Dissertation

submitted to the
combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by

Diplom-Biologe Marco Ferg
born in Karlsruhe
Oral examination: 30.06.2008

Large scale- and functional analysis for the requirement of TBP-function in early zebrafish development

Referees: Prof. Dr. Uwe Strähle
Prof. Dr. Jochen Wittbrodt

Zusammenfassung

Die differentielle Expression proteinkodierender Gene in spezifischen Zelltypen und während der Embryonalentwicklung beruht auf der Interaktion von Transkriptionsfaktoren mit regulatorischen Sequenzen im proximalen Promoter. Diese Interaktion gewährleistet die Erzeugung einer Vielzahl an Expressionsmustern. Kürzlich veröffentlichte Ergebnisse, die sich mit der Architektur des Kernpromoters auf Genomebene befassen, deuten jedoch darauf hin, das nicht nur der proximale Promoter, sondern auch der Kernpromoter eine aktive Komponente transkriptionsregulatorischer Prozesse ist, die zur differentiellen Genexpression beiträgt. TBP, das TATA bindende Protein ist ein Schlüsselelement der Polymerase II abhängigen Transkriptionsinitiation. Es wurde angenommen, das dieser generelle Transkriptionsfaktor zur Transkription aller Polymerase II transkribierten Gene benötigt wird. Neuere Studien deuten jedoch darauf hin, daß die Transkriptionsinitiation durch Proteine die in naher Verwandtschaft zu TBP stehen, wie zum Beispiel TLF oder TBP2, komplementiert werden kann. Es ist sehr wahrscheinlich, daß diese TBP verwandten Proteine Anwendung finden, um eine differenzierte Regulation der Genexpression zu ermöglichen.

In der vorliegenden Arbeit wird der Frage nach der regulatorischen Funktion von TBP während der frühen Entwicklung des Zebrafisches nachgegangen und die Architektur des Kernpromoters untersucht, der diese differenzierte Antwort bewerkstelligt.

Um die Regulation der Expression TBP unabhängiger Gene auf der Promoterebene zu untersuchen, wurde der Promoter von *notail*, einem Transkriptionsfaktor aus der T-box Familie funktionell analysiert. Um die Eigenschaften von Genen zu bestimmen, die TBP Abhängigkeit zeigen und die Charakteristika der ihnen zugehörigen Promotoren in großem Umfang zu analysieren, wurde das TBP abhängige Transkriptom durch eine Microarray Analyse untersucht und nachfolgend die Promotoren dieser Gene durch bioinformatische Methoden bestimmt und charakterisiert.

Die hier vorgestellte Arbeit kann zeigen, daß nur ein bestimmter Anteil aller untersuchten Gene funktionelles TBP zur Expression benötigt. TBP hat eine spezifische Funktion in der Degradation maternaler RNA, die über den miR-430-pathway abgebaut werden. Die erzielten Resultate deuten darauf hin, das TBP eine herausragende Rolle im Übergang von einem transkriptionell inaktiven Zustand zu einer transkriptionell aktiven Phase während der Zebrafischentwicklung zukommt und eindeutige Funktionen in der Transkriptionsregulation innerhalb der Zebrafischentwicklung wahrnimmt. Die bioinformatische Charakterisierung von Promotoren, die durch TBP reguliert werden, sowie die funktionelle Analyse des notail Promoters weisen darauf hin, das die TATA box, das DNA-Element, welches mit TBP in der Kernpromoterregion interagiert, kein bestimmendes Merkmal TBP abhängiger Transkriptionsinitiation darstellt. Obgleich die Sequenzeigenschaften TBP unabhängiger Transkriptionsmechanismen weiterhin unklar sind, weist die hier vorgelegte Studie darauf hin, daß Präinitiationskomplexe den Promoter durch dessen Architektur, das Vorhandensein von bestimmten Sequenzmotiven erkennen. Dies läßt vermuten, daß der Kernpromoter eine Schlüsselfunktion in der differenzierten Regulation der Genexpression während der frühen Embryonalentwicklung vertebrater Organismen einnimmt.

Abstract

The differential expression of protein coding genes in specific cell types and during development requires the interaction of transcription factors with regulatory sequences in the proximal promoter to generate diverse expression patterns. Recent approaches shedding light into the architecture of core promoters in large scale indicate however, that not only the proximal promoter, but also the core promoter is an active component of regulatory processes leading to differential gene expression. TBP, the TATA binding protein, a key regulator of Polymerase II transcription initiation was thought to be recruited to the promoter of all Polymerase II transcribed genes and required for transcriptional activity in vertebrates. Recent studies however suggest that transcription initiation can be complemented by TBP related proteins like TLF and TBP2. These factors are likely to be utilised to establish differential gene expression for the regulation of various developmental or physiological pathways.

In this thesis I address the differential regulatory function of TBP and the core promoter architecture facilitating this differential response in the complexity of the vertebrate organism by exploiting the experimental advantages of the zebrafish embryo model system. To better understand the promoter regulation of genes independent of TBP-function, the promoter of *notail*, a transcription factor of the T-box family was analysed in functional assays. To elucidate the nature of genes requiring TBP-function and to gain insight into the characteristics of proximal promoters and their regulation en masse in relation to the function of a general transcription factors, the zebrafish transcriptome was analysed for TBP-dependence in a microarray approach followed by a bioinformatic analysis to identify and analyse the promoters of a large number of zebrafish genes.

The work presented here demonstrates that only a proportion of genes require TBP-function in early zebrafish development and that TBP has a specific role in the clearance of maternal RNAs, that includes the miR-430 pathway. These results indicate that TBP plays a major role in the transition from a transcriptionally inactive state to a transcriptionally active phase of the zebrafish embryo and has distinct functions in regulating gene expression during development. Furthermore, the bioinformatic characterisation of promoters regulated by TBP, as well as the functional analysis of the *notail* promoter, indicate that the TATA box, the core promoter motif TBP binds to, is not the defining feature of TBP-dependent transcription initiation mechanisms. Although the sequence requirements for TBP-independent transcription initiation mechanism remain unclear, the study suggests that, based on the motif composition, pre-initiation complexes differentially recognize promoters, marking the core promoter a key component of transcriptional regulation in vertebrate development.

Parts of the work presented here have been published as:

Ferg M, Sanges R, Gehrig J, Kiss J, Bauer M, Lovas A, Szabo M, Yang L, Straehle U, Pankratz MJ, Olasz F, Stupka E, Muller F (2007) The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish. The EMBO journal 26: 3945-3956

Table of contents

Abbreviations

Abbreviation	-	Meaning
BLAST		basic local alignment tool
bp	-	basepair
BRE	-	TFII-B recognition element
CAGE	-	cap analysis of gene expression
CFP	-	cyan fluorescent protein
DBTSS	-	database of transcriptional startsites
DCE	-	downstream core element
downreg	-	downregulated RNA levels
DPE	-	downstream promoter element
DRE	-	DNA-replication related element
EST	-	encoded sequence tag
fc	-	fold change
FGF	-	fibroblast growth factor
GTF	-	general transcription factor
hpf	-	hours post fertilisation
inr	-	Initiator
kb	-	kilo-basepair
low var.	-	low variable changes in RNA levels
MBT	-	midblastula transition
miR	-	microRNA
MO	-	morpholino (antisense oligonucleotide)
MTE	-	motif ten element
ntl	-	*notail*
PIC	-	pre-initiation complex
Pol	-	Polymerase
PPR	-	putative promoter region
RACE	-	rapid amplification of cDNA ends
RT	-	reverse transcriptase
seq.	-	sequence
shh	-	*sonic hedgehog*
spec.	-	specific
TAF	-	TBP associated factor
TBP	-	TATA binding protein
TBP2	-	TATA binding protein 2
TF	-	transcription factor

TFTC	-	TBP free TAF containing complex
TLF	-	TBP like factor
TLS-box	-	TATA like sequence box
TRF	-	TBP related factor
TRF2	-	TBP related factor 2 (TLF)
TRF3	-	TBP related factor 3 (TBP2)
TSS	-	transcriptional startsite
upreg	-	upregulated RNA levels
UTR	-	untranslated region
wt	-	wildtype
xbra	-	*Xenopus brachyury*
xtbp	-	*Xenopus tbp* mRNA
YPF	-	yellow fluorescent protein
c MO	-	control morpholino
TBP MO	-	TBP morpholino

1.0 Introduction

1.1 Zebrafish as a model organism

In the late 1970s, Georg Streissinger from the University of Oregon was looking for a model organism that would allow him to study the genetic control of development in vertebrates and found it in the popular aquarium fish zebrafish (*Danio rerio*). The teleost has various advantages for studies in development and genetics in a vertebrate organism. The embryos are transparent, pigmentation starts not until 24 hours post fertilisation and, in conjunction with the *ex utero* development, makes the embryo easily accessible for microscopic observations and experimental manipulations. The development of the embryo from a zygote to a free-swimming larva is very fast. Within two days all major organs are developed and within 3 month the generation cycle is completed. Furthermore, zebrafish is highly reproductive and easy to breed. Under optimal conditions each female can lay up to 200 eggs per week.

1.1.1 Zebrafish - a model for genetic and gene regulation studies

Systematic genome-wide mutagenesis screens for embryonic phenotypes carried out in two big screens lead to the identification of almost 4000 mutations in genes affecting early zebrafish development (Driever *et al*, 1996; Haffter *et al*, 1996) and laid the foundation to study elementary processes in the early development of a vertebrate embryo. The majority of mutant screens in zebrafish have employed chemical mutagens like N-ethyl-N-nitrosourea (ENU) to induce mutations. The development of insertional mutagenesis for the zebrafish using mouse retroviral vectors made it easy to identify and clone the mutated gene through the viral tag (Amsterdam *et al*, 1999). To date, zebrafish lines carrying mutations in more than 10.000 genes are even commercially available. The large-scale identification of mutations also laid the groundwork to establish zebrafish as a model for human diseases. In the last decade zebrafish models have been established to elucidate the molecular mechanisms of human diseases like cardiovascular defects, muscle- and neural disorders, haematopoiesis and cancer (Amsterdam, 2006).

The post genomic era is characterised by approaches that try to elucidate the mechanisms of gene regulation on the promoter- and enhancer level. Comparative genomics has served as an essential guide in the identification of functional non-coding sequences in vertebrate genomes. As the sequencing and assembly of the zebrafish genome is nearing completion, human–zebrafish sequence comparisons are playing a fundamental role to uncover non-coding elements and efficiently test putative enhancer regions functionally in a vertebrate organism. In recent years techniques have been established that allow the efficient

study of gene function and regulation using transgenesis and enhancer traps through retroviral (Amsterdam & Becker, 2005; Gaiano *et al*, 1996) and transposon insertions (Kawakami, 2004; Kawakami *et al*, 1998; Parinov *et al*, 2004).

In addition, knockdown techniques using morpholino antisense oligonucleotides make it possible to inactivate genes of interest or do gain/loss of function experiments by the microinjection of mRNA. In conjunction with gene expression profiling tools like microarays and tiling arrays, zebrafish is an ideal model to study the regulation of gene expression.

The most powerful and unique feature of the zebrafish however is that it is a vertebrate model organism in which large-scale screens can be performed. These are not limited to the forward genetic screens for mutants, but also include chemical genetic screens on wildtype zebrafish to elucidate biological pathways or screening zebrafish disease models for drug development (Murphey & Zon, 2006). The emerging development of computer algorithms recognising tissues and changes of reporter signals within these tissues allows the automated screening for compounds having a specific effect on certain tissues (Tran *et al*, 2007) or to study promoter/enhancer interactions in large scale using high throughput systems (Kalmar *et al*, to be published).

Taken together, the model organism zebrafish, established to study embryonic- and organ development, has become a valuable resource to form disease models, drug targets and insight into pathways of gene regulation applicable to human development and disease.

1.1.2 Embryonic development of zebrafish

After fertilisation, the cytoplasm streams toward the animal pole to form the first cell of the blastodisc (Fig.1A). 45 minutes later the discoidal cleavage starts with the formation of the second cell. The first ten cleavages are synchronous, producing regular tiers of blastomeres (Fig.1B) that form a blastula sitting on top of the yolk sac (Fig.1C). At this stage of development, around the tenth cell cycle, 3 hours post fertilisation the mid blastula transition (MBT) takes place, which is characterised by a lengthening of the cell cycle, that until the MBT takes place, lacks gap phases, a loss of cell synchrony and the onset of cell motility (Kane & Kimmel, 1993). Subsequent cleavages and epiboly movements, where the cells of the blastodisc migrate over the yolk sac to cover it, leads to gastrulation around 5.5 hours post fertilisation (Fig.1D, E). During gastrulation involution at or near the blastoderm margin occurs. This movement folds the blastoderm into two cellular layers, the epiblast and hypoblast, within a ring (the germ ring) around its entire circumference (Warga & Kimmel, 1990). In addition, cells converge at the future dorsal side of the embryo and extend towards the animal pole to form the shield, the equivalent of the Speeman organiser. The formation

of the shield makes it possible to determine the dorsoventral axis of the embryo. The gastrulation process is giving rise to the 3 different germ layers of the embryo and is finished by reaching the tailbud stage 10 hours post fertilisation. After establishing the germ layers, segmentation starts (Fig.1F). This stage in development is characterised by the formation of the somites, which will give rise to muscle, blood and the sclerotome of the vertebra. Furthermore neurulation takes place, subdividing the ectodermal neural plate into a structured neural tube. At the end of segmentation, around 24 hours post fertilisation, the majority of the primordial organs are specified and the embryos start to move inside the chorion (Fig.1G). The segmentation is followed by the pharyngula stage were the bloodstream gets visible, pigmentation starts and the fins start to form. 48 hours post fertilisation the embryos hatch and 5 days post fertilisation the freely swimming larvae start to feed.

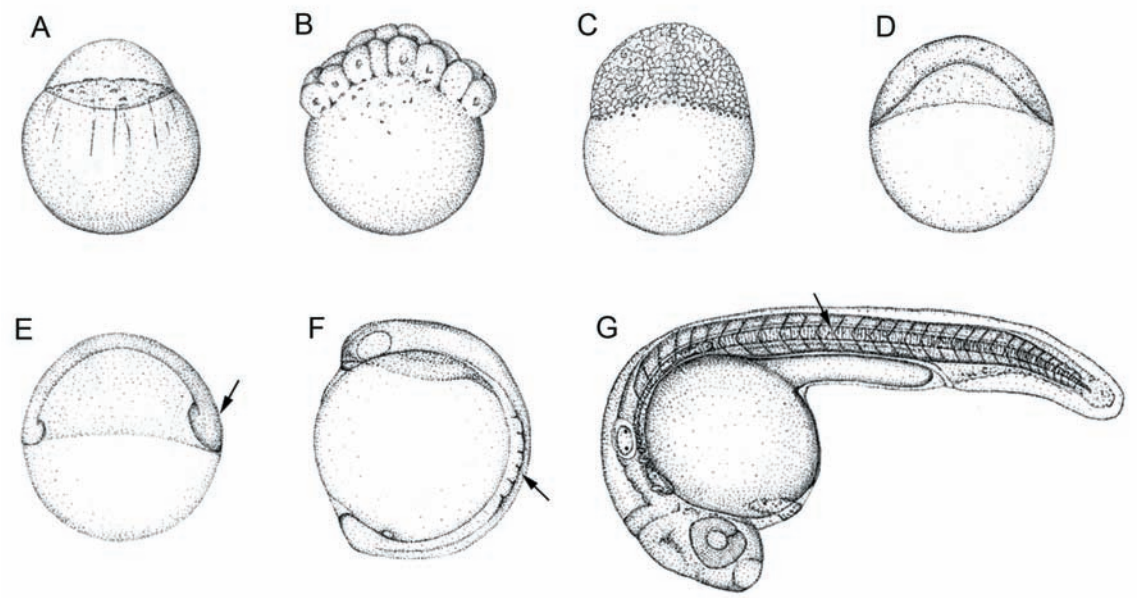

Figure 1. Selected stages of zebrafish embryonic development. (A) 1-cell stage (0.2 hpf). (B) 64-zell stage (1.75 hpf). (C) 1k-cell stage (3 hpf). (D) 30 % epiboly (4.7 hpf). (E) Shield stage (5.7 hpf), the arrow points to the embryonic shield, the dorsal organiser. (F) 6-somite stage (12 hpf), the arrow points at the forming somites. (G) Prim-5 stage (24 hpf), the arrow points to the notochord. Modified from Kimmel *et. al* 1995. Abbreviations, hpf, hours post fertilisation.

1.1.3 The mid blastula transition in zebrafish

To achieve the striking increase in cell number after fertilisation, many animals devote their early development to rapid and synchronous cell cycles (O'Farrell *et al*, 2004). These animals include model organisms like *Xenopus* (Newport & Kirschner, 1982a; Newport & Kirschner, 1982b) *Drosophila* (Robbins, 1984) and zebrafish (Kane & Kimmel, 1993). In zebrafish, the midblastula transition begins at cell cycle 10. It is characterised by cell cycle lengthening, loss of cell synchrony, appearance of cell motility and activation of the zygotic genome (Kane & Kimmel, 1993). Several mechanisms have been proposed to be involved in this process. MBT begins when the cleaving cells reach a particular nucleo-cytoplasmatic

ratio, because MBT is delayed in haploid embryos and early in tetraploid and polyploid embryos and suggested that the MBT is triggered by the DNA through titration of suppressor components present in the egg. (Newport & Kirschner, 1982a; Newport & Kirschner, 1982b). However, the key factors of this repressor model are still unknown. The activation of the zygotic genome has also been implicated to a large excess of histones repressing gene activity during early development through a dynamic competition between chromatin assembly and transcription complex assembly, as titration of chromatin components permits the establishment of stable transcription during early development (Prioleau *et al*, 1994). It was also suggested that gene-specific DNA methylation triggers MBT. In *Xenopus*, high levels of methylated DNA were observed in both paternally and maternally derived chromosomes. Between blastula and gastrula stages, a loss of methylation at individual *Xenopus* gene promoters that are activated at MBT could be detected (Stancheva *et al*, 2002). The TATA binding protein TBP has also been suggested to play a role in MBT. Pre-incubation of a reporter plasmid with TBP leads to relieve of the repression of transcription in pre-MBT stages, implicating that TBP is rate limiting in the activation of the zygotic genome (Prioleau *et al*, 1994). Although the above-mentioned findings account for some aspects of MBT, they do not explain why several genes are already activated in pre-MBT stages (Mathavan *et al*, 2005). These observations suggest that some of the mechanisms underlying zygotic genome activation remain to be discovered (Schier, 2007). Moreover, mammals do not seem to have a MBT. For example the mouse genome is activated at the two-cell stage (Thompson *et al*, 1998).

