

# Comparative analysis of metazoan chromatin organization

Joshua W. K. Ho<sup>1,2\*†</sup>, Youngsook L. Jung<sup>1,2\*</sup>, Tao Liu<sup>3,4\*†</sup>, Burak H. Alver<sup>1</sup>, Soohyun Lee<sup>1</sup>, Kohta Ikegami<sup>5†</sup>, Kyung-Ah Sohn<sup>6,7</sup>, Aki Minoda<sup>8,9†</sup>, Michael Y. Tolstorukov<sup>1,2,10</sup>, Alex Appert<sup>11</sup>, Stephen C. J. Parker<sup>12,13</sup>, Tingting Gu<sup>14</sup>, Anshul Kundaje<sup>15,16†</sup>, Nicole C. Riddle<sup>14†</sup>, Eric Bishop<sup>1,17</sup>, Thea A. Egelhofer<sup>18</sup>, Sheng'en Shawn Hu<sup>19</sup>, Artyom A. Alekseyenko<sup>2,20</sup>, Andreas Rechtsteiner<sup>18</sup>, Dalal Asker<sup>21,22</sup>, Jason A. Belsky<sup>23</sup>, Sarah K. Bowman<sup>10</sup>, Q. Brent Chen<sup>5</sup>, Ron A.-J. Chen<sup>11</sup>, Daniel S. Day<sup>1,24</sup>, Yan Dong<sup>11</sup>, Andrea C. Dose<sup>25</sup>, Xikun Duan<sup>19</sup>, Charles B. Epstein<sup>16</sup>, Sevinc Ercan<sup>5,26</sup>, Elise A. Feingold<sup>13</sup>, Francesco Ferrari<sup>1</sup>, Jacob M. Garrigues<sup>18</sup>, Nils Gehlenborg<sup>1,16</sup>, Peter J. Good<sup>13</sup>, Psalm Haseley<sup>1,2</sup>, Daniel He<sup>9</sup>, Moritz Herrmann<sup>11</sup>, Michael M. Hoffman<sup>27</sup>, Tess E. Jeffers<sup>5†</sup>, Peter V. Kharchenko<sup>1</sup>, Paulina Kolasinska-Zwierz<sup>11</sup>, Chitra V. Kotwaliwale<sup>9,28</sup>, Nischay Kumar<sup>15,16</sup>, Sasha A. Langley<sup>8,9</sup>, Erica N. Larschan<sup>29</sup>, Isabel Latorre<sup>11</sup>, Maxwell W. Libbrecht<sup>30</sup>, Xueqiu Lin<sup>19</sup>, Richard Park<sup>1,17</sup>, Michael J. Pazin<sup>13</sup>, Hoang N. Pham<sup>8,9,28</sup>, Annette Plachetka<sup>2,20</sup>, Bo Qin<sup>19</sup>, Yuri B. Schwartz<sup>21,31</sup>, Noam Shores<sup>16</sup>, Przemyslaw Stempor<sup>11</sup>, Anne Vielle<sup>11</sup>, Chengyang Wang<sup>19</sup>, Christina M. Whittle<sup>9,28</sup>, Hailing Xue<sup>1,2</sup>, Robert E. Kingston<sup>10</sup>, Ju Han Kim<sup>7,32</sup>, Bradley E. Bernstein<sup>16,28,33</sup>, Abby F. Dernburg<sup>8,9,28</sup>, Vincenzo Pirrotta<sup>21</sup>, Mitzi I. Kuroda<sup>2,20</sup>, William S. Noble<sup>30,34</sup>, Thomas D. Tullius<sup>17,35</sup>, Manolis Kellis<sup>15,16</sup>, David M. MacAlpine<sup>23</sup>, Susan Strome<sup>18</sup>, Sarah C. R. Elgin<sup>14</sup>, Xiaole Shirley Liu<sup>3,4,16</sup>, Jason D. Lieb<sup>5†</sup>, Julie Ahringer<sup>11</sup>, Gary H. Karpen<sup>8,9</sup> & Peter J. Park<sup>1,2,36</sup>

Genome function is dynamically regulated in part by chromatin, which consists of the histones, non-histone proteins and RNA molecules that package DNA. Studies in *Caenorhabditis elegans* and *Drosophila melanogaster* have contributed substantially to our understanding of molecular mechanisms of genome function in humans, and have revealed conservation of chromatin components and mechanisms<sup>1–3</sup>. Nevertheless, the three organisms have markedly different genome sizes, chromosome architecture and gene organization. On human and fly chromosomes, for example, pericentric heterochromatin flanks single centromeres, whereas worm chromosomes have dispersed heterochromatin-like regions enriched in the distal chromosomal ‘arms’, and centromeres distributed along their lengths<sup>4,5</sup>. To systematically investigate chromatin organization and associated gene regulation across species, we generated and analysed a large collection of genome-wide chromatin data sets from cell lines and developmental stages in worm, fly and human. Here we present over 800 new data sets from our ENCODE and modENCODE consortia, bringing the total to over 1,400. Comparison of combinatorial patterns of histone modifications, nuclear lamina-associated domains, organization of large-scale topological domains, chromatin environment at promoters and

