Supplemental Materials and Methods

Promoter microarray hybridization and data analyses

The MM8 RefSeq promoter microarrays (Roche NimbleGen) represent 22,036 genomic regions spanning on average 2,500 bp of which 500 bp are located downstream from the transcription start site (TSS) and 2,000 are upstream. Each promoter is covered by a set of 25 probes (on average) with a spacing of one 50-70 bp oligo per 100 bp. Sample labelling, hybridization to promoter microarrays and array scanning were performed by Roche NimbleGen Systems Inc. according to standard procedures.

For the analysis of DNA methylation the raw fluorescent intensity values were loess normalized using LIMMA package in R (Smyth 2005) and the log2 values MAP/input were calculated for each individual probe. Subsequently, for comparison between microarray experiments the probe values were scaled to have the same median absolute deviation. The log2 values of all probes located within a 1000bp window around the TSS (+500 to -500) were aggregated into a single log2 value per promoter (M promoter) by calculating the median. The log2 values of probes located -500 to -1500 from TSS were processed in the same way to obtain a single log2 value per upstream regions (M upstream). We discarded as non-informative all promoters and upstream regions which had log2 values for less that 5 probes within the 1000 bp region. The final log2 values ($M_{av)}$ for promoter and upstream regions represent the average from two hybridization experiments representing two independent biological replicas (two DNA purification preps and separate MAP purifications). The Speraman correlation coefficients between replicate experiments were ρ=0.89 for wild type MEFs and ρ=0.9 for *Hells*$^{-/-}$ MEFs.

The total number of methylated promoters in wild type and *Hells*$^{-/-}$ MEFs was determined by first grouping promoters into three separate categories: low CpG density

promoters (LCP), intermediate CpG density (ICP) and high CpG density (HCP) as previously described (Mohn et al. 2008) to account for differences in efficiency of enrichment on the MBD affinity column. A category-specific arbitrary cut offs of log2>0.8, log2>0.9 and log2>1 were used to calculate the number of methylated LCP, ICP and HCP respectively (supplemental Figure 1).

To identify differentially methylated regions between wild type and *Hells*$^{-/-}$ MEFs we calculated a delta methylation value ($\Delta M = M_{av}WT - M_{av}$ *Hells*$^{-/-}$) for each promoter and upstream region. We considered as significant only ΔM values that were $\geq$ +/-1 corresponding to a 2-fold difference. Promoter/upstream regions showing negative log2 values in all experiments were excluded from these analyses.

Definition of LSH-dependent domains

We designated as "clusters" or LSH-dependent domains sequential genomic regions, including promoters and upstream sequences, which showed $\Delta M = M_{av}WT - M_{av}$ *Hells*$^{-/-}$ $\geq$ 0.5. In cases when two clusters were separated by no more than 1 unmethylated region, they were considered as a single domain. LSH-dependent domains are shown in grey in the last column in the Supplemental Table 3.

Expression microarrays data analyses

The MM8 exprX4 microarrays (Roche NimbleGen) contain three 60bp probes for each of 25,631 RefSeq transcripts, representing 18,869 genes. Sample labelling, hybridization to promoter microarrays and array scanning were performed by Roche NimbleGen Systems Inc. according to standard procedures. The experiment included three replicates for wild type and five for *Hells*$^{-/-}$ MEFs.

Raw intensity values were quantile normalised using the BioConductor package LIMMA (Smyth 2005). The log2 values of the probes associated with each transcript were summarised into a single log2 value using Tukey's median polish procedure (Tukey 1977). A linear model was fit to the data with LIMMA, calculating the expression ratio M=log2 (WT/*Hells*$^{-/-}$) and moderated t-statistics, adjusting p-values for multiple testing. The false discovery rate (FDR) was obtained using the Benjamini-Hochberg method (Benjamini and Hochberg 1995)

Benjamini, Y. and Hochberg, Y. 1995. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**(1): 289-300.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. 2008. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* **30**(6): 755-766.

Smyth, G.K. 2005. Limma: linear models for microarray data. in *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (ed. R. Gentleman, Carey, V., Dudoit, S., Irizarry, R., Huber, W.), pp. 397--420. Springer, New York.

Tukey, J.W. 1977. *Exploratory data analysis*. Addison-Wesley, Reading, Mass.

Primer sequences

Primers for Methylation sensitive PCR

Ndufa1-Meth-F	GGAGAATCTCGAACCACATC
Ndufa1-Meth-R	GGACGACAGAAAGAATCCTG
Gm9-Meth-F	TGGGAGAGAATGGGCCAGAA
Gm9-Meth-R	ACTGCGCAGATTGCAAAGGG
Rhox2-Meth-F	TCCTCAGTGAACTGCCATCTC
Rhox2-Meth-R	TCGCTCCATGTCTGTAAAGCC
Rhox5-Meth-F	CATGGCATAAGGGGATCCTG
Rhox5-Meth-R	GAACCCTCAGCTTCCATACC

Rhox6&9-Meth-F	CTTTGGAGCTCTGAACTGCT
Rhox6&9-Meth-R	GCCTGTCCACTCTTCTTTCC
Rhox11-Meth-F	TCAGTGTACCCAAGATGCCA
Rhox11-Meth-R	ATATGCGCACAGGATTTTCC
Rhox13-Meth-F	ACTCCTCCGACGAGAGTGATG
Rhox13-Meth-R	GCCCCTCAACATGAGCCAAAG
Cul4b-Meth-F	TACTCTGGCCTCGAACTCAAC
Cul4b-Meth-R	GACCCACTTACAAGGCCCTAA
Mcts1-Meth-F	CGACAAAAGCGTCAACGTGGAGT
Mcts1-Meth-R	CTGGAAAGGCAGAACAGGGCTAA
Dppa3-Meth-F	AGAGCGGGGAATCCTACAGT
Dppa3-Meth-R	CAGTGCTGGGCTCAGTGATA
Elf5-Meth-F	TTGCCTTGAAAGCCTTCTGT
Elf5-Meth-R	TTCTTGAGCTGGCTATCCGT
Slc2a3-Meth-F	GGCCAAGTCCTACCCTCTTC
Slc2a3-Meth-R	TAGGACCCGAGGAACACTTG

