TBPL2/TFIIA complex overhauls oocyte transcriptome during oocyte growth

Changwei Yu[1,2,3,4], Nevena Cvetesic[5], Kapil Gupta[6], Tao Ye[1,2,3,4], Emese Gazdag[1,2,3,4], Vincent Hisler[1,2,3,4], Luc Negroni[1,2,3,4], Petra Hajkova[5], Boris Lenhard[5], Ferenc Müller[7], Imre Berger[6], Stéphane D. Vincent[1,2,3,4,*,#] and László Tora[1,2,3,4,*,#]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67404 Illkirch, France;

[2]Centre National de la Recherche Scientifique (CNRS), UMR7104, 67404 Illkirch, France;

[3]Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, 67404 Illkirch, France;

[4]Université de Strasbourg, 67404 Illkirch, France;

[5]Imperial College London, South Kensington Campus, London SW7 2AZ, UK;

[6]Max Planck Bristol Centre for Minimal Biology, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK.

[7]Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, UK;

[#] These authors contributed equally to this work

[*]Co-corresponding authors: S.D.V. vincent@igbmc.fr and L.T. laszlo@igbmc.fr

The first steps of oocyte development from primordial follicle are characterised by a growth phase, when unique RNA and protein reserves are created to achieve oocyte competence. During this growth, oocytes do not divide and the general transcription factor TATA binding protein (TBP) is replaced by its paralogue, TBPL2 (also called TBP2 or TRF3), which is essential for RNA polymerase II transcription (Pol II) [1,2]. However, the composition and function of transcription machinery and the regulatory mechanisms mediating Pol II transcription during this developmental stage remain unknown. In somatic cells, the general transcription factor TFIID, which contains TBP and 13 TBP-associated factors, is the first to bind gene promoters to nucleate Pol II transcription initiation[3]. Here, we show that in oocytes TBPL2 does not assemble into a canonical TFIID complex, while it stably associates with TFIIA via distinct TFIIA interactions when compared to TBP. Our transcript analyses in wild type and *Tbpl2*$^{-/-}$ oocytes demonstrates that TBPL2 mediates transcription of oocyte-expressed genes, including mRNA destabilisation factors genes, as well as specific endogenous retroviral elements (ERVs). Transcription start site (TSS) mapping from wild-type and *Tbpl2*$^{-/-}$ growing oocytes demonstrates that TBPL2 has a strong preference for TATA-like motif in gene core promoters driving specific sharp TSS selection. This is in marked contrast with TBP/TFIID-driven TATA-less gene promoters in preceding stages that have broad TSS architecture. We anticipate that our findings describing oocyte-specific transcription regulation will help to understand the mechanisms associated with primary ovarian insufficiency, which constitutes a frequent cause of infertility among women.

Pol II transcription requires the stepwise assembly of multi-protein complexes called general transcription factors (GTFs) and Pol II[4]. The evolutionary conserved TFIID complex plays a major role in transcription initiation as it is the first GTF to initiate the assembly of the

pre-initiation complex (PIC) by recognizing the core promoter[5,6]. TFIID is a large multi-protein complex composed of the TATA box-binding protein (TBP) and 13 TBP-associated factors (TAFs) in metazoa[7]. The model suggesting that transcription is always regulated by the same transcription complexes has been challenged in metazoans by the discovery of cell-type specific complexes containing specialized GTF-, TBP- or TAF-paralogs[3,8,9]. However, how alternative initiation complexes form and how they regulate cell type-specific transcription remains unknown. During oocyte growth TBP is absent and replaced by a vertebrate-specific TBP-related factor, TBPL2, required for female fertility[1,2,10]. During this phase the oocyte transcriptome is overhauled in a TBPL2-dependent manner. Importantly, Pol II transcription is blocked during oocyte growth in the absence of TBPL2[2].

TBPL2 forms an oocyte-specific complex with TFIIA different from TFIID

To characterize TBPL2-containing transcription complexes we prepared whole cell extracts (WCE) from 14 days post-natal (P14) mouse ovaries and analysed TBPL2-associated proteins by anti-mTBPL2 immunoprecipitation (IP) coupled to label free mass spectrometry (Fig. 1a, Extended Data Fig. 1a, b). We identified TFIIA-αβ and TFIIA-γ subunits as unique GTF subunits associated with TBPL2 (Fig. 1a, Supplementary Table 1). In parallel, anti-TBP IP from the same extracts showed that TBP assembles into the canonical TFIID complex in non-oocyte cells present in large excess in the ovary (Extended Data Fig. 1c, Supplementary Table 2). To determine the stoichiometry of the composition of the immunoprecipitated complexes, normalized spectral abundance factor (NSAF) values were calculated[11]. As growing oocytes represent only a tiny minority of ovary cells, we further tested the TBPL2-TFIIA interaction by a triple IP strategy: first, we depleted TAF7-containing TFIID complexes with an anti-TAF7 IP, second, the remaining TFIID and SAGA complexes, which contain also some shared TAFs[12], were depleted with an anti-TAF10 IP using the anti-TAF7 IP flow-through as

an input, third we performed an anti-TBPL2 IP on the anti-TAF7/anti-TAF10 flow-through fraction (Fig. 1b-d, Extended Data Fig. 1d, Supplementary Table 3). The analysis of this third consecutive IP further demonstrated that TBPL2 forms an unique complex with TFIIA-αβ, and TFIIFA-γ, but without any TFIID subunits. To further analyse the requirement of TFIID during oocyte growth, we carried out a conditional deletion of TFIID-specific *Taf7* gene during oocyte growth using the *Zp3-Cre* transgenic line[13]. The oocyte-specific deletion of *Taf7* did not affect the presence of secondary and antral follicles and the numbers of collected mature oocytes after superovulation (Fig. 1e, f, Extended Data Fig. 1e). Thus, the TBPL2-containing transcription complex does not require TFIID TAFs for its function during oocyte growth. These results together show that during oocyte growth a stable TBPL2-TFIIA complex forms, and may function differently from TBP/TFIID.

Structure/function characterization of the TBPL2/TFIIA complex

Next, we determined the X-ray crystal structure of the mouse TBPL2 core domain ($TBPL2^C$) (Fig. 2a, Supplementary Table 4). The 2Å structure reveals a $TBPL2^C$ dimer reminiscent of the dimer observed in TBP core (TBP^C) crystals[14,15]. $TBPL2^C$ forms a pseudo-symmetric bi-lobal fold similar to TBP^C (Extended Data Fig. 2a). The respective electrostatic surface charge distributions show that while $TBPL2^C$ and TBP^C exhibit similarly basic charges of their N-terminal lobes and DNA binding concave surfaces, their central regions and C terminal lobes are markedly different, with $TBPL2^C$ shifted towards acidic charge (Fig. 2b, c, Extended Data Fig. 2b, c). Based on the crystal structure of TBP^C in complex with TATA box DNA and a crystallisable core of TFIIA ($TFIIA^C$)[16], we modelled the $TBPL2^C$-$TFIIA^C$ interactions (Fig. 2d). The resulting model suggests that $TBPL2^C$ has a conserved, but not identical interaction interface with the TFIIA-γ chain when compared with TBP^C.

We next analysed the functional interactions of TBPL2^{C} with full-length processed TFIIA and adenovirus major late promoter TATA box DNA by electrophoretic mobility shift assay (EMSA) (Fig. 2e, f, Extended Data. Fig. 2d, e). Preformed TBPL2/DNA complex was supplemented with increasing amounts of processed TFIIA-α, -β and TFIIA-γ (Fig. 2e). Unexpectedly, we observed two shifted complexes in EMSA (labelled complex 1 and 2 in Fig. 2e), but complex 2 was barely noticeable when TBP^{C} was used (Fig. 2f). Intriguingly, when only low amounts of TFIIA were added, a prominent band corresponding to free DNA reproducibly re-emerged in the TBPL2^{C} EMSA experiment, disappearing again when increasingly more TFIIA was supplied (Fig. 2e). No such re-emergence of free DNA was observed with TBP^{C} (Fig. 2f). Our analyses together suggest that TBPL2^{C} can bind differently to TFIIA in the presence of canonical TATA box DNA, as compared to TBP^{C}.

TBPL2-dependent oocyte transcriptome

To characterize the growing oocyte-specific transcriptome and its dependence on TBPL2, we have performed a transcriptomic analysis of wild-type (WT) and *Tbpl2*$^{-/-}$ oocytes isolated from primary and secondary follicles [post-natal day (P) 7 and P14] (Fig. 3, Supplementary Table 5). We observed down-regulation of a number of oocyte-specific genes, such as *Bmp15* and *Gdf9*[17,18] (Fig. 3a, b and Extended Fig. 3a). Principal component analysis showed that the four distinct RNA samples clustered in individual groups and that the main explanation for the variance is the genotype, and then the stage (Extended Data Fig. 3b). Comparison of the RNA level fold changes between mutant and WT oocytes showed that in *Tbpl2*$^{-/-}$, there is a massive down-regulation of the most highly expressed transcripts, both at P7 and P14 (Extended Data Fig. 3c). The Pearson correlation between the P7 and P14 fold change data sets for transcripts expressed above 100 normalized reads was close to 0.8, indicating that *Tbpl2* loss of function

similarly altered RNA levels at P7 and P14 stages. We therefore focused on the P14 stage for the rest of the study.