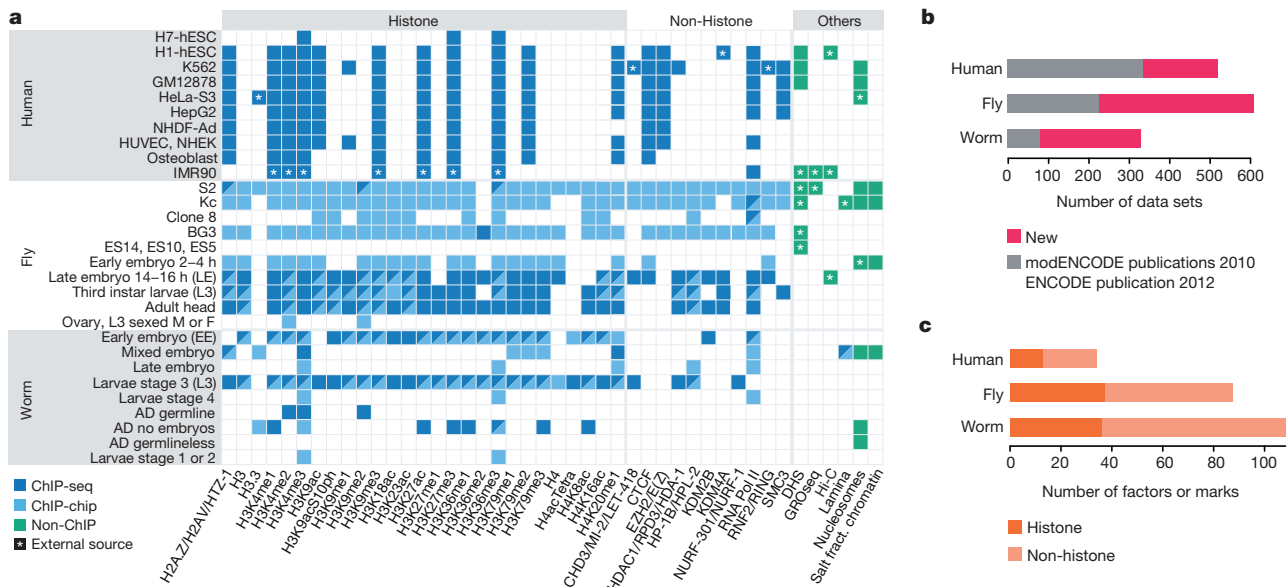
enhancers, nucleosome positioning, and DNA replication patterns reveals many conserved features of chromatin organization among the three organisms. We also find notable differences in the composition and locations of repressive chromatin. These data sets and analyses provide a rich resource for comparative and species-specific investigations of chromatin composition, organization and function.

We used chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) or microarray hybridization (ChIP-chip) to generate profiles of core histones, histone variants, histone modifications and chromatin-associated proteins (Fig. 1, Supplementary Fig. 1 and Supplementary Tables 1 and 2). Additional data include DNase I hypersensitivity sites in fly and human cells, and nucleosome occupancy maps in all three organisms. Compared to our initial publications<sup>1–3</sup>, this represents a tripling of available fly and worm data sets and a substantial increase in human data sets (Fig. 1b, c). Uniform quality standards for experimental protocols, antibody validation and data processing were used throughout the projects<sup>6</sup>. Detailed analyses of related transcriptome and transcription factor data are presented in accompanying papers<sup>7,8</sup>.

We performed systematic cross-species comparisons of chromatin composition and organization, focusing on targets profiled in at least two

<sup>1</sup>Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>2</sup>Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>3</sup>Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. <sup>4</sup>Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, 450 Brookline Avenue, Boston, Massachusetts 02115, USA. <sup>5</sup>Department of Biology and Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. <sup>6</sup>Department of Information and Computer Engineering, Ajou University, Suwon 443-749, Korea. <sup>7</sup>Systems Biomedical Informatics Research Center, College of Medicine, Seoul National University, Seoul 110-799, Korea. <sup>8</sup>Department of Genome Dynamics, Life Sciences Division, Lawrence Berkeley National Lab, Berkeley, California 94720, USA. <sup>9</sup>Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California 94720, USA. <sup>10</sup>Department of Molecular Biology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. <sup>11</sup>The Gurdon Institute and Department of Genetics, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. <sup>12</sup>National Institute of General Medical Sciences, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>13</sup>National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>14</sup>Department of Biology, Washington University in St. Louis, St. Louis, Missouri 63130, USA. <sup>15</sup>Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>16</sup>Broad Institute, Cambridge, Massachusetts 02141, USA. <sup>17</sup>Program in Bioinformatics, Boston University, Boston, Massachusetts 02215, USA. <sup>18</sup>Department of Molecular, Cell and Developmental Biology, University of California Santa Cruz, Santa Cruz, California 95064, USA. <sup>19</sup>Department of Bioinformatics, School of Life Science and Technology, Tongji University, Shanghai 200092, China. <sup>20</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>21</sup>Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854, USA. <sup>22</sup>Food Science and Technology Department, Faculty of Agriculture, Alexandria University, 21545 El-Shatby, Alexandria, Egypt. <sup>23</sup>Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA. <sup>24</sup>Harvard/MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139, USA. <sup>25</sup>Department of Anatomy Physiology and Cell Biology, University of California Davis, Davis, California 95616, USA. <sup>26</sup>Department of Biology, Center for Genomics and Systems Biology, New York University, New York, New York 10003, USA. <sup>27</sup>Princess Margaret Cancer Centre, Toronto, Ontario M6G 1L7, Canada. <sup>28</sup>Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA. <sup>29</sup>Department of Molecular Biology, Cellular Biology and Biochemistry, Brown University, Providence, Rhode Island 02912, USA. <sup>30</sup>Department of Computer Science and Engineering, University of Washington, Seattle, Washington 98195, USA. <sup>31</sup>Department of Molecular Biology, Umea University, 901 87 Umea, Sweden. <sup>32</sup>Seoul National University Biomedical Informatics, Division of Biomedical Informatics, College of Medicine, Seoul National University, Seoul 110-799, Korea. <sup>33</sup>Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. <sup>34</sup>Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. <sup>35</sup>Department of Chemistry, Boston University, Boston, Massachusetts 02215, USA. <sup>36</sup>Informatics Program, Children's Hospital, Boston, Massachusetts 02215, USA. †Present addresses: Victor Chang Cardiac Research Institute and The University of New South Wales, Sydney, New South Wales 2052, Australia (J.W.K.H.); Department of Biochemistry, University at Buffalo, Buffalo, New York 14203, USA (T.L.); Department of Molecular Biology and Lewis Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08540, USA (K.L., T.E.J.); Department of Human Genetics, University of Chicago, Chicago, Illinois 06037, USA (J.D.L.); Division of Genomic Technologies, Center for Life Science Technologies, RIKEN, Yokohama 230-0045, Japan (A.M.); Department of Genetics, Department of Computer Science, Stanford University, Stanford, California 94305, USA (A.K.); Department of Biology, The University of Alabama at Birmingham, Birmingham, Alabama 35294, USA (N.C.R.).

\*These authors contributed equally to this work.



