Large organized chromatin K9-modifications (LOCKs) distinguish differentiated from embryonic stem cells

Bo Wen, Hao Wu, Yoichi Shinkai, Rafael A. Irizarry, and Andrew P. Feinberg

Supplementary Methods

Cell Culture

A mouse embryonic stem cell line (ES25) was generated from day 13.5 C57BL/6 embryos by the ES Core Facility (ECF) at Johns Hopkins University according to their standard protocol (http://www.hopkinsmedicine.org/core/ES_Targeting/home.htm). The G9a knockout mouse ES and parental wild type lines were generated in the Shinkai laboratory. ES cells were cultured and maintained according to the standard protocols of the ES core laboratory at Johns Hopkins University (http://www.hopkinsmedicine.org/core/ES_Targeting/Protocol_Pages/protocolmain.htm). To differentiate ES cells, leukemia inhibitory factor (LIF, Chemicon) was withdrawn from the medium and cells were cultured in gelatin-coated plates without feeder layer. Cells were differentiated for 18-24 days in total.

Human lymphoblastoid cell lines GM06991 and GM06993 were obtained from the Coriell Cell Repositories (Camden, New Jersey). Cancer cell lines HCT116, HeLa, KG-1 and Ramos were from the American Type Culture Collection (ATCC, Manassas, VA). All the cells were cultured according to the protocols of the suppliers.
**Tissue specimens**

A male human placenta was obtained freshly from the obstetrics service at Johns Hopkins Hospital. Tissues from the fetal side were dissected into small pieces for further analysis. Liver and brain from two female C57BL/6 mice were cut into small pieces and homogenized using a glass Dounce homogenizer.

**ChIP-on-chip**

We prepared native chromatin as described\(^2\), followed by micrococcal nuclease digestion (MNase, GE Healthcare, Piscataway, NJ) to a size of 1-5 nucleosomes as described\(^2\). Chromatin immunoprecipitation (ChIP) was based on the protocol of Kim et al.\(^3\), except for the use of native chromatin and MNase digestion noted above, and excluding formaldehyde cross-linking and sonication. 200\(\mu g\) of chromatin was used for each ChIP reaction, using commercial monoclonal antibodies specific to H3 dimethyl lysine-9 (Abcam, ab1220), H3 trimethyl lysine-9 (Abcam, ab8898) and H3 trimethyl lysine-27 (Upstate, 07-449). ChIP and input DNA were amplified using the WGA2 kit (Sigma). Labeling of ChIP and input DNA, hybridization and scanning were conducted at NimbleGen in Iceland according to their standard protocols. The designs of tiling arrays and samples analyzed are provided in Supplementary Table 7.

We also validated this protocol by performing H3K27Me3 ChIP in undifferentiated ES cells, and examining the enrichment of ChIP DNA by quantitative real time PCR (qPCR) on six known regions that were found to be H3K27Me3-enriched in ES cells using a conventional ChIP protocol with formaldehyde cross-linking and sonication\(^4\). In two replicate cultures of ES cells, we obtained comparable results for all six genes using our protocol (Supplementary Fig. 1).
**Microarray data analysis**

The raw data consisted of intensities for the Cy3 channel (Input) and Cy5 channel (ChIP). Three steps were used to process the data. First, within-array normalization was performed to remove spatial and sequence-dependent effects using an approach similar to that described by Wu *et al.*. Between-array normalization was performed using a partial quantile normalization algorithm as follows. We first quantile normalized Cy3 (Input) intensities from all arrays to a common target. We used a smoothing procedure to improve the signal to noise ratio of the M (log\(_2\) ratio) values and fitted a function of M values to the chromosomal location using loess. We then picked the genomic regions with the lowest 20% fitted values as the reference region. The Cy5 intensities in the reference regions should have similar distributions in each sample. We therefore quantile normalized them to the same distribution target. The quantile procedure was used to construct a function to map pre-normalization Cy5 intensities to post-normalization Cy5 intensities. This map was then applied to all the Cy5 intensities. This procedure made the M values comparable from array to array. Finally the starts and ends of LOCKs were identified by running a loess smoother on the normalized M values using 15,000 bp windows. Genomic regions with smoothed values less than the 20th quantile of all smoothed values were used as reference regions. This distribution was used as a reference distribution (non-LOCKs). We used 97.5\(^{th}\) quantile of this distribution as the threshold to define LOCKs. Contiguous probes above this threshold were grouped into LOCKs.

To investigate the relationship between LOCK location and gene expression, we compared LOCK locations of liver and brain with the expression data from the gene atlas of mouse and human protein-encoding transcriptomes. We define a gene as present
within a LOCK if it overlapped the LOCK location. Significance of the expression difference between two groups of genes were calculated using Chi-square or Fisher’s exact tests by comparing numbers of high expressed genes (≥2) and low expressed genes (<2). All data analysis was performed in R, and all scripts are available on request.

