

# Integrative analysis of gene amplification in *Drosophila* follicle cells: parameters of origin activation and repression

Jane C. Kim,<sup>1,2</sup> Jared Nordman,<sup>1</sup> Fang Xie,<sup>1,2</sup> Helena Kashevsky,<sup>1</sup> Thomas Eng,<sup>1</sup> Sharon Li,<sup>1</sup> David M. MacAlpine,<sup>3</sup> and Terry L. Orr-Weaver<sup>1,2,4</sup>

<sup>1</sup>Whitehead Institute, Cambridge, Massachusetts 02142, USA; <sup>2</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA; <sup>3</sup>Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA

In metazoans, how replication origins are specified and subsequently activated is not well understood. *Drosophila* amplicons in follicle cells (DAFCs) are genomic regions that undergo rereplication to increase DNA copy number. We identified all DAFCs by comparative genomic hybridization, uncovering two new amplicons in addition to four known previously. The complete identification of all DAFCs enabled us to investigate these in vivo replicons with respect to parameters of transcription, localization of the origin recognition complex (ORC), and histone acetylation, yielding important insights into gene amplification as a metazoan replication model. Significantly, ORC is bound across domains spanning 10 or more kilobases at the DAFC rather than at a specific site. Additionally, ORC is bound at many regions that do not undergo amplification, and, in contrast to cell culture, these regions do not correlate with high gene expression. As a developmental strategy, gene amplification is not the predominant means of achieving high expression levels, even in cells capable of amplification. Intriguingly, we found that, in some strains, a new amplicon, *DAFC-22B*, does not amplify, a consequence of distant repression of ORC binding and origin activation. This repression is alleviated when a fragment containing the origin is placed in different genomic contexts.

[*Keywords:* DNA replication; ORC; transcription; H4 acetylation; chorion; oogenesis]

Supplemental material is available for this article.

Received February 17, 2011; revised version accepted May 20, 2011.

The initiation of DNA replication occurs from discrete genomic regions called replication origins. Although the protein complexes that bind DNA and license origins for replication are known and evolutionarily conserved, the properties that enable a DNA sequence to function as a replication origin in metazoans are poorly delineated (Gilbert 2004, 2010; Cvetic and Walter 2005; Mechali 2010). One reason is that analysis of metazoan replication origins has, until recently, proceeded from molecular characterization of a small number of identified origins (Aladjem 2007). For example, the DHFR locus in Chinese hamster ovary cells has been studied extensively using a variety of methods; replication at this locus initiates in a broad initiation zone containing multiple inefficient initiation sites (Hamlin et al. 2010). Recent genome-wide origin mapping studies in *Drosophila*, mouse, and human cell culture have greatly increased the number of identified metazoan origins (Cadoret et al. 2008; Sequeira-

Mendes et al. 2009; Karnani et al. 2010; MacAlpine et al. 2010; Mesner et al. 2011), revealing that most origins coincide with active transcription units—specifically, the transcription start site. These data confirm conclusions from analyses of individual replication origins that there is no sequence-specific motif for origin recognition complex (ORC) binding or origin specification in metazoans. However, how local transcription affects replication initiation remains unknown, apart from a few specific examples (Danis et al. 2004; Mesner and Hamlin 2005; Xie and Orr-Weaver 2008).

Developmentally regulated gene amplification is an important origin discovery tool and model for investigating the regulation of metazoan replication origins in vivo (Claycomb and Orr-Weaver 2005). Specific genomic regions are amplified through replication-based mechanisms to increase gene copy number via either chromosomal excision followed by extrachromosomal amplification or repeated bidirectional replication initiation from an endogenous chromosomal locus. Gene amplification increases the DNA template to allow for sufficient levels of gene products required at high levels, such as rRNA in frog oocytes and cocoon proteins in Sciarid fly salivary glands. In *Drosophila*, two chorion gene clusters on the third

<sup>4</sup>Corresponding author.

E-mail weaver@wi.mit.edu.

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.2043111>.

Freely available online through the *Genes & Development* Open Access option.

and X chromosomes are amplified in ovarian follicle cells, somatic epithelial cells that surround the oocyte and secrete the components of the eggshell (Spradling 1981). Amplification enables the eggshell proteins to be produced in a short developmental period. When gene amplification is abolished—as in female-sterile mutants of replication factors such as *ORC2*, *MCM6*, and *DBF4*—these flies lay inviable eggs with thin eggshells, attributed to the inadequate template for transcription of eggshell genes (Landis et al. 1997; Landis and Tower 1999; Schwed et al. 2002).

*Drosophila* follicle cell gene amplification occurs by a DNA rereplication-based mechanism, making it a powerful model for investigating metazoan origin function (Park et al. 2007). In the follicle cell amplicons, repeated rounds of initiation of DNA replication from specific amplification origins produce gradients of increased DNA copy number. DNA replication during amplification uses the same known initiation factors that act in the typical S phase, such as *ORC*, *Cdt1*, and the *MCM* complex. In addition, gene amplification is amenable to diverse experimental approaches to study replication initiation. First, the process occurs within the context of developing egg chambers that are morphologically distinct and can be isolated for experimental analysis, allowing replication events to be studied in the context of development. Gene amplification begins during follicle cell differentiation after genomic replication has ended. Therefore, methods to assay DNA replication, including quantitative PCR (qPCR) or immunofluorescence of the nucleotide analog bromodeoxyuridine (BrdU) to visualize newly replicated DNA, can be used to assess the precise timing of replication events and to separate initiation and elongation phases at individual amplicons. Finally, *Drosophila* genetic tools for introducing DNA at ectopic sites allow one to delineate *cis* requirements for gene amplification.

In the context of gene amplification, the relationship between replication and transcription has two facets: how replication and increased DNA copy number affect transcription, as well as how local transcription affects origin selection and activation. Although gene amplification is considered a strategy to augment gene expression to high levels, how frequently amplification is used to achieve this output and whether amplification always leads to high expression levels are unknown. This latter question is especially important for evaluating the consequences of gene amplification in cancer cells. Sequencing cancer genomes has revealed a high frequency of chromosomally integrated gene amplification (Meyerson et al. 2010). Investigating the relationship between increased DNA copy number and gene expression in the developmentally programmed context of follicle cell gene amplification may shed light on this relationship in cancer cells.

To investigate gene amplification as a developmental strategy and metazoan replication model, we isolated pure populations of amplification stage follicle cells to identify both the complete catalog of follicle cell amplicons with array-based comparative genomic hybridization (aCGH) and analyze transcription on a genome-wide scale using next-generation sequencing. This provides the first high-quality transcriptional profile of an amplifying differenti-

ated tissue. Because synchronous gene amplification in follicle cells begins at a specific developmental stage, these amplification origins can be analyzed precisely with respect to transcription, *ORC* localization, and histone modifications. Finally, we investigate the determinants of origin activation at one amplification origin by exploiting its property of strain-specific amplification.

## Results

### *CGH identifies two new follicle cell amplicons*

To identify all of the amplified regions in follicle cells, we employed an aCGH strategy, an expansion of a previous study that used cDNA microarrays containing <50% of all *Drosophila* genes and no intergenic regions (Claycomb et al. 2004). Before follicle cells begin gene amplification, they undergo three rounds of endoreduplication, or chromosomal replication without intervening mitoses, to reach 16C DNA copy levels. We used flow cytometry to isolate 16C nuclei, which encompass a mixed population enriched for amplification stages spanning exit from the endocycle, the start of amplification (stage 10), and the stage immediately prior to egg deposition. DNA isolated from 16C nuclei was competitively hybridized with diploid embryonic DNA on genome-wide tiling microarrays containing one 60-mer probe approximately every 600 base pairs (bp), representing all single-copy chromosomal sequences. Because of this high-density coverage, we are confident that we identified all of the amplicons present in euchromatin in these cells. Using this approach, we confirmed the presence of the four previously known follicle cell amplicons and identified two new follicle cell amplicons: *DAFC-22B* and *DAFC-34B* (*Drosophila* amplicon in follicle cell, followed by cytological position) (Fig. 1A).

Like the four previously identified amplicons, *DAFC-22B* and *DAFC-34B* show a gradient of replicated DNA that spans ~100 kb (Fig. 1B,C). The maximum aCGH enrichment ratios for *DAFC-22B* and *DAFC-34B* are less than those of the chorion amplicons *DAFC-7F* and *DAFC-66D* but are comparable with *DAFC-30B* and *DAFC-62D* (Table 1). As expected, the genomic regions in *DAFC-22B* and *DAFC-34B* contain genes. The 23 genes within the amplified region of *DAFC-34B* are generally <5 kb in length, and one gene, *Vm34Ca*, located in the peak of amplification, encodes a structural component of the vitelline membrane, the innermost layer underlying the *Drosophila* eggshell (Mindrinos et al. 1985). Notably, there is just one 60-kb gene of unknown function—*CG7337*—in the most amplified region of *DAFC-22B*, thus differing from the genomic organization of the other follicle cell amplicons, which contain small genes that are ~1 kb in length encoding eggshell proteins and enzymes.