As the zygotic genome in pre-MBT stages is inactive, early embryonic development depends on maternal gene products deposited in the egg. Studies in *Drosophila* using chromosomal ablation combined with microarrays indicate that only one third of zygotically active genes are not expressed maternally (De Renzis *et al*, 2007). The activation of the zygotic genome is accompanied by the degradation of these maternal mRNAs (Bashirullah *et al*, 1999; De Renzis *et al*, 2007; Giraldez *et al*, 2006; Mathavan *et al*, 2005). However the degradation of a large proportion of these transcripts already starts before the activation of the zygotic genome in several analysed organisms (Bashirullah *et al*, 1998; Mathavan *et al*, 2005). In mouse it was estimated that around 60% of the maternal transcripts are degraded before activation of the zygotic genome, a finding, which suggests for a transcription independent degradation mechanism (Alizadeh *et al*, 2005). In *Drosophila*, two pathways for maternal RNA degradation have been identified. The first pathway is driven by maternally encoded factors that are recruited by cis-acting RNA degradation elements independently of whether the transcript is translationally active or translationally repressed. This maternal degradation apparatus is conserved in *Xenopus* oocytes and early embryos. The second, zygotic pathway becomes active 2 hours after fertilization. Either pathway acting alone is

sufficient to eliminate maternal transcripts. However, the joint action of both pathways is necessary for elimination of transcripts prior to the MBT (Bashirullah *et al*, 1999). The transcription dependent degradation of maternal mRNA species is partially mediated by sequences in the 3' untranslated region (3' UTR) that represent binding sites for regulatory RNAs and proteins, inducing deadenylation by marking these RNAs for degradation by nucleases (Schier, 2007). Recently this decay has been linked to microRNAs. In zebrafish the microRNA miR-430 has been identified to bind to a target sequence in the 3' UTR of several hundred maternally provided transcripts and promotes their deadenylation and subsequent degradation (Giraldez *et al*, 2006).

One answer for the raised question, why the embryo transcribes genes when the corresponding maternal transcript is present, could be the need to control the precise temporal and spatial expression of these transcripts during development (De Renzis *et al*, 2007). Furthermore, the degradation pattern of maternal transcripts shows that not all maternally deposited mRNAs degrade before or soon after the activation of the zygotic genome, but persist until late segmentation stages, suggesting for a function in post MBT stages (Mathavan *et al*, 2005; Pelegri, 2003).

1.2 The general transcription machinery and general cofactors

1.2.1 General transcription factors and the assembly of the pre-initiation complex

DNA dependent RNA polymerase II (Pol II) in itself is not sufficient to initiate transcription, not even from the strongest promoter. To position Pol II at the core promoter and initiate accurate transcription, several additional factors are needed (Roeder, 1991). This requirement was first described in fractionation studies of human cell extracts (Matsui *et al*, 1980). As these factors are thought to be needed in most genes that are transcribed by Pol II, they are termed general transcription factors (GTF.) At present 6 GTF are characterised and isolated: TFII-A, -B, -D, -E, -F, and –H. They all show high conservation from yeast to man (reviewed in Orphanides *et al*, 1996). The assembly of these factors into the pre-initiation complex (PIC) is the first step in the transcription of a gene. In the most conventional model, transcription initiation is characterised by a series of events. The first step is the recognition of core promoter elements by TFII-D through its subunits TATA binding protein (TBP) and TBP associated factors (TAFs) (reviewed in Roeder, 1991). This binding of TFII-D can be stabilised by TFII-A, however TFII-A is not essential for the assembly of the PIC (Zawel & Reinberg, 1993). TFII-B binds to the complex through direct interaction with TBP and with DNA sequences adjacent to the TATA-box (Lagrange *et al*, 1998; Nikolov *et al*, 1995). TFII-B, like TFII-A is able to stabilise weak TBP-TATA interactions (Imbalzano *et al*, 1994) and is directly involved in RNA-Pol II-TFII-F recruitment as well as startsite selection by RNA-Pol II (Zawel &

Reinberg, 1995). A pre-formed complex of TFII-F and RNA-Pol II binds to the assembled platform of TFII-D and TFII-B present on the DNA through direct interactions of both TFII-F and the polymerase with TFII-B (Zawel & Reinberg, 1993). The next step in assembling the PIC is the binding of TFII-E through direct interactions with Pol II, having a function in promoter melting (Pan & Greenblatt, 1994). The last step is binding of TFII-H to the now complete complex and is mediated by direct interactions with TFII-E (Zawel & Reinberg, 1993). The function of TFII-H has been described to phosphorylate the carboxy-terminal domain (CTD) of RNA-Pol II, the activated form of the polymerase (Feaver *et al*, 1991) that is able to start transcription after the dissociation from the PIC (Roeder, 1996).

1.2.2 An unexpected diversity of promoter recognition complexes

In vitro transcription assays on chromatographic fractions from cell extracts identified TFII-D to be required for RNA polymerase II and binds to the TATA box in the promoter region (Hernandez, 1993; Matsui *et al*, 1980) TFII-D proved very difficult to purify, but in a reconstituted transcription assay, an activity purified in yeast was able to substitute for the human TATA box-binding factor BTF1 (Cavallini *et al*, 1989). This activity, which turned out to be TBP, is able to mediate basal RNA polymerase II transcription, but unlike TFII-D, is not able to respond to transcriptional co-activators (Hoey *et al*, 1990; Pugh & Tjian, 1990; Smale *et al*, 1990). This finding led to the hypothesis that transcriptional activation requires activities beside TBP present in the TFII-D fraction (Pugh & Tjian, 1990). Biochemical assays identified TFII-D to be composed of TBP and a number of TBP associated factors (TAFs) (Dynlacht *et al*, 1991). Since then, many TAFs have been shown to interact with transcriptional activators. Human TAF4 and TAF11 for example have been implicated in transcriptional activation by nuclear receptors (Caron *et al*, 1997; Mengus *et al*, 1997) and TAF7 has been shown to interact with multiple transcriptional activators like Sp1 and YY1 (Chiang & Roeder, 1995). However TAFs are not only co-activators of transcription, they are also able to bind specific elements on the DNA and recruit TFII-D to the promoter. So far a number of TAFs have been shown to interact with the promoter, underlining the importance of TFII-D as a promoter recognition complex especially in TATA-less promoters (Pugh & Tjian, 1991). TAF1 and TAF2 have been implicated in initiator recognition (Chalkley & Verrijzer, 1999). The downstream promoter element DPE can be cross-linked to both *Drosophila* TAF6 and TAF9 (Burke & Kadonaga, 1997) and it was demonstrated that TAF1 interacts with the downstream core element DCE in a sequence-dependent manner (Lee *et al*, 2005). Recently it could be shown that TFIID via its subunit TAF3 directly binds the trimethylated form of histone H3 (H3K4me3), which is indicated to be the hallmark of active promoters (Vermeulen *et al*, 2007). Furthermore asymmetric dimethylation of H3R2 selectively inhibits TFIID binding to H3K4me3. These findings indicate for a crosstalk between histone modifications and the

transcription machinery in the core promoter region. It has also been observed that some components of TFII-D are present in a substoichiometric ratio relative to other TAFs, suggesting that these subunits are expressed in a tissue specific manner (Bell & Tora, 1999) and have specific functions in the gonad (Falender *et al*, 2005; Freiman *et al*, 2001), male gametogenesis (Hiller *et al*, 2001), adipogenesis (Guermah *et al*, 2003) and apoptotic cell death (Bell *et al*, 2001). Furthermore, TFIID complexes exist that vary in the number of TAFs assembling the TFIID complex. *in vitro* studies observed TFIID complexes with and without TAF10 in a single human cell (Brou *et al*, 1993).

The above-mentioned studies illustrate the multi-protein complex TFIID as a highly variable and dynamic complex that is able to mediate differential cellular signals through promoter recognition and interaction with transcriptional activators.

1.2.3 Other TAF containing complexes

As the name says, TAFs were thought to be closely associated with TBP, which is thought to build the structural core of TFII-D. However a number of complexes containing TAFs but not TBP have been identified in recent years such as yeast SAGA (Grant *et al*, 1998) and its human counterpart STAGA (Martinez *et al*, 1998), TAF containing TBP free complex (TFTC) (Wieczorek *et al*, 1998) and the TFTC related complex PCAF/Gcn5 (Ogryzko *et al*, 1998). SAGA, TFTC and PCAF/Gcn5 complexes seam to be functional homologues. Unlike TFII-D, they never contain TAF1, the histone acetyltransferase subunit of TFII-D, but Gcn5/PCAF acetyltransferase subunits. A Comparison of the low resolution electron microscopy structures of TFII-D, TFTC and SAGA revealed similar features, suggesting that they share a common structural core that may be formed by homologous subunits (Brand *et al*, 1999; Wu *et al*, 2004). For both complexes, SAGA and TFTC, co-activator function has been demonstrated. Multiple subunits of TFTC interact directly with transcriptional activators (Hardy *et al*, 2002; Helmlinger *et al*, 2004; Palhan *et al*, 2005). Furthermore, both complexes are able to nucleate initiation of transcription. TFTC was demonstrated to replace TFII-D on both TATA-containing and TATA-less promoters in *in vitro* transcription assays (Wieczorek *et al*, 1998). A genome wide analysis for the promoter occupancy of TAF1, a subunit of TFIID that is not part of TFTC or STAGA, in human fibroblast cells revealed that 75% of active promoters tested are bound by TAF1 (Kim *et al*, 2005). It was implicated that this result could be due to a variant TFIID complex or a TBP free complex like TFTC or SAGA nucleating the PIC in the remaining 25% of these promoters (Muller *et al*, 2007). For SAGA, a role in regulating gene expression of stress inducible genes in yeast has been implicated (Huisinga & Pugh, 2004; Lee *et al*, 2000), however the role of TFTC in regulating gene expression *in vivo* remains to be investigated. Furthermore, the specific binding to DNA elements in the promoter region, which has been demonstrated for TFIID, and would strengthen the possible role of TFTC and

SAGA as promoter recognition complexes needs to be demonstrated.

1.2.4 TBP and TBP like factors

The TATA binding protein, which was first isolated in yeast as a factor mediating basal transcriptional activity in *in vitro* transcription assays (Cavallini *et al*, 1989) and has been described as the core component of TFIID, functioning as a scaffold TAFs can bind to to form holo-TFIID (Weinzierl *et al*, 1993). Because TBP is also required for RNA polymerase I (SL1) transcription and polymerase III (TFIIIB) transcription, where TBP associates with distinct TAFs to form SL1 and TFIIIB respectively, TBP was thought of as the 'universal' transcription factor (Davidson, 2003; Hernandez, 1993). The protein is characterised by a C-terminal core domain that forms a saddle like structure. This saddle is responsible to bind DNA via the concave underside and the interaction with other general transcription factors via the solvent exposed surface (Burley & Roeder, 1996). The C-terminal core of TBP shows a high conservation in all eukaryotes, whereas the N-terminal region shows less conservation and differs in length and sequence (Dantonel *et al*, 1999) and has been implicated in the modulation of DNA binding of TBP (Zhao & Herr, 2002) and in evading maternal immunorejection (Schmidt *et al*, 2003).

The identification of a protein in *Drosophila* that shows high similarity with TBP, TBP related factor 1 (TRF1) questioned the universality of TBP. TRF1 is a *Drosophila* specific TBP related factor that is mainly expressed in the central nervous system and in the gonads and exhibits 63% amino acid identity to TBP at its C-terminal core (Crowley *et al*, 1993). Using biochemical assays, TRF1 was confirmed to interact with TFIIA and TFIIB and direct RNA polymerase II transcription from TATA box containing promoter *in vitro* (Hansen *et al*, 1997). *in vivo* TRF1 has been implicated to bind to a TC-rich sequence and is likely interacting with neuron specific TRF1 associated factors (nTAFs). As TRF1 does not interact with TBP associated factors and is not interchangeable with TBP (Holmes & Tjian, 2000), it has been suggested that TRF1 may be a functional homolog of TBP that could have diversified to a function in tissue-specific or gene selective transcription. Interestingly, TRF1 has also been identified to be involved in RNA polymerase III transcription, as it forms a complex with BRF, which is part of the above-mentioned TFIIIB complex (Takada *et al*, 2000).

A search for other close homologues of TBP in the emerging EST (expressed sequence tags) libraries led to the identification of a third member of the TBP family and has been called TBP like protein, TLP (Ohbayashi *et al*, 1999), TBP related factor 2, TRF2 (Rabenstein *et al*, 1999) and TBP like factor, TLF (Dantonel *et al*, 1999). A sequence comparison and phylogenetic analysis of members of the TBP family in various organisms lead to the finding that TLF grouped distinct from TBP and was found in all analysed metazoans, but not in non-metazoans (Dantonel *et al*, 1999). Like TBP and TRF1, TLF has a bipartide repeat domain

that forms the saddle-like structure and has a function in DNA binding (Dantonel *et al*, 1999; Rabenstein *et al*, 1999). However a detailed comparison of the TLF core domain to the TBP core domain suggests that TLF does specifically bind DNA, but not to canonical TATA boxes although TLF shares 63% sequence homology in the core domain and 83% in the DNA binding domain were proposed (Rabenstein *et al*, 1999). This suggestion was confirmed by *in vitro* assays (Ohbayashi *et al*, 1999; Rabenstein *et al*, 1999) were no binding to the TATA box could be observed. TLF does however interact with TFIIA and TFIIB (Rabenstein *et al*, 1999; Teichmann *et al*, 1999) and is able to stimulate RNA polymerase II transcription via the assembly of a functional PIC, binding to a yet unknown DNA element (Chong *et al*, 2005; Ohbayashi *et al*, 2003). TLF has also been identified in a complex that contains components of the nucleosome remodelling factor (NURF) chromatin remodelling complex as well as the DNA replication-related element (DRE)-binding factor DREF and directs, among others, core promoter recognition of the proliferating cell nuclear antigen (PCNA) gene (Hochheimer *et al*, 2002). As TLF inhibited TBP-dependent basal transcription in an *in vitro* reconstituted system (Teichmann *et al*, 1999), TLF can also act as a repressor, likely due to competition for limiting amounts of TFIIA and TFIIB (Chong *et al*, 2005; Teichmann *et al*, 1999). The inactivation of TLF by RNAi in *C. elegans* (Dantonel *et al*, 2000; Kaltenbach *et al*, 2000) *Xenopus* laevis (Veenstra *et al*, 2000) and *Drosophila* (Kopytova *et al*, 2006) and microinjection of RNA encoding a dominant negative form of TLF in zebrafish (Muller *et al*, 2001) indicate that TLF is essential for embryonic development and contributes to transcription *in vivo*. In contrast to the embryonic lethality in these metazoan species, TLF-null mice are viable but show defects in spermiogenesis, suggesting that TLF may have species-specific functions (Martianov *et al*, 2001; Zhang *et al*, 2001).

More recently a fourth member of the TBP family was identified in a homology search on the initial draft of the human genome sharing 95% identity in the C-terminal core domain to TBP and has, in contrast to TLF, a N-terminal domain similar to TBP (Persengiev *et al*, 2003). The protein, named TRF3 or TBP2, because of its high similarity to TBP, is vertebrate specific and able to nucleate the PIC by its interaction with TFIIA and TFIIB, stimulating transcription initiation *in vitro*. Furthermore TBP2 was observed to bind to the canonical TATA-box (Bartfai *et al*, 2004; Jallow *et al*, 2004). The protein is most abundant in the oocyte, but low levels of TBP2 can be detected throughout early development of *Xenopus*. In mouse, TBP2 is highly expressed in growing oocytes during folliculogenesis and declines upon ovulation. Expression of TBP however is only detectable until folliculogenesis but not in later stages of oocyte development, suggesting different roles for the two proteins in establishing specialised programs of gene expression in this process (Gazdag *et al*, 2007). Recently TBP2 was shown to replace the canonical holo-TFIID complex in a novel TBP2/TAF3 complex in differentiating myoblasts to myotubes (Deato & Tjian, 2007). It was suggested that cell

types, like myoblasts, that are actively dividing contain an intact and presumably active canonical TFIID complex, whereas in terminally differentiated cell types such as myotubes or myofibers TFIID cornerstone subunits of TFIID like TBP and TAF4 are reduced by proteolytic degradation and replaced by alternative pre-initiation complexes like TBP2TAF3 (Deato & Tjian, 2007). Studies in zebrafish suggest that TBP2 is specifically required for the specification of haematopoietic lineage, by activating transcription of mespa, a transcription factor required for mesoderm specification (Hart *et al*, 2007).

In line with the above mentioned results, approaches to elucidate targets of TBP paralogs and their role during early vertebrate development in small scale using zebrafish and *Xenopus* indeed revealed that not all genes depend on TBP-function (Muller *et al*, 2001; Veenstra *et al*, 2000). In a recent approach the analysis of genes targeted by members of the TBP protein family using RNA antisense knockdown in combination with large-scale gene expression profiling in *Xenopus* concludes that TBP plays a limited and generic role in transcription, which is linked to widespread expression across developmental stages and tissues. TLF was indicated to be linked to the expression of preferentially embryonic transcripts and genes of the citric acid cycle and requirement for TBP2 function is linked to developmental processes with a specific enrichment for genes involved in dorso-ventral patterning (Jacobi *et al*, 2007).

1.3 The RNA polymerase II core promoter

1.3.1 Defining the core promoter

In its most common form, the RNA polymerase II core promoter is defined as the minimal stretch of DNA sequence that is sufficient to direct accurate initiation of transcription by the RNA polymerase II machinery and contains one single well-defined transcriptional startsite (TSS) (Butler & Kadonaga, 2002; Smale & Kadonaga, 2003). However, the characterisation of eukaryotic promoters in large scale and the identification of less abundant motives has been hampered by the lack of information about the 5' boundary of transcripts, as the majority of the so far identified core promoter elements show a strong bias in their position relative to the TSS. The availability of genome sequences of many model organisms and recent advances in large scale sequence analysis, methods like CAGE (cap analysis of gene expression), SAGE (serial analysis of gene expression) and PET (paired end tags) made it possible to identify the 5' boundaries of transcripts on a genome wide scale. Using Cage, a genome wide analysis of the mouse and human promoterome led to the identification of several distinct promoter classes in respect of their differential TSS usage. The clustering and mapping of CAGE tags to the genome identified promoters that have one dominant peak in the CAGE tag distribution mapped to the promoter region,

marking the TSS. These single peak promoters are more like associated with TATA boxes and tissue specific expression. A second class is characterised by a broad distribution of the CAGE tags over the promoter region, termed broad peak promoters and has been implicated to be strongly associated with CpG islands and ubiquitous expression and is less likely to contain a TATA box (Carninci *et al*, 2006).

1.3.2 A multitude of DNA-elements characterise the Pol II core promoter

In *Drosophila*, the RNA polymerase II core promoter is mainly implicated to contain either a TATA-box (TATA-containing promoter) or an initiator with or without a downstream promoter element (DPE) (TATA-less promoter). In vertebrates, the core promoter structure is much more diverse (Smale, 2001). Recent advances in large scale sequence analysis, both *in vivo* and *in silico* made it possible to identify a multitude of core promoter elements that, in conjunction with the diversity of promoter recognition complexes, adds another level to the complexity of gene regulation in eukaryotic organisms. To date 8 core promoter elements have been identified which have a characteristic DNA sequence and are required for the assembly and orientation of the pre-initiation complex in a subset of genes.