**Functional annotation analysis**

Gene Ontology (GO) and Tissue Expression annotation analysis was conducted using DAVID Tools. The analyses were conducted using genes within liver-specific or brain-specific LOCKs, comparing to all genes on the microarrays. Terms with P values < 0.01 are provided in Supplementary Tables 4-6.

**ChIP qPCR**

Quantitative real-time PCR (qPCR) was performed using SYBR Green PCR master mix (Applied Biosystem) on an ABI 7700 sequence detector. Primer sequences and locations are provided in Supplementary Table 8. 2ng of ChIP and input DNA were used for each PCR reaction, which was conducted in triplicate for each sample. The enrichment of ChIP over input was calculated by the difference in the number of threshold cycle (Ct) between IP and input, based on a two-fold change for each cycle.

**Quantitative RT-PCR**

Total RNA was prepared using the RNeasy Mini Kit (Qiagen, Valencia, CA), and then 2 µg of RNA was used for reverse transcription using QuantiTect Rev Transcription Kit (Qiagen) according to the manufacturer’s instruction. Quantitative real time PCR was performed as described under ChIP q-PCR. Primer sequences are provided in Supplementary Table 8.
References

1. Tachibana, M. et al. G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev* **16**, 1779-91 (2002).

2. Umlauf, D., Goto, Y. & Feil, R. Site-specific analysis of histone methylation and acetylation. *Methods Mol Biol* **287**, 99-120 (2004).

3. Kim, T.H. et al. A high-resolution map of active promoters in the human genome. *Nature* **436**, 876-80 (2005).

4. Bernstein, B.E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-26 (2006).

5. Wu, J., Irizarry, R., Gentleman, R., Murillo, F.M. & Spencer, F. A model based background adjustment for oligonucleotide expression arrays. *J Amer Statist Assoc* **99**, 909-917 (2004).

6. Bolstad, B.M., Irizarry, R.A., Astrand, M. & Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-93 (2003).

7. Cleveland, W.S., Grosse, E. & Shyu, W.M. Local regression models. in *Statistical Models in S* (eds. Chambers, J.M. & Hastie, T.J.) (Wadsworth and Brooks, Pacific Grove, 1992).

8. Su, A.I. et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* **101**, 6062-7 (2004).

9. Dennis, G., Jr. et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).
Supplementary Table 1. Coordinates of custom arrays

| Chromosome | Start   | End     |
|------------|---------|---------|
| chr1       | 3550000 | 3700000 |
| chr1       | 68000000| 68400000|
| chr2       | 15635000| 16785000|
| chr6       | 144216000| 144465000|
| chr6       | 160190000| 160556473|
| chr7       | 1259000001| 132000000|
| chr7       | 50287019| 51010342|
| chr7       | 92457810| 93003452|
| chr7       | 93601102| 94794489|
| chr7       | 95865658| 96583857|
| chr10      | 64879750| 69842996|
| chr11      | 32037172| 32579528|
| chr11      | 1000000| 3367849|
| chr11      | 6825689| 7014184|
| chr11      | 59567900| 59730000|
| chr13      | 46131964| 47737931|
| chr14      | 99783114| 100981613|
| Chromosome | Start (mm) | End (mm) |
|------------|------------|----------|
| chr15      | 17000001   | 31400000 |
| chr18      | 42649549   | 42969233 |
| chr19      | 61715851   | 62440220 |
| chr20      | 35426873   | 35741706 |
| chr20      | 41508900   | 41643913 |
| chr20      | 56692122   | 57018532 |

**Mouse custom array**

| Chromosome | Start (mm) | End (mm) |
|------------|------------|----------|
| chr6       | 2900000    | 8500000  |
| chr6       | 28500000   | 32500000 |
| chr7       | 4000000    | 8500000  |
| chr7       | 55500000   | 64500000 |
| chr7       | 140500000  | 145500000|
| chr11      | 21500000   | 24000000 |
| chr12      | 107000000  | 112500000|
| chr17      | 11000000   | 14500000 |

\(^a\)Coordinates based on HG17; \(^b\) coordinates based on MM8
| Cell type                | Number of LOCKs | Size of LOCKs (kb) | Percentage of genome in LOCKs |
|-------------------------|-----------------|--------------------|-------------------------------|
| Undifferentiated ES cells | 2601            | 20                 | 1,667                         | 4.3%                      |
|                         |                 | Minimum            | Maximum                       | Mean                       |                          |
| Differentiated ES cells  | 8611            | 20                 | 2,658                         | 93                        | 31.1%                    |
| Adult brain             | 4746<sup>a</sup> | 20                 | 3,483                         | 52                        | 9.8%                     |
| Adult liver             | 4916<sup>a</sup> | 20                 | 4,875                         | 235                       | 45.6%                    |

