The repetitive DNA that constitutes most of the heterochromatic regions of metazoan genomes has hindered the comprehensive analysis of gene content and other functions. We have generated a detailed computational and manual annotation of 24 megabases of heterochromatic sequence in the Release 5 Drosophila melanogaster genome sequence. The heterochromatin contains a minimum of 230 to 254 protein-coding genes, which are conserved in other Drosophilids and more diverged species, as well as 32 pseudogenes and 13 noncoding RNAs. Improved methods revealed that more than 77% of this heterochromatin sequence, including introns and intergenic regions, is composed of fragmented and nested transposable elements and other repeated DNAs. Drosophila heterochromatin contains “islands” of highly conserved genes embedded in these “oceans” of complex repeats, which may require special expression and splicing mechanisms.

The goal of genome annotation is to identify sequence features that have a biological role in the organism, but a telomere-to-telomere DNA sequence is not yet available for complex metazoans, including humans. The missing genomic “dark matter” is the heterochromatin, which is generally defined as repeat-rich regions concentrated in the centric and telomeric regions of chromosomes. Centric heterochromatin makes up at least 20% of human and 30% of fly genomes, respectively; thus, even for well-studied organisms such as Drosophila melanogaster, fundamental questions about gene number and global genome structure remain unanswered.

Once considered “junk” DNA, it is now clear that heterochromatin contains essential genes and also contributes to genome stability and chromosome segregation (1–3). In addition, heterochromatin participates in RNA interference mechanisms that epigenetically repress gene and transposable element (TE) expression, which may “immunize” genomes against the invasion and expansion of selfish DNA elements (4, 5). Despite differences in heterochromatin sequence composition between genomes, commonalities in the structures, chromatin modifications, and presence of genes suggest that D. melanogaster heterochromatin is an excellent model for studying repeat-rich genomic DNA in other species, including the 40% repetitive human euchromatin (6, 7).

Annotation overview. The Drosophila Heterochromatin Genome Project has generated 16 megabases (Mb) of finished or near-finished heterochromatin sequence from D. melanogaster, as well as 8 Mb of draft whole genome shotgun (WGS) heterochromatic assemblies (8, 9). We performed computational and manual curation to produce the Release 5.1 annotation of this 24 Mb of heterochromatin sequence, which excludes degenerate sequence reads not incorporated into the final sequence assembly (ArnUXtra) (10, 11). Several repeat-finding programs were implemented, including RepeatFinder (12) and Random Repeats Finder (TRF) (13). New data from the research community were incorporated, including GenBank third-party annotations and Heidelberg gene predictions (14). Lastly, the conservation of D. melanogaster heterochromatin genes was assessed by identifying putative gene orthologs in more than 16 species (supporting online material (SOM) text and data).

The annotations include protein-coding genes, non–protein-coding RNAs (ncRNAs), repetitive sequences, and other functional elements. The majority of nonrepeat annotations (64%) mapped to a chromosome arm, including regions contiguous with the euchromatic arms (chromosome arm h; e.g., 2Rh) and internal scaffolds that have been cytologically localized to an arm (chromosome arm H; e.g., 2RHet) (8, 9). The remaining annotations are not mapped to a chromosome arm (36%) and reside on Arm U (Unmapped). Although the highly repetitive simple sequence component of D. melanogaster heterochromatin remains unassembled (8, 9), few unique genes have been identified in these regions, and thus the current annotation is likely to include the majority of heterochromatic genes.

Protein-coding genes and comparative analysis. Previous studies indicated that at least 32 essential genetic loci are present in the heterochromatin (1). Using clone-based evidence [expressed sequence tag (EST) and cDNA] and predicted orthology, we annotated 613 protein-coding genes and gene fragments in the 24 Mb of Release 5 heterochromatin (Table 1, SOM text, and data) (11). Currently, 41% of these genes have supporting clone-based evidence, including 137 annotations with full-length cDNAs generated by the Berkeley Drosophila Genome Project (BDGP) (15). The incorporation of new EST sequences from Exelixis (16), which were generated with random primers, was instrumental for finding new internal splice sites and missed exons and was used to refine at least 43 gene models (Fig. 1) (11). New evidence resulted in 16 cases of Release 3.2b genes that were merged into seven larger Release 5.1 genes (e.g., CG41520 in Fig. 1) and two genes from Release 3.2b that were split into four smaller genes in Release 5.1.

