Cell PRESS
Developmental Cell
Resource

NA-Seq: A Discovery Tool for the Analysis of Chromatin Structure and Dynamics during Differentiation

Gaetano Gargiulo,[1,6] Samuel Levy,[2,6] Gabriele Bucci,[1] Mauro Romanenghi,[1] Lorenzo Fornasari,[3] Karen Y. Beeson,[2] Susanne M. Goldberg,[2] Matteo Cesaroni,[1] Marco Ballarini,[3] Fabio Santoro,[3] Natalie Bezman,[4] Gianmaria Frigè,[1] Philip D. Gregory,[4] Michael C. Holmes,[4] Robert L. Strausberg,[2] Pier Giuseppe Pelicci,[1] Fyodor D. Urnov,[4] and Saverio Minucci[1,5,*]

[1]Department of Experimental Oncology, IFOM-IEO Campus, European Institute of Oncology (IEO), Via Ripamonti 435, 20141 Milan, Italy
[2]J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850, USA
[3]Congenia, Genextra Group, Piazzetta Bossi 4, 20121 Milan, Italy
[4]Sangamo BioSciences, Richmond, CA 94804, USA
[5]Department of Biomolecular Sciences and Biotechnology, University of Milan, Via Celoria 26, 20133 Milan, Italy
[6]These authors contributed equally to this work
*Correspondence: saverio.minucci@ifom-ieo-campus.it
DOI 10.1016/j.devcel.2009.02.002

SUMMARY

It is well established that epigenetic modulation of genome accessibility in chromatin occurs during biological processes. Here we describe a method based on restriction enzymes and next-generation sequencing for identifying accessible DNA elements using a small amount of starting material, and use it to examine myeloid differentiation of primary human CD34+ cells. The accessibility of several classes of *cis*-regulatory elements was a predictive marker of in vivo DNA binding by transcription factors, and was associated with distinct patterns of histone post-translational modifications. We also mapped large chromosomal domains with differential accessibility in progenitors and maturing cells. Accessibility became restricted during differentiation, correlating with a decreased number of expressed genes and loss of regulatory potential. Our data suggest that a permissive chromatin structure in multipotent cells is progressively and selectively closed during differentiation, and illustrate the use of our method for the identification of functional *cis*-regulatory elements.

INTRODUCTION

Cellular development and differentiation rely on an integration of (static) genetic and (dynamic) epigenetic information (Bird, 2002; Turner, 2001). The "epigenome" of the cell emerges from the interplay between *trans*-acting factors, noncoding RNAs, chromatin modifiers and remodelers, and the preexisting state of the chromatin template (Bernstein et al., 2007).

High-throughput approaches are essential tools for studying the epigenome (Gargiulo and Minucci, 2009). High-resolution maps of histone posttranslational modifications (PTMs), nucleosome positioning, and DNaseI hypersensitive sites (DHSs) have been generated in lymphoid T cells (Barski et al., 2007; Boyle et al., 2008a; Schones et al., 2008). In addition, binding sites for several transcription factors (TFs) have been analyzed in cell lines. These analyses revealed that the DNA primary sequence is insufficient to determine whether a given factor will be bound in vivo. Furthermore, the number of binding sites for many TFs in living cell chromatin is considerably greater than expected, leading to the difficulty of distinguishing nonfunctional from bona fide *cis*-regulatory binding sites (Li et al., 2008).

Epigenetic mechanisms governing dynamic processes have begun to be elucidated, studying in vitro differentiation of murine and human embryonic stem (ES) cells. Human and murine hematopoietic stem and progenitor cells (HSPCs) provide a similarly attractive framework for studying epigenetic transitions during differentiation (Attema et al., 2007; Bottardi et al., 2003).

The accessibility of DNA in chromatin to exogenous nucleases (usually DNaseI) is a distinguishing epigenetic mark of gene regulatory elements (Elgin, 1988; Wu, 1980). The use of type II restriction enzymes (REs) to probe chromatin on a single-locus basis revealed the utility of this approach to map *cis*-regulatory elements, transcription factor binding sites (TFBSs), and domains of remodeled chromatin (Almer and Horz, 1986; Archer et al., 1992).

Here we describe a method for specifically sequencing RE-accessible regions from cell nuclei, and its application to map the "nuclease-accessible epigenome" in human HSPCs and their committed progeny. This analysis provided several insights into the general structure of chromatin and the changes that occur during the process of cellular differentiation, and a resource for the identification of hematopoietic regulatory elements in human cells.

RESULTS

High-Throughput Sequencing of Nuclease-Accessible Sites

Nuclease-accessible site sequencing (NA-Seq) starts with the treatment of isolated nuclei with appropriate REs to cleave

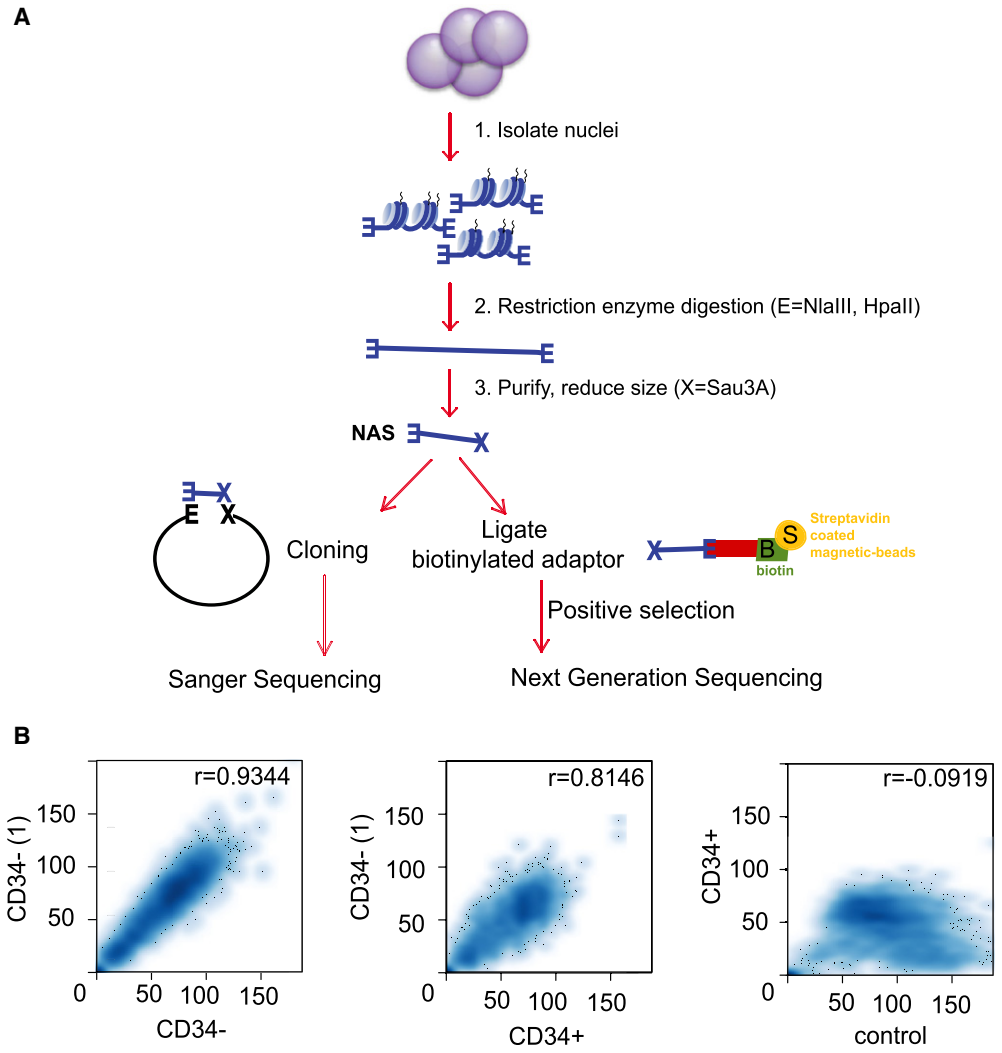


Figure 1. Genome-Wide Analysis of Nuclease-Accessible Sites

(A) Outline of the NA-Seq assay.

(B) Scatter analysis of NAS libraries generated using NlaIII as RE: two sequential 454 runs of CD34− libraries (left; $p < 0.001$), CD34+ versus CD34− libraries (middle; $p < 0.001$), and CD34+ versus the control library (naked DNA). The axes indicate the number of NASs within discrete 1.2 Mb genomic windows (data points). Darker paint means higher density (associated Pearson correlation is indicated).

accessible sites, leaving an "in vivo mark" due to its sequence specificity (Pingoud and Jeltsch, 2001). The combination of two REs (HpaII and NlaIII) allows an extensive coverage of the genome due to their different recognition motif (on average, one site every <300 bp). Next, genomic DNA is fractionated in vitro using a different enzyme (usually Sau3AI). Nuclease-accessible sites (NASs) will thus carry the sticky ends from both enzymes. NASs can be isolated by cloning, or by next-generation sequencing, after ligation of biotinylated linkers that enable the selective enrichment and sequencing of DNA fragments carrying NlaIII and HpaII sticky 5′ ends (Figure 1A). To reduce background, reads lacking the expected motif (CATG and CCG, for NlaIII and HpaII REs, respectively; see Table S1 available online) were not further analyzed.

We performed a pilot validation study using transformed cell lines (not limiting in number; see Supplemental Results), and then applied the procedure to primary human CD34+ HSPCs (Morrison et al., 1995). We differentiated CD34+ cells along the myeloid lineage by exposure to a cocktail of cytokines in vitro (Piacibello et al., 1999). Cells lose the CD34 marker and acquire myeloid-specific markers (CD34−/CD13+/CD33+); for brevity, we refer to them as CD34− cells (Figure S1). As an experimental control, we applied the same procedure to naked genomic DNA.

NAS libraries were obtained by considering only nonredundant NlaIII and HpaII reads, and then merging multiple reads that identify a region <600 bp to represent one continuously accessible stretch. We obtained 130,549 merged NASs in CD34+ cells and 338,316 in CD34− cells (the difference is due to the fact that we performed two consecutive runs of the CD34− library), and 177,465 merged sequence reads in the control library. To provide an immediate validation, we compared the distribution of NASs in the libraries, using a scanning interval of 1.2 Mb across all

chromosomes. As expected, we obtained a near-linear correlation between the distributions of NASs when comparing the two runs derived from the same CD34− NlaIII library ($r = 0.93$; $p < 0.001$; Figure 1B). CD34+ and CD34− libraries showed also a strong positive correlation ($r = 0.81$; $p < 0.001$); the decrease of the correlation coefficient is a measure of differential chromatin accessibility between CD34− and CD34+ cells, in a context of an overall similarity. In contrast, the library derived from naked DNA did not exhibit any significant correlation with the chromatin-derived libraries (control versus CD34+ cells: $r = -0.092$; $p >> 0.05$), showing that the chromatin-derived libraries share a specific accessibility landscape that is unrelated to the pattern of digestion of naked DNA (Figure 1B).

