Genomic analysis of *Drosophila* chromosome underreplication reveals a link between replication control and transcriptional territories

Stepan N. Belyakin*†, George K. Christophides†, Artyom A. Alekseyenko*‡, Evgenia V. Kriventseva†, Elena S. Belyaeva*, Roman A. Nanayev*, Igor V. Makunin*, Heidelberg Fly Array Consortium§, Fotis C. Kafatos†¶, and Igor F. Zhimulev*

*Institute of Cytology and Genetics of Siberian Division, Russian Academy of Sciences, Novosibirsk 630090, Russia; †European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany; and §Zentrum für Molekulare Biologie, Ruprecht-Karls-Universität, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany

Contributed by Fotis C. Kafatos, April 1, 2005

In *Drosophila* polytene chromosomes, most late-replicating regions remain underreplicated. A loss-of-function mutant of the suppressor of underreplication [*Su(UR)*] gene suppresses underreplication (UR), whereas extra copies of this gene enhance the level and number of regions showing UR. By combining DNA microarray analysis with manipulation of the number of *Su(UR)* gene copies, we achieved genomic-scale molecular identification of 1,036 genes that are arranged in clusters located in 52 UR chromosomal regions. These regions overlap extensively (96%) but are not completely identical with late-replicating regions of mitotically dividing Kc cells in culture. Reanalysis of published gene expression profiles revealed that genomic regions defined by replication properties include clusters of coordinately expressed genes. Genomic regions that are UR in polytene chromosomes and late replicated in Kc cell chromosomes show a particularly common association with transcriptional territories that are expressed in testis/males but not ovary/females or embryos. An attractive hypothesis for future testing is that factors involved in replication control, such as SU(UR), may interact physically with those involved in epigenetic silencing of transcription territories.

DNA replication | polytene chromosomes | suppressor of underreplication | transcriptional silencing

DNA replication in higher eukaryotes is tightly regulated in space and time during the S phase of the cell cycle. The chromosomes are organized into functional domains of DNA replication, the replication foci, within which replication begins simultaneously. Many foci are composed of clusters of replicons, which are considered stable units of the chromosome structure (1, 2). Attempts have been made to correlate specific replication programs with gene expression patterns that may establish the epigenetic chromosomal status (3–6).

The timing of replication in various genomic regions defines them as early-replicated (ER) or late-replicated (LR). Genomic regions that are LR include pericentric and other types of heterochromatin. Early replication strongly correlates with gene activity (3, 5, 7). However, when subjected to position-effect variegation, euchromatic regions that are normally ER become LR (8), and establishment of a LR state during development strongly correlates with gene silencing (8–11).

Polytene chromosomes of *Drosophila melanogaster* are a unique model for studying replication domains, because of their size and cytological properties and because of the availability of the genome sequence. The size of such domains, their chromosomal distribution, and their genetic and functional organization in specific differentiated cells are matters of substantial interest. Approximately 240 LR regions have been detected in these chromosomes, some of which (≈25% in *Oregon*-R WT flies) contain weak spots manifested as specific breaks that show incomplete local polytenization [underreplication (UR)]. The LR and UR regions of the chromosomal arms are called intercalary heterochromatin because they share several common features with pericentromeric heterochromatin, including chromatin condensation, frequent ectopic pairing, and location on the inner side of the nuclear envelope (12). However, the information content of these regions has remained unexplored to date.

At least two intercalary heterochromatin regions in polytene chromosomes, 89E and 84AB, contain silenced homeotic Polycomb-dependent genes of the *Bithorax* and *Antennapedia* complexes. These chromosome sites are known to contain trimethylated histone H3-K9 (13) and to bind Polycomb-Group silencer proteins (14). Therefore, there are grounds to believe that other intercalary heterochromatin regions may be also genetically silenced (12).

DNA replication in polytene chromosomes depends on the *suppressor of UR* [*Su*(*UR*)] gene (15). UR is suppressed in *Su*(*UR*) mutants, *Su*(*UR*)$^-$, but augmented in a 4x*Su*(*UR*)$^+$ transgenic line carrying two additional copies of the WT *Su*(*UR*) allele. Moreover, in the 4x*Su*(*UR*)$^+$ line some LR but normally fully replicating regions become UR, demonstrating that extra copies of this gene affect DNA replication, also within those LR regions, which are normally not UR. In addition, detection of SU(UR) protein in LR regions of polytene chromosomes (16) suggests a direct involvement of this protein in both LR and UR. Here, we have exploited these properties of the *Su*(*UR*) gene to define UR regions of the larval salivary gland polytene chromosomes at the level of DNA sequences and identified a large set of genes (1,036 or 7.5% of the genome) that are clustered in 52 UR regions. Of these regions, 50 (96%) are also LR in the nonpolytenic Kc cells, demonstrating a surprisingly consistent timing of replication in two unrelated tissues. Furthermore, a strong link between replication and specific transcription properties has been revealed. We have shown that specific types of transcriptional territories are preferentially located in certain specific types of replication-defined regions of chromosomal arms. Testis-specific territories are associated with UR regions,

GENETICS

Abbreviations: UR, underreplication; Su(UR), suppressor of UR; LR, late replication; LU$^-$R, late but not UR; LfUR, LR-flanking UR; ER, early replication.

‡Present address: Harvard Medical School–Partners Healthcare Center for Genetics and Genomics, 77 Louis Pasteur Avenue, Boston, MA 02115.

