Vol 464|8 April 2010|doi:10.1038/nature08866

LETTERS

Chromatin signature of embryonic pluripotency is established during genome activation

Nadine L. Vastenhouw[1]*, Yong Zhang[2,3]*, Ian G. Woods[1], Farhad Imam[1], Aviv Regev[4], X. Shirley Liu[2], John Rinn[4,5] & Alexander F. Schier[1,4,6,7]

After fertilization the embryonic genome is inactive until transcription is initiated during the maternal–zygotic transition[1–3]. This transition coincides with the formation of pluripotent cells, which in mammals can be used to generate embryonic stem cells. To study the changes in chromatin structure that accompany pluripotency and genome activation, we mapped the genomic locations of histone H3 molecules bearing lysine trimethylation modifications before and after the maternal–zygotic transition in zebrafish. Histone H3 lysine 27 trimethylation (H3K27me3), which is repressive, and H3K4me3, which is activating, were not detected before the transition. After genome activation, more than 80% of genes were marked by H3K4me3, including many inactive developmental regulatory genes that were also marked by H3K27me3. Sequential chromatin immunoprecipitation demonstrated that the same promoter regions had both trimethylation marks. Such bivalent chromatin domains also exist in embryonic stem cells and are thought to poise genes for activation while keeping them repressed[4–8]. Furthermore, we found many inactive genes that were uniquely marked by H3K4me3. Despite this activating modification, these monovalent genes were neither expressed nor stably bound by RNA polymerase II. Inspection of published data sets revealed similar monovalent domains in embryonic stem cells. Moreover, H3K4me3 marks could form in the absence of both sequence-specific transcriptional activators and stable association of RNA polymerase II, as indicated by the analysis of an inducible transgene. These results indicate that bivalent and monovalent domains might poise embryonic genes for activation and that the chromatin profile associated with pluripotency is established during the maternal–zygotic transition.

After fertilization, animals go through cleavage divisions that transform the one-cell egg into a multicellular embryo. During this phase, the genome is often inactive, and zygotic transcription is only initiated during the maternal–zygotic transition. In zebrafish (*Danio rerio*), the process of zygotic transcription begins 3 h post fertilization, when a period of rapid synchronous cleavages ends and cell cycles lengthen[9]. At this time, embryonic cells (blastomeres) are pluripotent and largely identical[9]. To study the changes in chromatin structure that accompany pluripotency and zygotic genome activation, we mapped the genomic locations of histone H3 modifications and RNA polymerase II (RNA pol II) during this transition. Previous studies have established that nucleosomes with H3K4me3 (catalysed by trithorax group proteins) and H3K27me3 (catalysed by polycomb group proteins) are associated with activation and repression of transcription, respectively[10]. Nucleosomes with H3K36me3 mark regions of transcriptional elongation[11]. By performing western blot analysis on embryos from early blastula stages to gastrulation, we found that RNA pol II phosphorylation and H3K4me3 appeared during the onset of zygotic transcription and were followed by the emergence of H3K36me3 (Fig. 1a). To determine the genomic locations of RNA pol II, H3K36me3, H3K4me3 and H3K27me3 before and after the maternal–zygotic transition, we performed chromatin immunoprecipitation (ChIP) and identified co-precipitated DNA fragments by hybridization to custom-designed NimbleGen tiling microarrays[12] covering ~31 megabases of the zebrafish genome (ChIP–chip). Peaks were detected using MA2C[13], a method that normalizes $\log_2$ fold enrichment by probe sequence and signal distribution between ChIP and control, and assigns each probe a normalized, window-smoothed MA2C score. Enriched domains were then defined and associated with RefSeq genes as described in the Methods and Supplementary Fig. 1. Consistent with the distribution profiles found in cultured cells[14], RNA pol II was enriched near transcription start sites (Fig. 1b), and H3K36me3 was detected in gene bodies, peaking towards the 3′ end (Fig. 1c). H3K4me3 was localized to promoter regions (Fig. 1d), whereas H3K27me3 covered promoters or entire genes (Fig. 1e and Supplementary Fig. 2). Notably, the comparison of ChIP–chip profiles before and after the maternal–zygotic transition (Fig. 1b–e and Supplementary Fig. 3) revealed that RNA pol II and the analysed histone modifications were not associated with the genome before zygotic genome activation. Inspection of specific genes and genomic regions confirmed the lack of H3 methylation marks before the maternal–zygotic transition as well as their presence thereafter (Fig. 1f and Supplementary Figs 4 and 5). These results indicate that RNA pol II and the histone modifications H3K36me3, H3K4me3 and H3K27me3 become associated with genomic loci only during the onset of zygotic genome activation.

We next analysed the histone modification patterns that emerged after the maternal–zygotic transition (Supplementary Fig. 6 and Supplementary Table 1; replicate analysis in Supplementary Fig. 7). Of 683 RefSeq genes, 82% were occupied by H3K4me3 at their promoter. These included 97% of the 367 genes that were zygotically expressed and 64% of the 316 genes that were inactive (that is, genes not occupied by H3K36me3; see Supplementary Discussion 1 and Supplementary Figs 1 and 6). The inactive genes were of particular interest, because studies in embryonic stem (ES) cells found that developmental regulatory genes are often inactive and associated with H3K4me3/H3K27me3 domains[4–8]. We found that in zebrafish blastomeres, 36% of inactive genes were associated with H3K4me3 and H3K27me3 (Fig. 2a, b and Supplementary Fig. 6). Gene Ontology (GO) analysis showed that these inactive genes were enriched for those annotated as 'DNA binding', 'transcription factor

[1]Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA. [2]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [3]School of Life Science and Technology, Tongji University, 1239 Siping Road, Shanghai 200092, China. [4]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [5]Department of Pathology, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. [6]Harvard Stem Cell Institute, [7]Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA.
*These authors contributed equally to this work.

