

OVO Positively Regulates Essential Maternal Pathways by Binding Near the Transcriptional Start Sites in the *Drosophila* Female Germline

Reviewed Preprint

v2 • June 21, 2024

Revised by authors

Reviewed Preprint

v1 • February 12, 2024

Leif Benner , Savannah Muron, Jillian G. Gomez, Brian Oliver

Section of Developmental Genomics, Laboratory of Biochemistry and Genetics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA • Department of Biology, Johns Hopkins University, Baltimore, MD, USA

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Abstract

Differentiation of female germline stem cells into a mature oocyte includes the expression of RNAs and proteins that drive early embryonic development in *Drosophila*. We have little insight into what activates the expression of these maternal factors. One candidate is the zinc-finger protein OVO. OVO is required for female germline viability and has been shown to positively regulate its own expression, as well as a downstream target, *ovarian tumor*, by binding to the transcriptional start site (TSS). To find additional OVO targets in the female germline and further elucidate OVO's role in oocyte development, we performed ChIP-seq to determine genome-wide OVO occupancy, as well as RNA-seq comparing hypomorphic and wild type rescue *ovo* alleles. OVO preferentially binds in close proximity to target TSSs genome-wide, is associated with open chromatin, transcriptionally active histone marks, and OVO-dependent expression. Motif enrichment analysis on OVO ChIP peaks identified a 5'-TAACNGT-3' OVO DNA binding motif spatially enriched near TSSs. However, the OVO DNA binding motif does not exhibit precise motif spacing relative to the TSS characteristic of RNA Polymerase II complex binding core promoter elements. Integrated genomics analysis showed that 525 genes that are bound and increase in expression downstream of OVO are known to be essential maternally expressed genes. These include genes involved in anterior/posterior/germ plasm specification (*bcd*, *exu*, *swa*, *osk*, *nos*, *aub*, *pgc*, *gcl*), egg activation (*png*, *plu*, *gnu*, *wisp*, *C(3)g*, *mtrm*), translational regulation (*cup*, *orb*, *bru1*, *me31B*), and vitelline membrane formation (*fs(1)N*, *fs(1)M3*, *clos*). This suggests that OVO is a master transcriptional regulator of oocyte development and is responsible for the expression of structural components of the egg as well as maternally provided RNAs that are required for early embryonic development.

eLife assessment

This **useful** manuscript extends prior work to identify OVO as a major transcriptional activator of the female germline gene expression program. Using a combination of **solid** genomic strategies, the authors demonstrate that OVO binds to the promoters of hundreds of genes in the female germline and promotes their expression.

<https://doi.org/10.7554/eLife.94631.2.sa1>

Introduction

Drosophila early embryonic development is directed by events that take place during oogenesis. Germline stem cells (GSCs) asymmetrically divide to renew the stem cell population and send one daughter cell towards oogenesis. In the germarium, oogenesis is preceded by four rounds of incomplete mitotic divisions resulting in a 16-cell egg chamber. One cell is specified as the oocyte which is arrested in prophase of meiosis I, while the rest of the 15 cells enter endoreplication cycles and become nurse cells (NCs). Once the 16 cell egg chamber buds from the germarium, the NCs begin to transcribe and translate a vast array of RNAs and proteins that serve diverse functional roles (Bastock and St Johnston 2008 [↗](#); Spradling et al. 2022 [↗](#)). These roles include positioning maternal mRNAs and proteins in the correct spatial orientation to support anterior/posterior, and dorsal ventral axis specification, as well negative regulators of translation to ensure that the maternal mRNAs are not translated before fertilization (Lasko 2012 [↗](#)). The oocyte also contains a number of proteins and mRNAs that are needed for egg activation, completion of meiosis and initiation of embryonic development after fertilization (Avilés-Pagán and Orr-Weaver 2018 [↗](#)). To repeat the process of oogenesis from generation to generation, germ cells in the developing embryo need to be specified and maintained separately from the rest of the developing somatic cell population. This requires maternally controlled localization of the germ plasm and early pole cell formation in the embryo (Mahowald 2001 [↗](#)). While the complex interactions between maternally supplied mRNAs and proteins have been well studied, transcriptional regulation driving the expression of these pathways are less well understood.

Few positive regulators of female specific germ cell transcription have been identified. Genes such as *grauzone* (*grau*) and *maternal gene required for meiosis* (*mamo*) have been shown to activate the transcription of *cortex* and *vasa*, respectively, in the female germline (Harms et al. 2000 [↗](#); S. Nakamura et al. 2019 [↗](#)). Active repression of male specific transcription through the activity of *egg*, *wde*, and *Su(var)205* (Smolko, Shapiro-Kulnane, and Salz 2018 [↗](#)), or global repression of non-ovarian transcriptional networks through the function of *sov* (Benner et al. 2019 [↗](#)), have shown the importance of heterochromatin formation in the female germline for cellular identity and oocyte development. In fact, recent work has shown the importance of transcriptional repression, mediated through changes in histone modifications, as a key regulator of egg chamber differentiation. GSCs have been shown to exist in a sort of ‘ground state’ of histone modifications. Characterized with modest non-canonical repressive H3K9me3 and H3K27me3 histone marks at many genes, as well as transcriptionally active H3K27ac histone marks and open chromatin at others (Pang et al. 2023 [↗](#); DeLuca et al. 2020 [↗](#)). As oocyte development continues, repressive histone marks associated with heterochromatin begin to increase in abundance resulting in fewer histone marks associated with transcription and open chromatin. This suggests that gene expression becomes more restricted throughout the differentiation process. However, it is unlikely that the female germline directs oocyte development solely through a repressive transcriptional model. Whether the female germline expresses paralogs of the RNA polymerase II complex, like

the male germline (M. Hiller et al. 2004 [DOI](#); M. A. Hiller et al. 2001 [DOI](#); Lu et al. 2020 [DOI](#)), or if there are pioneering transcription factors involved in determining the open chromatin status for female germ cell-specific expression, or something else entirely, has yet to be determined.

Although few female specific germline transcription factors have been identified, the conserved zinc-finger transcription factor OVO has long been known to be required for female germ cell viability. Female germ cells that are *ovo*[−] do not survive into adulthood (Oliver, Perrimon, and Mahowald 1987 [DOI](#); Oliver, Pauli, and Mahowald 1990 [DOI](#); Benner et al. 2023 [DOI](#)). Hypomorphic *ovo* alleles, specifically ones that disrupt the transcriptional activator OVO-B, show an arrested egg chamber phenotype, indicating that wild-type OVO-B activity is required for oocyte maturation (Salles et al. 2002 [DOI](#); Benner et al. 2023 [DOI](#)). Germline OVO is expressed at all stages of oogenesis, where it is eventually maternally loaded into the egg. Maternal OVO becomes specifically localized to the developing germline and persists throughout embryogenesis until zygotic OVO is expressed (Hayashi et al. 2017 [DOI](#); Benner et al. 2023 [DOI](#)). Thus, OVO is eternally expressed in the female germline, suggesting it may be a key regulator of female specific germline transcription. However, only two downstream targets have previously been identified for OVO. OVO has been shown to positively regulate the expression of its own transcription, therefore executing an autoregulatory loop, as well as positively regulating the transcription of the gene *ovarian tumor (otu)* (Lü et al. 1998 [DOI](#); Bielinska et al. 2005 [DOI](#); Lü and Oliver 2001 [DOI](#); Andrews et al. 2000 [DOI](#)). *otu* is also required in the female germline, where *otu*[−] germ cells show viability and germline tumor phenotypes (Bishop and King 1984 [DOI](#)). The *ovo* phenotype is epistatic to that of *otu*, and ectopic *otu* expression cannot rescue female germ cell death due to loss of *ovo* (Hinson, Pettus, and Nagoshi 1999 [DOI](#); Pauli, Oliver, and Mahowald 1993 [DOI](#)). Therefore, *ovo* must be responsible for activating the transcription of genes in addition to *otu* for female germ cell survival and differentiation.

We expanded our knowledge of OVO's role in the female germline by determining genome-wide OVO occupancy and global transcriptional changes downstream of OVO. This allowed us to determine which genes OVO binds, and which genes transcriptionally respond to OVO *in vivo*. We show that OVO is directly regulating essential maternal pathways such as axis specification, primordial germ cell formation, egg activation, and maternal mRNA translation regulation. Together, we show that OVO plays a pivotal role in the positive transcriptional regulation of oocyte and early embryonic development. We show that OVO likely carries out this regulation by binding at or in close proximity to the promoters of the genes it regulates and that OVO DNA binding motifs are enriched at or near the transcriptional start site (TSS) of OVO responsive genes, although the spacing of OVO binding sites suggests that it is not a component of the RNA Polymerase complex. OVO binding is also a signature of open chromatin status and active transcription throughout oocyte differentiation. Altogether, we suggest that OVO is required for the activity of a large number of female germline promoters and is likely a key regulator of oocyte maturation and RNAs and proteins that are required for early embryonic development.

Results

OVO Binds Promoters Genome-Wide

OVO-B, the predominant protein isoform expressed from the *ovo* locus in the female germline (Benner et al. 2023 [DOI](#)), is a positive regulator of transcription (Andrews et al. 2000 [DOI](#)) at both the *otu* and *ovo* locus. Transgenic reporter constructs of *otu* and *ovo* require OVO binding sites both at and upstream of the TSS in order to recapitulate full reporter expression (Bielinska et al. 2005 [DOI](#); Lü et al. 1998 [DOI](#); Lü and Oliver 2001 [DOI](#)). Females hemizygous for antimorphic dominant gain-of-function (*ovo*^D) or homozygous recessive (*ovo*^{D1rv}) *ovo* alleles lack germ cells in the adult ovary (Oliver, Perrimon, and Mahowald 1987 [DOI](#); Benner et al. 2023 [DOI](#)). True OVO-B null alleles created by deletion of the *ovo-B* promoter have the same germ cell loss phenotype (Benner et al. 2023 [DOI](#)). The phenotypes of *otu* females range from germ cell death to ovarian tumors depending on the allele

and undefined stochastic factors (Bishop and King 1984). It is possible that the germ cell death phenotype in *ovo*⁻ female germ cells can solely be explained by failure of OVO to activate *otu* expression, however, this is highly unlikely. The *ovo*^{D1rv} phenotype is epistatic to *otu*⁻, and ectopic *otu*⁺ expression in *ovo*⁻ germ cells does not rescue the germ cell death phenotype (Hinson, Pettus, and Nagoshi 1999). This suggests that OVO regulates the expression of additional genes in the female germline.

We wanted to identify OVO target genes in the female germline by using two complementary genome-wide approaches to test for OVO presence and function. Specifically, we determined OVO occupancy genome-wide with ChIP-seq, and determined *ovo* function by comparing the RNA expression profiles between *ovo*⁺ and *ovo* hypomorphs in the female germline. In order to determine OVO occupancy genome-wide, we performed ChIP-seq on adult ovaries in triplicate, using two C-terminally tagged alleles as affinity purification tools (*ovo*^{Cterm-3xFLAG} i.e. OVO-HA and *ovo*^{Cterm-GFP} i.e. OVO-GFP, **Figure 1A-C**, **Figure S1A**) (Benner et al. 2023) and called significantly enriched peaks from OVO-HA and OVO-GFP compared to their respective input controls.

We first compared the pulldown results with the OVO-HA versus OVO-GFP ChIP reagents. The GFP pull down appeared to be more efficient, but never-the-less we found excellent agreement, as most OVO-HA peaks were also found in the OVO-GFP dataset. The OVO-GFP ChIP dataset had 7,235 significant ChIP peaks according to peak enrichment analysis genome-wide, while OVO-HA ChIP dataset had 3,393 significant peaks genome-wide (Table S1). To determine the similarity in significant peak calling between the two datasets, we calculated a Jaccard index (intersection/union) between the significantly enriched peaks from the tagged *ovo* allele bearing ovaries. The Jaccard index between OVO-HA and OVO-GFP ChIP peaks was 0.64 (where 0 = no overlap and 1 = full overlap) with a total of 3,094 ChIP peaks overlapping. Thus, almost all significant OVO-HA ChIP peaks were also found within the OVO-GFP ChIP dataset (91% of OVO-HA peaks overlapped OVO-GFP peaks). OVO-GFP pulldown was either more effective, or less likely, promiscuous. Regardless, we decided to use the conservative intersection of the two datasets (3,094 peaks) for downstream OVO occupancy informatics (Table S2).

OVO is a sequence-specific DNA-binding protein, but many transcription factors also have a more general affinity for DNA. Additionally, immunoprecipitation can capture indirect interactions due to nuclear topology in addition to direct sequence-specific binding. For example, in the particular case of OVO TSS binding, this could be direct, as shown in the case of *ovo* and *otu* loci (Lü et al. 1998; Lü and Oliver 2001; Bielinska et al. 2005), or could be due to looping of an OVO-bound enhancer to the core promoter. Determining if there are canonical OVO binding sites at peaks can help to distinguish direct and indirect binding. If OVO is directly binding to the TSS, we would expect to find OVO binding site enrichment at that location. To examine the sequences enriched in peaks, we looked directly for the known OVO binding sites previously defined by footprinting (Lü et al. 1998) and SELEX-seq (Bielinska et al. 2005; Lee and Garfinkel 2000). We also did *de novo* motif finding on this substantial dataset to refine the sequence specific motifs bound by OVO and perhaps other motifs associated with other transcription factors preferentially bound by OVO enhancers, or OVO proximal sites in 3D nuclear space. We performed novel motif enrichment analysis using STREME (Bailey 2021) with our overlapping ChIP peaks and found a number of significant motifs within our dataset. The most significant motif, 5'-TAACGGTAAA-3', was found in 49% of significant OVO ChIP peaks (**Figure 1D**, 'Motif One'). This motif is highly similar to the OVO DNA binding motif that has been reported twice before, 5'-AGTAACNGT-3' (SELEX method, 'Garfinkel OVO Motif') and 5'-TGTAACNGT-3' (Footprinting method, 'Oliver OVO Motif'). The only differences between motif one in our dataset and the literature, is that the *de novo* motif is two nucleotides shorter than the previously described motifs at the 5' end, extends 3 nucleotides downstream, and includes a second G near the 3'-end (**Figure S1B**). Collectively, this is strong evidence that the core OVO binding sequence is 5'-TAACNGT-3'. Some binding sites can be recognized by multiple transcription factors. To determine if other characterized transcription

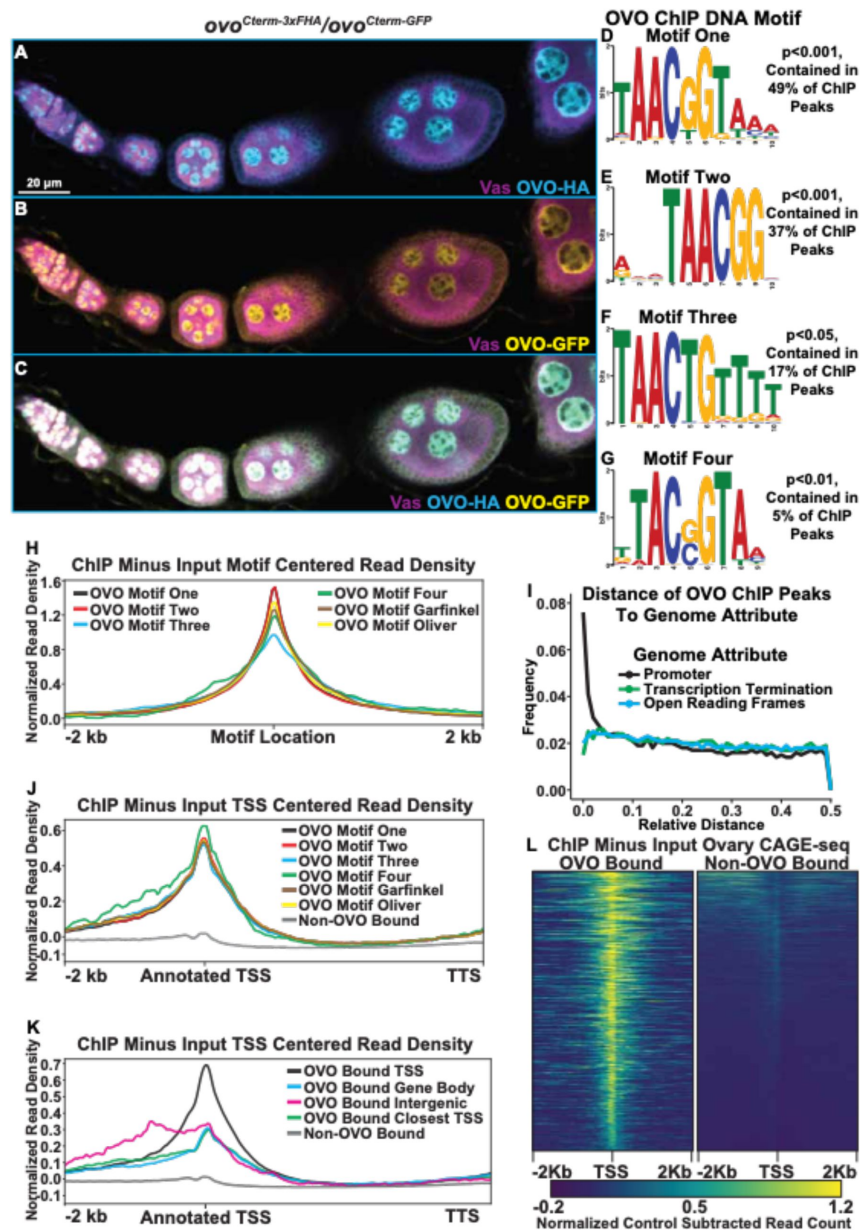


Figure 1

Significantly Enriched OVO DNA Binding Motifs and OVO ChIP Attributes Genome-Wide.

A-C) Immunofluorescent staining of adult ovarioles of *ovo*^{Cterm-3xHA/ovo}^{Cterm-GFP} females (20x, scale bar = 20 μ m). Ovarioles were stained for Vas (purple) to label the germline, HA (cyan) to label OVO-HA, and GFP (yellow) to label OVO-GFP. The homozygous version of these alleles were used to ChIP OVO. D-G) Significantly enriched motifs found within overlapping OVO ChIP peaks. The percentage of OVO ChIP peaks containing each motif and their corresponding p-value are indicated to the right. H) OVO ChIP minus input control ChIP-seq read coverage density centered on the location of the four *de novo* OVO DNA binding motifs and previously defined *in vitro* OVO DNA binding motifs (Lü et al. 1998, Bielinska et al. 2005, Lee and Garfinkel 2000). I) Relative distance of OVO ChIP peaks to gene level promoters, terminations sequences, and open reading frames genome-wide. J) OVO ChIP minus input control ChIP-seq read coverage density for genes containing significant OVO ChIP peaks and the corresponding OVO DNA binding motif. Genes are centered on the transcriptional start site. K) OVO ChIP minus input control ChIP-seq read coverage density for genes bound by OVO over the TSS, gene body, closest TSS in intergenic space, closest TSS for all, or not bound. Genes are centered on the transcriptional start site. L) OVO ChIP minus input control ChIP-seq heatmaps centered on the dominant significant ovary CAGE-seq TSSs overlapping or not overlapping OVO ChIP peaks.

factors recognize this sequence, we searched for significant matches to motif one in the Jasp database (Castro-Mondragon et al. 2022) of known *Drosophila* motifs using Tomtom (Gupta et al. 2007). The OVO DNA binding motif ('Garfinkel OVO Motif') was scored as a significant match (Figure S1B, $p < 0.05$). Along with the OVO DNA binding motif, other motifs were also significantly enriched in OVO ChIP peaks. The motif 5'-GWGMGAGMGAGABRG-3' (Figure S1C) was found in 18% of OVO ChIP peaks and is a significant match to the DNA binding motifs of the transcription factors GAF (*Trl*) (Omelina et al. 2011) and CLAMP (Sorucu et al. 2013). *Trl* germline clones are not viable, indicating that GAF activity is required in the germline during oogenesis (Chen et al. 2009). The possibility that OVO binds with and regulates genes alongside of GAF given the enrichment of both transcription factors DNA binding motifs is intriguing. Other significantly enriched motifs 5'-ACACACACACACA-3' (29% of peaks, Figure S1D), 5'-RCAACAACAACAACA-3' (26% of peaks, Figure S1E), and 5'-GAAGAAGAAGAAGAR-3' (17% of peaks, Figure S1F) were present in OVO ChIP peaks, however, these motifs did not significantly match known DNA binding motifs of other transcription factors. Determining the factors that bind to these sequences will certainly help elucidate our understanding of transcriptional control with relationship to OVO in the female germline.

Not every peak region had one of the consensus OVO DNA binding motifs. This does not mean *a priori* that they bound OVO indirectly. Motif enrichment can be driven by a few strongly enriched sequences, so that more minor enrichments of variants are missed. Therefore, we carefully examined the 51% of our overlapping OVO ChIP peaks where OVO DNA binding motif one was not found. This second round of *de novo* analysis revealed enrichment of OVO DNA binding motif derivatives. For example, the third most significant motif (found in 37% of peaks) was 5'-RWMTAACGGV-3' (Figure 1E, motif two). This motif had the core 5'-TAACNGT-3' sequence found in all three aforementioned methods, however, the last nucleotide in the core sequence is unspecific and lacks the three 3' nucleotides found in motif one. Two other derivative motifs, 5'-TAACTGTTTT-3' (found in 17% of sequences, Figure 1F, motif three), and 5'-TTACSGTAA-3' (found in 5% of sequences, Figure 1G, motif four), vary within the central core motif (at positions 5 and 2 of the core sequence, respectively) and upstream and downstream ends. Searching for all variations of the OVO DNA binding motif (Table S3) within our significant overlapping ChIP peaks indicated that 72% of peaks contained at least one variation of these four binding motifs. It is a reasonable hypothesis that OVO peaks are most often due to direct, rather than indirect OVO binding.

