Genome-scale mapping of DNase I sensitivity *in vivo* using tiling DNA microarrays

Peter J Sabo[1,6], Michael S Kuehn[1,2,6], Robert Thurman[1,2], Brett E Johnson[2], Ericka M Johnson[2], Hua Cao[2], Man Yu[2], Elizabeth Rosenzweig[2], Jeff Goldy[1], Andrew Haydock[1], Molly Weaver[1], Anthony Shafer[1], Kristin Lee[1], Fidencio Neri[1], Richard Humbert[1], Michael A Singer[3], Todd A Richmond[3], Michael O Dorschner[1], Michael McArthur[4], Michael Hawrylycz[5], Roland D Green[3], Patrick A Navas[2], William S Noble[1] & John A Stamatoyannopoulos[1]

Localized accessibility of critical DNA sequences to the regulatory machinery is a key requirement for regulation of human genes. Here we describe a high-resolution, genome-scale approach for quantifying chromatin accessibility by measuring DNase I sensitivity as a continuous function of genome position using tiling DNA microarrays (DNase-array). We demonstrate this approach across 1% (~30 Mb) of the human genome, wherein we localized 2,690 classical DNase I hypersensitive sites with high sensitivity and specificity, and also mapped larger-scale patterns of chromatin architecture. DNase I hypersensitive sites exhibit marked aggregation around transcriptional start sites (TSSs), though the majority mark nonpromoter functional elements. We also developed a computational approach for visualizing higher-order features of chromatin structure. This revealed that human chromatin organization is dominated by large (100–500 kb) 'superclusters' of DNase I hypersensitive sites, which encompass both gene-rich and gene-poor regions. DNase-array is a powerful and straightforward approach for systematic exposition of the *cis*-regulatory architecture of complex genomes.

The functional landscape of the human genome *in vivo* is reflected in the organization and topology of nuclear chromatin[1,2]. The finding, over 25 years ago[3–5], that active *cis*-regulatory sequences represent foci of heightened chromatin accessibility has been widely exploited to provide a window into the *cis*-regulatory environment of individual mammalian genes and selected loci. Enhanced chromatin accessibility, classically assayed by hypersensitivity to DNase I *in vivo*, is the *sine qua non* of activated mammalian *cis*-regulatory sequences including promoters, enhancers, insulators, boundary elements and locus control regions[6,7].

A major challenge facing modern human genomics is systematic identification of these transcriptional control elements over the entire genome and analysis of their relationship to the current annotation of human genes. Comprehensive delineation of the accessible chromatin compartment is expected to be of particular importance for identification of functional human genetic variants that mediate individual variation in gene expression and physiological phenotypes. On a broader level, human chromosomes have long been thought to be organized into discrete higher-order functional domains characterized by 'open' (active) and 'closed' (inactive) chromatin[8–10]. Experimental verification of this concept on a large scale, however, has been difficult using available technologies.

To address these challenges, we developed a method, DNase-array, capable of measuring chromatin accessibility *in vivo* at high resolution on a chromosomal or even genomic scale in a single experiment. We demonstrate this method on human B-lymphoblastoid cells, in which we comprehensively map DNase I hypersensitive sites across 1% of the human genome with very high sensitivity and specificity.

RESULTS

High-resolution mapping of accessible chromatin

Our approach, named DNase-array (**Fig. 1**), exploits the well-established phenomenon of chromatin-specific sensitivity to DNase I *in vivo*, and capitalizes on the continuously expanding potential of DNA microarrays to effect massively parallel quantification of genomic DNA samples.

The hallmark of increased chromatin accessibility is the occurrence of multiple cleavage events over a short distance on the same linear nuclear chromatin template when intact nuclei are exposed to the non-specific endonuclease DNase I. To assay this phenomenon in human lymphoblastoid cells (GM06990; Coriell), we devised an approach for isolating genomic DNA fragments released by two cleavage 'hits' occurring close to each other ($< \sim$1,200 bp)

[1]Department of Genome Sciences, University of Washington, 1705 NE Pacific St., Box 357730, Seattle, Washington 98195, USA. [2]Division of Medical Genetics, Department of Medicine, University of Washington, 1705 NE Pacific St., Box 357730, Seattle, Washington 98195, USA. [3]Nimblegen Systems, Inc., 1 Science Court, Madison, Wisconsin 53711, USA. [4]Department of Microbiology, John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH, UK. [5]Allen Institute for Brain Sciences, 551 N. 34th Street, Seattle, Washington 98103, USA. [6]These authors contributed equally to this work. Correspondence should be addressed to J.A.S. (jstam@u.washington.edu).

RECEIVED 12 APRIL; ACCEPTED 22 MAY; PUBLISHED ONLINE 21 JUNE 2006; DOI:10.1038/NMETH890

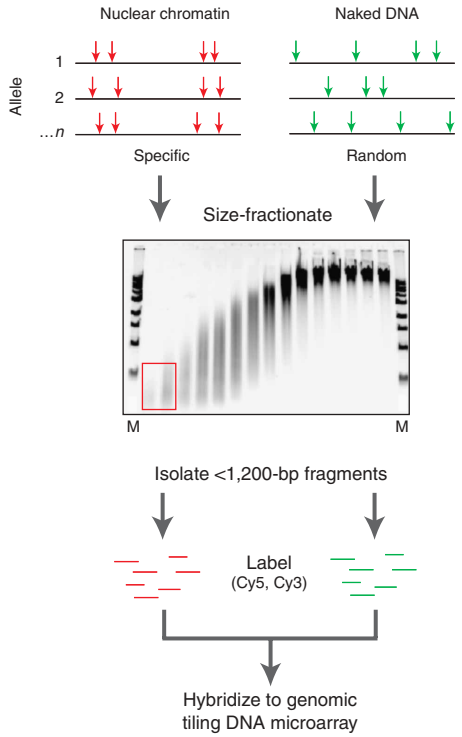

Figure 1 | Approach for high-resolution mapping of accessible chromatin in human cells using DNase-array. DNase I access to individual chromatin templates is captured with high specificity by isolating DNA fragments created by two DNase I–cutting 'hits' occurring in close proximity (<1,200 bp). These short fragments are isolated by size fractionation on a sucrose gradient (the red box marks lower-density fractions). To control for the possibility of intrinsic sequence preference, equally sized control fragments are isolated from DNase I–treated naked DNA. Chromatin- and nonchromatin–derived fragments are differentially labeled and hybridized to a genomic tiling DNA microarray. The resulting signal reflects chromatin accessibility ($\log_2$ relative DNase I sensitivity) as a function of genomic position. M, marker.