A subset of the annotations represents single-exon fragments of coding sequences present in small scaffolds, and these fragments should not yet be considered as complete genes. We classified gene annotations as either single-exon genes (186) or multi-exon genes (427) in order to more conservatively estimate higher-quality complete gene models without losing information about potentially incomplete genes (Fig. 2). Heterochromatin single-exon genes are currently less supported by clone-based evidence; roughly 50% of heterochromatic multi-exon genes and 61% of euchromatic single-exon genes had EST or cDNA support, compared to only 20% of single-exon genes.

Because many genes lacked EST or cDNA support, the translated basic local alignment sequence tool (TBLASTN) was used to identify putative orthologs for heterochromatin genes in 16 other insect genomes and more distantly related vertebrate species (17) (Fig. 3A). We defined orthologs as unique, top-scoring TBLASTN-identified sequences, although no evidence currently exists to support conserved function. Overall, more than 99% of the annotations had an ortholog identified in at least one species (Fig. 3A). Orthologs were identified for 86 to 98% of heterochromatin protein-coding genes in the four species (D. simulans, D. erecta, D. sechellia, and D. yakuba) most closely related to D. melanogaster, and 55 to 70% of genes are conserved in the more distantly related Drosophilids. In addition, 22 to 46% of genes are conserved in distantly related insects (such as the silkworm, mosquito, honeybee, wasp, and beetle), and 13% of all protein-coding genes had significant alignments to proteins in even more diverged species (Fig. 3A). We conclude that the majority of D. melanogaster heterochromatic genes are highly conserved in insect lineages that span 300 million years of evolution and that a surprising number share significant similarity with proteins from vertebrate species that diverged more than 900 million years ago (18).

The conservation patterns were bimodal, such that nearly 20% of protein-coding genes were conserved in fewer than 4 species, whereas more than 30% were conserved in all 16 insect species (Fig. 3B). This trend suggests the existence of

---

*Department of Biology, San Francisco State University, San Francisco, CA 94132, USA. 1Drosophila Heterochromatin Genome Project, Department of Genome and Computational Biology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. 2National Center for Biomedical Ontology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. 3Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA 94720, USA.

1To whom correspondence should be addressed. E-mail: karpen@fruitfly.org
distinct groups of Drosophila lineage-specific genes and another subset of ultraconserved insect genes. Closer examination of the 163 genes with orthologs in four or fewer species revealed that 93% had orthologs only in the four species most closely related to D. melanogaster (Fig. 3B, SOM text, and data). Conservation in only the Drosophilid lineage or melanogaster subgroup may identify more recently evolved genes (SOM text). Indeed, some open reading frames (ORFs), such as odorant-binding proteins or receptors, tend to evolve faster than do other genes and are often found in only a single species (19, 20). More in-depth study will be required to determine whether the lineage-specific Drosophilid genes are conserved because of function or because there has been insufficient time for ORFs to diverge.

Analysis of orthology data, gene prediction, and cDNA and EST clone evidence shows that 96% of heterochromatin multi-exon genes and 89% of single-exon genes are supported by two or more types of evidence (Fig. 2, A and B). These results demonstrate that single-exon genes probably represent parts of bona fide, full-length genes that will be merged into more complete genes as the sequence assemblies and cDNA resources are improved, as observed for fragmented annotations in prior releases (e.g., CG41520 in Fig. 1). Comparative genomic support for small single-exon genes also suggests that there may be many more conserved exons in other parts of less well-assembled Drosophilid genomes (21).