A prediction for direct OVO binding to motifs, is that the motif should be centered within the peak of fragmented input DNA sequences. Therefore, we plotted the significant ChIP (minus input) read density centered on the location of the motif. We found that the read density for all ChIP peaks that contain each one of the *de novo* OVO motifs, as well as the *in vitro* OVO motifs, are centered over the motif location (Figure 1H). This suggests that all of these motifs from our analysis are bonafide OVO DNA binding sites *in vivo*. While it is possible that OVO comes into contact with regions of DNA in three-dimensional nuclear space non-specifically, the presence of OVO motifs within a large percentage of significant ChIP peaks *in vivo* and enrichment of OVO ChIP read density at the location of the motifs, strongly reinforces the idea that our OVO ChIP dataset contains regions centered on sequences specifically bound by OVO in the ovary.

Given the clear function of OVO occupancy near the TSS of its two known targets, *ovo* and *otu* (Lü et al. 1998; Bielinska et al. 2005; Lü and Oliver 2001), we were interested in determining if OVO peaks are generally near the TSS of other target genes as well. In addition to informing the biology of OVO function, this simplifies the problem of associating peaks to potential functional target genes. As a preliminary test of this idea, we determined if the fully overlapping OVO-HA and OVO-GFP peaks were spatially enriched with respect to the currently annotated gene model elements such as TSS, open reading frames (ORFs), or transcription termination sequences (TTS). If the TSS association of OVO at the two known targets reflects a general propensity, then we expect OVO ChIP peaks to be more closely associated with TSS than other gene elements. To carry out this

analysis, we normalized the genome for these three gene elements, such that the distance between adjacent loci was 1. If there is no enrichment for OVO peaks to a specific gene element, then the peak location would have an equal frequency from 0.0 to 0.5 relative distance. Measuring the relative distance of our OVO ChIP peaks to TSS, ORFs, and TTS, showed that the OVO binding was highly enriched near TSS/promoter locations and was not correlated with ORF and TTS locations (**Figure 1I** [↗](#)). These results confirmed that OVO is characterized by core promoter proximal binding. Since OVO ChIP peaks as a class are associated with TSS, we plotted the ChIP minus input read density of genes that overlap significant ChIP peaks to examine the full distribution. We found that the OVO ChIP read density was highly enriched over the TSS and was not due to a few highly enriched examples (**Figure 1J** [↗](#)). This builds on the idea that OVO is binding directly over, or in close proximity to the TSS of its target genes genome-wide. In other words, the previous work showing OVO binding the *ovo-B* and *otu* TSS (Bielinska et al. 2005 [↗](#); Lü et al. 1998 [↗](#); Lü and Oliver 2001 [↗](#)) is typical. This very specific binding of OVO to the TSS is intriguing and unusual, as this region associates with the basal transcriptional machinery. It raises the possibility that OVO is not a typical transcription factor that acts primarily via enhancer binding but might be part of the core promoter binding complex or acts to precondition the core promoter region for example.

Although OVO ChIP peaks overlapping genes showed a strong read density enrichment over the TSS, we found that only 45% (1,394/3,094) of OVO ChIP peaks directly overlapped a TSS. 43% (1,339/3,094) of OVO ChIP peaks were found to overlap the gene body downstream of the TSS (intronic and exonic sequences) and 12% (366/3,094) did not overlap any gene elements, indicating that they were intergenic. We were interested in the differences between OVO binding directly over the TSS or at more distal upstream and downstream sites. We decided to plot the OVO ChIP read density of these different classes of OVO binding patterns and found that OVO bound over the TSS produced a sharp read density enrichment over the TSS which was consistent with what was found for all OVO bound genes (**Figure 1K** [↗](#)). OVO binding along the gene body surprisingly also showed a read density enrichment over the TSS, although the magnitude of read density enrichment was notably less than TSS OVO binding. Intergenic OVO binding also showed these same characteristics with a notable upstream read density enrichment possibly indicative of enhancer binding. This indicates that although the significantly called OVO ChIP peaks did not overlap the TSS, there was still a propensity for TSS sequences to be enriched with OVO ChIP over the input control. This could be due to weaker direct *in vivo* binding of OVO to these TSSs or indirect interactions between the upstream/downstream OVO bound sequences and the TSS, possibly through a looping enhancer-promoter interaction. However, regardless of the location of the OVO ChIP peak, OVO seemed to always be enriched at or in close proximity to TSSs.

The OVO ChIP read density was highly enriched over the annotated TSS of target genes, but TSS annotation is challenging and can be tissue specific. We were interested in empirically determining if the same enrichment was present in TSSs utilized specifically in ovarian tissue. The 5' ends of mRNA are capped. In order to determine where these caps mapped to the genome, we analyzed Cap Analysis of Gene Expression (CAGE-seq) data from adult *Drosophila* ovaries (Chen et al. 2014 [↗](#)) and extracted the dominant significant TSSs in the ovary. CAGER predicted 6,856 significant TSSs in the ovary dataset, of which 1,047 overlapped with OVO ChIP peaks. We plotted the OVO ChIP minus input read density centered on the significant ovary CAGE-seq TSSs for TSSs that overlapped or did not overlap OVO ChIP peaks (**Figure 1L** [↗](#)). We found that OVO ChIP read density was highly enriched over the location of TSSs from ovary CAGE-seq that overlapped OVO ChIP peaks when compared to TSSs that did not overlap OVO ChIP peaks. Thus, OVO TSS binding is not due to poor annotation of ovarian TSSs. Furthermore, OVO is binding at or near TSSs of genes actively being transcribed in the ovary.

OVO Binding is Associated with Open Chromatin and Transcriptionally Active Histone Marks

Our OVO ChIP data indicated that OVO was binding at or in close proximity to promoters genome-wide. OVO could have a positive and/or negative effect on transcription at these locations. For example, OVO could help recruit or sterically hinder RNA Polymerase binding to TSSs. However, previous *ovo* reporter constructs show positive effects of OVO binding near TSS (Lü and Oliver 2001 [↗](#); Lü et al. 1998 [↗](#); Bielinska et al. 2005 [↗](#)). If OVO binding is generally promoting transcription, then we hypothesize that it would be more closely associated with histone marks associated with active transcription, such as H3K27ac and H3K4me3, as well as lower nucleosome density that can be measured through ATAC-seq. In contrast, OVO binding would be expected to negatively correlate with repressive H3K9me3 and H3K27me3 histone marks and higher nucleosome density. It is technically difficult to determine changes in chromatin status and transcription in germ cells that lack OVO, as the phenotype is cell death (although we will return to this later for transcription profiling) but analyzing OVO binding in the context of ovarian chromatin was highly informative.

Recent work profiling nucleosome density and histone marks have shown that female GSCs have a ‘ground state’ chromatin profile (DeLuca et al. 2020 [↗](#)), similar to the histone mark profiles that are found in early embryos (Li et al. 2014 [↗](#)). This has been characterized to contain non-canonical H3K27me3 profiles and low H3K9me3 histone levels (Pang et al. 2023 [↗](#); DeLuca et al. 2020 [↗](#)). As egg follicles differentiate, nurse cells begin to accumulate H3K9me3 marks, and H3K27me3 histone marks begin to accumulate over more traditional polycomb domains. This in turn leads to a decrease in the number of open chromatin peaks as well as H3K27ac histone marks, which are generally associated with active transcription (Pang et al. 2023 [↗](#); DeLuca et al. 2020 [↗](#)). Essentially, these data support the idea that egg chambers restrict gene expression competency as they differentiate.

In order to determine the relationship in our OVO ChIP data and other chromatin marks, we analyzed GSC H3K27ac, H3K27me3, H3K9me3, H3K4me3, and ATAC-seq data (Pang et al. 2023 [↗](#); DeLuca et al. 2020 [↗](#)) with the same parameters used to establish significant OVO peaks in our OVO ChIP dataset. Our OVO ChIP data was from one day old ovaries, and we did not profile specific follicle stages. So we also analyzed 32c (roughly stage 5 egg chambers) H3K27ac, H3K27me3, ATAC-seq, and 8c H3K9me3 (32c was not available) histone marks (Pang et al. 2023 [↗](#); DeLuca et al. 2020 [↗](#)) to see if there were any stage specific differences in comparison to OVO DNA binding. We focused specifically on GSC and 32c egg follicles for these chromatin marks since that is when the *ovo* hypomorphic egg chambers arrest (Benner et al. 2023 [↗](#)). We first plotted the read density of each respective chromatin mark minus their input control centered on either the OVO ChIP peak local maximum (**Figure 2A** [↗](#)) or OVO DNA binding motifs contained within significant OVO ChIP peaks (**Figure 2B** [↗](#)). GSC ATAC and H3K27ac read density showed a high degree of enrichment over OVO ChIP peak maximums (**Figure 2A** [↗](#)) and OVO DNA binding motifs (**Figure 2B** [↗](#)), consistent with positive transcriptional activity. GSC H3K4me3 read density was, to a lesser extent, also enriched with OVO ChIP peak maximums and OVO DNA binding motifs. However, there was no read density enrichment for repressive GSC H3K27me3 and H3K9me3 histone marks. The same findings generally held true when looking at the overlap of OVO ChIP peaks and chromatin marks in differentiating egg chambers. Notably, there was an even higher read density enrichment over OVO ChIP peak maximum and DNA binding motifs for 32c ATAC-seq data, while read density enrichment decreased for 32c H3K27ac histone mark (**Figure 2A** [↗](#), B).

Since there was a high degree of read density enrichment between OVO ChIP and other chromatin marks/low nucleosome density we wanted to determine the extent of the overlap between significant OVO ChIP peaks and significantly called peaks from the different types and stages of histone marks and ATAC-seq data. To do this, we measured the relative distance of OVO ChIP peaks

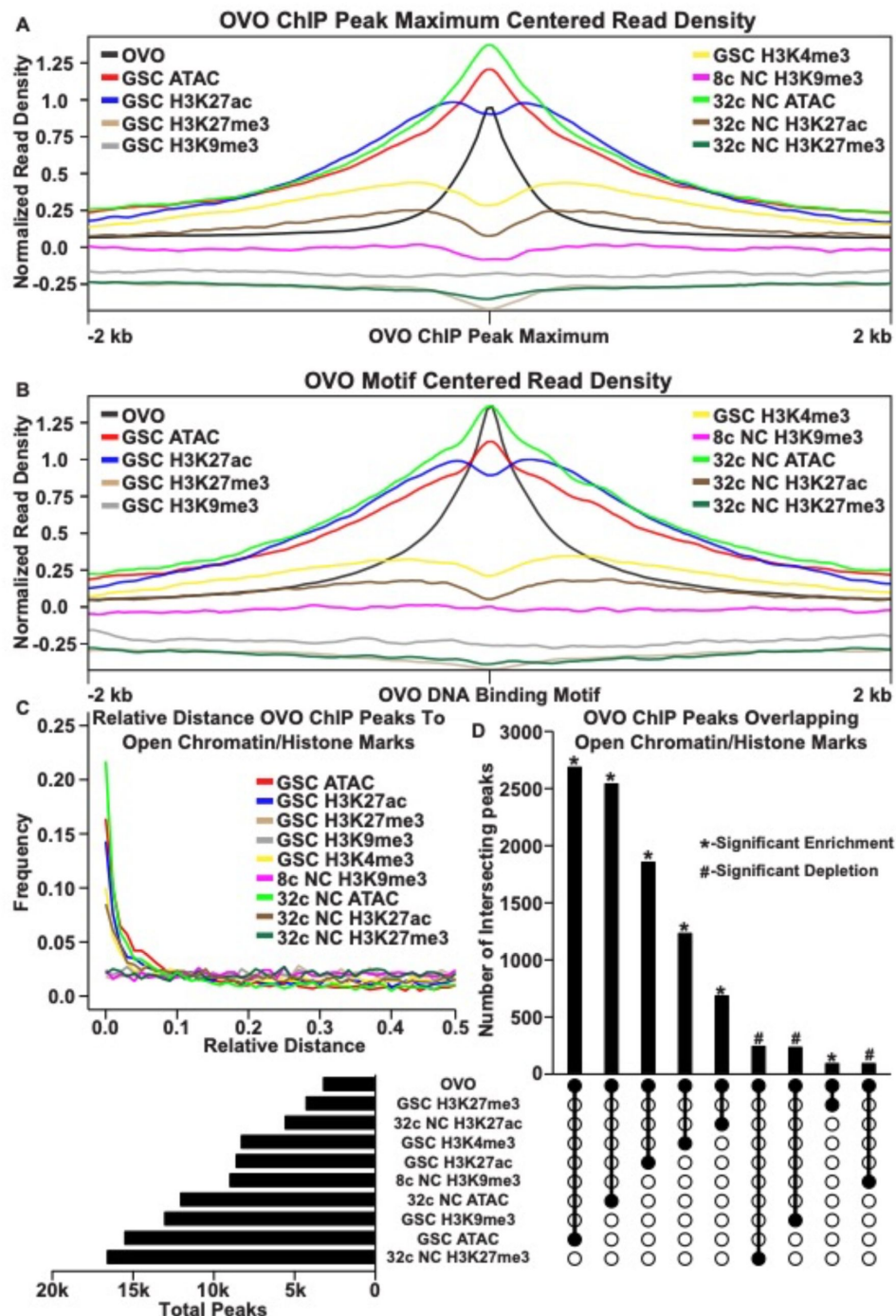


Figure 2

OVO DNA Binding is Associated with Open Chromatin and Transcriptionally Active Histone Marks.

A, B) OVO ChIP minus input control, GSC and 32c ATAC-seq, GSC H3K27ac, H3K4me3, H3K27me3, H3K9me3, 8c NC H3K9me3, 32c NC H3K27ac, and H3K27me3 ChIP-seq read coverage density centered on each OVO peak maximum or OVO DNA binding motif located within a significant OVO ChIP peak. C) Relative distance of OVO ChIP peaks to significantly called peaks for GSC and 32c ATAC-seq, GSC H3K27ac, H3K4me3, H3K27me3, H3K9me3, 8c NC H3K9me3, 32c NC H3K27ac, and H3K27me3 ChIP-seq genome-wide. D) Total number of significant peaks (left) and the total number of overlapping peaks (top) between OVO ChIP and GSC and 32c ATAC-seq, GSC H3K27ac, H3K4me3, H3K27me3, H3K9me3, 8c NC H3K9me3, 32c NC H3K27ac, and H3K27me3 ChIP-seq. Lines connecting solid dots indicates the amount of overlapping peaks between those two corresponding datasets. Asterisk indicates significantly enriched overlap while hashtag indicates significantly depleted overlap between datasets.

to the same datasets described above. We found that OVO ChIP peaks had a lower relative distance, and thus were spatially overlapping/closer in the genome, to 32c NC ATAC, GSC ATAC, GSC H3K27ac, GSC H3K4me3, and 32c NC H3K27ac peaks, in that order (**Figure 2C**). While the relative distance between OVO ChIP peaks and H3K9me3 and H3K27me3, regardless of stage, showed no spatial association. There was also further support for this association with transcriptionally active histone marks/open chromatin when measuring the overlap between significant OVO ChIP peaks and the respective significant histone ChIP and ATAC-seq peaks (**Figure 2D**). A Fisher's exact test indicated a significant enrichment in overlapping peaks genome-wide between OVO and GSC ATAC ($p < 0.001$, odds ratio = 75.9), 32c NC ATAC ($p < 0.001$, odds ratio = Infinite), GSC H3K27ac ($p < 0.001$, odds ratio = 31.7), GSC H3K4me3 ($p < 0.001$, odds ratio = 12.0), and 32c NC H3K27ac ($p < 0.001$, odds ratio = 7.9) peaks. While there was a significant depletion in overlapping peaks genome-wide between OVO and 32c NC H3K27me3 ($p < 0.001$, odds ratio = 0.6), GSC H3K9me3 ($p < 0.001$, odds ratio = 0.7), and 8c NC H3K9me3 ($p < 0.001$, odds ratio = 0.5).

The association of OVO binding with active histone marks and open chromatin was striking, but open chromatin is likely a general phenomenon of promoters (Haines and Eisen, 2008). Indeed, when measuring the read density for GSC and 32C ATAC-seq for OVO bound and OVO non-bound promoters, there is an enrichment for open chromatin at the TSS regardless of OVO binding. However, we did notice an increase in enrichment for OVO bound promoters compared to OVO non-bound promoters (**Figure S1G**), possibly suggesting that OVO bound promoters are more open or have an increase in accessibility when compared to non-OVO bound promoters. This same relationship held true for the transcriptionally active histone mark H3K27ac in GSCs (**Figure S1H**). Since only 45% of OVO ChIP peaks overlapped TSSs, we plotted the read density of the above chromatin marks over OVO ChIP peak maximums for OVO bound over the TSS, gene body, or intergenic regions (**Figure S2A-D**). We found that OVO bound regions that were not overlapping the TSS still showed the same propensity for enrichment of open chromatin and active histone marks. Intergenic regions were especially enriched for open chromatin measured through ATAC-seq. Altogether suggesting that OVO binding genome-wide is tightly associated with open chromatin regardless of germ cell stage, and active transcription in GSCs. In other words, chromatin state data suggests OVO is acting positively on its target genes and raises the possibility that OVO-binding and open chromatin are related.

OVO DNA Binding Motifs are Evenly Distributed Around Promoters and are Enriched for INR, DPE, and MTE Elements

Our data thus far clearly indicates that OVO binding occurs at or very near the core promoter, a region recognized by an enormous collection of factors that associate with RNA polymerase to initiate transcription (Aoyagi and Wassarman 2000; Vo Ngoc, Kassavetis, and Kadonaga 2019). The highly organized polymerase complex has sequence-specific DNA recognition sites with incredibly precise spacing between them, with an overall DNA footprint of a little less than 100bp (Rice, Chamberlin, and Kane 1993; FitzGerald et al. 2006; Ohler et al. 2002). There are upstream binding sites such as TATA, sites at transcription start, such as the initiator (INR), and downstream promoter elements (DPE) (Vo Ngoc, Kassavetis, and Kadonaga 2019). The combinations of these DNA motifs is not random in mammals and *Drosophila* (FitzGerald et al. 2006), and distinct combinations of different motifs at the TSS of genes expressed in *Drosophila* are conserved over tens of millions of years of evolution (Chen et al. 2014). The male germline expresses a number of TATA-associated factors that have been implicated in male-specific promoter usage for gene expression (M. Hiller et al. 2004; M. A. Hiller et al. 2001; Lu et al. 2020; V. C. Li et al. 2009). It is possible that OVO is a female germline specific TATA-associated factor, and if so, OVO binding sites at core promoters should share precise spacing with other core promoter elements, suggesting it is likely part of the complex. If not, then OVO is more likely to facilitate binding of the basal transcriptional machinery. Because of the extended footprint of engaged RNA polymerase, OVO and the basal machinery would not be likely to occupy the same region at the same time.

Like OVO chip peaks, OVO DNA binding motifs were highly enriched at or near the TSS (**Figure 3A**). We carefully analyzed the spacing of these sites relative to core promoter elements to see if spacing was precise at the nucleotide level. We first searched for the presence of previously defined DNA motifs that are enriched at promoters (FitzGerald et al. 2006; Ohler et al. 2002; Lim et al. 2004) using FIMO (Grant, Bailey, and Noble 2011). We defined promoters by using the DNA sequences 150 nucleotides upstream and downstream of the significant dominant TSSs in our previously analyzed ovary CAGE-seq datasets (Chen et al. 2014). After extracting these sequences and searching for significant scoring motifs, we plotted the density of each motif in relation to the empirically mapped TSSs (**Figure 3B**). We also searched for all OVO motifs found in our significant ChIP peaks within these promoter sequences. When plotting the density of DNA motifs found in ovary CAGE-seq promoters, we found that there were prominent peaks for INR and M1BP (M1BP (J. Li and Gilmour 2013) = Ohler 1 (Ohler et al. 2002) = DMv4 (FitzGerald et al. 2006)) near the TSS, and MTE (Lim et al. 2004) and DPE elements downstream of the TSS. This distribution and frequency are consistent with the constrained location of these DNA motifs (FitzGerald et al. 2006; Ohler et al. 2002; Chen et al. 2014). Significantly, the OVO DNA binding motifs showed a broad distribution upstream and downstream of the TSS.

The precise core promoter architecture of OVO bound TSSs is revealed in the CAGE-seq dataset. Plotting the distribution of classical core promoter sequence elements in OVO bound promoters showed a similar, but exaggerated, profile compared to all core promoters of the ovary CAGE-seq dataset. We found a significant enrichment for INR ($p < 0.01$, odds ratio = 1.70), DPE ($p < 0.01$, odds ratio = 1.81), MTE ($p < 0.01$, odds ratio = 1.65), and most importantly, OVO DNA binding motifs ($p < 0.01$, odds ratio = 4.83), in ovary promoters that overlapped an OVO ChIP peak in comparison to the subset of ovary promoters that did not overlap an OVO ChIP peak (**Figure 3C**, **3D**). This indicates that OVO bound promoters are more likely to contain these specific promoter elements than non-OVO bound promoters. As has been described before, promoters containing INR and DPE, but lacking TATA-box elements, are common among *Drosophila* gonad promoters compared to promoters of other tissue types (Chen et al. 2014). The presence of TATA-box elements is negatively associated with germline-specific gene expression (FitzGerald et al. 2006). We found that TATA-box elements were significantly depleted in ovary CAGE-seq promoters when compared to testes ($p < 0.01$, odds ratio = 0.78) (**Figure 3E**) or digestive system ($p < 0.01$, odds ratio = 0.50) (**Figure 3F**) CAGE-seq promoters. Indeed, both *ovo* and *otu* have TATA-less promoters. Briefly, OVO bound promoters are characterized by the presence of INR, DPE, MTE, and, of course, OVO DNA binding motifs. This could represent a functional class of promoters utilized for gene expression in the *Drosophila* ovary. Importantly, the distribution of OVO DNA binding motifs in ovary promoters is not fixed relative to TSSs or other core promoter elements. Thus, it is highly unlikely that OVO acts as a female germline RNA Polymerase complex member that anchors the complex to the core promoter and helps determine the +1 mRNA nucleotide. Rather, the imprecise location of OVO binding sites might suggest that OVO is more likely to facilitate the binding of other basal transcriptional factors.

