

## LETTERS

# Histone modifications at human enhancers reflect global cell-type-specific gene expression

Nathaniel D. Heintzman<sup>1,2\*</sup>, Gary C. Hon<sup>1,3\*</sup>, R. David Hawkins<sup>1\*</sup>, Pouya Kheradpour<sup>5</sup>, Alexander Stark<sup>5,6</sup>, Lindsey F. Harp<sup>1</sup>, Zhen Ye<sup>1</sup>, Leonard K. Lee<sup>1</sup>, Rhona K. Stuart<sup>1</sup>, Christina W. Ching<sup>1</sup>, Keith A. Ching<sup>1</sup>, Jessica E. Antosiewicz-Bourget<sup>7</sup>, Hui Liu<sup>8</sup>, Xinmin Zhang<sup>8</sup>, Roland D. Green<sup>8</sup>, Victor V. Lobanenko<sup>9</sup>, Ron Stewart<sup>7</sup>, James A. Thomson<sup>7,10</sup>, Gregory E. Crawford<sup>11</sup>, Manolis Kellis<sup>5,6</sup> & Bing Ren<sup>1,4</sup>

The human body is composed of diverse cell types with distinct functions. Although it is known that lineage specification depends on cell-specific gene expression, which in turn is driven by promoters, enhancers, insulators and other *cis*-regulatory DNA sequences for each gene<sup>1–3</sup>, the relative roles of these regulatory elements in this process are not clear. We have previously developed a chromatin-immunoprecipitation-based microarray method (ChIP-chip) to locate promoters, enhancers and insulators in the human genome<sup>4–6</sup>. Here we use the same approach to identify these elements in multiple cell types and investigate their roles in cell-type-specific gene expression. We observed that the chromatin state at promoters and CTCF-binding at insulators is largely invariant across diverse cell types. In contrast, enhancers are marked with highly cell-type-specific histone modification patterns, strongly correlate to cell-type-specific gene expression programs on a global scale, and are functionally active in a cell-type-specific manner. Our results define over 55,000 potential transcriptional enhancers in the human genome, significantly expanding the current catalogue of human enhancers and highlighting the role of these elements in cell-type-specific gene expression.

We performed ChIP-chip analysis as described previously<sup>5</sup> to determine binding of CTCF (insulator-binding protein) and the coactivator p300 (also known as EP300), and patterns of histone modifications in five human cell lines: cervical carcinoma HeLa, immortalized lymphoblast GM06690 (GM), leukaemia K562, embryonic stem cells (ES) and BMP4-induced ES cells (dES). We first investigated 1% of the human genome selected by the ENCODE consortium<sup>7</sup>, using DNA microarrays consisting of 385,000 50-base oligonucleotides that tile 30-million base pairs (bp) at 36 bp resolution. We examined mono- and tri-methylation of histone H3 lysine 4 (H3K4me1, H3K4me3) and acetylation of histone H3 lysine 27 (H3K27ac) at well-annotated promoters, reasoning that the state of these histone modifications would vary in a cell-type-specific manner. To our surprise, the chromatin signatures at promoters are remarkably similar across all cell types (Fig. 1a). Quantitative comparison of ChIP-chip enrichment (see Supplementary Information) revealed highly correlated histone modification patterns at promoters across all cell types, with an average Pearson correlation coefficient of 0.71 (Supplementary Fig. 1a). This observation also holds for the larger set of Gencode promoters (Supplementary Fig. 2).