©2010 Macmillan Publishers Limited. All rights reserved

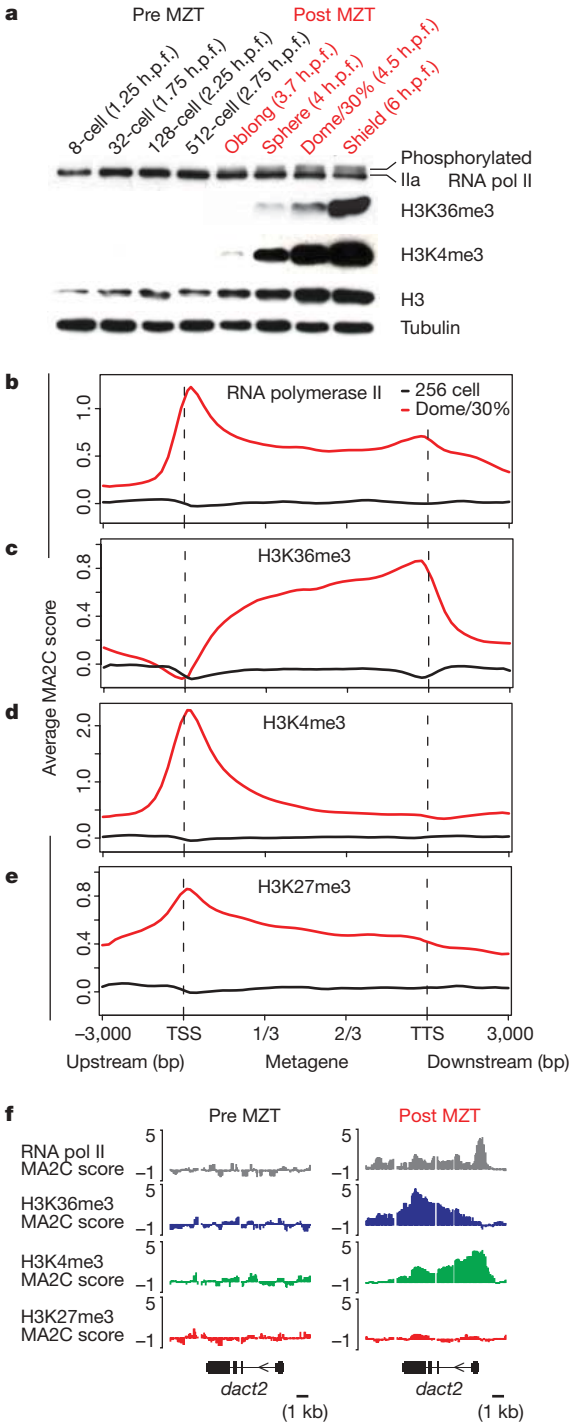

Figure 1 | Large-scale changes in chromatin modifications during maternal–zygotic transition. **a**, Western blot analysis of zebrafish embryos, from early blastula stages (8-cell; 1.25 h post fertilization (h.p.f.)) to the onset of gastrulation (shield; 6 h.p.f.). Note the shift of RNA pol II at oblong stage due to phosphorylation of the C-terminal domain, indicating engagement in transcription[30]. **b**–**e**, Average density profiles for RNA polymerase II (**b**), H3K36me3 (**c**), H3K4me3 (**d**) and H3K27me3 (**e**) before (256-cell: black line) and shortly after the maternal–zygotic transition (dome stage/30% epiboly (dome/30%): red line). These profiles show the average normalized and smoothed $\log_2$ ChIP enrichment (MA2C score) for all 822 analysed RefSeq genes on the array. Transcription units are shown as metagenes (that is, relative distance from transcription start site (TSS) to transcription termination site (TTS)), whereas upstream and downstream sequences are shown in absolute distance (bp). **f**, Example profiles for *dact2* before and after the maternal–zygotic transition (MZT). Absence of signal in all four traces indicates absence of probes due to repetitive sequences.

