A new portrait of constitutive heterochromatin: lessons from *Drosophila melanogaster*

René Massimiliano Marsano[1], Ennio Giordano[2], Giovanni Messina[3] and Patrizio Dimitri[3]

[1]Dipartimento di Biologia, Università di Bari, Bari, Italia;

[2]Dipartimento di Biologia, Università degli Studi di Napoli Federico II, Napoli, Italia;

[3]Dipartimento di Biologia e Biotecnologie "Charles Darwin", Sapienza Università di Roma, Istituto Pasteur Italia, Fondazione Cenci-Bolognetti, Roma, Italia

ABSTRACT

Constitutive heterochromatin represents a significant portion of eukaryotic genomes, but its functions still need to be elucidated. Even in the most updated Genetics and Molecular Biology textbooks, constitutive heterochromatin is portrayed mainly as the "silent" component of eukaryotic genomes. However, there may be more complexity to the relationship between heterochromatin and gene expression. In the fruit fly *Drosophila melanogaster*, a model for heterochromatin studies, about one-third of the genome is heterochromatic and is concentrated in the centric, pericentric and telomeric regions of the chromosomes. Recent findings indicate that hundreds of *D. melanogaster* genes can "live and work" properly within constitutive heterochromatin. The genomic size of these genes is generally very large and together they account for a significant fraction of the entire constitutive heterochromatin. Thus, this peculiar genome component in spite its ability to induce silencing, has in fact the means for being quite dynamic. A major scope of this review is to revisit the "dogma of silent heterochromatin".

Breaking the silence of constitutive heterochromatin

The term heterochromatin was introduced by Emil Heitz in 1928 to indicate chromosomal regions that retain a compact state throughout all stages of the cell cycle, as opposed to euchromatin that undergoes decondensation and condensation cycles [1].

Heterochromatin was further categorized into facultative and constitutive [2], with each one of these states being related to a "silenced" pattern (BOX 1). Over time, with the discovery of other properties of constitutive heterochromatin, this ubiquitous genomic component became increasingly synonymous with gene silencing (BOX 1).

More recently, the role of epigenetic modifications and non-coding RNAs in mediating the assembly and silencing properties of heterochromatin has been addressed in studies that have made use of *S. pombe*, *A. thaliana, C. elegans,* and *D. melanogaster* [6-9]. From these studies, a general scenario has emerged whereby constitutive heterochromatin shares similar structural and functional features in different organisms, in terms of compact organization, expression-repressive chromatin marks (HP1, Suvar3-9, H3 and H4 hypoacetylation, H3K9 methylation) and associated non-coding RNAs. These results have been extensively reviewed elsewhere [7-10] and will be not re-examined in detail here.

The notion that constitutive heterochromatin is incompatible with gene expression (Table 1; BOX 1) no longer seems to be a general rule. Indeed, pericentromeric regions of chromosomes express non-coding RNA transcripts in various organisms, which in turn contribute to promote heterochromatin formation and silencing of transposable elements (TEs) [7-11]. Some of these transcripts resemble the non-coding RNAs involved in the establishment of X-chromosome inactivation in mammalian females [4,12]. Moreover, transcription occurs at the centromeric level, albeit at low rates, and is thought to play a role in kinetochore organization and function [11].

Before the discovery of pericentromeric non-coding RNA transcripts, well-documented instances of heterochromatin activity have been reported in *D. melanogaster*. Combined genetics and cytological approaches have initially identified about forty essential genes embedded within constitutive heterochromatin [13-20]. Among these genes, *light* was the first protein-coding gene to be molecularly characterized [21]. Additionally, different Y chromosome satellite DNAs and heterochromatic copies of TEs were found to be transcriptionally active [22-24].

More recently, sequencing and annotation of the *D. melanogaster* genome have facilitated studies mapping the organization and function of constitutive heterochromatin [25-30]. Together, available results indicate that *D. melanogaster* constitutive heterochromatin contains

a minimum of 230 protein-coding genes [27], a significantly greater number than that defined by forward genetic analysis.

The mitotic chromosomes from larval neuroblasts of *D. melanogaster* provide a complete representation of the pericentric, centric, and Y chromosome heterochromatin with the mapping of genes (Figure 1; [13]).

Even with these results, the existence of functional *D. melanogaster* heterochromatic genes is often overlooked in Genetics and Molecular Biology textbooks (Table 1) and recent reviews. Instances of transcriptional activity associated with constitutive heterochromatin (both pericentromeric and centromeric) refer mainly – if not exclusively - to non-coding RNAs.

A possible explanation for this "lapse" may be due to at least two factors: i) a misleading use of the term "heterochromatic", often referred to euchromatic genes subjected to silencing (facultative heterochromatin) and ii) the idea that heterochromatic genes are merely a peculiarity of *D. melanogaster*, an exception to the rule. Conversely, heterochromatic genes have been found in genomes of different animal and plant species, such as yeast, *Arabidopsis*, tomato, rice, mouse and humans [31-39], but the ease of genetic manipulation has clearly enabled a wider identification in *Drosophila*.

On the whole, although silencing is the main recognized hallmark of constitutive heterochromatin, it may be only one face of the coin.

A major scope of this review is to revisit the dogma of "silent" heterochromatin and provide an updated picture of its peculiar genetic and molecular nature. Calling attention to experimental evidence for functional genetic elements resident in constitutive heterochromatin is important to help understanding both the epigenetic regulation of gene expression and the evolutionary dynamics shaping eukaryotic genomes.

The heterochromatic genes *of Drosophila melanogaster*

The Y chromosome single-copy genes

The Y chromosome of *D. melanogaster* contains about 40.9Mb of DNA [25] and has wide effects on male fertility and genome-wide gene expression [14, 17, 40, 41].

Previous studies identified a well-defined set of genetic functions on the Y chromosome, despite its heterochromatic nature: six fertility factors (*kl-1*, *kl-2*, *kl-3*, *kl-5*, *ks-1*, and *ks-2*) required for male fertility [14, 17, 40]; the *Mst77Y* gene cluster associated with male-related functions [42]; the *bobbed* locus (*bb*) which correspond to the rDNA gene cluster [43, 44] and the *crystal*/*Suppressor of Stellate* locus, whose deficiency causes accumulation of needle-like crystals in the primary spermatocytes and an abnormal meiosis [45-47].

The release 6 of the *D. melanogaster* reference genome sequence increased the Y chromosome assembly 10-fold, from 242 Kb to 3.4 Mb [30]. Recently, by assembling PacBio long single-molecule reads of heterochromatic sequences, Chang and Larracuente extended the overall Y chromosome size to 14.6 Mb [48], but large gaps are still present.

In addition to the fertility factors, sequence annotation identified 10 single-copy protein-coding genes on the Y chromosome, which escaped previous analyses (Figure 1; Table 2; Supplementary Tables 1 and 2). Many of these genes have paralogs on the autosomes, with a conserved exon-intron structure, suggesting their origin may be by DNA duplication, rather than by retrotransposition [40, 49].

The protein products of the Y chromosome fertility factors are required for proper spermiogenesis (Table S1; 40; 49-53). Cytologically, *kl-5, kl-3, kl-1* and *ks-1* were estimated to span at least 3.2, 4.3, 1.2, and 3.2 Mb of DNA, respectively [17, 54]. This prediction was confirmed at the molecular level, in that *kl-2*, *kl-3*, *kl-5* and *ks-1* were found to be very large, due to megabase-sized introns rich in repetitive sequences [40, 55]. In the Release 6 of the *Drosophila* genome, however, many of these Y-linked genes still lacked most of the intronic sequences, due to their repetitive nature [30, 56]. This problem has been partially overcome by Chang and Larracuente [48] allowing the identification of additional genomic portions of the Y chromosome genes.

