FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin

Paul G. Giresi,[1] Jonghwan Kim,[2] Ryan M. McDaniell,[2] Vishwanath R. Iyer,[2] and Jason D. Lieb[1,3]

[1]*Department of Biology and the Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3280, USA;* [2]*Institute for Cellular and Molecular Biology and Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas 78712-0159, USA*

DNA segments that actively regulate transcription in vivo are typically characterized by eviction of nucleosomes from chromatin and are experimentally identified by their hypersensitivity to nucleases. Here we demonstrate a simple procedure for the isolation of nucleosome-depleted DNA from human chromatin, termed FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements). To perform FAIRE, chromatin is crosslinked with formaldehyde in vivo, sheared by sonication, and phenol-chloroform extracted. The DNA recovered in the aqueous phase is fluorescently labeled and hybridized to a DNA microarray. FAIRE performed in human cells strongly enriches DNA coincident with the location of DNaseI hypersensitive sites, transcriptional start sites, and active promoters. Evidence for cell-type–specific patterns of FAIRE enrichment is also presented. FAIRE has utility as a positive selection for genomic regions associated with regulatory activity, including regions traditionally detected by nuclease hypersensitivity assays.

[Supplemental material is available online at www.genome.org.]

Chromatin at genomic loci that actively regulate transcription is distinguished from other chromatin types. The observation that the 5′ regions of genes became hypersensitive to both DNaseI and micrococcal nuclease upon gene activation in *Drosophila* was among the earliest demonstrations of this phenomenon (Wu et al. 1979; Wu 1980; Keene and Elgin 1981; Levy and Noll 1981). The appearance of these hypersensitive sites reflects a loss or destabilization of nucleosomes at the promoters of active genes (Boeger et al. 2003). Several mechanisms act in concert to achieve this result. Loss of nucleosomes can be caused directly by a protein bound to its cognate site on DNA (Yu and Morse 1999), facilitated in part by increased acetylation of the nucleosomes just before the activation of transcription (Reinke and Horz 2003), or mediated by the well-characterized SWI/SNF family of adenosine triphosphate-dependent nucleosome remodeling complexes (Tsukiyama and Wu 1995; Sudarsanam and Winston 2000; Varga-Weisz 2001). Regardless of the specific mechanisms employed at any individual promoter, achieving nucleosome clearance at active regulatory regions is a conserved mechanism among eukaryotes (Wallrath et al. 1994).

Because nucleosome disruption is a conserved hallmark of active regulatory chromatin throughout the eukaryotic lineage, a simple, high-throughput procedure to isolate and map chromatin depleted of nucleosomes would allow identification of regulatory regions in a broad range of organisms and cell types. The promise of one such procedure, which we now term FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements), was first demonstrated in *Saccharomyces cerevisiae* (hereafter "yeast") (Nagy et al. 2003). Following phenol-chloroform extraction of formaldehyde-crosslinked yeast chromatin, the genomic regions immediately upstream of genes were preferentially segregated into the aqueous phase (Fig. 1). The enrichment of regulatory regions in the aqueous phase was interpreted to indicate relatively inefficient crosslinking between proteins and DNA at these regions. Histones are by far the most abundant and readily crosslinkable protein component of chromatin and thus were likely to dominate the crosslinking profile (Brutlag et al. 1969; Solomon and Varshavsky 1985; Polach and Widom 1995). Therefore, it had been further hypothesized that FAIRE reflected heterogeneity in the occupancy and distribution of nucleosomes throughout the genome. Consistent with this hypothesis, the promoters of heavily transcribed yeast genes were more highly enriched by FAIRE than were promoters of genes with lower transcription initiation rates (Nagy et al. 2003). More recent experiments in yeast have shown that enrichment by FAIRE has a very strong negative correlation with nucleosome occupancy (Hogan et al. 2006), as measured by comparison with nucleosome ChIP–chip experiments (Bernstein et al. 2004; Lee et al. 2004) and high-resolution mapping of nucleosomes with micrococcal nuclease digestion (Yuan et al. 2005).

Human chromatin poses new challenges to FAIRE. Compared with the 12-million base-pair genome of yeast, the three-billion base-pair human genome is nearly 300 times as large. Only ~1.5% of human DNA is coding, with perhaps 30% of the genome transcribed (introns plus exons), relative to 50% coding for yeast, with 85% of the genome being transcribed under a single growth condition (Wong et al. 2001; Hurowitz and Brown 2003; Rao et al. 2005; David et al. 2006). In addition, mammalian chromatin is inherently more complex than that of yeast. Most mammalian genes contain introns, regulation can occur at much

[3]Corresponding author.
E-mail jlieb@bio.unc.edu; fax (919) 962-1625.
Article is online at http://www.genome.org/cgi/doi/10.1101/gr.5533506.
Freely available online through the *Genome Research* Open Access option.

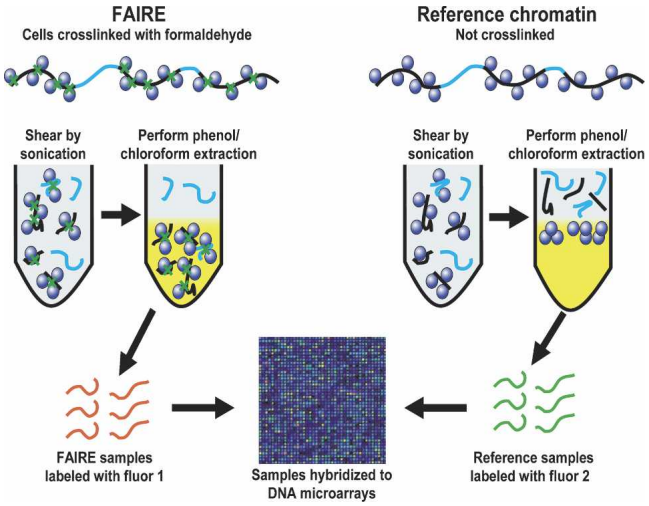

Figure 1. FAIRE in human cells is illustrated on the *left*, while preparation of the reference is illustrated on the *right*. For FAIRE, formaldehyde is added directly to cultured cells. The crosslinked chromatin is then sheared by sonication and phenol-chloroform extracted. Crosslinking between histones and DNA (or between one histone and another) is likely to dominate the chromatin crosslinking profile (Brutlag et al. 1969; Solomon and Varshavsky 1985; Polach and Widom 1995). Covalently linked protein–DNA complexes are sequestered to the organic phase, leaving only protein-free DNA fragments in the aqueous phase. For the hybridization reference, the same procedure is performed on a portion of the cells that had not been fixed with formaldehyde, a procedure identical to a traditional phenol-chloroform extraction. DNA resulting from each procedure is then labeled with a fluorescent dye, mixed, and comparatively hybridized to DNA microarrays. In this case, we used high-density oligonucleotide arrays that tile across the ENCODE regions of the human genome (30 Mb).

greater distances from the initiation of transcription, there are more repetitive and heterochromatic regions, and the baseline state of chromatin is more compact and repressive (Alberts et al. 2002). Therefore, it is reasonable to expect that a much smaller fraction of the genome will be in the "open" conformation representing regions of active chromatin. Moreover, it is not clear a priori whether the same physical properties of yeast chromatin that allow isolation of open regions by FAIRE can be successfully exploited for isolation of regulatory regions in human chromatin.