The TATA-box, the first eukaryotic core promoter element identified (Breathnach & Chambon, 1981; Goldberg, 1979), is typically located about 28-30bp upstream of the transcriptional startsite (Carninci *et al*, 2006). The consensus sequence has been described as TATA[AT]A[AT][AG] (Basehoar *et al*, 2004; Bucher, 1990), is recognised by TBP or the TBP subunit present in the TFIID complex, and is highly conserved from yeast to man. Because of the above mentioned limitations in the past, the use of bioinformatics tools to study the abundance of the TATA box in the core promoter region led to diverse results ranging from 51% in the eukaryotic promoter database (EPD) (Davuluri *et al*, 2000) to 2.6% estimated by a clustering analysis of 13010 putative promoter regions in the human genome (FitzGerald *et al*, 2004). Identification of functional TATA boxes *in vivo*, by assaying the sensitivity of gene expression to mutations along TBP's DNA binding surface in yeast suggested for about 20% of all yeast genes to contain a functional TATA box (Basehoar *et al*, 2004). Concluding from the above mentioned genome wide study together with a functional analysis of 400 promoters originating from the encyclopaedia of DNA elements project (ENCODE) (Cooper *et al*, 2006), and similar studies in *Drosophila (Ohler, 2006)* and *Arabidopsis (Molina & Grotewold, 2005)*, less than 10% of all genes in the human genome contain the TATA box motive, making it clear that TATA driven PIC nucleation is the exception rather than the rule (Kim *et al*, 2005; Sandelin *et al*, 2007).

A second well-characterised core promoter motive is the initiator, which encompasses the transcriptional startsite and was identified in a number of eukaryotes (Butler & Kadonaga, 2002; Smale & Baltimore, 1989). The consensus for the initiator was characterised

as [CT][CT]AN[AT][CT][CT] (Bucher, 1990; Corden *et al*, 1980) for human promoters and TCA[GT]T[CT] for *Drosophila* (Hultmark *et al*, 1986). The initiator is capable of directing accurate transcription initiation alone or in conjunction with the TATA box (Weis & Reinberg, 1997). The TAF1 and TAF2 components of TFIID have been implicated in initiator recognition, as binding of a dimeric complex of these two proteins in random DNA-binding site selection assays has been observed (Chalkley & Verrijzer, 1999). Although a number of initiator dependent PIC formation has been described (Smale *et al*, 1998; Weis & Reinberg, 1997), even in the absence of both TATA box and initiator a precise startpoint of transcription is not random (Sandelin *et al*, 2007). The analysis of CAGE tag data showed that highly used TSSs tend to use CG, TG and CA dinucleotides as TSS, whereas rarely used TSS tend to contain the dinucleotide GG, indicating that the initiator is not an absolute determinant of transcription initiation (Carninci *et al*, 2006). Furthermore, the initiator could not be identified in an analysis of Octamers clustering at the TSS, which might be explained by the degeneracy of the motive (FitzGerald *et al*, 2004).

A third core promoter element is the downstream promoter element (DPE) and was identified as a binding site for TFIID in assays using purified *Drosophila* TFIID (Burke & Kadonaga, 1996). The consensus sequence for the DPE is estimated as [AG]G[AT][CT][GAC] and has been indicated to be only functional in conjunction with an initiator separated by exactly 28 nucleotides in respect of the transcriptional startsite located in the initiator (usually the first adenine). This invariance of the spacing is explained by the cooperative binding of TFIID to the two motives. Photo-cross-linking analysis of purified TFIID with a TATA-less DPE-containing promoter revealed specific cross-linking of TAF6 and TAF9 to the DPE and it was suggested that TAF6 and TAF9 bind as a heterotetramer to the motive (Burke & Kadonaga, 1997). The analysis of a set of 205 *Drosophila* promoter revealed that 28% harbour the motive, typically in TATA-less promoters, and it was suggested that the DPE is as common as the TATA box (Kutach & Kadonaga, 2000). In human the analysis of active promoters in the human genome estimated the prevalence of the DPE to be around 48% (Kim *et al*, 2005). Interestingly, a repressor of TATA-dependent transcription, NC2, has been found to stimulate transcription from DPE-dependent, but TATA-less core promoters (Willy *et al*, 2000), which lead to the conclusion that there are fundamental differences in the mechanisms of transcription from DPE- versus TATA-dependent promoters (Butler & Kadonaga, 2002; Smale & Kadonaga, 2003).

In addition to the DPE, two other core promoter elements have been identified downstream of the transcription startsite. The MTE (motive ten element) has the consensus C[GC]A[AG]C[GC][GC]AACG[GC] and is typically located at position +18 to +28 relative to the transcriptional startsite. It was initially identified as an overrepresented sequence motive in a computational analysis of nearly 2000 *Drosophila* core promoters (Ohler *et al*, 2002)

and has been shown to enhance RNA polymerase II-mediated transcription in TATA-less and/or DPE-less promoters as well as synergistically with both motives in an initiator-dependent manner (Lim *et al*, 2004). However, the protein factors that act through the MTE have not been identified yet (Thomas & Chiang, 2006). The DCE (downstream core element) was identified through mutations in the human beta-globin promoter in beta-thalassemia patients (Lewis *et al*, 2000). The DCE consists of three subelements: SI is CTTC, SII is CTGT, and SIII is AGC. SI resides approximately from +6 to +11, SII from +16 to +21, and SIII from +30 to +34. The presence of SI and SIII does not correlate with the presence of the DPE, suggesting that the functions of these two subelements and the DPE are mutually exclusive. Crosslinking studies demonstrated that the DCE is contacted by TFIID through its subunit TAF1 (Lee *et al*, 2005).

Two more core promoter elements are known to date. The TFIIB–recognition element BRE is flanking the TATA box in both directions. The upstream part is called BRE^u and has the consensus [GC][GC][GA]CGCC. The downstream part of the motive is called BRE^d with the consensus [GA]T[GA][TG][GT] (Lagrange *et al*, 1998). Both motives are interacting with TFIIB. The BRE^u interaction can occur TBP-independent via the helix-turn-helix DNA binding motive of TFIIB (Lagrange *et al*, 1998), whereas the interaction with BRE^d appears to be TBP-dependent (Deng & Roberts, 2005). Initially both motives, BRE^d and BRE^u were identified in TATA-containing promoters. However a bioinformatics analysis of the eukaryotic promoter database EPD showed that the motives also occur in TATA-less promoters (Gershenzon & Ioshikhes, 2005; Lagrange *et al*, 1998).

Recently a new core promoter element, XCPE1 (X core promoter element 1) has been described, initially identified in the hepatitis B virus X gene promoter. XCPE1 has the consensus sequence [GAT][GC]G[TC]GG[GA]A[GC][AC] and is located -8 to +2 relative to the transcriptional startsite, encompassing the startsite of transcription. *In vitro* transcription assays suggest a TFIID independent transcriptional initiation mechanism, as the XCPE1 driven transcription can either utilise TFIID or free TBP. A bioinformatics search of the RefFull human promoter database identified 0.7% of the genes present in the database to contain the motive, preferentially in the category of TATA-less promoters (Tokusumi *et al*, 2007).

Another attribute of the promoter region is the higher number of CpG dinucleotides in the promoter region. These dinucleotides are underrepresented in vertebrate genomes due to a methylation of cytosine that, by deamination forms TpG, which is not repaired by the DNA repair machinery. However in the promoter region, these CpG dinucleotides are mostly unmethylated and form regions on the DNA that contain more than 50% CpG content over a stretch of 200bp, termed CpG islands (Bird, 1986). Promoters with CpG islands typically lack TATA boxes and are strongly associated with broad peak startsite selection and ubiquitously expressed genes (Carninci *et al*, 2006; Schug *et al*, 2005; Suzuki *et al*, 2001).

Taken together, the above mentioned results indicate that, although genome wide studies provided accurate information about the 5' boundary of transcripts and make it possible to provide an overview of the relative pervasiveness of core promoter elements in a given species, the biological function of these elements like the influence on transcriptional activity, PIC assembly and promoter/enhancer specificity, particularly with regards to the composition of core promoter elements, still remain elusive.

1.3.3 Bioinformatic approaches in computational promoter prediction

With the emerging sequence data from the human genome project and other large scale sequencing approaches the question arose how to identify open reading frames, intron/exon boundaries as well as 3'- and 5' ends. The latter, marking the TSS, was of special interest as with the prediction of the startsite of a gene, the number and location of genes on a genome wide scale could be predicted. Because transcription initiation seems to be brought about by the cooperative binding of a number of proteins to the DNA, the primary computational approach to promoter recognition has been to combine modules recognizing individual binding sites like the TATA box, initiator and nucleotide frequencies (Fickett & Hatzigeorgiou, 1997).

The underlying principle of current algorithms for promoter recognition is that the properties of the promoter regions are different from the properties of other functional regions and can be subdivided into three main categories: search by signal, search by content, and search by CpG island (Wu *et al*, 2007). Search by signal techniques use the above-mentioned method of recognizing individual binding sites like the TATA box or CAAT box and leads to a high ratio of false positives. Search by content algorithms are based on the differences in base- and word composition (often the frequency of 3mers) in the promoter region compared to non-regulatory regions. The algorithms are trained using two trainings sets: known promoter and non-regulatory regions to learn to discriminate promoter regions from other regions. The third technique makes use of the finding that many promoter regions contain CpG islands (Antequera & Bird, 1993; Saxonov *et al*, 2006). However it has also been reported that CpG islands are more frequently found in promoters of ubiquitously expressed genes (Schug *et al*, 2005), which introduces a bias in the search for promoter regions.

The most successful algorithms like PromoterInspector (Scherf *et al*, 2000) and Dragon Promoter Finder (Bajic *et al*, 2002) use the search by content method and reach a positive to false positive ratio of 2.3 (Scherf *et al*, 2000). However, these results have been deduced from a limited number of chromosomes or smaller data sets. When evaluated at the level of the whole genome, serious inaccuracy of predictions for non-CpG-island-related promoters occurs. Some promoter prediction programs even perform worse than, or close

to, pure random guessing (Bajic *et al*, 2004).

Promoter predicting algorithms can be a valuable tool to predict specific classes of promoters, but the heterogeneity of promoters makes it difficult to design algorithms and training sets to detect promoter regions on a global scale. As Promoters are key players in regulating gene expression by controlling the level of transcription initiation, these findings underline the need for large scale TSS determination by recently developed techniques like CAGE.

1.4 The Gene *notail*

In 1927 Nadine Dobrovolskaïa-Zavadskaïa described a mutation in mice that affected the length of the tail and the posterior vertebrae in heterozygous animals. Homozygous mutants are not viable and show defects in mesoderm formation and notochord differentiation and die around day ten in development (Dobrovolskaia-Zavadskaia, 1927). The mutation was called Brachyury because of the shortened tail and is derived from the Greek brakhus meaning short and oura meaning tail. According to human and mouse genome nomenclature, Brachyury now has the symbol and gene name T (for Tail) although Brachyury is maintained as the gene description. The cloning of the mouse T gene in 1990 (Herrmann *et al*, 1990) made it possible to describe its expression pattern. Initially mouse T is expressed throughout the primitive streak and is then maintained in those structures affected in the mutant embryo, notochord and tailbud. Shortly after this discovery, homologues of the T gene have been identified in Xenopus (xbra) (Smith *et al*, 1991), zebrafish (notail) (Schulte-Merker *et al*, 1992) and chick (Kispert *et al*, 1995b). Studies on Brachyury revealed it to be a transcriptional activator binding to the sequence T[GC]ACACCTAGGTGTGAAATT through a DNA binding protein domain, the so called T-box (Kispert & Hermann, 1993; Kispert *et al*, 1995a). A homology search for the T-box in other organisms revealed it to be the characteristic feature of a whole family of transcription factors, the so-called T-box genes. Many of them are involved in developmental processes in various stages and tissues like nervous system, lungs, kidney and muscle (Smith, 1999).
Mutations in the zebrafish homologue of mouse *T*, *notail*, leads, like in the mouse, to undifferentiated notochord and the most posterior 11-13 of their normal 30 somites are missing (Halpern *et al*, 1993).

1.4.1 Expressionpattern of *notail*

In all vertebrates, *brachyury* is expressed in presumptive mesoderm in early gastrulation embryos and later the expression is confined to the notochord and tailbud. *notail* mRNA can first be detected at dome stage in a point shaped area at the border of the blastodisc.

In mid blastula stages, when epiboly starts and the epiblasts start moving towards the vegetal pole, *notail* is expressed in a ring like area at the margin where the blastodisc meets the yolk (Fig. 14A). As epiboly commences, this expression in stable, cells at the margin express *notail*. When gastrulation starts and cells at the margin start to involute, building the epiblast (cells that have not involuted) and the hypoblast (cells that have involuted), forming the germring and the shield (Kimmel, 1989), *notail* is expressed in cells of both presumptive mesoderm and endoderm. Sections showed that the gene is active in the epiblast as well as the hypoblast. After the cells involuted, expression levels of *notail* decrease in the majority of hypoblast cells, but stay high in notochord precursors in the shield (Fig. 14B). This ring like expression of *notail* in cells just in the process of involution is maintained till the blastomeres have covered the yolk and somitogenesis is just about to start and the tailbud is forming. In tailbud stage *notail* mRNA is located in the tailbud, where the majority of posterior mesoderm is contained, and in the presumptive notochord (Fig. 14C). This pattern of expression is retained throughout somitogenesis (Fig. 14D). In later stages *notail* is only detectable in the notochord. In embryos older than 36 hours post fertilisation *notail* no longer can be detected (Schulte-Merker *et al*, 1992). It was suggested that the downregulation of zebrafish *notail* in late segmentation stages is correlated with the observed *de novo* methylation occurring at the CpG island of *notail* in the promoter region and the first exon (Yamakoshi & Shimoda, 2003).

1.4.2 Regulation of *notail* expression

A first step to gain insight into the regulation of *brachyury* comes from *xbra* the *Xenopus* homologue of *brachyury*. Cloning of the promoter into a reporter construct as well as deletion analysis showed that a 381 bp long promoter fragment of *xbra* is sufficient to drive mesoderm specific expression (Latinkic *et al*, 1997). It was suggested that Activin, a member of the TGF-beta super family, plays a major role in the regulation of *xbra* expression (Green *et al*, 1994; Smith *et al*, 1990). However, *in vivo* studies in zebrafish implicated that Activin acts only locally (Rodaway *et al*, 1999). These results suggest for other factors playing a major role in mesoderm induction and *brachyury* activation. One of the candidates could be the Nodal related morphogen Squint. It was shown that high protein levels of Squint induce *goosecoid* whereas low levels of Squint activate *notail*, *floating head* (*flt*) and *bhikhari* (*bik*) (Chen & Schier, 2001). However, in maternal zygotic One eye pinhead mutants (MZoep) a co-receptor of the Nodal signalling pathway, *notail* is still expressed in ventral and lateral margin and in the tailbud, but not in the dorsal margin and the notochord (Chen & Schier, 2001). This suggests for a Nodal signalling independent expression domain of *notail*.

As mentioned above, *brachyury* is also activated by FGF/MAPK signalling. Inhibition of FGF signalling leads to a loss of *notail* expression in the margin of early gastrulation

stages as well as a complete absence of trunk and tail in later stages (Griffin *et al*, 1995). In contrast to this, overexpression of FGFR leads to a loss of the spatial regulation of *notail* and expression is induced in the whole epiblast (Griffin *et al*, 1998). Further evidence for the link between FGF/MAPK pathway and *brachyury* was defined by the finding that Ets-2, a member of the ETS family of transcription factors is activated by MAPK dependent phosphorylation (Wasylyk *et al*, 1998) and can rescue the phenotype of dominant negative FGFR (Kawachi *et al*, 2003). Interestingly, Brachyury is able to activate FGF signalling (Casey *et al*, 1998) and it was suggested that expression of *brachyury* is regulated by an indirect autoregulatory loop involving FGF signalling (Isaacs *et al*, 1994; Schulte-Merker & Smith, 1995).

Canonical, beta-catenin dependent Wnt signalling has also been indicated in the regulation of *brachyury* expression in mouse (Arnold *et al*, 2000) and *Xenopus* (Vonica & Gumbiner, 2002). Blocking Wnt signalling through the use of C-catherin suppresses *brachyury* expression and it was suggested that this effect is cell autonomous and direct as putative TCF binding sites in the *brachyury* promoter respond to Wnt activation (Vonica & Gumbiner, 2002).

Extensive work on the *xbra* promoter showed that the correct spatial expression confined in the margin of early gastrulation stage in *Xenopus* embryos is mainly established by repressive signals rather than activation (Latinkic *et al*, 1997; Lerchner *et al*, 2000). A search for putative transcription factor binding sites in the proximal *xbra* promoter identified a deltaEF1 binding site that, in conjunction with an E2-box restricts expression of *xbra* to the marginal zone in early gastrulation stages. Point mutations in these binding sites of stable transgenic lines leads to an ectopic activation of the reporter in ectoderm and, to some extend, in the endoderm (Lerchner *et al*, 2000). As mentioned above, high levels of Activin and Squint are able to repress *xbra* whereas dorsal mesoderm markers like *goosecoid otx2* and *mix.1* expression increases dramatically, suggesting that these genes may regulate *xbra* expression (Green *et al*, 1994). Indeed mRNA injection of any of these genes represses *xbra* expression and where found to bind to bicoid- (Goosecoid and Otx2) and antennapedia (Mix.1) binding sites in the proximal promoter of *xbra* (Latinkic *et al*, 1997). Point mutation in this sites in stable transgenic lines leads to ectopic activation of the reporter in dorsal mesoderm (Lerchner *et al*, 2000).

In a morpholino knockdown approach, *notail* has been shown to be independent of TBP-function (Muller *et al*, 2001). Most likely, this independence of TBP-function is caused by the utilisation of an alternative pre-initiation complex that does not contain TBP. The gene promoter of the transcription factor *notail* is therefore an ideal model to study the mechanisms of alternative transcription initiation in the early development of a vertebrate organism.

1.4.3 Target genes of Brachyury

Brachyury plays an important role in mesoderm formation. Finding target genes of Brachyury would therefore be an important step towards understanding the mechanisms of early gastrulation processes. However, only view bona fide targets are known and described to date. The majority of these are the result of a substraction hybridization approach using hormone inducible *xbra* (Casey *et al*, 1999; Saka *et al*, 2000; Tada *et al*, 1998). The authors of this screen identified a group of genes that are direct targets of Brachyury, the so-called Bix genes (Brachyury inducible genes). They belong to the family of paired type homeobox transcription factors and are expressed in the marginal zone and the vegetal hemisphere of *Xenopus* embryos. Overexpression experiments suggest that Bix genes play a role in the formation of ventral mesoderm and endoderm. They further identified Wnt 11 to be a target of Brachyury. Wnt11 regulates convergent extension movements during gastrulation via non-canonical Wnt signalling. Injection of a dominant negative form of Wnt 11 leads to an impairment of gastrulation movements, although mesoderm and notochord are differentiated (Tada & Smith, 2000). Another target of Brachyury, Btg1, also plays a role in gastrulation movements and may act co-operatively with Wnt11 (Saka *et al*, 2000). Other genes identified in this screen, like egr-1, have been proposed to play a role in cell growth, division and differentiation. Taken together, the genes identified to be direct targets of Brachyury, have functions in mesoderm specification, control of gastrulation and cell cycle regulation (Saka *et al*, 2000).

2.0 Material

All materials used in this study were of *pro analysi* quality. Aqueous solutions were prepared using deionised water and sterile vessels and, if necessary, were autoclaved or sterile filtered before use.

