Review
TRENDS in Genetics Vol.22 No.6 June 2006
Full text provided by www.sciencedirect.com

Oxymoron no more: the expanding world of heterochromatic genes

Jiro C. Yasuhara and Barbara T. Wakimoto

Department of Biology, Box 351800, University of Washington, Seattle, WA 98195, USA

Heterochromatin has been oversimplified and even misunderstood. In particular, the existence of heterochromatic genes is often overlooked. Diverse types of genes reside within regions classified as constitutive heterochromatin and activating influences of heterochromatin on gene expression in *Drosophila* are well documented. These properties are usually considered paradoxical because heterochromatin is commonly portrayed as 'silent chromatin'. In the past, studies of heterochromatic genes were limited to a few *Drosophila* genes. However, the recent discovery of several hundred heterochromatic genes in *Drosophila,* plants and mammals through sequencing projects offers new opportunities to examine the variety of ways in which heterochromatin influences gene expression. Comparative genomics is revealing diverse origins of heterochromatic genes and remarkable evolutionary fluidity between heterochromatic and euchromatic domains. These features justify a broader view of heterochromatin, one that accommodates repressive, permissive and activating effects on gene expression, and recognizes chromosomal and evolutionary transitional states between heterochromatin and euchromatin.

Introduction

The term 'heterochromatic gene' is considered to be an oxymoron by some scientists, and there is a historical reason for this confusion. The inclination to equate heterochromatin with a lack of gene expression dates back to none other than Emil Heitz, the cytogeneticist who coined the term heterochromatin in 1928 [1]. Heitz characterized heterochromatin as the chromosomal component that appears darkly staining throughout the cell cycle, to distinguish it from the cycling euchromatin, which appears diffuse in interphase nuclei. He showed that heterochromatin is a common entity in plant and animal cells and imagined it to be the manifestation of functionally inactive regions of the genome. This notion was largely supported by observations of contemporaries of Heitz. When geneticists discovered and mapped the first mutations in *Drosophila*, they found that most map to euchromatin. Only a few genes mapped to the Y chromosome or near the spindle attachment site of other chromosomes, regions that were classified as 'constitutive' heterochromatin because they appeared consistently heterochromatic on both homologs in most if not all cell types in an organism. The ability of these heterochromatic regions to induce variegated expression of euchromatic genes when the two types of chromatin were abnormally juxtaposed by chromosome rearrangements made a striking impression on geneticists. This phenomenon, called position effect variegation (PEV), was discovered by H.J. Muller in 1930 [2]. Subsequent studies revealed that dozens of euchromatic genes are inactivated when placed near or in heterochromatin [3], irrespective of the time or tissue in which the gene is expressed, or the function of its product. The generality of PEV showed that heterochromatin not only lacks gene activity, but could routinely cause gene inactivation. Molecular biologists later showed that constitutive heterochromatin largely comprises highly repetitive satellite sequences and middle-repetitive transposable elements (TEs), so-called 'junk' DNA, whose transcription was considered either nonexistent or dispensable. Genetic tests revealed that the silencing effect of heterochromatin was subject to modification by altering the dose of heterochromatin (such as adding an extra Y chromosome) or by mutations in genes that became known as *Suppressors* (*Su(var)*s) or *Enhancers* (*E(var)*s) of PEV [4].

In the past decade, key insights into the molecular features of heterochromatin have been obtained. It is now widely recognized that heterochromatic domains in diverse organisms are often associated with particular chromosomal proteins and core histone modifications, and with the ability to silence euchromatic gene expression [5]. The discovery of the activities of two key highly conserved proteins, SU(VAR)3-9 and HP1, has been particularly informative and has led to a molecular model of heterochromatin formation [6]. The model proposes that SU(VAR)3-9, a methyltransferase, specifically methylates Lys9 of histone H3 (H3K9me) on heterochromatic DNA. HP1 (heterochromatin protein 1), a chromodomain protein that binds H3K9me as a dimer, is proposed to condense the chromatin fiber through protein–protein interaction, rendering it inaccessible for transcription [6]. Although SU(VAR)3-9 and HP1 are the most extensively studied SU(VAR) proteins, numerous genetic and molecular factors have been found to affect heterochromatin formation or maintenance, including components of the RNA interference (RNAi) machinery. The landmark discovery that small interfering RNAs (siRNAs) have a role in heterochromatin formation in some organisms has

Corresponding author: Wakimoto, B.T. (wakimoto@u.washington.edu).

www.sciencedirect.com 0168-9525/$ - see front matter © 2006 Elsevier Ltd. All rights reserved. doi:10.1016/j.tig.2006.04.008

highlighted the importance of the transcription of specific repetitive sequences in heterochromatin [7].

The molecular features described above are often considered the hallmarks of heterochromatin; however, none are absolute defining features. Nonetheless, the tendency to equate the term 'heterochromatin' with 'silent chromatin', 'H3K9me-enriched chromatin' or 'HP1-enriched chromatin' dominates the field. Like many generalizations, overly simplified views of heterochromatin can be misleading. In reality, some of the earliest geneticists recognized the diverse features and functions of heterochromatin. As we summarize in this review, the existence of active genes that normally reside within heterochromatin has important implications for understanding how chromosomal context and chromatin structure can regulate gene expression. We focus here on heterochromatic genes in *Drosophila*, plants and mammals that reside in or near constitutive heterochromatin. We refer the reader to other reviews for a discussion of genes expressed from the inactivated X chromosome of female mammals, the classic and best characterized example of facultative heterochromatin [8,9].

Drosophila heterochromatic genes

The study of heterochromatic genes began in the early days of *Drosophila* genetics [10]. In 1916, Calvin Bridges suggested that the Y chromosome must carry genes necessary for male fertility, although, as realized later, this chromosome was considered entirely heterochromatic by cytologic definition. In addition, several mutations, including the *bobbed*, *light* and *rolled* mutations, mapped to heterochromatin, an early finding that forced even Heitz to consider the possibility of heterochromatic genes [11]. Jack Schultz realized the regulatory implication of genes residing in heterochromatin in his pioneering work linking PEV to euchromatic–heterochromatic rearrangements [12]. When he reported an example of PEV in which several genes were affected, he was keen to point out that one of the variegating genes actually resided in the heterochromatic side of the rearrangement breakpoint. Even more fascinating was the finding that variegation of this heterochromatic gene, the *light* gene, was enhanced when the dosage of Y chromosome was increased, whereas variegation of genes located in the euchromatic side of the breakpoint was suppressed. These early insights turned out to be of broad significance when additional heterochromatic genes exhibited similar dependency on heterochromatin (Box 1). Moreover, genetic studies revealed that some heterochromatic genes showed opposite effects compared with variegating euchromatic genes when tested with the *Su(var)* and *E(var)* mutations [13,14].