In WT P14 oocytes transcripts corresponding to 10791 genes were detected. Importantly, many of these detected transcripts have been transcribed at earlier stages and are stored in growing oocytes[19]. As there is no Pol II transcription in *Tbpl2*$^{-/-}$ growing oocytes[2], RNAs detected in the *Tbpl2*$^{-/-}$ mutant oocytes represent mRNAs transcribed by a TBP/TFIID-dependent mechanism and deposited into the growing oocytes independently of TBPL2 activity at earlier stages. The proportion of genes (1396) upregulated following *Tbpl2* deletion (Fig. 3c) can be explained by two mutually not exclusive ways: i) the consequence of the normalization to the library size resulting in the over-estimation of up-regulated transcripts, and under-estimation of down-regulated transcripts and/or by transcript buffering mechanisms due to mRNA stabilization[20]. Validation of the up-regulation of some candidate transcripts levels (Extended Data Fig. 3d, e) strongly supports the latter hypothesis.

Nevertheless, we detected 1802 significantly downregulated transcripts in the *Tbpl2*$^{-/-}$ oocytes (Fig. 3c). Key genes known to be expressed during oocyte growth, such as *Bmp15, Eloc, Fgf8, Gdf9* and *Zar1*[17,18,21], were confirmed to be down-regulated (Extended Data Fig. 3f, g). These results suggest that TBPL2 has an important role in gene expression in the growing oocytes. Gene Ontology (GO) analyses of biological process of the identified down regulated categories of genes (Supplementary Table 6) indicated that many genes involved in meiosis II and distinct cell cycle processes were significantly down-regulated (Extended Data. Fig. 3h). The most enriched molecular function GO category was "poly(A)-specific ribonuclease activity" containing many genes coding for factors or subunits of complexes contributing to deadenylation/decapping/decay activity in eukaryotes (Fig. 3d) (i.e. CCR4-NOT, PAN2/PAN3[22]; DCP1A/DCP2[23], or BTG4[21]). Transcripts coding for these "poly(A)-specific ribonuclease activity" factors were significantly down regulated in *Tbpl2*$^{-/-}$ mutant P14 oocytes

(Fig. 3e, Extended Data. Fig. 3i). Thus, in P14 oocytes TBPL2 is regulating the transcription of many genes coding for factors, which are crucial in regulating the stability and translation of the mRNA stock deposited during early oogenesis, as well as transcription of meiosis II- and cell cycle-related genes to prepare the growing oocytes for the upcoming meiotic cell division.

A remarkable feature of oocyte is the very high expression of retrotransposons driven by Pol II transcription. There are three major classes of retrotransposons in mammals: long interspersed nuclear elements, short interspersed nuclear elements and ERVs. There are three main sub-classes of ERVs: ERV1, ERVK and endogenous retrovirus like ERVL-MaLR[24]. As expected, in WT P14 oocytes the expression of ERVs was found to be the most abundant[25,26] (Extended Data Fig. 3j, k). Importantly, the transcription of the vast majority of MaLR elements was the most affected in *Tbpl2*$^{-/-}$ mutant oocytes (Fig. 3f-i). Among them, three highly expressed members, *MT-int*, *MTA_Mm*, and *MTA_Mm-int*, were dramatically down-regulated in P14 *Tbpl2*$^{-/-}$ mutant oocytes (Extended Data Fig. 3l). As in P14 oocytes TBPL2 depletion is reducing transcription more than 4-fold from MaLR ERVs, which often serve as promoters for neighbouring genes[25,26], TBPL2 could seriously deregulate oocyte-specific transcription and consequent genome activation.

Therefore, TBPL2 is important for the *de novo* restructuration of the oocyte transcriptome and indirectly for silencing the translation of the TBP-dependent earlier deposited transcripts.

TBPL2-driven promoters are TATA box-containing with sharp TSS selection

The promoter usage changes during zebrafish maternal to zygotic transition revealed different rules of transcriptional initiation in oocyte and in embryo, driven by independent and often overlapping sets of promoter "codes"[27]. This switch has not yet been demonstrated in mammals and the role of TBPL2 during oogenesis remained to be investigated. To this end, we mapped the TSS usage by carrying out super-low input carrier-CAGE (SLIC-CAGE)[28] from

WT and *Tbpl2*$^{-/-}$ P14 oocytes. To characterize only the TBPL2-driven promoters we removed the CAGE tags present in the *Tbpl2*$^{-/-}$ dataset from the WT P14 dataset, to eliminate transcripts that have been deposited at earlier stages (hereafter called "TBPL2-specific-only"). Conversely, the *Tbpl2*$^{-/-}$ dataset corresponds to the TBP/TFIID or TBPL2-independent-only TSSs (hereafter called "TBPL2-independent-only").

Next, we analysed the genome-wide enrichment of T- and/or A-rich (WW) dinucleotide motifs within the -250/+250 region centred on the dominant TSSs of the TBPL2-specific-only and TBPL2-independent-only oocyte TSS clusters (Fig. 4a, b). TBPL2-specific-only TSS clusters are strongly enriched in a well-defined WW motif around their -30 bp region (Fig. 4a, read arrowhead) (see also Cvetesic *et al.*, in submission). In contrast, only about 1/3rd of the TBPL2-independent-only TSS clusters contained WW-enriched motifs at a similar position (Fig. 4b, read arrowhead), as would be expected from promoters that lack maternal promoter code determinants[27] (Cvetesic *et al.*, in submission). As canonical TATA boxes are often associated with tissue-specific gene promoters, we investigated whether the above observed WW motif densities correspond to TATA boxes using the TBP position weight matrix (PWM) from the JASPAR database as a reference. To this end the presence of TATA boxes was analysed in the TSS clusters of the two data sets and revealed that TBPL2- specific-only TSS clusters were enriched in high quality TATA boxes, including a clear increase in the proportion of canonical TATA boxes, when compared to TBPL2-independent-only TSS clusters (Fig. 4c). Depending on promoter sequence motifs, transcription initiation can occur within a narrow region with a major dominant TSS, or within a less defined region containing a wider distribution of TSSs, leading to the distinction of "sharp" and "broad" TSS-type initiation, respectively[29]. Genome browser view snapshots indicate that TSS clusters in P14 WT oocytes tend to be sharp and are associated with TATA-like motifs (Extended Data. Fig. 4a, b). Analysis of the global distribution of the number of TSSs and of the width of the TSS clusters in the

above defined two categories confirmed that TBPL2- specific-only TSS are sharper compared to the TBPL2-independent-only TSS clusters (Extended Data Fig. 4c, d).

In order to test whether TBPL2 controls transcription initiation from maternal promoter code determinants, we grouped the expression profiles corresponding to each consensus TSS clusters, to characterise promoter activity profiles among datasets by performing self-organizing maps (SOM) (Extended Data Fig. 4e). We then focussed on the two most distinct SOM groups: the down-regulated promoters (blue group, containing 9442 consensus TSS clusters) (Fig. 4d) and the up-regulated promoters (red group, with 6900 consensus TSS clusters) (Fig. 4e). Motif analyses of these two categories of promoters in their -35/+5 regions relative to the different dominant TSSs indicated that only the core promoters associated to TBPL2-specific-only dominant TSSs belonging to the down-regulated gene promoters contain a well-defined 7 bp long TATA box-like motif (W-box) in their -31 to -24 regions (Fig. 4f, g, Extended Data Fig. 4f-i). Importantly, W-box associated TSSs architecture usage distribution for these TBPL2-specific-only dominant TSSs was sharp (Extended data Fig. 4j, l), as expected for motif-dependent transcriptional initiation[27] (Cvetesic *et al.*, in submission). In contrast, TBPL2-independent-only TSSs belonging to the up-regulated promoters exert a much broader TSS pattern (Extended Data Fig. 4k, m). Interestingly, GO analyses of the genes associated with the down-regulated promoters revealed a strong association with deadenylation/decapping/decay activity (Extended Data Fig. 4n-p, Supplementary Table 7), further confirming our initial RNA-seq analysis observations (Fig. 3).

Importantly, TSS architecture analyses of the TBPL2-specific-only MaLR ERV TSSs indicated that the majority of MaLR core promoters contain high quality TATA box motif (median of the TATA box PWM match is 85%, Fig. 4i, j). These observations together demonstrate that the TBPL2-TFIIA complex drives transcription initiation primarily from core promoters that contain a TATA box-like motif in their core promoter and directs sharp

transcription initiation from the corresponding promoter regions to overhaul the growing oocyte transcriptome.