NASs Identify Regions of Enhanced Accessibility in Primary Cells

We validated our NAS libraries using two approaches: (1) experimental validation and (2) comparison of our NAS libraries with available data sets of accessible DNA elements.

To confirm that NASs derive from stretches of highly remodeled chromatin, we performed qPCR-based DNaseI assays (McArthur et al., 2001). We measured the sensitivity to DNaseI treatment of a cohort of 51 DNA regions from CD34+ and control libraries (representative of both data sets; see Table S2). The validation was performed on CD34+ cells derived from different donors, providing indications of the biological variability of the accessible DNA elements identified (see Discussion). Chromatin-derived NASs were found to be more sensitive to mild DNaseI treatment, if compared to control (Figure 2A). In fact, 80.3% of the NASs tested by real-time PCR (41/51) showed a marked sensitivity to DNaseI, either equivalent or higher to several positive control regions (Table S2; see in Figure 2A the CD34 gene-associated NASs found highly accessible in CD34+ and not in CD34− cells). Overall, the NAS and control populations were found to be significantly different, comparing average sensitivity to DNaseI degradation (t test for independent samples, unequal variance, $t = -9.64$, f.d. = 90; $p << 0.01$; Figure 2A), and in a receiving operator curve analysis (ROC analysis: area under the curve = 0.905; $Z = 13.86$; $p << 0.01$; Figure 2B). Together, these analyses demonstrate the high sensitivity and specificity of the NA-Seq libraries generated in primary cells.

We further characterized the structural properties of NASs using principal component analysis. Three discrete groups were partitioned according to their quantitative accessibility to DNaseI (Figure 2C). These results extend to a genome-wide analysis the demonstration of the existence of differential degrees of accessibility, as previously seen at a single-locus level (McArthur et al., 2001).

Next, we compared CD34+ and CD34− NASs with previously published data sets. NA-Seq can identify DHSs shared with T cells (Boyle et al., 2008a), as well as FAIRE sites in fibroblasts (genomic elements accessible in formaldehyde crosslinked chromatin) (Giresi et al., 2007). NASs are also enriched in transcription start sites (TSSs), expected to be accessible to exogenous nucleases (see Supplemental Results).

Our validation is therefore consistent with the expectation that NASs in our assay preferentially mark accessible loci in the genome.

NASs Are Associated with Specific Histone PTMs Reflecting Their Functional Activity and Developmental Potential

To correlate accessibility with histone PTMs, we selected ≈90 CD34+ NASs representative of promoters and nonpromoter regions, and analyzed them by quantitative chromatin immunoprecipitation (qChIP) for the presence of the nine best-characterized PTMs (Consortium et al., 2007). To overcome the limitation in cell number a myeloid progenitor cell line (U937) was used, and the accessibility of the selected NASs was experimentally verified by DNaseI qPCR.

Promoter regions and hypersensitivity to DNaseI tend to group together in an unsupervised clustering analysis, and are associated with di- and trimethylation of H3 lysine 4 (H3K4me2-me3), acetylated H3 (H3K9Ac and H3panAc), and dimethylation of H3 lysine 79 (Figure 3A). This pattern of PTMs (termed a "euchromatic island") was previously reported as a determinant for Myc binding to its target genes (Guccione et al., 2006); our results now extend this signature to the whole spectrum of NAS-containing promoters. We also noticed a small cluster of CD34+ NASs that was targeted by more repressive marks in U937 cells, resulting in a strongly reduced accessibility to DNaseI (Figure 3A, cluster 3; Figure 3D).

Furthermore, a subset of intergenic NASs simultaneously scored positive for euchromatic marks (H3/H4ac and H3K4me), accessibility to DnaseI, and the heterochromatic mark H3K27me3 (Figure 3A). In ES cells, the coexistence of activating and repressing marks was associated with a group of developmentally regulated genes, as well as the large domain containing the *HoxA* genes, named "bivalent" (Bernstein et al., 2006). Figure 3B shows an example of colocalization between H3K4me2 and H3K27me3 at an NAS in the intragenic region of the *Pak3* gene, observed also in resting T cells (Figure 3C), in agreement with a report showing co-occurrence of the bivalent mark in additional cell types other than ES cells (Barski et al., 2007).

To identify developmentally relevant genes, we generated gene expression profiles in CD34+ and CD34− cells, and selected NAS-associated genes with low expression levels in CD34+ cells and either strongly upregulated or further silenced in CD34− cells (i.e., poised for activation). Interestingly, these include >30% Polycomb target genes and bivalent promoters from ES cells (Figure S2). H3K4me3, H3K4me2, H3K27me3, and RNA polymerase II were then immunoprecipitated from CD34+ cells, and the state of seven of those promoters was analyzed (Figure 3E). We simultaneously found H3K4me3 and H3K27me3 at 2/7 promoters (validated by sequential ChIP; data not shown), indicating that accessible promoters of genes expressed at a low level include, but are not limited to, bivalent promoters. Interestingly, 7/7 promoters showed high levels of H3K4me2. In several cases, we noticed a pattern of H3K4me2 high/H3K4me3 low that is reminiscent of the H3K4me2+/H3K4me3− pattern previously found at lineage-specific gene promoters poised for activation, and associated with multilineage commitment in murine hematopoietic progenitors (Orford et al., 2008). This finding lends support to the hypothesis that similar subsets of genes may exist in primary human hematopoietic progenitors.

Together, our data clearly correlate distinct combinatorial patterns of histone PTMs with a structural and functional feature of the genome: its accessibility.

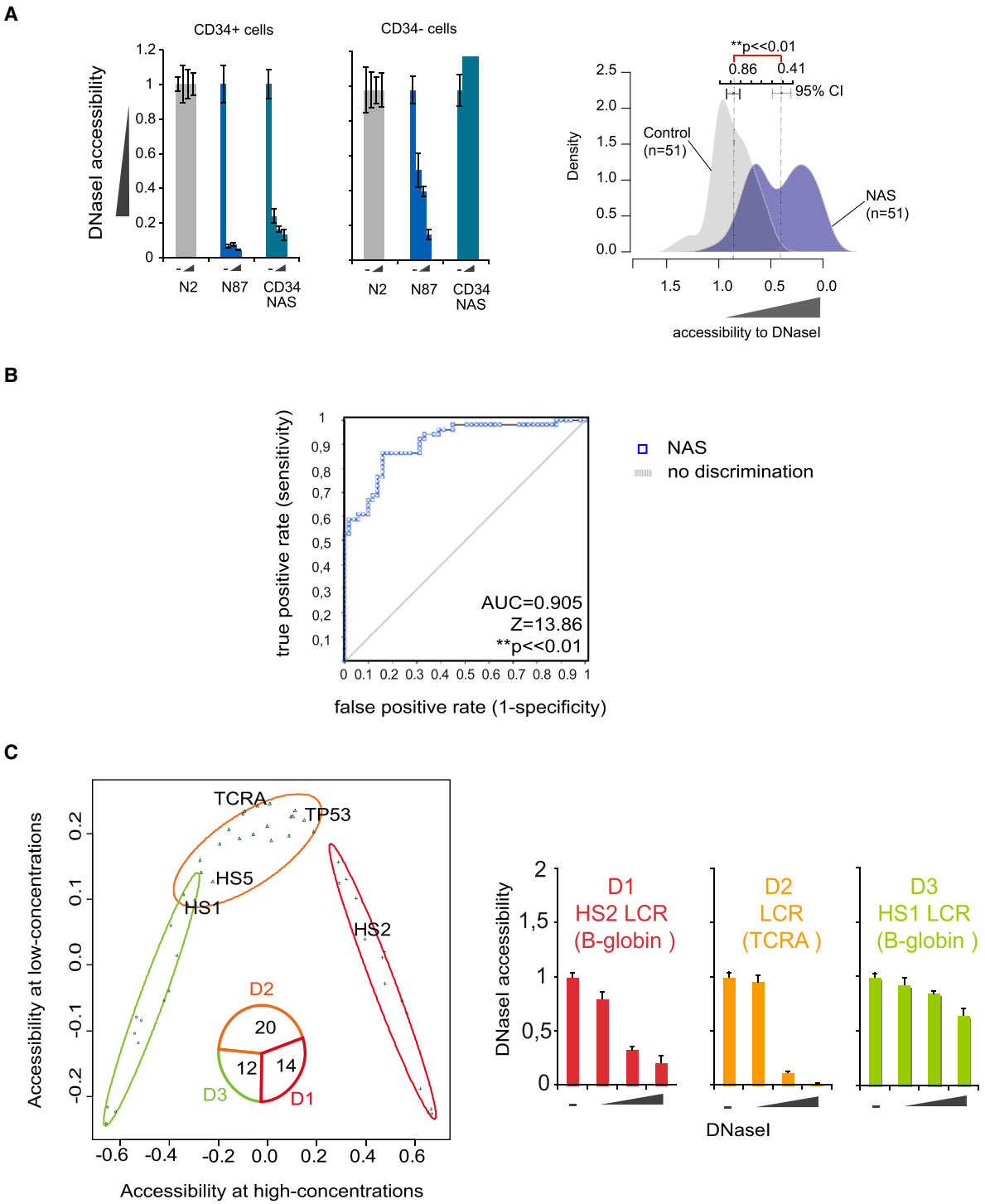


Figure 2. NASs Are Sites of Increased Genomic Accessibility in Vivo

(A) DNase PCR assay: CD34+ and CD34− cell nuclei were digested with DNaseI or left untreated, and the indicated genomic regions were amplified by PCR. Accessibility represents the amount of amplicons recovered. Left: bar plot showing one DnaseI-resistant locus (N2), one intergenic CD34+ NAS (N87), and a CD34+ NAS within the CD34 gene. Error bars represent the standard deviations of three independent experiments. Right: kernel density plot showing the comparison of 51 NASs from CD34+ NASs and 51 sites from control libraries (t test for independent samples, unequal variance, $p << 0.01$). One representative experiment (out of three) is shown.

(B) ROC plot of the data in (A) (right panel).

(C) Principal component analysis of data in (A) (right). Partitioning used two components (accessibility at high and low doses). Data points are summarized in the inset (D1, high sensitivity; D2, medium sensitivity; D3, low sensitivity). Known regulatory elements in the groups are shown in the right panels: HS of the TCRα LCR, and HS1 and HS2 from the β-globin LCR. N2 was used as a calibrator.

