DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*

William W. Fisher[a], Jingyi Jessica Li[b], Ann S. Hammonds[a], James B. Brown[b], Barret D. Pfeiffer[a,1], Richard Weiszmann[a], Stewart MacArthur[c], Sean Thomas[d], John A. Stamatoyannopoulos[d], Michael B. Eisen[c,e], Peter J. Bickel[b], Mark D. Biggin[f,2], and Susan E. Celniker[a,2]

[a]Department of Genome Dynamics, Division of Life Sciences, and [f]Division of Genome Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [b]Department of Statistics, [c]Department of Molecular and Cell Biology, and [e]Howard Hughes Medical Institute, University of California, Berkeley, CA 97420; and [d]Department of Genome Sciences, University of Washington, Seattle, WA 98195

Edited* by Kevin Struhl, Harvard Medical School, Boston, MA, and approved November 12, 2012 (received for review June 6, 2012)

In animals, each sequence-specific transcription factor typically binds to thousands of genomic regions in vivo. Our previous studies of 20 transcription factors show that most genomic regions bound at high levels in *Drosophila* blastoderm embryos are known or probable functional targets, but genomic regions occupied only at low levels have characteristics suggesting that most are not involved in the *cis*-regulation of transcription. Here we use transgenic reporter gene assays to directly test the transcriptional activity of 104 genomic regions bound at different levels by the 20 transcription factors. Fifteen genomic regions were selected based solely on the DNA occupancy level of the transcription factor Kruppel. Five of the six most highly bound regions drive blastoderm patterns of reporter transcription. In contrast, only one of the nine lowly bound regions drives transcription at this stage and four of them are not detectably active at any stage of embryogenesis. A larger set of 89 genomic regions chosen using criteria designed to identify functional *cis*-regulatory regions supports the same trend: genomic regions occupied at high levels by transcription factors in vivo drive patterned gene expression, whereas those occupied only at lower levels mostly do not. These results support studies that indicate that the high cellular concentrations of sequence-specific transcription factors drive extensive, low-occupancy, nonfunctional interactions within the accessible portions of the genome.

In vivo cross-linking experiments suggest that animal sequence-specific transcription factors each typically bind, at a minimum, to thousands or tens-of-thousands of genomic regions in every cell in which the transcription factor is active (1–4). Exponentially more genomic regions are cross-linked at low levels than are cross-linked at high levels, and the differences in levels of cross-linking between the DNA regions bound by a protein are only several ten-fold or several hundred-fold, depending on the transcription factor (1, 3). Importantly, a range of controls indicate that the levels of cross-linking in vivo are an accurate measure of the time averaged levels of DNA occupancy of each transcription factor at each location (5, 6).

It is challenging to determine which DNA binding events within these continua are functionally important, in part because of the complex and partially redundant interactions within animal transcription networks, as well as the prevalence of weak transcriptional regulatory events, the biological significance of which is unknown (2–4). Nevertheless, the more highly bound genomic regions include most known and probable functional targets, whereas the lowest-occupancy DNA binding events generally do not appear to be involved in the *cis*-regulation of transcription (1, 3, 7, 8). For example, in studies of *Drosophila* blastoderm patterning transcription factors we determined that genomic regions only bound at low occupancy have some of the following characteristics: proximity to genes whose biological functions are not associated with the bound transcription factors; proximity to genes that are not spatially regulated or not transcribed in early embryos; and mapping to poorly conserved sequences or protein coding sequences (6, 9). For simplicity, hereafter we will refer to DNA binding events where the transcription factor does not affect transcription of nearby genes in *cis* as nonfunctional, but recognize that other, nontranscriptional functions of DNA binding cannot be ruled out.

The concentrations of transcription factors and DNA in cells and the affinities of protein/DNA interactions measured in vitro are such that the majority of transcription factor molecules should be in direct contact with DNA in vivo (10–12). Fluorescence recovery after photo bleaching (FRAP), single molecule, in vivo footprinting, and other in vivo measurements of DNA binding generally support these thermodynamic predictions: most indicate that >90% of transcription factors molecules contact DNA (13–18), although estimates of only ~25% have been proposed in some FRAP studies (19). The sequence-independent, electrostatic affinity that all transcription factors have for DNA (K_d ~10^{-6} M) is sufficient to cause most molecules to be bound to DNA (10, 12, 15). However, DNA sequence-dependent interactions ($K_d < 10^{-8}$ M) mediate a high proportion of low-occupancy interactions because—for each transcription factor—tens-of-thousands of strong and weak matches to the factor's DNA recognition motifs in accessible parts of the genome are detectably bound in vivo (5, 20–22). The accessible regions are created by transcription factors competing nucleosomes off the DNA, which in turn allows DNA binding of additional transcription factor molecules (3). Animal transcription factors are each typically expressed at 10,000–300,000 molecules per cell (3). As a result, the ratios of the numbers of transcription factor molecules per cell to the length of accessible genome are much higher in animals than in prokaryotes (Table 1). Thus, the extensive low-occupancy DNA binding seen for animal transcription factors in in vivo cross-linking assays is not unexpected from a thermodynamic perspective.