In addition, we observed that TSS usage can shift within the promoter of individual genes depending on the genetic background (Extended Data Fig. 4b). To get more insights in these promoter architecture differences, we identified genome-wide 6429 shifting promoters by comparing either TBPL2-specific-only to TBPL2-independent-only TSS data. These results are consistent with TSS shifts between somatic and maternal promoter codes (Cvetesic *et al.*, in submission) occurring either in 5', or 3' directions (Fig. 4k, Extended Data Fig. 4q). WW motif analysis indicated that on each shifting promoter, TBPL2-specific-only dominant TSSs are associated with WW motifs, while TBPL2-independent-only dominant TSSs are not (Fig. 4l). In addition, the TATA box PWM match analyses indicated that these WW motifs are enriched in TATA box like elements compared to the corresponding TBPL2-independent-only shifting TSSs (Fig. 4m). Thus, our experiments provide a direct demonstration that TBP/TFIID and TBPL2/TFIIA machineries recognize two distinct sequences co-existing in promoters of the same genes with TBPL2 directing a stronger WW/TATA box-dependent TSS selection in them.

Discussion

In conclusion, our results demonstrate that in growing oocytes TFIID (TBP/TAFs) is replaced by a non-canonical TBPL2/TFIIA complex that uses canonical TATA box-containing core promoters to drive oocyte-specific sharp TSS-transcription initiation. Thus, the TBPL2/TFIIA complex is a key regulator of growing oocyte-specific transcription, including the high expression of MaLR ERVs, by creating a novel transcriptome pool, consisting of new TBPL2-dependent transcripts and TBP/TFIID-dependent transcripts deposited earlier during the primordial follicle stage (Fig. 5). Importantly, to block the translation of mRNAs deposited at the earlier developmental stage, TBPL2 is regulating the activity of several

deadenylation/decapping/decay complexes, which further contributes to establishing the novel TBPL2-dependent growing oocyte transcriptome and consequent proteome required for further development and oocyte competence for fertilization (Fig. 5). The indirect regulation of previously deposited mRNAs by a global transcription regulator resembles the well characterised maternal to zygotic transition (MZT), during which a clearance of inherited transcriptome is mediated by *de novo* gene products generated by newly activated transcription machinery[30-32]. At hundreds of gene promoters, two distinct TSS-defining "grammars" co-exist in close proximity genome-wide and are differentially utilised in either TBPL2/TFIIA in primary/secondary follicular oocytes or by TBP/TFIID in primordial follicular oocytes. This again shows a striking parallel to MZT[27], where multiple layers of information are embedded in the same promoter sequence, each representing a different type of regulatory grammar interpreted by dedicated transcription machinery depending on the cellular environment.

Main references

1. Gazdag, E., Rajkovic, A., Torres-Padilla, M.-E. & Tora, L. Analysis of TATA-binding protein 2 (TBP2) and TBP expression suggests different roles for the two proteins in regulation of gene expression during oogenesis and early mouse development. *Reproduction* **134,** 51–62 (2007).
2. Gazdag, E. *et al.* TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte. *Genes Dev* **23,** 2210–2223 (2009).
3. Goodrich, J. A. & Tjian, R. Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat Rev Genet* **11,** 549–558 (2010).

4. Roeder, R. G. 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. *Nat Struct Mol Biol* **26,** 783–791 (2019).

5. Vo Ngoc, L., Wang, Y.-L., Kassavetis, G. A. & Kadonaga, J. T. The punctilious RNA polymerase II core promoter. *Genes Dev* **31,** 1289–1301 (2017).

6. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol* **19,** 621–637 (2018).

7. Tora, L. A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev* **16,** 673–675 (2002).

8. Müller, F., Zaucker, A. & Tora, L. Developmental regulation of transcription initiation: more than just changing the actors. *Curr Opin Genet Dev* **20,** 533–540 (2010).

9. Akhtar, W. & Veenstra, G. J. C. TBP-related factors: a paradigm of diversity in transcription initiation. *Cell Biosci* **1,** 23 (2011).

10. Malecova, B. *et al.* TBP/TFIID-dependent activation of MyoD target genes in skeletal muscle cells. *Elife* **5,** 23 (2016).

11. Zybailov, B. *et al.* Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. *J. Proteome Res.* **5,** 2339–2347 (2006).

12. Helmlinger, D. & Tora, L. Sharing the SAGA. *Trends Biochem. Sci.* **42,** 850–861 (2017).

13. Lewandoski, M., Wassarman, K. M. & Martin, G. R. Zp3–cre, a transgenic mouse line for the activation or inactivation of loxP-flanked target genes specifically in the female germ line. *Current Biology* **7,** 148–151 (1997).

14. Chasman, D. I., Flaherty, K. M., Sharp, P. A. & Kornberg, R. D. Crystal structure of yeast TATA-binding protein and model for interaction with DNA. *Proceedings of the National Academy of Sciences* **90,** 8174–8178 (1993).

15. Nikolov, D. B. & Burley, S. K. 2.1 A resolution refined structure of a TATA box-binding protein (TBP). *Nat. Struct. Biol.* **1,** 621–637 (1994).

16. Bleichenbacher, M., Tan, S. & Richmond, T. J. Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes. *J Mol Biol* **332,** 783–793 (2003).

17. Dong, J. *et al.* Growth differentiation factor-9 is required during early ovarian folliculogenesis. *Nature* **383,** 531–535 (1996).

18. Galloway, S. M. *et al.* Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nat Genet* **25,** 279–283 (2000).

19. Bettegowda, A. Mechanisms of maternal mRNA regulation: implications for mammalian early embryonic development. *Frontiers in Bioscience* **12,** 3713 (2007).

20. Timmers, H. T. M. & Tora, L. Transcript Buffering: A Balancing Act between mRNA Synthesis and mRNA Degradation. *Mol Cell* **72,** 10–17 (2018).

21. Yu, C. *et al.* BTG4 is a meiotic cell cycle–coupled maternal-zygotic-transition licensing factor in oocytes. *Nat Struct Mol Biol* **23,** 387–394 (2016).

22. Wolf, J. & Passmore, L. A. mRNA deadenylation by Pan2-Pan3. *Biochem. Soc. Trans.* **42,** 184–187 (2014).

23. Ma, J., Flemr, M., Strnad, H., Svoboda, P. & Schultz, R. M. Maternally recruited DCP1A and DCP2 contribute to messenger RNA degradation during oocyte maturation and genome activation in mouse. *Biol Reprod* **88,** 11 (2013).

24. Thompson, P. J., Macfarlan, T. S. & Lorincz, M. C. Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. *Mol Cell* **62,** 766–776 (2016).

25. Peaston, A. E. *et al.* Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* **7,** 597–606 (2004).

26. Veselovska, L. *et al.* Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biol.* **16,** 20110330–0 (2015).

27. Haberle, V. *et al.* Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature* **507,** 381–385 (2014).

28. Cvetesic, N. *et al.* SLIC-CAGE: high-resolution transcription start site mapping using nanogram-levels of total RNA. *Genome Res.* **28,** 1943–1956 (2018).

29. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38,** 626–635 (2006).

30. Ferg, M. *et al.* The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish. *EMBO J* **26,** 3945–3956 (2007).

31. Giraldez, A. J. *et al.* Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312,** 75–79 (2006).

32. Vastenhouw, N. L., Cao, W. X. & Lipshitz, H. D. The maternal-to-zygotic transition revisited. *Development* **146,** dev161471 (2019).

Methods

Cell lines and cell culture

The NIH3T3-II10 line overexpressing TBPL2 and the control NIH3T3-K2 have already been described[1] and were maintained in high glucose DMEM supplemented with 10% of new-born calf serum at 37°C in 5% CO_2.

Whole cell extracts

NIH3T3-II10 and NIH3T3-K2 cells cultured in 15 cm dish were washed twice with 1x PBS, subsequently harvested by scrapping on ice. Harvested cells were centrifuged 1000 rcf at 4°C for 5 min and then resuspended in 1 packed cell volume of whole cell extraction buffer (20 mM Tris HCl pH7.5, 2 mM DTT, 20% Glycerol, 400 mM KCl, 1x Protease inhibitor cocktail (PIC, Roche)). Cell lysates were frozen in liquid nitrogen and thawed on ice for 3 times, followed by centrifugation at 20817 rcf, at 4°C for 15 min. The supernatant was collected and protein concentration was measured by Bradford protein assay (Bio-Rad). The cell extracts were used directly for immunoprecipitation and western blot, or stored at −80°C.

Ovaries collected from post-natal day 14 (P14) CD1 and C57BL/6N female mice were homogenized in whole cell extraction buffer (20 mM Tris HCl pH7.5, 2 mM DTT, 20% Glycerol, 400 mM KCl, 5x PIC (Roche)). Cell lysates were frozen in liquid nitrogen and thawed on ice for 3 times, followed by centrifugation at 20817 rcf, at 4°C for 15 min. The supernatant extracts were used directly for immunoprecipitation.