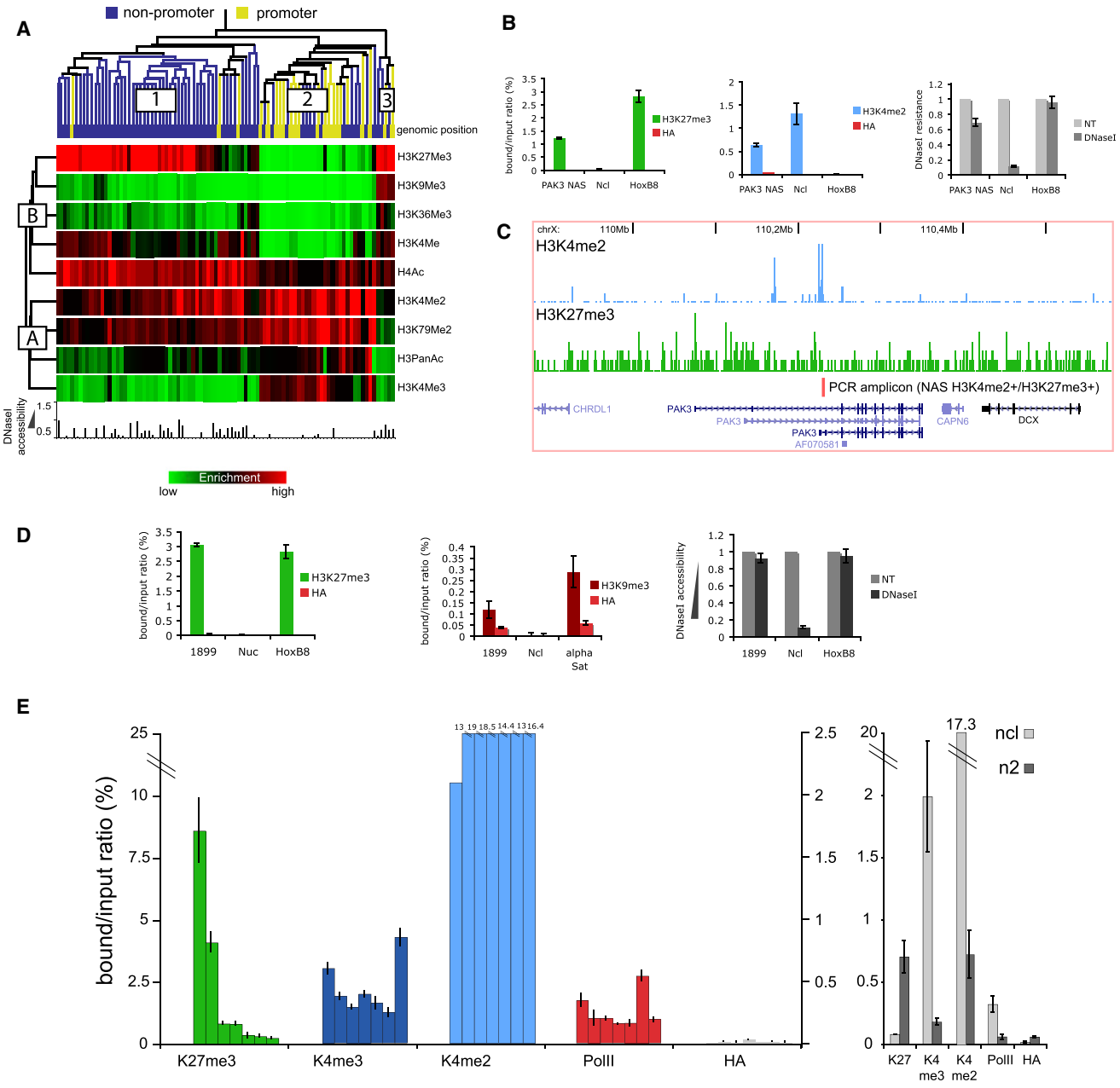


Figure 3. NASs Are Associated with Euchromatic Histone Marks

(A) Clustering analysis of histone marks (boxes A and B) at 90 NASs identified in the CD34+-derived library (boxes 1, 2, and 3). The histogram shows NAS accessibility measured in U937 cells (P/R-9 clone) by DNase PCR. ChIPs in P/R-9 cells were performed using the nine antibodies indicated, and normalized to the nonimmunoprecipitated DNA and to the relative H3 occupancy. The yellow and blue labels indicate whether NASs are promoters or not. The color-coded heat map represents recoveries of DNA for each NAS in each qChIP.

(B) NASs located in the intragenic region of *PAK3*, associated with H3K4me2 and H3K27me3 PTMs, show moderate accessibility in the DNaseI PCR assay. *NCL* and the *HoxB8* promoter served as positive controls for qChIPs.

(C) UCSC genome browser view of the *PAK3* NAS in mature CD4+ lymphocytes (Barski et al., 2007).

(D) Colocalization of H3K27me3 and H3K9me3 impairs DNaseI accessibility in P/R-9 cells of NAS number 1899, accessible in CD34+ cells. Error bars represent standard deviations of three independent experiments.

(E) NASs at promoters of poised genes in CD34+ cells are associated with H3K4me2. In the left panel, qChIP results for the indicated PTMs are reported for the promoter regions of seven genes (*HRK*, *NLF1*, *HLF*, *CRhBP*, *FLJ14213*, *pPP1R16B*, and *NPR3*, from left to right) associated with NASs in CD34+ cells poorly expressed, overlapping PcG target (Bracken et al., 2006) bivalent promoters in human embryonic stem cells (hES) (Pan et al., 2007), and regulated in CD34− cells. *NCL* (H3K4me3+/H3K27me3−) and N2 (H3K4me3−/H3K27me3+) served as controls (right panel). Note the low levels of PolII associated with these promoters, consistent with their low levels of expression in CD34+ cells.

NA-Seq Identifies Elements Involved in Tissue-Specific Gene Regulation and Expression

To evaluate the hypothesis that NA-Seq may be an appropriate methodology to identify functional regulatory elements, we performed (1) a bioinformatic and experimental evaluation of putative enhancer function by nonpromoter NASs, and (2) a correlative analysis of gene expression state with accessibility of its cognate locus.

We first made use of a data set of validated mammalian enhancers (http://enhancer.lbl.gov; Visel et al., 2007). We found 119/446 enhancers (26.7%) in our data set, and thus predicted to be functional in hematopoietic cells. We then identified a large subset of extremely evolutionarily conserved DNA stretches overlapping NASs, and among those, 123 out of 481 ultraconserved elements (UCEs), large-sized regions with the highest degree of conservation across species (Bejerano et al., 2004). To investigate their functional potential in hematopoietic cells, we cloned a subset of these regions upstream of the firefly luciferase gene driven by the SV40 minimal promoter, and assayed reporter activity in two hematopoietic cell lines. Of the tested regions, 5/7 (71.4%) had a considerably increased luciferase gene activity, whereas 2/2 UCEs not present in our data set of NASs failed to transactivate (Figure 4A).

Additionally, we found that 6096 NASs colocalize with distal H3K4me1 peaks recently described in HSPCs (Cui et al., 2009). Peaks of H3K4me1 are a distinctive feature of enhancers (Heintzman et al., 2007). Remarkably, 4434 NASs (72.6%) overlapping with H3K4me1 showed high sequence conservation across species, including the HS2 enhancer of the globin LCR and a distal enhancer of the PU.1 gene (Steidl et al., 2007).

To determine whether NASs identify tissue-specific elements, we checked the transcriptional state of those genes associated with NASs at their promoters. Promoter accessibility correlates with higher transcription than the average expression of all detectable transcripts in both cell types ($p << 0.01$; Figure 4B), confirming that transcriptionally active regions are enriched in the NAS libraries. A similar comparison starting from the naked DNA-derived library did not show significant differences (Figure 4C).

To establish a quantitative relationship between NASs and gene expression, we subdivided all genes transcribed in CD34+ cells (or CD34− cells; data not shown) into ten groups, according to their expression level. Groups of genes with the highest level of expression (tenth decile) presented an NAS in their promoter area more frequently than moderately transcribed genes (seventh decile; Figure 4D). In all cases, we observed enrichment in NASs in the promoter region of expressed genes when compared with the naked DNA-derived library, even for poorly expressed genes (second decile; Figure 4D). The average increase in expression observed for NAS-associated genes in hematopoietic cells is cell-type specific, because in unrelated cell types (for which the accessibility state is unknown), we could detect a much weaker (and not significant) increase in average expression (Figure 4C).

Interestingly, those genes identified in CD34+ cells by an associated NAS in their promoter regions were transcribed at the highest rate in CD34− cells (Figures 4B and 4C). Furthermore, genes expressed at a very low level in CD34+ cells are specifically regulated during differentiation, and include Polycomb target genes and genes involved in lymphoid development and function, and hence not linked to myeloid differentiation (Figure S2). Together with the observation that genes not actively elongating may be marked by H3K4me3 (Guenther et al., 2007), our data support the view that genome accessibility is not only a feature of active genes but also of genes poised for activation.

NASs Are Prevalent at Intergenic Sites, and Are Observed in Repetitive Elements

NASs are significantly located in "intergenic" sites by using the annotation settings previously employed by the ENCODE project (χ^2 = 3371.4, 2 d.f., 59.7%, CD34+; 63.1%, CD34−; 54.6%, naked; $p << 0.001$; Figure 5A).

In line with their ascribed regulatory role (previous section), intergenic NASs may also correspond to unannotated gene promoters, sites of noncoding transcription, and repeated DNA sequences. Given the much higher fraction of nonaccessible repetitive elements in the genome (e.g., centromeric repeats), the higher accessibility of a limited subset of repeated DNA elements may be indicative of a functional role. Independently of the RE used, the repeated sequences overrepresented in our libraries (35%–38%; Table S1) are almost exclusively composed of transposon repeats and Alu repeats (Figure 5B). We also found enrichment for satellite repeats, by weighted analysis of their abundance (not shown). Notably, only unmethylated repeats can be cleaved by the DNA methylation-sensitive HpaII, which usually targets nucleosome-free CpG islands (Tazi and Bird, 1990).

The role of Alu repeats in gene regulation is becoming increasingly more clear (Chen et al., 2008a). Alu elements carrying the DR2 consensus sequence (RARE) can bind RARα in vivo (Laperriere et al., 2007). Accordingly, NASs at these RARE-containing Alu repeats can bind the nuclear receptor RARα (Figure S3). Unlike transposons (Bernstein et al., 2006), accessible RARE-containing Alu repeats are found associated with developmentally regulated and lineage-specific genes (e.g., Figure 5C). Taken together, our data suggest that repeated elements are not accessible by chance, and support a role for Alu repeats in regulating the activity of nuclear receptors.