Beginning with work on the Y chromosome by Curt Stern in 1929 [10], *Drosophila* geneticists have used chromosomes carrying large deletions of the heterochromatin to map and identify heterochromatic genes. Systematic screens for mutations in heterochromatic genes showed that they are required at a variety of developmental stages and tissues [15]. Forty heterochromatic loci were defined by genetic studies [16]. These included functionally redundant loci identified by chromosomal deficiencies, specifically the X- and Y-chromosome-linked *bobbed* loci, which encode the abundantly transcribed 18+28S rRNA genes, and ABO loci, which are located on the sex chromosomes and autosomes. Most of the genetically defined heterochromatic loci were identified by spontaneous or ethyl methanesulfonate-induced alleles causing visible or lethal phenotypes [15]. Several of these genes, including *light* and *rolled*, are unique in function and protein coding [17]. Improved cytologic mapping of heterochromatic loci was possible using

Box 1. Documenting dependency on heterochromatin

Many genes from diverse organisms are known to be silenced by heterochromatin, but the expression of only a few genes is known to depend on heterochromatin. This is because until recently only a few heterochromatic genes were known and relatively few studies addressed their regulatory requirements. What evidence is required to demonstrate dependency on heterochromatin convincingly? Stringent proof requires demonstration that displacement from heterochromatin results in reduced gene expression, and return to heterochromatin restores expression. The test requires chromosome rearrangements or fully functional transgenes to change the chromosomal context of a gene, and phenotypic or molecular assays to monitor expression. Genes that are broadly expressed and produce easily scorable phenotypes with reduced expression have been most amenable to the test [70].

Schultz [12] provided the first evidence that expression of a heterochromatic gene, the *light* gene, depends on heterochromatin. Strong evidence for similar behavior exists for eight other *Drosophila* heterochromatic genes [71–73]. Howe *et al.* [74] demonstrated that the level of *light* gene expression is correlated with the size of the heterochromatin block adjacent to the gene. This gene is extremely sensitive to heterochromatin and is affected even with several megabases of heterochromatin retained adjacent to the gene. Eberl *et al.* [72] showed that the heterochromatic block containing the *rolled* gene had to be reduced substantially and located sufficiently distant from another block of heterochromatin for a detectable effect. One model to explain these long-range effects proposes that heterochromatic genes are dependent on heterochromatin-enriched factors that are compartmentalized in the nucleus [71]. The size of the surrounding heterochromatic block, complexity of the chromosome rearrangement and developmental stage of gene expression are factors that affect the probability that a displaced heterochromatic gene can compete favorably for factors required for its expression [29,70].

Demonstration that a *Su(var)* mutation decreases gene expression is consistent with dependency on heterochromatin but does not constitute proof. This is because most *Su(var)* mutations are pleiotropic and most SU(VAR) proteins are multifunctional. A more convincing case can be made if a consistent trend is observed with multiple mutations within the class of *Su(var)* or *E(var)* mutations, as has been best demonstrated for the *light* gene [13,14,75]. *Su(var)* mutants, including those in HP1a, reduce expression of heterochromatic genes [13], even in the absence of displacement from heterochromatin [35,76]. The effects of modifiers are visible at the level of the transcript on a single-cell basis [14] or as steady-state levels of mRNA isolated from whole animals [35]. Chromatin profiling studies have shown that HP1a and SU(VAR)3-9 are enriched in at least some portions of the *light* and *rolled* genes in cultured cells [39,42], consistent with the idea that these proteins affect expression of these genes by direct binding. However, higher-resolution studies are needed to determine the precise binding locations and consequences for gene activation. A major outstanding challenge is to understand how the molecular action of the SU(VAR) proteins on heterochromatic gene expression (Figure 1) can be reconciled with the current models of SU(VAR)- and heterochromatin-mediated euchromatic gene silencing.

a variety of chromosome banding and *in situ* hybridization techniques. The cytogenetic map of mitotic heterochromatin generated in these studies is now considered the gold standard for defining heterochromatic regions in *D. melanogaster*, and it provides resolution of the order of ~200 Kb [17,18]. Because the screens for mutations did not achieve saturation or identify redundant genes, there was a realistic expectation that the actual number of *Drosophila* heterochromatic genes would be significantly greater. This count did not include genes on the small chromosome 4, which has always been given a special 'quasi-heterochromatic' status by cytologists and is rich in repeated DNA sequences [19].

Expanding views of *Drosophila* heterochromatic genes: insights from genomics

Efforts to understand the organization and content of heterochromatin have been best rewarded through genome sequencing work. When the sequence and annotation of the euchromatic genome of *D. melanogaster* was announced in 2000 [20], it was with the full recognition that the formidable task of assembling and annotating the heterochromatin lay ahead. To determine the content of the 30% of the genome represented by heterochromatin, the *Drosophila* Heterochromatin Genome Project (DHGP) was launched.

The genomics approach to heterochromatic gene discovery relied on the whole-genome shotgun (WGS) approach to sequence assembly. The crucial advantage of WGS is its ability to sample essentially all regions of the genome that can be cloned in plasmid vectors and sequenced, allowing for the recovery of single-copy DNA fragments embedded in regions with a high repetitive DNA content. Such sequences are poorly represented in large-insert clone libraries, which provide the starting material for other approaches to large-scale genome sequencing. The first WGS release of the *D. melanogaster* genome provided superb coverage of the euchromatin [20]. Later efforts analyzed the remaining unassembled WGS sequences and filled coverage gaps, enabling extended coverage from the euchromatin into nearby heterochromatin [21]. Assignment of sequences to heterochromatin was based on chromosomal *in situ* hybridization of large clones to align the physical and cytogenetic maps [17,22,23]. Annotation of genes in the repeat-rich proximal euchromatin and heterochromatin required the use of repeat-masking, in addition to standard annotation tools such as alignment of genomic, cDNA and expressed sequence tag (EST) sequences, gene prediction programs, and searches for conserved coding sequences [22]. Altogether, 22.8 Mb of previously unmapped sequences were assigned to heterochromatin, within which 447 gene models were defined [21,22,24]. This coverage is estimated to be ~35% of the total *D. melanogaster* heterochromatin. Carvalho *et al.* [25,26] applied additional strategies that took advantage of the relative under-representation of Y chromosome sequence in the whole-genome shotgun sequence assembly and the staggered arrangement of scaffolds when aligned to cDNAs to identify additional heterochromatic genes.

The discovery of several hundred heterochromatic genes is an impressive accomplishment because only a few heterochromatic genes had been previously characterized at the sequence level. Continued investigations by the DHGP and other laboratories since 2002 suggest that 450 genes is an accurate accounting of *D. melanogaster* heterochromatic genes (C.D. Smith and G. Karpen, unpublished), indicating that they constitute ~2.7% of the estimated 14 715 genes in the genome. As expected from genetic studies, the distribution of heterochromatic genes is uneven, with greatest gene densities nearest to the euchromatin and between large blocks of highly repetitive satellite DNAs [22,27]. Overall, the average gene density in heterochromatin appears to be about one-sixth that of euchromatin, with the majority of heterochromatic sequences comprising repetitive sequences. The heterochromatic gene collection reveals diversity in predicted functions, association with repetitive sequences, and expanded gene size relative to euchromatic genes. Repetitive sequences typically reside within one or more introns of each gene, increasing the average intron size to more than fourfold that of euchromatic genes [22,28]. Y chromosome fertility factors are extreme examples, with introns that are highly enriched in simple satellite sequences and megabases long [25].