Although the above lines of evidence suggest that the majority of low-occupancy DNA binding detected by in vivo cross-linking is likely nonfunctional, this idea has not been systematically tested using reporter gene assays. Several published studies have used heterologous reporter assays to test many tens of genomic regions identified in in vivo cross-linking assays in transgenic *Drosophila* or mice (23–26). However, these studies either focused only on

Author contributions: J.B.B., S.M., M.B.E., P.J.B., M.D.B., and S.E.C. designed research; W.W.F., A.S.H., B.D.P., and R.W. performed research; W.W.F., J.J.L., A.S.H., J.B.B., S.T., J.A.S., M.D.B., and S.E.C. analyzed data; and W.W.F., J.J.L., A.S.H., M.D.B., and S.E.C. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

[1]Present address: Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147.

[2]To whom correspondence may be addressed. E-mail: mdbiggin@lbl.gov or celniker@fruitfly.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1209589110/-/DCSupplemental.

Table 1. Predicted density of low-occupancy, nonfunctional DNA binding by transcription factors in vivo

Species	DNA length per cell (Mb)*	Accessible genome (Mb)	Transcription factor	Molecules per cell	Nonfunctional molecules per kb accessible genome†
E. coli	4.6	4.6 (38)	Lac I	10 (10)	0.002 (13, 15)
E. coli	4.6	4.6 (38)	Median	300 (39)	0.065
D. melanogaster	360.0	24.0 (20)	Median	50,000 (3)	2.1
H. sapiens	6,400.0	64.0 (40)	Median	120,000 (3)	1.9

**E. coli* DNA content is for a haploid cell, *D. melanogaster* and *Homo sapiens* is for diploid cells.
†In *E. coli* cells, at any instant 20% of Lac I transcription factor molecules make functional contacts with target *cis*-regulatory DNA recognition sites, 10% are not bound to DNA, and 70% make low-occupancy, nonfunctional DNA interactions (13, 15). Assuming that the same ratios apply for other transcription factors, the predicted density of low-occupancy interactions at accessible portions of the genome are given for a transcription factor expressed at the median concentration typical for a transcription factor in *E. coli*, *D. melanogaster*, and *H. sapiens*.

genomic regions bound at high levels in vivo and found that nearly all of the genomic regions tested function as *cis*-regulatory regions, or they did not take the level of DNA occupancy into account. Therefore, to test the function of genomic regions occupied only at low levels, we have leveraged our extensively controlled in vivo DNA binding data and a transgenic system for assaying *cis*-regulatory regions at multiple stages throughout *Drosophila* embryogenesis. Our results support the idea that most low-occupancy DNA interactions by animal transcription factors are not involved in the *cis*-regulation of transcription.

Results

Genomic Regions Chosen Based only on Kruppel DNA Occupancy Levels. To determine the relationship between transcription factor DNA occupancy levels and the ability of bound genomic regions to act as *cis*-regulatory regions in a transgenic reporter assay, we defined a set of genomic regions confidently bound by one transcription factor, Kruppel (KR). We used two well-validated antibodies that each recognize nonoverlapping regions of the KR protein to generate two separate microarray based in vivo cross-linking (ChIP-chip) datasets from blastoderm (stage 5) embryos (9). These two independent datasets are highly correlated ($r = 0.97$) and the false-discovery rates (FDR) calculated from them had previously been confirmed separately by quantitative PCR (9). To calculate a FDR that combines data from both antibodies, we conservatively assigned each 675-bp window in the genome the lowest of the two ChIP scores from each antibody. We then used these "antibody-minimum" ChIP scores to define genomic regions bound by KR that had a cumulative FDR (9, 27) <5% and an irreproducible discovery rate (28) < 0.1% (*SI Materials and Methods*). Through this process, 5,713 such KR-bound genomic regions were identified. These regions were then ranked based on the local peak in the antibody-minimum ChIP signal in each region, with the most highly bound region designated as rank 1 and the most poorly bound region as rank 5,713 (Dataset S1). We also calculated the local FDR (29), which is especially useful as it gives the probability for each genomic region that it is falsely discovered. The 5,713 regions designated as "KR-bound" all had local FDRs <21% (Dataset S1).

We selected for transgenic analysis a set of 15 of the KR-bound genomic regions based only on KR ChIP scores (*SI Materials and Methods*). The local FDR of these 15 regions shows that less than one is expected to be falsely discovered (Dataset S2). Each of the 15 regions was expanded to 1.5 kb in length, centered around the location of the local peak in the ChIP-score. These 1.5-kb sequences were placed upstream of a universal promoter/GAL4 reporter gene fusion, and the resulting constructs were each integrated at a common AttP chromosomal integration site to eliminate position effect variability between transformed lines (30). We term the selected 1.5-kb genomic regions within these constructs as "KR-rank" transgenic test regions (KR-rank TTRs). We also use TTRs to refer to additional classes of genomic regions assayed by transgenic analysis, described later.

The six most highly bound KR-rank TTRs lie in the top 1,280 on our rank list of KR-bound genomic regions. Five of the six TTRs drive complex spatial patterns of reporter gene expression in embryos at stage 5 of development, and two of the reporter patterns resemble those of one of the genes that flank the TTRs normal chromosomal location (Fig. 1, Table 2, and Datasets S2, S3, and S4). The remaining nine KR-rank TTRs have scores that span ranks 1,151–5,124 of the KR-bound genomic regions. Only one of these TTRs drives detectable reporter expression at stage 5, the expression mimicking that of a nearby gene (Table 2 and Datasets S2, S3, and S4).

