The accessible chromatin landscape of the human genome

Robert E. Thurman1,*, Eric Rynes1,*, Richard Humbert1, Jeff Vierstra1, Matthew T. Maurano1, Eric Haugen1, Nathan C. Sheffield2, Andrew B. Stengachis1, Hao Wang1, Benjamin Vernet1, Kavita Garg1, Sam John1, Richard Sandstrom1, Daniel Bates1, Lisa Boatman1, Theresa K. Canfield1, Morgan Diegel1, Douglas Dunn1, Abigail K. Ebersol4, Tristan Frum1, Erika Giste1, Audra K. Johnson1, Ericka M. Johnson1, Tanya Kutyavin1, Bryan Lajoie5, Bum-Kyu Lee6, Kristen Lee1, Darin London2, Dimitra Lotakis4, Shane Nep1, Fidencio Neri1, Eric D. Nguyen6, Hongzhu Qu17, Alex P. Reynolds1, Vaughn Roach1, Alexias Saffi2, Minerva E. Sanchez4, Amartya Sanyal5, Anthony Shafer1, Jeremy M. Simon8, Lingyun Song2, Shinny Vong1, Molly Weaver1, Yongqi Yan4, Zhancheng Zhang8, Zhuzhu Zhang8, Boris Lenhard9, Muneesh Tewari3, Michael O. Dorschner10, R. Scott Hansen4, Patrick A. Navas4, George Stamatoyannopoulos4, Vishwanath R. Iyer6, Jason D. Lieb8, Shamil R. Sunyaev11, Joshua M. Akey1, Peter J. Sabo1, Rajinder Kaul1, Terrence S. Furey8, Job Dekker5, Gregory E. Crawford2 & John A. Stamatoyannopoulos1,12

DNase I hypersensitive sites (DHSs) are markers of regulatory DNA and have underpinned the discovery of all classes of cis-regulatory elements including enhancers, promoters, insulators, silencers and locus control regions. Here we present the first extensive map of human DHSs identified through genome-wide profiling in 125 diverse cell and tissue types. We identify ~2.9 million DHSs that encompass virtually all known experimentally validated cis-regulatory sequences and expose a vast trove of novel elements, most with highly cell-selective regulation. Annotating these elements using ENCODE data reveals novel relationships between chromatin accessibility, transcription, DNA methylation and regulatory factor occupancy patterns. We connect ~580,000 distal DHSs with their target promoters, revealing systematic pairing of different classes of distal DHSs and specific promoter types. Patterning of chromatin accessibility at many regulatory regions is organized with dozens to hundreds of co-activated elements, and the transcellular DNase I sensitivity pattern at a given region can predict cell-type-specific functional behaviours. The DHS landscape shows signatures of recent functional evolutionary constraint. However, the DHS compartment in pluripotent and immortalized cells exhibits higher mutation rates than that in highly differentiated cells, exposing an unexpected link between chromatin accessibility, proliferative potential and patterns of human variation.

Cell-selective activation of regulatory DNA drives the gene expression patterns that shape cell identity. Regulatory DNA is characterized by the cooperative binding of sequence-specific transcriptional regulatory factors in place of a canonical nucleosome, leading to a remodelled chromatin state characterized by markedly heightened accessibility to nucleases. DNase I hypersensitive sites (DHSs) in chromatin were first identified over 30 years ago, and have since been used extensively to map regulatory DNA regions in diverse organisms. DNase I hypersensitivity is central to all defined classes of active cis-regulatory elements including enhancers, promoters, silencers, insulators and locus control regions. Because DNase I hypersensitivity overlies cis-regulatory elements directly and is maximal over the core region of regulatory factor occupancy, it enables precise delineation of the genomic cis-regulatory compartment. DHSs are flanked by nucleosomes, which may acquire histone modification patterns that reflect the functional role of the adjoining regulatory DNA, such as the association of histone H3 lysine 4 trimethylation (H3K4me3) with promoter elements. Recent advances have enabled genome-scale mapping of DHSs in mammalian cells, laying the foundations for comprehensive catalogues of human regulatory DNA.

