Supplemental Figure Legends

Supplemental Figure 1: Multiple chromatin marks define the epigenomic border between heterochromatin and euchromatin.

Enrichments for multiple 'silent' (red) and 'active' (green) marks in BG3 cells are shown for the centromere-proximal 3Mb of chromosome arms 2R and 3L. Log intensity ratio values (Y-axis) are plotted for each mark relative to the chromosomal position (X-axis, bp). Boxes underneath the bar graphs demarcate genomic regions with significant enrichment ($p<0.001$). The cytogenomically-defined heterochromatin is shown in blue (solid bar), and the black arrows indicate the positions of the epigenomic borders. There is a striking congruence in the mapping patterns for heterochromatic marks.

Supplemental Figure 2: H3K9me3 is present in Drosophila and enriched in heterochromatin.

Using and interpreting data from antibodies generated against H3K9 methylated peptides can be problematic due to cross reactivity. Many H3K9me2 antibodies recognize H3K9me3, and vice versa, and can bind H3K27 methylated histones, due to similarities in the surrounding peptide sequences. Ebert et al. (Ebert et al. 2004) previously concluded that H3K9me3 is restricted to a small part of the chromocenter in polytene chromosomes, and is dramatically less abundant in flies than in mammalian cells, where H3K9me3 is a major heterochromatic mark (Peters et al. 2003). The specificity of the H3K9me3 antibodies used in our studies allows us to conclude that Drosophila does contain significant amounts of H3K9me3. It is likely that the antibodies used here recognize a broader spectrum of chromatin that contains this mark because they were generated against unbranched peptides; the antibody used by Ebert et al. (Ebert et al. 2004) was generated against a branched peptide (Peters et al. 2003), and its ability to recognize all forms of the modification *in vivo* may be more limited.

a. Distribution of H3K9me2 and H3K9me3 on polytene chromosomes. Polytene chromosomes were stained with antibodies to H3K9me2 (green) and H3K9me3 (red). Note the significant overlap at the chromocenter.

b. Western blot validation of the H3K9me3 antibody. Recombinant histone H3 expressed in *E. coli* and proteins from Drosophila embryo nuclei were assayed on a Western blot; both protein preparations contain equal amounts of H3. Only H3 from embryo nuclei is recognized (*) after probing with an antibody to H3K9me3, as recombinant H3 lacks post-translational modifications.

c. Slot blot peptide validation of the H3K9me3 antibody. Peptides with the listed modifications were blotted at 4 different concentrations (100, 30, 10, and 3 pmol), and the blot was incubated

with the H3K9me3 antibody. The highest reactivity is observed with the H3K9me3-containing peptide, and there is no apparent cross-reactivity with the H3K9me2 peptide. There is some cross-reactivity with the H3K27me3 peptide (signal intensity ratio for H3K9me3:H3K27me3 is approximately 10:1). Other assays have not found H3K27me3 in the pericentric heterochromatin.

d. The relative abundance of histone H3 peptides (amino acids 9-17) and unmethylated, mono-, di- or trimethylated H3 at lysine 9 in S2 and HeLa cells was estimated by mass spectrometry. Individual values for Drosophila and human peptides correspond to integrated areas under the extracted ion chromatograms of corresponding monoisotopic peaks. The ratios were normalized to that of a control peptide (amino acids 54-63, YQKSTELLIR), which does not carry any modifications.

e. Mapping of HP1c protein in chromatin from HP1c -/- confirms the majority of binding sites. The results of 2 ChIP experiments, each carried out with two biological replicates, are shown here. On the X-axis, the position along chromosome 2L (centromere to the right) in bp is shown. Y-axis shows smoothed M-value profiles for HP1c normalized to input. Regions of significant binding are indicated by boxes below the profile plots. To validate the binding profiles, we show enrichment profiles of HP1c in two different genotypes, wildtype (+/+) and an HP1c mutant (-/-; PBac{WH}HP1cf04929, stock #18819 from Bloomington). In HP1c mutants, (third instar larvae recovered from heterozygous parents), most HP1c signals are drastically decreased, with some signal remaining (see f. for quantitative analysis).