Here, we performed FAIRE in a human foreskin fibroblast cell line and assayed its performance within the genomic regions selected by the ENCODE Project Consortium (2004). Regions enriched by FAIRE were compared with functional genomic elements such as DNaseI hypersensitive sites, transcriptional start sites (TSSs), and active promoters. The results indicate that FAIRE is a simple genomic method for the isolation and identification of human functional regulatory elements, with broad utility for mammalian genomes.

Results

DNA isolated by FAIRE in human cells corresponds to regions of active chromatin

Fibroblasts were grown in culture, and formaldehyde was added directly to actively dividing cells to a final concentration of 1% (see Methods). The cells were then disrupted with glass beads. The resulting extract was sonicated to yield 0.5- to 1-kb chromatin fragments, and subjected to phenol-chloroform extraction (Fig. 1). The DNA fragments recovered in the aqueous phase were fluorescently labeled and hybridized to high-density oligonucleotide microarrays tiling the ENCODE regions at 38-bp resolution. The ENCODE regions represent 1% of the human genome (30 Mb), consisting of manually selected regions of particular interest and randomly selected regions of varying gene density and evolutionary conservation (The ENCODE Project Consortium 2004). As a reference, DNA prepared in parallel from uncrosslinked cells was labeled with a different fluor and simultaneously hybridized to the arrays.

We compared the genomic regions enriched by FAIRE to hallmarks of active chromatin, including localization of the general transcriptional machinery (Kim et al. 2005a,b), histone H3 and H4 acetylation and methylation (Koch et al. 2007), DNaseI hypersensitivity (Crawford et al. 2006; Sabo et al. 2006), and direct assays of promoter activity (Trinklein et al. 2003; Cooper et al. 2006). Genomic regions enriched by FAIRE correspond well with each of these indicators of active regulatory elements (Fig. 2, Table 1).

Active promoters are enriched by FAIRE

Earlier experiments performed in yeast had revealed that the regulatory regions of highly transcribed genes are preferentially isolated by FAIRE (Nagy et al. 2003). To determine whether this relationship holds in human cells, we compared FAIRE signal to measurements of promoter strength. Predicted promoters in the ENCODE regions have been analyzed for regulatory activity by cloning them upstream of reporters and measuring the resulting activity of the reporter gene in different cell types (Trinklein et al. 2003; Cooper et al. 2006). We assigned each probe on the microarray that mapped to a predicted promoter to one of four classes, based on the average activity of the corresponding promoter. Analysis revealed that probes mapping to the most active promoters have a higher FAIRE signal than those that do not map to a promoter or that map to a promoter of lower activity (Fig. 3A, $P < 10^{-100}$). Therefore, more active promoters are more strongly enriched by FAIRE in human cells.

FAIRE isolates DNA encompassing TSSs

Yeast experiments had also revealed that FAIRE isolated the nucleosome-free region located at yeast TSSs (Nagy et al. 2003; Yuan et al. 2005; Hogan et al. 2006). Alignment of DNase-chip signal (Crawford et al. 2006), FAIRE signal, and gene annotations suggested that a similar feature was enriched by FAIRE in human cells (Fig. 2). To assess the extent to which this was generally true, we aligned all TSSs for all annotated genes within the ENCODE regions and calculated the average FAIRE signal over a region spanning 1.5 kb upstream to 1.5 kb downstream of the TSS (Fig. 3B, solid line). This analysis revealed that, on average, the peak of enrichment by FAIRE occurs at the TSS. DNase hypersensitive sites are an indicator of DNA accessibility and a well-established characteristic of TSSs and regulatory DNA. We performed the same analysis using DNase-chip data (Crawford et al. 2006) and found that the pattern of DNA enrichment at TSSs was very similar to that generated by FAIRE (Fig. 3B, broken line).

Global comparison of FAIRE peaks to other annotated features

We also analyzed the overall concordance between the genomic regions enriched by FAIRE and other selected hallmarks of active chromatin (Fig. 3C; TSS [Ashurst et al. 2005; Harrow et al. 2006],

