Relics of repeat-induced point mutation direct heterochromatin formation
in *Neurospora crassa*

Zachary A. Lewis¹, Shinji Honda¹, Tamir K. Kh Lafallah¹, Jennifer K. Jeffress¹, Michael Freitag², Fabio Mohn³, Dirk Schübeler³ and Eric U. Selker¹,⁴

¹ Institute of Molecular Biology
University of Oregon
Eugene, OR, 97403-1229, US

² Center for Genome Research and Biocomputing
Department of Biochemistry and Biophysics
Oregon State University
Corvallis, OR 97331, US

³ Friedrich Miescher Institute for Biomedical Research
Maulbeerstrasse 66
4058 Basel, Switzerland

Corresponding author: selker@molbio.uoregon.edu

Running Title: Heterochromatin in *N. crassa*

keywords: DNA methylation, histone methylation, heterochromatin, repeat-induced point mutation, chromatin
SUMMARY

Both RNAi-dependent and -independent mechanisms have been implicated in the establishment of heterochromatin domains, which may be stabilized by feedback loops involving chromatin proteins and modifications of histones and DNA. *Neurospora crassa* sports features of heterochromatin found in higher eukaryotes, namely cytosine methylation (5mC), methylation of histone H3 lysine 9 (H3K9me) and HETEROCHROMATIN PROTEIN-1 (HP1), and is a model to investigate heterochromatin establishment and maintenance. We mapped the distribution of HP1, 5mC, H3K9me3 and H3K4me2 at 100bp-resolution and explored their interplay. HP1, H3K9me3 and 5mC were extensively colocalized and defined 44 heterochromatic domains on linkage group VII, all relics of repeat-induced point mutation (RIP). Interestingly, the centromere was found in a ~350kb heterochromatic domain with no detectable H3K4me2. 5mC was not found in genes, in contrast to the situation in plants and animals. H3K9me3 is required for HP1 localization and DNA methylation in *N. crassa*. In contrast, we found that localization of H3K9me3 was independent of 5mC or HP1 at virtually all heterochromatin regions. In addition, we observed complete restoration of DNA methylation patterns after depletion and reintroduction of the H3K9 methylation machinery. These data show that A:T-rich RIP’d DNA efficiently directs methylation of H3K9, which in turn, directs methylation of associated cytosines.
Supplemental material is available online at www.genome.org. The microarray data reported in this manuscript have been submitted to NCBI’s Gene Expression Omnibus under accession number GSE12690.
INTRODUCTION

Chromatin is the relevant substrate for all DNA-mediated processes in eukaryotes. Arrays of nucleosomes consisting of ~146 bp of DNA wrapped around an octamer of four histone proteins (H3, H4, H2A, H2B) represent the lowest level of chromatin organization. Interactions of nucleosomes with each other and with non-histone chromatin proteins and RNAs are thought to mediate functionally important higher-order chromatin structures. Gene-rich “euchromatin” exists in an open conformation during much of the cell cycle, facilitating DNA transactions such as transcription. In contrast, the densely staining “heterochromatin” remains highly condensed throughout the cell cycle, exhibits low levels of transcription and contains relatively few genes (Grewal and Jia 2007; Luger 2006).

The core histones are subject to extensive covalent modifications (e.g. by phosphorylation, acetylation, methylation and ubiquitination) that can impact chromatin structure by promoting or inhibiting nucleosome interactions, or by serving as binding sites for proteins or protein complexes such as chromatin remodelers. Generally, euchromatin is enriched for methylated H3K4 within active genes and is rich in acetylated histones. Some histone modifications in euchromatin are essential for transcriptional memory during development or for mounting appropriate and timely transcriptional responses to environmental stimuli. Conversely, heterochromatin is typically hypoacetylated and enriched for methylated H3K9 (Bhaumik et al. 2007). Though generally transcriptionally silent, heterochromatin is essential for proper centromere function and promotes genome stability by preventing illegitimate recombination between repeated DNA (Grewal and Jia 2007; Peng and Karpen 2008).
In addition to modification of histones, many organisms methylate some cytosines in DNA. Such DNA methylation in eukaryotes plays roles in development, genomic imprinting, X-chromosome inactivation, silencing of transposons and gene regulation (Heard and Disteche 2006; Miura et al. 2001; Reik et al. 2001; Reik and Walter 2001; Weber and Schubeler 2007). In animal genomes, 5mC is restricted to CpG dinucleotides whereas in plants and some fungi (e.g. *N. crassa*), DNA methylation occurs in both symmetric (CpG, CpHpG; H=A, C, or T) and asymmetric contexts (CHH) (Suzuki and Bird 2008). In plants and animals, hemi-methylated symmetrical sites produced during replication are recognized and methylated to propagate methylation patterns (Bostick et al. 2007; Saze et al. 2003; Sharif et al. 2007).

The mechanisms that contribute to maintenance of asymmetric DNA methylation and those that establish DNA methylation remain ill-defined; however, in some cases it is clear that histone methylation directs DNA methylation. The first, and most clear-cut example came from studies with the filamentous fungus *Neurospora crassa*, which showed that tri-methylation of H3K9 by the lysine methyltransferase (KMT) DEFECTIVE IN METHYLATION-5 (DIM-5) is essential for DNA methylation (Tamaru and Selker 2001). The *N. crassa* homolog of HETEROCHROMATIN PROTEIN-1 (HP1), which binds methylated H3K9, is also essential for DNA methylation (Freitag et al. 2004a). HP1 recruits DIM-2, the DNA methyltransferase (DNMT) responsible for all 5mC in vegetative cells (Honda and Selker 2008; Kouzminova and Selker 2001). Similarly, H3K9 methylation is important for some DNA methylation in both plants and animals (Jackson et al. 2002; Lehnertz et al. 2003). In addition, plants carry out RNA-directed DNA methylation, a process that involves the DNMT DRM2 and components of
the RNAi machinery (Henderson and Jacobsen 2007). Components of the RNAi pathway are dispensable for DNA methylation in *N. crassa* (Freitag et al. 2004b).