<sup>a</sup>Doubled to extrapolate from the half-genome measured
Supplementary Table 3. Genes within H3K9Me2 LOCKs show similar relationship to gene repression compared to genes individually marked by H3K9Me2

|                | High expression$^a$ | Low expression | P value |
|----------------|---------------------|----------------|---------|
| **Liver**      |                     |                |         |
| H3K9Me2 LOCKs  | 45                  | 2569           |         |
| Individual H3K9Me2 Marks | 15                | 652            | 0.456   |
| **Brain**      |                     |                |         |
| H3K9Me2 LOCKs  | 4                   | 693            |         |
| Individual H3K9Me2 Marks | 7                | 484            | 0.217   |

$^a$Number of genes with normalized log ratio ≥ 2 (see Full Methods)
## Supplementary Table 4. GO annotation of genes within liver-specific LOCKs

| Ontologies | Term                                                                 | Count | %    | P Value         | Fold Enrichment |
|------------|----------------------------------------------------------------------|-------|------|-----------------|-----------------|
| MF         | GO:0004930–G-protein coupled receptor activity                      | 213   | 19.8%| 2.86E-94        | 4.5             |
| BP         | GO:0007186–G-protein coupled receptor protein signaling             | 227   | 21.1%| 1.92E-89        | 4.0             |
| BP         | GO:0007606–sensory perception of chemical stimulus                   | 141   | 13.1%| 2.28E-89        | 6.5             |
| MF         | GO:0004888–transmembrane receptor activity                          | 245   | 22.8%| 4.61E-89        | 3.7             |
| MF         | GO:0001584–rhodopsin-like receptor activity                         | 190   | 17.7%| 3.24E-86        | 4.6             |
| MF         | GO:0004872–receptor activity                                        | 309   | 28.7%| 2.97E-82        | 2.9             |
| MF         | GO:0004984–olfactory receptor activity                              | 120   | 11.2%| 1.03E-77        | 6.7             |
| BP         | GO:0007608–sensory perception of smell                             | 122   | 11.4%| 1.37E-77        | 6.6             |
| MF         | GO:0004871–signal transducer activity                               | 318   | 29.6%| 2.69E-73        | 2.6             |
| MF         | GO:0060089–molecular transducer activity                            | 318   | 29.6%| 2.69E-73        | 2.6             |
| BP         | GO:0007600–sensory perception                                       | 155   | 14.4%| 1.30E-71        | 4.8             |
| BP         | GO:0050877–neurological system process                              | 185   | 17.2%| 3.53E-71        | 4.0             |
| BP         | GO:0007166–cell surface receptor linked signal transduction         | 269   | 25.0%| 2.00E-69        | 2.8             |
| BP         | GO:0003008–system process                                           | 190   | 17.7%| 5.23E-61        | 3.4             |
| BP         | GO:0007154–cell communication                                       | 347   | 32.3%| 1.85E-51        | 2.0             |
| BP         | GO:0007165–signal transduction                                      | 320   | 29.8%| 2.86E-45        | 2.0             |
| CC         | GO:0016021–integral to membrane                                     | 415   | 38.6%| 1.98E-38        | 1.7             |
| BP         | GO:0050896–response to stimulus                                     | 247   | 23.0%| 5.32E-38        | 2.1             |
| CC         | GO:0031224–intrinsic to membrane                                    | 415   | 38.6%| 5.50E-38        | 1.7             |
| CC         | GO:004425–membrane part                                            | 437   | 40.7%| 6.84E-34        | 1.6             |
| CC         | GO:0016020–membrane                                                | 481   | 44.7%| 1.62E-25        | 1.4             |
| MF         | GO:0016503–pheromone receptor activity                              | 33    | 3.1% | 1.59E-22        | 7.2             |
| BP         | GO:0032501–multicellular organismal process                         | 271   | 25.2%| 1.07E-20        | 1.6             |
| CC         | GO:0005576–extracellular region                                     | 203   | 18.9%| 4.53E-12        | 1.6             |
| CC         | GO:0005886–plasma membrane                                          | 152   | 14.1%| 6.72E-10        | 1.6             |
| BP         | GO:0007268–synaptic transmission                                    | 35    | 3.3% | 1.10E-08        | 2.9             |
| CC         | GO:0005615–extracellular space                                      | 172   | 16.0%| 1.42E-08        | 1.5             |
| CC         | GO:0044456–synapse part                                            | 25    | 2.3% | 1.89E-08        | 3.6             |
| CC         | GO:0044421–extracellular region part                               | 179   | 16.7%| 1.93E-08        | 1.5             |
| MF         | GO:0005549–odorant binding                                          | 14    | 1.3% | 2.22E-08        | 6.3             |
| MF         | GO:0005529–sugar binding                                            | 39    | 3.6% | 2.60E-08        | 2.6             |
| MF         | GO:0030594–neurotransmitter receptor activity                       | 23    | 2.1% | 9.12E-08        | 3.6             |
| BP         | GO:0019226–transmission of nerve impulse                            | 37    | 3.4% | 1.04E-07        | 2.6             |
| BP         | GO:0019236–response to pheromone                                    | 14    | 1.3% | 1.13E-07        | 5.7             |
| MF         | GO:0005550–pheromone binding                                        | 13    | 1.2% | 1.32E-07        | 6.1             |
| MF         | GO:0042165–neurotransmitter binding                                 | 23    | 2.1% | 1.72E-07        | 3.5             |
| BP         | GO:0050909–sensory perception of taste                             | 13    | 1.2% | 1.77E-07        | 6.0             |
| MF         | GO:0005179–hormone activity                                         | 26    | 2.4% | 2.54E-07        | 3.1             |
| MF         | GO:0030246–carbohydrate binding                                    | 45    | 4.2% | 4.29E-07        | 2.2             |
| MF         | GO:0022834–ligand-gated channel activity                            | 20    | 1.9% | 4.32E-07        | 3.7             |