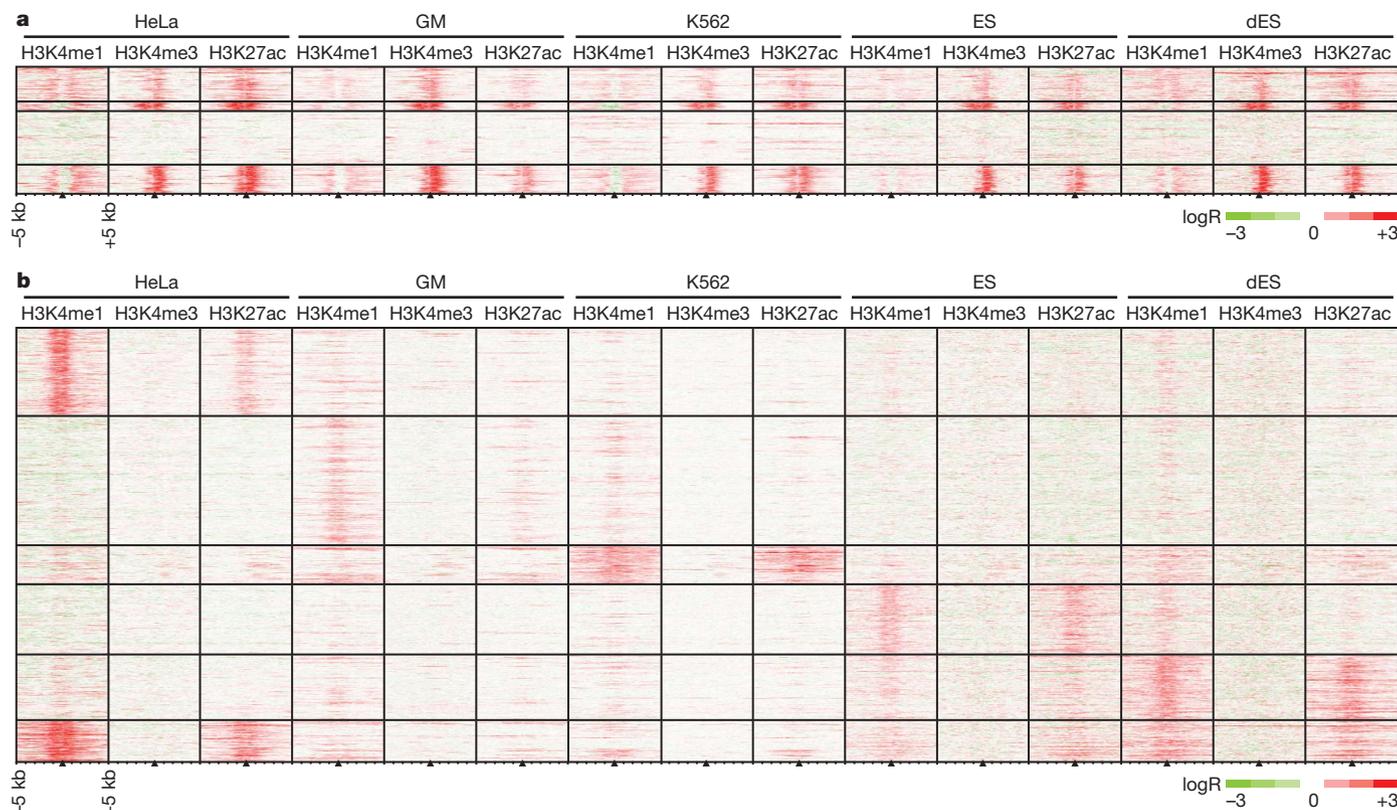
Next, we identified putative insulators in the ENCODE regions for these cell types based on CTCF binding, because mammalian insulators are generally understood to require CTCF to block promoter–enhancer interactions<sup>3</sup>. We observed nearly identical CTCF occupancy (Supplementary Table 1 and Supplementary Fig. 1e) and highly correlated CTCF enrichment patterns across all five cell types (Supplementary Fig. 1b), providing experimental support for the mostly cell-type-invariant function of CTCF as suggested by DNase hypersensitivity mapping results<sup>8</sup>.

We then investigated transcriptional enhancers in the ENCODE regions, performing ChIP-chip in HeLa, K562 and GM cells to locate binding sites for the transcriptional coactivator protein p300 (Supplementary Tables 2–4) because p300 is known to localize at enhancers<sup>9</sup>. We observed highly cell-type-specific histone modification patterns at distal p300-binding sites (Supplementary Fig. 1f), in marked contrast to the similarities in histone modifications across cell types at promoters. We then used our chromatin-signature-based prediction method<sup>5</sup> to identify additional enhancers in the ENCODE regions in these cell types (Fig. 1b and Supplementary Tables 5–9). In addition to the characteristic H3K4me1 enrichment, predicted enhancers are frequently marked by acetylation of H3K27, DNaseI hypersensitivity and/or binding of transcription factors and coactivators, and many contain evolutionarily conserved sequences (Supplementary Figs 3 and 4; see Supplementary Information). Unlike promoters and insulators, but similar to p300-binding sites, the histone modification patterns at predicted enhancers are largely cell-type-specific (Fig. 1b and Supplementary Fig. 1d), in agreement with observations that H3K4me1 is distributed in a cell-type-specific manner<sup>10</sup>.

These results indicate that enhancers are the most variable class of transcriptional regulatory element between cell types and are probably of primary importance in driving cell-type-specific patterns of gene expression. Knowledge of enhancers is therefore critical for understanding the mechanisms that control cell-type-specific gene expression, yet our incomplete knowledge of enhancers in the human genome has confined previous studies of gene regulatory networks mainly to promoters. To identify enhancers on a genome-wide scale and facilitate global analysis of gene regulatory mechanisms, we performed ChIP-chip throughout the entire human genome as described<sup>6</sup>, mapping enrichment patterns of H3K4me1 and H3K4me3 in HeLa cells. Using previously described chromatin signatures for enhancers<sup>5</sup>, we