To develop a more complete understanding of how 5mC is controlled, it is essential to determine the genomic location of this modification. In plants, Methylated DNA Immunoprecipitation (MeDIP) coupled with microarray analysis (Zhang et al. 2006; Zilberman et al. 2007), and more recently whole genome shotgun sequencing of bisulfite treated DNA (Cokus et al. 2008; Lister et al. 2008) revealed dense methylation of repeated sequences and transposons that are highly concentrated within the pericentric heterochromatin domains. Methylation was also found in the coding regions of genes, but not in the promoters, of over 30% of expressed genes, indicating that transcriptional repression is not necessarily an outcome of DNA methylation. In contrast, most CpG dinucleotides within the mammalian genome are methylated, though many promoters contain C:G rich “CpG islands” that lack DNA methylation (Suzuki and Bird 2008).

Isolation and characterization of a fraction of the methylated DNA from *N. crassa* suggested that most methylated sequences are relics of repeat-induced point mutation (RIP), a premeiotic genome defense system that results in C to T changes within duplicated sequences (Selker et al. 2003). Moreover, introduction of RIP’d or A:T rich DNA into *N. crassa* triggered DNA methylation *de novo*, suggesting that positive signals target methylation to specific DNA sequences (Miao et al. 2000; Selker and Stevens 1987; Tamaru and Selker 2003). Interestingly, lightly RIP’d sequences were unable to trigger *de novo* methylation but were able to maintain previously established methylation (Selker et al. 2002; Singer et al. 1995). This revealed that *N. crassa* has the capacity to
perform both *de novo* and maintenance methylation. Still, the extent to which each pathway contributes to total methylation levels is not clear.

Since chromatin modifications impact DNA methylation, it is important to determine what other features comprise the chromatin environments marked by 5mC. Analysis of H3K9me3 and HP1 localization at a small number of RIP’d regions suggested that these components are colocalized with 5mC in *N. crassa* (Honda and Selker 2008; Tamaru et al. 2003), but the extent of colocalization throughout the genome is unknown. Recent work in mammalian cells revealed a strong inverse correlation between H3K4 methylation and 5mC (Meissner et al. 2008; Weber et al. 2007). Despite advances, much remains to be learned about the chromosomal contexts that promote or inhibit DNA methylation.

Knowledge of the distribution of chromatin marks should also help elucidate interrelationships between them. Positive feedback loops have already been implicated in establishment and maintenance of heterochromatin domains in several organisms. In *S. pombe*, domains of H3K9 methylation are found at centromeres, telomeres, and the silent mating type loci (Cam et al. 2005). Maintenance of heterochromatin domains in this yeast involves the H3K9 binding protein Swi6p/HP1, which is essential for the spread of H3K9 methylation beyond RNAi-dependent nucleation sites (Hall et al. 2002). In addition, the Clr4 H3K9 KMT binds methylated H3K9 via its chromo domain, thereby facilitating methylation of adjacent nucleosomes (Zhang et al. 2008). Similarly, Drosophila and mammalian SU(VAR) 3-9 H3K9 KMTs interact with the methyl H3K9 binding protein HP1 (Aagaard et al. 1999; Schotta et al. 2002). In *Arabidopsis thaliana*, the chromo domain of the CMT3 DNMT interacts with H3 methylated at K9 and K27 to...
target 5methyl-cytosine (5mC) and KRYPTONITE/SUVH4, an H3K9 KMT, binds 5mC via its SRA domain to target K9 methylation (Johnson et al. 2007; Lindroth et al. 2004).

We mapped the chromosomal distribution of methylated H3K9, HP1 and 5mC and found that these co-localize to form discrete heterochromatin domains within the *N. crassa* genome. We also found that while positive feedback occurs within the DNA methylation pathway, it is not required for maintenance of most heterochromatin domains in Neurospora. Finally, we determined that relics of repeat-induced point mutation trigger efficient *de novo* H3K9 and cytosine methylation.
RESULTS

Distribution of 5mC in N. crassa

To investigate heterochromatin organization within the N. crassa genome, we mapped the distribution of 5mC, H3K9me3 and HP1 across an entire chromosome. The N. crassa genome is composed of seven chromosomes covering ~42Mb. We chose LGVII for these analysis because its current chromosome assembly contains few gaps, covers ~10% of the genome and optical mapping data for LGVII suggest that the current genome assembly covers almost the entire genetically mapped centromere (Centola and Carbon 1994, http://www.broad.mit.edu/annotation/genome/neurospora/Home.html).

We designed a custom microarray containing oligonucleotide probes covering N. crassa LGVII and 37 previously identified methylated regions (representing regions from all seven chromosomes) at 100 bp resolution and determined the distribution of 5mC in wild type N. crassa strain 74-OR-23IV by MeDIP coupled with microarray analysis (Weber et al. 2005). Because most “repeated” sequences in the N. crassa genome have been subjected to RIP, which typically causes 10-20% sequence divergence, virtually all of our oligos were unique sequences (see METHODS). Inspection of the MeDIP data revealed that DNA from all 37 previously identified methylated regions was enriched in the immunoprecipitated fraction, validating our experimental approach (Figure 1). We showed previously that these methylated sequences have been altered by RIP (Selker et al. 2003). The RIP machinery, which produces C to T mutations, exhibits a preference for CpA dinucleotides. Indeed, analyses of the frequencies of TpA and CpA, relative to the frequencies of control dinucleotides, can identify sequences that have been subjected
to RIP. Typically, sequences that have not been subjected to RIP exhibit values less than 0.8 for the “RIP product index” (TpA/ApT) and values greater than 1.1 for the “RIP substrate index” (CpA + TpG/ ApC + GpT). In contrast, RIP’d regions typically exhibit values greater than 1.1 and less than 0.9 for RIP product and substrate indices, respectively (Margolin et al. 1998; Selker et al. 2003). A composite RIP index (CRI) can be determined by subtracting the substrate index from the product index; thus, a positive CRI value implies that the DNA has been subjected to RIP. We plotted the %GC and the CRI across each previously identified region. 5mC was enriched throughout the RIP’d regions and did not extend significantly into unRIP’d DNA.