§Heidelberg Fly Array Consortium: M. Hild[a], B. Beckmann[b], S. A. Haas[c], B. Koch[a], V. Solovyev[d], C. Busold[b], K. Fellenberg[b], M. Boutros[e], M. Vingron[c], F. Sauer[a,f], J. D. Hoheisel[b], and R. Paro[a].

[a]Zentrum für Molekulare Biologie Heidelberg, University of Heidelberg, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany; [b]Division of Functional Genome Analysis, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany; [c]Max Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany; [d]Softberry, Inc., 116 Radio Circle, Suite 400, Mount Kisko, NY 10549; [e]Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany; and [f]Department of Biochemistry, University of California, Riverside, CA 92521.

¶To whom correspondence should be addressed. E-mail: kafatos@embl.de.

and, importantly, these regions lack ovary- and embryo-specific territories. This pattern appears to be associated with gene silencing because it occurs in Kc cells and salivary glands, where testis-specific genes are not expressed, and at chromosomal sites that bind the replication-related Su(UR) protein and one or more known silencing factors.

Materials and Methods

DNA Microarray Hybridization and Analysis. Labeling of genomic DNA was performed according to standard protocols with minor modifications. Three to 5 μg of genomic DNA were digested for 4 h with HaeIII before labeling. Labeled samples were purified through Qiagen (Valencia, CA) PCR purification columns; hybridizations were performed overnight at 42°C in a buffer containing 50% formamide, 6× SSC (1× SSC = 0.15 M sodium chloride/0.015 M sodium citrate, pH 7), 0.5% SDS, and 5× Denhardt's reagent (0.1% polyvinylpyrrolidone/0.1% Ficoll/0.1% molecular-biology-grade BSA); and washes were carried out at room temperature (twice for 15 min in 0.1× SSC/0.1% SDS and twice for 15 min in 0.1× SSC). Microarray scanning and analysis was performed with the GENEPIX PRO 3.0 scanner and software; data normalization was performed with the GENESPRING 7 software (Silicon Genetics, Redwood City, CA), and data clustering and visualization with the CLUSTER and TREEVIEW programs. We used two different DNA microarray platforms: the first version of the *Drosophila* Berkley EST collection (DGC1) and the entire *Drosophila* gene set (17). The sensitivity of the method to detect UR was first optimized by using the DGC1 platform with total genomic DNA isolated from male and female WT adults (*Oregon*-R) until we could reliably detect the 2-fold difference for euchromatic genes located on the X chromosome. A second optimization benchmark was provided by the experiments that compared gene dosage in salivary gland of 4x$Su(UR)^+$ males and $Su(UR)^-$ females. UR does not occur in the male X chromosome; thus, the comparison revealed a 2-fold difference for X-linked genes in addition to the difference attributed to UR at autosomal loci. The final experiments were performed with microarrays encompassing the entire gene set. Three independent biological samples from females of each strain were assayed. The results were *Lowess* normalized by using the GENESPRING software (Silicon Genetics), and the reproducibility of the results was checked as described in ref. 18. Features deviating >3 SD ($P < 0.01$) from the average were not further considered.

Identification of UR Regions. Normalized replication values (ratios of DNA representation in 4x$Su(UR)^+$ vs. $Su(UR)^-$ salivary gland chromosomes) of 11,673 genes showing statistically consistent hybridization signals between the three experimental replicates were sorted according to their position in the fly genome (FlyBase Release 3.1 database, available at www.flybase.org). A sliding averaging window of 10 genes, one gene per step, was applied to the normalized data. Resulting values were compared with the average value of all windows in each respective chromosomal arm. Gene windows with values significantly lower ($P < 0.05$) than the average of that chromosomal arm defined the UR region. Series of overlapping windows were considered part of one UR region, and the outer boundaries of each series were set as the boundaries of the respective UR region. Simulation runs with sliding windows of 5 or 20 genes demonstrated robustness of the applied procedure. LR regions were defined from the original data (7) in a similar manner.

Southern Blot Analysis. Total DNAs from 50 salivary glands and from 25 sets of larval brains and imaginal discs were digested with HindIII endonuclease. DNA was separated in agarose gel and transferred to Hybond-NX membrane (Amersham Pharmacia). DNA fragments were PCR-amplified from genomic DNA, cloned, and labeled with [^{32}P]dATP by random priming. Hybridizations were performed according to the protocol recommended by the manufacturer (Hybond-NX), and blots were exposed for various periods of time at −70°C with Agfa CP-BU x-ray film. Signal intensity was measured by using a Hewlett–Packard Scan Jet 4C/T scanner and the BAND LEADER 3.0 program. Relative DNA abundance was calculated as the ratio of hybridization intensity in salivary glands to imaginal discs after normalization to the *rosy* gene, which is fully replicated in polytene tissues.

Detection of Transcriptional Territories. The gene expression data of a previously defined developmental data set (19) were initially divided into seven transcriptional programs (see legend of Fig. 4). The relative expression data of each gene (as compared to the standard reference, which was a mixture of all developmental stages) within these programs were averaged, and arithmetic mean values over or below 2-fold were considered indicative of up- or down-regulation, respectively; in-between values were regarded as indicating no regulation. The obtained data were then arranged according to genomic positions, and a sliding nine-gene window (step one gene) across the genome was applied to detect regions enriched in coregulated genes.

