Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation

James A. Goodrich* and Robert Tjian

Abstract | The eukaryotic core promoter recognition complex was generally thought to play an essential but passive role in the regulation of gene expression. However, recent evidence now indicates that core promoter recognition complexes together with ‘non-prototypical’ subunits may have a vital regulatory function in driving cell-specific programmes of transcription during development. Furthermore, new roles for components of these complexes have been identified beyond development; for example, in mediating interactions with chromatin and in maintaining active gene expression across cell divisions.

Until recently, it was thought that a universal and highly conserved RNA polymerase II (RNAPII) core promoter recognition apparatus initiated transcription in all eukaryotic cells. Central components of the prototypical pre-initiation complex such as transcription factor IID (TFIID) — a complex of TATA-box-binding protein (TBP) and TBP-associated factors (TAFs) — were generally considered essential but passive partners that were destined to follow the regulatory instructions provided by sequence-specific activators and repressors. This view came in part from studying a limited set of cell types — for example, yeast, Drosophila melanogaster S2 cells and human HeLa cells — which divide rapidly and were preferred for practical reasons such as large-scale production for biochemical analysis or ease of genetic manipulation. In the few cases in which more differentiated cell types and tissues were used, they often comprised a mixture of cell types (for example, whole rat liver, calf thymus, Drosophila embryos). Furthermore, many experiments in the transcription field have used recombinant model genes and promoters and artificial regulators.

More recent studies have shifted towards an analysis of endogenous genes and physiologically relevant regulators observed in the context of nearly homogeneous populations of a single, specific, differentiated cell type and in distinct cell cycle stages. These studies have revealed the requirement for a number of ‘non-prototypical’ core promoter recognition factors for transcription, including cell-type-specific TAFs and TBP-related factors (TRFs). Furthermore, new functions of the prototypical core promoter recognition machinery have been identified. Here, we review these studies, which have implications for understanding gene regulation during somatic and germ cell development — for which these factors are increasingly being shown to regulate specific sets of genes — and reveal unexpected functions for core promoter recognition factors more generally in transcriptional regulation and the maintenance of gene expression states.

Prototypical core promoter recognition
Core promoter recognition is the first step in the general mechanism of transcription initiation. The primary prototypical core promoter recognition factor for eukaryotic mRNA genes is the general transcription factor TFIID, which binds multiple core promoter elements to begin the process of forming pre-initiation complexes containing RNAPII. RNAPII core promoters in higher eukaryotes are highly diverse and the core promoters of many genes do not contain any of the known core promoter elements. The most recognizable core promoter element is the TATA box, which divide rapidly and were preferred for practical reasons such as large-scale production for biochemical analysis or ease of genetic manipulation. In the few cases in which more differentiated cell types and tissues were used, they often comprised a mixture of cell types (for example, whole rat liver, calf thymus, Drosophila embryos).

More recent studies have shifted towards an analysis of endogenous genes and physiologically relevant regulators observed in the context of nearly homogeneous populations of a single, specific, differentiated cell type and in distinct cell cycle stages. These studies have revealed the requirement for a number of ‘non-prototypical’ core promoter recognition factors for transcription, including cell-type-specific TAFs and TBP-related factors (TRFs). Furthermore, new functions of the prototypical core promoter recognition machinery have been identified. Here, we review these studies, which have implications for understanding gene regulation during somatic and germ cell development — for which these factors are increasingly being shown to regulate specific sets of genes — and reveal unexpected functions for core promoter recognition factors more generally in transcriptional regulation and the maintenance of gene expression states.

Prototypical core promoter recognition
Core promoter recognition is the first step in the general mechanism of transcription initiation. The primary prototypical core promoter recognition factor for eukaryotic mRNA genes is the general transcription factor TFIID, which binds multiple core promoter elements to begin the process of forming pre-initiation complexes containing RNAPII (FIG. 1). RNAPII core promoters in higher eukaryotes are highly diverse and the core promoters of many genes do not contain any of the known core promoter elements. The most recognizable core promoter element is the TATA box, which divide rapidly and were preferred for practical reasons such as large-scale production for biochemical analysis or ease of genetic manipulation. In the few cases in which more differentiated cell types and tissues were used, they often comprised a mixture of cell types (for example, whole rat liver, calf thymus, Drosophila embryos). Furthermore, many experiments in the transcription field have used recombinant model genes and promoters and artificial regulators.