RT-PCR primers

GM9-RT-F	GGCAAACACAAGGGCTATGG
GM9-RT-R	ACCGGAGGTAAAGGTCGTTG
Rhox2-RT-F	CAAGCTTGATGTGGGACCTGAG
Rhox2-RT-R	GAATCAGGGACTGGCTCCTCTG
Rhox6-RT-F	AAACCCACATGTTCTGAATAGG
Rhox6-RT-R	GCCATCATGGGTGGCTCAGAAG
Rhox9-RT-F	CAGCCATTCCTGAGGGCATGA
Rhox9-RT-R	GCCTGCTGTGTCTGCGGAAA
Rhox11-RT-F	TGGAGGAGGACGAGTGTCTGTG
Rhox11-RT-R	TGGTGGCCCTCTTTCGGGATG
Rhox13-RT-F	AGCGAGGACGAGAATGAACC
Rhox13-RT-R	ATGCTCAGGAGCTCGGATAG
Cul4b-RT-F	GTGCAGGCAACAAAGAAGCTAC
Cul4b-RT-R	GCACATATGTTGGCCAGTAACC
Mkrn3-RT-F	AAGTGTGTGGCATCTGCATG
Mkrn3-RT-R	AGAATTCTCCAAATGGGCAGTG
Ndn-RT-F	GCTACAAGAAATGGTGCAGAAG
Ndn-RT-R	CCTCGGTGATTATCTTCCTCAC
Peg12-RT-F	TAACCGAATTGCCTTGCTGAG
Peg12-RT-R	CACGGTTGTGTGAGACTGTG

Ube3a-RT-F	ACCAGGAGAATCCCAGTCTGAG
Ube3a-RT-R	GACTTTCCGAAAGCTCAGAACC
HoxA5-RT-F	AGCCACAAATCAAGCACACA
HoxA5-RT-R	AGATCCATGCCATTGTAGCC
HoxA6-RT-F	TACGGGGCCTCATGTTTCTA
HoxA6-RT-F	CTGCATCCAGGGGTAAACAG
HoxA7-RT-F	CACATGGTTCCAGTTTGTGG
HoxA7-RT-R	CTGAACCCCAGATCCACTGT
HoxA10-RT-F	GGAAGCATGGACATTCAGGT
HoxA10-RT-R	CCAGGCAAGCAAGACCTTAG
HoxA11-RT-F	CCCCTGGTGGTTCACTCTTA
HoxA11-RT-F	CTTGGGGCACACAGTTTCTT
Actin-RT-F	GGCCCAGAGCAAGAGAGGTAT
Actin-RT-R	ACGCACGATTTCCCTCTCAGC
Gapdh-QRT-F	GGCTCATGACCACAGTCCATGCC
Gapdh-QRT-R	CACGGAAGGCCATGCCAGTGAG
Oct4-QRT-F	CCCACTTCACCACACTCTAC
Oct4-QRT-R	CCAAAGCTCCAGGTTCTCTT
Dppa2-QRT-F	GCATTCATTCAGCGGCTGCCTTT
Dppa2-QRT-R	TGCGTAGCGTAGTCTGTGTTTGG
Dppa3-QRT-F	GAACCGCATTGCAGCCGTACCT
Dppa3-QRT-R	GGCTCACTGTCCCGTTCAAACTCA
Dppa4-QRT-F	CAAGGGCTTTCCCAGAACAAATGC
Dppa4-QRT-R	GCAGGTATCTGCTCCTCTGGCAC
Gdf3-QRT-F2	CCTACTATAACCTGCGACCAGAG
Gdf3-QRT-R2	GTTGCTGCTCCAATCCTTAAGC

Bisulfite sequencing primers

Oct4-BS-F	TTTGAAGGTTGAAAATGAAGGTTTT
Oct4-BS-R	CATCACCCCCACCTAATAAAAATA
Rhox2-BS-F	TTTTTAGTGAATTGTTATTTTTTTT
Rhox2-BS-Fn	TTTTTTATTGTTTTGAGTTTTGTAG
Rhox2-BS-R	ATCCCACATCAAACTTATAATTAAC
Rhox6/9-BS1-F	AGAGGAAGTATAGGTAGGGGTTTTG
Rhox6/9-BS1-R	AAAATACAAATATCCCCCAAACAAC
Rhox6/9-BS1-Rn	CAAACAACACCTTAAAACCAAAAAT
Rhox6/9-BS2-F	GTTTGAGGAGGTTTTTGTGTTATTT
Rhox6/9-BS2-R	AAATCCAATCAAAACCAATTTACTC
Rhox6/9-BS2-Fn	TTATTTTTTGTGAAGGAATTTGAAT

