A unique chromatin signature uncovers early developmental enhancers in humans

Alvaro Rada-Iglesias[1], **Ruchi Bajpai**[1], **Tomek Swigut**[1], **Samantha A. Brugmann**[1], **Ryan A. Flynn**[1], and **Joanna Wysocka**[1,2]

[1]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, California 94305, USA.

[2]Department of Developmental Biology, Stanford University School of Medicine, Stanford, California 94305, USA.

Abstract

Cell-fate transitions involve the integration of genomic information encoded by regulatory elements, such as enhancers, with the cellular environment[1,2]. However, identification of genomic sequences that control human embryonic development represents a formidable challenge[3]. Here we show that in human embryonic stem cells (hESCs), unique chromatin signatures identify two distinct classes of genomic elements, both of which are marked by the presence of chromatin regulators p300 and BRG1, monomethylation of histone H3 at lysine 4 (H3K4me1), and low nucleosomal density. In addition, elements of the first class are distinguished by the acetylation of histone H3 at lysine 27 (H3K27ac), overlap with previously characterized hESC enhancers, and are located proximally to genes expressed in hESCs and the epiblast. In contrast, elements of the second class, which we term 'poised enhancers', are distinguished by the absence of H3K27ac, enrichment of histone H3 lysine 27 trimethylation (H3K27me3), and are linked to genes inactive in hESCs and instead are involved in orchestrating early steps in embryogenesis, such as gastrulation, mesoderm formation and neurulation. Consistent with the poised identity, during differentiation of hESCs to neuroepithelium, a neuroectoderm-specific subset of poised enhancers acquires a chromatin signature associated with active enhancers. When assayed in zebrafish embryos, poised enhancers are able to direct cell-type and stage-specific expression characteristic of their proximal developmental gene, even in the absence of sequence conservation in the fish genome. Our data demonstrate that early developmental enhancers are epigenetically pre-marked

Readers are welcome to comment on the online version of this article at www.nature.com/nature.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

in hESCs and indicate an unappreciated role of H3K27me3 at distal regulatory elements. Moreover, the wealth of new regulatory sequences identified here provides an invaluable resource for studies and isolation of transient, rare cell populations representing early stages of human embryogenesis.

Recent reports demonstrated that active enhancers can be identified by epigenomic profiling of p300 (ref. 4), H3K4me1 and H3K27ac[5,6]. To characterize the enhancer repertoire of hESCs we performed chromatin immunoprecipitation coupled to massively parallel DNA sequencing (ChIP-seq) using antibodies recognizing chromatin regulators (that is, p300, BRG1) and histone modifications (that is, H3K4me1, H3K27ac, H3K4me3, H3K27me3) that distinguish distal elements from proximal promoters[5,6] (Supplementary Fig. 1). As expected, previously characterized hESC enhancers (for example, *NANOG* (ref. 7) and *OCT4* (also called *POU5F1*)[8]) were bound by p300 and flanked by H3K4me1 and H3K27ac marked chromatin, but were not enriched for H3K27me3 or H3K4me3 (Fig. 1a and Supplementary Fig. 2a). Genome-wide analysis defined 5,118 genomic regions (hereafter referred to as class I elements) marked by a similar chromatin signature (that is, high p300, H3K4me1 and H3K27ac, low, if any, H3K4me3, and absence of H3K27me3), representing putative active hESC enhancers (Fig. 1b and Supplementary Data 1).

Interestingly, in the vicinity of many early developmental genes we noted promoter-distal p300-bound regions that were marked by H3K4me1 but, in contrast to the active hESC enhancers, lacked H3K27ac and were instead enriched for H3K27me3, a modification associated with polycomb silencing[9] (Fig. 1a). Overall, we identified 2,287 p300-bound regions devoid of H3K27ac and marked by H3K27me3, which we will hereafter refer to as class II elements (Fig. 1b and Supplementary Data 1). In general, class II elements showed enrichment of both H3K27me3 and H3K4me1 flanking p300 peaks (Fig. 1b). In contrast, analysis of previously described adult tissue-specific enhancers[10–13] revealed no enrichment for any of the interrogated modifications (Supplementary Fig. 2b–e).

p300 enrichment levels were comparable at class I and II elements (Supplementary Fig. 3a), both classes were bound by BRG1 (Supplementary Fig. 3b), and showed similar genomic distribution relative to annotated transcription start sites (TSS), with over 95% of regions located away from promoters (Fig. 1c). Moreover, only 1.7% and 3.9% of class I and class II elements, respectively, overlapped with CpG islands, in sharp contrast to the 50% overlap observed for promoters. Another property of enhancers is their relative nucleosomal depletion compared to the flanking regions[14,15]. Using FAIRE-seq (formaldehyde-assisted isolation of regulatory elements[16] coupled to sequencing) we showed that class I and II elements were comparably nucleosome-depleted (Supplementary Fig. 3c). Furthermore, examination of a reported DNA-methylation-sensitive restriction enzyme data set from hESCs[17] revealed similar levels of DNA hypomethylation at class I and class II elements (Supplementary Fig. 3d).

ChIP-seq results were validated by ChIP-qPCR at a representative subset of class I and class II elements (labelled with the name of their closest gene) (Supplementary Figs 4a–d and 5). Further examination of the H3K27ac and H3K27me3 enrichments showed a mutually exclusive marking pattern at class I and class II elements (Supplementary Fig. 6). Sequential

ChIP-qPCR demonstrated a simultaneous presence of H3K4me1/K27ac at class I regions, and H3K4me1/K27me3 at class II regions, indicating that the concurrent enrichments of H3K4me1 and H3K27me3 were not due to cell population heterogeneity (Fig. 2a, b). Moreover, consistent with H3K27me3, we observed enrichment of the PRC2 component, SUZ12, at class II elements (Supplementary Fig. 4e). We also detected preferential association of RNA POL2 with class I elements, as compared to class II elements, including its unphosphorylated, Ser 5 phosphorylated and Ser 2 phosphorylated forms (Supplementary Fig. 7a–c).