Antibodies and antibody purification

The 2B12 anti-TBPL2, the 3TF13G3 anti-TBP and the 15TF21D10 anti-GST mouse monoclonal antibodies have already been described[1,33,34]. The J7 rabbit polyclonal anti TFIIA

was a gift of H.G. Stunnenberg. The IGBMC antibody facility raised the anti-TBPL2 polyclonal 3024 serum against the CPDEHGSELNLNSNSSPDPQ peptide (amino acids 111-129) coupled to ovalbumin and injected into one rabbit. The resulting serum was affinity purified by using the Sulfolink Coupling Gel (Pierce) following the manufacturer's recommendations.

Immunoprecipitation

Ovary extract were incubated with anti-GST, anti-TBP, anti-TBPL2, anti-TAF7 or anti-TAF10 coated Dynabeads (Invitrogen) at 4°C overnight. After incubation, beads were washed 3 x 5min at 4°C with 500 mM KCl buffer [25 mM Tris-HCl (pH 7.9), 5 mM MgCl2, 10% glycerol, 0.1% NP40, 2 mM DTT, 500 mM KCl and 1x PIC (Roche)], then washed 3 x 5min at 4°C with 100 mM KCl buffer (25 mM Tris-HCl pH 7.9, 5 mM MgCl2, 10% glycerol, 0.1% NP40, 2 mM DTT, 100 mM KCl and 1x). Immunoprecipitated proteins were eluted with 0.1 M glycine pH 2.8 and neutralized with 1.5 M Tris-HCl pH 8.8. Immunoprecipitation performed from whole cell extracts of NIH3T3-II10 and NIH3T3-K2 cells were following the same procedures with protein G Sepharose beads (GE Healthcare).

Western blot

Protein samples (15-25 µg of cell extracts or 15 µL of IP elution) were mixed with 1/4th volume of loading buffer (100 mM Tris-HCl pH 6.8, 30% glycerol, 4% SDS, 0.2% bromophenol blue and freshly added 100 mM DTT) and boiled for 10 min. Samples were then resolved on a 10 % SDS-PAGE and transferred to nitrocellulose membrane (Protran, Amersham). Membranes were blocked in 3% non-fat milk in 1x PBS at room temperature (RT) for 30 min, and subsequently incubated with the primary antibody overnight at 4°C. Membranes were washed three times (10 min each) with 1x PBS - 0.05% Tween20. Membranes were then incubated with HRP-coupled secondary antibodies (Jackson ImmunoResearch) for 1 hour at RT, followed by

ECL detection (Thermo Fisher). The signal was acquired with the ChemiDoc imaging system (Bio-Rad).

Mass spectrometry analyzes and NSAF calculations

Samples were TCA precipitated, reduced, alkylated and digested with LysC and Trypsin at 37°C overnight. After C18 desalting, samples were analyzed using an Ultimate 3000 nano-RSLC (Thermo Scientific, San Jose, California) coupled in line with a linear trap Quadrupole (LTQ)-Orbitrap ELITE mass spectrometer via a nano-electrospray ionization source (Thermo Scientific). Peptide mixtures were loaded on a C18 Acclaim PepMap100 trap column (75 µm inner diameter × 2 cm, 3 µm, 100 Å; Thermo Fisher Scientific) for 3.5 min at 5 µL/min with 2% acetonitrile (ACN), 0.1% formic acid in H_2O and then separated on a C18 Accucore nano-column (75 µm inner diameter × 50 cm, 2.6 µm, 150 Å; Thermo Fisher Scientific) with a 240 minutes linear gradient from 5% to 50% buffer B (A: 0.1% FA in H_2O / B: 80% ACN, 0.08% FA in H_2O) followed with 10 min at 99% B. The total duration was set to 280 minutes at a flow rate of 200 nL/min.

Proteins were identified by database searching using SequestHT (Thermo Fisher Scientific) with Proteome Discoverer 1.4 software (Thermo Fisher Scientific) a combined Mus musculus database (Swissprot, release 2015_11, 16730 entries) where 5 interesting proteins sequences (TrEMBL entries) were added. Precursor and fragment mass tolerances were set at 7 ppm and 0.5 Da respectively, and up to 2 missed cleavages were allowed. Oxidation (M) was set as variable modification, and Carbamidomethylation© as fixed modification. Peptides were filtered with a false discovery rate (FDR) at 5 %, rank 1 and proteins were identified with 1 unique peptide. Normalized spectral abundance factor (NSAF)[11] were calculated using custom R scripts (R software version 3.5.3). Only proteins detected in at least 2 out of 3 of the technical or biological replicates were considered for further analyses.

1

2 ***Animal experimentation***

3 Animal experimentations were carried out according to animal welfare regulations and
4 guidelines of the French Ministry of Agriculture and procedures were approved by the French
5 Ministry for Higher Education and Research ethical committee C2EA-17 (project
6 n°2018031209153651). The Tg(Zp3-Cre), *Taf7*^{flox} and *Tbpl2*$^{-}$ lines have already been
7 described[2,13,35].

8

9 ***Histology analyses of ovaries***

10 Ovaries were collected from 6 weeks-old Tg(Zp3-Cre/+);*Taf7*$^{flox/+}$ and Tg(Zp3-Cre/+);*Taf7*$^{flox/\Delta}$
11 oocyte specific mutant females, fixed in 4% paraformaldehyde (Electron Microscopy Sciences)
12 over-night at 4°C, washed 3 times in PBS at room temperature and embedded in paraffin. Five
13 µm-thick sections were stained with hematoxylin and eosin and images were acquired using a
14 slide scanner Nanozoomer 2.0HT (Hamamatsu Photonics).

15

16 ***Supervovulation***

17 Five units of pregnant mare serum (PMS) was injected intraperitoneally in 4-week-old female
18 mice between 2-4 pm. After 44-46 hours, GV oocytes were collected from the ovaries by
19 puncturing with needles.

20

21 ***Expression and purification of TBPL2^{C} and TBP^{C}***

22 Mouse TBPL2^{C} (amino acids 170 to 349) fused to an N-terminal hexa-histidine tag was
23 expressed in Sf21 insect cells using the MultiBac system[36]. Cells were infected with composite
24 baculovirus, harvested and pelleted following published protocols[37]. Cell pellets were
25 resuspended and lysed by freeze-thaw in liquid nitrogen in Talon Buffer A (25 mM Tris pH

7.5, 1000 mM NaCl, 10 mM imidazole and compete protease inhibitor tablet (Roche)). Lysate was cleared by centrifugation at 40,000 rcf for 60 min. Cleared lysate was incubated with Talon Resin (Thermo Fisher) and unbound protein washed with 20 column volumes (CVs) of Talon Buffer A. Protein was eluted with Talon Buffer B (25 mM Tris pH 7.5, 1000 mM NaCl, 600 mM Imidazole) using a step gradient. The tag was removed by TEV protease cleavage during dialysis overnight into Dialysis Buffer (25 mM Tris pH 8.0, 300 mM NaCl, 2 mM β-MeSH) followed by a reverse IMAC step to remove uncleaved protein. TBP was then subjected to size exclusion chromatography (SEC) using a SuperdexS200 16/600 column equilibrated in SEC Buffer (25 mM Tris pH 7.5, 300 mM NaCl, 5 mM β-MeSH). Mouse TBP^C (amino acid residues 136 to 316) was expressed in *E. coli* Rosetta (DE3) cells at 18°C and purified as described for $TBPL2^C$.

Crystallization and structure determination of $TBPL2^C$

Purified $TBPL2^C$ was screened for crystallization in a range of conditions using the sitting drop method. Crystals obtained in 26% PEG3350, 0.2 M magnesium formate were harvested flash-frozen in liquid nitrogen and used for data collection at Diamond Light Source, (beamline I04-1). Diffraction data were processed and scaled using software XDS[38] and the structure solved by molecular replacement using software Phenix[39], with TBP^C (from PDBID 1NVP) as a starting model. The structure was iteratively rebuilt and refined. Data and refinement statistics are provided (Supplementary Table 5). Structure images were prepared using software PyMOL.

Expression and purification of TFIIA

The gene encoding for mouse TFIIA-γ was cloned in the pIDC plasmid from the MultiBac suite[37]. The gene encoding for mouse TFIIA-αβ fused to an N-terminal hexa-histidine tag and a TEV protease site was cloned in the pACEBac1 plasmid of the MultiBac system. A second

TEV protease site was also introduced in TFIIA-αβ at the TASPASE 1 cleavage site, to prepare processed TFIIA comprising native-like TFIIA- α and TFIIA-β polypeptides (Extended Data Fig. 1n). Plasmids were fused by Cre recombinase and expression cassettes inserted into the MultiBac baculoviral genome by Tn7 transposition. Sf21 cells were infected with composite baculovirus, harvested and pelleted. Cell pellets were lysed by freeze-thaw in liquid nitrogen in Talon Buffer A1 (25 mM Tris pH 7.5, 300 mM NaCl, 10 mM Imidazole and compete protease inhibitor tablet (Roche)). Lysate was cleared by centrifugation at 40,000 rcf for 60 min. Cleared lysate was incubated with Talon Resin (Thermo Fisher) and unbound protein washed with 20 CVs of Talon Buffer A. Resin was then washed with 20 CVs of Talon Buffer HS (25 mM Tris pH 7.5, 1000 mM NaCl, 10 mM imidazole). Protein was eluted from TALON resin with Talon Buffer B1 (25 mM Tris pH 7.5, 300 mM NaCl, 300 mM imidazole) using a step gradient. The TFIIA-αβ subunit was processed into α and β polypeptide chains by TEV protease cleavage during dialysis overnight in MonoQ buffer A (25 mM Tris pH 8.0, 100 mM NaCl, 5 mM β-MeSH). In this step, the histidine-tag was likewise removed. Thus, processed TFIIA was loaded on a MonoQ column followed by 15 CVs wash with MonoQ buffer A. Processed TFIIA was eluted with MonoQ buffer B (25 mM Tris pH 8, 1000 mM NaCl, 5 mM β-MeSH) using a continuous gradient. Protein complex was further purified using a Superdex S200 10/300 column equilibrated in SEC Buffer (25 mM Tris pH 7.5, 300 mM NaCl, 5 mM β-MeSH).