f. Fraction of the HP1c binding peaks disappearing in the HP1c mutant (-/-). The black line shows the fraction of the wild type HP1c binding positions (y-axis) that are not associated with a detectable peak in the HP1c mutant, as a function of the wild type HP1c peak magnitude (x-axis). The red line shows the fraction of wild type HP1c peaks that are significantly reduced (P value < 10^{-3}), as a function of wild type peak magnitude. The HP1c binding peaks were detected based on M=1.5 threshold, with the max distance of 600bp to any HP1c mutant peak positions. The curves show that while almost all HP1c peaks are significantly reduced (99.8%), some of the high-magnitude HP1c peaks are retained in the mutant samples. However, the number of such positions is very small (118 binding positions, 3.9%), and even these peaks show significant depletion in mutant samples. Only 6 HP1c binding positions (0.18%) are not significantly reduced in the HP1c mutant relative to the wild type. The rare remaining peaks could represent non-specific binding; however, there is no visible cross-reactivity observed with the HP1c antibody. Furthermore, the fact that high-magnitude wild type peaks are retained in mutants, and that these

peaks also show significant depletion in mutants, suggests that they result from small amounts of HP1c protein derived from maternal loading, and are likely to be *bona fide* HP1c binding sites.

Supplemental Figure 3: Heterochromatin-euchromatin borders differ among cell types.

H3K9me2 log intensity ratio values (Y-axis) in the proximal 3Mb region of the five chromosome arms (X-axis: sequence coordinates in bp) are shown for 2-4 hr embryos, 14-16 hr embryos, larvae, and adult heads, and for S2, BG3, Kc, and Clone 8 cells. Boxes underneath the bar graphs demarcate genomic regions with significant enrichment ($p<0.001$). The cytogenomically-defined heterochromatin is shown by the blue bar, and the blue arrowheads indicate the positions of the epigenomic border between euchromatin and heterochromatin. The 'Repeat Density' track shows the fraction of each 10kb window that consists of repeated DNAs, based on Repeat Masker (Release 3.28 (Smit et al. 1996-2010)). 'Gene coverage' plots the number of genes within 50kb windows, and individual genes are shown below with their orientations as indicated by the arrows. 'Histone' indicates the position of the histone gene cluster, which is not included on the arrays.

Supplemental Figure 4: Chromatin states of the centric heterochromatin and chromosome 4 in S2 and BG3 cells.

a. Prevalent combinatorial patterns of chromatin marks in BG3 cells. Pooled data from the heterochromatic regions and chromosome 4 were grouped into clusters based on similar patterns of chromatin mark enrichment (see Methods). Marks in panel 1 ('chromatin marks') were used for K-means clustering, and other properties shown in the remaining panels were assessed relative to these clusters. Due to the large combinatorial search space, some rarely occurring patterns may not be included, while others are represented by several similar states. Each column (panels 1 and 2) indicates average enrichment levels for a given histone modification or protein within the 15 identified clusters. Similarly, the third panel shows repeat content and RNA-seq levels relative to all genes. The 'genes' panel shows fractions of sequence within each cluster occurring within different parts of the annotated gene structures (white=0, black=1). The '% of sequence' column shows the fraction of sequence within BG3 heterochromatin associated within each cluster. The 'chromosomes' panel shows the fold over/under-representation of each cluster ($\log_2$ scale) in the heterochromatin from each chromosome arm (h and Het regions). The next two columns give the percentage of the cluster found in chromosome 4 ('% in chr4'), and the percentage of chromosome 4 that is accounted for by each cluster ('% of chr4'). '% in extensions' reports the percentage of each cluster present in the regions distal to the cytogenomic border that were

included in heterochromatin due to the localization of the epigenomic border (Figure 3b, Table 1). The cluster numbers (left-hand side) are colored green if they are associated with enhancer or promoter regions of active genes, light brown if they are associated with the elongating heterochromatic genes, and red if they show primarily heterochromatic 'silent' marks. The marks (top side) are colored red if they are generally associated with transcriptional silencing and green if they are associated with open chromatin regions.