Next we asked whether transcriptional status of nearby genes differs between the two classes. To this end, we analysed hESC transcriptome by RNA-seq and examined transcripts originating from TSS closest to the elements of each class. Class-I-associated gene expression was significantly higher than expression of all genes, or of class-II-associated genes, which were poorly expressed (Fig. 2c). In agreement, class-II-associated TSS were enriched for both H3K27me3 and H3K4me3, whereas class-I-associated TSS were marked by high H3K4me3 levels (Supplementary Fig. 8a, c). Thus, the two classes defined by unique chromatin signatures are also distinguished by the transcriptional status of associated genes.

To investigate whether the two classes are linked to genes of distinct functional annotations, we performed ontology analysis with the Genomic Regions Enrichment of Annotations Tool (GREAT)[18] (Fig. 2d, e and Supplementary Data 2 and 3). Class I elements showed association with genes expressed in the epiblast, whose mouse homologues exhibit knockout phenotypes with defects in pre- and periimplantation development (Fig. 2d). In contrast, class II elements are linked to genes expressed at, and essential for, gastrulation, germlayer formation, neurulation and early somitogenesis (including *NODAL*, *EOMES*, *LEFTY2*, *EN1*, as well as *FOX*, *SOX* and *WNT* family members) (Fig. 2e). Notably, we did not observe enrichment of adult-tissue categories among class-II-linked genes, indicating no association with late enhancers.

Taken together, our results suggest that class II elements represent poised enhancers, which reveal their cell-type-dependent activity during development. One prediction from this hypothesis is that upon differentiation to a specific fate, a subset of poised enhancers linked to genes induced in this fate should acquire an active, class I signature. To test this prediction, we differentiated hESCs into neuroectodermal spheres (hNECs)[19], generated p300, H3K4me1, H3K27ac and H3K27me3 profiles by ChIP-seq, and identified genomic elements that were marked by class II signature in hESCs, but acquired a strong enrichment of H3K27ac in hNECs (195 unique regions, Supplementary Data 1). Histone modification profiling over these regions showed concomitant decrease in H3K27me3 (Fig. 3a, b and Supplementary Fig. 9a) and we refer to them hereafter as class II→I elements. Ofnote, a large number of the remaining class II regions (that is, those that did not acquire H3K27ac) retained H3K4me1 and H3K27me3 signature in hNECs, but showed diminished p300 occupancy (Supplementary Fig. 9b–d).

The aforementioned observations were validated by ChIP-qPCR for a representative subset of enhancers (Fig. 3c–e). We further showed that class II→I elements acquired RNA POL2

enrichment in hNECs, whereas hESC-specific active enhancers showed diminished RNA POL2 binding (Supplementary Fig. 10a). In agreement with a report documenting short bidirectional transcripts originating from enhancers[20], we detected an increased level of bidirectional transcription from class II→I elements upon differentiation to hNECs, whereas transcripts originating from *NANOG* and *OCT4* enhancers were downregulated (Supplementary Fig. 10b, c).

GREAT annotation of class II→I elements showed association with genes expressed in neuroectoderm and related to abnormalities in nervous system development (Supplementary Fig. 11 and Supplementary Data 4). In agreement, hNEC RNA-seq transcriptome analysis revealed significant upregulation of class-II→I-associated genes upon differentiation, whereas expression of the remaining class-II-associated genes was persistently low (Fig. 3f). Moreover, H3K27me3 levels at class-II→I-associated TSS were diminished and H3K4me3 levels induced in hNECs as compared to hESCs, whereas modification profiles over TSS associated with the remaining class II elements were relatively unchanged (Supplementary Fig. 8b, d).

To examine if upon differentiation class II→I elements acquire the ability to drive gene expression, we infected hESCs with lentiviruses encoding a green fluorescent protein (GFP) reporter under the control of select class II→I (for example, *SOX2, HES1*), class I (for example, *CD9, JARID2*) and class II elements (for example, *EOMES, MYF5*) and monitored GFP fluorescence at day 1, 5 and 7 of differentiation to hNECs (Supplementary Table 1 and Supplementary Fig. 12). Class II→I reporters showed low, if any, fluorescence levels in hESCs, but were induced at day 5 of differentiation, whereas class I reporters displayed a reverse pattern. Our results are consistent with class II elements representing poised developmental enhancers, which upon differentiation acquire, in a cell-type-dependent manner, the properties of active enhancers.

To test whether class II elements indeed function as developmental enhancers, we examined their activity during embryogenesis. Sequence conservation analysis revealed that class II elements are evolutionarily constrained and display a higher degree of conservation than class I elements (Supplementary Fig. 13a). VISTA enhancer browser search[21] identified fourteen class II elements for which enhancer activity was previously assayed at embryonic day 11.5 of mouse development. In nine cases, highly specific expression patterns were noted (Supplementary Table 2). Interestingly, two enhancers (*WNT8B, CDH2*) belong to the class II→I and, in agreement, drive gene expression specifically in neuroectoderm-derived structures in the mouse (Supplementary Table 2).

Next we screened enhancer activity of a select set of class II elements using zebrafish embryo transgenic reporter assay[22,23]. Selected elements correspond to previously uncharacterized human genomic sequences (except for *WNT8B*) that are located in proximity to genes whose zebrafish homologues have known expression patterns, although the elements themselves are generally not well conserved in the zebrafish genome (Supplementary Figs 13 and 14). GFP reporters were injected into one-cell-stage embryos and fluorescence was monitored throughout fish embryogenesis (Supplementary Fig. 15). For eight out of nine assayed class II reporters, specific and reproducible GFP patterns were

observed at distinct developmental stages and anatomical locations (Fig. 4a–f, Supplementary Fig. 14 and Supplementary Table 3).