Electrophoretic mobility shift assay (EMSA)

Cy3 labelled Adenovirus major late promoter (AdMLP, coding strand sequence 5'CTGCTATAAAAGGCTG3') was purchased from Eurofins Genomics. Double-stranded DNA (dsDNA) substrate was generated by mixing coding and complimentary non-coding DNA oligonucleotides at equimolar ratio, followed by heating to 90°C and annealing by slowly cooling to room temperature (RT). Samples for EMSA experiments were prepared by mixing

annealed Cy3 labelled dsDNA (0.5 µM) with either TBP^C or $TBPL2^C$ (1 µM) in EMSA Reaction Buffer (10 mM Tris pH 8.0, 60 mM KCl, 10 mM MgCl2, 10% glycerol, 2.5 mM DTT). Purified processed TFIIA was added to aliquots of this mix at increasing concentrations (0.25 µM to 3 µM) followed by 90 min incubation on ice. Samples were analysed by non-denaturing 6% Tris-glycine gels (Invitrogen) using 1x Tris-Glycine native running buffer. Gels were imaged by recording fluorescence signal for Cy3 using a Typhoon FLA 9000 Gel Imager (GE Healthcare).

Oocytes collection

After dissection, ovaries are freed from adhering tissues in 1x PBS. Series of 6 ovaries were digested in 500 µL of 2 mg/mL Collagenase (SIGMA), 0.025% Trypsin (SIGMA) and 0.5 mg/mL type IV-S hyaluronidase (SIGMA), on a thermomixer at 600 rpm for 20 minutes. The digestion was then stopped by the addition of 1 mL of 37°C pre-warmed αMEM - 5% FBS. The oocytes were then size-selected under a binocular.

RNA preparation

Pool of 100-200 oocytes collected were washed through several M2 drops, and total RNA was isolated using NucleoSpin RNAXS kit (Macherey-Nagel) according to the user manual. RNA quality and quantity were evaluated using a Bioanalyzer. Between 5-10 ng of RNA was obtained from each pool of oocytes.

RNA-seq analyses

PolyA+ RNA seq libraries were prepared using the SMART-Seq v4 UltraLow Input RNA kit (Clonetch) followed by the Nextera XT DNA library Prep kit (Illumina) according to the manufacturer recommendations from 3 biological replicates for each condition (P7 wild-type

(WT), P7 *Tbpl2*$^{-/-}$ mutant, P14 WT and P14 *Tbpl2*$^{-/-}$ mutant oocytes) and sequenced 50 pb single end using an Illumina HiSeq 4000 (GenomEast platform, IGBMC).

Reads were preprocessed in order to remove adapter, polyA and low-quality sequences (Phred quality score below 20). After this preprocessing, reads shorter than 40 bases were discarded for further analysis. These preprocessing steps were performed using cutadapt version 1.10[40]. Reads were mapped to spike sequences using bowtie version 2.2.8[41], and reads mapping to spike sequences were removed for further analysis. Reads were then mapped onto the mm10 assembly of *Mus musculus* genome using STAR version 2.7.0f[42]. Gene expression quantification was performed from uniquely aligned reads using htseq-count version 0.9.1[43], with annotations from Ensembl version 96 and “union" mode. Read counts were normalized across samples with the median-of-ratios method[44], to make these counts comparable between samples and differential gene analysis were performed using the DESeq2 version 1.22.2[45]. All the figures were generated using R software version 3.5.3.

RT-qPCR

Complementary DNA was prepared using random hexamer oligonucleotides and SuperScript IV Reverse Transcriptase (Invitrogen) and amplified using LightCycler® 480 SYBR Green I Master (Roche) on a LightCycler® 480 II (Roche). Primers used for qPCR analysis are listed in Supplementary Table 8.

Repeat element analyses

Data were processed as already described[46] using Bowtie1[47] instead of Maq. The repeatMasker annotation was used to identified the different types of repeat elements (Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0. 2013-2015 http://www.repeatmasker.org). Differential

expression analyses were performed using DESeq2 (version 1.22.2)[45]. All the figures were generated using R custom scripts (version 3.5.3).

SLIC-CAGE analyses

Twenty-eight and 13 ng of total RNA isolated from P14 oocytes (biological replicate 1 and replicate 2, approximately 500-1000 oocytes pooled for each replicate) and 15 ng of total RNA isolated from P14 *Tbpl2*$^{-/-}$ mutant oocytes (approximately 550 pooled oocytes) were used for SLIC-CAGE TSS mapping[28]. Briefly, 5 µg of the carrier RNA mix were added to each sample prior to reverse transcription, followed by the cap-trapping steps designed to isolate capped RNA polymerase II transcripts. The carrier was degraded from the final library prior to sequencing using homing endonucleases. The target library derived from the oocyte RNA polymerase II transcripts was PCR-amplified (15 cycles for P14 WT, 16 cycles for P14 *Tbpl2*$^{-/-}$ mutant) and purified using AMPure beads (Beckman Coulter) to remove short PCR artifacts (< 200bp, size selection using 0.8 x AMPure beads to sample ratio). The libraries were sequenced using HiSeq2500 Illumina platform in single-end, 50 bp mode (Genomics Facility, MRC, LMS).

Sequenced SLIC-CAGE reads were mapped to the reference *M. musculus* genome (mm10 assembly) using the Bowtie2[41] with parameters that allow zero mismatches per seed sequence (22 nucleotides). Uniquely mapped reads were kept for downstream analyses using CAGEr Bioconductor package (version 1.20.0)[48] and custom R/Bioconductor scripts. Bam files were imported into R using the CAGEr package, where the mismatching additional G, if added through the template-free activity of the reverse transcriptase, was removed. Same samples sequenced on different lanes and biological replicates were merged prior to final analyses.

Promoter analyses

In order to consider only the CAGE TSS dependent only on TBPL2, we removed all the P14 WT CAGE tags at position where CAGE tags were also present in the P14 *Tbpl2*$^{-/-}$ mutant CAGE tags dataset: for the rest of the analysis, this dataset was called “TBPL2-specific-only” and we compared it to the P14 *Tbpl2*$^{-/-}$ mutant CAGE data (hereafter called “TBPL2-independent-only”). Briefly, a CAGE set object was created from the TBPL2-specific-only and TBPL2-independent-only CTSS files using CAGEr Bioconductor package (version 1.20.0)[48], data were normalized using normalizeTagCount (fitInRange = c(5,1000), alpha = 1.53, T = 1e6) and the powerLaw option. Cluster of CTSS were collected using clusterCTSS (threshold = 1, thresholdIsTpm = TRUE, nrPassThreshold = 1, method = "distclu", maxDist = 20, removeSingletons = TRUE, keepSingletonsAbove = 5). Width of the TSS regions was calculated using cumulativeCTSSdistribution and quantilePositions (clusters = "tagClusters", qLow = 0.1, qUp = 0.9): interquantile width corresponds to the 10th-90th percentile of the total tag cluster signal. In order to compare the different samples, consensus promoters were computed using aggregateTagCluster (tpmThreshold = 3, qLow = 0.1, qUp = 0.9, maxDist = 100). Self-organizing map (SOM) expression profiling was performed using getExpressionProfiles using a tpmThrshold of 3, the method “som”, xDim = 3 and yDim = 2. Shifting TSS were obtained after calculation of the cumulative distribution along the consensus clusters using cumulativeCTSSdistribution and calculation of the shift score using scoreShift with the Kolomogorov-Smirnov test. Shifting promoters were extracted using getShiftingPromoters (tpmThreshold = 3, scoreThreshold = -Inf, fdrThreshold = 0.01).

TSSs corresponding to the MaLR-ERVS were identified after annotation using HOMER (version 4.10)[49].

Sequences analyses were performed using seqPattern (version 1.14){Haberle:co} and R custom scripts. WW dinucleotides enrichment was computed using plotPatternDensityMap on -250/+250 regions centered on the dominant TSSs. TATA box position weight matrix (PWM)

matches analyses was performed using the MotifScanScores function applied on the -35/-20 sequences centered on the dominant TSSs, using the TBP PWM provided in the SeqPattern package (derived from the JASPAR data base). Distribution of the best match for each sequence was then plotted. Sequence Logo were created using SeqLogo (version 1.48.0)[50].