We also examined the activity of the KR-rank TTRs at later stages of embryogenesis, a period of ~18 h during which the 6,000 undifferentiated cells present at blastoderm undergo three further cell divisions, move extensively, and differentiate to form complex tissues (31). All six most highly bound KR-rank TTRs are active *cis*-regulatory regions at some time after stage 5 (Fig. 1, Table 2, and Datasets S2, S3, and S5). Five of the nine lowest-occupancy KR-rank TTRs become active during at least one stage after blastoderm, each driving a specific and unique pattern of reporter expression (Fig. 1, Table 2, and Datasets S2, S3, and S5). With two exceptions, the expression patterns driven in the late embryo resemble those of one of the flanking genes, where gene-expression data are available to allow this to be assessed (Datasets S2, S3, and S5). The remaining four low-occupancy TTRs do not drive embryonic patterned reporter expression (Fig. 1, Table 2, and Datasets S2 and S5).

The fact that nearly all TTRs bound at low levels by KR at stage 5 are not detectably active at this stage is consistent with the idea that low-occupancy interactions by transcription factors are nonfunctional. The fact that some low-occupancy TTRs become active after stage 5 cannot be taken as evidence that the binding of KR at stage 5 to these sequences contributes to this later activity. As discussed in more detail below, in the densely packed genome of *Drosophila melanogaster*, *cis*-regulatory regions are present so frequently that genomic regions picked at random will often function as enhancers in transgenic assays at some stage during embryogenesis.

Larger Survey of Genomic Regions. To further explore the relationship between the levels of transcription factor DNA occupancy and the ability of genomic regions to act as *cis*- regulatory regions, we expanded our analysis to test the activity of 137 additional genomic regions. In parallel to the analysis of KR-rank TTRs, we tested two additional sets of genomic regions in transgenic assays. One set was largely selected to identify *cis*-regulatory regions based on being bound in vivo at high levels by multiple transcription factors and mapping within 10 kb of genes expressed in spatial patterns at stage 5 (*SI Materials and Methods*). The second set was selected to identify *cis*-regulatory regions based on evolutionary conserved clustering of DNA recognition sites for five blastoderm transcription factors, Bicoid (BCD), Caudal (CAD), Knirps (KNI), Hunchback (HB), and KR, and being within 10 kb of genes spatially expressed at stage 5 (32).

Of these additional TTRs, 89 are bound by KR using the same criteria used to define the KR-rank TTRs; that is, they all have cumulative FDRs <5% and local FDRs <21% (*SI Materials and*

BIOCHEMISTRY

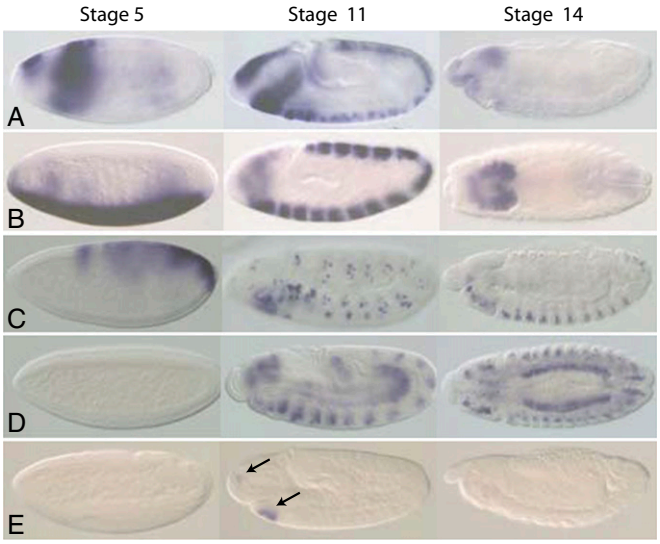

Fig. 1. Examples of reporter gene-expression patterns driven by KR-rank TTRs at stages 5, 11, and 14. *A*–*C* show data for three TTRs that are among the six most highly bound by KR at stage 5. These three TTRs are each active at stages 5–14. *D* and *E* shows data for two of the nine TTRs bound at low levels by KR at stage 5. Both of these TTRs are not detectably active at stage 5. The TTR shown in *D* becomes active at stages 11–14. The TTR in *E* shows faint staining at stage 11 just ventral of the stomodeum and at the anterior tip (arrows). This faint pattern is a variably penetrate reporter artifact in TTRs lines that are otherwise not detectably active, as it is also seen in a line bearing a basal promoter/GAL4 reporter construct that lacks a TTR. TTRs showing only this pattern are therefore classified as not active throughout embryogenesis. All embryos were imaged with differential interference contrast microscopy and are printed at 50-fold magnification.

Methods). These TTRs were then joined with the KR-rank TTRs to form a set of 104 KR-bound TTRs (Dataset S2). The remaining TTRs form a second set of 48 KR-unbound TTRs (Dataset S2). To permit the properties of TTRs bound at different levels by KR to be compared, we divided the KR-bound TTRs into three cohorts (top, middle, and bottom) based on the KR ChIP window score. From the local FDR, only one of the 104 KR-bound TTRs is expected to be falsely discovered (Dataset S2), almost certainly one of the lowest-ranked members of the bottom cohort. This finding indicates that our analysis of the KR-bound TTRs will not be significantly confounded by the presence of regions that are not in fact bound in vivo.

