptor binding events are opportunistic.

We next asked why, given the dominant influence of the chromatin structure, the glucocorticoid receptor occupied only a subset of DNase I–sensitive regions and why a small minority of glucocorticoid receptor binding events could escape the requirement for preexisting highly accessible chromatin. We first examined the relationship between GRBE motifs and glucocorticoid receptor occupancy patterns by developing an approach for quantifying the differential sensitivity of different GRBEs to their local chromatin environment. Of 2,296,115 GRBE (15 bp) matches[21] (**Fig. 2a**) in the non-repetitive mouse genome, only a very small fraction were actually occupied *in vivo* after hormone treatment. Standard position weight matrix matching[21] to the GRBE consensus was a poor predictor of glucocorticoid receptor binding, as many GRBEs with a high matching score were not occupied by a glucocorticoid receptor. However, we observed that many occupied GRBEs harbored distinct instantiations of the consensus sequence comprising specific combinations of non-degenerate bases (**Fig. 2b**).

To quantify the global relationship between these combinations and chromatin reprogramming, we partitioned the ~2.3-million candidate GRBEs into motif sequence classes such that all members of a given class shared identical non-degenerate consensus base sequences. Next, we computed a chromatin context coefficient (CCC) for each GRBE sequence class that quantified its relative dependence on pre-hormone chromatin accessibility as a prerequisite for post-hormone glucocorticoid receptor occupancy (**Fig. 2c–d** and **Supplementary Note**). High CCC values denote strong chromatin context dependence of glucocorticoid receptor binding, whereas low values mark classes with potential to override the dominant effect of chromatin structure and initiate local

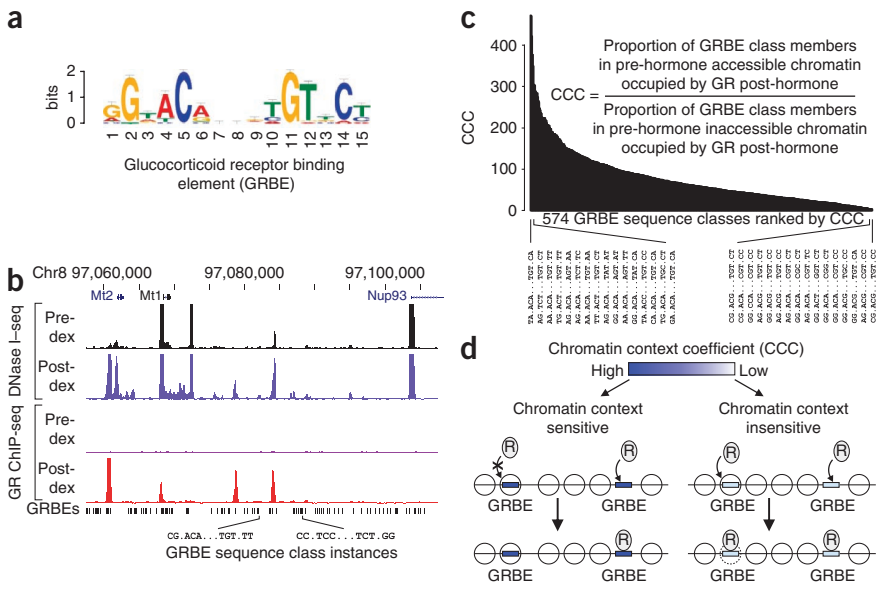

Figure 2 The quantitative effect of chromatin context on glucocorticoid receptor occupancy of GRBEs. (**a**) The top scoring motif recovered from *de novo* motif discovery performed on the top 500 glucocorticoid receptor occupancy sites by ChIP-seq tag density (MEME *E* value = 8.6×10^{-753}) closely matches the consensus glucocorticoid receptor binding element (GRBE). (**b**) A 50-kb genomic region comparing pre- and post-hormone chromatin accessibility and glucocorticoid receptor occupancy in relation to GRBE genomic sequence matches ($P < 10^{-3}$). Only a small fraction of the ~2.3×10^6 GRBE consensus sites are occupied *in vivo*, and occupied sites differ in their underlying combinations of consensus GRBE motif nucleotides. (**c**) GRBE sequence classes ranked by chromatin context coefficient (CCC). Genomic GRBE motif matches could be partitioned into discrete sequence classes, each comprising an identical, and distinct, combination of consensus nucleotides. Within each class of identical sequence elements, occurrence of member genomic sequences in a range of pre-hormone DNase I–sensitivity environments (from inaccessible to hyperaccessible) enabled quantification of the effect of chromatin context on the probability of post-hormone glucocorticoid receptor occupancy. Ranking specific GRBE sequence classes by CCC revealed graded sensitivity to chromatin context, from highly context-dependent elements that engender glucocorticoid receptor occupancy only when situated in accessible chromatin, to relatively context-independent elements associated with sites where glucocorticoid receptor occupancy induces chromatin remodeling. (**d**) Model illustrating the contribution of chromatin accessibility to transcription factor binding. CCC encodes the occupancy potential of different GRBE sequence classes relative to accessibility. GR, glucocorticoid receptor.