Models of heterochromatic gene expression

Heterochromatic genes reside in a chromosomal domain that silences euchromatic genes, suggesting that distinct mechanisms of gene regulation exist. A variety of models could be envisioned to accommodate heterochromatic gene expression. If one steadfastly adheres to the notion of heterochromatin as repressive, it follows that genes located in heterochromatin must somehow be shunning the surrounding environment so that they are transcribed. A formal possibility is that what appears to be heterochromatin at low resolution could actually be a mosaic of heterochromatin and euchromatin. In this case, an 'insulation model' could apply, with expressed genes protected from the spreading effects of repressive chromosomal factors by boundary elements or by depletion of the factors by distance. This situation predicts that whereas the bulk of the chromosomal region contains repressive factors, the gene is free of them (Figure 1a). Alternatively, the gene need not require insulation if it can recruit a factor that cripples the repressors. We refer to this as the 'denial model' (Figure 1b).

Experimental studies indicate that neither the 'insulation model' nor the 'denial model' apply for the *Drosophila* heterochromatic genes whose expression is compromised with progressive loss of surrounding heterochromatin and reduced dosage of SU(VAR) proteins (Box 1). To explain heterochromatin dependence, we have proposed that the genes take advantage of repetitive sequences and heterochromatin-enriched factors to facilitate essential, long-distance interactions between enhancers and promoters [29,30]. According to this 'integration model', the genes have adapted to heterochromatin and are now dependent on its components to form an active conformation (Figure 1c).

Heterochromatic genes could exploit heterochromatin-enriched components through direct action at the promoter. One type of 'exploitation model' proposes that

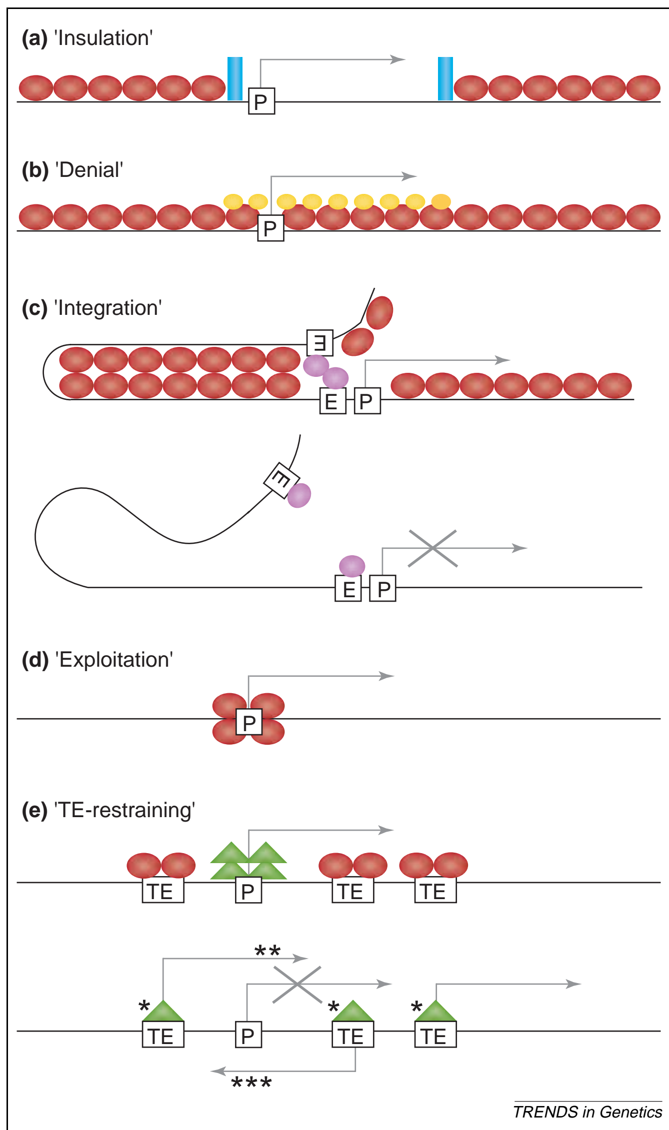

Figure 1. Different models of heterochromatic gene expression. Red ovals indicate heterochromatin-enriched chromosomal factors; 'P' denotes promoter, and gray arrow indicates transcription. **(a)** 'Insulation model': the gene is insulated from the surrounding repressive heterochromatin-enriched factors by boundaries (blue rectangles). **(b)** 'Denial model': the gene recruits a factor (yellow ovals) that neutralizes the effects of heterochromatin-enriched repressors. **(c)** 'Integration model': the heterochromatin-enriched factors facilitate long-range interaction between the promoter and enhancer (top), which is required for expression. If heterochromatin-enriched factors are depleted (bottom), the active configuration cannot form and transcription does not occur. 'E' denotes enhancer sequences, and purple ovals indicate enhancer-binding factors. **(d)** 'Exploitation' model: in this version, heterochromatin-enriched components are used as positive transcription factors at the promoter. **(e)** 'TE-restraining' model: heterochromatic gene transcription occurs and could require specific transcription factors (green triangles), whereas TE activity is repressed by heterochromatin-enriched factors. Depletion of heterochromatin proteins (bottom) results in decreased transcription of heterochromatic genes by competition for activating factors by TE sequences (*), promoter occlusion (**) or antisense formation (***).

heterochromatic genes use TE-derived sequences as promoters. There are several documented cases in which TEs have contributed promoters or other regulatory DNA sequences to host genes leading to altered expression patterns [31]. Systematic study of heterochromatic genes could increase the number of known examples substantially. Whether heterochromatic genes contain specialized or standard promoters, expression could depend on heterochromatin-enriched transcription factors (Figure 1d). Such factors could be bifunctional, positively regulating heterochromatic genes but negatively regulating euchromatic genes. Several transcription factors, such as yeast Rap1p and Abf1p and the *Drosophila* NC2, have been shown to activate or repress target genes depending on promoter type or context [32,33]. Especially interesting for heterochromatin are studies showing that HP1, which is best known for its ability to silence euchromatic genes, is a highly versatile protein. Recent studies have shown that *Drosophila*, like other organisms, has multiple members of the HP1 protein family [34]. Of these, HP1 (also known as HP1a) was the first HP1-like protein to be discovered and is encoded by the *Su(var)205* gene. In addition to heterochromatin-induced gene silencing, the activities of this *Drosophila* HP1 protein include: positive regulation of heterochromatic gene expression [13,35]; association with heat-shock and hormonally induced gene expression [36]; positive and negative effects on euchromatic gene expression [37–40]; dosage compensation [41,42]; and multiple functions at the telomere [43]. Not all of these functions depend on the capacity of HP1 to bind H3K9me [39,43,44]. Moreover, HP1 has been shown to bind to diverse genomic sites, including certain types and arrangements of repetitive DNAs, heterochromatic genes and euchromatic genes, and telomeres, with localization at some sites independent of its interaction with SU(VAR)3-9 [42,45].