The minimum size currently estimated for *kl-5, kl-3*, and *kl-2* is 0.354, 0.529, and 0.305 Mb, respectively; while *kl-1-WDY*, *ks-1/ORY* and *PPr-Y* could reach at least 2.4, 0.5, and 1.432 Mb. The genomic sequence of *kl5, PRY, kl-3, kl-2, ks-1/ORY* and *ks-2/CCY* genes are still incomplete, while only traces of sequences are found for other Y-linked genes in Release 6 (*Pp1-Y1, ARY, FDY, CG46191, CG46192, CG46193* and *Pp1-Y2*). A complete and updated estimation of the Y-linked gene size is shown in Tables S1 and S2.

Repetitive loci of the sex chromosome heterochromatin.

The rDNA gene clusters (18S, 5.8S, and 28S rRNAs) are found in both X heterochromatin and Y chromosome [57]. The X and Y clusters consist of about 2.2 Mb and 2.8 Mb, respectively mapping to mitotic regions h29 and h20 (Figure 1; Table 2; Tables S1 and S2). The basic rDNA unit is 11.5 Kb, but can differ in the intergenic spacer length (IGS), the internal transcribed spacer, and the distribution of R1 and R2 elements [58-60]. The rDNA copy number ranges from 80 to 600 [60-63], where individuals with <130 copies display the *bobbed* mutant phenotype [64]. Recently, Chang and Larracuente [48] identified 56 copies of a 1995 bp 18s rDNA gene (111.720 Kb), 238 copies of a 3945 bp 28s rDNA gene (938.910 Kb), and 721 copies

of IGS repeats on the Y chromosome (4.326 Mb, considering a weighted average length of 6Kb; [65]).

The *crystal (cry)* or *Suppressor of Stellate* [*Su(Ste)*] locus, together with the X-linked *Stellate* (*Ste*) sequences are components of a peculiar phenomenon of suppression of repetitive elements expression [45-47], mediated by the RNAi pathway in *D. melanogaste*r [66]. The cry/*Su(Ste*) locus maps to the Y chromosome cytogenetic region h11 (Figure 1; Supplementary Table 1) and is mainly composed of 2.5 and 2.8 Kb tandemly repeated sequences [45, 67].

Chang and Larracuente identified 627 *Su*(*Ste*) tandemly repeated copies on the Y chromosome which are frequently flanked by TEs [48]. Given that, the size of the cry/*Su*(*Ste*) region is approximately 1.6 Mb (Table 2; Tables S1 and S2).

The *Mst77Y* gene cluster is an interesting case of genomic evolution [42] and its organization was recently re-evaluated. Using long-read single molecule sequencing technology, Krsticevic et al. [42], found that the *Mst77Y* region spans 96 Kb and originated from a 3.4 Kb transposition from chromosome 3L to the Y chromosome, followed by tandem duplications targeted by TEs. The current prediction includes 10 functional copies and 8 pseudogenes, all showing the same orientation (Table 2; Tables S1 and S2).

Two *Ste* clusters map to the X chromosome, one in euchromatin (polytene chromosome region 12E) and the other in the region h26 of mitotic heterochromatin (Figure 1; [47]).

The X chromosome heterochromatin also carries a group of still molecularly uncharacterized loci, whose genetic behavior suggests a repeated nature: *Abnormal Oocyte* (*ABO*), *compensatory response* (*cr*), and *Ribosomal exchange* (*Rex*). These loci, similarly to cry/*Su*(*Ste*)*,* are part of genetic systems that involve specific interactions between heterochromatic and euchromatic genetic elements [18, 46, 47].

Single-copy genes, non-coding RNAs genes and pseudogenes in the heterochromatin of chromosomes X, 2, and 3

Based on the euchromatic/heterochromatic borders defined in Figure 2, we have analysed the current heterochromatin sequence annotation and retrieved different classes of sequences, i.e. protein coding genes, non-coding RNA genes and pseudogenes, respectively, from chromosomes X, 2, and 3 (Table 2 and Tables S2, S3, S4, S5, S6 and S7).

Data obtained by bioinformatic and reverse genetics approaches indicate that most of the annotated genes encode evolutionarily conserved products involved in important cellular and developmental processes, which are shared by euchromatic genes [27, 28, 69]. However, in many cases the molecular function of their protein product still needs to be clarified. A well-

known difference between *D. melanogaster* heterochromatic and euchromatic single-copy genes lies in their size and molecular structure [21, 27, 70, 71]. Notably the *Myosin 81F* and *Piezo-like* genes in 3Rh, occupy a genomic region of about 1.96 and 0.7 Mb, respectively, while *CG45782* in 3Lh and *CG17684* in 2Rh are both about 400Mb large (Tables S5, S6 and S7). These autosomal genes are among the largest found in *D. melanogaster* constitutive heterochromatin, together with the above-mentioned Y-linked genes (Figure 1; Tables S2 and S7). As for the Y chromosome fertility genes [40, 55], the large genomic size of autosomal heterochromatic genes is due to the presence of long introns composed of nested TE remnants [21, 27, 70, 71]. For example, the *Myosin 81F* gene introns account for 99.6% of its genomic region [30] (Table S7). The introns of heterochromatic genes have been estimated to be on average at least five-fold longer than those present in euchromatic genes [27], a value which is likely to be an underestimate since it does not take into account most of the mega-sized Y-linked genes.

Experimental data suggested that over evolutionary time, TEs not only have contributed to the build-up of the large introns of *Drosophila* heterochromatin, but might have even shaped the structural and functional organization of heterochromatic genes [72, 73].

Deciphering how heterochromatic gene pre-mRNA carrying megabase-sized introns can be transcribed and processed is an important challenge. In that respect, the HEPH protein has been recently suggested to play a crucial role in the splicing of *kl-5* fertility factor [74].

The heterochromatin of chromosome 2 also contains the *Responder* (*Rsp*) locus made up by tandem arrays of 120bp AT-rich DNA repeats targeted by the *Segregation Distorter* (*Sd*) gene [Figure 1; 75,76]. Interestingly, the *Rsp* region originates precursor piRNA transcripts in ovaries and testes, which are bound by *Piwi, Aubergine* and *Argonaute-3* proteins involved in the *Piwi*-interacting RNA (piRNA) pathway [76].

The chromosome 4

The dot chromosome 4 shows a peculiar genetic and molecular organization with both euchromatic and heterochromatic domains. Cytologically, chromosome 4 exhibits the banding pattern characteristic of euchromatin (bands and interbands). Moreover, the 1.2 Mb distal portion carrying about 80 genes was described as containing interspersed euchromatic and heterochromatic domains [77]. Such a peculiar genetic and molecular nature of chromosome 4 has been defined as a “quasi heterochromatic status" [73]. Since it is not obvious to distinguish between euchromatic and heterochromatic components of chromosome 4, we decided to not include it in our analysis.

The genomic features of chromosome 4 and its associated genes have been discussed in depth by Riddle and Elgin [78] and will be not re-examined here further.

A significant portion of the *D. melanogaster* constitutive heterochromatin is occupied by active genes

Overall, at least 205 protein-coding genes, 78 non-coding RNA genes and 105 pseudo-genes, together with *rDNA* and *Mst77* gene clusters, cry/*Su*(*Ste*) and *Ste* repeated sequences can be assigned to the constitutive heterochromatin of *D. melanogaster* (Figure 1; Table 2; Tables S1-S7).

