Molecular and Genetic Structure of Polytene Chromosome Banding Pattern in *Drosophila melanogaster*

Tatyana Zykova Vatolina  
Researcher  
Siberian Branch of the Russian Academy of Sciences  
Russia  

Darya Demidova  
Researcher  
Siberian Branch of the Russian Academy of Sciences & Novosibirsk State University  
Russia  

Viktor Levitsky  
Researcher  
Siberian Branch of the Russian Academy of Sciences  
Russia
Varvara Khoroshko  
Researcher  
Siberian Branch of the Russian Academy of Sciences  
Russia

Elena Belyaeva  
Researcher  
Siberian Branch of the Russian Academy of Sciences  
Russia

Elena Kokoza  
Researcher  
Siberian Branch of the Russian Academy of Sciences  
Russia

Igor Zhimulev  
Researcher  
Siberian Branch of the Russian Academy of Sciences &  
Novosibirsk State University  
Russia
An Introduction to
ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. This paper has been peer reviewed by at least two academic members of ATINER.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:

Molecular and Genetic Structure of Polytene Chromosome Banding Pattern in *Drosophila melanogaster*

Tatyana Zyкова Vatolina  
Darya Demidova  
Viktor Levitsky  
Varvara Khoroshko  
Elena Belyaeva  
Elena Kokoza  
Igor Zhimulev

Abstract

Classic polytene chromosomes from dipteran insects are extensively used as a model for interphase chromosomes. The morphology of polytene chromosomes is formed by the alternating densely packed chromatin of bands and less compact interbands. Novel approaches were developed to tag and position interband regions at a molecular map. We used available datasets from several recent *Drosophila* genome-wide projects and compared the molecular organization of 32 interband regions, which were accurately mapped previously. We demonstrate that in the interphase chromosomes of *Drosophila* cell lines, the interband regions are enriched with proteins of the “open” chromatin. Polytene chromosome interbands contain 5’-ends of housekeeping genes. As a rule, interbands display the “head-to-head” orientation of genes. They are enriched for “broad” class promoters. Comparison of expression patterns of genes mapping to late-replicating dense bands vs genes whose promoter regions map to interbands shows that the former are generally tissue-specific, whereas the latter are represented by ubiquitously active genes.

**Keywords:** *Drosophila melanogaster*, bands, housekeeping genes, interbands, interphase chromosomes, ModEncode project, tissue-specific genes, open chromatin

**Acknowledgments:** This research was funded by the grants from the Russian Fund for Basic Research # 13-04-40137H and # 15-04-03898. The authors are grateful to Dr. A. Gorchakov for invaluable discussions and advice.
Introduction

Local structural organization and genetic activity of interphase chromosome regions are intimately related. Yet, many fascinating aspects of this interplay are still poorly explored. *Drosophila* polytene chromosomes represent one of the best models of interphase chromosomes available so far. One prominent feature of polytene chromosomes is their banding pattern, which is based on the alternation of dark and compacted bands with interbands that appear light and decondensed. Banding pattern is stable across generations and is known to reflect the functional state of an interphase chromosome. Exactly how the structure and functions are coupled in this context is still an open question.

Our group developed an approach to accurately relate molecular and cytology maps: it combines electron microscopy (EM) mapping of P-based transposons inserted in interband regions with sequencing the adjacent DNA (Demakov et al., 1993; Semeshin et al., 2008). Figure 1 features examples of such insertions in polytene chromosomes. Two scenarios are formally possible: 1) insertion of a transposon into an interband, which results in the formation of a novel band (A); 2) insertion does not result in a novel band (C). EM images of the region 84E from the chromosome arm 3R of wild-type (top) and transgenic (bottom) larvae are shown. Transgenic material forms a novel band (black arrow), which is absent in the chromosomes from the control stock (white arrow) (B). Figure D shows an EM image of the region 12E of the chromosome X from wild-type (top) and transgenic (bottom) larvae. Chromosome morphology remains unaltered (black arrow) in the transgenic strain and is similar to that observed in the wild-type X chromosome (white arrow). Selected "reference" bands are indicated by arrowheads. The bar corresponds to 1 mkm.