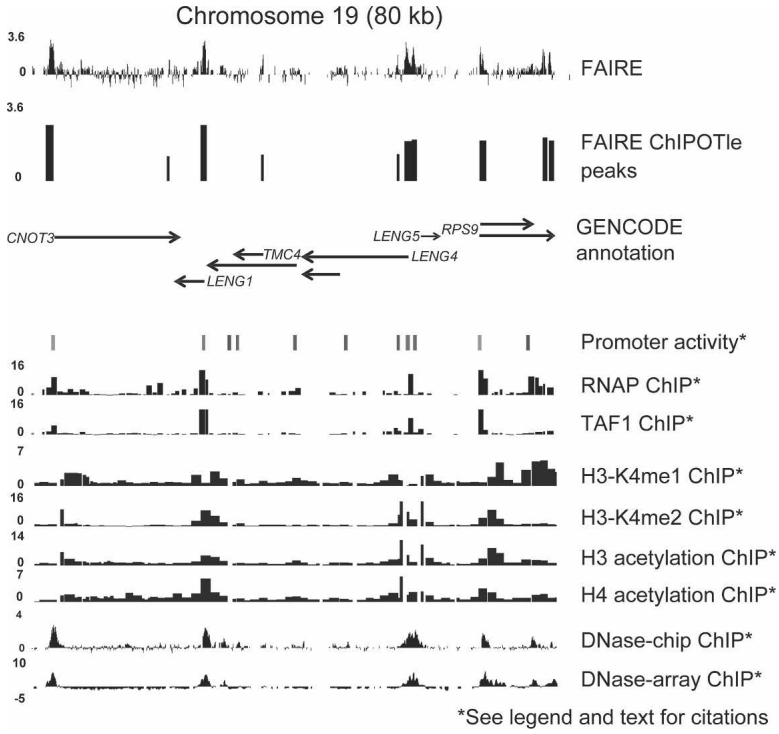

Figure 2. FAIRE enrichment of regulatory DNA across 80 kb of human chromosome 19. FAIRE data were loaded into the UCSC Genome Browser along with data sets generated by other ENCODE Consortium members (labeled on the *right*). The *top* track represents the average $\log_2$ ratios for the FAIRE data from four independent cultures (biological replicates), each of which were crosslinked separately (for 1, 2, 4, and 7 min). The second track shows FAIRE peaks (cutoff = $P < 10^{-25}$) as determined by ChIPOTle (Buck et al. 2005). The GENCODE annotations represent experimentally verified transcribed segments (Ashurst et al. 2005; Harrow et al. 2006). "Promoter activity" represents the average activity of a reporter construct driven by each of the indicated regions and measured across 16 cell lines, where light gray bars indicate high activity and black bars no activity (Trinklein et al. 2003; Cooper et al. 2006). ChIP–chip data for RNAP and TAF1 from lung fibroblast cells (IMR90) are displayed as the $-\log_{10}$ of the *P*-value for each probe, scaled to 0–16 (Kim et al. 2005a,b). ChIP–chip data for histone H3 and H4 acetylation and H3K4 mono-, di-, and trimethylation in embryonic lung fibroblast cells (HFL-1) are shown as the ratio of ChIP signal over background (Koch et al. 2007). Finally, data on DNaseI hypersensitivity are shown for two different techniques, DNase-chip and DNase-array. Both techniques isolate DNA fragments flanking DNaseI cleavage sites and map them back to the genome using microarrays (Crawford et al. 2006; Sabo et al. 2006). The data shown for DNase-chip are the average $\log_2$ ratio for nine replicates (3 biological at 3 different enzyme concentrations), whereas the DNase-array data are the $\log_2$ ratios scaled so that a $\log_2$ ratio of 0 represents the 99% confidence bound on the experimental noise. The region shown corresponds to chromosome 19 coordinates 59,330,000 to 59,409,000.

DNaseI hypersensitivity [Crawford et al. 2006; Sabo et al. 2006], 75th percentile of promoter activity [Trinklein et al. 2003; Cooper et al. 2006], RNA polymerase II ChIP–chip, or TAF1 ChIP–chip [Kim et al. 2005a,b]). The concordance of FAIRE peaks with these marks is very strong, in most cases over 10 times the frequency observed with permuted data (Table 1). Furthermore, 21% of all FAIRE peaks overlap multiple marks of active chromatin (Fig. 3C). Forty-three percent of the FAIRE peaks are "orphans," which do not correspond to any of the annotations selected for comparison. These likely arise because of a number of factors, most significantly the difference in cell types used among the experiments being compared and the sparse state of current human genome annotations (see Discussion).

qPCR verification

To determine the extent to which the DNA microarray signals accurately reflect the identity of DNA fragments isolated by FAIRE, we performed real-time quantitative PCR (qPCR) analysis on samples from independently grown fibroblasts. We designed 85 primer pairs spanning three genomic loci within the ENCODE regions, each of which contained several FAIRE peaks. At each position covered by a pair of primers, we determined FAIRE enrichment by calculating the ratio of signal from the FAIRE sample relative to the uncrosslinked control sample. All ratios were normalized to an unlinked locus. The data were concordant with the regions that were strongly enriched by FAIRE according to tiling microarrays, even in the case of "orphan" FAIRE peaks like those shown in Figure 3D. These data indicate that the signal measured by the microarrays faithfully represents the population of DNA fragments isolated by FAIRE and is not an artifact of amplification, labeling, or microarray hybridization.

FAIRE isolates regulatory elements specific to individual cell types

Although all somatic cells in an organism contain the same genomic DNA, different cell types express different genes. These differences reflect differential utilization of regulatory information encoded in the genome. To determine whether FAIRE could detect regulatory elements specific to a certain cell type, we compared FAIRE data derived from fibroblasts with DNase-chip data derived from lymphoblastoid cells (Fig. 4). The data are concordant at most promoters (Fig. 4A, black circle), and there was very little signal from either assay as one moved away from the proximal promoter (Fig. 4B, black circle). However, there were a number of probes that detected differences between the assays in the different cell types (Fig. 4, A and B, gray circles).

Differences between FAIRE and DNase hypersensitivity could result from either (1) similar underlying chromatin but differences in what FAIRE and DNase hypersensitivity detect or (2) real differences in the chromatin state between the different cell types. To determine which was more likely, we examined loci that contained a FAIRE peak but not a DNase-chip peak, were within 500 bp of a TSS, and were covered by probes over at least 100 contiguous bases. Forty-one (5%) of the GENCODE annotated genes (1.4% of TSS) met this definition. The largest and most pronounced locus mapped to one of the fibroblast growth factor 1 (*FGF1*) TSSs. Examination of data collected in lung fibroblast cells (IMR90) revealed that this promoter was indeed occupied by RNAP (currently known as POLR2A) and TAF1 in fibroblasts (Kim et al. 2005a,b), consistent with our isolation of that promoter by FAIRE using fibroblast cells (Fig. 4C). However, in a lymphoblast cell line that does not express the *FGF1* gene, no DNaseI hypersensitivity was detected (Fig. 4C). Furthermore, in HeLa S3 cells (which also do not express *FGF1*), the promoter was

Table 1. The peak-finding algorithm ChIPOTle yielded 1008 FAIRE peaks (cutoff = $P < 10^{-25}$, see Methods)

How many of the. . .

308 RNAP sites	281 TAF1 sites	3150 DNase hypersensitive sites	75th %tile of promoter activity (162)	2888 GENCODE TSSs	
. . .overlap with the 1008 identified FAIRE peaks?					
157 (51%)	194 (69%)	677 (22%)	109 (67%)	390 (14%)	
16 ± 5	16 ± 4	123 ± 13	5 ± 3	89 ± 16	Permuted

How many of the 1008 FAIRE peaks overlapped with the. . .

308 RNAP sites	281 TAF1 sites	3150 DNase hypersensitive sites	75th %tile of promoter activity (162)	2888 GENCODE TSSs	
144 (14%)	189 (19%)	492 (49%)	107 (11%)	169 (17%)	
14 ± 4	13 ± 4	91 ± 9	7 ± 3	52 ± 7	Permuted

The location of each FAIRE peak was compared with hallmarks of active chromatin, taking into account the width of the features reported by the authors Kim et al. (2005a,b); Cooper et al. (2006); Crawford et al. (2006); Harrow et al. (2006); Koch et al. (2007). The number of features reported for each data set is shown in parenthesis in the *top* panel. The overlap between data sets was calculated by searching 250 bp on either side of a FAIRE peak. Overlap using other window sizes (including zero) and increasing or decreasing peak-finding stringency was calculated with no substantive change in results. The *top* panel shows the number of features that fall within 250 bp of a FAIRE peak, whereas the *bottom* panel shows the number of FAIRE peaks with a corresponding feature within 250 bp on either side. To assess significance, we generated 1008 peaks of the same width as those observed for FAIRE, randomized their genomic location within the ENCODE regions, and calculated overlap with genomic features as described above. This permutation was performed 1000 times. The distributions (overlap with permuted peaks) were compared to a Gaussian distribution using a Q-Q plot and found to be normal. *P*-values were then calculated in R; with the observed overlap compared with the distribution generated using permuted peaks. All *P*-values were $<10^{-100}$.

not bound by RNAP or TAF1 (Fig. 4C). These data indicate that FAIRE can detect biologically relevant, cell type–specific differences in chromatin.

