

Latent Regulatory Potential of Human-Specific Repetitive Elements

Michelle C. Ward,^{1,5} Michael D. Wilson,^{1,5,6,*} Nuno L. Barbosa-Morais,² Dominic Schmidt,¹ Rory Stark,¹ Qun Pan,² Petra C. Schwalie,³ Suraj Menon,¹ Margus Lukk,¹ Stephen Watt,¹ David Thybert,³ Claudia Kutter,¹ Kristina Kirschner,¹ Paul Flicek,^{3,4} Benjamin J. Blencowe,² and Duncan T. Odom^{1,4,*}

¹University of Cambridge, Cancer Research UK-Cambridge Institute, Robinson Way, Cambridge CB2 0RE, UK

²Banting and Best Department of Medical Research and Department of Molecular Genetics, Donnelly Centre, Toronto, ON M5S 3E1, Canada

³European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton CB10 1SD, UK

⁴Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK

⁵These authors contributed equally to this work

⁶Present address: SickKids Research Institute and Department of Molecular Genetics, University of Toronto, Toronto, ON M5G 1L7, Canada

*Correspondence: michael.wilson@sickkids.ca (M.D.W.), duncan.odom@cruk.cam.ac.uk (D.T.O.)

<http://dx.doi.org/10.1016/j.molcel.2012.11.013>

Open access under [CC BY license](#).

SUMMARY

At least half of the human genome is derived from repetitive elements, which are often lineage specific and silenced by a variety of genetic and epigenetic mechanisms. Using a transchromosomal mouse strain that transmits an almost complete single copy of human chromosome 21 via the female germline, we show that a heterologous regulatory environment can transcriptionally activate transposon-derived human regulatory regions. In the mouse nucleus, hundreds of locations on human chromosome 21 newly associate with activating histone modifications in both somatic and germline tissues, and influence the gene expression of nearby transcripts. These regions are enriched with primate and human lineage-specific transposable elements, and their activation corresponds to changes in DNA methylation at CpG dinucleotides. This study reveals the latent regulatory potential of the repetitive human genome and illustrates the species specificity of mechanisms that control it.

INTRODUCTION

Between one-half and possibly up to two-thirds of the human genome is derived from repetitive sequences, most of which are classified as transposable elements (TEs) (de Koning et al., 2011). TEs can serve as regulatory DNA contributing to tissue-specific transcriptional evolution (Bourque et al., 2008; Faulkner et al., 2009; Lowe et al., 2007; Oliver and Greene, 2009), and their activity has altered the regulatory circuitry of embryonic stem cells (Kunarso et al., 2010), mammalian pregnancy pathways (Lynch et al., 2011; Xie et al., 2010), and the deployment of CTCF binding sites across mammalian genomes (Bourque et al., 2008; Schmidt et al., 2012). The rapid increase in sequenced mammalian genomes (Lindblad-Toh et al., 2011), in vivo multivertebrate transcription factor binding maps

(Kunarso et al., 2010; Schmidt et al., 2010, 2012), and computational tools to dissect repeat-based genomes (Treangen and Salzberg, 2012) has uncovered lineage-specific genome innovations whose biological functions are not known.

Relative to the mobility of TEs in species such as maize, *Drosophila*, mice, and some primates (Maksakova et al., 2006), the activity of retrotransposons has declined in hominids (Lander et al., 2001), yet TEs continue to shape the human genome. The insertion of TEs underlies at least 65 human diseases (Cordaux and Batzer, 2009; Goodier and Kazazian, 2008) and can provide a substrate for nonallelic homologous recombination, resulting in structural changes (Beck et al., 2011). As transposons rapidly acquire mutations and as their activity can damage a genome, multiple mechanisms have evolved to silence them in mammals (Levin and Moran, 2011), including specific histone modifications, DNA methylation, and targeted small RNAs (De Fazio et al., 2011; Maksakova et al., 2008; Rebollo et al., 2011; Reuter et al., 2011).

One consequence of the arms race between transposons and transposon silencing mechanisms is that the regulatory potential of a transposon-derived sequence is difficult to evaluate in its host which has coevolved mechanisms to repress it. Thus, functional in vivo studies of TEs often employ heterologous strategies such as placing human retroelements into other vertebrate species including mouse, rat (Kano et al., 2009), and zebrafish (Pi et al., 2004). Using an aneuploid mouse that stably transmits a majority (42 of 46.9 Mb) of human chromosome 21 (HsChr21) through the female germline (Tc1) (O'Doherty et al., 2005), we previously demonstrated that the nonrepetitive portion of mammalian genomes is largely transcriptionally directed by *cis*-acting regulatory elements (Wilson et al., 2008). This study employed genome-tiling microarrays, which by design did not permit the analysis of repetitive regions. Using high-throughput sequencing, which allows us to explore a greater percentage of the TE-derived genome, we here explore the in vivo transcriptional control of a human chromosome placed in a heterologous mouse environment. This unique system revealed unexpected regulatory and transcriptional potential in many recently evolved human sequences which are associated with changes in DNA methylation and chromatin state.

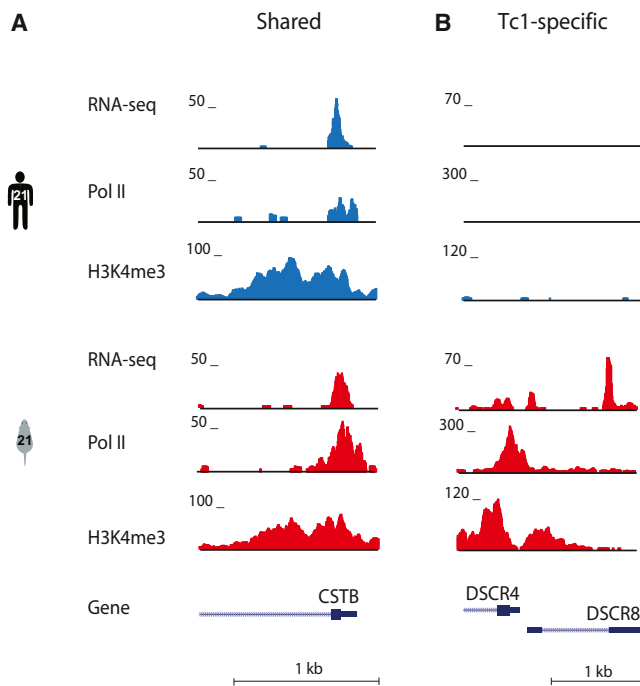


Figure 1. In a Mouse Carrying Human Chromosome 21, Most Locations of Human Chromosome 21 Are Transcribed in Liver Largely as in Human Tissues, Yet Specific Loci Show Differences

(A) High-throughput sequencing of chromatin immunoprecipitations and poly(A) mRNA enrichment revealed that the *CSTB* locus is occupied by RNA polymerase II, enriched for H3K4me3, and transcribed into RNA in both human and Tc1 mouse liver.

(B) The *DSCR4/8* locus on HsChr21 shows similar evidence of transcription in Tc1 mouse liver tissue, which is not evident in normal human liver.

RESULTS

Many HsChr21 Regions Transcriptionally Active in Tc1 Mouse Tissues Are Silent in Human

To compare in vivo gene regulation of both repetitive and non-repetitive regions of HsChr21 between human and Tc1 mouse livers, we experimentally profiled the following using high-throughput sequencing methodologies: (1) polyadenylated (poly[A])-containing mRNA transcripts, (2) regions enriched for trimethylation of H3K4 (H3K4me3) as a marker for transcriptional initiation (Bernstein et al., 2005; Guenther et al., 2007; Heintzman et al., 2007), and (3) genomic occupancy of the RNA polymerase II (Pol II) basal machinery (Figure 1; Experimental Procedures; see Table S1A, Document S2, and Figure S1 online).

Using H3K4me3 as a proxy for transcription initiation, we found that most regions activated on HsChr21 (214/383) were not significantly different in quality and quantity between the two species (we henceforth refer to these sites as Shared; Figure 1A, Figure S2A, Experimental Procedures). By using sequencing-based methods instead of microarrays, we also identified specific human sequence locations uniquely activated in the mouse nucleus. For example, in Tc1 mouse livers, the bidirectional promoter of the Down's Syndrome critical regions

4 and 8 (*DSCR4* and *DSCR8*) is bound by Pol II, enriched for H3K4me3, and generates poly(A) transcripts (Figure 1B). Normally, these two long noncoding RNAs are transcriptionally driven in human placenta by a primate-specific LTR retrotransposon but are silenced in liver (Dunn et al., 2006).

