

Regulated chromatin domain comprising cluster of co-expressed genes in *Drosophila melanogaster*

Alla I. Kalmykova, Dmitry I. Nurminsky¹, Dmitry V. Ryzhov and Yuri Y. Shevelyov*

Department of Molecular Genetics of Cell, Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia and ¹Department of Anatomy and Cell Biology, Tufts University School of Medicine, Boston, MA 02111, USA

Received September 7, 2004; Revised January 24, 2005; Accepted February 15, 2005

DDBJ/EMBL/GenBank accession nos⁺

ABSTRACT

Recently, the phenomenon of clustering of co-expressed genes on chromosomes was discovered in eukaryotes. To explore the hypothesis that genes within clusters occupy shared chromatin domains, we performed a detailed analysis of transcription pattern and chromatin structure of a cluster of co-expressed genes. We found that five non-homologous genes (*Crtp*, *Yu*, *CK2βtes*, *Pros28.1B* and *CG13581*) are expressed exclusively in *Drosophila melanogaster* male germ-line and form a non-interrupted cluster in the 15 kb region of chromosome 2. The cluster is surrounded by genes with broader transcription patterns. Analysis of DNase I sensitivity revealed 'open' chromatin conformation in the cluster and adjacent regions in the male germ-line cells, where all studied genes are transcribed. In contrast, in somatic tissues where the cluster genes are silent, the domain of repressed chromatin encompassed four out of five cluster genes and an adjacent non-cluster gene *CG13589* that is also silent in analyzed somatic tissues. The fifth cluster gene (*CG13581*) appears to be excluded from the chromatin domain occupied by the other four genes. Our results suggest that extensive clustering of co-expressed genes in eukaryotic genomes does in general reflect the domain organization of chromatin, although domain borders may not exactly correspond to the margins of gene clusters.

INTRODUCTION

It has been suggested that eukaryotic genome is functionally divided into domains that contain all regulatory elements

necessary for full, correct and positionally independent expression of resident gene(s) (1,2). These domains are evident at the level of local chromatin structure, where transcriptionally active genes are associated with 'open' and acetylated chromatin, whereas silent genes are embedded in compacted and hypermethylated chromatin [reviewed in (3–5)]. The assumption that not only single genes such as chicken glyceraldehyde-3-phosphate dehydrogenase (6), human *apoB* (7), and chicken lysozyme (8,9) genes, but also multigenic loci may be regulated at the level of common chromatin domains, originated from the data obtained on several intensively studied gene clusters. The examples include, but are not limited to, the cluster of β -globin genes in chicken, mouse and human (10–17); growth hormone genes (18,19), *IL-4/IL-13* genes (20) and *PRM1-PRM2-TNP* genes in human (21); ovalbumin genes in chicken (22); and heat-shock genes in *Drosophila* (23). All these clusters comprise paralogous genes that most likely have originated from ancestral genes through local duplications. It is not clear whether the inferences from the studies on these models can be extrapolated to the organization of entire genome, where clusters of paralogous genes occur rather rarely.

Recently, a number of reports have demonstrated non-random clustering of co-expressed genes on chromosomes. First observations of this phenomenon, that we are aware of, date back to 1991 (24), but in 2002 an influx of publications based on the analysis of whole-genome transcription data from different organisms indicated that mechanisms of transcriptional co-regulation, that operate with chromatin domains, are common from yeast to higher eukaryotes [reviewed in (25)]. Moreover, according to the data of Spellman and Rubin (26) over 20% of all *Drosophila* genes are clustered on chromosomes according to their expression patterns and thus may share common chromatin domains. However, direct evidence that would link the observed gene clusters to the chromatin domains was still missing. To address this issue, we thoroughly characterized the cluster of non-homologous testes-specific

*To whom correspondence should be addressed. Tel: +7 095 1960809; Fax: +7 095 1960221; Email: shevelev@img.ras.ru

⁺AF197938, AY078076 and AY078077

genes, and analyzed the chromatin structure in the region. The cluster of five testes-specific genes in the cytological region 60D1-2 of *Drosophila melanogaster* chromosome 2 includes new genes *Crtp*, *Yu* and *CG13581*, along with previously identified *Pros28.1B* (27) and *CK2βtes* (aka *Ssl*) (28,29). Analysis of chromatin sensitivity to DNase I revealed that four out of five genes in the cluster occupy shared chromatin domain which appears to be in the 'closed' configuration in somatic tissues (embryos and larval brain) and in 'open' configuration in larval testes. This finding of the regulated chromatin domain that encompasses cluster of non-homologous co-expressed genes implies that the 'domain' hypothesis may be applicable to a large portion of the eukaryotic genome.

MATERIALS AND METHODS

Isolation and sequencing of cDNA clones

The λZapII cDNA library (Stratagene) custom made from testes of *D. melanogaster* was provided by Dr Tulle Hazelrigg. About 6×10^5 phage plaques were screened on the nitrocellulose lifts with the ^{32}P -labeled probes indicated in Figure 1, yielding numerous phage corresponding to the genes *Crtp*, *Yu* and *CG13581*. After *in vivo* excision from λZapII, both strands of cDNA inserts were sequenced using the Sequenase 2.0 kit (United States Biochemicals). Corresponding genomic regions were subcloned from the cosmid clone #9 (29) into pBlueScriptII SK⁻ vector and also sequenced. The following fragments of the cosmid clone #9 were used as the probes for screening the library: probe **a**, a mixture of 880 bp BamHI–BamHI and 1461 bp BamHI–HindIII fragments; probe **b**, 1113 bp AvrII–PstI fragment; probe **c**, 3103 bp NsiI–NsiI fragment; and probe **d**, 308 bp fragment amplified by PCR using the primers CTCGAATTCGGACCCAGCACTTTTGCATTCCCG and CTCAAGCTTTGACTCGCGGTGGAACCACCCATA.

Northern analysis

For developmental northern analysis, 30 μg of total RNA isolated by TRIzol (Invitrogen) extraction from adult or larval testes, ovaries, embryos, larvae, pupae, gonadectomized adult males or females, and from cell culture, were fractionated by electrophoresis in denaturing formaldehyde-agarose gel and transferred by blotting onto the HyBond-N membrane (Amersham). For northern analysis of mutants, total RNA

was isolated from testes of the *bam*⁸⁶ (30) and *aly*⁵ (31) homozygous adult males. Total RNA isolated from the testes of the *Df(I)w*^{67c23}, *y* strain with normally proceeding spermatogenesis was used as a positive control. Hybridizations and washes were performed according to standard protocols (32). ^{32}P -labeled antisense riboprobes were synthesized with the T7 RNA polymerase and [α - ^{32}P]UTP (3000 Ci/mmol) on the linearized plasmid templates, using the pBlueScriptII SK⁻ T7 promoter. For the templates generated by PCR, the T7 promoter sequence was embedded in one of the PCR primers. Plasmid templates were as follows: full-size *Crtp* cDNA #321 (29) lacking poly(A) tail; *Yu*-specific probe, 1113 bp AvrII–PstI fragment of cosmid #9 (29) cloned into the XbaI–PstI digested pBlueScriptII SK⁻; *CG13581*-specific probe, 308 bp PCR fragment cloned in the EcoRI–HindIII digested pBlueScriptII SK⁻ (primers used: CTCGAATTCGGACCCAGCACTTTTGCATTCCCG and CTCAAGCTTTGACTCGCGGTGGAACCACCCATA); and probe **c**, 3103 bp NsiI–NsiI fragment of cosmid #9 cloned in both orientations into the PstI site of pBlueScriptII SK⁻. PCR-generated templates included 184 bp *CK2βtes*-specific probe, (primers used: CGATAGCCAGACACTGGAGGTAGTCT and GAATTAATACGACTCACTATAGGGAGAGATCCCCCCTTTCGTACTTTAATCG); 364 bp *Pros28.1B*-specific probe (primers used: GGCCACCCTGATCCATTCTG and GAATTAATACGACTCACTATAGGGAGACCAATCAAGGAAACCATCTCG); and 738 bp *anon60Da*-specific probe, (primers used: CCACCATCCGGCGAATGAAG and GAATTAATACGACTCACTATAGGGAGACCTTCGCTTTATCGGGCACA). In all cases, DNA of cosmid #9 was used as a PCR template. As a control for RNA loading, the same filters were hybridized with the probe specific to constitutively expressed gene *Rp49* (33).