| MF         | GO:0015276–ligand-gated ion channel activity                        | 20    | 1.9% | 4.32E-07        | 3.7             |
| CC         | GO:0045211–postsynaptic membrane                                   | 22    | 2.1% | 4.37E-07        | 3.4             |
| MF         | GO:0016917–GABA receptor activity                                  | 13    | 1.2% | 5.02E-07        | 5.6             |
| MF         | GO:0004890–GABA-A receptor activity                                | 13    | 1.2% | 5.02E-07        | 5.6             |
| GO ID          | GO Term                                      | Count | FDR   | p-Value    | Log10 p-Value |
|---------------|----------------------------------------------|-------|-------|------------|---------------|
| GO:0005230    | extracellular ligand-gated ion channel activity | 17    | 1.6%  | 6.34E-07   | 2.8           |
| GO:0005887    | integral to plasma membrane                  | 66    | 6.1%  | 1.11E-06   | 2.1           |
| GO:0044459    | plasma membrane part                          | 110   | 10.2% | 1.94E-06   | 1.9           |
| GO:0043168    | anion binding                                | 15    | 1.4%  | 2.13E-06   | 1.4           |
| GO:0031404    | chloride ion binding                          | 15    | 1.4%  | 2.13E-06   | 1.4           |
| GO:0031226    | intrinsic to plasma membrane                  | 66    | 6.1%  | 2.52E-06   | 1.8           |
| GO:0045202    | synapse                                      | 33    | 3.1%  | 3.19E-06   | 2.4           |
| GO:0005254    | chloride channel activity                     | 14    | 1.3%  | 3.87E-06   | 4.5           |
| GO:0005253    | anion channel activity                         | 14    | 1.3%  | 1.88E-05   | 3.9           |
| GO:0009952    | defense response                              | 45    | 4.2%  | 2.47E-05   | 1.9           |
| GO:0007214    | gamma-aminobutyric acid signaling pathway     | 9     | 0.8%  | 2.74E-05   | 6.1           |
| GO:0022836    | gated channel activity                         | 31    | 2.9%  | 3.77E-05   | 2.2           |
| GO:0007267    | cell-cell signaling                           | 40    | 3.7%  | 5.32E-05   | 1.9           |
| GO:0050913    | sensory perception of bitter taste            | 7     | 0.7%  | 5.52E-05   | 8.0           |
| GO:0008066    | glutamate receptor activity                   | 12    | 1.1%  | 6.65E-05   | 4.1           |
| GO:0009986    | cell surface                                  | 26    | 2.4%  | 8.78E-05   | 2.3           |
| GO:0006821    | chloride transport                            | 12    | 1.1%  | 9.68E-05   | 4.0           |
| GO:007269     | neurotransmitter secretion                    | 12    | 1.1%  | 1.84E-04   | 3.7           |
| GO:0005216    | ion channel activity                           | 35    | 3.3%  | 2.30E-04   | 1.9           |
| GO:001505     | regulation of neurotransmitter levels          | 15    | 1.4%  | 3.28E-04   | 2.9           |
| GO:0031644    | regulation of neurological process             | 10    | 0.9%  | 3.72E-04   | 4.1           |
| GO:0001580    | detection of chemical stimulus                | 6     | 0.6%  | 3.82E-04   | 7.7           |
| GO:0050912    | detection of chemical stimulus                | 6     | 0.6%  | 3.82E-04   | 7.7           |
| GO:0022838    | substrate specific channel activity            | 35    | 3.3%  | 3.92E-04   | 1.9           |
| GO:0042742    | defense response to bacterium                 | 14    | 1.3%  | 4.00E-04   | 3.0           |
| GO:009897     | external side of plasma membrane              | 18    | 1.7%  | 4.79E-04   | 2.5           |
| GO:0004175    | endopeptidase activity                        | 47    | 4.4%  | 5.84E-04   | 1.7           |
| GO:0008067    | metabotropic glutamate, GABA-B-like receptor   | 9     | 0.8%  | 6.51E-04   | 4.2           |
| GO:008233     | peptidase activity                            | 63    | 5.9%  | 0.0102     | 1.5           |
| GO:0051704    | multi-organism process                        | 25    | 2.3%  | 0.00124    | 2.0           |
| GO:0008037    | cell recognition                              | 10    | 0.9%  | 0.00128    | 3.5           |
| GO:005907     | detection of chemical stimulus                | 6     | 0.6%  | 0.00145    | 6.1           |
| GO:007185     | transmembrane receptor protein signaling       | 6     | 0.6%  | 0.00145    | 6.1           |
| GO:008509     | anion transmembrane transporter activity       | 16    | 1.5%  | 0.00146    | 2.5           |
| GO:0022803    | passive transmembrane transporter activity     | 36    | 3.4%  | 0.00169    | 1.7           |
| GO:0015267    | channel activity                              | 36    | 3.4%  | 0.00169    | 1.7           |
| GO:0051606    | detection of stimulus                         | 13    | 1.2%  | 0.00175    | 2.8           |
| GO:0045055    | regulated secretory pathway                   | 12    | 1.1%  | 0.00180    | 2.9           |
| GO:0050906    | detection of stimulus during sensory perception | 10  | 0.9%  | 0.00215    | 3.3           |
| GO:0009617    | response to bacterium                         | 15    | 1.4%  | 0.00215    | 2.5           |
| GO:0051057    | positive regulation of small GTPase signaling  | 5     | 0.5%  | 0.00248    | 7.3           |
| GO:0009593    | detection of chemical stimulus                | 7     | 0.7%  | 0.00289    | 4.5           |
| GO:042734     | presynaptic membrane                          | 5     | 0.5%  | 0.00303    | 6.9           |
| GO:0019835    | cytolysis                                     | 7     | 0.7%  | 0.00410    | 4.2           |
| GO:0007565    | female pregnancy                              | 8     | 0.7%  | 0.00499    | 3.6           |
| GO:0019233    | sensory perception of pain                    | 7     | 0.7%  | 0.00565    | 4.0           |
| GO:0042923    | neuropeptide binding                          | 9     | 0.8%  | 0.00661    | 3.1           |
| GO:0008188    | neuropeptide receptor activity                | 9     | 0.8%  | 0.00661    | 3.1           |
| GO:0030054    | cell junction                                 | 33    | 3.1%  | 0.00803    | 1.6           |
| GO:0042221    | response to chemical stimulus                 | 39    | 3.6%  | 0.00848    | 1.5           |
| BP        | GO:0007610~behavior       | 32  | 3.0%  | 0.00879 | 1.6 |
|-----------|---------------------------|-----|-------|---------|-----|
| BP        | GO:0015698~inorganic anion transport | 16  | 1.5%  | 0.00967 | 2.0 |
| BP        | GO:0051969~regulation of transmission of nerve impulse | 7   | 0.7%  | 0.00995 | 3.6 |