OVO Activates Gene Expression in the Female Germline

Occupancy is a requirement for activity, but occupancy does not equal activity. Understanding the transcriptional consequences of OVO occupancy genome-wide would allow us to investigate mechanisms. However, as we mentioned earlier, the fact that *ovo* is absolutely required for female germline viability greatly complicates this analysis. Measuring gene expression in dead or dying germ cells was unlikely to be informative. Previous work from our lab has identified a transheterozygous *ovo* allelic combination (*ovo^{ovo-GAL4}/ovo^{ΔBP}*) that greatly reduces OVO activity resulting in sterility, however, female germ cells are able to survive up until at least stage 5 of oogenesis (Benner et al. 2023). *ovo^{ovo-GAL4}* is a CRISPR/Cas9 derived T2A-GAL4-3xSTOP insertion upstream of the splice junction of exon 3 in the *ovo*-RA transcript (**Figure S1A**). Importantly, this insertion in the extended exon 3 would disrupt roughly 90% of the *ovo*-B transcripts. However, since about 10% of *ovo*-B transcripts utilize an upstream splice junction in exon 3, these transcripts would not be disrupted with the T2A-GAL4-3xSTOP insertion and thus

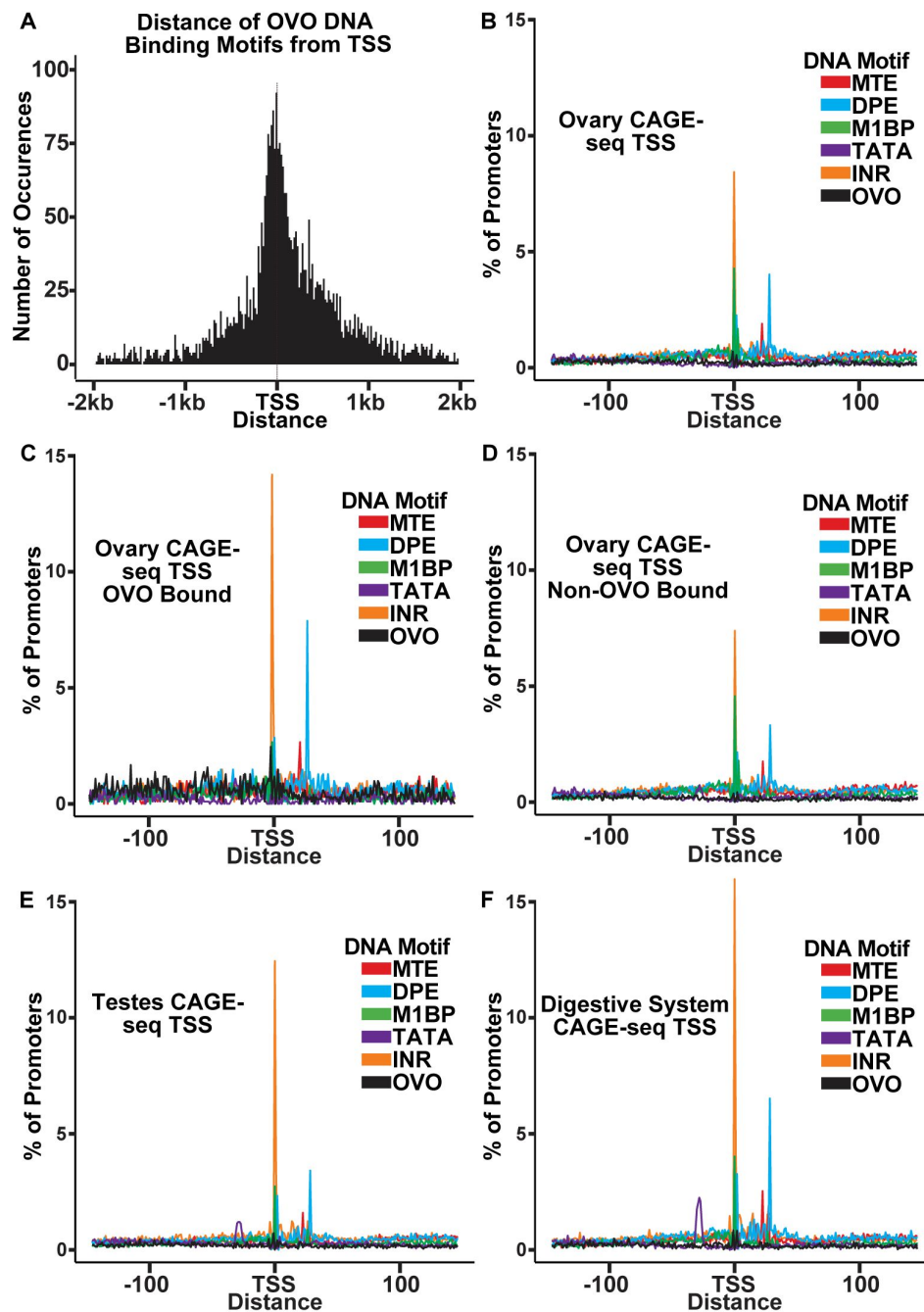


Figure 3

OVO Bound Promoters are Enriched for INR, DPE, and MTE elements.

A) Histogram of the distance of *in vivo* and *in vitro* OVO DNA binding motifs within significant overlapping OVO ChIP peaks from the closest genes TSS. B-F) Histogram of the percent of promoters from tissue specific CAGE-seq analysis of common promoter motif elements centered on the dominant significant TSS.

allow for enough OVO activity for germ cell survival (Benner et al. 2023 [\[4\]](#)). Since *ovo^{ovo-GAL4}* expresses GAL4 in place of full length OVO due to the T2A sequences, we can drive expression of a rescuing OVO-B construct downstream of *UASp* to generate OVO⁺ female germ cells, which in fact does rescue the arrested germ cell phenotype of *ovo^{ovo-GAL4}/ovo^{ΔBP}* ovaries. Therefore, in order to determine genes that are transcriptionally responsive to OVO, we compared the gene expression profiles in sets of ovaries that had the *ovo* hypomorphic phenotype with a negative control rescue construct (*ovo^{ovo-GAL4}/ovo^{ΔBP}; UASp-GFP*) (Figure 4A [\[4\]](#)) versus those that drive expression of the rescue construct expressing OVO-B (*ovo^{ovo-GAL4}/ovo^{ΔBP}; UASp-3xFHA-OVO-B*) (Figure 4B [\[4\]](#)).

Since *ovo^{ovo-GAL4}/ovo^{ΔBP}; UASp-3xFHA-OVO-B* females have full rescue of the arrested germ cell phenotype seen in *ovo^{ovo-GAL4}/ovo^{ΔBP}; UASp-GFP* females, we needed to take further measures to ensure our analysis of gene expression was stage comparable between the two sets of ovaries. The adult female ovary contains somatic cells, germline stem cells, and germline derived nurse cells that would be profiled in a bulk ovary tissue RNA-seq experiment. Although OVO is only required and expressed in germline derived cell types, we chose to dissect one day old post-eclosion *ovo^{ovo-GAL4}/ovo^{ΔBP}; UASp-3xFHA-OVO-B* female ovaries to enrich for early stages of oogenesis and collected only ovarioles containing the germarium through previtellogenic egg chambers. *ovo^{ovo-GAL4}/ovo^{ΔBP}; UASp-GFP* ovaries were collected at the same age post-eclosion and we specifically collected ovaries that contained a visible ovariole structure (and therefore contained germ cells) to minimize comparing germ cells to somatic ovary structures, but rather germ cells to germ cells. We then performed RNA-seq in quadruplicate and measured the changes in gene expression between ectopic rescue OVO and hypomorphic OVO ovaries. We used a significance level of $p\text{-adj} < 0.05$ and a log2 fold change cutoff of $> |0.5|$ to call differential expression between these two sets of ovaries. We utilized these log2 fold change cutoffs for two reasons. Our control ovary genotype (*ovo^{ovo-GAL4}/ovo^{ΔBP}; UASp-GFP*) has hypomorphic OVO activity, hence germ cells can survive but are arrested. With the addition of ectopic rescue OVO in *ovo^{ovo-GAL4}/ovo^{ΔBP}; UASp-3xFHA-OVO-B* ovaries, we predicted that genes that were directly regulated by OVO would transcriptionally respond, however, we were unsure as to what degree the response would be in comparison to hypomorphic OVO. We reasoned that if the changes were not significant between genotypes, then minor changes in gene expression would not matter. Our second reason for using these cutoffs is we had an internal control between the two genotypes. We knew through immunostaining that Vas protein was present in the germline of both genotypes (Figure 4A [\[4\]](#), 4B) and therefore was likely expressed at similar levels in the germline of both genotypes. Both genotypes also expressed *GAL4* under the control of *ovo* in the germline. We examined the expression levels of *vas* and *GAL4* and found that *vas* had a log2 fold change of 0.15 ($p\text{-adj} = 0.03$) and *GAL4* had a log2 fold change of 0.33 ($p\text{-adj} = 0.18$) (Figure 4C [\[4\]](#)). These data suggest a slight underrepresentation of germline expression in *ovo* hypomorphic ovaries. Therefore, by using the greater than 0.5 and less than -0.5 log2 fold change cutoffs, and a less than 0.05 $p\text{-adj}$ value cutoff, we would be conservative to not call genes differentially expressed due to differences in the relative abundance of germ cells and somatic cells.

We were able to reliably detect the expression of 10,804 genes in these early ovarioles (Table S4). The differential expression analysis indicated that 1,994 genes primarily expressed in the germline (see next paragraph) significantly increased in expression with ectopic rescue OVO (Figure 4C [\[4\]](#), cyan/purple dots) and 2,924 genes primarily expressed in the soma (see next paragraph) significantly decreased in expression with ectopic rescue OVO expression (Figure 4C [\[4\]](#), yellow/blue dots). 5,886 genes were not considered to be differentially expressed in our analysis (Figure 4C [\[4\]](#), gray dots). OVO is expressed in the germline, not the soma, and previous work has shown that OVO-B is a transcriptional activator (Andrews et al. 2000 [\[4\]](#)), so we hypothesized that many of the genes increasing in expression in the presence of rescuing OVO were direct downstream targets. We found that 2,298 genes that were expressed in our RNA-seq data overlapped an OVO ChIP peak. 666 genes significantly increased in expression and were bound by OVO, which is a significant enrichment according to a Fisher's exact test (Figure 4C [\[4\]](#), cyan dots, $p < 0.01$, odds ratio = 2.21). While conversely, 564 genes decreased in expression and were bound by

OVO, indicating a significant depletion according to a Fisher's exact test (**Figure 4C**, blue dots, $p < 0.01$, odds ratio = 0.85). This strongly suggests that genes that are bound by OVO, transcriptionally respond in a positive manner. This finding is fully consistent with our meta-analysis comparing OVO ChIP-seq and histone ChIP/ATAC-seq data (**Figure 4E**). OVO binding was highly associated with transcriptionally active histone marks such as H3K27ac and H3K4me3, open chromatin, and increased expression.

There are genes that showed decreased expression in the OVO rescued ovaries, but we believe this is technical rather than biological. OVO is expressed only in the germline, but ovarioles contain germ cells and somatic cells. The presence of empty ovarioles, containing leftover strings of somatic cells, are evident even in *ovo^{ovo-GAL4/ovo^{ABP}}; UASp-GFP* ovaries that contain germ cells. Conversely, *ovo^{ovo-GAL4/ovo^{ABP}}; UASp-3xFHA-OVO-B* ovaries are fully rescued, and therefore possess more germ cell containing ovarioles than *ovo^{ovo-GAL4/ovo^{ABP}}; UASp-GFP* ovaries (Benner et al. 2023). Despite our best efforts to dissect individual ovarioles with a full complement of germ cells and egg chambers, we wondered if there might be fewer germ cells and egg chambers in the *ovo^{ovo-GAL4/ovo^{ABP}}; UASp-GFP* ovaries. To confirm that genes increasing in expression in ectopic rescue OVO were germline derived, we cross-referenced the significantly expressed genes in our RNA-seq datasets with the modENCODE developmental RNA-seq datasets (Graveley 2010). We extracted the gene names of all genes that were considered to be 'moderately expressed' in 0-2 hour old embryos, which are produced during oogenesis and are deposited into the early embryo. We found that 71% of genes (1,409/1,994) that had a significant increase in expression in the presence of ectopic rescue OVO were found to be expressed in 0-2 hour old embryos (**Figure 4D**, green dots), while only 21% of genes (625/2,924) that had a significant decrease in expression were found in the same embryo dataset (**Figure 4D**, yellow dots). 3,448 genes from the 0-2 hour old embryo dataset were not differentially expressed in our RNA-seq dataset (**Figure 4D**, purple dots). A Fisher's exact test confirmed that there was a significant enrichment for genes that significantly increased in expression and were present in 0-2 hour old embryos ($p < 0.01$, odds ratio = 2.8). In comparison, there was a significant depletion for genes that significantly decreased in expression and were present in 0-2 hour old embryos ($p < 0.01$, odds ratio = 0.17). This result indicated that genes that significantly increased in expression were more likely to be expressed in the germline and that the presence of ectopic rescue OVO significantly increased the expression of genes that were maternally deposited in the early embryo. While the set of genes that significantly decrease in expression are not enriched in the embryo and are more likely specific to somatic cell gene expression. These genes are unlikely to be direct OVO targets due to the absence of OVO in those cells, although we certainly cannot rule out the possibility of a non-autonomous effect of OVO on somatic gene expression. In terms of the germline proper, OVO appears to be a positively acting transcription factor.

OVO Positively Regulates Essential Oogenesis Genes

We wanted to examine a subset of the OVO target genes in detail, and began with the known OVO targets, *ovo* itself and *otu* (Bielinska et al. 2005; Lü et al. 1998; Lü and Oliver 2001). Since the relationship between OVO binding to these two genes has been well-characterized, we validated the OVO ChIP, histone ChIP/ATAC-seq, and RNA-seq datasets by examining these two genes first. Since OVO positively regulates the expression of both these genes, then we would expect OVO to be physically bound at OVO motifs required for high transcription, the presence of transcriptionally active histone marks and open chromatin at these loci, as well as a positive transcriptional response in the presence of rescuing OVO-B. This is exactly what we observed. A significant OVO ChIP peak was found overlapping the TSS of *ovo-B*, with four significant OVO DNA binding motifs present (**Figure 5A**). ATAC-seq, H3K27ac, and H3K4me3 peaks overlapped the *ovo* promoter. Transcriptionally, *ovo* RNA-seq reads are likely derived from the *UASp-3xFHA-OVO-B* cDNA rescue or are indistinguishable between the genomic locus and rescuing cDNA transgene. We found a non-significant increase in exon 3 to exon 4 intronic *ovo* reads with the expression of ectopic rescue OVO (log2 fold change = 0.76, $p\text{-adj} = 0.26$). These intronic reads would be derived from the endogenous *ovo* locus, but it is difficult to conclusively determine if the endogenous *ovo* locus

would respond transcriptionally to ectopic OVO downstream of *UASp* (for example, the pathway for *ovo* is no longer autoregulatory in *ovo^{ovo-GAL4/ovo^{ΔBP}}*; *UASp-3xFHA-OVO-B* germ cells, there is an additional GAL4>*UASp* activation step). So, we could not confidently assess whether *ovo* responded transcriptionally to ectopic rescue OVO. However, when looking at the *otu* locus (**Figure S3** [↗](#)), we found OVO occupancy over the TSS of both annotated *otu* promoters, with significant OVO DNA binding motifs overlapping and in close proximity to the TSSs. The *otu* locus contained similar ATAC-seq and activating histone mark peaks overlapping the TSS found at the *ovo* locus. It was also evident that *otu* had a positive transcriptional response to the presence of OVO rescue (log2 fold change = 2.41; p-adj < 0.001). These results confirm that OVO binds and positively regulates both itself and *otu* *in vivo*, as previous work has indicated.

Since our overlapping OVO ChIP-seq and RNA-seq data suggests that hundreds of genes that are bound by OVO increase in expression in the presence of ectopic rescue OVO, we wanted to know more about the functions of those genes. To do this, we performed Gene Ontology enrichment analysis with gProfiler software (Raudvere et al. 2019 [↗](#)). To be especially stringent, we focused on the genes that contained an OVO ChIP peak overlapping the transcriptional start site and significantly increased in expression in the presence of rescue OVO. 525 genes met these criteria. Biological process GO term enrichment analysis on these 525 genes showed a significant enrichment for 45 GO terms (Table S5). The significant GO terms were almost exclusively related to female reproduction and maternal control of early embryonic development (**Figure 6A** [↗](#)). Genes that are required for processes such embryonic axis specification, mRNA localization, egg activation, and translational regulation were found within these significantly enriched GO terms. These GO terms are well understood in the context of oogenesis and broadly suggest that OVO expression in adult gonads is essential for constructing an egg and depositing maternal RNAs to support zygotic embryonic development.

GO term enrichment analysis of genes that are bound by OVO and increase in expression in the presence of ectopic rescue OVO suggested that OVO is likely a main transcriptional regulator of oogenesis. These genes are the subject of decades of work on *Drosophila* oogenesis, but essentially all the work on them has focused on what they do, not on how they are transcriptionally regulated. For example, *bicoid* (**Figure 6B** [↗](#)), and *bicoid* mRNA binding proteins *exuperantia* (*exu*) and *swallow* (*swa*), are essential for anterior specification of the embryo (Lasko 2012 [↗](#)). All these genes were occupied by OVO *in vivo*, significantly upregulated by OVO-B, and had OVO motifs in close proximity to the TSS. Genes involved in posterior patterning (*oskar* and *nanos*) (**Figure 6C** [↗](#)), as well as pole cell specification genes (*polar granule component*, *germ cell-less*, and *aubergine*) (Lasko 2012 [↗](#); L. Benner, Deshpande, and Lerit 2018 [↗](#)), also showed similar RNA-seq, ChIP-seq, and OVO DNA binding motif profiles as *ovo* and *otu*. Genes that are involved in translational silencing and regulation of maternally provided mRNAs, such as *cup* (**Figure 6D** [↗](#)), *maternal expression at 31B* (*me31B*), *oo18 RNA-binding protein* (*orb*), and *bruno 1* (*bru1*), as well as essential genes involved in meiosis completion and egg activation after fertilization (*giant nuclei* (*gnu*), *pan gu* (*png*), *plutonium* (*plu*), *wispy* (*wisp*), *C(3)G*, and *matrimony* (*mtrm*)) (**Figure 6E** [↗](#)) (Lasko 2012 [↗](#); Avilés-Pagán and Orr-Weaver 2018 [↗](#)) all show this stereotypic pattern of promoter proximal OVO occupancy and DNA binding motifs, and OVO-dependent transcription. These data indicate that the OVO is a central transcription factor activating the expression of essential maternal and early embryonic development pathways in the female germline.

We also found that the genes *fs(1)N*, *fs(1)M3*, and *closca*, were all bound by OVO and responded transcriptionally to the presence of ectopic rescue OVO. These genes are significant because they constitute a set of genes that are expressed in the germline and the encoded proteins are eventually incorporated into the vitelline membrane providing the structural integrity and impermeability of the egg (Mineo, Furriols, and Casanova 2017 [↗](#); Ventura et al. 2010 [↗](#)). Loss-of-function of these three genes results in flaccid eggs that are permeable to dye and fail to develop. The loss-of-function phenotype of *fs(1)N*, *fs(1)M3*, and *closca* closely resembles the dominant antimorph *ovo^{D3}* phenotype. The *ovo^{D3}* allele is the weakest of the original dominant-negative *ovo*

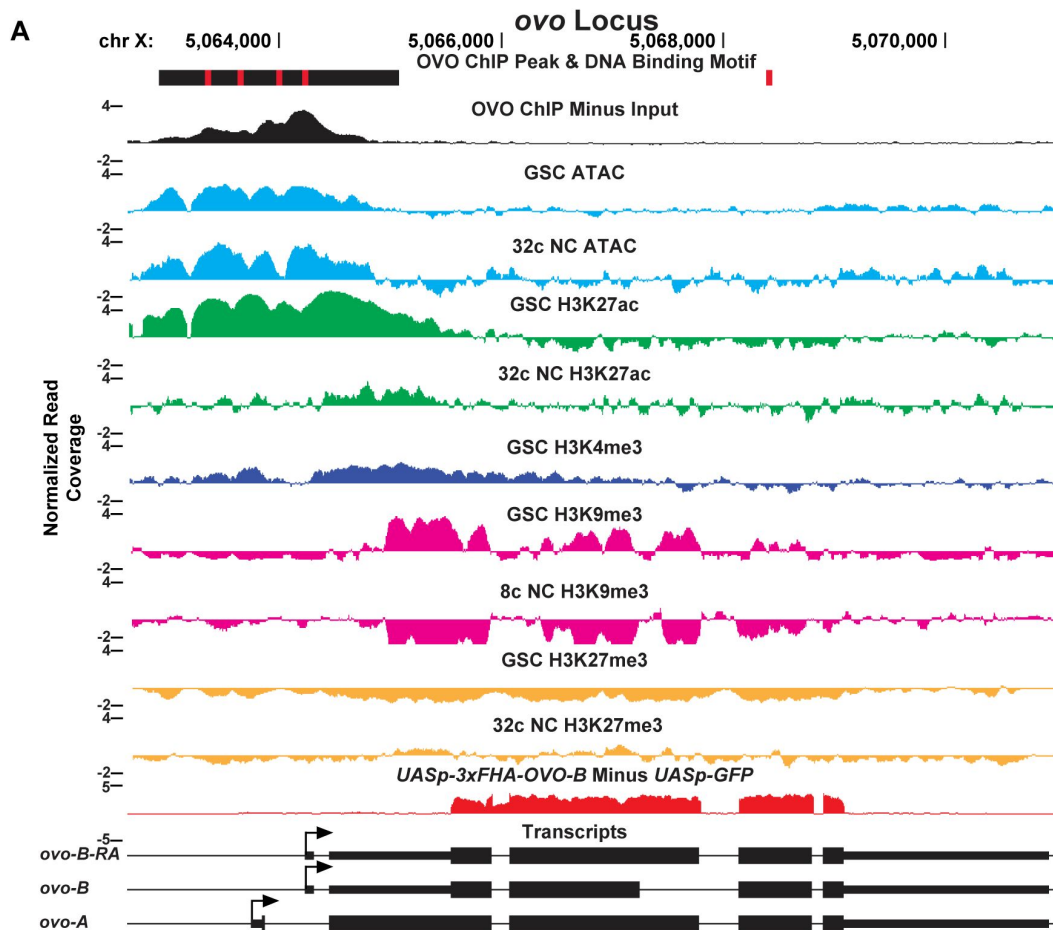


Figure 5

OVO ChIP-seq, ATAC/Histone ChIP-seq, RNA-seq, and DNA Binding Motifs at the *ovo* Locus.

ovo gene level read coverage tracks for OVO ChIP minus input (black), GSC and 32c ATAC-seq (light blue), GSC and 32c H3K27ac (green), H3K4me3 (dark blue), GSC and 32c H3K27me3 (orange), and GSC and 8c H3K9me3 (pink) ChIP-seq, and *ovo*^{ΔBP}/*ovo*^{ovo-GAL4}; *UASp-3x FHA-OVO-B* minus *ovo*^{ΔBP}/*ovo*^{ovo-GAL4}; *UASp-GFP* RNA-seq (red). Red rectangles and black rectangles represent significant OVO DNA binding motifs and OVO ChIP peaks, respectively. Gene models are represented at bottom. Small rectangles represent untranslated regions, large rectangles represent translated regions. Arrows indicate transcriptional start sites.