remodeling. Notably, we observed no CCC values <1, indicating that glucocorticoid receptor occupancy was universally enhanced by residence of GRBEs within pre-hormone accessible chromatin. Five hundred twenty-six out of one thousand one hundred statistically well-defined GRBE sequence classes lacked any occupancy at GRBE instances in pre-hormone closed chromatin (that is, CCC = ∞), indicating an absolute requirement of preexisting chromatin accessibility for glucocorticoid receptor occupancy (**Supplementary Note** and **Supplementary Table 6**). Ranking the remaining 574 GRBE sequence classes with finite CCC values revealed a hierarchy of chromatin dependence among GRBE elements, with the quantitative effect of preexisting chromatin accessibility on the probability of glucocorticoid receptor occupancy ranging from twofold to 473-fold (**Fig. 2c** and **Supplementary Table 6**). CCC values and GRBE class size were uncorrelated ($R^2 = 0.15$).

We next profiled both DNase I sensitivity and glucocorticoid receptor binding pre- and post-dexamethasone treatment in a highly divergent cell type (mouse pituitary cell line AtT-20) (**Fig. 3**, **Supplementary Fig. 8a–c** and **Supplementary Tables 7–10**). In pituitary cells, we found an even tighter targeting of *de novo* glucocorticoid receptor occupancy to pre-hormone accessible chromatin, with 95% (3,079 out of 3,242) of glucocorticoid receptor occupancy sites occurring within pre-hormone DNase I–sensitive regions (**Fig. 3c**). As in mammary cells, we observed no pre-hormone glucocorticoid receptor occupancy, and almost all (99%) post-hormone glucocorticoid receptor occupancy was accompanied by increased DNase I sensitivity. Pre-hormone chromatin accessibility patterns in mammary cells compared to pituitary cells were highly discordant (~30% overlap), consistent with cell type–specific *cis*-regulatory landscapes (**Fig. 3d**). The cell-selectivity of glucocorticoid receptor occupancy was even more pronounced, with only 11.4% (371 out of 3,242) of glucocorticoid receptor occupancy sites shared between pituitary and mammary cells (**Fig. 3e**).

Eighty-three percent (473 out of 572) of the GRBE sequence classes with well-defined CCC values in both 3134 and AtT-20 cells showed statistically significant enhancement of glucocorticoid receptor binding in both cell types (CCC > 1; **Supplementary Fig. 8d**). In AtT-20 cells, enhancement of GRBE occupancy by chromatin context ranged from threefold to 596-fold (**Supplementary Table 6**). The effects associated with specific GRBE classes were largely stable between cell types ($R = 0.48$, $P < 0.01$; **Supplementary Fig. 8e**). Notably, we were unable to identify a unique or specific GRBE sequence class that functioned exclusively to render closed chromatin more accessible.

In 3134 cells, ~25% of baseline accessible DHSs contained GRBEs, yet only 23% were occupied by a glucocorticoid receptor, suggesting additional requirements for glucocorticoid receptor binding. Glucocorticoid receptors have been reported to interact with a number of cell-restricted and ubiquitous transcriptional regulators[22]. We therefore examined glucocorticoid receptor sites for evidence of accessory factor motifs by performing *de novo* motif discovery on preprogrammed compared to reprogrammed sites from each cell type. This analysis revealed distinct complements of highly significant ($P < 10^{-5}$) motifs enriched in conjunction with classical GRBEs (**Fig. 4** and **Supplementary Fig. 9**). In mammary preprogrammed sites, these included AP-1 most prominently, as well as AML1, NF-κB and a previously unassigned motif (**Fig. 4a**). In pituitary preprogrammed sites, we recovered the canonical GRBE plus consensus motifs for HNF3, TAL1 and NF1 (**Fig. 4b**). Notably, both HNF3 and NF1 have previously been connected with both nuclear receptor binding generally and with glucocorticoid receptor interaction specifically[23,24]. ChIP analyses confirmed that at least a proportion of the identified sequence motifs were occupied by their cognate factors (**Supplementary Fig. 10a–e**).