<sup>1</sup>Ludwig Institute for Cancer Research, <sup>2</sup>Biomedical Sciences Graduate Program, <sup>3</sup>Bioinformatics Program, and <sup>4</sup>Department of Cellular and Molecular Medicine, UCSD School of Medicine, 9500 Gilman Drive, La Jolla, California 92093-0653, USA. <sup>5</sup>MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar Street, Cambridge, Massachusetts 02139, USA. <sup>6</sup>Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. <sup>7</sup>Morgridge Institute for Research, Madison, Wisconsin 53707-7365, USA. <sup>8</sup>Roche NimbleGen, Inc., 500 South Rosa Road, Madison, Wisconsin 53719, USA. <sup>9</sup>National Institutes of Allergy and Infectious Disease, 5640 Fishers Lane, Rockville, Maryland 20852, USA. <sup>10</sup>University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin 53706, USA. <sup>11</sup>Institute for Genome Sciences and Policy, and Department of Pediatrics, Duke University, 101 Science Drive, Durham, North Carolina 27708, USA.

\*These authors contributed equally to this work.



**Figure 1 | Chromatin modifications at promoters are generally cell-type-invariant whereas those at enhancers are cell-type-specific.** We used ChIP-chip to map histone modifications (H3K4me1, H3K4me3 and H3K27ac) in the ENCODE regions in five cell types (HeLa, GM, K562, ES, dES). **a**, We performed *k*-means clustering on the chromatin modifications found  $\pm 5$  kb from 414 promoters, and observe them to be generally

invariant across cell types. **b**, As in **a**, but clustering on 1,423 non-redundant enhancers predicted on the basis of chromatin signatures, revealing the cell-type-specificity of enhancers. LogR is the log ratio of enrichment of each marker as determined by ChIP-chip. Promoters and predicted enhancers are located at the centre of 10-kb windows as indicated by black triangles.

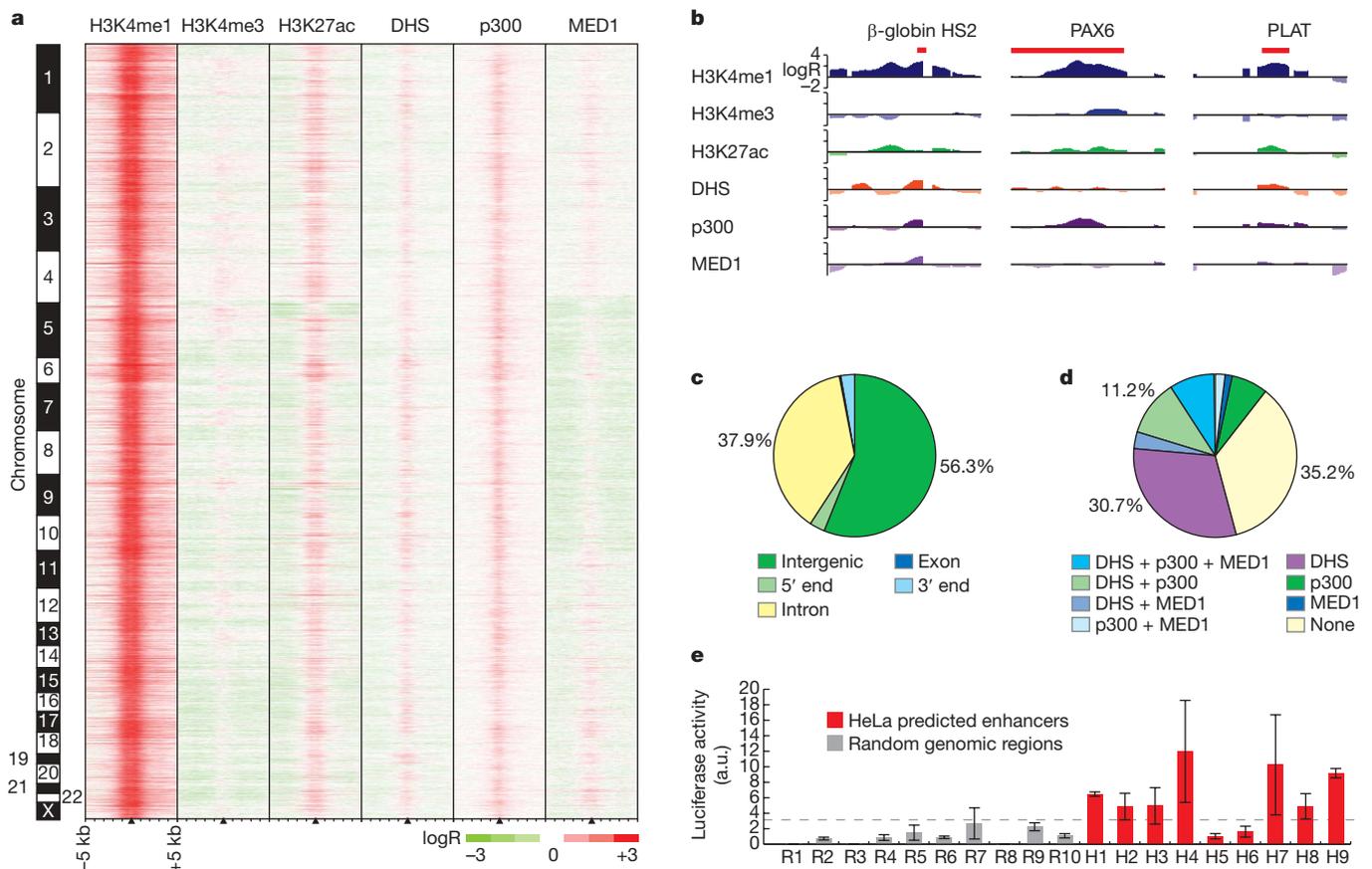
predicted 36,589 enhancers in the HeLa genome (Fig. 2a and Supplementary Table 10; see Supplementary Information). This method correctly located several previously characterized enhancers, including the  $\beta$ -globin HS2 enhancer<sup>11</sup> and distal enhancers for the PAX6 (ref. 12) and PLAT<sup>13</sup> genes (Fig. 2b). Most predicted enhancers are distal to promoters (Fig. 2c), have strong evolutionary conservation (see Supplementary Information) and are marked by histone acetylation (H3K27ac), binding of coactivator proteins (p300, MED1) or DNaseI hypersensitivity (DHS) (Fig. 2a, d; see Supplementary Information). We verified the functional potential of predicted HeLa enhancers using luciferase reporter assays as described<sup>5</sup> (see Supplementary Methods). Out of nine predicted enhancers that we evaluated, seven (78%) were active in reporter assays (Fig. 2e and Supplementary Table 11), with median activity significantly different from that in random genomic regions ( $P = 3.25 \times 10^{-4}$ ). These results support the suitability of using chromatin signatures to identify genomic regions with enhancer function.