2.1 Chemicals, enzymes and kits

3-(N-Morpholino)-propansulfonsäure	Roth, Karlsuhe
Acetic acid	Merck, Darmstadt
Agarose	Sigma, Taufkirchen
Ammonium acetate	Merck, Darmstadt
Ampicillin	Roche, Mannheim
Bacto-Agar	Roth, Karlsruhe
Bacto-Trypton	Roth, Karlsruhe
Bacto-Yeast extract	Roth, Karlsruhe
Boric acid	Roth, Karlsruhe
BSA	Serva, Heidelberg
Calcium chloride	Merck, Darmstadt
Calf intestine alkaline phosphatase	Promega, Mannheim
Chloroform	Merck, Darmstadt
Cyscribe cDNA labelling kit	GE Healthcare, Munchen
Disodium hydrogen phosphate	Roth, Karlsruhe
DNA-Ladder (1 kb)	NEB, Frankfurt a.M
DNA-Ladder (100 bp)	NEB, Frankfurt a.M
DNA-Ladder (Mix)	Peqlab, Erlangen
dNTP	Promega, Heidelberg
EDTA	Roth, Karlsruhe
Ethanol	Roth, Karlsruhe
Ethidium bromide	Roth, Karlsruhe
Formaldehyde	Merck, Darmstadt
Gentamicin	Sigma, Taufkirchen
Isoamyl alcohol	Roth, Karlsruhe
Isopropanol	Merck, Darmstadt

Magnesium sulphate	Merck, Darmstadt
Methanol	Roth, Karlsruhe
MLV reverse transcriptase	Promega, Mannheim
Morpholinos	GeneTools, Philomath, USA
Nuclease free water	Ambion, Huntigdon, UK
NucleoSpin RNA-L kit	Macherey-Nagel,
Oligonucleotides	Metabion, Planegg
Paraformaldehyd	Merck, Darmstadt
PBS	Invitrogen, Karlsruhe
Phenol	Roth, Karlsruhe
Phenol red	Roth, Karlsruhe
Potassium acetate	Roth, Karlsruhe
Proteinase K	Sigma, Taufkirchen
PureYield Plasmid Midiprep System	Promega, Mannheim
QIAGEN Plasmid Maxi Kit	Qiagen, Hilden
QIAquick Gel Extraction Kit	Qiagen, Hilden
QIAquick PCR Purification Kit	Qiagen, Hilden
QuickLyse Miniprep Kit	Qiagen, Hilden
Restriction endonucleases	Promega, Mannheim
RLM-RACE kit	Ambion, Huntigdon, UK
SDS	Roth, Karlsruhe
Sodium acetate	Roth, Karlsruhe
Sodium chloride	Roth, Karlsruhe
Sodium dihydrogen phosphate	Roth, Karlsruhe
Sodium hydrogen carbonate	Roth, Karlsruhe
Sodium hydroxide	Sigma, Taufkirchen
T4 DNA ligase	Promega, Mannheim
T4 DNA polymerase	Promega, Mannheim
GoTaq DNA polymerase	Promega, Mannheim
TOPO TA Cloning Kit	Invitrogen, Karlsruhe
Triple Master PCR System	Eppendorf, Hamburg
Tris-Base	Roth, Karlsruhe
Tris-HCl	Roth, Karlsruhe
Trizol	Roth, Karlsruhe

2.2 Equipment and materials

Bacteria incubators	Heraeus, Hanau
Borosilicate glass capillaries	Harvard Ltd., Kent, UK
Cool centrifuge J2-HS	Beckman, Stuttgart
Digital camera DFC300 FX,	Leica, Bensheim
Electrophorese chambers	Peqlab, Erlangen
Eppendorf microcentrifuge tubes	Eppendorf, Hamburg
Falkon tubes	Greiner, Nürtingen
FemtoJet microinjector	Eppendorf, Hamburg
Flaming-Brown Needle puller	Sutter Instruments, USA
Fluorescent stereomicroscope MZ FLI- II	Leica, Bensheim
Gas microinjector	Tritech research inc., L.A., USA
Incubator for fish embryos	Heraeus, Hanau
Magnetic thermomixer	Heidolph, Rosenfeld
Microcentrifuge 5417 R and C	Eppendorf, Hamburg
Microcentrifuge Biofuge pico	Heraeus, Hanau
Microfiltration columns	Pall, Ann Arbor, USA
NanoDrop ND-1000	Peqlab, Erlangen
Omnigrid 100	Genemachines
PCR-Thermocycler,	MJ Research Biozym, Oldendorf
Petri dishes	Greiner, Nürtingen
Pipette tips	Corning, Corning
Spectrophotometer	Eppendorf, Hamburg
Spin-X-Filter	Costar, Corning, USA
Stereomicroscope SMZ645	Nikon, Düsseldorf
Sterile filters	Renner, Darmstadt
Thermomixer	Eppendorf, Hamburg
UV Transilluminator	Saur, Reutlingen
Vac-Man Vacuum manifold	Promega, Mannheim
Vortex	Bender & Hohbein, Karlsruhe
Water bath	Kötterman, Uetze-Hänigsen

2.3 Oligonucleotides and morpholinos

All oligonucleotides have been designed using the software "primer3" version 0.40 and purchased from Metabion, Planegg. A full list of all PCR-primer used in this study can be found on the enclosed CD in supplementary table II.

Morpholino antisense oligonucleotides have been purchased from Gene Tools, Philomath, USA. The sequences are as follows:

TBP MO: 5'-GAGGTAGGCTGTTGTTATGTTCCAT-3'
TBP control MO: 5' -GACGTACGCTGTTCTTCTCCTCGAT
TBP MO2: 5'-CAAAAGACGTAAACGATAATTCGCA-3'

2.4 Bacterial strains

All clonings have been done using the *E. coli* strain "Top 10", purchased from Invitrogen.

2.5 Zebrafish lines

All zebrafish embryos used in this study have been of the wildtype Tubingen AB line.

2.6 Solutions and buffers

If not specified differently, all solutions were prepared using deionised water.

20x SSC-Buffer
3M NaCl, 0.3M sodiumcitrate

Blocking buffer
0.2% BSA, 1% DMSO, 0.1% Tween 20 in PBS

BT-Fix
4% paraformaldehyde, 4% sucrose, 0.12 mM KCl, 0.1 M phosphate, pH 7.3

Hank's solution
0.14 M NaCl, 5.4 mM KCl, 0.25 mM Na_2HPO_4, 0.44 mM KH_2PO_4, 1.3 mM $CaCl_2$, 1 mM $MgSO_4$, 4.2 mM $NaHCO_3$.

Hybridisation buffer
50% formamide, 10 mM EDTA, 0.1% bromophenolblue, 0.1% xylencyanol

LB-Agar
1.5% Bacto-Agar in LB-media

LB-Media
1% Bactotrypton, 0.5% Yeast extract, 1% NaCl; pH 7.0

Lysis buffer for genomic DNA extraction
10 mM Tris-HCl, 0.1 M EDTA, 0.5% SDS; pH 8.0

Methylene blue solution (2000x)
0.1% methylene blue in distilled water

Phenol red solution (10x)
10% phenolred, 0.2 M KCl; pH 7.5

Pronase-Solution
1% pronase, 10% Hank's-Solution, incubate 30 minutes at 37°

Proteinase K stock solution
10 mg/ml proteinase K in PBS

PTW
0.1% Tween 20 in PBS

SOC-Media
2% Bactotrypton, 0.5% Yeast extract, 10 mM NaCl, 25 mM KCl

Staining buffer
0.2% BSA, 1% DMSO, 0.1% Tween 20 in PBS

System water in the fish facility
120 mg/l „Ocean Sea Salt“, 45 mg/l $NaHCO_3$ in desalted water

TAE Buffer
40 mM Tris-Base, 1 mM EDTA, 5 mM Acetic acid; pH 7.8

TBE-Buffer
90 mM Tris-Base, 1 mM EDTA, 44 mM Boric acid; pH 8.0

TE-Buffer
10 mM Tris-HCl (pH = 7.4), 1mM EDTA, pH 8.0

3.0 Methods

3.1 Biomolecular Methods

3.1.1 Phenol chloroform extraction of nucleic acids

To remove proteins from nucleic acid solution, an equal volume of buffer-saturated phenol:chloroform (1:1) was added to the nucleic acid solution and vortexed for 20 seconds followed by a centrifugation for 5 minutes at 13.000 rpm to separate phases. The aqueous layer was removed to a new tube and an equal volume of chloroform was added, mixed briefly and centrifuged for 5 minutes at 13.000 rpm. The aqueous top layer again was removed to a new tube and ethanol precipitated.

3.1.2 Precipitation of nucleic acids

To a nucleic acid containing solution 1/10 volume of sodium acetate, pH 5.2, (final concentration of 0.3 M) in case of DNA or an equal volume of 5 M ammonium acetate (final concentration of 2.0-2.5 M) in case of RNA was added to adjust the salt concentration. After mixing 2 to 2.5 volumes of cold 100% ethanol were added and placed at -20° C for >1 hour. The solution was then centrifuged for 30 minutes at 13.000 rpm. The supernatant was discarded and the precipitated nucleic acid washed with 70% ethanol, briefly centrifuged and after discarding the supernatant, the pellet was air dried and resuspended in the appropri-

ate volume of TE or water.

3.1.3 Isolation of plasmid DNA

Plasmid DNA was isolated using Quiagen or Promega kits for Mini, Midi and Maxi Plasmid preparations from over night bacterial cultures in LB- medium following the manufacturer's instructions. Both of the above mentioned kits are based on the alkali-lysis method by which the plasmid DNA is separated from the genomic DNA and most of the proteins (they form white precipitate, which is removed by centrifugation or filtration). The remaining solution, containing the plasmid DNA, is subsequently purified on anion exchange or silica membrane columns to ensure complete removal of remaining proteins, RNA and bacterial endotoxins. The column-bound DNA is then eluted in an appropriate volume of nuclease free water. The concentration was determined using a NanoDrop spectro-photometer.

3.1.4 Isolation of genomic DNA

Genomic DNA was isolated from cerebral tissue of adult zebrafish. The tissue samples (~500 mg) were homogenized on ice and immediately afterwards incubated in lysis buffer containing 10 µg/ml proteinase K overnight at 55° C. To remove peptides and remaining proteins, the cell lysate was extracted two times with an equal volume of phenol and once with phenol/chloroform. Finally, the DNA was recovered from the remaining solution by ethanol precipitation, washed with 70% ethanol and, after air-drying, dissolved in an appropriate volume of TE buffer.

3.1.5 Restriction digest of DNA

The digestion of DNA with restriction endonucleases was performed according to the instructions of the enzyme supplier. Approximately one unit of enzyme per 1 µg DNA in appropriate buffered digestion reaction was used. If not otherwise specified by the manufacturer, the reaction was incubated for 1-4 hours on 37° C, depending on the amount of DNA.

3.1.6 Agarose Gel Electrophoresis

For size and quality check, as well as the separation of DNA fragments, agarose gel electrophoresis was performed. Depending on the size of the analysed DNA fragments,

agarose gels with a concentration varying from 0.8% to 1.5% were prepared. For visualization of the DNA on a UV-transilluminator, ethidium bromide in final concentration 0.5 µg/ml was added to the agarose gel. Before loading, the samples were then supplemented with loading buffer and the electrophoresis carried out in TAE or TBE electrophoretic buffers with 3-5 V/cm intensity of the electric field. An appropriate DNA marker (DNA ladder) was loaded in parallel for determining the size and approximate quantity of the DNA samples.

3.1.7 Isolation of DNA from agarose gels

For cloning purposes or microinjections, DNA fragments resulting from a restriction digest or PCR were separated by agarose gel electrophoresis. The band containing the desired DNA fragment was cut out from the gel and the DNA was extracted using the QIAquick Gel Extraction kit according to the manufacturer's instructions.

3.1.8 Polymerase Chain reaction (PCR)

The amplification of DNA fragments from genomic DNA or plasmids was performed by PCR. Two enzyme systems were used depending on the purpose. Ordinary Taq polymerase (GoTaq, Promega) was used when proof reading activity was not needed (for example colony tests). Triple Master PCR System (Eppendorf) was utilised for the amplification of DNA fragments for cloning purposes. This is an enzyme mixture (Taq polymerase with proof-reading polymerases), which is optimised for amplification of long targets with relatively high speed and proof reading activity. The PCR was performed according to the user manuals provided with the enzymes, with adjustment of the annealing temperature and elongation time according to the used primers and the size of the amplified fragments. All PCRs were performed on a MJ Research thermocycler.

3.1.9 Blunting and ligation of DNA fragments

Before ligating DNA fragments with incompatible cohesive ends, a blunting (filling of a 5'-overhang or removing of a 3'-overhang) of the cohesive ends was performed, using T4 DNA polymerase, according to the supplier's instructions. The ligation of DNA fragments was performed with T4 DNA ligase as described in the user manual provided with the enzyme. Approximately 100 ng of vector DNA and 1-3 units of ligase were used in a 20 µl reaction. The molar ratio of free DNA ends of vector and insert was 1:3. In case of ligation of DNA fragments with cohesive ends, the reaction was incubated for 3 hours at room tem-

perature and in case of fragments with blunt ends over night at 16°C.

3.1.10 TOPO-cloning

TOPO TA cloning kit (Invitrogen) was used for fast direct cloning of PCR-ampli- fied fragments with Thymidine overhangs (fragments amplified with Taq polymerase or Taq polymerase based enzyme blends). When the PCR amplification resulted in one specific band 2-4 µl from the PCR were used directly (without any purification) for the cloning reaction; in the other cases the desired DNA fragment was purified from the PCR by agarose gel electrophoresis. Subsequently the purified fragment was adenylated (addition of an adenine on the 3' end of the DNA fragment) before using it in the TOPO-cloning reaction. This adenylation step was necessary to increase the amount of the adenylated fragments (critical for the efficiency of the cloning reaction), which significantly decreases during the purification steps. The adenylation was performed by adding PCR buffer (to 1x final concentration), 1-2 units Taq polymerase and dATP to 250 µM into the solution, containing the purified fragment. The reaction was incubated for 15 minutes at 72° C and 2-4 µl were used for TOPO-cloning reaction. After 5-10 minutes incubation at room temperature, the cloning reaction was transformed into TOP10 chemically competent cells (see below), provided with the kit.

3.1.11 Transformation of competent *E. coli*

10-50 ng plasmid DNA or 10 µl of a ligation reaction (see blunting and ligation of DNA fragments) were used to transform chemically competent *E. coli*. The cells were incubated with the DNA for 10 minutes on ice, heat shocked at 42° C for 45 seconds, placed again on ice for 2 minutes and incubated for 1 hour at 37° C in SOC- or LB-medium. Finally the transformed bacteria were plated on LB-agar plates with the respective antibiotic and incubated over a night on 37° C. The concentration of the used antibiotics was 100 µg/ml for ampicillin and 50 µg/ml for kanamycin and chloramphenicol.

3.1.12 Extraction of total RNA

Embryos were collected at the desired stage, snap frozen in liquid nitrogen and stored at -80° C to avoid degradation of the RNA. To extract total RNA for cDNA synthesis and 5' Race, embryos were homogenized in 1 ml Trizol reagent (Invitrogen) per 100 embryos using a 20-gauge needle and syringe. The homogenized samples were incubated for 5 minutes at

room temperature, 0.2 ml of chloroform per 1 ml of Trizol added and briefly vortexed.

After centrifugation at 12.000 rpm at 4° C for 15 minutes, the upper, aqueous phase, was transferred into a new tube. Precipitation of the RNA was performed by mixing with 0.5 ml of isopropyl alcohol per 1 ml Trizol used for the initial homogenization and Incubation of the samples at room temperature for 10 minutes followed by a centrifugation at 13.000 rpm at 4° C for 10 minutes. After centrifugation the supernatant was removed and the RNA pellet washed with 75% ethanol. At the end of the procedure, the RNA was briefly air-dried for 5 minutes and dissolved in RNase –free water and stored at -80°C.

3.1.13 First strand cDNA synthesis

cDNA templates for RT-PCR were synthesized by using 1 µg of a total RNA preparation and 1 µg of random hexamer primer per microgram RNA sample in a total volume of 14 µl in water. The tube was heated to 70° C for 5 minutes to melt secondary structures within the template and immediately afterwards cooled on ice for 5 minutes to prevent secondary structure from reforming. After the annealing of the random hexamer primer, 5 µl M-MLV RT 5x Reaction Buffer, dNTPs to a final concentration of 0.5 mM and 200 units M-MLV RT (H–) enzyme was added, incubated for 10 minutes at room temperature and finally the reaction mix was incubated for 50 minutes at 50° C for cDNA synthesis.

3.1.14 Site directed mutagenesis and deletion

A PCR based approach ("Higuchi Method", described in (Higuchi *et al*, 1988)) was utilised to generate mutation and deletions in the notail promoter. This method allows mutation, deletion and insertion of sequences at any position in the DNA fragment. The method is based on two PCR rounds. In the first round, two primary PCRs produce two overlapping DNA fragments, both bearing the same mutation introduced via primer mismatch in the region of overlap. In the second round, the products of the first two reactions are mixed (after gel purification) and used as a template in a sec-

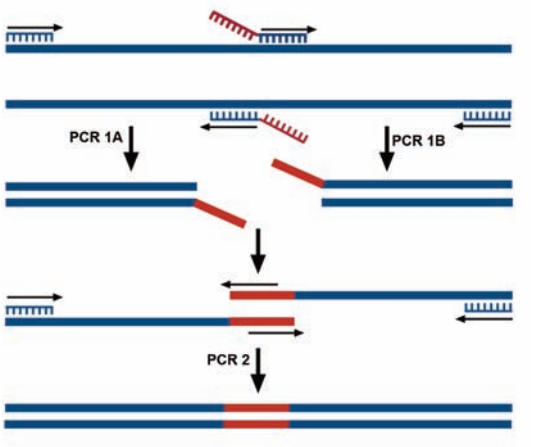

Figure 2. Schematic representation of the "Higuchi-method", used for site-directed mutagenesis by PCR.

ond PCR. The overlap in sequence allows the two fragments to anneal after their denaturation and renaturation and produce a structure with recessed 3' OH ends that can be extended by a DNA polymerase to produce a complete duplex fragment. These extended segments can then serve as a template for the secondary reamplification of the combined sequences using only the outermost two of the four primers used to produce the primary fragments.

3.1.15 Assement of transcriptional startsites

The start site of transcription (TSS) of genes was determined by 5'-RACE using the RLM-RACE kit from Ambion. The kit is designed to amplify cDNA only from full-length, capped mRNA through a series of enzymatic reactions that involves dephosphorylation of truncated transcripts lacking the cap-structure followed by a replacement of the cap by an adaptor oligonucleotide which is used as priming site in subsequent PCR steps. After ligation of the adaptor oligonucleotide, a standard PCR was carried out using a primer specific for the adaptor and a primer specific for the gene of interest. As in most of the cases, the resulting PCR product is too weak to be visualised on a agarose gel, a "nested" PCR was carried out, using the adaptor specific primer and a gene specific primer annealing 5' of the primer used in the first PCR reaction. The procedure followed the manufacturer's protocol. PCR products were then Topo cloned and sequenced.