**Figure 1 | Overview of the data set.** **a**, Histone modifications, chromosomal proteins and other profiles mapped in at least two species (see Supplementary Fig. 1 for the full data set and Supplementary Table 1 for detailed descriptions). Different protein names for orthologues are separated by ‘/’ (see Supplementary Table 2). **b**, The number of all data sets generated by this

organisms (Fig. 1). Sample types used most extensively in our analyses are human cell lines H1-hESC, GM12878 and K562; fly late embryos, third instar larvae and cell lines S2, Kc, BG3; and worm early embryos and stage 3 larvae. Our conclusions are summarized in Extended Data Table 1.

Not surprisingly, the three species show many common chromatin features. Most of the genome in each species is marked by at least one histone modification (Supplementary Fig. 2), and modification patterns are similar around promoters, gene bodies, enhancers and other chromosomal elements (Supplementary Figs 3–12). Nucleosome occupancy patterns around protein-coding genes and enhancers are also largely similar across species, although we observed subtle differences in H3K4me3 enrichment patterns around transcription start sites (TSSs) (Extended Data Fig. 1a and Supplementary Figs 12–14). The configuration and composition of large-scale features such as lamina-associated domains (LADs) are similar (Supplementary Figs 15–17). LADs in human and fly are associated with late replication and H3K27me3 enrichment, suggesting a repressive chromatin environment (Supplementary Fig. 18). Finally, DNA structural features associated with nucleosome positioning are strongly conserved (Supplementary Figs 19 and 20).

Although patterns of histone modifications across active and silent genes are largely similar in all three species, there are some notable differences (Extended Data Fig. 1b). For example, H3K23ac is enriched at promoters of expressed genes in worm, but is enriched across gene bodies of both expressed and silent genes in fly. H4K20me1 is enriched on both expressed and silent genes in human but only on expressed genes in fly and worm (Extended Data Fig. 1b). Enrichment of H3K36me3 in genes expressed with stage or tissue specificity is lower than in genes expressed broadly, possibly because profiling was carried out on mixed tissues (Supplementary Figs 21–23; see Supplementary Methods). Although the co-occurrence of pairs of histone modifications is largely similar across the three species, there are clearly some species-specific patterns (Extended Data Fig. 1c and Supplementary Figs 24 and 25).

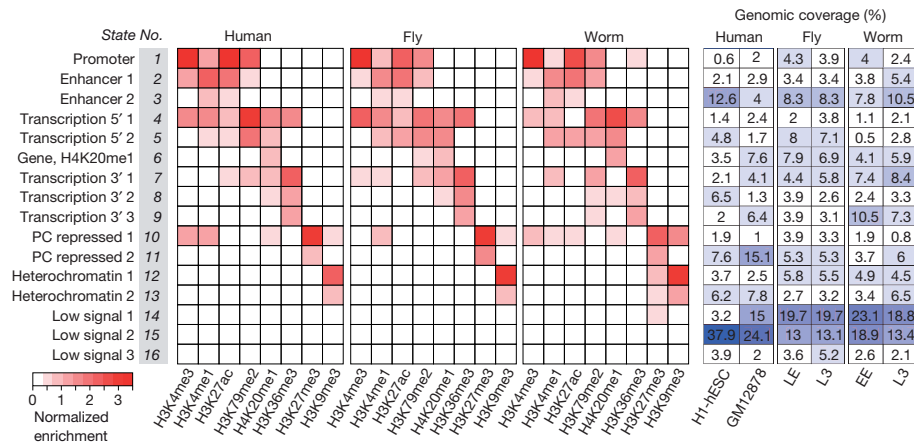
Previous studies showed that in human<sup>9,10</sup> and fly<sup>11,11</sup> prevalent combinations of marks or ‘chromatin states’ correlate with functional features such as promoters, enhancers, transcribed regions, Polycomb-associated domains, and heterochromatin. ‘Chromatin state maps’ provide a concise and systematic annotation of the genome. To compare chromatin

and previous consortia publications<sup>1–3</sup> (new, 815; old, 638). Each data set corresponds to a replicate-merged normalized profile of a histone, histone variant, histone modification, non-histone chromosomal protein, nucleosome or salt-fractionated nucleosome. **c**, The number of unique histone marks or non-histone chromosomal proteins profiled.

states across the three organisms, we developed and applied a novel hierarchical non-parametric machine-learning method called hiHMM (see Supplementary Methods) to generate chromatin state maps from eight histone marks mapped in common, and compared the results with published methods (Fig. 2 and Supplementary Figs 26–28). We find that combinatorial patterns of histone modifications are largely conserved. Based on correlations with functional elements (Supplementary Figs 29–32), we categorized the 16 states into six groups: promoter (state 1), enhancer (states 2 and 3), gene body (states 4–9), Polycomb-repressed (states 10 and 11), heterochromatin (states 12 and 13), and weak or low signal (states 14–16).

Heterochromatin is a classically defined and distinct chromosomal domain with important roles in genome organization, genome stability, chromosome inheritance and gene regulation. It is typically enriched for H3K9me3 (ref. 12), which we used as a proxy for identifying heterochromatic domains (Fig. 3a and Supplementary Figs 33 and 34). As expected, the majority of the H3K9me3-enriched domains in human and fly are concentrated in the pericentromeric regions (as well as other specific domains, such as the Y chromosome and fly 4th chromosome), whereas in worm they are distributed throughout the distal chromosomal ‘arms’<sup>11,13,14</sup> (Fig. 3a). In all three organisms, we find that more of the genome is associated with H3K9me3 in differentiated cells and tissues compared to embryonic cells and tissues (Extended Data Fig. 2a). We also observe large cell-type-specific blocks of H3K9me3 in human and fly<sup>11,14,15</sup> (Supplementary Fig. 35). These results suggest a molecular basis for the classical concept of ‘facultative heterochromatin’ formation to silence blocks of genes as cells specialize.