We evaluated the predicted enhancers for conserved motif-like sequences using several-hundred shuffled TRANSFAC motifs across ten mammals in a phylogenetic framework that tolerates motif movement, partial motif loss and sequencing or alignment discrepancies (see Supplementary Methods). Predicted enhancers showed conservation for 4.3% of instances (at branch-length-score  $> 50\%$ , see Supplementary Methods), which is substantially greater than for the remaining intergenic regions (2.9%,  $P < 1 \times 10^{-100}$ ) and even promoter regions (3.9%,  $P = 1 \times 10^{-57}$ ). Additionally, testing a list of 123 unique TRANSFAC motifs as described<sup>14</sup> (see Supplementary Information), we found that 67 (54%) are over-conserved and 39 (32%) are enriched in predicted enhancers (Supplementary Table 12). We also performed *de novo* motif discovery in enhancer regions using multiple alignments of 10 mammalian genomes (see Supplementary Methods),

revealing 41 enhancer motifs, 19 of which match known transcription factor motifs whereas 22 are new (Supplementary Table 13). These motifs show conservation rates between 7% and 22% in enhancers (median 9.3%), compared to only 1.1% for control shuffled motifs of identical composition. Furthermore, over 90% of these motifs seem to be unique to enhancers, as only 4 motifs are enriched in promoter regions and 12 are in fact depleted in promoters (Supplementary Table 13), indicating that predicted enhancers contain unique regulatory sequences that may be specific to enhancer function.

To investigate the association of predicted enhancers with HeLa-specific gene expression, we used Shannon entropy<sup>15</sup> to rank genes by the specificity of their expression levels in HeLa compared to that in three other cell lines (K562, GM06990, IMR90) (Supplementary Fig. 5; see Supplementary Information), and then plotted the distribution of enhancers around genes within insulator-delineated domains (as defined by CTCF-binding sites in Supplementary Fig. 6; see Supplementary Information). Predicted enhancers are markedly enriched near HeLa-specific expressed genes (Fig. 3a), particularly within 200 kb of promoters. We observed a 1.83-fold enrichment ( $P = 4.71 \times 10^{-279}$ ) of predicted enhancers around HeLa-specific expressed genes relative to a random distribution (see Supplementary Information) and significant depletion of enhancers around non-specific expressed genes ( $P = 5.43 \times 10^{-15}$ ) and HeLa-specific repressed genes ( $P = 4.63 \times 10^{-2}$ ).

To investigate more directly the relationship between chromatin modification patterns at enhancers and cell-type-specific gene expression, we expanded our global analysis to another cell type. We performed genome-wide ChIP-chip for H3K4me1 and H3K4me3 in K562 cells and identified 24,566 putative enhancers in this cell type using our chromatin-signature-based enhancer-prediction method (Supplementary Table 14; see Supplementary Information).