The 'insulation' and 'denial' models of heterochromatic gene expression are consistent with the conventional view of repressive heterochromatin, whereas the 'integration' and 'exploitation' models are fundamentally different in proposing permissive or activating influences on gene expression. The potential exists for more-complex regulatory relationships between heterochromatic genes, repetitive sequences and their transcripts. For instance, heterochromatin might be permissive for heterochromatic gene transcription but repressive for transcription of TEs or other repetitive DNAs. This incorporates a 'TE-restraining' activity for heterochromatin. Reduced levels of heterochromatin components would then result in TE activation and consequently disrupt nearby gene expression via mechanisms such as dilution of transcriptional activators, promoter occlusion or antisense transcription (Figure 1e).

Further studies of the large set of genes identified by the DHGP will reveal which of the above mechanisms are relevant. In a recent study, Yasuhara *et al.* [30] identified the promoters of seven heterochromatic genes, including the *light* gene, and eliminated the possibility of a heterochromatin-specific promoter type for these genes. The 'integration model' remains favored because it proposes a role for TEs (or other repetitive DNAs) that would confer a selective advantage to maintain gene expression. There is evidence for fixation of specific TEs or TE-like fragments in heterochromatin [46] and within heterochromatic genes [46,47]. In the heterochromatic poly(ADP-ribose) polymerase (*PARP*) gene, dozens of TEs exist in and near the gene, and nearly all are oriented in the same direction and opposite to that of the gene. Moreover, the retrotransposons lack certain sequences, such as long terminal repeats and insulator sequences,

that typically disrupt gene expression. These biased properties of such TEs are suggestive of adaptive significance to permit, or perhaps promote, heterochromatic gene expression [48].

Other evidence supports different models of heterochromatic gene expression. Sun *et al.* [49] interpreted differences in the expression of transgenes inserted in various locations within chromosome 4 of *Drosophila* as functional evidence for a mosaic of euchromatic and heterochromatic domains existing throughout the chromosome. Interestingly, the vast majority of sites that induced transgene silencing were located within or near genes, consistent with the idea that euchromatic and heterochromatic genes differ in their requirements for expression. It is not yet known if transcription of any of the ~80 genes on chromosome 4 actually depends on a heterochromatic environment.

The discovery of heterochromatic genes of plants and mammals

In the past, the existence of diverse heterochromatic genes in *Drosophila* did not draw sufficient interest from investigators focused on other organisms. However, the picture is beginning to change with the discovery of heterochromatic genes and candidate heterochromatic genes in plants and mammals. Many of these genes have been identified in studies of centromeres, which reside in cytologically defined heterochromatin.

A variety of approaches have been used to define the organization of the centromere and its flanking pericentromeric DNA. The finding of expressed single-copy genes residing in these regions, which are highly enriched in repetitive DNAs, has been considered surprising. Copenhaver *et al.* [50] reported 160 genes within the 4.3 Mb and 2.7 Mb regions that include the centromeres of chromosomes 2 and 4 of *Arabidopsis thaliana*. This number indicates that the gene density in centromeric and pericentromeric heterochromatin is 20–50% that typical of euchromatin. At least 25 of the annotated genes are transcribed, based on cDNA evidence. Based on sequence, several dozen of the 160 annotated genes are predicted to be nonfunctional [50]. A rice centromere, *Cen8*, which has a low content of highly repetitive CentO satellite sequences but is highly enriched for other repeated sequences, contains 16 active genes [51]. *Cen8* flanking regions contain an estimated 136 active genes, with an abundance of transposable elements in intergenic regions [52]. Chromatin profiling indicates enrichment for H3K9me in the nontranscribed region overall, but the majority of genes have H4 acetylation and H3 methylation profiles that are typical of other active genes [52]. Additional heterochromatic genes are being identified as genome sequences and transcripts are correlated with cytogenetic maps in rice [53] and other plants. Recent data indicate that tomato and sorghum, which have >50% of their karyotype appearing heterochromatic, could have a greater fraction of genes within or near heterochromatin compared with *Arabidopsis* or rice [54,55]. Therefore, these species might be particularly interesting for studies of plant heterochromatic genes [54].

Progress has also been made in the discovery of genes that map near the human centromere. Genome-wide studies have provided information about the organization of an estimated 76% of pericentromeric regions, defined as the 2 Mb on either side of the centromere for each chromosome [56]. These regions exhibited an overall density of exons that was half that of the genome average [56], with an uneven distribution of the >300 annotated genes. Several transcribed genes are located within 100 Kb of satellite sequences [57], although it is not known if expression of these genes is influenced by proximity to highly repetitive DNAs.

The presence in *Arabidopsis* of a cytologically and molecularly well-characterized interstitial block of heterochromatin known as the hk4S knob, has provided an opportunity to study heterochromatic genes. Sequence analysis of the knob region shows an abundance of repetitive DNAs and a low density of expressed genes [58]. In an elegant study, Lippman *et al.* [59] performed high-resolution chromatin profiling and expression analysis of the entire heterochromatic knob region. Although the region as a whole was enriched for TEs, H3K9me and DNA methylation, which are considered markers of heterochromatin, the genes were devoid of these attributes. These results are consistent with the 'insulation' model (Figure 1a).

Evolutionary histories of heterochromatic genes

Several studies have shown that heterochromatic genes are fascinatingly diverse in their evolutionary histories (Figure 2). Comparison of the location of the *light* gene among ten species of *Drosophila* revealed that this autosomal gene was heterochromatic in seven species but euchromatic in three species [30]. Phylogenetic relationships indicate that the gene was transformed into a heterochromatic gene <20 million years ago in the lineage leading to the melanogaster subgroup. Furthermore, six of the ten heterochromatic genes located in the vicinity of the *light* gene in *D. melanogaster* have orthologs that reside within ~200 Kb of the euchromatic *light* genes of *D. pseudoobscura* [30] and *D. virilis* (J.C. Yasuhara, unpublished). Hence, the evolutionary transition from euchromatin to heterochromatin did not occur on a gene-by-gene basis but rather at the level of a large chromosomal segment, probably by 'infiltration of heterochromatin' (Figure 2a). These genes are single-copy in the ten *Drosophila* species analyzed so there is no evidence for duplication events accompanying the change in chromosomal position.