*Drosophila* genome sequencing has opened vast opportunities for analyzing the structural and functional elements of chromosomes. A comprehensive analysis of genome organization has been performed for human, fly, and worm, with major efforts from ENCODE (Encyclopedia of DNA Elements) and modENCODE (model organism ENCODE) projects. Massive datasets were generated and it was important to convert this information into functional maps that would inform of the changes in transcriptional, replication, and splicing programmes (modENCODE Consortium, 2010). Recent studies using Dam-ID and ChIP-chip approaches have identified different types of chromatin (Filion et al, 2010; Kharchenko et al, 2011). Nonetheless, the exact functions of the chromatin types identified remain to be explored. It is also important to understand how these chromatin types relate to the bands, interbands, and intercalary heterochromatin regions observed in polytene chromosomes.

The domain organization of *drosophila* larval salivary gland polytene chromosomes has been recently demonstrated to be quite similar to that of diploid cells (Vatolina et al., 2011b; Zhimulev et al., 2012). This makes it straightforward to compare the structure and functioning of the diploid and
polytene chromosomes, which is of great value, because many interesting discoveries using microscopy and molecular genetic tools were made on polytene chromosomes.

Based on the interbands specific protein composition a special algorithm has been developed (Zhimulev et al., 2014). This algorithm allows us to detect the positions of the interbands in the whole Drosophila’s genome. The analysis of the correlation relationships of modEncode proteins made it possible to build a hierarchical clustering and pick out the cluster containing interband specific proteins. These proteins were used to create 4 types of chromatin with the help of the Hidden Markov model. These types were conditionally named Cyan, Blue, Green and Magenta. Cyan chromatin is characterized by the presence of the opened chromatin proteins and it corresponds to the cytological interbands. Magenta chromatin corresponds to black transcriptionally inactive bands. Blue and green chromatins correspond to cytological grey bands.

In the present work, we give a short review of our data on the genetic and molecular organization of bands and interbands, which have been molecularly mapped in the Drosophila melanogaster genome.

**Figure 1. Morphology of P-element Insertions in Polytenic Chromosomes According to Demakov et al., (2011)**
Results

*Molecular and Genetic Organization of Interbands*

We used a set of 32 interbands, for which molecular and cytology localization data were available (Demakov et al., 1993; Schwartz et al., 1998; Demakov et al., 2001; Demakov et al., 2004; Vatolina et al., 2011a; Zhimulev et al., 2014). Using a custom mathematical algorithm, we mapped the span of these interband regions on the physical map. As an example, Figure 2 demonstrates the molecular and genetic organization of the interband 10B1-2/10B3: A – physical map, FlyBase genes and FISH probe (green rectangle). B – positions of promoter types according to Hoskins et al, 2011. C – localization of chromatin types according to Zhimulev et al., 2014. D – five chromatin states, as found in Kc cells by Filion et al, 2010. E – 30 chromatin states in S2 cells by Kharchenko et al., 2011. F – 30 chromatin states in BG3 cells Kharchenko et al., 2011. G – Nucleosome density by Henikoff et al., 2011. H – localization of histone H1 dips in Kc cells by Braunschweig et al., 2009. I – DNase I hypersensitivity sites (DHSs) in S2, BG3, and Kc cells by Kharchenko et al., 2011. J – ORC-binding sites in S2, BG3, Kc cells, and salivary glands by Eaton et al., 2011; Sher et al., 2012. K – Proteins of the NSL complex: NSL1 and MCRC2 in salivary glands by Raja et al., 2010; NSL3 in S2 cells by Lam et al., 2012, NSL1 in S2 cells by Feller et al., 2011. L – interband-specific and active chromatin specific proteins (fly modEncode project).