Rhox6-2.7kb-BS-F	TATAGAGATAAAGATAGAGAGAAGAGAGG
Rhox6-2.7kb-BS-R	AAAAAACTTTTCCCAAATTTAAACC
Rhox6-2.7kb-BS-Rn	TAATCAACCTTATTAAAAACTCTTAATATT
Rhox6+1.5kb-BS-F	ATTTGGTTTGTGGAAGAAATTTATT
Rhox6+1.5kb-BS-Fn	TGGGTTTAAATATATAAAATTAAAT
Rhox6+1.5kb-BS-R	TAAAAAACCAACAAAACACCTTAAC
Elf5-BS-F	GGAGTTGGGTATAAAAGTAGGAGAAA
Elf5-BS-R	TAAAAACAAAAACTCAATCCCAAAC
Elf5-BS-Rn	AAATCCAAACAAAAAACTTTCAAAA
GM9-BS-F	GAGAATGGGTTAGAATTTTGAGGTA
GM9-BS-Fn	TAAGAGAGTTGAGTTTGGAGAAATT
GM9-BS-R	AATAAAAAATTAAACAAACTAAAAAACACA
Cul4b-BS-F	GGTGAATGTTGGGATAAAGTTTTAA
Cul4b-BS-R	GGTGAATGTTGGGATAAAGTTTTAA
Cul4b-Rn	CTAACCCACCTTTCTTAAAAAACC
Peg12-BS-F	TGTTGTAGTAGGAGGAAGTTGTTG T
Peg12-BS-R	CTTAACCTCCACCCAACCTAA T
Ndn-BS-F	ATATAGGAGATTAGGAAA T
Ndn-BS-R	CTATCCTACATCTCACAAT
Mkrn3-BS-F	AGATAATATTTGTAGAGG AT
Mkrn3-BS-R	CCTACCTATAAACAAATA

ChIP primers

Ndufa-CH-F (-73)	TCCGCACCGTTACTCGCACG
Ndufa-CH-R (+52)	AGCCACCGTCGCTTCCTCCT
Rhox2-CH-F1 (+211)	CGCACCAGCCTAGGCCCTGA
Rhox2-CH-R1 (+327)	GCAGGCCCCGTCCACTCTCT
Rhox6/9-CH-F1 (-285)	GGTGTCCCCCAAGCAGCACC
Rhox6/9-CH-R1 (-164)	GAGGCGCACAGGCAGAAGCA
Rhox11-CH-F1 (+413)	GCTCACTGGTCTCGGCCACC
Rhox11-CH-R1 (+299)	CCGAAAGAGGGCCACCAGAGC
Elf5-CH-F2 (+51)	GCGCGCTTGCCTTCTCTTGC
Efl5-CH-R2 (+169)	CTGAGCCCCGACACCCCCTT
Wfdc15CH-F1 (-445)	GGAGCTTGGGGACAGACGGC
Wfdc15-CH-R1 (-336)	CGCTGACACAAACCTTCGCCCT
Oct4-ChIP-F (-303)	GGGTGCAGTGCCAACAGGCTT
Oct4-ChIP-R (-132)	TCCAGACGGAGGTTGGGGACG
Dppa2-ChIP-F (+178)	GGCATTCATTCAGCGGCTGCCT
Dppa2-ChIP-R (+340)	CCCACGCAGTTCCTCGGAAGC

Dppa3-CHIP-F (+76)	TGAGGCTTCTGCCCATCGCATC
Dppa3-ChIP-R (+224)	CCAAAACCCCGGGCAGAAAGGC
Dppa4-ChIP-F (-50)	GACCAGAAGGGGAGGGGAACGG
Dppa4-ChIP-R (+110)	ACCTTCTTGTCTCCAGCAGTCTCC
Gdf3-ChIP-F (-300)	GAATTGCTCTCTTCTCAGCCCCTC
Gdf3-ChIP-R (-131)	TACACTGCTGACCAACCCGCGG

Supplemental tables

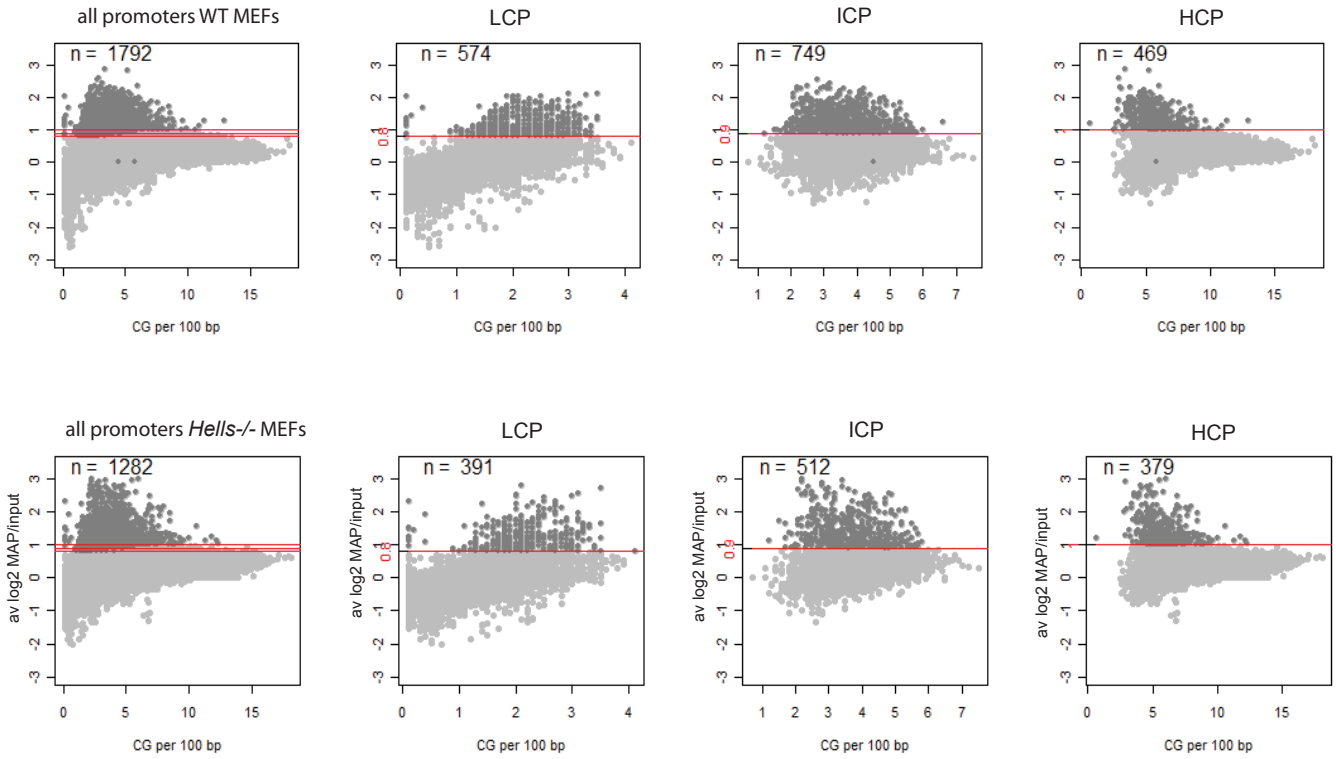
Supplemental Table 1 - Promoters showing either loss or gain of DNA methylation in *Hells*$^{-/-}$ MEFs