Studies of different *Drosophila* species have provided evidence for several distinct evolutionary origins and fates for genes on the Y chromosome. In most cases, heterochromatinization of the Y chromosome has been associated with genetic erosion over time. Based on studies of genes on the neo-Y chromosome of *D. miranda*, Steinemann and Steinemann [60] have proposed that accumulation of TEs is an early step that typically leads to gene inactivation or loss. Active genes might have somehow tolerated TE insertions or they might be relatively recent arrivals to the Y chromosome. Carvalho *et al.* [61] discovered that nearly all of the single-copy Y-linked

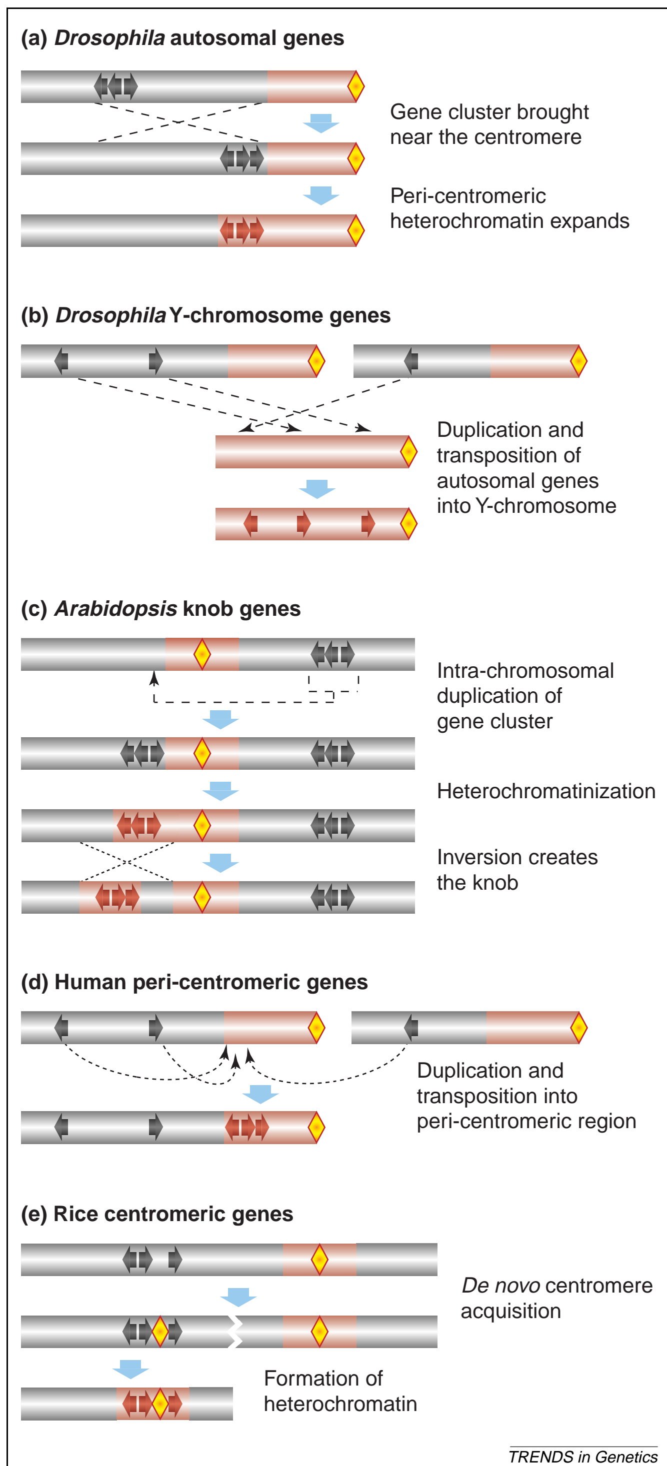

Figure 2. Proposed evolutionary origins of heterochromatic genes. Red and gray bars indicate heterochromatic and euchromatic chromosomal segments, respectively. Resident genes are shown as block arrows with the same color codes. Yellow diamond denotes the centromere. **(a)** A gene cluster containing the *Drosophila light* gene is originally located in euchromatin, then placed proximally by a chromosome rearrangement. The region transitions to a heterochromatic domain by 'infiltration of heterochromatin' owing to the gradual accumulation of transposable elements and expansion of other repetitive DNA elements near and within the gene cluster [30]. **(b)** Individual *D. melanogaster* autosomal genes are duplicated and transposed onto the Y chromosome [61]. Heterochromatinization of the Y chromosome can begin before or after gene arrival [60]. **(c)** *Arabidopsis* knob genes are derived from a segmental duplication that becomes heterochromatic [59,64]. **(d)** Human pericentromeric regions contain genes arising from intra- or interchromosomal insertions of duplicated euchromatic segments [65]. **(e)** In rice, a region containing active genes acquires centromere activity, then becomes heterochromatic by acquisition of centromeric and pericentromeric repetitive DNAs [51].

genes in *D. melanogaster* have autosomal paralogs. This finding is consistent with derivation through duplication and transposition of individual genes or large autosomal euchromatic regions onto the Y chromosome (Figure 2b). These data are also relevant for considering the origin of the Y chromosome itself. Unlike the mammalian Y chromosome, which is a degenerate version of the X chromosome, the *D. melanogaster* Y chromosome might have arisen from a supernumerary 'B' chromosome that acquired genes from the autosomes [62]. These duplicate genes have consistently acquired male function and some are now required for male fertility.

Carvalho and Clark [63] also investigated the origin of the Y-linked genes in *D. pseudoobscura*. Their data suggest a more complicated scenario and are consistent with the movement of formerly Y-linked genes back to an autosome in the *D. pseudoobscura* lineage. Whereas the Y-linked genes in *D. melanogaster* have characteristic megabase-sized introns and intergenic regions, the derived autosomal counterparts in *D. pseudoobscura* show reduced size more typical of euchromatic genes. This surprising observation is consistent with the notion that the *D. pseudoobscura* genes are evolving toward a euchromatic nature. Taken together, the interspecific studies of *Drosophila* heterochromatic genes provide evidence for remarkable evolutionary fluidity between euchromatic and heterochromatic compartments.

The heterochromatic hk4S knob is present in only some ecotypes of *Arabidopsis thaliana* [59,64]. The region probably arose by an intrachromosomal segmental duplication and a pericentromeric inversion, resulting in the incorporation of centromeric heterochromatic sequences adjacent to the duplication (Figure 2c). Euchromatic and heterochromatic paralogs for at least a subset of the genes coexist in the genome, so the gene pairs are evolving under conditions of functional redundancy.

Human pericentromeric regions have the intriguing property of being highly prone to accepting segmental duplications [65]. This is an unexpected property that accounts for the derivation of at least 30% of the pericentromeric sequences from intra- or interchromosomal duplications of euchromatin [56,57] (Figure 2d). These duplicated segments contain nearly 50% of the annotated pericentromeric genes. Curiously, a majority of these genes are expressed primarily in the testis [56,57] and are also activated in cancer tissues. The significance of this biased transcription is unknown but it might reflect a global chromatin feature of the duplicated segments that results in repression in most tissues, but relaxation in the germline and cancer cells [57].