RNA *in situ* hybridization

For RNA *in situ* hybridization the same antisense probes were used as for the northern analysis. Antisense RNA digoxigenin-labeled probes were synthesized using T7 RNA polymerase and templates as described above, and hybridized with testes according to conventional protocols (34) with some modifications. Testes were manually dissected, fixed for 1 h on ice in phosphate-buffered saline (PBS) containing 4% paraformaldehyde, and treated with Proteinase K (50 μg/ml) for 8 min. Prehybridization was performed at 60°C in the HS buffer (50% formamide, 5× SSC, 0.1% Tween-20, 100 μg/ml salmon sperm DNA and 50 μg/ml heparin). Hybridization was performed at 60°C overnight in the HS buffer, and was followed by washes at 60°C: HS buffer for 1.5 h; 2× SSC, 0.1% Tween-20 for 30 min; and 0.2× SSC, 0.1% Tween-20 for 30 min. Blocking was performed in PBS containing 0.1% Tween-20 and 0.3% Triton X-100. Incubation with anti-DIG-alkaline-phosphatase-conjugated antibodies (Roche Diagnostics) was performed for 1 h in the same solution, followed by mounting in glycerol/PBS (9:1). Samples were observed using the Leica MZ9-5 microscope.

RT-PCR analysis

Total RNA was extracted from manually dissected adult testes, ovaries and heads, from larval salivary glands and brains, and from 2 to 10 h embryos of the laboratory strain *Df(I)w*^{67c23}, *y*

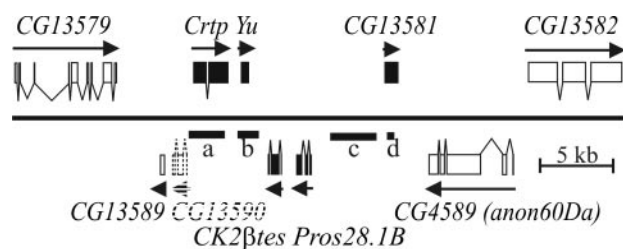


Figure 1. Structure of the 60D1-2 region including the cluster of five testes-specific genes, and surrounding genes with broader expression pattern. Exon-intron structure and the location of genes from the region is according to the GadFly (release 3.1) with addition of the *Yu* transcription unit, according to our data. Testis-specific genes are in black. Gene *CG13590* (dashed) is predicted in *Drosophila* genome annotation, but its existence is not supported by our data. The horizontal bars (labeled a through d) indicate the probes used for testes cDNA library screening and for *in situ* hybridizations.

Table 1. Primers used for real-time RT-PCR and for chromatin analysis

Amplicon	Upper primer	Lower primer
A8	AATGCTCGGCGGAACCTCGGTAATG	CGGGACTGCTGAAGTTGTAAGGA
A10	TCAATCGCCGCAAACATTCCCACA	CACAAAGCGGTACACAATAATCC
A12	ACAGCAGACCACCTCCTCAGCAA	ATCCTTGATCCCAGTGCCCTCTGT
A15	GCATATCATGTGCTGCTGTGGAAC	TTCGACGCAATGGTAGTTATTACA
A16	AAACTGACGGTGTACTGTAATAA	CTGCCATCTGGGGACTACTACTAA
A17	CGGCTCCACGAACGTGACATTTAT	TTCCACTTCCGGCTATTTCGATTTA
A17.5	ACCCTTGCCGAATCAACAACATTAC	CCGCTCGTCCGCCAACTTTGTGTCC
A18	GGATCGGGCTCAGTACAAGTGTCT	GCTTCCGCTCCAATTTCTGATACG
A19	TTTGAGCAGGCGGTGAGATCATGG	GGCGACCGTAGGAGTATCCGTATAG
A21	CTTTCTGTATTTTCGATGGGGTCTCT	CCGCGTCTGTACTGTTCTCTGTTT
A24	GTGTAGGTGGAGCAGGAAATCGCAA	GTAATCAGCAGGAAAAATGTGTAAA
A25	CCTCCATTGCTCTGCGTGTAGTT	GCCACCCGCACTTGCCATTAGTT
A27	AACCCCGTCCGCTCCAAACTAA	TTGGAGGTCTGGCAACATTTTAC
A30	CTTCTCCAGTCTTCTTAATCTATC	TGTAAGTGTGCTGCTGGGCTGCGTTC
A32	TTCCCGTCCGCTCCATCCAGCATA	TCTTTGCCTGCCGCTGACCCATA
A36	CACTCGCCAAAGGAGCTTAGAGCA	GTCACAGATGAACCCGCCAAAGAA
A37	GCAGGTGCCTCTTTGCTGGTCTT	GTGTGTTAGCGGGCATTTTGTT
<i>Actin5C</i>	AGTTGCTGCTCTGGTTGTGCG	CGTAGGACTTCTCCAACGAGG
<i>Crtp</i>	CACCAGGCGCTCATTACAG	TGGGACAGGAGCGACACAG
<i>CK2βtes</i>	GTCGTGTCCAGGAGCATCG	CAGCGTGGACCGAATGGTGT
<i>Pros28.1B</i>	TCCGAAGGGGTCCACTGTG	GTGATTTGCCCGCTTTACAC
<i>CG13581</i>	GCCGGTCTATCCCTTTTCCA	TCGTCTGGGCTGCGTTCGTA
<i>CG13589</i>	AAATCGTGGGTGGTTGTTCACTACT	CTTATACCCATTGGCCTTCTTCATC
<i>CG13579</i>	TTGAGGGCCTCTCCGAGAAGTACTG	CTGCTGCGGCTGGTGGATCTG
<i>CG13582</i>	ATTGAATCGTGTGGGAGCAGAAGT	GAAAAGCCCGAAAGCTGTAGACGAA

of *D.melanogaster* using the TRIzol reagent (Invitrogen), and purified on the RNeasy columns (Qiagen) according to the manufacturer's protocols. Reverse transcription of 1 µg samples was performed with the OmniScript enzyme (Qiagen) in the presence of oligo(T) primer, reaction products were treated with RNase H, and diluted 10-fold with water. Mock-reactions run in the absence of reverse transcriptase were used as negative controls. An aliquot of 2 µl of diluted samples served as templates for 25 µl real-time PCRs, which were performed in the ABI 5700 Sequence Detector using the SYBR Green chemistry (Applied Biosystems). The constitutive transcript of the *Actin5C* gene was used as a cDNA template loading reference. Primer sequences are presented in Table 1.