We inferred 1030 conserved introns between well-conserved exons of the orthologs identified in the 16 insect genomes. Heterochromatin gene introns are, on average, five times longer than introns in euchromatic genes [4949 versus 1149 base pairs (bp)], and may be as long as 1 Mb (22). The longest contiguous intron we have identified is 224,977 bp in the Snap25 gene (11). Introns lengths are highly conserved among euchromatic gene orthologs (Fig. 3C), whereas heterochromatin intron lengths were correlated in species closely related to D. melanogaster but were poor-

Table 1. Annotation summary. ND, not done.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotated sequence (Mb)</td>
<td>3.8</td>
<td>12.1</td>
<td>14.2</td>
<td>24</td>
</tr>
<tr>
<td>Sequence length of repeats (Mb/%)</td>
<td>ND</td>
<td>6.3/52</td>
<td>6.3/75</td>
<td>18/77</td>
</tr>
<tr>
<td>Sequence length of exons (Mb/%)</td>
<td>0.15/4</td>
<td>0.33/2.7</td>
<td>0.43/3.0</td>
<td>1.33/5.5</td>
</tr>
<tr>
<td>Repeat nest fragments (number/Mb)</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>10084/10</td>
</tr>
<tr>
<td>Full-length TEs</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>202</td>
</tr>
<tr>
<td>Total annotations</td>
<td>130</td>
<td>447</td>
<td>556</td>
<td>11038</td>
</tr>
<tr>
<td>Protein-coding genes</td>
<td>130</td>
<td>297</td>
<td>472</td>
<td>613</td>
</tr>
<tr>
<td>Single-exon genes</td>
<td>43</td>
<td>58</td>
<td>195</td>
<td>187</td>
</tr>
<tr>
<td>Genes with finished cDNAs</td>
<td>48</td>
<td>58</td>
<td>92</td>
<td>137</td>
</tr>
<tr>
<td>Protein-coding genes with any EST/cDNA clone evidence</td>
<td>ND</td>
<td>80</td>
<td>142</td>
<td>250</td>
</tr>
<tr>
<td>Pseudogenes</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>32</td>
</tr>
<tr>
<td>ncRNAs</td>
<td>0</td>
<td>3</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Recursive splice sites</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>16</td>
</tr>
<tr>
<td>Miscellaneous annotations</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>9</td>
</tr>
<tr>
<td>Unassembled ribosomal DNA fragments</td>
<td>0</td>
<td>6</td>
<td>52</td>
<td>67</td>
</tr>
</tbody>
</table>

Fig. 1. Computational pipeline results used for the Release 5.1 annotation. An Apollo (36) screenshot of the evidence (black region) and Release 5.1 annotations (light blue region) for scaffold CP000218, which was produced by merging, extending, and finishing the Release 3 WGS scaffolds (AABU01002750 and AABU010027230) (9). New cDNA evidence was used to merge the Release 3.2b annotations CG40067, CG17443, and CG40109 into one Release 5.1 gene (CG41520) and to identify an alternative exon for CG41250-RB (asterisk). ncRNA CG40375 is shown on the opposite strand to illustrate that it is nested within CG41520. CG40388 represents a Release 3.2b gene that is now annotated as a TE fragment. Complete annotation and evidence are shown in (11).
ly correlated for more distantly related species. For example, intron lengths for *D. erecta* orthologs were correlated in both euchromatin (correlation coefficient \( r = 0.48 \)) and heterochromatin (\( r = 0.58 \)), whereas *D. virilis* gene intron lengths were correlated in euchromatin (\( r = 0.37 \)) but not in heterochromatin (\( r = 0.14 \)) (Fig. 3C). A list of conserved introns, their average lengths, and the percent conservation of flanking coding exons is provided (SOM text and data).

One of the features of heterochromatic genes is that introns and intergenic regions are com-

---

**Fig. 2.** Evidence of *D. melanogaster* heterochromatin protein-coding gene annotations. Venn diagrams show the percentage of protein-coding genes supported by gene prediction (pink), EST or cDNA (yellow), and/or BLASTX/TBLASTN comparative genomic evidence (blue). (A) Multi-exon genes are likely to be complete, whereas (B) single-exon genes are likely to represent genes that are fragmented across multiple scaffolds. The number of genes measured for each class are indicated, as well as the number and percent of genes with putative orthologs in melanogaster group species, nonmelanogaster group Drosophilids, or other insect species.