More recent studies have shifted towards an analysis of endogenous genes and physiologically relevant regulators observed in the context of nearly homogeneous populations of a single, specific, differentiated cell type and in distinct cell cycle stages. These studies have revealed the requirement for a number of ‘non-prototypical’ core promoter recognition factors for transcription, including cell-type-specific TAFs and TBP-related factors (TRFs). Furthermore, new functions of the prototypical core promoter recognition machinery have been identified. Here, we review these studies, which have implications for understanding gene regulation during somatic and germ cell development — for which these factors are increasingly being shown to regulate specific sets of genes — and reveal unexpected functions for core promoter recognition factors more generally in transcriptional regulation and the maintenance of gene expression states.

Prototypical core promoter recognition
Core promoter recognition is the first step in the general mechanism of transcription initiation. The primary prototypical core promoter recognition factor for eukaryotic mRNA genes is the general transcription factor TFIID, which binds multiple core promoter elements to begin the process of forming pre-initiation complexes containing RNAPII (FIG. 1). RNAPII core promoters in higher eukaryotes are highly diverse and the core promoters of many genes do not contain any of the known core promoter elements. The most recognizable core promoter element is the TATA box, which divide rapidly and were preferred for practical reasons such as large-scale production for biochemical analysis or ease of genetic manipulation. In the few cases in which more differentiated cell types and tissues were used, they often comprised a mixture of cell types (for example, whole rat liver, calf thymus, Drosophila embryos). Furthermore, many experiments in the transcription field have used recombinant model genes and promoters and artificial regulators.

More recent studies have shifted towards an analysis of endogenous genes and physiologically relevant regulators observed in the context of nearly homogeneous populations of a single, specific, differentiated cell type and in distinct cell cycle stages. These studies have revealed the requirement for a number of ‘non-prototypical’ core promoter recognition factors for transcription, including cell-type-specific TAFs and TBP-related factors (TRFs). Furthermore, new functions of the prototypical core promoter recognition machinery have been identified. Here, we review these studies, which have implications for understanding gene regulation during somatic and germ cell development — for which these factors are increasingly being shown to regulate specific sets of genes — and reveal unexpected functions for core promoter recognition factors more generally in transcriptional regulation and the maintenance of gene expression states.
Core promoter
The region of a gene to which RNA polymerase II and the general transcription factors (TFs) bind to initiate transcription. Core promoters span from approximately 40 base pairs upstream to 40 base pairs downstream of the transcription start site and are composed of DNA elements to which subunits of TFIID (or TFIIB) bind.

Pre-initiation complex
The assembly of general transcription factors and RNA polymerase II on core promoter DNA. This complex, which can be assembled in the absence of nucleotide triphosphates in vitro, is competent to initiate transcription in the presence of nucleotides.

Transcription factor IID (TFIID). A transcription factor for RNA polymerase II that binds core promoters. The TFIID complex is composed of the TATA-box-binding protein (TBP) and 1 or 14 TBP-associated factors.

TATA-box-binding protein (TBP). The central subunit of transcription factor IID. TBP binds TATA boxes found in the core promoters of some eukaryotic mRNA genes.

TBP-associated factor (TAF). All subunits of the transcription factor IID (TFIID) complex other than TATA-box-binding protein (TBP) are TAFs. There are 1 or 14 TAFs in the prototypical TFIID complex. There are also several proteins with sequence similarity to the prototypical TAFs, which are referred to as non-prototypical TAFs.

Core promoter recognition factor
A protein or multi-subunit complex that binds with sequence specificity to core promoter elements. The prototypical core promoter recognition factor for mRNA genes in eukaryotes is transcription factor IID, subunits of which recognize multiple core promoter elements.

Figure 1 | Core promoter recognition by TFIID. Multiple subunits of the transcription factor IID (TFIID) complex bind core promoter elements. TATA-box-binding protein (TBP) binds TATA boxes. TBP-associated factor 1 (TAF1) and TAF2 bind the initiator element (Inr). TAF6 and TAF9 bind the downstream promoter element (DPE).

Germ cell differentiation

**TRF2 functions in germ cell differentiation.** TRF2 (also known as TLF, TLP, TRP or TBPL1) was independently discovered by sequence similarity to TBP in organisms ranging from Caenorhabditis elegans to humans. Although highly similar in sequence to TBP, TRF2 does not bind to the TATA box. In mammals, TRF2 protein is most highly expressed in testes, although it is also found at lower levels in other tissues and a primary defect of Trf2 knockout mice is a deficiency in spermatogenesis. Expression of TRF2 is tightly controlled in developing spermatocytes, with high levels of expression first occurring in step 6 to 7 spermatids. In Trf2 knockout mice, defects in spermatogenesis first appear at step 7, differentiation to elongated spermatids does not occur and transcription of multiple post-meiotic, testis-specific genes is severely decreased. Further investigation showed that early-stage spermatids in these mice have a defect in formation of the chromocenter, which is a structure containing condensed centromeric heterochromatin that is involved in chromatin organization. It is not yet clear whether the deficiency of spermatogenesis in the Trf2 knockout mice is due entirely to altered gene expression and chromocenter formation.