Results and Discussion

We used the experimental protocol summarized in Fig. 1 to identify UR regions in the *Drosophila* polytene chromosomes. Total DNAs prepared from late larval salivary glands of the $Su(UR)^-$ mutant strain (where UR is suppressed) and from late larval salivary glands of the 4x$Su(UR)^+$ strain (where UR is enhanced) were labeled with two different fluorophores, respectively, mixed, and hybridized to DNA microarrays containing all predicted genes of the *Drosophila* genome (17). Comparison of DNAs from those two strains was of utmost importance, as it magnified the UR signal and permitted its unambiguous detection. In the example shown in Fig. 1*A*, pseudored fluorescence corresponds to UR in 4x$Su(UR)^+$ vs. $Su(UR)^-$ DNAs, whereas pseudoyellow fluorescence corresponds to equal levels of polytenization. Three independent experiments were performed, and 11,673 sequences corresponding to unique genes of the fly genome that yielded highly reproducible profiles ($P < 0.01$) were further processed. By using the genome annotation (FlyBase Release 3.1), we sorted data according to the position of each gene in the genome and generated a whole-genome polytenization profile for the salivary gland chromosomes (Data Set 1, which is published as supporting information on the PNAS web site). We report below the analysis of 52 genomic regions, which encompass genes showing statistically significant ($P < 0.05$) UR (Table 1, which is published as supporting information on the PNAS web site). As expected, most chromosomal regions, which are always replicated completely, showed similar polytenization levels in $Su(UR)^-$ and 4x$Su(UR)^+$ strains.

We validated the microarray-based statistically significant replication profiles by comparing them with corresponding Southern blot-based profiles from three different genomic regions: 19E and 11A on the X chromosome and 89DE on the 3R chromosomal arm (Fig. 1*B* and data not shown). The profile of region 89DE was reported in ref. 20. The boundaries, length, and main features of the three types of profiles proved quite comparable (given experimental fluctuations and differences in exact coordinates of assessed sequences).

The 52 statistically significant UR regions share several common characteristics: (*i*) they are all located at known cytological sites of LR in the salivary glands cells, (*ii*) all but one region (35B) were shown previously to colocalize with SU(UR) in WT chromosomes (16), and (*iii*) their chromosomal positions predominantly coincided with cytologically defined weak spots (which are a morphological criterion of UR). However, some

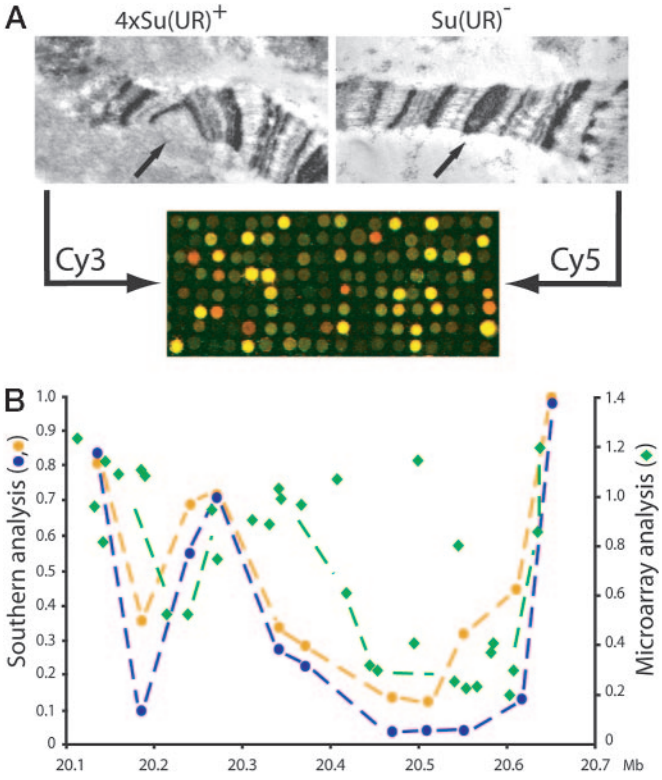

Fig. 1. Principle of detecting UR regions with DNA microarray analysis. (*A*) (*Upper*) In the salivary gland polytene chromosomes of the 4x*Su*(*UR*)$^+$ line (*Left*) UR is augmented and results in weak spots, yielding characteristic breaks (region 89E arrow), whereas in *Su*(*UR*) mutants (*Right*) UR and breaks are no longer detected. Genomic DNAs prepared from salivary glands of female third instar larvae of the *Su*(*UR*)$^-$ and 4x*Su*(*UR*)$^+$ lines were fluorescently labeled with Cy5 and Cy3, respectively, mixed, and hybridized to glass slide DNA microarrays. (*Lower*) Red spots in the microarray image represent genes that are overrepresented in the *Su*(*UR*)$^-$ line and correspondingly underrepresented in the 4x*Su*(*UR*)$^+$ line. (*B*) Abundance profile of DNA from the 19E region of the polytene X chromosome of female larvae of the 4x*Su*(*UR*)$^+$ strain, obtained by using microarrays (diamonds) and Southern blot validation of differences in DNA abundance in the WT and 4x*Su*(*UR*)$^+$ strains (yellow and blue dots, respectively). DNA fragments spaced 30–90 kb apart were used as probes for Southern blot analysis; therefore, the data points do not necessarily correspond to those of the microarray analysis. Abscissa: Genomic physical map according to the *Drosophila* genome annotation 3.1 (www.ensembl.org). Ordinates: Normalized Cy-3/Cy-5 signal ratios in the microarray experiments (right axis) and relative DNA abundance in Southern blots using the *rosy* gene as a calibrator (left axis).