We next examined the distribution of 5mC across N. crassa LGVII (Figure 2). Interestingly, a large region (~346kb) comprising the putative pericentric/centromeric DNA was enriched in the MeDIP fraction (Figure 3A). In addition, we observed numerous methylated domains along both chromosome arms including short methylated subtelomeric domains (Figure 3B). To determine the location of 5mC in an unbiased fashion, peaks of enrichment were identified using ChIPOTle (Chromatin ImmunoPrecipitation On Tiled microarrays) ChIP-chip data analysis software (Buck et al. 2005). Forty-six peaks were identified by ChIPOTle on LGVII ranging in size from 0.6 to 138.5 kb (Supplemental Table 1). Consistent with visual inspection of the data, peaks 29, 30 and 31 covered virtually all 346 kb of the putative pericentromere/centromere. The remaining 43 peaks were distributed on both arms of LGVII and ranged in size from 0.6 to 21.4kb (average of 7.1 kb excluding centromeric region). Overall, methylated regions on LGVII included 652kb, or 16.5% of the 3.9Mb chromosome. This is somewhat higher than expected since previous studies indicated that ~2% of cytosines are methylated in N.
crassa (Foss et al. 1993), but is consistent with previous reports of heterogeneous methylation within N. crassa cultures (Selker et al. 1993).

Co-localization of 5mC and H3K9me3

H3K9 methylation in wild type N. crassa exists predominantly, if not exclusively, as tri-methyl K9 (Tamaru et al. 2003) and this chromatin modification has been shown to co-localize with 5mC at several loci (Honda and Selker 2008). We determined the extent of co-localization by examining the distribution of H3K9me3 using our high-resolution tiled microarray. Enrichment of H3K9me3 was found within all regions that had been previously examined, once again validating the experimental approach (8:B1, 5:H8, 8:F10, Figure 1). Importantly, experiments using two different antibodies that recognize H3K9me3 yielded virtually identical results (see METHODS). Visual inspection of the data revealed striking co-localization of 5mC and H3K9me3 across LGVII (Figure 2). As with 5mC, H3K9me3 was enriched within the putative centromeric region (Figure 3A) and in dispersed regions along both arms, including subtelomeric domains (Figure 3B). We used ChIPOTle to identify peaks of H3K9me3 enrichment; 45 peaks were identified ranging in size from 0.5 to 151.1kb (Supplementary Table 2). Peaks 27 through 31 covered virtually all 346 kb of the putative centromere, whereas the remaining 40 peaks were distributed on both arms of LGVII, ranging in size from 0.5 to 21.6kb. H3K9me3 regions included 646kb, or 16.4% of the 3.9Mb chromosome, close to that detected by MeDIP (see above).

We examined the degree of overlap between the 5mC peaks and the H3K9me3 peaks. Of the 46 5mC peaks, 41 of these coincided with an H3K9me3 peak
(Supplementary Table 1). It is likely that the five 5mC regions that did not overlap with H3K9me3 peaks contain low levels of the modified histone that were not detected by ChIPOTle. Indeed, three of these 5mC regions (peak 10, 27 and 44, Supplementary Table 1) appeared to contain H3K9me3 by visual inspection and were identified as H3K9me3 peaks in one experiment, suggesting that low levels of H3K9me3 are present in these 5mC regions (data not shown). The fourth peak that did not overlap with an H3K9me3 peak displayed obvious enrichment of HP1 and appeared to have low levels of H3K9me3 by visual inspection and conventional ChIP (peak 34, Supplementary Table 1; Figure 4C and see below). The fifth peak that did not overlap with H3K9me3 had low levels of 5mC enrichment and may contain H3K9me3 that was not detectable under our experimental conditions. Notably, all 45 H3K9me3 peaks coincided with 5mC peaks (Supplementary Table 2). Taken together, these data show that 5mC and H3K9me3 are co-localized within the *N. crassa* genome.

**Complex localization of HP1**

HP1 interacts directly with both H3K9me3 and the DIM-2 DNMT via its chromo domain and chromo shadow domain, respectively (Freitag et al. 2004a; Honda and Selker 2008). These interactions are essential for DNA methylation in *N. crassa*. To determine if HP1 is present at all methylated regions and if HP1 is localized to both methylated and unmethylated regions, we examined the distribution of an HP1-GFP fusion protein expressed from the native HP1 locus and compared its localization to that of 5mC and H3K9me3 by ChIP-chip. HP1 was generally co-localized with these two modifications (Figures 1, 2, and 3). Surprisingly, however, variable enrichment of HP1-GFP was
detected, both among H3K9me3 regions and within some regions. It seems unlikely that this distribution resulted from altered protein function due to the GFP tag because the strain expressing this fusion protein displays wild type levels of DNA methylation (Honda and Selker 2008). Moreover, we performed conventional ChIP experiments at four heterochromatic regions using a strain expressing FLAG-tagged HP1. Like HP1-GFP, the HP1-3XFLAG fusion protein was expressed from the native HP1 locus and this strain displayed wild type levels of DNA methylation (Honda, S. and Selker, E.U., unpublished). Peak 35 (Supplementary Table 1) exhibited high levels of both H3K9me3 and HP1 throughout the RIP’d region (Figure 4A). In contrast, HP1 exhibited significantly higher enrichment near the edges of peak 33 (Supplementary Table 1), whereas H3K9me3 was relatively evenly distributed (Figure 4B). In contrast, peak 34 (Supplementary Table 1) was significantly enriched for HP1 even though H3K9me3 levels were low at this region (Figure 4C); the enrichment value for H3K9me3 and HP1 were of 1.9 and 2.6, respectively. Although these enrichment values are similar, HP1 levels were ~45% of the maximum enrichment (6 fold) while H3K9me3 levels were less than 20% of the maximal enrichment (10 fold) suggesting that while HP1 and H3K9me3 colocalize, they differ quantitatively. For example, peak 19 had low levels of HP1 even though H3K9me3 enrichment levels were high at this region (Figure 4D). Thus, conventional ChIP experiments confirmed the results seen in the microarray experiments. In addition, similar results were obtained for ChIP-chip experiments using strains expressing GFP-tagged and FLAG-tagged HP1 (data not shown), confirming that the location and amount of HP1 binding is not a direct function of H3K9me3.
DNA sequences in heterochromatic regions are products of RIP

A plot of the CRI for LGVII revealed that heterochromatin, defined by H3K9me3 and 5mC, is predominantly associated with RIP’d DNA. Major peaks of 5mC and H3K9me3 enrichment coincided with CRI peaks (Figure 2). We next asked if all methylated DNA on LGVII is present within relics of RIP. We calculated the CRI for all 46 methylated regions described above. The average CRI for all 46 methylated regions was 0.69 and 44 regions had a positive CRI. Both observations suggest that methylated DNA, in general, had been extensively RIP’d. We asked if methylated regions also had low levels of G:C, also expected of sequences exposed to RIP. Indeed, 80% of methylated regions had a G:C content of 40% or lower and the average GC content for all 46 methylated regions was 34.4% (compared to an average G:C content of 54.7% for predicted LGVII genes).