A first subgroup of assayed elements (for example, *NODAL, EOMES, LEFTY2*) drove gastrulation-specific expression at the shield, the fish equivalent of mouse primitive groove (Fig. 4a and Supplementary Fig. 14). Although none of the three tested sequences is well conserved in fish, proximal genes *NODAL, EOMES* and *LEFTY2* are conserved across vertebrates, with shield-specific expression pattern of zebrafish *NODAL* and *LEFTY2* homologues[24] (Supplementary Figs 14 and 16a). From mice to frogs, *EOMES* expression is initially restricted to the primitive groove and blastopore lip, respectively[25,26], but the zebrafish *EOMES* homologue is only expressed at later stages (ZFIN database, identifier ZDB-PUB-051025-1). Remarkably, the element representing a putative *EOMES* enhancer drives shield-specific expression, indicating responsiveness of this human sequence to zebrafish gastrulation circuitry.

A second subgroup of class II reporters (for example, *SOX2, NKX2-1, EN1, WNT8B, MIXL1*) drove GFP expression at later developmental stages (24–28 h post fertilization (h.p.f.)) (Fig. 4b–f); this expression was restricted to specific anatomical structures such as the midbrain–hindbrain boundary (*EN1*)[27] or the ventral diencephalon/hypothalamus (*NKX2-1*)[28]. Again, despite the low degree of sequence conservation in fish (Supplementary Fig. 13), observed GFP patterns were generally consistent with the reported expression of the putative target gene homologues[24,29] (Supplementary Fig. 16b–d).

Importantly, specificity of our results was validated with an extensive set of control regions, including: (1) five class I elements; (2) four non-conserved genomic regions flanking select analysed class II elements; (3) four human adult tissue-specific enhancers; (4) three randomly selected intergenic non-conserved regions; (5) empty vector (Supplementary Table 4). All control regions showed only weak, diffused and nonspecific GFP patterns from 6 h.p.f. to 5 d.p.f. (Supplementary Figs 17–21). It is worth mentioning that based on our limited analysis, class I elements active in hESCs do not appear to drive pre-specification expression in zebrafish. Finally, to address whether expression patterns driven by class II elements are dynamic, we monitored several reporters (*LEFTY2, SOX2, EN1, NKX2-1*) throughout embryogenesis for up to 5 d.p.f. In all cases, GFP patterns were transient in nature, with fluorescence signals barely detectable after 3 d.p.f. (Supplementary Figs 17–21), further underscoring that class II regions represent dynamically regulated developmental enhancers.

We uncovered a unique chromatin signature that bookmarks early developmental enhancers in pluripotent cells, likely to prime them for a response to signalling and developmental cues (Fig. 4g). In addition to novel insights into gene regulation, our study identified a set of over 2,000 putative regulatory sequences, thereby creating an invaluable resource for lineage tracking and isolation of transient cell populations representing early steps of human development.

METHODS

hESC culture

hESCs (H9 line, Wi-Cell) were expanded in feeder-free, serum-free medium, mTESR-1 from StemCell technologies. Cells were passaged 1:7 every 5–6 days by incubation with accutase (Invitrogen) and resultant small cell clusters (50–200 cells) were subsequently re-plated on tissue culture dishes coated overnight with growth-factor-reduced matrigel (BD Biosciences). hESC quality was regularly tested by evaluating the expression of a panel of hESC markers (for example, alkaline phosphatase, OCT4) and the capacity to differentiate into cell types derived from the three germ layers.

Neuroectoderm cell (NEC) differentiation

hESCs were differentiated into hNECs using a previously described differentiation protocol[21]. Briefly, hESCs were incubated with 2 mg ml^{-1} collagenase. Once detached, cells were plated in NEC differentiation media: 1:1 neurobasal medium/DMEM F-12 medium (Invitrogen), 0.5× B-27 supplement minus vitamin A (50× stock, Invitrogen), 0.5× N-2 supplement (100× stock, Invitrogen), 20 ng ml^{-1} bFGF (Peprotech), 20 ng ml^{-1} EGF (Sigma-Aldrich), 5 μg ml^{-1} bovine insulin (Sigma-Aldrich), 0.1 μg ml^{-1} recombinant human NOGGIN (Peprotech), 1× Glutamax-I supplement (100× stock, Invitrogen). Cells were differentiated for 7 days, changing media every other day.

Chromatin immunoprecipitation (ChIP), sequential ChIP, FAIRE and antibodies

ChIP assays were performed from approximately 10^7 hESCs or hNECs per experiment, according to previously described protocol with slight modifications[31]. Briefly, cells were crosslinked with 1% formaldehyde for 10 min at room temperature and formaldehyde was quenched by addition of glycine to a final concentration of 0.125 M. Chromatin was sonicated to an average size of 0.5–2 kb, using Bioruptor (Diagenode). A total of 3–5 μg of antibody was added to the sonicated chromatin and incubated overnight at 4 °C. 10% of chromatin used for each ChIP reaction was kept as input DNA. Subsequently, 75 μl of protein A or protein G Dynal magnetic beads (depending of antibody species and Ig isotype) were added to the ChIP reactions and incubated for four additional hours at 4 °C. Magnetic beads were washed and chromatin eluted, followed by reversal of the crosslinkings and DNA purification. Resultant ChIP DNA was dissolved in water.

Sequential ChIPs were performed as previously described with slight modifications[32]. Chromatin was prepared as described above for ChIP and after addition of the first antibody (3–5 μg) and corresponding washes, magnetic beads were resuspended in 75 μl TE/10 mM DTT. Samples were diluted 20 times with dilution buffer (1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8, 150 mM NaCl) and second antibody (3–5 μg) was added to each reaction. Beads were then washed, crosslinking reversed and DNA purified and dissolved in water.

