

Published in final edited form as:

*Cell*. 2006 April 21; 125(2): 301–313. doi:10.1016/j.cell.2006.02.043.

## Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells

Tong Ihn Lee<sup>1,8</sup>, Richard G. Jenner<sup>1,8</sup>, Laurie A. Boyer<sup>1,8</sup>, Matthew G. Guenther<sup>1,8</sup>, Stuart S. Levine<sup>1,8</sup>, Roshan M. Kumar<sup>1</sup>, Brett Chevalier<sup>1</sup>, Sarah E. Johnstone<sup>1,2</sup>, Megan F. Cole<sup>1,2</sup>, Kyo-ichi Isono<sup>3</sup>, Haruhiko Koseki<sup>3</sup>, Takuya Fuchikami<sup>4</sup>, Kuniya Abe<sup>4</sup>, Heather L. Murray<sup>1</sup>, Jacob P. Zucker<sup>6</sup>, Bingbing Yuan<sup>1</sup>, George W. Bell<sup>1</sup>, Elizabeth Herbolsheimer<sup>1</sup>, Nancy M. Hannett<sup>1</sup>, Kaiming Sun<sup>1</sup>, Duncan T. Odom<sup>1</sup>, Arie P. Otte<sup>5</sup>, Thomas L. Volkert<sup>1</sup>, David P. Bartel<sup>1,2</sup>, Douglas A. Melton<sup>6</sup>, David K. Gifford<sup>1,7</sup>, Rudolf Jaenisch<sup>1,2</sup>, and Richard A. Young<sup>1,2,\*</sup>

<sup>1</sup>Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

<sup>2</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>3</sup>Developmental Genetics Group, RIKEN Center for Allergy and Immunology, 1-7-22, Suehiro, Tsurumiku, Yokohama, Kanagawa 230-0045, Japan

<sup>4</sup>Technology and Development Team for Mammalian Cellular Dynamics, BioResource Center, RIKEN Tsukuba Institute, 3-1-1, Koyadai, Tsukuba, Ibaraki 230-0045, Japan

<sup>5</sup>Swammerdam Institute for Life Sciences, University of Amsterdam, 1098 SM Amsterdam, The Netherlands

<sup>6</sup>Howard Hughes Medical Institute, Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA

<sup>7</sup>MIT CSAIL, 32 Vassar Street, Cambridge, MA 02139, USA

### SUMMARY

Polycomb group proteins are essential for early development in metazoans, but their contributions to human development are not well understood. We have mapped the Polycomb Repressive Complex 2 (PRC2) subunit SUZ12 across the entire nonrepeat portion of the genome in human embryonic stem (ES) cells. We found that SUZ12 is distributed across large portions of over two hundred genes encoding key developmental regulators. These genes are occupied by nucleosomes trimethylated at histone H3K27, are transcriptionally repressed, and contain some of the most highly conserved noncoding elements in the genome. We found that PRC2 target genes are preferentially activated during ES cell differentiation and that the ES cell regulators OCT4, SOX2, and NANOG cooccupy a significant subset of these genes. These results indicate that PRC2 occupies a special set of developmental genes in ES cells that must be repressed to maintain pluripotency and that are poised for activation during ES cell differentiation.

© 2013 Elsevier Inc. All rights reserved.

\*Contact: young@wi.mit.edu.

<sup>8</sup>These authors contributed equally to this work.

Supplemental Data

Supplemental Data include fifteen figures, fifteen tables, Experimental Procedures, and References and can be found with this article online at <http://www.cell.com/cgi/content/full/125/2/301/DC1/>.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## INTRODUCTION

Embryonic stem (ES) cells are a unique self-renewing cell type that can give rise to the ectodermal, endodermal, and mesodermal germ layers during embryogenesis. Human ES cells, which can be propagated in culture in an undifferentiated state but selectively induced to differentiate into many specialized cell types, are thought to hold great promise for regenerative medicine (Thomson et al., 1998; Reubinoff et al., 2000; Mayhall et al., 2004; Pera and Trounson, 2004). The gene expression program of ES cells must allow these cells to maintain a pluripotent state but also allow for differentiation into more specialized states when signaled to do so. Learning how this is accomplished may be key to realizing the therapeutic potential of ES cells and further understanding early development.

Among regulators of development, the Polycomb group proteins (PcG) are of special interest. These regulators were first described in *Drosophila*, where they repress the homeotic genes controlling segment identity in the developing embryo (Lewis, 1978; Denell and Frederick, 1983; Simon et al., 1992; Orlando and Paro, 1995; Pirrotta, 1998; Kennison, 2004). The initial repression of these genes is carried out by DNA binding transcriptional repressors, and PcG proteins modify chromatin to maintain these genes in a repressed state (Duncan, 1986; Bender et al., 1987; Strutt et al., 1997; Horard et al., 2000; Hodgson et al., 2001; Mulholland et al., 2003).

The PcG proteins form multiple Polycomb Repressive Complexes (PRCs), the components of which are conserved from *Drosophila* to humans (Franke et al., 1992; Shao et al., 1999; Birve et al., 2001; Tie et al., 2001; Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Levine et al., 2002). The PRCs are brought to the site of initial repression and act through epigenetic modification of chromatin structure to promote gene silencing (Pirrotta, 1998; Levine et al., 2004; Lund and van Lohuizen, 2004; Ringrose and Paro, 2004). PRC2 catalyzes histone H3 lysine-27 (H3K27) methylation, and this enzymatic activity is required for PRC2-mediated gene silencing (Cao et al., 2002; Czermin et al., 2002; Kuzmichev et al., 2002; Muller et al., 2002; Kirmizis et al., 2004). H3K27 methylation is thought to provide a binding surface for PRC1, which facilitates oligomerization, condensation of chromatin structure, and inhibition of chromatin remodeling activity in order to maintain silencing (Shao et al., 1999; Francis et al., 2001; Cao et al., 2002; Czermin et al., 2002).

Components of PRC2 are essential for the earliest stages of vertebrate development (Faust et al., 1998; O'Carroll et al., 2001; Pasini et al., 2004). PRC2 and its related complexes, PRC3 and PRC4, contain the core components EZH2, SUZ12, and EED (Kuzmichev et al., 2004; Kuzmichev et al., 2005). EZH2 is a H3K27 methyltransferase, and SUZ12 (Suppressor of zeste 12) is required for this activity (Cao and Zhang, 2004; Pasini et al., 2004). ES cell lines cannot be established from Ezh2-deficient blastocysts (O'Carroll et al., 2001), suggesting that PRC2 is involved in regulating pluripotency and self-renewal. Although the PRCs are known to repress individual *HOX* genes (van der Lugt et al., 1996; Akasaka et al., 2001; Wang et al., 2002; Cao and Zhang, 2004), it is not clear how these important PcG regulators contribute to early development in vertebrates.

Because the nature of PRC2 target genes in ES cells might reveal why PRC2 is essential for early embryonic development, pluripotency, and self-renewal, we have mapped the sites occupied by the SUZ12 subunit throughout the genome in human ES cells. This genome-wide map reveals that PRC2 is associated with a remarkable cadre of genes encoding key regulators of developmental processes that are repressed in ES cells. The genes occupied by PRC2 contain nucleosomes that are trimethylated at histone H3 lysine-27 (H3K27me3), a modification catalyzed by PRC2 and associated with the repressed chromatin state. Both PRC2 and nucleosomes with histone H3K27me3 occupy surprisingly large genomic

domains around these developmental regulators and are frequently associated with highly conserved non-coding sequence elements previously identified by comparative genomic methods. The transcription factors OCT4, SOX2, and NANOG, which are also key regulators of ES cell pluripotency and self-renewal, occupy a significant subset of these genes. Thus, the model of epigenetic regulation of homeotic genes extends to a large set of developmental regulators whose repression in ES cells appears to be key to pluripotency. We suggest that PRC2 functions in ES cells to repress developmental genes that are preferentially activated during differentiation.

## RESULTS AND DISCUSSION

### Mapping Genome Occupancy in ES Cells

We mapped the location of both RNA polymerase II and the SUZ12 subunit of PRC2 genome-wide in human ES cells (Figure 1). The initiating form of RNA polymerase II was mapped to test the accuracy of the method and provide a reference for comparison with sites occupied by PRC2. The SUZ12 subunit of PRC2 is critical for the function of the complex and was selected for these genome-wide experiments. Human ES cells (H9, NIH code WA09) were analyzed by immunohistochemistry for characteristic stem cell markers, tested for their ability to generate cell types from all three germ layers upon differentiation into embryoid bodies, and shown to form teratomas in immunocompromised mice (Supplemental Data; Figures S1–S3).