In total, we identified 118 regions on HsChr21 enriched for trimethylation of H3K4 in Tc1 mouse liver compared to human (henceforth Tc1-specific), which we defined as having at least 4-fold greater normalized read counts, with an FDR of <0.1 (Experimental Procedures, Figure 2A, Figures S2A and S2B). We confirmed that the Tc1-specific regions were robust and could not be explained by misalignment of sequencing reads with control experiments including paired-end mapping of chromatin immunoprecipitation (ChIP) experiments (Table S1A), different alignment strategies (Tables S1B–S1D, Figure S2C), and were also validated by qPCR (Table S1E and Figure S2D).

Only 51 locations on HsChr21 activated in human liver were found enriched relative to Tc1 mouse liver (henceforth Human-specific) using these criteria. Notably, in contrast to most Tc1-specific activated regions, which largely appeared unique to the mouse, regions identified as containing H3K4me3 specifically in human often showed measurable signal in the Tc1 mouse (Figure 2, Figures S2E and S2F).

The design and sensitivity of genome tiling microarrays used by the prior study that compared human and Tc1 mouse liver gene regulation (Wilson et al., 2008) prevented the identification of the majority of these Tc1-specific regions, discovered here by ChIP-seq. Because repetitive sequences are excluded from microarray design (Bertone et al., 2006), 40% of the Tc1-specific regions ($n = 47$) overlapped less than three microarray probes, which was the minimum criteria to identify regions bound by ChIP experiments. Indeed, only 907 of the 74,901 chromosome 21 probes on the Agilent microarray overlap with sequences in the RepeatMasker library. Thirty-seven percent ($n = 44$) of the Tc1-specific regions that overlapped three or more probes had insufficient signal to be called in either species in our previous study. Of the remaining 27 Tc1-specific H3K4me3 regions that were identified on microarrays, only nine were classified as unique to the Tc1 mouse. In summary, our current study validates over 94% of the Shared H3K4me3 regions originally identified by ChIP-chip; however, due to the increased coverage and sensitivity afforded by sequencing, we newly identify Tc1-specific genomic regions associated with activated chromatin and repetitive elements.

Consequences of Tc1-Specific Transcriptional Activation

H3K4me3 has been shown to serve as an anchorage for the basal transcriptional machinery (Bernstein et al., 2005; Guenther et al., 2007; Heintzman et al., 2007; Vermeulen et al., 2007). We looked for evidence for transcriptional activation near HsChr21 regions associated with H3K4me3. Tc1-specific H3K4me3 regions both have higher overlap with Pol II in Tc1 mouse liver than in human liver (74% versus 28%, respectively; Figure S2G) and are located near genes with higher gene expression in the Tc1 mouse (Figure 2 and Figure S2G). Many Tc1-specific regions

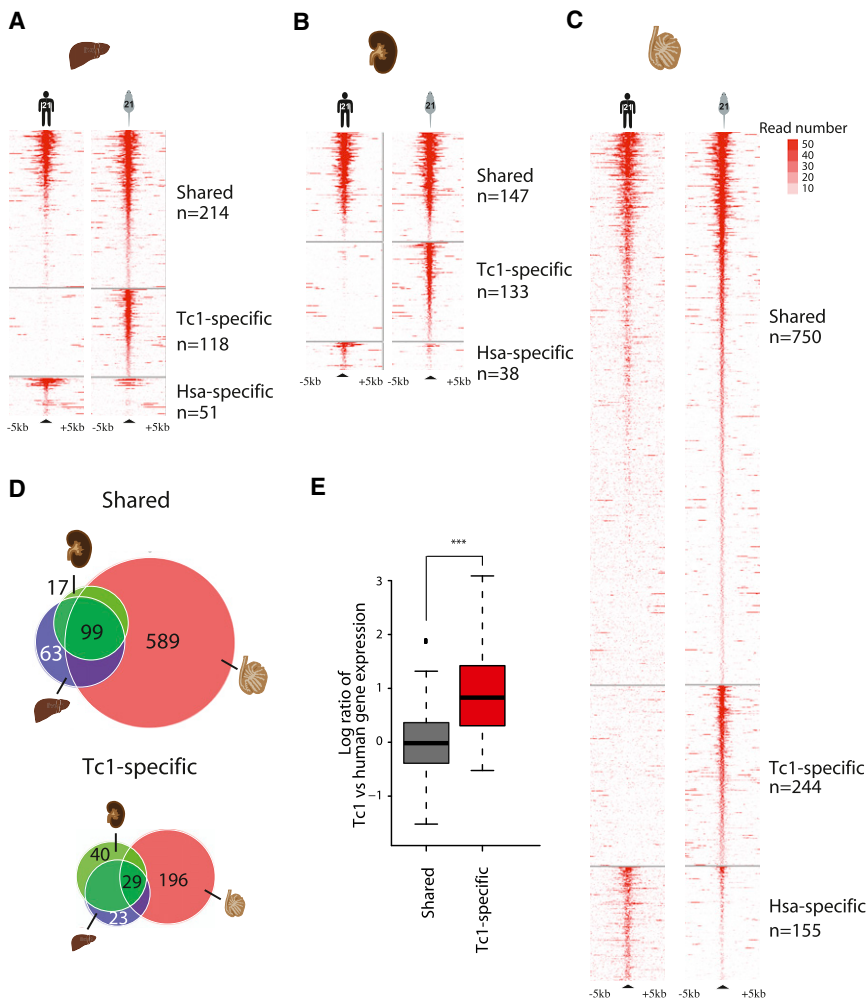


Figure 2. Tc1-Specific Transcription Initiation Occurs in Somatic and Germline Tissues and Influences Gene Expression

(A–C) Chromosome-wide identification of all regions enriched for H3K4me3 on HsChr21 in mouse and human liver, kidney, and testes; heatmap is sorted by descending signal strength and by species specificity.

(D) Similar sets of loci on HsChr21 are enriched for H3K4me3 in both mouse and human, and these are largely a subset of loci found in testes in both species. In liver and kidney, Tc1-specific loci enriched for H3K4me3 were found in similar locations, which were distinct from those found in Tc1 mouse testes.

(E) H3K4me3-enriched sites in the Shared and Tc1-specific categories in liver were associated with the nearest transcription start site and the expression levels of genes on Tc1-HsChr21 determined relative to human genes (Shared, n = 86; Tc1 specific, n = 17) (***) (***p ≤ 0.0005, one-sided Mann-Whitney U test).

were identified on HsChr21 in testes from either Tc1 mouse (n = 994) or human (n = 905), compared to somatic tissues (Figure 2C). In testes, the majority of regions enriched in H3K4me3 were shared between human and mouse (n = 750) and appear to be a superset encompassing those found in liver and kidney (Figure 2D). The regions enriched for H3K4me3 specifically in Tc1 mouse testes were largely distinct from those found specifically in Tc1 mouse liver and kidney (Figure 2D and Table S3). In sum, our data indicate that latent regulatory

regions in the human genome can become transcriptionally activated in somatic and germline tissues of a heterologous species.

Young, Primate-Specific Repeats Can Be Transcriptionally Activated in the Tc1 Mouse

We asked whether the TE composition differed between the Tc1-specific H3K4me3 regions and those H3K4me3 regions that showed no significant differences between species. To do this, we collected all the repeat elements that were significantly enriched for H3K4me3 in liver, kidney, and testes and then subdivided them based on whether they were shared between species or unique to the Tc1 mouse (Figures 3A–3C). We found that repeat elements were significantly more likely to be specifically enriched for H3K4me3 in Tc1 liver and kidney tissues (p = 7.6 × 10^{−9} and 6.7 × 10^{−9}, respectively) (Figures 3A and 3B, leftmost panels). In contrast to the somatic tissues, the testes do not show significant repetitive element enrichment between Tc1-specific and Shared H3K4me3 regions (p = 0.27; Tables S2A and S2B) (Figure 3C, leftmost panel).

To identify whether specific types of repeats were responsible for the transcriptional activation identified in the somatic tissues of the Tc1 mouse, we further subdivided the H3K4me3 regions

have therefore resulted in changes in gene expression, highlighting their latent regulatory potential.