Chromatin sensitivity to DNase I

The assay was performed essentially as described by Kramer *et al.* (35), with some modifications. Dechorionized 2–4 h embryos and manually dissected larval testes and larval brains were homogenized in the freezing storage buffer (50 mM Tris-HCl, 5 mM MgCl₂, 10 mM NaCl and 25% glycerol, pH 7.5) and frozen at –80°C. After thawing, homogenized tissues were permeabilized in 0.05% NP-40 for 5 min on ice, then pelleted by centrifugation at 1500 g for 8 min at 4°C. Pellets were resuspended in the DNase I-buffer (40 mM Tris-HCl, 0.4 mM EDTA, 10 mM MgCl₂, 10 mM CaCl₂ and 0.1 mg/ml bovine serum albumin) and treated with 0.5 U/ml DNase I (Promega). Two digestion timepoints (10 and 15 min at room temperature) were analyzed and similar results were observed. Reactions were stopped by adding EDTA to 20 mM, genomic DNAs were purified on DNeasy columns (Qiagen) according to the manufacturer's protocol and eluted in 200 µl. An aliquot of 2 µl of the eluates served as templates for 25 µl real-time PCRs, which were performed

as described above. Each reaction was repeated three to six times, and each DNase I-digestion was repeated twice (for embryos and brains) or four times (for testes). Overall, each data point presented in Figure 3 is an average of 6–12 repeated measurements. C_t values for each reaction are presented in Supplementary Table S1. Data were analyzed using the Excel 2001 (Microsoft) and SPSS 10.1 (SPSS) software packages. Primer sequences are presented in Table 1.

Analysis of the chromatin assay data

To quantify chromatin sensitivity to DNase I-digestion, we first measured the relative quantitative PCR (QPCR) yield for each amplicon as the ratio of yields observed with DNase I-treated sample ('output') versus untreated sample ('input'). Results were adjusted to account for the differences in amplicon length, as described in (35). The regions flanking the *60D* cluster genes show high accessibility to DNase I in all tissues (low relative QPCR yields), similar to the constitutively expressed (36) gene *Actin5C* (Table 3). We therefore assumed that these highly accessible regions possess the 'open' chromatin structure characteristic of actively transcribed genes, and chose the level of resistance to DNase I in these regions as a reference. For each tissue, the average relative QPCR yield observed for the amplicons A8 through A12 (corresponding to *CG13579*) and A37 (corresponding to *CG4589*) was calculated, and the relative QPCR yields observed for individual amplicons were normalized against this value to produce the normalized relative yields (NRYs) shown in Figure 3. As a result, 'open' chromatin regions show NRYs close to 1.0, while 'closed' chromatin regions, more resistant to DNase I, show significantly higher NRY values.

The plot data points (NRY versus genomic position) for each tissue were fitted into a panel of simple curve models using the SPSS 10.1 software package. The subset of data

points corresponding to the regulated chromatin domain (amplicons A15 through A25) was analyzed against the rest of the data (treated as the second subset) to evaluate the significance of differences in NRY between the two subsets, also using SPSS 10.1.

HSS mapping using indirect end-labeling

Approximately 0.6 g of 0–18 h embryos were homogenized in 0.3 M sucrose-A1 buffer [60 mM KCl, 15 mM NaCl, 15 mM Tris-HCl (pH 7.5), 5 mM MgCl₂, 0.1 mM EGTA, 0.5 mM DTT, 0.5% Triton X-100 and Complete protease inhibitor cocktail (Roche Diagnostics GmbH)], filtered through two layers of Miracloth (Calbiochem) on the step gradient of 0.3 M sucrose-A1/0.8 M sucrose-A1. Nuclei were pelleted by centrifugation at 3300 g for 5 min at 4°C and resuspended in 0.4 ml of DNase I-digestion buffer [60 mM KCl, 15 mM NaCl, 15 mM Tris-HCl (pH 7.5), 3 mM MgCl₂, 0.1 mM CaCl₂, 0.5 mM DTT and 0.3 M sucrose]. Four 0.1 ml aliquots were digested by 0, 0.05, 0.15 and 0.45 U of DNase I (RQ1, Promega) for 5 min at 37°C. Digestion was stopped by addition of 2 µl 0.5 M EDTA and 10 µl 30% Sarcosyl. Proteinase K (40 µg) was then added and samples were incubated at 50°C overnight. DNA was phenol and chloroform extracted and then ethanol precipitated. Isolated DNA was digested by HindIII and BamHI, separated in 0.75% agarose gel overnight, blotted onto the HyBond-XL membrane (Amersham) and hybridized with ³²P-labeled 839 bp ClaI-HindIII fragment adjoining the promoter region of *CG13581* gene. Hybridization and washes were performed according to standard protocols (32). Filter was exposed with the film for 2 weeks.

RESULTS

Our initial observation was that two testes-specific genes, *CK2βtes* (28,29) and *Pros28.1B* (27), are located <1 kb apart from each other in the chromosomal cytological region 60D1-2. Both genes have constitutively expressed paralogs in the *Drosophila* genome, namely the *CK2β* (29) and *Pros28.1* (27) genes. We hypothesized that duplicated copies of *CK2β* and *Pros28.1* may have integrated into a testes-specific chromatin domain, and this event in part shaped the path for evolution of these copies into testes-specific *CK2βtes* and *Pros28.1B*. This suggestion, which also implies that other testes-specific genes pre-existed in the region before integration of duplicated copies, prompted us to examine the transcription patterns and chromatin structure in the vicinity of the *CK2βtes* and *Pros28.1B*.

Three more testes-expressed genes are located in the *CK2βtes/Pros28.1B* region

To search for another transcription units in the genomic regions adjacent to *CK2βtes* and *Pros28.1B* genes, the hybridization probes were prepared from the cosmid clone #9 (29) covering several kilobases upstream and downstream from these genes (Figure 1) and used for screening of the testes cDNA library. While each of the probes **a**, **b** and **d** revealed numerous positive plaques, probe **c** gave no hybridization signals when 6 × 10⁵ phage plaques were screened. Strand-specific RNA probes for the fragment **c** were hybridized with northern blots and gave no detectable signal in embryos,

larvae, pupae, adult flies, testes, ovaries and heads (data not shown). We conclude that region corresponding to the probe **c** is not transcribed in *Drosophila* and may harbor regulatory functions, e.g. carrying a boundary element (see below).

Partial sequencing of five cDNAs detected by the probe **a** showed that they were identical in the nucleotide sequence, differing only by the 5'- and 3'-truncation points. Sequencing of the longest cDNA and of corresponding genomic region (DDBJ/EMBL/GenBank accession no. AF197938) from cosmid clone #9 revealed 2235 bp open reading frame (ORF) interrupted by 52 bp intron. Encoded 86 kDa polypeptide shows weak (20% amino acid identity) similarity with the human smooth muscle caldesmon across the whole length, hence it was named Crtp (caldesmon-related testes protein). However, despite the name it is not likely that Crtp is related to caldesmon in function, because the level of identity between the proteins in the actin-, myosin-, calmodulin- and tropomyosin-binding domains essential for caldesmon does not exceed the average 20%.

Sequencing of two cDNA clones detected by the probe **d** (DDBJ/EMBL/GenBank accession no. AY078077) revealed ORF region corresponding to the gene *CG13581* predicted in the annotated *Drosophila* genome database. The encoded 23 kDa Arg/Ser-rich polypeptide has no significant similarities in protein databases, and lacks any known protein domain signatures.