DNA sequences bound by RNA polymerase II were identified in replicate chromatin-immunoprecipitation (ChIP) experiments using DNA microarrays that contain over 4.6 million unique 60-mer oligonucleotide probes spanning the entire nonrepeat portion of the human genome (Figure 1 and Supplemental Data). To obtain a probabilistic assessment of binding events, an algorithm was implemented that incorporates information from multiple probes representing contiguous regions of the genome, and threshold criteria were established to identify a data-set with minimal false positives and false negatives. RNA polymerase II was associated with the promoters of 7,106 of the approximately 22,500 annotated human genes, indicating that one-third of protein-coding genes are prepared to be transcribed in ES cells. Three lines of evidence suggest this dataset is of high quality. Most of the RNA polymerase II sites (87%) occurred at promoters of known or predicted genes. Transcripts were detected for 88% of the genes bound by RNA polymerase II in previous expression experiments in ES cells. Finally, independent analysis using gene-specific PCR (Supplemental Data) indicated that the frequency of false positives was approximately 4% and the frequency of false negatives was approximately 30% in this dataset. A detailed analysis of the RNA polymerase II dataset, including binding to miRNA genes, can be found in Supplemental Data (Tables S1–S6 and Figures S4 and S5).

The sites occupied by SUZ12 were then mapped throughout the entire nonrepeat genome in H9 ES cells using the same approach described for RNA polymerase II (Figure 1C). SUZ12 was associated with the promoters of 1,893 of the approximately 22,500 annotated human genes, indicating that ~8% of protein-coding genes are occupied by SUZ12 in ES cells (Supplemental Data; Tables S7 and S8). Independent site-specific analysis indicated that the frequency of false positives was approximately 3% and the frequency of false negatives was approximately 27% in this dataset.

Comparison of the genes occupied by SUZ12 with those occupied by RNA polymerase II revealed that the two sets were largely exclusive (Figure 1D; Supplemental Data; Table S8). There were, however, genes where SUZ12 and RNA polymerase II cooccupied promoters. At these genes, PRC complexes may fail to block assembly of the preinitiation complex (Dellino et al., 2004), consistent with the observation that Polycomb group proteins can

associate with components of the general transcription apparatus (Breiling et al., 2001; Saurin et al., 2001).

The vast majority of SUZ12 bound sites were found at gene promoters (Figure 1E). Ninety-five percent of the SUZ12 bound regions were found within 1 kb of known or predicted transcription start sites (Supplemental Data and Table S7). This suggests that SUZ12 functions in human ES cells primarily at promoters rather than at distal regulatory elements. It is interesting that 40% of all SUZ12 bound regions are within 1 Kb of CpG islands (Table S7), given the recent discovery of a mechanistic link between PcG proteins and DNA methyltransferases (Vire et al., 2006).

### Global Transcriptional Repression by PRC2

PRC2 is composed of three core subunits, SUZ12, EED, and EZH2, and has been shown to mediate histone H3K27 methylation at specific genes in vivo. To confirm that SUZ12 is associated with active PRC2 at target genes, we used chromatin immunoprecipitation with antibodies against EED and the histone H3K27me3 mark and analyzed the results with promoter microarrays. We found that EED and the histone H3K27me3 mark cooccurred with SUZ12 at most genes using a high-confidence binding threshold (Figure 2). The false negative rates of thresholded data can lead to an underestimate of the similarity between different datasets. Plotting raw enrichment ratios for genes associated with SUZ12, EED, or H3K27me3 demonstrates that SUZ12 binding represents PRC2 binding at almost all target genes (Figure S6).

Genetic and biochemical studies at selected genes indicate that PRC2-mediated H3K27 methylation represses gene expression, but it has not been established if it acts as a repressor genome-wide. If genes occupied by SUZ12 are repressed by PRC2, then transcripts from these genes should generally be present at lower levels in ES cells than in differentiated cell types. To test this prediction, we compared the expression levels of PRC2-occupied genes in four different ES cell lines with the expression level of these genes in 79 differentiated human cell and tissue types (Sato et al., 2003; Abeyta et al., 2004; Su et al., 2004). We found that PRC2 occupied genes were generally underexpressed in ES cells relative to other cell types (Figure 2C). A small fraction of the genes occupied by PRC2 were relatively overexpressed in ES cells (Figure 2C); these tended to show less extensive SUZ12 occupancy and were more likely to be cooccupied by RNA polymerase II (Supplemental Data). These results are consistent with the model that PRC2-mediated histone H3K27 methylation promotes gene silencing at the majority of its target genes throughout the genome in ES cells.

### Key Developmental Regulators Are Targets of PRC2

Examination of the targets of SUZ12 revealed that they were remarkably enriched for genes that control development and transcription (Figure 3) and that SUZ12 tended to occupy large domains at these genes (Figure 4). Although only 8% of all annotated genes were occupied by SUZ12, ~50% of those encoding transcription factors associated with developmental processes were occupied by SUZ12. By comparison, RNA polymerase II preferentially occupied genes involved in a broad spectrum of cell proliferation functions such as nucleic acid metabolism, protein synthesis, and cell cycle (Figure 3A and examples in Figure 1B; Supplemental Data; Table S10).

It was striking that SUZ12 occupied many families of genes that control development and transcription (Figures 3B and S7 and Table S11). These included 39 of 40 of the homeotic genes found in the *HOX* clusters and the majority of homeodomain genes. SUZ12 bound homeodomain genes included almost all members of the *DLX*, *IRX*, *LHX*, and *PAX* gene

families, which regulate early developmental steps in neurogenesis, hematopoiesis, axial patterning, tissue patterning, organogenesis, and cell-fate specification. SUZ12 also occupied promoters for large subsets of the *FOX*, *SOX*, and *TBX* gene families. The forkhead family of *FOX* genes is involved in axial patterning and tissue development from all three germ layers (Lehmann et al., 2003). Mutations in members of the *SOX* gene family alter cell-fate specification and differentiation and are linked to several developmental diseases (Schepers et al., 2002). The *TBX* family of genes regulates a wide variety of developmental processes such as gastrulation, early pattern formation, organogenesis, and limb formation (Showell et al., 2004). Thus, the genes preferentially bound by SUZ12 have functions that, when expressed, promote differentiation. This is likely to explain, at least in part, why PRC2 is essential for early development and ES cell pluripotency.

A remarkable feature of PRC2 binding at most genes encoding developmental regulators was the extensive span over which the regulator occupied the locus (Figures 4, S8, and S9). For the majority (72%) of bound sites across the genome, SUZ12 occupied a small region of the promoter similar in size to regions bound by RNA polymerase II (Figure 1). For the remaining bound regions, SUZ12 occupancy encompassed large domains spanning 2–35 kb and extending from the promoter into the gene. A large portion of genes encoding developmental regulators (72%) exhibited these extended regions of SUZ12 binding. In some cases, binding encompassed multiple contiguous genes. For instance, SUZ12 binding extended ~100 kb across the entire *HOXA*, *HOXB*, *HOXC*, and *HOXD* clusters but did not bind to adjacent genomic sequences, yielding a highly defined spatial pattern (Figure 4B). In contrast, clusters of unrelated genes, such as the interleukin 1- cluster, were not similarly bound by SUZ12. Thus, genes encoding developmental regulators showed an unusual tendency to be occupied by PRC2 over much or all of their transcribed regions.

### PRC2 and Highly Conserved Elements

Previous studies have noted that many highly conserved noncoding elements of vertebrate genomes are associated with genes encoding developmental regulators (Bejerano et al., 2004; Siepel et al., 2005; Woolfe et al., 2005). Given SUZ12's strong association with this class of genes, we investigated the possibility that SUZ12 bound regions are associated with these highly conserved elements. Inspection of individual genes suggested that SUZ12 occupancy was associated with regions of sequence conservation (Figure 5A). Eight percent of the approximately 1,400 highly conserved noncoding DNA elements described by Woolfe and colleagues (Woolfe et al., 2005) were found to be associated with the SUZ12 bound developmental regulators (p-value  $10^{-14}$ ). Using entries from the PhastCons database of conserved elements (Siepel et al., 2005), we found that SUZ12 occupancy of highly conserved elements was highly significant (using highly conserved elements with a LoD conservation score of 100 or better, the p-value for significances was less than  $10^{-85}$ ). Since PRC2 has not been shown to directly bind DNA sequences, we expect that specific DNA binding proteins occupy the highly conserved DNA sequences and may associate with PRC2, which spreads and occupies adjacent chromatin. Thus, the peaks of SUZ12 occupancy might not be expected to precisely colocalize with the highly conserved elements, even if these elements are associated with PRC2 recruitment.