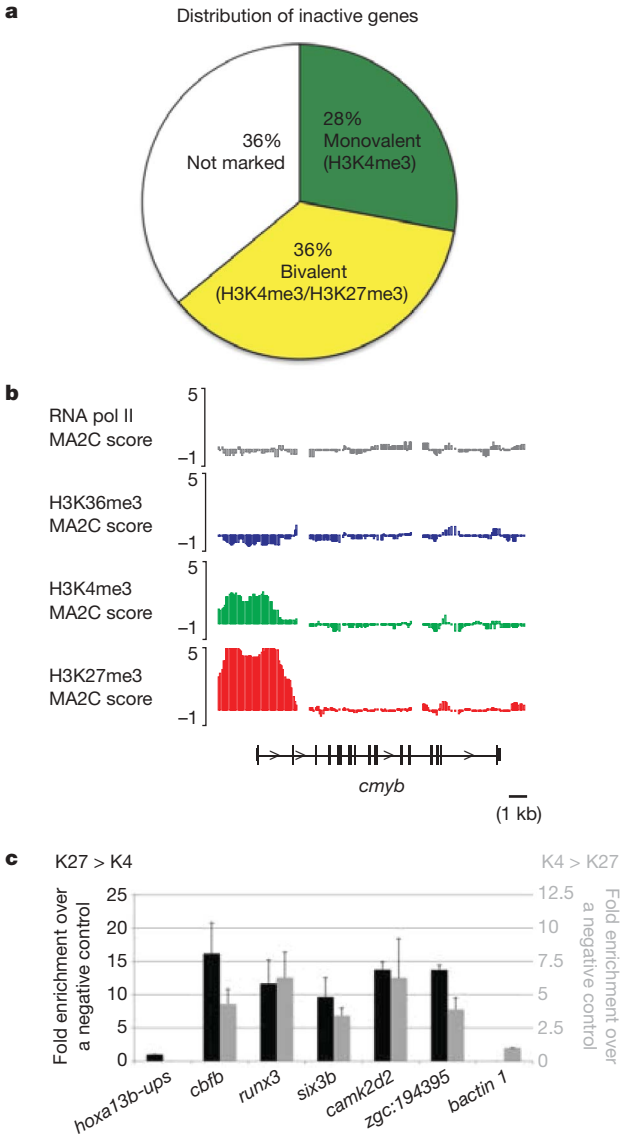

Figure 2 | Bivalent chromatin domains in zebrafish embryos. **a**, Pie chart showing all inactive genes (see Supplementary Discussion 1 and Supplementary Fig. 6 for criteria) and the relative contributions of bivalent, monovalent and not marked genes. **b**, Example profiles for *cmyb*, an inactive gene with a bivalent promoter after the maternal–zygotic transition (dome stage/30% epiboly). Absence of signal in all four traces indicates absence of probes due to repetitive sequences. **c**, Sequential ChIP analysis. Shown is fold enrichment over a negative control. Black: H3K27me3 (first ChIP) > H3K4me3 (second ChIP); *hoxa13b-ups* served as a negative control because this region is occupied by H3K27me3 but not H3K4me3. Grey: H3K4me3 > H3K27me3; *bactin1* served as a negative control because this region is occupied by H3K4me3 but not H3K27me3. Enrichment values are shown as the mean of two or three independent experiments (+s.e.m.). See Methods for details.

activity' and 'multicellular organismal development', with *P*-values of 3.3×10^{-13}, 8.9×10^{-12} and 4.1×10^{-6}, respectively, similar to that seen for ES cells[4,6–8]. Among the H3K4me3/H3K27me3 genes were the *Hox* clusters (Supplementary Fig. 4), *cmyb* (Fig. 2b), *evx1* and *irx3*. These results indicate that H3K4me3 and H3K27me3 mark regulatory genes in both mammalian ES cells and zebrafish blastomeres.

The presence of H3K4me3 and H3K27me3 on a subset of genes might indicate the simultaneous association of both chromatin marks with a given promoter in the same cells. Such 'bivalent' domains were found in ES cells and have been postulated to poise genes for activation (H3K4me3) while keeping them repressed

©2010 Macmillan Publishers Limited. All rights reserved

(H3K27me3)[4]. Alternatively, H3K4me3 marks might occupy a given promoter in only a subset of cells, whereas H3K27me3 marks might be present in a different subpopulation. Indeed, recent experiments in *Xenopus* have suggested that H3K4me3 and H3K27me3 do not co-occupy the same promoters[15]. In that study, H3K4me3 marks were found at blastula stages but, in contrast to our results (Fig. 2a, b and Supplementary Figs 4–6), H3K27me3 marks were only detected after the onset of gastrulation. This temporal separation of histone modifications would preclude bivalency in *Xenopus* blastomeres. That study also identified few, if any, bivalently marked promoters during gastrulation[15]. To determine whether bivalent domains exist in zebrafish blastomeres, we performed sequential ChIP analysis (Supplementary Fig. 8 and Supplementary Discussion 2). Using this assay, we found that H3K4me3/H3K27me3 genes such as *cbfb*, *runx3* and *six3b* are truly bivalent (Fig. 2c). These results establish that bivalent chromatin domains are present in pluripotent embryonic cells *in vivo*.

In addition to the 36% bivalent genes, 28% of inactive genes were associated with H3K4me3 but not H3K27me3 marks (Figs 2a and 3a, b). Examples of such monovalent genes (H3K4me3 only) included *psat1* (phosphoserine aminotransferase 1), which is ultimately expressed in the central nervous system and the eye, and *zgc:110784*, which is expressed later in the neural crest (Fig. 3a). GO term analysis did not reveal any enrichment within this class. The existence of such monovalent inactive genes was surprising because H3K27me3 had been postulated to prevent transcription from H3K4me3-bound promoters[4,5,16]. However, closer analysis of available human and mouse ES cell data showed that many non-expressed genes also carried monovalent H3K4me3 marks (Supplementary Discussion 3 and Supplementary Fig. 9). In contrast to the large fraction of monovalent H3K4me3 genes, very few genes carried monovalent H3K27me3 marks in zebrafish blastomeres (Supplementary Fig. 6) and ES cells[6]. These results reveal remarkably similar chromatin profiles between embryonic blastomeres and ES cells.