---

**Fig. 3.** Comparative analysis summary for *D. melanogaster* heterochromatin genes. (A) Number of heterochromatin protein-coding genes with a predicted ortholog in a given species, ordered (left to right) by increasing evolutionary distance from *D. melanogaster*. (B) Frequency histograms showing the percentage of heterochromatin protein-coding genes with a predicted ortholog in the 16 insect species tested. (C) Scatter plots of intron lengths (bp) for euchromatin and heterochromatin protein-coding gene introns conserved in either *D. erecta* or *D. virilis*. Each data point refers to a single conserved intron. Correlation coefficients \( r \) for intron lengths are indicated.
posed almost entirely of repeated sequences, predominantly fragmented TEs (Fig. 1). There was no appreciable difference in the average repeat density or composition of intronic (56%) versus intergenic sequences (63%) (11). Changes in intron length are most often due to TE insertions and excisions and simple repeat expansions. The high repeat content of introns and regulatory regions suggests that regulation of heterochromatic gene expression may differ from euchromatic genes. We identified 16 recursive splice sites (RSSs) shown to aid in the splicing of long introns (23), including 8 RSSs located in long heterochromatin gene introns ranging from 11 to 166 kilobases (kb). Of particular interest are three RSSs predicted in the 23.6-kb intron of CG40120, only one of which was previously predicted because of gaps in the sequence assembly (23). Fifty-six percent of the RSS motifs were embedded within retrotransposons, suggesting that cis TE sequences may be used to splice out TEs that invade heterochromatin genes.

Non–protein-coding genes. We identified 13 putative non–protein-coding genes in the heterochromatin (Table 1), defined as single-copy genes with EST or cDNA support that contained protein-coding ORFs that were substantially shorter than the length of the transcript. Spliced ESTs and cDNAs were identified for 11 of the ncRNA annotations, excluding the possibility of false positives generated by the priming of polymerase chain reaction products from adenine (A)-rich regions. Analysis of the two unspliced ncRNAs suggested that the clones were not primed from genomic

---

**Repeats and transposable elements.** Repeats were defined by significant alignment to known RepBase repetitive sequences identified by RepeatMasker (28), BLASTX homology to TE proteins, or TRF (29) results (10). The application of these improved methods demonstrated that 18 Mb (77%) of the annotated heterochromatin could be classified as repetitive or transposable elements, a 4-Mb (20%) increase compared with the previous annotation (1). About 50 previously identified protein-coding genes have been reannotated as repetitive features. For example, CG40388 was annotated as a protein-coding gene in Release 3.2b but annotated as a 1731 repeat in Release 5.1 (Fig. 1). The heterochromatin sequence was previously reported to consist of 3.86 to 6% of repetitive sequence (30, 31). Our methods provide better recognition of TE fragments, small tandem repeats, and short ORFs from degenerate TEs, and they identified 7% of the Release 5 euchromatin as repetitive (Fig. 4), similar to other estimates (32). The sequenced D. melanogaster heterochromatin has an overall repeat and TE content more than 10 times that of fly euchromatin and is more similar to the repeat density in human euchromatin (40%) (6, 7).

We measured the repeat content and gene distribution across the heterochromatin sequence in 100-kb sections and calculated the average density for each region (Fig. 4). Lower overall repeat content was observed for regions more distal from the centromere, especially for the heterochromatic regions on chromosome arm 3R (Fig. 4). In addition, the unmapped scaffolds had a higher repeat content than the heterochromatic regions that have been mapped to the chromosome arms (Fig. 4). In general, there was a strong inverse correlation between the repeat and gene content of a region ($r = -0.89$) (Fig. 4). Euchromatin had an average gene density of 12.6 genes per 100 kb, whereas heterochromatin contained 1.8 to 4.4 genes per 100 kb (2.9 genes per 100 kb overall). Exceptions include Xh and XHet, with 9.2 and 6.4 genes per 100 kb. Chromosome arm 3R had an average gene content even higher than that of typical euchromatin (19 genes per 100 kb) and a 20% average repeat content, which is significantly lower than the rest of the heterochromatin (Fig. 4).