Three poly(A)-containing cDNAs detected by the probe **b** and corresponding genomic region were also sequenced (DDBJ/EMBL/GenBank accession no. AY078076). These cDNAs represent an intronless transcription unit *Yu* that lacks obvious ORFs beginning with AUG, but may encode several peptides of <100 amino acid residues starting with the unconventional GUG or CUG initiator codons. Alternatively, *Yu* transcript may be a precursor of miRNA, a representative of the small ~22 nt RNA molecules that are involved in the regulation of expression of protein-coding genes [reviewed in (37)]. Further experiments are necessary to examine which of these possibilities is realized *in vivo*.

Thus, three testes-expressed genes are located in close vicinity to *CK2βtes* and *Pros28.1B*. We further asked whether these genes share additional aspects of transcription with previously characterized *CK2βtes* and *Pros28.1B*, such as testes specificity and expression in primary spermatocytes.

Crtp, *Yu* and *CG13581* are testes specific

Northern analysis demonstrates the presence of *Crtp*, *Yu*, *CK2βtes*, *Pros28.1B* and *CG13581* transcripts exclusively in adult or larval testes (Figure 2A). No signals were detected in the RNA samples isolated from embryos, whole larvae, pupae, gonadectomized males and females, ovaries, heads and cell culture. For each gene tested, the size of the transcripts matched the size of the corresponding cDNA. The observed lack of the hybridization signal in whole larvae or pupae is probably due to excessive dilution of testes transcripts by the RNA from other body parts, since *Drosophila* male larvae and pupae do contain developing testes where most of the male germ-line-specific genes are already transcribed [reviewed in (38)]. Testes specificity of transcription of *Crtp*, *CK2βtes*, *Pros28.1B* and *CG13581* was confirmed by real-time RT-PCR analysis (Table 2).

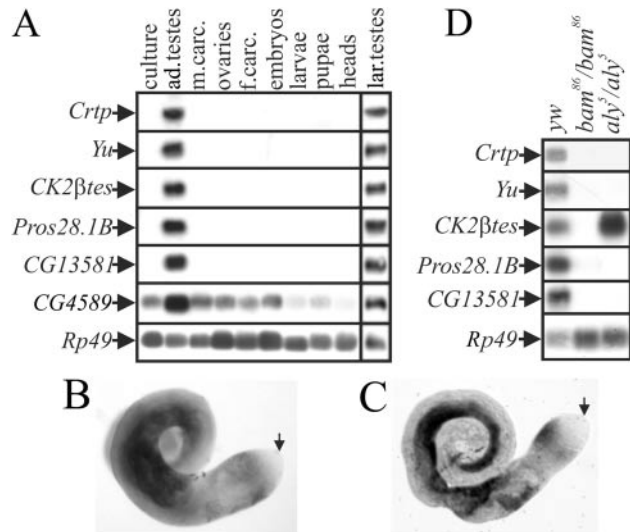


Figure 2. Transcriptional analysis of genes from the *60D1-2* region. (A) Developmental northern analysis shows testis specificity of expression of genes from the *60D* cluster. Total RNA sources are indicated on top; culture corresponds to the *Schneider-2* cells; m.carc. and f.carc. represent gonadectomized males and females, respectively; ad.testes and lar.testes represent adult and larvae testes, respectively. 32 P-labeled antisense RNA probes were used. Hybridization with the antisense RNA probe for the constitutively expressed gene *Rp49* served as the loading control. RNA *in situ* hybridization of DIG-labeled *CG13581* (B) or *Crtp* (C) probes with whole-mount adult testes show onset of transcription in spermatocytes; transcripts persist until late post-meiotic stages of spermatogenesis. Arrow marks tip of testis occupied by stem cells and spermatogonia, where no hybridization was detected. (D) Northern analysis of the transcription of genes from the *60D* cluster in the *bam*⁸⁶ and *aly*⁵ mutants shows universal requirement for the *bam* function, but diverse response to the *aly* deficiency. Total RNA was isolated from testes of males with the genotype indicated on top; the *Df(1)w^{67c23}(2)*, *y* males with normally proceeding spermatogenesis were used as positive control.

Table 2. Gene expression profiles assayed by the real-time RT-PCR analysis

Gene name	Adult heads ^a	Adult testes	Adult ovaries ^a	Embryos ^a	Larval brains ^a	Larval salivary glands ^a
<i>Crtp</i>	0.16	1.00	0.01	0.09		
<i>CK2βtes</i>	0.09	1.00	0.04	0.01		
<i>Pros28.1B</i>	0.06	1.00	0.00	0.00		
<i>CG13581</i>	0.05	1.00	0.00	0.00		
<i>CG13589</i>	0.26	1.00	0.51	0.00	0.00	2.31
<i>CG13579</i>	5.01	1.00	0.95	1.25		
<i>CG13582</i>		1.00		0.00	0.68	192.67

^aAmounts of transcripts in tissues are shown, relative to the amount of transcripts in adult testes.

We also determined transcription patterns of other genes in the region, including *CG4589* (*anon60Da*) and *CG13590* (for gene locations see Figure 1). We detected *anon60Da* transcripts by northern analysis at all stages of *Drosophila* development, although they were noticeably more abundant in testes (Figure 2A). No transcripts were detected for the predicted gene *CG13590* at any stage of development by northern analysis or by real-time RT-PCR (data not shown). The gene *CG13589*, located next to *CG13590*, appears to be transcribed in heads, testes, ovaries and salivary glands according to our real-time RT-PCR data, although no significant transcription

was observed in embryos and larval brains (Table 2). The gene *CG13579*, located adjacent to *CG13589*, is transcribed at similar levels in heads, testes, ovaries and embryos as determined by real-time RT-PCR (Table 2). We also analyzed transcription of *CG13582* which is located next to *CG4589*, and found it to be expressed at similar levels in testes and larval brains and up-regulated in salivary glands, but silent in embryos (Table 2). Therefore, the five genes (*Crtp*, *Yu*, *CK2βtes*, *Pros28.1B* and *CG13581*) form cluster of testes-specific genes (hereafter called the *60D* cluster) flanked by more broadly expressed *CG13589* and *CG13579* on one side, and *CG4589* (*anon60Da*) and *CG13582* on the other side.

Simultaneous onset of transcription of all genes in the *60D* cluster during spermatogenesis

All stages of male germ line cell differentiation are represented in a single adult testis of *Drosophila* [reviewed in (38)]. The primary spermatogonial cells, born at the apical tip of testis from the stem cells, undergo four rounds of mitotic divisions to produce cysts of 16 interconnected spermatogonia. Maturing cysts of spermatogonia are gradually displaced from the tip as they soon enter the meiotic program and become primary spermatocytes. After extended G₂ phase associated with robust gene expression, primary spermatocyte undergo meiotic divisions, producing haploid spermatids. Since spermatocyte cysts continue to be displaced away from the apical tip as they grow, the cells enter meiotic division near the start of the coiled part of the testis.

Figure 2B and C show the results of RNA *in situ* hybridization of DIG-labeled *CG13581* and *Crtp* probes to whole-mount adult testes. The pattern of hybridization with other genes comprising the *60D* cluster was in general the same (data not shown). No labeling was detected at the apical tip occupied by stem cells and spermatogonia. Transcription of each gene starts in spermatocytes and transcripts disappear at the late post-meiotic stages of spermatogenesis. This transcription pattern resembles that of other male germ-line-specific genes, such as the β 2-*tubulin* gene (39), the *don juan* gene (40) and the *Sdic* gene (41).