FAIRE isolates intragenic transcription start sites specific to individual cell types

The transcription of the lymphocyte-specific protein 1 gene (*LSP1*) is regulated in a tissue-specific manner, whereby alternative promoters are utilized in lymphocyte or fibroblast cells. This alternative promoter usage is controlled by differential utilization of regulatory elements in the two cell-types (Gimble et al. 1993; Misener et al. 1994; Thompson et al. 1996). The promoter that produces the longer *LSP1* transcript is utilized in lymphocyte cells, whereas the promoter producing the shorter fragment is utilized in fibroblasts (Fig. 5). We examined the *LSP1* locus to determine whether FAIRE (performed in fibroblasts) could detect alternative promoter usage in comparison with DNaseI hypersensitivity signal (performed in lymphocytes). Both FAIRE and DNaseI hypersensitivity signals were detected at the *LSP1* locus but were localized to the alternative TSSs unique to each cell type (Fig. 5). Specifically, the DNaseI hypersensitivity peak derived from lymphoblasts was found only at the promoter of the lymphocyte-specific transcript, and the FAIRE signal was found only at the promoter of the fibroblast-specific transcript. Additional data from lung fibroblast cells (IMR90) (Kim et al. 2005a,b) confirm that the general transcriptional machinery is localized to the fibroblast-specific TSS and that the fibroblast TSS harbors histone modifications characteristic of an active TSS. Therefore, FAIRE can isolate TSSs specific to individual cell types.

Discussion

FAIRE as a method for identification of active regulatory elements

Several aspects of FAIRE make it a powerful genome-wide approach for detecting functional in vivo regulatory elements in mammalian cells. First, FAIRE requires no treatment of the cells before the addition of formaldehyde. Formaldehyde is applied directly to the growing cells and enters quickly because of its small size (HCHO). In yeast, 1% formaldehyde immediately stops cell growth and results in 50% lethality in just 100 sec, with 99% lethality achieved in 360 sec (data not shown). Therefore, the state of chromatin just before the addition of the formaldehyde is likely to be captured. In contrast, nuclease sensitivity assays often require that cells be permeabilized, or that nuclei be prepared, both of which allow time for artifacts based on these preparations to occur.

Second, each time a nuclease-sensitivity assay is performed, the appropriate enzyme concentration and incubation time must be determined, because of lot-to-lot variations in commercial DNase activity and variations in individual nuclei preparations. With FAIRE, a wide range of incubation times (1, 2, 4, and 7 min) at a single formaldehyde concentration (1%) appears to be equally effective. FAIRE involves few steps, few variables and takes less than an hour, making the method easy to control and develop. Few reagents other than formaldehyde, phenol, and chloroform are required. These properties make FAIRE amenable to high throughput. Third, in contrast with ChIP, there is no dependence on antibodies, supplies of which may be limited, or on tagged proteins, which may be difficult to construct, impaired in function, or expressed at inappropriate levels. FAIRE can analyze any cells: wild type, mutant, or those that contain transgenes that would make histone ChIPs technically difficult (e.g., those containing Protein-A–based tags).

Another important advantage of FAIRE is that it positively selects genomic regions at which nucleosomes are disrupted. These same regions would be degraded in nuclease sensitivity assays and require identification by their absence or by cloning and identification of flanking DNA (Crawford et al. 2004). In contrast, DNA isolated by FAIRE is the DNA of interest, allowing the use of direct detection methods like DNA microarrays.