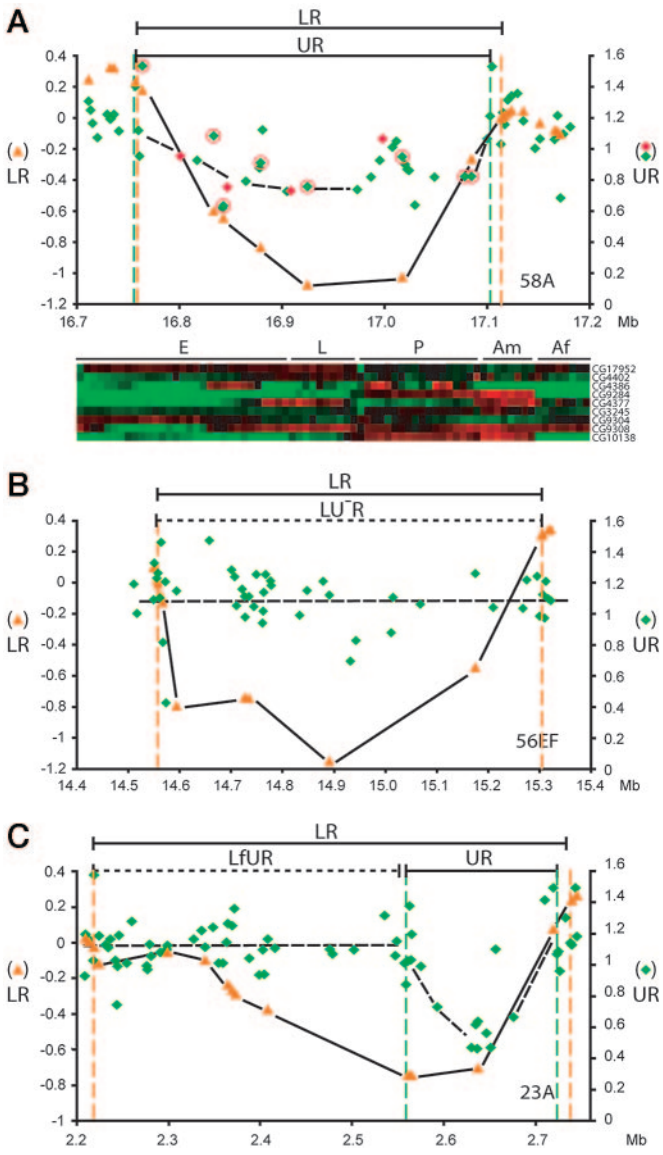

Fig. 2. Comparisons of UR profiles in salivary glands (diamonds) and LR in Kc cells (7) (triangles) in the indicated chromosomal regions. The detected boundaries of the UR (green) and LR (red) regions are indicated with vertical dashed lines. (*A*) (*Upper*) Coincidence of UR in salivary glands and LR in Kc cells in UR region 58A. Genes that are highly expressed in testis (26) are indicated with red diamonds. (*Lower*) The known developmental expression profile of nine genes (encircled diamonds) from the developmental data set (19). E, embryo; L, larva; P, pupa; Am, adult male; Af, adult female. (*B*) DNA abundance profile in LU$^-$R region 78E, which shows LR in Kc cells and not UR in salivary glands. ER is an adjacent ER region in Kc cells. (*C*) Comparison of the LR in Kc cells and UR in salivary gland in the region 64D: LR in Kc cells clearly extends beyond the UR region in salivary gland polytene chromosomes (LfUR). Abscissa: Genomic physical map according to the *Drosophila* genome annotation 3.1 (www.ensembl.org). Ordinates: LR, replication timing in Kc cells presented as $\log_2$-transformed ratios of DNA abundance at early vs. late S phase of the cell cycle (7); UR, DNA polytenization levels presented as ratios of DNA abundance in 4x*Su*(*UR*)$^+$ vs. *Su*(*UR*)$^-$ larval salivary glands.

weak spots were not represented among the 52 UR regions, possibly because of low degree of UR and therefore difficulty of detection, or because of absence of their sequences from the microarrays, if the UR regions are unusually short or predominantly intergenic.

We noted a striking coincidence of observed UR patterns in the polytenic salivary glands with recently reported (7) LR patterns in cultured nonpolytenic Kc cells (Data Set 1 and Table 1). Of the 52 UR regions, 50 (96%) also replicate late in Kc cells (938 of 1,036 detected genes) (Fig. 2*A*). However, some LR regions in Kc cells are completely polytenized in the salivary glands; these regions were named "late but not UR" (L$^-$UR) regions (Fig. 2*B*). In other cases, the regions of UR in the salivary gland and those of the LR in the Kc cells overlap but are not coterminous; in such cases, the LR regions flanking an UR region were designated as LR-flanking UR (LfUR) regions (Fig. 2*C*). Altogether, the data indicate substantial similarity albeit not full coincidence of replication programs among different *D. melanogaster* cell types, polytenic salivary gland cells, and mitotically dividing cultured cells of embryonic origin.

We examined in detail the genetic organization of the most interesting class, the UR regions. These regions ranged in length from 114 to 618 kb and collectively encompassed 1,036 predicted genes or 7.5% of the *D. melanogaster* genes (Fig. 5, which is published as supporting information on the PNAS web site).