Analysis of reprogrammed glucocorticoid receptor sites revealed a strikingly different picture. In 3134 cells, we found only the canonical

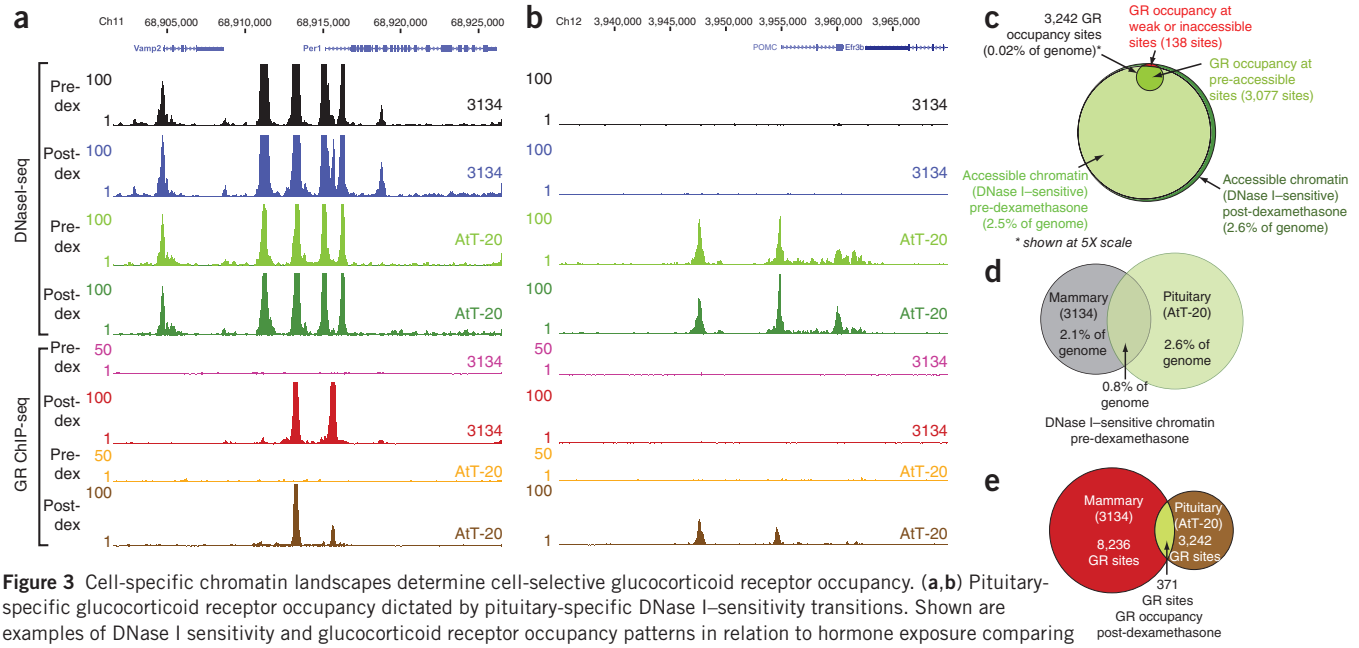

Figure 3 Cell-specific chromatin landscapes determine cell-selective glucocorticoid receptor occupancy. (**a**,**b**) Pituitary-specific glucocorticoid receptor occupancy dictated by pituitary-specific DNase I–sensitivity transitions. Shown are examples of DNase I sensitivity and glucocorticoid receptor occupancy patterns in relation to hormone exposure comparing mouse mammary (3134) and pituitary (AtT-20) cells (see **Fig. 1** legend and **Supplementary Fig. 8a–c** for additional examples). (**c**) Global glucocorticoid receptor occupancy compared to the chromatin accessibility landscape in pituitary cells. In pituitary cells, virtually all sites of glucocorticoid receptor occupancy (94.9%, or 3,079 out of 3,242 sites) occurred within pre-hormone accessible chromatin. The small fraction of reprogrammed glucocorticoid receptor sites (138 glucocorticoid receptor ChIP peaks, 4.2% of total) is shown in red. As in mammary cells, only a small fraction of pre-hormone accessible chromatin was occupied (for legibility, the glucocorticoid receptor circle is shown at ×5 scale). (**d**) Significant differences in the genomic distribution of pre-hormone DNase I sensitivity in mammary (gray) compared to pituitary (green) cells; only 0.78% of the genome (20.5 Mb) was accessible in both cell types. (**e**) Glucocorticoid receptor occupancy is highly cell selective. Only 371 glucocorticoid receptor occupancy sites are shared between mammary and pituitary cells (4.5% of 3134 cell sites and 11.4% of AtT-20 cell sites).