Two ENCODE production centres (University of Washington and Duke University) profiled DNase I sensitivity genome-wide using massively parallel sequencing in a total of 125 human cell and tissue types including normal differentiated primary cells (n = 71), immortalized primary cells (n = 16), malignancy-derived cell lines (n = 30) and multipotent and pluripotent progenitor cells (n = 8) (Supplementary Table 1). The density of mapped DNase I cleavages as a function of genome position provides a continuous quantitative measure of chromatin accessibility, in which DHSs appear as prominent peaks within the signal data from each cell type (Fig. 1a and Supplementary Figs 1 and 2). Analysis using a common algorithm (see Methods) identified 2,890,742 distinct high-confidence DHSs (false discovery rate (FDR) of 1%; see Methods), each of which was active in one or more cell types. Of these DHSs, 970,100 were specific to a single cell type, 1,920,642 were active in 2 or more cell types, and a

*These authors contributed equally to this work.
small minority (3,692) was detected in all cell types. The relative accessibility of DHSs along the genome varies by >100-fold and is highly consistent across cell types (Supplementary Figs 1 and 2). To estimate the sensitivity and accuracy of the sequencing-derived DHS maps, one ENCODE production centre (University of Washington) performed 7,478 classical DNase I hypersensitivity experiments by cleavage sites for selected cell types, shown for an example that the overall sensitivity for DHSs of the combined cell type maps is >98%.

Approximately 3% (n = 75,575) of DHSs localize to transcriptional start sites (TSSs) defined by GENCODE and 5% (n = 135,735, including the aforementioned) lie within 2.5 kilobases (kb) of a TSS. The remaining 95% of DHSs are positioned more distally, and are roughly evenly divided between intronic and intergenic regions (Fig. 1b). Promoters typically exhibit high accessibility across cell types, with the average promoter DHS detected in 29 cell types (Fig. 1c, second column). By contrast, distal DHSs are largely cell selective (Fig. 1c, third column).

MicroRNAs (miRNAs) comprise a major class of regulatory molecules and have been extensively studied, resulting in consensus annotation of hundreds of conserved miRNA genes, approximately one-third of which are organized in polycistronic clusters. However, most predicted promoters driving microRNA expression lack experimental evidence. Of 329 unique annotated miRNA TSSs (Supplementary Methods), 300 (91%) either coincided with or closely approximated (<500 base pairs (bp)) a DHS. Chromatin accessibility at miRNA promoters was highly promiscuous compared with GENCODE TSSs (Fig. 1c, fourth column), and showed cell lineage organization, paralleling the known regulatory roles of well-annotated lineage-specific miRNAs (Supplementary Fig. 3).

The 20–50-bp read lengths from DNase I-seq experiments enabled unique mapping to 86.9% of the genomic sequence, allowing us to interrogate a large fraction of transposon sequences. A surprising number contain highly regulated DHSs (Fig. 1c, fifth column and Supplementary Figs 4 and 5), compatible with cell-specific transcription of repetitive elements detected using ENCODE RNA sequencing data. DHSs were most strongly enriched at long terminal repeat (LTR) elements, which encode retroviral enhancer structures (Supplementary Table 2). Two such examples are shown in Supplementary Fig. 4, which also illustrates the strong cell-selectivity of chromatin accessibility seen for each major repeat class. We also documented numerous examples of transposon DHSs that displayed enhancer activity in transient transfection assays (Supplementary Table 3).

Comparison with an extensive compilation of 1,046 experimentally validated distal, non-promoter cis-regulatory elements (enhancers, insulators, locus control regions, and so on) revealed the overwhelming majority (97.4%) to be encompassed within DNase I hypersensitive chromatin (Supplementary Table 4), typically with strong cell selectivity (Supplementary Fig. 2b).