Consistent with previous results (6, 20), the levels of in vivo DNA occupancy by KR and 19 other sequence specific transcription factors to each TTR and the degree of accessibility of each region to DNaseI enzyme digestion in nuclei isolated from blastoderm nuclei are highly correlated (Fig. 2) (6, 20). More importantly, the levels of transcription factor DNA occupancy and DNaseI accessibility generally predict the likelihood that a TTR will be active in the transgenic reporter assay (Fig. 2, Table 3, and Datasets S4 and S5). For example, 91% of the top KR-bound cohort are active at stage 5, whereas only 46% of the bottom KR-bound cohort are active at stage 5 (Table 3). The difference in activity at stage 5 between the top cohort and the bottom cohort are highly significant (Bonferroni-corrected P value $= 1 \times 10^{-4}$). Thus, consistent with the results obtained with the KR-rank TTRs, low-occupancy DNA binding events tend to be nonfunctional, whereas the majority of genomic regions bound at high levels are functional in the transgenic assay.

An important caveat when interpreting the results in Fig. 2 is that—with the exception of the KR-rank TTRs—criteria other than the level of transcription factor DNA occupancy in vivo were used to help select genome regions for testing in the transgenic reporter assay. To determine what bias these additional selection criteria introduced, we first identified new cohorts of genomic regions that were randomly selected from the set of all 5,713 genomic regions bound by KR to yield cohorts with similar distributions of KR DNA occupancy levels as the top, middle, and bottom TTRs. No other criteria were used in the selection of these "occupancy-matched" cohorts. We then compared properties of the top, middle, and bottom KR-bound TTRs to those of the occupancy-matched cohorts and also to those of regions picked randomly from the entire genome (Fig. 3). If the TTR cohorts showed similar properties to the equivalent occupancy-matched cohorts, then this would suggest that the criteria used to choose the TTRs had not introduced a bias. Alternatively, if the TTRs cohorts differed in their properties from the occupancy-matched cohorts, then the analysis would indicate the degree and direction of the bias.

Consistent with earlier results (6), the DNA sequences of the top cohort of occupancy-matched regions are more strongly conserved between *Drosophila* species than are the sequences of the bottom occupancy-matched cohort (Fig. 3*D*, red bars). Also as expected (6), the members of the top occupancy-matched cohort are generally closer to genes with Gene Ontology annotations, suggesting that they are developmental regulators, genes that are transcribed by RNA polymerase II, and genes that are expressed in spatial patterns at stage 5 than are the bottom occupancy-matched cohort (Fig. 3 *A*–*C*, red bars). Interestingly, the properties of the bottom occupancy-matched cohort are similar to regions picked randomly from the entire genome, consistent with low-occupancy DNA binding by KR being nonfunctional (Fig. 3, compare bottom cohort red bars to green bars).

Paralleling the properties of occupancy-matched regions, the members of the top TTRs are closer on average to developmental regulatory genes and to genes transcribed by polymerase II than are bottom TTRs (Fig. 3 *A* and *B*, blue bars). In contrast to the occupancy-matched cohorts, however, the DNA sequences of bottom TTRs tend to be almost as well-conserved as those of top TTRs, and bottom TTRs are typically nearly as close to spatially expressed genes as top TTRs (Fig. 3 *C* and *D*, blue bars). These biases in the bottom TTRs are not unexpected, however: these TTRs are dominated by TTRs whose selection criteria included proximity to spatially patterned gene and evolutionary conservation, but did not include the biological functions of nearby genes nor proximity to RNA polymerase II transcribed gene.

More importantly, although the selection criteria used to identify the non–KR-rank TTRs have introduced some biases, these biases are such that the estimates using the data in Fig. 2 to assess the proportion of low KR occupancy regions that are nonfunctional will likely be conservative. In other words, it is probable that a cohort of genomic regions selected only on the basis of low KR DNA occupancy will include either the same number or fewer active *cis*-regulatory regions than found in the bottom KR-bound cohort because many members of this cohort share some properties that are more commonly found in highly bound, active *cis*-regulatory regions.

Table 2. Functional activity of KR rank TTRs

Cohort	Rank by KR ChIP score	TTRs	TTRs active at stage 5	TTRs active at stages 9–14	TTRs inactive all stages
KR rank top	1–1,280	6	5 (83%)	6 (100%)	0 (0%)
KR rank bottom	1,551–5,124	9	1 (11%)	5 (56%)	4 (44%)