In addition to function in germ cell differentiation, TRF2 has been found to function in embryogenesis in some organisms. For example, TRF2 is required for early embryogenesis and for the proper transcription of many genes in C. elegans and Xenopus laevis. TRF2 functions both during germ cell differentiation and during the onset of metamorphosis in Drosophila, in which mutations in Trf2 result in alterations in the levels and temporal delays in transcription of genes normally regulated by edysone during early metamorphosis. Indeed, it seems likely that as more studies are carried out, the specific functions of these factors may turn out to be utilized in diverse contexts in different organisms, although germ cell development may initially emerge as a primary process regulated by atypical TRFs and TAFs.

**TRF3 functions in germ cell differentiation.** TRF3 (also known as TBPL2) was first identified during a search of the initial draft of the human genome for predicted proteins with sequence similarity to human TBP. The similarity between TRF3 and TBP is limited to their carboxy-terminal DNA-binding domains, which are nearly identical. By contrast, the amino-terminal domains share little sequence similarity. Genes encoding TRF3 have also been found in the mouse, frog and zebrafish genomes, but not those of Drosophila spp. and C. elegans. TRF3 was characterized biochemically and was found to bind TATA boxes, to interact with TFIID and TFIIH, and to direct basal transcription, properties that it shares with TBP. Hence, TRF3 can be thought of as a replacement for TBP, with respect to these activities.

In early studies, TRF3 protein was shown to be expressed at various levels in many human and mouse tissues and cell lines. However, other groups have found TRF3 mRNA to be most highly expressed in the testis and ovary of zebrafish and X. laevis, and in oocytes...
of mice. The reason for these different observed expression levels and patterns is not known and requires further investigation, although the use of different antibodies and extraction methods in different laboratories may partly account for this.

In mice, it is clear that TRF3 is highly expressed in the ovary. Specifically, the TRF3 protein is found in the nuclei of mouse oocytes during folliculogenesis and levels decrease during ovulation. Interestingly, TBP was not detected in oocytes during the stages of folliculogenesis where TRF3 was found in the nucleus. A significant proportion of the most strongly downregulated genes are oocyte specific, indicating that TRF3 is required for normal levels of transcription in developing oocytes. Chromatin immunoprecipitation (ChIP) assays confirmed that TRF3 occupies the promoters of a number of these genes in developing oocytes in wild-type mice (FIG. 2b), suggesting that it effectively replaces TBP in directing transcription during folliculogenesis.

Non-prototypical TAFs in germ cell differentiation. Several non-prototypical TAFs also have roles in germ cell differentiation. The protein encoded by the Drosophila cannonball gene, which is a homologue of Drosophila TAF5, is selectively expressed in primary spermatocytes and is needed to establish proper levels of transcription of multiple genes involved in spermatid differentiation. An additional four alternative TAFs are expressed in primary spermatocytes: no hitter, a homologue of TAF4; meiosis I arrest, a homologue of TAF6; spermatocyte arrest, a homologue of TAF8; and ryan express, a homologue of TAF12 (REF. 34). All four of these testis-specific TAFs are required for proper spermatocyte differentiation, and they control the expression of a common set of genes involved in the differentiation process.

Further investigation showed that testis-specific TAFs are localized at the promoters of genes involved in spermatid development. The presence of wild-type
testis-specific TAFs at the target promoters in these studies correlated with reduced occupancy of polycomb repressive complex 1 (PRC1), a chromatin modifying complex that mediates transcriptional silencing, and increased H3K4me3, a mark for active transcription. Together, these observations suggest that Drosophila testis-specific TAFs function during transcriptional activation of developmental genes in primary spermatocytes, probably by reducing levels of PRC1 found at the promoter. Exactly how the TAFs reduce PRC1 levels is not fully understood. However, much of the pool of testis-specific TAFs localizes with PRC1 in a subcompartment of spermatocyte nucleioli, and this localization requires the testis-specific TAFs\(^{36}\). A proportion of the testis-specific TAFs might function to re-localize PRC1 away from genes required for spermatid development, thereby de-repressing their transcription.