Significant levels of transcription in diverse tissues and developmental stages were originally assessed for *light* [21], *rolled* [79], *Yeti* [80], *Nipped A* and other 12 single-copy heterochromatic genes of chromosomes 2 and 3 [28]. A more detailed scenario of the expression profile of genes in constitutive heterochromatin is given by the heatmaps visualization of modENCODE RNA-Seq data (Figure 3). It appears that most protein-coding genes are expressed at appreciable levels during developmental stages and in different tissues. The annotated non-coding RNAs (mostly lncRNAs) are also transcribed, although at a lower extent (Figure S1).

Considering the maximum genomic size estimated for the *kl-5*, *kl-1*, and *ks-1* (Tables 2 and S2), the fraction of constitutive heterochromatin occupied by active genes account for about 40% (30 Mb) and 20% (11 Mb) of Y-carrying gametes and X-carrying gametes, respectively, and 30% (41Mb) and 20% (22 Mb) of the male and female diploid genome, respectively (Table S8). However, the actual amount of constitutive heterochromatin occupied by active domains could be even greater. First, as previously discussed, due to the gaps in the Y-chromosome assembly the sequence of many Y-linked genes is incomplete or even scarce (Tables S1 and S2). Second, other loci such as ABO, *cr* and *Rex* (see above) remain molecularly unknown. Finally, transcription of heterochromatic satellite DNAs, which have been estimated to be 20% of the *Drosophila melanogaster* genome, may not be limited to the known examples [11, 22, 76, 82]..

Overall, although the functional heterochromatic genes are clearly fewer compared to the euchromatic ones, their genomic regions together accounts for a significant fraction of constitutive heterochromatin: a surprising conclusion that should finally contribute to dispel the dogma of silent constitutive heterochromatin.

Heterochromatic genes, piRNAs and circRNAs : a functional link?

A large fraction of Piwi-interacting RNAs (piRNAs), involved in the epigenetic silencing of TEs, are transcribed from multiple genomic loci named piRNA clusters [83]. Roughly 140 piRNA

clusters have been identified in *D. melanogaster*, many of which map to constitutive heterochromatin [83, 84], whose expression requires the H3K9me3 mark [85]. Most of these clusters, such as *Flamenco*, have been annotated as lncRNA in Release 6 (Table S3). It is worth noting that some piRNA clusters overlap with the genomic coordinates of single-copy genes found in pericentric heterochromatin [81]. For example, among 14 piRNA clusters of 3Rh, 8 fall within the large introns of *Myosin 81F* and 2 in those of *Piezo-like*.

Moreover, single copy genes from constitutive heterochromatin of *D. melanogaster* generate 90 circular RNAs (circRNAs) [86]. It has been hypothesized that circRNAs could act as regulators of many cellular processes [87], thus the observation that they also arise from heterochromatic genes adds additional functional value to constitutive heterochromatin.

It is tempting to speculate that the expression of the above-mentioned piRNAs clusters and circRNAs loci might depend on the transcriptional activity of heterochromatin genes within which they are found. If this was true, then mutations impairing the expression of a given heterochromatic gene harbouring piRNAs and circRNAs, might result in a pleiotropic effect on different cellular pathways e.g., expression of the gene product and biogenesis of the corresponding piRNAs and circRNAs, resulting in activation of TEs and impairment of gene expression modulation.

How can genes properly function in constitutive heterochromatin?

The presence of essential genes that "live and work" properly, albeit being located within a genomic environment with the long-term reputation of being inhospitable to transcription, clearly seemed a paradox [72, 73], raising questions on the mechanism(s) allowing gene expression in constitutive heterochromatin. The central questions are: how might their expression be compatible with the known silencing properties of heterochromatin, and what factors are involved? Does constitutive heterochromatin change its structural/functional state during development, thus making the genes accessible to transcription factors/machinery? Different models have been proposed to reconcile the repressive properties of heterochromatin with gene expression [73].

Notably, *D. melanogaster* heterochromatic genes are transcribed from promoter regions sharing basic similarities with those of euchromatic genes [71, 88]. However, contrary to what happens for euchromatic genes, the pericentric environment is a crucial regulatory requirement for heterochromatic genes such as *light* and *rolled*, because their expression is compromised when moved away from their native location to euchromatin by chromosome rearrangements [89, 90], presumably because they lose heterochromatic marks. From this

perspective, it can be hypothesized that factors required for constitutive heterochromatin formation may control the expression of heterochromatic genes.

Genetic and molecular studies have shown that the proper expression of heterochromatic genes *light, rolled, Rpl15,* and *Dbp80* depends on the HP1a protein [91-94], a well-known epigenetic regulator controlling the heterochromatic silencing [95]. In accord, large-scale mapping experiments carried out in *Drosophila* Kc cells have shown that both HP1a and SUVAR3-9 associate with pericentric genes that are preferentially expressed during embryogenesis [96]. Together, these findings concur in supporting the idea that HP1a and other epigenetic regulators are required to ensure the expression of genes located in constitutive heterochromatin. A role of HP1 as positive regulator has been also shown for euchromatic genes [97, 98].

Consistently, the HP1 paralog *rhino* marks the regions of dual-strand heterochromatic piRNA clusters through the binding with H3K9me3 and is required for piRNAs biogenesis in the female germline [99].

Histone modifications are also likely to be crucial players in heterochromatic gene expression. Yasuhara and Wakimoto found that H3K9me2 is depleted at the 5' end, but enriched throughout the body of heterochromatin genes [100]. Simultaneously, a strong peak of H4 acetylation at the gene promoters was observed. This profile is clearly different from that shown by euchromatic genes, where H3-di-methylated-at lysine 9 (H3K9me2) is found at a lesser level throughout the gene length. The authors suggested that heterochromatin genes are integrated into, rather than insulated from, the H3K9me2-enriched domain.

Subsequently, ChIP-chip analyses of histone modifications carried out on heterochromatic genes in different *Drosophila* cell lines and embryos revealed complex and dynamic patterns of chromosomal proteins and histone modifications [101]. Transcriptionally silent heterochromatic genes are characterized by the enrichment of HP1a, H3K9me2, and H3K9me3 silencing marks, while the majority of the active heterochromatic genes show both ‘‘active’’ marks (e.g., H3K4me3 and H3K36me3) and ‘‘silent’’ marks (e.g., H3K9me2 and HP1a). Overall, depletion of H3K9 methylation at the transcription start site is the hallmark of active heterochromatic genes, in agreement with previous findings [100].

Recently, using ChIP-seq data from modENCODE [102], the histone modification pattern of 60 heterochromatic pericentromeric genes of *D. melanogaster* was analysed during different developmental stages [103]. The results confirmed dynamic changes of histone modifications occurring on heterochromatic genes [96, 100, 101]. On average, both activation (H3K4me1/3, H3K9/27ac) and silencing (H3K9me3) marks occurred across heterochromatic genes, with

predominant levels of H3K9me3 on the gene body and a peak of activation marks at TSS. Heterochromatic genes were grouped in three classes that differ on the distribution/presence of H3K9m3 and activation marks in the stages greatest and lowest expression [103]. Representative of these classes are *light*, *rolled* and *Nipped A*, three essential genes of 2Rh defined by genetics and molecular analyses. The *light* gene (group I) showed both inactive and active histone marks at the greatest expression stage, and only inactive marks at the lowest stage, while *rolled* (Group II) had mainly active marks at both expression stages, a pattern characteristic of most euchromatic genes. Intriguingly, *Nipped A* (class III) had inactive marks at both expression stages and exhibited a higher expression compared to *rolled* (Fig. S1).