Recent studies have indicated that genes can be poised for expression by association of RNA pol II with transcription start sites[17–21]. By examining RNA pol II occupancy, we found that only 17% of monovalent transcription units had detectable RNA pol II associated with their start sites (Fig. 3b and Supplementary Fig. 6). Comparing the ChIP signals for H3K4me3 and RNA pol II (Fig. 3c) showed that even monovalent genes with high levels of H3K4me3 had undetectable levels of RNA pol II. Similarly, many genes in ES cells are associated with H3K4me3 but not RNA pol II (Supplementary Discussion 4). Taken together, these results reveal that many genes are initially inactive after genome activation yet marked with H3K4me3.

The presence of monovalent genes (H3K4me3 only) without detectable RNA pol II raised the possibility that sequence-specific transcription factors might direct the methylation of H3K4 before recruitment of RNA pol II[22,23]. Alternatively, H3K4me3 might mark monovalent genes even before the binding of sequence-specific transcription factors and recruitment of RNA pol II. Because only a subset of transcription factors is present at the maternal–zygotic transition, the latter possibility is more likely. To distinguish between these models, we analysed an inducible transgene that encodes EGFP under the control of the heterologous transcription factor Gal4–VP16. The transgene contained 14 upstream binding sites for Gal4–VP16 joined to a minimal viral promoter (Fig. 4a). Expression of the transgene was dependent on Gal4–VP16 and was not detected before zygotic genome activation (data not shown). This experimental system allowed us to compare occupancy by RNA pol II and H3K4me3 in the presence or absence of Gal4–VP16. As expected, injection of *GAL4–VP16* resulted in the occupancy of the reporter gene by both H3K4me3 and RNA pol II after the maternal–zygotic transition (Figs 4b, c). In the absence of Gal4–VP16, H3K4me3 still marked the promoter, but neither RNA pol II nor H3K27me3 were detected (Fig. 4b, c and data not shown). These results suggest that monovalent H3K4me3 marks can be established in the absence of sequence-specific activators and without the stable association of RNA pol II.

The chromatin profiles before and after the maternal–zygotic transition lead to three major conclusions about genome activation and pluripotency. First, genome activation is accompanied by the appearance of specific H3 trimethylation marks, indicating that these histone modifications begin to provide regulatory information only during the maternal–zygotic transition. This finding is pertinent to the suggestion that histone marks associated with developmental regulators in human sperm prepare genes for embryonic activation or repression[24]. Our results suggest that if such modifications exist in zebrafish sperm, they are replaced by other marks, or they are erased or diluted during the rapid cleavage divisions before the maternal–zygotic transition. Our data are consistent with a report that identified H3K4me3 marks during the maternal–zygotic transition in *Xenopus* embryos[15], but in contrast to that study, we detect H3K27me3 modifications already in blastomeres.

Second, our study provides evidence in embryos for the bivalent domains found in ES cells[4–8]. This finding reveals a shared chromatin landscape between transiently pluripotent cells in the embryo and permanently pluripotent cells in culture[4,5]. In contrast to the recent proposal that bivalent domains might not exist in embryos[15], we observe clear evidence for bivalency. This discrepancy could be due to species-specific modes of gene regulation but might also be caused by experimental differences (Supplementary Discussions 2 and 5). The function of bivalent domains has not been fully established. In the context of ES cells, such marks might poise genes for activation, maintain gene repression and pluripotency and control lineage commitment[25]. In the developing embryo, where pluripotency is short lived, such marks might also ensure the proper timing and quantity of developmental regulators[26].

Third, we find that many genes are monovalently marked by H3K4me3 but inactive. This subset of genes has received little attention

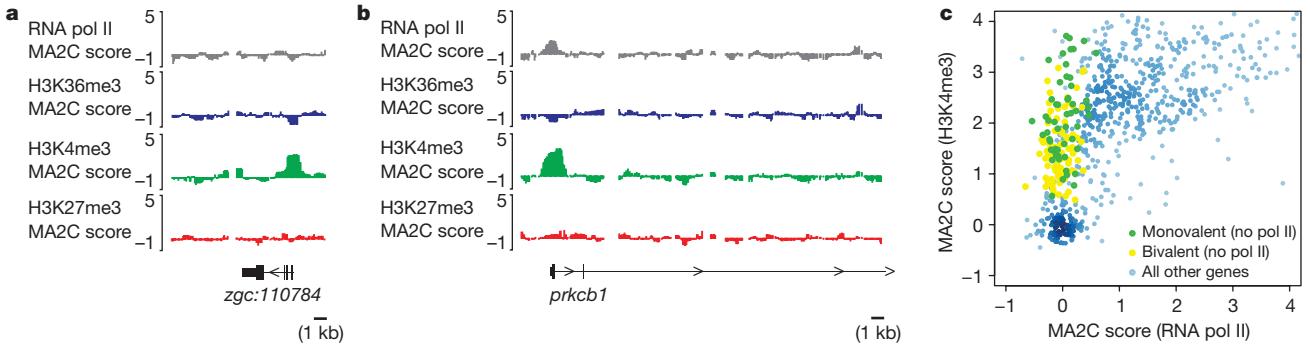

Figure 3 | Many inactive genes are monovalently marked with H3K4me3. **a**, ChIP–chip profiles for *zgc:110784*, an inactive gene with a monovalent promoter. **b**, ChIP–chip profiles for *prkcb1*, an inactive gene with a monovalent promoter that is associated with RNA pol II around the TSS. **c**, Plot of normalized ChIP signals for H3K4me3 and RNA pol II reveals genes with undetectable levels of RNA pol II but high levels of H3K4me3.