Transcription factor drivers of chromatin accessibility
DNase I hypersensitive sites result from cooperative binding of transcription factors in place of a canonical nucleosome. To quantify the relationship between chromatin accessibility and the occupancy of regulatory factors, we compared sequencing-depth-normalized DNase I sensitivity in the ENCODE common cell line K562 to normalized chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq) signals from all 42 transcription factors mapped by ENCODE ChIP-seq14 in this cell type (Fig. 2). Simple summation of the ChIP-seq signals markedly parallels quantitative DNase I sensitivity at individual DHSs (Fig. 2a) and across the genome (r = 0.79, Fig. 2b). For example, the β-globin locus control region contains a major enhancer element at hypersensitive site 2 (HS2), which appears to be occupied by dozens of transcription factors (Supplementary Fig. 6a). Such highly overlapping binding patterns have been interpreted to signify weak interactions with lower-affinity recognition sequences potentiated by an accessible DNA template. However, HS2 is a compact element with a functional core spanning ~110 bp that contains 5–8 sites of transcription factor–DNA interaction in vivo depending on the cell type. The fact that the cumulative ChIP-seq signal closely parallels the degree of nuclease sensitivity at HS2 and elsewhere is thus most readily explained by interactions between DNA-bound factors.
targeted mass spectrometry assays for KAP1 and three factors binding within compacted heterochromatin. To test this, we developed that a proportion of the occupancy sites of these factors represented TRIM28), SETDB1 and ZNF274 (refs 20, 21) (Fig. 2c). We hypothesized (Fig. 2c and Supplementary Fig. 7a), with the median factor having accessibility landscape.

that regulators acting through these sequences are key drivers of the all classes of sites and all cell types (Supplementary Fig. 6c), indicating consistently linked with elevated chromatin accessibility across found that the recognition sequences for a small number of factors were others may interact nonspecifically with the remodelled state. We also in establishment and maintenance of chromatin remodelling, whereas and other interacting factors that collectively potentiate the accessible chromatin state (Supplementary Fig. 6b). Given the relatively limited number of factors studied, it may seem surprising that such a close correlation should be evident. However, most of the factors selected for ENCODE ChIP-seq studies have well-described or even fundamental roles in transcriptional regulation, and many were identified originally based on their high affinity for DNA. Alternatively, as originally proposed in ref. 19, a limited number of factors may be involved in establishment and maintenance of chromatin remodelling, whereas others may interact nonspecifically with the remodelled state. We also found that the recognition sequences for a small number of factors were consistently linked with elevated chromatin accessibility across all classes of sites and all cell types (Supplementary Fig. 6c), indicating that regulators acting through these sequences are key drivers of the accessibility landscape.

Overall, 94.4% of a combined 1,108,081 ChIP-seq peaks from all ENCODE transcription factors fall within accessible chromatin (Fig. 2a and Supplementary Fig. 7a), with the median factor having 98.2% of its binding sites localized therein. Notably, a small number of factors diverged from this paradigm, including known chromatin repressors, such as the KRAB-associated factors KAP1 (also called TRIM28), SETDB1 and ZNF274 (refs 20, 21) (Fig. 2c). We hypothesized that a proportion of the occupancy sites of these factors represented binding within compacted heterochromatin. To test this, we developed targeted mass spectrometry assays for KAP1 and three factors localizing almost exclusively within accessible chromatin (GATA1, c-Jun, NRF1), and quantified their abundance in biochemically defined heterochromatin against a total chromatin fraction (Supplementary Fig. 7b). This analysis confirmed that factors such as KAP1 show a significant level of heterochromatin occupancy (Supplementary Fig. 7c).

An invariant directional promoter chromatin signature

The annotation of sites of transcription origination continues to be an active and fundamental endeavour. In addition to direct evidence of TSSs provided by RNA transcripts, H3K4me3 modifications are closely linked with TSSs. We therefore explored systematically the relationship between chromatin accessibility and H3K4me3 patterns at well-annotated promoters, its relationship to transcription origination, and its variability across ENCODE cell types.

We performed ChIP-seq for H3K4me3 in 56 cell types using the same biological samples used for DNase I data (Supplementary Table 1, column D). Plotting DNase I cleavage density against ChIP-seq tag density around TSSs reveals highly stereotyped, asymmetrical patterning of these chromatin features with a precise relationship to the TSS (Fig. 3a, b). This directional pattern is consistent with a rigidly positioned nucleosome immediately downstream from the promoter DHS, and is largely invariant across cell types (Fig. 3b and Supplementary Fig. 8).