Remarkably, the degree of the association between SUZ12 binding and conserved sequences increases when considering sequences with an increasing degree of conservation (Figure 5B). By comparison, RNA polymerase II showed no such enrichment. These results suggest that the subset of highly conserved noncoding elements at genes encoding developmental regulators may be associated with PcG-mediated silencing of these regulators.

## Signaling Genes Are among PRC2 Targets

The targets of SUZ12 were also enriched for genes that encode components of signaling pathways (Figure 3A and Table S12). There is evidence that transforming growth factor- (TGF ), bone morphogenic protein (BMP), wingless-type MMTV integration site (Wnt), and fibro-blast growth factor (FGF) signaling pathways, which are required for gastrulation and lineage differentiation in the embryo, are also essential for self-renewal and differentiation of ES cells in culture (Loebel et al., 2003; Molofsky et al., 2004). SUZ12 generally occupied the promoters of multiple components of these pathways, but it occupied larger domains within a group of signaling genes that contained highly conserved elements. This group contained members of the Wnt family (*WNT1*, *WNT2*, *WNT6*) as well as components of the TGF superfamily (*BMP2*, *GDF6*). Recent studies have shown that Wnt signaling plays a role in pluripotency and self-renewal in both mouse and human ES cells (Sato et al., 2004), and our results suggest that it is important to maintain specific family members in a repressed state in ES cells.

## Activation of PRC2 Target Genes during Differentiation

PRC2 is associated with an important set of developmental regulators that must be silent in ES cells but activated during differentiation. This observation suggests that PRC2 ultimately functions to repress occupied genes in ES cells and that these genes may be especially poised for transcriptional activation during ES cell differentiation. We reasoned that if this model is correct, genes bound by SUZ12 should be preferentially activated upon ES cell differentiation or in cells that lack SUZ12. Furthermore, in differentiated cells, SUZ12 might continue to be observed at silent genes but must be removed from genes whose expression is essential for that cell type.

We first examined gene expression in ES cells stimulated to undergo differentiation (Sato et al., 2003). We found that genes occupied by SUZ12 were more likely to be activated during ES cell differentiation than genes that were not occupied by SUZ12 (Figure 6A; Supplemental Data; Table S13), indicating that SUZ12-occupied genes show preferential activation during differentiation under these conditions. Thirty-six percent of genes bound by SUZ12 showed greater than 2-fold increases in expression during ES cell differentiation, whereas only 16% of genes not bound by SUZ12 showed such an increase. This effect was particularly striking at the set of developmental regulators (Figure 6B). SUZ12 occupied most (83%) of the developmental regulators that were induced more than 10-fold during ES cell differentiation.

We next examined the expression of SUZ12 target genes in *Suz12*-deficient cell lines derived from homozygous mutant blastocysts (Supplemental Data). We reasoned that genes bound by SUZ12 in human ES cells have orthologs in mice that should be upregulated in *Suz12*-deficient mouse cells, although we expected the overlap in these sets of genes to be imperfect because of potential differences between human and mouse ES cells, the possible repression of PRC2 target genes by additional mechanisms, and pleiotropic effects of the *Suz12* knockout on genes downstream of *Suz12*-target genes. Differences in gene expression between *Suz12* homozygous mutant cells and wild-type ES cells were measured using gene expression microarrays and the human SUZ12 binding data mapped to orthologous mouse genes using HomoloGene ([www.ncbi.nlm.nih.gov/HomoloGene](http://www.ncbi.nlm.nih.gov/HomoloGene)). We found that a significant portion of mouse genes whose counterparts were bound by SUZ12 in human ES cells were upregulated in *Suz12*-deficient mouse cells (70 of 346 genes,  $p = 6 \times 10^{-4}$ ); these genes are listed in Table S14. Orthologs of genes occupied by SUZ12 in human ES cells were more likely to be activated and less likely to be repressed in *Suz12*-deficient mouse cells than orthologs of genes not occupied by SUZ12 (Figure 6C). Furthermore, we found that orthologs of *Suz12* target genes that were induced upon human

ES cell differentiation were generally also induced upon loss of Suz12 in mouse cells (Figure 6D). Genes that were activated during ES cell differentiation and in Suz12-deficient cells included those encoding transcriptional regulators (*GATA2*, *GATA3*, *GATA6*, *HAND1*, *MEIS2*, and *SOX17*) signaling proteins (*WNT5A*, *DKK1*, *DKK2*, *EFNA1*, *EFNB1*, *EPHA4*, and *EPHB3*) and the cell-cycle inhibitor *CDKN1A*. These data indicate that Suz12 is necessary to fully repress the genes that are occupied by PRC2 in wild-type ES cells and have since been confirmed with binding data and knockout studies of a second PRC subunit in mouse (Boyer et al., 2006).

If PRC2 functions to repress genes in ES cells that are activated during differentiation, then in differentiated tissues SUZ12 occupancy should be diminished at genes encoding developmental regulators that have a role in specifying the identity of that tissue, similar to results seen with Ezh2 at specific genes in mouse (Caretti et al., 2004). To test this, we designed an array focused on the promoters of developmental regulators and used ChIP-Chip to investigate SUZ12 occupancy at these promoters in primary differentiated muscle cells. The results demonstrated that genes encoding key regulators of muscle differentiation, including *MYOD1*, displayed greatly diminished SUZ12 occupancy when compared to ES cells (Figure 6E). *MYOD1* is a master regulator for muscle differentiation (Tapscott, 2005), and the gene encoding this transcription factor displayed no significant SUZ12 occupancy when compared to the levels of SUZ12 occupancy observed in ES cells. Genes encoding other transcriptional regulators that play a central role in muscle development, such as *PAX3* and *PAX7* (Brand-Saberi, 2005), showed reduced levels of SUZ12 occupancy in muscle cells relative to ES cells (Supplemental Data and Figure S11). In contrast, other developmental regulators important for differentiation of nonmuscle tissues remained occupied by SUZ12 in differentiated muscle cells (Figure 6F and Table S15). These data support a model where PRC2 binding in ES cells represses key developmental regulators that are later expressed during differentiation.

### Targets of PRC2 Are Shared with Key ES Cell Regulators

The transcription factors OCT4, SOX2, and NANOG have essential roles in early development and are required for the propagation of undifferentiated ES cells in culture (Nichols et al., 1998; Avilion et al., 2003; Chambers et al., 2003; Mitsui et al., 2003). We recently reported that these transcription factors occupied promoters for many important developmental regulators in human ES cells (Boyer et al., 2005). This led us to compare the set of genes encoding developmental regulators and occupied by OCT4, SOX2, and NANOG with those occupied by PRC2 (Figure 7 and Supplemental Data). We found that each of the three DNA binding transcription factors occupied approximately one-third of the PRC2-occupied genes that encode developmental transcription factors (Figure 7A; Supplemental Data; Table S11). Remarkably, we found that the subset of genes encoding developmental regulators that were occupied by OCT4, SOX2, and NANOG and repressed in the regulatory circuitry highlighted in Boyer et al. were almost all occupied by PRC2 (Figure 7B). These included genes for transcription factors known to be important for differentiation into extraembryonic, endodermal, mesodermal, and ectodermal lineages (e.g., *ESX1L*, *ONECUT1*, *HAND1*, *HOXB1*). As expected, active genes encoding ES cell transcription factors (e.g., *ZIC3*, *STAT3*, *OCT4*, *NANOG*) were occupied by OCT4, SOX2, NANOG, and RNA polymerase II but not by PRC2 (Figure 7B).

The observation that OCT4, SOX2, and NANOG are bound to a significant subset of developmental genes occupied by PRC2 supports a link between repression of developmental regulators and stem cell pluripotency. Like PRC2, OCT4 and NANOG have been shown to be important for early development and ES cell identity. It is possible, therefore, that inappropriate regulation of developmental regulators that are common targets of OCT4, NANOG, and PRC2 contributes to the inability to establish ES cell lines in OCT4,

NANOG, and EZH2 mutants (Nichols et al., 1998; O'Carroll et al., 2001; Chambers et al., 2003; Mitsui et al., 2003).

### Concluding Remarks

We have mapped the sites occupied by SUZ12 throughout the genome to gain insights into how PRC2 contributes to pluripotency in human embryonic stem cells. ES cells proliferate in an undifferentiated state yet remain poised to respond to development cues. Genes encoding the transcriptional regulators that promote differentiation must therefore be repressed in ES cells but activated upon receiving signals to differentiate. We found that PRC2 occupies large domains at genes encoding a key set of repressed developmental regulators that are preferentially activated upon cellular differentiation, thus implicating this complex directly in the maintenance of the pluripotent state.