We next asked if all RIP’d DNA is associated with DNA methylation. We used ChIPOTle to call peaks of RIP’d DNA, yielding a total of 587 kb of DNA for LGVII. 96% of this predicted RIP’d DNA overlapped with a peak of 5mC. Interestingly, peaks of RIP’d DNA that did not overlap with 5mC exhibited one or more of the following properties: 1) the RIP’d region was less than 300 bp in length, 2) the RIP’d region overlapped with a region that contained H3K4me2, 3) the RIP’d region had a G:C content higher than 45%. These data indicate that in N. crassa, A:T rich, RIP’d sequences that lack H3K4me2 and are greater than 300 bp in length are targeted for H3K9me3 and 5mC.
In contrast to the A:T-rich, RIP’d sequences, *N. crassa* genes are G:C-rich. We found that only 0.5% of the predicted ORFs on LGVII (5/1008) significantly overlap (>90%) with 5mC regions and these appeared to be pseudogenes that resulted from RIP. They show a positive CRI, their predicted coding regions are short (<100 amino acids) and BLAST searches failed to identify similar sequences in the NCBI non-redundant database. Other apparent overlap between genes and 5mC turned out to be simply a consequence of the limited resolution MeDIP assay, which uses DNA sheared to ~500 bp. For example, the gene to the left of the 8:B1 region (Figure 1) appears to overlap with the edge of the 5mC region but Southern hybridizations demonstrated that it is not methylated (data not shown). Taken together, our data show that genes are not targets of DNA methylation in Neurospora, unlike the situation in plants and animals. DNA methylation is a widespread feature of repeated sequences of plants and animals, and this is also true for Neurospora. *N. crassa* repeated sequences are divergent, with numerous C:G to T:A nucleotide changes indicative of RIP. We found that 89% of the annotated repeats on LGVII overlapped with called 5mC peaks. Figure 5 summarizes the distribution of 5mC and H3K9me3 within LGVII genes and repeated sequences.

**DNA methylation is not affected in a double argonaute mutant**

Some DNA methylation in plants depends on RNAi machinery (Henderson and Jacobsen 2007) but this does not appear to be the case in *N. crassa*. Both maintenance and *de novo* DNA methylation occurred normally in a dicer double mutant, a triple RNA dependent RNA polymerase mutant, and two single argonaute mutants (Chicas et al. 2005; Freitag et al. 2004b). It remained a formal possibility, however, that the two
argonaute proteins have redundant roles in DNA methylation. To test this, we performed a MeDIP experiment with a double argonaute mutant and determined the distribution of 5mC using our high-resolution microarray. DNA methylation was detected at all methylated regions identified in the wild type strain (Supplementary Figure 1). Southern blots also revealed that DNA methylation was equivalent to that in wild type at all the heterochromatic loci examined (data not shown). Thus, our findings strengthen the conclusion that the RNAi machinery is not required for DNA methylation in *N. crassa*.

**Distribution of a euchromatic mark**

In contrast to heterochromatin, actively transcribed genes are methylated at K4 of histone H3 in eukaryotes that have been examined (Bhaumik et al. 2007). Because levels of H3K4me2 are inversely correlated with DNA methylation in mammals (Meissner et al. 2008; Weber et al. 2007), we performed ChIP-chip experiments to examine the distribution of this modification in *N. crassa*. Similar to the case in other organisms, H3K4me2 was enriched in the coding regions of genes. Of 1008 predicted ORFs on LGVII, 42.5% were enriched for H3K4me2 within the gene bodies. In contrast, 5mC enrichment peaks displayed negative log2 values, indicative of H3K4me2 depletion (Figure 2, 5C). Indeed, the average log2 enrichment value for H3K4me2 within the 46 5mC regions was -0.61 suggesting that heterochromatin and H3K4me2 are mutually exclusive in *N. crassa*. In many organisms, the centromeric core is enriched for H3K4me2 and for a centromere-specific H3 variant (CenH3/CENP-A) (Cam et al. 2005;
Sullivan and Karpen 2004). Interestingly, we did not detect any enrichment of H3K4me2 within the centromeric core of *N. crassa* LGVII (Figure 3A).

**Limited positive feedback from HP1 contributes to normal H3K9 methylation**

In mammals, DNA methyl-binding domain (MBD) proteins interact with H3K9 KMTs (Fujita et al. 2003; Sarraf and Stancheva 2004), whereas in plants, the SRA domain present in the KRYPTONITE/SUVH4 H3K9 KMT binds 5mC (Johnson et al. 2007). To examine whether DNA methylation might direct chromatin modification in *N. crassa*, we compared the distribution of H3K9me3 in the dim-2 DNMT mutant with that in a wild type strain. As a negative control, we examined the distribution of H3K9me3 in the dim-5 KMT mutant (strain N3436). We found that H3K9me3 persists at all LGVII heterochromatic regions in the dim-2 mutant (Figure 6). We also examined the distribution of H3K9me3 in the *hpo* strain because HP1 is known to interact with numerous proteins that modify chromatin. Most of the H3K9me3 detected in wild type strains was also present in the *hpo* mutant (Figure 6A) but we did identify three regions that lost H3K9me3 in the *hpo* mutant. These included one region on LGVII (peak 33) and two previously identified methylated clone sequences (8:A6 and 2:C9). HP1-sensitive regions were short (< 2.0 kbp) and had higher G:C content than HP1-insensitive regions, although these regions had lower G:C contents than gene sequences. The data for a typical heterochromatin region (HP1-insensitive) and two of these atypical (HP1-sensitive) regions are shown in Figure 6B. We conclude that 5mC and HP1 are dispensable for normal H3K9me3 at most regions of the genome but that HP1 is required for H3K9me3 at a minority of heterochromatic regions.
Efficient de novo methylation in N. crassa.

Previous studies demonstrated that both maintenance and de novo DNA methylation occur in N. crassa (Selker et al. 2002). To assess the fraction of methylated regions that are capable of rapid de novo methylation, we crossed different mutant strains that lack all DNA methylation and isolated a wild type daughter strain. Thus, all 5mC present in the daughter strain must be a result of de novo methylation. The distributions of 5mC from wild type, the two parental strains (dim-2 and hpo) and the wild type daughter strain (strain N3668) are shown for LGVII (Figure 7). As expected, 5mC was not detected in the two parental strains. In contrast, extensive methylation was detected in the daughter strain. Indeed, 5mC was detected at all of the methylated regions detected in the wild type strain. This indicates that de novo methylation occurs rapidly in Neurospora and that DNA methylation does not, in general, depend on preexisting methylation.