The table contains the data from two individual hybridization experiments performed with wild type and *Hells*$^{-/-}$ MEFs as well as the average methylation values and Δlog2 = (log2MAP/input WT) – (log2MAP/input *Hells*$^{-/-}$). Δlog2 $\geq$ +/- 1 (2-fold) was used as a cut off to identify differentially methylated promoters.

Supplemental Table 2 – Upstream regions showing either loss or gain of DNA methylation in *Hells*$^{-/-}$ MEFs

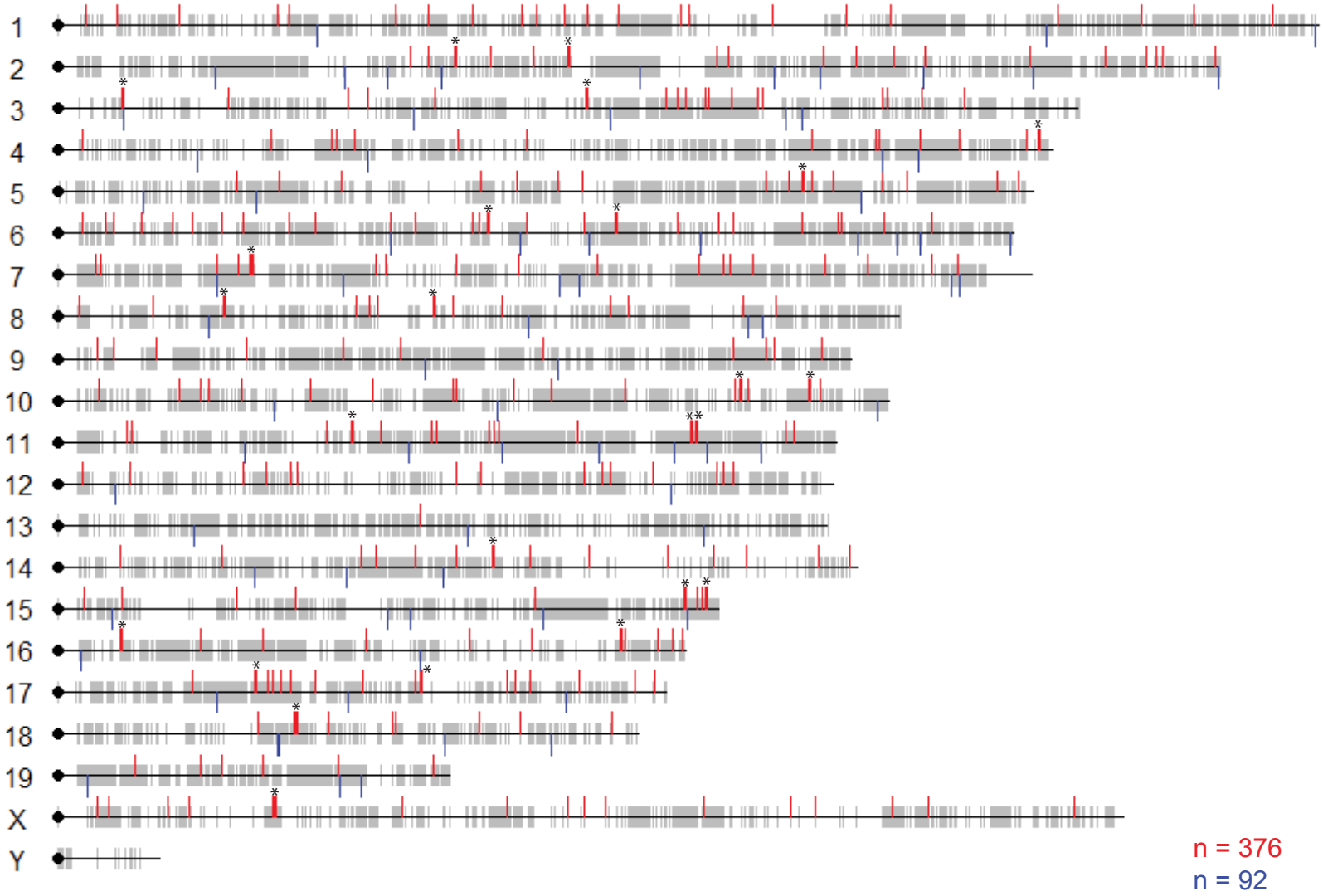
The table contains the data from two individual hybridization experiments performed with wild type and *Hells*$^{-/-}$ MEFs as well as the average methylation values and Δlog2 = (log2MAP/input WT) – (log2MAP/input *Hells*$^{-/-}$). Δlog2 $\geq$ +/- 1 (2-fold) was used as a cut off to identify differentially methylated upstream regions.