Accessibility Predicts Transcription Factor Occupancy

We hypothesized that the identification of NASs could be a predictive tool in the analysis of TF binding to *cis*-regulatory elements. As a test case, we exploited the existence of a previously reported genome-wide study of binding sites of the insulator-binding protein CTCF (Kim et al., 2007). Functional CTCF binding sites are shared in 67% of cases among primary fibroblasts (IMR90 cells) and a hematopoietic cell line (Kim et al., 2007). We searched for CTCF binding sites that colocalize with NASs identified in CD34+ or CD34− cells and found a highly significant overlap between these two data sets (total nonredundant sites: 5098, $p << 0.01$; Figure 6A), suggesting an enrichment of CTCF binding sites in our NAS libraries, as previously found at DHSs (Xi et al., 2007). To avoid the limitation in cell number of primary cells, we confirmed the binding in vivo of CTCF by qChIP in KG1 (CD34+) and U937 (CD34−) cell lines. All (27/27) of the predicted CTCF binding sites scored positively in qChIP, providing strong evidence for all 5098 identified NASs being associated with functional CTCF binding sites (Figure 6B; data not shown). For example, CTCF binds at the upstream region of both *HoxA7* and

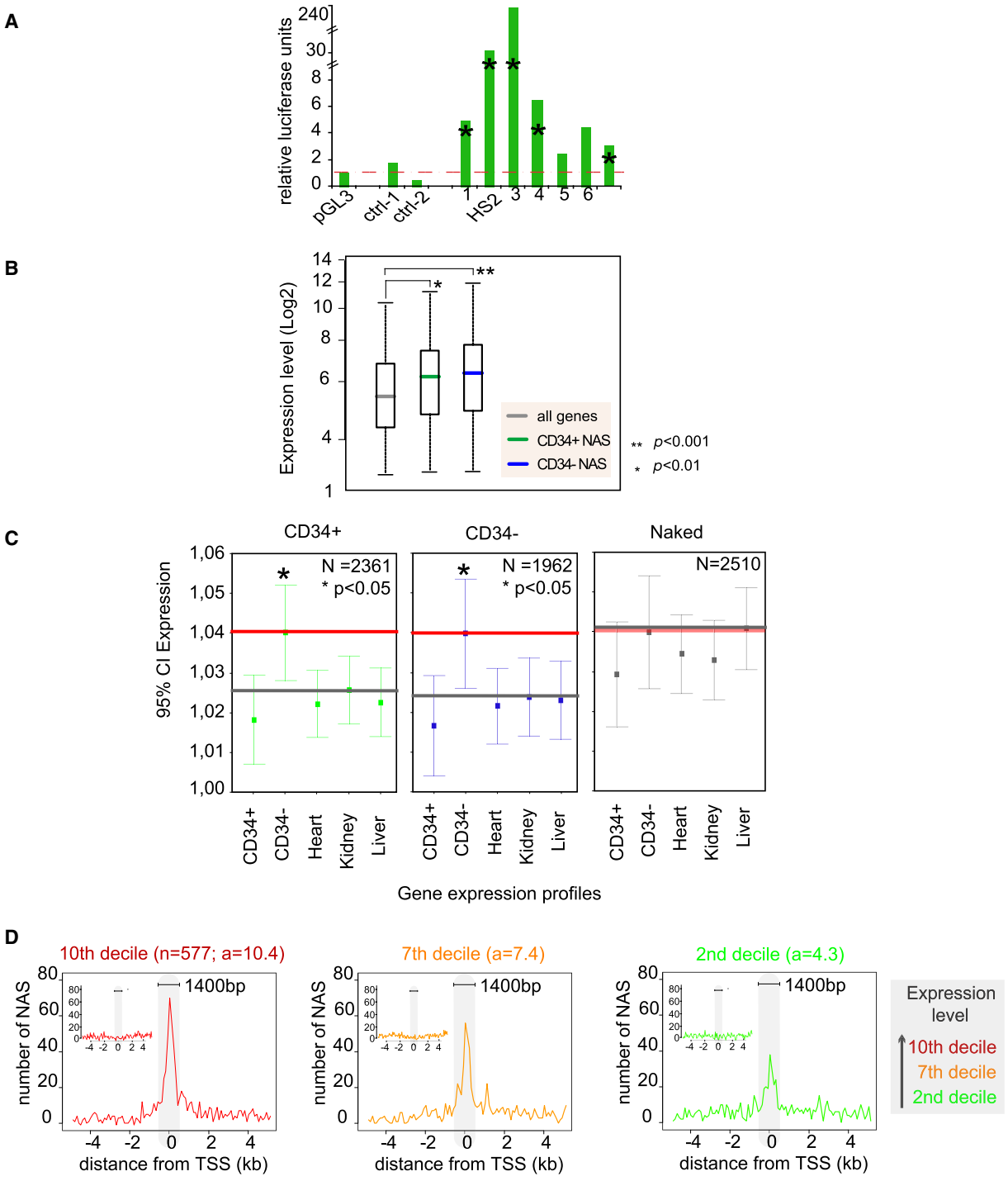


Figure 4. NASs Correlate with Active Gene Expression

(A) Reporter constructs containing conserved noncoding sequences (CNSs) overlapping with CD34+ NASs were used for luciferase assays in KG1a cells. HS2 is the β-globin LCR enhancer (NASs in CD34+) used as positive control; an enhancer-free vector (pGL3) and nonaccessible CNSs served as negative controls. Asterisks mark CNSs/NASs that scored positive also in 32D murine hematopoietic progenitor cells.

(B) The 2361 transcripts occurring near NASs in CD34+ and CD34− cells (<2.5 kb in either direction from the TSS) had an average expression higher than the full set of transcripts (Mann-Whitney test; $p << 0.01$).

(C) The average expression level of genes associated with NASs—within the indicated data set—was normalized by the average expression value of all other genes, and this ratio compared with the relative ratios in nonrelated cell-type expression profiles (from the GEO database). Naked DNA is provided as negative control; asterisks label significantly different values (ANOVA; $p < 0.05$). The highest average is highlighted with a red line, the following highest average with a gray line.

(D) Total number of NASs and their distribution surrounding the promoter for genes expressed in CD34+ cells was shown according to the decile of expression in the different panels (tenth decile being the highest). For each decile, the average normalized level is shown above the relative panel.

HoxA9 genes in correspondence with NASs in CD34+ cells, and in several cell lines in correspondence with DHSs (Figure S4B).

Strikingly, the presence of NASs provides a solid prediction of cell-type-specific CTCF binding. We found 2316 CTCF binding sites predicted by computational analysis though not bound by CTCF in IMR90 cells (Figure 6A) (Kim et al., 2007), and located within NASs identified in CD34+ (1345) and CD34− (1428) cells. qChIP analysis confirmed functional binding of CTCF in 19/22 candidate sites (86.3%; Figure 6B). The binding of CTCF to a selection of putative binding sites was also confirmed in primary CD34+ cells (Figure 6C). In contrast, the sole bioinformatic prediction resulted in CTCF binding in only two out of ten cases (20%).

Among tissue-specific CTCF binding sites, potential insulator sequences are found, such as region 46 (Figures 6B and 6C; showing CTCF binding in KG1a, U937, and CD34+ cells) that is most likely insulating the coding sequence of the TF *Hlf* (downregulated in CD34− cells) from the transactivation driven by transcription of a myeloid-specific gene, *Mmd* (upregulated in CD34− cells; Figure 6C).

To identify an independent example of TF occupancy concordant with the detection of NASs, we examined a subset of neuronal-specific genes showing NASs in their promoters but a low level of expression in CD34+ cells. Neuronal genes are negatively regulated in nonneuronal cells through the zinc finger TF NRSF (Ballas et al., 2005). We found that 232/2217 (10.4%) CD34+ NASs, but not control sites, overlap with a previously published data set of NRSF target promoters, obtained in Jurkat T cell leukemia cells (Johnson et al., 2007). These NASs act presumably as silencers in the hematopoietic compartment, and qChIP analyses in KG1a and CD34+ cells confirmed NRSF binding to 11/12 (91.2%) promoter regions (Figure 6D). In contrast, we found evidence of NRSF binding in only 1 out of 15 cases (6.7%) when bioinformatic prediction was used as the sole predictor of NRSF binding to regions not containing NASs (Figure 6E). Intriguingly, a known NRSF target gene (*Agpat5*) showing higher NRSF binding in CD34+ cells than in cell lines (Figure 6E; number 7), is transcribed at higher levels than other NRSF targets in hematopoietic cells, and is induced in CD34− cells, suggesting a role in myeloid differentiation (Figure S5).

Taken together, these results show that one structural feature of the genome (accessibility) provides a global picture of protein binding for several classes of transcription factors (including repressors).

CD34+ NAS Libraries Are Enriched for Transcription Factor Binding Sites

We searched for TFBS occurrence in CD34+ and CD34− chromatin libraries. We found enrichment of TFBSs within ±600 bp of each NAS (Figure S6A). The overall density of TFBSs in CD34+ NASs was higher than in CD34− cells, supporting the idea of a differentially accessible epigenome in cells endowed with multilineage potential (CD34+) relative to the committed state (CD34−; Figure 6F).

Analysis of individual TFBSs revealed a specific enrichment of binding sites for certain factors in CD34+ cells (Figure S6B). Among those TFBSs enriched in CD34+ cells we found AP-2, which plays a role in the maintenance of an undifferentiated state in mammary progenitors, and therefore could play a similar role in hematopoietic cells (Jäger et al., 2003). In some cases, TFBSs were enriched in both cell types, suggesting a broader role in hematopoietic cells, such as the transcription factor NF-Y (Table S5). To confirm that NF-Y is active in hematopoietic cells, we made a bioinformatic prediction of NF-Y binding in our NAS libraries and could detect a strong enrichment for NF-Y binding in all the loci we tested by qChIP in KG1 or U937 cells (7/7; Figure S7). In the absence of accessibility, 0/6 predicted binding sites showed enrichment for NF-Y, further confirming and extending the validity of our predictive approach to forecast DNA-binding proteins that may act as regulators of cellular differentiation (Figure S7).

Global Changes in Chromatin Accessibility Occur during Myeloid Differentiation

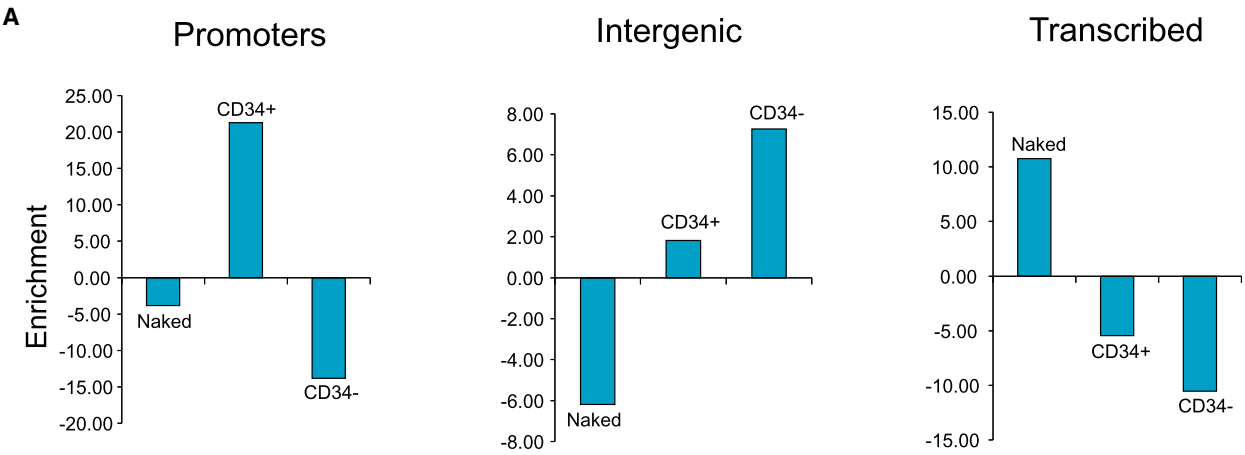
Hematopoietic cells display gross differences in their nuclear structure and chromatin compaction at each step of maturation, highlighting the occurrence of global changes in chromatin organization during differentiation (Löffler et al., 2004).