We carried out a similar experiment to determine whether H3K9me3 is efficiently reestablished de novo. In particular, we crossed dim-5 (Tamaru and Selker 2001) and dim-8 (Lewis, Z.A., Adhvaryu, K.K., Honda, S. and Selker, E.U., unpublished) strains, which are both defective in H3K9me3 (and 5mC), isolated a wild type recombinant and performed a MeDIP experiment to determine which regions were able to undergo de novo H3K9me3 and subsequent DNA methylation. Remarkably, DNA methylation was restored at all LGVII heterochromatin regions described above. Southern analyses were fully consistent with the MeDIP data (Supplementary Figure 2). We conclude that both
H3K9me3 methylation and DNA methylation are efficiently re-established *de novo* in *N. crassa*.
DISCUSSION

Colocalization of 5mC and H3K9me3

We examined the detailed genomic distribution and hierarchical relationships of DNA methylation, H3K9 methylation, and HP1 in the filamentous fungus *N. crassa*. H3K9 methylation is required for DNA methylation in this fungus and consistent with this observation, we found that the location of DNA methylation was highly correlated with the location of H3K9me3. Indeed, all regions that were enriched for H3K9me3 were also enriched for 5mC. This suggests that H3K9 methylation is sufficient to trigger DNA methylation in *N. crassa*.

Complex localization of HP1

In *N. crassa*, HP1 appears to directly link H3K9me3 with 5mC (Honda and Selker 2008). Two PXVXL-related motifs in DIM-2 interact directly with the HP1 chromo shadow domain. Interaction of HP1 with methylated H3K9 is well established (Bannister et al. 2001; Lachner et al. 2001). Consistent with this, we found that HP1 was generally colocalized with H3K9 methylation at *N. crassa* heterochromatic domains. Interestingly, HP1 was not highly enriched at all H3K9me3 regions, however, and HP1 enrichment was found to be high at one region that showed low levels of H3K9me3. Furthermore, in some regions HP1 displayed a non-uniform distribution relative to H3K9me3 and 5mC (Figure 4). Several possible models could account for the complex pattern of HP1 localization observed in our experiments. 1) HP1-interacting proteins may contribute to the pattern of HP1 localization by recognizing additional structural features such as histone modifications or DNA sequences. Accessory factors that promote chromatin
association of HP1 have been identified in mammals and Drosophila (Nielsen et al. 2001; Shi et al. 2008). 2) Additional chromo domain proteins could compete for binding of H3K9me3 and thereby, influence the distribution of HP1. Mammals encode multiple HP1 proteins that show distinct localization patterns and the fission yeast *S. pombe* encodes several H3K9-binding, chromo domain proteins that each display preferential binding at distinct genomic locations (Grewal and Jia 2007; Maison and Almouzni 2004). *N. crassa* encodes several chromo domain proteins (Borkovich et al. 2004). 3) The ChIP procedure may preferentially enrich for sites where HP1 forms a stable complex with the chromatin fiber. Both high and low mobility populations of HP1 exist within cells (Cheutin et al. 2004; Schmiedeberg et al. 2004). Importantly, the non-uniform distribution of HP1 relative to that of H3K9me3 (and 5mC) may indicate that only transient association with chromatin is required for HP1-directed DNA methylation by DIM-2. Alternatively, HP1 may be more uniformly distributed with H3K9me3 during a specific phase of the cell cycle, such as during replication. It will be important to determine how a somewhat variable pattern of HP1 localization can lead to the observed pattern of DNA methylation.

**RIP and DNA methylation**

Previous studies revealed a close association between the process of RIP and DNA methylation (Selker et al. 1993; Selker et al. 2003). The remaining cytosines within a duplicated sequence are typically methylated following RIP and isolation of methylated DNA revealed that most methylated sequences appear to be relics of RIP. Our results strongly support and extend prior evidence of a tight connection between RIP
and DNA methylation. In this study, we have shown that both 5mC and H3K9me3 are exclusively associated with relics of RIP all along LGVII and in other previously identified methylated regions.

Structure of LGVII centromere and telomeres

The most notable chromosomal domain of RIP’d DNA includes the genetically mapped centromere. In *S. pombe*, Drosophila and humans, the centromeric core domain is comprised of CENP-A containing nucleosomes interspersed with nucleosomes that contain H3K4me2 (Cam et al. 2005; Sullivan and Karpen 2004). In contrast, pericentric domains in these organisms typically contain H3 methylated at K9, and in plants and animals, pericentric sequences contain 5mC (Amor et al. 2004; Suzuki and Bird 2008). Recent work in *S. pombe* revealed that heterochromatic modifications within the pericentric domains are required for CENP-A deposition (Folco et al. 2008). Interestingly, we failed to identify any H3K4 methylation within the putative LGVII centromere. Moreover, 5mC and H3K9me3 were distributed broadly throughout the regions believed to comprise the centromeric core suggesting that *N. crassa* possesses a novel centromere organization. Although it is conceivable that the current genome sequence annotation lacks the *bona fide* LGVII centromeric core sequence, this is unlikely since ChIP-sequencing experiments with the *N. crassa* CENP-A protein show enrichment of the H3 variant in sequences included on our microarray (Smith, K.M. and Freitag, M., unpublished). Thus, our data indicate that enrichment of H3K4me2 within the centromeric core may not be a general feature of eukaryotic centromeres. Future
work to determine the distribution of *N. crassa* centromere proteins should provide additional insight into centromere organization in *N. crassa*.

Silencing of genes that reside near telomeres, termed “telomere position effect”, has been attributed to repressive chromatin that spreads from telomere-associated sequences into adjacent subtelomeric chromatin (Perrod and Gasser 2003). In *S. pombe*, a *cenH*-like repeat near telomere IL is believed to nucleate a nearly 40 kb telomeric heterochromatin domain containing H3K9 methylation and Swi6/HP1 (Cam et al. 2005). Similarly, we observed localization of H3K9me3 and HP1, as well as 5mC, at the chromosome ends in *N. crassa*. In contrast to the extended domains of subtelomeric heterochromatin in *S. pombe*, the *N. crassa* LGVII 5mC/H3K9me3 telomeric domain was short and highly correlated with A:T rich sequences. These data suggest that heterochromatin does not spread beyond the short RIP’d sequences found at telomeres in *N. crassa* (Wu et al. in press). Moreover, genes that are located close to subtelomeric heterochromatin were found enriched for H3K4 methylation suggesting that gene expression is not silenced in these regions.