*BP: biological processes; MF: molecular function; CC: cellular components.*
| Ontologies<sup>a</sup> | Term                                                      | Count | %     | P Value      | Fold Enrichment |
|----------------------|-----------------------------------------------------------|-------|-------|--------------|-----------------|
| MF                   | GO:0004091~carboxylesterase activity                       | 11    | 15.3% | 4.94E-11     | 21.3            |
| MF                   | GO:0016712~oxidoreductase activity                         | 7     | 9.7%  | 6.24E-09     | 44.1            |
| MF                   | GO:0053081~unspecific monoxygenase activity                | 6     | 8.3%  | 6.92E-08     | 50.4            |
| CC                   | GO:0005792~microsome                                       | 8     | 11.1% | 3.56E-07     | 16.7            |
| MF                   | GO:0004867~serine-type endopeptidase inhibitor             | 8     | 11.1% | 3.87E-07     | 16.5            |
| CC                   | GO:0005615~extracellular space                             | 24    | 33.3% | 3.92E-07     | 3.0             |
| CC                   | GO:0044421~extracellular region part                      | 24    | 33.3% | 1.06E-06     | 2.9             |
| MF                   | GO:0030414~protease inhibitor activity                     | 8     | 11.1% | 1.77E-06     | 13.3            |
| MF                   | GO:0004866~endopeptidase inhibitor activity                | 8     | 11.1% | 1.77E-06     | 13.3            |
| MF                   | GO:0004497~monoxygenase activity                           | 7     | 9.7%  | 2.40E-06     | 17.3            |
| CC                   | GO:0005576~extracellular region                            | 14    | 19.4% | 9.40E-07     | 5.3             |
| MF                   | GO:0016705~oxidoreductase activity                         | 7     | 9.7%  | 3.45E-06     | 2.7             |
| CC                   | GO:0000267~cell fraction                                  | 11    | 15.3% | 1.16E-05     | 5.8             |
| MF                   | GO:0016788~hydrolase activity, acting on ester bonds       | 13    | 18.1% | 1.79E-05     | 4.5             |
| CC                   | GO:0005624~membrane fraction                              | 10    | 13.9% | 2.23E-05     | 6.2             |
| MF                   | GO:0004857~enzyme inhibitor activity                       | 8     | 11.1% | 2.80E-05     | 8.8             |
| MF                   | GO:0020037~heme binding                                   | 7     | 9.7%  | 4.45E-05     | 10.5            |
| MF                   | GO:0046906~tetrapyrrole binding                            | 7     | 9.7%  | 4.45E-05     | 10.5            |
| BP                   | GO:0006091~generation of precursor metabolites             | 9     | 12.5% | 2.46E-04     | 5.1             |
| BP                   | GO:0006118~electron transport                             | 8     | 11.1% | 3.17E-04     | 5.8             |
| MF                   | GO:0047760~butyrate-CoA ligase activity                    | 3     | 4.2%  | 5.90E-04     | 75.6            |
| MF                   | GO:0005506~iron ion binding                               | 7     | 9.7%  | 0.00156      | 5.4             |
| CC                   | GO:0044444~cytoplasmic part                               | 23    | 31.9% | 0.00210      | 1.8             |
| MF                   | GO:0016878~acid-thiol ligase activity                      | 3     | 4.2%  | 0.00259      | 37.8            |
| MF                   | GO:0016877~ligase activity, carbon-sulfur bonds           | 3     | 4.2%  | 0.00442      | 29.1            |
| MF                   | GO:0030234~enzyme regulator activity                      | 9     | 12.5% | 0.00471      | 3.3             |
| MF                   | GO:0003824~catalytic activity                             | 33    | 45.8% | 0.00952      | 1.4             |