Supplemental Table 3 – Summary of all DNA methylation and expression data



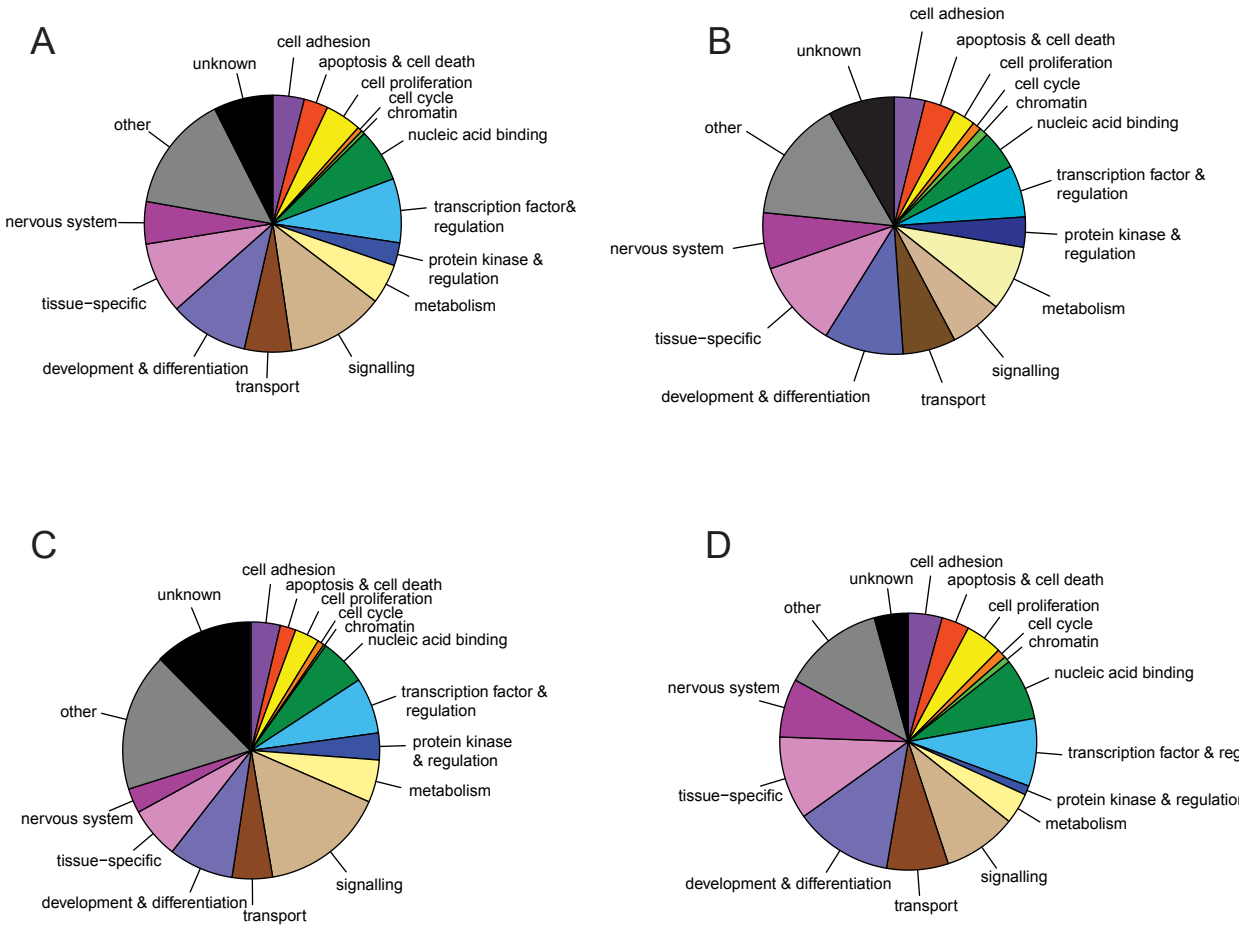
Supplemental Figure 1 Methylated promoters in wild type and *Hells-/-* MEFs

Promoters were subdivided into low CpG density (LCP), intermediate CpG density (ICP) and high CpG density (HCP) sequences as described[28, 29] and the cut off values log2 MAP/input were asigned to each class as shown. The number of promoters in each class displaying log2 MAP/input value above the cut off were calculated for wild type and *Hells-/-* MEFs. Note that in the Lsh deficient cells all three classes of promoters show loss of DNA methylation.