For FAIRE, sonicated chromatin was prepared as for ChIP and DNA was extracted as previously described[16]

All antibodies used in this study have been previously reported as suitable for ChIP: p300 (sc-585, Santa Cruz Biotechnology)[5], BRG1 (clone JA1, a gift from G. Crabtree)[33], H3K4me1 (ab8895, Abcam)[5], H3K27ac (ab4729, Abcam)[5], H3K4me3 (39159, Active Motif)[34], H3K27me3 (39536, Active Motif)[35], RNA POL2 unphosphorylated (8WG16 clone, MMS-126R, Covance)[36], RNA POL2 ser5P (ab5131, Abcam)[37], RNA POL2 ser2P (ab5095, Abcam)[38], normal rabbit IgG (12-370, Millipore).

ChIP-qPCR

All primers used in qPCR analysis are shown in Supplementary Data 5. Primers are named after proximal putative target genes of the investigated enhancers. For each tested genomic element, two sets of primers were used, one set overlapping the peak of maximal p300 enrichment (central primers) and another set overlapping flanking regions with histone modification enrichments (flanking primers). This strategy was used because p300 peaks typically occurred within nucleosome-poor regions. qPCR analysis was performed in a Light Cycler 480II machine (Roche), using technical triplicates and ChIP-qPCR signals were calculated as percentage of input. Standard deviations were measured from the technical triplicate reactions and represented as error bars.

RT-qPCR of enhancer RNAs

To assess levels of enhancer-associated transcription, total RNA from hESCs and hNECs differentiated for 7 days was isolated using Trizol reagent followed by ethanol precipitation according to the manufacture's protocol (Invitrogen). To remove genomic DNA contaminants, the Turbo DNA-Free kit was used following rigorous DNase treatment (two times, 30 min incubations at 37 °C). cDNA was generated from 100 ng of DNA-free RNA using the QuantiTech Reverse Transcription Kit (Qiagen) with two modifications: (1) The gDNA elimination reaction was extended for 5 min and (2) the reverse transcription elongation time was 30 min. Quantitative PCR (qPCR) primers were designed (Supplementary Data 5) to target regions surrounding the p300 peaks that defined each tested enhancer. qPCR runs and analysis were preformed on the Light Cycler 480II machine (Roche). To calculate fold change between the hESCs and hNECs, the C_t method was used and the 18S rRNA transcripts were used as a loading control. Standard deviations were measured from technical triplicate reactions and were represented as error bars. Biological replicate experiments for hNECs were performed and very similar results were obtained (data not shown).

ChIP-seq

Libraries were prepared from: hESC and hNEC p300 ChIP, hESC BRG1 ChIP, hESC FAIRE, hESC and hNEC H3K4me3 ChIP, hESC and hNEC H3K4me1 ChIPs, hESC and hNEC H3K27me3 ChIPs, hESC and hNEC H3K27ac ChIPs, hESC and hNEC input DNAs. ChIP-seq, FAIRE-seq and input libraries were prepared according to Illumina protocol and sequenced using Illumina Genome Analyser. All sequences were mapped by ELAND software (Illumina Inc.) and analysed by QuEST 2.4 software[30,35]. ChIP-seq enrichment regions for the following profiled proteins were determined using the indicated settings, according to QuEST recommendations: hESC p300: KDE (kernel density estimation)

bandwidth = 30, ChIP seeding fold enrichment = 30, ChIP extension fold enrichment = 3, ChIP-to-background fold enrichment = 3; hESC H3K4me3: KDE bandwidth = 60, ChIP seeding fold enrichment = 30, ChIP extension fold enrichment = 3, ChIP-to-background fold enrichment = 3; hESC H3K4me1: KDE bandwidth = 100, ChIP seeding fold enrichment = 10, ChIP extension fold enrichment = 3, ChIP-to-background fold enrichment = 2.5; hESC H3K27me3: KDE bandwidth = 100, ChIP seeding fold enrichment = 10, ChIP extension fold enrichment = 8, ChIP-to-background fold enrichment = 2.5; hESC and hNEC H3K27ac: KDE bandwidth = 100, ChIP seeding fold enrichment = 10, ChIP extension fold enrichment = 3, ChIP-to-background fold enrichment = 2.5.

For all ChIP-seq data sets, WIG files were generated with QuEST, which were subsequently used for visualization purposes and for obtaining average signal profiles.

RNA-seq

RNAs from hESCs and NECs were extracted with Trizol (Invitrogen), following the manufacturer's recommendations. 10 μg of total RNA were subjected to two rounds of oligo-dT purification using Dynal oligo-dT beads (Invitrogen). 100 ng of the purified RNA were fragmented with $10\times$ fragmentation buffer (Ambion). Fragmented RNA was used for first-strand cDNA synthesis, using random hexamer primers (Invitrogen) and SuperScript II enzyme (Invitrogen). Second strand cDNA was obtained by adding RNaseH (Invitrogen) and DNA Pol I (New England Biolabs) to the first strand cDNA mix. The resulting double-stranded cDNA was used for Illumina library preparation as described for ChIP-seq experiments.

RNA-seq libraries were sequenced with Illumina Genome Analyser and both mapping and analysis of resulting reads were performed with DNAnexus software tools (https://dnanexus.com). Reads per kilobase per million mapped reads (RPKM) were calculated for all human Ensembl genes. The specificity and quality of our RNA-seq data can be visualized at several hESC- or hNEC-specific genes (Supplementary Fig. 22).