b. Chromatin states in S2 cells. Average enrichment levels within euchromatin, heterochromatin and chromosome 4 are shown for the S2 cell data (top panel, as in Figure 4a of the main manuscript), followed by the 15-state clustering of S2 data described in **a**.

c. A simplified view of the 15-state clustering, grouping similar states as done for BG3 cells in Figure 4b.

Supplemental Figure 5: Chromatin states associated with heterochromatic and chromosome 4 genes in S2 cells.

Average chromatin states of expressed and silent genes are compared between euchromatic, heterochromatic and chromosome 4 regions in S2 cells. See Figure 5 of the main text for a detailed description of such analysis in BG3 cells. Note that the total number of genes analyzed in S2 cells (340) is greater than in BG3 cells (252; Figure 5) due to the more distal location of the epigenomic borders on all chromosome arms in S2 cells (Figure 3b and Table 1).

Supplemental Figure 6: Differences in correlation of heterochromatic marks between pericentric heterochromatin and chromosome 4.

a. To illustrate the overall relationship between HP1a enrichment and other chromatin marks, the plot shows Pearson correlation coefficients between HP1a and 'active' (green) or 'silent' (red) marks. The correlation is estimated separately in pericentric heterochromatin (grey bars) and chromosome 4 (blue bars). On chromosome 4, significant positive correlation is observed between HP1a and several marks associated with open chromatin and transcribed genes: H3K36me3, H2B-ubiquitination and H3K79me1. On the other hand, chromosome 4 shows lower correlation between H3K9me2 and H3K9me3.

b. Analogous comparison of correlation coefficients of H3K9me3 enrichment with other marks, illustrating a pattern of differences similar to that seen for HP1a.

c, d. Analogous comparison of heterochromatic and chromosome 4 correlation coefficients for HP1a and H3K9me3 for the S2 cell line. While the same trends are observed in the correlation

differences between chromosome 4 and pericentric heterochromatin, the magnitude of the differences is smaller than in BG3 cells.

Supplemental Figure 7: Average enrichment profiles of genes with S2-unique euchromatic H3K9me2 domains and heterochromatic genes.

The average profiles of enrichment in H3K9me2, HP1a and H3K36me3 are shown for **a.** all non-overlapping heterochromatic genes (n=542), **b.** expressed heterochromatic genes (n=307), **c.** genes associated with S2-unique and S2+Kc-unique euchromatic domains of H3K9me2 (n=167).

Supplemental Figure 8: Changes in chromatin states of the variable extension regions between S2 and BG3 cells.

a. The first panel shows the average changes (fold difference) between S2 and BG3 cells in the enrichment levels of different marks within the cytogenomically-defined heterochromatin (cytogenomic) and S2-specific extension regions (beyond the BG3 extensions). The orange colors indicate higher levels of enrichment in S2 cells compared to BG3, and green colors indicate lower levels of enrichment in S2 cells. The second and third panels show the average enrichments in S2 and BG3 cells separately.

b. Prevalent patterns of chromatin differences within the S2 extension regions. The S2 extension regions were broken into 500bp tiling blocks and clustered according to the overall pattern of differences in the enrichment levels between S2 and BG3 cells. Combinatorial chromatin state analysis identified regions that deviate from the average pattern. Approximately half (49%) of the S2-specific extension sequences (clusters 8-10), including 90% of the chromosome X extension, display only weak increases in both ‘silent’ and ‘active’ marks in S2 relative to BG3 cells. In contrast, 51% of the S2-specific extensions show significant increases in ‘silent’ marks (clusters 1-7). Some of these clusters reveal specific difference patterns relative to the corresponding BG3 cell sequences, including loss of H3K27me3 enrichment in S2 cells accompanied by gains in ‘active’ marks such as H3K36me3 (clusters 4-5, 18% of sequence), or loss of marks associated with active transcription (clusters 1-2; 9% of sequence).