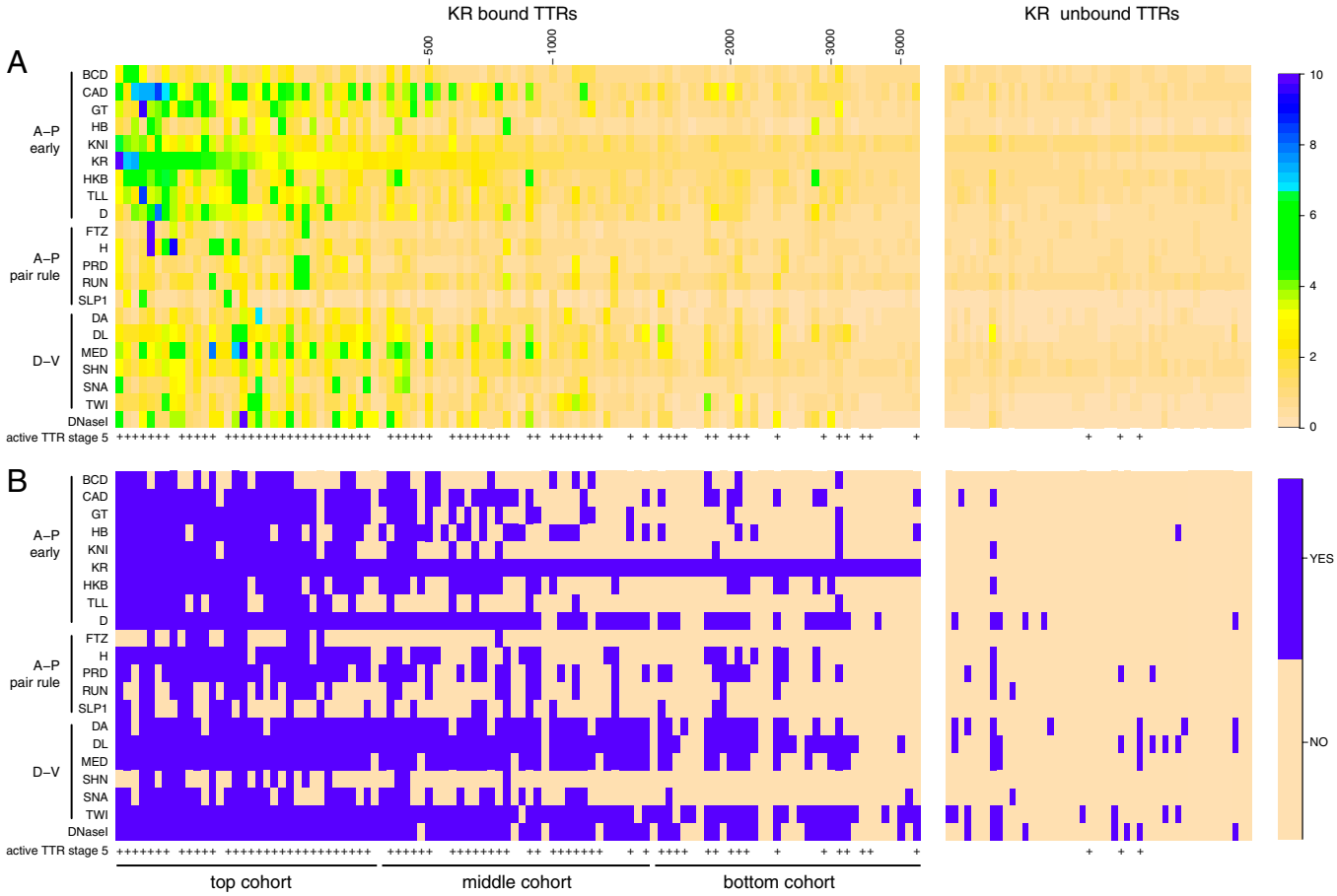

Fig. 2. Correlation between *cis*-regulatory activity and transcription factor DNA binding and genome accessibility. Each of the 152 columns displays data for one TTR. The columns are divided into KR-bound TTRs (*Left*) and KR-unbound TTRs (*Right*) and are ranked by the antibody-miniumum KR ChIP score (*SI Materials and Methods* and Dataset S2). Locations of TTRs along the rank list of 5,713 genomic regions bound by KR are shown (*Top Left*) and the KR-bound TTRs belonging to the top, middle and bottom cohorts are indicated (*Bottom Left*). (*A*) The upper heat map rows show in vivo DNA binding ChIP scores for 20 transcription factors in stage 5 embryos. The lowest heat map row indicates the degree of DNaseI accessibility at stage 5. The row beneath this indicates if the TTR drives reporter expression in stage 5 embryos (+ symbols). (*B*) The colored rows show if in vivo DNA binding or DNaseI accessibility is confidently detected for each TTR (blue) or is not confidently detected (pink) (*SI Materials and Methods*). The lowest row shows if the TTR drives reporter transcription in stage 5 embryos (+ symbols).

Of the bottom KR-bound TTRs, 69% are active in at least one stage postblastoderm (Table 3). Of the 48 KR-unbound TTRs, 38% are active *cis*-regulatory regions after stage 5 (Table 3), and of the 29 that are not bound by any of the 20 transcription factors, 19% are active in later embryogenesis (Dataset S2). Similar percentages of *cis*-regulatory activity during embryogenesis have also been reported for randomly selected genomic regions in other transgenic studies (26). In addition, a subset of transcription factor interactions on *cis*-regulatory regions are probably nonfunctional, even when the regulatory region is actively regulating transcription (3), and there are at least 100 transcription factors expressed in every animal cell, most molecules of which will likely engage in extensive, low-occupancy DNA binding. Thus, there is no compelling evidence that activity in late embryogenesis is a consequence of binding of either KR or other transcription factors to these TTRs at stage 5.

Discussion

Multiple lines of evidence suggest that transcription factors in all organisms make extensive, low-occupancy, nonfunctional interactions that are thermodynamically driven by the concentrations of transcription factors and DNA in cells (10–18; reviewed in ref. 3). For most animal transcription factors, the ratio of protein molecules per cell to the length of accessible genome is particularly high (Table 1). These ratios, combined with the high frequencies of DNA recognition sites for each transcription factor throughout the genome (33), imply only modest differences in DNA binding levels between high-occupancy functional interactions

Table 3. Functional activity of all TTRs combined

Cohort	Rank by KR ChIP score	TTRs	TTRs active at stage 5	TTRs active at stages 9–14	TTRs inactive all stages
KR-bound top	1–372	34	31 (91%)	30 (88%)	1 (3%)
KR-bound middle	385–1,567	35	25 (71%)	29 (83%)	4 (11%)
KR-bound bottom	1,602–5,460	35	16 (46%)	24 (69%)	10 (29%)
KR-unbound	NA	48	3 (6.3%)	18 (38%)	30 (63%)