3.2 Immunohistochemistry – *in situ* hybridisation

3.2.1 Synthesis and labeling of DIG-RNA probes

To 1 µg linerised plasmid DNA, 2 µl 10x reactionbuffer, 1 µl RNAse inhibitor, 10 µl DIG labelling mix and 2 µl enzyme mix (T7 or SP6 respectively) was added to a final volume of 20 µl in water. After incubation for two hours at 37° C, the synthesised RNA probe was ethanol precipitated and solved in hybridisation buffer.

3.2.2 Fixation of zebrafish embryos

Embryos were collected at the desired stage, fixed in BT-fix over night at 4° C and stored in Methanol at -20° C until further procedure.

3.2.3 Hybridisation of DIG labelled RNA probes

In methanol stored embryos were re-hydrated by incubation for 5 minutes in 75% methanol, 5 minutes in 50% methanol and 5 minutes in 25% methanol. Subsequently the embryos were washed 4 x 5 minutes in PTW and incubated for 4 hours in hybridisation buffer following a incubation with the DIG labelled RNA probe in a 1/400 dilution over night. After the incubation with the probe, embryos were washed 2 x 30 minutes in 50% formamide/50% (2 x SSC; 0.1% Tween 20), 1 x 15 minutes in 2 x SSC, 0.1% Tween 20 and 2 x 30 minutes in 0.2% x SSC, 0.1% Tween 20 to remove unhybridised probe. All steps have been carried out at 65°C.

3.2.4 Blocking of unspecific binding

To block unspecific binding of the anti-DIG antibody, embryos were washed 1 x 5 minutes in blocking buffer followed by incubation in blocking buffer for 4 hours at room temperature. In parallel, the anti-DIG antibody was pre absorbed in fish powder in a 1/400 dilution in blocking buffer.

3.2.5 Staining of embryos

After blocking, the anti-DIG antibody was added to the embryos in a 1/400 dilution and incubated over night at 4°C. To remove unbound antibody, the embryos were washed 6 x 20 minutes in PTW and 2 x 5 minutes in staining buffer. To start the staining reaction, the chromogenic substrates BCIP (0.5mg/ml final concentration) and NBT (0.188 mg/ml final concentration was added and stained in the dark.

3.3 cDNA Microarrays

The microarrays used in this study, were spotted on an Omnigrid 100 robot (Genemachines). The spotted 65mer oligonucleotides were purchased from Compugen and represent 10501 individual zebrafish genes.

3.3.1 mRNA preparation for microarray hybridisation

The extraction of total RNA for microarray hybridision was done using the NucleoSpin RNA-L kit (Macherey-Nagel). 500 snap frozen sphere stage embryos were thawed on ice and mechanically homogenized in 1,8 ml lysis buffer (included in the kit). Further steps

followed the manufacturer's protocol. Finally, mRNA was isolated from 300 µg total RNA per treatment group using the MicroPoly(A) Purist kit from Ambion also following the manufacturer's instruction.

3.3.2 Preparation of fluorescent-labelled cDNA

The preparation of fluorescent labelled cDNA was performed using the Cyscribe direct cDNA labelling kit, (Amersham) where the fluorescent-labelled nucleotide analogs Cy3-dUTP and Cy5-dUTP are incorporated into the synthesized cDNA. 1.5 mg mRNA was used in each subsequent fluorescence-labelling reaction. After cDNA synthesis, the mRNA was degraded by NaOH, the labelled cDNA purified by Millipore Microcon YM-30 columns (Millipore) and eluted in 16,5 µl TE buffer. The concentration and quality of the fluorescent-labelled cDNA was determined photometrically utilising the Nanodrop device.

3.3.3 Hybridisation of the Microarrays

The labelled cDNA was denatured by heating to 95° C for 2-5 minutes, cooled on ice and mixed with 15 µl 2x hybridisation buffer (DigEasy Hyb, pre warmed to 42° C). The solution was then carefully applied onto the microarray and sealed with a coverslip. The microarrays were then placed into a hybridisation chamber and incubated for 16-18 hours at 42° C in a hybridisation oven.

After the hybridisation the microarray slides were washed 5 minutes in 2 x SSC/0,1% SDS, 10 minutes in 0,1 x SSC/0,1 % SDS and 4 times 1 minute in 0.001 x SSC. After drying the hybridised microarrays are ready for scanning.

3.3.4 Scanning and statistical analysis of the microrrays

To assess the fluorescent signal of labelled cDNA hybridised to the probes on the microarray, the microaray slides were scanned on a laser scanner Axon 400 B (Axon instruments) taking care that both channels (550 nm for Cy5, 649 nm for Cy3) are adjusted to equal overall signal intensities.

To determine and calculate signal intensities derived from each spot of the microarray, the software GENEPIX was used. The emerged data file of signal intensitites was then fed to KAMAAN, an R-program developed by Dr.-Ing. Jens Jäkel (Bauer *et al*, 2006), for normalisation and statistical analysis.

3.4 Fish husbandry and injection techniques

3.4.1 Fish husbandry and care

The adult zebrafish stocks are maintained in an aquarium system built by Aquarienbau Schwarz (Göttingen). Approximately 10 pairs are kept in each tank (30 liter) under the following water conditions: conductivity 400-500 µS; hardness 5° dH; pH 7,0-7,5 temperature: 26 <=> 28° C. The ammonium, nitrate, nitrite and phosphate levels are checked once per week to ensure a good water quality The light/dark cycle in the facility is set to 14 hours light and 10 hours dark. Wild type zebrafish from the AB strain were used for the experiments. The crossing of fishes was performed in 1 litre crossing cages, filled with system water, containing one fish pair. To avoid parental cannibalism the cage contains a sieve, which separates the eggs from the parents after the laying. The laying starts the next morning with the switching on of the facility light, which is one of the main breeding stimuli for the fishes. The eggs were collected shortly after the light came up, transferred into petri-dishes and used for experiments.

3.4.2 Preparation of injection solution

For generating transient transgenic zebrafish, circular plasmid DNA in a concentration from 20 to 50 ng/µl was used to assay reporter gene activity. The injection solution was prepared by dilution of the plasmid DNA to the desired concentration in distilled water and addition of phenol red to a final concentration of 1%. The phenol red serves as colour marker, to distinguish injected- from uninjected embryos. To prevent blocking of injection needles by phenol red aggregates, the solution was filtered through a spin filter column (0,2 µm) and stored at -20° C until further usage.

3.4.3 DNA microinjections

The microinjection experiments were performed using Gas micro injectors and Nikon SMZ645 stereomicroscopes. The needles for the microinjection were prepared from borosilicate glass capillaries (0,7 mm inner and 1,0 mm outer diameter) on a Flaming-Brown needle puller. Before injection, the needles were filled with 1-3 µl injection solution (see Preparation of injection solution) using Eppendorf microloader pipette tips. The zebrafish eggs were collected shortly after fertilisation (zygote stage), dechorionated using 5 mg/ml pronase solution, transferred to a agarose coated petri-dish and injected into the cytoplasm with approximately 2-3 nl injection solution. After injection, the embryos were incubated at

28° C until they reached desired stage.

3.5 Bioinformatic methods

The bioinformatic analysis in this study is based on the use of the programming language Perl version 5.8.1-RC3 and BioPerl 1.4.0. All scripts used in this study can be found on the enclosed supplemental CD.

3.5.1 Storage of genomic sequence information

The Ensembl zebrafish core database, assembly Zv7 release 46_6 has been downloaded as MySQL text format dumps and imported into the relational database management software MySQL version 4.1.22. This database stores the unmasked genome sequence, the non-redundant set of Ensembl gene, transcript and protein models as well as external references to other databases like UniGene, RefSeq or GenBank. The standard Perl Application Programme Interfaces (APIs) provided by Ensembl were used to retrieve and process data from the database. These APIs serve as a middle-layer between the underlying database scheme and custom perl scripts that were written to specifically retrieve sequence information.

3.5.2 Isolation of putative promoter regions

In the Ensembl database, a gene is defined as a grouping of transcripts, which share overlapping exons. Transcripts also have an associated translation object, which defines the UTR and coding sequence. Furthermore, each nucleotide in the genome can be defined by its position. Having this information makes it possible to parse the Ensembl database for genes, which transcripts have a 5' UTR and extract sequence up- and downstream of the 5' end of the transcript based on its position (database version 46_7). (see perl script "promoter_extraction" in the supplement).

3.5.3 Permutation of putative promoter regions

Each sequence of the Ensembl PPR set (7809 sequences) was segmented into 10bp intervals. The segments were then permutated x100 times using the "Fisher-Yates shuffle" (Fisher & Yates, 1938) and re-assembled into a permutated control sequence showing high similarity in the nucleotide composition as the original dataset (see perl script "promoter_shuffle" in the supplement).

3.5.4 Annotation and mapping of sequence identifiers

To map GenBank identifiers to the UniGene database, UniGene build 105 was used. The database utilises sequence information submitted to GenBank to build a comprehensive and non redundant sequence archieve by clustering cognate sequences and holds detailed information about the gene and the sequence identifiers that are part of a cluster. To assign GenBank identifiers representing oligonucleotides on the Compugen microarray to a UniGene EST cluster, the flat file was parsed by using the Perl module "Bio::Cluster::UniGene" (see perl script "UniGene_parser" in the supplement).

Annotation of Affymetrix sequence identifiers to other databases was established by utilisation of the flat file "netaffix" provided by Affymetrix.

3.5.5 Basic local alignment (BLAST)

To identify sequences present in different datasets on the level of sequence similarity, BLAST (Basic local alignment search tool) version 2.2.12 (Altschul *et al*, 1990) was carried out on a "Formatdb" formatted source database. The BLAST output was parsed by utilising the Perl Module "Bio::Search::IO", which is part of the BioPerl package of modules. If not specified differently, a threshold of 95% identity determined sequences to be of identical origin (see perl script "TSS_validation").

3.5.6 Sequence motive and nucleotide frequency analysis

The putative promoter regions extracted from the Ensembl core database have been searched for published core promoter motives and CpG islands. The analysis was carried out using the published consensus sequences (see 4.2.1) in a pattern-match approach, allowing no mismatch (Table 1). In the position-restricted analysis, only hits in the defined window (Table 1) were counted and calculated as percentages over the whole set of promoters (see perl script "findpattern_restricted"). For the clustering analysis of core promoter elements, each hit was counted and plotted over the whole sequence region in a histogram utilising the statistic software "R" version 2.2.0 (see perl script "findpattern_restricted").
The search for CpG islands has been carried out by using "newcpgreport", which is part of the EMBOSS package of bioinformatics tools version 4.1.0. The definition of a CpG island was chosen as follows: 50% CG content in a 200bp window with an observed/expected ratio of >0.6 (Saxonov *et al*, 2006).

The nucleotide frequency in the core promoter region has been dtermined by sim-

ply counting the occurrences of the nucleotides A, T, C and G at a specific position in respect of the putative transcriptional startsite (see perl script "determine_nucfreq"). The resulting table was then plotted as a histogram utilising the statistic software "R" version 2.2.0.

3.5.7 Compilation of maternal mRNA degradation patterns

Maternal genes present in the TBP knockdown microarray and miR-430 targets gene sets (Giraldez *et al*, 2006) were identified by utilizing an existing dataset of transcripts accumulated in the unfertilized egg (Mathavan *et al*, 2005). The variation of steady state mRNA levels over developmental time was determined by comparing fold changes as described in (Mathavan *et al*, 2005) and was visualised by plotting their values in the unfertilised egg and at 3, 4.5 and 6h post fertilization utilising the statistic software "R" version 2.2.0. For a clustering of maternal mRNAs into fast-, medium- and slow degrading subgroups the following criteria were used: fast degrading: >200% decrease of fold change from 0 hpf to 3 hpf, medium degrading: <100% from 0 hpf to 3 hpf, >100% from 3 hpf to 4.5 hpf, slow degrading: <50% decrease of fold change till 4.5 hpf).

3.5.8 Quantification of RT-PCR signal intensities

The signal intensities of cDNA bands separated by agarose gel electrophoresis have been quantified by calculating the number of grey scale pixels of the respective band in three independent repeats per gene using the software ImageJ (version 1.3.7). Single Intensity values of each band have then been imported into Excel (version 11.3.5) to calculate the mean signal intensity for each gene and the standard deviation resulting from the independent repeats. Blotting of the mean signal intensities as a bar chart was also carried out using Excel.

4.0 Results

4.1 Large scale analysis for the requirement of TBP-function

TBP is a key factor in eukaryotic transcription and has a central role in mediating the nucleation of the pre-initiation complex in all three polymerases. The presence of TBP like factors sharing high homology with TBP and their ability to interact with members of the PIC raised the question of how general TBP-function is in transcriptional regulation. In previous studies only a small set of genes could be analysed of their requirement for TBP (Muller *et al*, 2001; Veenstra *et al*, 2000), which did not allow for the assessment of the generality and specificity of TBP.

What is the proportion and the nature of genes requiring TBP for their activity at a stage in development when the zygotic genome becomes active? A genome level analysis for the requirement of TBP-function of a vertebrate embryo using morpholino antisense oligonucleotides would provide the opportunity to elucidate which genes require TBP-function in large scale and was addressed by a microarray experiment carried out by Dr. Monica Szlabo, Dr. Janosh Kiss and Dr. Lixin Yang in the laboratory of Prof. Dr. Uwe Strähle and Dr. Ferenc Müller. The microarray analysis was carried out at dome/sphere stage in embryos in which TBP was blocked by injecting morpholino antisense oligonucleotides (MO) into zebrafish embryos (published in Ferg *et al*, 2007). The basis of the work presented here, is the unprocessed data file originating from this TBP knockdown microarray experiment.

4.1.1 Bioinformatic microarray data processing suggests that only a fraction of genes expressed in late blastula stages require TBP-function

Transcriptional activity of developmentally regulated genes under TBP knockdown conditions suggests that not all genes require TBP-function (Muller *et al*, 2001; Veenstra *et al*, 2000). To elucidate the proportion of genes that require TBP for their activity in dome stage, shortly after the activation of the zygotic genome, the data file of the above mentioned microarray experiment has been processed using “Excel”. Genes represented by less than 10 data points were removed to strengthen the statistical argument of the data. False positives and false negatives were removed by applying a false discovery rate (Benjamini & Hochberg, 1995)(FDR) cut-off of <= 0.05. This filtering selected 1927 genes that show a differential response to TBP knockdown compared to the c MO injected control (supplementary table I). The 1927 genes showing differential response to TBP depletion were then

grouped according to the change of expression levels compared to the control by establishing three different response classes. Genes that show a fold change of <= -2 were grouped into the class of downregulated genes. The class of upregulated genes was defined as showing a fold change of >= 2. The third class was established by grouping genes that show low variable, insignificant changes in RNA levels (fold change <2 and >-2).

The largest group of the 1927 genes selected for the analysis falls into the class of low variable genes, as they show no significant change in RNA levels (n= 1259, 65.3%) comparing TBP MO injected embryos to the c MO injected control. This finding suggests that TBP is not required for the steady state mRNA levels of these genes at the analysed stage in development. A smaller proportion of genes show elevated RNA levels in TBP depleted embryos in comparison to the control (n= 330, 17.1%), indicating that TBP-function is directly or indirectly required to alleviate the steady state mRNA level of these transcripts. About the same proportion of genes displays a decrease of RNA levels in TBP morphants compared to the c MO injected control (n = 338, 17.5%). These genes likely require TBP for their activity. Although the selected stage in development, shortly after MBT, minimises the potential of secondary effects, the elevated mRNA levels observed on the microarray, could be of direct or indirect origin.

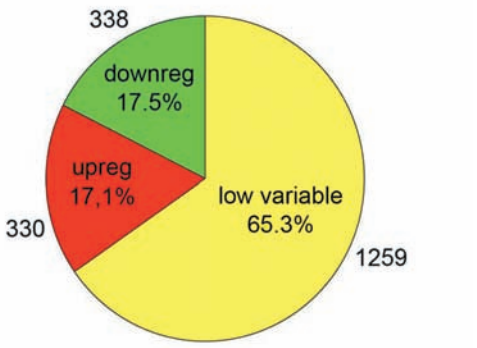

Figure 3. Summary of expression profiling data from the TBP knockdown microarray experiment represented as a pie chart diagram. Abbreviations, downreg, downregulated on the TBP knockdown microarray; upreg, upregulated on the TBP knockdown microarray; low variable, low variable changes in gene expression on the TBP knockdown microarray.

Taken together these results suggest that only a proportion of genes expressed at in the dome stage embryo require TBP for their activity.

4.1.2 Validation and specificity control of TBP antisense morpholino oligonucleotides by independent techniques

cDNA microarray experiments are a qualitative approach to study gene expression profiles on a global scale and are associated with many sources of errors associated with the microarray technology itself, like image collection, validation, normalization and analysis of the data (Wang *et al*, 2006). To validate the microarray result by a more quantitative method, semi quantitative RT-PCR was carried out in a subset of randomly chosen genes. To represent

the three described response classes, 10 upregulated-, 10 downregulated- and 8 genes showing low variable changes in RNA levels on the TBP knockdown microarray where analysed to determine if comparable changes in mRNA levels can be observed by RT-PCR. Total RNA of TBP MO or c MO injected embryos was extracted from dome-stage embryos in three parallel samples and reverse transcribed for cDNA synthesis. Semi-quantitative RT-PCR was carried out using gene specific PCR parameter and primer pairs specified in supplementary table II. Using *beta-actin* as internal control, a gene whose steady state mRNA level does not exhibit a change in TBP morphants, the signal intensities of a total of 28 genes comparing TBP MO injected embryos to c MO injected embryos in three independent repeats where analysed and quantified using the gel analysis tool of the imageJ software (see 4.5.8 for details). A comparison of signal intensities indicates that the entirety of the 28 analysed genes show a comparable response to TBP knockdown as on the microarray (Fig. 4), suggesting for a validation of the microarray results.