Transcription factors and chromatin regulators contribute to the transcriptional regulatory circuitry responsible for pluripotency and self-renewal in human ES cells. Understanding this circuitry is fundamental to understanding human development and realizing the therapeutic potential of these cells. In this context, we find it exciting that the outlines of the core transcriptional regulatory circuitry of human ES cells are emerging. The transcription factors OCT4, SOX2, and NANOG are associated with actively transcribed genes that contribute to growth and self-renewal (Boyer et al., 2005). These factors also occupy genes encoding key developmental regulators that are transcriptionally repressed, due at least in part to their association with PRC2 and nucleosomes modified at histone H3K27me3. Further study of transcription factors and chromatin regulators genome-wide will allow investigators to produce a more comprehensive map of transcriptional regulatory circuitry in ES cells and to test models that emerge from the circuitry. This information may provide insights into approaches by which pluripotent cells can be stimulated to differentiate into different cell types.

## EXPERIMENTAL PROCEDURES

### Cells and Cell Culture

Human H9 ES cells (WiCell, Madison, WI) were cultured as described (Boyer et al., 2005). Primary human skeletal muscle cells were obtained from Cell Applications (San Diego, CA) and expanded and differentiated into myotubes according to the supplier's protocols. Suz12  $-/-$  mouse cell lines were derived from blastocysts from crosses between heterozygous Suz12 mutant animals, as described in Supplemental Data.

### Chromatin Immunoprecipitation and DNA Microarray Analysis

ChIP was combined with DNA microarray analysis as described (Boyer et al., 2005). The antibodies used here were specific for hypophosphorylated RNA polymerase II (8WG16) (Thompson et al., 1989), SUZ12 (Upstate, 07-379), EED (Hamer et al., 2002), H3K27me3 (Abcam, AB6002), and total histone H3 (Abcam, AB1791). The design of the oligobased arrays, which were manufactured by Agilent Technologies, is described in detail in Supplemental Data. A whole-chip error model was used to calculate confidence values from the enrichment ratio and the signal intensity of each probe (probe p-value) and of each set of three neighboring probes (probe-set p-value). Probe-sets with significant probe-set p-values ( $p < 0.001$ ) and significant individual probe p-values were judged to be bound (see Supplemental Data for additional information). Bound regions were assigned to genes if they were within 1 kb of the transcription start site from one of five genomic databases; RefSeq, MGC, Ensembl, UCSC Known Gene, or H-Inv. All microarray data is available at ArrayExpress under the accession designation E-WMIT-7.



## Gene Expression Analysis

Gene expression data were collated from H1 ES cells (Sato et al., 2003), H9, HSF1, and HSF6 ES cells (Abeyta et al., 2004), and 79 differentiated human cell and tissue types (Su et al., 2004) and analyzed as described in detail in Supplemental Data. Replicate gene expression data was obtained for wild-type mouse ES cells and Suz12-deficient cells using Agilent Mouse Development arrays and were analyzed as described in Supplemental Data.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Elizabeth Jacobsen for technical assistance and Robert Brady for help with array design. L.A.B. and H.L.M. were supported by NRSA postdoctoral fellowships. M.G.G. is an Amgen Fellow of LSRF. R.M.K. was supported by a fellowship from the ACS. D.T.O. was supported by NIH award DK070813. This work was supported by NIH grants HG002668 and GM069400. T.L., T.L.V., D.K.G., and R.A.Y. consult for Agilent Technologies.

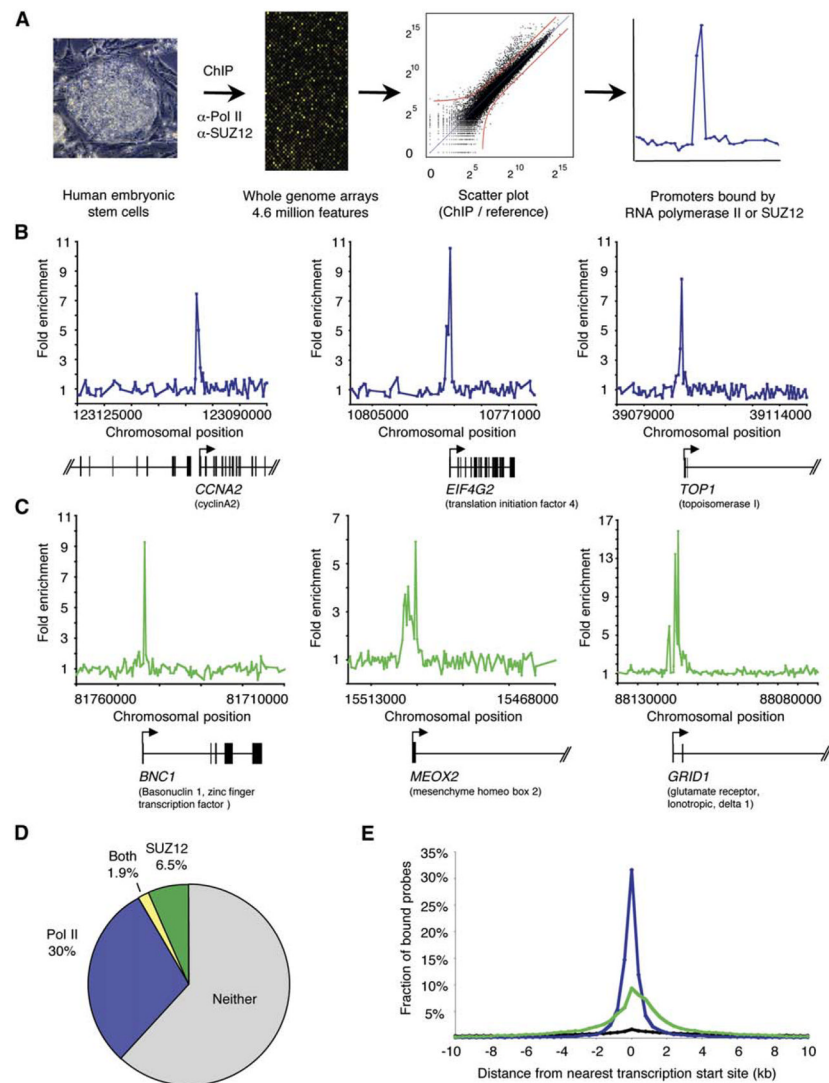
## References

- Abeyta MJ, Clark AT, Rodriguez RT, Bodnar MS, Pera RA, Firpo MT. Unique gene expression signatures of independently-derived human embryonic stem cell lines. *Hum Mol Genet.* 2004; 13:601–608. [PubMed: 14749348]
- Akasaka T, van Lohuizen M, van der Lugt N, Mizutani-Koseki Y, Kanno M, Taniguchi M, Vidal M, Alkema M, Berns A, Koseki H. Mice doubly deficient for the Polycomb Group genes *Mei18* and *Bmi1* reveal synergy and requirement for maintenance but not initiation of Hox gene expression. *Development.* 2001; 128:1587–1597. [PubMed: 11290297]
- Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* 2003; 17:126–140. [PubMed: 12514105]
- Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D. Ultraconserved elements in the human genome. *Science.* 2004; 304:1321–1325. [PubMed: 15131266]
- Bender M, Turner FR, Kaufman TC. A development genetic analysis of the gene regulator of postbithorax in *Drosophila melanogaster*. *Dev Biol.* 1987; 119:418–432. [PubMed: 3100363]
- Birve A, Sengupta AK, Beuchle D, Larsson J, Kennison JA, Rasmuson-Lestander A, Muller J. *Su(z)12*, a novel *Drosophila* Polycomb group gene that is conserved in vertebrates and plants. *Development.* 2001; 128:3371–3379. [PubMed: 11546753]
- Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell.* 2005; 122:947–956. [PubMed: 16153702]
- Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, Otte AP, Vidal M, Gifford DK, Young RA, Jaenisch R. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature.* 2006 in press.
- Brand-Saberi B. Genetic and epigenetic control of skeletal muscle development. *Ann Anat.* 2005; 187:199–207. [PubMed: 16130819]
- Breiling A, Turner BM, Bianchi ME, Orlando V. General transcription factors bind promoters repressed by Polycomb group proteins. *Nature.* 2001; 412:651–655. [PubMed: 11493924]
- Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science.* 2002; 298:1039–1043. [PubMed: 12351676]
- Cao R, Zhang Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol Cell.* 2004; 15:57–67. [PubMed: 15225548]