Class I and class II element selection criteria

ChIP-seq enrichment regions as determined by QuEST were used to define class I and class II elements (Supplementary Data 1). To this end, operations (intersection, subtraction, and so on) between genomic data sets were performed with GALAXY (http://main.g2.bx.psu.edu/) and the following selection criteria were used: class I elements (5,518 regions): genomic regions with hESC p300 enrichment (ChIP seeding fold enrichment >30), located within 2 kb of regions enriched in hESC H3K4me1 and H3K27ac (ChIP seeding fold enrichment >10 for both modifications), and, to distinguish these elements from proximal promoters, we demanded that these regions do not overlap with hESC H3K4me3 (ChIP seeding fold enrichment >30); class II elements (2,287 regions): genomic regions with hESC p300 enrichment (ChIP seeding fold enrichment >30), located within 2 kb of regions enriched in hESC H3K27me3 (ChIP seeding fold enrichment >8). These regions were further required not to overlap with hESC H3K4me3 (ChIP seeding fold enrichment >30) or hESC H3K27ac (ChIP seeding fold enrichment >10). Class II→I elements (195 regions): class II elements (as determined in hESCs) which in hNECs acquired enrichment in

H3K27ac (H3K27ac ChIP seeding fold enrichment >10, within 2 kb of p300 peaks defining class II elements).

In total, we identified 11,543 regions marked by p300 and H3K4me1 in hESCs, of which 1,639 did not contain H3K27ac, H3K27me3 or H3K4me3 enrichment. A total of 3,531 regions were enriched for p300, H3K4me1 and H3K4me3 (those generally corresponded to proximal promoters).

Please note that although our definition of class II elements does not use an H3K4me1 enrichment filter, about 55% of class II regions are enriched for H3K4me1 at ChIP seeding fold enrichment >10; when lower cutoff is allowed, the overlap is significantly more substantial. Thus, the vast majority, if not all, class II elements probably contains above-background levels of H3K4me1, as exemplified by the observation that class II elements with ChIP-seq H3K4me1 levels below the seeding fold enrichment >10 cutoff are still substantially enriched for H3K4me1 when assayed by ChIP-qPCR (see Supplementary Fig. 5, for example, *CHD2, EPHA4, GPR19, ADRA2A, KLF5, EML1* regions).

Other sequencing data analyses

Average ChIP-seq signal profiles around the centre of p300-enriched regions were generated with the Sitepro tool, part of the Cistrome Analysis pipeline (http://cistrome.dfci.harvard.edu/ap/), using the corresponding WIG files generated with QuEST. Similarly, ChIP-seq signal profiles were generated around gene TSS. For genes associated with the different classes of distal elements, each element was linked to its closest gene, based on the distance to TSS, and considering a maximum distance of 100 kb.

Average PhastCons scores profiles around the centre of p300-enriched regions were generated with the Conservation/Aggregate Datapoints tool, part of the Cistrome Analysis pipeline (http://cistrome.dfci.harvard.edu/ap/).

Distance between enhancers and their closest Ensembl gene TSS was calculated using PinkThing software (http://pinkthing.cmbi.ru.nl/) and Ensembl 52 assembly. With this information, it was possible to calculate the overall genomic distribution, based on distance to TSS, for the different enhancer groups and to assign enhancers to their closest genes.

Functional annotation of enhancers was obtained with GREAT (http://great.stanford.edu/public/html/input.php), using the Basal plus extension association rules and the whole human genome as background.

For RNA-seq data analysis, each enhancer was assigned to its closest gene based on distance to TSS considering a maximum distance of 100 kb, resulting in various gene groups each corresponding to an enhancer class (for example, class I, class II, class II→I). Statistical significance (*P*-values) of the difference in expression levels between different gene groups was calculated using two-sample one-sided Wilcoxon-test (R software, http://www.r-project.org). Paired or non-paired tests were performed when the same or different genes were compared, respectively. Box plots representing RPKM distribution were generated with R (http://www.r-project.org).

MRE-seq (methylation-sensitive restriction enzyme) data for hESCs was obtained from the GEO data set public repository under accession number GSM450236.

In vitro enhancer reporter assays in hESCs and hNECs

Representative class I, class II→I and class II elements (Supplementary Table 1) were cloned into a lentiviral vector (Sin-minTK-eGFP) in front of a minimal TK promoter driving GFP expression. hESC colonies were transduced with the appropriate lentiviruses and GFP fluorescence levels were subsequently monitored in undifferentiated hESCs, as well as in the course of hNEC differentiation (at day 1, 5 and 7 after induction of differentiation).

Zebrafish reporter assays

The biological relevance of the identified human enhancers was evaluated using Tol2 transposon-mediated transgenesis in zebrafish[24]. Selected human enhancers were PCR amplified and cloned in the pT2HE vector (gift from D. M. Kingsley), upstream of the *hsp70* promoter and eGFP. Tol2 transposase was *in vitro* transcribed using mMessage mMachine Sp6 kit (Ambion), according to the manufacturer's instructions. It is worth mentioning that the *hsp70* promoter independently drives robust and stable expression in the lens after 28–38 h.p.f.[39]. This lens signal is also observed when additional sequences are placed upstream of the minimal *hsp70* promoter, acting as a positive control for correct transgenesis. Vector DNA, with corresponding enhancers, and transposase RNA were mixed and injected in one-cell-stage zebrafish embryos as previously described. eGFP expression patterns were typically monitored at three different developmental times: 6–8 h.p.f., 10–14 h.p.f. and 24–28 h.p.f. According to ref. 24, using the described reporter assay method, 10–20% of the injected embryos are expected to display consistent and representative expression patterns. Because 50 embryos were typically injected, expression patterns were considered as representative for a given enhancer if displayed by at least 5–10 embryos within each batch (the remaining embryos typically showed a nonspecific or lack of fluorescence pattern). For those enhancers with identifiable and consistent expression patterns, a second set of injections (biological replicate) were performed for 50 additional embryos and in all cases similar results were obtained compared to the first injections.

Initial monitoring and embryo imaging were performed with a Leica M205 FA fluorescent stereoscope. High-resolution images presented in Fig. 4 were obtained with a Leica DM4500 B upright compound microscope.