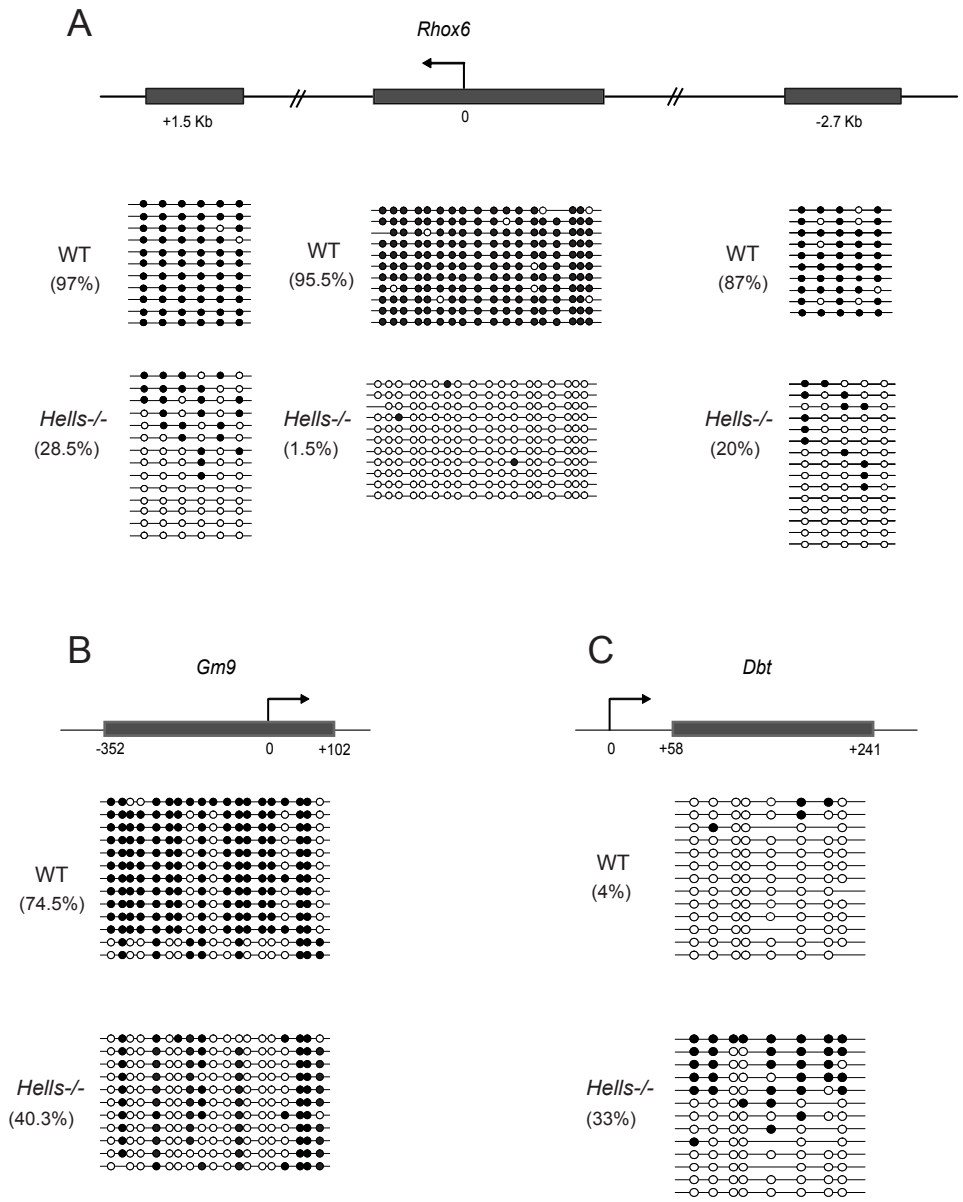
Supplemental Figure 2 Map of chromosomal position of hypo- and hypermethylated promoters in *Hells-/-* MEFs

Promoters that loose DNA methylation in *Hells-/-* MEFs (n=376) are shown in red. The asterisks indicate the largest domains containing several sequential hypomethylated promoters. Promoters which gain DNA methylation (n=96) in *Hells-/-* MEFs are shown in blue.



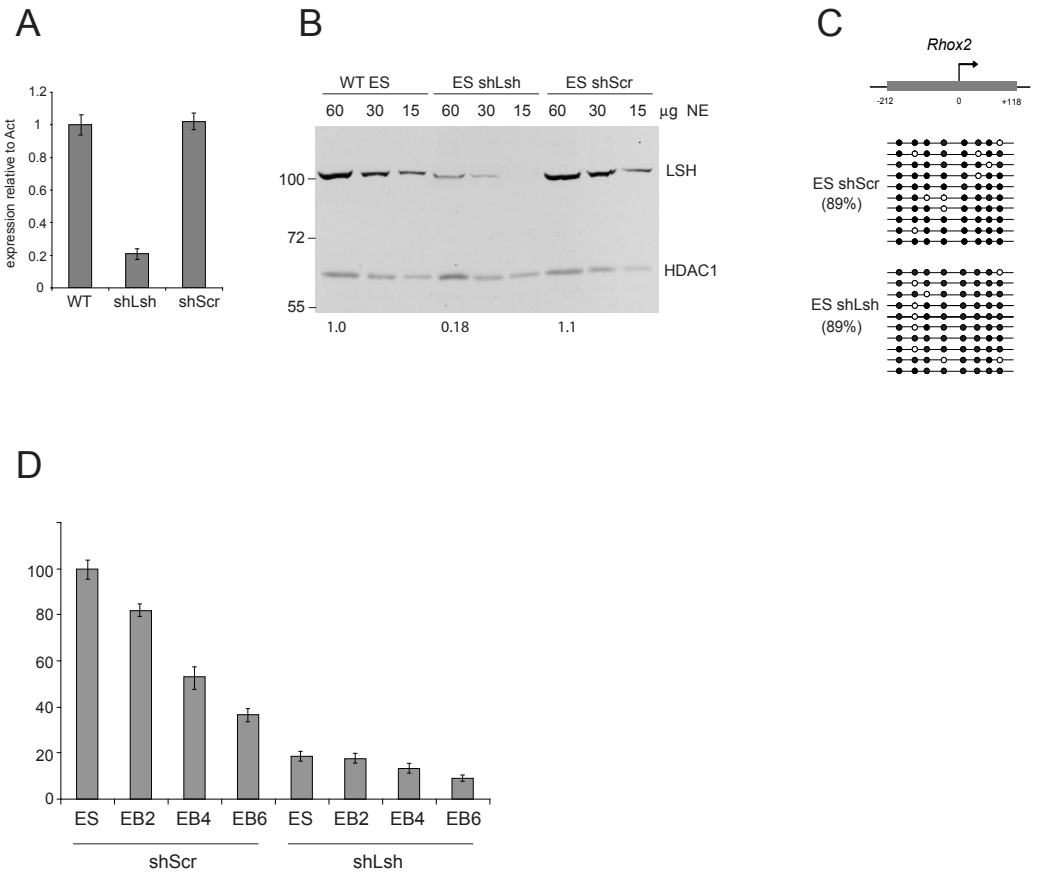
Supplemental Figure 3 GO annotation of genes with hypo- and hypermethylated promoters in *Hells-/-* MEFs

Generic GO Term finder and GO Tern mapper (http://go.princeton.edu/) were used to annotate genes with hypo- and hypermethylated promoters in *Hells-/-* MEFs. (**A**) All genes showing more than 2-fold loss of DNA methylation at promoters (n=376). (**B**) All genes showing gain of DNA methylation at promoters relative to wild type MEFs and ES cells (n=92; Group III in Figure 2G). (**C**) Group I genes, normally methylated in ES cells and MEFs, but not in *Hells-/-* MEFs (n=223).(D) Group II genes, normally methylated in MEFs, but not in ES cells and *Hells-/-* MEFs (n=143).