- Caretti G, Di Padova M, Micales B, Lyons GE, Sartorelli V. The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. *Genes Dev.* 2004; 18:2627–2638. [PubMed: 15520282]
- Chambers I, Colby D, Robertson M, Nichols J, Lee S, Tweedie S, Smith A. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell.* 2003; 113:643–655. [PubMed: 12787505]
- Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell.* 2002; 111:185–196. [PubMed: 12408863]
- Dellino GI, Schwartz YB, Farkas G, McCabe D, Elgin SC, Pirrotta V. Polycomb silencing blocks transcription initiation. *Mol Cell.* 2004; 13:887–893. [PubMed: 15053881]
- Denell RE, Frederick RD. Homoeosis in *Drosophila*: a description of the Polycomb lethal syndrome. *Dev Biol.* 1983; 97:34–47. [PubMed: 6404678]
- Duncan I. Control of bithorax complex functions by the segmentation gene fushi tarazu of *D. melanogaster*. *Cell.* 1986; 47:297–309. [PubMed: 2876778]
- Faust C, Lawson KA, Schork NJ, Thiel B, Magnuson T. The Polycomb-group gene *eed* is required for normal morphogenetic movements during gastrulation in the mouse embryo. *Development.* 1998; 125:4495–4506. [PubMed: 9778508]
- Francis NJ, Saurin AJ, Shao Z, Kingston RE. Reconstitution of a functional core polycomb repressive complex. *Mol Cell.* 2001; 8:545–556. [PubMed: 11583617]
- Franke A, DeCamillis M, Zink D, Cheng N, Brock HW, Paro R. Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of *Drosophila melanogaster*. *EMBO J.* 1992; 11:2941–2950. [PubMed: 1353445]
- Hamer KM, Sewalt RG, den Blaauwen JL, Hendrix T, Satijn DP, Otte AP. A panel of monoclonal antibodies against human polycomb group proteins. *Hybrid Hybridomics.* 2002; 21:245–252. [PubMed: 12193277]
- Hodgson JW, Argiropoulos B, Brock HW. Site-specific recognition of a 70-base-pair element containing d(GA)(n) repeats mediates bithoraxoid polycomb group response element-dependent silencing. *Mol Cell Biol.* 2001; 21:4528–4543. [PubMed: 11416132]
- Horard B, Tatout C, Poux S, Pirrotta V. Structure of a polycomb response element and in vitro binding of polycomb group complexes containing GAGA factor. *Mol Cell Biol.* 2000; 20:3187–3197. [PubMed: 10757803]
- Kennison JA. Introduction to Trx-G and Pc-G genes. *Methods Enzymol.* 2004; 377:61–70. [PubMed: 14979018]
- Kirmizis A, Bartley SM, Kuzmichev A, Margueron R, Reinberg D, Green R, Farnham PJ. Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes Dev.* 2004; 18:1592–1605. [PubMed: 15231737]
- Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* 2002; 16:2893–2905. [PubMed: 12435631]
- Kuzmichev A, Jenuwein T, Tempst P, Reinberg D. Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol Cell.* 2004; 14:183–193. [PubMed: 15099518]
- Kuzmichev A, Margueron R, Vaquero A, Preissner TS, Scher M, Kirmizis A, Ouyang X, Brockdorff N, Abate-Shen C, Farnham P, Reinberg D. Composition and histone substrates of poly-comb repressive group complexes change during cellular differentiation. *Proc Natl Acad Sci USA.* 2005; 102:1859–1864. [PubMed: 15684044]
- Lehmann OJ, Sowden JC, Carlsson P, Jordan T, Bhattacharya SS. Fox's in development and disease. *Trends Genet.* 2003; 19:339–344. [PubMed: 12801727]
- Levine SS, Weiss A, Erdjument-Bromage H, Shao Z, Tempst P, Kingston RE. The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol Cell Biol.* 2002; 22:6070–6078. [PubMed: 12167701]
- Levine SS, King IF, Kingston RE. Division of labor in polycomb group repression. *Trends Biochem Sci.* 2004; 29:478–485. [PubMed: 15337121]

- Lewis EB. A gene complex controlling segmentation in *Drosophila*. *Nature*. 1978; 276:565–570. [PubMed: 103000]
- Loebel DA, Watson CM, De Young RA, Tam PP. Lineage choice and differentiation in mouse embryos and embryonic stem cells. *Dev Biol*. 2003; 264:1–14. [PubMed: 14623228]
- Lund AH, van Lohuizen M. Polycomb complexes and silencing mechanisms. *Curr Opin Cell Biol*. 2004; 16:239–246. [PubMed: 15145347]
- Mayhall EA, Paffett-Lugassy N, Zon LI. The clinical potential of stem cells. *Curr Opin Cell Biol*. 2004; 16:713–720. [PubMed: 15530786]
- Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*. 2003; 113:631–642. [PubMed: 12787504]
- Molofsky AV, Pardal R, Morrison SJ. Diverse mechanisms regulate stem cell self-renewal. *Curr Opin Cell Biol*. 2004; 16:700–707. [PubMed: 15530784]
- Mulholland NM, King IF, Kingston RE. Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA. *Genes Dev*. 2003; 17:2741–2746. [PubMed: 14630938]
- Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA. Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell*. 2002; 111:197–208. [PubMed: 12408864]
- Nichols J, Zevnik B, Anastasiadis K, Niwa H, Klewe-Nebenius D, Chambers I, Scholer H, Smith A. Formation of pluri-potent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*. 1998; 95:379–391. [PubMed: 9814708]
- O'Carroll D, Erhardt S, Pagani M, Barton SC, Surani MA, Jenuwein T. The polycomb-group gene *Ezh2* is required for early mouse development. *Mol Cell Biol*. 2001; 21:4330–4336. [PubMed: 11390661]
- Orlando V, Paro R. Chromatin multiprotein complexes involved in the maintenance of transcription patterns. *Curr Opin Genet Dev*. 1995; 5:174–179. [PubMed: 7613086]
- Pasini D, Bracken AP, Jensen MR, Denchi EL, Helin K. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J*. 2004; 23:4061–4071. [PubMed: 15385962]
- Pera MF, Trounson AO. Human embryonic stem cells: prospects for development. *Development*. 2004; 131:5515–5525. [PubMed: 15509763]
- Pirrotta V. Polycomb-ing the genome: PcG, trxG, and chromatin silencing. *Cell*. 1998; 93:333–336. [PubMed: 9590168]
- Reubinoff BE, Pera MF, Fong CY, Trounson A, Bongso A. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nat Biotechnol*. 2000; 18:399–404. [PubMed: 10748519]
- Ringrose L, Paro R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet*. 2004; 38:413–443. [PubMed: 15568982]
- Sato N, Sanjuan IM, Heke M, Uchida M, Naef F, Brivanlou AH. Molecular signature of human embryonic stem cells and its comparison with the mouse. *Dev Biol*. 2003; 260:404–413. [PubMed: 12921741]
- Sato N, Meijer L, Skaltsounis L, Greengard P, Brivanlou AH. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat Med*. 2004; 10:55–63. [PubMed: 14702635]
- Saurin AJ, Shao Z, Erdjument-Bromage H, Tempst P, Kingston RE. A *Drosophila* Polycomb group complex includes Zeste and dTAFII proteins. *Nature*. 2001; 412:655–660. [PubMed: 11493925]
- Schepers GE, Teasdale RD, Koopman P. Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. *Dev Cell*. 2002; 3:167–170. [PubMed: 12194848]
- Shao Z, Raible F, Mollaaghababa R, Guyon JR, Wu CT, Bender W, Kingston RE. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell*. 1999; 98:37–46. [PubMed: 10412979]
- Showell C, Binder O, Conlon FL. T-box genes in early embryogenesis. *Dev Dyn*. 2004; 229:201–218. [PubMed: 14699590]

- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 2005; 15:1034–1050. [PubMed: 16024819]
- Simon J, Chiang A, Bender W. Ten different Polycomb group genes are required for spatial control of the *abdA* and *AbdB* homeotic products. *Development.* 1992; 114:493–505. [PubMed: 1350533]
- Strutt H, Cavalli G, Paro R. Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J.* 1997; 16:3621–3632. [PubMed: 9218803]
- Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA.* 2004; 101:6062–6067. [PubMed: 15075390]
- Tapscott SJ. The circuitry of a master switch: MyoD and the regulation of skeletal muscle gene transcription. *Development.* 2005; 132:2685–2695. [PubMed: 15930108]
- Thompson NE, Steinberg TH, Aronson DB, Burgess RR. Inhibition of in vivo and in vitro transcription by monoclonal antibodies prepared against wheat germ RNA polymerase II that react with the heptapeptide repeat of eukaryotic RNA polymerase II. *J Biol Chem.* 1989; 264:11511–11520. [PubMed: 2472398]
- Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. Embryonic stem cell lines derived from human blastocysts. *Science.* 1998; 282:1145–1147. [PubMed: 9804556]
- Tie F, Furuyama T, Prasad-Sinha J, Jane E, Harte PJ. The *Drosophila* Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. *Development.* 2001; 128:275–286. [PubMed: 11124122]
- van der Lugt NM, Alkema M, Berns A, Deschamps J. The Polycomb-group homolog Bmi-1 is a regulator of murine Hox gene expression. *Mech Dev.* 1996; 58:153–164. [PubMed: 8887324]
- Vire E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, Morey L, Van Eynde A, Bernard D, Vanderwinden JM, Bollen M, Esteller M, Di Croce L, de Launoit Y, Fuks F. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature.* 2006; 439:871–874. [PubMed: 16357870]
- Wang J, Mager J, Schnedier E, Magnuson T. The mouse PcG gene *eed* is required for Hox gene repression and extra-embryonic development. *Mamm Genome.* 2002; 13:493–503. [PubMed: 12370779]
- Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, et al. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* 2005; 3:e7. [PubMed: 15630479]



### Figure 1. Genome-Wide ChIP-Chip in Human Embryonic Stem Cells

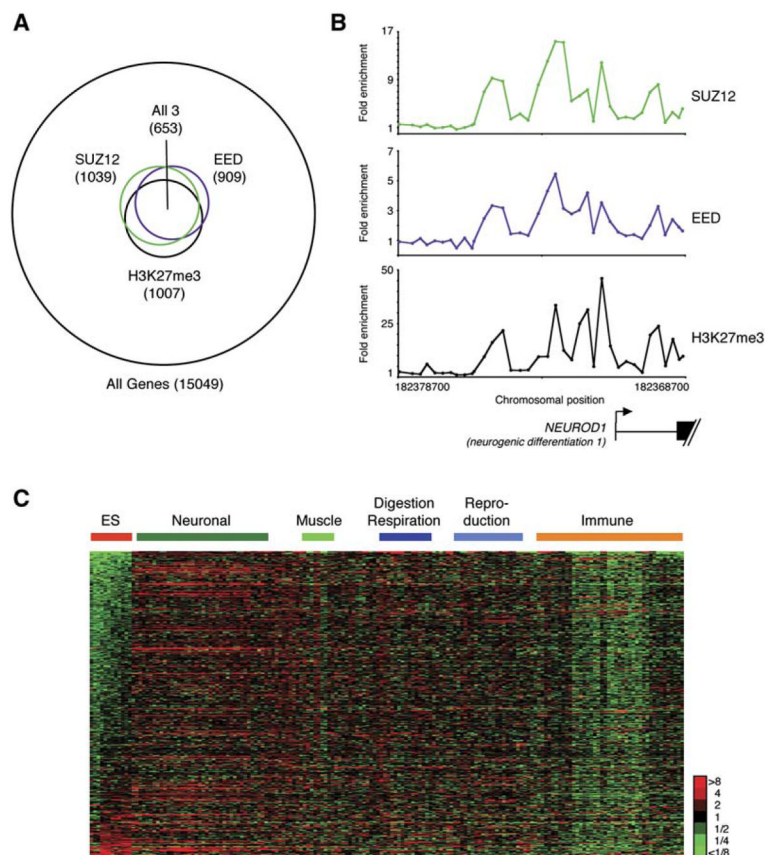
(A) DNA segments bound by the initiation form of RNA polymerase II or SUZ12 were isolated using chromatin-immunoprecipitation (ChIP) and identified with DNA microarrays containing over 4.6 million unique 60-mer oligonucleotide probes spanning the entire nonrepeat portion of the human genome. ES cell growth and quality control, the antibodies, ChIP protocol, DNA microarray probe design, and data analysis methods are described in detail in Supplemental Data.

(B) Examples of RNA polymerase II ChIP signals from genome-wide ChIP-Chip. The plots show unprocessed enrichment ratios (blue) for all probes within a genomic region (ChIP versus whole genomic DNA). Chromosomal positions are from NCBI build 35 of the human genome. Genes are shown to scale below plots (exons are represented by vertical bars). The start and direction of transcription are noted by arrows.

(C) Examples of SUZ12 ChIP signals from genome-wide ChIP-Chip. The plots show unprocessed enrichment ratios (green) for all probes within a genomic region (ChIP versus whole genomic DNA). Chromosomal positions, genes, and notations are as described in (B).

(D) Chart showing percentage of all annotated genes bound by RNA polymerase II (blue), SUZ12 (green), both (yellow), or neither (gray).

(E) Distribution of the distance between bound probes and the closest transcription start sites from RefSeq, Ensembl, MGC, UCSC Known Genes and H-Inv databases for SUZ12 (green line), and RNA polymerase II (blue line). The number of bound probes is given as the percentage of total probes and is calculated for 400 bp intervals from the start site. The null-distribution of the distance between all probes and the closest transcription are shown as a black line.



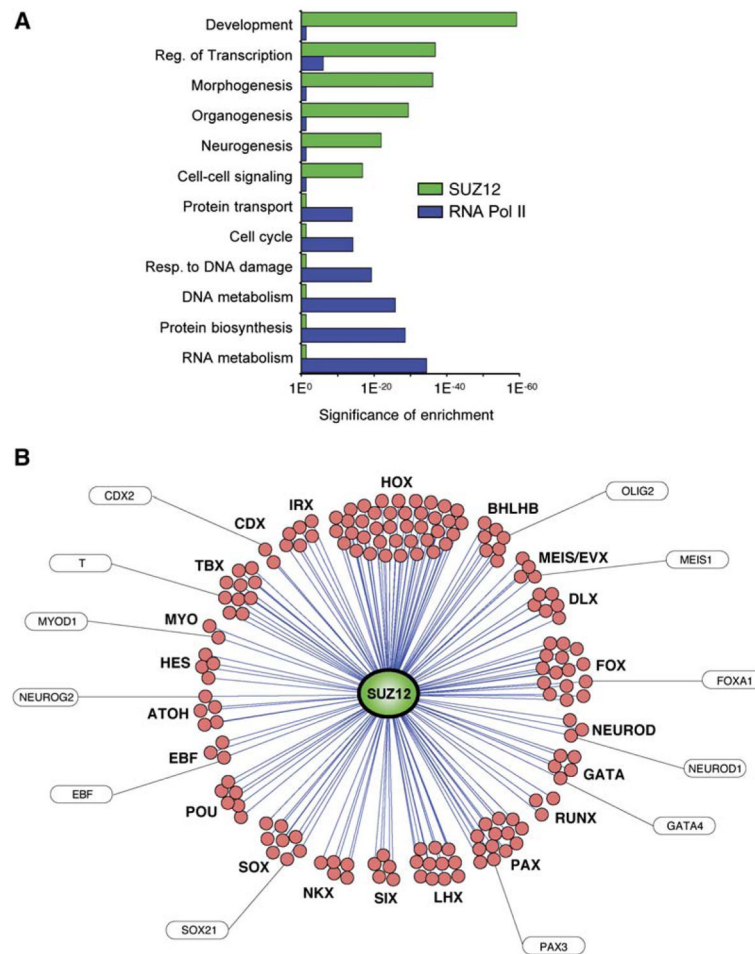
**Figure 2. SUZ12 Is Associated with EED, histone H3K27me3 Modification, and Transcriptional Repression in ES Cells**

(A) Venn diagram showing the overlap of genes bound by SUZ12 at high-confidence, genes bound by EED at high-confidence, and genes trimethylated at H3K27 at high-confidence.

The data are from promoter micro-arrays that contain probes tiling  $-8$  kb and  $+2$  kb around transcription start. 72% of the genes bound by SUZ12 at high-confidence are also bound by EED at high-confidence; others are bound by EED at lower confidence (Figure S6).

(B) SUZ12 (top), EED (middle), and H3K27me3 (bottom) occupancy at *NEUROD1*. The plots show unprocessed enrichment ratios for all probes within this genomic region (SUZ12 ChIP versus whole genomic DNA, EED ChIP versus whole genomic DNA, and H3K27me3 ChIP versus total H3 ChIP). Chromosomal positions are from NCBI build 35 of the human genome. *NEUROD1* is shown to scale below plots (exons are represented by vertical bars). The start and direction of transcription are noted by arrows.