*Amplified genes are not necessarily highly expressed, and highly expressed gene regions are not typically amplified*

We performed high-throughput RNA sequencing (RNA-seq) to assess expression of genes in the precisely defined

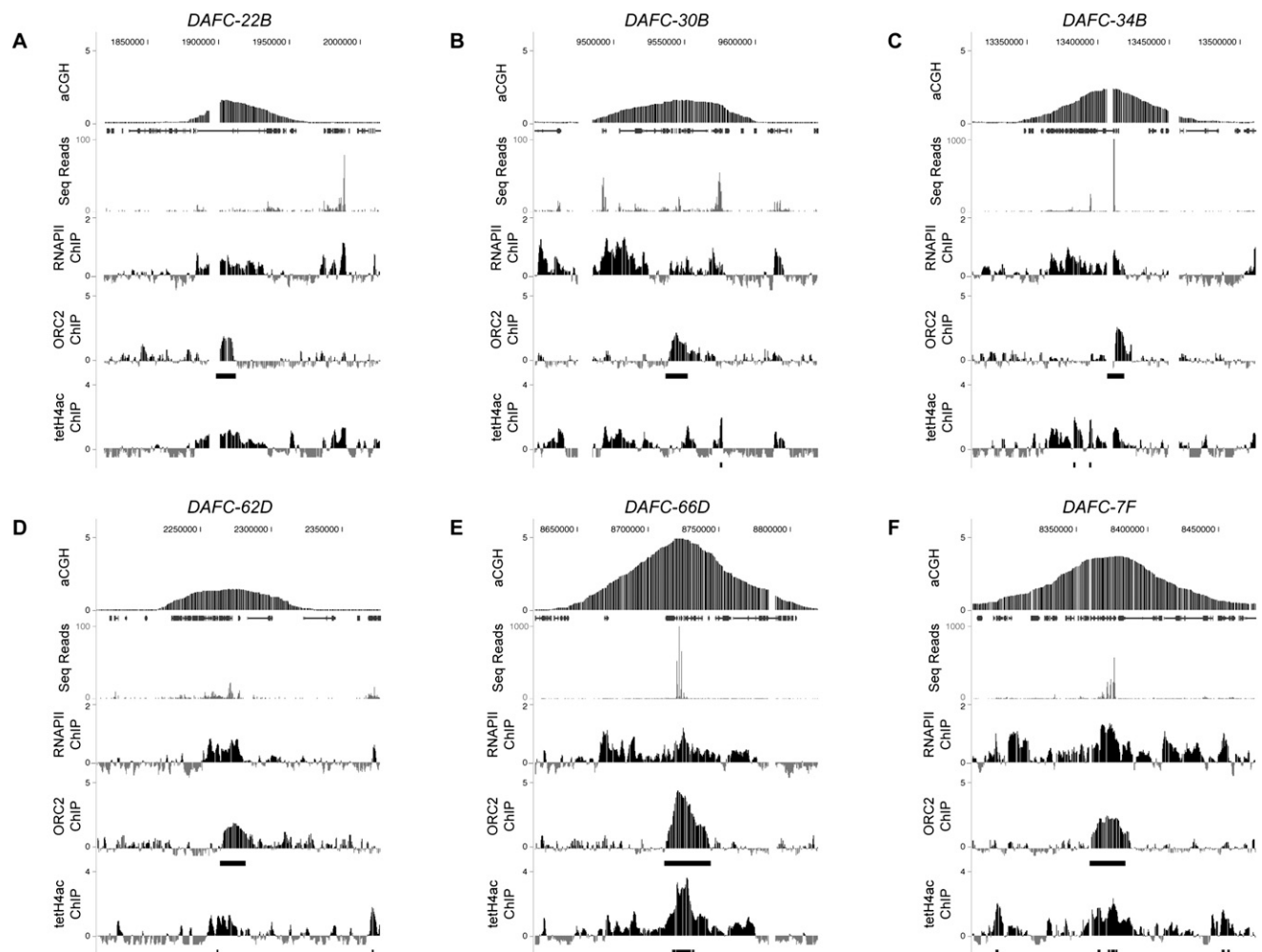


**Table 1.** *DAFCs*

Genomic position	CGH max (log <sub>2</sub> )	ORC coordinates	ORC zone
<i>DAFC-22B</i>	1.5503	chr2L: 1,898,141–1,912,322	14182 bp
<i>DAFC-30B</i>	1.5354	chr2L: 9,536,621–9,552,287	15667 bp
<i>DAFC-34B</i>	2.338	chr2L: 13,406,581–13,418,558	11978 bp
<i>DAFC-62D</i>	1.4083	chr3L: 2,263,821–2,281,806	17986 bp
<i>DAFC-66D</i>	4.9087	chr3L: 8,711,291–8,743,880	32590 bp
<i>DAFC-7F</i>	3.6786	chrX: 8,359,161–8,384,368	25208 bp

correspond precisely with gene expression studies using developmental Northern blots and in situ hybridization, validating the accuracy of our RNA-seq quantification (Griffin-Shea et al. 1982). Other genes within the 100-kb amplified gradients of *DAFC-66D* or *DAFC-7F* were expressed at low levels. Thus, being located within an

amplicon does not, by default, lead to abundant expression, nor does it, for expressed genes, enhance expression to uniform levels. Although gene amplification can promote high expression of some genes, additional regulatory mechanisms fine-tune the levels, developmental timing, and spatial control of gene expression.



**Figure 2.** Gene expression and localization of RNAPII, ORC, and tetra-acetylated H4 at the six amplicons. (A–F) Two-hundred-kilobase regions of all follicle cell amplicons with sequence reads from RNA-seq of purified follicle cells, as well as localization of RNAPII, ORC, and tetra-acetylated H4 by ChIP–chip from stage 10 egg chambers. Maximum sequence read is 100 for *DAFC-22B*, *DAFC-30B*, and *DAFC-62D*, whereas maximum sequence read is 1000 for *DAFC-34B*, *DAFC-66D*, and *DAFC-7F*. ChIP–chip experiments depict the log<sub>2</sub> ratios of immunoprecipitated DNA compared with input DNA, which contains increased DNA copy levels at the amplicons. (E,F) The chorion amplicons show the greatest increase in DNA copy number, indicating the most rounds of origin firing, which correspond to the largest zones of ORC enrichment and overlap of these regions with tetra-acetylated H4 enrichment. Bars below the ORC and tetra-acetylated H4 values represent significant regions of enrichment, as calculated by the MA2C program, with a *P*-value cutoff of 0.002 and a bandwidth of 5000 bp for ORC, and *P*-value cutoff of 0.00001 and a bandwidth of 500 bp for tetra-acetylated H4.

The four nonchorion amplicons show distinct transcript profiles. Like the chorion amplicons, *DAFC-34B* shows high expression levels for at least two genes in the amplified region (Fig. 2C). The vitelline membrane gene *Vm34Ca* had the highest expression value in our RNA-seq data set, although, unlike the chorion amplicons, there is not a cluster of highly expressed genes in the 10-kb most amplified region of *DAFC-34B*. In contrast, *DAFC-30B* and *DAFC-62D* contain genes that are only moderately expressed by RNA-seq (Fig. 2B,D), which may reflect the very narrow developmental window in which they are expressed. Claycomb et al. (2004) showed that genes within *DAFC-30B* and *DAFC-62D*—such as *CG13113* and *yellow-g2*, respectively—display high expression levels across all follicle cells in only one or two stages of egg chamber development, unlike the chorion genes, which show more sustained high expression levels throughout several stages (Parks et al. 1986; Parks and Spradling 1987). Additionally, *DAFC-30B* contains genes at the edge of the amplified gradient that are more highly expressed than genes in the central region, a unique feature among the six amplicons. For *DAFC-22B*, we found that *CG7337* is expressed at low levels (Fig. 2A). It is possible that gene amplification is necessary for even these low expression levels in follicle cells. However, *CG7337* is expressed at comparable and even higher levels in other tissues where the *22B* region is not amplified (Chintapalli et al. 2007). Another possibility is that amplification of this region does not augment gene expression. Instead, amplification may be the indirect consequence of transcription, which results in chromatin modifications or displacement of nucleosomes, permitting the region to serve as a replication origin (Eaton et al. 2010).