We further categorized the average percentage of retrotransposons, DNA transposons, and other repeats (Fig. 4). Roughly two-thirds (16 Mb, 66%) of the heterochromatin is composed of retrotransposon sequences [33% long terminal repeats (LTRs) and 33% long interspersed nuclear elements (LINEs)]. DNA transposons are overrepresented on the relatively repeat-rich euchromatin regions of the fourth chromosome of D. melanogaster (33) but constitute only 15% of the Release 5 heterochromatin, which does not include the fourth chromosome heterochromatin (8, 9). TRF-identified tandem repeats and satellite repeats make up ~10% of the available heterochromatin sequence, which is significantly higher than in the euchromatin sequence (~3%), espe-
cially in the proximal centric regions of chromosomes 2 and 3 and the Y chromosome (Fig. 4). Calculations of tandem-repeat content are likely to be an underestimate, because WGS3 sequence underrepresents the difficult-to-clone satellite DNA and tandem-repeat regions. Unlike the Y chromosome and autosomal heterochromatin, the available X chromosome sequence is not enriched for tandemly repeated sequences. Our repeat analysis indicates that nearly all of the ArmU and ArmUExtra (10) sequence is composed of repetitive sequences, further suggesting that we have identified most of the unique sequence available in the D. melanogaster heterochromatin.

The majority of repetitive TE-like sequences in heterochromatin is not intact. We found 202 full-length TEs in the heterochromatin (2% of heterochromatic TEs), compared with 361 full-length TEs reported for the nonpericentromeric euchromatin (20.6% of euchromatic TEs). The most recent annotation of the Release 4 euchromatin identified nests of TEs that were fragmented, interdigitated, and transposed into one another (32). Our manual curation identified 846 repeat nests in newly sequenced regions of the Release 5 heterochromatin, compared with 112 nested TEs in the euchromatin. We annotated 117 instances where there were two nested TEs (i.e., a TE jumped into a TE that itself had jumped into a TE) and 17 instances where four or more TEs were nested (Fig. 1).

Conclusions. The assembly of a more complete genome sequence (9) and an integrated annotation set for D. melanogaster provides a reference set of genes and other features that will be useful for investigating the biological functions of heterochromatin in flies and other organisms. These results are now fully integrated, with information about the euchromatin, in FlyBase (34) and GenBank (35).

These results demonstrate that repetitive and TE sequences constitute at least 77% of the 24 Mb of heterochromatin sequence. Although there are more full-length TEs and TE nests in the centric heterochromatin relative to the euchromatin, it appears that most heterochromatic TEs are fragmented and not capable of autonomous transposition. As with the euchromatin, the repetitive sequences in heterochromatin are dominated by LTR- and LINE-like retrotransposons. We have identified a substantial amount of tandemly repeated sequences in the most proximal centric heterochromatin of the second, third, Y, and unmapped chromosomes, but not in the currently sequenced X chromosome. These carefully annotated repetitive regions provide opportunities for more in-depth analysis of their functions and evolution. For example, a recent study showed that at least 80% of piwi-associated small RNAs, which regulate transposon activity, map to the euchromatin. We have identified 32 pseudogenes in the heterochromatin, including 8 in the poorly represented Y chromosome sequences, representing a density of pseudogenes that is at least three times that of euchromatin. The high repeat content of heterochromatin may provide recombination substrates that increase the frequency of tandem and segmental duplications.

Despite differences in gene density, there are many similarities between the basic structures and putative functions of euchromatic and het-
erosomal genes. Based on cDNA-supported genes, it appears that euchromatin and heterochromatin genes have, on average, a similar number of exons and transcript variants per gene. In general, heterochromatic and euchromatic genes appear to encode a similar spectrum of functions, based on gene ontology (GO) analysis (Fig. 5). Some classes of genes are overrepresented in the heterochromatin, relative to the euchromatin. For example, heterochromatin genes are enriched 35-fold for putative membrane cation transporter domains (4 out of 308 heterochromatin domains versus 5 out of 13,500 euchromatin domains). Heterochromatic genes are also enriched for domains involved in DNA (53 domains) or protein binding (122 domains) that may regulate chromatin structure or function, including histone variants and proteins (Fig. 5, SOM text, and data) (11). This raises the intriguing possibility that heterochromatin may encode genes involved in its own establishment or maintenance.

Heterochromatin genes can reside in regions that approach 90% repeat content. Heterochromatin gene introns are usually composed of fragmented TE sequences (Fig. 1), are on average five times longer than euchromatin gene introns, and display less length conservation in interspecies comparisons. We found nine recursive splice site motifs nested in the long introns of heterochromatin genes, which may regulate splicing in repeat-rich regions. The underlying mechanisms that allow essential genes to be expressed and regulated in otherwise silent chromatin remain unknown. Studying heterochromatin in other species promises to shed light on whether there are cis sequences that define or regulate boundaries between euchromatin and heterochromatin and if there are genic and non-genic regions of heterochromatin in other repeat-rich regions, including human euchromatin.