(C) Relative expression levels of 604 genes occupied by PRC2 and trimethylated at H3K27 in ES cells. Comparisons were made across four ES cell lines and 79 differentiated cell types. Each row corresponds to a single gene that is bound by SUZ12, associated with EED and H3K27me3, and for which Affymetrix expression data are available. Each column corresponds to a single expression microarray. ES cells are in the following order: H1, H9, HSF6, HSF1. For each gene, expression is shown relative to the average expression level of that gene across all samples, with shades of red indicating higher than average expression and green lower than average expression according to the scale on the right. Cell types are grouped by tissue or organ function, and genes are ranked according to the significance of their relative level of gene expression in ES cells.

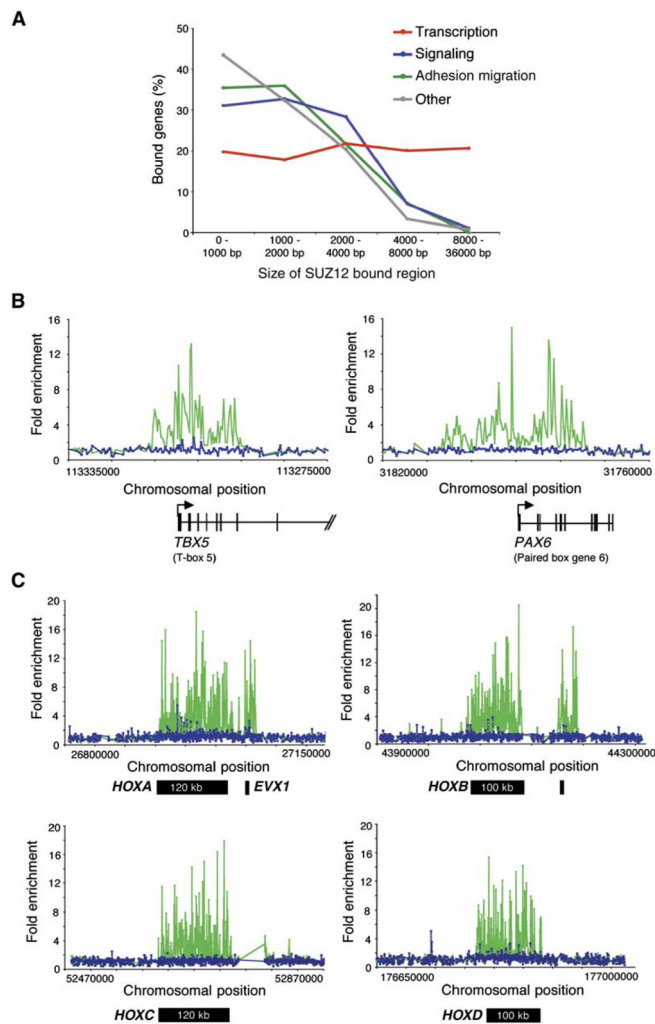


### Figure 3. Cellular Functions of Genes Occupied by SUZ12

(A) Genes bound by SUZ12 or RNA polymerase II were compared to biological process gene ontology categories; highly represented categories are shown. Ontology terms are shown on the *y* axis; p-values for the significance of enrichment are graphed along the *x* axis (SUZ12 in green, RNA polymerase II in blue).

(B) Selected examples of developmental transcription factor families bound by SUZ12. SUZ12 is represented by the green oval; individual transcription factors are represented by circles and grouped by family as indicated. Examples of transcription factors with defined roles in development are labeled. Transcription factor families include homeobox protein (HOX), basic helix-loop-helix domain containing, class B (BHLHB), HOX cofactors (MEIS/EVX), distal-less homeobox (DLX), Forkhead box (FOX), NEUROD, GATA binding protein (GATA), runt related transcription factor (RUNX), paired box and paired-like (PAX), LIM homeobox (LHX), sine oculis homeobox homolog (SIX), NK transcription factor related (NKX), SRY box (SOX), POU domain containing, classes 3 and 4 (POU), early B-cell factor (EBF), atonal homolog (ATOH), hairy and enhancer of split protein (HES), myogenic basic domain (MYO), T-box (TBX), caudal type homeobox (CDX), and iroquois homeobox protein (IRX).



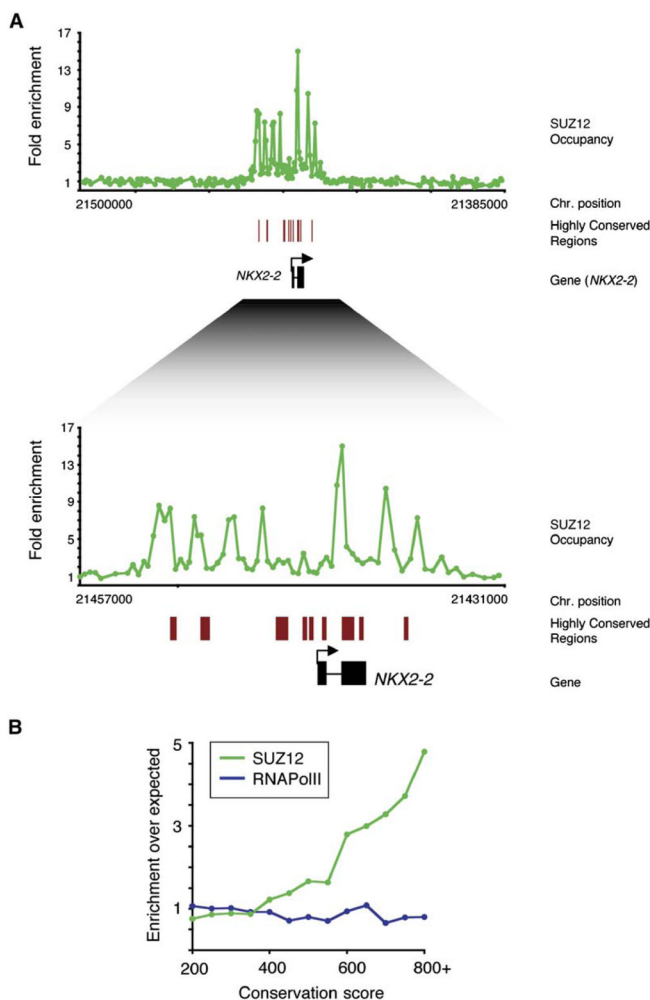


**Figure 4. SUZ12 Occupies Large Portions of Genes Encoding Transcription Factors with Roles in Development**

(A) The fraction of SUZ12 target genes associated with different sizes of binding domains. Genes are grouped into four categories according to their function: Signaling, Adhesion/migration, Transcription, and Other.

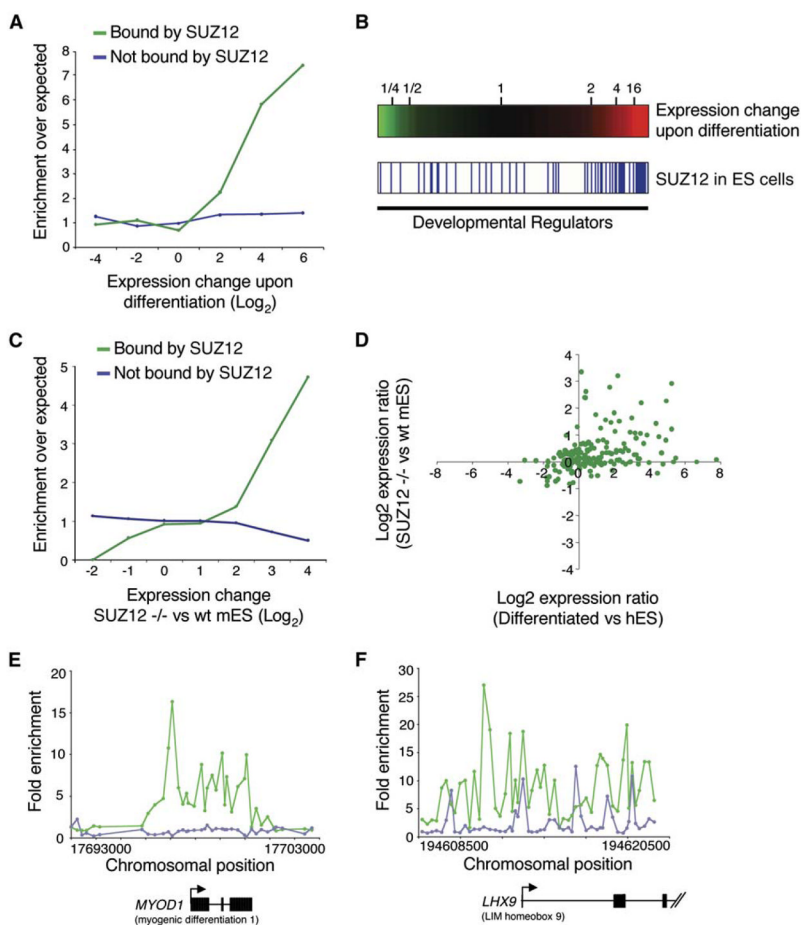
(B) Examples of SUZ12 (green) and RNA poly-merase II (blue) binding at the genes encoding developmental regulators TBX5 and PAX6. The plots show unprocessed enrichment ratios for all probes within a genomic region (ChIP versus whole genomic DNA). Genes are shown to scale below plots (exons are represented by vertical bars). The start and direction of transcription are noted by arrows.