We performed FISH localization of a handful of interbands (Vatolina et al., 2011b; Zhimulev et al., 2014), and observed a perfect overlap between the FISH-mapped DNA fragments mapping to the 5'-ends of genes, corresponding interbands found in polytene chromosomes and Cyan chromatin fragments (see Figure 2). Furthermore, transposon-tagged interbands and Cyan chromatin were found to perfectly match, too. Thus, our data argues in favor of the idea that Cyan chromatin may correspond to the 5'-ends of the genes on the physical map and also to the interbands of the cytology map (see Figure 3).

Cyan, Blue, Green and Magenta chromatin types occupy 12.7, 16.8, 22.5, and 48.0% of genome total, respectively. Despite the fact that Cyan chromatin occupies the smallest fraction of the genome, it turned out to be significantly enriched for P-element insertions, DHSes, histone H1 dips, Broad-class promoters characteristic of housekeeping genes, head-to-head positioned genes, and ORC components (Zhimulev et al., 2014).
Figure 2. Localization of Proteins and Elements of Genome (modEncode) around the Interband Region 10A1-2/10A3

Of the 32 interbands analyzed here, 26 (81%) encompass broad-type promoters. Hoskins et al, 2011 reported that genes having these promoters are active both in drosophila embryos and adults. We estimate that the vast majority (84%) of all broad promoters map to the Cyan chromatin regions (see Figure 3).

Next, we observe that Cyan chromatin appears to perfectly match the gene regulatory regions (type 1 chromatin, as defined by Kharchenko et al., 2011). Comparison of positions of interband regions with colored chromatin types by Filion et al., (2010) shows that interbands are enriched for active chromatin, YELLOW and RED (see Figure 3).

All the interbands analyzed here appear enriched for the "open chromatin" proteins, proteins involved in transcription and nucleosome remodelling, such as RNA polymerase II, promoter-restricted active histone mark H3K4me3, WDS, NURF301, and ISWI (see Figure 3). Notably, interbands show the enrichment for an interband-specific protein CHRIZ (Gorchakov et al., 2005), whose exact role in the biology of interbands is still unknown. Insulator protein BEAF is found in 84% of the interbands from our list. Among others, the following genomic features are characteristic of interbands: pronounced
depletion for nucleosomes and histone H1, enrichment in DHSes and ORC components (see Figure 2, Figure 3).

We also observed that several NSL complex components are found in the 32 interbands studied here. This complex was reported to be specific for the promoters of the housekeeping genes (Feller et al., 2011; Lam et al., 2012) (see Figure 2, Figure 3). This fact is in good agreement with gene expression data in interbands (see below).

Figure 3 summarizes the frequencies of the chromatin states in 32 selected interbands: A - cyan (1), state 1 (red chromatin) in 9 state chromatin model in BG3 (2) and state 1 (red chromatin) in S2 cells (3-according to Kharchenko et al., 2011, RED, YELLOW, RED/YELLOW and BLACK/BLUE chromatin types (4–7) according to Filion et al, 2010. B – Occurrence of various chromatin states, proteins and other genomic features in interbands. X axis - proteins or genomic features found in different cell cultures. Y axis shows the fraction of interbands demonstrating these characteristics.

All features of interbands (see Figure 2 and Figure 3) indicate that they correspond to an open chromatin state and contain 5’ regulatory parts of genes.

**Figure 3. Positions of Protein Enriched Regions, Chromatin States and Other Genomic Features in 32 Cytologically Defined Interbands According to Zhimulev et al. (2014).**

*Polytene Chromosome Interbands Correspond to Promoter Regions of Housekeeping Genes*  
We proceeded to compare the expression levels for the genes mapping to 18 intercalary heterochromatin (IH) bands and 32 interbands. The exact
molecular borders of these bands and interbands were established, and gene expression data were exported from the FlyAtlas Anatomical Expression Data resource (Chintapalli et al., 2007) (see Figure 4). On average, the expression level of interband-resident genes is 27 times higher than that of the genes found in IH (median values 112 vs. 4.2). We compared the expression of genes across 4 chromatin types on a genome-wide scale. The average transcription level of the genes overlapping with Cyan chromatin is 21.6-fold higher than that of the Magenta-resident genes (median values 110.2 vs. 5.1) and 13.5-fold higher than what is found for Blue- and Green-resident genes (data not shown). This data is indicative of a permanently strong expression of genes, whose 5-UTRs map to Cyan chromatin or interbands (Zhimulev et al., 2014).