when intact nuclei are treated with DNase I. Consideration of only 'two-hit' DNase I cutting events in the genome has the potential to impart great specificity, compared with analysis of individual free DNA ends, as the latter cannot be reliably distinguished *a priori* from random shear sites. This approach also maintains a 1:1 relationship between each DNase I–released fragment and each chromatin template. We labeled these chromatin-specific fragments with Cy5 and labeled similarly sized control fragments (created by DNase I–treatment of naked genomic DNA) with Cy3. To visualize the chromatin accessibility pattern as a function of genomic position at high resolution, we hybridized the mixture to a high-density oligonucleotide DNA microarray comprising 390,000 50-mer oligonucleotides consecutively tiled with a 12-bp overlap (net resolution of 38 bp) over nonrepetitive regions of 44 genomic loci (the ENCODE regions[11]), which collectively span ~30 Mb, or 1% of the human genome. For each genomic position interrogated by the microarray, we quantified chromatin accessibility by computing the ratio of the resulting signals, expressed as the ratio of chromatin-specific (Cy5) versus nonspecific (Cy3) DNase I activity. This provided a continuous quantitative picture of chromatin accessibility as a function of genome position. Exemplary results from a 1-Mb region of chromosome 5 containing the Th2 cytokine gene cluster are shown in **Figure 2a**.

DNase I hypersensitive sites result from binding of *trans*-acting factors in place of a canonical nucleosome, with consequent alteration of local chromatin structure resulting in increased accessibility of both the core functional element and the flanking regions[1–6,12]. Previously, we showed that DNase I hypersensitive sites could be distinguished reliably as statistical outliers against a continuous trend of DNase I sensitivity along the genome[13]. We identified DNase I hypersensitive sites by resolving regions containing at least four contiguous microarray probes with significant enrichment over background (empirical $P < 0.01$, corresponding to signal ratio >2.1-fold over background), and then applying a peak-finding algorithm to delineate the sub-region(s) with maximum accessibility (see Methods). Using this approach, we localized 2,690 DNase I hypersensitive sites within the 30-Mb experimental regions, and classified these into 1,191 'major' sites defined by highly significant enrichment ($P < 0.001$, enrichment >6.2-fold over background), and 1,499 less intense 'minor' DNase I hypersensitive sites (**Supplementary Table 1** online). We observed wide dynamic range in DNase I sensitivity (>200-fold) between the least-intense and most-intense genomic regions, indicative of marked diversity in nuclear chromatin microenvironments.

To assess reproducibility, we compared results from independent DNase-array experiments performed more than two months apart using chromatin preparations derived from separate cell cultures expanded from frozen cell pellets (**Supplementary Fig. 1** online). Of the 1,191 major DNase I hypersensitive sites detected in the first sample, we identified 94% (1,113) as hypersensitive sites in the second sample (732 at the $P < 0.001$ level and 381 at the $P < 0.01$ level). It is possible that the difference of ~6% reflects natural biological variability between different chromatin preparations from independent cell expansions, nuclear harvests and DNase I treatments.

To determine whether DNase I hypersensitive sites identified by DNase-array correspond with classical DNase I hypersensitive sites, we performed extensive conventional DNase I hypersensitivity mapping in lymphoblast chromatin using the well-established Southern end-label strategy[3,6,12] (**Fig. 2b–g** and **Supplementary Fig. 2** online). Each conventional assay interrogates a restriction fragment within which DNase I hypersensitive sites appear as subbands on a Southern blot (for example, **Fig. 2b–g** and **Supplementary Fig. 1**). To ensure that the conventional assays were unbiased, we designed a restriction fragment tiling path to provide end-to-end coverage across two large genomic regions on chromosomes 5 and 11 (**Supplementary Table 2** online). We successfully performed and analyzed 151 conventional DNase I hypersensitivity assays, collectively interrogating ~836 kb of genomic terrain, and successfully collected data over large (≫100 kb) contiguous blocks of sequence (**Supplementary Table 2**). Critically, the expansive genomic coverage of these studies subsumed a substantial territory (~538 kb) in which no DNase I hypersensitive sites were detected (that is, true negative regions; **Supplementary Table 2**); such information is critical for correct determination of specificity. These studies collectively allowed us to define the proportion of true and false positives, and true and false negatives in the DNase-array data relative to an accepted gold standard (**Table 1** and **Supplementary Table 2**). Taken together, the conventional assays confirmed that the DNase-array method delineated classical DNase