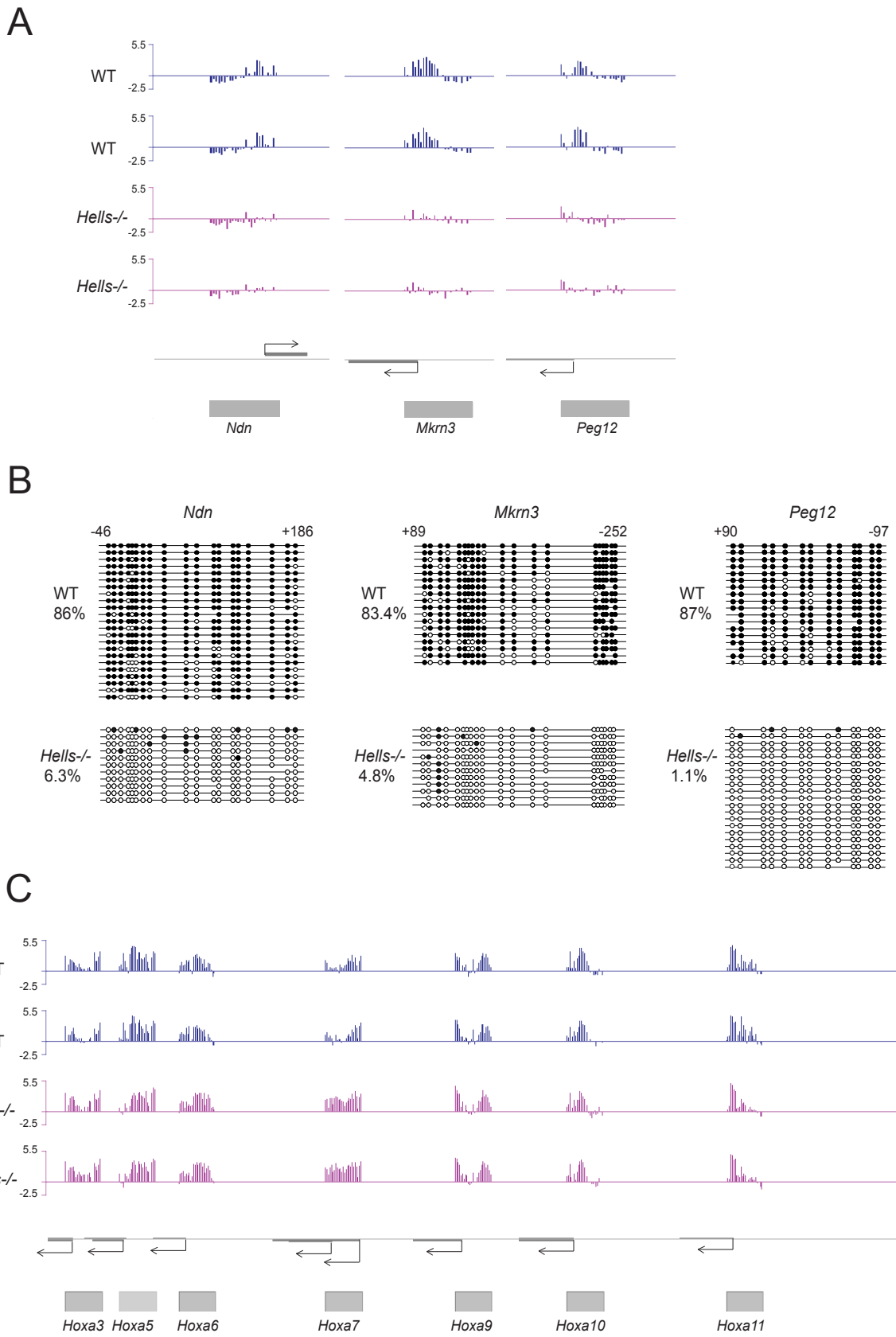
Supplemental Figure 4 Additional bisulfite DNA sequencing data for wild type and *Hells-/-* MEFs

(**A**) *Rhox6* promoter and sequences upstream and downstream from the *Rhox6* promoter are hypomethylated in *Hells-/-* MEFs. (**B**) DNA methylation at promoters of genes adjacent to the *Rhox* cluster. The promoter of *Gm9* gene, located immediately upstream of the *Rhox* locus, shows 2-fold decrease in DNA methylation according to the microarray data and 1.85-fold according to bisulfite sequencing. (**C**) Bisulfite DNA sequencing of *Dbt* promoter shows gain of DNA methylation in *Hells-/-* MEFs.