**Figure 2 | Genome-wide enhancer predictions in human cells.** **a**, We predict 36,589 enhancers in HeLa cells on the basis of chromatin signatures for H3K4me1 and H3K4me3 as determined by ChIP-chip using genome-wide tiling microarrays and condensed enhancer microarrays (see Supplementary Information). Enhancer predictions are located at the centre of 10-kb windows as indicated by black triangles, and ordered by genomic position. Enrichment data are shown for histone modifications (H3K4me1, H3K4me3 and H3K27ac), DNaseI hypersensitivity (DHS), and binding of p300 and MED1. **b**, ChIP-chip enrichment profiles at several known enhancers (indicated in red) recovered by prediction:  $\beta$ -globin HS2 (chromosome 11: 5,258,371–5,258,665)<sup>11</sup>, PAX6 (chromosome 11: 31,630,500–31,635,000)<sup>12</sup>, PLAT (chromosome 8: 42,191,500–42,192,400)<sup>13</sup> (5-kb windows centred on enhancer predictions; images generated in part at the UCSC Genome Browser). **c**, Predicted enhancer distribution relative to UCSC Known

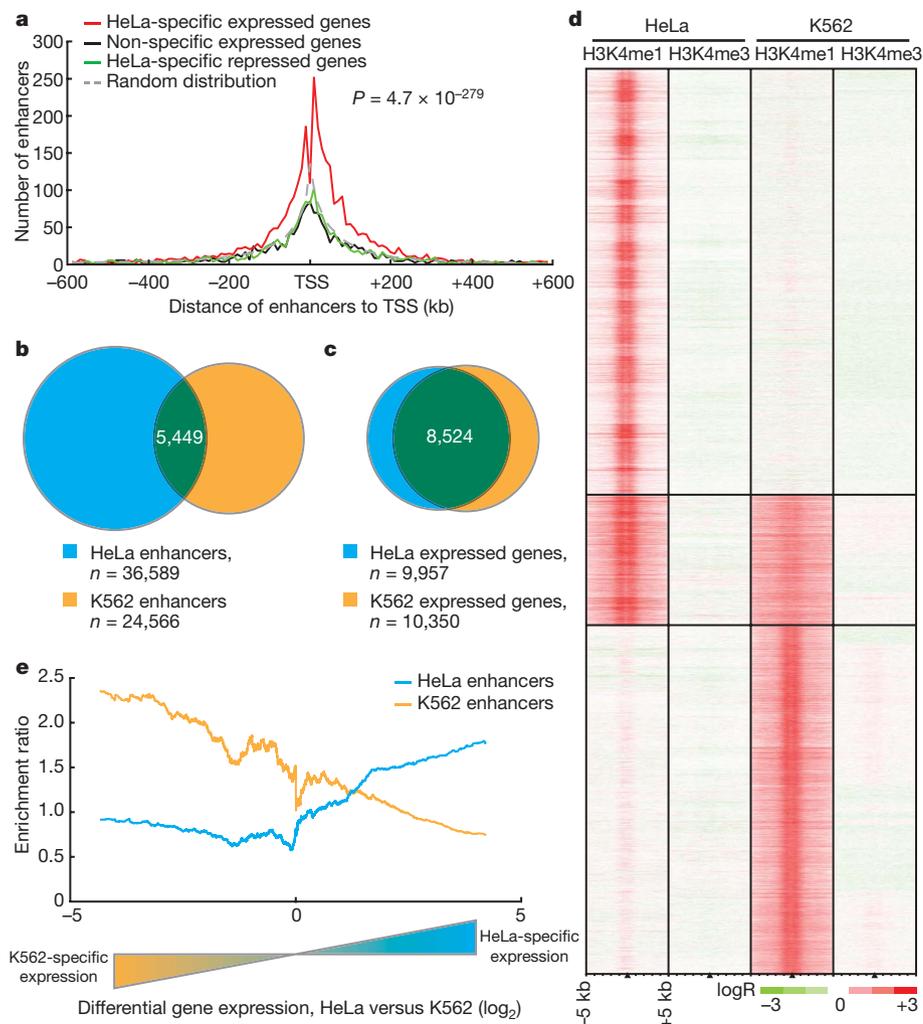
Consistent with results in the ENCODE regions, most enhancers predicted in K562 and HeLa cells are unique to either cell type (Fig. 3b), even though most expressed genes are common between the cell types (Fig. 3c). Chromatin modification profiles at predicted enhancers throughout the genome are highly cell-type-specific (Fig. 3d), with a Pearson correlation coefficient of  $-0.32$ . Furthermore, these differences seem to have regulatory implications, because domains with HeLa-specific expressed genes are enriched in HeLa enhancers but depleted in K562 enhancers, and vice versa (Fig. 3e; see Supplementary Information). These observations hold across all five cell types in the ENCODE regions (see Supplementary Information). To assess the cell-type-specificity of enhancer activity, we cloned enhancers predicted specifically in K562 cells (and not in HeLa cells) and subjected them to reporter assays in HeLa cells as described above. Out of nine K562-specific enhancers tested, only two (22%) were active in HeLa cells (Supplementary Fig. 7), and the median activity of the K-562-specific enhancers was not significantly different from random ( $P = 0.11$ ), indicating that the enhancer chromatin signature is a reliable marker of cell-type-specific enhancer function.