©2010 Macmillan Publishers Limited. All rights reserved

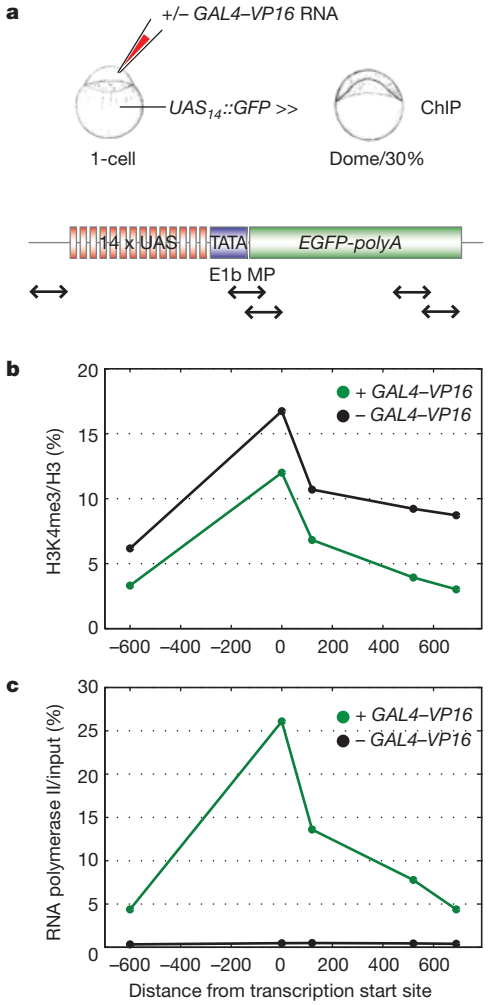

Figure 4 | H3K4me3 occupancy in the absence of a sequence-specific transcriptional activator. a, Embryos carrying an integrated transgene (*UAS_{14}::EGFP*) were either injected with *GAL4–VP16* mRNA or left uninjected. Quantitative RT–PCR analysis detected *GFP* RNA only after the maternal–zygotic transition, and only in the presence of *GAL4–VP16* (data not shown). At dome stage/30% epiboly, embryos were subjected to ChIP–quantitative PCR (ChIP–qPCR). E1b MP is a minimal viral promoter. Arrows indicate qPCR amplicons. **b**, **c**, Representative ChIP–qPCR analysis of H3K4me3 (**b**) and RNA polymerase II (**c**) in the presence (green line) and absence (black line) of *GAL4–VP16*. H3K27me3 was not detected at the transcription start site (data not shown).

in ES cell studies, but our analysis of published data sets suggests that many non-expressed genes in ES cells also carry monovalent H3K4me3 marks. Most of these H3K4me3 domains are not associated with detectable levels of RNA pol II, and our experiments with an inducible transgene suggest that H3K4me3 occupancy can be independent of sequence-specific transcriptional activators or stable association with RNA pol II. These H3K4me3 domains might poise genes for activation by creating a platform for the transcriptional machinery[27]. Consistent with this idea, basal transcription factor TFIID has a higher affinity for H3K4me3 than for unmethylated H3K4[23], indicating that H3K4me3-poised genes might be activated more efficiently or synchronously[21]. Detailed functional studies are needed to test these models and determine how the chromatin signature of embryonic pluripotency is established and regulates development.

METHODS SUMMARY

ChIP. ChIP was performed as described[28] with minor modifications. For sequential ChIP, two changes were made to the protocol: (1) antibody was crosslinked to tosylactivated magnetic Dynabeads (Invitrogen); and (2) chromatin was eluted from the beads in a milder elution buffer. After removal of the beads, 10 μl of each sample was used for analysis of the first ChIP. The remaining 100 μl was used for sequential ChIP. The eluate was diluted tenfold in immunoprecipitation reconstitution buffer to reconstitute the original immunoprecipitation buffer composition. Samples were then pre-cleared by incubation with empty beads. After removal of the beads, half of each sample was subjected to a second round of immunoprecipitation with the tosylactivated beads crosslinked to the antibody against the 'other' modification (that is, the one that was not used in the first ChIP) and the other half was incubated with empty beads as a negative control. Washing, elution and reversal of the crosslinks were performed as in single ChIP. (See also Supplementary Fig. 8.) All sequential ChIPs were normalized to their input (that is, the signal of the modification analysed in the first round of ChIP), and fold enrichment was calculated over a negative control.

ChIP–chip analysis. For analysis on tiling arrays, ChIP material was amplified as previously described[29] using the Sigma GenomePlex Kit to perform one round of amplification. Sample labelling, hybridization, washing and scanning were performed as described in the NimbleChip Arrays user's guide. Each array was normalized by MA2C[13]. Each probe was assigned a MA2C score to reflect the normalized and window-averaged $\log_2$ ratio of ChIP enrichment over control. See Supplementary Fig. 1 for decision trees used to determine the chromatin and RNA pol II status of genes.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 August 2009; accepted 27 January 2010.
Published online 24 March 2010.