Two distinct types of transcriptionally repressed chromatin have been described. As discussed above, classical ‘heterochromatin’ is generally concentrated in specific chromosomal regions and enriched for H3K9me3 and also H3K9me2 (ref. 12). In contrast, Polycomb-associated silenced domains, involved in cell-type-specific silencing of developmentally regulated genes<sup>11,14</sup>, are scattered across the genome and enriched for H3K27me3. We found that the organization and composition of these two types of transcriptionally silent domains differ across species. First, human, fly and worm display significant differences in H3K9 methylation patterns. H3K9me2 shows a stronger correlation with H3K9me3 in fly than in worm ( $r = 0.89$  versus  $r = 0.40$ , respectively), whereas H3K9me2



**Figure 2 | Shared and organism-specific chromatin states.** Sixteen chromatin states derived by joint segmentation using hiHMM (see Supplementary Methods) based on enrichment patterns of eight histone

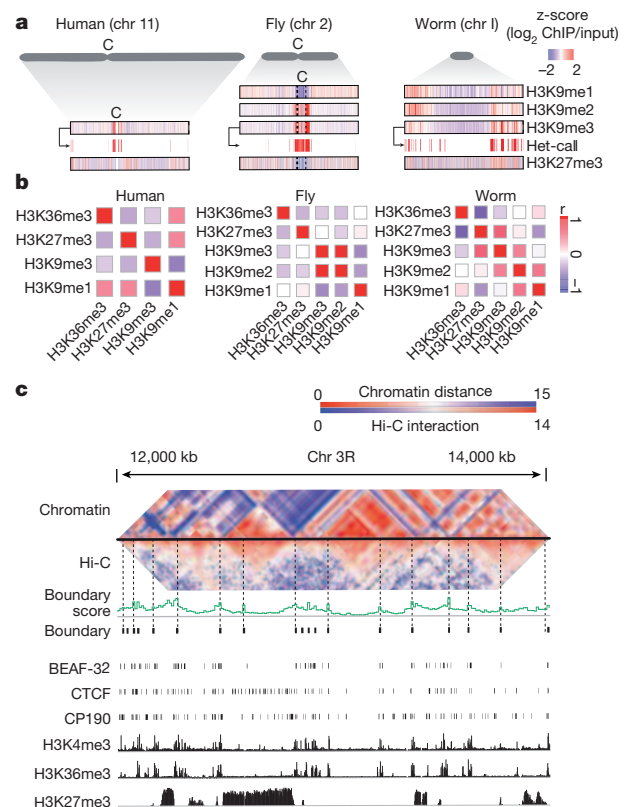
is well correlated with H3K9me1 in worm but not in fly ( $r = 0.44$  versus  $r = -0.32$ , respectively) (Fig. 3b). These findings suggest potential differences in heterochromatin in the three organisms (see below). Second, the chromatin state maps reveal two distinct types of Polycomb-associated repressed regions: strong H3K27me3 accompanied by marks for active genes or enhancers (Fig. 2, state 10; perhaps due to mixed tissues in whole embryos or larvae for fly and worm), and strong H3K27me3 without active marks (state 11) (see also Supplementary Fig. 31). Third, we observe a worm-specific association of H3K9me3 and H3K27me3. These two marks are enriched together in states 12 and 13 in worm but not in human and fly. This unexpectedly strong association between H3K9me3 and H3K27me3 in worm (observed with several validated antibodies; Extended Data Fig. 2b) suggests a species-specific difference in the organization of silent chromatin.

We also compared the patterns of histone modifications on expressed and silent genes in euchromatin and heterochromatin (Extended Data Fig. 2c and Supplementary Fig. 36). We previously reported prominent depletion of H3K9me3 at TSSs and high levels of H3K9me3 in the gene bodies of expressed genes located in fly heterochromatin<sup>14</sup>, and now find a similar pattern in human (Extended Data Fig. 2c and Supplementary Fig. 36). In these two species, H3K9me3 is highly enriched in the body of both expressed and silent genes in heterochromatic regions. In contrast, expressed genes in worm heterochromatin have lower H3K9me3 enrichment across gene bodies compared to silent genes (Extended Data Fig. 2c and Supplementary Figs 36, and 37). There are also conspicuous differences in the patterns of H3K27me3 in the three organisms. In human and fly, H3K27me3 is highly associated with silent genes in euchromatic regions, but not with silent genes in heterochromatic regions. In contrast, consistent with the worm-specific association between H3K27me3 and H3K9me3, we observe high levels of H3K27me3 on silent genes in worm heterochromatin, whereas silent euchromatic genes show modest enrichment of H3K27me3 (Extended Data Fig. 2c and Supplementary Fig. 36).

Our results suggest three distinct types of repressed chromatin (Extended Data Fig. 3). The first contains H3K27me3 with little or no H3K9me3 (human and fly states 10 and 11, and worm state 11), corresponding to developmentally regulated Polycomb-silenced domains in human and fly, and probably in worm as well. The second is enriched for H3K9me3 and lacks H3K27me3 (human and fly states 12 and 13), corresponding to constitutive, predominantly pericentric heterochromatin in human and fly, which is essentially absent from the worm genome. The third contains both H3K9me3 and H3K27me3 and occurs predominantly in worm (worm states 10, 12 and 13). Co-occurrence of these marks is consistent with the observation that H3K9me3 and H3K27me3 are both required for silencing of heterochromatic transgenes in worms<sup>16</sup>. H3K9me3 and H3K27me3 may reside on the same or adjacent nucleosomes in individual cells<sup>17,18</sup>; alternatively the two marks may occur in different

marks. The genomic coverage of each state in each cell type or developmental stage is also shown (see Supplementary Figs 26–32 for detailed analysis of the states). States are named for putative functional characteristics.

cell types in the embryos and larvae analysed here. Further studies are needed to resolve this and determine the functional consequences of the overlapping distributions of H3K9me3 and H3K27me3 observed in worm.



**Figure 3 | Genome-wide organization of heterochromatin.** **a**, Enrichment profiles of H3K9me1, H3K9me2, H3K9me3 and H3K27me3, and identification of heterochromatin domains based on H3K9me3 (illustrated for human H1-hESC, fly L3 and worm L3). For fly chr 2, 2L, 2LHet, 2RHet and 2R are concatenated (dashed lines). C, centromere; Het, heterochromatin. **b**, Genome-wide correlation among H3K9me1, H3K9me2, H3K9me3, H3K27me3 and H3K36me3 in human K562 cells, fly L3 and worm L3; no H3K9me2 profile is available for human. **c**, Comparison of Hi-C-based and chromatin-based topological domains in fly LE. Heat maps of similarity matrices for histone modification and Hi-C interaction frequencies are juxtaposed (see Supplementary Fig. 40).