To map novel promoters (and their directionality) not encompassed by the GENCODE consensus annotations, we applied a pattern-matching approach to scan the genome across all 56 cell types (Supplementary Methods). Using this approach we identified a total of 113,622 distinct putative promoters. Of these, 68,769 correspond to previously annotated TSSs, and 44,853 represent novel predictions (versus GENCODE v7). Of the novel sites, 99.5% are supported by evidence from spliced expressed sequence tags (ESTs) and/or cap analysis of gene expression (CAGE) tag clusters (Fig. 3c and Supplementary Fig. 9, P < 0.0001; see Supplementary Methods). We found novel sites in every configuration relative to existing annotations (Fig. 3d–f and Supplementary Fig. 10). For example, 29,203 putative promoters are contained in the bodies of annotated genes, of which 17,214 are oriented antisense to the annotated direction of transcription, and 2,794 lie immediately downstream of an annotated gene’s 3’ end, with 1,638 in antisense orientation. The results indicate that chromatin data can systematically inform RNA transcription analyses, and suggest the existence of a large pool of cell-selective transcriptional promoters, many of which lie in antisense orientations.

Chromatin accessibility and DNA methylation patterns

CpG methylation has been closely linked with gene regulation, based chiefly on its association with transcriptional silencing. However, the relationship between DNA methylation and chromatin structure has not been clearly defined. We analysed ENCODE reduced-representation bisulphite sequencing (RRBS) data, which provide quantitative methylation measurements for several million CpGs (K. E. Varley et al., manuscript submitted; see Gene Expression Omnibus accession GSE27584). We focused on 243,037 CpGs falling within DHSs in 19 cell types for which both data types were available from the same sample. We observed two broad classes of sites: those with a strong inverse correlation across cell types between DNA methylation and chromatin accessibility (Fig. 4a and Supplementary Fig. 11a), and those with variable chromatin accessibility but constitutive hypomethylation (Fig. 4a, right). To quantify these trends globally, we performed a linear regression analysis between chromatin accessibility and DNA methylation at the 34,376 CpG-containing DHSs (see Supplementary Methods). Of these sites, 6,987 (20%) showed a significant association (1% FDR) between methylation and accessibility (Supplementary Fig. 11b). Increased methylation was almost uniformly negatively associated with chromatin accessibility (>97% of cases). The magnitude of the association between methylation and accessibility was strong, with the latter on
Figure 3 | Identification and directional classification of novel promoters. a, DNase I (blue) and H3K4me3 (red) tag densities for K562 cells around annotated TSS of ACTR3B. b, Averaged H3K4me3 tag density (red, right y axis) and log DNase I tag density (blue, left y axis) across 10,000 randomly selected GENCODE TSSs, oriented $5' \rightarrow 3'$. Each blue and red curve is for a different cell type, showing invariance of the pattern. c, Relation of 113,615 promoter average 95% lower in cell types with coinciding methylation versus cell types lacking coinciding methylation (Supplementary Fig. 11c). Fully 40% of variable methylation was associated with a concomitant effect on accessibility.

The role of DNA methylation in causation of gene silencing is presently unclear. Does methylation reduce chromatin accessibility by evicting transcription factors? Or does DNA methylation passively ‘fill in’ the voids left by vacating transcription factors? Transcription factor expression is closely linked with the occupancy of its binding site sequences should be negatively correlated with transcription factor gene expression. If the latter, methylation at transcription factor recognition sequences should be independent of transcription factor occupancy. c, Relationship between transcription factor transcript levels and overall methylation at cognate recognition sequences of the same transcription factors. Lymphoid regulators in B-lymphoblastoid line GM06990 (left) and erythroid regulators in the erythroleukaemia line K562 (right). Negative correlation indicates that site-specific DNA methylation follows transcription factor vacation of differentially expressed transcription factors.
A map of distal DHS–to–promoter connections