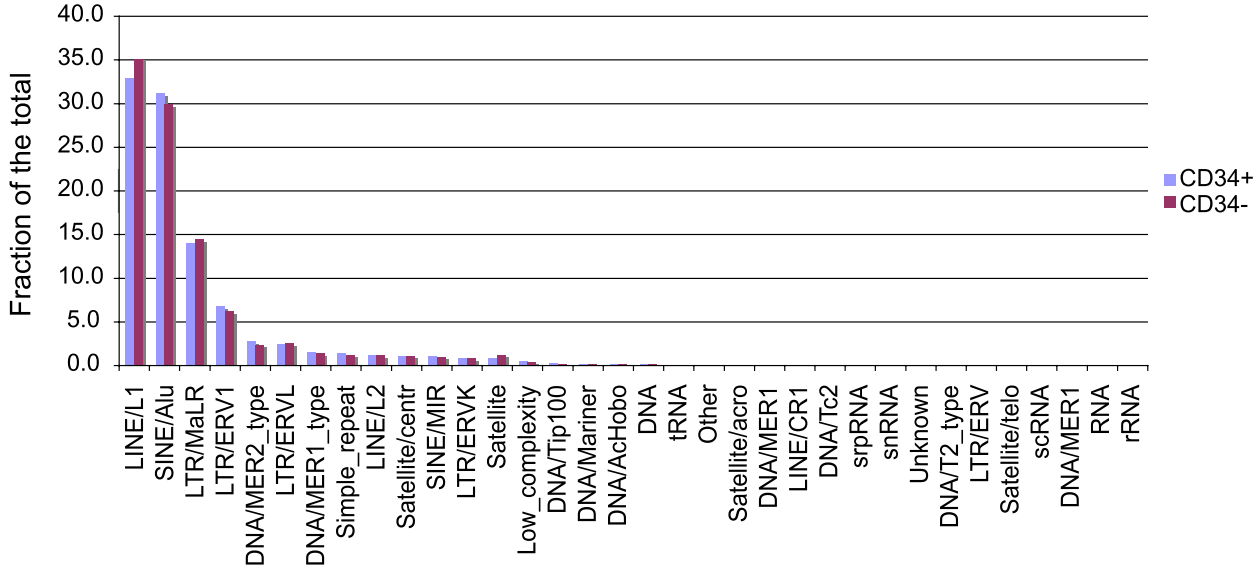
We therefore decided to look at the dynamics of genome accessibility in CD34+ cells triggered to myeloid differentiation. In our gene expression profile, maturing CD34− cells express a decreased number of genes compared to CD34+ cells, though at a higher extent (Figure S8). Importantly, CD34− cells showed strongly decreased accessible promoter regions compared to CD34+ cells (Figure 5A; $\chi^2 = 878.9$, 2 d.f., $p << 0.001$), and likewise the accessibility of the annotated CpG islands (CD34+ cells: 1203/2492, 48.3%; CD34− cells: 869/2492, 34.9%; data not shown). This analysis suggests that CD34+ cells present a broad subset of genes awaiting full activation.

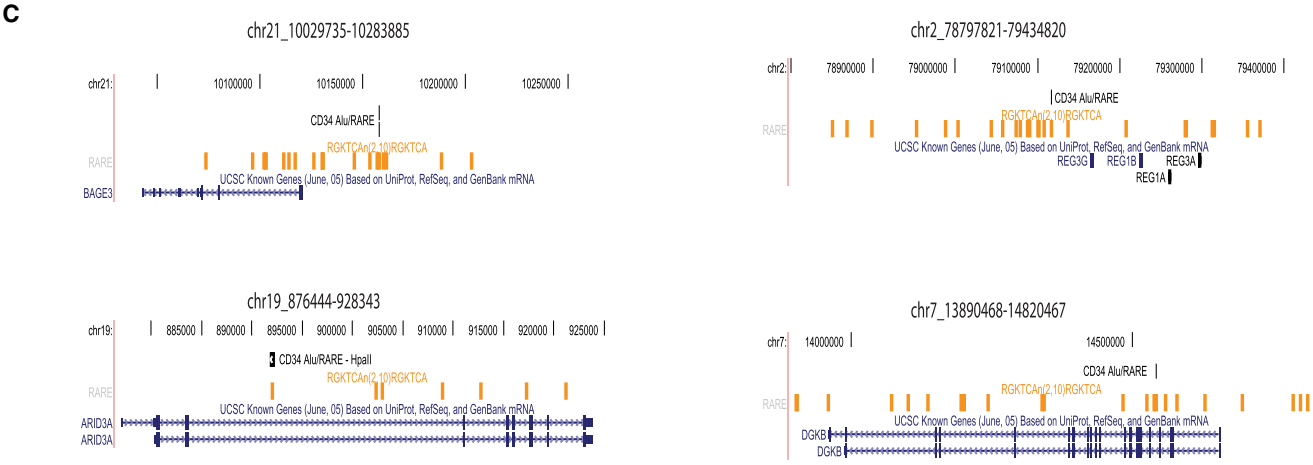
To identify genomic regions showing differential accessibility between CD34+ and CD34− cells, we compared all paired frequencies of NA-Seq libraries with high statistical confidence ($p < 0.001$) using a genomic window of 1.2 Mb that allowed us to overcome potential limitations imposed by the differential degree of sequencing saturation (Figure 1C; see statistical procedures in Supplemental Experimental Procedures). To locate these domains precisely, we set a 10 kb sliding window approach and 307 regions were found undergoing statistically significant changes in accessibility (Figure S9; $p < 0.001$) ranging from 510 kb up to 5 Mb and representing 10.7% of the total genome (see Figure 7A for representative regions; Tables S8 and S9). Of these regions, 191 showed higher accessibility in CD34+ cells, and 116 were found more accessible in CD34− cells. No significant changes were found comparing two runs from the same NAS library (Figure S10). The smaller number of accessible domains found in CD34− cells suggests that cells being committed to a differentiated state along one specific lineage restrict their chromatin accessibility globally, consistent with our results discussed above (Figure S8; Figure 5A).

Next, we constructed a map by weighted differences of chromosomal distribution of the domains accessible in CD34+ and CD34− cells (Figure 7B). Intriguingly, these domains were nonrandomly distributed in the genome and most chromosomes showed reduced accessibility during differentiation, with the relevant exceptions of chromosomes 1, 9, 15, 17, and 19 (Figure 7B).

To characterize these loci, we analyzed whether the genes that are contained within the differentially accessible domains (2047 in CD34+ cells and 737 in CD34− cells) can be linked functionally by means of Gene Ontology categories (GO) and ingenuity


A
Promoters
Intergenic
Transcribed
Enrichment
25.00
20.00
15.00
10.00
5.00
0.00
-5.00
-10.00
-15.00
-20.00
CD34+
Naked
CD34-
8.00
6.00
4.00
2.00
0.00
-2.00
-4.00
-6.00
-8.00
CD34-
CD34+
Naked
15.00
10.00
5.00
0.00
-5.00
-10.00
-15.00
Naked
CD34+
CD34-


B
Fraction of the total
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
CD34+
CD34-
LINE/L1
SINE/Alu
LTR/MaLR
LTR/ERV1
DNA/MER2_type
LTR/ERVL
DNA/MER1_type
Simple_repeat
LINE/L2
Satellite/centr
SINE/MIR
LTR/ERVK
Satellite
Low_complexity
DNA/Tip100
DNA/Mariner
DNA/AcHobo
DNA
tRNA
Other
Satellite/acro
DNA/MER1
LINE/CR1
DNA/Tc2
srpRNA
snRNA
Unknown
DNA/T2_type
LTR/ERV
Satellite/telo
scRNA
DNA/MER1
RNA
rRNA


C
chr21_10029735-10283885
chr21:
10100000
10150000
10200000
10250000
CD34 Alu/RARE
RGKTCAn{2,10}RGKTCA
RARE
UCSC Known Genes (June, 05) Based on UniProt, RefSeq, and GenBank mRNA
BAGE3
chr2_78797821-79434820
chr2:
78900000
79000000
79100000
79200000
79300000
79400000
CD34 Alu/RARE
RGKTCAn{2,10}RGKTCA
RARE
UCSC Known Genes (June, 05) Based on UniProt, RefSeq, and GenBank mRNA
REG3G
REG1B
REG3A
REG1A
chr19_876444-928343
chr19:
885000
890000
895000
900000
905000
910000
915000
920000
925000
CD34 Alu/RARE - HpaII
RGKTCAn{2,10}RGKTCA
RARE
UCSC Known Genes (June, 05) Based on UniProt, RefSeq, and GenBank mRNA
ARID3A
ARID3A
chr7_13890468-14820467
chr7:
14000000
14500000
CD34 Alu/RARE
RGKTCAn{2,10}RGKTCA
RARE
UCSC Known Genes (June, 05) Based on UniProt, RefSeq, and GenBank mRNA
DGKB
DGKB

pathway analysis (IPA). Genes contained in CD34+ open domains are predicted to be mainly involved in metabolic activity, but also in T cell and death receptor signaling (Figure S11). Genes associated with open domains in CD34− cells were, instead, strongly enriched ($p < 0.01$) for nuclear receptor signaling and retinoid metabolism (Figure S11). Consistently, these genes encode for different molecular classes of proteins: CD34+-accessible domains contain genes mainly involved in signal transduction, whereas genes included in CD34− open domains were significantly enriched for TFs (Figure 7C).

To explore the possibility that large-scale changes in genome accessibility correlate with clustered gene activation (known to occur during hematopoietic differentiation; Kosak et al., 2007), we focused on chromosome 19 (chr.19), which showed a marked increase in accessibility during differentiation.

CD34+- and CD34−-derived NAS libraries (using NlaIII as the nuclease probe) were hybridized on a high-resolution DNA array representing the entire nonrepetitive fraction of chr.19. Despite the different technology, we retrieved also in this case a higher fraction of accessible regions in CD34− cells (e.g., Figure 7A, right). Next, we identified clusters of coregulated genes. Most of these coregulated regions containing three or more genes were found upregulated in CD34− cells (Table S10). Importantly, one out of four gene clusters coregulated in chr.19 was part of one of the chromatin domains identified as more accessible in CD34− cells (Figure 7D). None of the coregulated genes (*Dpy19*, *Ankrd27*, and *Nutd19*) was previously reported as involved in myeloid differentiation. Additionally, we found in this accessible domain genes relevant to cell-cycle progression (*Ccne1*) and hematopoietic development, such as *Cebpα* and *Cebpγ*, encoding TFs involved in granulocytic differentiation and deregulated in acute myeloid leukemias (Rosenbauer and Tenen, 2007).

To investigate whether the newly identified domain is coordinately regulated in other cellular contexts, we used an established model of myeloid differentiation of a leukemia cell line. NB4 cells express the leukemia-associated fusion protein PML/RARα: retinoic acid (RA) treatment leads to PML/RARα degradation, triggering terminal differentiation (Di Croce et al., 2002). Analysis of expression levels of five genes located in the domain revealed that three of them were induced following RA treatment (Figure 7D). Therefore, the analysis of one open domain (1) identifies a gene cluster coregulated during differentiation of different cell types; (2) suggests that coordinate opening of chromatin and clustered gene activation may occur at several locations during hematopoietic differentiation; and (3) provides a roadmap of genomic "hot spots" to be further investigated functionally.