Supplemental Figure 9: Changes in chromatin states of genes within S2 extension regions.

a. Average differences in the chromatin states of genes that are transcriptionally silent in both S2 and BG3 cell lines and lie within the cytogenomically-defined heterochromatic regions and S2 extensions. Similar to Supp. Figure 8, the first panel shows the difference in the enrichment levels between S2 and BG3 cells, with the two subsequent panels showing average chromatin profile of these genes in the two cell lines.

b. Average differences in the chromatin state between S2 and BG3 cells are shown for genes within cytogenomically-defined heterochromatin and S2 extension regions that are transcribed in both cell lines.

c. Prevalent patterns of chromatin state differences for genes located within the S2 extension regions. K-means clustering was used to group genes based on the difference of chromatin state between S2 and BG3 cell lines (first panel), with the second and third panels showing average chromatin state of genes within each cluster in S2 and BG3 cell lines. Examination of the differences in combinatorial chromatin profiles shows that 94% of the genes are in clusters with no overall change in RNA-seq signals in S2 versus BG3 cells (Supp. Figure 8c, clusters 4-10). The majority of genes (64%, clusters 8-10, 56 genes) show only minor changes in both 'silent' and 'active' mark enrichments in the two cell types, despite significant changes in the intergenic chromatin marks. Changes in clusters 4-7 (27 genes, 31%) are predominantly restricted to HP1a, H3K9me2 and H3K9me3 enrichment levels, with very few, minor changes in 'active' marks. Only five S2-specific extension genes (6%, clusters 1-3) show significant changes between the two cell lines in levels of marks associated with active transcription, including RNA-seq signals. Cluster 1 shows a single gene (*CG11426*) that lost all marks normally associated with transcription and acquired heterochromatic marks in S2 cells. However, clusters 2 and 3 show higher levels of expression in S2 cells compared to BG3. Cluster 3 includes two genes under PcG-mediated silencing in BG3 cells, but in S2 cells the PcG marks are absent and heterochromatic marks are present, along with RNA Pol II enrichment and other marks associated with active transcription. The two genes (*tsh* and *CG11629*) are in the same domain on arm 2L, separated by ~82kb. Cluster 2 contains two genes (*CG31619*, ~100kb distal to *tsh* on 2L, and *CG2201*, ~60kb distal to *CG31619*) that acquired heterochromatic marks in S2 cells, as well as increased expression and levels of other chromatin marks associated with transcription, relative to BG3 cells.

Supplemental Figure 10: Chromatin states of intergenic regions in BG3 and S2 cells.

a. The average level of enrichment across different chromatin marks is shown for the intergenic regions with euchromatic, heterochromatic and chromosome 4 regions. The intergenic regions

were defined as > 2Kbp away from any annotated gene. Compared to the pericentric heterochromatin, chromosome 4 intergenic regions exhibit significantly lower enrichment in 'silent' heterochromatic marks. '% of sequence' reports the percentage of all intergenic sequences in each region.

b. Prevalent combinatorial patterns of histone marks within the intergenic regions of pericentric heterochromatin and chromosome 4. The chromatin states of the intergenic regions were classified in the same way as described in Supp. Figure 4a. The state numbers (left-hand side) are colored to indicate states with signatures dominated by 'silent' marks (red), 'silent' marks and H3K36me3 (light brown), and those corresponding to open chromatin (green). '% of sequence' reports the percentage of heterochromatic intergenic sequences present in each cluster.

c, d. Analogous plots showing average enrichment levels and prevalent chromatin states of intergenic regions in S2 cells.