Genome-wide chromatin conformation capture (Hi-C) assays have revealed prominent topological domains in human<sup>19</sup> and fly<sup>20,21</sup>. Although their boundaries are enriched for insulator elements and active genes<sup>19,20</sup> (Supplementary Fig. 38), the interiors generally contain a relatively uniform chromatin state: active, Polycomb-repressed, heterochromatin, or low signal<sup>22</sup> (Supplementary Fig. 39). We found that chromatin state similarity between neighbouring regions correlates with chromatin interaction domains determined by Hi-C (Fig. 3c, Supplementary Fig. 40 and Supplementary Methods). This suggests that topological domains can be largely predicted by chromatin marks when Hi-C data are not available (Supplementary Figs 41 and 42).

*C. elegans* and *D. melanogaster* have been used extensively for understanding human gene function, development and disease. Our analyses of chromatin architecture and the large public resource we have generated provide a blueprint for interpreting experimental results in these model systems, extending their relevance to human biology. They also provide a foundation for researchers to investigate how diverse genome functions are regulated in the context of chromatin structure.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 November 2013; accepted 29 April 2014.


- modENCODE Consortium *et al.* Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1787–1797 (2010).
- Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
- The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Gassmann, R. *et al.* An inverse relationship to germline transcription defines centromeric chromatin in *C. elegans*. *Nature* **484**, 534–537 (2012).
- Blower, M. D., Sullivan, B. A. & Karpen, G. H. Conserved organization of centromeric chromatin in flies and humans. *Dev. Cell* **2**, 319–330 (2002).
- Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
- Gerstein, M. B. *et al.* Comparative analysis of the transcriptome across distant species. *Nature* <http://dx.doi.org/10.1038/nature13424> (this issue).
- Boyle, A. P. *et al.* Comparative analysis of regulatory information and circuits across distant species. *Nature* <http://dx.doi.org/10.1038/nature13668> (this issue).
- Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
- Hoffman, M. M. *et al.* Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res.* **41**, 827–841 (2013).
- Kharchenko, P. V. *et al.* Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**, 480–485 (2011).
- Elgin, S. C. & Reuter, G. Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harb. Perspect. Biol.* **5**, a017780 (2013).
- Liu, T. *et al.* Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Res.* **21**, 227–236 (2011).
- Riddle, N. C. *et al.* Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin. *Genome Res.* **21**, 147–163 (2011).
- Hawkins, R. D. *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010).
- Towbin, B. D. *et al.* Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* **150**, 934–947 (2012).
- Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. & Young, R. A. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* **23**, 2484–2489 (2009).
- Voigt, P. *et al.* Asymmetrically modified nucleosomes. *Cell* **151**, 181–193 (2012).
- Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
- Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* **48**, 471–484 (2012).
- Zhu, J. *et al.* Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152**, 642–654 (2013).

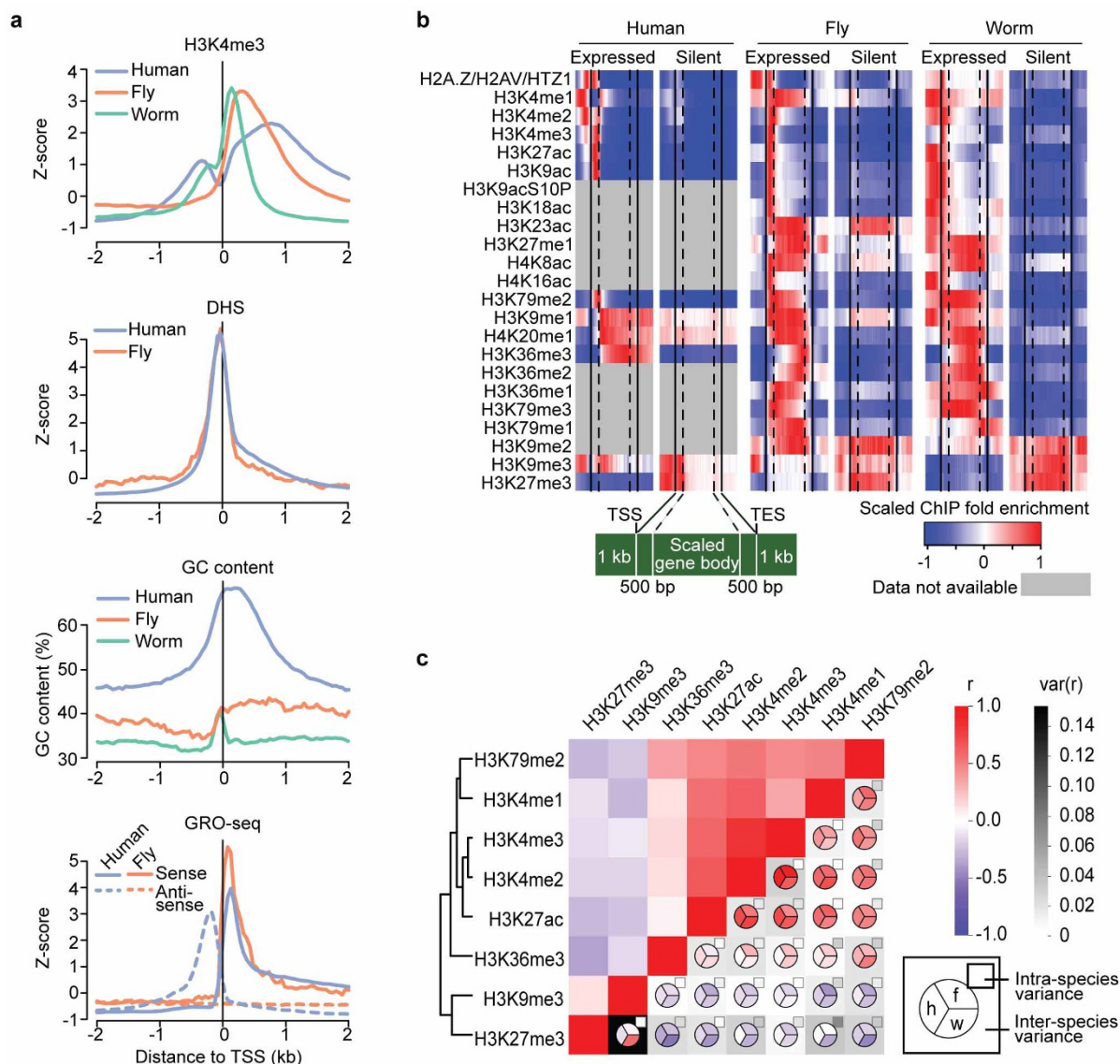
**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This project was mainly funded by NHGRI U01HG004258 (G.H.K., S.C.R.E., M.I.K., P.J.P., V.P.), U01HG004270 (J.D.L., J.A., A.F.D., X.S.L., S.S.), U01HG004279 (D.M.M.), U54HG004570 (B.E.B.) and U01HG004695 (W.S.N.). It is also supported by NHBIB 5RL9EB008539 (J.W.K.H.), NHGRI K99HG006259 (M.M.H.), NIGMS fellowships (S.C.J.P., E.N.L.), NIH U54CA121852 (T.D.T.), NSF 1122374 (D.S.D.), National Natural Science Foundation of China 31028011 (X.S.L.), MEST Korea MHW-2013-HI13C2164 (J.H.K.), NRF-2012-0000994 (K.-A.S.), and Wellcome Trust 54523 (J.A.). We thank D. Acevedo and C. Kennedy for technical assistance.