We asked whether the transcriptional activation observed in Tc1 mouse liver also occurred in other somatic tissues. We mapped H3K4me3-associated DNA regions in Tc1 mouse and human kidney samples and identified a similar set of Tc1-specific transcriptionally active regions, many of which were shared with liver (Figure 2B and Table S3). H3K4me3 profiling in Tc1 mouse brain, spleen, and muscle tissues revealed that at least 50% of Tc1-specific liver regions are enriched across all profiled somatic tissues (Table S3 and Figure S2H). Thus, the occurrence of Tc1-specific H3K4me3 regions on HsChr21 occurs broadly across somatic mouse tissues.

Relative to somatic tissues, testes transcribe a greater proportion of the mouse genome (Shima et al., 2004), and this transcription is accompanied by an increased number of H3K4me3-associated regions (Smagulova et al., 2011). We tested the possibility that this global transcriptional upregulation in germline tissues might further unmask additional regulatory regions in the human genome. Consistent with these reports, more than twice as many H3K4me3-associated regions

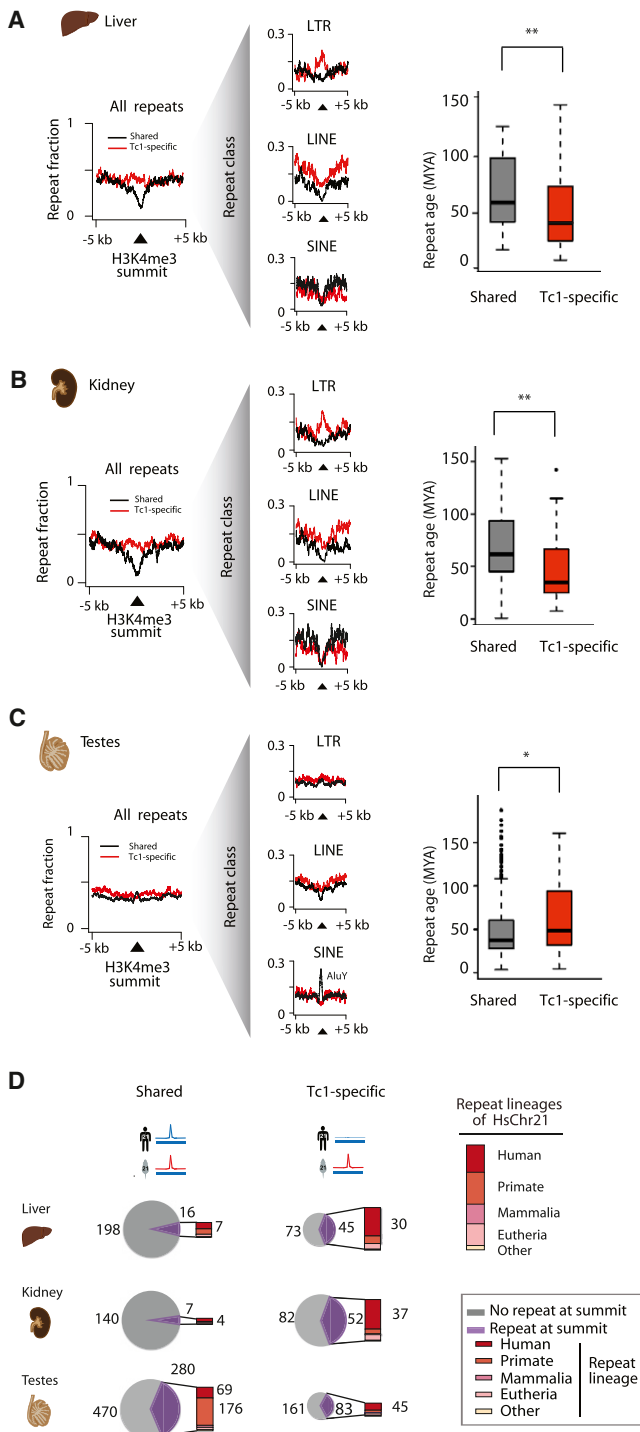


Figure 3. Tc1-Specific Locations of Transcription Initiation Are Enriched for Young, Lineage-Specific Repetitive Elements

(A) Fraction of repeat elements within a 10 kb window around the H3K4me3 peak summit in Shared (black) and Tc1-specific (red) events in liver. Shown is age of the repeats in Tc1-specific and Shared sites as determined by nucleotide substitution rates of repeat instances at H3K4me3 peak summits.
 (B) As in (A) for kidney.
 (C) As in (A) for testes.

by their component repeat class (Figures 3A–3C, middle panels). Most notably, LTR elements showed a clear, Tc1-specific enrichment of H3K4me3 in the liver and kidney ($p < 10^{-5}$). Similarly, H3K4me3 in Tc1 mouse liver and kidney was enriched at LINE repeats, despite a more diffuse architecture ($p < 10^{-3}$) (Figures 3A and 3B; Tables S2A and S2B, Figure S3A, Document S2). Five of the 20 human transposable SVA elements on chromosome 21 (Tc1-HsChr21) were significantly enriched for Tc1-specific H3K4me3 in liver (and 6/20 in kidney) ($p < 0.03$) (Table S2B).

In total, there are 41,877 repeat instances across Tc1-HsChr21, of which 1,043 were associated with H3K4me3. The distribution of the repeat classes captured in these H3K4me3 regions is similar to the repeat class distribution in the entire Tc1 HsChr21 chromosome (9% versus 8%, respectively, for DNA elements; 33% versus 22% for LINEs; 23% versus 25% LTRs; and 35% versus 43% for SINEs) (Tables S2C, Document S2). This suggests that the observed Tc1-specific repeat enrichments are not due to biases in repeat content in promoter regions.

At least 60% of Tc1-specific H3K4me3-associated regions enriched for repeats were identified in all somatic tissues profiled (Table S3). Of the most significantly enriched repeat types identified in the Tc1 mouse liver, the LTR elements LTR12C (10/11 on Tc1-HsChr21) and LTR12D (5/5 on Tc1-HsChr21) were constitutively activated across kidney, brain, muscle, and spleen Tc1 mouse tissues (Table S3A and Figure 4). Other significant liver-enriched repeats of the LINE, SVA, and SINE classes showed variable activation across tissues (Table S3).

Many of these repeat elements were primate and human lineage-specific and were transcriptionally silent in human liver and kidney (Figure 3D, Figure S3B). Indeed, based on analysis of the nucleotide substitution and mutation rates, these repetitive elements are significantly younger than those shared between human and Tc1 mice (Experimental Procedures, Figures 3A and 3B, righthand panels; Figure S3A), consistent with a mechanism wherein mice may lack the regulatory machinery needed to silence human-specific repetitive elements.

In the testes, the repeat enrichments were not significantly different between the Tc1-specific and Shared categories ($p = 0.27$) (Table S2B). In contrast to the human somatic tissues, many repetitive elements found in the Tc1 mouse testes are also enriched for H3K4me3 in human testes (e.g., 2/12 L1s, 1/5 SVAs, 7/8 AluYs, and 14/17 LTRs).

Between somatic and germline tissues, the most striking difference in transcription initiation among classes of repeat regions was observed for the SINE class (especially the AluY subfamily), which were significantly more often enriched for H3K4me3 in both human and Tc1 mouse testes (Figure 3C, middle panel;

(D) Number and lineage of repeat elements at H3K4me3 peak summits in liver, kidney, and testes. Pie charts are scaled relative to the total number of H3K4me3 binding events in the Shared category where the purple proportion represents the fraction of H3K4me3 events that have a repeat element at the H3K4me3 peak summit. Bar charts represent the lineage in which the repeat elements originated; the numbers of human-specific (and primate-specific for Shared testes sites) are reported beside the bar charts. The composition of repeat lineages on HsChr21 is shown in the bar graph on the right panel (* $p \leq 0.05$, ** $p \leq 0.005$, Wilcoxon rank-sum test).