1. Newport, J. & Kirschner, M. A major developmental transition in early *Xenopus* embryos: II. Control of the onset of transcription. *Cell* **30,** 687–696 (1982).
2. Schier, A. F. The maternal-zygotic transition: death and birth of RNAs. *Science* **316,** 406–407 (2007).
3. Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* **136,** 3033–3042 (2009).
4. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
5. Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nature Cell Biol.* **8,** 532–538 (2006).
6. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).
7. Zhao, X. D. *et al.* Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1,** 286–298 (2007).
8. Pan, G. *et al.* Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* **1,** 299–312 (2007).
9. Schier, A. F. & Talbot, W. S. Molecular genetics of axis formation in zebrafish. *Annu. Rev. Genet.* **39,** 561–613 (2005).
10. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128,** 735–745 (2007).
11. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128,** 707–719 (2007).
12. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129,** 1311–1323 (2007).
13. Song, J. S. *et al.* Model-based analysis of two-color arrays (MA2C). *Genome Biol.* **8,** R178 (2007).
14. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).
15. Akkers, R. C. *et al.* A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in *Xenopus* embryos. *Dev. Cell* **17,** 425–434 (2009).
16. Stock, J. K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nature Cell Biol.* **9,** 1428–1435 (2007).
17. Core, L. J. & Lis, J. T. Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science* **319,** 1791–1792 (2008).
18. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130,** 77–88 (2007).
19. Zeitlinger, J. *et al.* RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nature Genet.* **39,** 1512–1516 (2007).
20. Muse, G. W. *et al.* RNA polymerase is poised for activation across the genome. *Nature Genet.* **39,** 1507–1511 (2007).
21. Boettiger, A. N. & Levine, M. Synchronous and stochastic patterns of gene activation in the *Drosophila* embryo. *Science* **325,** 471–473 (2009).
22. Dreijerink, K. M. *et al.* Menin links estrogen receptor activation to histone H3K4 trimethylation. *Cancer Res.* **66,** 4929–4935 (2006).
23. Vermeulen, M. *et al.* Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131,** 58–69 (2007).

©2010 Macmillan Publishers Limited. All rights reserved

24. Hammoud, S. S. *et al.* Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460,** 473–478 (2009).
25. Sha, K. & Boyer, L. A. The chromatin signature of pluripotent cells. The Stem Cell Research Community, StemBook doi:10.3824/stembook.1.45.1 ⟨http://www.stembook.org⟩ (2009).
26. Yuzyuk, T., Fakhouri, T. H., Kiefer, J. & Mango, S. E. The polycomb complex protein mes-2/E(z) promotes the transition from developmental plasticity to differentiation in *C. elegans* embryos. *Dev. Cell* **16,** 699–710 (2009).
27. Barski, A. *et al.* Chromatin poises miRNA- and protein-coding genes for expression. *Genome Res.* **19,** 1742–1751 (2009).
28. Wardle, F. C. *et al.* Zebrafish promoter microarrays identify actively transcribed embryonic genes. *Genome Biol.* **7,** R71 (2006).
29. O'Geen, H., Nicolet, C. M., Blahnik, K., Green, R. & Farnham, P. J. Comparison of sample preparation methods for ChIP-chip assays. *Biotechniques* **41,** 577–580 (2006).
30. Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20,** 2922–2936 (2006).

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

Acknowledgements We thank members of the Schier laboratory for help and advice; H. G. Shin, L. Taing and Z. J. Wu for computational analysis and discussions; N. Follmer and B. Lilley for technical advice; and J. Dubrulle, N. Francis, R. Losick, S. Mango, T. van Opijnen and W. Talbot for discussions and critical reading of the manuscript. This work was supported by NIH grants to X.S.L. (1R01 HG004069) and A.F.S. (5R01 GM56211), and by EMBO and HFSP (LT-00090/2007) fellowships to N.L.V.

Author Contributions N.L.V. and A.F.S. designed the study. N.L.V. performed the experiments. Y.Z. performed computational analysis. N.L.V., Y.Z., J.R., X.S.L. and A.F.S. designed and performed data analysis. I.G.W. provided technical support. F.I. provided RNA profiling data. A.R. provided analytical advice. N.L.V. and A.F.S. interpreted the data and wrote the paper with support from co-authors.

Author Information ChIP–chip data is available under GEO accession number GSE20023; custom designed array platform is under accession number GPL9970. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to X.S.L. (xsliu@jimmy.harvard.edu) or A.F.S (schier@fas.harvard.edu).

©2010 Macmillan Publishers Limited. All rights reserved

METHODS

Zebrafish. Zebrafish were maintained and raised under standard conditions. Wild-type and UAS_{14}::GFP[31] embryos were collected at the 1-cell stage, synchronized and allowed to develop to the desired stage at 28 °C. *GAL4–VP16* RNA was injected at the 1-cell stage at 48 pg per embryo.

Western blotting. For western blotting, embryos were dechorionated immediately after fertilization, allowed to develop to the desired stage, deyolked either manually or, in the case of H3K36me3 mechanically[32], and snap frozen in liquid nitrogen. For every antibody, equal numbers of embryos were loaded for each developmental stage (RNA pol II ($n = 2$), H3K36me3 ($n = 30$), H3K4me3 ($n = 5$) and H3 ($n = 1$)). Samples were run on 6% (RNA pol II) or 15% (histones) acrylamide gels. Tubulin was analysed on all blots as a loading control. Western blots with antibodies that recognize phosphoserine 5 or 2 on the C-terminal domain of RNA pol II indicated that the bandshift in Fig. 1a is due to phosphorylation of the C-terminal domain (data not shown).

Antibodies. Antibodies used were histone H3 (Abcam ab1791), H3K4me3 (Abcam ab8580 and Millipore 07-473), H3K27me3 (Upstate 07-449), H3K36me3 (Abcam ab9050), RNA pol II (8WG16/Covance MMS-126R), phosphoserine 5 version of RNA pol II (H14/Abcam ab24759), phosphoserine 2 version of RNA pol II (H5/Abcam ab24758) and α-tubulin (Sigma T6074).