From examination of DNase I profiles across many cell types we observed that many known cell-selective enhancers become DHSs synchronously with the appearance of hypersensitivity at the promoter of their target gene (Supplementary Fig. 13). To generalize this, we analysed the patterning of 1,454,901 distal DHSs (DHSs separated from a TSS by at least one other DHS) across 79 diverse cell types (Supplementary Methods and Supplementary Table 6), and correlated the cross-cell-type DNase I signal at each DHS position with that at all promoters within ±500 kb (Supplementary Fig. 14a). We identified a total of 578,905 DHSs that were highly correlated (r > 0.7) with at least one promoter ($P < 10^{-10}$), providing an extensive map of candidate enhancers controlling specific genes (Supplementary Methods and Supplementary Table 7). To validate the distal DHS/ enhancer–promoter connections, we profiled chromatin interactions using the chromosome conformation capture carbon copy (5C) technique.$^{30}$ For example, the phenylalanine hydroxylase (PAH) gene is expressed in hepatic cells, and an enhancer has been defined upstream of its TSS (Fig. 5a). The correlation values for three DHSs within the gene body closely parallel the frequency of long-range chromatin interactions measured by 5C. The three interacting intrinsic DHSs cloned downstream of a reporter gene driven by the PAH promoter all showed increased expression ranging from three- to tenfold over a promoter-only control, confirming enhancer function.

We next examined comprehensive promoter-versus–all 5C experiments performed over 1% of the human genome$^{31}$ in K562 cells. DHS–promoter pairings were markedly enriched in the specific cognate chromatin interaction ($P < 10^{-13}$, Supplementary Fig. 14b). We also examined K562 promoter–DHS interactions detected by polymerase II chromatin interaction analysis with paired-end tag sequencing (ChiA-PET)$^{24}$, which quantifies interactions between promoter-bound polymerase and distal sites. The ChiA-PET interactions were also markedly enriched for DHS–promoter pairings ($P < 10^{-15}$ Supplementary Fig. 14c). Together, the large-scale interaction analyses affirm the fidelity of DHS–promoter pairings based on correlated DNase I sensitivity signals at distal and promoter DHSs.

Most promoters were assigned to more than one distal DHS, indicating the existence of combinatorial distal regulatory inputs for most genes (Fig. 5b and Supplementary Table 7). A similar result is forthcoming from large-scale 5C interaction data$^{31}$. Surprisingly, roughly half of the promoter-paired distal DHSs were assigned to more than one promoter (Fig. 5b and Supplementary Methods), indicating that human cis-regulatory circuitry is significantly more complicated than previously anticipated, and may serve to reinforce the robustness of cellular transcriptional programs.

The number of distal DHSs connected with a particular promoter provides, for the first time, a quantitative measure of the overall regulatory complexity of that gene. We asked whether there are any systematic functional features of genes with highly complex regulation. We ranked all human genes by the number of distal DHSs paired with the promoter of each gene, then performed a Gene Ontology analysis on the rank-ordered list. We found that the most complexly regulated human genes were markedly enriched in immune system functions (Supplementary Fig. 14d), indicating that the complexity of cellular and environmental signals processed by the immune system is directly encoded in the cis-regulatory architecture of its constituent genes.

Next, we asked whether DHS–promoter pairings reflected systematic relationships between specific combinations of regulatory factors (Supplementary Methods). For example, KLF4, SOX2, OCT4 (also called POU5F1) and NANOG are known to form a well-characterized transcriptional network controlling the pluripotent state of embryonic stem cells. We found significant enrichment ($P < 0.05$) of the KLF4, SOX2 and OCT4 motifs within distal DHSs correlated with promoter DHSs containing the NANOG motif; enrichment of NANOG, SOX2 and OCT4 distal motifs co-occurring with promoter motif OCT4; and enrichment of distal SOX2 and OCT4 motifs with promoter SOX2 motifs (Supplementary Fig. 15a). By contrast, promoters containing KLF4 motifs were associated with KLF4-containing distal DHSs, but not with DHSs containing NANOG, SOX2 or OCT4 motifs (Supplementary Fig. 15a, bottom).

We also tested for significant co-associations between promoter types (defined by the presence of cognate motif classes; see Supplementary Methods) and motifs in paired distal DHSs (Fig. 5c and Supplementary Fig. 15b, c). For example, when a member of the ETS domain family (motifs ETS1, ETS2, ELF1, ELK1, NRF2 (also called ELF2), SP1, and others) is present within a promoter DHS, motif PU.1 (also called SPI1) is significantly more likely to be observed in a correlated distal DHS ($P < 10^{-7}$). These results suggest that a limited set of general rules may govern the pairing of co-regulated distal DHSs with particular promoters.