**Author Contributions** J.W.K.H., Y.L.J., T.L., B.H.A., S.L., K.-A.S., M.Y.T., S.C.J.P., A.K., E.B., S.S.H. and A.R. led the data analysis. K.I., A.M., A.A., T.G., N.C.R., T.A.E., A.A.A. and D.A. led the data production. J.A.B., D.S.D., X.D., F.F., N.G., P.H., M.M.H., P.V.K., N.K., E.N.L., M.W.L., R.P., N.S., C.W. and H.X. analysed data. S.K.B., Q.B.C., R.A.-J.C., Y.D., A.C.D., C.B.E., S.E., J.M.G., D.H., M.H., T.E.J., P.K.-Z., C.V.K., S.A.L., I.L., X.L., H.N.P., A.P., B.Q., P.S., Y.B.S., A.V. and C.M.W. produced data. E.A.F., P.J.G. and M.J.P. carried out NIH scientific project management. The role of the NIH Project Management Group was limited to coordination and scientific management of the modENCODE and ENCODE consortia. J.W.K.H., Y.L.J., T.L., B.H.A., S.L., K.-A.S., M.Y.T., S.C.J.P., S.S.H., A.R., K.I., T.D.T., M.K., D.M.M., S.S., S.C.R.E., X.S.L., J.D.L., J.A., G.H.K. and P.J.P. wrote the paper. The group leaders for data analysis or production were R.E.K., J.H.K., B.E.B., A.F.D., V.P., M.I.K., W.S.N., T.D.T., M.K., D.M.M., S.S., S.C.R.E., J.A., X.S.L., G.H.K., J.D.L. and P.J.P. The overall project managers were D.M.M., S.S., S.C.R.E., X.S.L., J.D.L., J.A., G.H.K. and P.J.P.

**Author Information** All data are available at the project data portals (<http://data.modencode.org>; <http://encodeproject.org>), modMine (<http://intermine.modencode.org>), or our database and web application ([http://encode-x.med.harvard.edu/data\\_sets/chromatin/](http://encode-x.med.harvard.edu/data_sets/chromatin/)). More information and annotated files are available at <http://encodeproject.org/comparative>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. B.H.A., S.L., K.I., K.-A.S., A.M., M.Y.T., A.A., S.C.J.P., T.G., A.K., N.C.R., E.B., T.A.E., S.S.H., A.A.A., A.R., D.A. contributed equally and should be considered co-second authors. Correspondence and requests for materials should be addressed to D.M.M. ([david.macalpine@duke.edu](mailto:david.macalpine@duke.edu)), S.S. ([ssstrom@ucsc.edu](mailto:ssstrom@ucsc.edu)), S.C.R.E. ([selgin@biology2.wustl.edu](mailto:selgin@biology2.wustl.edu)), X.S.L. ([xslu@jimmy.harvard.edu](mailto:xslu@jimmy.harvard.edu)), J.D.L. ([jdlieb@uchicago.edu](mailto:jdlieb@uchicago.edu)), J.A. ([ja219@cam.ac.uk](mailto:ja219@cam.ac.uk)), G.H.K. ([ghkarpen@lbl.gov](mailto:ghkarpen@lbl.gov)), or P.J.P. ([peter\\_park@hms.harvard.edu](mailto:peter_park@hms.harvard.edu)). Questions about worm data should be directed to S.S., J.D.L. or J.A., fly data to D.M.M., S.C.R.E. or G.H.K., and analysis to X.S.L. and P.J.P.

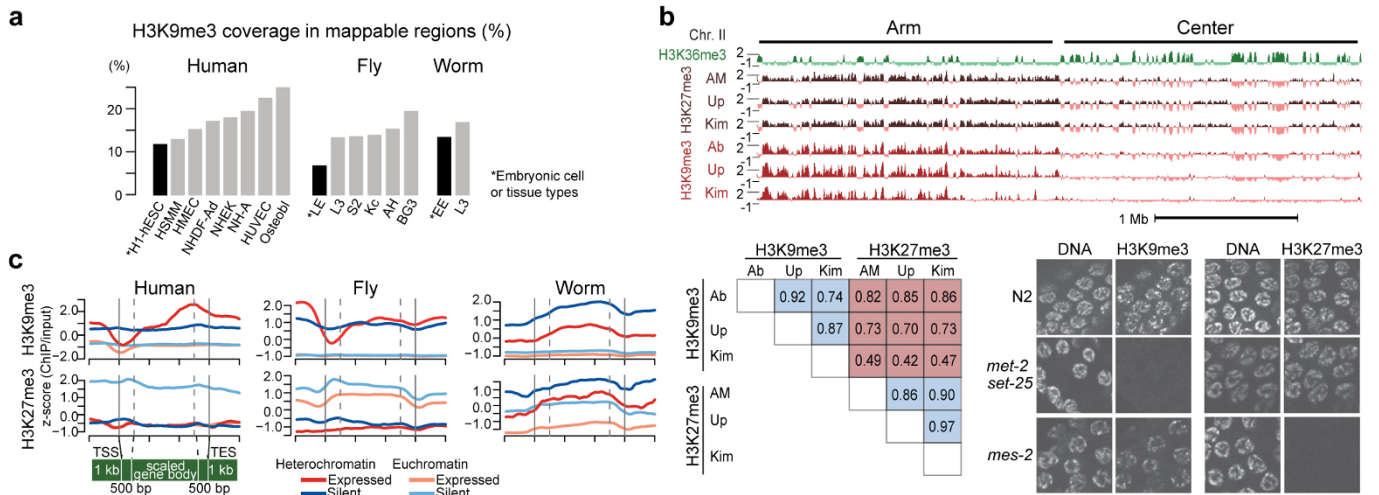
 This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>