As previously recalled, a common feature of autosomal heterochromatin genes is that they need a native heterochromatic context in order to function [89, 90] and depend on HP1a protein [90-93]. Thus, although the observed differences in histone marks may contribute to modulate the expression of heterochromatic genes during development, the primary factors/conditions ensuring gene expression in constitutive heterochromatin still need to be elucidated. HP1a and other epigenetic regulators are possible candidates. Notably, HP1a was also found to bind to TEs remnants in repeat-dense heterochromatic regions [104]. TE sequences located nearby heterochromatic genes might act as cis-regulatory elements and recruit HP1a to promote transcription [72, 73].

The role of HP1a as crucial heterochromatin player has been recently further investigated by coupling biochemical purification with an image-based genome-wide RNAi screen [105]. This approach allowed the identification of a large number of novel heterochromatic proteins with diverse and dynamic localization patterns during the cell cycle that may contribute to regulate heterochromatin organization and function.

The involvement of chromatin remodelling complexes in heterochromatin regulation has been also highlighted by several studies. In yeast, the chromatin-remodeling factor FACT contributes to centromeric heterochromatin [106]. In mammals, the SWI/SNF-like protein SMARCAD1 was suggested to promote the establishment of pericentric heterochromatin [107]. In *D. melanogaster* mutations in genes encoding Tip60 chromatin remodelling complex subunits are dominant suppressors of PEV [108]. In accord, HP1a was found to interact with YETI, a subunit of *Drosophila* Tip60 complex [80] and with other chromatin remodelling factors [106]. Together, these results strongly suggest that chromatin remodelling has an impact on the dynamic changes of constitutive heterochromatin observed during developmental stages and may play a role in modulating heterochromatic gene expression. Finally, transcription of heterochromatin may also be subjected to dynamic regulation during the cell cycle, as shown

for centromeric repeats in *S. pombe* and mouse minor satellite that increase in S/G2 phases [109, 110].

In conclusion, the expression of genes in constitutive heterochromatin of *D. melanogaster* during development or cell cycle may be differentially regulated, possibly depending on changes in accessibility to transcription machinery mediated by the combined action of epigenetics players and chromatin remodelling complexes.

The evolutionary history of the single-copy genes of *Drosophila melanogaster* heterochromatin

Few studies have investigated the evolutionary history of *D. melanogaster* heterochromatin genes by comparing putative orthologous genes. Yasuhara et al. [71] analysed a 2L heterochromatin gene cluster spanning 594 Kb, including the *light* gene, while Shultze et al. [94] studied *RPL15* and *Dbp80,* located in chromosome 3 heterochromatin. *D. melanogaster* orthologs *D. pseudoobscura* and *D. virilis* mapped to euchromatin, suggesting gene repositioning during genome evolution in the Drosophilidae lineage.

More recently, Caizzi et al., [111] carried out a comparative genomic analysis on a group of 53 single-copy protein genes located in pericentric heterochromatin of chromosome 2, spanning several Mb. Orthologs of *D. melanogaster* heterochromatic genes were found clustered at three main syntenic regions in the *D. virilis* and *D. pseudoobscura* genomes. In *D. virilis* the clusters are located in the euchromatin, across a few hundred Kb region with low repeat content, while in *D. pseudoobscura* the clusters are associated with the distal portions of pericentric heterochromatin enriched in repeated sequences.

These results indicate that the *D. melanogaster* chromosome 2 heterochromatin genes mainly arose through an evolutionary repositioning of ancestral gene clusters located in the euchromatin of progenitor species. Later, the structure of the genes would have dramatically changed due to a remarkable increase in the intron size following recurrent insertions of TEs. A similar trend was suggested for the evolution of the *light* and other heterochromatic genes of chromosome 3 [71, 94]. Remarkably, in both *D. virilis* and *D. pseudoobscura* the clusters show an unexpected association with the evolutionarily conserved HP1a protein [111]. This conserved association suggests the intriguing possibility that an ancestral HP1-like protein (and other epigenetic regulators) may have contributed to the success of gene clusters repositioning into pericentromeric regions.

Concluding Remarks and Future Perspectives

A significant challenge in current genomic research is represented by the elusive nature of heterochromatin structural and functional organization in model organisms and in humans. Due to the development of whole genome sequencing and annotation, we have now a deeper understanding of the molecular and functional organization of *D. melanogaster* constitutive heterochromatin. The emerging picture is that a variety of functional sequences unexpectedly account for a large portion of constitutive heterochromatin (at least to 40% of Y-carrying gametes and 30% of diploid male genome), showing that this peculiar component of the *D. melanogaster* genome has in fact the means for being quite "eloquent".

Despite the significant progresses achieved by genomic analysis of *D. melanogaster* constitutive heterochromatin, the function of many single-copy genes still need elucidating. In addition, some heterochromatic loci may have remained undisclosed due to the gaps still present in the sequence assembly and to their peculiar genetic organization. In fact, as we have learned from the genetic nature of *Responder* and *crystal*/*Su(Ste)* elements, constitutive heterochromatin also contains "cryptic loci" that escape both forward and reverse genetic analyses [47, 75, 76]. Therefore, more studies are required to get a complete picture on the multiplicity of functional genetic elements harboured by *Drosophila* constitutive heterochromatin.

Noteworthy, functional genes in constitutive heterochromatin are not restricted to *D. melanogaster*, having been also identified in other species [31-39]. Once the current large-scale comparative analysis of thousand animal and plant genomes will provide a more detailed scenario of constitutive heterochromatin, the presence of functional genes or unconventional genetic elements in this ubiquitous component of eukaryotic cells is likely to be more widespread than we may now imagine.

Overall, available studies suggest that constitutive heterochromatin and euchromatin may be regarded as two different and yet dynamic genomic domains, both of which can be active or repressed during developmental stages or cell cycle phases, as a consequence of different regulatory strategies they have acquired to control gene expression.

Acknowledgements

We are grateful to Patrizia Lavia and Amanda Larracuente for helpful comments and suggestions, and to Pietro D'Addabbo for helpful suggestion in the management of large datasets, and to the Flybase Staff for sharing bed files of the chromatin tracks implemented in GBrowse. P. Dimitri and Giovanni Messina were supported by grants from Istituto Pasteur-Fondazione Cenci Bolognetti.