Transcription of genes comprising the *60D* cluster is diversely affected in the *bam* and *aly* mutants

We analyzed the transcription of genes comprising the *60D* cluster against the background of mutations *bag-of-marbles* (*bam*) and *always early* (*aly*), which affect spermatogenesis in *Drosophila*. Testes of the *bam*⁸⁶ homozygous males are filled with amplified mitotically dividing spermatogonia, with no spermatocytes present (30). Figure 2D demonstrates the absence of transcripts of all testes-specific genes from the *60D* cluster in *bam*⁸⁶ mutants. This is consistent with our RNA *in situ* hybridization data indicating that transcription of these genes is initiated in spermatocytes but not in spermatogonia (Figure 2B and C).

The meiotic arrest gene *aly* controls transcription of numerous genes in testes (31). It encodes a chromatin-associated protein suspected of having a function in chromatin remodeling (42). The transcription of the four genes from the *60D* cluster (*Crtp*, *Yu*, *Pros28.1B* and *CG13581*) appears to be *aly*-dependent, as no hybridization signals were detected in the RNA isolated from testes of homozygous *aly*⁵ mutant

males (Figure 2D). Surprisingly, transcripts of the fifth gene of the cluster (*CK2βtes*) are even accumulated in the *aly*⁵ mutant testes. Since *CK2βtes* is located in the same chromatin domain as the *Crtp*, *Yu* and *Pros28.1B* genes (see below), it seems unlikely that the *aly* protein is involved in 'potentiation' (35) of chromatin domains during spermatogenesis.

Regulated chromatin domain in the 60D cluster region

To study the organization of chromatin in the 60D cluster region, we examined the accessibility of chromatin to the DNase I using a modified PCR-based nuclease resistance assay (35). The segment of genome covered by analysis included the 60D cluster region and surrounding sequences. Seventeen ~0.5 kb PCR amplicons (named according to their positions on the x-axis of Figure 3) were designed to provide a dense coverage across this 29 kb region of interest. Larval testes that were used for the assay contain mostly spermatocytes (where the 60D cluster genes are expressed), and also some spermatogonia and somatic cells [reviewed in (38)]. Larval brains and embryos were chosen as examples of somatic tissue where the 60D cluster genes are silent. Cells from larval testes, larval brain and embryos were permeabilized with detergent. Part of the cell suspension was set aside, and later used to estimate the amount of 'input' DNA. The rest of the suspension was incubated with DNase I. Then, both undigested and partially digested DNA samples were isolated and the quantity of DNA that escaped cleavage by DNase I was estimated by real-time PCR. Since cutting the DNA renders it an ineffective template for PCR, the less accessible is the chromatin region to DNase I, the higher is the yield of PCR product in the DNase I-treated sample ('output'). The PCR yield was measured relative to the sample not treated with DNase I ('input'), and normalized against the values observed for constitutively transcribed genes *CG13579* and *CG4589* flanking the region. The resulting NRYs were plotted along the region. The data (Figure 3) shows variability of chromatin structure of the genomic segment including amplicons A15 through A25 in different tissues.

As shown in the Figure 3, upper panel (see also Table 3), in larval testes the chromatin exhibits nearly uniform high sensitivity to DNase I across the entire region studied. The distribution fits well to a linear model ($P = 0.019$) with the minor slope of 0.022 NRY U/kb and intercept at 0.62 NRY U, i.e. the line very similar to a constant of 1.0. This indicates that in larval testes, the chromatin in the entire region of the 60D cluster and neighboring genes is in the 'open' configuration, similar to the chromatin over *Actin5C* gene.

In contrast, much greater variation in chromatin sensitivity to DNase I across the region was observed in embryos (Table 3) (Figure 3, lower panel). An ~10 kb segment containing testes-specific genes *Crtp*, *Yu*, *CK2βtes* and *Pros28.1B* from the 60D cluster, as well as the non-cluster gene *CG13589*, (amplicons A15 through A25, marked black in Figure 3) showed higher resistance to DNase I, while the rest of the region still demonstrated NRY values not so different from the 1.0 reading representative for the 'open' chromatin. The average NRY for the DNase I-resistant segment is 2.28, this is about two times higher than the average NRY of 1.25 for the rest of the region. The high significance of this difference is

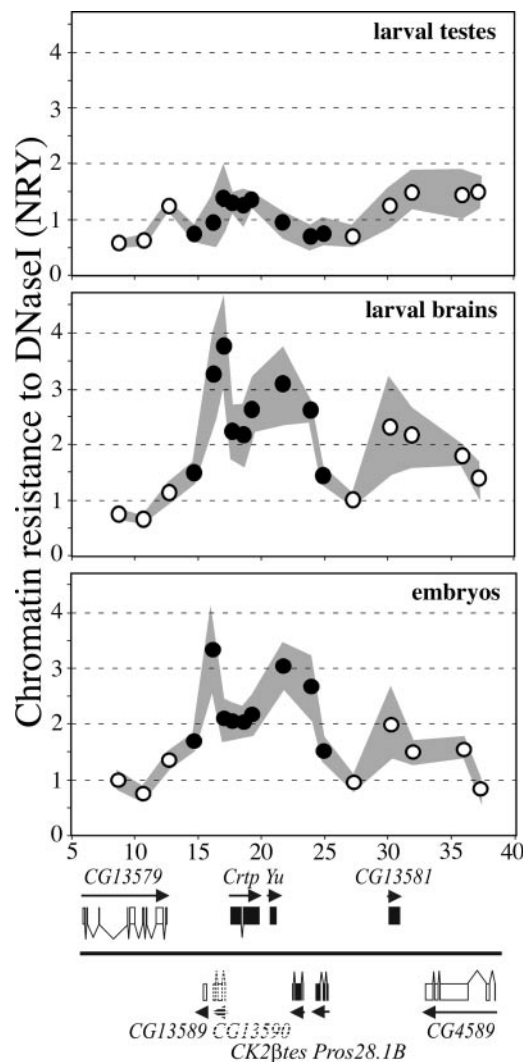


Figure 3. Profiles of chromatin DNase I-resistance across the 60D cluster region and surrounding sequences. Chromatin resistance to DNase I [as normalized relative yield (NRY) for each amplicon] is plotted on the vertical axis; the length of 60D1-2 genomic region (in kb) is plotted on the horizontal. Positions of the genes in the region are shown at bottom. Circles (black for the regulated chromatin domain, and white for the rest of the region) indicate average NRY for each amplicon, the grey area corresponds to the calculated 95% confidence interval. Upper panel: in larval testes, the entire region shows nearly uniform low resistance to DNase I typical for the 'open' chromatin. In contrast, in larval brains (middle panel) and in embryos (lower panel) regulated chromatin domain that contains the genes *CG13589* through *Pros28.1B* shows significantly higher resistance to DNase I, indicative of 'closed' chromatin configuration.

supported by non-parametric tests (the Mann-Whitney U-test and Kolmogorov-Smirnov two-sample test showed P values of 0.002 and 0.012, respectively). For comparison, in larval testes the average NRY for the amplicons A15 through A25 is 1.08, which is not different from the average NRY of 1.13 observed for the rest of the region (0.95-fold difference, P values for the two aforementioned tests are 0.85 and 0.59). The distribution of chromatin resistance to DNase I in embryos is no longer supported by the linear model ($P = 0.78$), the best fit was observed for the quadratic ($P = 0.031$) and cubic ($P = 0.039$) models which can accommodate the 'hill' of resistance to DNase I in the middle of the region.