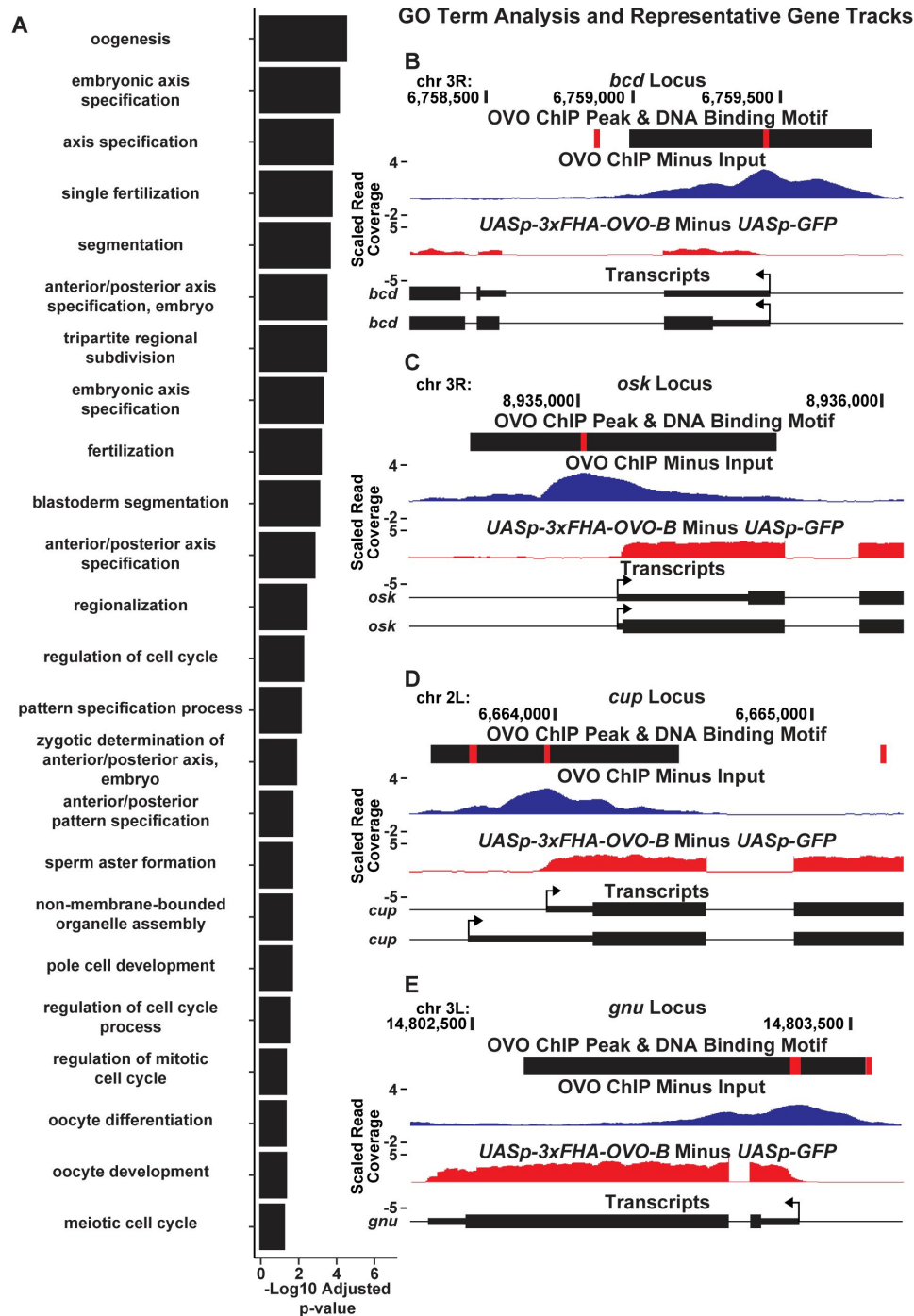


Figure 6

OVO Binds and Significantly Increases the Expression of a Number of Genes Involved in Essential Maternal Processes.

A) Significantly enriched GO biological process terms for genes bound by OVO and significantly increase in expression in the presence of ectopic rescue OVO. GO terms are restricted to the 24 smallest term sizes. B-E) Example GO term gene level read coverage tracks for OVO ChIP minus input and *ovo*^{ΔBP}/*ovo*^{ovo-GAL4}; *UASp-3xFHA-OVO-B* minus *ovo*^{ΔBP}/*ovo*^{ovo-GAL4}; *UASp-GFP*. Red rectangles and black rectangles represent significant OVO DNA binding motifs and OVO ChIP peaks, respectively. Gene models are represented at bottom. Small rectangles represent untranslated regions, large rectangles represent translated regions. Arrows indicate transcriptional start sites.

alleles and produces defective eggs allowing us to explore the role of OVO in late stages (Busson et al. 1983 [↗](#); Komitopoulou et al. 1983 [↗](#)). *ovo*^{D3}/*ovo*⁺ transheterozygous females express a repressive form of OVO that results in dominant sterility, and importantly, these females lay flaccid eggs with compromised vitelline membranes that are permeable to the dye neutral red (Oliver, Pauli, and Mahowald 1990 [↗](#)). Since OVO⁺ is bound at the TSS of *fs(1)N*, *fs(1)M3*, and *closca*, and these three genes respond transcriptionally to OVO⁺, then it is plausible that the repressive OVO^{D3} is negatively regulating these three genes that are required for vitelline membrane formation. This is evidence that OVO is not only involved in regulating the expression of numerous essential maternal pathways for embryonic development, but it is also essential for regulating genes that are required for egg integrity and maturation.

Discussion

Since its original isolation as a dominant female sterile locus in *Drosophila* (Busson et al. 1983 [↗](#); Komitopoulou et al. 1983 [↗](#)), *ovo* has long been known as an essential gene in oogenesis. Female germ cells require *ovo* for survival and differentiation, while it has no described roles or functions in the male germline (Oliver, Perrimon, and Mahowald 1987 [↗](#); Oliver, Pauli, and Mahowald 1990 [↗](#)). OVO has also been found to be eternally present in the female germline, attesting to its likely continual requirement for female germ cell viability and identity (Benner et al. 2023 [↗](#)). Our work here significantly expands our knowledge on OVO function in the female germline, showing that OVO binds and positively regulates a large array of genes required to build an egg and pattern the resulting embryo after fertilization. OVO accomplishes this by directly binding to the promoters of its targets, as well as more distant sites that could represent enhancers. Altogether, we suggest that OVO is a master transcriptional regulator coordinating a number of essential maternal pathways involved in oocyte and early embryonic development. Hints of these functions can be found in the hypomorphic and antimorphic *ovo* alleles which show egg chamber arrest, ventralized eggs, and permeable vitelline membranes. It is clear that OVO is required to activate multiple pathways involved in oocyte and early embryonic development.

The GO term enrichment analysis on genes that were bound by OVO and transcriptionally responded to OVO surprisingly indicated a large degree of overlap in oocyte and early embryonic developmental pathways. Also, OVO seemed to reinforce these pathways at multiple key genes within each pathway. For example, OVO bound to the promoters and increased the expression of *bcd*, and the *bcd* mRNA binding proteins *exu* and *swa*, all involved in ensuring correct anterior specification of the embryo (Lasko 2012 [↗](#)). Genes that are essential for egg activation were coordinately regulated by OVO as well. OVO downstream target genes *gnu*, *png*, *plu*, and *wisp* all belong in the same interconnected pathway ensuring egg activation (Avilés-Pagán and Orr-Weaver 2018 [↗](#)). A similar story was found for genes such as *cup*, *Me31b*, *bru1*, and *orb*, indicating that OVO controls a battery of genes involved in the positive regulation of RNA binding proteins that negatively regulate translation (Lasko 2012 [↗](#)). OVO also bound and positively regulated a number of posterior and germ plasm specification genes such as *osk*, *nos*, *aub*, *gcl*, and *pgc* (Lasko 2012 [↗](#); Mahowald 2001 [↗](#)). Given this plethora of famous maternal effect loci, it might be tempting to suggest that OVO is sufficient for egg production, but there are important exceptions. For example, other important germ plasm factors such as *staufen* and *tudor* were not bound by or transcriptionally responsive to OVO. This observation suggest that other transcription factors are responsible for regulating these genes.

OVO binds in close proximity to the TSS of genes it positively regulates, however, it is still unclear precisely how it regulates gene expression. The possibilities include integration into the RNA Polymerase complex itself, a short distance sigma factor like function, a core promoter conditioning function (pioneering), and/or garden variety transcription factor. Although active core promoters specifically in the ovary are enriched for OVO DNA binding motifs, we did not find a strict spatial orientation for these motifs in relation to the TSS, such as is found with other DNA

elements such as INR, DPE, and MTE (Ohler et al. 2002 [↗](#); FitzGerald et al. 2006 [↗](#); Lim et al. 2004 [↗](#)). It is therefore unlikely that OVO is a core component of the RNA Polymerase complex in the female germline. This suggests that it is unlikely to be analogous to male specific TATA-associated factors that have been shown to activate gene expression in the male germline (M. Hiller et al. 2004 [↗](#); M. A. Hiller et al. 2001 [↗](#); Lu et al. 2020 [↗](#); V. C. Li et al. 2009 [↗](#)). It is therefore possible, and previously well supported, that OVO is an activator of transcription (Lü et al. 1998 [↗](#); Lü and Oliver 2001 [↗](#); Bielinska et al. 2005 [↗](#)). One aspect of OVO DNA binding that showed differences with stage-specific histone ChIP and ATAC-seq, was OVO's strong association with open chromatin. The role of repressors of transcription such as the polycomb complex, *egg*, *wde*, and *Su(var)205* in restricting gene expression through promoting heterochromatin formation in differentiating egg chambers is well established (Smolko, Shapiro-Kulnane, and Salz 2018 [↗](#); DeLuca et al. 2020 [↗](#)). OVO might ensure that the chromatin status of maternally expressed genes remains open. Evidence from our work points to OVO fulfilling that role.

In GSCs OVO ChIP peaks largely overlap open chromatin and transcriptionally active histone marks. However, in stage 5 egg chambers, there was an even higher degree of association with open chromatin (almost all OVO ChIP peaks overlapped ATAC-seq peaks), while the significant association with H3K27ac marks was greatly reduced. This difference is likely significant. As GSCs differentiate they accumulate repressive chromatin marks while the number of ATAC-seq and H3K27ac peaks are reduced. This increase in association with open chromatin and OVO binding, even as the amount of open chromatin is reduced throughout egg chamber differentiation, might indicate that OVO binding helps to maintain chromatin accessibility, even when the locus in question is no longer actively transcribed. The loss of histone marks of active transcription at OVO-bound open chromatin in later differentiating egg chambers might mean that OVO does not influence the transcriptional potential of target genes as strongly as it influences the chromatin status in this second phase. Therefore, OVO might be more similar in function to pioneer factors/chromatin remodelers than it is to a transcription factor that is only involved in activating transcription.

The requirement for OVO at the TSS of target genes has been well characterized at its own locus as well as its downstream target *otu*. Our OVO ChIP and expression data confirm findings from previous work that OVO is binding to these target promoters, and in the case of *otu*, strongly responds transcriptionally to the presence of OVO. Although we did not test the requirement for OVO DNA binding motifs at other OVO bound genes in this work, this has been extensively explored before, showing that removal of OVO DNA binding sites overlapping the TSS results in a strong decrease in reporter expression (Lü et al. 1998 [↗](#); Bielinska et al. 2005 [↗](#); Lü and Oliver 2001 [↗](#)). Removal of more distal upstream OVO DNA binding sites also reduces reporter expression to a lesser degree. However, for most cases tested, removal of OVO DNA binding sites while leaving the rest of the enhancer regions intact, never totally abolished reporter expression. These dynamics are highly similar to work that has been completed on the pioneer factor *zelda* (*zld*). Adding *zld* DNA binding motifs to a stochastically expressed transcriptional reporter increases the activity and response of the reporter (Dufourt et al. 2018 [↗](#)). Distally located *zld* DNA binding motifs influenced reporter expression to a lesser degree than proximal sites. A single *zld* DNA binding site adjacent to the TSS produced the strongest reporter activity. Importantly, just like the activity of OVO transgenic reporters, there is not an absolute requirement for *zld* DNA binding to activate reporter expression, however, the addition of TSS adjacent *zld* DNA binding motifs does strongly influence reporter response. We know that *zld* achieves this reporter response through its pioneering activity (Xu et al. 2014 [↗](#); Harrison et al. 2011 [↗](#)), whether OVO achieves this similar effect on gene expression through a shared mechanism, or in cooperation with other transcription factors needs to be further explored.

Methods

All reagents used in this study can be found in the FlyBase recommended supplementary ART table (Table S6).

Fly Husbandry

All fly crosses were conducted at 25°C with 65% relative humidity and constant light unless otherwise noted. Flyfood consisted of premade flyfood (D20302Y) from Archon Scientific (Durham, NC).

Immunofluorescence and Image Analysis

All immunostaining procedures were done as previously described (Benner et al. 2023 [\[1\]](#)). Antibodies and their respective dilutions are indicated in Table S6. Imaging was also completed as previously described (Benner et al. 2023 [\[1\]](#)).

RNA-seq library preparation and sequencing

Twenty, one day old post-eclosion *ovo* ^{ΔBP} /*ovo*^{*ovo-GAL4*}, *UASp-GFP* and *ovo* ^{ΔBP} /*ovo*^{*ovo-GAL4*}, *UASp-3xFHA-OVO-B* ovaries were dissected and germariums through previtellogenic egg chambers were removed with microdissection scissors and placed in ice cold PBS making up one biological replicate. RNA was then extracted from four biological replicates with a Qiagen RNeasy Plus Kit (Qiagen) according to the manufacturer's protocol, eluted in dH₂O, and RNA concentrations were measured with Quant-iT RiboGreen RNA Assay Kit (ThermoFisher Scientific). 500 ng of total RNA was then used to make RNA-seq libraries with an Illumina Stranded mRNA Prep Kit according to the manufacturer's protocol (Illumina). IDT for Illumina RNA UD Indexes Set A were used. Library concentrations were measured with Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher Scientific), pooled, and then 50 nucleotide paired-end DNA sequencing was completed on an Illumina NovaSeq 6000 system using a S1 flow cell (Illumina). Raw RNA-seq reads are available at the SRA (SAMN38284748-SAMN38284755).

ChIP-seq library preparation and sequencing

Adult *ovo*^{*Cterm-3xFHA*} and *ovo*^{*Cterm-GFP*} females were collected and fed for 24 hours before ovaries were dissected. 50 dissected ovaries were placed in ice cold phosphate buffered saline (PBS, Gibco, ThermoFisher Scientific) and then incubated in 1 mL crosslinking solution containing 2% formaldehyde (Pierce, ThermoFisher Scientific) (50mM HEPES Buffer, 1mM EDTA, 0.5 mM EGTA, 100mM NaCl), and rotated at 37°C for 20 minutes. Ovaries were then incubated in 1 mL stop solution (125mM Glycine, 0.01% Triton X-100 (Millipore Sigma), diluted in PBS) and rotated for 5 minutes at room temperature.

Ovaries were then washed twice with 1 mL ice cold wash buffer (0.01% Triton X-100 in PBS) for 5 minutes. The last wash was removed, and ovaries were stored at -80°C until future processing. Once all samples were collected, 4×50 ovaries were then homogenized in 250 μ L RIPA lysis buffer (Pierce, ThermoFisher Scientific) containing 1x protease inhibitor cocktail (cOmplete Mini Protease Inhibitor Cocktail, Roche, Millipore Sigma) and 1 mM PMSF (Roche, Millipore Sigma) and kept on ice for 10 minutes. 40mg of 212-300 μ m acid-washed glass beads (Millipore Sigma) were then added to homogenized ovary lysate. Samples were then sonicated with a Bioruptor Pico sonication device (Diagenode) at 4C for 15 cycles of 30 seconds on and 30 seconds off. Sonicated lysate was then transferred to a new tube and centrifuged at 13,300 rpm for 10 minutes at 4°C. Three supernatants were then combined to form one biological replicate. 100 μ L for each biological replicate was removed and stored at -80°C for input control. To pull down C-terminally

tagged OVO, 100 μ L of monoclonal anti-HA-agarose (Millipore Sigma) or 50 μ L of ChromoTek GFP-Trap agarose (Proteintech) were washed three times with RIPA lysis buffer and spun down at 1,200 RPMs for one minute at 4°C. 550 μ L of *ovo*^{Cterm-3xFHA} supernatant was added to monoclonal anti-HA-agarose and 550 μ L of *ovo*^{Cterm-GFP} supernatant was added to ChromoTek GFP-Trap agarose. Samples were supplemented with 1x protease inhibitor cocktail and 1 mM PMSF and incubated on a rotator at 4°C overnight.

The next day, agarose was washed in a stepwise fashion with solutions from a Chromatin Immunoprecipitation Assay Kit (Millipore Sigma), beginning with 1 mL of a low salt wash buffer, high salt wash buffer, LiCl buffer, and ending with 2 washes in 0.1x TE buffer. 300 μ L of freshly prepared ChIP elution/decrosslinking solution (1% SDS, 100mM NaHCO₃, 250 mM NaCl, 10mM EDTA, 50 mM Tris-HCl, 200 μ g/mL Proteinase K) was added to the pelleted agarose, or 200 μ L of chip elution/decrosslinking solution was added to 100 μ L input control and incubated at 65°C overnight. DNA was extracted by adding 300 μ L phenol:chloroform:iso-amyl alcohol (125:24:1) (Millipore Sigma). The samples were vortexed for 30 seconds then centrifuged at 13,300 RPMs for 5 minutes at 4°C. The aqueous layer was extracted, and this process was repeated once more. 1 μ L glycogen (20 mg/mL), 30 μ L 1M sodium acetate, and 750 μ L 100% EtOH was added to the extracted aqueous layer, vortexed, and incubated at -20°C for 30 minutes. Solution was spun at 13,300 RPMs for 20 min at 4°C. Supernatant was removed and the pellet was washed with 500 μ L 70% EtOH and spun down at 13,300 RPMs for 20 minutes at 4°C. This step was repeated but with 100% EtOH. The resulting pellet was briefly speedvacced and resuspended in 50 μ L dH₂O.

To make ChIP-seq libraries, DNA concentration for immunoprecipitated and input control samples were measured with a Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher Scientific). 5 ng of DNA for each sample was then used with the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) and completed according to the manufacturer's protocol. ChIP-seq library concentrations were then measured with a Quant-iT PicoGreen dsDNA Assay Kit, pooled, and then 50 nucleotide paired-end DNA sequencing was performed on an Illumina NovaSeq 6000 system using the XP workflow (Illumina). Raw ChIP-seq reads are available at the SRA (SAMN38284736-SAMN38284747).

RNA-seq, ChIP-seq, CAGE-seq and Gene Ontology Analysis

For RNA-seq analysis of *ovo* ^{Δ BP/*ovo*^{*ovo*-GAL4}}; *UASp-GFP* and *ovo* ^{Δ BP/*ovo*^{*ovo*-GAL4}}; *UASp-3xFHA-OVO-B* ovaries, 50 nucleotide paired-end reads were mapped to the FlyBase r6.46 genome (Gramates et al. 2022) for differential expression analysis and the BDGP Release 6 Drosophila Genome (dos Santos et al. 2015 [\[1\]](#)) for read level genome browser tracks using Hisat2 (-k 1 --rna-strandness RF --dta)(Kim et al. 2019 [\[2\]](#)). DNA sequences for *GAL4* and *GFP* were added to the FlyBase r6.46 genome as separate chromosomes. Mapped reads were then sorted and indexed with Samtools (samtools sort and samtools index)(Daneczek et al. 2021 [\[3\]](#)). Gene level readcounts were then derived with htseq-count (-s reverse -r pos)(Anders, Pyl, and Huber 2015 [\[4\]](#)) and used for differential expression analysis with DESeq2 (Love, Huber, and Anders 2014 [\[5\]](#)). Genes with 0 mapped reads were removed from the DESeq2 analysis.

For ChIP-seq analysis of OVO-HA, OVO-GFP, OVO-HA input, and OVO-GFP input samples, 50 nucleotide paired-end reads were mapped to the FlyBase r6.46 genome for peak calling analysis and the BDGP Release 6 Drosophila Genome for read level genome browser tracks using Hisat2 (-k 1 --no-spliced-alignment -X 900). Mapped reads were sorted using Samtools (samtools sort and samtools index) and duplicate reads were removed with Picard (REMOVE_DUPLICATES=true) (broadinstitute, n.d.).