GENETICS

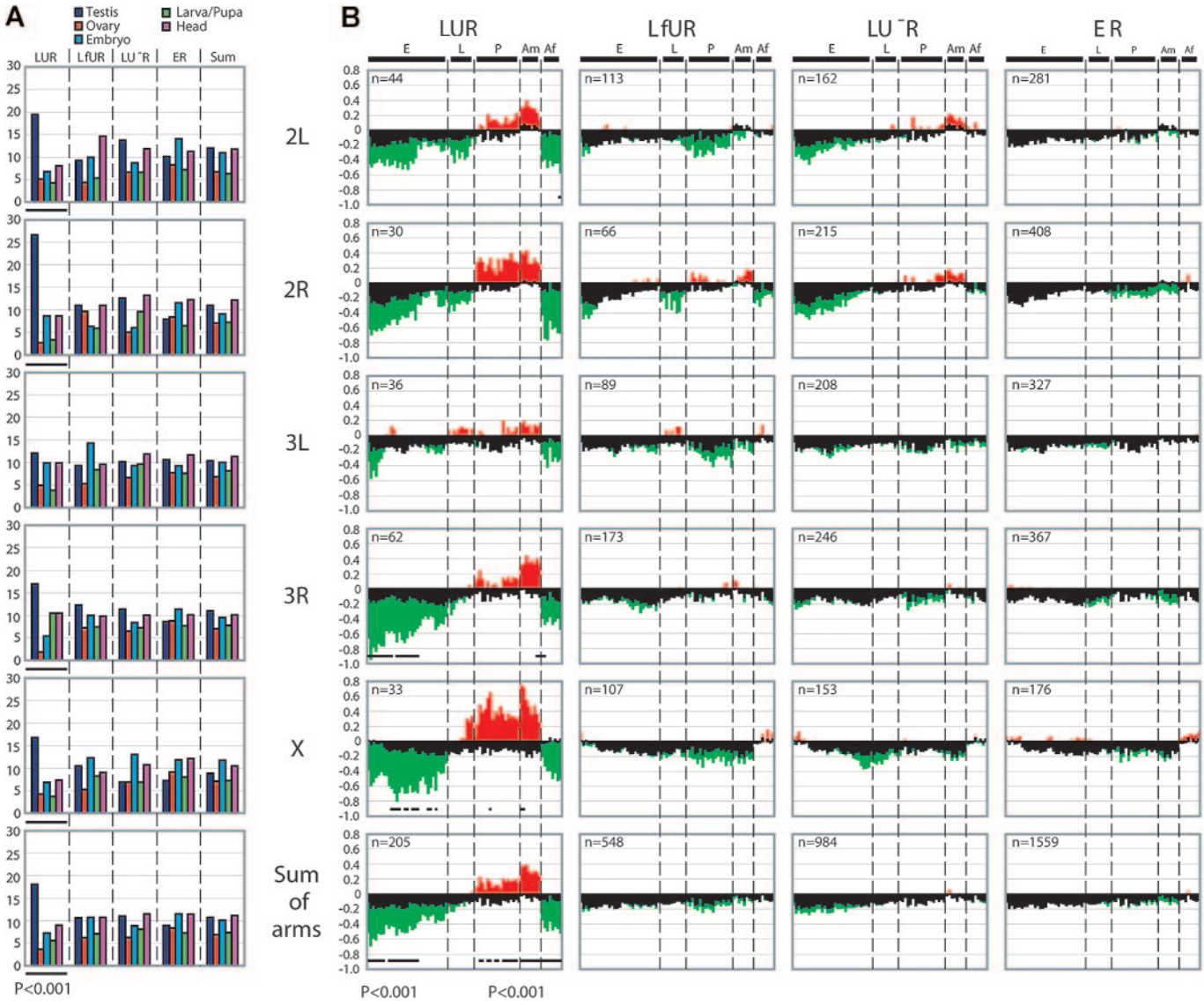

Fig. 3. Most UR regions show enrichment in genes that are highly expressed in the adult male (testis) and silenced in the adult female (ovary) and embryo. A similar but less prominent pattern is detected in the LU¯R regions of 2L and 2R. In the five rows of panels, genes are classified according to their location in the *D. melanogaster* chromosomal arms, and the vertical columns represent the various replication-related regions. (*A*) The ordinate shows the percentage of genes categorized as specific in the cDNA data set (26) (blue, testis; red, ovary; cyan, embryo; green, larva/pupa; and violet, head) in the four different replication-related regions and in the entire chromosomal arm (Sum). Horizontal bars below each diagram indicate statistically significant differences (contingency table test) with $P < 0.001$ cutoffs. (*B*) Average expression of genes ($\log_2$-transformed) from the developmental data set (19) (red, up-regulation; green, down-regulation) in each of the four types of replication-related regions (ordinate). The baseline expression of the entire data set is indicated in black in all panels. In the original data set, the expression of each gene was referenced to a mixture of all developmental stages, normalized for the median, and log-transformed. This total intensity normalization probably explains a small bias toward down-regulation. Horizontal bars that are shown below the diagrams indicate areas of the averaged expression profile that exhibit significant (Student's *t* test, $P < 0.001$) deviation from the average expression profile of the entire chromosomal arm. Abscissa: 75 developmental time points starting from unfertilized eggs to adult flies. E, Embryo; L, larvae; P, pupae; Am, adult male; Af, adult female.