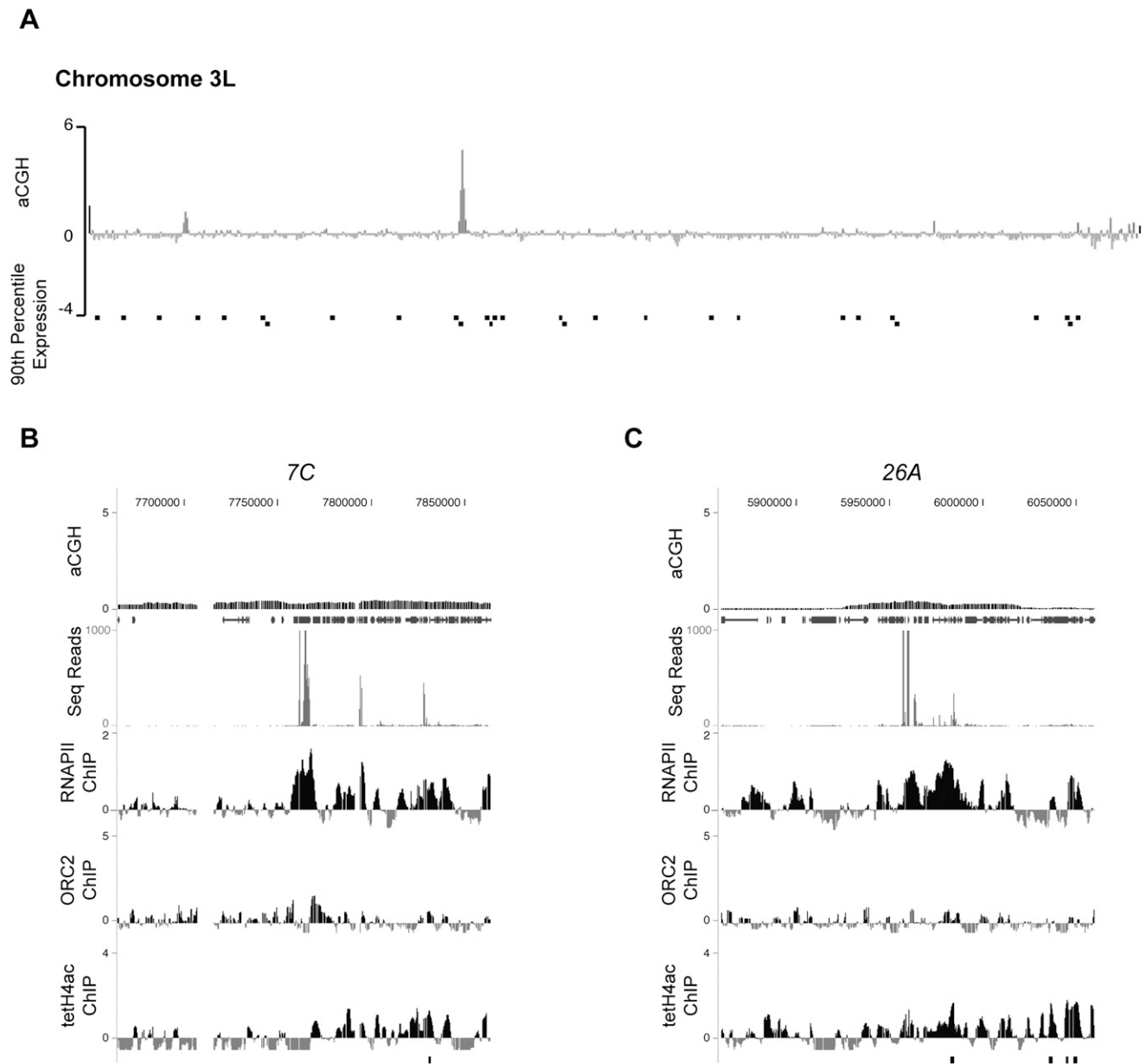
If high levels of transcription can create a chromatin domain that promotes gene amplification specifically in follicle cells, we might expect other genomic regions with highly expressed genes to become amplified. However, this does not appear to be the case, as there are hundreds of highly expressed genes and gene clusters, yet only six amplified regions (Fig. 3A). 16C follicle cells contain cells from stage 9 egg chambers, which have exited the endocycle but have not yet begun gene amplification. To address whether highly expressed genes measured by RNA-seq reflect robust stability of mRNAs transcribed at or prior to stage 9 rather than active transcription during gene amplification stages, we determined the localization of RNA polymerase II (RNAPII) by chromatin immunoprecipitation (ChIP) followed by hybridization to a microarray (ChIP-chip). We examined RNAPII localization in stage 10 egg chambers with the nurse cells manually removed, rather than in a mixed population of follicle cells, to assess whether RNAPII was active or poised for transcription at the precise onset of gene amplification. Although many genomic regions, including all six amplicons, show RNAPII localization at stage 10 and expression by RNA-seq (Fig. 2), only a limited number become amplified, demonstrating that active transcription is not sufficient for origin activation and gene amplification in follicle cells. Furthermore, these data reveal that gene amplification is not the exclusive means by which genes become highly expressed in follicle cells.

### *ORC binding in amplicons localizes in a domain to the most amplified region*

Although we found that abundant and active transcription is not sufficient for gene amplification, it was nevertheless possible that transcription influences the localization of ORC. For example, recent studies in *Drosophila* cell culture have shown that ORC associates predominantly with the promoters of active genes (MacAlpine et al. 2004, 2010). To investigate a possible correlation between transcription and ORC localization in follicle cells, as well as determine the localization of ORC with respect to the 100-kb amplified gradients of the six amplicons, we performed ChIP-chip experiments using an antibody specific to ORC2. We used hand-sorted stage 10 egg chambers to obtain a stage-specific profile of ORC localization at the start of gene amplification (stage 10) and examined ORC binding in regions showing high expression and RNAPII localization. The *dec-1* locus and vitelline membrane gene cluster at *26A* are highly expressed in follicle cells (Fig. 3B,C). These regions do not exhibit significant levels of ORC binding or an increase in DNA copy number, emphasizing that high expression levels can be, and are most frequently, achieved in follicle cells without gene amplification.

We computationally identified the top scoring regions of ORC enrichment using the MA2C program (Song et al. 2007) and found that, although the amplicons were among the top scoring of the 99 regions, ORC localized to other discrete genomic regions that were not amplified (Fig. 4). ORC localization alone is not sufficient for gene amplification in follicle cells. Other regulatory mechanisms are likely necessary to establish full helicase loading and activate replication initiation. Alternatively, ORC localization at these nonamplifying sites may play a role independent of replication initiation, such as cohesin loading or maintenance of heterochromatic silencing (Pak et al. 1997; Huang et al. 1998; Takahashi et al. 2004). Additionally, although roughly two-thirds of the ORC-bound sites overlapped with transcription units (65 of 99), only one-tenth corresponded to regions that were expressed at a reads per kilobase per million (RPKM) of >3 (10 of 99), marking a notable difference from *Drosophila* cell culture, where ORC-binding sites are associated predominantly with active promoters (MacAlpine et al. 2010).

The six follicle cell amplicons show differences in the number of rounds of replication initiation. To ask whether differences in ORC localization might be responsible for differences in amplification levels, we examined the ORC-binding profile at the six amplicons. We found that ORC localizes to all follicle cell amplicons in a region centered at the peak of amplification, corresponding to a large zone ~10–30 kb in length, depending on the amplicon (Fig. 2; Table 1). Previous studies by ChIP-qPCR suggested that ORC was bound to specific sites in the *DAFC-62D* and *DAFC-66D* amplicons (Austin et al. 1999; Xie and Orr-Weaver 2008); thus, it was striking to find ORC localized to broad domains. As the status of ORC binding for *DAFC-7F* and *DAFC-30B* was previously unknown, these studies contribute four new ORC-bound amplification origins to the catalog of metazoan replication origins. The

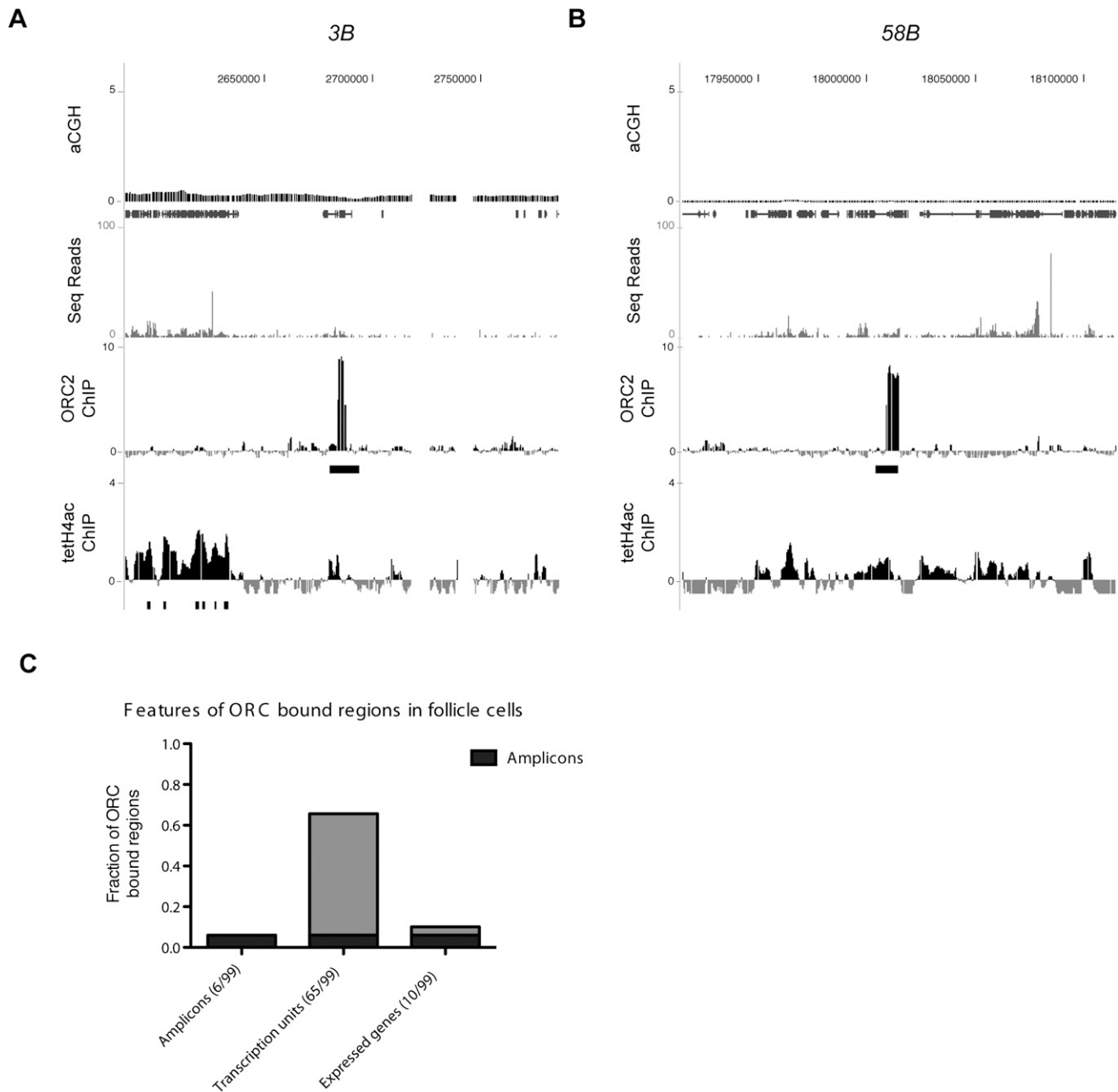


**Figure 3.** Gene amplification is not required for high follicle cell gene expression. (A) Gene expression by RNA-seq was analyzed in 100-kb windows, and the 90th percentile of highly expressed 100-kb regions was plotted and is designated by black boxes. Chromosome 3L is depicted in A. Many genomic regions show high expression but are not amplified. (B,C) Two-hundred-kilobase regions of two nonamplified genomic loci with highly expressed genes. Sequence reads from RNA-seq of purified follicle cells as well as localization of RNAPII, ORC, and tetra-acetylated H4 by ChIP-chip from stage 10 egg chambers are shown as in Figure 2. 7C contains the *defective chorion 1* gene (*dec-1*), and 26A contains a cluster of genes encoding components of the vitelline membrane. Bars below the tetra-acetylated H4 values represent significant regions of enrichment, as calculated by the MA2C program, using a *P*-value cutoff of 0.00001 and a bandwidth of 500 bp.

chorion amplicons, which reach the highest copy number levels, also show the largest domain of ORC localization. However, the relationship between the width of ORC binding and magnitude of gene amplification is not precisely correlated for the other four amplicons.