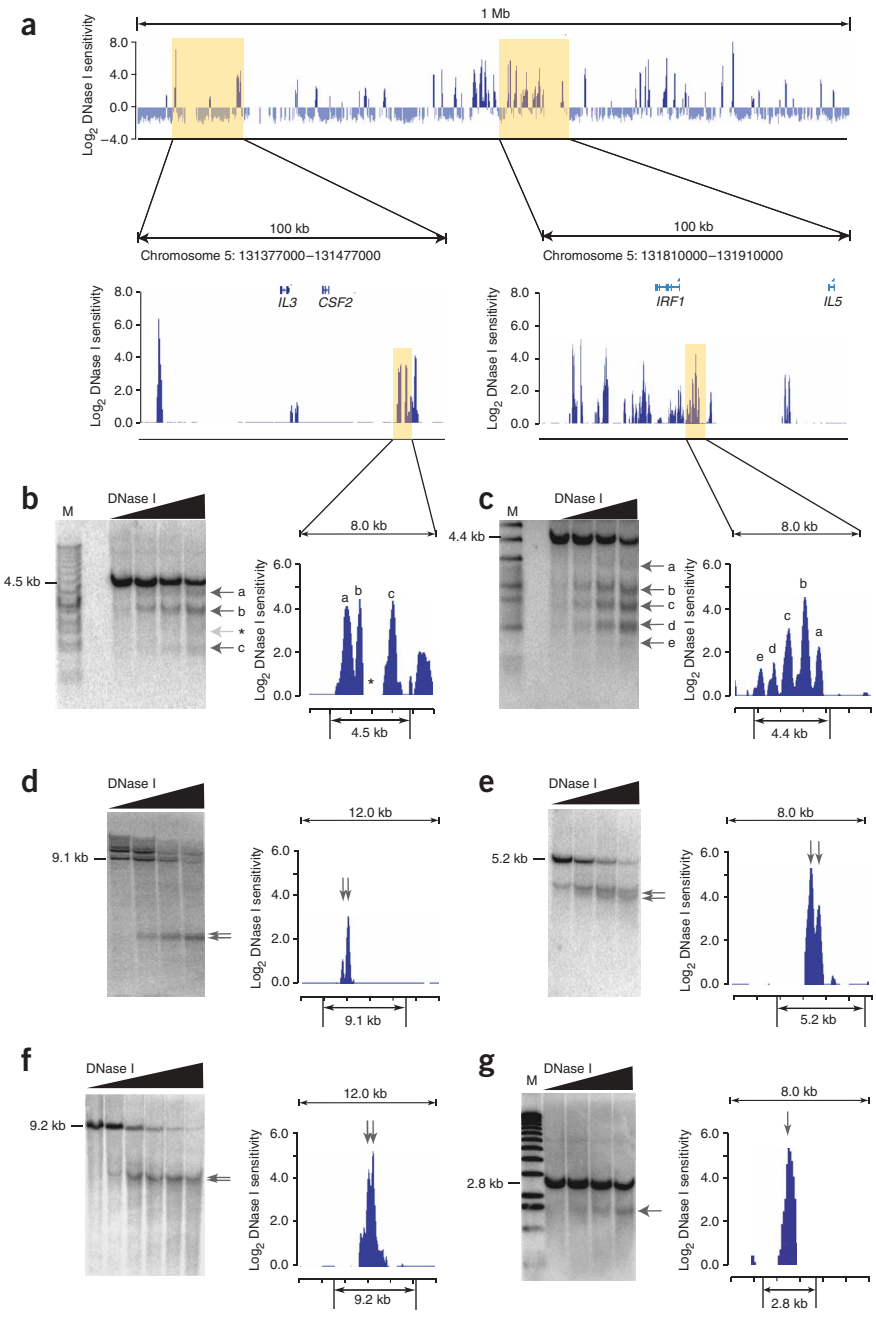

Figure 2 | Chromatin accessibility at both megabase and fine scale. (**a**) Chromatin accessibility (DNase I sensitivity) across a 1-Mb domain at chromosome 5: 131284314–132284313 (horizontal axis) and two 100-kb sub-domains. DNase I sensitivity is expressed as $\log_2$ over background, where $\log_2 = 0$ is scaled to a significance threshold of $P < 0.01$. (**b**–**g**) Comparison with conventional hypersensitivity experiments, which are visualized as a Southern blot showing the pattern of cleavage within a restriction fragment of indicated size with increasing (left to right) DNase I concentrations. DNase-array results for the corresponding region are shown to the right of each blot. M, marker. The following regions were examined: *IL3-CSF2* enhancer region (4.5-kb *Sac*I fragment; **b**), region upstream of the *IRF1* gene (4.4-kb *Sac*I fragment; **c**), chromosome 5: 132180000–132192000 (9.1-kb *Sap*I fragment; **d**), chromosome 11: 5658000–5666000 (5.2-kb *Bgl*II fragment; **e**), chromosome 11: 116143000–116155000 (9.2-kb *Eco*RV fragment; **f**), and chromosome 21: 33652000–33660000 (2.8-kb *Hind*III fragment; **g**). *, DNase I hypersensitive site on Southern blot falls into gap as observed by tiling microarray.

I hypersensitive sites with high sensitivity (91.7%), specificity (>99.5%), positive predictive value (89.3%) and low false positive rate (0.47%) over genomic distances (**Supplementary Table 2**). The actual sensitivity over unique genomic regions is likely to be higher because Southern end-label assays interrogate all genomic sequences within a probed restriction fragment, including repetitive elements (some of which harbor DNase I hypersensitive sites) that are not represented on the tiling microarray (for example, **Fig. 2b** and **Supplementary Fig. 2**). The estimate of positive predictive value is also conservative because it does not take into consideration the potential biological variability between different nuclear harvests and chromatin preparation protocols. The large tissue requirements of conventional hypersensitivity assays required pooling of nuclear samples harvested at log phase (versus from a single preparation of synchronized cells for DNase-array). Such pooling may have the effect of diluting weaker sites or even more prominent sites that appear variably between different preparations.

A unique additional example of functional validation is provided by the T helper-cell type 2 (Th2) cytokine locus on chromosome 5 (**Fig. 3**). The mouse homolog of this locus has been subjected to intensive investigation for *cis*-regulatory sequences, including comprehensive mapping of DNase I hypersensitive sites in mouse T-lymphoid cells and experimental studies to define the functional roles of these sites[14–17]. The human homologs of these elements remain largely undefined. To test whether mouse DNase I hypersensitive sites in this region were functionally conserved in the human, we identified the DNA sequences corresponding to 20 previously defined mouse DNase I hypersensitive sites and *cis*-regulatory elements and mapped these to the human genome. We found striking correspondence between DNase I hypersensitive sites mapped by DNase-array in human B-lymphoid cells and the orthologous position of mouse DNase I hypersensitive sites, and functional elements mapped in mouse T cells (**Fig. 3**). This revealed that all the major *cis*-regulatory elements including the Th2 locus control region[14,15], and long-range enhancer[16] and silencer[17] elements controlling the Th2 cytokine genes are functionally conserved as DNase I hypersensitive sites between two species. This observation reinforces the potential of using DNase-array to delineate critical functional elements controlling human genes.

We conclude that DNase-array accurately measures DNase I sensitivity over genomic distances at high resolution, and

Table 1 | Validation of DNase-array with conventional assays

Parameter	Result
Sensitivity	91.67%
Specificity	99.53%
Positive predictive value	89.34%
Negative predictive value	99.64%
False positive rate	0.47%

We assumed conventional Southern assays to represent a gold standard and used standard definitions. We computed sensitivity as the number of true positive DNase I hypersensitive sites (TP = DNase I hypersensitive sites detected by both DNase-array and conventional assays) divided by the number of true positives plus false negatives (sensitivity = TP / (TP + FN)) (**Supplementary Table 2** and **Supplementary Methods**). We computed the specificity as the number of corresponding non–DNase I hypersensitive site segments identified in both assays (= true negatives) divided by the number of true negatives plus false positives (specificity = TN / (TN + FP); see **Supplementary Methods**). Positive predictive value (PPV) was computed as the number of true positive DNase I hypersensitive sites divided by the number of true positives plus false positives (PPV = TP / (TP + FP)). Negative predictive value was computed analogously (NPV = TN / (TN + FN)). False positive rate (FPR) represents the probability that a true negative region will be classified incorrectly as positive (FPR = FP/ (FP + TN)); see **Supplementary Table 2** and **Supplementary Methods**).

reproducibly allows accurate and sensitive mapping of thousands of classical DNase I hypersensitive sites and associated functional elements in a single experiment.