<sup>a</sup>BP: biological processes; MF: molecular function; CC: cellular components.
## Supplementary Table 6. Tissue-specific expression of genes within LOCKs

| Expression Tissue | Number of genes | %   | P Value  |
|-------------------|-----------------|-----|---------|
| **Genes in liver-specific LOCKs** |                 |     |         |
| Cerebellum        | 96              | 8.9%| < 10⁻³ |
| Hypothalamus      | 33              | 3.1%| < 10⁻³ |
| Brain cortex      | 32              | 3.0%| 0.00284|
| Epididymis        | 14              | 1.3%| 0.00285|
| Diencephalon      | 27              | 2.5%| 0.00603|
| **Genes in brain-specific LOCKs** |                 |     |         |
| Liver             | 31              | 43.1%| < 10⁻⁸|
| Plasma            | 4               | 5.6%| 0.00387|
### Supplementary Table 7. Arrays and samples analyzed

| Array          | Design | Feature | Samples analyzed                      |
|----------------|--------|---------|---------------------------------------|
| Human ENCODE   | HG17   | 385K    | Placenta, GM06993, Hela, HCT116       |
| Mouse ENCODE   | MM7    | 385K    | Liver, differentiated WT and G9a knockout ES |
| Placenta       | Custom | 385K    | Placenta                              |
| Mouse imprinted| Custom | 385K    | Liver                                 |
| Mouse Whole genome arrays | Economy Tiling Set HX1 | 2.1M | ES25, diff.ES, Liver\textsuperscript{b} and Brain\textsuperscript{b} |

\textsuperscript{a}All arrays were manufactured at NimbleGen; \textsuperscript{b}Half of the genome was investigated.
### Supplementary Table 8. Primer sequences