Less is known about evolutionary origins of the pericentromeric genes in plants. Nagaki *et al.* [51] speculate that the rice *Cen8* resulted from acquired centromere activity relatively recently (Figure 2e). On a related note, Saffery *et al.* [66] found that the genes that lie within a human neocentromere are transcribed at a level indistinguishable from their counterparts in euchromatin. *Cen8* and the human neocentromere both lack long stretches of centromeric satellite DNA arrays, but they are fully functional centromeres. Hence, in these two cases,

the acquisition of centromeric chromatin has not hindered gene transcription.

Implications for adaptation

The evolutionary origin, copy number and function of a gene are factors that will influence how the organism responds to alterations in the chromosomal context of the gene. Nonredundant heterochromatic genes that are dispensable for cellular function could be free to degenerate without consequence to the organism. Those that arise through duplication events could also degenerate provided one or more functional paralogs are retained in the genome. By contrast, some heterochromatic genes, such as the *Drosophila light* and *rolled* genes, are single-copy and essential for viability. Requirement for activity of these genes might have resulted in a robust adaptation strategy, one that enables, or even promotes, continued expression in spite of the bombardment of TEs.

Species can differ in their ability to tolerate TE insertions or expansion of repetitive DNAs. For instance, *Arabidopsis* seems to have constraints on intron size and hence gene size. By contrast, large introns (exceeding 10 Kb, the typical size of a TE) are well tolerated in *Drosophila* and humans. DNA methylation is also a factor that can influence levels of gene expression, TE regulation and heterochromatin in some organisms. However, it does not seem to have a major role in these processes in *Drosophila*.

Genome sequencing and comparative genomic approaches will identify additional genes that have euchromatic and heterochromatic paralogs or orthologs. Depending on the direction of gene movement and the age of the initiating event, we can expect to find genes and chromatin domains that are 'more' or 'less' euchromatic or heterochromatic. Overall, this predicts a dynamic continuum of transitional states between what is traditionally viewed as euchromatin and heterochromatin existing among species. These 'intermediate states' could be structurally similar to the euchromatin–heterochromatin 'transition zones' and pericentromeric regions that occur naturally on chromosomes. We suggest that further study of the regulation of genes caught in these evolutionary and chromosomal transition zones will be especially informative for discovering the variety of ways that chromatin can affect gene expression.

Concluding remarks

When Heitz invented the term 'heterochromatin' he knew something about genes and their relative locations, but nothing about the chromosomal distributions of repetitive DNAs, chromosomal proteins, or histone modifications. His cytologic definition recognized a visible distinction in the behavior of chromosomes and chromosomal segments so the historically correct definition of heterochromatin remains naturally limited by low resolution of microscopy. As high-resolution molecular maps of different

Table 1. Summary of views on heterochromatin

Widely cited views	Diverse properties of 'constitutive' heterochromatin	Refs
Heterochromatic genes are rare	*Drosophila melanogaster* has hundreds of protein-coding heterochromatic genes Heterochromatic genes have been identified in plants and mammals	[22]; this review
Heterochromatin is antagonistic to gene expression	Heterochromatin has repressive, permissive and activating effects on gene expression It contains protein-coding genes, rRNA genes, noncoding RNA genes and expressed TE sequences in diverse organisms	[18,29]
Heterochromatin-enriched proteins repress gene expression	SU(VAR) proteins function in heterochromatin-induced silencing of euchromatic genes Many are also required for expression of several *Drosophila* euchromatic and heterochromatic genes	[4,13,37]
Heterochromatin is late replicating compared with euchromatin	The replication timing of some heterochromatic regions is coincident with that of euchromatin	[67]
Heterochromatin is characterized by uniformly spaced nucleosome arrays and is inaccessible to the transcriptional machinery	Ordered nucleosomal arrays and decreased accessibility to protein probes has been demonstrated for several heterochromatic regions 'Open' conformation has been demonstrated within two expressed heterochromatic sequences Dynamic exchange of heterochromatin proteins has been demonstrated	[5,68,77,78]
Heterochromatin is deeply staining chromatin that is visible throughout the cell cycle	The cytologic definition of Heitz is limited by low resolution Boundaries between heterochromatin and euchromatin are not always precise; intermediate or transition zones can exist Underlying molecular composition of heterochromatin can vary depending on specific region and cell type	[1,11]; this review
Constitutive heterochromatin is distinct from facultative heterochromatin	Underlying DNA sequence has a deterministic role in constitutive heterochromatin, but not facultative heterochromatin Even 'constitutively' heterochromatic regions can change cytologic appearance and activity depending on the cell type and developmental stage Common to all heterochromatin is epigenetic control and, especially for repeat-rich heterochromatin, there is a capacity for 'long-range' propagation	[69]; this review

chromosomal regions emerge, it is clear that there is no single set of molecular markers common to all regions designated by various investigators as 'heterochromatin'. Nonetheless, there is currently a tendency to restrict the term heterochromatin and equate it with 'silent' or 'H3K9me-enriched' chromatin. In extreme cases, heterochromatin has been loosely used to refer to any transcriptionally inactive chromatin in differentiated cells and a single transposable element in euchromatin has also been referred to as 'cryptic' heterochromatin [59]. Such 'redefinitions' of heterochromatin have neglected its historical significance and its current utility as a 'higher-order' chromatin term (i.e. appropriate for defining broad categories of substances). We prefer to use heterochromatin in a manner that preserves the notion of an epigenetically controlled chromatin state with a capacity for 'long-range' propagation. In repeat-rich heterochromatin, this distance can be kilobases or even megabases of DNA, depending on the organism and specific region of heterochromatin, but it probably reflects a cooperative assembly process. Our use of the term 'heterochromatin' accommodates molecularly diverse forms of heterochromatin, including known heterochromatic domains with repressive, permissive or activating effects on gene expression. The majority of functional studies of heterochromatin have so far focused on its repressive effects; however, the discovery of a wide variety of genes in or near heterochromatin and new tools for probing heterochromatic sequences at higher molecular resolution provides opportunities to determine if apparent 'exceptions' to the commonly held views of heterochromatin are more widespread in nature (Table 1). At minimum, clarity in the use of the term heterochromatin will improve if investigators make it clear which chromatin attributes are confirmed for the region being studied and avoid assuming that experimentally untested features exist based on a monolithic view of heterochromatin.

We predict that further studies of the organization and regulation of actively transcribed heterochromatic genes will broaden views of how chromosomal context and chromatin structure affect gene expression. In addition, a more accurate assessment of gene content of heterochromatic regions will contribute toward the actual completion of genome projects in organisms in which heterochromatin occupies a sizable fraction of the genome. Heterochromatin has presented geneticists and genome scientists with more than its share of technical challenges, and repeat-rich sequences remain largely underrepresented in genome projects. However, improved coverage of heterochromatin is expected with the application of a whole-genome shotgun approach for sequencing genomes, directed projects that specifically focus on the heterochromatic regions, and the generation of new strategies to gain genetic and molecular footholds in the repeat-rich chromosomal regions [22,65]. If evolutionary fluidity between euchromatic and heterochromatic regions is widespread among species, as suggested by the studies reviewed here, then comparative genomics will continue to provide an especially profitable avenue for heterochromatic gene discovery.