Morpholino antisense oligonucleotides are widely used for targeted knockdown of

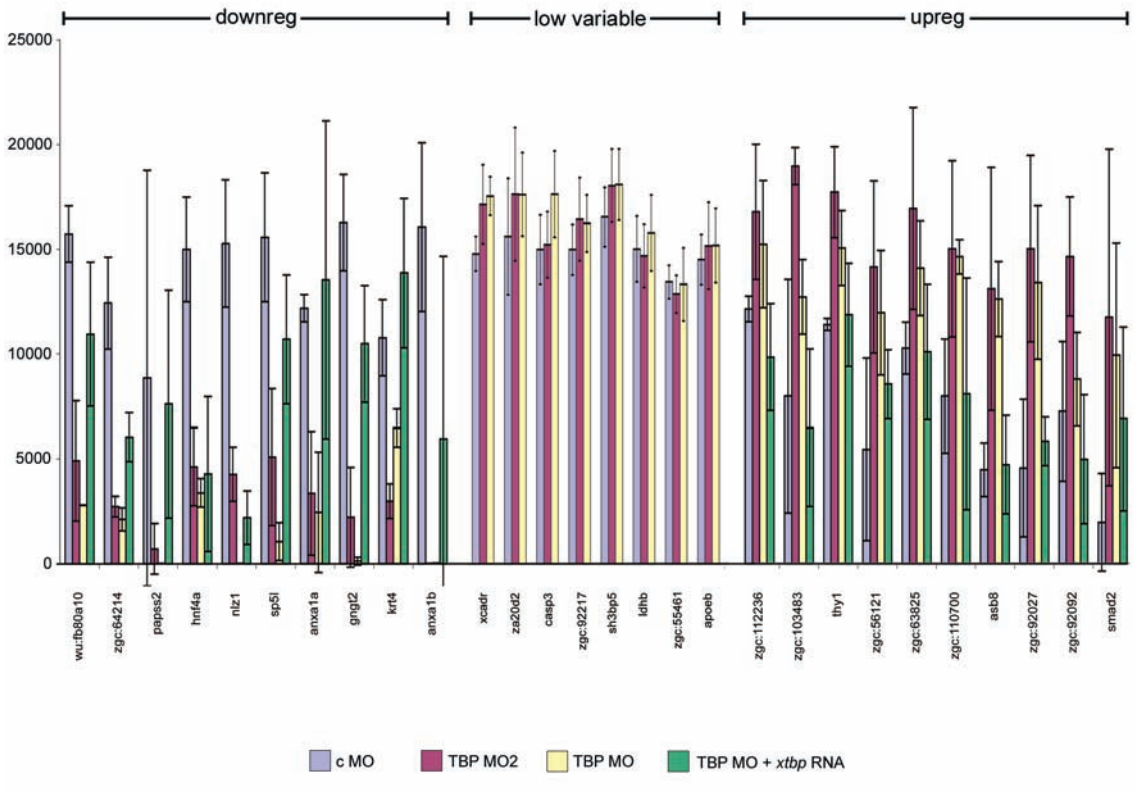

Figure 4. Quantification of semi quantitative RT-PCR to assay gene expression changes in TBP MO (yellow bars), TBP MO2 (red bars) and TBP MO + TBP mRNA (green bars) injected embryos in comparison to c MO (purple bars) injected embryos using the gel analysis tool of the ImageJ software. Averages of triplicates are given with standard deviations.

gene function. However the observed phenotype can also be caused by non specific effects of morpholino oligonucleotides, or effects due to unexpected complementary of the oligonucleotide to other genes (Nasevicius & Ekker, 2000). To confirm the specificity of the TBP-MO used in the microarray study, a morpholino targeting a non-overlapping sequence more upstream in the *tbp* mRNA was designed (TBP MO2) and injected into one-cell stage embryos parallel to c MO. To address if TBP MO2 recapitulates the change of RNA levels in the subset of genes selected to verify the microarray results, RT-PCR was carried out on cDNA templates originating from RNA of TBP MO2 and c MO injected embryos collected at dome-stage. Using the same PCR conditions and primer pairs as in the above mentioned analysis, the same set of 28 genes were analysed for their response to TBP knockdown by TBP MO2 in three independent repeats. A comparison of signal intensities quantified by imageJ shows that 27 of the 28 genes analysed display comparable changes in mRNA levels in TBP MO2 injected embryos as in TBP MO injected embryos compared to the c MO injected control (Fig. 4). The only exception is *zgc:92092*, whose mRNA levels are increased in TBP MO injected embryos compared to the c MO injected control, but show equal mRNA levels in TBP MO2 injected embryos when compared to the c MO injected control. As the majority of analysed genes displays similar changes in RNA levels of TBP MO and TBP MO2 injected embryos compared to the c MO injected control, the analysis indicates that TBP MO specifically blocks TBP-function.

Furthermore, the specificity of the MOs used in this study was ascertained by the ability to rescue the changes in gene expression caused by TBP MO in comparison to c MO injected embryos. *Xenopus tbp* RNA (*xtbp*), truncated in the 5' UTR (Veenstra *et al*, 2000) and lacking the sequence essential for the interaction with TBP MO, was in vitro transcribed to rescue the lack of endogenous TBP protein. To prevent unspecific binding of TBP MO to the *xtbp* RNA, a double injection was carried out: After injection of TBP MO, the embryos were split and separated into two batches. Subsequently the embryos were injected a second time with either *xtbp* RNA or a RNA coding for the bacterial transposase Is30. Subsequent semi-quantitative RT-PCR using the same PCR conditions and primer pairs as in the above mentioned analysis on a subset of 20 down- and upregulated genes was carried out. As no significant change in the expression levels in the class of low variable genes was expected, this set was excluded from the analysis. A comparison of signal intensities of the treatment groups quantified by imageJ indicates that in 18 of the 20 analysed genes, TBP MO + *xtbp* injected embryos, but not TBP MO + *is30* injected embryos shows similar changes in RNA levels compared to the c MO injected control. This finding suggests for a rescue of

steady state mRNA levels as the signal strength in the class of upregulated genes as well as the downregulated class approximate to the signal intensities of the c MO injected control (Fig. 4). Exceptions are *hnf4b* and *nlz1* from the group of downregulated genes. The signal intensities of both genes do not demonstrate a rescue and should be excluded from further analysis, as the reduced mRNA levels in TBP morphants could be caused by an unspecific effect of the MO.

Taken together, these results indicate that in 90% of cases, the observed changes in gene expression levels on the microarray are specifically attributable to the knockdown of TBP using antisense morpholino oligonucleotides.

4.1.3 TBP morphants utilise transcriptional startsites in close proximity to the wildtype TSS

Recent studies revealed that many vertebrate genes have functional alternative promoters and that a single TSS is rather the exception (Carninci *et al*, 2006; Cooper *et al*, 2006; Sandelin *et al*, 2007). Therefore, depletion of TBP could lead to the utilisation of alternative pre-initiation complexes nucleating at alternative promoters and transcriptional startsites, which, in return, could lead to an ectopic activation of transcription and constitute as low variable change in gene expression on the microarray in the comparison to c MO injected embryos. To address this question, rapid amplification of cDNA ends on the 5' end of transcripts (5' Race) was carried out on 4 genes representing the class of low variable genes on the microarray. Beforehand, as an utilisation of alternative promoters can only be addressed in zygotically active genes, RT-PCR analysis of these 4 genes was carried out in pre and post MBT stages to exclude a possible interference with maternally deposited gene products. Subsequently, 5'Race was carried out to assess the transcriptional startsite in TBP MO injected embryos compared to c MO injected embryos. As a specificity control for the primer

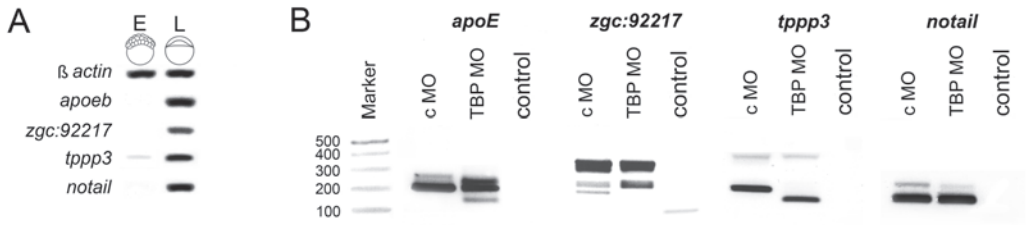

Figure 5. Utilisation of transcriptional startsites in TBP depleted embryos. (A) RT-PCR of genes selected for 5'RACE in pre MBT- and post MBT stages to exclude maternal contribution. (B) Analysis of TSS selection in c MO and TBP MO injected embryos by 5'RACE. Abbreviations, E, early embryos before MBT; L, late embryos after MBT

pairs in the PCR reaction, cDNA templates were used lacking the adaptor oligonucleotide. In 3 out of 4 analysed genes, transcription preferentially starts in close proximity (Fig. 5B), demonstrating that in TBP-depleted embryos, the formed pre-initiation complexes are able to utilise the same startsite of transcription as in c MO injected embryos. It also demonstrates that in these genes a PIC is nucleated that does not contain TBP. Subsequently this result suggests that the low variable class comprises genes, which do not require TBP to initiate transcription. However there is also evidence for differences in the utilisation of transcriptional startsites. In TBP depleted embryos, transcription of the gene *tppp3* utilises a startsite that is not detectable in c Mo injected embryos. This could suggest for an alternative pre-initiation complex formed in the 5' region of *tppp3* that does not contain TBP and the low variability in gene expression observed on the microarray could be due to the usage of this alternative PIC and to an ectopic activation.

4.1.4 Maternal transcripts are enriched in the class of upregulated Genes

As the zygotic genome in pre-MBT stages is inactive, early embryonic development depends on maternal gene products deposited in the egg. The activation of the zygotic genome is accompanied by the degradation of these maternal mRNAs and involves both transcription dependent and –independent processes (Bashirullah *et al*, 1999; Mathavan *et al*, 2005) (Fig. 6). If TBP-function is required in this process, maternal RNA degradation would be impaired in TBP MO injected embryos and, in the comparison to c MO injected embryos, these RNA species are likely to be contained in the class of upregulated- and low variable genes on the TBP knockdown microarray. To address this question, the intersection of a database of 622 transcripts showing highest abundance in the unfertilised egg (Mathavan *et al*, 2005) to the TBP knockdown microarray was built to identify maternally deposited transcripts in the three classes of the TBP knockdown microarray. As both datasets are based on EST sequences, the datasets were mapped to the UniGene database of clustered EST sequences to maximise the number of genes overlapping both datasets (see 3.5.4 for details). This resulted

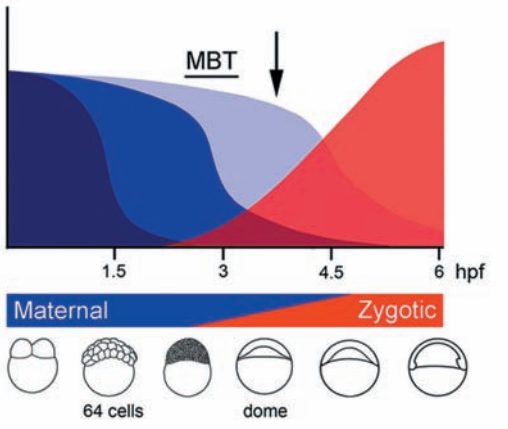

Figure 6. Schematic representation of the dynamics of maternal RNA degradation and zygotic gene activation. The arrow marks the time point embryos were collected for microarray analysis. Abbreviations, hpf, hours post fertilisation.

in the identification of 1733 UniGene clusters in the TBP microarray dataset and 571 Unigene clusters in the dataset of preferentially maternal transcripts. The intersection was then built using Unigene identifiers and led to the identification of 143 maternal transcripts overlapping the TBP knockdown dataset and the dataset of maternal transcripts (Supplementary table III). Figure 7A shows that 15.8% of the upregulated class of genes are primarily maternal, whereas only 1.5% of the genes present in the downregulated class fall into that category (The statistical analysis of the intersection by Remo Sanges could demonstrate that the increase of maternal transcripts in the class of upregulated genes is significant (Ferg *et al*, 2007)). The class of low variable genes holds 6.8% maternal transcripts. This finding suggests for an impaired degradation of maternally deposited transcripts in TBP depleted embryos.

The dataset of maternal transcripts (Mathavan *et al*, 2005) is based on transcripts showing highest abundance in the unfertilised egg, including maternally deposited transcripts, which are degraded before MBT, suggesting for a transcription independent degradation mechanism, that does not require TBP. To elucidate, if maternal transcripts degraded by a transcription dependent process are enriched in the class of genes upregulated in TBP morphants, a microarray experiment was performed on dome stage embryos in which transcription dependent processes were blocked by injecting alpha-amanitin into one cell stage embryos. Alpha-amanitin treatment impairs epiboly (Baumann & Sander, 1984; Kane *et al*, 1996) and causes a block of polymerase II transcription (Kedinger *et al*, 1970). As a consequence, maternal mRNAs, which are degraded shortly after MBT by a transcription dependent process, accumulate and, compared to the water injected control, show el-

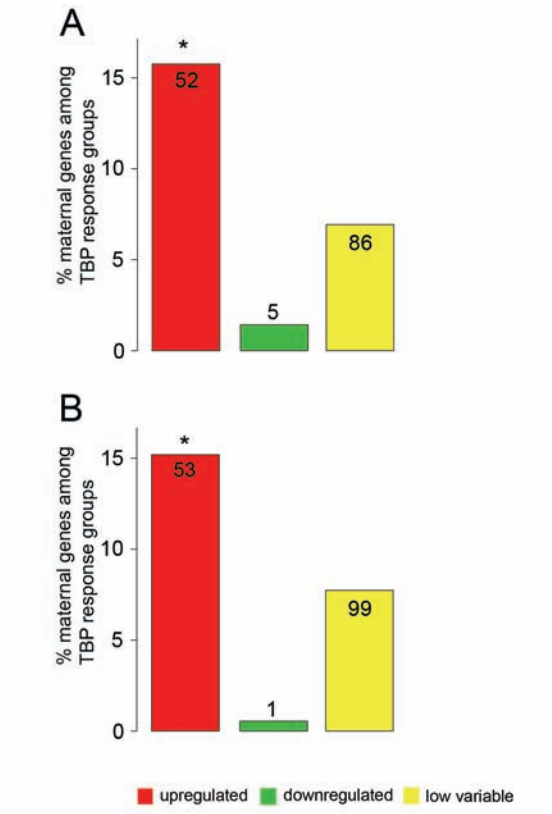

Figure 7. Distribution of maternally expressed genes among significantly regulated genes present on the TBP knockdown microarray given as percentages. The number of overlapping genes are given in the respective bars. (A) Intersection of the TBP knockdown microarray with genes showing highest abundance in the unfertilised egg (Mathavan *et al*, 2005). (B) Intersection of the TBP knockdown microarray with genes significantly upregulated on the alpha-amanitin microarray. Bars marked with an asterisk show statistical significant enrichment of maternal genes.

evated levels of this RNA species. To ascertain that transcriptional processes were blocked in alpha amanitin injected embryos, expression levels of the zygotically active gene *notail* were analysed by RT-PCR on a significant proportion of alpha-amanitin- and water injected embryos (data not shown). The experimental setup of the alpha-amanitin microarray was identical to the TBP MO microarray and was done in 2 biological repeats by Dr. Lixin Yang in the laboratory of Prof. Dr. Uwe Strähle. The exclusion of genes represented by less than 3 data points and applying an FDR cut-off of ffi 0.05 to eliminate false positives and false negatives, resulted in 409 supposedly maternal transcripts that show a more than 2-fold upregulation in alpha-amanitin injected embryos compared to the water injected control (Supplementary table IV). The intersection of the TBP-MO microarray dataset to the set of upregulated genes in the alpha-amanitin microarray is based on UniGene identifiers (see 3.5.4 for details) and shows that 15.4% in the class of genes upregulated through TBP knockdown are likely maternal, whereas only 0.3% of the transcripts present in the class of genes downregulated through TBP knockdown are of maternal origin (Fig. 7B, Supplementary table V). This result indicates for an enrichment of maternally deposited transcripts degraded by a transcription dependent process in the class of genes upregulated by TBP knockdown and suggests for an impaired degradation of maternally inherited transcripts in TBP morphants and a requirement for TBP-function in this process.

4.1.5 TBP is required for the degradation of maternal transcripts

The bioinformatic analysis suggested that the degradation of maternal transcripts could be dependent on TBP-function. To scrutinise the predicted involvement of TBP in the degradation of maternal transcripts, the steady state mRNA levels of 2 individual genes, *zorba* (Bally-Cuif *et al*, 1998) and *smad2* (Muller *et al*, 1999) were analysed. Both transcripts have been reported to be present in the oocyte and are upregulated on the TBP knockdown microarray 2.5 fold and 2.3 fold respectively. Therefore, *zorba* and *smad2* are ideal candidates to address the question if maternal RNA degradation is perturbed in TBP morphants. The steady state RNA levels of *zorba* and *smad2* were analysed by RT-PCR in pre- and post MBT stages in TBP MO and c MO injected embryos (Fig. 8). In pre MBT stages, the maternal mRNA component of *zorba* and *smad2* can be detected in TBP-MO as well as c MO injected embryos and is of comparable strength (Fig. 8, comparing lane 1 to 3). In post MBT stages however, TBP morphants exhibit elevated *zorba* and *smad2* RNA levels in TBP MO injected embryos compared to c MO injections (Fig. 8, comparing lane 2 to 4). The control gene *beta-actin* did not show a change in expression levels in post MBT stages. This result suggests that TBP-func-

tion is required for the degradation of maternally deposited transcripts.

smad2, like many other genes, is expressed both maternally and zygotically (De Renzis *et al*, 2007; Muller *et al*, 1999). Therefore, elevated mRNA levels on the TBP knockdown microarray could not only be due to a block of mRNA degradation, but also due to a premature activation of the zygotic genome. To address this question, synthetic *smad2* RNA (*smad2(s)*) was injected into one-cell stage embryos. If a loss of degradation of this synthetic RNA in TBP morphants can be observed in comparison to the c MO injected control, would indicate for a loss of degradation of maternal RNAs. However, it does not exclude a premature activation of the zygotic genome. To determine if RNA degradation is perturbed in TBP morphants, synthetic *smad2* RNA was co-injected with either c MO or TBP MO. The reduction of the cycle

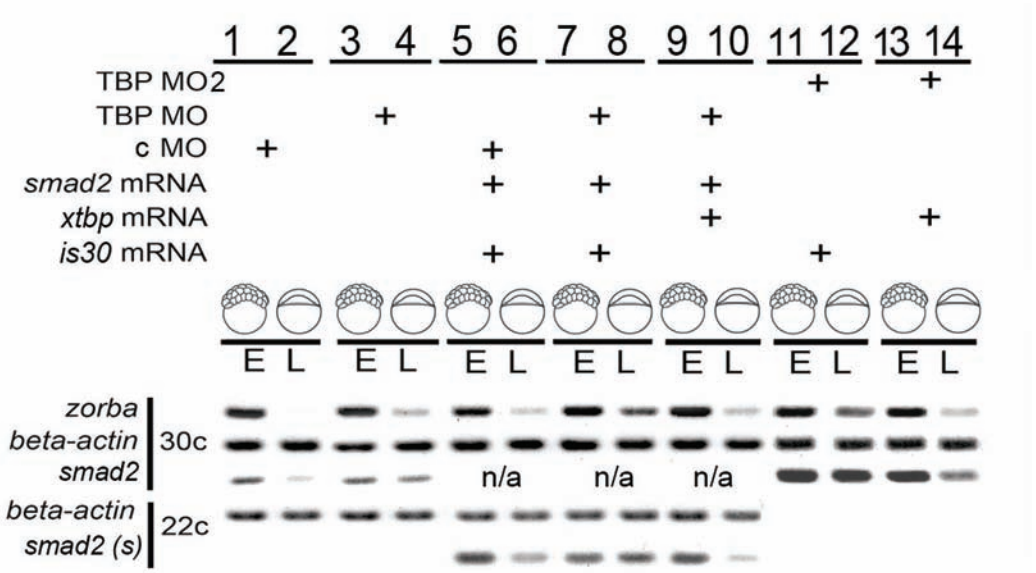

Figure 8. RT-PCR analysis of selected maternally expressed genes before MBT and after MBT. Injection of MOs, *xtbp* or *is30* transposase mRNAs are indicated by "+" symbols. Abbreviations, hpf, hours post fertilisation; up, upregulated; down, downregulated; low var, low variable expression in TBP MO injected embryos; c, cycle number of PCR reactions; s, synthetic mRNA; E, early embryos before MBT; L, late embryos after MBT.

numbers in the subsequent RT-PCR ensured that only the signal of the synthetic *smad2* RNA was detected. In pre MBT stages of c MO- as well as TBP MO co-injected embryos, the signal of *smad2(s)* RNA is present in comparable strength (Fig. 8, comparing lane 5 to 7). In contrast to this, post MBT stages show a higher *smad2(s)* RNA level in TBP MO injected embryos compared to c MO injections (Fig. 8, comparing lane 6 to 8). This result suggests it to be unlikely that the elevated mRNA levels observed in TBP morphants is caused by a premature

activation of the zygotic genome, but a lack of degradation of maternally provided mRNAs through the knockdown of TBP.