| Locus | Upper primer | Lower primer |
|-------|--------------|--------------|
| **qPCR primers for human ChIP** | | |
| R1    | TTTATGAAGTCAACCCACGAC | GGGGTATCATATAATCTGACCTG |
| R2    | CGAGTGTGATAATTGGGGCTAGC | TCCACTCCGTACCTGCTTTACT |
| R3    | TAACCCCTTGTTTCCAGGTATGG | AAGCTGCTGATGAGAAGAAAACC |
| R4    | CACAGCATAATGTCTTTCGATT | TGGCAACTTTGCTATGGTGTC |
| R5    | AGGCCGACCAATTTCTAAAAA | GCCATTCATCCACTGACACTCA |
| R6    | CCTATGAATTTCACGTAGTC   | CCTGAGATGGGGCAGTATAGTC |
| R7    | CAATGGACCAAGACATTGATA | ATAGGGTATGAAACCCCCGAGT |
| R8    | TACATCGGTTGATTGGCTAGTC | ACCCCCTAAATACCGATCCTT |
| R9    | TGAGTCACCAAGGAAAGTTTTT | ACATTCAAAGAGGCAAGGACATT |
| R10   | TCCCTATGTACTGCTTTCCTC | TACCTGGGAGGTGATAGGAAAA |
| R11   | CTTTACCTTTGCTCCCTAGATT | AATTGCAAGGCCACTTTAAGTCA |
| R12   | TTTGCTTCTCAGGAAGCTCATAAA | TCCGGAAAGCATGACATATAAC |
| R13   | CATCATTACTGCTTCTTCTCCCT | CACTCAATACCTGAACACCAA |
| R14   | ACTTTTAGGATCTGGCCACCTTGGT | CTAATCGCTGCTCTCTGTTTT |
| R15   | AGTGGGAGGCACAGAGGAAGA | TGTGACCCCTAGAAACCTG |
| R16   | TCTGAATTTGTCGCAACAGC | GCAAATAGCTGCTCCCTTGG |
| R17   | GTGGTCGACAGAAGGCTCCTTC | GCATCCCGTCTTCTGAGAC |
| R18   | GGCTTAGGGTGGGAAATATCGC | CTCACCCCTAGTGCCTACCT |
| R19   | ATAGTGAAGGACGGCGAGGTTG | CGCAGGCTTCTTCTACCTTG |
| **qPCR primers for mouse ChIP** | | |
| Actb  | GAACCCCAACACACCTAGCA | GCCTGGAATTGAAATGGACAGA |
| Gapdh | CACTTCTCCATTTCCCTGTG | GGTCCAGGATAGGACACTCA |
| Wnt2  | GGTTGACAGGAGGCTCAGAG | TCCCAGGCTATTCTCTCCTT |
| Gm8   | CTCCCTGACTGAGGATACAGCTC | TCTCTCTTCTCCACCCAGCA |
| Hnt_A | TAACACAGGGAAAAACGTGT | TCAACACAGCAGAGGTTGGA |
| Hnt_B | CCTACATTTTGATGCCAGAGC | TACCGACACAGTGGAGG |
| BpiL2  | TAGAGCCAGGCGAGTATCGC | TTTGCTGACTACAGGACTT |
| Lrrk2 | GTTGGGCTTCTTGGGCTGTG | GACCAGACCTCTTCCCTCAAT |
| **qRT-PCR primers for mouse** | | |
| Actb  | GGTCACTCCTATTGGCAACG | ACGGATGTCACACGTCACACT |
| Ncam2 | TGCCCAAGGCAGCACAAG | ATCCCTCCCTTGGTTAAGAAT |
| Nefl  | GGACAAACGCGAAGTGACAGA | AGGCCATCTTGACATTGAGG |
| ANP   | GGGGTTAGGATTGACAGGAT | ACACACACAGGCTTTAGG |
| Myog  | GTGCCAGTGAATGCAACCTCA | AGATTTGTTGGCTGCTTGAAG |
| Afp   | CTCAGCGAGGAGAAAAATGGTC | GAGTTCAAGGCTTTGCTTC |
| Pdx1  | GAAATCCACCAAGGCTCACG | ACGGCTCTCTTGGTTTCTT |
| Hnt   | AGTCCAGCAGGTACGAGTGC | CCTGTCCTCCCTAGCCTGAG |
| Gene  | Primer 1                  | Primer 2                  |
|-------|--------------------------|--------------------------|
| Lrrk2 | TCCCCACCAATGAAAAACATC    | TGCACCTCGTTCAACACCAG     |
| AK016497 | CATCACCAGACACCTACTGG    | AAGAAGAGGACGCAGGTTA     |
| Bpi12 | GAGAACAGCCAACGAGATGC    | AGGGGTGGAAGAGGAAAT      |
| Cdh2  | GACAAAGATCAGCCCCCACAC   | AATGGCAGTTGTTCTGCG     |
| Grm8  | CTCGCGCAGTGATTATGTTT    | GAAAATGCCCACCTGTTT     |
| Spp2  | CTGAAGACGCTGGCTTTTGT    | GCCCGAAACAGGTAAGGACT   |
| Trpm8 | CGGGGACATTCTAGTTTGAGA  | GCTGGGTCAACAGTCCAAGAG   |
| Ubqln3 | TCACAGTCCACCTGTCATC     | AAGAAGGAGACCCATCCACA   |
| Wnt2  | AGCTGGAAGGAAGGCTGTAA    | GTCGCTGTTTCTGAAAGT     |
Supplementary Figure 1

Supplementary Figure 1 | Proof of principle for our ChIP protocol. ChIP experiments were performed on two independent cultures of mouse ES cells using antibody to H3K27Me3, a well-studied heterochromatin marker. We then tested the enrichment using real time quantitative PCR on two housekeeping gene (negative control) and six known loci marked by H3K27Me3 in ES cells.
Supplementary Figure 2 | Additional examples of H3K9Me2 LOCKs in human placenta. **a**, In the 5 Mb region examined in chromosome 10, a 2.8 Mb LOCK covers the 3’ end and extends downstream of the imprinted gene *CTAA3*. **b**, A 800 kb LOCK is observed in the intergenic region between *HTR2A* and *SUCLA2* of chromosome 13. In both cases, CTCF binding sites are distributed at the boundaries of the LOCKs.
Supplementary Figure 3

**Supplementary Figure 3 | Validation of H3K9Me2 LOCK array data by quantitative real-time PCR (qPCR).**

**a.** ChIP-on-chip data on human placenta (red) and location of regions (blue bars) selected for ChIP qPCR validation. R1, R2, R12, R13, and R15 are in non-LOCK regions, and the other 14 sites are within LOCKs.