GRBE and AP-1 motifs. We found GRBEs in >80% of reprogrammed sites compared to only 29% of preprogrammed sites ($P < 10^{-100}$) (**Fig. 4c** and **Supplementary Fig. 9**), compatible with direct engagement of DNA following chromatin penetration. By contrast, we found consensus AP-1 sites in ~10% of reprogrammed sites compared to 26% of preprogrammed sites ($P < 10^{-80}$), and AP-1 and glucocorticoid receptor motifs were mutually exclusively distributed such that only 4.8% of preprogrammed sites had both (data not shown). In AtT-20 cells, we identified consensus HNF3 motifs in 34% of preprogrammed sites compared to 21% in reprogrammed glucocorticoid receptor sites ($P < 0.003$) (**Fig. 4c** and **Supplementary Fig. 9**). We observed mutual exclusivity between GRBEs and HNF3 in preprogrammed sites (only 5.8% of sites had both; $P < 10^{-11}$), which is analogous to the results with AP-1 in 3134 cells (data not shown). Taken together, these data

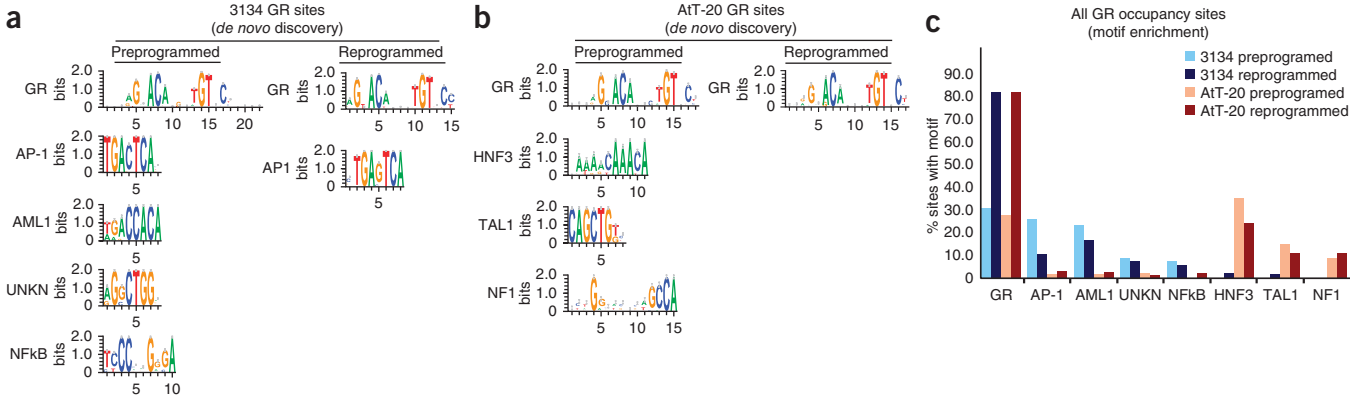

Figure 4 Regulatory motifs in glucocorticoid receptor–occupied regions differ substantially between cell types. (**a**,**b**) Results of *de novo* motif discovery (**Supplementary Note**) performed on the top 500 glucocorticoid receptor occupancy sites identified in 3134 (**a**) and AtT-20 (**b**) cells. The glucocorticoid receptor sites were further separated into preprogrammed (glucocorticoid receptor occupancy within pre-hormone accessible chromatin) and reprogrammed (glucocorticoid receptor occupancy within pre-hormone inaccessible chromatin) sites. Shown are motifs with highly significant enrichment ($P < 10^{-5}$). In all cases, the GRBE was the most highly enriched single motif ($E = 8.6 \times 10^{-753}$). Notably, AP1 and AML1 motifs were enriched in 3134 cells (**a**), whereas HNF3 and NF1 were correspondingly enriched in AtT-20 cells (**b**). (**c**) Motif occurrence patterns across all glucocorticoid receptor occupancy sites. Bar plots show percentage of all glucocorticoid receptor occupancy sites (8,236 sites in 3134 cells compared to 3,242 sites in AtT-20 cells) that harbor significant matches to the *de novo*–identified motifs from **a** and **b**. Note that canonical GRBEs are highly enriched in reprogrammed sites compared to preprogrammed sites (>80% of reprogrammed sites compared to <30% of preprogrammed sites; $P < 10^{-4}$).

suggest that in both cell types, common regulatory factors including AP-1 (3134 cells) and HNF3 (AtT-20 cells)—or possibly other factors acting through the same cognate motifs—may be mediating glucocorticoid receptor occupancy within a subset of pre-hormone accessible chromatin. However, this effect is quantitatively minor compared with that conferred by chromatin accessibility itself. For example, of the 34,587 positions in the mouse genome where AP-1 motifs and GRBEs co-occur, only 1.8% are occupied by glucocorticoid receptor post-hormone in 3134 cells, compared with the ~80% of glucocorticoid receptor binding that occurs within accessible chromatin generically (**Supplementary Fig. 10f–g**).