As mentioned above, the observed phenotype in antisense morpholino oligonucleotide injections can be caused by non-specific effects. To test whether the lack of maternal mRNA degradation is directly attributable to a loss of TBP-function and not caused by a side effect of the morpholino antisense oligonucleotide, synthetic *xtbp* RNA was injected that, as described in 4.1.2, lacks the TBP MO interaction site. The injection followed the principle described in 4.1.2. After co-injection of TBP MO with *smad2* RNA, the embryos were separated into two batches and subsequently injected a second time with either *xtbp* RNA or *is30* RNA. In pre-MBT stages signal originating from maternal *zorba* mRNA as well as synthetic *smad2* RNA of equal strength can be detected in both treatment groups (Fig. 8, comparing lane 7 to 9). In post-MBT stages only the TBP MO + *xtbp* injected embryos show reduced levels of maternal *zorba* mRNA and synthetic *smad2* RNA. A signal reduction in the TBP MO + *is30* mRNA injected embryos is not detectable (Fig. 8, comparing lane 8 to 10). The same experimental setup using TBP MO2 led to equal results in endogenous *zorba*- and *smad2* mRNA levels (Fig. 8, comparing lane 11 to 13 and 12 to 14).

As the injection of *xtbp*, but not of the *is30* control RNA can rescue the phenotype of perturbed maternal RNA degradation, the observed loss of maternal RNA degradation is directly attributable to the loss of TBP-function.

4.1.6 A subclass of maternal transcripts is degraded by a transcription- and TBP-dependent mechanism

The degradation of a large proportion of maternal transcripts already starts before the activation of the zygotic genome in several analysed organisms and involves both transcription dependent and –independent processes (Bashirullah *et al*, 1999). Furthermore, a proportion of maternally deposited transcripts persist until late gastrulation stages and beyond (Mathavan *et al*, 2005). The intersection of the TBP knockdown microarray with genes upregulated on the alpha-amanitin microarray showed that maternally deposited transcripts degraded by a transcription dependent process are also present in the class displaying low variable changes in gene expression and, to a lesser extend, in the downregulated class of genes. This suggests that the TBP-dependent degradation of maternal mRNA affects only a subclass of transcripts, as the maternal component of genes regulated by TBP knockdown could have already been eliminated before MBT or persists until later stages in development. This raises the question of the degradation dynamics of maternal RNA degra-

dation requiring TBP-function. To address this question, the analysis of steady state RNA levels representing 12 time points of zebrafish development (Mathavan *et al*, 2005) was utilised to form the degradation kinetics of maternal RNA degradation during early zebrafish development. Based on the time point when degradation of maternal transcripts intensifies, three classes of mRNA degradation patterns were established: *Fast degrading* transcripts, which start to degrade in pre-MBT stages, suggesting for a transcription independent degradation mechanism, *medium degrading* transcripts, which start to degrade at- or after MBT, indicating for a zygotic mechanism and a *late degrading* class of maternal mRNAs that do persist until somitogenesis and beyond (For details see 3.5.7) (Fig. 9A). Utilising the intersection between the TBP knockdown microarray and transcripts showing the highest abundance in the unfertilised egg established in 4.1.4, the degradation dynamics of maternal transcripts upregulated on the TBP knockdown microarray were established (Fig.

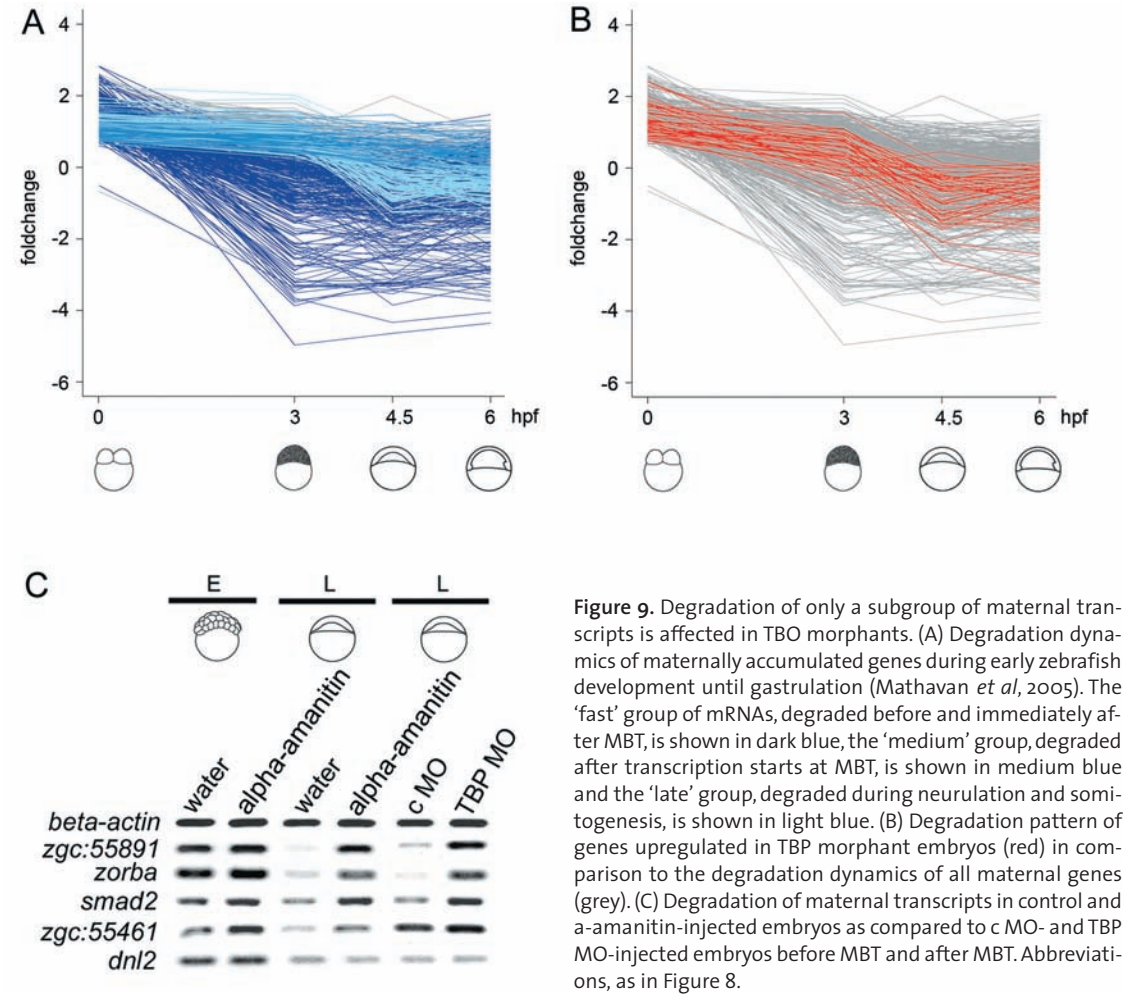

Figure 9. Degradation of only a subgroup of maternal transcripts is affected in TBO morphants. (A) Degradation dynamics of maternally accumulated genes during early zebrafish development until gastrulation (Mathavan *et al*, 2005). The 'fast' group of mRNAs, degraded before and immediately after MBT, is shown in dark blue, the 'medium' group, degraded after transcription starts at MBT, is shown in medium blue and the 'late' group, degraded during neurulation and somitogenesis, is shown in light blue. (B) Degradation pattern of genes upregulated in TBP morphant embryos (red) in comparison to the degradation dynamics of all maternal genes (grey). (C) Degradation of maternal transcripts in control and a-amanitin-injected embryos as compared to c MO- and TBP MO-injected embryos before MBT and after MBT. Abbreviations, as in Figure 8.

9B). A comparison of TBP-dependent degradation dynamics and general degradation dynamics suggests that maternal transcripts degraded by a TBP-dependent mechanism follow the degradation pattern of the *medium degrading* class (Fig. 9B). The transcripts show minimal changes in mRNA levels until the zygotic genome becomes active (3 hpf). These slight changes are followed by a sharp decrease by early gastrulation (4.5 hpf). In later stages the RNA levels are increasing again, suggesting that the zygotic activity of these genes is higher than the rate of degradation.

As the degradation of maternal transcripts upregulated in the TBP knockdown microarray preferentially starts in post MBT stages of development, this result indicates that a subclass of maternal transcripts degraded by a transcription dependent mechanism, require TBP-function degradation of maternal transcripts preferentially starts soon after MBT.

The bioinformatically established result addressing the degradation dynamics of maternal transcripts and the intersection of the TBP knockdown microarray with genes upregulated on the alpha-amanitin microarray suggests that only a subclass of maternally deposited transcripts degraded by a transcription dependent process soon after MBT requires TBP-function. To address if TBP indeed is only required for the degradation of only a subclass of maternally deposited transcripts and to validate the requirement for zygotic activity suggested by the alpha-amanitin microarray, RT-PCR was carried out on a set of genes representing the class of *medium degrading* transcripts in alpha-amanitin injected embryos in pre- and post MBT stages (Fig. 9C). In post MBT stages, the majority of genes analysed exhibit no significant signal reduction in alpha-amanitin or TBP MO injected embryos compared to pre-MBT stages, but elevated RNA levels compared to the water or c MO injected control. In contrast to the observed loss of degradation in the majority of genes, no significant effect of alpha-amanitin- or TBP MO injection can be detected in RNA levels of *dnl2*. Maternal transcripts of this gene are degraded in alpha-amanitin as well as TBP MO injected embryos.

These results indicate that the TBP-dependent degradation of maternal mRNAs depends on transcriptional activity. However, it also demonstrates that not all degradation processes require zygotic transcription or TBP as has been shown with the transcription- and TBP-independent degradation of maternally deposited *dnl2* transcripts.

Taken together, the so far gathered results demonstrate that TBP-function is required for the degradation of a subclass of maternally deposited mRNAs that are degraded shortly after MBT by a transcription dependent mechanism.

4.1.7 TBP-function is required for a specific subset of transcripts degraded by the miR-430 pathway

The transcription dependent degradation of maternal mRNA species is partially mediated by sequences in the 3' untranslated region that represent binding sites for regulatory RNAs and proteins, marking these RNAs for degradation by nucleases (Schier, 2007). Recently this decay has been linked to micro RNAs. In zebrafish, the micro RNA miR-430 has been identified to bind to a target sequence in the 3' UTR of several hundred maternally provided transcripts and promotes their deadenylation and subsequent degradation (Giraldez *et al*, 2006). Among the targets of miR-430 are the two maternally deposited transcripts of *zorba* and *smad2*. As the degradation of these transcripts is also affected in TBP morphants, TBP-dependent degradation processes could be linked to miR-430 function. To address this question, the dataset of maternally deposited miR-430 targets was compared to the TBP knockdown microarray dataset (supplementary table VI) and was established by mapping the Affymetrix identifiers to the UniGene database (see 3.5.4 for details), leading to 279 genes overlapping both datasets. The intersection indicates for an enrichment of miR-430 targets in the class of genes upregulated by TBP knockdown (Fig. 10A) (The statistical analysis of the intersection by Remo Sanges could demonstrate that the increase of miR-430 targets in the class of upregulated genes is significant (Ferg *et al*, 2007)), suggesting that miR-430 dependent degradation of maternal transcripts requires TBP-function.

The functional analysis of the requirement of TBP-function in the miR-430 dependent degradation pathway carried out by Simone Schindler could show, that in embryos injected with either c MO or TBP MO and a synthetic *gfp* RNA that contains a 3'-UTR with or without a miR-430 target site, *gfp* RNA is only degraded in the presence of a miR-430 target site in the 3' UTR of the *gfp*-RNA in c MO injected embryos, but not in the absence of a miR-430 target site or in TBP MO injected embryos. As the *gfp* RNA comprising a miR-430 target site is degraded in c MO injected embryos, but not in TBP morphants or a *gfp* RNA that does not contain a miR-430 target, these results demonstrate that miR-430 specifically promotes the degradation of these transcripts and that TBP-function is required in this process (Ferg *et al*, 2007).

The comparison of the TBP knockdown microarray dataset with the dataset of transcripts degraded by the miR-430 pathway (Giraldez *et al*, 2006) could show that a substantial number of transcripts degraded by the miR-430 dependent degradation pathway overlap with the class of genes downregulated by TBP knockdown. As downregulation can only occur if TBP-function is required for the transcription of the gene and the maternal

component is degraded by a TBP-independent mechanism, it could imply for a miR-430 dependent degradation in a TBP-independent manner. To address if miR-430 dependent degradation follows similar degradation dynamics as TBP-dependent degradation of maternal transcripts, the dataset of steady state RNA levels representing 12 time points of zebrafish development (Mathavan *et al*, 2005) was utilised to build the intersection with the dataset of transcripts degraded by the miR-430 pathway (Giraldez *et al*, 2006). As the miR-430 dataset is based on Affymetrix identifiers, the dataset was mapped to the UniGene database to build the intersection with the dataset of maternal degradation dynamics and plot the resulting overlap in a similar approach as described in 4.1.7. Figure 10B suggests that, al-

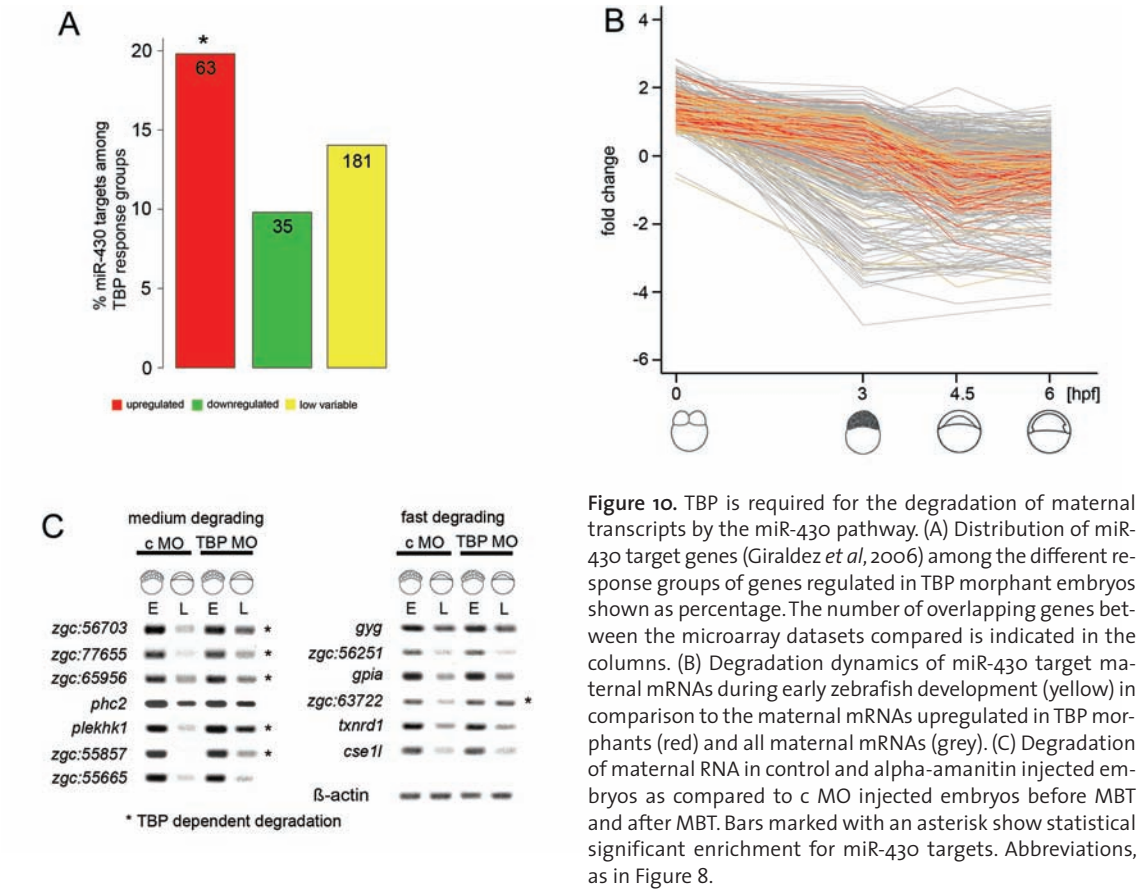

Figure 10. TBP is required for the degradation of maternal transcripts by the miR-430 pathway. (A) Distribution of miR-430 target genes (Giraldez *et al*, 2006) among the different response groups of genes regulated in TBP morphant embryos shown as percentage. The number of overlapping genes between the microarray datasets compared is indicated in the columns. (B) Degradation dynamics of miR-430 target maternal mRNAs during early zebrafish development (yellow) in comparison to the maternal mRNAs upregulated in TBP morphants (red) and all maternal mRNAs (grey). (C) Degradation of maternal RNA in control and alpha-amanitin injected embryos as compared to c MO injected embryos before MBT and after MBT. Bars marked with an asterisk show statistical significant enrichment for miR-430 targets. Abbreviations, as in Figure 8.

though the majority of miR-430 targets follows a similar degradation pattern as the TBP-dependent mRNA degradation, a number of transcripts is following the group of fast degrading transcripts that are degraded before- or at MBT. To address if these fast- and miR-430 dependent degradation are TBP-dependent, RT-PCR was carried out to elucidate the steady state mRNA levels of a number of miR-430 dependent, fast- and medium degrading

transcripts in c MO or TBP MO injected embryos (Fig. 10C). In pre-MBT stages, all analysed genes in both groups display comparable steady state mRNA levels in c MO and TBP MO injected embryos. In post MBT-stages however, 5 out of 7 analysed medium degrading transcripts show a stronger signal in TBP MO injected embryos compared to the c MO injected control, suggesting that the degradation of these transcripts is depending on TBP-function. In contrast to the finding that the majority of medium degrading transcripts require TBP-function, in the group of fast degrading transcripts, only 1 out of the 6 analysed genes displays elevated levels of maternally deposited mRNAs in the TBP MO injected embryos compared to the c Mo injected control in post MBT stages, suggesting, that the degradation mechanism of medium degrading transcripts are more likely to be dependent on TBP-function.

Taken together, these results indicate that TBP-function is preferentially required for the degradation of a specific subset of transcripts, which are degraded by a miR-430 dependent mechanism in a narrow time window starting in late blastula stages.

4.2 Bioinformatic analysis of gene promoter sequences in relation to their requirement for TBP-function

As only the minority of genes present on the microarray exhibit significant changes in gene expression levels, suggests that alternative, TBP-independent, promoter recognition complexes regulate the transcriptional activity of many genes. However, the recognition sites for protein-complexes on the DNA that enable the directional assembly of a pre-initiation complex are yet unknown. Furthermore, recent approaches in elucidating the frequency of the TATA box (Basehoar *et al*, 2004; FitzGerald *et al*, 2004; Ohler, 2006) suggest that TATA-driven nucleation of the pre-initiation complex is the exception rather than the rule (reviewed in Sandelin *et al*, 2007).