Genomic distribution of DNase I hypersensitive sites

We next examined the global distribution of DNase I hypersensitive sites and associated *cis*-regulatory sequences with respect to genes and to one another. The genomic distribution of DNase I hypersensitive sites relative to known protein-coding genes is summarized in **Figure 4a**. Compared with random expectation, DNase I hypersensitive sites are enriched in introns and in regions proximal to TSSs and transcription termination sites (TTSs), and are depleted in distal intergenic regions (**Fig. 4b**). We also examined the distribution of DNase I hypersensitive sites relative to the TSS and TTS of known genes and of mRNA transcripts and spliced expressed sequence tags (ESTs; **Fig. 4c–g**). We found marked aggregation of DNase I hypersensitive sites in a nearly symmetrical distribution around the TSSs of known genes (**Fig. 4c**). We found 6.4% of lymphoid DNase I hypersensitive sites situated within the proximal promoter region (first 500 bp upstream of the TSS[18]) of known genes (8.1% of major DNase I hypersensitive sites, and 5.1% of minor DNase I hypersensitive sites). This suggests that only a small fraction of transcriptional regulatory information may be encoded within currently defined proximal promoter regions of known genes. If a larger region from –2,500 bp upstream to +2,500 bp downstream of the TSS is considered, 29.8% of DNase I hypersensitive sites are encompassed (35% major, 25.7% minor). This suggests that the majority of classical *cis*-regulatory sequences associated with DNase I hypersensitive sites are long-range regulatory elements[19], with ~50% expected to be located >10 kb away from the nearest TSS of known genes (**Fig. 4c**).

We observed increased density of DNase I hypersensitive sites in the region immediately 3′ of gene TTSs, a region from which antisense transcripts may originate[20], and which may be bound *in vivo* by transcription factors typically associated with promoters[21]. This finding suggests that a subset of DNase I hypersensitive sites may be involved either in the process of transcription termination, or in the origination and regulation of antisense transcripts[22].

We also considered a broader range of transcript data, including databases of human mRNAs and spliced ESTs[23]. We found the proportions of DNase I hypersensitive sites juxtaposed with either the 5′ or the 3′ ends of mRNAs and spliced ESTs to be substantially greater than for known genes (**Fig. 4c–g**). Given that many mRNAs and spliced ESTs are incomplete (that is, lack a true 5′ end), the actual correspondence with DNase I hypersensitive sites is likely to be even higher than observed. Notably, the increasing proportion of 5′-proximal elements observed with respect to spliced ESTs was paralleled by a decrease in elements that were far removed (>25 kb) from the TSSs of known genes (**Fig. 4c,e**). Given that the human transcript map is far from complete, it may be the case that truly isolated distal DNase I hypersensitive sites and *cis*-regulatory elements are substantially less common than suggested by their relationship to known genes.

Marked short-range clustering of DNase I hypersensitive sites

We next examined the short-range distribution of DNase I hypersensitive sites relative to one another and found evidence for marked intrinsic clustering along the genome. A surprisingly high proportion (80.1%) of major DNase I hypersensitive sites are located within 2.5 kb of another major DNase I hypersensitive site. This relationship is found across all regions examined, and is not dependent on location with respect to TSSs or TTSs as it is equally evident for promoter DNase I hypersensitive sites and for those located >20 kb away from any annotated or putative TSS. Approximately 10% of DNase I hypersensitive sites occur as isolated singletons, >10 kb from the nearest DNase I hypersensitive site. The co-occurrence of DNase I hypersensitive sites encoding *cis*-regulatory elements has been observed anecdotally in the context of several major mammalian gene regulatory systems, most notably locus control regions[7]. The present data suggest that such short-range cluster formations may be expected as a regular feature of the genome.

Visualization of higher-order chromatin features

Next we asked whether we could detect higher-order organizational features in the lymphoid chromatin accessibility profile. We adapted wavelet analysis[24], a mathematical tool pioneered in the

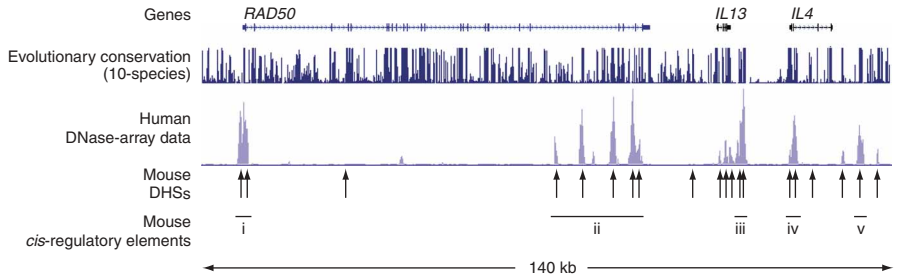

Figure 3 | Regulatory elements of the Th2 cytokine cluster. Shown is a 140-kb region from the Th2 cytokine cluster on Chr5 containing three genes (*RAD50*, *IL13* and *IL4*). Shown are genes (top), evolutionary conservation[23], human lymphoblastoid DNase-array results and locations of DNase I hypersensitive sites (DHSs) mapped in mouse lymphoid cells (bottom). Clearly defined are the *RAD50* promoter (i), the human homologues of the mouse Th2 locus control region[14,15] (ii), the long-range enhancer located 3′ to *IL13* (ref. 16; iii), the *IL4* promoter and proximal intronic enhancer (iv), and a recently defined long-range silencer element[17] (v).