In summary, our results reveal the marked dominant effect of pre-existing chromatin structure on *de novo* regulatory-factor binding. This effect may be secondarily modulated by local sequence features such as variations in regulatory factor recognition elements or the presence of accessory sequence motifs for well-known regulators. However, even considered collectively, these additional sequence features likely account for only a minority of the overall effect.

Because of the dramatic dependence of regulatory factor binding on preexisting chromatin architecture, substantial variations in the baseline pattern of chromatin accessibility between different cell types are expected to expose distinct patterns and genomic locations of regulatory factor recognition sequences. The distribution of such exposed binding elements should, in turn, dictate the genomic distribution of *de novo* regulatory factor binding.

Corticosteroids are one of the most commonly used pharmaceuticals, and they exhibit widely differing effects on different tissues despite the fact that most human cell types contain the same glucocorticoid-response machinery[4]. Our results provide a simple explanation for these effects, namely, that they are a direct consequence of cell type–specific patterns of baseline (pre-hormone) chromatin accessibility and exposed glucocorticoid receptor recognition sequences.

A further implication of our results is that sequential factor occupancy during development and differentiation may be largely prespecified by the chromatin landscape as a form of cellular memory. Reprogramming of the chromatin structure at a limited number of sites may incrementally alter this pattern and create new potential occupancy sites for subsequently available factors, resulting in a directional process that is difficult to reverse without extraordinary measures such as the simultaneous introduction of multiple potent regulators[25].

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes. All DNase I and ChIP-seq data are available through the UCSC browser (http://genome.ucsc.edu/) and through NCBI Sequence Read Archive (SRA) under study number SRP004871 and the following accession codes: SRX034804, SRX034802, SRX034811, SRX034818, SRX034860, SRX034861, SRX034862, SRX034863, SRX034864, SRX034865, SRX034837, SRX034838, SRX034867, SRX034868, SRX034869, SRX034870, SRX034871 and SRX034872. All expression array data are available from the Gene Expression Omnibus (GEO) database under study number GSE26189 and the following accession codes: GSM642864, GSM642865, GSM642866, GSM642867, GSM642868, GSM642869, GSM642870, GSM642871, GSM642872, GSM642873, GSM642874, GSM642875, GSM642876, GSM642877, GSM642878.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We would like to thank T. Miranda, S. Morris, K. Nalley and L. Grontved for critical reading of the manuscript. We also thank M. Weaver, K. Lee, F. Neri, D. Bates and M. Diegel for technical assistance with the DNase I library preparation and sequencing. This research was supported in part by the Intramural Research Program of the US NIH, National Cancer Institute, Center for Cancer Research and funding from US NIH grant 1RC2HG005654 to J.A.S.