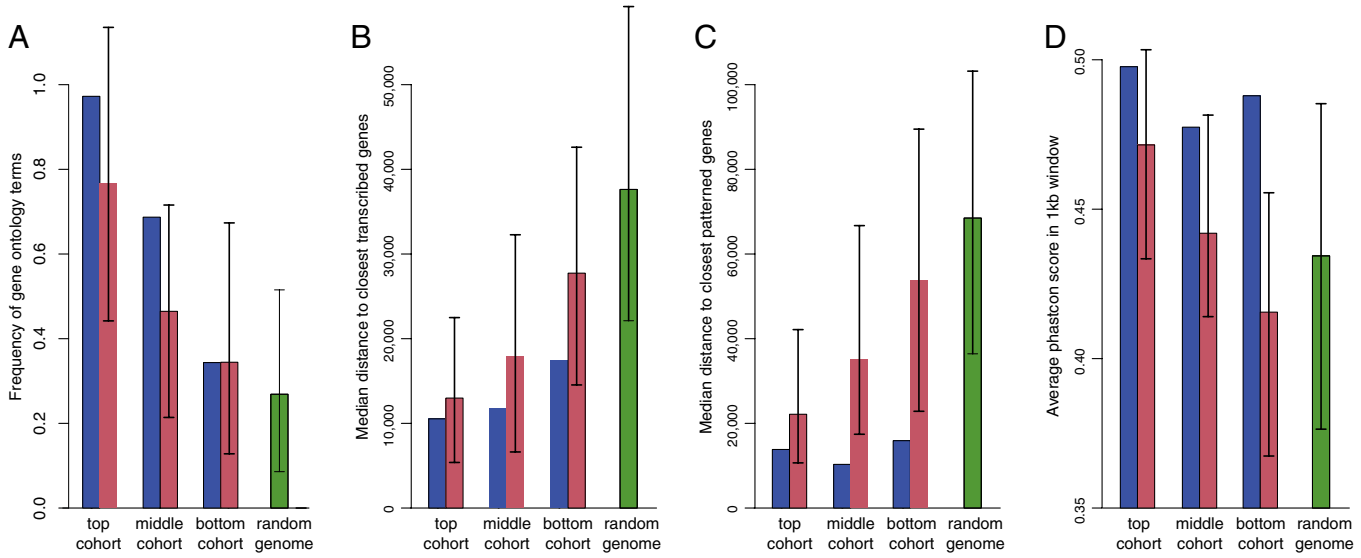

Fig. 3. Properties of different KR occupancy cohorts. Different properties of KR-bound TTRs (blue bars), KR occupancy-matched genomic regions (red bars), and randomly selected genomic regions (green bars) are shown. The height of the bar shows the mean. For the top, middle, and bottom occupancy-matched cohorts and the random genome cohort, 95% confidence intervals indicate the 2.5% and 97.5% quantiles from 100 samplings (*SI Materials and Methods*). (*A*) The frequency of the eight most common Gene Ontology terms for the genes closest to the top KR-bound TTR and top occupancy-matched cohorts. (*B*) The median distance to the start site of the closest gene cross-linked by the actively transcribing, phosphorylated form of RNA polymerase II. (*C*) The median distance to the start site of the closest gene whose mRNA is expressed in spatial patterns in stages 4–6. (*D*) The mean PhastCons DNA sequence conservation score for *Drosophila* species in 1-kb windows centered in the middle of each genomic region.

and lower-occupancy nonfunctional interactions. In vivo cross-linking studies support this prediction, showing a continuum of DNA occupancies that stretches from high levels at known *cis*-regulatory regions to lower levels at many thousands of regions that have characteristics, suggesting that they are not involved with the *cis*-regulation of transcription (1, 6–9). The differences in DNA occupancy levels across this spectrum of interactions are typically only several ten-fold or several hundred-fold, depending on the transcription factor.

Here we have directly tested the functional activity of high- and low-occupancy genomic regions in transgenic reporter assays in *Drosophila* embryos. Although 91% of regions bound at high levels in blastoderm embryos act as *cis*-regulatory regions at this early stage of development (Table 3), only 46% of genomic regions bound at low occupancy are active at this time, and 29% are not detectably active at any stage of embryogenesis (Tables 2 and 3). Although a number of genomic regions bound at low levels and not active in the blastoderm do become active *cis*-regulatory regions later in development, we have no evidence that this later activity requires the DNA binding of the transcription factors observed at blastoderm. Approximately one-third of genomic regions not bound by any transcription factor at blastoderm have *cis*-regulatory activity at some later stage of embryogenesis (Table 3 and Dataset S2) (26), and the highly promiscuous binding by multiple transcription factors at most accessible genomic regions suggest that these interactions do not lead either immediately or eventually to significant changes in transcription of nearby genes.

For simplicity, we use the term "nonfunctional" to refer to DNA binding events that do not affect transcription of nearby genes *in cis*, either at the time DNA binding is measured or as a later consequence of the binding event. It has long been appreciated, however, that low-occupancy interactions that do not affect transcription *in cis* will affect the system *in trans* by lowering the concentration of unbound transcription factor molecules, which will in turn necessarily reduce the occupancy levels of transcription factors at highly bound, functional *cis*-regulatory regions (10–18). Low-occupancy DNA binding could in this sense be said to be functional, but it is a very different function from that of regulating nearby genes via *cis*-regulatory regions, and most low-occupancy interactions are not under strong natural selection in the same way that direct regulatory interactions in *cis*-regulatory regions are (Fig. 3*D*) (6, 9).