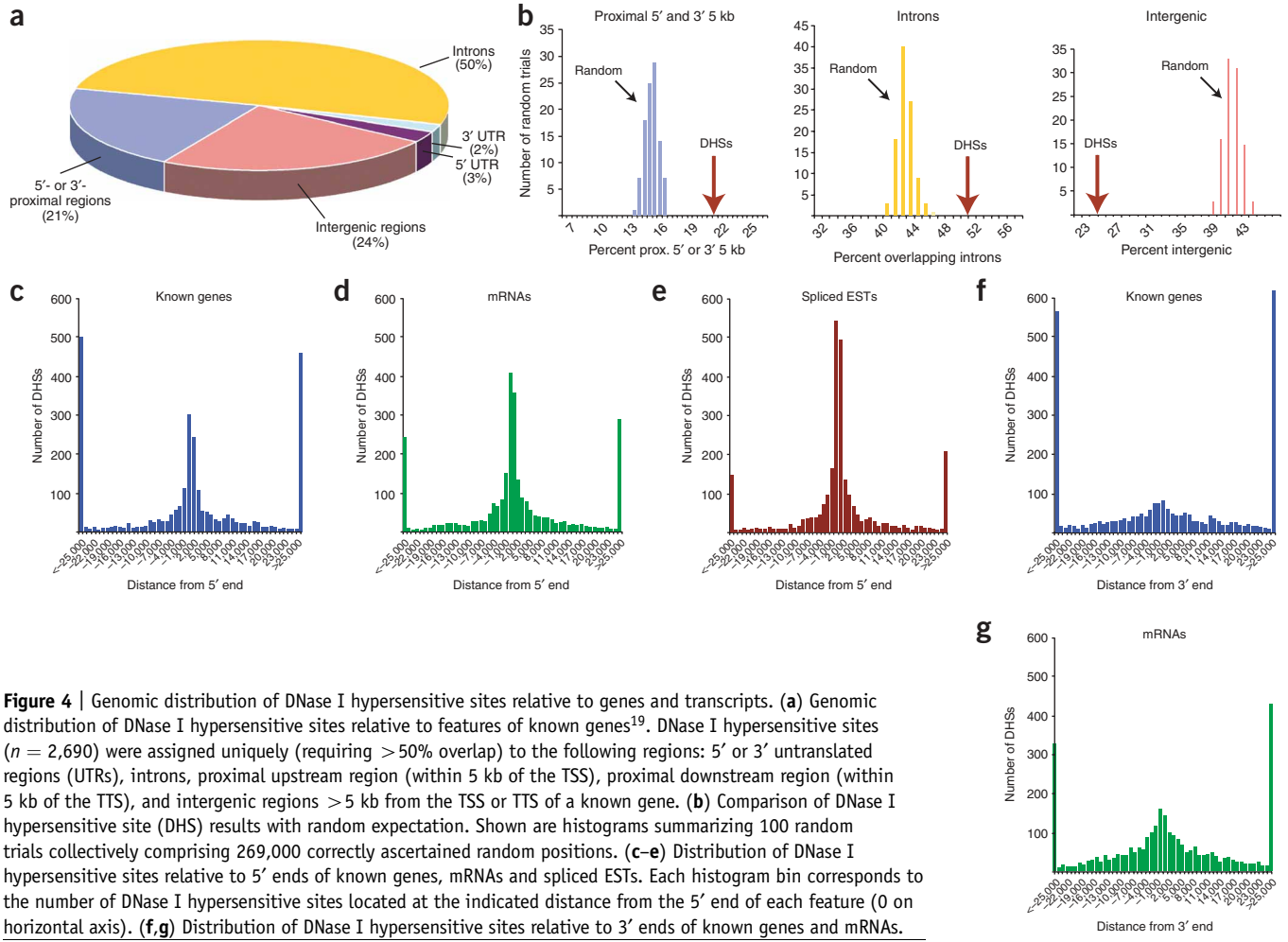

Figure 4 | Genomic distribution of DNase I hypersensitive sites relative to genes and transcripts. (**a**) Genomic distribution of DNase I hypersensitive sites relative to features of known genes[19]. DNase I hypersensitive sites ($n = 2{,}690$) were assigned uniquely (requiring >50% overlap) to the following regions: 5′ or 3′ untranslated regions (UTRs), introns, proximal upstream region (within 5 kb of the TSS), proximal downstream region (within 5 kb of the TTS), and intergenic regions >5 kb from the TSS or TTS of a known gene. (**b**) Comparison of DNase I hypersensitive site (DHS) results with random expectation. Shown are histograms summarizing 100 random trials collectively comprising 269,000 correctly ascertained random positions. (**c**–**e**) Distribution of DNase I hypersensitive sites relative to 5′ ends of known genes, mRNAs and spliced ESTs. Each histogram bin corresponds to the number of DNase I hypersensitive sites located at the indicated distance from the 5′ end of each feature (0 on horizontal axis). (**f**,**g**) Distribution of DNase I hypersensitive sites relative to 3′ ends of known genes and mRNAs.

field of signal processing, to quantify the rate of change in DNase I sensitivity as a function of both genomic scale (~200 bp to ~200 kb) and position along the chromosome. Wavelets enable decomposition of a given data type into increasingly coarse scales, allowing broader and broader trends in the data to reveal themselves[24].

We encoded continuous DNase I sensitivity measurements in wavelet coefficients and represented these in the form of a continuous wavelet transform heatmap in which the color scale denotes the strength of the wavelet coefficients at a particular position and scale (**Fig. 5**). Such heatmaps represent, for a given genomic position and analysis scale, the degree to which the average value of the DNase I signal, measured over the scale interval, is changing at that genomic position (**Fig. 5**). Small accessible chromatin features (for example, individual DNase I hypersensitive sites) will become evident at their physiological scale (~200–400 bp), whereas larger chromatin features will emerge only at higher scales (for example, ~25–50 kb).

An exemplary DNase I sensitivity wavelet transform heatmap for a 1.7-Mb territory on chromosome 21 within the Down Syndrome critical region[25] is illustrated in **Figure 5a**. It reveals multiple sharply demarcated large (~150–500 kb) active and inactive chromatin territories. The former correspond largely to regions containing 'superclusters' of DNase I hypersensitive sites, and the latter to regions characterized by a very low density or absence of DNase I hypersensitive sites over long intervals (~100–400 kb), and absence of signal at higher scales. DNase I hypersensitive site superclusters may encompass both gene-rich and gene-poor domains (**Fig. 5**). Such high-level features are evident across the entire 30-Mb region that we surveyed and appear to reflect a fundamental organizing principle of the human genome.