*5mC is not found in N. crassa genes*

In addition to centromeres and telomeres, H3K9me3, HP1 and 5mC mapped to over 40 dispersed sites across LGVII that also show hallmarks of RIP. None of the heterochromatic domains on LGVII appear to contain functional genes. Conversely, coding sequences and virtually all promoters were found to be free of 5mC, H3K9me3 and HP1. Thus, our data do not support an extensive role for DNA methylation, H3K9 methylation or HP1 in regulation of transcription under standard growth conditions.
is also notable that H3K4 methylation was found in several heterochromatin-adjacent genes (see Figure 4), suggesting that heterochromatic domains do not necessarily suppress expression of flanking genes. This observation is consistent with transcript profiling experiments that do not support a role for 5mC in regulating gene expression in Neurospora crassa (Shiver, A., Lewis, Z.A., and Selker, E.U., unpublished data). In plants and animals, DNA methylation within gene bodies has been proposed to prevent initiation at cryptic promoters, a function apparently relegated to methylation of H3K36 in the yeast S. cerevisiae (Carrozza et al. 2005; Zilberman et al. 2007). Neurospora crassa does not appear to rely on DNA methylation to prevent such aberrant transcription initiation (Rountree and Selker 1997) but does perform H3K36 methylation (Adhvaryu et al. 2005). Perhaps gene body methylation is important for inhibiting aberrant transcription initiation in organisms with larger coding sequences while H3K36 methylation may be sufficient to provide this function in organisms with shorter genes.

Limited feedback within the N. crassa 5mC pathway

Although we found that DNA methylation was dispensable for H3K9 methylation at all RIP’d regions examined, it is possible that DNA methylation recruits factors that reinforce a repressive chromatin environment. A methylated-DNA binding activity has been identified in N. crassa nuclear extracts, consistent with this possibility (Selker et al. 2002). Interestingly, transcriptional activation of a silent copy of the bacterial hph gene was observed following depletion of DNA methylation (Irelan and Selker 1997). Moreover, loss of H3K9 methylation at a RIP’d allele of the N. crassa am gene (amRIP8) was observed in a dim-2 mutant (Tamaru et al. 2003). This may indicate that DNA
methylation is required for H3K9 methylation of transcriptionally competent sequences but not at sequences that are not typically transcribed. The extensive RIP of LGVII heterochromatic sequences presumably inactivated promoter sequences within these domains.

Similarly, HP1 had only a modest influence on the distribution of H3K9me3. Although we found that HP1 is required to maintain H3K9 methylation at three loci (i.e. one on LGVII and two previously identified regions), most heterochromatin domains remain unaffected in an hpo mutant. Specifically, 99.7% of H3K9me3 on LGVII persists in the hpo mutant. This observation was surprising because current models implicate HP1 in the spreading and maintenance of heterochromatin domains in other organisms.

In S. pombe, HP1 is required for spreading of H3K9 methylation across the silent mating type locus and Drosophila HP1 is required for normal localization of the SU(VAR)3-9 H3K9 KMT (Hall et al. 2002; Schotta et al. 2002). HP1 homologs in Drosophila and mammals interact with H3K9 KMTs in vivo (Aagaard et al. 1999; Schotta et al. 2002). Unlike its homologs in S. pombe, Drosophila and mammals, DIM-5, the N. crassa H3K9 KMT, does not have a chromo domain and apparently does not interact directly with HP1 (Honda and Selker 2008; Tamaru and Selker 2001). Therefore, it is likely that the limited effects of HP1 on H3K9 methylation are indirect. HP1 may act to recruit histone deacetylases since treatment of N. crassa with the histone deacetylase inhibitor Trichostatin-A results in limited loss of DNA methylation (Selker 1998).

Epigenetic inheritance reflects the stable maintenance of a chromatin state through replication cycles independently of the associated DNA sequence. Reporter genes inserted into the S. pombe mat locus are stably inherited in either the ‘on’ or ‘off’
state through mitosis and meiosis (Grewal and Klar 1996). In plants and animals, faithful inheritance of DNA methylation patterns requires maintenance methylation of hemi-methylated symmetrical sites, such as CpGs, following replication. Indeed, loss of methylation at CpG sites in *A. thaliana* MET1 mutants is not completely restored following reintroduction of a wild type MET1 gene, highlighting the importance of existing methylation (Saze et al. 2003). In contrast, we found no chromosomal regions of *N. crassa* that rely on maintenance methylation. This result fits nicely with our observation that H3K9me3 is normally distributed in a *dim-2* mutant (Fig. 6).

Interestingly, we observed complete restoration of DNA methylation patterns after depletion and reintroduction of the H3K9 methylation machinery, indicating that the signals for *de novo* heterochromatin formation lie upstream of H3K9 methylation. These data indicate that A:T-rich RIP’d DNA efficiently directs methylation of H3K9 and in turn, directs methylation of associated cytosines.

Taken together, these results support a model for heterochromatin formation in *N. crassa* that involves efficient targeting of H3K9 methylation and DNA methylation to A:T-rich relics of RIP. In contrast to heterochromatin formation in *S. pombe* and some DNA methylation in plants, this process appears to be independent of RNAi and probably relies on DNA binding factors that specifically recognize A:T-rich DNA. Indeed, the AT-hook analog Distamycin can inhibit DNA methylation in *N. crassa* (Tamaru and Selker 2003). Identification of the factors that direct heterochromatin formation at these regions will be critical for developing a complete model of heterochromatin formation in *N. crassa*.
METHODS