In addition to ORC localization at the onset of gene amplification (stage 10), we also examined ORC-binding profiles by ChIP-chip at the six amplicons in stages

immediately prior (stage 9 and early stage 10) to examine potential dynamics of ORC localization that could not be identified using 16C DNA alone. The ORC-binding profiles of the four nonchorion amplicons were remarkably similar between the two samples (Supplemental Fig. 1). In contrast, ORC enrichment dramatically increased from preamplification to amplification stages at the two chorion amplicons, suggesting that an active mechanism promotes ORC



**Figure 4.** ORC localizes to genomic regions that do not become amplified. (A,B) Two-hundred-kilobase regions of two nonamplified, ORC-bound regions. Sequence reads from RNA-seq of purified follicle cells as well as localization of ORC and tetra-acetylated H4 by ChIP-chip from stage 10 egg chambers are shown as in Figure 2. ORC localizes to the transcription unit of *white* in genomic region 3B and *mei-S332* in 58B, although neither gene is highly expressed (RPKM, <5). (C) Features of ORC-bound regions showing the fraction of the 99 ORC-bound regions that overlap with amplified domains, transcription units, and expressed genes (RPKM, >3).

localization at the start of amplification for the amplicons that will achieve the highest copy number increases.

#### *H4 acetylation corresponds to the magnitude of gene amplification*

To investigate further the molecular determinants of differential ORC localization and origin activation among the six amplicons, we examined histone modifications on

a genome-wide scale. Two independent studies have found enrichment of hyperacetylated H3 and H4 at follicle cell amplicons (Aggarwal and Calvi 2004; Hartl et al. 2007). Loss-of-function mutant clones of the histone deacetylase *Rpd3* resulted in increased acetylation levels and showed inappropriate genomic replication in amplification stage egg chambers. Furthermore, follicle cell amplification using a reporter construct of *DAFC-66D* could be inhibited by tethering *Rpd3* to the region (Aggarwal and Calvi 2004).

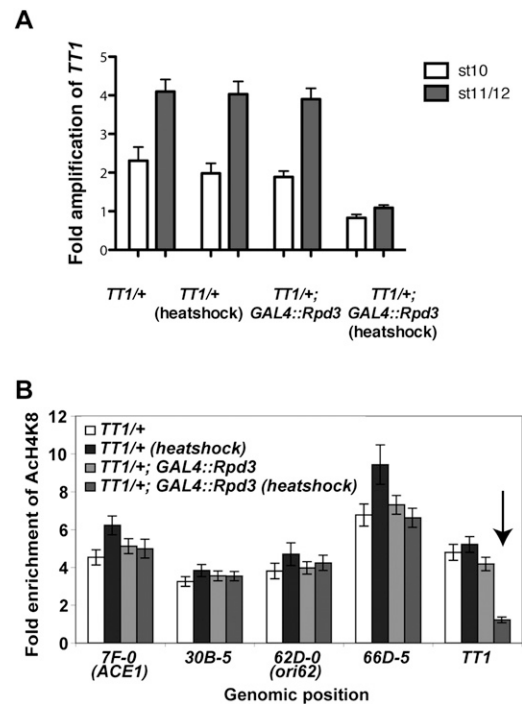
We performed ChIP–chip on stage 10 egg chambers using an antibody against tetra-acetylated H4, which recognizes acetylated Lys 5, Lys 8, Lys 12, and Lys 16. Enrichment of tetra-acetylated H4 was found at significant levels only at the two chorion amplicons (Fig. 2). Notably, tetra-acetylated H4 was enriched across many sites in the genome but was not sufficient for ORC localization or gene amplification. Because this antibody recognizes all four acetylated lysines, we used antibodies specific for single residues to test their specificity at the amplicons. Whereas acetylated H4K5, H4K12, and H4K16 antibodies showed only modest enrichment at the amplicons by ChIP–qPCR, acetylated H4K8 was most significantly enriched at the chorion amplicons (Supplemental Fig. 2; data not shown).

Given that there are significant peaks of H4 acetylation only at the chorion amplicons, we wanted to test for functionality of this modification. To analyze whether H4 acetylation levels could influence amplification levels, we used the amplification reporter system developed by the Calvi laboratory (Aggarwal and Calvi 2004). The reporter designated *TT1* contains the 3.8-kb minimal origin with *ACE3* and *Ori $\beta$*  from *DAFC-66D* next to UAS, which can be bound by GAL4 and GAL4 DNA-binding domain (DBD) fusion proteins. We expressed the Rpd3 fusion protein via a heat-shock promoter to determine quantitatively the effect of tethering the histone deacetylase to *TT1* on DNA copy number and enrichment of H4K8. We used qPCR with transgene-specific primers to get a more accurate measure of DNA copy number changes than previous analyses, which used quantification of immunofluorescence images and Southern blots. We specifically examined H4K8 acetylation because this mark showed the highest levels of enrichment at the amplicons by ChIP–qPCR, and we reasoned that we could more confidently observe differences in acetylation levels due to experimental treatment.

Following 1-h heat-shock induction of a GAL4<sup>DBD</sup>::Rpd3 fusion, amplification was completely abolished in stage 10 as well as pooled egg chambers of stages 11 and 12, without affecting the endogenous amplicons (Fig. 5A). We examined stage 10 egg chambers specifically for acetylated H4K8 after induction of the GAL4<sup>DBD</sup>::Rpd3 fusion and found that this histone mark was significantly reduced at *TT1* but was essentially unchanged at the endogenous amplicons (Fig. 5B). Thus, these results indicate that acetylation of H4K8 is necessary for amplification of the *TT1* transgene, an effect that may be specific to *TT1* and the chorion amplicons because the other four amplicons amplify without showing significant enrichment of H4 acetylation.

#### *DAFC-22B* exhibits strain-specific amplification

In the process of characterizing the amplification properties of *DAFC-22B*, we examined gene amplification at this locus in a number of genetic backgrounds. Surprisingly, we found that *DAFC-22B* displays strain-specific amplification. By aCGH, we found that, even in two closely related strains where all five other amplicons are common, *DAFC-22B* amplifies in *OrR<sup>TOW</sup>* (an Oregon R strain maintained in our laboratory) and is not ampli-