DISCUSSION

Here we presented a powerful new approach, DNase-array, for large-scale, high-resolution mapping of chromatin structure and localization of DNase I hypersensitive sites and associated *cis*-regulatory sequences in complex genomes. The method is simple, reproducible and direct. In marked contrast to other widely applied chromatin analysis techniques such as ChIP-on-chip[26], complicated manipulations, ligations and amplification of DNA fragments before hybridization (with attendant increase in sample noise) are not required. Analysis of DNase-array data is straightforward: the clarity of the resulting DNase I sensitivity signal permits visualization as a standard log ratio (versus a *P* value[26]), against which DNase I hypersensitive sites can be identified with a simple algorithm. The 'two-hit' DNase-array approach has considerable advantages over a 'one-hit' approach based on end-capture[27,28], which may be subject to artifacts from DNA shearing occasioned in

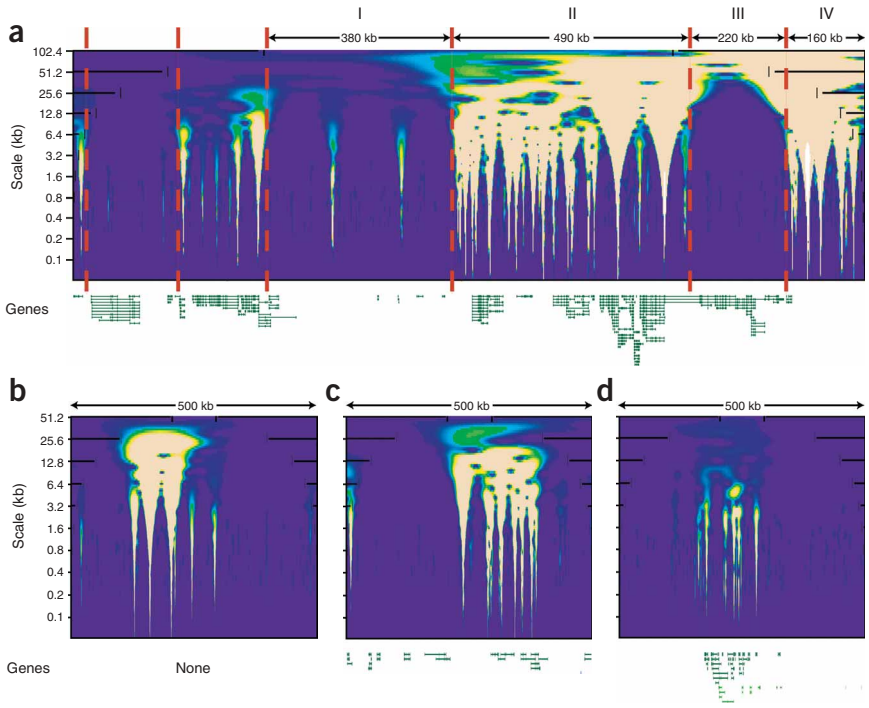

Figure 5 | Higher-order chromatin features revealed by DNase-array. (**a**) Continuous wavelet transform heatmap of chromatin accessibility across a 1.7-Mb segment of chromosome 21 containing the Down Syndrome critical region[29] (*x* axis, genomic position; *y* axis, wavelet scale). Evident are marked transitions between large (160–490 kb) 'superclusters' of DNase I hypersensitive sites (regions II and IV) and intervening inactive chromatin territories. Four broad classes of chromatin domains are thus distinguished based on TSS density and chromatin activity: I, TSS-poor, inactive chromatin; II, TSS-rich, DNase I hypersensitive site–rich active chromatin; III, TSS-rich, inactive chromatin; IV, TSS-poor, DNase I hypersensitive site–rich active chromatin. (**b**–**d**) Heatmaps showing DNase I hypersensitive site superclusters within three diverse 500-kb regions with varying TSS or gene density.

the course of nuclear release and DNA purification. Shear forces are maximal over the midsections of large DNA molecules and are virtually nonexistent at ends; as such, the small fragments exploited by DNase-array are highly specific. At a granular level, DNase I hypersensitive sites detected by DNase-array analysis represent the aggregate of very large numbers of specific, two-hit DNase I–cleavage events occurring over a short distance. The vast majority of the fragments that generate the DNase-array signal are of smaller size than individual DNase I hypersensitive sites. This feature minimizes any potential for bias toward the potential arrangement of DNase I hypersensitive sites relative to one another in the genome. This is well illustrated by results such as those in **Figure 2** and **Supplementary Figure 2**, which show that the approach is capable of detecting a wide variety of different *in vivo* configurations of DNase I hypersensitive sites.

Using conventional chromatin accessibility assays, we have demonstrated DNase-array to be highly sensitive and specific for detection of classical DNase I hypersensitive sites over unique regions of the genome that are interrogable using tiling DNA microarrays. To achieve comprehensive high-throughput mapping of functional elements encoded by DNase I hypersensitive sites, DNase-array may be combined with methods such as Quantitative Chromatin Profiling[13], which is capable of interrogating most repetitive regions that are intractable to hybridization-based approaches. The application of these high-throughput methodologies, however, must be placed into perspective. For focused interrogation of selected genomic DNA segments on a smaller scale ($< \sim 25$ kb), or for targeted high-resolution exploration of a limited number of functional elements (for example, promoters), classical DNase I hypersensitivity assays remain the gold standard. The utility of these conventional assays has been convincingly demonstrated, and they should be regarded as an essential component of the functional genomic researcher's repertoire.

In summary, our results collectively reveal the degree to which quantitative analysis of chromatin accessibility can illuminate the complexities of higher genome organization. Computational analysis of higher-order chromatin features exposed by the long-range DNase I sensitivity profile may highlight additional unexpected features of functional genome architecture. We anticipate that large-scale, high-resolution mapping of DNase I hypersensitive sites across the genome of multiple cell types will greatly accelerate discovery of the genomic DNA signals that encode tissue-specific chromatin accessibility and the activity potential of human genes. With the advent of new high-density DNA microarrays, a comprehensive genome-wide map of DNase I hypersensitive sites and associated *cis*-regulatory sequences for major human tissues should be within reach. Such maps are expected to have a substantial impact on the search for functional noncoding variations that modulate gene regulation, human disease and quantitative phenotypes.

METHODS

Nuclear extraction and DNase I digestion. We performed nuclear extraction, permeabilization and DNase I (Roche) digests using a standard approach as described previously[13]. After cultivation, we pelleted cells and washed them with phosphate-buffered saline. We resuspended cell pellets in 19 ml of 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, 0.15 mM spermine). We then released nuclei by dropwise addition of 6 ml of 2× NP-40 (0.08% in Buffer A) to the cells followed by incubation on ice for 10 min. We centrifuged nuclei at 1,000*g* for 3 min, and then resuspended and washed the pellets with 25 ml of fresh Buffer A. We resuspended nuclei in Buffer A to a concentration of $\sim 10^8$ nuclei/ml. We performed DNase I (10–80 U/ml) digests for 3 min at 37 °C in 1-ml volumes of DNase I buffer (60 mM $CaCl_2$, 750 mM NaCl). We digested 10^7 nuclei in each DNase I treatment. We terminated reactions by adding an equal volume of stop buffer (1 M Tris-Cl (pH 8.0), 5 M NaCl, 20% SDS, 0.5 M EDTA (pH 8.0), 10 μg/ml RNase A) followed by incubation at 55 °C. After 15 min, we added Proteinase K (25 μg/ml final concentration) to each

digest reaction and incubated them overnight at 55 °C. After DNase I treatments, we performed careful phenol-chloroform extraction. We used an untreated aliquot of nuclei to prepare naked DNA for use as a control sample (see below).