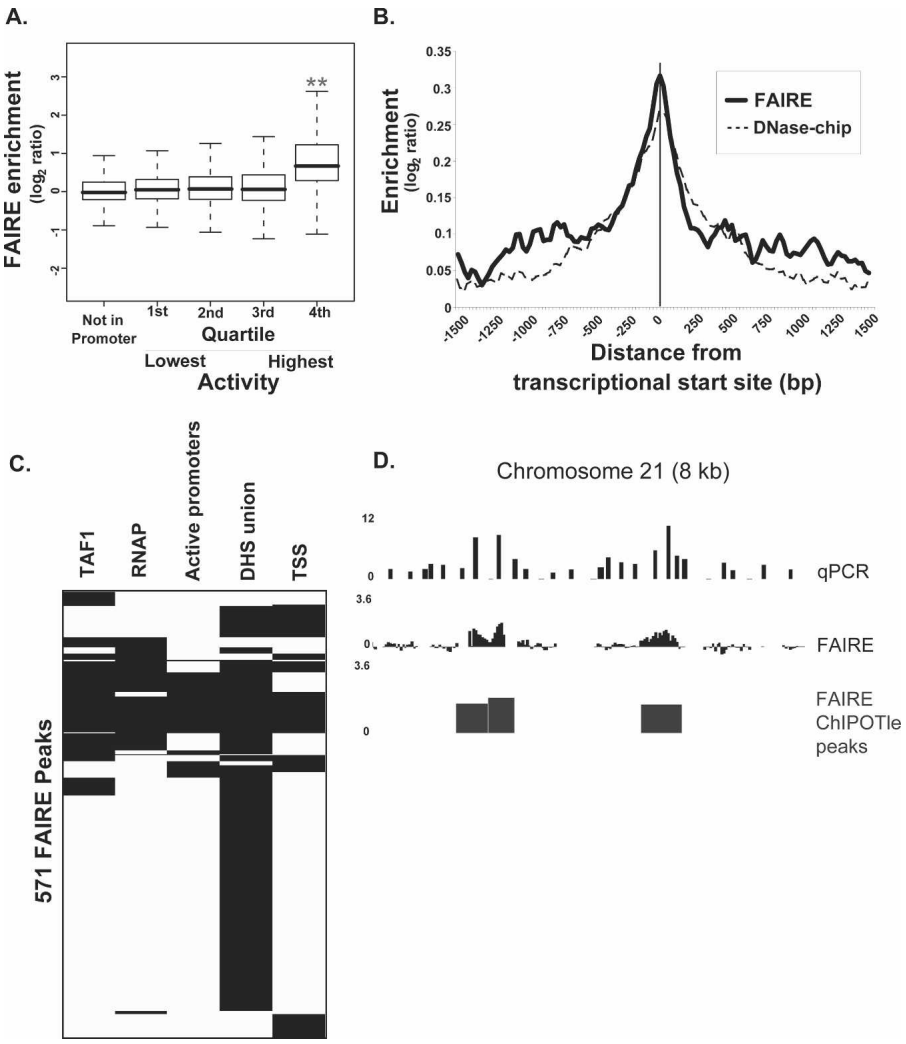

Figure 3. FAIRE isolates DNA at the TSSs of genes. (*A*) Probes that mapped to predicted promoters were divided into quartiles based on the level of activity for each promoter, which was measured by using it to drive a reporter construct (Trinklein et al. 2003; Cooper et al. 2006). The reported activity represents an average from the 16 different cell types assayed. Boxes represent the 25th to the 75th percentile of the FAIRE data (interquartile range, IQR), the black line in the *middle* of the box is the median, and the dotted lines extend out 1.5 times the IQR. Probes within the regions of highest regulatory activity (fourth quartile, *right* side), represent the most active promoters and correspond to regions most efficiently isolated by FAIRE (**$P < 10^{-100}$). (*B*) Probes from the high-density oligonucleotide tiling array were mapped relative to GENCODE annotated TSSs (Ashurst et al. 2005; Harrow et al. 2006). A sliding window (50 bp, 1-bp steps) was then used to calculate the average FAIRE enrichment from 1.5 kb upstream to 1.5 kb downstream of the TSS (solid line). For comparison, the same analysis was performed using the DNase-chip data set (broken line); DNase-chip samples were hybridized to the same design of high-density oligonucleotide tiling array as was used for FAIRE. (*C*) A representation of the relationship between FAIRE peaks and other annotated features. Each row corresponds to one of the 571 FAIRE peaks that overlap with at least one of the following: a TSS (Ashurst et al. 2005; Harrow et al. 2006); union of DHS (Crawford et al. 2006; Sabo et al. 2006); 75th percentile of promoter activity (Trinklein et al. 2003; Cooper et al. 2006); RNAP ChIP–chip; or TAF1 ChIP–chip (Kim et al. 2005a,b). A black bar represents overlap with the FAIRE signal, whereas white represents no overlap ("overlap" defined in Table 1 legend). Not shown are the 437 FAIRE peaks that do not overlap with any of these marks. Data were clustered for display (Eisen et al. 1998). (*D*) qPCR validation of the microarray data was performed over three 8-kb regions. The height of the bars from the qPCR analysis represents the enrichment of the FAIRE samples relative to the uncrosslinked reference; the FAIRE data and peaks are the same as described in Figure 2. A representative region corresponding to chromosome 21 coordinates 32,813,792–32,820,968 is shown. Note that this region contains no annotated genes and that these were "orphan" FAIRE peaks, unassigned to any other active chromatin mark.

Orphan FAIRE peaks

A substantial fraction of FAIRE peaks do not correspond to any of the annotations selected for comparison (Table 1). This is not simply a consequence of using relaxed criteria for defining FAIRE peaks, since more stringent peak definitions do not substantially increase the percentage of FAIRE peaks that overlap with the selected marks (data not shown). Furthermore, a number of orphan FAIRE peaks were reproducibly isolated and verified by qPCR. Rather, a number of factors unrelated to the FAIRE procedure itself are likely to contribute to the appearance of orphan FAIRE signals, including: (1) The data used for comparison were derived from different cell lines. As more ChIP–chip data become available in additional human cell lines (or if a superset of data from all cell types were available), the number of FAIRE peaks assigned to other active marks will expand significantly. (2) It is certain that current annotations represent only a fraction of the activities encoded by the human genome (Margulies et al. 2006) and are heavily biased toward those associated with transcription. For example, 48% of the FAIRE peaks shown in Figure 3C are coincident with a DNaseI hypersensitivity peak but none of the other marks of transcriptional activity. These regions may correspond to an unannotated genomic activity. (3) The marks selected for comparison with FAIRE are not likely to fully encompass even a single category (transcription) of genomic activity. For example, in the alpha- and beta-globin locus control regions, which would not necessarily be represented in any of the categories used for comparison, distinct FAIRE peaks exist at the HS40 and HS2 enhancer elements, respectively (data not shown). Finally, (4) FAIRE may detect regions that correspond to hallmarks of genomic activity that are not captured by traditional nuclease sensitivity assays or the currently available ChIP–chip data. Future studies will be required to determine what other genomic activities are associated with FAIRE and the extent to which data from additional cell lines link FAIRE to other active marks.