**Extended Data Figure 1 | Chromatin features at TSSs and gene bodies, and co-occurrence of histone modifications.** **a**, Comparative analysis of promoter architecture at transcription start sites (TSSs). From the top, H3K4me3 (human GM12878, fly L3 and worm L3), DNase I hypersensitivity sites (DHSs), GC content, and nascent transcript (GRO-seq in human IMR90 and fly S2 cells). Human promoters, and to a lesser extent worm promoters (as defined using recently published capRNA-seq data<sup>23</sup>), exhibit a bimodal enrichment for H3K4me3 and other active marks around TSSs. In contrast, fly promoters clearly exhibit a unimodal distribution of active marks, downstream of TSSs. As genes that have a neighbouring gene within 1 kb of a TSS or TES (transcription end site) were removed from this analysis, any bimodal histone modification pattern cannot be attributed to nearby genes. This difference is also not explained by chromatin accessibility determined by DHS, or by fluctuations in GC content around the TSSs, although the GC profiles are highly variable across species. **b**, Average gene body profiles of histone modifications on

protein-coding genes in human GM12878, fly L3 and worm L3. **c**, Genome-wide correlations between histone modifications show intra- and inter-species similarities and differences. Top left, pairwise correlations between marks in each genome, averaged across all three species. Bottom right, pairwise correlations, averaged over cell types and developmental stages, within each species (pie chart), with inter-species variance (grey-scale background) and intra-species variance (grey-scale small rectangles) of correlation coefficients for human, fly and worm. Modifications enriched within or near actively transcribed genes are consistently correlated with each other in all three organisms. In contrast, we found a major difference in the co-occurrence pattern of two key repressive chromatin marks (black cell in bottom left): H3K27me3 (related to Polycomb (Pc)-mediated silencing) and H3K9me3 (related to heterochromatin). These two marks are strongly correlated at both developmental stages analysed in worm, whereas their correlation is low in human ( $r = -0.24$  to  $-0.06$ ) and fly ( $r = -0.03$  to  $-0.1$ ).

23. Chen, R. A. *et al.* The landscape of RNA polymerase II transcription initiation in *C. elegans* reveals promoter and enhancer architectures. *Genome Res.* **23**, 1339–1347 (2013).



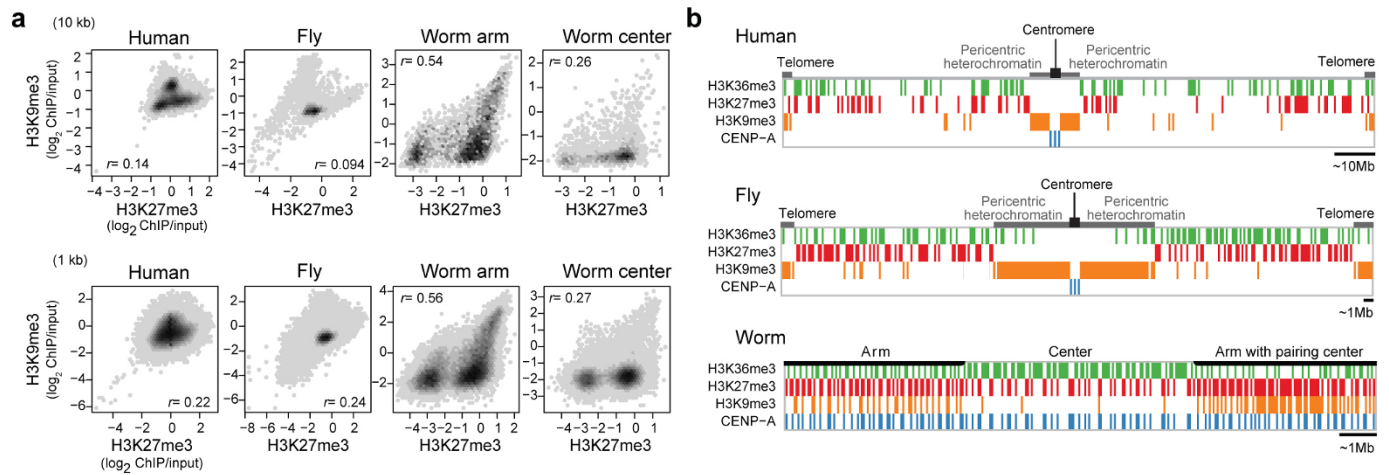
### Extended Data Figure 2 | Histone modifications in heterochromatin.

**a**, Genomic coverage of H3K9me3 in multiple cell types and developmental stages. Embryonic cell lines or stages are marked with an asterisk and a black bar. **b**, Evidence that overlapping H3K9me3 and H3K27me3 ChIP signals in worm are not due to antibody cross-reactivity. ChIP-chip experiments were performed from early embryo (EE) extracts with three different H3K9me3 antibodies (from Abcam, Upstate and H. Kimura) and three different H3K27me3 antibodies (from Active Motif, Upstate and H. Kimura). The H3K9me3 antibodies show similar enrichment profiles (top panel) and high genome-wide correlation coefficients (bottom left). The same is true for H3K27me3 antibodies. There is significant overlap between the H3K9me3 and H3K27me3 ChIP signal, especially on chromosome arms, resulting in relatively high genome-wide correlation coefficients (Extended Data Fig. 1c). The Abcam and Upstate H3K9me3 antibodies showed low level cross-reactivity with H3K27me3 on dot blots<sup>24</sup>, and the Abcam H3K9me3 ChIP signal overlapped with H3K27me3 on chromosome centres. The Kimura monoclonal antibodies against H3K9me3 and H3K27me3 showed the least overlap and smallest genome-wide correlation. In enzyme-linked immunosorbent assay (ELISA) using histone H3 peptides containing different modifications, each Kimura H3K9me3 or H3K27me3 antibody recognized the modified tail against which it