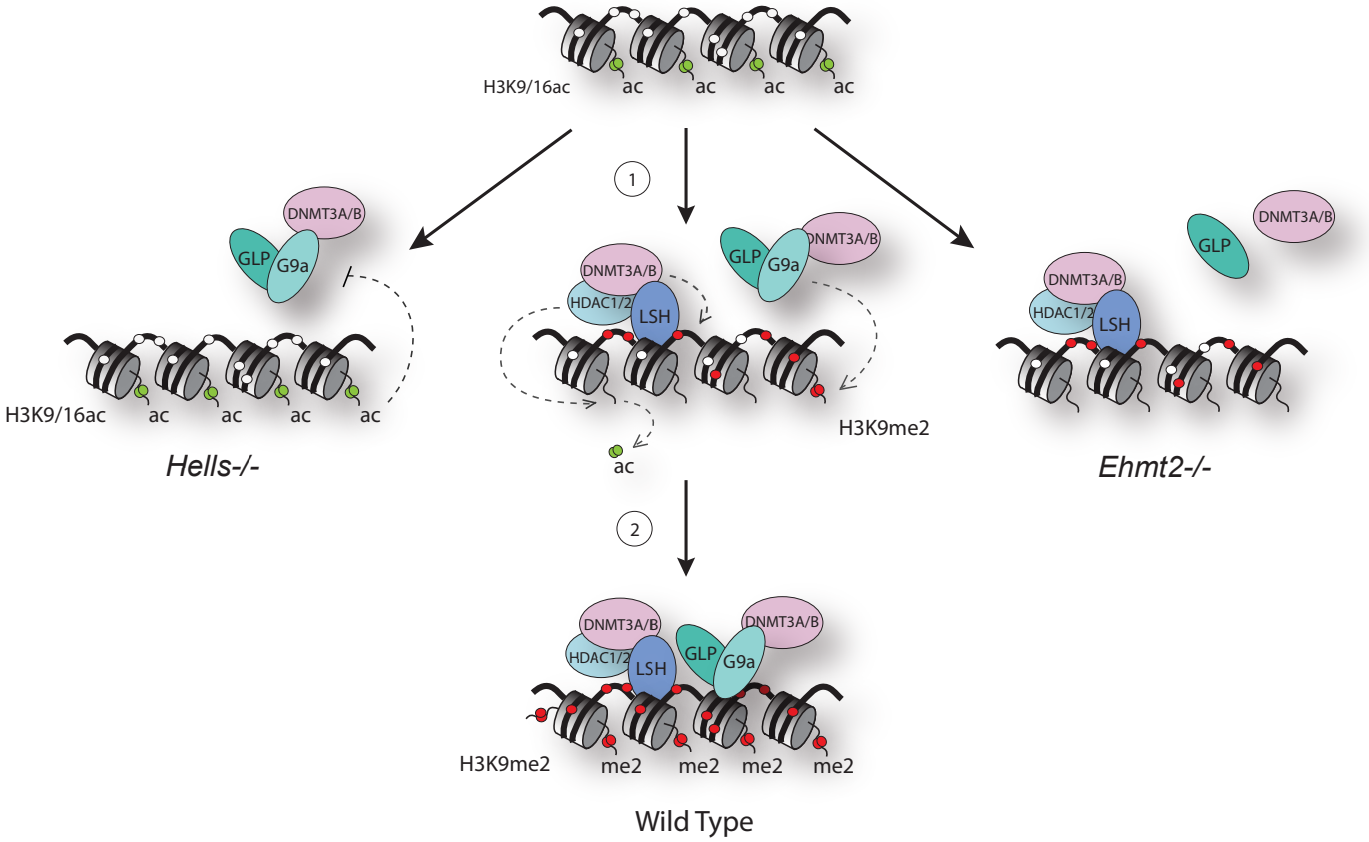
Supplemental Figure 5 Stable knock down of LSH in ES cells

(**A**) Real time PCR detects 80% knock down of Lsh RNA relative to Actin in cells stably expressing Lsh shRNA (Athanasiadou et al, 2010; Zhu et al, 2006). Wild type ES cells and cells expressing scrambled (Scr) Lsh shRNA were used as controls. (**B**) Quantitative Western blots detect 82% reduction of LSH protein in cells stably expressing Lsh shRNA. HDAC1 was used as a loading control. The numbers below the graph represent LSH levels normalized to HDAC1. (**C**) Bisulfite sequencing of *Rhox2* promoter in ES cells stably expressing either Lsh or scrambled shRNA. (**D**) LSH protein levels relative to HDAC1in shScr and shLsh ES cells and differentiating embryoid bodies (EB) at day 2, 4 and 6. The graph represents the average values from three independent quantitative Western blots.



Supplemental Figure 6 DNA methylation patterns at PWS and Hoxa loci

(**A**) Signal map view of *Ndn*, *Mkrn3* and *Peg12*, the imprinted genes from the mouse Prader-Willi Syndrome (PWS) locus on chromosome 7. These genes are normally expressed from the paternally derived chromosome in brain and methylated at promoter sequences and silenced on the maternal chromosome. (**B**) Bisulfite DNA sequencing shows that all three promoters display loss of DNA methylation in *Hells-/-* MEFs. The numbers above bisulfite plots represent base pair position relative to transcription start site. (**C**) Hoxa locus on chromosome 6 is heavily methylated in wild type and *Hells-/-* MEFs. The scale on the left of each plot indicates the log2 MAP/input values. Each bar represents a single probe on the microarray.



Supplemental Figure 7 A model of LSH-dependent recruitment of G9a/GLP and *de novo* DNA methylation

(1) The initial binding of LSH and LSH-associated DNMT3B/HDAC1,2 complex results in removal of acetyl groups (green circles) from histone H3 tail and partial DNA methylation. Such patterns can be observed in *Ehmt2-/-* (G9a-null) ES cells (right). (2) Deacetylation of histone H3 is likely to promote the recruitment of G9a/GLP complex and additional DNM3A/3B resulting in complete DNA methylation throughout the locus and appearance of dimethylated K9 (red circles) at histone H3. In the absence of LSH (*Hells-/-,* left), neither the initial recruitment of DNMT3A/3B (1) nor the LSH-facilitated accumulation of G9a/GLP complex (2) occur efficiently resulting in loss of DNA methylation and presence of acetylated chromatin. H3K9/K14 acetylation may inhibit H3K9me2 by G9a/GLP as both modifications occur on the dame amino acid (K9). Different loci may vary in their requirement for LSH and G9a. Promoters that are already silenced and deacetylated by other repressors may not require LSH for binding of G9a/GLP and subsequent DNA methylation.