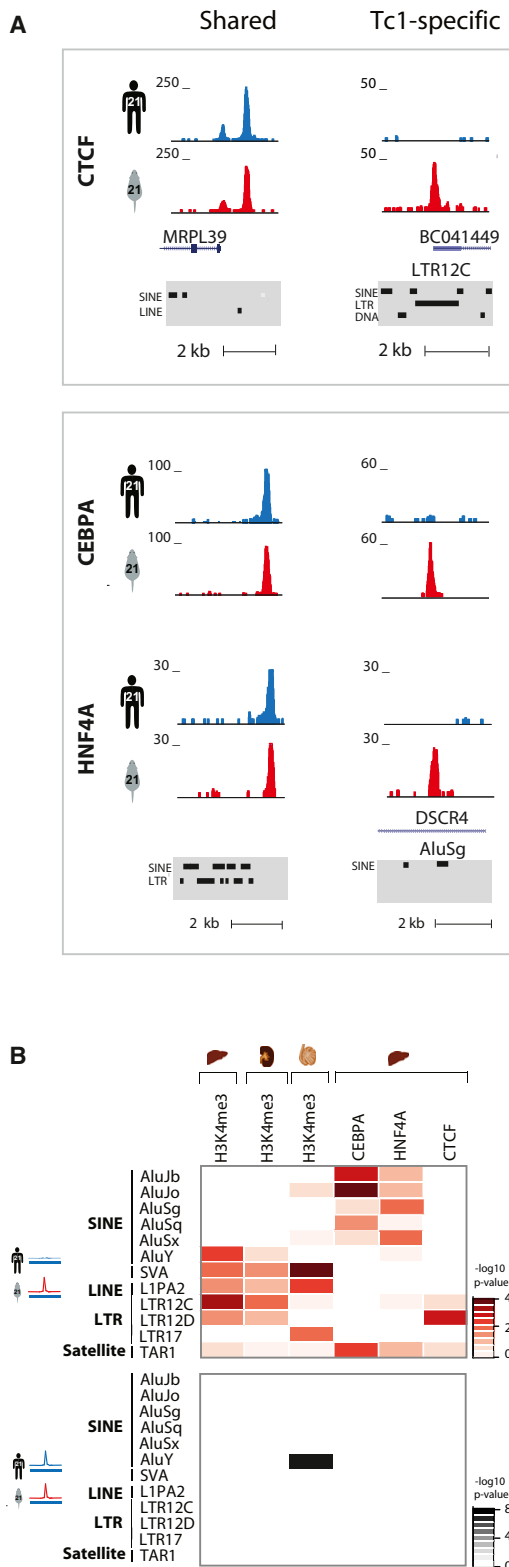


Figure 4. Repetitive Elements Contain Latent Transcriptional Regulator Binding Sites

(A) Individual CEBPA, HNF4A, and CTCF binding sites can be carried by specific repeat elements. An LTR12C repetitive element upstream of the *SOD1*

(Figure S3, Table S3B). AluY is the youngest Alu subfamily member in human, originating less than 35 million years ago (Jurka et al., 2002), after the mammalian radiation. The many recently evolved SINE elements transcribed in both species' testes shifted the repeat age distribution (Figure 3C, right panel); removal of this specific class of repeats results in age distributions similar to those found in somatic tissues (Figure S3A).

In contrast to mouse B1 SINE elements, which are often bound by Pol II at testes-specific promoters (Ichiyangi et al., 2011), primate-specific Alu SINE elements are regulated by RNA polymerase III (Pol III) (Rogers, 1983). We performed Pol III ChIP-seq in Tc1 testes, and using multiply mapping reads, we observed 165 AluY-associated Pol III peaks that occurred within regions enriched for H3K4me3 in the same tissue (Table S4). While these results support the role of Pol III in regulating AluY elements in mouse testes, the low mappability of AluY repeats prevented us from assessing binding differences with other Tc1 mouse tissues. Overall, these results indicate that Pol III and its regulatory machinery can, at least in part, accurately interpret human AluYs in the heterologous mouse testes.

Transcriptional Regulator Binding in the Tc1 Mouse

Given the widespread transcriptional activation of human-specific repeat elements in Tc1 mouse tissues, we asked whether they might also harbor latent transcription factor binding sites. Recent studies have demonstrated that up to a quarter of the OCT4 and NANOG stem-cell-specific transcription factor binding events in the human genome contain TEs (Kunarsko et al., 2010); similarly, the binding evolution of neural restrictive silencing factor (NRSF) (Johnson et al., 2006) and the insulator protein CTCF binding can depend on repetitive elements (Bourque et al., 2008; Kunarsko et al., 2010; Schmidt et al., 2012).

We investigated the genome-wide binding of CTCF (Schmidt et al., 2012) as well as the tissue-specific transcription factors CEBPA and HNF4A (Schmidt et al., 2010) in both human and Tc1 mouse liver tissue (Figure 4A and Figure S4A) and found a number of primate-specific repeats bound by these transcription factors (Figure S4B). For instance, an LTR12C repeat, comprising part of the long noncoding RNA *BC041449* that is located directly upstream of the amyotrophic lateral sclerosis gene *SOD1*, is occupied by CTCF in Tc1-mouse liver, but not in human. This CTCF binding event also shows trimethylation of H3K4 as well as Pol II occupancy, illustrating that repeat-driven latent regulatory potential could be biologically important (Figure S4C).

The set of TE families significantly enriched in a Tc1-specific manner for each transcription factor varied and was distinct from those revealed in our H3K4me3 ChIP-seq experiments

locus reveals a latent CTCF binding site in Tc1 mouse liver. Upstream of the Shared H3K4me3 site at the *CSTB* gene, there is a Shared HNF4A and CEBPA binding event, while Alu-associated CEBPA and HNF4A sites are revealed in the Tc1 mouse.

(B) Heatmap representation of repeats enriched in Tc1-specific events (red) or enriched in Shared (black) at H3K4me3 and transcription factor peak summits in liver, kidney, and testes. p values as calculated by chi-square test are presented in $-\log_{10}$ scale. Only repeat names that are significant in at least one data set are shown ($p = 0.05$).

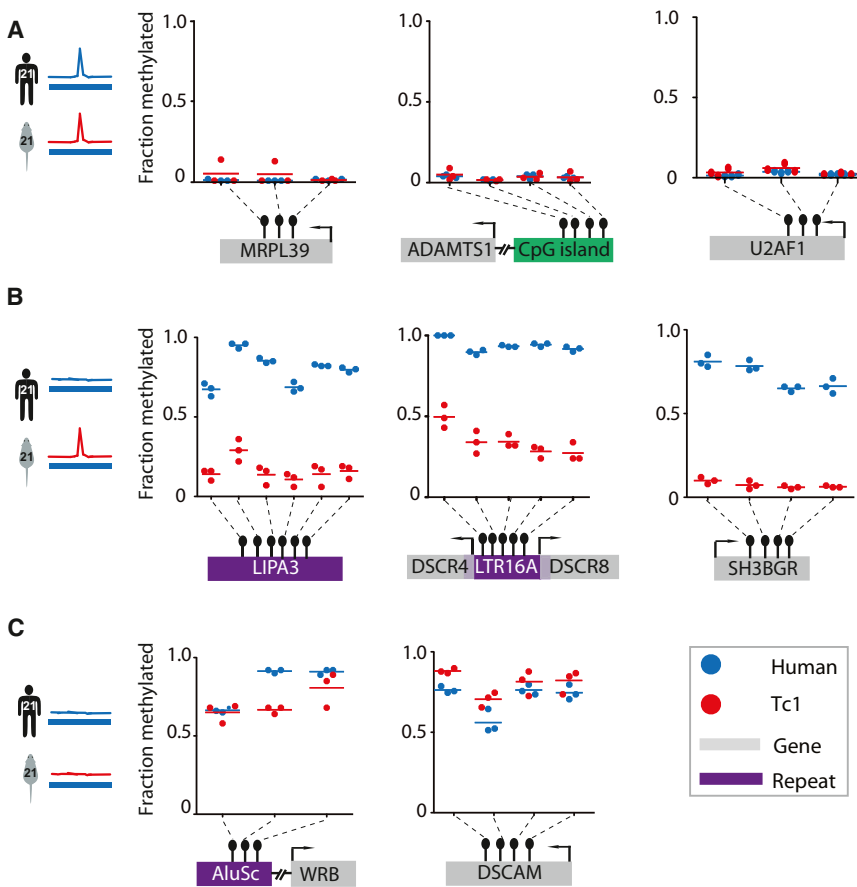


Figure 5. Tc1-Specific Sites of Transcription Initiation Are Depleted in DNA Methylation

(A) Fraction of methylated DNA at H3K4me3-enriched sites Shared between HsChr21 (blue) and Tc1-HsChr21 (red) in liver. (B) Fraction of methylated DNA at Tc1-specific H3K4me3-enriched sites. (C) Fraction of methylated DNA at regions where there is no H3K4me3 enrichment in Human or Tc1 mouse. Interrogated CpG sites are shown within LTR and LINE repetitive elements (purple) or genes (gray). Each experiment was performed using three biological replicates.

tions have noted that methylation of cytosines in CpG dinucleotides can cause transcriptional silencing of repetitive elements across eukaryotes (Zemach et al., 2010). Indeed, cytosine methylation is typically anticorrelated with H3K4me3 (Cedar and Bergman, 2009). Thus, we first asked if differences in the methylation of cytosines on HsChr21 associate with the transcriptional changes we observed.