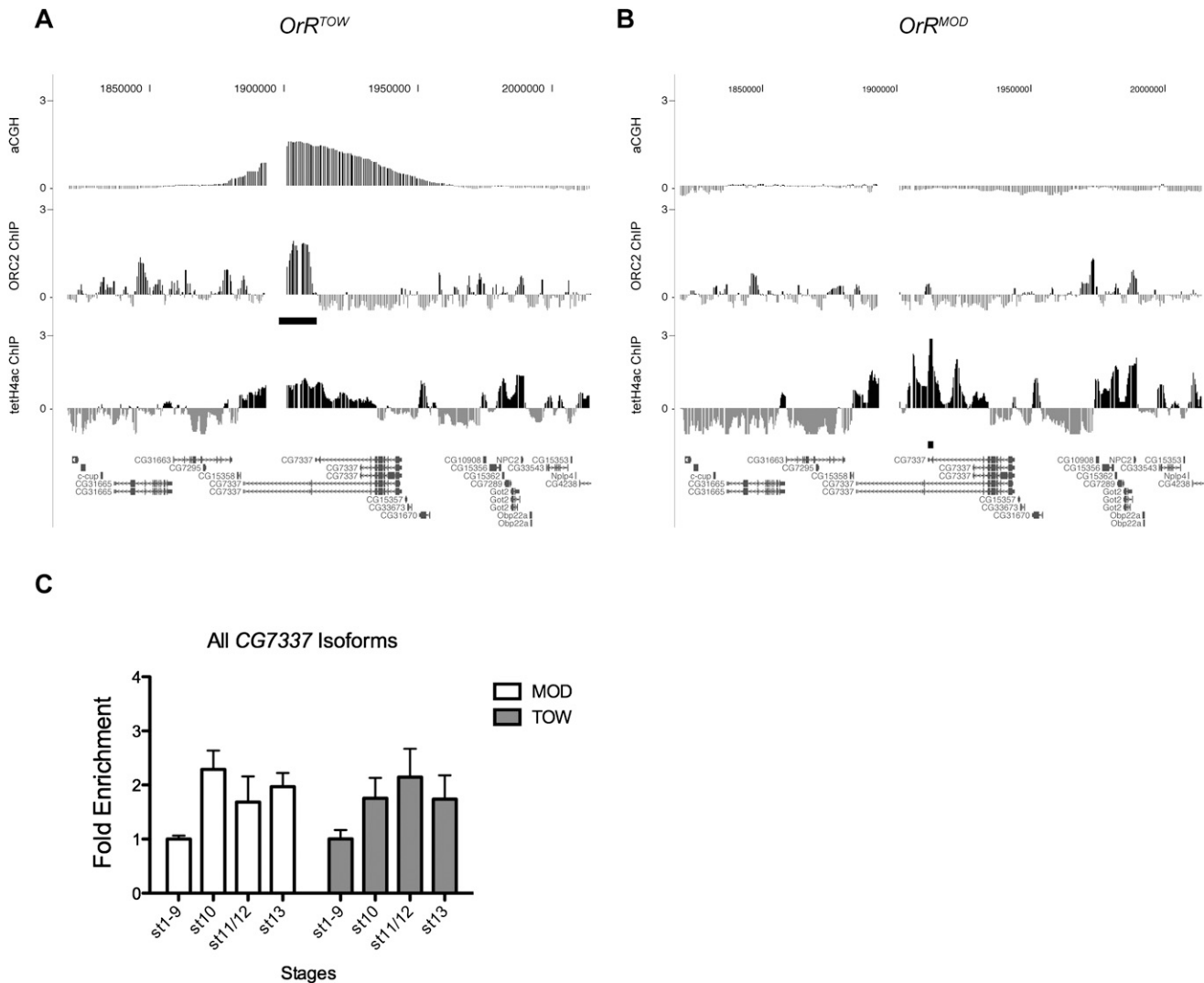


**Figure 5.** Tethering Rpd3 to the amplification reporter *TT1* eliminates its amplification and reduces H4K8 acetylation. (A) qPCR with transgene-specific primers was used to determine copy number of the *TT1* transgene in different genetic backgrounds and experimental conditions. *TT1* amplification is repressed upon *Rpd3* expression (heat shock) and tethering to *TT1* at stage 10 and pooled stages 11 and 12 egg chamber samples. (B) Levels of H4K8 acetylation at *TT1* and representative primers from endogenous amplicons in stage 10 by ChIP–qPCR. The arrow points to reduction in acetylated H4K8 levels upon Rpd3 expression and tethering to *TT1*.

fied in *OrR<sup>MOD</sup>* (an Oregon R strain adopted for the modENCODE project) (Fig. 6A,B). This finding enabled us to test the link between amplification, transcription, and histone acetylation, as well as define the requirements for ORC localization at this locus.

To characterize the amplification properties of *DAFC-22B* more closely, we examined the timing of origin activation as well as expression of the gene in the amplified locus. We determined the stage-specific replication profile of *DAFC-22B* by hand-sorting *OrR<sup>TOW</sup>* egg chambers to isolate genomic DNA and performing qPCR quantification of DNA copy levels. We found that replication initiation increased from stages 10B through 13, resulting in approximately fourfold amplification by stage 13 (Supplemental Fig. 3). Because *DAFC-22B* coincides with a gene of which the 5' end of one isoform (D-isoform) is in the peak of amplification, we examined stage-specific gene expression of *CG7337* in both strains for possible differences that might explain the presence or absence of amplification at the region. Intriguingly, despite the presence of four times the amount of DNA template in *OrR<sup>TOW</sup>* compared with *OrR<sup>MOD</sup>*, we found that there was no difference in overall *CG7337* expression levels between the amplifying and





**Figure 6.** *DAFC-22B* exhibits strain-specific amplification that correlates with the ability of the region to bind ORC. (A,B) Follicle cell DNA copy levels as well as ORC and tetra-acetylated H4 by ChIP-chip from stage 10 egg chambers are shown as in Figure 2 for two different strains: *OrR<sup>TOW</sup>* and *OrR<sup>MOD</sup>*. *DAFC-22B* is not amplified in the *OrR<sup>MOD</sup>* strain, and this region does not show ORC localization. Tetra-acetylated H4 is observed in similar patterns in both amplifying and nonamplifying strains. (C) cDNA was generated from different egg chamber stages and analyzed by qPCR. Fold enrichment is relative to *Rps17* and normalized to the stage 1–9 sample. There is no difference in expression levels for the gene in *DAFC-22B* (*CG7337*) for any developmental stage.

nonamplifying strains for any developmental stage of dissected egg chambers (Fig. 6C; Supplemental Fig. 4).

To elucidate the mechanism of differential gene amplification at *DAFC-22B*, we performed ChIP–chip with the ORC2 antibody on the strain that does not amplify *DAFC-22B* in stage 10 egg chambers and found that ORC is absent at *DAFC-22B* despite the same ORC localization for the other five amplicons (Fig. 6A,B). Because the determinants of ORC binding are poorly understood in metazoans, we sought to uncover the difference between the two strains responsible for the difference in amplification to gain a better understanding of how ORC binds to specific DNA regions to promote replication initiation. We performed ChIP–chip on stage 10 egg chambers to examine H4 tetra-acetylation at *DAFC-22B* and observed similar patterns between the two strains (Fig. 6A,B). Thus, although

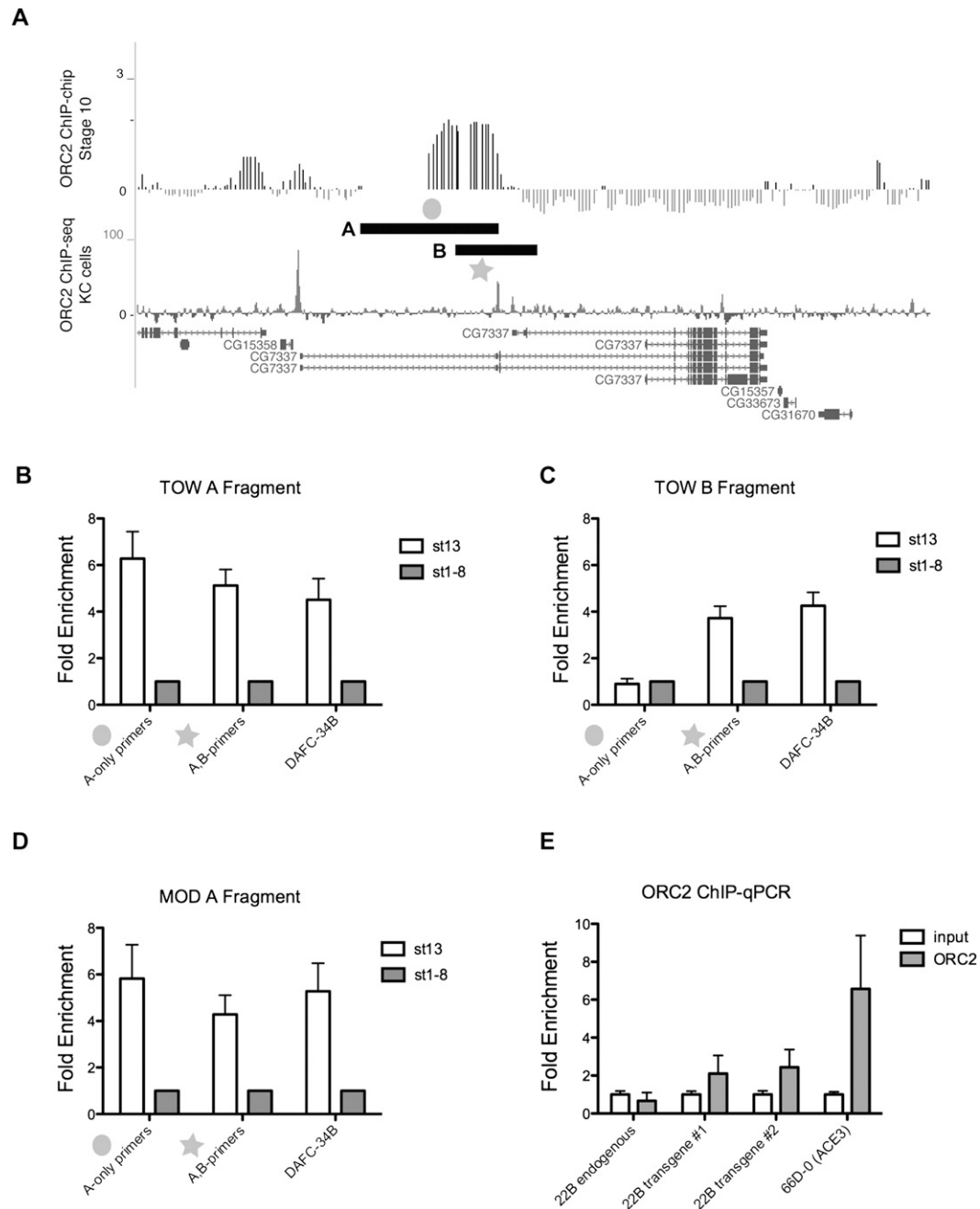
levels of H4 acetylation influence the magnitude of gene amplification at the chorion amplicons, acetylated H4 alone is not sufficient for replication initiation and does not appear to be responsible for differential *DAFC-22B* amplification.