**TAF4b** was initially discovered as a cell-type-specific TAF, uniquely found in a human B-cell line\(^{36}\). However, when its expression was analysed in mouse tissues, it was found to be most highly expressed in the testes and ovary\(^{37}\). In the mouse ovary, TAF4b mRNA is uniquely localized to the granulosa cells of the ovarian follicle, and knockout of TAF4b resulted in infertile females with smaller ovaries that lack mature follicles\(^{37}\) (FIG. 2b). Further studies showed that TAF4b promotes granulosa cell proliferation and is required for the survival of these cells\(^{36}\). *Taf4b* knockout mice have reduced expression of many ovarian-specific genes, and the overall gene expression programme in ovaries from young knockout mice appear similar to those of aged wild-type mice, which is consistent with an observed acceleration in ovary ageing in the knockout mice\(^{37,39}\). Overexpression of TAF4b in a spontaneously immortalized rat granulosa cell line significantly increased the expression of a range of genes, notably including c-Jun\(^{40}\). This increase seems to be cell-type specific, as it was not observed in NIH/3T3 cells overexpressing TAF4b. ChIP assays showed that TAF4b and c-Jun co-localized to the promoters of several genes, including that of c-Jun itself. Moreover, TFIID complexes containing TAF4b have higher transcriptional activity on the c-Jun promoter in a reconstituted transcription system than is achieved with the prototypical TFIID complex\(^{41}\). Together, these studies suggest that TAF4b works with a specific transcriptional activator, c-Jun, to control the transcription of genes involved in follicle growth.

TAF4b is also involved in spermatogenesis in the mouse, in which it is expressed in spermatids in adults\(^{42}\) (FIG. 2a). Male *Taf4b* knockout mice are initially fertile, but become infertile as they age; the testes degenerate and germ cells are lost. Furthermore, multiple genes involved in spermatogenesis are expressed at lower
levels in Taf4b knockout mice. It will be interesting to determine whether TAF4b also interacts with a specific transcriptional activator to control spermatogenesis.

Finally, a paralogue of TAF7, TAF7L, is also involved in male germ cell development in mice. TAF7L is expressed in male germ cells throughout differentiation and is found in the nucleus of spermatids. As germ cell development progresses, TAF7L expression increases, which correlates with increased TBP expression and decreased TAF7 expression. In spermatocytes, TAF7L is associated with TBP, whereas TAF7 is not. Knockout of the Taf7l gene in mice leads to the development of deformed sperm, although the mice are fertile. Gene expression profiling revealed six transcripts that were markedly decreased in testis from the knockout mice, although their relevance to spermatogenesis was not studied.

**Somatic cell differentiation**

How are transcriptional programmes established in the diverse differentiation pathways that occur in multicellular organisms? And how do these transcriptional programmes direct specific differentiation pathways? Research aimed at answering these questions has largely focused on transcriptional activators and repressors, which are known to play crucial roles in directing cell-type-specific transcription. More recently, however, non-prototypical core promoter recognition factors have also been shown to be important for directing differentiation.

**TRF3 and TAF3 function in embryogenesis.** Decreased Trf3 levels in zebrafish embryos were found to cause defects in mesoderm patterning. In addition, during X. laevis embryogenesis, TRF3 is required for gastrulation for the normal expression of nearly 900 genes in the developing X. laevis embryo. Thus, a protein considered to be a non-prototypical core promoter recognition factor controls a transcriptional programme involving a large number of genes.

Further studies of Trf3 during zebrafish embryogenesis revealed that it is required for haematopoiesis (Fig. 3a). Expression profiling and ChIP assays showed that Trf3 was necessary for the developmentally regulated transcription of the mespa gene, which encodes a basic helix–loop–helix transcription factor that is vital for haematopoiesis. The mespa promoter was bound by Trf3, but not by TBP. Depletion of mespa resulted in developmental defects that were similar to those observed with depletion of Trf3 and, importantly, ectopic expression of mespa in the Trf3-depleted embryo restored normal development. Ultimately it was shown that the differentiation of mesoderm into the haematopoietic lineage involved binding of Trf3 to the mespa promoter. A more recent study found that Trf3 functions in conjunction with Taf3 to initiate haematopoiesis in the zebrafish embryo. Taf3 binds Trf3, associates with the mespa promoter, and is required for haematopoiesis. Extension of this analysis to mice showed that TRF3 and TAF3 bind the promoter of the Mesp1 gene, the mouse orthologue of the zebrafish mespa gene, and that TRF3 is also required to initiate haematopoiesis in a mouse embryonic stem cell model system.