AUTHOR CONTRIBUTIONS

S.J., P.J.S., G.L.H. and J.A.S. designed the experiments. S.J., P.J.S., S.C.B. and T.A.J. conducted the DNase-seq, ChIP-seq and expression array experiments. S.J., P.J.S., R.E.T., M.-H.S. and J.A.S. analyzed the data. S.J., P.J.S., R.E.T., M.-H.S., G.L.H. and J.A.S. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Britten, R.J. & Davidson, E.H. Gene regulation for higher cells: a theory. *Science* **165**, 349–357 (1969).
2. McKenna, N.J. & O'Malley, B.W. Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* **108**, 465–474 (2002).
3. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007).
4. Evans, R.M. The steroid and thyroid hormone receptor superfamily. *Science* **240**, 889–895 (1988).
5. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448–453 (2003).
6. Wu, C. The 5′ ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. *Nature* **286**, 854–860 (1980).
7. Gross, D.S. & Garrard, W.T. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* **57**, 159–197 (1988).
8. Htun, H., Barsony, J., Renyi, I., Gould, D.L. & Hager, G.L. Visualization of glucocorticoid receptor translocation and intranuclear organization in living cells with a green fluorescent protein chimera. *Proc. Natl. Acad. Sci. USA* **93**, 4845–4850 (1996).
9. So, A.Y.-L., Chaivorapol, C., Bolton, E.C., Li, H. & Yamamoto, K.R. Determinants of cell- and gene-specific transcriptional regulation by the glucocorticoid receptor. *PLoS Genet.* **3**, e94 (2007).
10. Reddy, T.E. *et al.* Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res.* **19**, 2163–2171 (2009).
11. Richard-Foy, H. & Hager, G.L. Sequence-specific positioning of nucleosomes over the steroid-inducible MMTV promoter. *EMBO J.* **6**, 2321–2328 (1987).
12. Becker, P., Renkawitz, R. & Schütz, G. Tissue-specific DNaseI hypersensitive sites in the 5′-flanking sequences of the tryptophan oxygenase and the tyrosine aminotransferase genes. *EMBO J.* **3**, 2015–2020 (1984).
13. Hager, G.L. *et al.* Influence of chromatin structure on the binding of transcription factors to DNA. *Cold Spring Harb. Symp. Quant. Biol.* **58**, 63–71 (1993).
14. Hesselberth, J.R. *et al.* Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nat. Methods* **6**, 283–289 (2009).
15. Sekimata, M. *et al.* CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-gamma locus. *Immunity* **31**, 551–564 (2009).
16. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4**, 651–657 (2007).
17. Johnson, D.S., Mortazavi, A., Myers, R.M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
18. Stalder, J. *et al.* Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. *Cell* **20**, 451–460 (1980).
19. von der Ahe, D. *et al.* Glucocorticoid and progesterone receptors bind to the same sites in two hormonally regulated promoters. *Nature* **313**, 706–709 (1985).
20. Diamond, M.I., Miner, J.N., Yoshinaga, S.K. & Yamamoto, K.R. Transcription factor interactions: selectors of positive or negative regulation from a single DNA element. *Science* **249**, 1266–1272 (1990).
21. Bailey, T.L. & Gribskov, M. Concerning the accuracy of MAST E-values. *Bioinformatics* **16**, 488–489 (2000).
22. Beck, I.M.E. *et al.* Crosstalk in inflammation: the interplay of glucocorticoid receptor-based mechanisms and kinases and phosphatases. *Endocr. Rev.* **30**, 830–882 (2009).
23. Rigaud, G., Roux, J., Pictet, R. & Grange, T. *In vivo* footprinting of rat TAT gene: dynamic interplay between the glucocorticoid receptor and a liver-specific factor. *Cell* **67**, 977–986 (1991).
24. Cordingley, M.G. & Hager, G.L. Binding of multiple factors to the MMTV promoter in crude and fractionated nuclear extracts. *Nucleic Acids Res.* **16**, 609–628 (1988).
25. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).

ONLINE METHODS

Cell lines and culture conditions. The 3134 cell line was derived by transformation of the C127 cell line, which was originally isolated from a mammary adenocarcinoma tumor of the RIII mouse. The AtT-20 cell line is an anterior pituitary corticotroph of mouse origin (ATCC). Both cell lines were maintained in Dulbecco's Modified Eagle Medium (DMEM) (Invitrogen) supplemented with 10% FBS (Gemini), 2 mM L-glutamine, 1 mM sodium pyruvate, 0.1 mM non-essential amino acids and 5 mg/ml penicillin-streptomycin (Invitrogen) and kept at 37 °C in an incubator with 5% CO_2. Cells were transferred to 10% charcoal-dextran–treated, heat-inactivated FBS for 48 h before hormone treatment (1 h with 100 nM dexamethasone)[26].

ChIP assays. Chromatin immunoprecipitations were performed as per standard protocols (Upstate)[27]. Briefly, cells were treated with either vehicle or 100 nM dexamethasone for 1 h. Cells were crosslinked for 10 min at 37 °C in 1% formaldehyde followed by a quenching step for 10 min with 150 mM glycine. A single chromatin immunoprecipitation contained 400 µg of sonicated, soluble chromatin and a cocktail of antibodies to the glucocorticoid receptor (7.5 µg of PA1-511A antibody, ABR, 15 ug of MA1-510 antibody, ABR and 3 µg of sc-1004; Santa Cruz). The ChIP reaction was scaled ×5 for ChIP-seq. DNA isolates from immunoprecipitates were used as templates for real-time quantitative PCR amplification or sequenced as described below. All ChIP experiments were performed at least two times.