To identify the CpG methylation state of representative HsChr21 regions, we performed bisulphite conversion of isolated genomic DNA from human and Tc1 mouse livers, followed by locus-specific pyrosequencing analysis. In regions of

(Figure 4B). For example, the Tc1-specific HNF4A binding events are particularly enriched for Alu SINE elements, an observation that is supported by recent analyses showing that these elements contain HNF4A binding motifs (Bolotin et al., 2011).

A greater fraction of the Tc1-specific CTCF binding sites are enriched for H3K4me3 (22/39) than are the Shared CTCF sites (90/358) (Table S5). In contrast, 40%–50% of CEBPA and HNF4A binding is associated with H3K4me3, regardless of whether the binding is Shared or Tc1-specific (Table S4). All CTCF-bound LTR elements are associated with H3K4 trimethylation (4/4) compared to a minority of CEBPA-bound (2/16) and HNF4A-bound (3/18) Alu elements. The absence of H3K4me3 at these transcription factor binding sites does not exclude the possibility that their binding was facilitated by the differential regulation of other epigenetic modifications, such as those found at enhancer elements. In sum, many different classes of repetitive elements in the human genome contain latent regulatory instructions for transcription factor binding, transcriptional activation, and polymerase occupancy that is revealed in vivo when placed in a heterologous mouse environment.

Mouse-Nucleus-Mediated Changes in Human Chromosome 21 DNA Methylation Correspond to Regions of Unmasked Regulatory Potential

We sought to identify the mechanism underlying the activation of these normally latent human regulatory elements. Prior observa-

Shared H3K4me3 enrichment, CpG dinucleotides were consistently hypomethylated in both species (five regions with two to four CpG sites per region) (Experimental Procedures, Figure 5A, Tables S6A and S6B). Conversely, in regions lacking H3K4me3 enrichment, CpGs were uniformly methylated in both species (two regions with three CpG sites per region) (Figure 5C, Tables S6A and S6B). These results are consistent with the above-mentioned anticorrelation between DNA methylation and H3K4 trimethylation. Importantly, CpGs falling in Tc1-specific, H3K4me3-enriched regions in Tc1 mouse livers showed less CpG methylation compared to human livers. This trend was observed for multiple CpGs in both repeat-associated (LINE and LTR; $n = 4$), and nonrepetitive Tc1-specific H3K4me3 regions ($n = 3$) (Figure 5B, Tables S5A and S5B).

We extended these CpG methylation experiments to the entire HsChr21 by using Illumina Human Methylation 450k BeadArrays to assess CpG methylation in two tissues (liver and testes) obtained from human, Tc1 mice, and (as a hybridization control) wild-type mice (Experimental Procedures). We identified 3,174 human CpG probes on HsChr21 that did not crosshybridize with mouse DNA and used these to compare the methylation state of HsChr21 in mouse and human. Consistent with the locus-specific results above, we found that in the Tc1 mouse, CpG sites in Tc1-specific H3K4me3 regions in liver ($n = 15$) and testes ($n = 12$) were depleted of DNA methylation, compared to human (liver median fraction methylated CpG sites 0.6 [Hsa]

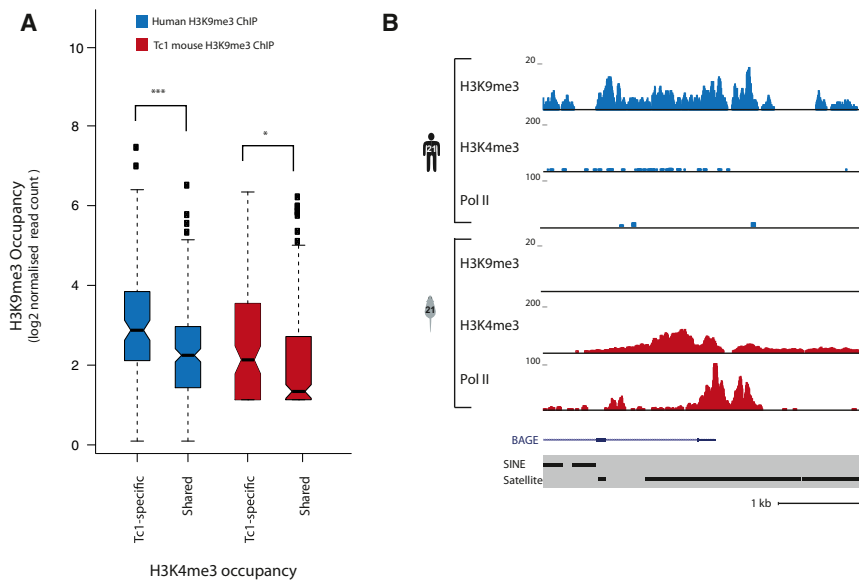


Figure 6. Changes in Transcription Status between Human and Tc1 Mouse Are Associated with Changes in the Repressive Histone Mark H3K9me3

(A) The genome-wide occupancy of H3K9me3 was determined in human liver (blue) and Tc1 mouse liver (red) by ChIP-seq and then compared with the transcriptional activation status of these regions. The vertical axis shows the log₂ normalized read counts for H3K9me3 ChIP experiments averaged across three individuals from each species. (* $p \leq 0.05$, *** $p \leq 0.0005$, Wilcoxon-matched pairs test).

(B) The occupancy of H3K9me3, H3K4me3, and Pol II is shown for the *BAGE* gene on HsChr21 located in human liver (blue) and Tc1 mouse liver (red). Human liver shows enrichment of the repressive histone mark H3K9me3 over the *BAGE* gene, which is missing in Tc1 mouse liver. Conversely, the mouse liver shows hallmarks of transcription (H3K4me3 and Pol II occupancies) that are absent from human.

versus 0.45 [Tc1], $p < 0.05$; testes median methylation 0.83 [Hsa] versus 0.39 [Tc1], $p < 0.05$) (Figure S5A and Table S6C). Interestingly, CpG sites within Shared H3K4me3 regions in liver ($n = 43$) and testes ($n = 92$) indicated an elevation of DNA methylation in the Tc1 mouse relative to human (liver median fraction methylated CpG sites was 0.06 [Hsa] versus 0.25 [Tc1], $p < 0.005$); testes median methylation was 0.11 [Hsa] versus 0.40 [Tc1], $p < 0.005$) (Figure S5A). Although this array-based analysis was limited by the small number of CpG probes in Tc1-specific H3K4me3 regions, these results suggest that DNA methylation plays a role in the differential epigenetic regulation of human DNA sequences in a heterologous mouse environment (Experimental Procedures, Figure S5A, Table S6A).

If DNA methylation changes in the Tc1 mouse were the mechanism that unmasked latent regulatory information in human repetitive sequences, then treating human cells with an agent that globally demethylates DNA should result in transcriptional upregulation of the same regions revealed by the Tc1 mouse. We treated cultured HepG2 liver cancer cells with 5-Aza-2'-deoxycytidine-5'-triphosphate (5-Aza-dCTP) to globally demethylate cytosines (Dannenber and Edenberg, 2006) and indeed observed that this subsequently resulted in trimethylation of H3K4 in the regions activated in Tc1 mouse tissues ($n = 4$, Figure S5B).

For instance, this treatment altered the DNA methylation of the LTR16A repeat element found at the *DSCR4/8* promoter, which is also upregulated in Tc1 mice (Figure 1B, Figure S5B). Consistent with the anticorrelation of DNA methylation and transcription, this promoter is hypomethylated in human placenta, where it is normally expressed, but hypermethylated in blood cells (Du et al., 2011). Importantly, no increase in H3K4me3 enrichment was observed for regions silenced in both human and mouse. However, regions not capable of further DNA demethylation ($n = 4$) also showed enrichment in H3K4me3, suggesting that the effect we observe could be due in part to indirect effects (Figure S5B). Indeed, a related cytosine analog designed to

inhibit DNA methylation has been shown to additionally affect genomic organization (Komashko and Farnham, 2010).