ChIP. ChIP was performed essentially as described[28] with some minor modifications. Briefly, for each condition, ~200 embryos were carefully staged, dechorionated and fixed in 1.85% formaldehyde for 15 min at 20 °C. Formaldehyde was quenched by adding glycine to a final concentration of 0.125 M. Embryos were rinsed three times in ice-cold PBS, immediately resuspended in cell lysis buffer (10 mM Tris-HCl pH7.5/10 mM NaCl/0.5% NP40) and lysed for 15 min on ice. Nuclei were collected by centrifugation, resuspended in nuclei lysis buffer (50 mM Tris-HCl pH 7.5/10 mM EDTA/1% SDS) and lysed for 10 min on ice. Samples were diluted three times in IP dilution buffer (16.7 mM Tris-HCl pH 7.5/167 mM NaCl/1.2 mM EDTA/0.01% SDS) and sonicated to obtain fragments of ~500 bp. Triton X-100 was added to a final concentration of 0.75% and the lysate was incubated overnight while rotating at 4 °C with 25–50 µl of protein G magnetic Dynabeads (Invitrogen) that had been pre-bound to an excess amount of antibody. Bound complexes were extensively washed with RIPA (50 mM HEPES pH 7.6/1 mM EDTA/0.7% DOC/1% Igepal/0.5 M LiCl) and TBS and then eluted from the beads with elution buffer (50 mM $NaHCO_3$/1% SDS). Crosslinks were reversed overnight at 65 °C, RNA degraded by a 30 min RNaseA treatment and DNA purified by the QIAquick PCR purification kit (Qiagen).

Sequential ChIP. The first round of ChIP was essentially as described above, except for two changes: (1) tosylactivated magnetic Dynabeads (Invitrogen) were used to which antibody was crosslinked according to the protocol provided by the manufacturer; and (2) chromatin was eluted from the beads in a milder elution buffer (50 mM $NaHCO_3$/0.1% SDS). Elution was performed at 20 °C for 15 min in 120 µl elution buffer. After removal of the beads, 10 µl was removed from each sample for subsequent analysis of the first ChIP. The volume was adjusted to 300 µl with standard elution buffer (1% SDS), and this sample was then used in the single ChIP protocol. The remaining 100 µl was used for sequential ChIP. The eluate was diluted tenfold in IP reconstitution buffer (30.9 mM Tris pH 7.5/123 mM NaCl/4.6 mM EDTA/0.37% SDS) and Triton X-100 was added to a final concentration of 0.75% to reconstitute the original IP buffer composition. Samples were then pre-cleared by incubating them with empty beads for 30 min at 4 °C. After removal of the beads, one-half of each sample was subjected to a second round of immunoprecipitation with the tosylactivated beads crosslinked to the antibody against the 'other' modification (that is, the one that was not used in the first ChIP) and the other half was incubated with empty beads as a negative control. Washing, elution and reversal of the crosslinks were performed as in single ChIP. (See also Supplementary Fig. 8.) All sequential ChIPs were normalized to their input (that is, the signal of the modification analysed in the first round of ChIP). Fold enrichment was calculated over a negative control: *hoxa13b-ups* served as a negative control for H3K27me3 > H3K4me3 sequential ChIPs, because this *hoxa13b* upstream region is occupied by H3K27me3 but not H3K4me3; *β-actin 1* served as a negative control for H3K4me3 > H3K27me3 sequential ChIPs, because this region is occupied by H3K4me3 but not H3K27me3. See Supplementary Discussion 2 for details.

ChIP–chip. For analysis on tiling arrays, ChIP material was amplified as previously described[29] using the Sigma GenomePlex Kit to perform one round of amplification. Sample labelling, hybridization, washing and scanning were performed as described in the NimbleChip Arrays User's Guide. H3K4me3 (Abcam ab8580), H3K27me3 (Upstate 07-449) and H3K36me3 (Abcam ab9050) ChIPs were normalized to H3 (ab1791) and RNA pol II (8WG16/Covance MMS-126R) ChIPs were normalized to input.

Tiling array design and hybridization. Sequences for each genomic region were retrieved from the Sanger Institute Zv7 sequence via the UCSC genome browser. Repetitive sequence elements were removed by RepeatMasker. A total of 385,000 50mer probes were printed using NimbleGen technology (NimbleGen) and were spaced at 30 bp, resulting in 80-bp resolution. Tiled regions include specifically selected genes (that is, maternally provided genes, genes that are transcribed at zygotic genome activation, genes important in development and genes known to be expressed in later stages) as well as two large contiguous genomic regions on chromosomes 3 and 11. A comparison of the data obtained using all genes on the array with the data obtained using only the genes on the two contiguous regions resulted in very similar histone modification profiles showing that the array is a good representation of the genome. All co-precipitated DNA fragments from ChIP experiments were labelled and hybridized to the custom-designed array using the standard NimbleGen protocol as described[12]. Probe intensities were extracted using the standard ChIP analysis settings provided in the Nimblescan software (NimbleGen).