Digital DNase I mapping. Digital DNase I mapping was performed essentially as described in reference 28. Briefly, 3134 and AtT-20 cells were grown as described above. We pelleted 1×10^8 cells and washed them with cold phosphate-buffered saline. We resuspended cell pellets in Buffer A (15 mM Tris-Cl (pH 8.0), 15 mM NaCl, 60 mM KCl, 1 mM EDTA (pH 8.0), 0.5 mM EGTA (pH 8.0), 0.5 mM spermidine and 0.15 mM spermine) to a final concentration of 2×10^6 cells/ml. Nuclei were obtained by dropwise addition of an equal volume of Buffer A containing 0.04% NP-40 to the cells, followed by incubation on ice for 10 min. Nuclei were centrifuged at 1,000*g* for 5 min and then resuspended and washed with 25 ml of cold Buffer A. Nuclei were resuspended in 2 ml of Buffer A at a final concentration of 1×10^7 nuclei/ml. We performed DNase I (Roche, 10–80 U/ml) digests for 3 min at 37 °C in 2 ml volumes of DNase I buffer (13.5 mM Tris-HCl pH 8.0, 87 mM NaCl, 54 mM KCl, 6 mM $CaCl_2$, 0.9 mM EDTA, 0.45 mM EGTA, 0.45 mM Spermidine). Reactions were terminated by adding an equal volume (2 ml) of stop buffer (1 M Tris-Cl (pH 8.0), 5 M NaCl, 20% SDS and 0.5 M EDTA (pH 8.0), 10 µg/ml RNase A, Roche) and incubated at 55 °C. After 15 min, we added Proteinase K (25 µg/ml final concentration) to each digest reaction and incubated for one hour at 55 °C. After DNase I treatments, careful phenol-chloroform extractions were performed. Control (untreated) samples were processed as above except for the omission of DNase I. DNase I double-cut fragments and sequencing libraries were constructed as described in references 29 and 30.

High-throughput sequencing data analysis. High-throughput sequencing output was processed similarly for both DNase I and ChIP data. Twenty-seven base pairs of Illumina sequence reads were mapped to the human genome (UCSC HG18), and only uniquely mapping read positions were considered. For DNase I sequence tags, the 5′ ends represent *in vivo* cleavage events. Significantly enriched regions were identified in both DNase I and glucocorticoid receptor ChIP-seq datasets using a version of the Hotspot algorithm[31] (and Thurman *et al.*, in preparation; see also description below). Motifs in GR ChIP peaks were identified using the MEME algorithm[32].

Delineation of DNase I–sensitive regions. DNase I cleavage sites were represented computationally as the single base pair from the 5′ end of each sequence tag. Enrichment of tags along the genome was gauged in a small window (200–300 bp) relative to a local background model based on the binomial distribution and using the observed tags in a 50-kb surrounding window. Each mapped tag was given a *z*-score (explained below) relative to the surrounding small and background windows centered on the tag. A 'hotspot' was defined as a succession of neighboring tags within a 250-bp window, each of whose *z*-score was greater than 2. Once a hotspot was identified, the hotspot itself was assigned a *z*-score relative to the small and background windows centered on the average position of the tags forming the hotspot.

z-score calculation. Suppose n observed tags are mapped to the small window, and N total tags are mapped to the 50-kb surrounding background window ($N \geq n$). Each tag in the background window is considered an 'experiment' with a favorable outcome if it falls in the smaller window. Assuming each base in the 50-kb window is equally likely, the probability of success for each tag is therefore $P = 250/50{,}000$. Not all bases in the 50-kb window may be uniquely mappable by 27 mers (the tag length for our data), however, so p was adjusted to account for the number of uniquely mappable bases for that window. Under these assumptions, the binomial distribution applies, and the expected number of tags falling in the smaller window is $\mu = Np$.

The standard deviation of this expected value is $\sigma = \sqrt{Np(1-p)}$. Finally, the *z*-score for the observed number of tags in the smaller window is $z = n - \mu/\sigma$.

We also computed the expected number of tags and the *z*-score using the entire genome as background, rather than the 50-kb window, and, to be conservative, reported the lower of the two *z*-scores.

Correction for regional DNase I–sensitivity background. In regions of very high enrichment, the resulting hotspots can inflate the background for neighboring regions and deflate neighboring *z*-scores. The effect of this is that regions of otherwise high enrichment can be shadowed by a neighboring extreme hotspot. To address this problem, we implemented a two-pass procedure. After the first round of hotspot detection, we deleted all tags falling in the first-pass hotspots. We then computed a second round of hotspots with this deleted background. The hotspots from the first and second passes were combined, and all hotspots were rescored using the deleted background: the number of tags in each hotspot was computed using all tags, but the 50-kb background windows used only the deleted background.