Table 3. Results of the chromatin assay

Amplicon name	Position in scaffold (bp) ^a	Larval testes				Larval brains				Embryos			
		Average NRY ^b	N ^c	SD	95% CI ^d	Average NRY ^b	N ^c	SD	95% CI ^d	Average NRY ^b	N ^c	SD	95% CI ^d
A8	108405.5	0.54	12	0.14	0.08	0.77	6	0.07	0.05	1.01	12	0.28	0.16
A10	110492.5	0.62	12	0.25	0.14	0.69	6	0.09	0.07	0.76	12	0.11	0.06
A12	112175	1.24	12	0.20	0.11	1.16	6	0.21	0.17	1.29	12	0.23	0.13
A15	114444	0.83	12	0.26	0.14	1.46	6	0.25	0.20	1.67	12	0.20	0.11
A16	115868.5	0.98	12	0.89	0.50	3.22	6	1.02	0.82	3.32	6	0.85	0.68
A17	116704	1.42	12	1.11	0.63	3.82	6	0.89	0.71	2.08	12	0.62	0.35
A17.5	117459.5	1.32	12	0.37	0.21	2.22	6	0.56	0.45	2.02	6	0.34	0.28
A18	118236.5	1.26	12	0.60	0.34	2.15	6	0.71	0.57	2.01	12	0.38	0.22
A19	118939	1.34	12	0.39	0.22	2.69	6	0.64	0.51	2.16	6	0.42	0.33
A21	121367	0.95	12	0.33	0.19	3.09	6	0.94	0.75	3.04	6	0.58	0.46
A24	123583	0.75	12	0.32	0.18	2.62	6	0.31	0.25	2.67	6	0.75	0.60
A25	124541.5	0.84	12	0.36	0.21	1.42	6	0.19	0.15	1.49	6	0.36	0.29
A27	126990	0.76	12	0.29	0.16	0.99	6	0.10	0.08	0.96	6	0.10	0.08
A30	130001.5	1.24	12	0.60	0.34	2.26	6	1.30	1.04	1.99	6	0.97	0.77
A32	131511	1.57	12	0.66	0.37	2.15	6	0.62	0.49	1.43	6	0.29	0.23
A36	135594	1.52	12	0.75	0.42	1.88	6	0.15	0.12	1.67	6	0.20	0.16
A37	136941.5	1.57	12	0.47	0.27	1.37	6	0.40	0.32	0.84	6	0.25	0.20
<i>Actin5C</i>		0.75	6	0.25	0.20	0.62	6	0.22	0.18	0.58	6	0.15	0.12

^aPositions of the center points of amplicons in the AE003464 scaffold are shown.

^bAverage normalized relative yield (NRY) is proportional to efficiency of amplification from the DNase I-digested chromatin template.

^cNumber of real-time PCRs used to produce the average NRY.

^dCalculated 95% confidence interval.

Analysis of chromatin sensitivity to DNase I in larval brains (Table 3) and (Figure 3, middle panel) revealed the pattern almost identical to that observed in embryos. DNase I-resistant 10 kb segment containing testes-specific genes *Crtp*, *Yu*, *CK2βtes* and *Pros28.1B* from the *60D* cluster, as well as the non-cluster gene *CG13589*, (amplicons A15 through A25) showed average NRY of 2.52, highly significantly different from the average NRY of 1.41 observed for the rest of the region (1.79-fold difference, *P* values for the the Mann-Whitney U-test and Kolmogorov-Smirnov two-sample test are 0.007 and 0.054, respectively). The distribution of chromatin resistance to DNase I in brains is also not supported by the linear model (*P* = 0.77), the best fit was observed for the quadratic (*P* = 0.06), cubic (*P* = 0.04) and S-shaped curve (*P* = 0.05) models.

Altogether, the data are indicative of a regulated chromatin domain that includes testes-specific genes *Crtp*, *Yu*, *CK2βtes* and *Pros28.1B* from the *60D* cluster, as well as the non-cluster gene *CG13589*.

DISCUSSION

Eukaryotic genome has been intensively studied in efforts to unravel the relationship between its structure and function. One attractive hypothesis is that chromosomes are organized in independent chromatin domains that contain individual genes or gene clusters, and these domains play the regulatory role, being either permissive or repressive towards gene expression (1,2). The 'domain' hypothesis is supported mainly by the data from thoroughly investigated models such as the β-globin (10–17), growth hormone (18,19), interleukin (20) and some other gene clusters. These clusters are composed of paralogous genes, which are probably the products of local gene duplications. Because of the implied mechanism of formation of these clusters, occurrence of the related genes

in a common chromatin domain may merely be an automatic consequence of gene duplications within the pre-existing domain. Whether or not the resulting structure is employed in co-regulation of the paralogous genes in the cluster, it is not clear that similar mechanism(s) operates to regulate gene expression elsewhere in genome.

Recently, we (43) and others [reviewed in (25)] have revealed the abundant presence of clusters of co-expressed genes in the genomes of multicellular eukaryotes, leading to a tempting suggestion that these clusters outline chromatin domains. For example, according to our data (43), about one-third of the tissue-specific genes form clusters in *Drosophila* genome. Spellman and Rubin (26) have shown that more than 20% of *Drosophila* genes (both constitutively and tissue-specifically expressed) are clustered according to their expression patterns. Here, we present the evidence for the regulated chromatin domain that encompasses cluster of non-homologous co-expressed genes. Considering that majority of the clusters of co-expressed genes consist of non-homologous genes (26,43–45), this finding implies that the 'domain' hypothesis may be applicable to the substantial portion of eukaryotic genome.

We have analyzed in detail the 15 kb genomic region of *Drosophila* genome where five non-homologous testes-specific genes are localized next to each other, thus forming a non-interrupted cluster of co-expressed genes (the *60D* cluster). All genes in the *60D* cluster are activated simultaneously in spermatocytes, as shown by RNA *in situ* hybridizations and by northern analysis of *bam*⁸⁶ mutants that are arrested at the spermatogonial stage. On each side, the cluster is flanked by genes with broader tissue specificity of expression.

Our analysis of chromatin structure indicates that four testes-specific genes in the cluster share a regulated chromatin domain. While sensitivity to DNase I in larval testes was high across the entire region that included the *60D* cluster and adjacent sequences, a segment showing high resistance to

DNase I was observed in larval brains and in embryos. Sensitivity to DNase I is a conventional gauge for the compactness of chromatin. The actively transcribed genes are associated with the DNase I-sensitive, unfolded chromatin, while the 'closed' chromatin structure, typical for the repressed genes, is more resistant to DNase I (6–8,11,12,14,16,21–23). Therefore, our interpretation of the observed patterns of sensitivity to DNase I is that the chromatin of the *60D* cluster and surrounding genes is 'open' in testes (where all these genes are expressed), but a segment encompassing most of the *60D* cluster is 'closed' in somatic tissues represented by embryos and brain. This segment apparently represents a regulated domain of repressed chromatin, and it includes the genes *CG13589*, *Crip*, *Yu*, *CK2βtes* and *Pros28.1B*. All these genes, except for *CG13589*, are transcribed in the male germ line only. Chromatin structure of the domain becomes 'open', i.e. 'potentiated' (35), before or at the primary spermatocyte stage, at which transcription of the cluster genes is initiated.

It is quite intriguing that while the regulated chromatin domain does contain most of the testes-specific genes from the *60D* cluster, the borders of the domain and of the cluster are not identical. On one side, the gene *CG13589*, which is included in the regulated domain, is not a member of the *60D* cluster of testes-specific genes. Our RT-PCR data show that *CG13589* is transcribed at low level in testes, ovaries and heads, but not in embryos and larval brains. Thus, some aspects of tissue specificity of *CG13589* appear to be similar to that of the genes in the *60D* cluster (namely, expression in testes and absence of expression in embryos and in larval brains). The shared chromatin domain may, at least partially, underlie this similarity. This indicates that genes included in common chromatin domain ought to show significantly overlapping, but not necessarily identical transcription patterns.