Methods references

33. Brou, C. *et al.* Distinct TFIID complexes mediate the effect of different transcriptional activators. *EMBO J* **12,** 489–499 (1993).

34. Nagy, Z. *et al.* The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. *Cell Mol Life Sci* **67,** 611–628 (2010).

35. Gegonne, A. *et al.* The general transcription factor TAF7 is essential for embryonic development but not essential for the survival or differentiation of mature T cells. *Mol Cell Biol* **32,** 1984–1997 (2012).

36. Sari, D. *et al.* The MultiBac Baculovirus/Insect Cell Expression Vector System for Producing Complex Protein Biologics. *Advances in Experimental Medicine and Biology* **896,** 199–215 (2016).

37. Bieniossek, C., Richmond, T. J. & Berger, I. MultiBac: multigene baculovirus-based eukaryotic protein complex production. *Curr Protoc Protein Sci* **Chapter 5,** Unit 5.20–5.20.26 (2008).

38. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 125–132 (2010).

39. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75,** 861–877 (2019).

40. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).

41. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

42. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2012).

43. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).

44. Anders, S. *et al.* Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat Protoc* **8,** 1765–1786 (2013).

45. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 31 (2014).

46. Fadloun, A. *et al.* Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat Struct Mol Biol* **20,** 332–338 (2013).

47. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).

48. Haberle, V., Forrest, A. R. R., Hayashizaki, Y., Carninci, P. & Lenhard, B. CAGEr: precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. - PubMed - NCBI. *Nucleic Acids Res* **43,** e51–e51 (2015).

49. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38,** 576–589 (2010).

50. *seqLogo: Sequence logos for DNA sequence alignments*. doi:10.18129/B9.bioc.seqLogo

Data availability

The datasets generated during the current study are available in different repositories: structure factors and coordinates for TBPL2^{C}; Protein Data Bank (PDB ID 6TH9), proteomic data; ProteomeXchange PRIDE database (PXD0316347), RNA-seq data; Gene Expression Omnibus database (GSE140090) and SLIC-CAGE data; ArrayExpress (submission in progress).

Acknowledgements

We thank D. Singer and A. Gegonne for the gift of the *Taf7*^{flox} mouse line and H. Stunnenberg for TFIIA antibodies. We would also like to thank D. Devys for critically reading the manuscript, all members of the Tora lab for thoughtful discussions and suggestions throughout the course of the work. We are grateful to I. Kukhtevich, M. Borsos, M.E. Torres Padilla, T. Gupta and L. Casini for advice and preliminary experiments. We thank C. Hérouard and M. Jung from the GenomEAST platform for library preparation and preliminary analyses, P. Eberling for peptide synthsis, F. Ruffenach for proteomic analyses, G. Duval for polyclonal antibody generation, the histology platform, the IGBMC cell culture facility and S. Falcone, M. Poirot and F. Memedov of the IGBMC animal facility for animal care taking. This work was supported by funds from CNRS, INSERM, and Strasbourg University. This study was also supported by the European Research Council (ERC) Advanced grant (ERC-2013-340551, Birtoaction) and grant ANR-10-LABX-0030-INRT (to LT) and a French State fund managed by the Agence Nationale de la Recherche under the frame program Investissements d'Avenir ANR-10-IDEX-0002-02. IB and FM acknowledge support by Wellcome Trust Senior Investigator awards (106115/Z/14/Z and 106955/Z/15/Z, respectively).

Authors' contribution

CY, SDV and LT designed the study; SDV and LT supervised the project; CY performed all molecular lab and mouse experiments, EG generated the anti-TBP2 polyclonal antibodies, LN carried

out the proteomic analyses, PH organized the SLIC-CAGE, NC carried out SLIC-CAGE analyses, NC and BL analysed the SLIC-CAGE data, and SDV analysed the proteomic, RNA-seq and SLIC-CAGE data. FM oriented the promoter analyses. KG performed X-ray crystallography and biochemical experiments, supervised by IB. CY, SDV, KG, IB, FM, and LT wrote the manuscript with contributions to manuscript text and figure legends from all authors. All authors gave final approval for publication.

Competing interests

The authors declare that they have no competing interests.

Additional Information

Supplementary information is available for this paper.

Correspondence and request for materials should be addressed to S. D. Vincent (vincent@igbmc.fr) or L. Tora (laszlo@igbmc.fr).

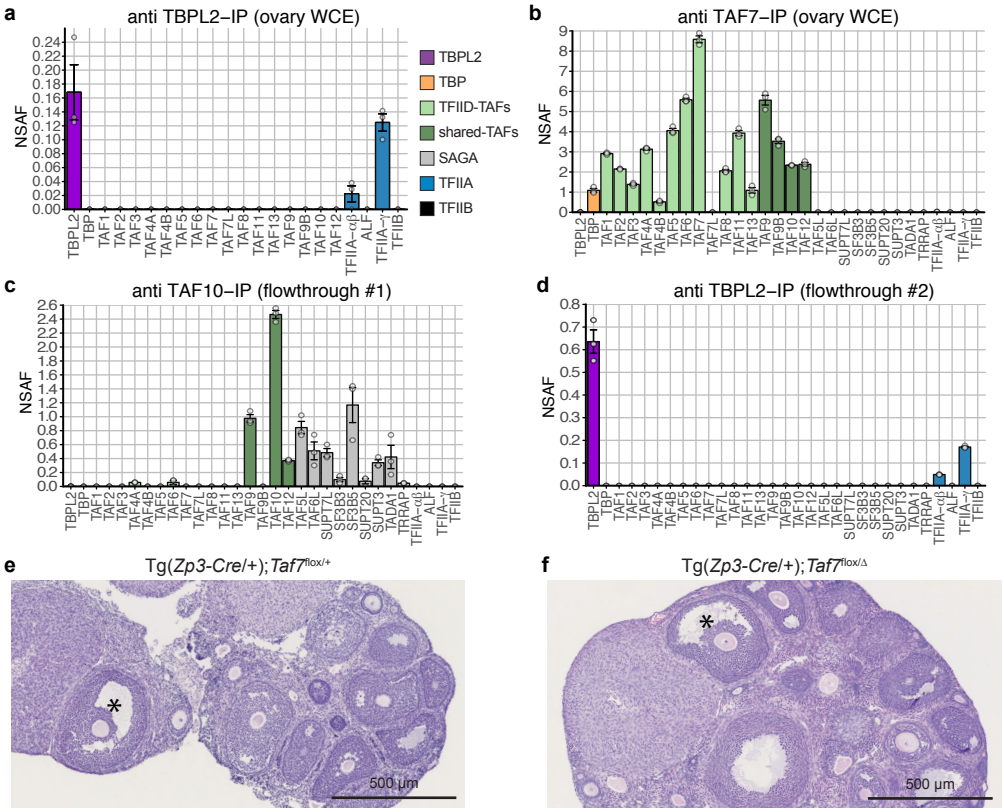

Fig. 1 | TBPL2 does not assemble in a TFIID-like complex during oocyte growth. a, anti TBPL2 immunoprecipitation followed by mass spectrometry (IP-MS) analysis from three biological replicates of mouse ovarian whole cell extracts (WCE). The colour code for the different proteins or complexes is indicated on the right. NSAF; normalized spectral abundance factor. **b-d,** Sequential IP-MS experiment from ovarian WCE (technical triplicates). Anti TAF7 IP-MS (**b**), followed by an anti TAF10 IP-MS (**c**) and then an anti TBPL2 IP-MS (**d**). The colour code is the same as in (**a**). **e, f,** Hematoxylin and eosin stained ovaries section from control (**e**) and oocyte-specific *Taf7* mutant (**f**) ovaries. The presence of antral follicles is indicated by an asterisk. Scale bars: 500 µm.