Supplemental Figure 11: Chromatin context of repetitive elements in BG3 cells.

An extended version of Figure 7a shows average chromatin states associated with instances of repetitive elements integrated within euchromatic and heterochromatic regions for BG3 cells. The name of the repeat type corresponding to each row is shown on the left. The number of repeat instances found in the euchromatic and heterochromatic regions is shown to the right of each panel. The fraction of the heterochromatic repeat instances found in the extension regions is illustrated by the grayscale column on the right.

Supplemental Figure 12: Chromatin context of repetitive elements in S2 cells.

A plot analogous to Supplemental Figure 11 is shown here for S2 cells.

Supplemental Figure 13: Genome-wide patterns of H3K9me2 enrichment common and unique to different cell types.

The figure shows a whole-genome analysis analogous to that shown for only euchromatic sequences in Figure 8a. Tiled genome regions were grouped into sets based on the H3K9me2 enrichment status across multiple cell types. Some regions exhibit H3K9me2 enrichment in nearly all cell types (cluster 1), while others are not H3K9me2-enriched in any of the examined cell types (cluster 6). In addition to these universal classes, we find regions with H3K9me2 enrichment specific to individual cell types (clusters 2-5). No specific regions of enrichment are found in the embryo samples, indicating that embryo cells exhibit a minimal set of heterochromatic regions, as observed for the euchromatin-heterochromatin borders. The bar plot on the left shows the fraction of the tiled genome associated with each enrichment type cluster, separating sequence that falls

within euchromatic and heterochromatic regions (based on the most distal border positions across all cell types). A significant fraction of H3K9me2 enrichment states is present within the euchromatin.

Supplemental Figure 14: H3K9me2 enrichment profiles of heterochromatic and "common"-domain euchromatic genes.

a. The average enrichment profiles of H3K9me2 are shown for 542 non-overlapping heterochromatic genes in different cell types.

b. The average enrichment profiles of 12 non-overlapping genes belonging to "common" euchromatic regions of H3K9me2 enrichment found primarily on chromosome X.

Supplemental Figure 15: Chromatin states within BG3-specific euchromatic H3K9me2 domains.

The plot shows four prevalent patterns of chromatin marks found within the BG3-specific euchromatic H3K9me2 domains. The first two panels show the chromatin state of these regions in S2 cells and BG3 cells. The regions are devoid of active chromatin marks in both cell lines, but marks associated with heterochromatic silencing are present only in BG3 cells. Three patterns of 'silent' marks are apparent: clusters 1 and 3 display strong enrichment for all four heterochromatin marks (example in Supp. Figure 16a, top panel, coordinates 15,100,000 to 16,250,000), cluster 4 contains moderate enrichment for H3K9me3 but much lower levels of HP1a and SU(VAR)3-9 enrichment (example in Supp. Figure 16a, red box and bottom panel: coordinates 14,707,500 to 14,973,000), and cluster 2 shows no enrichment or depletion for HP1a, and low enrichment in SU(VAR)3-9 (example in Supp. Figure 16a, top panel, coordinates 16,925,000 to 17,100,000).

Supplemental Figure 16: Genomic rearrangements are not associated with the majority of euchromatic H3K9me2 domains

a. Region of chromosome 2L with multiple H3K9me2-enriched domains in BG3 cells. Note that these domains differ in their enrichment for other heterochromatic marks: The domain(s) ranging from ~ 15,100,000 to 16,250,000 were enriched for all four heterochromatin marks included in this analysis, HP1a, SU(VAR)3-9, H3K9me2 and H3K9me3, while the H3K9me2 and H3K9me3 enriched domain ranging from 14,707,500 to 14,973,000 showed a much lower level of HP1a and SU(VAR)3-9 enrichment, and the domain from 16,925,000 to 17,100,000 lacked the enrichment for these two proteins completely while still maintaining high levels of H3K9me2 and me3. X-axis: Coordinates along chromosome 2L with centromere to the right. Genes are indicated in green with their orientations as indicated by the arrows. Y-axis: log intensity ratio values, with boxes marking