Although live embryos were typically monitored and imaged, in order to obtain flat whole-embryo images, selected embryos were fixed and the yolk removed. Briefly, 24–28 h.p.f. embryos were dechorionated and transferred to 4% paraformaldehyde solution in PBS. After overnight rocking at 4 °C, fixed embryos were washed and stored in methanol at 20 °C until ready to use.

Specificity of our reporter assays was validated by assaying an extensive set of negative controls (Supplementary Table 4): (1) five class I elements; (2) four non-conserved genomic regions in proximity of four of the tested class II elements; (3) four human adult-tissue-

specific enhancers that should not drive expression during early developmental stages; (4) three randomly selected intergenic non-conserved regions; (5) empty vector.

In addition, four selected class II elements were followed up to 5 days post-fertilization, together with their corresponding flanking non-conserved regions and additional negative controls. GFP patterns were monitored after 6 h.p.f., 24 h.p.f., 3 d.p.f. and 5 d.p.f. In these cases and for the class II elements, embryos showing specific patterns at the corresponding stage (for example, 6 h.p.f. for LEFTY2 and 24 h.p.f. for SOX2, EN1 and NKX2-1) were selected and their GFP patterns subsequently monitored. For the negative controls, once lens signal appeared (that is, transgenic embryos), such embryos were subsequently followed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Bulger M, Groudine M. Enhancers: the abundance and function of regulatory sequences beyond promoters. Dev. Biol. 2010; 339:250–257. [PubMed: 20025863]

2. Hallikas O, et al. Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. Cell. 2006; 124:47–59. [PubMed: 16413481]

3. Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. Nature. 2009; 461:199–205. [PubMed: 19741700]

4. Visel A, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature. 2009; 457:854–858. [PubMed: 19212405]

5. Heintzman ND, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009; 459:108–112. [PubMed: 19295514]

6. Heintzman ND, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genet. 2007; 39:311–318. [PubMed: 17277777]

7. Chan KK, et al. KLF4 and PBX1 directly regulate NANOG expression in human embryonic stem cells. Stem Cells. 2009; 27:2114–2125. [PubMed: 19522013]

8. Yeom YI, et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development. 1996; 122:881–894. [PubMed: 8631266]

9. Kerppola TK. Polycomb group complexes–many combinations, many functions. Trends Cell Biol. 2009; 19:692–704. [PubMed: 19889541]

10. Cockerill PN, et al. Human granulocyte-macrophage colony-stimulating factor enhancer function is associated with cooperative interactions between AP-1 and NFATp/c. Mol. Cell. Biol. 1995; 15:2071–2079. [PubMed: 7891702]

11. Nakabayashi H, et al. Functional mapping of tissue-specific elements of the human α-fetoprotein gene enhancer. Biochem. Biophys. Res. Commun. 2004; 318:773–785. [PubMed: 15144905]

12. Itani HA, Liu X, Pratt JH, Sigmund CD. Functional characterization of polymorphisms in the kidney enhancer of the human renin gene. Endocrinology. 2007; 148:1424–1430. [PubMed: 17158202]

13. Segawa K, et al. Identification of a novel distal enhancer in human adiponectin gene. J. Endocrinol. 2009; 200:107–116. [PubMed: 18931025]

14. Mito Y, Henikoff JG, Henikoff S. Histone replacement marks the boundaries of cis-regulatory domains. Science. 2007; 315:1408–1411. [PubMed: 17347439]

15. He HH, et al. Nucleosome dynamics define transcriptional enhancers. Nature Genet. 2010; 42:343–347. [PubMed: 20208536]

16. Giresi PG, Lieb JD. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). Methods. 2009; 48:233–239. [PubMed: 19303047]

17. Harris RA, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nature Biotechnol. 2010; 28:1097–1105. [PubMed: 20852635]

18. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. Nature Biotechnol. 2010; 28:495–501. [PubMed: 20436461]

19. Bajpai R, et al. Molecular stages of rapid and uniform neuralization of human embryonic stem cells. Cell Death Differ. 2009; 16:807–825. [PubMed: 19282867]

20. Kim TK, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010; 465:182–187. [PubMed: 20393465]

21. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser–a database of tissue-specific human enhancers. Nucleic Acids Res. 2007; 35:D88–D92. [PubMed: 17130149]

22. Fisher S, et al. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. Nature Protocols. 2006; 1:1297–1305.

23. Navratilova P, et al. Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes. Dev. Biol. 2009; 327:526–540. [PubMed: 19073165]

24. Sprague J, et al. The Zebrafish Information Network: the zebrafish model organism database. Nucleic Acids Res. 2006; 34:D581–D585. [PubMed: 16381936]

25. Hancock SN, Agulnik SI, Silver LM, Papaioannou VE. Mapping and expression analysis of the mouse ortholog of *Xenopus Eomesodermin*. Mech. Dev. 1999; 81:205–208. [PubMed: 10330501]

26. Ryan K, Garrett N, Mitchell A, Gurdon JB. Eomesodermin, a key early gene in *Xenopus* mesoderm differentiation. Cell. 1996; 87:989–1000. [PubMed: 8978604]

27. Danielian PS, McMahon AP. Engrailed-1 as a target of the Wnt-1 signalling pathway in vertebrate midbrain development. Nature. 1996; 383:332–334. [PubMed: 8848044]

28. Marin O, Baker J, Puelles L, Rubenstein JL. Patterning of the basal telencephalon and hypothalamus is essential for guidance of cortical projections. Development. 2002; 129:761–773. [PubMed: 11830575]

29. Robb L, et al. Cloning, expression analysis, and chromosomal localization of murine and human homologues of a *Xenopus* mix gene. Dev. Dyn. 2000; 219:497–504. [PubMed: 11084649]