REFERENCES

1. Heitz, E. (1928) Das Heterochromatin der Moose. Jb Wiss Bot 69, 762– 818.
2. Brown, S.W. (1966) Heterochromatin. Science 151 (3709), 417-25.
3. Elgin, S.C. and Reuter, G. (2013) Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. Cold Spring Harb Perspect Biol 5 (8), a017780.
4. Robert Finestra, T. and Gribnau, J. (2017) X chromosome inactivation: silencing, topology and reactivation. Curr Opin Cell Biol 46, 54-61.
5. Nur, U. (1990) Heterochromatization and euchromatization of whole genomes in scale insects (Coccoidea: Homoptera). Dev Suppl, 29-34.
6. Sun, F.L. et al. (2004) cis-Acting determinants of heterochromatin formation on Drosophila melanogaster chromosome four. Mol Cell Biol 24 (18), 8210-20.
7. Elgin, S.C. and Grewal, S.I. (2003) Heterochromatin: silence is golden. Curr Biol 13 (23), R895-8.
8. Guzzardo, P.M. et al. (2013) The piRNA pathway in flies: highlights and future directions. Curr Opin Genet Dev 23 (1), 44-52.
9. Martienssen, R. and Moazed, D. (2015) RNAi and heterochromatin assembly. Cold Spring Harb Perspect Biol 7 (8), a019323.
10. Allshire, R.C. and Madhani, H.D. (2018) Ten principles of heterochromatin formation and function. Nat Rev Mol Cell Biol 19 (4), 229-244.
11. Talbert, P.B. and Henikoff, S. (2018) Transcribing Centromeres: Noncoding RNAs and Kinetochore Assembly. Trends Genet 34 (8), 587-599.
12. Migeon, B.R. et al. (2001) Identification of TSIX, encoding an RNA antisense to human XIST, reveals differences from its murine counterpart: implications for X inactivation. Am J Hum Genet 69 (5), 951-60.
13. Gatti, M. and Pimpinelli, S. (1992) Functional elements in Drosophila melanogaster heterochromatin. Annu Rev Genet 26, 239-75.
14. Brosseau, G.E. (1960) Genetic Analysis of the Male Fertility Factors on the Y Chromosome of Drosophila Melanogaster. Genetics 45 (3), 257-74.
15. Hilliker, A.J. (1976) Genetic analysis of the centromeric heterochromatin of chromosome 2 of Drosophila melanogaster: deficiency mapping of EMS-induced lethal complementation groups. Genetics 83 (4), 765-82.
16. Marchant, G.E. and Holm, D.G. (1988) Genetic Analysis of the Heterochromatin of Chromosome 3 in Drosophila Melanogaster. II. Vital Loci Identified through Ems Mutagenesis. Genetics 120 (2), 519-32.
17. Gatti, M. and Pimpinelli, S. (1983) Cytological and genetic analysis of the Y chromosome of Drosophila melanogaster. Chromosoma 88 (5), 349-373.
18. Pimpinelli, S. et al. (1985) On biological functions mapping to the heterochromatin of Drosophila melanogaster. Genetics 109 (4), 701-24.
19. Dimitri, P. (1991) Cytogenetic analysis of the second chromosome heterochromatin of Drosophila melanogaster. Genetics 127 (3), 553-64.
20. Koryakov, D.E. et al. (2002) Cytogenetic analysis of the third chromosome heterochromatin of Drosophila melanogaster. Genetics 160 (2), 509-17.
21. Devlin, R.H. et al. (1990) The organization and expression of the light gene, a heterochromatic gene of Drosophila melanogaster. Genetics 125 (1), 129-40.
22. Bonaccorsi, S. et al. (1990) Transcription of a satellite DNA on two Y chromosome loops of Drosophila melanogaster. Chromosoma 99 (4), 260-6.

23. Bucheton, A. et al. (1992) I elements and the Drosophila genome. Genetica 86 (1-3), 175-90.
24. Pardue, M.L. et al. (1996) Drosophila telomeres: new views on chromosome evolution. Trends Genet 12 (2), 48-52.
25. Hoskins, R.A. et al. (2002) Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. Genome Biol 3 (12), RESEARCH0085.
26. Corradini, N. et al. (2003) FISH analysis of Drosophila melanogaster heterochromatin using BACs and P elements. Chromosoma 112 (1), 26-37.
27. Smith, C.D. et al. (2007) The Release 5.1 annotation of Drosophila melanogaster heterochromatin. Science 316 (5831), 1586-91.
28. Rossi, F. et al. (2007) Cytogenetic and molecular characterization of heterochromatin gene models in Drosophila melanogaster. Genetics 175 (2), 595-607.
29. Hoskins, R.A. et al. (2007) Sequence finishing and mapping of Drosophila melanogaster heterochromatin. Science 316 (5831), 1625-8.
30. Hoskins, R.A. et al. (2015) The Release 6 reference sequence of the Drosophila melanogaster genome. Genome Res 25 (3), 445-58.
31. Kuhn, R.M. et al. (1991) Clustered tRNA genes in Schizosaccharomyces pombe centromeric DNA sequence repeats. Proc Natl Acad Sci U S A 88 (4), 1306-10.
32. Copenhaver, G.P. et al. (1999) Genetic definition and sequence analysis of Arabidopsis centromeres. Science 286 (5449), 2468-74.
33. Nagaki, K. et al. (2004) Sequencing of a rice centromere uncovers active genes. Nat Genet 36 (2), 138-45.
34. Brun, M.E. et al. (2003) Juxtacentromeric region of human chromosome 21: a boundary between centromeric heterochromatin and euchromatic chromosome arms. Gene 312, 41-50.
35. Guyot, R. et al. (2005) Complex organization and evolution of the tomato pericentromeric region at the FER gene locus. Plant Physiol 138 (3), 1205-15.
36. She, X. et al. (2004) The structure and evolution of centromeric transition regions within the human genome. Nature 430 (7002), 857-64.
37. Mudge, J.M. and Jackson, M.S. (2005) Evolutionary implications of pericentromeric gene expression in humans. Cytogenet Genome Res 108 (1-3), 47-57.
38. Ryan, D.P. et al. (2010) Mutations in potassium channel Kir2.6 cause susceptibility to thyrotoxic hypokalemic periodic paralysis. Cell 140 (1), 88-98.
39. Ostromyshenskii, D.I. et al. (2018) Mouse chromocenters DNA content: sequencing and in silico analysis. BMC Genomics 19 (1), 151.
40. 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 (23), 13225-30.
41. Branco, A.T. et al. (2013) Chromatin-associated proteins HP1 and Mod(mdg4) modify Y-linked regulatory variation in the drosophila testis. Genetics 194 (3), 609-18.
42. Krsticevic, F.J. et al. (2015) Long-Read Single Molecule Sequencing to Resolve Tandem Gene Copies: The Mst77Y Region on the Drosophila melanogaster Y Chromosome. G3 (Bethesda) 5 (6), 1145-50.
43. Ritossa, F.M. et al. (1966) A molecular explanation of the bobbed mutants of Drosophila as partial deficiencies of "ribosomal" DNA. Genetics 54 (3), 819-34.
44. Williams, S.M. and Robbins, L.G. (1992) Molecular genetic analysis of Drosophila rDNA arrays. Trends Genet 8 (10), 335-40.
45. Hardy, R.W. et al. (1984) Cytogenetic analysis of a segment of the Y chromosome of Drosophila melanogaster. Genetics 107 (4), 591-610.
46. Livak, K.J. (1990) Detailed structure of the Drosophila melanogaster stellate genes and their transcripts. Genetics 124 (2), 303-16.