**Figure 4.** Box-and-whiskers Diagram Showing Mean Number of Larval and Imaginal Tissues where the Activity of “Interband” (A) and “Band” (B) Genes was Found According to Zhimulev et al (2014)

**Discussion**

The approach originally developed in our group combines the EM mapping of P-transposon insertions in interbands and sequencing of the adjacent interband DNA (see Figure 1). It allowed us to compare side-by-side molecular and cytology data for these regions (Demakov et al., 1993).

We showed that the interbands are typically 1-3 kb long and have unique DNA sequences that are organized into autonomously “open” chromatin. Overall, the interbands share many features, such as reduced nucleosome density, preferential localization histone H1 dips, DHSes, P-element insertions, ORCs, as well as open chromatin proteins (see Figure 2, Figure 3) (Demakov et al., 2004; Semeshin et al., 2008; Vatolina et al., 2011a,b; Demakov et al., 2011; Zhimulev et al., 2014).
A special mathematical algorithm was developed to process protein localization data generated by the modENCODE consortium (Zhimulev et al., 2014), and so we accurately mapped the borders of the interbands on the physical map of the drosophila genome, as well as providing a description of their molecular and genetic organization (see Figure 2, Figure 3). Clearly, interbands were demonstrated to be key to the control of the transcription and replication of the house-keeping genes whose 5'-ends they typically host (see Figure 3). One of the most prominent features shared by interbands and house-keeping genes is localization of CHRIZ (see Figure 2, Figure 3) (Gortchakov et al., 2005). This protein is likely part of the complex that helps maintain the promoters of house-keeping genes in a permanently open configuration.

Moreover, there are three basic states of bands in the polytene chromosomes: band material is densely packed (black bands), partially decondensed (marginally transcribing, grey loose bands), and entirely decondensed (highly transcriptionally active puffs) (see Zhimulev, 1999 for more details). The status of genes found in the bands may be subject to developmental or tissue-specific control, and so the bands should be considered as dynamic regions of the genome. Their transcriptional plasticity depends on the differential activity of the band-resident genes.

Several recent genome-wide chromatin profiling projects have provided localization data for a wide range of chromatin proteins and histone modifications across various cell types and developmental stages. These efforts have greatly expanded our choice of analysis tools and provided unprecedented opportunities to look into the domain organization of drosophila polytene chromosomes. Using these datasets, 60 IH regions were accurately positioned on the molecular map (Belyaeva et al., 2012). This allowed us to analyze the molecular properties of these regions and to compare their organization in salivary glands and in Kc cell line. Differences in the IH replication timing of polytene and diploid cells were highly local, and were likely due to distinct expression of genes in Kc cells. Overall, the IH regions share conserved organization in both polytene and non-polytene cells. They form a special class of chromatin domains and are enriched with tissue-specific genes separated by extremely long intergenic spacers.

Large bands of IH are of special interest. They tend to span several thousand kilobasepairs, and are found scattered across chromosome arms. It was noted that these domains contain clusters of unique tissue-specific genes with shared expression profiles. The IH is generally known to be transcriptionally inert, late-replicating and tightly compacted. In the context of polytene chromosomes, these regions demonstrate the underreplication of DNA, which is manifested as the chromosome breaks and ectopic contacts between heterochromatic regions (Belyaeva et al., 2012). One of the prominent markers of pericentric and intercalary heterochromatin in the salivary glands of D. melanogaster is a product of the SuUR gene (Suppressor of Underreplication) (Makunin et al., 2002). The IH bands were reported to have a low ORC density (McAlpine et al., 2010), which may help explain why underreplication occurs in these regions. Taken together, these data
characterize IH as clusters of repressed single-copy genes having narrow developmental expression profiles, and so while inactive, they tend to be similarly organized across many different cell types. It is the large size of these tightly packed domains that underlies the major properties of IH: late replication and underreplication. Recent body of evidence supports classification of IH bands as a special group of sequences with particular genetic organization and regulation (Belyaeva et al., 2012).