Significant ChIP peaks were called for OVO-HA and OVO-GFP versus their respective input controls separately using Macs3 callpeak software (-g 1.2e8 -q 0.0001)(Zhang et al. 2008 [\[6\]](#)). Overlapping ChIP peaks for OVO-HA and OVO-GFP were then determined with bedtools intersect software (Quinlan and Hall 2010 [\[7\]](#)). Peak calling for GSC ATAC-seq (SRR24203655), 32c ATAC-seq

(SRR24203650), GSC H3K27ac (SRR11084657), H3K4me3 (SRR11084658), H3K27me3 (SRR11084656), H3K9me3 (SRR24203631), 8c NC H3K4me3 (SRR24203629), 32c NC H3K27ac (SRR24203635), and H3K27me3 (SRR11084652) ChIP-seq versus their respective input controls (SRR11084655, SRR11084651, SRR24203634, SRR24203637) was conducted in the same manner as OVO ChIP-seq.

In order to generate gene-level read coverage tracks, deepTools' bamCompare software was used to generate a single bigWig file comparing all replicates versus input controls (-bs 5 --effectiveGenomeSize 142573017 --normalizeUsing BPM --exactScaling --scaleFactorsMethod None) (Ramírez et al. 2016 [\[1\]](#)). The bigWig file was then uploaded to UCSC genome browser for visualization (Kent et al. 2002).

To generate read coverage plots centered on the motif location or OVO peak maximums, genomic locations of significant scoring motifs or peak maximums within overlapping OVO ChIP peaks were determined and used as input for deepTools' computeMatrix reference-point (-a 2000 -b 2000 -bs 25 --missingDataAsZero). Read density profiles for each motif or OVO peak maximum were then visualized with deeptools plotProfile. In order to generate read coverage plots centered on the TSS, the same methods as above were conducted except genes overlapping OVO ChIP peaks containing the respective significant OVO DNA binding motifs were used as input instead.

In order to generate gene level read coverage heatmaps, deepTools' computeMatrix scale-regions software was used to generate a single matrix for genes that were bound by OVO (-bs 25 --missingDataAsZero -m 4000 --metagene). This matrix was then used as input for deepTools' plotHeatmap software to generate heatmaps of ChIP-seq and RNA-seq for the given OVO binding profiles centered on the TSS (--sortUsing max).

For CAGE-seq analysis, CAGE-seq libraries for ovary (SRR488283, SRR488282), testes (SRR488284, SRR488308, SRR488285, SRR488309), and male and female digestive system (SRR488289, SRR488288) tissues were downloaded and combined for each tissue type from the SRA. Reads were mapped to the BDGP Release 6 Drosophila Genome with Hisat2 (-k 1). Mapped reads were sorted with Samtools. Significant dominant TSSs were then determined with CAGER software (Haberle et al. 2015 [\[2\]](#)) from sorted BAM files with getCTSS and annotated with annotateCTSS using the dm6.ensGene.gtf file (Hubbard et al. 2002 [\[3\]](#)) downloaded from UCSC (Kent et al. 2002). CAGE-seq reads were normalized with normalizeTagCount (ce, method = "simpleTpm", fitInRange = c(5, 40000), alpha = 1.15, T = 1*10^6) and then TSS clusters were determined with clusterCTSS (ce, threshold = 1, thresholdIsTpm = TRUE, nrPassThreshold = 1, method = "paraclu", maxDist = 20, removeSingletons = TRUE, keepSingletonsAbove = 5) in order to determine the dominant significant TSS for each respective tissue.

Gene ontology enrichment analysis was completed with g:Profiler's g:GOST software (Raudvere et al. 2019 [\[4\]](#)) on the set of genes overlapping OVO ChIP peaks over the TSS and significantly upregulated in the presence of ectopic OVO (525 genes in total). All genes that were considered to be expressed in our RNA-seq datasets were used as a background control (10,801 genes in total). Default parameters were used for the enrichment analysis except for 'statistical domain scope' was set to 'custom' (our control background genes were uploaded here), 'significance threshold' was set to 'Bonferroni correction', and only GO biological process terms were searched for enrichment with the gene list. The GO terms listed in **Figure 6** [\[5\]](#) represent the 24 smallest GO term sizes according to Table S5.

Fisher's exact test was conducted for each respective analysis with the fisher.test() command in R (R Core Team 2021 [\[6\]](#)).

***de novo* Motif Enrichment and Promoter Motif Analysis**

DNA sequences from significant overlapping OVO ChIP peaks were extracted from the *Drosophila* r6.46 genome and submitted to STREME software (Bailey 2021 [↗](#)) of the MEME suite (Bailey et al. 2015 [↗](#)). The default parameters were used for *de novo* motif enrichment analysis, including the use of shuffled input sequences as a control. After identifying ‘OVO Motif One’, OVO ChIP peaks that contained that sequence were removed and the resulting ChIP peaks were resubmitted for STREME analysis deriving derivative OVO DNA binding motifs like above. Significant OVO DNA binding motifs and *in vitro* OVO DNA binding motifs were searched in the BDGP Release 6 *Drosophila* Genome using FIMO (Grant, Bailey, and Noble 2011 [↗](#)). In order to find significant DNA binding motif matches for ‘OVO Motif One’, this motif from STREME was submitted to Tomtom software (Gupta et al. 2007 [↗](#)) of the MEME suite and searched within the JASPAR Core Insect database (2022)(Castro-Mondragon et al. 2022 [↗](#)). XSTREME software was used with the default parameters to identify previously characterized and published DNA binding motifs within OVO ChIP peaks (Grant and Bailey 2021 [↗](#)).

Promoter motif analysis was conducted by extracting the DNA sequences 200 nucleotides upstream and downstream of the significant dominant TSSs from CAGE-seq analysis for each respective tissue type. All common core promoter motifs (FitzGerald et al. 2006 [↗](#); Ohler et al. 2002 [↗](#)) were then searched in these sequences depending on their strand specificity with the use of FIMO from the MEME suite using a p-value of < 0.003 for all non-OVO promoter motifs. All OVO motifs found in this study and through *in vitro* methods were also searched with the same method, except a p-value of < 0.0002 was used.

Data Availability

Drosophila strains used for this study are available upon request. All sequence information and datasets used in this study are in Table S6.

Acknowledgements

We would like to thank previous and current members of the Oliver lab, Laboratory of Biochemistry and Genetics at NIH, and L.B. committee members Mark Van Doren and Allan Spradling for insightful discussion and comments on this work throughout. Monoclonal antibodies were obtained from the Developmental Studies Hybridoma Bank, created by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) of the NIH and maintained at the Department of Biology, University of Iowa, Iowa City, IA 52242. Genetic and genomic information was obtained from FlyBase (U41 HG-000739). This work utilized the computational resources of the NIH High-Performance Computing Biowulf cluster (<http://hpc.nih.gov> [↗](#)). Sequencing was completed by The National Heart, Lung, and Blood Institute (NHLBI) DNA Sequencing and Genomics Core.

Funding

This research was supported in part by the Intramural Research Program of the NIH, The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) (awarded to B.O.). L.B. was supported by the NIH Graduate Partnership Program.

Table S1: OVO ChIP-seq Results by Chromosome.

Table S2: Genomic Locations of Significant Overlapping OVO ChIP Peaks

Table S3: Significant OVO DNA Binding Motifs in MEME Format

Table S4: Differential Expression Analysis Results for All Genes

Table S5: Significantly Enriched GO Biological Process Terms

Table S6: Flybase ART Table.

References

1. Anders Simon, Pyl Paul Theodor, Huber Wolfgang (2015) **HTSeq--a Python Framework to Work with High-Throughput Sequencing Data** *Bioinformatics* **31**:166–69
2. Andrews J., Garcia-Estefania D., Delon I., Lü J., Mével-Ninio M., Spierer A., Payre F., Pauli D., Oliver B. (2000) **OVO Transcription Factors Function Antagonistically in the Drosophila Female Germline** *Development* **127**:881–92
3. Aoyagi N., Wassarman D. A. (2000) **Genes Encoding Drosophila Melanogaster RNA Polymerase II General Transcription Factors: Diversity in TFIIA and TFIID Components Contributes to Gene-Specific Transcriptional Regulation** *The Journal of Cell Biology* **150**:F45–50
4. Avilés-Pagán Emir E., Orr-Weaver Terry L. (2018) **Activating Embryonic Development in Drosophila** *Seminars in Cell & Developmental Biology* **84**:100–110
5. Bailey Timothy L (2021) **STREME: Accurate and Versatile Sequence Motif Discovery** *Bioinformatics* **37**:2834–40
6. Bailey Timothy L., Johnson James, Grant Charles E., Noble William S. (2015) **The MEME Suite** *Nucleic Acids Research* **43**:W39–49
7. Bastock Rebecca, Johnston Daniel St (2008) **Drosophila Oogenesis** *Current Biology: CB* **18**:R1082–87
8. Benner L., Deshpande G., Lerit D. A. (2018) **“Primordial Germ Cells of Drosophila Melanogaster.”**
9. Benner Leif, Castro Elias A., Whitworth Cale, Venken Koen J. T., Yang Haiwang, Fang Junnan, Oliver Brian, Cook Kevin R., Lerit Dorothy A. (2019) **Drosophila Heterochromatin Stabilization Requires the Zinc-Finger Protein Small Ovary** *Genetics* **213**:877–95
10. Benner Leif, Muron Savannah, Oliver Brian (2023) **Female Germline Expression of OVO Transcription Factor Bridges Drosophila Generations** *bioRxiv* <https://doi.org/10.1101/2023.08.25.554887>
11. Bielinska Beata, Lü Jining, Sturgill David, Oliver Brian (2005) **Core Promoter Sequences Contribute to Ovo-B Regulation in the Drosophila Melanogaster Germline** *Genetics* **169**:161–72

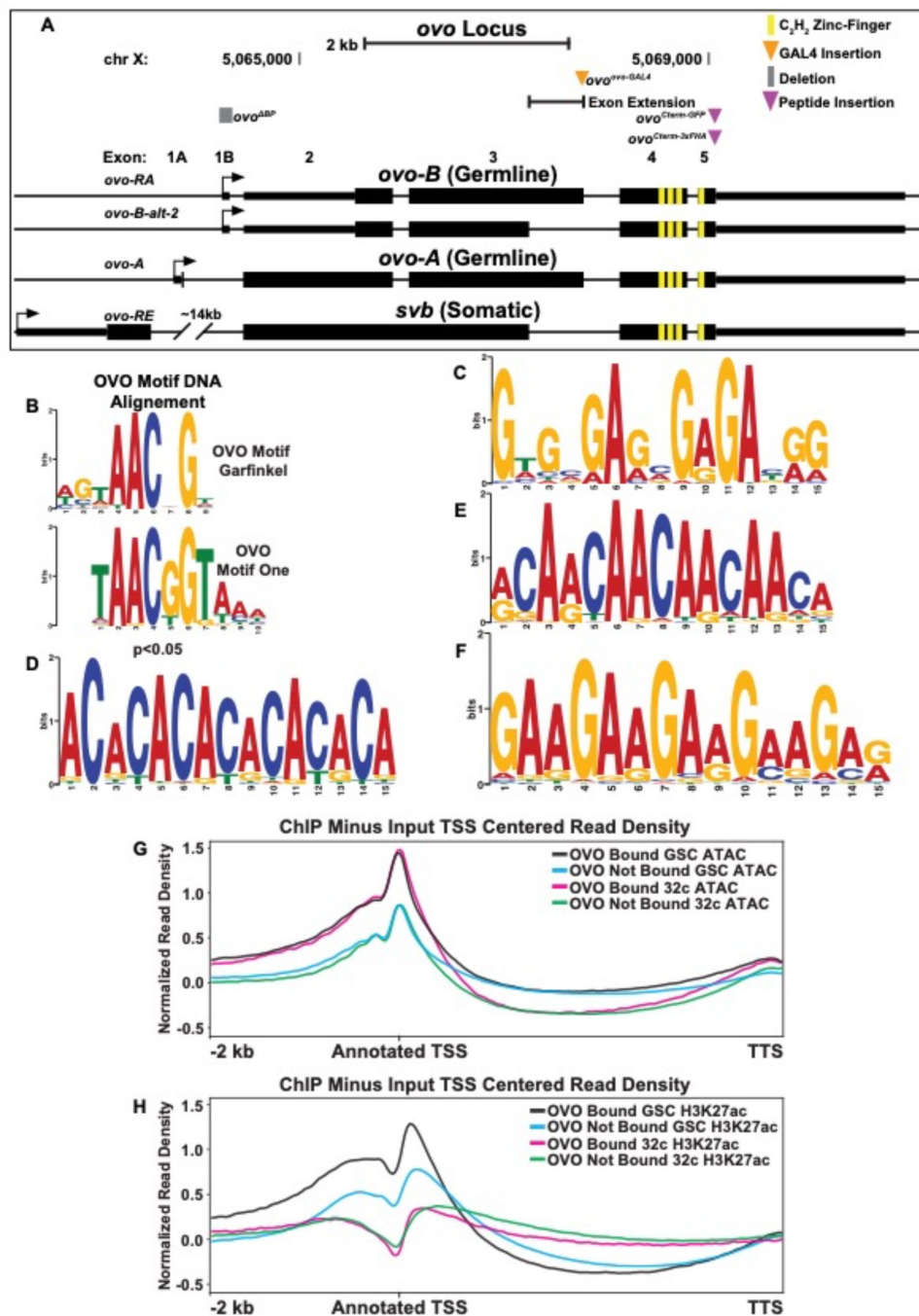


Figure S1

Significant OVO DNA Binding Motif.

A) Cartoon of the transcripts expressed from the *ovo* locus based on Benner et al. 2023 [\[1\]](#). Gray box represents endogenously generated deletions, orange triangle represents the location of the *T2A-GAL4-3xSTOP* insertion, purple triangles represent the location of small peptide insertions and yellow boxes represent the location of the shared C₂H₂ zinc-fingers. Small rectangles represent untranslated regions, large rectangles represent translated regions and arrows indicate TSSs. B) Significant alignment of the *in vivo* OVO DNA binding 'motif one' and *in vitro* OVO DNA binding motif (Lee and Garfinkel 2000 [\[2\]](#)). C-F) Significantly enriched DNA motifs within OVO ChIP peaks. C is a significant match for the DNA binding motif of GAF and CLAMP (Omelina et al. 2011 [\[3\]](#); Soruco et al. 2013 [\[4\]](#)). G, H. GSC and 32c ATAC-seq and H3K27ac ChIP-seq minus input control read coverage density for genes bound by OVO over the TSS, gene body, closest TSS in intergenic space, closest TSS for all, or not bound. Plots are centered on the TSS.

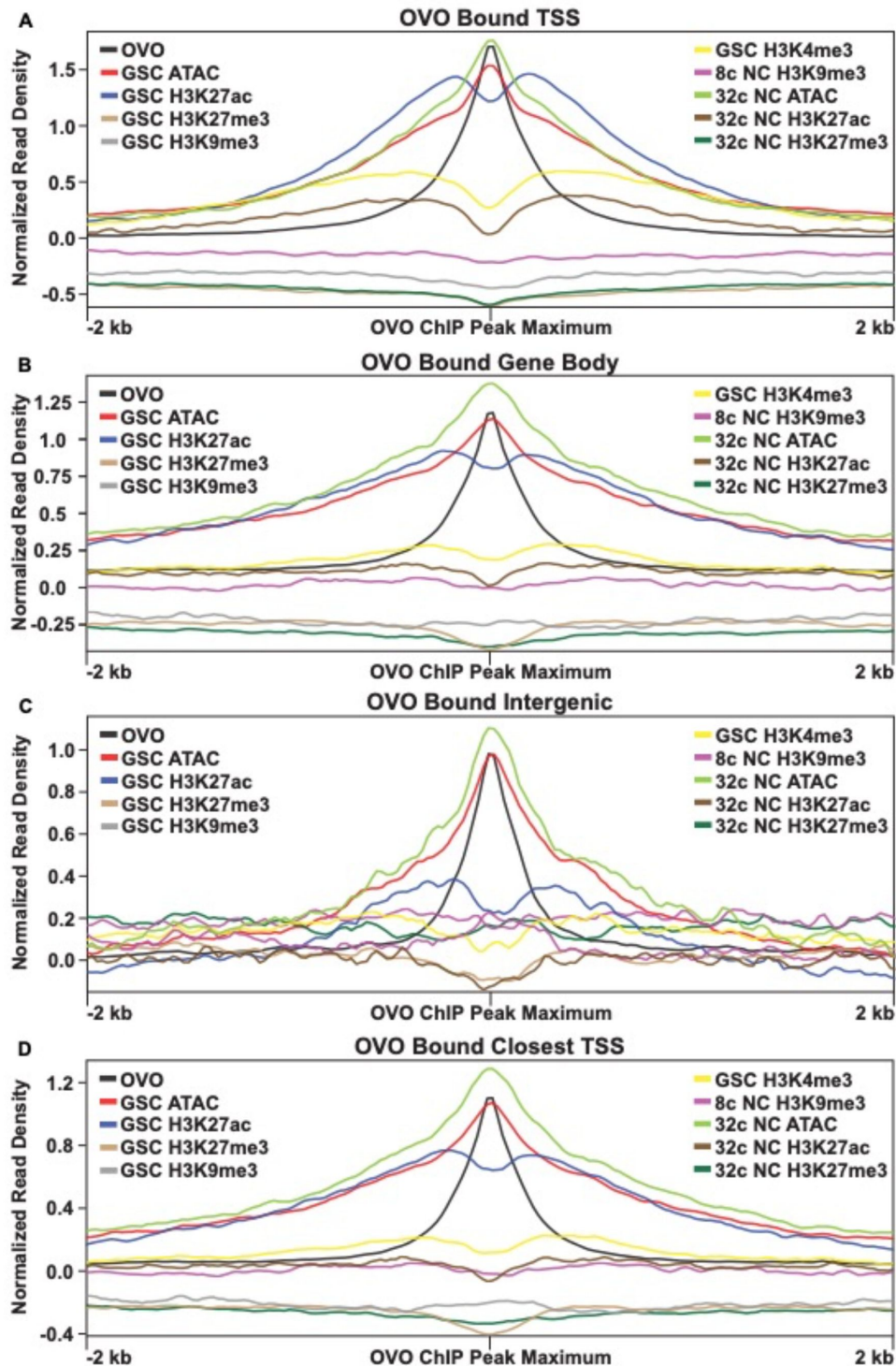


Figure S2

OVO DNA Binding is Associated with Open Chromatin and Transcriptionally Active Histone Marks Across Variations of Gene Binding Patterns.

A-D) OVO ChIP minus input control, GSC and 32c ATAC-seq, GSC H3K27ac, H3K4me3, H3K27me3, H3K9me3, 8c NC H3K9me3, 32c NC H3K27ac, and H3K27me3 ChIP-seq read coverage density centered on OVO peak maximums located within a significant OVO ChIP peak for genes bound by OVO over the TSS, gene body, closest TSS in intergenic space, or closest TSS for all.

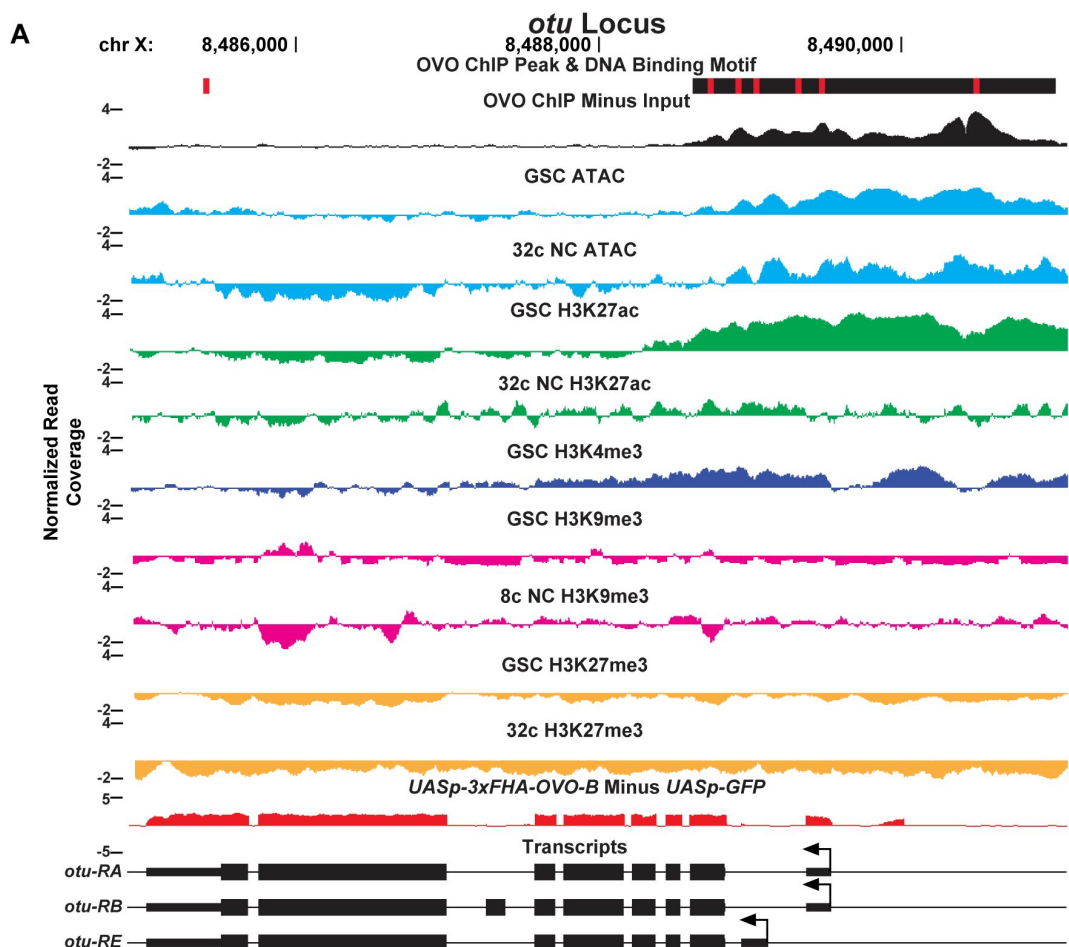


Figure S3

OVO ChIP-seq, ATAC/Histone ChIP-seq, RNA-seq, and DNA Binding Motifs at the *otu* Locus.