47. Palumbo, G. et al. (1994) Interaction systems between heterochromatin and euchromatin in Drosophila melanogaster. Genetica 94 (2-3), 267-74.
48. Chang, C.H. and Larracuente, A.M. (2019) Heterochromatin-Enriched Assemblies Reveal the Sequence and Organization of the Drosophila melanogaster Y Chromosome. Genetics 211 (1), 333-348.
49. Carvalho, A.B. et al. (2015) Birth of a new gene on the Y chromosome of Drosophila melanogaster. Proc Natl Acad Sci U S A 112 (40), 12450-5.
50. Goldstein, L.S. et al. (1982) Structural genes on the Y chromosome of Drosophila melanogaster. Proc Natl Acad Sci U S A 79 (23), 7405-9.
51. Gepner, J. and Hays, T.S. (1993) A fertility region on the Y chromosome of Drosophila melanogaster encodes a dynein microtubule motor. Proc Natl Acad Sci U S A 90 (23), 11132-6.
52. 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 (24), 13239-44.
53. Vibranovski, M.D. et al. (2008) Two new Y-linked genes in Drosophila melanogaster. Genetics 179 (4), 2325-7.
54. Bonaccorsi, S. et al. (1988) Y chromosome loops in Drosophila melanogaster. Genetics 120 (4), 1015-34.
55. Kurek, R. et al. (2000) Molecular aspects of intron evolution in dynein encoding mega-genes on the heterochromatic Y chromosome of Drosophila sp. Genetica 109 (1-2), 113-23.
56. Carvalho, A.B. et al. (2016) Improved assembly of noisy long reads by k-mer validation. Genome Res 26 (12), 1710-1720.
57. Ritossa, F. (1976) The Genetics and Biology of Drosophila.
58. Wellauer, P.K. et al. (1978) X and Y chromosomal ribosomal DNA of Drosophila: comparison of spacers and insertions. Cell 14 (2), 269-78.
59. Tautz, D. et al. (1988) Complete sequences of the rRNA genes of Drosophila melanogaster. Mol Biol Evol 5 (4), 366-76.
60. Lyckegaard, E.M. and Clark, A.G. (1991) Evolution of ribosomal RNA gene copy number on the sex chromosomes of Drosophila melanogaster. Mol Biol Evol 8 (4), 458-74.
61. Lyckegaard, E.M. and Clark, A.G. (1989) Ribosomal DNA and Stellate gene copy number variation on the Y chromosome of Drosophila melanogaster. Proc Natl Acad Sci U S A 86 (6), 1944-8.
62. Tartof, K.D. (1973) Regulation of ribosomal RNA gene multiplicity in Drosophila melanogaster. Genetics 73 (1), 57-71.
63. Terracol, R. and Prud'homme, N. (1986) Differential elimination of rDNA genes in bobbed mutants of Drosophila melanogaster. Mol Cell Biol 6 (4), 1023-31.
64. Mohan, J. and Ritossa, F.M. (1970) Regulation of ribosomal RNA synthesis and its bearing on the bobbed phenotype in Drosophila melanogaster. Dev Biol 22 (3), 495-512.
65. Mateos, M. and Markow, T.A. (2005) Ribosomal intergenic spacer (IGS) length variation across the Drosophilinae (Diptera: Drosophilidae). BMC Evol Biol 5, 46.
66. Aravin, A.A. et al. (2001) Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr Biol 11 (13), 1017-27.
67. McKee, B.D. and Satter, M.T. (1996) Structure of the Y chromosomal Su(Ste) locus in Drosophila melanogaster and evidence for localized recombination among repeats. Genetics 142 (1), 149-61.
68. Kharchenko, P.V. et al. (2011) Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature. 471:480-5.
69. Coulthard, A.B. et al. (2010) Essential loci in centromeric heterochromatin of Drosophila melanogaster. I: the right arm of chromosome 2. Genetics 185 (2), 479-95.

70. Dimitri, P. et al. (2003) Colonization of heterochromatic genes by transposable elements in Drosophila. Mol Biol Evol 20 (4), 503-12.
71. Yasuhara, J.C. et al. (2005) Evolution of heterochromatic genes of Drosophila. Proc Natl Acad Sci U S A 102 (31), 10958-63.
72. Dimitri, P. et al. (2005) The paradox of functional heterochromatin. Bioessays 27 (1), 29-41.
73. Yasuhara, J.C. and Wakimoto, B.T. (2006) Oxymoron no more: the expanding world of heterochromatic genes. Trends Genet 22 (6), 330-8.
74. Fingerhut, J.M. et al. (2018) Satellite DNA-containing gigantic introns in a unique gene expression program during Drosophila spermatogenesis. bioRxiv, 493254.
75. Hartl, D.L. and Hiraizumi, Y. (1976) Segregation Distortion. In: Novitsky E, Ashburner M (eds) The genetics and biology of Drosophila melanogaster, vol. 1b. Academic Press, New York, pp 615–666.
76. Larracuente, A.M. and Presgraves, D.C. (2012) The selfish Segregation Distorter gene complex of Drosophila melanogaster. Genetics 192 (1), 33-53.
77. 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 (10), 5340-5.
78. Riddle, N.C. and Elgin, S.C.R. (2018) The Drosophila Dot Chromosome: Where Genes Flourish Amidst Repeats. Genetics 210 (3), 757-772.
79. Berghella, L. and Dimitri, P. (1996) The heterochromatic rolled gene of Drosophila melanogaster is extensively polytenized and transcriptionally active in the salivary gland chromocenter. Genetics 144 (1), 117-25.
80. Messina, G. et al. (2014) Yeti, an essential Drosophila melanogaster gene, encodes a protein required for chromatin organization. J Cell Sci 127 (Pt 11), 2577-88.
81. Chintapalli, V.R. et al. (2007) Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet 39 (6), 715-20.
82. Usakin, L. et al. (2007) Transcription of the 1.688 satellite DNA family is under the control of RNA interference machinery in Drosophila melanogaster ovaries. Genetics 176 (2), 1343-9.
83. Brennecke, J. et al. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128 (6), 1089-103.
84. Brennecke, J. et al. (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing. Science 322 (5906), 1387-92.
85. Rangan, P. et al. (2011) piRNA production requires heterochromatin formation in Drosophila. Curr Biol 21 (16), 1373-9.
86. Ashwal-Fluss, R. et al. (2014) circRNA biogenesis competes with pre-mRNA splicing. Mol Cell 56 (1), 55-66.
87. Barrett, S.P. and Salzman, J. (2016) Circular RNAs: analysis, expression and potential functions. Development 143 (11), 1838-47.
88. Schulze, S.R. et al. (2006) Heterochromatic genes in Drosophila: a comparative analysis of two genes. Genetics 173 (3), 1433-45.
89. Wakimoto, B.T. and Hearn, M.G. (1990) The effects of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of Drosophila melanogaster. Genetics 125 (1), 141-54.
90. Eberl, D.F. et al. (1993) The role of heterochromatin in the expression of a heterochromatic gene, the rolled locus of Drosophila melanogaster. Genetics 134 (1), 277-92.
91. 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 (4), 785-97.
92. Clegg, N.J. et al. (1998) Suppressors of position-effect variegation in Drosophila melanogaster affect expression of the heterochromatic gene light in the absence of a chromosome rearrangement. Genome 41 (4), 495-503.

93. Lu, B.Y. et al. (2000) Heterochromatin protein 1 is required for the normal expression of two heterochromatin genes in Drosophila. Genetics 155 (2), 699-708.
94. Schulze, S.R. et al. (2005) A genetic and molecular characterization of two proximal heterochromatic genes on chromosome 3 of Drosophila melanogaster. Genetics 169 (4), 2165-77.
95. Eissenberg, J.C. and Elgin, S.C. (2014) HP1a: a structural chromosomal protein regulating transcription. Trends Genet 30 (3), 103-10.
96. 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 (22), 2825-38.
97. Piacentini, L. et al. (2003) Heterochromatin protein 1 (HP1) is associated with induced gene expression in Drosophila euchromatin. J Cell Biol 161 (4), 707-14.
98. Cryderman, D.E. et al. (2005) Role of Drosophila HP1 in euchromatic gene expression. Dev Dyn 232 (3), 767-74.
99. Klattenhoff, C. et al. (2009) The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell 138 (6), 1137-49.
100. Yasuhara, J.C. and Wakimoto, B.T. (2008) Molecular landscape of modified histones in Drosophila heterochromatic genes and euchromatin-heterochromatin transition zones. PLoS Genet 4 (1), e16.
101. Riddle, N.C. et al. (2011) Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. Genome Res 21 (2), 147-63.
102. mod, E.C. et al. (2010) Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science 330 (6012), 1787-97.
103. Saha, P. et al. (2018) Epigenomic and genomic landscape of Drosophila melanogaster heterochromatic genes. Genomics.
104. de Wit, E. et al. (2005) Genome-wide HP1 binding in Drosophila: developmental plasticity and genomic targeting signals. Genome Res 15 (9), 1265-73.
105. Swenson, J.M. et al. (2016) The composition and organization of Drosophila heterochromatin are heterogeneous and dynamic. Elife 5.
106. Lejeune, E. et al. (2007) The chromatin-remodeling factor FACT contributes to centromeric heterochromatin independently of RNAi. Curr Biol 17 (14), 1219-24.
107. Rowbotham, S.P. et al. (2011) Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. Mol Cell 42 (3), 285-96.
108. Qi, D. et al. (2006) Drosophila Reptin and other TIP60 complex components promote generation of silent chromatin. Genetics 174 (1), 241-51.
109. Chen, E.S. et al. (2008) Cell cycle control of centromeric repeat transcription and heterochromatin assembly. Nature 451 (7179), 734-7.
110. Ferri, F. et al. (2009) Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. Nucleic Acids Res 37 (15), 5071-80.
111. Caizzi, R. et al. (2016) Comparative Genomic Analyses Provide New Insights into the Evolutionary Dynamics of Heterochromatin in Drosophila. PLoS Genet 12 (8), e1006212.