Collectively, our results show that NA-Seq is suitable for uncovering large-scale changes in accessibility and that—indeed—global changes in chromosomal organization are a frequent event observed during myeloid differentiation (and probably other biological processes).

DISCUSSION

Several novel findings emerge from the present work: (1) the feasibility of profiling the nuclease-accessible epigenome (including but not limited to "hypersensitive sites") of both transformed and primary cells using REs; (2) the use of information on accessibility to derive notions on the regulatory circuitry imposed on individual loci, and on the genome more broadly; (3) the association between distinct patterns of histone PTMs and DNA accessibility; (4) the discovery—and most notably, their location in the genome—of large-scale transitions in the accessibility of chromosomal domains during cell-fate specification in hematopoiesis; and (5) the relevance of these transitions in imposing spatially constrained patterns of gene regulation during normal differentiation and in leukemic cells.

Mapping Chromatin Accessibility Using NA-Seq Offers a Window on Genome Regulatory Circuitry

We propose that chromatin accessibility to exogenous nucleases offers a synthetic view of nuclear structure, and highlights genomic regions where regulatory processes take place. Accessibility mapping is uniquely poised to present "snapshots" of the regulatory DNA in a given cell, at a given stage, in a sensitive and cost-effective way.

The use of REs abates the background noise frequently found in high-throughput studies using nonspecific nucleases (Boyle et al., 2008a) (Crawford et al., 2004). Sticky-end-mediated identification of NASs allows the use primary cells (available in limited number).

We believe that NA-Seq should be considered as a specific way of looking at chromatin rather than an alternative to DNase-Seq (see also Supplemental Discussion). In fact, NA-Seq identifies elements with no bias for gene promoters (see Supplemental Results), presumably owing to its high signal-to-noise ratio, whereas DHSs are preferentially associated with promoter regions (Boyle et al., 2008b; Crawford et al., 2006). Remarkably, functional enhancers tend to lie at considerable distances from promoters (Carroll et al., 2006).

Nuclease sensitivity provides a continuous pattern of digestion and a measurement of relative, rather than absolute, sensitivity (Hebbes et al., 1994; Weintraub and Groudine, 1976). DHSs represent the highest peak of this continuous state, but regions of extended nuclease sensitivity related to the presence of regulatory elements and of functional events (transcription, in the first place) have been described in most of the systematic studies of single genomic regions reported so far (Hebbes et al., 1994; Lawson et al., 1982). In this study, we have confirmed and extended on a genome-wide scale the original observation that these areas are optimally suited to RE analysis.

Our initial results are consistent with the view that nuclease-accessible elements are in most cases functional elements, because several NASs (though located in intergenic regions) may behave as enhancer regions, or bind the CTCF "insulating"

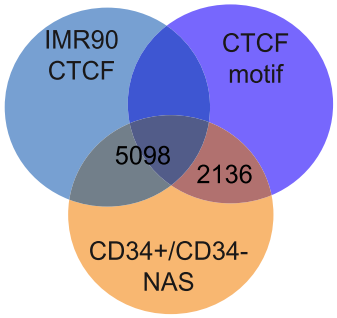
Figure 5. Distribution of NASs in CD34+ and CD34− Cells

(A) Relative enrichment for NASs within each library was calculated by χ^2 distributions.

(B) Classes of repeats accessible in CD34+ and CD34− cells.

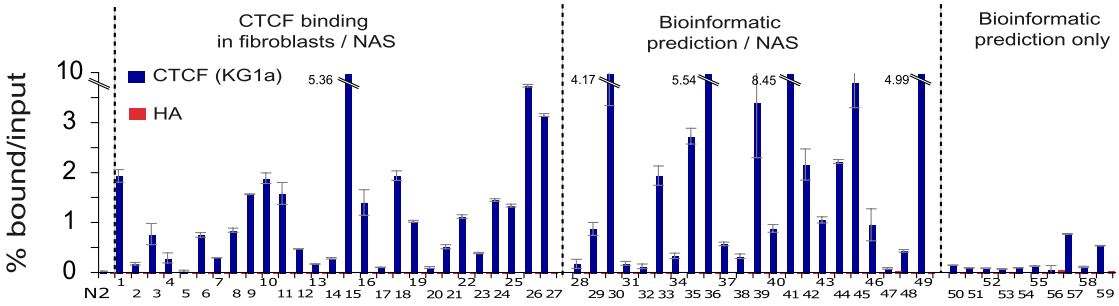
(C) UCSC view of clustered Alu repeats containing the retinoic acid responsive element (RARE; orange vertical lines) close to genes involved in different molecular functions. NASs containing Alu repeats in CD34+ and CD34− cells are also indicated.

A

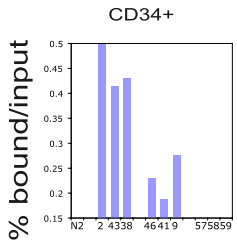


	CD34+	CD34-
CTCF IMR90	3794	2185
CTCF MOTIF	3749	2223

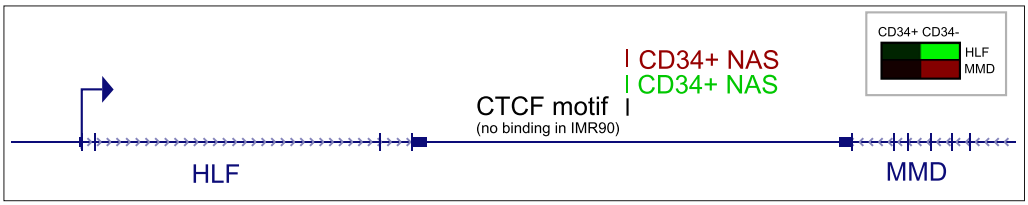
B



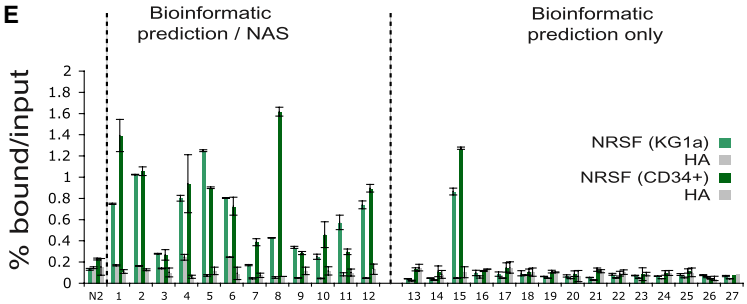
C



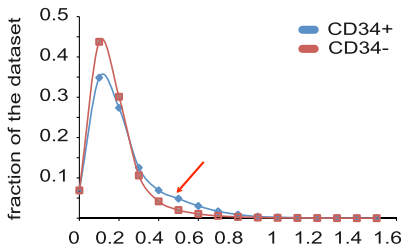
D



E



F



factor. We have shown using several examples, including canonical TFs (NF-Y), insulator-binding factors (CTCF), and transcriptional repressors (NRSF), that in the vast majority of cases analyzed, accessibility of chromatin is coupled to actual binding of TFs to DNA. We propose therefore that for a characterization of any cell state, in terms of occupancy of regulatory DNA by TFs, it is possible to achieve in one single test (accessibility) coupled to DNA sequence analysis the visualization (with an excellent approximation) of the entire pattern of DNA binding of TFs (the "regulome"). The finding that the neuronal gene repressor NRSF binds to NASs is of great interest (Figure 6D), confirming on a genome-wide scale the finding that a few previously analyzed NRSF binding sites are hypersensitive in cell types other than neurons (Loo and Rine, 1994). It is possible that to maintain stable repression of neuronal genes in loci where heterochromatin formation (although more stable and less expensive in terms of energy) would affect surrounding genes, cells have evolved mechanisms to control their expression in a manner that is compatible with maintenance of an accessible state. Furthermore, we found that also a specific family of accessible repeats (Alu) can bind TFs such as RARα. The limited number of accessible and unmethylated repeated elements discovered using NA-Seq might suggest a functional role in genome regulation.

One important nuclear function (gene expression) has been clearly linked to the presence of accessible chromatin. We have confirmed that transcriptionally active regions (like gene promoters) are being selected in our assay, in a tissue-specific manner (Figure 4C; see examples in Figure S12).

Importantly, we observed quantitative differences in the number of promoters accessible in CD34+ and CD34− cells. More immature CD34+ cells—endowed with multilineage potential—have a statistically significant higher number of accessible promoters (including CpG islands) than in vitro differentiated myeloid cells. Cell-type-specific gene ontology analyses reveal that accessible promoters in CD34+ cells are more consistently associated with developmental pathways and with hematopoiesis than in CD34− cells (Figure S2); these observations might be related to the relevance of instructive signals for the "multipotent" CD34+ cells to start differentiation along specific lineages, whereas CD34− cells (which are restricted in their fate) need to execute selected transcriptional programs.

One possible explanation for the higher number of accessible promoters in less differentiated cells is the coexistence of both activating (H3K4me2/3) and repressive histone marks on specific gene promoters (H3K27me3). This "bivalent" signature in ES cells is believed to maintain chromatin in a poised state for activation (Bernstein et al., 2006). Bivalent promoters exist in CD34+ cells (Cui et al., 2009), and we show that they are accessible to nucleases. We also showed that genes previously associated with lineage commitment in mouse multipotent progenitors bearing the histone PTMs H3K4me2+/H3K4me3− (Orford et al., 2008) may exist and be accessible in HSPCs (Figure 3D).

Notably, we report that also NAS nonoverlapping promoters can be simultaneously associated with "bivalent" histone PTMs, in agreement with a large study of histone PTMs in CD4+ cells, where a number of enhancers can be concurrently associated with DHSs, high levels of expression of their associated genes, and the repressive histone mark H3K27me3 (Wang et al., 2008). Indeed, TF binding (e.g., CTCF) and H3K27me3 were not necessarily mutually exclusive (Chen et al., 2008b). One potential explanation for these data is that one single repressive modification (like the Polycomb-dependent H3K27me3) may not be sufficient to counteract chromatin accessibility; indeed, we observed greatly reduced accessibility in the cases of overlapping H3K27me3 and H3K9me3 (Figures 3A and 3D).

Cellular Differentiation Is Associated with Large-Scale Transitions in Genome Accessibility

Higher-order chromatin structures and their dynamics during the execution of various nuclear processes are poorly characterized, due to the relative lack of appropriate methods (Fraser and Bickmore, 2007). To this end, the ability of NA-Seq for measuring differential accessibility is of particular value.