A) *otu* gene level read coverage tracks for OVO ChIP minus input (black), GSC and 32c ATAC-seq (light blue), GSC and 32c H3K27ac (green), H3K4me3 (dark blue), GSC and 32c H3K27me3 (orange), and GSC and 8c H3K9me3 (pink) ChIP-seq, and *ovo*^{ΔBP}/*ovo*^{ovo-GAL4}; *UASp-3xFHA-OVO-B* minus *ovo*^{ΔBP}/*ovo*^{ovo-GAL4}; *UASp-GFP* RNA-seq (red). Red rectangles and black rectangles represent significant OVO DNA binding motifs and OVO ChIP peaks, respectively. Gene models are represented at bottom. Small rectangles represent untranslated regions, large rectangles represent translated regions. Arrows indicate transcriptional start sites.

12. Bishop D. L., King R. C. (1984) **An Ultrastructural Study of Ovarian Development in the *otu7* Mutant of *Drosophila Melanogaster*** *Journal of Cell Science* **67**:87–119
13. broadinstitute. “**Picard.**” **Picard**
14. Busson D., Gans M., Komitopoulou K., Masson M. (1983) **Genetic Analysis of Three Dominant Female-Sterile Mutations Located on the X Chromosome of *DROSOPHILA MELANOGASTER*** *Genetics* **105**:309–25
15. Castro-Mondragon Jaime A., Riudavets-Puig Rafael, Rauluseviciute Ieva, Lemma Roza Berhanu, Turchi Laura, Blanc-Mathieu Romain, Lucas Jeremy, et al. (2022) **JASPAR 2022: The 9th Release of the Open-Access Database of Transcription Factor Binding Profiles** *Nucleic Acids Research* **50**:D165–73
16. Chen Yu-Chien, Lin Suewei I., Chen Ying-Kuan, Chiang Chuen-Sheue, Liaw Gwo-Jen (2009) **The Torso Signaling Pathway Modulates a Dual Transcriptional Switch to Regulate Tailless Expression** *Nucleic Acids Research* **37**:1061–72
17. Chen Zhen-Xia, Sturgill David, Qu Jiaxin, Jiang Huaiyang, Park Soo, Boley Nathan, Suzuki Ana Maria, et al. (2014) **Comparative Validation of the *D. Melanogaster* modENCODE Transcriptome Annotation** *Genome Research* **24**:1209–23
18. Danecek Petr, Bonfield James K., Liddle Jennifer, Marshall John, Ohan Valeriu, Pollard Martin O., Whitwham Andrew, et al. (2021) **Twelve Years of SAMtools and BCFtools** *GigaScience* **10** <https://doi.org/10.1093/gigascience/giab008>
19. DeLuca Steven Z., Ghildiyal Megha, Pang Liang-Yu, Spradling Allan C. (2020) **Differentiating *Drosophila* Female Germ Cells Initiate Polycomb Silencing by Regulating PRC2-Interacting Proteins** *eLife* **9** <https://doi.org/10.7554/eLife.56922>
20. Dufourt Jeremy, Trullo Antonio, Hunter Jennifer, Fernandez Carola, Lazaro Jorge, Dejean Matthieu, Morales Lucas, et al. (2018) **Temporal Control of Gene Expression by the Pioneer Factor Zelda through Transient Interactions in Hubs** *Nature Communications* **9**:1–13
21. FitzGerald Peter C., Sturgill David, Shyakhtenko Andrey, Oliver Brian, Vinson Charles (2006) **Comparative Genomics of *Drosophila* and Human Core Promoters** *Genome Biology* **7**
22. Gramates L. Sian, Agapite Julie, Attrill Helen, Calvi Brian R., Crosby Madeline A., Santos Gilberto Dos, Goodman Joshua L., et al. (2022) **FlyBase: A Guided Tour of Highlighted Features** *Genetics* **220** <https://doi.org/10.1093/genetics/iyac035>
23. Grant Charles E., Bailey Timothy L. (2021) **XSTREME: Comprehensive Motif Analysis of Biological Sequence Datasets** *bioRxiv* <https://doi.org/10.1101/2021.09.02.458722>
24. Grant Charles E., Bailey Timothy L., Noble William Stafford (2011) **FIMO: Scanning for Occurrences of a given Motif** *Bioinformatics* **27**:1017–18
25. Graveley Brenton (2010) **The Developmental Transcriptome of *Drosophila Melanogaster*** *Genome Biology* **11**
26. Gupta Shobhit, Stamatoyannopoulos John A., Bailey Timothy L., Noble William Stafford (2007) **Quantifying Similarity between Motifs** *Genome Biology* **8**

27. Haberle Vanja, Forrest Alistair R. R., Hayashizaki Yoshihide, Carninci Piero, Lenhard Boris (2015) **CAGEr: Precise TSS Data Retrieval and High-Resolution Promoterome Mining for Integrative Analyses** *Nucleic Acids Research* **43**
28. Haines Jenna E., Eisen Michael B. (2018) **Patterns of Chromatin Accessibility along the Anterior-Posterior Axis in the Early Drosophila Embryo** *PLoS Genetics* **14**
29. Harms E., Chu T., Henrion G., Strickland S. (2000) **The Only Function of Grauzone Required for Drosophila Oocyte Meiosis Is Transcriptional Activation of the Cortex Gene** *Genetics* **155**:1831–39
30. Harrison Melissa M., Li Xiao-Yong, Kaplan Tommy, Botchan Michael R., Eisen Michael B. (2011) **Zelda Binding in the Early Drosophila Melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition** *PLoS Genetics* **7**
31. Hayashi Makoto, Shinozuka Yuko, Shigenobu Shuji, Sato Masanao, Sugimoto Michihiko, Ito Seiji, Abe Kuniya, Kobayashi Satoru (2017) **Conserved Role of Ovo in Germline Development in Mouse and Drosophila** *Scientific Reports* **7**:1–10
32. Hiller M. A., Lin T. Y., Wood C., Fuller M. T. (2001) **Developmental Regulation of Transcription by a Tissue-Specific TAF Homolog** *Genes & Development* **15**:1021–30
33. Hiller Mark, Chen Xin, Jodeane Pringle M., Suchorolski Martin, Sancak Yasemin, Viswanathan Sridhar, Bolival Benjamin, Lin Ting-Yi, Marino Susan, Fuller Margaret T. (2004) **Testis-Specific TAF Homologs Collaborate to Control a Tissue-Specific Transcription Program** *Development* **131**:5297–5308
34. Hinson S., Pettus J., Nagoshi R. N. (1999) **Regulatory and Functional Interactions between Ovarian Tumor and Ovo during Drosophila Oogenesis** *Mechanisms of Development* **88**:3–14
35. Hubbard T., Barker D., Birney E., Cameron G., Chen Y., Clark L., Cox T., et al. (2002) **The Ensembl Genome Database Project** *Nucleic Acids Research* **30**:38–41
36. Sugnet Charles W., Furey Terrence S., Roskin Krishna M., Pringle Tom H., Zahler Alan M., Haussler David (2002) **The Human Genome Browser at UCSC** *Genome Research* **12**:996–1006
37. Kim Daehwan, Paggi Joseph M., Park Chanhee, Bennett Christopher, Salzberg Steven L. (2019) **Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype** *Nature Biotechnology* **37**:907–15
38. Komitopoulou K., Gans M., Margaritis L. H., Kafatos F. C., Masson M. (1983) **Isolation and Characterization of Sex-Linked Female-Sterile Mutants in DROSOPHILA MELANOGASTER with Special Attention to Eggshell Mutants** *Genetics* **105**:897–920
39. Lasko Paul (2012) **mRNA Localization and Translational Control in Drosophila Oogenesis** *Cold Spring Harbor Perspectives in Biology* **4** <https://doi.org/10.1101/cshperspect.a012294>
40. Lee S., Garfinkel M. D. (2000) **Characterization of Drosophila OVO Protein DNA Binding Specificity Using Random DNA Oligomer Selection Suggests Zinc Finger Degeneration** *Nucleic Acids Research* **28**:826–34
41. Li Jian, Gilmour David S. (2013) **Distinct Mechanisms of Transcriptional Pausing Orchestrated by GAGA Factor and M1BP, a Novel Transcription Factor** *The EMBO Journal* **32**:1829–41

42. Lim Chin Yan, Santoso Buyung, Boulay Thomas, Dong Emily, Ohler Uwe, Kadonaga James T. (2004) **The MTE, a New Core Promoter Element for Transcription by RNA Polymerase II** *Genes & Development* **18**:1606–17
43. Li Victor C., Davis Jerel C., Lenkov Kapa, Bolival Benjamin, Fuller Margaret T., Petrov Dmitri A. (2009) **Molecular Evolution of the Testis TAFs of Drosophila** *Molecular Biology and Evolution* **26**:1103–16
44. Li Xiao-Yong, Harrison Melissa M., Villalta Jacqueline E., Kaplan Tommy, Eisen Michael B. (2014) **Establishment of Regions of Genomic Activity during the Drosophila Maternal to Zygotic Transition** *eLife* **3**
45. Love Michael I., Huber Wolfgang, Anders Simon (2014) **Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2** *Genome Biology* **15**
46. Lu Dan, Sin Ho-Su, Lu Chenggang, Fuller Margaret T. (2020) **Developmental Regulation of Cell Type-Specific Transcription by Novel Promoter-Proximal Sequence Elements** *Genes & Development* **34**:663–77
47. Lü J., Andrews J., Pauli D., Oliver B. (1998) **Drosophila OVO Zinc-Finger Protein Regulates Ovo and Ovarian Tumor Target Promoters** *Development Genes and Evolution* **208**:213–22
48. Lü J., Oliver B. (2001) **Drosophila OVO Regulates Ovarian Tumor Transcription by Binding Unusually near the Transcription Start Site** *Development* **128**:1671–86
49. Mahowald A. P. (2001) **Assembly of the Drosophila Germ Plasm** *International Review of Cytology* **203**:187–213
50. Mineo Alessandro, Furriols Marc, Casanova Jordi (2017) **Transfer of Dorsoventral and Terminal Information from the Ovary to the Embryo by a Common Group of Eggshell Proteins in Drosophila** *Genetics* **205**:1529–36
51. Nakamura Shoichi, Hira Seiji, Fujiwara Masato, Miyagata Nasa, Tsuji Takuma, Kondo Akane, Kimura Hiroshi, et al. (2019) **A Truncated Form of a Transcription Factor Mamo Activates Vasa in Drosophila Embryos** *Communications Biology* **2**:1–13
52. Ohler Uwe, Liao Guo-Chun, Niemann Heinrich, Rubin Gerald M. (2002) **Computational Analysis of Core Promoters in the Drosophila Genome** *Genome Biology* **3**
53. Oliver B., Pauli D., Mahowald A. P. (1990) **Genetic Evidence That the Ovo Locus Is Involved in Drosophila Germ Line Sex Determination** *Genetics* **125**:535–50
54. Oliver B., Perrimon N., Mahowald A. P. (1987) **The Ovo Locus Is Required for Sex-Specific Germ Line Maintenance in Drosophila** *Genes & Development* **1**:913–23
55. Omelina E. S., Baricheva E. M., Yu Oshchepkov D., Merkulova T. I. (2011) **Analysis and Recognition of the GAGA Transcription Factor Binding Sites in Drosophila Genes** *Computational Biology and Chemistry* **35**:363–70
56. Pang Liang-Yu, DeLuca Steven, Zhu Haolong, Urban John M., Spradling Allan C. (2023) **Chromatin and Gene Expression Changes during Female Drosophila Germline Stem Cell Development Illuminate the Biology of Highly Potent Stem Cells** *eLife* **12**

57. Pauli D., Oliver B., Mahowald A. P. (1993) **The Role of the Ovarian Tumor Locus in *Drosophila Melanogaster* Germ Line Sex Determination** *Development* **119**:123–34
58. Quinlan Aaron R., Hall Ira M. (2010) **BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features** *Bioinformatics* **26**:841–42
59. Ramírez Fidel, Ryan Devon P., Grüning Björn, Bhardwaj Vivek, Kilpert Fabian, Richter Andreas S., Heyne Steffen, Dündar Friederike, Manke Thomas (2016) **deepTools2: A next Generation Web Server for Deep-Sequencing Data Analysis** *Nucleic Acids Research* **44**:W160–65
60. Raudvere Uku, Kolberg Liis, Kuzmin Ivan, Arak Tambet, Adler Priit, Peterson Hedi, Vilo Jaak (2019) **g:Profiler: A Web Server for Functional Enrichment Analysis and Conversions of Gene Lists (2019 Update)** *Nucleic Acids Research* **47**:W191–98
61. R Core Team (2021) **“R: A Language and Environment for Statistical Computing.”**
62. Rice G. A., Chamberlin M. J., Kane C. M. (1993) **Contacts between Mammalian RNA Polymerase II and the Template DNA in a Ternary Elongation Complex** *Nucleic Acids Research* **21**:113–18
63. Salles Cathy, Mével-Ninio Maryvonne, Vincent Alain, Payre François (2002) **A Germline-Specific Splicing Generates an Extended Ovo Protein Isoform Required for *Drosophila* Oogenesis** *Developmental Biology* **246**:366–76
64. Santos Gilberto dos, Schroeder Andrew J., Goodman Joshua L., Strelets Victor B., Crosby Madeline A., Thurmond Jim, Emmert David B., Gelbart William M., Consortium FlyBase (2015) **FlyBase: Introduction of the *Drosophila Melanogaster* Release 6 Reference Genome Assembly and Large-Scale Migration of Genome Annotations** *Nucleic Acids Research* **43**:D690–97
65. Smolko Anne E., Shapiro-Kulnane Laura, Salz Helen K. (2018) **The H3K9 Methyltransferase SETDB1 Maintains Female Identity in *Drosophila* Germ Cells** *Nature Communications* **9**:1–11
66. Soruco Marcela M. L., Chery Jessica, Bishop Eric P., Siggers Trevor, Tolstorukov Michael Y., Leydon Alexander R., Sugden Arthur U., et al. (2013) **The CLAMP Protein Links the MSL Complex to the X Chromosome during *Drosophila* Dosage Compensation** *Genes & Development* **27**:1551–56
67. Spradling Allan C., Niu Wanbao, Yin Qi, Pathak Madhulika, Maurya Bhawana (2022) **Conservation of Oocyte Development in Germline Cysts from *Drosophila* to Mouse** *eLife* **11**
68. Ventura Gemma, Furriols Marc, Martín Nicolás, Barbosa Vitor, Casanova Jordi (2010) **Closca, a New Gene Required for Both Torso RTK Activation and Vitelline Membrane Integrity. Germline Proteins Contribute to *Drosophila* Eggshell Composition** *Developmental Biology* **344**:224–32
69. Vo Ngoc Long, Kassavetis George A., Kadonaga James T. (2019) **The RNA Polymerase II Core Promoter in *Drosophila*** *Genetics* **212**:13–24
70. Xu Zhe, Chen Hongtao, Ling Jia, Yu Danyang, Struffi Paolo, Small Stephen (2014) **Impacts of the Ubiquitous Factor Zelda on Bicoid-Dependent DNA Binding and Transcription in *Drosophila*** *Genes & Development* **28**:608–21

71. Zhang Yong, Liu Tao, Meyer Clifford A., Eeckhoutte Jérôme, Johnson David S., Bernstein Bradley E., Nusbaum Chad, et al. (2008) **Model-Based Analysis of ChIP-Seq (MACS)** *Genome Biology* 9

Editors

Reviewing Editor

Yukiko Yamashita

Whitehead Institute/MIT, Cambridge, United States of America

Senior Editor

Utpal Banerjee

University of California, Los Angeles, Los Angeles, United States of America

Reviewer #1 (Public Review):

Summary:

In this manuscript, Benner et al. identify OVO as a transcriptional factor instrumental in promoting expression of hundreds of genes essential for female germline identity and early embryo development. Prior data had identified both ovo and otu as genes activated by OVO binding to the promoters. By combining ChIP-seq, RNA-seq and analysis of prior datasets, the authors extend these data to hundreds of genes and therefore propose that OVO is a master transcriptional regulator of oocyte development. They further speculate that OVO may function to promote chromatin accessibility to facilitate germline gene expression. Overall, the data compellingly demonstrate a much broader role for OVO in activation of genes in the female germline than previously recognized. By contrast, the relationship between OVO, chromatin accessibility and the timing of gene expression is only correlative, and more work will be needed to determine the mechanisms by which OVO promotes transcription.

Strengths

Here Benner et al. convincingly show that OVO is a transcriptional activator that promotes expression of hundreds of genes in the female germline. The ChIP-seq and RNA-seq data included in the manuscript are robust and the analysis is compelling.

Importantly, the set of genes identified are essential for maternal processes, including egg production and patterning of the early embryo. Together, these data identify OVO as a major transcriptional activator of the numerous genes expressed in the female germline, deposited into the oocyte and required for early gene expression. This is an important finding as this is an essential process for development and prior to this study the major drivers of this gene expression program were unknown.

Weaknesses

The novelty of the manuscript is somewhat limited as the authors show that, like two prior, well-studied OVO target genes, OVO binds to promoters of germline genes and activates transcription. The fact that OVO performs this function more broadly is not particularly surprising.

A major challenge to understanding the impact of this manuscript is the fact that the experimental system for the RNA-seq, the tagged constructs, and the expression analysis that provides the rationale for the proposed pioneering function of OVO are all included in a separate manuscript.

<https://doi.org/10.7554/eLife.94631.2.sa0>

Author response:

The following is the authors' response to the original reviews.

Public Reviews:**Reviewer #1 (Public Review):***Summary:*

In this manuscript, Benner et al. identify OVO as a transcriptional factor instrumental in promoting the expression of hundreds of genes essential for female germline identity and early embryo development. Prior data had identified both ovo and otu as genes activated by OVO binding to the promoters. By combining ChIP-seq, RNA-seq, and analysis of prior datasets, the authors extend these data to hundreds of genes and therefore propose that OVO is a master transcriptional regulator of oocyte development. They further speculate that OVO may function to promote chromatin accessibility to facilitate germline gene expression. Overall, the data compellingly demonstrate a much broader role for OVO in the activation of genes in the female germline than previously recognized. By contrast, the relationship between OVO, chromatin accessibility, and the timing of gene expression is only correlative, and more work will be needed to determine the mechanisms by which OVO promotes transcription.

We fully agree with this summary.

Strengths:

Here Benner et al. convincingly show that OVO is a transcriptional activator that promotes expression of hundreds of genes in the female germline. The ChIP-seq and RNA-seq data included in the manuscript are robust and the analysis is compelling.

Importantly, the set of genes identified is essential for maternal processes, including egg production and patterning of the early embryo. Together, these data identify OVO as a major transcriptional activator of the numerous genes expressed in the female germline, deposited into the oocyte and required for early gene expression. This is an important finding as this is an essential process for development and prior to this study, the major drivers of this gene expression program were unknown.

We are delighted that this aspect of the work came across clearly. Understanding the regulation of maternal effect genes has been something of a black-box, despite the importance of this class of genes in the history of developmental genetics. The repertoire of essential oogenesis/embryonic development genes that are bound by and respond to OVO are well characterized in the literature, but nothing is known about how they are transcriptionally regulated. We feel the manuscript will be of great interest to readers working on these genes.

Weaknesses:

The novelty of the manuscript is somewhat limited as the authors show that, like two prior, well-studied OVO target genes, OVO binds to promoters of germline genes and activates transcription. The fact that OVO performs this function more broadly is not particularly surprising.

Clearly, transcription factors regulate more than one or two genes. Never-the-less we were surprised at how many of the aspects of oogenesis *per se* and maternal effect genes were OVO targets. It was our hypothesis that OVO would have a transcriptional effect genome-wide,

however, it was less clear whether OVO would always bind at the core promoter, as is with the case of *ovo* and *otu*. Our results strongly support the idea that core promoter proximal binding is essential for OVO function; a conclusion of work done decades ago, which has not been revisited using modern techniques.

A major challenge to understanding the impact of this manuscript is the fact that the experimental system for the RNA-seq, the tagged constructs, and the expression analysis that provides the rationale for the proposed pioneering function of OVO are all included in a separate manuscript.

This is a case where we ended up with a very, very long manuscript which included a lot of revisiting of legacy data. It was a tough decision on how to break up all the work we had completed on *ovo* to date. In our opinion, it was too much to put everything into a single manuscript unless we wanted a manuscript length supplement (we were also worried that supplemental data is often overlooked and sometimes poorly reviewed). We therefore decided to split the work into a developmental localization/characterization paper and a functional genomics paper. As it stands both papers are long. Certainly, readers of this manuscript will benefit from reading our previous OVO paper, which we submitted before this one. The earlier manuscript is under revision at another journal and we hope that this improved manuscript will be published and accessible shortly.

Reviewer #2 (Public Review):

Summary:

In this manuscript, Benner et al. interrogate the transcriptional regulator OVO to identify its targets in the Drosophila germline. The authors perform ChIP-seq in the adult ovary and identify established as well as novel OVO binding motifs in potential transcriptional targets of OVO. Through additional bioinformatic analysis of existing ATAC-seq, CAGE-seq, and histone methylation data, the authors confirm previous reports that OVO is enriched at transcription start sites and suggest that OVO does not act as part of the core RNA polymerase complex. Benner et al. then perform bulk RNA-seq in OVO mutant and "wildtype" (GAL4 mediated expression of OVO under the control of the ovo promoter in OVO mutants) ovaries to identify genes that are differentially expressed in the presence of OVO. This analysis supports previous reports that OVO likely acts at transcription start sites as a transcriptional activator. While the authors propose that OVO activates the expression of genes that are important for egg integrity, maturation, and for embryonic development (nanos, gcl, pgc, bicoid), this hypothesis is based on correlation and is not supported by in vivo analysis of the respective OVO binding sites in some of the key genes. A temporal resolution for OVO's role during germline development and egg chamber maturation in the ovary is also missing. Together, this manuscript contains relevant ChIP-seq and RNA-seq datasets of OVO targets in the Drosophila ovary alongside thorough bioinformatic analysis but lacks important in vivo experimental evidence that would validate the high-quality datasets.