We performed a comprehensive analysis of the available localization data for structural proteins, histone modifications, DHSes, and replication origins in the drosophila genome. These genomic features were compared to the cytological structures found in polytene chromosomes. We thus showed that the identical pattern of alternating bands and interbands is a common theme for the two types of interphase chromosomes studied – the salivary gland polytene chromosomes and the diploid chromosomes from the mitotically dividing cell lines. Thus, the organization of both polytene and non-polytene interphase chromosomes in D. melanogaster appears to be guided by the same principles. Therefore, polytene chromosomes may likely serve as an accurate model of interphase chromosomes (Vatolina et al., 2011b; Zhimulev et al., 2012).

Interbands have a key role in the initiation of transcription and replication and contain 5’ UTR of active genes. In contrast, massive IH bands are depleted for ORC subunits and are polygenic. Band-mapping genes are typically tissue-specific. Thus, pronounced differences in gene density and the sets of associating proteins in interbands, IH bands and smaller decompacted bands may serve as convenient markers to establish the positions of these structures on a molecular map. When combined with EM, this approach allows building a unified molecular and cytogenetic map of bands and interbands in the drosophila polytene chromosomes. Our results argue in favor of a highly ordered organization of interphase chromosomes. The banding pattern observed in the context of polytene chromosomes is also present in the interphase chromosomes of cells that undergo regular mitotic divisions. Importantly the span of these structures in either chromosome type is very similar if not identical across different cell types.

Conclusions

Here we performed a special analysis and discovered the main principles of genetic organization of the interphase chromosome bands and interbands. We used a large set of cytology-mapped interbands (32) and annotated them with various genomic features and chromatin proteins (modEncode datasets). As a result, we obtained highly accurate positions of interband borders on a physical map, and so provided a detailed description of their molecular and genetic organization.

We can confidently state that the polytene chromosome interbands are formed by an open chromatin comprising 5’UTR of a gene and, as a rule, its regulatory region, first exon and a part of the first long intron. Overall, the
interbands associate mainly with RNA polII, nucleosome remodelers WDS, NURF, ISWI, and an interband-specific protein Chriz/CHROMATOR. Since the interbands encompass promoter regions of active genes, they are likewise characterized by low nucleosome density, histone H1 dips and DNaseI hypersensitive sites. Interbands are enriched in ORC-biding sequences, they are hotspots of P-element insertions and signals of house-keeping genes (NSL). Interbands tend to harbor regulatory regions for bidirectionally transcribed genes (found in “head-to-head” orientation), they encompass a “broad” promoter class typically found in house-keeping genes. The protein-coding parts of genes typically maps to the so-called “grey” bands, permanently decondensed structures on the flanks of interbands. Genes, whose promoters map to interbands, are highly active across the vast majority of the tissues and organs studied.

We developed a special algorithm to computationally the process protein-localization data generated by the modEncode project. Using this algorithm, we show that Drosophila genome has about 5700 sites that show all the features shared by the cytologically mapped interbands (Zhimulev et al., 2014).

The comparison of the expression patterns for genes mapping to dense bands and away from the interbands with genes whose promoter regions map to interbands, shows that the former are generally tissue-specific, whereas the latter are represented by ubiquitously active house-keeping genes. Analysis of RNA-seq data (modEncode-FlyBase) indicates that the transcripts from interband-mapping genes are present in most tissues and cell lines studied, across most developmental stages and upon various treatment conditions.

References