In general, as DNase I digestion time (or units) increases, DNase I hypersensitive sites become manifest, followed ultimately by an increase in nonspecific background cutting. The goal is to select a sample in which adequate digestion has occurred to reveal DNase I hypersensitive sites systematically, but which is not overdigested with attendant elevation of background cutting. In principle, this could be accomplished by using the same amount of DNase I each time. However, we found that, irrespective of manufacturer, neither the number of units of DNase I used in a particular nuclear digestion nor the time of exposure is a reliable predictor of the amount of specific and nonspecific chromatin digestion. The advent of real-time PCR–based methods for interrogating DNase I sensitivity at specific genomic positions[13,29] solves this dilemma by providing a platform for quantitative *a posteriori* monitoring of the degree of specific and nonspecific DNase I digestion in individual aliquots of treated nuclei. To monitor DNase I digestion quantitatively, and to select an optimum sample for evaluation by DNase-array, we prepared several aliquots from the same nuclear preparation, and then used a real-time PCR–based strategy to identify optimum DNase I–treated samples. We quantified, for each aliquot, the degree of cutting in five known lymphoid DNase I hypersensitive sites relative to a reference sequence from the rhodopsin locus on chromosome 3. The latter region is known to be DNase I–insensitive in lymphoid cells[13] and was interrogated using a primer pair with outstanding amplification efficiency (see primer pair RDSN in **Supplementary Table 3** online). The former regions comprised four evolutionarily conserved noncoding sequences encoding DNase I hypersensitive sites found in lymphoid cells and several additional tissue types (including K562 (ref. 27); and CaCo2, HeLa, HL60 and HepG2; P.J.S. *et al.*; unpublished data), which were first uncovered during a genomic survey[27], together with a lymphoid DNase I hypersensitive site in the *DAD1* gene identified previously[13]. We interrogated control DNase I hypersensitive sites using primer pairs Dad, CNS8966, CNS28539, CNS28516 and CNS28540, the full sequences of which are available in **Supplementary Table 3**. Using this approach we identified DNase I–treated samples in which the average amount of cleavage (copy-number loss) among the five DNase I hypersensitive sites fell within the range of 50–70% concomitant with no detectable DNase I–cutting in the control non–DNase I hypersensitive site. Previous work had shown that samples selected using this approach comprehensively expose DNase I hypersensitive sites with high experimental signal-to-noise ratio[13]. We then selected for study the sample that exhibited the highest average cleavage within DNase I hypersensitive sites with no copy number loss as the reference.

DNase I digestion of control DNA. We performed nuclear isolation as described above, and purified DNA by phenol-chloroform extraction and ethanol precipitation, followed by overnight resuspension of purified DNA in 500 µl of TE buffer (pH 8.0). We then treated the purified naked DNA with 0.5, 0.25, 0.125 and 0.625 units/100 µl DNase I at 37 °C for 5 min to generate pools of random control fragments.

Isolation of chromatin-specific and nonspecific DNase I fragments. To size-fractionate both the control and treated samples, we adjusted the concentration of NaCl to 0.8 M by adding 5 M NaCl. We then loaded DNA onto sucrose step gradients. We layered 1-ml aliquots of sucrose solutions (10–40% sucrose, 20 mm Tris-Cl (pH 8.0), 5 mM EDTA, 1 M NaCl) in open-top Polyclear ultracentrifuge tubes (Seton Scientific). We then ran gradients for 24 h at 25,000 r.p.m. at 25 °C in a SW21 swinging bucket rotor Beckman LE-80 Ultracentrifuge 77,002*g*. We pipetted fractions of 600 µl each from the top of the six gradients and pooled the respective fractions. To determine the DNA fragment sizes in the fractions, we mixed 10 µl from the pooled fractions with 2 µl of loading dye and 2 µl of a 1:1,000 dilution of SYBR Green (Invitrogen). We loaded the sample on a 1% Tris acetate-EDTA (TAE) agarose gel, ran it at 5 V/cm for 60 min, and then imaged it on a Typhoon 9200 imager (Amersham Biosciences). We pooled fractions with fragments smaller than 1.2 kb and cleaned the DNA using Qiagen PCR purification columns according to the manufacturer's protocol. We eluted DNA with 1 mM Tris-Cl (pH 8) and determined the concentration. Finally, we precipitated 2 µg of purified DNA by adding 1/10 volume of 3 M sodium acetate (pH 5.0) and 2 volumes of ethanol, centrifuged the reactions at 10,000*g* for 10 min, washed them with 70% ethanol, dried them in a desiccator; and resuspended them in TE buffer.

Microarray design. We manufactured (Nimblegen) a DNA microarray comprising ~390,000 50-mer probes tiled with 12-bp overlap across non-RepeatMasked regions of 44 genomic segments defined by the ENCODE regions[11]. The tiling path of probes is available as a publicly accessible track in the UCSC genome browser (http://genome.ucsc.edu).

Labeling and hybridization. We used 1 µg of each of chromatin-specific and nonspecific DNase I–released fragments for labeling and microarray hybridization. We mixed each sample with 40 µl each of 1 µM Cy5 or Cy3 end-labeled random nonamer oligonucleotides (TriLink Biotechnologies) together with bacterial control DNA in a total volume of 88 µl. To anneal random primers we heated the sample to 98 °C for 5 min, and cooled rapidly in ice water for 2–3 min. To create labeled fragments, we added 100 units of *E. coli* DNA polymerase Klenow fragment and 10 µl of a 10 mM equimolar mixture of dATP, dTTP, dCTP and dGTP. We incubated the reactions at 37 °C for 2 h, and terminated the reaction by adding 10 µl of 0.5 M EDTA. We then precipitated the DNA with 110 µl of isopropanol and 11 µl of 5 M NaCl, collected the precipitate by centrifugation, and washed the pellet with 80% ethanol. We then dried the pellet and resuspended in 10 µl of dH_2O.