We thank reviewer 2 for the appreciation of the genomics data and analysis. Some of the suggested *in vivo* experiments are clear next steps, which are well underway. These are beyond the scope of the current manuscript.

Temporal analysis of *ovo* function in egg chamber development is not easy, as only the weakest *ovo* alleles have any egg chambers to examine. However, we will also point out the long-known phenotypes of some of those weak alleles in the text (e.g. ventralized chambers in *ovoD3/+*). We will need better tools for precise rescue/degradation during egg chamber maturation.

The manuscript contains relevant ChIP-seq and RNA-seq datasets of OVO targets in the Drosophila ovary alongside thorough bioinformatic analysis

Thank you. We went to great lengths to do our highly replicated experiments in multiple ways (e.g. independent pull-down tags) and spent considerable time coming up with an optimized and robust informatic analysis.

Weaknesses:

(1) The authors propose that OVO acts as a positive regulator of essential germline genes, such as those necessary for egg integrity/maturation and embryonic/germline development. Much of this hypothesis is based on GO term analysis (and supported by the authors' ChIP-seq data). However accurate interpretation of GO term enrichment is highly dependent on using the correct background gene set. What control gene set did the authors use to perform GO term analysis (the information was not in the materials and methods)? If a background gene set was not previously specified, it is essential to perform the analysis with the appropriate background gene set. For this analysis, the total set of genes that were identified in the authors' RNA-seq of OVO-positive ovaries would be an ideal control gene set for which to perform GO term analysis. Alternatively, the total set of genes identified in previous scRNA-seq analysis of ovaries (see Rust et al., 2020, Slaidina et al., 2021 among others) would also be an appropriate control gene set for which to perform GO term analysis. If indeed GO term analysis of the genes bound by OVO compared to all genes expressed in the ovary still produces an enrichment of genes essential for embryonic development and egg integrity, then this hypothesis can be considered.

We feel that this work on OVO as a positive regulator of genes like *bcd*, *osk*, *nos*, *png*, *gnu*, *plu*, etc., is closer to a demonstration than a proposition. These are textbook examples of genes required for egg and early embryonic development. Hopefully, this is not lost on the readers by an over-reliance on GO term analysis, which is required but not always useful in genome-wide studies.

We used GO term enrichment analysis as a tool to help focus the story on some major pathways that OVO is regulating. To the specific criticism of the reference gene-set, GO term enrichment analysis in this work is robust to gene background set. We will update the GO term enrichment analysis text to indicate this fact and add a table using expressed genes in our RNA-seq dataset to the manuscript and clarify gene set robustness in greater detail in the methods of the revision. We will also try to focus the reader's attention on the actual target genes rather than the GO terms in the revised text.

We have updated the GO term analysis by including all the expressed genes in our RNA-seq datasets as a background control. Figure 6 has been updated to include the significant GO terms. We have outlined changes in the methods section below.

Lines 794-801:

“Gene ontology enrichment analysis was completed with g:Profiler's g:GOST software (Raudvere et al. 2019) on the set of genes overlapping OVO ChIP peaks over the TSS and significantly upregulated in the presence of ectopic OVO (525 genes in total). All genes that were considered to be expressed in our RNA-seq datasets were used as a background control (10,801 genes in total). Default parameters were used for the enrichment analysis except for ‘statistical domain scope’ was set to ‘custom’ (our control background genes were uploaded here), ‘significance threshold’ was set to ‘Bonferroni correction’, and only GO biological process terms were searched for enrichment with the gene list. The GO terms listed in Figure 6 represent the 24 smallest GO term sizes according to Table S5.”

(2) The authors provide important bioinformatic analysis of new and existing datasets that suggest OVO binds to specific motifs in the promoter regions of certain germline genes. While the bioinformatic analysis of these data is thorough and appropriate, the authors do not perform any *in vivo* validation of these datasets to support their hypotheses. The authors should choose a few important potential OVO targets based on their analysis, such as *gcl*, *nanos*, or *bicoid* (as these genes have well-studied phenotypes in embryogenesis), and perform functional analysis of the OVO binding site in their promoter regions. This may include creating CRISPR lines that do not contain the OVO binding site in the target gene promoter, or reporter lines with and without the OVO binding site, to test if OVO binding is essential for the transcription/function of the candidate genes.

Exploring mechanism using *in vivo* phenotypic assays is awesome, so this is a very good suggestion. But, it is not essential for this work -- as has been pointed out in the reviews, *in vivo* validation of OVO binding sites has been comprehensively done for two target genes, *ovo* and *otu*. The “rules” appear similar for both genes. That said, we are already following up specific OVO target genes and the detailed mechanism of OVO function at the core promoter. We removed some of our preliminary *in vivo* figures from the already long current manuscript. We continue to work on OVO and expect to include this type of analysis in a new manuscript.

(3) The authors perform *de novo* motif analysis to identify novel OVO binding motifs in their ChIP-seq dataset. Motif analysis can be significantly strengthened by comparing DNA sequences within peaks, to sequences that are just outside of peak regions, thereby generating motifs that are specific to peak regions compared to other regions of the promoter/genome. For example, taking the 200 nt sequence on either side of an OVO peak could be used as a negative control sequence set. What control sequence set did the authors use as for their *de novo* motif analysis? More detail on this is necessary in the materials and methods section. Re-analysis with an appropriate negative control sequence set is suggested if not previously performed.

We apologize for being unclear on negative sequence controls in the methods. We used shuffled OVO ChIP-seq peak sequences as the background for the *de novo* motif analysis, which we will better outline in the methods of the revision. This is a superior background set of sequences as it exactly balances GC content in the query and background sequences. We are not fond of the idea of using adjacent DNA that won't be controlled for GC content and shadow motifs. Furthermore, the *de novo* OVO DNA binding motifs are clear, statistically significant variants of the characterized *in vitro* OVO DNA binding motifs previously identified (Lu et al., 1998; Lee and Garfinkel, 2000; Bielinska et al., 2005), which lends considerable confidence. We also show that the OVO ChIP-seq read density are highly enriched for all our identified motifs, as well as the *in vitro* motifs. We provide multiple lines of evidence, through multiple methods, that the core OVO DNA binding motif is 5'-TAACNGT-3'. We have high confidence in the motif data.

We have added the below text to the methods section for further clarity on motif analysis parameters.

Lines 808-812

“The default parameters were used for *de novo* motif enrichment analysis, including the use of shuffled input sequences as a control. After identifying ‘OVO Motif One’, OVO ChIP peaks that contained that sequence were removed and the resulting ChIP peaks were resubmitted for STREME analysis deriving derivative OVO DNA binding motifs like above.”

(4) The authors mention that OVO binding (based on their ChIP-seq data) is highly associated with increased gene expression (lines 433-434). How many of the 3,094 peaks (conservative OVO binding sites), and what percentage of those peaks, are associated with a significant increase in gene expression from the RNA-seq data? How many are associated with a decrease in gene expression? This information should be added to the results section.

Not including the numbers of the overlapping ChIP peaks and expression changes in the text was an oversight on our part. The numbers that relate to this (666 peaks overlapping genes that significantly increased in expression, significant enrichment according to Fishers exact test, 564 peaks overlapping genes that significantly decreased in expression, significant depletion according to Fishers exact test) are found in figure 4C and will be added to the text.

We have modified the results section to include the overlap between the RNA-seq and ChIP-seq data.

Lines 463-468

“We found that 2,298 genes that were expressed in our RNA-seq data overlapped an OVO ChIP peak. 666 genes significantly increased in expression and were bound by OVO, which is a significant enrichment according to a Fisher’s exact test (Figure 4C, cyan dots, $p < 0.01$, odds ratio = 2.21). While conversely, 564 genes decreased in expression and were bound by OVO, indicating a significant depletion according to a Fisher’s exact test (Figure 4C, blue dots, $p < 0.01$, odds ratio = 0.85).”

(5) The authors mention that a change in endogenous OVO expression cannot be determined from the RNA-seq data due to the expression of the OVO-B cDNA rescue construct. Can the authors see a change in endogenous OVO expression based on the presence/absence of OVO introns in their RNA-seq dataset? While intronic sequences are relatively rare in RNA-seq, even a 0.1% capture rate of intronic sequence is likely to be enough to determine the change in endogenous OVO expression in the rescue construct compared to the OVO null.

This is a good point. The *GAL4* transcript is downstream of *ovo* expression in the hypomorphic *ovoovo-GAL4* allele. We state in the text that there is a nonsignificant increase in *GAL4* expression with ectopic rescue OVO, although the trend is positive. We calculated the RPKM of RNA-seq reads mapping to the intron spanning exon 3 and exon 4 in *ovo-RA* and found that there is also a nonsignificant increase in intronic RPKM with ectopic rescue OVO (we will add to the results in the revision). We would expect OVO to be autoregulatory and potentially increase the expression of *GAL4* and/or intronic reads, but the *ovoovoGAL4>UASp-OVOB* is not directly autoregulatory like the endogenous locus. It is not clear to us how the intervening *GAL4* activity would affect *OVOB* activity in the artificial circuit. Dampening? Feed-forward? Is there an effect on *OVOA* activity? Regardless, this result does not change our interpretation of the other OVO target genes.

We have added the analysis of intronic *ovo* RNA-seq to the results as outlined below.

Lines 512-520

“Transcriptionally, *ovo* RNA-seq reads are likely derived from the *UASp-3xFHA-OVO-B* cDNA rescue or are indistinguishable between the genomic locus and rescuing cDNA transgene. We found a nonsignificant increase in exon 3 to exon 4 intronic *ovo* reads with the expression of ectopic rescue OVO (log2 fold change = 0.76, $p\text{-adj} = 0.26$). These intronic reads would be derived from the endogenous *ovo* locus, but it is difficult to conclusively determine if the endogenous *ovo* locus would respond transcriptionally to ectopic OVO downstream of *UASp*

(for example, the pathway for *ovo* is no longer autoregulatory in *ovoovo-GAL4/ovoΔBP; UASp-3x FHA-OVO-B* germ cells, there is an additional GAL4>UASp activation step). So, we could not confidently assess whether *ovo* responded transcriptionally to ectopic rescue OVO.”

(6) The authors conclude with a model of how OVO may participate in the activation of transcription in embryonic pole cells. However, the authors did not carry out any experiments with pole cells that would support/test such a model. It may be more useful to end with a model that describes OVO's role in oogenesis, which is the experimental focus of the manuscript.

We did not complete any experiments in embryonic pole cells in this manuscript and base our discussion on the potential dynamics of OVO transcriptional control and our previous work showing maternal and zygotic OVO protein localization in the developing embryonic germline. Obviously, we are highly interested in this question and continue to work on the role of maternal OVO. We agree that we are extended too far and will remove the embryonic germ cell model in the figure. We will instead focus on the possible mechanisms of OVO gene regulation in light of the evidence we have shown in the adult ovary, as suggested.

We have removed figure 7 and have re-written the last two paragraphs of the discussion as below.

Lines 645-663

“The requirement for OVO at the TSS of target genes has been well characterized at its own locus as well as its downstream target *otu*. Our OVO ChIP and expression data confirm findings from previous work that OVO is binding to these target promoters, and in the case of *otu*, strongly responds transcriptionally to the presence of OVO. Although we did not test the requirement for OVO DNA binding motifs at other OVO bound genes in this work, this has been extensively explored before, showing that removal of OVO

DNA binding sites overlapping the TSS results in a strong decrease in reporter expression (Lü et al. 1998; Bielinska et al. 2005; Lü and Oliver 2001). Removal of more distal upstream OVO DNA binding sites also reduces reporter expression to a lesser degree. However, for most cases tested, removal of OVO DNA binding sites while leaving the rest of the enhancer regions intact, never totally abolished reporter expression. These dynamics are highly similar to work that has been completed on the pioneer factor *zelda* (*zld*). Adding *zld* DNA binding motifs to a stochastically expressed transcriptional reporter increases the activity and response of the reporter (Dufourt et al. 2018). Distally located *zld* DNA binding motifs influenced reporter expression to a lesser degree than proximal sites. A single *zld* DNA binding site adjacent to the TSS produced the strongest reporter activity. Importantly, just like the activity of OVO transgenic reporters, there is not an absolute requirement for *zld* DNA binding to activate reporter expression, however, the addition of TSS adjacent *zld* DNA binding motifs does strongly influence reporter response. We know that *zld* achieves this reporter response through its pioneering activity (Xu et al. 2014; Harrison et al. 2011), whether OVO achieves this similar effect on gene expression through a shared mechanism, or in cooperation with other transcription factors needs to be further explored.”

Recommendations for the authors:

Reviewer #1 (Recommendations For The Authors):

The Results section could be streamlined by limiting the discussion of analysis to only those details that are unusual or essential for understanding the science. For example, the fact that MACS3 was used to call peaks seems most suitable for the Methods section.

We have removed the below excerpts from the results section to streamline the text.

‘We compared immuno-purified OVO associated DNA with input DNA as a control, for a total of 12 ChIPseq libraries, which we sequenced using the Illumina system. After quality control and alignment to the *Drosophila* r6.46 genome (Gramates et al. 2022), we used MACS3 (Zhang et al. 2008)’

The Supplemental Tables are referred to out of order. Table S2 is referred to on line 143 while Table S1 is not referred to until the Methods section.

We have reorganized the order of the tables in the manuscript text.

In the analysis of CAGE-seq data, it is unclear whether there is anything distinctive about the ~2000 regions bound by OVO but that is not near TSS in the ovary dataset. Are these TSS that are not active in the ovary or are these non-promoter bound OVO sites? If they are TSS of genes not in the CAGE-seq data set, are these genes expressed in other tissues or just expressed at lower levels in the ovary?

This was a good point that prompted us to take a closer look at the characteristics of OVO binding and its relationships to promoters and other gene elements. 45% of OVO ChIP peaks overlapped the TSS while 55% were either non-overlapping downstream or upstream of the TSS. When plotting OVO ChIP read density, there was still a striking enrichment of OVO binding over the TSS, even though the ChIP peak was not overlapping the TSS (new figure 1K). This is possibly due to weaker direct OVO binding at the TSS that was not considered significant in the peak calling software or were indirect interactions of the distal OVO binding and the TSS. We outline this in the below text added to the results section on the OVO ChIP. To showcase these results, we have included a new panel in figure 1K. We removed the panel showing the enrichment over the cage-seq TSS, but this same data remains in the heatmap shown in figure 1L, so no information is lost. To directly answer the Cage-seq questions considering the OVO bound over the annotated TSS results, we found that 1,047 chip peaks overlapped CAGE-seq TSS, which is only 347 fewer than the annotated TSS overlap (1,394). Of the 1,394 genes that were bound by over the TSS, all of them were considered to be expressed in our RNA-seq dataset, indicating that these might just be more lowly expressed genes that for whatever reason were not considered to be enriched TSSs in the CAGE-seq data. This difference is likely not significant.

Lines 235-251

“Although OVO ChIP peaks overlapping genes showed a strong read density enrichment over the TSS, we found that only 45% (1,394/3,094) of OVO ChIP peaks directly overlapped a TSS. 43% (1,339/3,094) of OVO ChIP peaks were found to overlap the gene body downstream of the TSS (intronic and exonic sequences) and 12% (366/3,094) did not overlap any gene elements, indicating that they were intergenic.

We were interested in the differences between OVO binding directly over the TSS or at more distal upstream and downstream sites. We decided to plot the OVO ChIP read density of these different classes of OVO binding patterns and found that OVO bound over the TSS produced a sharp read density enrichment over the TSS which was consistent with what was found for all OVO bound genes (Figure 1K). OVO binding along the gene body surprisingly also showed a read density enrichment over the TSS, although the magnitude of read density enrichment was notably less than TSS OVO binding. Intergenic OVO binding also showed these same characteristics with a notable upstream read density enrichment possibly indicative of enhancer binding. This indicates that although the significantly called OVO ChIP peaks did not overlap the TSS, there was still a propensity for TSS sequences to be enriched with OVO ChIP over the input control. This could be due to weaker direct *in vivo* binding of OVO to these TSSs or indirect interactions between the upstream/downstream OVO bound sequences

and the TSS, possibly through a looping enhancer-promoter interaction. However, regardless of the location of the OVO ChIP peak, OVO seemed to always be enriched at or in close proximity to TSSs.”

It would be helpful for the authors to provide a bit more detailed analysis of chromatin states of OVObound regions in GSC, 8c NC, and 32c NC (or some more clarity in the current analysis). Are the regions that are bound by OVO accessible in all these cell types or specifically enriched for accessibility in a subset? The authors state that OVO binding is correlated with open chromatin, but whether these are regions that are open in all cell types analyzed or a subset is not clear from the data presented. Promoters are often accessible regardless of cell type, so it is unclear what exactly is to be concluded from this association. Also, is the proximity to open chromatin features for OVO-bound promoters (as shown in Figure 2C) different than non-OVO-bound promoters (the two classes shown Figure 1L, for example)?

We utilized previously published datasets of staged germ cell chromatin status to look at the association of chromatin status and OVO binding. Unfortunately, not all the same germ cell stages were profiled for each chromatin mark from the datasets derived for these two papers. For example, only H3K4me3 data exists for GSCs, and only gsc and 8c data exists for H3K9me3, while the other chromatin marks had more profiles, even including later stages. We focused specifically on gsc and 32c (essentially stage 5 egg chambers) for the other chromatin marks since that is when the *ovo* hypomorphic egg chambers arrest. A nice control would have been chromatin states in somatic follicle cells of the ovary, since we know germ cell genes such as *ovo* and *otu* are not expressed and presumably the chromatin states in somatic cell types would be different than germ cells. However, chromatin states for somatic follicle cells were not published in these two papers and we are not aware of any other existing datasets to compare too. Essentially, we need to determine the changes in chromatin states with and without OVO, which we are currently working on.

We did further analyze chromatin states and differential OVO binding in respect to gene elements, and found that OVO binding, regardless of the relationship to the gene element, is always open (gsc and 32c ATAC). OVO binding over the gene body shows the same enrichment for open chromatin and transcriptionally active histone marks. We compared the profiles of these chromatin marks and the promoters of OVO bound and not bound genes and consistent with the suggestion that promoters are generally open, we found that this was the case. However, there is an enrichment for open chromatin and transcriptionally active histone marks for OVO bound genes compared to non-OVO bound genes. This could be a consequence of OVO binding or indirect consequence of a downstream OVO target. Regardless, as has been suggested, future experiments directly measuring chromatin status and OVO needs to be performed. The below excerpts have been added to the text to supplement the comments provided above.

Lines 328-343

“The association of OVO binding with active histone marks and open chromatin was striking, but open chromatin is likely a general phenomenon of promoters (Haines and Eisen, 2008). Indeed, when measuring the read density for GSC and 32C ATAC-seq for OVO bound and OVO non-bound promoters, there is an enrichment for open chromatin at the TSS regardless of OVO binding. However, we did notice an increase in enrichment for OVO bound promoters compared to OVO non-bound promoters (Figure S1G), possibly suggesting that OVO bound promoters are more open or have an increase in accessibility when compared to non-OVO bound promoters. This same relationship held true for the transcriptionally active histone mark H3K27ac in GSCs (Figure S1H). Since only 45% of OVO ChIP peaks overlapped TSSs, we plotted the read density of the above chromatin marks over OVO ChIP peak maximums for OVO bound over the TSS, gene body, or intergenic regions (Figure S2A-D). We found that OVO

bound regions that were not overlapping the TSS still showed the same propensity for enrichment of open chromatin and active histone marks. Intergenic regions were especially enriched for open chromatin measured through ATAC-seq. Altogether suggesting that OVO binding genome-wide is tightly associated with open chromatin regardless of germ cell stage, and active transcription in GSCs. In other words, chromatin state data suggests OVO is acting positively on its target genes and raises the possibility that OVO-binding and open chromatin are related.”

For clarity, it would help the reader if the authors mentioned the male-specific TATA-associated factors as a rationale for testing the role of OVO binding in core promoter function. This is currently mentioned in the Discussion on lines 575-577, but would help in understanding the motivation behind the detailed analysis of the promoter binding of OVO in the Results and make the negative result more clearly impactful.

We have introduced the male specific tata factors as suggested and have condensed the two intro paragraphs in this section into one, as shown below.

Lines 347-363

“Our data thus far clearly indicates that OVO binding occurs at or very near the core promoter, a region recognized by an enormous collection of factors that associate with RNA polymerase to initiate transcription (Aoyagi and Wassarman 2000; Vo Ngoc, Kassavetis, and Kadonaga 2019). The highly organized polymerase complex has sequence-specific DNA recognition sites with incredibly precise spacing between them, with an overall DNA footprint of a little less than 100bp (Rice, Chamberlin, and Kane 1993; FitzGerald et al. 2006; Ohler et al. 2002). There are upstream binding sites such as TATA, sites at transcription start, such as the initiator (INR), and downstream promoter elements (DPE) (Vo Ngoc, Kassavetis, and Kadonaga 2019). The combinations of these DNA motifs is not random in mammals and *Drosophila* (FitzGerald et al. 2006), and distinct combinations of different motifs at the TSS of genes expressed in *Drosophila* are conserved over tens of millions of years of evolution (Chen et al. 2014). The male germline expresses a number of TATA-associated factors that have been implicated in male-specific promoter usage for gene expression (M. Hiller et al. 2004; M. A. Hiller et al. 2001; Lu et al. 2020; V. C. Li et al. 2009). It is possible that OVO is a female germline specific TATA-associated factor, and if so, OVO binding sites at core promoters should share precise spacing with other core promoter elements, suggesting it is likely part of the complex. If not, then OVO is more likely to facilitate binding of the basal transcriptional machinery. Because of the extended footprint of engaged RNA polymerase, OVO and the basal machinery would not be likely to occupy the same region at the same time.”