Conclusion

We have presented evidence that FAIRE is capable of isolating nucleosome-

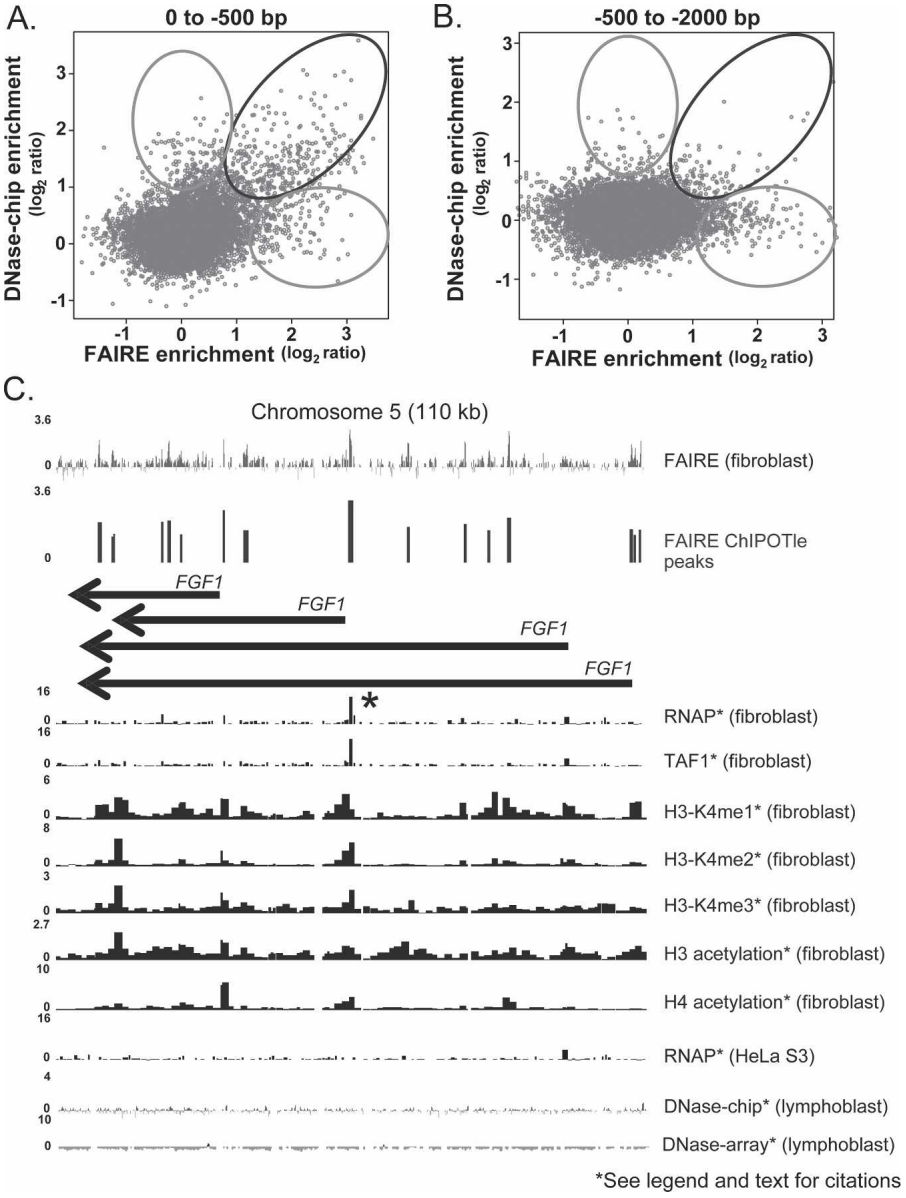

Figure 4. Cell-type specific differences identified by FAIRE. (*A*) A scatterplot of the $\log_2$ values for individual 50-mer probes from the DNase-chip (Crawford et al. 2006) and FAIRE data sets that mapped between 0 and 500 bp upstream of a GENCODE TSS (Harrow et al. 2006) are plotted. The black oval indicates probes that had high enrichment values in both data sets, whereas the gray ovals indicate probes with enrichment values that were high in only one of the data sets. (*B*) Same as *A*, but probes that mapped from 500 to 2000 bp upstream of a GENCODE TSS are plotted. (*C*) The fibroblast growth factor 1 (*FGF1*) gene, which has several annotated TSSs, exhibits extensive FAIRE signal (performed in fibroblast cells) but no detectable DNaseI signal (performed in lymphoblastoid cells). The asterisk indicates the presence of RNAP and TAF1 ChIP signal over this region in lung fibroblast (IMR90) cells (Kim et al. 2005a,b). The units of data for each track are described in Figure 2. The region shown corresponds to chromosome 5 coordinates 141,950,000 to 142,060,000.

depleted DNA, a hallmark of active regulatory elements, from human chromatin. Genome-wide maps of DNA accessibility will allow a better understanding of how the availability of sequence-based regulatory elements is coordinated with the regulation of factors that utilize them in a given cellular environment. Understanding this relationship will be critical to constructing realistic models of gene regulation in eukaryotic cells.

Methods

FAIRE procedure

Four independent cultures (biological replicates) of human foreskin fibroblast (ATCC CRL 2091) cells were grown in 245 × 245-mm plates to 90% confluence. Formaldehyde was added directly to the plates at room temperature (22–25°C) to a final concentration of 1% and incubated for 1, 2, 4, or 7 min, respectively. Glycine was added to a final concentration of 125 mM for 5 min at room temperature to quench the formaldehyde. Cells were rinsed with phosphate buffered saline containing phenylmethylsulphonylfluoride, and the plate was scraped and rinsed two more times. The cells were spun at 2000 rpm for 4 min and snap frozen. Cells were resuspended in 1 mL of lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-Cl at pH 8.0, 1 mM EDTA) per 0.4 g of cells and lysed using glass bead disruption for five 1-min sessions with 2-min incubations on ice between sessions. Samples were then sonicated for five sessions of sixty pulses (1 sec on/1 sec off) using a Branson Sonifier at 15% amplitude. Cellular debris was cleared by spinning at 15,000 rcf for 5 min at 4°C.

DNA was isolated by adding an equal volume of phenol-chloroform (Sigma #P3803 phenol, chloroform, and isoamyl alcohol 25:24:1 saturated with 10 mM Tris at pH 8.0, 1 mM EDTA), vortexing, and spinning at 15,000 rpm for 5 min at 4°C. The aqueous phase was isolated and stored in a separate tube. An additional 500 µl of TE was added to the organic phase, vortexed, and spun again at 15,000 rpm for 5 min at 4°C. The aqueous phase was isolated and combined with the first aqueous fraction, and a final phenol-chloroform extraction was performed on the pooled aqueous fractions to ensure that all protein was removed. The DNA was precipitated by addition of sodium acetate to 0.3 M, glycogen to 20 µg/mL, and two times the volume of 95% ethanol, and incubated at −20°C overnight. The precipitate was spun at 15,000 rpm for 10 min at 4°C, then the pellet was washed with 70% ethanol and dried in a Speed-Vac. The pellet was resuspended in dH_2O and treated with RNase A (100 µg/mL) and incubated at 37°C for 2 h. Crosslinked samples were incubated at 65°C overnight to ensure that any DNA–DNA crosslinks did not interfere with downstream enzymatic steps.

Sample amplification, labeling, hybridization, and quantitation

Samples were amplified using ligation-mediated PCR (Ren et al. 2000). Briefly, DNA fragments in a sample from each time-point