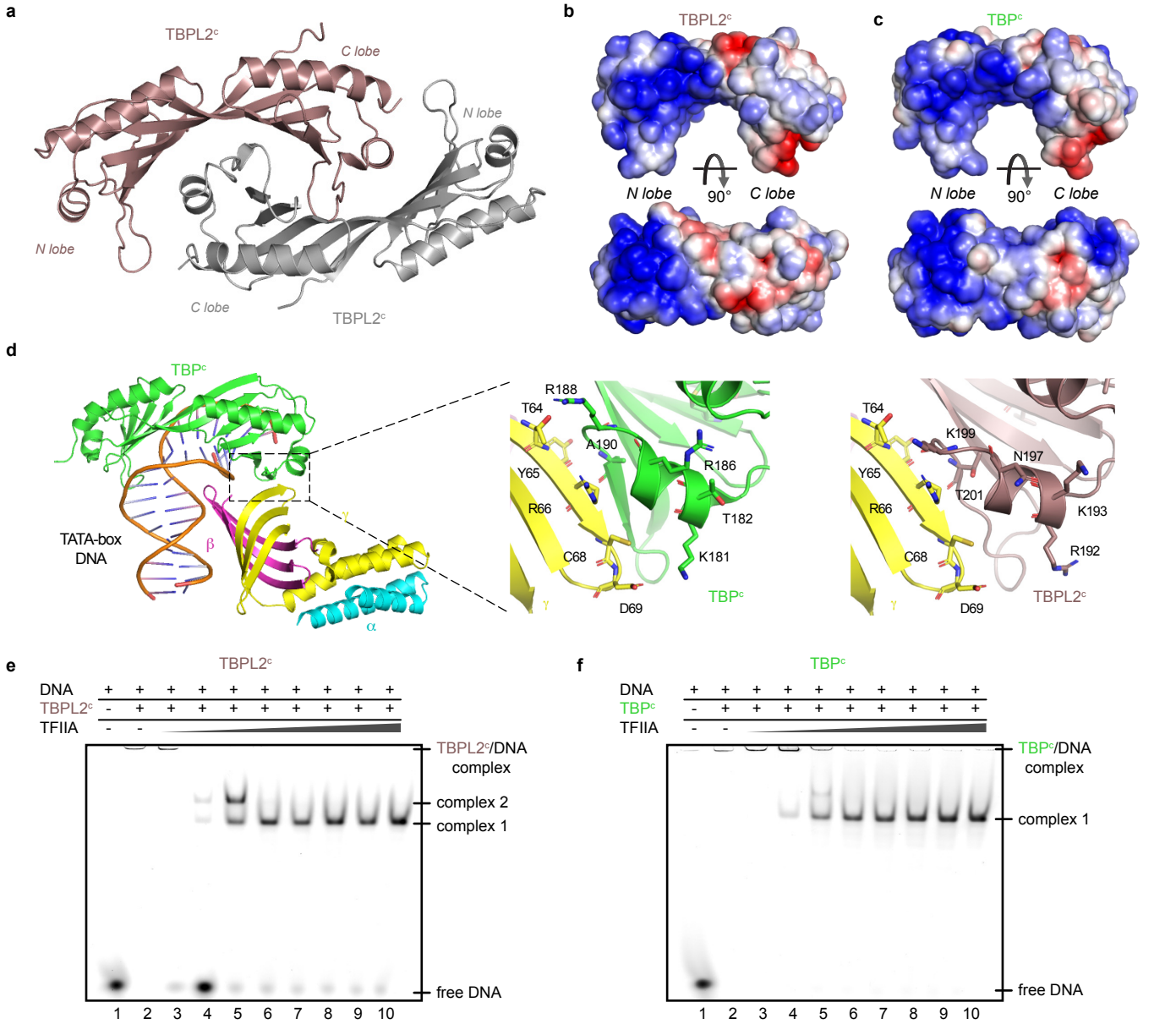

Fig. 2 | TBPL2 and TBP core domains are very similar but display distinct properties. a, X-ray crystal structure at 2Å resolution of mouse TBPL2 core (residues 170-349; TBPL2^{C}) is shown (brown). TBPL2^{C}, similar to TBP core (TBP^{C}), exists as a dimer in the crystal (monomers in brown and grey) . The N and C lobes of TBPL2^{C} are labelled. **b, c,** TBPL2^{C} (**b**) and TBP^{C} (from PDBID 1NVP, **c**) are shown in two views in an electrostatic surface charge representation (blue, basic; red, acidic) revealing significant differences. **d,** The crystal structure of a TBP/TFIIA/DNA complex (PDBID 1NVP) is shown on the left. This complex was used to model TBPL2^{C} interactions with the crystallized TFIIA core (TFIIA^{C}). The segments corresponding to the boxed-in section are depicted in zoom-ins on the right. TBP^{C} is coloured in green, TBPL2^{C} in brown, TFIIA^{C} α, β and γ chains in blue, magenta and yellow, respectively. Amino acids at the interfaces are labelled by their residue numbers. TATA DNA is shown as a ladder. **e, f,** Interactions of TATA DNA,full-length processed TFIIA and TBPL2^{C} (**e**) or TBP^{C} (**f**) analysed by EMSA . Bands corresponding to free DNA, TBPL2^{C}/DNA (**e**) or TBP^{C}/DNA (**f**) complexes, and the different complexes (complex 1, complex 2) are marked (n=2).

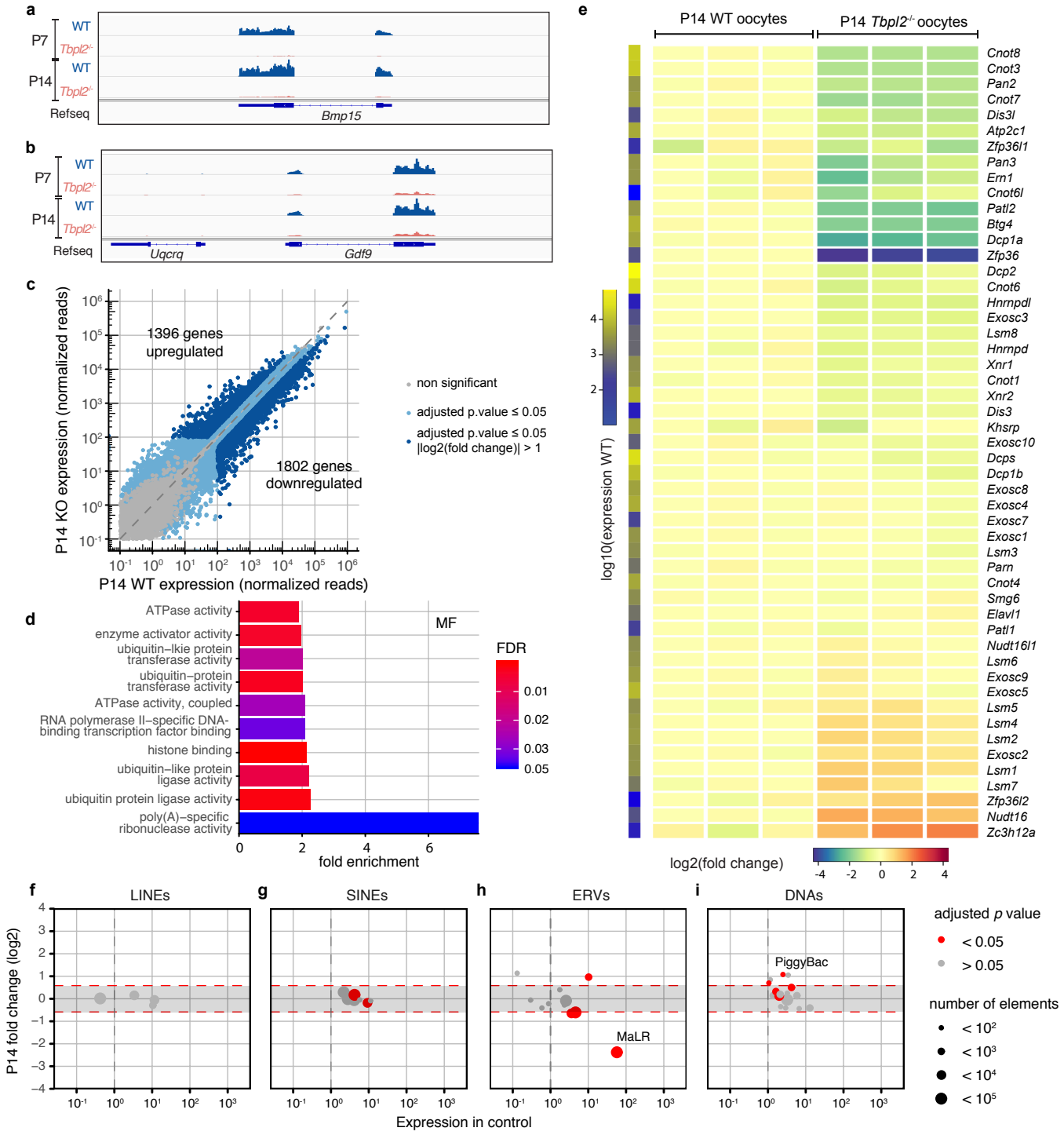

Fig. 3 | Expression of genes related to the mRNA deadenylation/decapping/decay pathways and of the MaLR endogenous retroviral elements are down regulated in growing *Tbpl2*$^{-/-}$ mutant oocytes. a, b, IGV genomic snapshots of *Bmp15* (**a**) and *Gdf9* (**b**). **c,** Expression comparison between wild-type (WT) and *Tbpl2*$^{-/-}$ mutant post-natal 14 (P14) oocytes (biological triplicates). Expression has been normalized to the median size of the transcripts in kb. Grey dots correspond to non-significant genes and genes with high Cook's distance, light-blue dots to significant genes for an adjusted *p* value ≤ 0.05 and dark-blue dots to significant genes for an adjusted *p* value ≤ 0.05 and an absolute log2 fold change > 1, after Wald test and Benjamini-Hochberg correction (DESeq2). The number of up- or down-regulated genes is indicated on the graph. **d,** Down-regulated genes GO category analyses for the molecular functions (MF). The top ten most enriched significant GO categories for a FDR ≤ 0.05 are represented. **e,** Heatmap of selected genes involved in mRNA decay, decapping or deadenylation pathways. Expression levels in fold-change (compared to the mean of WT) of three biological replicates of P14 WT and P14 *Tbpl2*$^{-/-}$ mutant oocytes are indicated. The fold change colour legend is indicated at the bottom. **f-i,** Differential expression between wild type and *Tbpl2*$^{-/-}$ mutant P14 oocytes of the different transposon classes; RNA transposon classes (LINEs (**f**), SINEs (**g**) and ERVs (**h**)) and DNA transposons (DNAs (**i**)). The ERV sub-class III mammalian apparent LTR retrotransposon (MaLR) family is the most severely affected in *Tbpl2*$^{-/-}$ mutant oocytes at P14.