On the other side of the *60D* cluster we observed that a continuous stretch of co-expressed genes on the chromosome may occupy more than one chromatin domain. The border of regulated domain is clearly positioned to the left of *CG13581*, thus leaving *CG13581* outside of the domain shared by other testes-specific genes of the cluster. 'Closing' of chromatin over *CG13581* was observed in somatic tissues, indicating the possible presence of a second domain of repressed chromatin in the region, however, the effect was not very significant. Whether or not the chromatin domain containing *CG13581* is a regulated one, it is distinct from the domain that contains other genes of the *60D* cluster, being separated by a constitutively DNase I-sensitive region that contains amplicon A27 (Figure 3). It is important to mention here that the constitutive DNase I sensitivity of the amplicon A27 is not due to the presence of DNase I-hypersensitive site (HSS). We have studied this by Southern analysis of the 5 kb fragment between the genes *Pros28.1B* and *CG13581* (encompassing amplicon A27) in embryos, using indirect end-labeling. We found nine 'weak' sites with increased sensitivity to DNase I, however, none of these mapped within A27 (Supplementary Figure S1).

This 'border' region may represent a fixed domain boundary established, e.g. by insulator sequence [reviewed in (46–48)], although other types of boundary elements were described (49,50). Taken together, our studies indicate that while clusters of co-expressed genes do outline chromatin domains, the exact domain pattern can not be deduced solely from the expression data.

Shared chromatin domains provide the tool for co-regulation of genes with similar expression patterns. For clusters of paralogous genes, possible mechanism of creation of multigenic domains is the internal 'stretching' due to local gene duplications within domains. In case of clusters of non-homologous co-expressed genes, shared domains could be created by integration of genes into pre-existing domains. This could partially canalize the evolution of newly integrated genes by pre-setting general aspects of their transcription patterns.

Finding of the regulated chromatin domain implies that there are factors that control its repressive state in somatic tissues, and potentiation in male germ line as it was shown, e.g. for the cluster of neuron-specific genes (51). One potential candidate for such role would be the gene *aly*, which is required for expression of wide spectrum of genes during male gametogenesis, and has been implicated in regulation of chromatin structure (42). We analyzed the effect of *aly* deficiency on expression of testes-specific genes in the *60D* cluster, and found that one gene (*CK2βtes*) does not require the *aly* function for its transcription. *CK2βtes* is located in the middle of regulated chromatin domain, being surrounded by other testes-specific genes that depend on *aly* heavily. Failure to 'open' the regulated domain due to the *aly* mutation would have similar effect on all genes included in the domain. Therefore, the Aly protein is probably involved in more downstream steps of transcriptional regulation in testes.

The results presented here suggest that extensive clustering of co-expressed genes observed recently in eukaryotic genomes does in general reflect the domain organization of chromatin, although exact domain borders may not correspond to the margins of gene clusters. 'Opening' or 'closing' of chromatin domains would outline tissue specificity of expression of genes included in these domains. The chromatin organization then provides a tool for coordinated regulation of multiple genes. Pattern of 'open' and 'closed' chromatin domains along the chromosomes, if maintained through cell divisions, can shape the transcriptome of the cell lineage. Alternatively, modifications of this pattern during development may lead to the coordinated changes in the expression of gene batteries, thus establishing new cell lineages.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr T. Hazelrigg for providing testes cDNA library, Prof. V. A. Gvozdev for his interest in this work and for critical reading of manuscript. This work was supported by Russian Foundation for Basic Research Grants 01-04-48420 and 04-04-48718, by Grant to Scientific Schools from Russian Ministry of Science 2074.2003.4, by Grant of Molecular and Cellular Biology from Russian Academy of Sciences, and by NIH Grant GM61549. The Open Access publication charges for this article were waived by Oxford University Press.