Book	Authors	Sentences
Genetics: From genes to genomes, IV edition (2011)	Hartwell, Hood, Goldberg, Reynolds, Silver	1) Heterochromatin appears to be transcriptionally inactive for the most part, probably because it is so tightly packaged that the enzymes required for transcription of the few genes it contains cannot access the correct DNA sequences. 2) Most of the genes contained in heterochromatin are transcriptionally inactive or silenced.
Molecular Biology of the Cell VI edition (2014)	Alberts, Lewis, Raff and Roberts	The DNA in heterochromatin typically contains few genes, and when euchromatic regions are converted to a heterochromatic state, their genes are generally switched off as a result. However, we know now that the term heterochromatin encompasses several distinct modes of chromatin compaction that have different implications for gene expression. Thus, heterochromatin should not be thought of as simply encapsulating "dead" DNA, but rather as a descriptor for compact chromatin domains that share the common feature of being unusually resistant to gene expression.
Introduction to Genetic Analysis XI edition (2014)	Griffiths, Wessler, Carrol and Doebley	It is important to note that chromatin is not uniform over all chromosomes: certain regions of chromosomes are bundled in highly condensed chromatin called heterochromatin. Other domains are packaged in less-condensed chromatin called euchromatin. After cell division, regions forming heterochromatin remain condensed, especially around the centromeres and telomeres (called constitutive heterochromatin), whereas the regions forming euchromatin become less condensed. The densely packed nucleosomes of heterochromatin are said to form a "closed" structure that is largely inaccessible to regulatory proteins and inhospitable to gene activity.
Genes XII (2018)	B. Lewin	1) Heterochromatin is the term used to describe regions of chromosomes that are permanently tightly coiled up and inert, in contrast with the euchromatin that represents most of the genome. 2) It often consists of multiple repeats of a few sequences of DNA that are not transcribed or are transcribed at very low levels. Genes that reside in het-erochromatic regions are generally less transcriptionally active than their euchromatic counterparts, but there are exceptions to this general rule.

Table 1 Constitutive heterochromatin in Genetics and Molecular Biology textbooks

	protein coding genes/size	non coding RNAs/size	miRNA/size	pseudogenes/size	rDNA	cry-Su(Ste)	mst77Y/size	total size	total size[1] (no pseudo)	genomic size[2]
X	40/0.35292	13 /0.314621	1/0.000144	14/0.012500	2.8	nd	-	3.480185	3.467685	1.666933[3]
Y	16/5.858102 16/15.131233[4]	1/0.000855[5]	0	60/0.038223[6]	5.37663[7]	627/1.6[8]	10/0.096[9]	12.96981 22.24294[4]	12.931587 22.242941[3]	40.9[10]
2L	44/0.461670	26/0.045514	1/0.000074	5/0.003480	-	.	-	0.510738	0.507258	1.182357
2R	54/2.190338	13/0.027385	1/0.000112	12/0.012557	-	.	-	2.230392	2.217835	5.326047
3L	37/1.767208	16/0.012926	1/0.000110	10/0.024960	-	.	-	1.805024	1.780244	5.014227
3R	14/2.930834	5/0.002932	0	4/0.009975	-	.	-	2.943741	2.933766	3.637496
Tot	205/13.561072 205/22.834203[4]	74/0.404233	4/0.000440	105/0.101695	8.17663	1.6	0.096	23.94007 33.213201[4]	23.838375 33.149729[4]	

Table 2 Genes of *Drosophila melanogaster* constitutive heterochromatin (genomic size expressed in Mb)

[1] Genomic portion of constitutive heterochromatin occupied by genes (excluding pseudogenes) defined by Release 6 coordinates.
[2] Size of genomic portions analyzed is defined by the genomic coordinates of Release 6. The Satellite DNAs blocks are not included, since the junctions between them and complex DNA regions are still unknown.
[3] The Ribosomal DNA cluster is not included in Release 6.
[4] Considering the maximum size estimated for *kl-5, kl3, kl-1* and *ks-1* fertility genes (Table S2) [17,54].
[5] Excluding the *cry/Su(Ste)* repeats located in region h11 (Figure 1).
[6] 52 fertility factors pseudogenes plus 8 *Mst77* pseudogenes [42].
[7] Based on 18S, 28S, and IGS copies number determined by Chang and Larracuente [48], we calculated the overall size by multiplying for 1995 bp the 18S copies, for 3945 bp the 28S copies (59), and for an average length of 6kb for each IGS repeat [65].
[8] Based on *cry/Su(Ste)* copies number determined by Chang and Larracuente [48], we calculated the overall length of *Su(ste)* cluster considering the size of each repeat as a mean between the 2,5 and 2,8 kb types of repeats (transposable elements associated to each *cry/Su(Ste)* repeat were not included).
[9] Krsticevic et al [42].
[10] Y chromosome size estimated by Hoskins et al [22].

BOX 1

Classical properties of heterochromatin described in eukaryotic genomes
The “sounds of silence” of heterochromatin emerged as far back as in 1930, when Herman Muller discovered the *white mottled* eye phenotype in *Drosophila melanogaster*, which is observed when the *white+* gene is moved from euchromatin next to the pericentric heterochromatin of the X-chromosome following a chromosomal inversion. As a consequence, the *white+* gene is aberrantly “heterochromatinized” and inactivated in some cells during eye differentiation, giving rise to the mottled phenotype. That phenomenon, called position effect variegation (PEV), is regarded as a signature feature of constitutive heterochromatin, and represents the first classical example of epigenetic silencing to be identified [2], later found in evolutionarily distant organisms. Facultative heterochromatin corresponds to euchromatin that undergoes silencing (chromosome regions, entire chromosomes or even whole genomes) [3,4,5]. In contrast, constitutive heterochromatin consists mostly of repetitive sequences and shows the same cytological and molecular characteristics on both homologous chromosomes. Constitutive heterochromatin represents a significant portion of the eukaryotic genomes, (10% in *Arabidopsis*, 20% in humans, 30% in *Drosophila melanogaster*, up to 85% in certain nematodes), where it contributes to centromere and telomere functions. Genetic and molecular features differ between facultative and constitutive and heterochromatin. First, their DNA composition clearly differs. In almost all eukaryotes, constitutive heterochromatin is enriched in repetitive sequences (mainly satellite DNAs and defective transposons), compared to facultative heterochromatin. Among these sequences, specific remnants of transposable elements (TEs) have been proposed to act as docking sites for the formation and organization of heterochromatin [6]. Second, they differ for their distinctive repressive epigenetic marks: typically, constitutive heterochromatin is marked by di/trimethylation of histone H3 on lysine 9 (H3K9me2/me3), while facultative heterochromatin is mainly characterized by the enrichment in H3K27me3 and H2AK119Ub [11]. Most importantly, a major difference between facultative and constitutive heterochromatin is that the latter can be transcribed. In addition to PEV, properties historically assigned to constitutive heterochromatin are: 1) absence of meiotic recombination; 2) absence of genes or extremely low gene density; 3) late replication during S phase; 4) transcriptional inactivity; 5) enrichment in highly repetitive satellite DNAs and transposable element remnants. 6) Presence of silent epigenetic marks (mainly H3K9 methylation). Together, these properties are antithetical to those of euchromatin and have led to the view that constitutive heterochromatin is a "wasteland" made up by junk DNA.