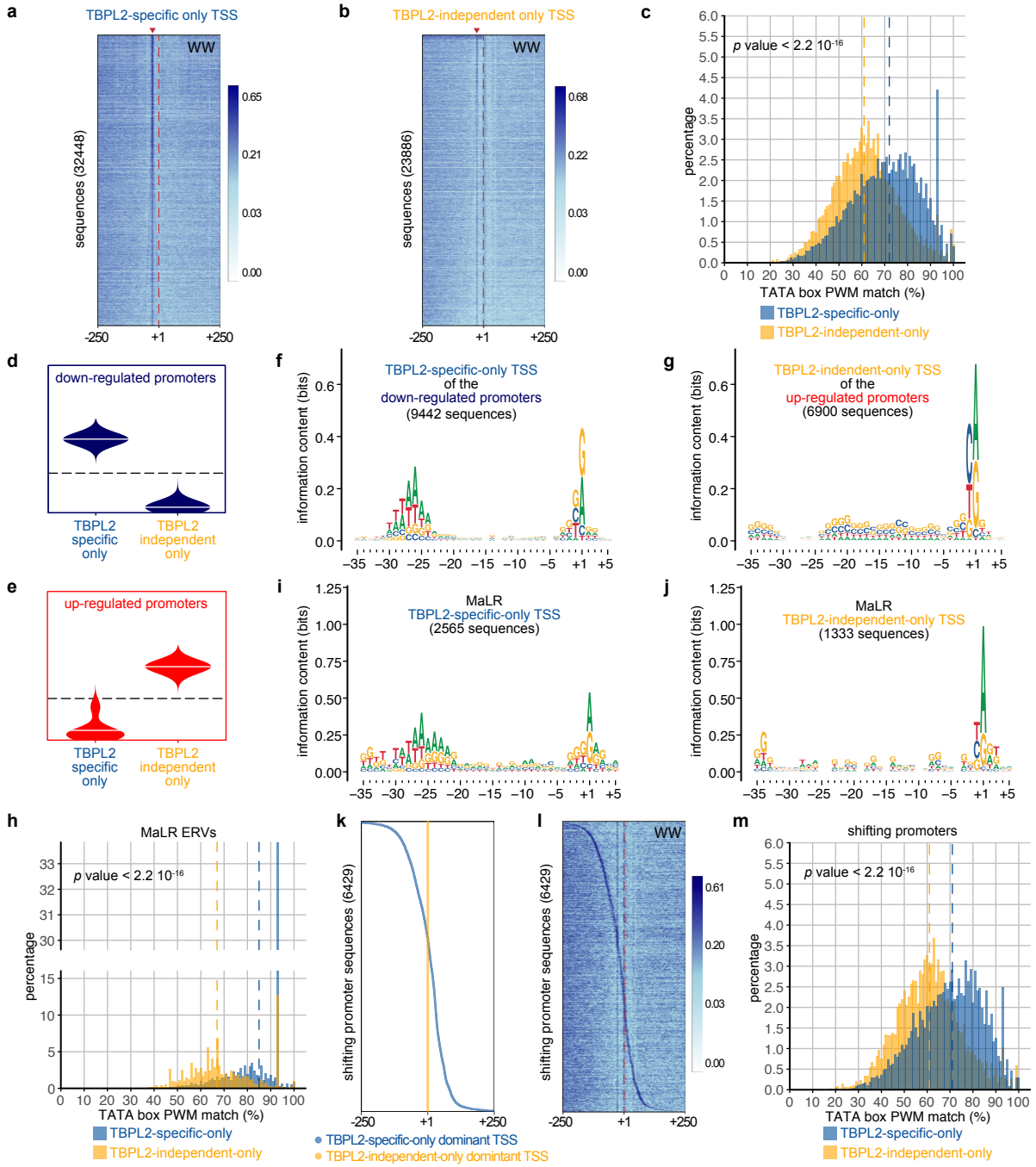

Fig. 4 | Core promoter regions of TBPL2 specific transcription units in P14 oocyte are enriched in TATA-like elements and are sharp. a, b, Genome-wide A/T-rich dinucleotide (WW) motif analyses of -250/+250 sequences centred on the dominant transcription start sites (TSS, position +1, dashed red line) of TBPL2-specific-only (**a**) and TBPL2-indendent-only (**b**) TSS clusters. The sequences have been ordered by increasing size of the width of each cluster (corresponding to the 10th-90th percentile of the total tag cluster signal). The red arrowheads indicate the WW enrichment at position -30 in the TBPL2-specific-only TSS clusters (**a**) and the equivalent position in the TBPL2-independent-only TSS clusters (**b**). The number of TSS clusters is indicated in brackets. **c,** Distribution of the best TATA box position weight matrix (PWM) matches within a -35 to -20 region upstream of the dominant TSSs (+1) of TBPL2-specific-only (blue) compared to the TBPL2-independent-only (orange) TSS clusters. The dashed lines indicate the median of the TATA box PWM matches for the TBPL2-specific-only (blue) and the TBPL2-independent-only (orange) TSS clusters (*p* value after a two-tailed Wilcoxon rank-sum test). **d, e,** Two selected self-organizing map (SOM) groups of the consensus TSS clusters: the down-regulated promoters (blue, **d**) and the up-regulated promoters (red, **e**) groups. **f, g,** Sequence logos of the -35/+5 sequence of the TBPL2-specific-only dominant TSSs from the down-regulated promoters (**f**) and of the TBPL2-independent-only dominant TSSs from the up-regulated promoters (**g**). **h,** Distribution of the best TATA box PWM matches within a -35 to -20 region upstream of the TBPL2-specific-only (blue) and TBPL2-independent-only (orange) MaLR ERVs dominant TSS. The dashed lines indicate the median of the TATA box PWM matches for the TBPL2-specific-only (blue) and the TBPL2-independent-only (orange) TSS clusters (*p* value after a two-tailed Wilcoxon rank-sum test). **i, j,** Sequence logo of the -35/+5 sequence of the MaLR ERVs TBPL2-specific only (**i**) and TBPL2-independent-only (**j**) dominant TSSs. **k, l,** Analysis of the TBPL2-specific-only versus TBPL2-independent-only shifting promoters within a -250/+250 region centred on the position of the TBPL2-independent-only TSS clusters (position +1 in (**k, l**) and red dashed line in (**l**)). Position of the TBPL2-specific-only (blue) and of the TBPL2-independent-only (orange) dominant TSSs for each shifting promoter sequence (**k**) and WW dinucleotide enrichment heat map (**l**) from the same set of sequences ordered as in (**k**). **m**, Distribution of the best TATA box PWM matches within a -35 to -20 region upstream of the TBPL2-specific-only (blue) and TBPL2-independent-only (orange) dominant TSS of the shifting promoters. The dashed lines indicate the median of the TATA box PWM matches for the TBPL2-specific-only (blue) and the TBPL2-independent-only (orange) shifting TSS clusters (*p* value after a two-tailed Wilcoxon rank-sum test).

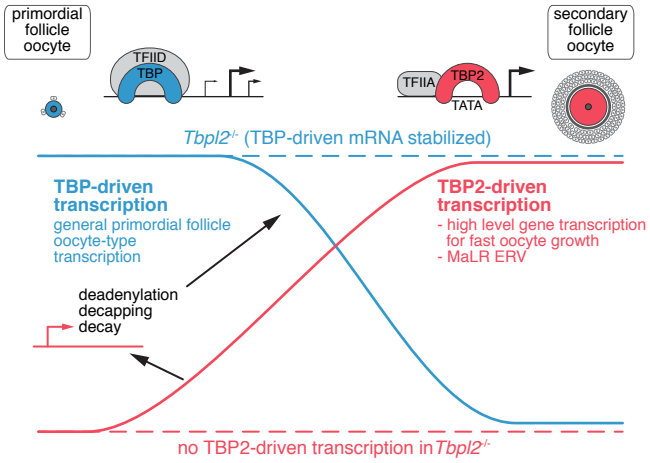

Fig. 5 | Transcriptome overhaul controlled by TBPL2/TFIIA during oocyte growth.