Although most enhancers are cell-type-specific, the presence of predicted enhancers shared by HeLa and K562 (Fig. 3b, d) indicates

Genes. Most enhancers have intergenic (56.3%) or intronic (37.9%) localization relative to UCSC Known Gene 5'-ends. **d**, Most enhancers (64.8%) are significantly marked by DNaseI hypersensitivity, binding of p300, binding of MED1, or some combination thereof. **e**, Seven out of nine enhancers predicted in HeLa cells were active in reporter assays (red bars) as compared to none of the random fragments selected as controls (grey), where activity is defined as relative luciferase value greater than 2.33 standard deviations ( $P = 0.01$ ) above the median random activity (grey dashed line). Error bars, standard deviation. Regions of  $\sim 1$ –2 kb in size were randomly selected for validation in reporter assays based on histone modification patterns as in **a**, overlap with features in **d**, and sequence features amenable to cloning by means of polymerase chain reaction (see Supplementary Information). a.u., arbitrary units.

that some enhancers may be active in multiple cell types or conditions. We compared the HeLa enhancer predictions with the results of several genome-wide studies of binding sites for sequence-specific transcription factors in different cell types, namely oestrogen receptor<sup>16</sup> (ER, also known as ESR1), p53 (TP53; ref. 17) and p63 (TP63; ref. 18) in MCF7, HCT116 and ME180 cells, respectively. Interestingly, significant percentages of binding sites for each transcription factor (from 21.4% to 32.6%) overlap with predicted enhancers in HeLa cells (Fig. 4a and Supplementary Table 15), in contrast to a significant depletion of the repressor NRSF (also known as REST)<sup>19</sup> at predicted enhancers and minimal overlap with CTCF-binding sites (see Supplementary Information).

To examine the potential role of enhancers in regulating inducible gene expression, we treated HeLa cells with the cytokine interferon- $\gamma$  (IFN- $\gamma$ ) and identified binding sites for the transcription factor STAT1 throughout the genome using ChIP-chip. STAT1 generally binds its target DNA sequences only after IFN- $\gamma$  induction<sup>20</sup>, with a small fraction of binding possible before induction<sup>21</sup>. In IFN- $\gamma$ -treated HeLa cells, we identified 1,969 STAT1-binding sites (Supplementary Table 16), with 85.8% of STAT1-binding sites occurring distal to Known Gene 5'-ends. Comparison of these distal STAT1-binding sites with recent