#### *Repression of origin activation is responsible for strain-specific differences in DAFC-22B amplification*

In *Drosophila*, *cis* requirements for amplification can be determined using P-element-mediated transformation to examine amplification of a test sequence at an ectopic site. Because of this capability, we generated transposons containing the 12-kb ORC-binding zone from *OrR<sup>TOW</sup>* as well as the equivalent region from *OrR<sup>MOD</sup>* by PCR (fragment A in Fig. 7A) to test amplification. In addition, *DAFC-22B*

is the only amplified region that, in cell culture, also shows ORC binding, although the peaks of ORC localization in follicle cells and cell culture do not overlap. Thus, we also

amplified the 10-kb sequence containing the cell culture ORC-binding sites from *OrR<sup>TOW</sup>* to test for amplification (fragment B in Fig. 7A). We flanked the *22B* sequences with



**Figure 7.** Genetic analysis of *cis* control elements for differential *DAFC-22B* amplification reveals distant repression of origin activation. (A) The *DAFC-22B* regions for P-element-mediated transformation and testing sufficiency for amplification are shown in black bars. Fragment A was PCR-amplified from *OrR<sup>TOW</sup>* (amplifying) and *OrR<sup>MOD</sup>* (nonamplifying) flies. Fragment B was amplified from *OrR<sup>TOW</sup>* flies. These sequences integrated into random sites in the genome of a strain that does not amplify the endogenous *22B* locus. ORC2 ChIP-chip data are shown for *OrR<sup>TOW</sup>*. ORC2 ChIP-seq data from KC cells are also shown (scale is sequence tag density). The circle represents the position of qPCR primers that detect amplification of fragment A. The star represents the position of qPCR primers that detect amplification of fragment A or B. (B) The TOW A fragment shows amplification using both sets of primers. The stage 1–8 egg chambers serve as a control, because amplification does not begin until stage 10. (C) The TOW B fragment shows amplification using only the star set of primers because the endogenous *22B* sequence recognized by the circle primers does not amplify. (D) The MOD A fragment shows amplification using both sets of primers. (E) ORC2 ChIP-qPCR was performed on stage 10 egg chambers from transgenic flies carrying the MOD A fragment. The primers designated *22B* endogenous do not amplify the MOD A transgene.

Suppressor of Hairy wing [Su(HW)]-binding sites to minimize position effect variability (Lu and Tower 1997). Sequencing the 22B transposon constructs revealed that no major rearrangements were present between the two strains, and no cytological rearrangements were observed in polytene squashes of chromosome 2 heterozygous for the two strain backgrounds that could account for repression in *OrR<sup>MOD</sup>* (data not shown). By restriction mapping, we confirmed that the transposon constructs from each strain contained unrearranged DNA relative to the genomic strain. The *DAFC-22B* transposons were injected into a genetic background, *w<sup>1118</sup>*, where the endogenous *DAFC-22B* locus does not amplify. Thus, in addition to using transposon-specific primers to detect amplification, the *DAFC-22B* sequence that is on the transposon can be used to test amplification, whereas the *DAFC-22B* sequence that is not on the transposon serves as a control for no endogenous *DAFC-22B* amplification.

When we tested for *DAFC-22B* amplification in transgenic lines where the test sequence had integrated randomly into the genome, we found that the fragments bearing the egg chamber and cell culture ORC-binding sequences derived from *OrR<sup>TOW</sup>* were both sufficient to support amplification at an ectopic site (Fig. 7B,C). As the two sequences are partially overlapping, these results indicate that, minimally, a 5-kb sequence is sufficient for amplification. Remarkably, we found that, in the two independent transformants recovered, the egg chamber ORC-binding sequence derived from the *OrR<sup>MOD</sup>* strain, which does not amplify *DAFC-22B* at the endogenous locus, was also capable of supporting amplification to levels comparable with the transgene derived from *OrR<sup>TOW</sup>* (Fig. 7D). To determine whether there is a peak of amplification within the 12-kb transgene derived from the *OrR<sup>MOD</sup>* strain, we assessed DNA copy levels by qPCR using primers spaced ~2 kb apart. We observed a modest increase at the center of the transgene compared with the distal ends (Supplemental Fig. 5). Importantly, we observed an increase in DNA copy levels from stage 10B to stage 13 egg chambers, recapitulating the developmental program of replication initiation at the endogenous *22B* locus in *OrR<sup>TOW</sup>*. In addition, we examined ORC localization at the transgene derived from the *OrR<sup>MOD</sup>* strain and found levels of ORC enrichment consistent with that observed by ChIP-chip (Fig. 7E).

Thus, we uncovered an example of a sequence that is competent to function as a replication origin but is typically repressed for ORC localization and origin activation at its endogenous location. Because the transposon boundaries are ~5 kb away from the peak of ORC binding, repression of origin activation appears to be acting from a distance, and not directly at the central region of ORC binding. This repression can be reset when the DNA sequence is placed in new genomic contexts.

## Discussion

We used *Drosophila* follicle cell gene amplification as a model system to study the molecular properties of metazoan replication origins. Although cell culture stud-

ies have shown a strong relationship between actively transcribed promoters and replication origins, we found that high expression output and active transcription are not sufficient for ORC localization in the context of follicle cell amplification, an *in vivo* replication model system. Furthermore, ORC localization by itself is not sufficient for origin activation in amplifying follicle cells.

Our studies unexpectedly uncovered an example in which chromatin or DNA elements repress ORC binding and replication initiation at a distance. *DAFC-22B* provides a powerful opportunity to study the determinants of ORC binding in metazoans because the region displays strain-specific amplification that is correlated with ORC localization. By testing the amplification status of sequences generated from strains that do and do not amplify *DAFC-22B* at ectopic sites in transgenic flies, we found that sequences from both strains are capable of amplification. Thus, sequence differences within 5 kb of the ORC-binding zone are likely not responsible for differential ORC localization and origin activation. Instead, in strains that do not amplify *DAFC-22B* at the endogenous site, the replication origin-competent sequence is repressed, possibly by an epigenetic mechanism, from a distance of at least 5 kb away. The repression is not due to a *trans*-acting factor because *DAFC-22B* is not amplified at the endogenous locus in the genetic background in which the *DAFC-22B* transposons were introduced and found to amplify. We do not think that placing DNA in the context of Su(HW) insulator elements artificially makes any sequence replication-competent. Molecular analysis of *DAFC-62D* using the same transposon system demonstrated that various deletions in a 10-kb sequence, which was sufficient for ectopic amplification, eliminated origin activation (Xie and Orr-Weaver 2008).

A striking feature of the ORC-binding profiles for the six amplicons is that ORC localizes over broad domains of 10–30 kb. This contrasts with the localization pattern seen in diploid cell culture and with the other sites of ORC localization in the follicle cells (MacAlpine et al. 2010). ORC binding in a zone in the amplicons also has been observed in the laboratories of Michael Botchan and Brian Calvi (pers. comm.). Notably, the chorion amplicons (*DAFC-7F* and *DAFC-66D*) show the largest domains of ORC binding, and they also undergo the most rounds of initiation. How might this molecular snapshot of ORC binding correspond to increased origin activation at the amplicons? The large domains of ORC binding may reflect ORC coating the DNA strands over tens of kilobases, and this could increase the probability of origin activation. Alternatively, because follicle cells are polyploid and have 16 copies of each DNA duplex when amplification initiates, ORC could be bound at slightly different sites within the amplicons on each of the 16 duplexes. Likewise, there may be cell-to-cell variability in the precise position of ORC localization, resulting in an apparent zone of binding. Differences in the levels of amplification could reflect the number of DNA helices that initiated DNA replication.

A distinct interpretation is that the increased ORC zone at the amplicons is a consequence of initiation rather than

a cause. If more ORC molecules bind to the newly replicated DNA strands at the amplicons, this would explain the higher levels of ORC enrichment as well as the additional rounds of origin activation at the amplicons—in a sense, a feed-forward mechanism for replication. An active process of continuous ORC recruitment is an attractive model, particularly for the chorion amplicons, as there is a marked increase in ORC binding from preamplification to amplification stages at the chorion amplicons that is not observed at the other four loci (Supplemental Fig. 1).

We investigated the relationship between histone acetylation and gene amplification in *Drosophila* follicle cells and found that acetylation of H4—and, specifically, H4K8—quantitatively correlated with levels of gene amplification, as demonstrated by amplification of a reporter upon Rpd3 tethering. At the endogenous amplicons, only the two chorion amplicons show overlap of the ORC-binding region with significant enrichment of H4 acetylation. Thus, H4 acetylation is not sufficient or necessary for ORC localization or origin firing. Instead, H4 acetylation levels may be required specifically for chorion gene amplification or may influence the number of rounds of origin firing in the context of regions undergoing replication initiation. Recent studies in Brian Calvi's laboratory also found that H4 acetylation correlates with the level of origin firing at the follicle cell amplicons. This work shows that multiple acetylated forms of H3 and H4 are enriched at amplicons, and these marks are regulated in a dynamic manner (B Calvi, pers. comm.).

One of the most consistent findings among recent genome-wide origin mapping studies is the significant number of origins that correspond to gene regions and, in particular, the transcription start sites of active genes (Cadoret et al. 2008; Sequeira-Mendes et al. 2009; MacAlpine et al. 2010). Additionally, a high-resolution study of replication timing in multiple cell lines using next-generation sequencing methods reported that genes expressed solely in one cell type were early-replicating exclusively in that cell type, a finding also supported by analysis of individual model replicons, suggesting a causal effect of transcription on early replication (Hansen et al. 2010). Although there are just six follicle cell amplicons to examine, compared with thousands of origins activated in the canonical S phase, all six ORC-binding amplification origins correspond to gene regions that show active transcription. However, there are hundreds of genes and gene clusters expressed at higher levels in follicle cells than even the moderately expressed genes of *DAFC-30B* and *DAFC-62D*, and these regions do not become amplified. Furthermore, highly expressed gene regions do not show a greater enrichment of ORC localization. Thus, active transcription is not sufficient for origin activation or ORC localization in amplifying follicle cells.