In sum, DNA methylation changes mechanistically contribute to transcriptional activation of regions identified as latent regulatory elements by the Tc1 mouse, as global DNA demethylation in human cells can lead to transcriptional activation of these same regions.

Changes in the Level of H3K9 Trimethylation Repressive Mark Occur at Human Regions that Are Specifically Activated in the Tc1 Mouse

Mechanistic studies in mouse embryonic stem cells have implicated the regulation of H3K9me3 in silencing proviral ERV elements through proteins such as KAP-1, the histone deacetylase (HDAC1), and ESET (Macfarlan et al., 2011; Matsui et al., 2010; Reichmann et al., 2012; Rowe et al., 2010). Trimethylation of H3K9 is associated with transcriptional repression, in contrast to trimethylation at H3K4, which is associated with transcriptional activation (Barski et al., 2007). We considered the possibility that H3K9me3 is important for silencing the active Tc1-specific H3K4me3-associated repeats in human tissue (Figure 6A).

We therefore profiled the genome-wide occupancy of H3K9me3 in human and Tc1 mouse liver using ChIP-seq (Experimental Procedures). We found that in human liver, Tc1-specific H3K4me3 regions showed elevated levels of H3K9me3 when compared to Shared H3K4me3 regions (Wilcoxon $p = 7.9 \times 10^{-7}$, Kolmogorov-Smirnov $p = 1.5 \times 10^{-5}$, distance = 0.28) (Figure 6A, blue boxplots). We also observed modest changes in the enrichment of H3K9me3 at Tc1-specific H3K4me3 regions relative to Shared regions in the Tc1 mouse, but there is less evidence for this enrichment (Wilcoxon $p = 0.001$, Kolmogorov-Smirnov $p = 1.2 \times 10^{-1}$, distance = 0.18) (Figure 6A, red boxplots). This result is consistent with the idea of trimethylation of H3K9 playing a role in sustaining transcriptional repression in human tissues; however, this mark is not anticorrelated with the

trimethylation of H3K4 in our system (Figure 6A). It has previously been reported that there are site-specific effects of H3K9me3-mediated repression (Matsui et al., 2010; Karimi et al., 2011), and our data likely reflect this. Alternatively, the lost repression observed in the Tc1 mouse may be heterogeneous, such that the presence of apparently coexisting H3K9 and H3K4 trimethylation may actually indicate fluctuating activation and repression within a population of cells. This may also help explain why depletion of DNA methylation appears to transit to intermediate values, as opposed to complete demethylation.

At a few specific, repeat-containing loci, the occupancy of H3K9me3 in human appeared to be replaced by hallmarks of active transcription (Figure 6B). For instance, in human liver, the *BAGE* locus is strongly enriched for H3K9me3 across a genomic region containing a large REP522 satellite repetitive region; no evidence of occupancy by H3K4me3 or Pol II was apparent in human liver. Conversely, in Tc1 mouse liver, the *BAGE* locus appears to be actively transcribed and lacking the repressive histone mark H3K9me3. However, the majority of corresponding activated Tc1-specific regions in human liver show more modest changes.

We also attempted to perturb the H3K9 trimethylation-mediated silencing pathway identified in mouse embryonic stem cells, which has been shown to reactivate human-silent regions by inhibiting the activity of histone deacetylases (HDACs) using Trichostatin A in human HepG2 cells. However, this treatment was insufficient to activate latent sites, unlike treatment with 5-Aza-dCTP (Figure S6B).

DISCUSSION

The dynamic mechanisms underlying the ongoing evolution of the regulatory human genome are of profound interest, and the widespread involvement of repeat elements is only beginning to be understood (Bourque et al., 2008; Faulkner et al., 2009; Kunarso et al., 2010; Lowe et al., 2007; Lynch et al., 2011; Oliver and Greene, 2009; Schmidt et al., 2012; Xie et al., 2010). Although heterologous systems are routinely employed in cell culture and humanized mouse models (Devoy et al., 2012), it remains unknown how heterologous nuclear environments globally interact with species-specific repetitive elements *in vivo*. Using an entire human chromosome carried in an aneuploid mouse, we demonstrated that hundreds of normally silenced human loci have previously unsuspected regulatory potential. These loci show tissue-specific activation and are enriched for primate- and human-specific repeat elements utilized by transcription factors, insulator elements, and transcriptional machinery. We experimentally determined that the mechanism underlying this activation involves depletion of DNA methylation at LINE and LTR elements and a global decrease in repressive histone modifications in the Tc1 mouse.

These discoveries support a model wherein the regulatory impact of certain repeat elements in somatic tissues is diminished by DNA methylation, and possibly H3K9 trimethylation (De Fazio et al., 2011; Kondo and Issa, 2003; Levin and Moran, 2011; Matsui et al., 2010; Rebollo et al., 2011; Reuter et al., 2011). It follows that accurate regulation of recently evolved human-specific repeat elements depends on coevolved, species-specific silencing

mechanisms. This can involve targeted repression by rapidly evolving zinc finger DNA-binding proteins (Huntley et al., 2006) or small RNA molecules that direct epigenetic machinery to specific genomic loci (Saito and Siomi, 2010). Indeed, the binding of tissue-specific transcription factors can establish regions of diminished DNA methylation (Stadler et al., 2011), and the subsequent, direct establishment of activating epigenetic modifications has been demonstrated using artificial unmethylated CpG clusters, which can recruit a CpG binding protein (Cfp1) and induce H3K4me3 *de novo* in the absence of Pol II and other transcription machinery (Thomson et al., 2010).

Our results also contribute to our understanding of germ cells as a highly active and transcriptionally exceptional tissue. The Tc1 mouse activates primate-lineage-specific AluY elements in the testes in a similar manner to what we observed in human testes, suggesting that the testes is uniquely suited to handling this class of elements. Thus, the testes-specific mechanisms of transcriptional activation seem to be more conserved than are the mechanisms of transcriptional silencing linked to changes in DNA methylation within somatic tissues.

Large-scale efforts using cells from a diverse range of primary human tissues are beginning to reveal the structure and function of the noncoding human genome (Consortium, 2011; Farnham, 2012; Dunham et al., 2012). Using a different strategy, we have exploited a single human chromosome transplanted into a heterologous regulatory environment to assess the transcriptional potential of most known human repeat families. The complete human genome is a hundred times larger, indicating that substantial latent regulatory potential remains to be discovered.

EXPERIMENTAL PROCEDURES

Tissue Preparation

Mouse Material

The Tc1 mouse line was generated and maintained as previously described (O'Doherty et al., 2005). Tc1 mice were bred by crossing female Tc1 mice to male (129S8 × C57BL/6J) F1 mice and were housed in the Biological Resources Unit under UK Home Office licensing. Tissue was obtained from at least two independent males. Sibling Tc0 mice, which do not carry HsChr21, or C57BL/6J mice were used as a control. Fresh tissue was either flash frozen or crosslinked with 1% formaldehyde as previously described (Schmidt et al., 2009).

Human Material

Male and female human tissue samples were obtained from biopsied tissue collected at Addenbrooke's Hospital, Cambridge, and were provided by the Biobank under human tissue license 08/H0308/117. Liver tissue was also obtained from the Liver Tissue Distribution Program (NIDDK contract number N01-DK-9-2310) at the University of Pittsburgh. For ChIP-seq, these samples were thawed in 1% formaldehyde and processed equivalently to the fresh crosslinked liver material used in this study.

ChIP-seq

ChIP-seq was performed for H3K4me3, Pol II, Pol III, and H3K9me3 as described in Schmidt et al. (2009). The data for CEBPA and HNF4A (Schmidt et al., 2010) and CTCF (Schmidt et al., 2012) have been previously described and are deposited under accession numbers E-MTAB-722 and E-MTAB-437, respectively.