significantly enriched regions. Marks assayed from top: H3K9me2, H3K9me3, HP1a, SU(VAR)3-9, and RNA pol II. Top panel: Overview of the region. Bottom panel: Close-up on the H3K9me2-enriched domain investigated by PCR and Southern blot. PCR fragments assayed are shown below the enrichment graph (blue boxes). The arrow shows the approximate location of the rearrangement identified by Southern analysis.

b. PCR analysis of four H3K9me2-enriched domain borders shows a possible rearrangement in the proximity of the chr3R: 2,472,000 domain, indicated by the missing PCR fragment in BG3 cells.

c. Southern analysis indicates the presence of a second genomic rearrangement near a heterochromatin-like domain in BG3 cells on chromosome 2L. The left and right boundary of the heterochromatin-like domain on chromosome 2L from 14,707,500 - 14,973,000 was assayed using ~ 5kb PCR fragments as a probe. DNA from BG3 cells, S2 cells, and OR wild-type flies was digested with three enzymes, *Bam*HI, *Eco*RI, and *Hind*III. The fragments corresponding to the boundary region are marked by *. Note that none of the fragments detected show any length polymorphisms in the Southern on the left, while BG3 cells show a distinct fragment on the right for the *Bam*HI digest. Fragment sizes: (left) *Bam*HI - 3178bp and 9799bp; *Eco*RI - >15120bp; *Hind*III - 5824bp, 1403bp, 289bp (not visible), 958bp, and 4006bp; (right) *Bam*HI - >13,559bp; *Eco*RI - 2690bp (not visible as it overlap with the probe by only 70bp), 1819bp, 1663bp, 1033bp, 240bp (not visible), and 6736bp (not visible due to minimal overlap with the probe); *Hind*III - 1623bp and 6424bp.

Supplemental Figure 17: Antibody validation

Western blots and peptide blots for validation of the antibodies used in the manuscript are shown (ID numbers corresponding to those in Supplementary Table 1 are used). See the 'Antibody validation' section in Materials and Methods for validation criteria. For Western blots, equivalent amounts (based on Coomassie blue staining) of the histone protein of interest from embryo nuclei (Nuc. Ext.) and the unmodified recombinant histone (rH3 or rH4) were run on the same gel. Either extracts from RNAi knockdowns in S2 cells, or mutant protein extracts, were used to validate antibodies designed to be specific for a given protein. For peptide blots, 100, 33, 10, and 3.3 pM of the desired peptides were spotted onto membranes. The intensities of the bands were analyzed by Image J, and percent specificity was calculated relative to total intensity of all spots.

Supplemental Table 1: Quality control measurements for ChIP datasets

Overlap of top 40% peaks – overlap between the top 40% of peaks in the replicate experiments, with peaks determined based on p-value enrichment scores.

Overlap based on size-adjusted threshold – overlap between the peaks in the replicate experiments with peaks determined based on p-value enrichment scores and adjusted for the number of peaks called in each experiment.

Correlation coefficient r – correlation coefficient observed between the log10 p-value enrichment scores of the two replicate experiments. When three replicate datasets were available, the average correlation coefficient is reported.

Supplemental Data

Supplemental Data 1: Positions of BG3 clusters A-E (Figure 4b)

Supplemental Data 2: Positions of S2 clusters A-E (Supp. Figure 4c)

Supplemental Data 3: List of gene cluster memberships (Figure 5d, Supp. Figure 5c, Supp. Figure 9c)

Supplemental Data 4: Positions of K-means euchromatic H3K9 domains (Figure 8a)

Supplemental Data 5: Positions strictly-defined euchromatic H3K9 domains (Figure 8b)