References and Notes
10. Materials and methods are available as supporting material on Science Online.
37. We thank E. Frise for maintaining the hardware and software used in these studies; M. Yandell for providing the specialized comparative genomics library–based software used in our analyses; A. Denburg, D. Acededo, J. Carlson, S. Celniker, R. Hoskins, and C. Kennedy for their helpful comments on the manuscript and input on annotations; and the members of the BDGP for cDNA sequencing. This work was supported by the National Human Genome Research Institute grant R01 HG000747 to C.D.S. and G.H.K. and NIH grant U54 HG004028-01 to S.S. and C.J.M.

Supporting Online Material
www.sciencemag.org/cgi/content/full/316/5831/1586/DC1
Materials and Methods
SOM Text
SOM Tables (D to F, I, and J)
References
11 January 2007; accepted 7 May 2007
10.1126/science.1139815

Tracking Solar Gravity Modes: The Dynamics of the Solar Core
Rafael A. García,* Sylvaine Turck-Chièze,® Sebastian J. Jiménez-Reyes,® Jérome Ballot,® Pere L. Pallè,® Antonio Eff-Darwich,® Savita Mathur,® Janine Provost®

Solar gravity modes have been actively sought because they directly probe the solar core (below 0.2 solar radius), but they have not been conclusively detected in the Sun because of their small surface amplitudes. Using data from the Global Oscillation at Low Frequency instrument, we detected a periodic structure in agreement with the period separation predicted by the theory for gravity dipole modes. When studied in relation to simulations including the best physics of the solar interior (9), the sound speed down to 0.05 solar radius (R⊙) are not well defined. For example, large uncertainties still remain in the solar rotation profile below 0.2R⊙ (Fig. 1) because of the lack of sensitivity and the poor spatial resolution of the modes toward the deep interior (9).

To progress at greater depths and down into the solar core requires the study of another type of waves—the gravity-driven modes (g modes), for which the driving force is buoyancy. These modes are trapped within the radiative region of the Sun and become evanescent in the convective zone, reaching the solar surface with amplitudes that could be very small (10). Even considering their low surface amplitudes, g modes remain below 0.3R⊙. They are not well defined. For example, large uncertainties still remain in the solar rotation profile below 0.2R⊙ (Fig. 1) because of the lack of sensitivity and the poor spatial resolution of the modes toward the deep interior (9).

To progress at greater depths and down into the solar core requires the study of another type of waves—the gravity-driven modes (g modes), for which the driving force is buoyancy. These modes are trapped within the radiative region of the Sun and become evanescent in the convective zone, reaching the solar surface with amplitudes that could be very small (10). Even considering their low surface amplitudes, g modes remain below 0.3R⊙. They are not well defined. For example, large uncertainties still remain in the solar rotation profile below 0.2R⊙ (Fig. 1) because of the lack of sensitivity and the poor spatial resolution of the modes toward the deep interior (9).

To progress at greater depths and down into the solar core requires the study of another type of waves—the gravity-driven modes (g modes), for which the driving force is buoyancy. These modes are trapped within the radiative region of the Sun and become evanescent in the convective zone, reaching the solar surface with amplitudes that could be very small (10). Even considering their low surface amplitudes, g modes remain
ERRATUM

Post date 7 September 2007

Reports: “The Release 5.1 annotation of Drosophila melanogaster heterochromatin” by C. D. Smith et al. (15 June 2007, p. 1586). The affiliation for ShengQiang Shu and Christopher J. Mungall was listed incorrectly. They are affiliated with Berkeley Bioinformatics and Ontologies Project, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. In addition, two funding sources were omitted from the acknowledgements note. The work in this paper was also supported by NIH grant P41 HG000739-15 (S.S.) and by the Howard Hughes Medical Institute (C.J.M.).
The Release 5.1 Annotation of Drosophila melanogaster Heterochromatin
Christopher D. Smith, ShengQiang Shu, Christopher J. Mungall and Gary H. Karpen

Science 316 (5831), 1586-1591.
DOI: 10.1126/science.1139815