Sequence Alignment, Peak Calling, and Repeat Identification

ChIP-seq and input reads of 36–50 bp were aligned to the reference genome, human NCBI36 (hg18), the mouse genome (NCBI37) (mm9) with the addition of HsChr21 (Tc1 genome), or a composite human + mouse genome, using the

default parameters of the MAQ short DNA read alignment tool (<http://maq.sourceforge.net/maq-man.shtml>). Regions with a mapping quality score of 0 were removed, and uniquely mapping reads were used for subsequent enrichment analysis. Regions of H3K4me3 enrichment were determined using the Control-based ChIP-seq Analysis Tool (CCAT2.0) (Xu et al., 2010). Peaks were called at an FDR of 0.001 with a minimum 5-fold enrichment over input. As previously described, ChIP-enriched regions of CEBPA and HNF4A (Schmidt et al., 2010) and CTCF (Schmidt et al., 2012) were called using SWEMBL (<http://www.ebi.ac.uk/~swilder/SWEMBL/>) with the parameters $-R\ 0.005 -i -S$.

Regions that mapped to known deleted or alpha satellite regions DNA on HsChr21 in the Tc1 mouse (O'Doherty et al., 2005) were excluded from analysis (precise regions were kindly provided by Dr. Susan Gribble). Peaks were overlapped with RepeatMasker (A.F.A. Smit, R. Hubley, and P. Green, RepeatMasker Open-3.0, 1996–2010, <http://www.repeatmasker.org/>) using custom Galaxy workflows (Blankenberg et al., 2010) (<http://main.g2.bx.psu.edu/u/mdwilson/w/wilsonwardetarepeatitemizationccat>, <http://main.g2.bx.psu.edu/u/mdwilson/w/wilsonwardetarepeatitemizationswembla>).

Differential Binding Analysis

Enrichment analysis for all ChIP-seq experiments was performed as in Ross-Innes et al. (2012) using the DiffBind R/Bioconductor package (version 1.0) (R. Stark and G.D. Brown, DiffBind: differential binding analysis of ChIP-seq peak data, Bioconductor [2011] <http://bioconductor.org/packages/release/bioc/html/DiffBind.html>) for analysis of differential binding.

RNA-Seq

Library Preparation and Sequencing

For the Tc1 mouse, mRNA-seq libraries were prepared from total RNA of three liver samples and sequenced on an Illumina Genome Analyzer II (50 bp single-end reads). For human liver, HiSeq 50 bp paired-end sequence reads from Illumina Human BodyMap 2.0 project were used. RefSeq transcripts were downloaded from NCBI (ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/).

Alignment and Gene Expression Analysis

RNA-seq reads were mapped to RefSeq transcript sequences using Blat (Kent, 2002), and mapped reads were filtered as previously described (Pan et al., 2008). Gene expression levels were quantified in reads per kilobase of transcript per million mapped reads (RPKM) (Mortazavi et al., 2008).

DNA Methylation Analysis

Locus-Specific DNA Methylation Assays

DNA was extracted from three Tc1, Tc0, and human liver tissues using the Easy-DNA Kit (Invitrogen) and bisulphite converted using the EZ DNA Methylation-Gold Kit (Zymo) according to manufacturers' instructions. Regions of interest were amplified in a nested PCR reaction and the ratio of C:Ts determined using the PyroMark Q96 MD pyrosequencer (QIAGEN) and Pyro Q-CpG software (QIAGEN).

Illumina Infinium Human Methylation450K Bead Arrays

DNA was extracted and bisulphite treated from liver and testes flash frozen material from four human, Tc1, and Tc0 individuals as described above. Methylation profiling was performed using Illumina Infinium Human Methylation450K beadarrays according to the manufacturer's standard protocol. Data were normalized using the Lumi BioConductor package (Du et al., 2008). Probes that showed significant detectable signal in Tc1 littermates that do not harbor Tc1-HsChr21 were excluded from further analysis. Arrays were run and analyzed by the Cambridge Genomics Service.

Abrogation of DNA Methylation in a Human Cell Line

HepG2 cells were treated with 1.5 μ M 5-Aza-2'-deoxycytidine-5'-triphosphate (5-Aza-dCTP) (Jena Bioscience) for 48 hr and retreated after 24 hr. Treated and untreated control cells were crosslinked with 1% formaldehyde 48 hr after initial treatment and H3K4me3 ChIP-qPCR performed as described above. DNA methylation levels were assayed to verify that DNA demethylation had occurred.

Inhibition of HDAC Activity in a Human Cell Line

HepG2 cells were treated with DMSO or 500 nM Trichostatin A (TSA) (Sigma) for 24 hr prior to formaldehyde crosslinking followed by H3K4me3 ChIP-qPCR analysis as described above.

ACCESSION NUMBERS

The ArrayExpress accession number for the chromatin immunoprecipitation sequence reported in this paper is E-MTAB-1104. The ArrayExpress accession number for the human gene expression sequence reported in this paper is E-MTAB-513. The GEO accession number for the Tc1 gene expression sequence reported in this paper is GSE31454. The ArrayExpress accession number for the Illumina Infinium HumanMethylation450k beadarray data reported in this paper is E-MTAB-1304.

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, six tables, supplemental Excel file, Supplemental Experimental Procedures, and Supplemental References and can be found with this article at <http://dx.doi.org/10.1016/j.molcel.2012.11.013>.

ACKNOWLEDGMENTS

We are grateful to J. Hadfield and the University of Cambridge Cancer Research UK Cambridge Institute Genomics Core, the Bioinformatics and Biological Resources Cores, and the Cambridge Genomics Service for running the human DNA methylation arrays. Thanks to G. Brown for computational assistance; V. Tybulewicz and E. Fisher for providing Tc1 mice and mice tissues; and Dr. Hiroshi Kimura, Dr. Masashi Narita, and Dr. Robert J. White for providing antibodies. This research was supported by ERC Starting Grant and EMBO Young Investigator Award (D.T.O.), Wellcome Trust Awards WT062023 and WT098051 (D.T.O., P.F.), University of Cambridge (D.T.O., M.C.W., M.D.W., D.S., P.C.S.), Cancer Research UK (D.T.O., M.C.W., M.D.W., D.S., R.S., S.M., M.L., S.W., C.K., K.K.), EMBL (P.C.S., D.T., P.F.), Canadian Institutes for Health Research (N.L.B.-M., B.J.B.), Commonwealth Scholarship Commission (M.C.W.), European Commission Marie Curie Actions (N.L.B.-M.), and Swiss National Science Foundation (C.K.). M.C.W., M.D.W., and D.T.O. designed experiments; M.C.W., M.D.W., D.S., Q.P., C.K., S.W., and K.K. performed experiments; M.C.W., M.D.W., N.L.B.-M., D.S., P.C.S., R.S., Q.P., D.T., M.L., and S.M. analyzed data; M.C.W., M.D.W., and D.T.O. wrote the manuscript. M.D.W., B.J.B., P.F., and D.T.O. oversaw the work.

Received: May 30, 2012

Revised: September 28, 2012

Accepted: November 9, 2012

Published: December 13, 2012

REFERENCES

- Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837.
- Beck, C.R., Garcia-Perez, J.L., Badge, R.M., and Moran, J.V. (2011). LINE-1 elements in structural variation and disease. *Annu. Rev. Genomics Hum. Genet.* 12, 187–215.
- Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120, 169–181.
- Bertone, P., Trifonov, V., Rozowsky, J.S., Schubert, F., Emanuelsson, O., Karro, J., Kao, M.Y., Snyder, M., and Gerstein, M. (2006). Design optimization methods for genomic DNA tiling arrays. *Genome Res.* 16, 271–281.
- Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol. Chapter 19*, Unit 19.10, 1–21.
- Bolotin, E., Chellappa, K., Hwang-Verslues, W., Schnabl, J.M., Yang, C., and Sladek, F.M. (2011). Nuclear receptor HNF4 α binding sequences are