Figure 5 | A genome-wide map of distal DHS-to-promoter connectivity. a. Cross-cell-type correlation (red arcs, left y axis) of distal DHSs and PAH promoter closely parallels chromatin interactions measured by 5C-seq (blue arcs, right y axis); black bars indicate HindIII fragments used in 5C assays. Known (green) and novel (magenta) enhancers confirmed in transfection assays are shown below. Enhancer at far right is not separable by 5C as it lies within the HindIII fragment containing the promoter. b. Left: proportions of 69,965 promoters correlated (r > 0.7) with 0 to 3 promoter DHSs within 500 kb. Right: proportions of 578,905 non-promoter DHSs (out of 1,454,901) correlated with 1 to 3 promoters within 500 kb. c. Co-association of motif families with specific motifs in distal DHSs.
Stereotyped chromatin accessibility parallels function

In addition to the synchronized activation of distal DHSs and promoters described above, we observed a surprising degree of patterned co-activation among distal DHSs, with nearly identical cross-cell-type patterns of chromatin accessibility at groups of DHSs widely separated in trans (Supplementary Figs 16 and 17). For many patterns, we observed tens or even hundreds of like elements around the genome. The simplest explanation is that such co-activated sites share recognition motifs for the same set of regulatory factors. We found, however, that the underlying sequence features for a given pattern were surprisingly plastic. This suggests that the same pattern of cell-selective chromatin accessibility shared between two DHSs can be achieved by distinct mechanisms, probably involving complex combinatorial tuning.

We next asked whether distal DHSs with specific functions such as enhancers exhibited stereotypical patterning, and whether such patterning could highlight other elements with the same function. We examined one of the best-characterized human enhancers, DNase I HS2 of the β-globin locus control region16–18. HS2 is detected in many cell types, but exhibits potent enhancer activity only in erythroid cells24. Using a pattern-matching algorithm (see Supplementary Methods) we identified additional DHSs with nearly identical cross-cell-type accessibility patterns (Fig. 6a). We selected 20 elements across the spectrum of the top 200 matches to the HS2 pattern, and tested these in transient transfection assays in K562 cells (Supplementary Methods). Seventy per cent (14 of 20) of these displayed enhancer activity (mean 8.4-fold over control) (Fig. 6a, f). Of note, one (E3) showed a greater magnitude of enhancement (18-fold versus control) than HS2, which is itself one of the most potent known enhancers24. Next we selected three elements from the 14 HS2-like enhancers, applied pattern matching (Methods) to each to identify stereotyped elements, and tested samples of each pattern for enhancer activity, revealing additional K562 enhancers (total 15 of 25 positive) (Fig. 6b–d, f). In each case, therefore, we were able to discover enhancers by simply anchoring on the cross-cell-type DHS pattern of an element with an enhancer activity. Collectively, these results show that co-activation of DHSs reflected in cross-cell-type patterning of chromatin accessibility is predictive of functional activity within a specific cell type, and suggest more generally that DHSs with stereotyped cellular patterning are likely to fulfil similar functions.

To visualize the qualities and prevalence of different stereotyped cross-cellular DHS patterns, we constructed a self-organizing map of a random 10% subsample of DHSs across all cell types and identified a total of 1,225 distinct stereotyped DHS patterns (Supplementary Figs 18 and 19). Many of the stereotyped patterns discovered by the self-organizing map encompass large numbers of DHSs, with some containing >1,000 elements (Supplementary Fig. 20).

Taken together, the above results show that chromatin accessibility at regulatory DNA is highly choreographed across large sets of co-activated elements distributed throughout the genome, and that DHSs with similar cross-cell-type activation profiles probably share similar functions.

Variation in regulatory DNA linked to mutation rate

The DHS compartment as a whole is under evolutionary constraint, which varies between different classes and locations of elements44, and may be heterogeneous within individual elements34. To understand the evolutionary forces shaping regulatory DNA sequences in humans, we estimated nucleotide diversity (π) in DHSs using publicly available whole-genome sequencing data from 53 unrelated individuals (Supplementary Methods). We restricted our analysis to nucleotides outside of exons and RepeatMasked regions. To provide a comparison with putatively neutral sites, we computed π for synonymous positions (third positions) of coding exons. This analysis showed that, taken together, DHSs exhibit lower π than fourfold degenerate sites, compatible with the action of purifying selection.