It is challenging to prove that a given molecular binding event has no biological function. The transgenic assay that we have used may not detect certain classes of bona fide transcriptional *cis*-regulatory regions, including insulators, pure silencers, DNA sequences that augment the activity of nearby regulatory regions but which are not active on their own, and sequences that can act autonomously when coupled with a proximal promoter but drive expression patterns that are too weak to be detected in a standard transgenic assay. There is no reason, however, to suppose that silencers, insulators, or other strongly acting *cis*-regulatory region preferentially use only low-occupancy interactions. It is a general finding that transgenic assays in *Drosophila* and other model organisms detect *cis*-regulatory regions, the sum of whose activities approximates the wild-type transcription patterns of the associated genes (34–37). These observations suggest that transgenic assays are sensitive enough to detect most *cis*-regulatory regions.

Low-occupancy interactions might have a nontranscriptional function. For example, the muscle specification transcription factor MyoD induces modest quantitative increases in histone acetylation at tens-of-thousands of the genomic regions to which it binds in vivo, which could lead to higher-order changes in chromatin structure without directly regulating the transcription of genes in the vicinity of the DNA binding event (22).

We suggest, however, that along the continuum of transcription factor DNA occupancy levels in vivo, a point must be reached where most interactions have no specific functional impact, other than the reduction of unbound protein concentrations *in trans*. Where that point lies will likely differ for different transcription factors and developmental contexts. Some proteins may have important biological functions in regulating modest quantitative transcriptional responses for thousands of genes (3) and may also act more diffusely by modulating large-scale changes in chromatin architecture (22). Other proteins may significantly regulate only tens or hundreds of genes. However, extensive, nonfunctional interactions appear to be an unavoidable consequence of whatever selective pressure drove animals to adopt transcriptional regulators

that have relatively broad DNA sequence specificities, and which are expressed at high concentrations.

What may initially appear to be a different perspective on low-occupancy DNA binding in vivo has been suggested by Tanay (38). This report describes a weak correlation between in vivo ChIP-chip scores and transcriptional function across the 90% of genes that have the lowest ChIP-chip scores in yeast. These in vivo cross-linking scores are well below those we consider as statistically significant here, and their correlation with function is only observed after the ChIP scores have been adjusted using a novel noise-removal model, limiting confidence in Tanay's conclusions. Nonetheless, if it is assumed that the weak correlation is biologically relevant, Tanay's analysis is not inconsistent with ours, because only a small minority of low-occupancy interactions need be functional to produce the poor correlation observed. The possibility that some weak interactions may have a function does not argue that all are functional.

In summary, although much work needs to be done to determine the fraction of DNA interactions that are biological significant in controlling transcription in *cis*, the broad sweep of the data imply that a high proportion of low-occupancy interactions are nonfunctional. Extensive, nonfunctional interactions have also been proposed for other classes of proteins, including kinases and RNA polymerase, and thus may be common to many biological processes (3, 39, 40).

Materials and Methods

DNA Constructs, Transgenics, and Reporter Gene Expression. Genome coordinates of the DNA sequences of TTRs are given in Dataset S2 and the selection criteria and ChIP-chip and other data used to identify them are described in *SI Materials and Methods*. All TTRs were PCR-amplified from *y:cn bw sp* (the *D. melanogaster* sequenced reference strain) and their DNA sequences were verified after cloning. KR-rank and ChIP TTRs were cloned into pBPGUw and the resulting DNA constructs were integrated at an attP2 integration site located on the third chromosome at 68A4, as described previously (41). Correct insertion at the AttP site was verified by crossing *w+* F1 males to *y w*; *Dr*, *e/TM3*, *Sb* females. Gal4 reporter gene expression was detected by in situ RNA hybridization as previously described (41). "Cluster" TTRs were cloned into an *eve* basal promoter/β-galactosidase reporter gene construct and transformed into w^{1118} embryos, and reporter gene expression was detected as described previously (32). Only those cluster constructs that showed consistent expression patterns in at least two separate transformed lines were considered further. As previously described (42), in cluster TTR that we class as not active, we occasionally detect a weak anterior stripe of expression because of the sequences in the basal promoter/b-galactosidase/P-transposon sequences. For images showing gene-expression patterns in embryos, the horizontal and vertical aspect ratios were occasionally altered to allow alignment (Fig. 1 and Datasets S3, S4, S5, and S6).

ACKNOWLEDGMENTS. We thank the members of the Berkeley *Drosophila* Transcription Network Project for their advice and encouragement. This work was supported by National Institutes of Health Grants R01 GM704403 (to M.D.B., S.E.C., and M.B.E.), P01 GM009655 (to M.D.B. and S.E.C.), and R01 GM076655 (to S.E.C.). Work at the Lawrence Berkeley National Laboratory was conducted under Department of Energy Contract DEAC02-05CH11231.