Acknowledgements
We thank Roger Hoskins, Chris Smith and the anonymous reviewers for helpful comments. Our work on heterochromatic genes is supported by a National Institutes of Health Predoctoral Training Grant Fellowship T32HD007183–26 (to J.C.Y.), a National Science Foundation grant and a Washington Research Foundation Professorship Award (to B.T.W.).

References
1 Heitz, E. (1928) Das heterochromatin der moose. *I. Jb. wiss. Bot.* 69, 762–818
2 Muller, H.J. (1930) Types of visible variations induced by X-rays in *Drosophila*. *J. Genet.* 22, 299–335
3 Baker, W.K. (1968) Position-effect variegation. *Adv. Genet.* 14, 133–169
4 Schotta, G. *et al*. (2003) Position-effect variegation and the genetic dissection of chromatin regulation in *Drosophila*. *Semin. Cell Dev. Biol.* 14, 67–75
5 Grewal, S.I.S. and Elgin, S.C.R. (2002) Heterochromatin: new possibilities for the inheritance of structure. *Curr. Opin. Genet. Dev.* 12, 178–187
6 Jenuwein, T. (2001) Re-SET-ing heterochromatin by histone methyltransferases. *Trends Cell Biol.* 11, 266–273
7 Lippman, Z. and Martienssen, R.A. (2004) The role of RNA interference in heterochromatic silencing. *Nature* 431, 364–370
8 Disteche, C. *et al*. (2002) Escape from X inactivation. *Cytogenet. Genome Res.* 99, 36–43
9 Brown, C.J. and Greally, J.M. (2003) A stain upon silence: genes escaping X inactivation. *Trends Genet.* 19, 432–438
10 Sturtevant, A.H. (1965) *A History of Genetics*, Harper and Row
11 Zacharias, H. (1995) Emil Heitz (1892–1965): Chloroplasts, heterochromatin, and polytene chromosomes. *Genetics* 141, 7–14
12 Schultz, J. (1936) Variegation in *Drosophila* and the inert chromosomal regions. *Proc. Natl. Acad. Sci. U. S. A.* 22, 27–33
13 Hearn, M.G. *et al*. (1991) The effect of modifiers of position effect variegation on the variegation of heterochromatic genes of *Drosophila melanogaster*. *Genetics* 128, 785–797
14 Weiler, K.S. and Wakimoto, B.T. (2002) Suppression of heterochromatic gene variegation can be used to distinguish and characterize *E(var)* genes potentially important for chromosome structure. *Mol. Genet. Genomics* 266, 922–932
15 Hilliker, A.J. *et al*. (1980) The genetic analysis of *D. melanogaster* heterochromatin. *Cell* 21, 607–619
16 Pimpinelli, S. *et al*. (1986) The peculiar organization of *Drosophila* heterochromatin. *Trends Genet.* 13, 17–20
17 Dimitri, P. *et al*. (2005) The paradox of functional heterochromatin. *BioEssays* 27, 29–41
18 Gatti, M. and Pimpinelli, S. (1992) Functional elements in *Drosophila melanogaster* heterochromation. *Annu. Rev. Genet.* 26, 239–275
19 Sun, F-L. *et al*. (2000) The fourth chromosome of *Drosophila melanogaster:* Interspersed euchromatic and heterochromatic domains. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5340–5345
20 Adams, M.D. *et al*. (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195
21 Celniker, S.E. *et al*. (2002) Finishing a whole-genome shotgun: Release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biology.* DOI: 10.1186/gb-2002-3-12-research0079 (http://genomebiology.com/2002/3/12/research/0079)
22 Hoskins, R.A. *et al*. (2002) Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. *Genome Biology.* DOI: 10.1186/gb-2002-3-12-research0086 (http://genomebiology.com/2002/3/12/research/0086)
23 Yasuhara, J.C. *et al*. (2003) A strategy for mapping the heterochromatin of chromosome 2 of *Drosophila melanogaster*. *Genetica* 117, 217–226
24 Misra, S. *et al*. (2002) Annotation of the *Drosophila* euchromatic genome: a systematic review. *Genome Biology.* DOI: 10.1186/gb-2002-3-12-research0083 (http://genomebiology.com/2002/3/12/research/0083)
25 Carvalho, A.B. *et al*. (2000) Y chromosomal fertility factors kl-2 and kl-3 of *Drosophila melanogaster* encode dynein heavy chain polypeptides. *Proc. Natl. Acad. Sci. U. S. A.* 97, 13239–13244
26 Carvalho, A.B. *et al*. (2003) Y chromosome and other heterochromatic sequences of the *Drosophila melanogaster* genome: how far can we go? *Genetica* 117, 227–237