There is evidence to suggest that the coupling of TRF3 and TAF3 to facilitate a specific differentiation pathway is not limited to haematopoiesis. In studies of TFIID in skeletal muscle differentiation, the levels of prototypical TFIID subunits (TBP, TAF1 and TAF4) were dramatically reduced when mouse C2C12 myoblasts were induced to differentiate into myotubes. These observations held true in myoblasts and myotubes and are found together at the promoter of the myogenin gene (Myog), which encodes a transcriptional activator that is critical for myogenesis.
TRF3, which is worth noting that in zebrafish, mouse embryonic stem cells and F9 cells, TRF3 could normally function in myogenesis and haematopoiesis, nor in zebrafish haematopoiesis.

ChIP assays showed that TAF3 and TRF3 occupy the myogenin promoter in myotubes, but not in myoblasts. Subsequent studies showed that TAF3 and TRF3 can mediate activation of transcription from the myogenin promoter by MyoD in a reconstituted transcription system. The ability of TAF3/TRF3 to direct transcription from the myogenin promoter was independent of Mediator subunits, which are also depleted in myotubes. Together, these results led to a model in which there is a profound reorganization of the RNAPII promoter recognition machinery during myogenesis; TFIID subunits are decreased and TAF3/TRF3 directs transcription of genes encoding key myogenic transcriptional regulators.

There is an ongoing debate with regard to the significance of these results from cell lines. Knockout of Trf3 in mice did not seem to affect the development of skeletal muscle or blood as might have been anticipated given the documented functions of TRF3 in myogenesis in the C2C12 cell model, and its function in haematopoiesis in zebrafish, mouse embryonic stem cells and F9 cells. There are multiple possible explanations for these apparently contradictory results. TRF3 could normally function in myogenesis and haematopoiesis in developing mice, but when knocked out another factor could compensate for the loss of TRF3 in myogenesis and haematopoiesis but not in embryogenesis. The compensatory factor, however, would not be able to replace the function of TRF3 in the cellular model systems for myogenesis and haematopoiesis, nor in zebrafish haematopoiesis.

Additional experiments (for example, expression of potential compensatory factors in cells or inactivation of potential factors in the Trf3 knockout mice) could test this possibility. With respect to this possibility it is worth noting that in X. laevis oocytes, ectopically expressed TBP can replace TRF3 in driving transcription from specific promoters. Alternatively, muscle and blood phenotypes might only be displayed in the Trf3 knockout mice under highly specific conditions. For example, the Trf3 knockout mice could have impaired wound healing capabilities, which requires muscle stem cells to differentiate in a timely manner. It is also possible that the commonly used cellular model systems are not reliable models for some aspects of myogenesis and haematopoiesis. In general, it can be difficult to prove that a specific factor is not involved in a particular complex process given the redundancies of living systems. Clearly, additional experiments using different strategies need to be performed to test the various possibilities regarding the role of TRF3 in mouse myogenesis, haematopoiesis and other developmental pathways.

**Unique functions of TAF8 and TAF10 in development.** A non-prototypical TAF has a key role during adipogenesis, although in this case a wholesale rearrangement of the core promoter recognition machinery does not seem to occur. Roeder and colleagues found that levels of many prototypical TFIID TAFs decrease when 3T3-L1 preadipocytes are induced to differentiate into adipocytes, with the exception of the non-prototypical factor TAF8 [Ref 50]. By contrast, TAF8 was not detected in preadipocytes, but was upregulated during adipogenesis and was associated with TFIID.

TAF8 contains a histone fold domain in its N-terminal region that interacts with other TAFs that contain histone folds, which probably facilitates the association of TAF8 with the TFIID complex. Overexpression of the TAF8 histone fold domain blocked the differentiation of 3T3-L1 preadipocytes to adipocytes, perhaps by blocking the association of endogenous TAF8 with TFIID. This effect could be reversed by overexpression of full-length TAF8 [Ref 50]. Importantly, expression of the TAF8 histone fold domain repressed expression of peroxisome proliferator-activated receptor-γ and CCAAT enhancer binding protein-α, two regulators of adipogenesis. Together, these observations support a model in which TAF8 association with TFIID during the differentiation of preadipocytes to adipocytes stimulates transcription of genes required for adipogenesis, although the mechanism by which this occurs is not understood.

Knockout of Taf10, a well-characterized subunit of TFIID, leads to embryonic lethality in mice. Conditional inactivation of TAF10 in the embryonic liver caused a dramatic reduction in the size of the liver, suggesting that TAF10 is required for liver development. When TAF10 was conditionally inactivated in the livers of adult mice, the TFIID complex disassembled, although subunits other than TAF10 were still present in liver cells. Transcript profiling revealed that inactivation of TAF10 in the developing liver affected expression of only 11% of genes, the majority of which are hepatocyte specific, which is the likely cause of the defects observed in liver development upon TAF10 inactivation.