Microarray design

Custom microarrays produced by Agilent Technologies (http://www.agilent.com) contained 39,386 probes that span LGVII and 2,622 control probes corresponding to previously identified methylated regions. A complete LGVII sequence file was assembled based on the physical map assembled by the *Neurospora crassa* genome project (http://www.broad.mit.edu/annotation/genome/neurospora/Home.html). The relevant contig numbers, contig orientations, and position on the LGVII sequence are given in Supplementary Table 3. This assembly includes sequences for the telomere repeats and the adjacent subtelomeric heterochromatin for *telVIII*, but the telomere repeats and adjacent sequences are currently not available for *telVIIR*. Isothermal probes between 45 and 60 bp were designed at intervals of ~100bp from start to start as previously described (Thibaud-Nissen et al. 2006; Zilberman et al. 2007). The microarray contained probes corresponding to both unique and repeated sequences. It is important to note that virtually all repeated DNA within the *N. crassa* genome has been subjected to RIP. Therefore, repeated sequences typically display < 80% similarity (Galagan and Selker 2004). Of the 39,386 probes covering LGVII, 38,350 contained at least two or more mismatches when compared to the most similar repeated sequence. Our expectation that the vast majority of our microarray oligonucleotides would not cross-hybridize with other regions of the genome was supported by preliminary results of massive sequencing of DNA isolated by MeDIP or ChIP with the antibodies used in this study (Smith, K.M. and Freitag, M., unpublished).
**MeDIP and ChIP**

Genomic DNA was isolated for MeDIP as previously described (Freitag et al. 2004a). MeDIP was performed essentially as described except typically 1 ul of antibody specific for 5mC (Diagenode) was used to precipitate methylated DNA from 4ug of total genomic DNA (Weber et al. 2005). The procedure for ChIP analysis and antibodies that recognize H3K4me2 (#07-030 Upstate) and H3K9me3 (a gift from Prim Sing) were previously described (Tamaru et al. 2003). In addition, ChIP was performed using a second antibody specific for H3K9me3 (#39161 Active Motif). All primers used for conventional ChIP analysis are listed in supplementary Table 4.

**Sample labeling and microarray hybridization**

For MeDIP experiments, approximately 200 ng of sheared input DNA was labeled with Cy3 and 25% of the immunoprecipitated fraction (from 4ug input) was labeled with Cy5 using the Bioprime Array CGH Genomic Labeling system (Invitrogen) as described (Lee et al. 2006). For ChIP-chip experiments, 10 ng of input DNA and 50% of the immunoprecipitated fraction was amplified using a whole genome amplification as described (O'Geen et al. 2006). After two rounds of amplification, 300 ng of input DNA was labeled with Cy3 and 1 ul of amplified immunoprecipitated DNA was labeled with Cy5 using the Bioprime Array CGH Genomic Labeling system. Microarray hybridizations were performed for ~40 hours at 42°C in hybridization buffer (Lee et al. 2006).
Data analysis

Microarrays were scanned at 5 Micron resolution using a Genepix 4000a scanner (Axon) and pixel intensities were calculated using Genepix Pro 6.0 (Axon). The data were median-normalized and the log_2 [Cy5/Cy3] ratio was calculated for each spot. Data were plotted as log_2 values for each spot covering the entire LGVII sequence using the Argo genome browser available at http://www.broad.mit.edu/annotation/argo/. The positions of predicted open reading frames and repeats were obtained from the Neurospora Genome Project (http://www.broad.mit.edu/annotation/fungi/neurospora/).

To identify statistically significant peaks of enrichment, we first calculated median values for 500 bp sliding windows (5 spots) to reduce noise. We then utilized the ChIP-chip data analysis program ChIPoTle v1.11 (Buck et al. 2005) to call peaks of enrichment. Peaks of 5mC and H3K9me3 were determined for two biological replicate experiments using a window size of 500 base pairs and a step size of 100 base pairs. Replicate experiments for H3K9me3 were performed with two different antibodies that recognize H3K9me3 (see above). Only peaks that were detected by ChIPoTle in both replicate experiments were considered for subsequent analysis. Similarly, we used ChIPoTle to identify peaks of RIP’d DNA (i.e. regions with high CRI values) using a window size of 500pb and a step size of 100bp. For all analyses, only peaks with a p value of 0.05 or better were considered significant.

Strains and culture conditions

_N. crassa_ strains were grown and crossed using standard culture conditions (http://www.fgsc.net/Neurospora/neurospora.html). Wild type strain 74-OR-23-IVA
(Selker lab strain N150), dim-2 (strain N1850), hpo (strain N2556), dim-5 hpo-gfp dim-2-3xflag (strain N3436), hpo-gfp (N3415), and hpo-flag (N3320) strains were described previously (Freitag et al. 2004a; Honda and Selker 2008; Kouzminova and Selker 2001). The double argonaute mutant was generously provided by Dr. Yi Liu (UT Southwestern, Dallas, TX). A dim-5 knockout strain (N3074) was created by replacing the dim-5 gene with a basta-resistance cassette by homologous recombination as described (Colot et al. 2006). Primers used to amplify the dim-5 flanking sequence and create a knockout vector using yeast homologous recombination are listed in supplementary Table 4. The dim-8 gene was isolated by selecting for mutants that are defective in methylation (Lewis, Z.A., Adhvaryu, K.K., Honda, S., and E.U. Selker, unpublished). Since dim-5 strains are female sterile, a forced heterokaryon between dim-5 and the “helper 2” strain (FGSC #8745) was used as a female parent in crosses with dim-8. Helper 2 harbors a deletion of the mating type locus and therefore does not undergo productive meiosis (Metzenberg and Sachs 2002).

ACKNOWLEDGEMENTS

We would like to thank Nicholas Stiffler (UO) for microarray oligo design, Scott Givan (OSU) for help with oligo design, Reinhard Engels (MIT) for help with the Argo browser, and Yi Liu (UT Southwestern) for the double argonaute mutant. This work was supported by GM025690-22 to E.U.S. from the National Institutes of Health. Z.A.L. was supported by fellowships from the American Cancer Society (PF-04-122-01-GMC) and the Leukemia and Lymphoma Society (3295-09).
FIGURE LEGENDS

Figure 1. Chromatin modification profile of previously identified methylated regions. Data from five representative control regions are shown. A scale bar indicating 2 kb is shown at the top right of each plot. The base composition (%GC) and a composite RIP index (CRI) of each region, plotted as the moving average for 500bp windows with 100bp steps, are shown at the top of each plot. Enrichment values for MeDIP and ChIP-chip experiments are shown as log2 values (y-axis) for immunoprecipitation experiments with antibodies to 5mC, H3K9me3, Green Fluorescent Protein for HP1-GFP (HP1) and H3K4me2 for each region. The positions of predicted open reading frames (genes) and the previously identified methylated DNA clones, with their identification numbers, are shown at the bottom.