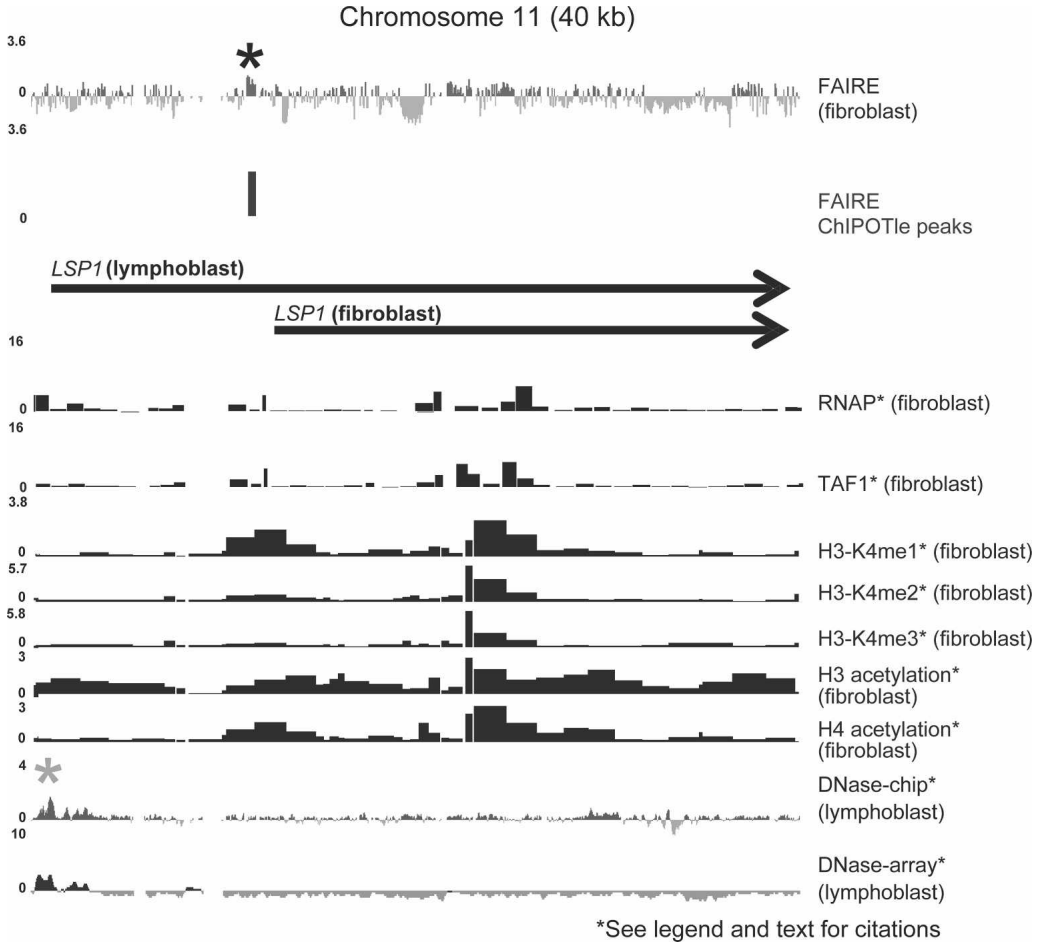

Figure 5. Tissue-specific accessibility of the *LSP1* promoter at alternative TSSs FAIRE from fibroblasts and the DNaseI hypersensitivity data (Crawford et al. 2006; Sabo et al. 2006) from lymphoblastoid cells correspond to alternative, tissue-specific promoter usage at the *LSP1* gene. On the *top* track, an asterisk marks the peak in the raw FAIRE data that corresponds to the TSS shown to be active in fibroblast cells. Data corresponding to RNAP, TAF1, and the histone modifications from adult and embryonic lung fibroblast cells are shown in the tracks *below* (Kim et al. 2005a,b; Koch et al. 2007). These tracks are also consistent with the utilization of this TSS in fibroblast cells. The *bottom* two tracks show DNaseI hypersensitivity results from lymphoblast cells, with a peak that corresponds only to the TSS for the lymphoblast-specific transcript (gray asterisk). An unannotated TSS about 10 kb downstream of the second TSS is suggested by the FAIRE signal (*upper* track, just below the 10^{-25} cutoff for peak detection) and the strong ChIP–chip signals. The units of data for each track are described in Figure 2. The region shown corresponds to chromosome 11 coordinates 1,830,000 to 1,870,000.

were made blunt using T4 DNA polymerase. Asymmetric linkers (5′-GCGGTGACCCGGGAGATCTGAATTC-3′ and 5′-GAATTCAGATC-3′) were ligated to the blunt ends, and the samples were amplified by PCR with a primer complementary to the linker.

Sample labeling and hybridization were performed at NimbleGen Systems, Inc. Samples were labeled by incorporation of cyanine dyes by polymerization with Klenow fragment primed by random nonomers. FAIRE samples were labeled with Cy5, and genomic DNA (to be used as a reference) was labeled with Cy3. The labeled samples were mixed and hybridized to high-density oligonucleotide microarrays tiling the ENCODE regions (NimbleGen Systems, Inc.). The microarray contains ~385,000 50-mer probes, sharing 6 bp with each of the adjacent probes, allowing measurements at 38-bp resolution across the nonrepetitive sequence in the ENCODE regions. Hybridizations were performed in a MAUI hybridization station for 16 h at 42°C. Arrays were washed and scanned with an Axon Scanner 4000B. Spot intensities were quantitated using GenePix software and normalized by NimbleGen's in-house software. Data from all four crosslinking times, which were prepared from four independent biological samples, were averaged for all analyses.

qPCR validation

Portions of three ENCODE regions were selected for validation: chr8:119189349–119195557, chr21:32,813,792–32,820,968, and chr7:26,978,053–26,987,656. Ninety-six primer pairs were designed for qPCR and divided between the three regions, spaced as evenly apart as possible. DNA used in the qPCR validation was obtained independently using an identical protocol and cell line as for the microarray analysis. PCR was performed using SYBR green chemistry on an ABI 7900 instrument. Relative enrichment of each amplicon in the FAIRE-treated DNA was calculated using the comparative cT method (Livak and Schmittgen 2001). DNA from untreated fibroblast cells served as the control for the calculations.

Data analysis

The signal generated by FAIRE is similar to that generated by a conventional ChIP–chip experiment. Therefore, we used the

peak-finding algorithm ChIPOTle (Buck et al. 2005) (http://sourceforge.net/projects/chipotle-perl/) to identify regions isolated with FAIRE. Briefly, ChIPOTle uses a sliding window (300 bp) to identify statistically significant signals that comprise a peak. The null distribution is determined by reflecting the negative data from the region of interest about zero and fitting a Gaussian distribution. For the analysis presented, values calculated from the average of four FAIRE experiments were input to ChIPOTle. Displayed peaks correspond to a *P*-value of $<10^{-25}$, after using the Benjamini-Hochberg correction to adjust for multiple tests (Benjamini and Hochberg 1995). All of the feature sets used for comparison with FAIRE peaks were downloaded from the UCSC Genome Browser. For the DNase-chip data, we excluded peaks found in only one of the three DNase concentrations reported (Crawford et al. 2006).

For visualization, data were loaded to the UCSC Genome Browser (Hinrichs et al. 2006). Genomic annotations including TSSs were produced by the GENCODE project (Ashurst et al. 2005; Harrow et al. 2006), whose goal is to provide high-quality annotation of all protein-coding DNA sequences that have been experimentally verified. All coordinates reported are based on human genome sequence release "hg17" (NCBI build 35). Each annotation track presented is available for download, along with the raw FAIRE data for each microarray (ftp://hgdownload.cse.ucsc.edu/goldenPath/hg17/encode/datafiles/UncFaire/). The FAIRE data are also available from GEO (GSM109841, GSM109842, GSM109843, GSM109844, and series GSE4886).

Acknowledgments

We thank the ENCODE Project Consortium for making their data publicly available. We especially thank Roland Green and Mike Singer at NimbleGen Systems, Inc., for performing labeling and microarray hybridization. We thank Greg Crawford, Francis Collins, Nathan Trinklein, Rick Myers, and Ian Dunham for allowing use of unpublished data for comparison with FAIRE. This work was supported by ENCODE technology development grant HG3532-2 from the National Human Genome Research Institute.

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. 2002. *Molecular Biology of the Cell*. Garland Science, New York.

Ashurst, J.L., Chen, C.K., Gilbert, J.G., Jekosch, K., Keenan, S., Meidl, P., Searle, S.M., Stalker, J., Storey, R., Trevanion, S., et al. 2005. The Vertebrate Genome Annotation (Vega) database. *Nucleic Acids Res.* **33:** D459–D465.

Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57:** 289–300.

Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., and Schreiber, S.L. 2004. Global nucleosome occupancy in yeast. *Genome Biol.* **5:** R62.

Boeger, H., Griesenbeck, J., Strattan, J.S., and Kornberg, R.D. 2003. Nucleosomes unfold completely at a transcriptionally active promoter. *Mol. Cell* **11:** 1587–1598.