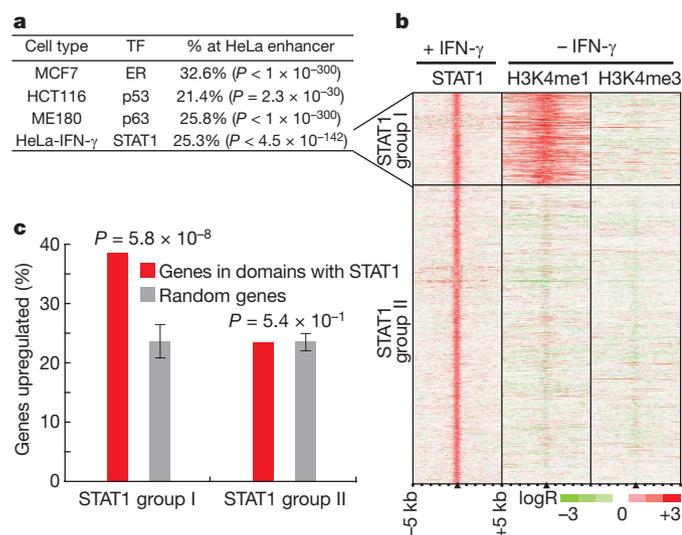


**Figure 3 | Chromatin modifications at enhancers are globally related to cell-type-specific gene expression.** **a**, Enhancer localization relative to HeLa-specific expressed genes compared to K562, GM06990 and IMR90 cells (red), non-specific expressed (green), HeLa-specific repressed (black), and a random distribution (dashed grey). Predicted enhancers are enriched around HeLa-specific expressed genes within insulator-defined domains and depleted in domains of ubiquitous or non-expressed genes (*P*-value reflects significance of enhancer enrichment in domains of HeLa-specific expressed genes, see Supplementary Information). TSS, transcription start site. **b, c**, Most enhancers predicted in HeLa and K562 cells are cell-type-specific (**b**) whereas most genes in HeLa and K562 cells are not specifically expressed (**c**); *n* = integer number of enhancers or genes in each set. **d**, Chromatin modification patterns are cell-type-specific at most of the 55,454 enhancers predicted in HeLa and K562 cells. **e**, Comparison of enhancer enrichment and differential gene expression between HeLa cells and K562 cells revealed that HeLa enhancers are enriched near HeLa-specific expressed genes (blue line) whereas K562 enhancers are enriched near K562-specific expressed genes (orange line).

analysis of STAT1 binding in uninduced HeLa cells<sup>21</sup> shows that only 6.5% of IFN- $\gamma$ -induced STAT1-binding sites are occupied by STAT1 before induction. We observed that 429 distal STAT1-binding sites overlapped enhancers predicted in HeLa cells before induction (Fig. 4a and Supplementary Table 15). The H3K4me1 enhancer chromatin signature is present before induction at these STAT1-binding sites, which we designated as STAT1 group I, whereas no evidence of this signature was visible at the remaining 1,260 distal STAT1-binding sites, designated STAT1 group II (Fig. 4b). Intriguingly, we

observed significant relative induction of expression of genes in the domains of STAT1-group-I-binding sites after just 30 min of IFN- $\gamma$  induction, whereas induction levels remained relatively unchanged for genes in the domains of other distal STAT1-group-II-binding sites during this time (Fig. 4c). These findings indicate that an enhancer chromatin signature confers increased regulatory responsiveness to a STAT1-binding site, in agreement with our previous discovery of functional enhancers in HeLa cells that were marked by the enhancer chromatin signature but were not active until they were bound by STAT1 (ref. 5).

Our findings offer, to our knowledge, the first genome-wide evaluation of the relationship between chromatin modifications at transcriptional enhancers and global programs of cell-type-specific gene expression. We determined over 55,000 potential enhancers in the human genome and show that the chromatin modifications at the



**Figure 4 | Chromatin modifications are associated with an increased regulatory response of transcription-factor-binding sites at enhancers.** **a**, Predicted enhancers in steady-state HeLa cells overlap with significant fractions of transcription-factor-binding sites (ER, p53, p63) in diverse cell types (MCF7, HCT116, ME180), as well as with STAT1-binding sites in HeLa cells treated with the cytokine interferon- $\gamma$  (HeLa-IFN- $\gamma$ ) (TF, transcription factor; TFBS, transcription factor binding sites). **b**, Hundreds of STAT1-binding sites after treatment (+IFN- $\gamma$ ) are marked by the enhancer chromatin signature in HeLa cells even before treatment (-IFN- $\gamma$ ). **c**, In HeLa cells treated with IFN- $\gamma$  (upper panel), gene expression is significantly ( $P = 5.8 \times 10^{-8}$ ) more likely to be induced by STAT1 binding at sites with the enhancer chromatin signature (red, STAT1 group I) than by STAT1 binding at other distal sites (red, STAT1 group II) relative to a random distribution (grey). Error bars, standard deviation.

enhancers correlate with cell-type-specific gene expression and functional enhancer activity. Perhaps the most intriguing observation is the large number of enhancers identified from the investigation of just two cell lines. Because enhancers are mostly cell-type-specific, our data indicate the existence of a vast number of enhancers in the human genome, on the order of  $10^5$ – $10^6$ , that are used to drive specific gene expression programs in the 200 cell types of the human body. Future experiments with diverse cell types and experimental conditions will be necessary to comprehensively identify these regulatory elements and understand their roles in the specific gene expression program of each cell type.

## METHODS SUMMARY

HeLa, K562 and IMR90 cells were obtained from ATCC. GM06990 cells were acquired from Coriell. All cells were cultured under recommended conditions. Passage 32 H1 cells were cultured as described<sup>22</sup> with/without  $200 \text{ ng ml}^{-1}$  BMP4 for 6 days (RND Systems). Chromatin preparation, ChIP, DNA purification and ligation-mediated PCR were performed as described using commercially available and custom-made antibodies, and ChIP samples were hybridized to tiling microarrays and to custom-made condensed enhancer microarrays (NimbleGen Systems, Inc.) as described<sup>5,6</sup>. DNase-chip was performed and the data analysed as described<sup>23</sup>. Cloning and reporter assays were performed as described<sup>5</sup>. Data were normalized as described<sup>5</sup>, and ChIP-chip targets for CTCF, p300, MED1 and STAT1 were selected with the Mpeak program. We used MA2C (ref. 24) to normalize and call peaks on Nimblegen HD2 arrays. Enhancers were predicted, and *k*-means clustering, intersection analysis and evolutionary conservation analysis were performed as described<sup>5</sup>. Motif analysis was performed as described<sup>25</sup>. Gene expression was analysed using HGU133 Plus 2.0 microarrays (Affymetrix) as described<sup>5</sup>. Specificity of expression was determined using a function of Shannon entropy<sup>15</sup>. We use the MAS5 algorithm from the Bioconductor R package to generate gene expression present/absent calls. Detailed methods can be found in Supplementary Information. Supplementary data for the microarray experiments have been formatted for viewing in the UCSC genome browser via <http://bioinformatics-renlab.ucsd.edu/enhancer>.