- widespread in Alu repeats. *BMC Genomics* 12, 560. <http://dx.doi.org/10.1186/1471-2164-12-560>.
- Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., and Liu, E.T. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* 18, 1752–1762.
- Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* 10, 295–304.
- Consortium, E.P.; ENCODE Project Consortium. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* 9, e1001046. <http://dx.doi.org/10.1371/journal.pbio.1001046>.
- Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* 10, 691–703.
- Dannenberg, L.O., and Edenberg, H.J. (2006). Epigenetics of gene expression in human hepatoma cells: expression profiling the response to inhibition of DNA methylation and histone deacetylation. *BMC Genomics* 7, 181.
- De Fazio, S., Bartonicek, N., Di Giacomo, M., Abreu-Goodger, C., Sankar, A., Funaya, C., Antony, C., Moreira, P.N., Enright, A.J., and O'Carroll, D. (2011). The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* 480, 259–263.
- de Koning, A.P., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7, e1002384. <http://dx.doi.org/10.1371/journal.pgen.1002384>.
- Devoy, A., Bunton-Stasyshyn, R.K., Tybulewicz, V.L., Smith, A.J., and Fisher, E.M. (2012). Genomically humanized mice: technologies and promises. *Nat. Rev. Genet.* 13, 14–20.
- Du, P., Kibbe, W.A., and Lin, S.M. (2008). lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 24, 1547–1548.
- Du, Y., Zhang, J., Wang, H., Yan, X., Yang, Y., Yang, L., Luo, X., Chen, Y., Duan, T., and Ma, D. (2011). Hypomethylated DSCR4 is a placenta-derived epigenetic marker for trisomy 21. *Prenat. Diagn.* 31, 207–214.
- Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Fritze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.
- Dunn, C.A., Romanish, M.T., Gutierrez, L.E., van de Lagemaat, L.N., and Mager, D.L. (2006). Transcription of two human genes from a bidirectional endogenous retrovirus promoter. *Gene* 366, 335–342.
- Farnham, P.J. (2012). Thematic minireview series on results from the ENCODE Project: Integrative global analyses of regulatory regions in the human genome. *J. Biol. Chem.* 287, 30885–30887.
- Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T., et al. (2009). The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* 41, 563–571.
- Goodier, J.L., and Kazazian, H.H., Jr. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* 135, 23–35.
- Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130, 77–88.
- Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39, 311–318.
- Huntley, S., Baggott, D.M., Hamilton, A.T., Tran-Gyamfi, M., Yang, S., Kim, J., Gordon, L., Branscomb, E., and Stubbs, L. (2006). A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res.* 16, 669–677.
- Ichihyanagi, K., Li, Y., Watanabe, T., Ichihyanagi, T., Fukuda, K., Kitayama, J., Yamamoto, Y., Kuramochi-Miyagawa, S., Nakano, T., Yabuta, Y., et al. (2011). Locus- and domain-dependent control of DNA methylation at mouse B1 retrotransposons during male germ cell development. *Genome Res.* 21, 2058–2066.
- Johnson, R., Gamblin, R.J., Ooi, L., Bruce, A.W., Donaldson, I.J., Westhead, D.R., Wood, I.C., Jackson, R.M., and Buckley, N.J. (2006). Identification of the REST regulon reveals extensive transposable element-mediated binding site duplication. *Nucleic Acids Res.* 34, 3862–3877.
- Jurka, J., Krnjajic, M., Kapitonov, V.V., Stenger, J.E., and Kokhanyy, O. (2002). Active Alu elements are passed primarily through paternal germlines. *Theor. Popul. Biol.* 61, 519–530.
- Kano, H., Godoy, I., Courtney, C., Vetter, M.R., Gerton, G.L., Ostertag, E.M., and Kazazian, H.H., Jr. (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* 23, 1303–1312.
- Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., and Lorincz, M.C. (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* 8, 676–687.
- Kent, W.J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664.
- Komashko, V.M., and Farnham, P.J. (2010). 5-azacytidine treatment reorganizes genomic histone modification patterns. *Epigenetics* 5, 5. <http://dx.doi.org/10.4161/epi.5.3.11409>.
- Kondo, Y., and Issa, J.P. (2003). Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells. *J. Biol. Chem.* 278, 27658–27662.
- Kunarse, G., Chia, N.Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.S., Ng, H.H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* 42, 631–634.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al.; International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Levin, H.L., and Moran, J.V. (2011). Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* 12, 615–627.
- Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., et al.; Broad Institute Sequencing Platform and Whole Genome Assembly Team; Baylor College of Medicine Human Genome Sequencing Center Sequencing Team; Genome Institute at Washington University. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478, 476–482.
- Lowe, C.B., Bejerano, G., and Haussler, D. (2007). Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc. Natl. Acad. Sci. USA* 104, 8005–8010.
- Lynch, V.J., Leclerc, R.D., May, G., and Wagner, G.P. (2011). Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat. Genet.* 43, 1154–1159.
- Macfarlan, T.S., Gifford, W.D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., Andrews, S.E., Franco, L., Rosenfeld, M.G., Ren, B., and Pfaff, S.L. (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* 25, 594–607.
- Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., and Mager, D.L. (2006). Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet.* 2, e2. <http://dx.doi.org/10.1371/journal.pgen.0020002>.
- Maksakova, I.A., Mager, D.L., and Reiss, D. (2008). Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cell. Mol. Life Sci.* 65, 3329–3347.
- Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* 464, 927–931.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628.
- O'Doherty, A., Ruf, S., Mulligan, C., Hildreth, V., Errington, M.L., Cooke, S., Sesay, A., Modino, S., Vanes, L., Hernandez, D., et al. (2005). An aneuploid

- mouse strain carrying human chromosome 21 with Down syndrome phenotypes. *Science* 309, 2033–2037.
- Oliver, K.R., and Greene, W.K. (2009). Transposable elements: powerful facilitators of evolution. *Bioessays* 31, 703–714.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415.
- Pi, W., Yang, Z., Wang, J., Ruan, L., Yu, X., Ling, J., Krantz, S., Isaacs, C., Conway, S.J., Lin, S., and Tuan, D. (2004). The LTR enhancer of ERV-9 human endogenous retrovirus is active in oocytes and progenitor cells in transgenic zebrafish and humans. *Proc. Natl. Acad. Sci. USA* 101, 805–810.
- Rebollo, R., Karimi, M.M., Bilenky, M., Gagnier, L., Miceli-Royer, K., Zhang, Y., Goyal, P., Keane, T.M., Jones, S., Hirst, M., et al. (2011). Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genet.* 7, e1002301. <http://dx.doi.org/10.1371/journal.pgen.1002301>.
- Reichmann, J., Crichton, J.H., Madej, M.J., Taggart, M., Gautier, P., Garcia-Perez, J.L., Meehan, R.R., and Adams, I.R. (2012). Microarray analysis of LTR retrotransposon silencing identifies Hdac1 as a regulator of retrotransposon expression in mouse embryonic stem cells. *PLoS Comput. Biol.* 8, e1002486. <http://dx.doi.org/10.1371/journal.pcbi.1002486>.
- Reuter, M., Berninger, P., Chuma, S., Shah, H., Hosokawa, M., Funaya, C., Antony, C., Sachidanandam, R., and Pillai, R.S. (2011). Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* 480, 264–267.
- Rogers, J. (1983). Retroposons defined. *Nature* 301, 460.
- Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481, 389–393.
- Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* 463, 237–240.
- Saito, K., and Siomi, M.C. (2010). Small RNA-mediated quiescence of transposable elements in animals. *Dev. Cell* 19, 687–697.
- Schmidt, D., Wilson, M.D., Spyrou, C., Brown, G.D., Hadfield, J., and Odom, D.T. (2009). ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* 48, 240–248.
- Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., et al. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328, 1036–1040.
- Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Gonçalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 148, 335–348.
- Shima, J.E., McLean, D.J., McCarrey, J.R., and Griswold, M.D. (2004). The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis. *Biol. Reprod.* 71, 319–330.
- Smagulova, F., Gregoret, I.V., Brick, K., Khil, P., Camerini-Otero, R.D., and Petukhova, G.V. (2011). Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature* 472, 375–378.
- Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490–495.
- Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464, 1082–1086.
- Treangen, T.J., and Salzberg, S.L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* 13, 36–46.
- Vermeulen, M., Mulder, K.W., Denisov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* 131, 58–69.
- Wilson, M.D., Barbosa-Morais, N.L., Schmidt, D., Conboy, C.M., Vanes, L., Tybulewicz, V.L., Fisher, E.M., Tavaré, S., and Odom, D.T. (2008). Species-specific transcription in mice carrying human chromosome 21. *Science* 322, 434–438.
- Xie, D., Chen, C.C., Ptaszek, L.M., Xiao, S., Cao, X., Fang, F., Ng, H.H., Lewin, H.A., Cowan, C., and Zhong, S. (2010). Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Res.* 20, 804–815.
- Xu, H., Handoko, L., Wei, X., Ye, C., Sheng, J., Wei, C.L., Lin, F., and Sung, W.K. (2010). A signal-noise model for significance analysis of ChIP-seq with negative control. *Bioinformatics* 26, 1199–1204.
- Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328, 916–919.