REFERENCES

- Goldman, M.A. (1988) The chromatin domain as a unit of gene regulation. *BioEssays*, **9**, 50–55.
- Bodnar, J.W. (1988) A domain model for eukaryotic DNA organization: a molecular basis for cell differentiation and chromosome evolution. *J. Theor. Biol.*, **132**, 479–507.
- Dillon, N. and Sabbattini, P. (2000) Functional gene expression domains: defining the functional unit of eukaryotic gene regulation. *BioEssays*, **22**, 657–665.
- Xin, L., Liu, D.-P. and Ling, C.-C. (2003) A hypothesis for chromatin domain opening. *BioEssays*, **25**, 507–514.
- Razin, S.V., Farrell, C.M. and Recillas-Targa, F. (2003) Genomic domains and regulatory elements operating at the domain level. *Int. Rev. Cytol.*, **226**, 63–125.
- Alevy, M.C., Tsai, M.-J. and O'Malley, B.W. (1984) DNase I sensitive domain of the gene coding for the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase. *Biochem.*, **23**, 2309–2314.
- Levy-Wilson, B. and Fortier, C. (1989) The limits of the DNase I-sensitive domain of the human apolipoprotein B gene coincide with the locations of chromosomal anchorage loops and define the 5' and 3' boundaries of the gene. *J. Biol. Chem.*, **264**, 21196–22204.
- Jantzen, K., Fritton, H.P. and Igo-Kemenes, T. (1986) The DNase I sensitive domain of the chicken lysozyme gene spans 24 kb. *Nucleic Acids Res.*, **14**, 6085–6099.
- Loc, P.V. and Stratling, W.H. (1988) The matrix attachment regions of the chicken lysozyme gene co-map with the boundaries of the chromatin domain. *EMBO J.*, **7**, 655–664.
- Hebbes, T.R., Clayton, A.L., Thorne, A.W. and Crane-Robinson, C. (1994) Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken β -globin chromosomal domain. *EMBO J.*, **13**, 1823–1830.
- Prioleau, M.-N., Nony, P., Simpson, M. and Felsenfeld, G. (1999) An insulator element and condensed chromatin region separate the chicken β -globin locus from an independently regulated erythroid-specific folate receptor gene. *EMBO J.*, **18**, 4035–4048.
- Bender, M.A., Bulger, M., Close, J. and Groudine, M. (2000) β -globin gene switching and DNase I sensitivity of the endogenous β -globin locus in mice do not require the locus control region. *Mol. Cell*, **5**, 387–393.
- Forsberg, C.E., Downs, K.M., Christensen, H.M., Im, H., Nuzzi, P.A. and Bresnick, E.H. (2000) Developmentally dynamic histone acetylation pattern of a tissue-specific chromatin domain. *Proc. Natl Acad. Sci. USA*, **97**, 14494–14499.
- Forrester, W.C., Thompson, C., Elder, J.T. and Groudine, M. (1986) A developmentally stable chromatin structure in the human β -globin gene cluster. *Proc. Natl Acad. Sci. USA*, **83**, 1359–1363.
- Schübeler, D., Francastel, C., Cimbara, D.M., Reik, A., Martin, D.I.K. and Groudine, M. (2000) Nuclear localization and histone acetylation: a pathway for chromatin domain opening and transcriptional activation of the human β -globin locus. *Genes Dev.*, **14**, 940–950.
- Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R. and Fraser, P. (2000) Intergenic transcription and developmental remodelling of chromatin subdomains in the human β -globin locus. *Mol. Cell*, **5**, 377–386.
- Plant, K., Routledge, S.J.E. and Proudfoot, N.J. (2001) Intergenic transcription in the human β -globin gene cluster. *Mol. Cell Biol.*, **21**, 6507–6514.
- Elefant, F., Cooke, N.E. and Liebhaber, S.A. (2000) Targeted recruitment of histone acetyltransferase activity to a locus control region. *J. Biol. Chem.*, **275**, 13827–13834.
- Elefant, F., Su, Y., Liebhaber, S.A. and Cooke, N.E. (2000) Patterns of histone acetylation suggest dual pathways for gene activation by bifunctional locus control region. *EMBO J.*, **19**, 6814–6822.
- Yamashita, M., Ukai-Tadenuma, M., Kimura, M., Omori, M., Inami, M., Taniguchi, M. and Nakayama, T. (2002) Identification of a conserved GATA3 response element upstream proximal from the interleukin-13 gene locus. *J. Biol. Chem.*, **277**, 42399–42408.
- Choudhary, S.K., Wykes, S.M., Kramer, J.A., Mohamed, A.N., Koppitch, F., Nelson, J.E. and Krawetz, S.A. (1995) A haploid expressed gene cluster exists as a single chromatin domain in human sperm. *J. Biol. Chem.*, **270**, 8755–8762.
- Lawson, G.M., Knoll, B.J., March, C.J., Woo, S.L.C., Tsai, M.-J. and O'Malley, B.W. (1982) Definition of 5' and 3' boundaries of the chromatin domain containing the ovalbumin multigene family. *J. Biol. Chem.*, **257**, 1501–1507.
- Udvardy, A., Maine, E. and Schedl, P. (1985) The 87A7 chromomere: identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *J. Mol. Biol.*, **185**, 341–358.
- Hager, E.J. and Miller, O.L., Jr (1991) Ultrastructural analysis of polytene chromatin of *Drosophila melanogaster* reveals clusters of tightly linked co-expressed genes. *Chromosoma*, **100**, 173–186.
- Hurst, L.D., Pal, C. and Lercher, M.J. (2004) The evolutionary dynamics of eukaryotic gene order. *Nature Rev. Genet.*, **5**, 299–310.
- Spellman, P.T. and Rubin, G.M. (2002) Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J. Biol.*, **1**, 5.
- Yuan, X., Miller, M. and Belote, J.M. (1996) Duplicated proteasome subunit genes in *Drosophila melanogaster* encoding testes-specific isoforms. *Genetics*, **144**, 147–157.
- Kalmykova, A.I., Dobritsa, A.A. and Gvozdev, V.A. (1997) The Su(Ste) repeat in the Y chromosome and β CK2tes gene encode predicted isoforms of regulatory β -subunit of protein kinase CK2 in *Drosophila melanogaster*. *FEBS Lett.*, **416**, 164–166.
- Kalmykova, A.I., Shevelyov, Y.Y., Dobritsa, A.A. and Gvozdev, V.A. (1997) Acquisition and amplification of a testis-expressed autosomal gene, SSL, by the *Drosophila* Y chromosome. *Proc. Natl Acad. Sci. USA*, **94**, 6297–6302.
- McKearin, D.M. and Spradling, A.C. (1990) bag-of-marbles: a *Drosophila* gene required to initiate both male and female gametogenesis. *Genes Dev.*, **4**, 2242–2251.
- Lin, T.-Y., Viswanathan, S., Wood, C., Wilson, P.G., Wolf, N. and Fuller, M.T. (1996) Coordinate developmental control of the meiotic cell cycle and spermatid differentiation in *Drosophila* males. *Development*, **122**, 1331–1341.
- Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual, 3rd edn*. Cold Spring Harbor Laboratory Press, Plainview, NY.
- O'Connell, P.O. and Rosbash, M. (1984) Sequence, structure, and codon preference of the *Drosophila* ribosomal protein 49 gene. *Nucleic Acids Res.*, **12**, 5495–5513.
- Tautz, D. and Pfeifle, C. (1989) A non-radioactive *in situ* hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene hunchback. *Chromosoma*, **98**, 81–85.
- Kramer, J.A., McCarrey, J.R., Djakiew, D. and Krawetz, A. (1998) Differentiation: the selective potentiation of chromatin domains. *Development*, **125**, 4749–4755.
- Tobin, S.L., Cook, P.J. and Burn, T.C. (1990) Transcripts of individual *Drosophila* actin genes are differentially distributed during embryogenesis. *Dev. Genet.*, **11**, 15–26.
- Pasquinelli, A.E. and Ruvkun, G. (2002) Control of developmental timing by microRNAs and their targets. *Annu. Rev. Cell Dev. Biol.*, **18**, 495–513.
- Fuller, M.T. (1993) Spermatogenesis. In Bate, M. and Martinez-Arias, A. (eds), *The Development of Drosophila melanogaster*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY Vol. 1, pp. 71–147.
- Michiels, F., Gasch, A., Kaltschmidt, B. and Renkawitz-Pohl, R. (1989) A 14 bp promoter directs the testis specificity of the *Drosophila* β 2 tubulin gene. *EMBO J.*, **8**, 1559–1565.
- Santel, A., Winhauer, T., Blümer, N. and Renkawitz-Pohl, R. (1997) The *Drosophila* don juan (*dj*) gene encodes a novel sperm specific protein component characterized by an unusual domain of a repetitive amino acid motif. *Mech. Dev.*, **64**, 19–30.
- Nurminsky, D.I., Nurminskaya, M.V., Aguiar, D. and Hartl, D.L. (1998) Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature*, **396**, 572–575.
- White-Cooper, H., Leroy, D., MacQueen, A. and Fuller, M.T. (2000) Transcription of meiotic cell cycle and terminal differentiation genes depends on a conserved chromatin associated protein, whose nuclear localisation is regulated. *Development*, **127**, 5463–5473.
- Boutanaev, A.M., Kalmykova, A.I., Shevelyov, Y.Y. and Nurminsky, D.I. (2002) Large clusters of co-expressed genes in the *Drosophila* genome. *Nature*, **420**, 666–669.
- Roy, P.J., Stuart, J.M., Lund, J. and Kim, S.K. (2002) Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature*, **418**, 975–979.
- Lercher, M.J., Urrutia, A.O. and Hurst, L.D. (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nature Genet.*, **31**, 180–183.
- Udvardy, A. (1999) Dividing the empire: boundary chromatin elements delimit the territory of enhancers. *EMBO J.*, **18**, 1–8.

47. Zhan,H.-C., Liu,D.-P. and Liang,C.-C. (2001) Insulator: from chromatin domain boundary to gene regulation. *Hum. Genet.*, **109**, 471–478.
48. West,A.G., Gaszner,M. and Felsenfeld,G. (2002) Insulators: many functions, many mechanisms. *Genes Dev.*, **16**, 271–288.
49. Donze,D., Adams,C.R., Rine,J. and Kamakaka,R.T. (1999) The boundaries of the silenced HMR domain in *Saccharomyces cerevisiae*. *Genes Dev.*, **13**, 698–708.
50. Fourel,G., Boscheron,C., Revardel,E., Lebrun,E., Hu,Y.F., Simmen,K.C., Müller,K., Li,R., Mermod,N. and Gilson,E. (2001) An activation-independent role of transcription factors in insulator function. *EMBO Rep.*, **2**, 124–132.
51. Lunyak,V.V., Bugess,R., Prefontaine,G.G., Nelson,C., Sze,S.-H., Chenoweth,J., Schwartz,P., Pevzner,P.A., Glass,C., Mandel,G. *et al.* (2003) Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science*, **298**, 1747–1752.