**Non-prototypical functions beyond development.** Several additional examples of new functions involving core promoter recognition factor has emerged in the past several years. The common theme that ties the three examples below together is that a single core promoter recognition factor was discovered to have a new activity. The three factors, which have already been discussed in this Review, are TFR2, TAF3 and TBP, the primary prototypical core promoter recognition factor.

**Drosophila TRF2 controls transcription of many genes.** Tjian and colleagues immunopurified Drosophila TRF2 and found that it associates with NURF, a nucleosome remodelling factor, and DREF, a transcription factor that controls the expression of cell cycle and proliferation genes. Biochemical and cell-based experiments showed that TRF2 and DREF direct transcription from the proliferating cell nuclear antigen (PCNA) gene, which encodes a protein required for high fidelity DNA replication; the promoter of the PCNA gene contains a DRE (the binding site for DREF). In a later study, Drosophila TRF2 was found to be required for histone H1 transcription and TRF2 occupied the TATA-less histone H1 promoter. Interestingly, TRF2 did not...
occupy the promoters of core histone genes that are directly adjacent to the histone H1 gene within the same repeating histone cluster; instead, the promoters of the four core histone genes were occupied by TBP. This is interesting because the amounts of the linker histone H1 relative to the core histones can vary dramatically in cells during development.65,66.

ChIP-chip experiments identified more than 1,000 genes occupied by TRF2 in D. melanogaster S2 cells, including a cluster of ribosomal protein genes for which co-regulation requires TRF2 (REF 54). Comparison with TBP ChIP-chip data showed that the genes bound by TRF2 and those bound by TBP were largely non-overlapping (80%). Moreover, while the majority of the TBP-bound genes contained a TATA box, the TRF2-bound genes were nearly all devoid of a TATA box. Knock down of Trf2 in salivary glands caused severe growth defects, which were mostly consistent with TRF2 normally functioning to promote cell growth during development. These studies show that TRF2 has a major role in controlling mRNA transcription throughout the Drosophila genome to regulate genes functioning in cell growth, arguing that the term non-prototypical or specialized might not be entirely appropriate for this factor.

TAF3 anchors TFIIID to nucleosomes. H3K4me3 at promoters is considered a mark of actively transcribed genes.67,68. Timmers and colleagues discovered that the human TFIIID complex binds H3K4me3 (REF 59), an interaction that is mediated by the TAF3 subunit, which contains a PHD finger that selectively binds H3K4me3 even when present in nucleosomes. Functional assays showed that the PHD finger of TAF3 can mediate transcriptional activation in a histone methyltransferase-dependent manner.69. Other modifications on histone H3 can affect the association of TFIIID with H3K4me3. For instance, asymmetric dimethylation of R2 inhibits binding of TFIIID to H3 with the K4me3 mark, whereas acetylation of K9 and K14 augments TFIIID binding to H3 with the K4me3 mark. The latter case is interesting, as H3K9Ac and H3K14Ac are associated with active promoters, and TAF1 has a double bromodomain known to bind acetyl-lysine, with a preference for diacetylated histone H4 (REFS 60,61).

These observations provide strong support for TFIIID binding histone modifications in promoter regions, and highlight the possibility that TFIIID is in part recruited to promoters containing specific histone modifications through the combined actions of TAF1 and TAF3. Perhaps the core promoter should be considered as not just DNA, but instead an ensemble of DNA elements and associated histones with specific modifications. In other words, it is this chromatin nucleoprotein complex containing the core promoter DNA that must be recognized by TFIIID.

**TBP bookmarks active genes during mitosis.** During mitosis transcription ceases, but the chromatin at the promoters of genes that were active before mitosis remains uncompacted or ‘bookmarked’ so that when cells enter G1 these same genes again become transcriptionally active. TBP, perhaps as part of TFIIID, is involved in the process of mitotic bookmarking. Early studies showed that at least some portion of the TBP and TAFs remain bound to chromatin during mitosis, whereas other factors such as RNAPII do not.65–66.

Recently, Sarge and colleagues further investigated the role of TBP in bookmarking (FIG. 4). They found that in extracts from mitotic cells TBP binds both protein phosphatase 2A (PP2A) and the CAP-G subunit of condensin, which inactivates condensin and keeps the core promoters from being tightly compacted in mitotic chromatin. This allows transcription of the bookmarked genes to be reactivated when cells exit mitosis and enter G1. During mitosis RNA polymerase II (RNAPII) leaves chromatin and genes are transcriptionally inactive. Whether any general transcription factors (GTFs) other than TBP remain at bookmarked genes is unknown.