Figure 7a shows π for the DHSs of all analysed cell types, with colour coding to indicate the origin of each cell type. Particularly striking is the distribution of diversity relative to proliferative potential. DHSs in cells with limited proliferative potential have uniformly lower average diversity than immortal cells, with the difference most pronounced in malignant and pluripotent lines. This ordering is identical when highly mutable CpG nucleotides are removed from the analysis.

If differences in π are due to mutation rate differences in different DHS compartments, the ratio of human polymorphism to human–chimpanzee divergence should remain constant across cell types. By contrast, differences in π due to selective constraint should result in pronounced differences. To distinguish between these alternatives, we first compared polymorphism and human–chimpanzee divergence for DHSs from normal, malignant and pluripotent cells (Fig. 7b).

Differences in polymorphism and divergence between these three groups are nearly identical, compatible with a mutational cause. Second, raw mutation rate is expected to affect rare and common genetic variation equally, whereas selection is likely to have a larger impact on common variation. We consistently observe ~62% of single nucleotide polymorphisms (SNPs) in DHSs of each group to have derived-allele frequencies below 0.05. DHSs in different cell
lines exhibit differences in SNP densities but not in allele frequency distribution (Fig. 7c). Collectively, these observations are consistent with increased relative mutation rates in the DHS compartment of immortal cells versus cell types with limited proliferative potential, exposing an unexpected link between chromatin accessibility, proliferative potential and patterns of human variation.

Discussion
Since their discovery over 30 years ago, DNase I hypersensitive sites have guided the discovery of diverse cis-regulatory elements in the human and other genomes. Here we have presented by far the most comprehensive map of human regulatory DNA, revealing novel elements that shape cell-type identity.

METHODS SUMMARY
DNase I hypersensitivity mapping was performed using protocols developed by Duke University7 or University of Washington8 on a total of 125 cell types (Supplementary Table 1). Data sets were sequenced to an average depth of 30 million tags. Sequence reads were mapped using the Bowtie aligner, allowing a maximum of two mismatches. Only reads mapping uniquely to the genome were used in our analyses. Mappings were to male or female versions of hg19/GRCCh37, depending on cell type, with random regions omitted. Data were analysed jointly using a single algorithm9 (Supplementary Methods) to localize DNase I hypersensitive sites. H3K4me3 ChIP-seq was performed using antibody 9751 (Cell Signaling) on 1% formaldehyde crosslinked samples sheared by Diagenode Bioruptor. Gene expression measurements for each cell type were performed on Affymetrix human exon microarrays. 5C experiments were performed as describedref.10. Transcription factor recognition motif occurrences within DHSs were defined with FIMO 11 at significance P < 10−5 using motif models from the TRANSFAC database.

Received 15 December 2011; accepted 15 May 2012.

1. Felsenfeld, G., Boyes, J., Chung, J., Clark, D. & Studitsky, V. Chromatin structure and gene expression. Proc. Natl Acad. Sci. USA 93, 9384–9388 (1996).
2. Gross, D. S. & Garrard, W. T. Nucleosome hypersensitive sites in chromatin. Annu. Rev. Biochem. 57, 159–197 (1988).
3. Gaszner, M. & Felsenfeld, G. Insulators: exploiting transcriptional and epigenetic mechanisms. Nature Rev. Genet. 7, 703–713 (2006).
4. Li, Q., Harji, S. & Peterson, K. R. Locus control regions: coming of age at a decade plus. Trends Genet. 15, 403–408 (1999).
5. Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genet. 39, 311–318 (2007).
6. Hesselberth, J. R. et al. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nature Methods 6, 283–289 (2009).
7. Boyle, A. P. et al. High-resolution mapping and characterization of open chromatin across the genome. Cell 132, 311–322 (2008).
8. John, S. et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nature Genet. 43, 264–268 (2011).
9. Song, L. et al. Open chromatin defined by DNase I and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res. 21, 1757–1767 (2010).
10. Harrow, J. et al. GENCODE. The reference human genome annotation for the ENCODE project. Genome Res. (in the press).
11. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. Nucleic Acids Res. 36, D154–D158 (2008).
12. Farazi, T. A., Spitzer, J. L. Morozov, P. & Tuschl, T. miRNAs in human cancer. J. Pathol. 223, 102–115 (2011).