Some UR regions contain tandemly duplicated genes. Examples are a cluster of 20 closely related genes encoding a family of transmembrane proteins (*Osiris* cluster or *Tpl*-locus) (21) at 83DE and a cluster of repeated histone genes located in region 39DE. We tested the overprevalence or underprevalence of Gene Ontology (GO-Slim) terms associated with genes in the UR regions by using the GOTOOLBOX software that is based on a hypergeometric test with Bonferroni correction (22). The analysis detected significant overprevalence of genes with unknown biological function and significant underprevalence of genes involved in biosynthetic processes (both at $P < 0.001$; Table 2, which is published as supporting information on the PNAS web site). However, no clear correlation between UR and gene function could be unambiguously established.

A recent study has shown that ≈20% of the *Drosophila* genome is represented by groups of 10–30 adjacent and similarly expressed genes, which are not otherwise functionally related (23). These gene groups were defined as genomic transcriptional territories. Although the 52 UR regions encompass only 7.5% of the *Drosophila* genome, 30 of them correlate with such transcriptional territories (3-fold enrichment, $P < 0.01$); all are LR in the Kc cells (Table 1). The Kc cell study also reported a link between ER and transcriptional activation (7), suggesting that genes located in LR regions not only become active synchro-

nously but may also be coordinately inactivated in somatic cells. In addition, among the 52 UR regions, 32 (61%) have been previously shown to bind antibodies to Polycomb-Group proteins (Table 1), which repress homeotic gene expression (24). An example is the well known cluster of homeotic genes known as the *Bithorax* complex, located in 89E. It has been suggested that the presence of silencing complexes may result in condensed chromatin structure, thus delaying and suppressing replication (25).

We inspected our data for possible correlation between replication properties and coordinate gene expression/silencing by using two different expression data sets: the microarray-based developmental data set of expression profiles (19) and an independent data set consisting of genes that are differentially overexpressed ("specific") in particular tissues or developmental stages, according to subtraction analysis of corresponding cDNA libraries (26). The developmental and cDNA data sets encompassed 3,296 and 5,401 genes, respectively; we have assigned 205 and 451 of these to UR regions, respectively.

By consolidating the developmental expression profiles of genes associated with UR regions, we noted that genes within the same region often show similar expression (Fig. 6, which is published as supporting information on the PNAS web site). In particular, genes from UR regions are often up-regulated in males during metamorphosis but not in females (e.g., Fig. 2 and Fig. 7, which is published as supporting information on the PNAS web site). Interestingly, these male-specific genes appear to be expressed in the male germ line, because they are no longer expressed in mutants of the *tudor* gene (19), which plays a key role in male germ line development.

To explore further the possibility that such coordinate gene expression is associated with DNA replication properties, we first mapped the cDNA data set (26) to the four types of regions (UR, LfUR, LU$^-$R, and putative ER regions) in each of the five major chromosomal arms (Fig. 3*A*). Indeed, this analysis revealed that genes specific for certain tissues or stages (testis, ovary, embryo, larva/pupa, and head) are unequally distributed among these four replication-related types of regions. Statistical tests performed on the original binary (specific/nonspecific) data set using χ^2 contingency tables showed that the unequal distributions in each of the chromosomal arms except for 3L, and in all five arms combined, were highly significant ($P < 0.001$). In particular, UR regions (especially in 2L and 2R) are highly enriched in testis-specific genes, compared with each respective chromosomal arm as a whole. In contrast, ovary- and embryo-specific genes are significantly underrepresented in UR regions. Neither the LU$^-$R regions nor the LfUR regions show a similar significant enrichment or underrepresentation.

We mapped in a similar manner the developmental data set (19), which has a much higher temporal resolution although fewer genes. We first averaged the temporal profiles in all four types of regions (same as in Fig. 3*A*) in each chromosomal arm and in the sum of all five arms combined (black profiles in Fig. 3*B*). Next, we averaged the profiles in each type of region within a chromosomal arm and in all arms combined and displayed all deviations of these profiles from the respective baseline in red or green for overexpression or underexpression, respectively (Fig. 3*B*). The UR regions of all chromosomal arms combined showed strong underexpression in embryos and adult females, consistent with the observed deficit of embryo- and ovary-specific genes in the cDNA data set. Conversely, the UR regions showed strong overexpression in males and late pupae, again consistent with the overabundance of testis-specific genes in the cDNA data set. In general, male-specific up-regulation was often accompanied by pupal overexpression (see also Figs. 2*A*, 6, and 7), consistent with the fact that gonads develop extensively during the pupal stages. These findings were statistically significant at the level of $P < 0.001$ by Student's t test. Similar overexpression or underexpression patterns were evident in the URs of individual arms to different degrees: Those in the 3R and X were statistically significant at the same level for embryos, males, and females. In 2L and 2R, similar conclusions could be supported, but at lower levels of confidence ($P < 0.01$), partly because of the lower number of genes analyzed.

The other three types of regions (LfUR, LU$^-$R, and ER) did not show significant deviations from the baseline in the sum of all arms. However, LfUR regions in 2R and LU$^-$R regions in 2R and 2L showed a similar pattern to UR regions, contrasting with underexpression of LfUR regions in pupae and males in the 2L, 3L, and X chromosomes. These interesting patterns apart of LU$^-$R in 2L were supported at a lower level of statistical confidence ($P < 0.05$). In summary, the analyses presented in Fig. 3 clearly indicated that developmentally regulated genes tend to cluster in a coordinate manner, in chromosomal regions defined by their DNA replication properties.