NA-Seq allowed us to define specific chromosomal domains of differential accessibility between CD34+ cells undergoing myeloid-lineage commitment (Figure S9). Several pieces of evidences in our study suggest that the structural organization of the genome within "blocks" reflects their developmental role and potential (see also Figure S2). Some of the differentially accessible domains identified in this study may be linked to clustered gene regulation, usually observed in hematopoietic differentiation (Kosak et al., 2007). However, our data suggest that large-scale changes may not solely reflect clustered gene activation. Other chromosomal dynamics may be dictating changes in nuclear architecture over time: multistep mechanisms regulating the interactions between distinct chromatin loci that occur prior to gene activation (Spilianakis et al., 2005), and additional events other than gene expression such as DNA replication (Farkash-Amar et al., 2008) and repositioning among chromosomal territories (Takizawa et al., 2008).

Figure 6. Different Classes of Transcription Factors Bind NASs In Vivo

(A) Venn diagram showing the overlap between CTCF binding sites in human fibroblast IMR90 cells (light blue), CD34+ and CD34− NASs (orange), and in silico predicted CTCF motifs (blue). The number of actual CD34+ and CD34− NASs containing a CTCF site is reported in the accompanying table.

(B) qChIP analysis of CTCF occupancy in KG1a cells. Vertical dashed lines separate the indicated groups. Enrichments were normalized to the levels of input chromatin, before immunoprecipitation. ChIP with an antibody against hemagglutinin (HA) served as negative control.

(C) A selection of CTCF sites from the second and third groups was validated by ChIP in primary CD34+ cells.

(D) Schematic representation of a candidate CD34+/CD34− tissue-specific insulator (number 46). The heat map shows the normalized mRNA levels for HLF and MMD during myeloid differentiation.

(E) Promoter regions of neuronal genes validated by qChIP in KG1a and primary CD34+ cells for NRSF binding. The vertical dashed line separates the indicated groups. N2 and HA served as negative controls. Error bars in KG1a and CD34+ cells represent standard deviations in three and two independent ChIP experiments, respectively.

(F) Enrichment of TFBSs in CD34+ or CD34− NASs using a positional weight matrix. Note the higher density of TFBSs observed in CD34+ cells (arrow).

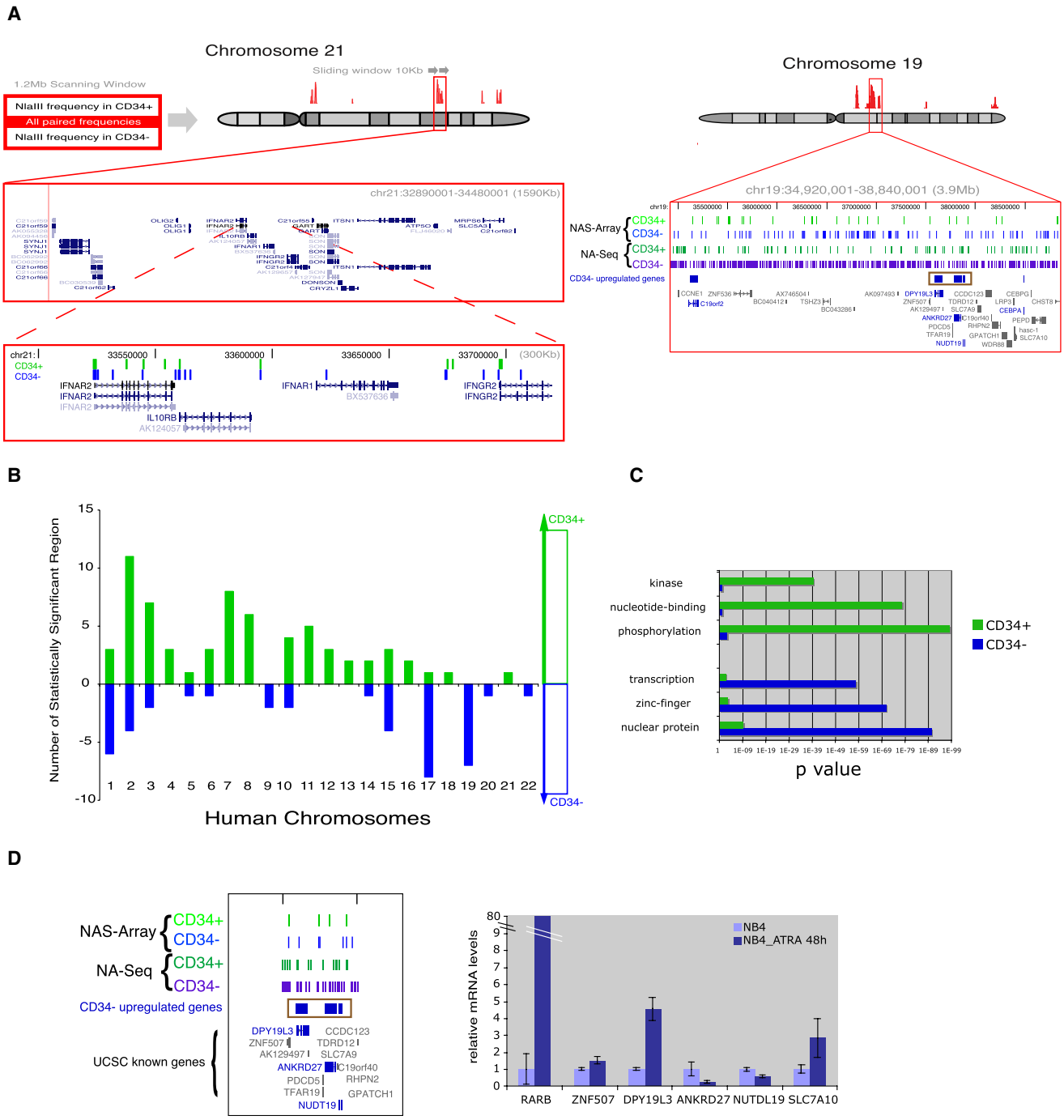


Figure 7. Characterization of Differentially Accessible Chromatin Domains

(A) (Top) Schematic representation of the 1.2 Mb window scanning in 10 kb steps along each chromosome the frequency of NlaIII NASs in CD34+ and CD34− cells (NlaIII distribution along the genome is nonbiased). The red smoothed track is the unifying p value resulting from the χ^2 test. Below, a UCSC genome browser view of a region of increased accessibility in CD34+ cells (left) in chr.21, and a region showing increased accessibility in CD34− cells (right) in chr.19. CD34+ NASs (green) and CD34− NASs (blue) are reported. Note in the 300 kb inset that the TSS of each gene is accessible in both cell types. In the right panel, we also report the DNA microarray hybridization of independent CD34+ and CD34− NAS libraries (NAS array, dark green and violet track), showing patterns of accessibility consistent with NA-Seq results.

(B) Chromosomal distribution of regions with differential accessibility between CD34+ and CD34− cells. Weighted difference of the number of accessible regions per chromosome at $p < 0.01$. Chromosomes with a higher number of regions more accessible in CD34+ versus CD34− cells are ranked and then plotted with a positive value on the y axis. The X chromosome was not included in the analysis.

(C) Top categories (vertical axis) obtained using molecular function annotation of DAVID for genes present in differentially accessible chromatin domains present in CD34+/CD34− cells. The horizontal axis reports the p value.

NA-Seq as a Tool to Study Epigenomic Regulation in Development and Disease

From these studies, we conclude that mapping chromatin accessibility constitutes a fundamental tool for the study of the epigenome, and will help to tackle several unresolved issues.

One important consideration to account for when studying the epigenome of human primary cells is that, in principle, these cells may be heterogeneous between donors. Although our own study does not directly address this point, we note that the genome-wide accessibility landscape observed during differentiation of CD34+ cells is unlikely to change significantly among individuals because of the stringent statistics employed. In addition, experimental validation analyses performed on samples from independent donors showed highly consistent results among samples. Nevertheless, it remains possible that accessibility of specific loci may differ among individuals, as a consequence of genetic and epigenetic variations.

Global changes in gene expression accompanied by chromatin reorganization must also occur in biological processes other than cellular differentiation, including tumorigenesis. Our technology could potentially be applied to the integration of regulated gene transcription and chromatin dynamics also in response to other stimuli, notably oncogene activation. The limited background of NA-Seq experiments suggests that next-generation sequencing approaches could in principle be used to compare chromatin accessibility for several cell types in a single experiment. This represents an important opportunity for epigenomic studies.

EXPERIMENTAL PROCEDURES

A detailed description of experimental methods and materials can be found in Supplemental Data. All raw and processed data are being deposited at the UCSC genome browser at http://genome.ucsc.edu and the Gene Expression Omnibus at http://www.ncbi.nlm.nih.gov/geo (accession number GSE11092). Crossanalyses of all data sets were conducted by using the UCSC table browser, and can be readily accessed for use there.

Cell Culture

Primary cells were obtained from healthy donors (stimulated with G-CSF) according to local ethical guidelines. A total of seven donors was used to derive independent samples for the experiments. CD34+ cells were isolated from leukaphereses using the EasySep magnetic positive selection procedure (StemCell, Vancouver, BC, Canada). Cell purity was confirmed by FACS analysis, and revealed a purity of >98% (Figure S1A).

Preparation of NAS Libraries

Isolated nuclei were incubated at 37°C for 1 min, and then treated with 1/10th v/v of NlaIII or HpaII (New England Biolabs, 50,000 units/ml) for 5 min at 37°C. The reaction was stopped by addition of EDTA (25 mM). 454 adaptor oligonucleotides were ligated to the DNA fragments and subjected to the 454 amplification/sequencing protocol (Margulies et al., 2005). The modified 454 adaptor sequences were prepared from the A and B adaptor sequences employed by standard 454 technology, but contained either an NlaIII or HpaII recognition sequence on the A adaptor and a Sau3A or NlaIII recognition sequence on the B adaptor.

Motif Search for Transcription Factor Binding Sites

The method to identify the location of TFBSs was employed previously (Levy and Hannenhalli, 2002), using positional weight matrices found in TRANSFAC 7.4 (Matys et al., 2003).

DNaseI PCR and qChIP Assays

qPCR DNaseI analyses of NASs derived from the various libraries was carried out as described (McArthur et al., 2001). qChIP was performed by standard procedures: protocol details and primary antibodies are included in Supplemental Experimental Procedures. Genomic region primers used and outcomes of DNaseI PCR, qChIP, and reporter assays are listed in Supplemental Tables.

Software

Pathway analysis was performed using Ingenuity Pathway Analysis software (http://www.ingenuity.com). Molecular categories were analyzed using DAVID (http://www.david.niaid.nih.gov).

SUPPLEMENTAL DATA

Supplemental Data include 21 figures, 11 tables, Supplemental Results, Supplemental Discussion, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at http://www.cell.com/developmental-cell/supplemental/S1534-5807(09)00076-8.