Brutlag, D., Schlehuber, C., and Bonner, J. 1969. Properties of formaldehyde-treated nucleohistone. *Biochemistry* **8:** 3214–3218.

Buck, M.J., Nobel, A.B., and Lieb, J.D. 2005. ChIPOTle: A user-friendly tool for the analysis of ChIP-chip data. *Genome Biol.* **6:** R97.

Cooper, S.J., Trinklein, N.D., Anton, E.D., Nguyen, L., and Myers, R.M. 2006. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.* **16:** 1–10.

Crawford, G.E., Holt, I.E., Mullikin, J.C., Tai, D., Blakesley, R., Bouffard, G., Young, A., Masiello, C., Green, E.D., Wolfsberg, T.G., et al. 2004. Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc. Natl. Acad. Sci.* **101:** 992–997.

Crawford, G.E., Davis, S., Scacheri, P.C., Renaud, G., Mohamad, J.H., Erdos, M.R., Green, R.D., Meltzer, P.S., Wolfsberg, T.G., and Collins, F.S. 2006. DNase-chip: A high resolution method to identify DNaseI hypersensitive sites using tiled microarrays. *Nat. Methods* **3:** 503–509.

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. 2006. A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci.* **103:** 5320–5325.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.* **95:** 14863–14868.

The ENCODE Project Consortium. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306:** 636–640.

Gimble, J.M., Dorheim, M.A., Youkhana, K., Hudson, J., Nead, M., Gilly, M., Wood Jr., W.J., Hermanson, G.G., Kuehl, M., Wall, R., et al. 1993. Alternatively spliced pp52 mRNA in nonlymphoid stromal cells. *J. Immunol.* **150:** 115–121.

Harrow, J., Denoeud, F., Frankish, A., Reymond, A., Chen, C.K., Chrast, J., Lagarde, J., Gilbert, J.G., Storey, R., Swarbreck, D., et al. 2006. GENCODE: Producing a reference annotation for ENCODE. *Genome Biol.* S4.1–9.

Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., et al. 2006. The UCSC Genome Browser Database: Update 2006. *Nucleic Acids Res.* **34:** D590–D598.

Hogan, G.J., Lee, C.K., and Lieb, J.D. 2006. Cell-cycle specified fluctuation of nucleosome occupancy at gene promoters. *PLoS Genet.* **2:** e158.

Hurowitz, E.H. and Brown, P.O. 2003. Genome-wide analysis of mRNA lengths in *Saccharomyces cerevisiae*. *Genome Biol.* **5:** R2.

Keene, M.A. and Elgin, S.C. 1981. Micrococcal nuclease as a probe of DNA sequence organization and chromatin structure. *Cell* **27:** 57–64.

Kim, T.H., Barrera, L.O., Qu, C., Van Calcar, S., Trinklein, N.D., Cooper, S.J., Luna, R.M., Glass, C.K., Rosenfeld, M.G., Myers, R.M., et al. 2005a. Direct isolation and identification of promoters in the human genome. *Genome Res.* **15:** 830–839.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. 2005b. A high-resolution map of active promoters in the human genome. *Nature* **436:** 876–880.

Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaöz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P., et al. 2007. The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res.* (this issue) doi: 10.1101/gr.5704207.

Lee, C.K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* **36:** 900–905.

Levy, A. and Noll, M. 1981. Chromatin fine structure of active and repressed genes. *Nature* **289:** 198–203.

Livak, K.J. and Schmittgen, T.D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Δ ΔC(T)) method. *Methods* **25:** 402–408.

Margulies, E.H., Chen, C.W., and Green, E.D. 2006. Differences between pair-wise and multi-sequence alignment methods affect vertebrate genome comparisons. *Trends Genet.* **22:** 187–193.

Misener, V.L., Hui, C., Malapitan, I.A., Ittel, M.E., Joyner, A.L., and Jongstra, J. 1994. Expression of mouse LSP1/S37 isoforms. S37 is expressed in embryonic mesenchymal cells. *J. Cell Sci.* **107:** 3591–3600.

Nagy, P.L., Cleary, M.L., Brown, P.O., and Lieb, J.D. 2003. Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. *Proc. Natl. Acad. Sci.* **100:** 6364–6369.

Polach, K.J. and Widom, J. 1995. Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation. *J. Mol. Biol.* **254:** 130–149.

Rao, B., Shibata, Y., Strahl, B.D., and Lieb, J.D. 2005. Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Mol. Cell. Biol.* **25:** 9447–9459.

Reinke, H. and Horz, W. 2003. Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Mol. Cell* **11:** 1599–1607.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. 2000. Genome-wide location and function of DNA binding proteins. *Science* **290:** 2306–2309.

Sabo, P.J., Kuehn, M.S., Thurman, R., Johnson, B.E., Johnson, E.M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., et al. 2006. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat. Methods* **3:** 511–518.

Solomon, M.J. and Varshavsky, A. 1985. Formaldehyde-mediated DNA-protein crosslinking: A probe for in vivo chromatin structures. *Proc. Natl. Acad. Sci.* **82:** 6470–6474.

Sudarsanam, P. and Winston, F. 2000. The Swi/Snf family

nucleosome-remodeling complexes and transcriptional control. *Trends Genet.* **16:** 345–351.
Thompson, A.A., Omori, S.A., Gilly, M.J., May, W., Gordon, M.S., Wood Jr., W.J., Miyoshi, E., Malone, C.S., Gimble, J., Denny, C.T., et al. 1996. Alternatively spliced exons encode the tissue-specific 5′ termini of leukocyte pp52 and stromal cell S37 mRNA isoforms. *Genomics* **32:** 352–357.
Trinklein, N.D., Aldred, S.J., Saldanha, A.J., and Myers, R.M. 2003. Identification and functional analysis of human transcriptional promoters. *Genome Res.* **13:** 308–312.
Tsukiyama, T. and Wu, C. 1995. Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83:** 1011–1020.
Varga-Weisz, P. 2001. ATP-dependent chromatin remodeling factors: Nucleosome shufflers with many missions. *Oncogene* **20:** 3076–3085.
Wallrath, L.L., Lu, Q., Granok, H., and Elgin, S.C. 1994. Architectural variations of inducible eukaryotic promoters: Preset and remodeling chromatin structures. *Bioessays* **16:** 165–170.
Wong, G.K., Passey, D.A., and Yu, J. 2001. Most of the human genome is transcribed. *Genome Res.* **11:** 1975–1977.
Wu, C. 1980. The 5′ ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. *Nature* **286:** 854–860.
Wu, C., Wong, Y.C., and Elgin, S.C. 1979. The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. *Cell* **16:** 807–814.
Yu, L. and Morse, R.H. 1999. Chromatin opening and transactivator potentiation by RAP1 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19:** 5279–5288.
Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. 2005. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309:** 626–630.

Received May 21, 2006; accepted in revised form August 15, 2006.