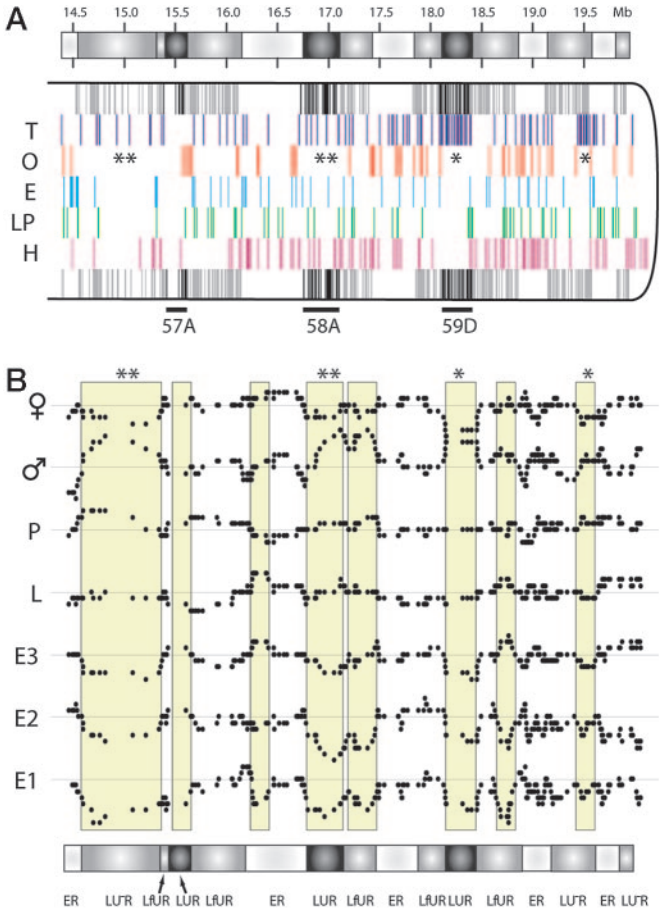

Fig. 4. Correlation of replication-related regions with transcriptional territories in a 5.8-megabase fragment of the chromosomal arm 2R. On the top and bottom are the genomic scales with the regions of different replication timing presented as shaded boxes. Asterisks (* or **) indicate transcriptional territories discussed in the text. (*A*) Schematic representation of the four replication-related related regions (first and last rows of the vertical bars: UR, black; LU$^-$R and LfUR, gray; ER, no color) and the five types of specifically expressed genes from the cDNA data set (26). Characterized genes are shown in five vertical differently colored bars (testis, blue; ovary, red; embryo, cyan; larva/pupa, green; and head, violet). (*B*) Detection of transcriptional territories (boxed) using the developmental gene expression data (19) and a sliding window across the genome. E1, embryo aged 0–3 h; E2, embryo aged 3–10 h; E3, embryo aged 10–24 h; L, larva; P, pupa; Am, adult male; Af, adult female. Each dot reflects the degree of gene activation: below the horizontal gray line indicates down-regulation, and above the line indicates up-regulation for each of seven selected stages.

The obtained data prompted us to perform more detailed genome-wide analysis of transcriptional territories by using the expression profiles of the cDNA data set (26). We displayed schematically each chromosomal arm and mapped onto it the replication-related related regions and the five categories of specifically expressed genes (Fig. 8, which is published as supporting information on the PNAS web site). In the magnified segment at the end of 2R that is shown in Fig. 4*A*, two purely testis-specific territories are evident (Fig. 4*A*, single asterisks), one encompassing the UR 59D and the other located at the distal end of an LU$^-$R, at 60A. Two additional territories (one UR and one LU$^-$R Fig. 4*A*, double asterisks) also showed enrichment in testis-specific genes and absence of ovary- and embryo-specific genes, although this was less pronounced. In contrast, ER regions encompass genes of all five expression types, some showing internal subclustering. We also analyzed transcriptional territories by using the independent data of the developmental data set (19) and a sliding nine-gene window across the genome. As shown in Fig. 4*B*, this method confirmed the existence of the four UR and LU$^-$R-associated testis-specific territories that were illustrated in Fig. 4*A*.

Conclusion

This study is a clear example of how genome-wide studies conducted by different authors can be combined by using the genome sequence as the reference framework, leading to unexpected, discovery-driven hypotheses. Here, we have used the unique ability of the SU(UR) protein to modulate UR in *D. melanogaster* polytene chromosomes and thus identified for the first time a large set of genes (1,036 or 7.5% of the genome) that are clustered in 52 UR regions. Most of these (96%) are also LR in the nonpolytenic Kc cells, which are of embryonic origin. Our first finding of general interest is the surprisingly consistent timing of replication in these two unrelated tissues. The second major finding is the strong association between LR (in salivary gland cells and Kc cells) and testis-specific genomic territories, which are deficient in embryo- and ovary-specific gene clusters. LR of testis-specific territories appears to be associated with transcriptional silencing: These territories occur in Kc cells and salivary glands, where testis-specific genes are not expressed, and at chromosomal sites that bind the replication-related SU(UR) protein and the known silencing factors Polycomb-Group. A recent genomic study has revealed an association between genomic binding of the chromatin repressor, suppressor of variegation 3-9, and male-specific gene expression (27). However, none of the male-specific genes detected in that study are located in the UR regions, suggesting two different mechanisms for male-specific gene silencing. Further studies are needed to test the obvious hypothesis that emerges from these discoveries: that molecular machines implicated in LR share key factors with molecular machines implicated in epigenetic silencing of certain types of transcriptional territories. The revealed regions of UR, which are enriched in coordinately expressed genes that form transcriptional territories, represent a good model for studying the relationship between genome replication programs and epigenetic gene silencing.