Received 17 October 2008; accepted 26 January 2009.

Published online 18 March 2009.

1. Heintzman, N. D. & Ren, B. The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome. *Cell. Mol. Life Sci.* **64**, 386–400 (2007).
2. Nightingale, K. P., O'Neill, L. P. & Turner, B. M. Histone modifications: signalling receptors and potential elements of a heritable epigenetic code. *Curr. Opin. Genet. Dev.* **16**, 125–136 (2006).
3. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.* **7**, 29–59 (2006).
4. Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876–880 (2005).
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. Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231–1245 (2007).
7. ENCODE Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
8. Xi, H. *et al.* Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet.* **3**, e136 (2007).

9. Wang, Q., Carroll, J. S. & Brown, M. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol. Cell* **19**, 631–642 (2005).
10. Koch, C. M. *et al.* The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res.* **17**, 691–707 (2007).
11. King, D. C. *et al.* Evaluation of regulatory potential and conservation scores for detecting *cis*-regulatory modules in aligned mammalian genome sequences. *Genome Res.* **15**, 1051–1060 (2005).
12. Kleinjan, D. A. *et al.* Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. *Hum. Mol. Genet.* **10**, 2049–2059 (2001).
13. Wolf, A. T., Medcalf, R. L. & Jern, C. The t-PA -7351C>T enhancer polymorphism decreases Sp1 and Sp3 protein binding affinity and transcriptional responsiveness to retinoic acid. *Blood* **105**, 1060–1067 (2005).
14. Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**, 338–345 (2005).
15. Schug, J. *et al.* Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.* **6**, R33 (2005).
16. Carroll, J. S. *et al.* Genome-wide analysis of estrogen receptor binding sites. *Nature Genet.* **38**, 1289–1297 (2006).
17. Wei, C. L. *et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207–219 (2006).
18. Yang, A. *et al.* Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. *Mol. Cell* **24**, 593–602 (2006).
19. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
20. Brivanlou, A. H. & Darnell, J. E. Jr. Signal transduction and the control of gene expression. *Science* **295**, 813–818 (2002).
21. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods* **4**, 651–657 (2007).
22. Ludwig, T. E. *et al.* Feeder-independent culture of human embryonic stem cells. *Nature Methods* **3**, 637–646 (2006).
23. Crawford, G. E. *et al.* DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nature Methods* **3**, 503–509 (2006).
24. Song, J. S. *et al.* Model-based analysis of two-color arrays (MA2C). *Genome Biol.* **8**, R178 (2007).
25. Kheradpour, P., Stark, A., Roy, S. & Kellis, M. Reliable prediction of regulator targets using 12 *Drosophila* genomes. *Genome Res.* **17**, 1919–1931 (2007).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank members of the Ren laboratory for comments. This work was supported by funding from American Cancer Society (R.D.H.), NIAID Intramural Research Program (V.V.L.), LICR (B.R.), NHGRI (B.R.), NCI (B.R.) and CIRM (B.R.).

**Author Contributions** R.D.H., N.D.H., G.C.H. and B.R. designed the experiments; R.D.H., N.D.H., L.F.H., Z.Y., L.K.L., R.K.S., C.W.C., H.L. and X.Z. conducted the ChIP-chip experiments; G.C.H. and K.A.C. analysed the ChIP-chip data; G.C.H. predicted enhancers; R.D.H. and L.K.L. conducted the reporter assays; J.E.A.-B., R.S. and J.A.T. provided human ES cells and expression data; V.V.L. provided advice and antibodies for CTCF-ChIP experiments; P.K., A.S. and M.K. analysed the transcription factor motifs; G.E.C. performed and analysed the DNaseI-chip experiments; and N.D.H., G.C.H., R.D.H. and B.R. wrote the manuscript.

**Author Information** Microarray data have been submitted to the GEO repository under accession numbers GSE14083, GSE8098, GSE7872 and GSE7118. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to B.R. ([biren@ucsd.edu](mailto:biren@ucsd.edu)).