(C) Binding profiles of SUZ12 (green) and RNA polymerase II (blue) across  $\sim 500$  kb regions encompassing *HOX* clusters A–D. Unprocessed enrichment ratios for all probes within a genomic region are shown (ChIP versus whole genomic DNA). Approximate *HOX* cluster region sizes are indicated within black bars.



### Figure 5. SUZ12 Binding Is Associated with Highly Conserved Regions

(A) SUZ12 occupancy (green) and conserved elements are shown at *NKX2-2* and adjacent genomic regions. The plots show unprocessed enrichment ratios for all probes within this genomic region (SUZ12 ChIP versus whole genomic DNA). Conserved elements (red) with LoD scores > 160 derived from the PhastCons program (Siepel et al., 2005) are shown to scale above the plot. Genes are shown to scale below plots (exons are represented by vertical bars). A higher resolution view is also shown below. (B) Enrichment of conserved noncoding elements within SUZ12 (green) and RNA polymerase II (blue) bound regions. The maximum nonexonic PhastCons conservation score was determined for each bound region. For comparison, the same parameter was determined using a randomized set of genomic regions with the same size distribution. The graph displays the ratio of the number of bound regions with that score versus the number of randomized genomic regions with that score.



### Figure 6. Preferential Activation of PRC2 Target Genes during ES Cell Differentiation

(A) Fold enrichment in the number of genes induced or repressed during ES cell differentiation. The change in gene expression is given as the  $\log_2$  transformed ratio of the signals in differentiated H1 cells versus pluripotent H1 cells and is binned into six groups. The upper limit of each bin is indicated on the  $x$  axis. The two lines show genes transcriptionally inactive in ES cells (absence of RNA polymerase II) and bound by SUZ12 (green) and genes transcriptionally inactive in ES cells and repressed by other means (blue). In both cases, fold enrichment is calculated against the total population of genes and normalized for the number of genes present in each group.

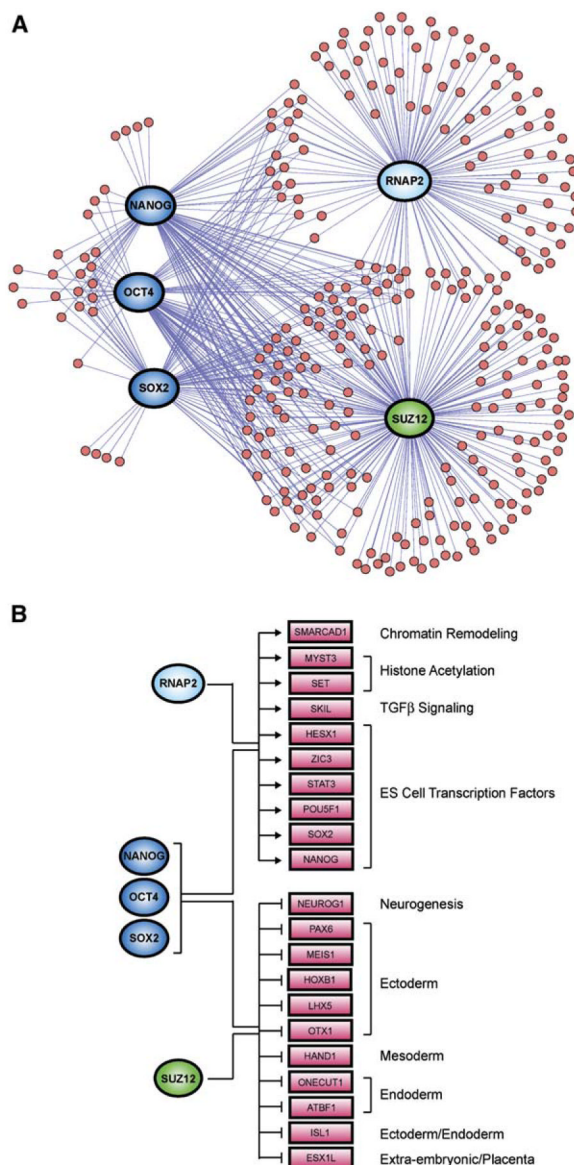
(B) Expression changes of genes encoding developmental regulators during ES cell differentiation. Expression ratio (differentiated/pluripotent) is represented by color, with shades of red indicating upregulation and shades of green downregulation according to the scale shown above. Genes are ordered according to change in gene expression, with genes exhibiting higher expression in pluripotent ES cells to the left and genes exhibiting higher expression in differentiated cells to the right. Genes bound by SUZ12 in undifferentiated ES cells are indicated by blue lines in the lower panel.

(C) Fold enrichment in the number of genes induced or repressed in SUZ12-deficient mouse cells. The change in gene expression is given as the  $\log_2$  transformed ratio of the signals in *Suz12*-deficient cells versus wild-type ES cells. The two lines show genes transcriptionally inactive in human ES cells (absence of RNA polymerase II) and bound by SUZ12 (green) and genes transcriptionally inactive in human ES cells and repressed by other means (blue). In both cases, fold enrichment is calculated against the total population of genes.

(D) Gene expression ratios (log base 2) of Suz12 target genes in differentiated human H1 ES cells relative to pluripotent H1 ES cells ( $x$  axis) and in Suz12-deficient mouse cells relative to wild-type mouse ES cells ( $y$  axis). Upper right quadrant: genes upregulated during human ES cell differentiation and in Suz12-deficient mouse cells; lower right: genes upregulated during ES cell differentiation and downregulated in Suz12-deficient cells; lower left: genes downregulated during ES cell differentiation and in Suz12-deficient cells; upper left: genes downregulated during ES cell differentiation and upregulated in Suz12-deficient cells.

(E) SUZ12 binding profiles across the gene encoding muscle regulator MYOD1 in H9 human ES cells (green) and primary human skeletal myotubes (gray). The plots show unprocessed enrichment ratios for all probes within a genomic region (ChIP versus whole genomic DNA). Genes are shown to scale below plots (exons are represented by vertical bars). The start and direction of transcription are noted by arrows.

(F) Suz12 binding profiles across the gene encoding LHX9 in H9 human ES cells (green) and primary human skeletal myotubes (gray). The plots show unprocessed enrichment ratios for all probes within a genomic region (ChIP versus whole genomic DNA). Genes are shown to scale below plots (exons are represented by vertical bars). The start and direction of transcription are noted by arrows.



**Figure 7. SUZ12 Is Localized to Genes also Bound by ES Cell Transcriptional Regulators**  
 (A) Transcriptional regulatory network model of developmental regulators governed by OCT4, SOX2, NANOG, RNA polymerase II, and SUZ12 in human ES cells. The ES cell transcription factors each bound to approximately one-third of the PRC2-occupied, developmental transcription factor genes. Developmental regulators were selected based on gene ontology. Regulators are represented by dark blue circles; RNA polymerase II is represented by a light blue circle; SUZ12 is represented by a green circle; gene promoters for developmental regulators are represented by small red circles.  
 (B) SUZ12 occupies a set of repressed developmental regulators also bound by OCT4, SOX2, and NANOG in human ES cells. Genes annotated as bound by OCT4, SOX2, and NANOG previously and identified as active or repressed based on expression data (Boyer et al., 2005) were tested to see if they were bound by SUZ12 or RNA polymerase II. Ten of eleven previously identified active genes were found to be bound by RNA polymerase II at known promoters, while eleven of twelve previously identified repressed genes were bound

by SUZ12. Regulators are represented by dark blue circles, RNA polymerase II by a light blue circle, and SUZ12 by a green circle. Gene promoters are represented by red rectangles.