The identification of all of the amplified regions in *Drosophila* follicle cells, combined with genome-wide assessment of transcript levels, reveals that the simple model of gene amplification being a developmental strategy to promote high levels of gene expression, as seems to be the case for some genes within the two chorion amplicons, may not be an absolute one. When examined by RNA-seq

of 16C follicle cells, the genes in *DAFC-30B* and *DAFC-62D* are expressed to moderate levels, and the genes in *DAFC-22B* are expressed to low levels. For the chorion genes and genes in *DAFC-62D* and *DAFC-30B*, the extra gene copies are necessary for full expression levels, as gene expression is reduced in amplification mutants (Landis et al. 1997; Schwed et al. 2002; Claycomb et al. 2004). As a gene regulatory strategy, gene amplification is absolutely essential, because these mutants lay inviable eggs with thin eggshells. However, amplification appears to have no effect on overall *CG7337* expression levels between the amplifying and nonamplifying strains despite a fourfold difference in DNA template. Gene amplification and DNA copy number variations are significant features of cancer genomes. Although there is a strong correlation between these genetic alterations and cancer progression, our data highlight the importance of considering that increased DNA copy number may not always be coupled to enhanced gene expression.

Our studies have generated important insights regarding gene amplification as a developmental strategy and metazoan replication model. First, gene amplification is not always a driving force for abundant transcription. Our results of highly expressed genes that are not amplified demonstrate that promoter strength is the primary strategy for achieving high expression levels, even in a cell that is capable of gene amplification. Additionally, gene amplification does not uniformly lead to abundant expression levels. Gene amplification can augment gene expression, as is the case for the chorion genes, but it can also have no effect on expression, as for *CG7337* at *DAFC-22B*. Second, we discovered an example of distant repression of ORC binding and origin activation. This finding validates the importance of replication origin discovery combined with close molecular analysis of individual origins. The strain specificity of *DAFC-22B* amplification likely reflects the widespread plasticity of origin specification and activation in metazoans, and will be a useful model replicon for future studies of poorly understood areas such as the regulation of dormant origins and cell type specificity of replication origin usage.

## Materials and methods

### CGH

Two strains were used: *OrR<sup>TOW</sup>* and *OrR<sup>MOD</sup>*. 16C nuclei were isolated by FACS from *OrR<sup>TOW</sup>* and *OrR<sup>MOD</sup>* fattened females as described previously (Lilly and Spradling 1996). Genomic DNA was prepared using the DNeasy Blood and Tissue kit (Qiagen), digested with AluI and RsaI, and labeled using Invitrogen's BioPrime labeling kit. Slides were hybridized to custom Agilent tiling arrays with probes every 600 or 400 bp and washed as per Agilent's recommendations. Array intensities were median-normalized across channels and smoothed by genomic windows of 10 kb using the Ringo package in R (Toedling et al. 2007).

### Follicle cell isolation

16C follicle cells were isolated as described previously (Bryant et al. 1999; Cayirlioglu et al. 2003) with the following modifications. Ovaries from ~100 females were dissected in Grace's

unsupplemented medium containing 10  $\mu\text{g}/\text{mL}$  Hoechst 33342. After three washes in PBS to remove residual calcium, ovaries were trypsinized (0.5% final; Gibco) in PBS supplemented with 10  $\mu\text{g}/\text{mL}$  Hoechst 33342 and filtered through 100- $\mu\text{m}$  mesh into Grace's unsupplemented medium with 10  $\mu\text{g}/\text{mL}$  Hoechst 33342. Four rounds of trypsinization were necessary to completely dissociate all follicle cells. FACS was used to isolate follicle cells based on ploidy using a MoFlo flow cytometer. Two-hundred-fifty-thousand to 750,000 16C follicle cells were routinely recovered from 100 fattened females.

#### RNA isolation and RNA-seq

Total RNA was extracted from purified 16C follicle cells (from *OrR<sup>TOW</sup>*) using TRIzol reagent (Invitrogen) according to the manufacturer's recommendations. One-hundred nanograms of RNA was used to generate an RNA-seq library using the mRNA-seq Sample Preparation kit from Illumina with the exception that no selection was used to isolate polyA RNA. Rather, RNA-seq libraries generated from total RNA were subjected to Duplex-Specific thermostable nuclease enzyme (DSN) normalization according to Illumina recommendation. Briefly, 100 ng of the RNA-seq library was treated with DSN after heat denaturation and subsequent renaturation to remove highly abundant rRNA and tRNA-derived transcripts. Two biological replicates with a correlation coefficient of 0.89 were used in the analysis. Expression is not normalized to DNA copy number.

RNA-seq reads were aligned to BDGP release 5/DM3 using ELAND and counts were normalized by the mRNA length to obtain the final RPKM values as in Nordman et al. (2011).

For the analysis of *CG7337* expression, purified follicle cells were isolated using a protocol modified from Bryant et al. (1999). Two-hundred stage 13 egg chambers were dissected in ice-cold Schneider's medium supplemented with 10% FBS. Tissue was digested with 0.9 mL of 0.25% trypsin/EDTA and 0.1 mL of 50 mg/mL collagenase for 12 min at room temperature. The supernatant was strained through a 40- $\mu\text{m}$  mesh and spun at 1000g for 7 min in the cold. TRIzol was added to the pellet for RNA isolation. RNA samples were reverse-transcribed with AMV reverse transcriptase (Promega). For cDNA analysis, *Rps17* was used as the endogenous control.

#### ChIP

ChIP was performed with the following antibodies: RNAPII (Upstate Biotechnology, clone CTD4H8), which recognizes both phosphorylated and nonphosphorylated forms; ORC2 (Steve Bell); tetra-acetylated H4 (Active Motif, 39179); acetylated H4K5 (Upstate Biotechnology, 07-327); acetylated H4K8 (Upstate Biotechnology, 07-328); acetylated H4K12 (Upstate Biotechnology, 07-595); and acetylated H4K16 (Upstate Biotechnology, 07-329). The commercial histone modification antibodies used were ones validated by the modENCODE consortium for specificity. ChIP-qPCR was performed on 300 staged egg chambers per experiment as described (Xie and Orr-Weaver 2008). Likewise, for ChIP-chip, we isolated stage 10 egg chambers by hand when follicle cell gene amplification commences, and performed ChIP-chip. The starting number of egg chambers (~1200) produced enough material so that no amplification had to be performed on the samples. For ChIP-chip with the RNAPII antibody, stage 10 egg chambers were further dissected in half to remove the nurse cell population. All ChIP experiments were compared with input DNA. For hybridization to arrays, DNA was labeled using Invitrogen's BioPrime Total for Agilent aCGH labeling kit.

Peak calling for ORC ChIP-chip and TetraH4 ChIP-chip was done using the python MA2C package (Song et al. 2007). ORC

peaks were identified using a *P*-value cutoff of 0.002 and a bandwidth of 5000 bp. TetraH4 peaks were identified using a *P*-value cutoff of 0.00001 and a bandwidth of 500 bp.

#### Reduction of histone acetylation on DAFC-66D transgenes

Transgenic lines carrying the *TT1* and the *hsp::GAL4<sup>DBD</sup>::Rpd3* transposons were a gift from Brian Calvi (Aggarwal and Calvi 2004). Flies were crossed to introduce the two transposons into the same line; siblings that contained only *TT1* (*TT1/+*) were kept as controls. One hour of heat shock at 37°C was used to overexpress the GAL4 fusion protein, followed by a 6-h recovery period. DNA copy levels were assessed for dissected stage 10 as well as pooled stages 11 and 12 egg chambers to ensure that expression of the fusion protein occurred prior to or during replication initiation of the reporter at stage 10.

#### qPCR

Genomic DNA was isolated from staged egg chambers and quantified using absolute quantitative PCR as described (Claycomb et al. 2004) or relative quantification as described (Xie and Orr-Weaver 2008). Absolute quantification was used for the *DAFC-22B* replication profile, whereas relative quantification was used for all other experiments.

#### Transgenic fly construction

To test the *cis* requirements for amplification at *DAFC-22B*, we constructed transposons with the 12-kb ORC-binding region (stage 10) from *OrR<sup>MOD</sup>* (MOD4-9, MOD A in text), the 12-kb ORC-binding region (stage 10) from *OrR<sup>TOW</sup>* (TOW4-9, TOW A in text), and the 10-kb ORC-binding region (cell culture) from *OrR<sup>TOW</sup>* (TOW8-10, TOW B in text). The gap in probes in the *DAFC-22B* aCGH amplification profile represents a repeated DNA sequence not present in either *OrR* strain. In the transposons, the ORC-binding fragments were flanked by Su(HW)-binding sites (SHWBS) to control for genomic position-specific integration effects. The sequences were PCR-amplified using exTaq DNA polymerase (Takara) and primers with *AscI* and *AvrII* sites on the forward and reverse sequences, respectively. These products were cloned into a modified PCRA vector with *AscI* and *AvrII* sequences engineered into the multiple cloning site (Lu et al. 2001). These plasmids are called PCRA\_22B\_MOD4-9, PCRA\_22B\_TOW4-9, and PCRA\_22B\_TOW8-10. These plasmids were digested with *NotI* and subjected to a partial *XhoI* digest to transfer the 12-kb or 10-kb inserts to the *NotI* and *XhoI* sites of Big Parent to generate BP\_22B\_MOD4-9, BP\_22B\_TOW4-9, and BP\_22B\_TOW8-10. These constructs were sent to BestGene, Inc., for injection. Two independent transformants were recovered for the MOD A line and one transformant each was recovered for the TOW A and TOW B lines.

#### Data Accession

All data have been deposited at Gene Expression Omnibus: GSE29517, ORC2 ChIP-chip from *Drosophila* egg chambers; GSE29518, RNAPII ChIP-chip from *Drosophila* egg chambers; GSE29520, tetra-acetylated H4 ChIP-chip from *Drosophila* egg chambers; GSE29526, RNA-seq 16C ovarian follicle cells.