Identification of DNase I hypersensitive peaks. Hotspots were resolved into discrete 150-bp peaks using a peak-finding procedure. First, neighboring hotspots within 150 bp of each other were merged. We computed a sliding window tag density (tiled every 20 bp in 150-bp windows) and then perform peak finding of the density in each merged hotspot region. Each 150-bp peak was assigned the *z*-score from the unmerged hotspot that contains it. Peak finding proceeded in two phases so that each hotspot has at least one peak. Phase-I peaks are local maxima occurring in regions above the ninety-ninth percentile of the density and satisfying certain *ad hoc* criteria for ensuring a sustained increase to or decrease from the local maxima. For each hotspot that does not contain at least one phase-I peak, a phase-II peak was simply defined as the maximum density value in the hotspot. For details, see the code available from the authors.

False discovery rate (FDR) calculations. We assigned FDR *z*-score thresholds to a given hotspot set using random data. As a null model, we computationally generated tags uniformly over the uniquely mappable bases of the genome. We used the same number of tags for observed and random data. The random data also coalesced into hotspots, which we identified and scored as usual. For a given *z*-score threshold T, the FDR for the observed hotspots with *z*-score greater than T was estimated as

$$FDR(T) \cong \frac{\#\text{ of random hotspots with } z \geq T}{\#\text{ of observed hotspots with } z \geq T}$$

As the numerator, which was calculated on a dataset that is entirely null, likely overestimates the number of false positives in the observed data, this is likely a conservative estimate of the FDR. Hotspots with FDR of 0% were constructed by taking all hotspots with a *z*-score greater than the maximum *z*-score attained in the paired random set. We constructed FDR-thresholded peak sets by performing peak finding in FDR-thresholded hotspots.

Generation of tables of DNase I sensitive regions and DHSs for pre- and post-hormone datasets. We observed that Dex DNase I hotspots (DNase I sensitive regions) that occur outside of Dex$^+$ DNase I hotspots are generally of low intensity and significance. We therefore restricted our published tables of Dex hotspots and peaks to those that also intersect Dex$^+$ hotspots. For 3134

doi:10.1038/ng.759

cells, we pooled samples from two replicates for each condition (Dex^- and Dex^+), whereas for AtT-20 cells, we used a single replicate per condition. See, however, the section on replicate concordant sets below, which details methods for defining DNase I sets for CCC analysis and aggregate plots.

Analysis of ChIP-seq data. The preceding sections describe procedures for handling DNase I tag data. Modifications were made to this process to account for unique properties of ChIP data. For one, duplicate tags (tags mapped to the same location) were used for DNase I but unique tags were only retained for ChIP calculations. This was because multiple tags mapping to the same position for DNase I provided biological meaning (the more tags at a given position, the more locally accessible the chromatin is at that location), whereas for ChIP data we expected the relevant information to be only the locations of measured binding. The most important difference between the processing of DNase I and ChIP data is the use of sequence data for the ChIP input experiment, which gives, for each ChIP experiment, a measure of the non-binding background signal, which can be significant. We used input tags at the scoring phase for ChIP hotspots. Once two-pass hotspots had been identified as usual, we scored each hotspot by first subtracting the number of tags in the paired input experiment from the observed ChIP tags in the hotspot window before applying the binomial model. We normalized the number of input tags subtracted in each window by a factor that brings the total number of input tags to the same number of ChIP tags. We did not subtract input tags from the surrounding 50-kb background window, so the scoring should be conservative.

Adjusted scoring for maximum sensitivity analyses using deep sequencing data. When scoring the deeper, 100-million-tag datasets, we strove for maximum sensitivity in detecting accessible chromatin, and therefore we made two adjustments in scoring hotspots. First, instead of taking the lower of the two *z*-scores from using a 50-kb local background and the genome-wide background, we used the greater of the two; and second, we lowered the initial *z*-score threshold for hotspot detection from two to one.

For additional methods, see the **Supplementary Note**.

26. John, S. *et al.* Kinetic complexity of the global response to glucocorticoid receptor action. *Endocrinology* **150**, 1766–1774 (2009).
27. John, S. *et al.* Interaction of the glucocorticoid receptor with the chromatin landscape. *Mol. Cell* **29**, 611–624 (2008).
28. Sekimata, M. *et al.* CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon-gamma locus. *Immunity* **31**, 551–564 (2009).
29. Sabo, P.J. *et al.* Genome-scale mapping of DNase I sensitivity *in vivo* using tiling DNA microarrays. *Nat. Methods* **3**, 511–518 (2006).
30. Hesselberth, J.R. *et al.* Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nat. Methods* **6**, 283–289 (2009).
31. Sabo, P.J. *et al.* Discovery of functional noncoding elements by digital analysis of chromatin structure. *Proc. Natl. Acad. Sci. USA* **101**, 16837–16842 (2004).
32. Bailey, T.L., Williams, N., Misleh, C. & Li, W.W. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* **34**, W369–W373 (2006).