We thank Prof. V. F. Semeshin (Russian Academy of Science, Novosibirsk, Russia) for the gift of the EM photographs presented in Fig. 1. S.N.B. and G.K.C. were supported in part by European Molecular Biology Organization Short Term and Marie Curie Fellowships, respectively. This work was funded by Russian Foundation for Basic Research Grant 02-04-48222, the Russian State Program for Scientific Schools, Russian State Program for Frontiers in Genetics Grants 2-02PNG and PD02-1.4-74, and by the European Molecular Biology Laboratory.

1. Berezney, R., Dubey, D. D. & Huberman, J. A. (2000) *Chromosoma* **108,** 471–484.
2. Jackson, D. A. & Pombo, A. (1998) *J. Cell Biol.* **140,** 1285–1295.
3. MacAlpine, D. M., Rodriguez, H. K. & Bell, S. P. (2004) *Genes Dev.* **18,** 3094–3105.
4. Gilbert, D. M. (2002) *Curr. Opin. Cell Biol.* **14,** 377–383.
5. McNairn, A. J. & Gilbert, D. M. (2003) *BioEssays* **25,** 647–656.
6. Pasero, P. & Schwob, E. (2000) *Curr. Opin. Genet. Dev.* **10,** 178–186.
7. Schubeler, D., Scalzo, D., Kooperberg, C., van Steensel, B., Delrow, J. & Groudine, M. (2002) *Nat. Genet.* **32,** 438–442.
8. Zhimulev, I. F. (1998) *Adv. Genet.* **37,** 1–566.
9. Lima-de-Faria, A. & Jaworska, H. (1968) *Nature* **217,** 138–142.
10. Arcos-Teran, L. (1972) *Chromosoma* **37,** 233–296.
11. Kalisch, W. E. & Hagele, K. (1976) *Chromosoma* **57,** 19–23.
12. Zhimulev, I. F. & Belyaeva, E. S. (2003) *BioEssays* **25,** 1040–1051.
13. Cowell, I. G., Aucott, R., Mahadevaiah, S. K., Burgoyne, P. S., Huskisson, N., Bongiorni, S., Prantera, G., Fanti, L., Pimpinelli, S., Wu, R., *et al*. (2002) *Chromosoma* **111,** 22–36.
14. Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A. & Pirrotta, V. (2002) *Cell* **111,** 185–196.
15. Belyaeva, E. S., Zhimulev, I. F., Volkova, E. I., Alekseyenko, A. A., Moshkin, Y. M. & Koryakov, D. E. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 7532–7537.
16. Makunin, I. V., Volkova, E. I., Belyaeva, E. S., Nabirochkina, E. N., Pirrotta, V. & Zhimulev, I. F. (2002) *Genetics* **160,** 1023–1034.
17. Hild, M., Beckmann, B., Haas, S. A., Koch, B., Solovyev, V., Busold, C., Fellenberg, K., Boutros, M., Vingron, M., Sauer, F., *et al*. (2003) *Genome Biol.* **5,** R3.
18. Yang, I. V., Chen, E., Hasseman, J. P., Liang, W., Frank, B. C., Wang, S., Sharov, V., Saeed, A. I., White, J., Li, J., *et al*. (2002) *Genome Biol.* **3,** research0062.
19. Arbeitman, M. N., Furlong, E. E., Imam, F., Johnson, E., Null, B. H., Baker, B. S., Krasnow, M. A., Scott, M. P., Davis, R. W. & White, K. P. (2002) *Science* **297,** 2270–2275.
20. Moshkin, Y. M., Alekseyenko, A. A., Semeshin, V. F., Spierer, A., Spierer, P., Makarevich, G. F., Belyaeva, E. S. & Zhimulev, I. F. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 570–574.
21. Dorer, D. R., Rudnick, J. A., Moriyama, E. N. & Christensen, A. C. (2003) *Genetics* **165,** 613–621.
22. Martin, D., Brun, C., Remy, E., Mouren, P., Thieffry, D. & Jacq, B. (2004) *Genome Biol.* **5,** R101.
23. Spellman, P. T. & Rubin, G. M. (2002) *J. Biol.* **1,** 5.
24. Ross, J. M. & Zarkower, D. (2003) *Dev. Cell* **4,** 891–901.
25. Zhimulev, I. F., Belyaeva, E. S., Makunin, I. V., Pirrotta, V., Volkova, E. I., Alekseyenko, A. A., Andreyeva, E. N., Makarevich, G. F., Boldyreva, L. V., Nanayev, R. A. & Demakova, O. V. (2003) *Chromosoma* **111,** 377–398.
26. Boutanaev, A. M., Kalmykova, A. I., Shevelyov, Y. Y. & Nurminsky, D. I. (2002) *Nature* **420,** 666–669.
27. Greil, F., van der Kraan, I., Delrow, J., Smothers, J. F., de Wit, E., Bussemaker, H. J., van Driel, R., Henikoff, S. & van Steensel, B. (2003) *Genes Dev.* **17,** 2825–2838.