#### Acknowledgments

We thank Steve Bell for the ORC2 antibody and Brian Calvi for transgenic flies and for providing unpublished information. George Bell provided bioinformatics advice and Inma Barrasa assisted with

the 100-kb window analysis for expression. We thank Steve Bell, Mary Gehring, and Hannah Blitzblau for critical comments on the manuscript. We thank Michael Botchan for sharing unpublished results. J.N. is a Damon Runyon Cancer Foundation post-doctoral fellow. This work was initiated by support from the Mathers Charitable Foundation, and the aCGH studies were funded by NIH grant 1U01HG004279 as part of the modENCODE project. The project was supported by NIH grant GM57960 and a Research Professor grant from the American Cancer Society to T.O.-W.

## References

- Aggarwal BD, Calvi BR. 2004. Chromatin regulates origin activity in *Drosophila* follicle cells. *Nature* **430**: 372–376.
- Aladjem MI. 2007. Replication in context: dynamic regulation of DNA replication patterns in metazoans. *Nat Rev Genet* **8**: 588–600.
- Austin RJ, Orr-Weaver TL, Bell SP. 1999. *Drosophila* ORC specifically binds to *ACE3*, an origin of DNA replication control element. *Genes Dev* **13**: 2639–2649.
- Bryant Z, Subrahmanyam L, Tworoger M, LaTray L, Liu CR, Li MJ, van den Engh G, Ruohola-Baker H. 1999. Characterization of differentially expressed genes in purified *Drosophila* follicle cells: toward a general strategy for cell type-specific developmental analysis. *Proc Natl Acad Sci* **96**: 5559–5564.
- Cadoret JC, Meisch F, Hassan-Zadeh V, Luyten I, Guillet C, Duret L, Quesneville H, Prioleau MN. 2008. Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc Natl Acad Sci* **105**: 15837–15842.
- Cayirlioglu P, Ward WO, Silver Key SC, Duronio RJ. 2003. Transcriptional repressor functions of *Drosophila* E2F1 and E2F2 cooperate to inhibit genomic DNA synthesis in ovarian follicle cells. *Mol Cell Biol* **23**: 2123–2134.
- Chintapalli VR, Wang J, Dow JA. 2007. Using FlyAtlas to identify better *Drosophila* models of human disease. *Nat Genet* **39**: 715–720.
- Claycomb JM, Orr-Weaver TL. 2005. Developmental gene amplification: insights into DNA replication and gene expression. *Trends Genet* **21**: 149–162.
- Claycomb JM, Benasutti M, Bosco G, Fenger DD, Orr-Weaver TL. 2004. Gene amplification as a developmental strategy: isolation of two developmental amplicons in *Drosophila*. *Dev Cell* **6**: 145–155.
- Cvetic C, Walter JC. 2005. Eukaryotic origins of DNA replication: could you please be more specific? *Semin Cell Dev Biol* **16**: 343–353.
- Danis E, Brodolin K, Menut S, Maiorano D, Girard-Reydet C, Mechali M. 2004. Specification of a DNA replication origin by a transcription complex. *Nat Cell Biol* **6**: 721–730.
- Eaton ML, Galani K, Kang S, Bell SP, MacAlpine DM. 2010. Conserved nucleosome positioning defines replication origins. *Genes Dev* **24**: 748–753.
- Gilbert DM. 2004. In search of the holy replicator. *Nat Rev Mol Cell Biol* **5**: 848–855.
- Gilbert DM. 2010. Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat Rev Genet* **11**: 673–684.
- Griffin-Shea R, Thireos G, Kafatos FC. 1982. Organization of a cluster of four chorion genes in *Drosophila* and its relationship to developmental expression and amplification. *Dev Biol* **91**: 325–336.
- Hamlin JL, Mesner LD, Dijkwel PA. 2010. A winding road to origin discovery. *Chromosome Res* **18**: 45–61.
- Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, Dorschner MO, Gartler SM, Stamatoyannopoulos JA. 2010. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc Natl Acad Sci* **107**: 139–144.
- Hartl T, Boswell C, Orr-Weaver TL, Bosco G. 2007. Developmentally regulated histone modifications in *Drosophila* follicle cells: initiation of gene amplification is associated with histone H3 and H4 hyperacetylation and H1 phosphorylation. *Chromosoma* **116**: 197–214.
- Huang DW, Fanti L, Pak DT, Botchan MR, Pimpinelli S, Kellum R. 1998. Distinct cytoplasmic and nuclear fractions of *Drosophila* heterochromatin protein 1: their phosphorylation levels and associations with origin recognition complex proteins. *J Cell Biol* **142**: 307–318.
- Karnani N, Taylor CM, Malhotra A, Dutta A. 2010. Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol Biol Cell* **21**: 393–404.
- Landis G, Tower J. 1999. The *Drosophila* *chiffon* gene is required for chorion gene amplification, and is related to the yeast Dbf4 regulator of DNA replication and cell cycle. *Development* **126**: 4281–4293.
- Landis G, Kelley R, Spradling AC, Tower J. 1997. The *k43* gene, required for chorion gene amplification and diploid cell chromosome replication, encodes the *Drosophila* homolog of yeast origin recognition complex subunit 2. *Proc Natl Acad Sci* **94**: 3888–3892.
- Lilly MA, Spradling AC. 1996. The *Drosophila* endocycle is controlled by Cyclin E and lacks a checkpoint ensuring S-phase completion. *Genes Dev* **10**: 2514–2526.
- Lu L, Tower J. 1997. A transcriptional insulator element, the su(Hw) binding site, protects a chromosomal DNA replication origin from position effects. *Mol Cell Biol* **17**: 2202–2206.
- Lu L, Zhang H, Tower J. 2001. Functionally distinct, sequence-specific replicator and origin elements are required for *Drosophila* chorion gene amplification. *Genes Dev* **15**: 134–146.
- MacAlpine DM, Rodriguez HK, Bell SP. 2004. Coordination of replication and transcription along a *Drosophila* chromosome. *Genes Dev* **18**: 3094–3105.
- MacAlpine HK, Gordan R, Powell SK, Hartemink AJ, MacAlpine DM. 2010. *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res* **20**: 201–211.
- Mechali M. 2010. Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat Rev Mol Cell Biol* **11**: 728–738.
- Mesner LD, Hamlin JL. 2005. Specific signals at the 3' end of the DHFR gene define one boundary of the downstream origin of replication. *Genes Dev* **19**: 1053–1066.
- Mesner LD, Valsakumar V, Karnani N, Dutta A, Hamlin JL, Bekiranov S. 2011. Bubble-chip analysis of human origin distributions demonstrates a genomic scale significant clustering into zones and significant association with transcription. *Genome Res* **21**: 377–389.
- Meyerson M, Gabriel S, Getz G. 2010. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* **11**: 685–696.
- Mindrinos MN, Scherer LJ, Garcini FJ, Kwan H, Jacobs KA, Petri WH. 1985. Isolation and chromosomal location of putative vitelline membrane genes in *Drosophila melanogaster*. *EMBO J* **4**: 147–153.
- Nordman J, Li S, Eng T, MacAlpine DM, Orr-Weaver TL. 2011. Developmental control of the DNA replication and transcription programs. *Genome Res* **21**: 175–181.
- Pak DT, Pflumm M, Chesnokov I, Huang DW, Kellum R, Marr J, Romanowski P, Botchan MR. 1997. Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. *Cell* **91**: 311–323.

- Park EA, Macalpine DM, Orr-Weaver TL. 2007. *Drosophila* follicle cell amplicons as models for metazoan DNA replication: a *cyclinE* mutant exhibits increased replication fork elongation. *Proc Natl Acad Sci* **104**: 16739–16746.
- Parks S, Spradling A. 1987. Spatially regulated expression of chorion genes during *Drosophila* oogenesis. *Genes Dev* **1**: 497–509.
- Parks S, Wakimoto B, Spradling A. 1986. Replication and expression of an X-linked cluster of *Drosophila* chorion genes. *Dev Biol* **117**: 294–305.
- Schwed G, May N, Pechersky Y, Calvi BR. 2002. *Drosophila* minichromosome maintenance 6 is required for chorion gene amplification and genomic replication. *Mol Biol Cell* **13**: 607–620.
- Sequeira-Mendes J, Diaz-Uriarte R, Apedaile A, Huntley D, Brockdorff N, Gomez M. 2009. Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet* **5**: e1000446. doi: 10.1371/journal.pgen.1000446.
- Song JS, Johnson WE, Zhu X, Zhang X, Li W, Manrai AK, Liu JS, Chen R, Liu XS. 2007. Model-based analysis of two-color arrays (MA2C). *Genome Biol* **8**: R178. doi: 10.1186/gb-2007-8-8-r178.
- Spradling AC. 1981. The organization and amplification of two chromosomal domains containing *Drosophila* chorion genes. *Cell* **27**: 193–201.
- Takahashi TS, Yiu P, Chou MF, Gygi S, Walter JC. 2004. Recruitment of *Xenopus* Scc2 and cohesin to chromatin requires the pre-replication complex. *Nat Cell Biol* **6**: 991–996.
- Toedling J, Skylar O, Krueger T, Fischer JJ, Sperling S, Huber W. 2007. Ringo—an R/Bioconductor package for analyzing ChIP-chip readouts. *BMC Bioinformatics* **8**: 221. doi: 10.1186/1471-2105-8-221.
- Xie F, Orr-Weaver TL. 2008. Isolation of a *Drosophila* amplification origin developmentally activated by transcription. *Proc Natl Acad Sci* **105**: 9651–9656.