Figure 2. Chromatin modification profile for N. crassa LGVII. Base composition of the 3.9Mb LGVII is shown as the moving average of %GC and the CRI calculated for 500bp windows with 100bp steps at the top of the plot. Enrichment values for MeDIP and ChIP-chip experiments are shown as log2 values indicated on the y-axis (right) for immunoprecipitation experiments with antibodies to 5mC, H3K9me3, Green Fluorescent Protein for HP1-GFP (HP1) and H3K4me2. The positions of predicted open reading frames (predicted genes) and repeats are indicated below. The scale bar on the top right indicates 0.5Mb.
Figure 3. Chromatin modification profile for the LGVII centromere and telomeres.

For *N. crassa* LGVII centromere (A) and *N. crassa* chromosome ends (B), nucleotide composition is shown as the moving average of %GC and CRI calculated for 500bp windows with 100bp steps at the top of the plot. Enrichment values for MeDIP and ChIP-chip experiments are shown as log$_2$ values indicated on the y-axis (right) for immunoprecipitation experiments using antibodies to 5mC, H3K9me3, Green Fluorescent Protein for HP1-GFP (HP1) and H3K4me2. The position of predicted open reading frames (genes) and repeats are indicated below. The scale bars on the top indicate 30kb and 2kb for (A) and (B), respectively. The black dots above each plot in (B) indicate the position of the telomere. The broken line above *telVIIR* indicates that sequences containing the telomere repeats for *telVIIR* are missing from the current LGVII sequence assembly.

Figure 4. Complex localization patterns of HP1. Microarray data from immunoprecipitation experiments using antibodies to 5mC, H3K9me3, Green Fluorescent Protein (for HP1-GFP; HP1) and H3K4me2 are shown as log$_2$ values for 5mC peaks 35, 33, 34 and 19 (Supplemental Table 1). The position of predicted open reading frames (genes) and repeats are indicated below. Conventional ChIP was performed for each region using antibodies to H3K4me2, H3K9me3 and FLAG (for HP1-FLAG) to validate the microarray data. PCR products obtained using whole cell extracts (WCE) or the indicated immunoprecipitate fraction as the template were resolved by gel electrophoresis. The black bars labeled P1 through P5 depict the position of the expected PCR products within each region. The coding region of histone H4 was used as
an internal euchromatin control. The labels to the right of each autoradiograph identify specific PCR products. The enrichment of each PCR product relative to the PCR product for H4 to is shown below.

**Figure 5. Genes and repeats reside in distinct chromatin environments.** Scatter plots depicting the relationship between RIP and chromatin modifications are shown for LGVII genes (gray dots) and repeats (black dots). The CRI value calculated for each predicted open reading frame and annotated repeat is plotted on the y-axis. The median enrichment value for 5mC (A), H3K9me3 (B), and H3K4me2 (C) for each gene and annotated repeat is plotted on the x-axis.

**Figure 6. Distribution of H3K9 methylation in wild type, hpo and dim-2 strains.** A) The distribution of H3K9me3 across LGVII is shown for wild type, *dim-5, hpo* and *dim-2* strains. B) The distribution of H3K9me3 is plotted for wild type, *hpo* and *dim-2* strains for LGVII 5mC peak 40, the previously identified methylated region 8:B1 (C) and LGVII 5mC peak 33 (D) (Supplemental Table 1 and Selker *et al.* 2003). Conventional ChIP was performed using antibodies to H3K9me3 for the indicated strains. For each strain, PCR products obtained using the immunoprecipitated fraction (H3K9me3, top) and whole cell extract (WCE, bottom) as a template were resolved by gel electrophoresis. The black bars numbered P2 through P7 depict the position of the expected PCR products within each region. The coding region of histone H4 (hH4) was used as an internal euchromatic control. The labels to the right of each autoradiogram indicate specific PCR products. The relative enrichment of each PCR product is shown below.
**Figure 7.** *N. crassa* performs efficient de novo methylation of heterochromatic regions. The distribution of 5mC is shown across LGVII for wild type (wt), the methylation deficient strains *dim-2* and *hpo*, and wild type daughter strains from two crosses between methylation deficient strains. Enrichment values for MeDIP experiments are shown as log2 values on the y-axis. The wild type strain N3668 was obtained by crossing two strains that lack 5mC but retain H3K9me3, *dim-2* and *hpo*. The wild type strain N3669 was obtained by crossing two strains that lack both H3K9me3 and 5mC, *dim-5* and *dim-8*. Thus, all methylation in these strains is a result of de novo methylation.
References


*Current Opinion in Genetics & Development*: 8.


Figure 1

- %GC
- CRI
- 5mC
- K9me3
- HP1
- K4me2
- Genes
- Clone

8:B1 5:H8 8:F10 9:E1 2:A10

2kb
Figure 2
Figure 3

A

Cen VII

%GC
CRI
5mC
K9me3
HP1
K4me2
genes
repeats

B

telVIIL
telVIIR

%GC
CRI
5mC
K9me3
HP1
K4me3
genes
repeats

Cold Spring Harbor Laboratory Press on October 27, 2017 - Published by genome.cshlp.org
Figure 4

Figure showing the results of the ChIP-seq experiment. The figure displays the enrichment of specific histone modifications (5mC, K9me3, HP1-GFP, K4me2) at different peaks labeled as p1 to p5. The histograms represent the enrichment of these modifications across the genome, with peaks at 35, 33, 34, 19, and 19. The bands on the gel images correspond to the enrichment levels at these peaks.
Figure 5

(A) log2 [5mC/input] vs. CRI

(B) log2 [K9me3/input] vs. CRI

(C) log2 [K4me2/input] vs. CRI
Figure 7
Relics of repeat-induced point mutation direct heterochromatin formation in *Neurospora crassa*

Zachary A Lewis, Shinji Honda, Tamir K Khlaifallah, et al.

*Genome Res.* published online December 17, 2008
Access the most recent version at doi:10.1101/gr.086231.108

**Supplemental Material**

http://genome.cshlp.org/content/suppl/2009/02/06/gr.086231.108.DC1

**P<P**

Published online December 17, 2008 in advance of the print journal.

**Accepted Manuscript**

Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

**License**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](http://genome.cshlp.org/content/suppl/2009/02/06/gr.086231.108.DC1).

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *Genome Research* go to: [http://genome.cshlp.org/subscriptions](http://genome.cshlp.org/subscriptions)

Copyright © 2008, Cold Spring Harbor Laboratory Press