**b.** qPCR results. The Y axis shows the fold enrichment of ChIP over input by comparing the difference of threshold cycles (C_t), with a two-fold change equal to one cycle. The results of ChIP-on-chip were confirmed by qPCR in all 19 cases examined.

**c.** correlation of qPCR results with microarray intensities. The black and red dots denote sites on chromosomes 15 and 19, for which we qPCR was performed on separate plates. The correlations are 0.77 and 0.9 for chromosomes 15 and 19, respectively, with an overall correlation of 0.71.
Supplementary Figure 4

**Supplementary Figure 4** | **Locations of LOCKs are conserved between human and mouse.** ChIP-on-chip data on human placenta and mouse liver. **a,** human ENCODE region ENr321 (top) and syntenic mouse region (bottom). **b,** human ENCODE region ENr121 (top) and syntenic mouse region (bottom). The X-axis is drawn according to the genome annotation but in (b) the human and mouse sequences are inverted. The LOCKs are in corresponding location to the genome sequence. Other annotation information is the same as in Fig. 2 in the main text.
Supplementary Figure 5

Supplementary Figure 5 | Expression of lineage-specific markers in undifferentiated ES25 and day-24 differentiated cells. We differentiated ES cells randomly as described1. In order to provide better documentation of the nature of these cells, we examined six lineage-specific markers (ectoderm: Ncam2 and Nefl; mesoderm: Anp and Myog; endoderm: Afp and Pdx1)1-3 in ES25 and day 24 differentiated cells. These data show that random differentiation of these ES cells led to overexpression of mesodermal and endodermal markers, with a ~100-fold change in Afp, a marker of pre-liver lineages, but not ectodermal markers.
Supplementary Figure 6

Supplementary Figure 6| LOCK formation is G9a-dependent. Two additional examples comparing wild type differentiated ES cells (black), to differentiated G9a knockout ES cells (red). Most LOCKs are absent in the G9a knockout cells but some persist in the absence of G9a.
Supplementary Figure 7

Supplementary Figure 7 | Expression of lineage markers in WT and *G9a*/*- ES and day-18 differentiated cells. Legend as in Supplementary Figure 5.
Loss of LOCK in G9a-/ ES was related to differential gene expression. Gene expression level was examined using quantitative real-time RT-PCR of 10 genes within LOCKs lost in G9a-/ ES, in day 18-differentiated WT and G9a-/ ES cells. Eight of 10 genes were over-expressed in G9a-/ compared to WT cells.
Supplementary Figure 9

H3K9Me2 LOCKs were independent of H3K27Me3. H3K27Me3 ChIP was performed on day 18-differentiated WT and G9a-/- ES cells, followed by quantitative real-time RT-PCR. Actb and Pax5 were negative and positive controls, respectively, for the enrichment of H3K27Me3. No substantial enrichment of H3K27Me3 was found at any of 6 loci located within H3K9Me2 LOCKs (Wnt2, Grm8, Hnt at 2 sites 125 kb apart, Bpil2, and Lrrk2).
Supplementary Figure 10 | Genes within LOCKs are generally CpG-poor in their promoters. Histogram of gene promoter CpG density measured by observed to expected ratio of CpGs.
**Supplementary Figure 11**

**Supplementary Figure 11| Loss of H3K9Me2 LOCKs in human cancer cell lines.**
ChIP qPCR on four cancer cell lines (KG-1 leukemia, Ramos lymphoma, HCT116 colon cancer, and HeLa cervical cancer) and two lymphoblastoid cell lines (GM06991 and GM06993), measured at three sites in LOCKs in ENCODE region ENr313 (R20, 21 and 22, numbered consecutively from the validation sites described earlier). In all the cases, the H3K9Me2 ChIP DNA is less enriched in cancer cells than in the non-cancer cells, validating loss of the LOCKs in cancer.

**References**

1. Ward, C.M., Barrow, K., Woods, A.M. & Stern, P.L. The 5T4 oncofoetal antigen is an early differentiation marker of mouse ES cells and its absence is a useful means to assess pluripotency. *J Cell Sci* **116**, 4533-42 (2003).

2. Czyz, J. & Wobus, A. Embryonic stem cell differentiation: the role of extracellular factors. *Differentiation* **68**, 167-74 (2001).

3. Keller, G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev* **19**, 1129-55 (2005).