30. Valouev A, et al. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nature Methods. 2008; 5:829–834. [PubMed: 19160518]

31. Boyer LA, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005; 122:947–956. [PubMed: 16153702]

32. Furlan-Magaril M, Rincon-Arano H, Recillas-Targa F. Sequential chromatin immunoprecipitation protocol:ChIP-reChIP. Methods Mol. Biol. 2009; 543:253–266. [PubMed: 19378171]

33. Ho L, et al. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc. Natl Acad. Sci. USA. 2009; 106:5187–5191. [PubMed: 19279218]

34. Ieda M, et al. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell. 2010; 142:375–386. [PubMed: 20691899]

35. Peng JC, et al. Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell. 2009; 139:1290–1302. [PubMed: 20064375]

36. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell. 2007; 130:77–88. [PubMed: 17632057]

37. Rahl PB, et al. c-Myc regulates transcriptional pause release. Cell. 2010; 141:432–445. [PubMed: 20434984]

38. Hargreaves DC, Horng T, Medzhitov R. Control of inducible gene expression by signal-dependent transcriptional elongation. Cell. 2009; 138:129–145. [PubMed: 19596240]

39. Blechinger SR, et al. The heat-inducible zebrafish *hsp70* gene is expressed during normal lens development under non-stress conditions. Mech. Dev. 2002; 112:213–215. [PubMed: 11850198]
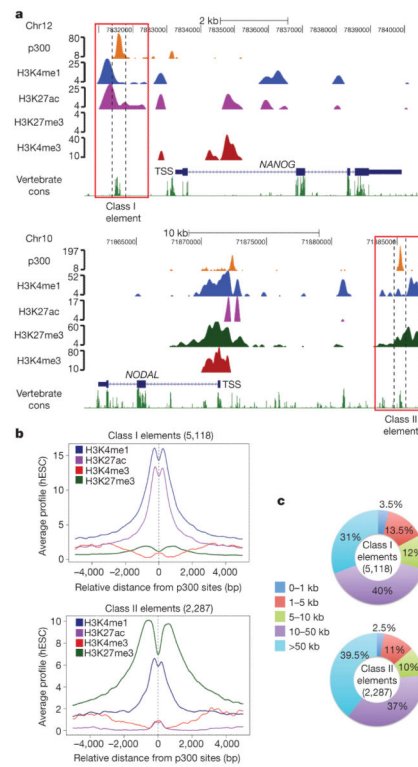
Figure 1. Unique chromatin signatures distinguish two classes of regulatory elements in hESCs
a, Genome browser representations of p300, H3K4me1, H3K27ac, H3K27me3 and H3K4me3 enrichment profiles in hESCs are shown for a representative class I (for example, *NANOG*, top) and class II (for example, *NODAL*, bottom) element and its flanking regions. The peak height corresponds to normalized fold enrichments as calculated by QuEST. **b**, Average hESC ChIP-seq signal profiles were generated for the indicated histone modifications around the central position of p300-bound regions, over class I (top) and class II (bottom) elements, respectively. **c**, Class I and II elements were mapped to their closest Ensembl gene TSS and the distribution of distances between elements and TSS is shown.
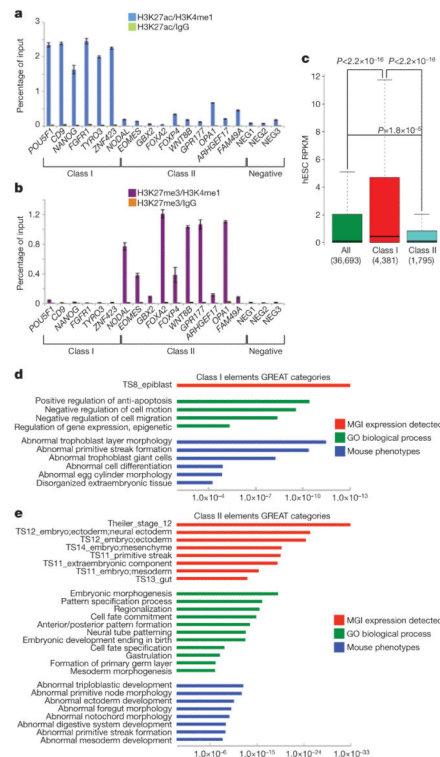
Figure 2. Functional and molecular characterization of class I and II elements