27 Dimitri, P. (1991) Cytogenetic analysis of the second chromosome heterochromatin of *Drosophila melanogaster*. *Genetics* 127, 553–564
28 Dimitri, P. *et al*. (2003) Colonization of heterochromatic genes by transposable elements in *Drosophila*. *Mol. Biol. Evol.* 20, 503–512
29 Weiler, K.S. and Wakimoto, B.T. (1995) Heterochromatin and gene expression in *Drosophila*. *Annu. Rev. Genet.* 29, 577–605
30 Yasuhara, J.C. *et al*. (2005) Evolution of heterochromatic genes of *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 102, 10958–10963
31 Marino-Ramirez, L. *et al*. (2005) Transposable elements donate lineage-specific regulatory sequences to host genomes. *Cytogenet. Genome Res.* 110, 333–341
32 Ma, J. (2005) Crossing the line between activation and repression. *Trends Genet.* 21, 54–59
33 Willy, P.J. *et al*. (2000) A basal transcription factor that activates or represses transcription. *Science* 290, 982–985
34 Smothers, J.F. and Henikoff, S. (2001) The hinge and chromo shadow domain impart distinct targeting of HP1-like proteins. *Mol. Cell. Biol.* 21, 2555–2569
35 Lu, B.Y. *et al*. (2000) Heterochromatin protein 1 is required for the normal expression of two heterochromatic genes in *Drosophila*. *Genetics* 155, 699–708
36 Piacentini, L. *et al*. (2003) Heterochromatin protein 1 (HP1) is associated with induced gene expression in *Drosophila* euchromatin. *J. Cell Biol.* 161, 707–714
37 Cryderman, D.E. *et al*. (2005) Role of *Drosophila* HP1 in euchromatic gene expression. *Dev. Dyn.* 232, 767–774
38 Hwang, K.K. *et al*. (2001) Transcriptional repression of euchromatic genes by *Drosophila* heterochromatin 1 and histone modifiers. *Proc. Natl. Acad. Sci. U. S. A.* 98, 11423–11427
39 Greil, F. *et al*. (2003) Distinct HP1 and Su(var)3-9 complexes bind to sets of developmentally coexpressed genes depending on chromosomal location. *Genes Dev.* 17, 2825–2838
40 Liu, L-P. *et al*. (2005) Sex-specific role of *Drosophila melanogaster* HP1 in regulating chromatin structure and gene transcription. *Nat. Genet.* 37, 1361–1366
41 Spierer, A. *et al*. (2005) Loss of the modifiers of variegation *Su(var)3-7* or *HP1* impacts male X polytene chromosome morphology and dosage compensation. *J. Cell Sci.* 118, 5047–5057
42 de Wit, E. *et al*. (2005) Genome-wide HP1 binding in *Drosophila*: developmental plasticity and genomic targeting signals. *Genome Res.* 15, 1265–1273
43 Perrini, B. *et al*. (2004) HP1 controls telomere capping, telomere elongation and telomere silencing by two different mechanisms in *Drosophila*. *Mol. Cell* 15, 467–476
44 Hiragami, K. and Festenstein, R. (2005) Heterochromatin protein 1: a pervasive controlling influence. *Cell. Mol. Life Sci.* 62, 2711–2726
45 Schotta, G. *et al*. (2002) Central role of *Drosophila* SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J.* 21, 1121–1131
46 Maside, X. *et al*. (2005) Fixation of transposable elements in the *Drosophila melanogaster* genome. *Genet. Res.* 85, 195–203
47 McCollum, A.M. *et al*. (2002) Evidence for the adaptive significance of an LTR retrotransposon sequence in a *Drosophila* heterochromatic gene. *BMC Evol. Biol.* 2, 5–11
48 Tulin, A. *et al*. (2002) The *Drosophila* heterochromatic gene encoding poly(ADP-ribose) polymerase (PARP) is required to modulate chromatin structure during development. *Genes Dev.* 16, 2108–2119
49 Sun, F-L. *et al*. (2004) *cis*-acting determinants of heterochromatin formation on *Drosophila melanogaster* chromosome 4. *Mol. Cell. Biol.* 24, 8210–8220
50 Copenhaver, G.P. *et al*. (1999) Genetic definition and sequence analysis of *Arabidopsis* centromeres. *Science* 286, 2468–2474
51 Nagaki, K. *et al*. (2004) Sequencing of a rice centromere uncovers active genes. *Nat. Genet.* 36, 138–145
52 Yan, H. *et al*. (2005) Transcription and histone modifications in the recombination-free region spanning a rice centromere. *Plant Cell* 17, 3227–3238
53 Li, L. *et al*. (2005) Tiling microarray analysis of rice chromosome 10 to identify the transcriptome and relate its expression to chromosome architecture. *Genome Biol*. DOI: 10.1186/gb-2005-6-6-r52 (http://genomebiology.com/2005/6/6/R52)
54 Guyot, R. *et al*. (2005) Complex organization and evolution of the tomato pericentric region at the *FER* gene locus. *Plant Physiol.* 138, 1205–1215
55 Kim, J-S. *et al*. (2005) Comprehensive molecular cytogenetic analysis of sorghum genome architecture: distribution of euchromatin, heterochromatin, genes, and recombination in comparison to rice. *Genetics* 171, 1963–1976
56 She, X. *et al*. (2004) The structure and evolution of centromeric transition regions in the human genome. *Nature* 430, 857–864
57 Mudge, J.M. and Jackson, M.S. (2005) Evolutionary implications of pericentromeric gene expression in humans. *Cytogenet. Genome Res.* 108, 47–57
58 McCombie, W.R. *et al*. (2000) The complete sequence of a heterochromatic island from a higher eukaryote. *Cell* 100, 377–386
59 Lippman, Z. *et al*. (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430, 471–476
60 Steinemann, S. and Steinemann, M. (2005) Y chromosomes: born to be destroyed. *BioEssays* 27, 1076–1083
61 Carvalho, A.B. *et al*. (2001) Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13225–13230
62 Carvalho, A.B. (2002) Origin and evolution of the *Drosophila* Y chromosome. *Curr. Opin. Genet. Dev.* 12, 664–668
63 Carvalho, A.B. and Clark, A.G. (2005) Y chromosome of *D. pseudoobscura* is not homologous to the ancestral *Drosophila*. *Science* 307, 108–110
64 Fransz, P.F. *et al*. (2000) Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. *Cell* 100, 367–376
65 Eichler, E.E. *et al*. (2004) An assessment of the sequence gaps: unfinished business in a finished human genome. *Nat. Rev. Genet.* 5, 345–354
66 Saffery, R. *et al*. (2003) Transcription within a functional human centromere. *Mol. Cell* 12, 509–516
67 Schubeler, D. *et al*. (2002) Genome-wide DNA replication profile for *Drosophila melanogaster*: a link between transcription and replication timing. *Nat. Genet.* 32, 438–442
68 Festenstein, R. *et al*. (2003) Modulation of Heterochromatin Protein 1 dynamics in primary mammalian cells. *Science* 299, 719–721
69 Grigoryev, S.A. *et al*. (2004) Dynamic relocation of epigenetic chromatin markers reveals an active role of constitutive heterochromatin in the transition from proliferation to quiescence. *J. Cell Sci.* 117, 6153–6162
70 Weiler, K.S. and Wakimoto, B.T. (1998) Chromosome rearrangements induce both variegated and reduced, uniform expression of heterochromatic genes in a development-specific manner. *Genetics* 149, 1451–1464
71 Wakimoto, B.T. and Hearn, M.G. (1990) The effects of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of *D. melanogaster*. *Genetics* 125, 141–154
72 Eberl, D. *et al*. (1993) The role of heterochromatin in the expression of a heterochromatic gene, the *rolled* gene of *Drosophila melanogaster*. *Genetics* 134, 277–292
73 Schulze, S.R. *et al*. (2005) A genetic and molecular characterization of two proximal heterochromatic genes on chromosome 3 of *Drosophila melanogaster*. *Genetics* 169, 2165–2177
74 Howe, M. *et al*. (1995) *Cis*-effects of heterochromatin on euchromatic and heterochromatic gene expression in *Drosophila melanogaster*. *Genetics* 140, 1033–1045
75 Delattre, M. *et al*. (2000) The genomic silencing of position effect variegation in *Drosophila melanogaster*: interaction between the heterochromatin associated proteins Su(var)3-7 and HP1. *J. Cell Sci.* 113, 4253–4261
76 Clegg, N.J. *et al*. (1998) Suppressors of position-effect variegation in *Drosophila melanogaster* affect expression of the heterochromatic *light* gene in the absence of chromosome rearrangement. *Genome* 41, 495–503
77 Cheutin, T. *et al*. (2003) Maintenance of stable heterochromatic domains by dynamic HP1 binding. *Science* 229, 721–725
78 Sun, F.-L. *et al*. (2001) Long-range nucleosome ordering is associated with gene silencing in *Drosophila melanogaster* pericentric heterochromatin. *Molecular and Cellular Biology* 21, 2867–2879