---

**Mitotic bookmarking**
The process by which genes that are active before mitosis are marked such that transcription begins again at these genes when cells exit mitosis and enter the G1 phase of the cell cycle.

---

Figure 4 | TBP bookmarks genes during mitosis. In mitotic nuclei, TATA-box-binding protein (TBP) binds protein phosphatase 2A (PP2A) and the CAP-G subunit of condensin at the core promoters of genes that were active before mitosis. PP2A dephosphorylates CAP-G, which inactivates condensin and keeps the core promoters from being tightly compacted in mitotic chromatin. This allows transcription of the bookmarked genes to be reactivated when cells exit mitosis and enter G1. During mitosis RNA polymerase II (RNAPII) leaves chromatin and genes are transcriptionally inactive. Whether any general transcription factors (GTFs) other than TBP remain at bookmarked genes is unknown.
These observations raise many intriguing questions. Is TBP in this context working as part of TFIIID or in a different partnership with some other, as yet, undefined core promoter recognition complex? How many promoters require TBP for bookmarking and do they all contain TATA boxes? Does TBP, which is also involved in RNAPI and RNAPIII transcription, function in bookmarking promoters of genes that are transcribed by these polymerases? Do promoters of inactive genes that naturally harbour promoter-proximal paused polymerases also utilize TBP for bookmarking? Perhaps understanding the nature of the active pre-initiation complex at the histone loci will provide insight into the mechanism of bookmarking.

Conclusions

Two unanticipated themes have emerged from studies of non-prototypical core promoter recognition factors. The first is the different roles that cell-type-specific TAFs and TRFs have in potentiating developmental gene regulation and cellular differentiation. The second is the apparent dispensability of canonical core promoter recognition components in some terminally differentiated mature cells that continue to actively synthesize cell and tissue-specific mRNAs as was seen in the C2C12 model of skeletal muscle\(^1\).\(^{18}\). In the next few years, it will be interesting to see how many different mature cell types utilize these mechanisms to determine cell identity. It is possible that novel core promoter recognition factors might function to guide cell determination and direct programmes of gene expression in terminally differentiated adult cell types.

Indeed, recent studies suggest that loss of prototypical TFIIID subunits and potential replacement by non-prototypical TAFs may be a prevalent mechanism that can be observed in both liver and fat cell differentiation \textit{in vitro} and \textit{in vivo} (J. D’Alessio, H. Zhou and R.T., unpublished observations). Future studies using a combination of loss of function (short hairpin RNA depletion) or gain of function (ectopic expression) for non-prototypical core promoter recognition factors will shed light on this issue. For example, to gain more insight into the differential roles of variant core promoter machinery in different cell types, a combination of short hairpin RNA knock down, mouse knock out and ectopic expression of key factors (for example, TRFs and TAFs) would be used. In some cases, one might even imagine that ectopic overexpression of one set of TAFs/core factors may drive transdifferentiation of one precursor cell type into a different fate. Which cell systems and which \textit{in vitro} models will best serve to establish the physiological functions of cell-specific core components is difficult to predict, but almost certainly a more comprehensive picture will require purified cell types, \textit{in vitro} biochemical systems and animal studies. Clearly, a significant challenge will be the availability of homogeneous and scalable cell types. Here, we expect that rapidly evolving single cell biochemistry technology and high-resolution imaging might provide a useful experimental avenue.

One can imagine that not all cell types will necessarily use both of these mechanisms in the dramatic manner that has been observed for differentiated myotubes, which involves both the loss of TFIIID and the retention and utilization of one select TAF (TAF3), possibly in conjunction with the non-prototypical TRF3. Indeed, we would expect that with hundreds, if not thousands of cell types in higher animals, there might very well be various alternative mechanisms that take advantage of unique combinations of the TRFs and TAFs to work in concert with specific activators and repressors. When we drill down into multiple cell types to access their core transcription apparatus, we may see a complex palate of pre-initiation complexes at work. In some cases the prototypical TFIIID subunits may have become largely dispensable, whereas in other cases select cell-type-specifically expressed TAFs or TRFs may have assumed a more commanding role, while still functioning in collaboration with prototypical TFIIID components and other cofactors such as Mediator. For example, in ovarian follicles, TAF4 is replaced by TAF4b in the mostly intact TFIIID complex\(^4\).\(^{55}\). We might also imagine that in some mature cell types that may need to be rapidly reactivated to replicate (that is, immune activation of B cells and T cells) TFIIID levels may stay largely unchanged but one or more TAF subunit or TRF could become functionally more important.