ChIP–chip normalization and peak calling. After scanning, each array was normalized by MA2C[13]. Each probe was assigned a MA2C score to reflect the normalized and window-averaged $\log_2$ ratio of ChIP enrichment over control. One replicate was performed at the 256-cell stage and the absence of signal was confirmed by ChIP–qPCR analysis at that stage (Supplementary Fig. 7a). Two biological replicates were performed at dome stage/30% epiboly. The overall Pearson correlations for MA2C scores of all probes between two biological replicates at dome stage /30% epiboly were 0.91 (H3K4me3), 0.92 (H3K27me3), 0.91 (H3K36me3) and 0.78 (RNA pol II), respectively. Because there may be minor staging differences between replicates, the data from one replicate (the first) was used in all subsequent analyses. Analysis of both replicates gave very similar results (Supplementary Fig. 7b). Enriched histone modification/RNA pol II domains were determined with the following approach: First, MA2C was applied to detect enriched peaks under two different threshold cutoffs: FDR 5% and FDR 20%; that is, each peak with a MA2C score of x is assigned an FDR based on (the number of control peaks over ChIP with a score above x) divided by (the number of ChIP peaks over control with a score above x). Second, adjacent peaks (under FDR 20%) were concatenated if the gap size separating them was smaller than 5 kb and there were either no probes in the gap, or if at least 75% of the probes in the gap had a MA2C score larger than 0. Third, domains that were smaller than 1.5 kb and did not contain peaks under FDR 5% were discarded.

Histone modification/RNA pol II status of genes. A total of 822 RefSeq genes were used in the analysis (this included all the RefSeq genes larger than 1.5 kb that were tiled on the array from at least 1.5 kb upstream of the transcription start site up to the transcription termination site (−1.5 kb to TSS, TTS)). Enriched chromatin domains (as defined above) were associated with these RefSeq genes. By following the decision trees shown in Supplementary Fig. 1, this resulted in three groups of genes for RNA pol II, H3K36me3 and H3K4me3 (high, moderate and low), and four groups of genes for H3K27me3 (high whole gene (W), high promoter (P), moderate and low).

Metagene profiles. The metagene profiles were generated from the average MA2C scores across all RefSeq genes (except in Supplementary Fig. 2, where only genes with the indicated profile are included) by CEAS (Cis-regulatory Element Annotation System)[33]. In the profiles, transcription units are shown as metagenes (that is, relative distance from TSS to TTS), whereas immediate upstream and downstream sequences are shown in absolute distance (bp).

Heat maps. Mean MA2C scores were used to generate the heat map. For H3K36me3, the mean MA2C score of the last two-thirds of the gene body was used, whereas for H3K4me3, and H3K27me3, the mean MA2C scores of the promoter regions ((−1.5 kb, 1.5 kb) to TSS) were used. To ensure unambiguous grouping of genes, genes with a moderate status in any histone modification (H3K4me3, H3K27me3, H3K36me3) were filtered out.

Gene Ontology. Gene Ontology (GO) analysis was performed using the DAVID web server (http://david.abcc.ncifcrf.gov/)[34,35]. Only GO terms with a P-value $< 1 \times 10^{-5}$ were considered enriched and shown.

Pairwise correlation between ChIP–chip data sets. Scatter plots were generated to display the pairwise correlation between ChIP–chip data. For this, the mean MA2C score was used. For H3K36me3, the mean MA2C score of the last two-thirds of the gene body was used for each gene, whereas for the other data, the mean MA2C scores of the promoter region ((−1.5 kb, 1.5 kb) to TSS) were used.

Expression analysis. Expression analysis was performed using a NimbleGen expression array containing 392,778 probes with a maximum of 12 probes from 37,157 known and predicted zebrafish genes. Groups of 12 embryos from the indicated time points were homogenized in Trizol and RNA was extracted using the standard Trizol protocol. One microgram of the resulting total RNA was amplified into aRNA using the Ambion MessageAmpII aRNA kit. Fluorescent labelling of samples was performed by random-prime reverse transcription of 10 µg input aRNA with either Cy3- or Cy5-dNTPs. Arrays were hybridized and washed as per standard NimbleGen protocol. Microarray scanning was performed on an Axon Genepix scanner. Expression microarray quality control

©2010 Macmillan Publishers Limited. All rights reserved

and data analysis was performed using NimbleScan and R/Bioconductor with RMA (quantile) normalization. To determine if a gene was maternally provided, the averaged expression data from 1- and 4-cell stage embryos was used. If a gene's average expression level at these stages was larger than the mean expression value of all genes, the gene was regarded as maternally provided, and removed from the expression analysis of the dome stage/30% epiboly sample. A different cutoff that eliminated fewer genes (that is, defined a maternal gene if its expression was larger than twice the mean expression value of all genes) was also tested. This analysis produced similar results, demonstrating that the more stringent cutoff we used in our analysis eliminates most or all maternally provided genes.

31. Sagasti, A., Guido, M. R., Raible, D. W. & Schier, A. F. Repulsive interactions shape the morphologies and functional arrangement of zebrafish peripheral sensory arbors. *Curr. Biol.* **15,** 804–814 (2005).
32. Link, V., Shevchenko, A. & Heisenberg, C. Proteomics of early zebrafish embryos. *BMC Dev. Biol.* **6,** 1 (2006).
33. Shin, H., Liu, T., Manrai, A. K. & Liu, X. S. CEAS: Cis-regulatory Element Annotation System. *Bioinformatics* **25,** 2605–2606 (2009).
34. Dennis, G. Jr *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4,** 3 (2003).
35. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4,** 44–57 (2009).

©2010 **Macmillan Publishers Limited. All rights reserved**