Next we mixed 12 µg each of Cy5-labeled (chromatin-specific) and Cy3-labeled (control) DNA samples, and added 4 µl of a 2.94 nM equimolar mixture of control oligos C1 and C2 (**Supplementary Table 3**). We concentrated the sample by drying it under vacuum and low heat, to a volume <14.4 µl, and adjusted the final volume to 14.4 µl with dH_2O. We then added 11.25 µl of 20× SSC, 18 µl 100% formamide, 0.45 µl 10% SDS, 0.45 µl 10× TE (100 mM Tris-Cl (pH 8.0), 10 mM EDTA), and 0.45 µl of an equimolar mixture of Cy3- and Cy5-labeled CPK6 oligonucleotides CPK61 and CPK62 (**Supplementary Table 3**). We heated the sample to 95 °C, applied it to the microarray slide, and incubated

in a MAUI Hybridization Station (BioMicro Systems) at 42 °C for 16–20 h. After hybridization, we washed the samples once in 0.2× SSC with 0.2% SDS and 0.1 mM DTT for 10–15 s, followed by another 2-min wash in the same solution with gentle agitation. We then washed in 0.2× SSC with 0.1 mM DTT for 1 min, and in 0.05× SSC with 0.1 mM DTT for 15 s. After washes we dried the slides by centrifugation.

Identification of DNase I hypersensitive sites in DNase-array data. We identified DNase I hypersensitive sites using a straightforward generative algorithm. In summary, this consisted of: (i) minimizing probe-to-probe variability; (ii) computing confidence bounds on signal outliers; and (iii) finding local maxima (peaks) in the signal that exceed the confidence bounds. See **Supplementary Methods** online for details.

Additional information. Descriptions of cell culture, microarray data acquisition and processing, conventional DNase I assays, comparison of conventional assays with DNase-array, calculation of null distributions for genomic feature analysis, and wavelet analysis of DNase I sensitivity are available in **Supplementary Methods**. DNase I–sensitivity and DNase I hypersensitivity data tracks are publicly available through the UCSC genome browser under the track labeled 'UW DNase-array' in the 'ENCODE Chromosome Chromatin' subsection. All software used to analyze DNase-array data is available on request.

Accession codes. Gene Expression Omnibus (GEO): GSE4334.

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

ACKNOWLEDGMENTS

This work was supported by grants from the US National Institute of General Medical Sciences and the National Human Genome Research Institute to J.A.S. and W.S.N.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Methods* website for details).

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

1. Felsenfeld, G. Chromatin as an essential part of the transcriptional mechanism. *Nature* **355**, 219–224 (1992).
2. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448–453 (2003).
3. Wu, C. The 5′ ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase. *Nature* **286**, 854–860 (1980).
4. Keene, M.A., Corces, V., Lowenhaupt, K. & Elgin, S.C. DNase I hypersensitive sites in *Drosophila* chromatin occur at the 5′ ends of regions of transcription. *Proc. Natl. Acad. Sci. USA* **78**, 143–146 (1981).
5. McGhee, J.D., Wood, W.I., Dolan, M., Engel, J.D. & Felsenfeld, G. A 200 base pair region at the 5′ end of the chicken adult β-globin gene is accessible to nuclease digestion. *Cell* **27**, 45–55 (1981).
6. Gross, D.S. & Garrard, W.T. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* **57**, 159–197 (1988).
7. Li, Q., Peterson, K.R., Fang, X. & Stamatoyannopoulos, G. Locus control regions. *Blood* **100**, 3077–3086 (2002).
8. Weintraub, H. & Groudine, M. Chromosomal subunits in active genes have an altered conformation. *Science* **193**, 848–856 (1976).
9. Felsenfeld, G. *et al.* Chromatin boundaries and chromatin domains. *Cold Spring Harb. Symp. Quant. Biol.* **69**, 245–250 (2004).
10. Sproul, D., Gilbert, N. & Bickmore, W.A. The role of chromatin structure in regulating the expression of clustered genes. *Nat. Rev. Genet.* **6**, 775–781 (2005).
11. The Encode Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
12. Elgin, S.C. The formation and function of DNase I hypersensitive sites in the process of gene activation. *J. Biol. Chem.* **263**, 19259–19262 (1988).
13. Dorschner, M.O. *et al.* High-throughput localization of functional elements by quantitative chromatin profiling. *Nat. Methods* **1**, 219–225 (2004).
14. Lee, G.R., Fields, P.E., Griffin, T.J. & Flavell, R.A. Regulation of the Th2 cytokine locus by a locus control region. *Immunity* **19**, 145–153 (2003).
15. Lee, G.R., Spilianakis, C.G. & Flavell, R.A. Hypersensitive site 7 of the TH2 locus control region is essential for expressing TH2 cytokine genes and for long-range intrachromosomal interactions. *Nat. Immunol.* **6**, 42–48 (2005).
16. Loots, G.G. *et al.* Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* **288**, 136–140 (2000).
17. Ansel, K.M. *et al.* Deletion of a conserved Il4 silencer impairs T helper type 1-mediated immunity. *Nat. Immunol.* **5**, 1251–1259 (2004).
18. Smale, S.T. & Kadonaga, J.T. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* **72**, 449–479 (2003).
19. West, A.G. & Fraser, P. Remote control of gene transcription. *Hum. Mol. Genet.* **14** (Spec. No. 1), R101–111 (2005).
20. Yelin, R. *et al.* Widespread occurrence of antisense transcription in the human genome. *Nat. Biotechnol.* **21**, 379–386 (2003).
21. Cawley, S. *et al.* Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499–509 (2004).
22. Katayama, S. *et al.* Antisense transcription in the mammalian transcriptome. *Science* **309**, 1564–1566 (2005).
23. Hinrichs, A.S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
24. Percival, D.B. *Wavelet methods for time series analysis* (Cambridge University Press, Cambridge, 2000).
25. Korenberg, J.R. *et al.* Down syndrome phenotypes: the consequences of chromosomal imbalance. *Proc. Natl. Acad. Sci. USA* **91**, 4997–5001 (1994).
26. Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
27. 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).
28. Sabo, P.J. *et al.* Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. *Proc. Natl. Acad. Sci. USA* **101**, 4537–4542 (2004).
29. McArthur, M., Gerum, S. & Stamatoyannopoulos, G. Quantification of DNaseI-sensitivity by real-time PCR: quantitative analysis of DNaseI-hypersensitivity of the mouse beta-globin LCR. *J. Mol. Biol.* **313**, 27–34 (2001).