1. Walter J, Dever CA, Biggin MD (1994) Two homeo domain proteins bind with similar specificity to a wide range of DNA sites in *Drosophila* embryos. *Genes Dev* 8(14):1678–1692.
2. Farnham PJ (2009) Insights from genomic profiling of transcription factors. *Nat Rev Genet* 10(9):605–616.
3. Biggin MD (2011) Animal transcription networks as highly connected, quantitative continua. *Dev Cell* 21(4):611–626.
4. Walhout AJ (2011) What does biologically meaningful mean? A perspective on gene regulatory network validation. *Genome Biol* 12(4):109.
5. Kaplan T, et al. (2011) Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early *Drosophila* development. *PLoS Genet* 7(2):e1001290.
6. MacArthur S, et al. (2009) Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol* 10(7):R80.
7. Rey G, et al. (2011) Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver. *PLoS Biol* 9(2):e1000595.
8. Yu HB, Johnson R, Kunarso G, Stanton LW (2011) Coassembly of REST and its cofactors at sites of gene repression in embryonic stem cells. *Genome Res* 21(8):1284–1293.
9. Li XY, et al. (2008) Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol* 6(2):e27.
10. Lin S, Riggs AD (1975) The general affinity of lac repressor for *E. coli* DNA: Implications for gene regulation in procaryotes and eucaryotes. *Cell* 4(2):107–111.
11. Yamamoto KR, Alberts BM (1976) Steroid receptors: Elements for modulation of eukaryotic transcription. *Annu Rev Biochem* 45:721–746.
12. von Hippel PH, Revzin A, Gross CA, Wang AC (1974) Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: Equilibrium aspects. *Proc Natl Acad Sci USA* 71(12):4808–4812.
13. Kao-Huang Y, et al. (1977) Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound *Escherichia coli* lac repressor in vivo. *Proc Natl Acad Sci USA* 74(10):4228–4232.
14. Yang SW, Nash HA (1995) Comparison of protein binding to DNA in vivo and in vitro: Defining an effective intracellular target. *EMBO J* 14(24):6292–6300.
15. Elf J, Li GW, Xie XS (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316(5828):1191–1194.
16. Phair RD, et al. (2004) Global nature of dynamic protein-chromatin interactions in vivo: Three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Mol Cell Biol* 24(14):6393–6402.
17. Janssen S, Cuvier O, Müller M, Laemmli UK (2000) Specific gain- and loss-of-function phenotypes induced by satellite-specific DNA-binding drugs fed to *Drosophila melanogaster*. *Mol Cell* 6(5):1013–1024.
18. Liu X, Wu B, Szary J, Kofoed EM, Schaufele F (2007) Functional sequestration of transcription factor activity by repetitive DNA. *J Biol Chem* 282(29):20868–20876.
19. Mueller F, Wach P, McNally JG (2008) Evidence for a common mode of transcription factor interaction with chromatin as revealed by improved quantitative fluorescence recovery after photobleaching. *Biophys J* 94(8):3323–3339.
20. Li XY, et al. (2011) The role of chromatin accessibility in directing the widespread, overlapping patterns of *Drosophila* transcription factor binding. *Genome Biol* 12(4):R34.
21. Rhee HS, Pugh BF (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* 147(6):1408–1419.
22. Cao Y, et al. (2010) Genome-wide MyoD binding in skeletal muscle cells: A potential for broad cellular reprogramming. *Dev Cell* 18(4):662–674.
23. Visel A, et al. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457(7231):854–858.
24. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE (2009) Combinatorial binding predicts spatio-temporal *cis*-regulatory activity. *Nature* 462(7269):65–70.
25. Junion G, et al. (2012) A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* 148(3):473–486.
26. Kvon EZ, Stampfel G, Yáñez-Cuna JO, Dickson BJ, Stark A (2012) HOT regions function as patterned developmental enhancers and have a distinct *cis*-regulatory signature. *Genes Dev* 26(9):908–913.
27. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100(16):9440–9445.
28. Li Q, Brown JB, Huang H, Bickel PJ (2011) Measuring reproducibility of high-throughput experiments. *Annals of Applied Statistics* 5(3):1752–1779.
29. Efron B (2007) Size, power and false discovery rates. *Ann Stat* 35(4):1351–1377.
30. Groth AC, Fish M, Nusse R, Calos MP (2004) Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. *Genetics* 166(4):1775–1782.
31. Campos-Ortega JA, Hartenstein V (1997) *The Embryonic Development of Drosophila melanogaster* (Springer, Berlin), , Second Ed.
32. Berman BP, et al. (2004) Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*. *Genome Biol* 5(9):R61.
33. Wunderlich Z, Mirny LA (2009) Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet* 25(10):434–440.
34. Davidson EH (2006) *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution* (Academic, Burlington, MA).
35. Fujioka M, Emi-Sarker Y, Yusibova GL, Goto T, Jaynes JB (1999) Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development* 126(11):2527–2538.
36. Castelli-Gair J, Müller J, Bienz M (1992) Function of an Ultrabithorax minigene in imaginal cells. *Development* 114(4):877–886.
37. Wimmer EA, Simpson-Brose M, Cohen SM, Desplan C, Jäckle H (1995) *Trans*- and *cis*-acting requirements for blastodermal expression of the head gap gene buttonhead. *Mech Dev* 53(2):235–245.
38. Tanay A (2006) Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Res* 16(8):962–972.
39. Struhl K (2007) Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* 14(2):103–105.
40. Levy ED, Landry CR, Michnick SW (2009) How perfect can protein interactomes be? *Sci Signal* 2(60):pe11.
41. Pfeiffer BD, et al. (2008) Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proc Natl Acad Sci USA* 105(28):9715–9720.
42. Small S, Blair A, Levine M (1992) Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *EMBO J* 11(11):4047–4057.