The description of the system used for the RNA-seq would benefit from additional clarity. It is not clear as written why it is "Lucky" that there is an mRNA isoform with extended exon 2 required for egg chamber development beyond stage 5. How does this requirement compare to the global requirement for OVO, which seems to be required for germ cell development even before stage 5? Understanding this system is essential for interpreting the RNA-seq results. Indeed, the authors have a separate manuscript (currently on bioRxiv) that explains the details of this system. As such, the current description requires that the reader refer to this additional pre-print. Could the authors include a diagram to better illustrate this system? Furthermore, since this RNA-seq is being performed on tissue that includes nurse cells, follicle cells, and germ cells from multiple stages of development, it is important for the authors to clearly state in which cell types OVO is expressed and likely functional. (While this is well beyond this manuscript, this analysis is the type that might benefit from the use of single-cell sequencing as a means to deconvolute the phenotypic effects of OVO loss.)

We have rewritten the text to better describe the system for RNA-seq. We have also included a figure (Figure S1A) showing the alleles used that should help provide clarity for the readers. We agree that moving forward single cell experiments will be critical to have a better understanding of the transcriptional changes and chromatin dynamics with and without OVO. We have included the below changes to the text.

Lines 409-423

“Previous work from our lab has identified a transheterozygous *ovo* allelic combination (*ovoovo-GAL4/ovoΔBP*) that greatly reduces OVO activity resulting in sterility, however, female germ cells are able to survive up until at least stage 5 of oogenesis (Benner et al. 2023). *ovoovo-GAL4* is a CRISPR/Cas9 derived T2A-GAL43xSTOP insertion upstream of the splice junction of exon 3 in the *ovo-RA* transcript (Figure S1A).

Importantly, this insertion in the extended exon 3 would disrupt roughly 90% of the *ovo-B* transcripts. However, since about 10% of *ovo-B* transcripts utilize an upstream splice junction in exon 3, these transcripts would not be disrupted with the T2A-GAL4-3xSTOP insertion and thus allow for enough OVO activity for germ cell survival (Benner et al. 2023). Since *ovoovo-GAL4* expresses GAL4 in place of full length OVO due to the T2A sequences, we can drive expression of a rescuing OVO-B construct downstream of *UASp* to generate OVO+ female germ cells, which in fact does rescue the arrested germ cell phenotype of *ovoovo-GAL4/ovoΔBP* ovaries. Therefore, in order to determine genes that are transcriptionally responsive to OVO, we compared the gene expression profiles in sets of ovaries that had the *ovo* hypomorphic phenotype with a negative control rescue construct (*ovoovo-GAL4/ovoΔBP; UASp-GFP*)(Figure 4A) versus those that drive expression of the rescue construct expressing OVO-B (*ovoovo-GAL4/ovoΔBP; UASp-3xFHAOVO-B*)(Figure 4B).”

Lines 427-432

“The adult female ovary contains somatic cells, germline stem cells, and germline derived nurse cells that would be profiled in a bulk ovary tissue RNA-seq experiment. Although OVO is only required and expressed in germline derived cell types, we chose to dissect one day old post-eclosion *ovoovoGAL4/ovoΔBP; UASp-3xFHA-OVO-B* female ovaries to enrich for early stages of oogenesis and collected only ovarioles containing the germarium through previtellogenic egg chambers.”

*On lines 526-532, it is unclear why the genes *fs(1)N*, *fs(1)M3*, and *closca* are particularly sensitive to the *ovoD3* allele. What is this allele trans heterozygous with in the assay that allows development through egg laying? Why might these genes be unique in their sensitivity?*

These genes are not particularly sensitive, the transheterozygous hypomorphic *ovo* ovaries are weak enough to reveal the role of OVO for these genes. We rewrote this paragraph to try and provide more clarity to the relationship between OVO+ binding at these vitelline membrane genes and the phenotype of OVOD3 expressing females.

Lines 562-577

“We also found that the genes *fs(1)N*, *fs(1)M3*, and *closca*, were all bound by OVO and responded transcriptionally to the presence of ectopic rescue OVO. These genes are significant because they constitute a set of genes that are expressed in the germline and the encoded proteins are eventually incorporated into the vitelline membrane providing the structural integrity and impermeability of the egg (Mineo, Furriols, and Casanova 2017; Ventura et al. 2010). Loss-of-function of these three genes results in flaccid eggs that are permeable to dye and fail to develop. The loss-of-function phenotype of *fs(1)N*, *fs(1)M3*, and *closca* closely resembles the dominant antimorph *ovoD3* phenotype. The *ovoD3* allele is the

weakest of the original dominant-negative *ovo* alleles and produces defective eggs allowing us to explore the role of OVO in late stages (Busson et al. 1983; Komitopoulou et al. 1983). *ovoD3/ovo+* transheterozygous females express a repressive form of OVO that results in dominant sterility, and importantly, these females lay flaccid eggs with compromised vitelline membranes that are permeable to the dye neutral red (Oliver, Pauli, and Mahowald 1990). Since OVO+ is bound at the TSS of *fs(1)N*, *fs(1)M3*, and *closca*, and these three genes respond transcriptionally to OVO+, then it is plausible that the repressive OVO^{D3} is negatively regulating these three genes that are required for vitelline membrane formation. This is evidence that OVO is not only involved in regulating the expression of numerous essential maternal pathways for embryonic development, but it is also essential for regulating genes that are required for egg integrity and maturation.”

*The Discussion of OVO as a pioneer factor is highly speculative and based only on correlative data. In fact, the expression data in the embryonic germline is not included in this manuscript, but rather in a separate bioRxiv preprint. This makes it challenging to understand, why this is extensively discussed here. However, there are experiments that could begin to test this proposal. OVO could be expressed in an exogenous tissue and test whether it promotes accessibility. Also, mutations could be made (using gene editing) to identify previously known OVO binding sites in the *otu* and/or other promoters and these could be assayed for accessibility. By selecting promoters of genes that are not essential for germline development, the authors could directly test the role of OVO in promoting chromatin accessibility. Alternatively, are there reasons that the system used for RNA-seq couldn't be similarly used for ATACseq? It is imperfect but could provide insights into chromatin accessibility in the absence of OVO.*

We have largely removed the speculation on pioneering activity, reference to embryonic germline OVO dynamics included in the previous work, and Figure 7. These are excellent suggestions for experiments and ones we are currently pursuing. Below is the modified discussion.

Lines 645-663

“The requirement for OVO at the TSS of target genes has been well characterized at its own locus as well as its downstream target *otu*. Our OVO ChIP and expression data confirm findings from previous work that OVO is binding to these target promoters, and in the case of *otu*, strongly responds transcriptionally to the presence of OVO. Although we did not test the requirement for OVO DNA binding motifs at other OVO bound genes in this work, this has been extensively explored before, showing that removal of OVO

DNA binding sites overlapping the TSS results in a strong decrease in reporter expression (Lü et al. 1998; Bielinska et al. 2005; Lü and Oliver 2001). Removal of more distal upstream OVO DNA binding sites also reduces reporter expression to a lesser degree. However, for most cases tested, removal of OVO DNA binding sites while leaving the rest of the enhancer regions intact, never totally abolished reporter expression. These dynamics are highly similar to work that has been completed on the pioneer factor *zelda* (*zld*). Adding *zld* DNA binding motifs to a stochastically expressed transcriptional reporter increases the activity and response of the reporter (Dufourt et al. 2018). Distally located *zld* DNA binding motifs influenced reporter expression to a lesser degree than proximal sites. A single *zld* DNA binding site adjacent to the TSS produced the strongest reporter activity. Importantly, just like the activity of OVO transgenic reporters, there is not an absolute requirement for *zld* DNA binding to activate reporter expression, however, the addition of TSS adjacent *zld* DNA binding motifs does strongly influence reporter response. We know that *zld* achieves this reporter response through its pioneering activity (Xu et al. 2014; Harrison et al. 2011), whether OVO achieves this similar effect on gene expression through a shared mechanism, or in cooperation with other transcription factors needs to be further explored.”

The authors suggest that OVO binding is essential for transcriptional activation, but that this may be indirect and that expression of other transcription factors might be necessary for activating gene expression. Did the motif analysis of the OVO-bound regions suggest additional transcription factors that might provide this function?

We did find other motifs significantly enriched in OVO ChIP peaks. We performed XSTREME analysis on the same set of OVO ChIP peaks which allowed us to determine if any of these motifs were significant matches to DNA binding motifs of known transcription factors. Notably, the DNA binding motifs of GAF and CLAMP were enriched in OVO ChIP peaks. GAF is required in germline clones and the potentially for co-regulation of genes is possible. Other enriched motifs did not match any known binding motifs of other transcription factors but we reported some of the most significantly enriched motifs that were alongside of OVO in Figure S1C-F. The below text outlines changes made to the text incorporating these findings.

Lines 170-182

“Along with the OVO DNA binding motif, other motifs were also significantly enriched in OVO ChIP peaks. The motif 5'-GWGMGAGMGAGABRG-3' (Figure S1C) was found in 18% of OVO ChIP peaks and is a significant match to the DNA binding motifs of the transcription factors GAF (*Trl*) (Omelina et al. 2011) and CLAMP (Soruco et al. 2013). *Trl* germline clones are not viable, indicating that GAF activity is required in the germline during oogenesis (Chen et al. 2009). The possibility that OVO binds with and regulates genes alongside of GAF given the enrichment of both transcription factors DNA binding motifs is intriguing. Other significantly enriched motifs 5'-ACACACACACACACA-3' (29% of peaks, Figure S1D), 5'RCAACAACAACAACA-3' (26% of peaks, Figure S1E), and 5'-GAAGAAGAAGAAGAR-3' (17% of peaks,

Figure S1F) were present in OVO ChIP peaks, however, these motifs did not significantly match known

DNA binding motifs of other transcription factors. Determining the factors that bind to these sequences

will certainly help elucidate our understanding of transcriptional control with relationship to OVO in the female germline.”

The figures would benefit from a bit more detail in the legends (see comments below).

Minor comments:

In multiple places throughout the document, the citations are inadvertently italicized (see lines 57-59, 91, and 327 as examples.)

We have changed this in these locations and other instances in the text.

On line 76, when discussing OVO as a transcription factor this is referencing the protein and not the gene. Thus, should be written OVO and not ovo.

We have made the correction ovo to OVO.

On line 349, "core" promoters is likely what is meant rather than "care" promoters.

We have corrected 'care' to 'core' in the text.

On line 404, the authors state that they wanted to use a "less conservative log2 fold change" but it is not clear what they are comparing to. This is important to understand

| the motivation.

We are talking about the gene expression comparison between the ectopic *ovo* rescue and *ovo* hypomorphic ovaries. “less conservative” was an unfortunate phrasing. We have rewritten the text to state this directly to the reader.

Lines 435-444

“We then performed RNA-seq in quadruplicate and measured the changes in gene expression between ectopic rescue OVO and hypomorphic OVO ovaries. We used a significance level of $p\text{-adj} < 0.05$ and a \log_2 fold change cutoff of $> |0.5|$ to call differential expression between these two sets of ovaries. We utilized these \log_2 fold change cutoffs for two reasons. Our control ovary genotype (*ovoovo-GAL4/ovoΔBP; UASp-GFP*) has hypomorphic OVO activity, hence germ cells can survive but are arrested. With the addition of ectopic rescue OVO in *ovoovo-GAL4/ovoΔBP; UASp-3xFHA-OVO-B* ovaries, we predicted that genes that were directly regulated by OVO would transcriptionally respond, however, we were unsure as to what degree the response would be in comparison to hypomorphic OVO. We reasoned that if the changes were not significant between genotypes, then minor changes in gene expression would not matter.”

| On line 615, it is unclear what is meant by “showing expression with only 10s of bp of sequence in reporters.”

This is in reference to some of the previously studied *ovo* reporter deletion lines, however, we have decided to remove the below text in the revised discussion.

“, despite being remarkably compact. The OVO-dependent *ovo* core promoter is very compact; showing expression with only 10s of bp of sequence in reporters.”

| It would be useful to cite and discuss Dufourt et al. Nature Communications 2018 (PMID30518940) regarding the role of Zelda in potentiating transcriptional activation when mentioned on line 624.

We have added this and the relationship to previous similar work on OVO in the discussion.

Lines 645-663

“The requirement for OVO at the TSS of target genes has been well characterized at its own locus as well as its downstream target *otu*. Our OVO ChIP and expression data confirm findings from previous work that OVO is binding to these target promoters, and in the case of *otu*, strongly responds transcriptionally to the presence of OVO. Although we did not test the requirement for OVO DNA binding motifs at other OVO bound genes in this work, this has been extensively explored before, showing that removal of OVO

DNA binding sites overlapping the TSS results in a strong decrease in reporter expression (Lü et al. 1998; Bielinska et al. 2005; Lü and Oliver 2001). Removal of more distal upstream OVO DNA binding sites also reduces reporter expression to a lesser degree. However, for most cases tested, removal of OVO DNA binding sites while leaving the rest of the enhancer regions intact, never totally abolished reporter expression. These dynamics are highly similar to work that has been completed on the pioneer factor *zelda* (*zld*). Adding *zld* DNA binding motifs to a stochastically expressed transcriptional reporter increases the activity and response of the reporter (Dufourt et al. 2018). Distally located *zld* DNA binding motifs influenced reporter expression to a lesser degree than proximal sites. A single *zld* DNA binding site adjacent to the TSS produced the strongest reporter activity. Importantly, just like the activity of OVO transgenic reporters, there is not an absolute requirement for *zld* DNA binding to activate reporter expression, however, the addition of TSS adjacent *zld* DNA

binding motifs does strongly influence reporter response. We know that *zld* achieves this reporter response through its pioneering activity (Xu et al. 2014; Harrison et al. 2011), whether OVO achieves this similar effect on gene expression through a shared mechanism, or in cooperation with other transcription factors needs to be further explored.”

| On line 1006 (Figure 1 legend), it is unclear what is meant by "The percentage of OVO ChIP peaks each motif was found". Is a word missing?

This was unclear, we have revised the sentence below.

Lines 1035-1036

“The percentage of OVO ChIP peaks containing each motif and their corresponding p-value are indicated to the right.”

| In the Figure 1 legend, please include citations for the Garfinkel motif and Oliver motif.

Included, as below.

Lines 1036-1039

“H) OVO ChIP minus input control ChIP-seq read coverage density centered on the location of the four *de novo* OVO DNA binding motifs and previously defined *in vitro* OVO DNA binding motifs (Lü et al. 1998, Bielinska et al. 2005, Lee and Garfinkel 2000).”

| In Figure 2 legend, it is unclear if B is all instances of a given motif or the DNA motifs that are bound by ChIP. Please clarify.

We meant only the OVO DNA binding motifs that were within significant OVO ChIP peaks. We have revised the legend below.

Lines 1049-1052

“A, B) OVO ChIP minus input control, GSC and 32c ATAC-seq, GSC H3K27ac, H3K4me3, H3K27me3, H3K9me3, 8c NC H3K9me3, 32c NC H3K27ac, and H3K27me3 ChIP-seq read coverage density centered on each OVO peak maximum or OVO DNA binding motif located within a significant OVO ChIP peak.”

| The Figure legend for 2D could use more explanation. What do the lines and circles indicate?

These lines and circles indicate the amount of overlapping peaks measured between the two datasets with solid circles. We have included a better description of what these indicate in the figure legend.

Lines 1054-1058

“D) Total number of significant peaks (left) and the total number of overlapping peaks (top) between OVO

ChIP and GSC and 32c ATAC-seq, GSC H3K27ac, H3K4me3, H3K27me3, H3K9me3, 8c NC H3K9me3, 32c NC H3K27ac, and H3K27me3 ChIP-seq. Lines connecting solid dots indicates the amount of overlapping peaks between those two corresponding datasets.”

| In Figure 4C, bring the 564 blue dots forward so they are not masked by the yellow dots.

We have brought the colored dots forward in both figure 4C and 4D.

| *In Figure 4E, what is the order of the heatmaps?*

The order is genes with the highest to lowest OVO read density enrichment. We have included this in the figure 4 legend.

Lines 1086-1087

“The order of the heatmap is genes with the highest to lowest amount of OVO ChIP read density.”

| *In Figure 5, the order of the tracks is not immediately obvious. It appears to be those chromatin features most associated with OVO ChIP and those less correlated. Additional clarity could be provided by showing these tracks (and in Supplemental Figure S2) in different colors with a reference to the figure legend about what the colors might indicate.*

We have changed the colors and order of the tracks to be more similar and consistent in both figures.

Lines 1090-1093

“ovo gene level read coverage tracks for OVO ChIP minus input (black), GSC and 32c ATAC-seq (light blue), GSC and 32c H3K27ac (green), H3K4me3 (dark blue), GSC and 32c H3K27me3 (orange), and GSC and 8c H3K9me3 (pink) ChIP-seq, and *ovo*ΔBP/*ovoovo*-GAL4; *UASp*-3x*FHA*-*OVO-B* minus *ovo*ΔBP/*ovoovo*-GAL4; *UASp*-GFP RNA-seq (red).”

| *In Figure S1 legend, what is the reference to the da-GAL4 X UAS transgene in the title?*

This was an error on our part and we have removed it.

Reviewer #2 (Recommendations For The Authors):

Overall, the manuscript would benefit from revisions of the writing style. At times it is difficult to distinguish between hypothesis and results. The use of colloquial phrases/prose was distracting while reading, which the authors may consider revising. Some sentences were confusing or extraneous, and the authors may consider revising those. Occasionally sentences within the results sections seem more appropriate for the materials and methods.

(1) The manuscript is generally clear; however, it is at times difficult to distinguish between hypothesis and results. The use of colloquial phrases/prose was distracting while reading, which the authors may consider revising. Examples include:

a) Lines 48-49 "While thematic elements of this complex orchestration have been well studied, coordinate regulation of the symphony has not."

We have edited this sentence below.

Lines 48-50

“While the complex interactions between maternally supplied mRNAs and proteins have been well studied, transcriptional regulation driving the expression of these pathways are less well understood.”

| *b) Lines 232-233 "In other words, where exactly does transcription start at these genes."*

We have removed this sentence.

| c) Line 385, the word "sham" could be changed to "negative control" or "GFP control"

We have rewritten this sentence below.

Lines 419-423

"Therefore, in order to determine genes that are transcriptionally responsive to OVO, we compared the gene expression profiles in sets of ovaries that had the *ovo* hypomorphic phenotype with a negative control rescue construct (*ovoovo-GAL4/ovoΔBP; UASp-GFP*)(Figure 4A) versus those that drive expression of the rescue construct expressing OVO-B (*ovoovo-GAL4/ovoΔBP; UASp-3xFHA-OVO-B*)(Figure 4B)"

| d) Line 490 "For the big picture"

We have removed this and revised with the below sentence.

Lines 530-531

"To do this, we performed Gene Ontology enrichment analysis with gProfiler software ([Raudvere et al. 2019](#))."

| (2) Some sentences were confusing or extraneous, and the authors may consider revising them. Examples include:

| a) Lines 195-196 "Therefore, we plotted the significant ChIP (minus input) read density peaks centered on the location of the motif itself."

We have removed the word 'peaks' and 'itself', as below.

Lines 200-201

"Therefore, we plotted the significant ChIP (minus input) read density centered on the location of the motif."

| b) Lines 201-203 "... over the location of the motifs, strongly reinforces the idea that our dataset contains regions centered on sequence-specifically bound OVO transcription factor in the ovary."

We have edited this sentence to clarify below.

Lines 204-208

"While it is possible that OVO comes into contact with regions of DNA in three-dimensional nuclear space non-specifically, the presence of OVO motifs within a large percentage of significant ChIP peaks *in vivo* and enrichment of OVO ChIP read density at the location of the motifs, strongly reinforces the idea that our OVO ChIP dataset contains regions centered on sequences specifically bound by OVO in the ovary."

| c) Lines 326-328 "The combinations of these elements...tens of millions of years of evolution."

We have revised this sentence below.

Lines 354-357

“The combinations of these DNA motifs is not random in mammals and *Drosophila* (FitzGerald et al. 2006), and distinct combinations of different motifs at the TSS of genes expressed in *Drosophila* are conserved over tens of millions of years of evolution (Chen et al. 2014).”

d) Lines 444-446 “To address this directly, we tested the idea that genes with... and thus downstream of OVO.”

We have removed this sentence in its entirety.

e) Line 579-580 “Where OVO binding in close proximity, in any ...activates transcription”

We have removed this sentence in its entirety.

(3) Occasionally sentences within the results sections seem more appropriate for the materials and methods. For example, lines 213-218.

(4) At the end of line 375, do the authors mean “only” instead of “also”?

We have modified this sentence below.

Lines 411-414

“*ovoovo-GAL4* is a CRISPR/Cas9 derived T2A-GAL4-3xSTOP insertion upstream of the splice junction of exon 3 in the *ovo-RA* transcript (Figure S1A). Importantly, this insertion in the extended exon 3 would disrupt roughly 90% of the *ovo-B* transcripts. However, since about 10% of *ovo-B* transcripts utilize an upstream splice junction in exon 3, these transcripts would not be disrupted with the T2A-GAL4-3xSTOP insertion and thus allow for enough OVO activity for germ cell survival (Benner et al. 2023).”

(5) In line 392 the authors say that they dissected ovaries “one day post-eclosion” but the methods section says that ovaries were 3-5 days old. Please clarify.

We meant one day old for the RNAseq experiments. We have changed this in the text.

Lines 679-681

“Twenty, one day old post-eclosion *ovoΔBP/ovoovo-GAL4; UASp-GFP* and *ovoΔBP/ovoovo-GAL4; UASp-3x FHAOVO-B* ovaries were dissected and germariums through previtellogenic egg chambers were removed with microdissection scissors and placed in ice cold PBS making up one biological replicate.”

(6) In line 668 the authors mention CRISPR/Cas9 in the methods, but no such experiment was described.

We have removed this from the Methods header.

<https://doi.org/10.7554/eLife.94631.2.sa2>