was raised and did not cross-react with the other modified tail<sup>25,26</sup>, providing support for their specificity. Specificity of the Kimura antibodies was further analysed by immunostaining germlines from wild type, *met-2 set-25* mutants (which lack H3K9 histone methyltransferase (HMT) activity<sup>16</sup>), and *mes-2* mutants (which lack H3K27 HMT activity<sup>27</sup>) in the bottom right panel. Staining with anti-HK9me3 was robust in wild type and in *mes-2*, but undetectable in *met-2 set-25*. Staining with anti-HK27me3 was robust in wild type and in *met-2 set-25*, but undetectable in *mes-2*. Finally, we note that the laboratories that analysed H3K9me3 and H3K27me3 in other systems used Abcam H3K9me3 (for human and fly) and Upstate H3K27me3 (for human), and in these cases observed non-overlapping distributions. Another paper also reported non-overlapping distributions of H3K9me3 and H3K27me3 in human fibroblast cells using the Kimura antibodies<sup>26</sup>. The overlapping distributions that we observe in worms using any of those antibodies suggest that H3K9me3 and H3K27me3 occupy overlapping regions in worms. Those overlapping regions may exist in individual cells or in different cell sub-populations in embryo and L3 preparations. **c**, Average gene body profiles of H3K9me3 and H3K27me3 on expressed and silent genes in euchromatin and heterochromatin in human K562 cells, fly L3 and worm L3.

24. Egelhofer, T. A. *et al.* An assessment of histone-modification antibody quality. *Nature Struct. Mol. Biol.* **18**, 91–93 (2011).
25. Hayashi-Takanaka, Y. *et al.* Tracking epigenetic histone modifications in single cells using Fab-based live endogenous modification labeling. *Nucleic Acids Res.* **39**, 6475–6488 (2011).

26. Chandra, T. *et al.* Independence of repressive histone marks and chromatin compaction during senescent heterochromatic layer formation. *Mol. Cell* **47**, 203–214 (2012).
27. Bender, L. B., Cao, R., Zhang, Y. & Strome, S. The MES-2/MES-3/MES-6 complex and regulation of histone H3 methylation in *C. elegans*. *Curr. Biol.* **14**, 1639–1643 (2004).



**Extended Data Figure 3 | Organization of silent domains.** **a**, The correlation of H3K27me3 and H3K9me3 enrichment for human K562 (left panel), fly L3 (second panel), and worm EE chromosome arms (third panel) and centres (right panel) with a 10-kb bin (top) and a 1-kb bin (bottom). The density was calculated as a frequency of bins that fall in the area in the scatter plot (darker grey at a higher frequency).  $r$  indicates Pearson correlation coefficients between binned H3K27me3 fold enrichment (log<sub>2</sub>) and H3K9me3 fold enrichment (log<sub>2</sub>). Worm chromosome arms have a distinctly high correlation between H3K27me3 and H3K9me3. The lower correlation in worm chromosome centres is due to the overall absence of H3K9me3 in these regions. **b**, Schematic diagrams of the distributions of silent domains along the chromosomes in human (H1-hESC), fly (S2) and worm (EE). In human and fly,

the majority of the H3K9me3-enriched domains are located in the pericentric regions (as well as telomeres), while the H3K27me3-enriched domains are distributed along the chromosome arms. H3K27me3-enriched domains are negatively correlated with H3K36me3-enriched domains, although in human, there is some overlap of H3K27me3 and H3K36me3 in bivalent domains. CENP-A resides at the centromere. In contrast, in worm the majority of H3K9me3-enriched domains are located in the arms, whereas H3K27me3-enriched domains are distributed throughout the arms and centres of the chromosomes and are anti-correlated with H3K36me3-enriched domains. In arms and centres, domains that are permissive for CENP-A incorporation generally reside within H3K27me3-enriched domains.

**Extended Data Table 1 | Summary of key shared and organism-specific chromatin features in human, fly and worm**

Chromatin features	Human	Fly	Worm	Figures
<b>Promoters</b>				
H3K4me3 enrichment pattern around TSS	Bimodal peak	Unimodal peak*	Weak bimodal peak	ED1a,b,S12
Well positioned +1 nucleosome at expressed genes	Yes	Yes	Yes	S13
<b>Gene bodies</b>				
Lower H3K36me3 in specifically expressed genes	Yes	Yes	Yes	S21-S23
<b>Enhancers</b>				
High H3K27ac sites are closer to expressed genes	Yes	Yes	Yes	S5-6
Higher nucleosome turnover at high H3K27ac sites	Yes	Yes	ND	S7
<b>Nucleosome positioning</b>				
10-bp periodicity profile	Yes	Yes	Yes	S19a
Positioning signal in genome	Weak	Weak	Less weak	S19b
<b>LADs</b>				
Histone modification in short LADs	H3K27me3	H3K27me3	H3K27me3	S17
Histone modification in long LADs	H3K9me3 internal, H3K27me3 borders	ND	H3K9me3 + H3K27me3	S15
Associated with late replication in S-phase	Yes	Yes	ND	S18
<b>Genome-wide correlation</b>				
Correlation between H3K27me3 and H3K9me3	Low	Low	High (in arms)	ED1c,ED3a
<b>Chromatin state maps</b>				
Similar marks and genomic features at each state	Yes	Yes	Yes	2,S29-32
<b>Silent domains: constitutive heterochromatin</b>				
Composition	H3K9me3	H3K9me3	H3K9me3 + H3K27me3	2,ED3b
Predominant location	Pericentric + chrY	Pericentric + chr4/Y	Arms	3a,ED3b
Depletion of H3K9me3 at TSS of expressed genes	Yes	Yes	Weak	ED2c
<b>Silent domains: Polycomb-associated</b>				
Composition	H3K27me3	H3K27me3	H3K27me3	2
Predominant location	Arms	Arms + chr4	Arms + centers	3a,ED3b
<b>Topological domains</b>				
Active promoters enriched at boundaries	Yes	Yes	ND	S38
Similar chromatin states are enriched in each domain	Yes	Yes	ND	S39

\* Unimodal peak enriched downstream of TSS; ND, no data.