ACKNOWLEDGMENTS

This work is dedicated to the memory of Alan P. Wolffe, who was an inspiration for the beginning of this study. We thank Gioacchino Natoli, Henk Stunnenberg, and Robert O.J. Weinzierl for critical reading of the manuscript, and Tim Stockwell, Cynthia Pfannkoch, Hamilton O. Smith, and Erica Dumais for helping us in the pilot phase. We also thank Marco Cirò, Simona Segalla, and Ernesto Guccione for helpful comments throughout the work. G.G. is supported by a predoctoral fellowship of the SEMM Foundation. This work is supported by EEC (Epitron), AIRC, MIS, and MIUR (PRIN) funding to S.M. The Congenia-Genextra Group is funded by the EEC and MIUR (FIRB). Author contributions: G.G., S.L., F.D.U., and S.M. conceived the work and developed the methodology. G.G. and S.M. designed the experiments. G.G., M.R., M.B., G.F., S.M.G., F.D.U., N.B., and F.S. performed the experiments. G.B., L.F., M.C., G.G., and S.L. performed the bioinformatic and statistical analyses. G.G., S.L., and S.M. analyzed the data. G.G. and S.M. wrote the manuscript, with comments from all of the coauthors. Competing interests: the authors declare that they have no competing interests.

Received: August 9, 2008
Revised: December 19, 2008
Accepted: February 6, 2009
Published: March 16, 2009

REFERENCES

Almer, A., and Horz, W. (1986). Nuclease hypersensitive regions with adjacent positioned nucleosomes mark the gene boundaries of the PHO5/PHO3 locus in yeast. EMBO J. *5*, 2681–2687.

Archer, T.K., Lefebvre, P., Wolford, R.G., and Hager, G.L. (1992). Transcription factor loading on the MMTV promoter: a bimodal mechanism for promoter activation. Science *255*, 1573–1576.

Attema, J.L., Papathanasiou, P., Forsberg, E.C., Xu, J., Smale, S.T., and Weissman, I.L. (2007). Epigenetic characterization of hematopoietic stem cell differentiation using miniChIP and bisulfite sequencing analysis. Proc. Natl. Acad. Sci. USA *104*, 12371–12376.

(D) Relative mRNA levels (RT-qPCR) of five genes found in a domain more accessible in CD34− cells that showed coordinated upregulation in differentiated CD34− cells versus CD34+ cells ([A], left panel, dark blue boxes). Expression levels were analyzed in NB4 cells (following retinoic acid treatment [ATRA] or left untreated). Error bars represent standard deviations in three independent RT-PCR experiments.

Ballas, N., Grunseich, C., Lu, D.D., Speh, J.C., and Mandel, G. (2005). REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. Cell *121*, 645–657.

Barski, A., Cuddapah, S., Cui, K., Roh, T., Schones, D., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. (2004). Ultraconserved elements in the human genome. Science *304*, 1321–1325.

Bernstein, B., Mikkelsen, T., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bernstein, B., Meissner, A., and Lander, E. (2007). The mammalian epigenome. Cell *128*, 669–681.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Bottardi, S., Aumont, A., Grosveld, F., and Milot, E. (2003). Developmental stage-specific epigenetic control of human β-globin gene expression is potentiated in hematopoietic progenitor cells prior to their transcriptional activation. Blood *102*, 3989–3997.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008a). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311–322.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., and Crawford, G.E. (2008b). High-resolution mapping and characterization of open chromatin across the genome. Cell *132*, 311–322.

Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. *20*, 1123–1136.

Carroll, J., Meyer, C., Song, J., Li, W., Geistlinger, T., Eeckhoute, J., Brodsky, A., Keeton, E., Fertuck, K., Hall, G., et al. (2006). Genome-wide analysis of estrogen receptor binding sites. Nat. Genet. *38*, 1289–1297.

Chen, L., Decerbo, J., and Carmichael, G. (2008a). Alu element-mediated gene silencing. EMBO J. *27*, 1694–1705.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008b). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

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., and Collins, F.S. (2004). Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. Proc. Natl. Acad. Sci. USA *101*, 992–997.

Crawford, G.E., Holt, I.E., Whittle, J., Webb, B.D., Tai, D., Davis, S., Margulies, E.H., Chen, Y., Bernat, J.A., Ginsburg, D., et al. (2006). Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res. *16*, 123–131.

Cui, K., Zang, C., Roh, T., Schones, D., Childs, R.W., Peng, W., and Zhao, K. (2009). Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell *4*, 80–93.

Di Croce, L., Raker, V.A., Corsaro, M., Fazi, F., Fanelli, M., Faretta, M., Fuks, F., Lo Coco, F., Kouzarides, T., Nervi, C., et al. (2002). Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor. Science *295*, 1079–1082.

Elgin, S.C. (1988). The formation and function of DNase I hypersensitive sites in the process of gene activation. J. Biol. Chem. *263*, 19259–19262.

ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J., Dutta, A., Guigó, R., Gingeras, T., Margulies, E., Weng, Z., Snyder, M., Dermitzakis, E., Thurman, R.E., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816.

Farkash-Amar, S., Lipson, D., Polten, A., Goren, A., Helmstetter, C., Yakhini, Z., and Simon, I. (2008). Global organization of replication time zones of the mouse genome. Genome Res. *18*, 1562–1570.

Fraser, P., and Bickmore, W. (2007). Nuclear organization of the genome and the potential for gene regulation. Nature *447*, 413–417.

Gargiulo, G., and Minucci, S. (2009). Epigenomic profiling of cancer cells. Int. J. Biochem. Cell Biol. *41*, 127–135.

Giresi, P., Kim, J., McDaniell, R.M., Iyer, V., and Lieb, J. (2007). FAIRE (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin. Genome Res. *17*, 877–885.

Guccione, E., Martinato, F., Finocchiaro, G., Luzi, L., Tizzoni, L., Dall' Olio, V., Zardo, G., Nervi, C., Bernard, L., and Amati, B. (2006). Myc-binding-site recognition in the human genome is determined by chromatin context. Nat. Cell Biol. *8*, 764–770.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77–88.

Hebbes, T.R., Clayton, A.L., Thorne, A.W., and Crane-Robinson, C. (1994). Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken β-globin chromosomal domain. EMBO J. *13*, 1823–1830.

Heintzman, N., Stuart, R., Hon, G., Fu, Y., Ching, C., Hawkins, R., Barrera, L., Van Calcar, S., Qu, C., Ching, K., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Jäger, R., Werling, U., Rimpf, S., Jacob, A., and Schorle, H. (2003). Transcription factor AP-2γ stimulates proliferation and apoptosis and impairs differentiation in a transgenic model. Mol. Cancer Res. *1*, 921–929.

Johnson, D.S., Mortazavi, A., Myers, R., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497–1502.

Kim, T., Abdullaev, Z.K., Smith, A.D., Ching, K., Loukinov, D.I., Green, R., Zhang, M., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231–1245.

Kosak, S., Scalzo, D., Alworth, S., Li, F., Palmer, S., Enver, T., Lee, J.S., and Groudine, M. (2007). Coordinate gene regulation during hematopoiesis is related to genomic organization. PLoS Biol. *5*, e309.

Laperriere, D., Wang, T.T., White, J.H., and Mader, S. (2007). Widespread Alu repeat-driven expansion of consensus DR2 retinoic acid response elements during primate evolution. BMC Genomics *8*, 23.

Lawson, G.M., Knoll, B.J., March, C.J., Woo, S.L., Tsai, M.J., and O'Malley, B.W. (1982). Definition of 5′ and 3′ structural boundaries of the chromatin domain containing the ovalbumin multigene family. J. Biol. Chem. *257*, 1501–1507.

Levy, S., and Hannenhalli, S. (2002). Identification of transcription factor binding sites in the human genome sequence. Mamm. Genome *13*, 510–514.

Li, X., Macarthur, S., Bourgon, R., Nix, D., Pollard, D., Iyer, V., Hechmer, A., Simirenko, L., Stapleton, M., Hendriks, C., et al. (2008). Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. PLoS Biol. *6*, e27.

Löffler, H., Rastetter, J., Haferlach, T., Heilmeyer, L., and Begemann, H. (2004). Atlas of Clinical Hematology, Sixth Edition (Berlin: Springer).

Loo, S., and Rine, J. (1994). Silencers and domains of generalized repression. Science *264*, 1768–1771.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376–380, Erratum: (2006). Nature *441*, 120.

Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V., et al. (2003). TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res. *31*, 374–378.

McArthur, M., Gerum, S., and Stamatoyannopoulos, G. (2001). Quantification of DNaseI-sensitivity by real-time PCR: quantitative analysis of DNaseI-hypersensitivity of the mouse β-globin LCR. J. Mol. Biol. *313*, 27–34.

Morrison, S.J., Uchida, N., and Weissman, I.L. (1995). The biology of hematopoietic stem cells. Annu. Rev. Cell Dev. Biol. *11*, 35–71.

Orford, K., Kharchenko, P., Lai, W., Dao, M.C., Worhunsky, D.J., Ferro, A., Janzen, V., Park, P.J., and Scadden, D.T. (2008). Differential H3K4 methylation identifies developmentally poised hematopoietic genes. Dev. Cell *14*, 798–809.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell *1*, 299–312.

Piacibello, W., Sanavio, F., Severino, A., Danè, A., Gammaitoni, L., Fagioli, F., Perissinotto, E., Cavalloni, G., Kollet, O., Lapidot, T., and Aglietta, M. (1999). Engraftment in nonobese diabetic severe combined immunodeficient mice of human CD34(+) cord blood cells after ex vivo expansion: evidence for the amplification and self-renewal of repopulating stem cells. Blood *93*, 3736–3749.

Pingoud, A., and Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. Nucleic Acids Res. *29*, 3705–3727.

Rosenbauer, F., and Tenen, D. (2007). Transcription factors in myeloid development: balancing differentiation with transformation. Nat. Rev. Immunol. *7*, 105–117.

Schones, D., Cui, K., Cuddapah, S., Roh, T., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887–898.

Spilianakis, C., Lalioti, M.D., Town, T., Lee, G., and Flavell, R. (2005). Interchromosomal associations between alternatively expressed loci. Nature *435*, 637–645.

Steidl, U., Steidl, C., Ebralidze, A., Chapuy, B., Han, H.J., Will, B., Rosenbauer, F., Becker, A., Wagner, K., Koschmieder, S., et al. (2007). A distal single nucleotide polymorphism alters long-range regulation of the PU.1 gene in acute myeloid leukemia. J. Clin. Invest. *117*, 2611–2620.

Takizawa, T., Meaburn, K.J., and Misteli, T. (2008). The meaning of gene positioning. Cell *135*, 9–13.

Tazi, J., and Bird, A. (1990). Alternative chromatin structure at CpG islands. Cell *60*, 909–920.

Turner, B. (2001). Chromatin and Gene Regulation: Mechanisms in Epigenetics (Oxford: Blackwell Science).

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser—a database of tissue-specific human enhancers. Nucleic Acids Res. *35*, D88–D92.

Wang, Z., Zang, C., Rosenfeld, J., Schones, D., Barski, A., Cuddapah, S., Cui, K., Roh, T., Peng, W., Zhang, M., and Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. *40*, 897–903.

Weintraub, H., and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. Science *193*, 848–856.

Wu, C. (1980). The 5′ ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. Nature *286*, 854–860.

Xi, H., Shulha, H.P., Lin, J., Vales, T.R., Fu, Y., Bodine, D.M., McKay, R.D., Chenoweth, J.G., Tesar, P.J., Furey, T.S., et al. (2007). Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. PLoS Genet. *3*, e136.