FIGURE LEGENDS

Figure 1 Cytological map of mitotic heterochromatin of chromosomes X, Y, 2 and 3 showing the locations of single-copy and repetitive genes.

Constitutive heterochromatin is generally considered to represent about bout one-third, 60 Mb of DNA, of the Drosophila melanogaster haploid genome. However, according to Hoskins et al., [25], the total amount of constitutive heterochromatin of a female and male gamete is 58.8 Mb and 79,8 Mb, respectively (Table S8). This difference is due to the presence of the entirely heterochromatic Y chromosome, which is about 40.9 Mb, while the X heterochromatin consists of 19.9Mb [25]. Using banding techniques, these heterochromatic portions of chromosomes have been subdivided into a map of 61 regions with diverse cytological features, designated h1 to h61 [13]. Filled areas represent the Hoechst 33258 or DAPI-bright regions; the shaded boxes represent regions of intermediate fluorescence and the open boxes are regions of dull fluorescence. X = X chromosome; Y = Y chromosome; 2L = left arm of chromosome 2; R = right arm of chromosome 2. 3L = left arm of chromosome 3; 3R = right arm of chromosome 3. C = centromeric region. 4; 4R = right arm of four 3. Black horizontal lines indicates mapping of the genes. a = the ABO elements have been mapped to regions h11 and h26 (on Y and X, respectively) and are still molecularly unknown.

Figure 2 Euchromatin/heterochromatin borders, density of genes vs repeats and chromatin marks along constitutive heterochromatin of chromosomes X, 2 and 3.

The borders between pericentric heterochromatin and proximal euchromatin have been defined previously by cytogenomic and epigenomics approaches [25-29, 101]. Other studies have linked such borders to a sharp drop in the density of repetitive sequences [100]. To define the borders between pericentric heterochromatin and proximal euchromatin we have considered cytogenetics, genomics and epigenomics criteria: 1) The cytogenetic map of polytene chromosomes where divisions 20A-F (X), 40A-F (2L), 41A-F (2R), 80A-F (3L), and 81A-F (3R) are generally considered heterochromatic; 2) The mitotic chromosome mapping by FISH [26,28,29]; Figure 1); 3) The density of genes and repeats, and 4) The presence of active/silent chromatin marks. The genomic coordinates were assigned considering the minimum overlapping of at least three of the four criteria. Notably, the coordinates defined with the above mentioned criteria (Tables S3, S4, S5, S6 and S7) are in good agreement with the cytogenomic borders defined by Hoskins et al. [25] and with the epigenomics borders described by Riddle et al [101]. Vertical dashed lines show the borders. Frequencies are depicted in logarithmic scale and are calculated using windows size of 100 Kb (overlapping by 50 Kb). Chromatin states have been defined by Kharchenko et al.[68]. The active (states 1-6, green lines) or repressive (states 7-9, red lines) chromatin states are depicted on the top of each plot. Black: State 1 H3K4me3/me2 and H3K9ac (Active promoter and transcription start site proximal regions); White: State 2 H3K36me3 (transcriptional elongation signature - exonic regions); Red: State 3 H3K27ac, H3K4me1, and H3K18ac (introns); Lime: State 4 H3K36me1 lacks H3K27ac (other open chromatin); Blue: State 5 H4K16ac (Actively transcribed exon on the male X chromosome-dosage compensation); Yellow: State 6 H3K27me3 and Pc (regions of Polycomb-mediated repression); Cyan: State 7 H3K9me2/me3 (pericentromeric heterochromatin); Magenta: State 8 H3K9me2/me3; Silver: State 9 slightly enriched in H3K27me3 (transcriptionally silent intergenic

euchromatin). Chromosome coordinates, genes and repeats position, and chromatin state markers data were obtained from Flybase.

Figure 3 Heatmaps showing of the expression profile of single-copy coding genes of Xh, 2Lh, 2Rh, 3Lh and 3Rh.
Shades of color from red to green indicate the expression bin classification from 1 (no/extremely low expression) to 7 (very high expression). The genes encoding ribosomal proteins show the highest expression levels, either during development or in different tissues. Constitutive high expression levels are also shown, for example, by Tim 23 in 2Lh, Yeti, CG17691, RNAseq, d4, CG10465, Fis1 and CG17337 in 2Rh, CkIIalpha, ND.MLR, CG40045 in 3Lh and Tim17b and Gfat1 in 3Rh. Only few genes (CG40813 and CG41562 in Xh, CG40439, CG17715, CG17490 and GpB5 in 2Lh, CG46302, CG41241 and RyA in 2Rh, or Myosin 81F in 3Rh) show no or very low expression levels. However, most of these genes have different annotated transcripts confirmed by the presence of cDNA clones. For example, CG40439 has 2 transcripts supported by 15 cDNA clones, while CG17715 has 8 transcripts supported by 70 cDNA clones (see FlyBase) and both shows moderate expression in FlyAtlas [81]. Finally, RyA is expressed in Drosophila cell lines and Myo81F is expressed in the larval and pupal stages, most abundantly in white prepupae aged 12 h and 24 h with RPKM expression levels of 2.9 and 2.7, respectively [30].
Developmental stages and tissues expression data were obtained from Flybase. Tissues (from left to right): A_1d_carcass, A_1d_dig_sys, A_20d_carcass, A_20d_dig_sys, A_4d_carcass, A_4d_dig_sys, A_MateF_1d_head, A_MateF_20d_head, A_MateF_4d_head, A_MateF_4d_ovary, A_MateM_1d_head, A_MateM_20d_head, A_MateM_4d_acc_gland, A_MateM_4d_head, A_MateM_4d_testis, A_VirF_1d_head, A_VirF_20d_head, A_VirF_4d_head, A_VirF_4d_ovary, L3_CNS, L3_Wand_carcass, L3_Wand_dig_sys, L3_Wand_fat, L3_Wand_imag_disc, L3_Wand_saliv, P8_CNS, P8_fat, WPP_fat, WPP_saliv. Developmental stages (from left to right): AdF_Ecl_1days,AdF_Ecl_30days, AdF_Ecl_5days, AdM_Ecl_1days, AdM_Ecl_30days, AdM_Ecl_5days, em0-2hr, em10-12hr, em12-14hr, em14-16hr, em16-18hr, em18-20hr, em2-4hr, em20-22hr, em22-24hr, em4-6hr, em6-8hr, em8-10hr, L1, L2, L3_12hr, L3_PS1-2, L3_PS3-6, L3_PS7-9, P15, P5, P6, P8, P9-10, WPP