a, b, Sequential ChIP experiments were performed from hESCs with the indicated pairs of histone modification antibodies. ChIP material was analysed by qPCR for select class I and class II elements, as well as negative control regions (NEG1–3). The y axis shows per cent input recovery; error bars represent standard deviation (s.d.) from three technical replicates. **c**, RNA-seq data set was obtained from hESC poly(A)-RNA and reads per kilobase per million mapped reads (RPKM) were calculated for all human Ensembl genes. RPKMs for all annotated genes (green) or for those closest to class I (red) or class II (blue) elements are represented as box plots. P-values were calculated using non-paired Wilcoxon tests. In the box plots, bottom and top of the boxes correspond to the 25th and 75th percentiles and the internal band is the 50th percentile (median). The plot whiskers extending outside the boxes correspond to the lowest and highest datum within 1.5 interquartile range of the lower and upper quartiles, respectively. **d, e**, Functional annotation of class I (**d**) and class II (**e**) elements was performed using GREAT. The top over-represented categories belonging to three different ontologies are shown: Mouse Genome Informatics (MGI) expression detected (red) contains information on tissue- and developmental-stage-specific expression in mouse; Gene Ontology (GO) biological process (green) describes the biological processes associated with gene function; mouse phenotypes (blue) ontology contains data about mouse genotype–phenotype associations. The x axes values (in logarithmic scale) correspond to the binomial raw (uncorrected) P-values.
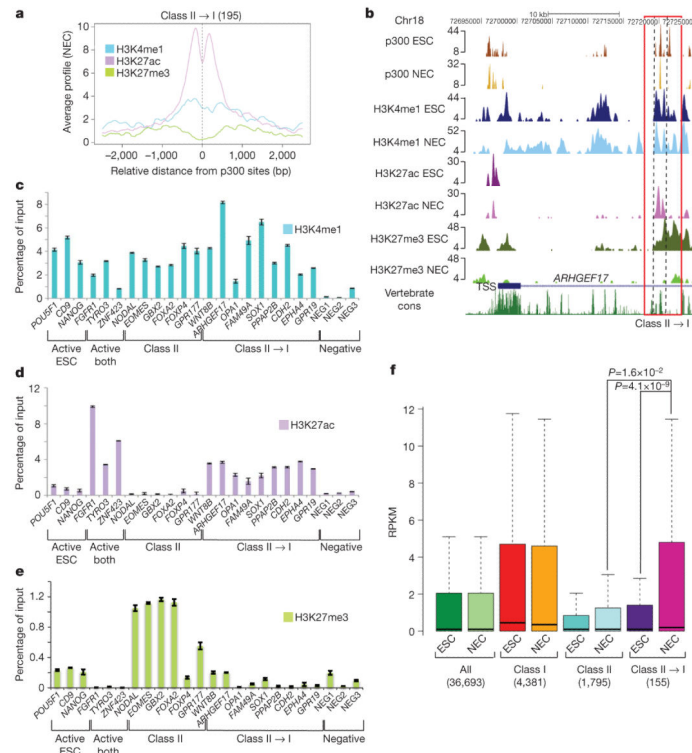
Figure 3. A subset of class II elements acquires active enhancer chromatin signature upon neuroectodermal differentiation

a, Average hNEC ChIP-seq signal profiles were generated for the indicated histone modifications around the central position of those p300-bound regions (as determined in hESC) that acquired H3K27ac enrichment in hNECs (that is, class II→I elements). **b**, Genome browser representation of p300, H3K4me1, H3K27ac and H3K27me3 (in hESCs and hNECs) binding profiles at a representative class II→I element. The peak height corresponds to normalized fold enrichments as calculated by QuEST. **c–e**, ChIP-qPCR analyses from hNECs with indicated histone modification antibodies at select elements including: class I elements that were only active in hESCs (active ESC), or in both hESCs and hNECs (active both), or class II elements that did not acquire H3K27ac in hNEC (class II), or class II→I elements. The *y* axis shows per cent input recovery; error bars represent s.d. from three technical replicates. ChIPs used in these qPCRs represent biological replicates of those samples used in ChIP-seq. **f**, RNA-seq data sets from hESC and hNEC poly(A)-RNA were used to calculate the RPKM for all human Ensembl genes. RPKMs in both cell types are represented as box plots for all genes (All), genes linked to class I elements, genes linked to class II elements, and genes linked to class II→I elements. *P*-values were calculated using paired (NEC class II→I versus ESC class II→I) or non-paired (NEC class II→I versus NEC class II) Wilcoxon tests.
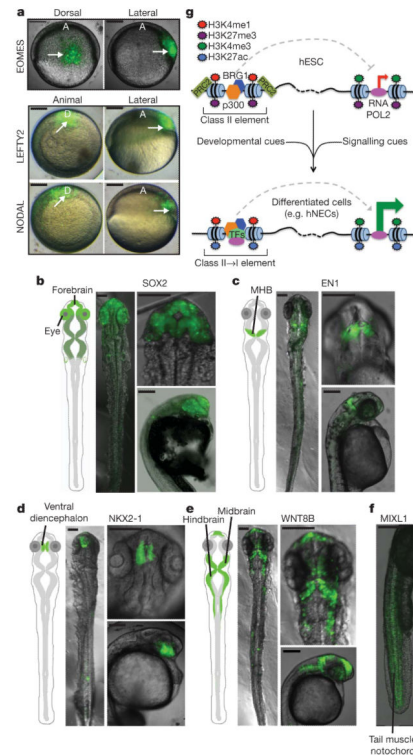
Figure 4. Class II elements have developmental enhancer activity *in vivo*

a, Merged bright-field and GFP images are shown for representative shield stage zebrafish embryos injected with class II elements proximal to human *EOMES, LEFTY2* and *NODAL*. For the EOMES enhancer, dorsal (anterior to top) and lateral (shield to right) views are presented in the left and right panels, respectively. For LEFTY2 and NODAL, animal pole (shield to top) and lateral (shield to right) views are presented in the left and right panels, respectively. White arrows indicate the location of the shield in each image. A, anterior; D, dorsal. Scale bar, 150 μm. **b–f**, Merged bright-field and GFP images are shown for representative 24–28 h.p.f. zebrafish embryos injected with class II elements proximal to *SOX2* (**b**), *EN1* (**c**), *NKX2-1* (**d**), *WNT8B* (**e**) and *MIXL1* (**f**) genes. In **b–e**, schematics highlighting the relevant anatomical structures where GFP expression was reproducibly observed are shown on the left, and three images correspond, from left to right and top to bottom, to whole-embryo flattened dorsal views, dorsal anterior views and lateral anterior views, respectively. In **f**, a lateral posterior view is shown. In **b–f**, scale bar = 150 μm. MHB, midbrain–hindbrain boundary. **g**, Proposed model for enhancer bookmarking during early embryonic development. Poised developmental enhancers (class II) are marked by a unique chromatin signature, involving occupancy of chromatin modifiers p300, BRG1 and PRC2 and nucleosomal regions marked by H3K4me1 and H3K27me3. During differentiation, appropriate developmental and signalling cues are able to rapidly transition these poised, pre-marked enhancers into an active state represented by the acquisition of H3K27ac, RNA POL2 binding, recruitment of tissue-specific transcription factors (TFs) and loss of H3K27me3, leading to the establishment of tissue-specific gene expression patterns.