It will be intriguing to see whether some terminally differentiated cells that no longer replicate continue to use the prototypical TFIIID as part of the pre-initiation complex. In such cases, might these differentiated cells then modify the canonical core machinery and highjack their usual promoter recognition functions by substituting or adding new cell-specific subunits, as has been documented in the case of TAF4b in the granulose cells of the mammalian ovary? In other situations in which mature cells have exited the cell cycle, if one or two TAF/TRF factors are retained or even upregulated, while the bulk of the TFIIID complex becomes largely eliminated, might different cell types enlist distinct sets of orphan TAFs and TRFs to meet their specific transcriptional needs? For example, instead of TAF3, which was used in myotubes, could some other orphan TAF become the key player in liver, fat, neurons, and so on?

We might also envisage that the coordination between transcription factors of all types, including those we have described here, must integrate with chromatin remodeling and modifying factors to initiate cell-specific programmes of transcription and maintain the specified differentiated state throughout the life of the mature cell. In this respect, we are particularly curious about the presence and activities of core promoter recognition factors during the formation of induced pluripotent stem (iPS) cells. Here, terminally differentiated cells that in some cases have probably jettisoned subunits of the prototypical TFIIID complex must subsequently re-activate the myriad sets of genes encoding core promoter factors as well as DNA replication genes in order to become self-renewing iPS cells. Specifically, it will be interesting to learn the fate of TFIIID, TAFs and TRFs as distinct differentiated cells are forced to become iPS cells.
Hopefully, over the next several years, some of the questions raised above will be addressed, as core promoter recognition in specific cell types is elucidated using a more powerful repertoire of modern molecular and cellular imaging tools, such as super high-resolution microscopy for single cell analysis, and in vitro biochemistry and cell-based assays in microfluidic chambers. If we are to understand what defines a cell type and how the many different functional cell types are derived, we must gain a more complete picture of the molecular players and transcriptional mechanisms controlling cell fate and identity. We anticipate that the core promoter recognition factors and functions discussed here as non-prototypical will themselves be considered prototypical as additional studies are performed in an expanding set of cells and differentiation pathways.

Presumably, the mechanisms by which non-prototypical core machinery might direct cell-specific programs of transcription have not been determined in any detail. A potential model might be reminiscent of the bacterial sigma hypothesis, in which new core promoter recognition subunits of the holo RNA polymerase can recognize and bind to a specific subset of promoters designated for sporation or specific phase gene expression. It would not be difficult to imagine that novel TAF and TRF complexes can recognize and bind a distinct set of promoters containing unique, yet unidentified, elements and that various cell-type-specifically expressed activators or repressors have evolved to target co-activator domains of these alternative core complexes as a means of expanding and diversifying the repertoire of combinatorial transactions. It is also possible that non-prototypical TAFs have the capacity to interact with specifically marked chromatin components and thereby regulate transcription activity by yet unknown mechanisms.
REVIWES


43. Pointud, J. C. et al. The intracellular localisation of TAF7L, a paralogue of transcription factor TFIID Initiation of zebrafish haematopoiesis by the

44. Cheng, Y. male germ-cell differentiation. TAF7L, a paralogue of transcription factor TFIID

45. Falender, A. C. correlaute with promoter selectivity.


57. Demonstrated that TAF3 binds a specific post-translational modified form of histone H5 (H5K4me3) and recruits TFIIID to chromatin containing this modification. Importantly, H5K4me3 is a known mark for transcriptionally active genes. These studies prompt us to reconsider the definition of a core promoter to include not just DNA elements, but also the specific post-translational modifications on histones associated with core promoter regions.


66. Showed that TBP functions in mitotic bookmarking by binding PP2A and the CAP-G subunit of condensin at core promoters of genes that are active prior to mitosis. This assembly causes dephosphorylation of CAP-G, inactivation of condensin and blocks the compaction of chromatin containing these core promoters. Once cells exit mitosis, transcription of the bookmarked genes is reactivated.


Acknowledgements

We apologize to those whose relevant research we were unable to discuss owing to space limitations. J.A.G. was partly funded by the Bookmarking Group at the National Cancer Institute.

Competing interests statement

The authors declare no competing financial interests.

DATABASES


TAFL TAFL TAFL TAFL TAFL TAFL TAFL TAFL TAFL TAFL TAFL TAFL TAFL TAFL TAFL

TAFL TAFL TAFL

FURTHER INFORMATION


ALL LINKS ARE ACTIVE IN THE ONLINE PDF