Drawing a concluding from these results, suggests that the sequence requirements for TBP-dependent activation of transcription are also largely unidentified. The obtained results of the large scale gene expression profiling in TBP depleted embryos establish the ideal basis to address the question of the sequence requirements and motif composition of gene promoters transcribed by TBP-dependent and -independent mechanisms.

4.2.1 Identification of DNA elements in the Ensembl promoterome

The majority of known core promoter motifs display a strong positional bias in respect to the transcriptional startsite. Therefore, characterisation of core promoters in detail makes it necessary to validate the TSS of the gene. The resource for genomic data, the Ensembl project (Flicek *et al*, 2008), holds information of 17330 cDNAs mapped to the zebrafish genome. However the sources of these cDNAs are mainly EST clusters, suggesting that only a fraction of cDNAs have a complete 5' end. As a means of validating the integrity of the 5' end of transcripts mapped to the zebrafish genome by the Ensembl project, the 5' UTR-containing transcriptome was extracted using the Ensembl core database "danio_rerio_core_46_7". This database was queried for all stable Ensembl protein coding genes containing transcripts with an annotated 5' UTR and led to the identification of 7809 genes with a putative TSS marked by the 5' end of the transcript. The retrieved sequences were than compared to the database of verified transcriptional startsites (DBTSS) (Wakaguri *et al*, 2007), which holds experimentally gained information of 15.198 transcriptional startsites mapped to 3417 zebrafish genes. The promoters of this database include a region 1000 bp upstream- and 200 bp downstream of the validated TSS. The basic local alignment tool (BLAST) (Altschul *et al*, 1990) was used to identify congenial sequences which share >95% identity over a region of >100 bp in the overlapping 200 bp window and led to the identification of 985 genes present in both datasets. The deviation of the TSS-position of the two aligned sequences was than calculated by the 5'-overhang of either the query-sequence (Ensembl) or the hit-sequence (DBTSS). By arbitrarily clustering TSS positions into two classes, where startsites originating from the query extending more 5' than DBTSS, obtain the algebraic sign plus (Fig. 11B) and startsites extending less 5' than DBTSS obtain the algebraic sign minus (Fig. 11C). This way, a comparative TSS-distribution could be formed as a histogram (Fig. 11A).

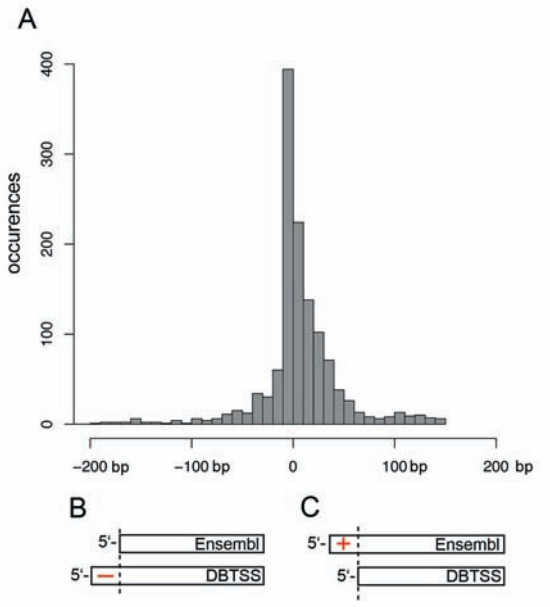

Figure 11. Variance analysis of experimentally verified 5' ends (DBTSS)(Wakaguri *et al*, 2007) and predicted 5' ends (Ensembl) (Flicek *et al*, 2008). (A) Distribution of 5' ends depicted as a histogram. (B) Schematic representation of TSS clustering. Transcriptional startsites extending less 5' than DBTSS obtain the algebraic sign minus, startsites extending more 5' than DBTSS obtain the algebraic sign plus (C).

The comparison of 5' ends depicted in the graph led to the finding that the 5' end of 643 Ensembl sequences (65%) lie within a ±20 bp window of the experimentally verified TSS, suggesting that it is possible to characterise promoters based on predicted startsites present in the Ensembl database with ±20 bp accuracy.

To estimate the pervasiveness of core promoter elements in the Ensembl promoterome, the promoters of the above-mentioned database of 7809 protein coding genes with an annotated 5'UTR were extracted. The putative promoter region (PPR) includes sequence information 1000 bp upstream and 500 bp downstream of the putative TSS and was stored in a Fasta formatted file (Supplementary file I). The majority of core promoter elements are highly degenerate, which tends to lead to a high false discovery rate when attempting to detect them in promoter sequences. To estimate the random occurrence of core promoter motifs in the core promoter region, a set of control sequences was generated by bootstrapping the Ensembl PPR database. The PPR database was permutated by re-sampling the nucleotides in 10bp windows (Supplementary file II) (Frith *et al*, 2003; Jin *et al*, 2006). Compared to the Ensembl PPR database and the DBTSS promoters, this control set of randomized promoters shows the same nucleotide frequencies as the Ensembl PPR database (Fig. 12A, B), but the control sequences have no preference for a DNA motif to cluster at a certain position, making it possible to ascertain the random occurrence of a DNA-motif underlying the nucleotide composition around the startsite of transcription.

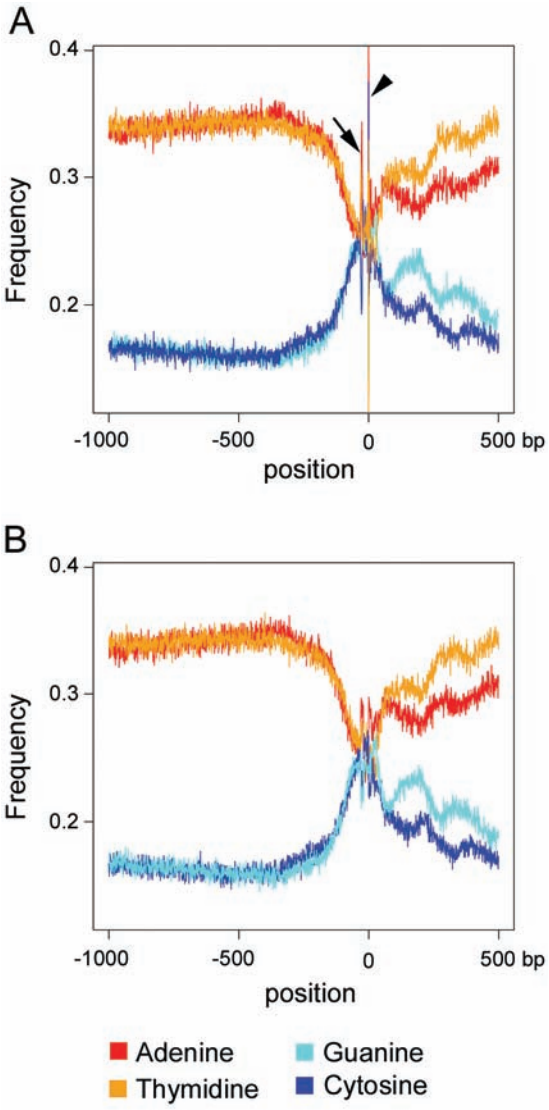

Figure 12. Frequency of occurrences of nucleotides in the promoter region of the Ensembl PPR dataset (A) and the bootstrapped control sequences (B). The arrow points to an A-T rich region, which could represent the TATA box. The arrowhead points to the dinucleotide CA, overrepresented at the putative TSS.

It was reasoned that if a DNA sequence clusters relative to the TSS, the DNA sequences that are in the cluster have a high likelihood of being biologically significant (FitzGerald *et al*, 2004). To test if core promoter elements cluster at a certain position, a finding that would underline the significance of the motif, the distribution of the DNA-motifs over the whole promoter region was examined using the consensus sequences listed in table 1 and plotted as histograms. Compared to this permutated control set, the human (Corden *et al*, 1980)- and *Drosophila* (Hultmark *et al*, 1986) initiator as well as the TATA box (Bucher, 1990) display a well-defined peak at the consensus position (Fig. 13), suggesting that these DNA motifs preferentially occur at a distinct position relative to the TSS. The Downstream promoter element DPE (Burke & Kadonaga, 1996) however does not show clustering in the analysed Ensembl PPR sequences (Fig. 13). The DPE exhibits a high frequency of occurrences and forms a distribution-pattern very similar to the permutated control, suggesting that the distribution of the DNA element reflects the nucleotide composition in the core promoter region. The lack of an occurrence-peak at the consensus position indicates that the random distribution of the DPE, most likely caused by the degeneracy of the consensus sequence, covers the positional constrained abundance, which is more likely to be functional.

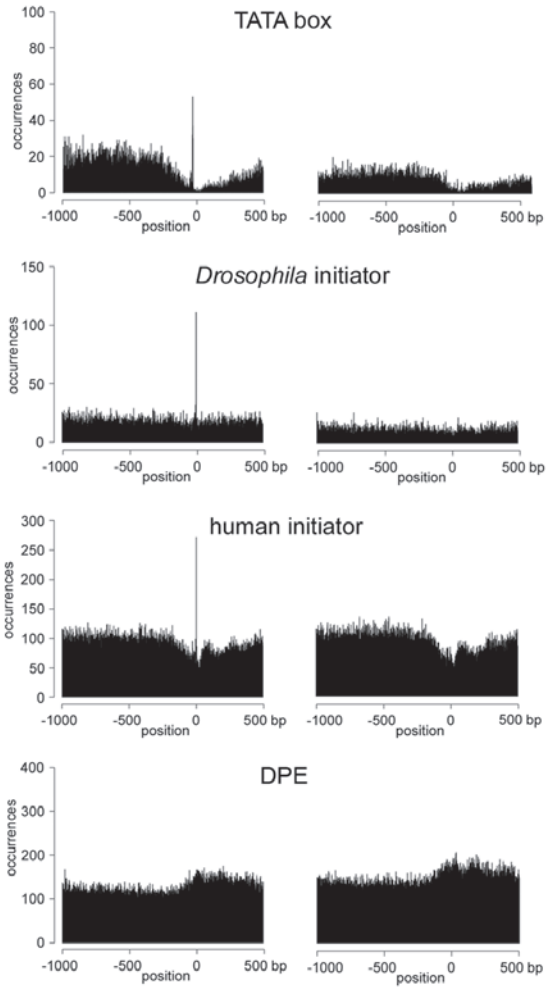

Figure 13. Clustering analysis of selected core promoter elements on the Ensembl PPR dataset and the bootstrapped control sequences in a region from -1000 bp to +500 bp represented as histograms (1 bp resolution).

The Ensembl PPR database was than searched for the same core promoter elements and CpG islands (Bird, 1986) in a position restricted approach using the consensus sequences and search windows listed in table 1, allowing no mismatch. Table 2 and Figure 14 show the total number of occurrences and the resulting percentages of core promoter motifs found

in the Ensembl PPR database and in the permutated control set. As already suggested by the distribution of motifs in the putative promoter region (Fig. 14), the percentage of found DPEs in the Ensembl PPR dataset shows no differs only slightly from the permutated control. A two-sided statistical test for equality suggests that these differences are statistically insignificant (p = 1). Compared to the permutated control set, the TATA box is significantly increased in the Ensembl PPR dataset (p < 2.2e-16) and is present in 4.1%

Motive	Consensus sequence	Consensus position	search window
TATA box	TATA[AT]A[AT][AG]	-28 to -21	-38 to 10
Initiator-human	[CT][CT]AN[AT][CT][CT]	-2 to 5	-5 to 5
Initiator-Drosophila	TCA[GT]T[CT]	-2 to 4	-5 to 5
DPE	[AG]G[AT][CT][GAC]	28 to 32	8 to 52

Table 1. Consensus sequences and search windows of selected core promoter elements. References to the consensus sequences are given in the text.

of the 7809 analysed promoter sequences, a finding which underlines the low frequency of the motif and is in line with previous reports (Cooper *et al*, 2006; FitzGerald *et al*, 2004; Kim *et al*, 2005). The initiator of the *Drosophila* consensus sequence is found in 4% of the analysed sequences compared to 1.2% in the permutated control, a significant increase of the motif compared to the bootstrapped control sequences (p < 2.2e-18). The human initiator also shows a significant enrichment in the promoter sequences compared to the permutated control (p = 1.5e-8). The number of found CpG islands however is not significantly increased compared to the control (p = 1). This result could be explained by the unchanged nucleotide composition the permutated control set is based on.

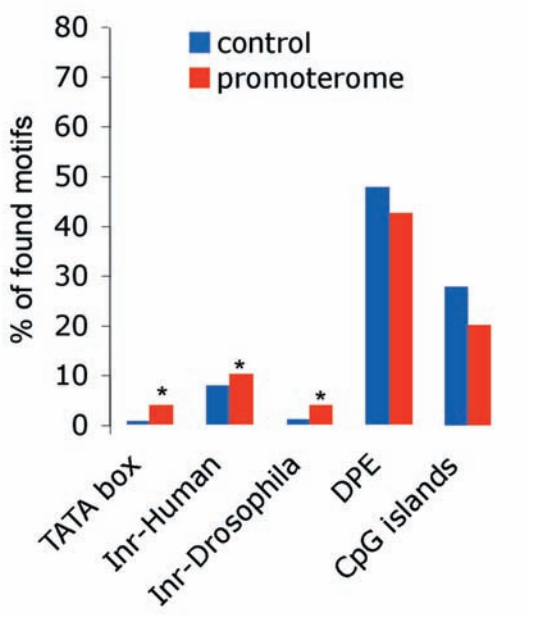

Figure 14. Position restricted search for selected core promoter elements in the Ensembl PPR dataset and the bootstrapped control sequences. The asterisks indicate significant increase of the respective motif in comparison to the bootstrapped control sequences.

Taken together, these results indicate that the initiator motifs and the TATA box can be detected over background and show an enriched occurrence at the consensus position compared to the permutated control set. The DPE shows a distribution similar to the control and does not cluster above background, suggesting that the applied bioinformatics tools are not able to dis-

criminate the DPE signal from background, probably due to a lack of DPE elements in zebrafish.

	Ensembl promoterome		permutated control promoter	
Motiv	occurances	percentage	occurances	percentage
TATA box	319	4,1	64	0,8
Inr-Human	802	10,3	618	7,9
Inr-Drosophila	311	4	96	1,2
DPE	3334	42,7	3730	47,8
CpG islands	1575	20,17	2182	27,94

Table 2. Pervasiveness of core promoter elements in the Ensembl PPR dataset and the permutated control sequences

4.2.2 Comparison of core promoter elements in the TBP knockdown microarray classes

Several subunits of TFIID are known to specifically interact with DNA motifs in the core promoter region (reviewed in Butler & Kadonaga, 2002). Recent approaches in large-scale promoter characterisation suggest that the TATA box is preferentially present in promoters of differentially expressed genes (Kim *et al*, 2005), which have been associated with single peak promoters (Carninci *et al*, 2006). However, the sequence requirements for alternative transcription initiation complexes that are likely to play a role in TBP-independent transcription mechanisms are still unknown (Muller *et al*, 2007). To elucidate aspects of the sequence requirements for TBP-independent and -dependent transcription initiation represented by the three generated TBP-knockdown microarray classes, the putative promoter region of these genes were extracted *in silico* in a similar approach as in the assembly of the Ensembl PPR dataset.

The comparison of the TBP-knockdown microarray with datasets containing primarily maternal transcripts could show that a large proportion of the class of genes upregulated by TBP knockdown are of maternal origin. As these transcripts are not expressed zygotically at the analysed stage in development, a requirement for TBP-function in regulating transcription of maternal genes cannot be addressed in the experimental system presented here. Transcripts overlapping the dataset of upregulated genes present on the alpha-amanitin microarray and the dataset of transcripts showing the highest abundance in the unfertilised egg (Mathavan *et al*, 2005) have therefore been excluded from the analysis. The resulting promoter dataset contained 120 promoters in the class of downregulated genes (Supplementary file III), 563 promoters in the class of low variable genes (Supplementary file IV) and 122 promoters in the class of genes upregulated by TBP knockdown (Supplementary file V).

Based on the results of the large-scale analysis of core promoter elements in the Ensembl PPR database containing 7809 sequences, the resulting promoters have been searched in a position restricted approach for the consensus sequences of the human- and *Drosophila* initiator, the TATA box, as well as CpG islands. As the DPE cannot be discriminated from background, the motif was excluded from the analysis. Table 3 and Figure 15 show that, in comparison to the pervasiveness of the motifs in the Ensembl promoterome, the initiator of both consensus sequences is less abundant in the upregulated class of genes, whereas the downregulated class and the class showing low variable changes in gene expression show a similar frequency of occurrences as the Ensembl promoterome. The application of a 2-sample test for equality of proportions could confirm the observation of an under representation of the human initiator in the upregulated class of genes in comparison to the Ensembl promoterome (p = 0.02056) and the downregulated class (p = 0.01957), but not in comparison to the low variable class of genes (p = 0.0928). Surprisingly, the down- and upregulated class of genes show a similar distribution of the TATA box, but the motif is less abundant in the class of low variable genes. A comparison to the large-scale analysis of promoter sequences in the Ensembl-PPR dataset also suggests for an under representation of the TATA box in the low variable class. A test for equality of proportions could confirm this observation (p = 0.01443). The analysis of the frequency of CpG islands shows an equal distribution in the TBP knockdown response classes, suggesting that the occurrence of CpG islands does not characterise gene promoters showing a differential response to TBP knockdown.

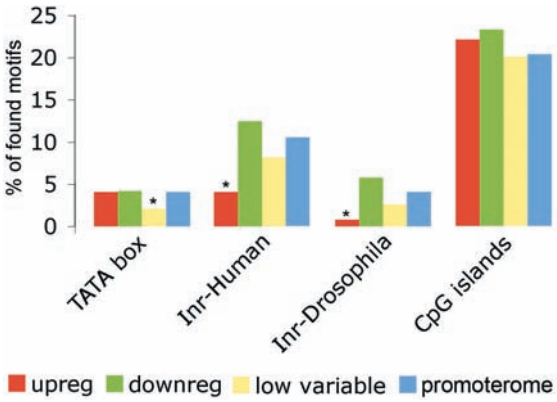

Figure 15. Comparison of selected core promoter elements in the TBP knockdown microarray classes. The asterisks indicate significant decrease of the respective motif in comparison to the Ensembl PPR dataset.

	upregulated genes (120 promoters)		Downregulated genes (122 promoters)		low variable genes (563 promoters)	
Motiv	occurances	percentage	occurances	percentage	occurances	percentage
TATA box	5	4,1	5	4,2	12	2,1
Inr-Human	5	4,1	15	12,5	46	8,2
Inr-Drosophila	1	0,8	7	5,8	15	2,6
CpG islands	27	22,1	28	23,3	113	20,1

Table 3. Pervasiveness of selected core promoter elements in the TBP knockdown microarray classes.

Taken together, these results indicate a tendency of genes showing low variable changes in gene expression by TBP knockdown to less frequently contain the TATA box motif and a decrease of initiator occurrences in the class of genes upregulated by TBP knockdown. The downregulated class of genes are characterised by a frequency of core promoter elements similar to the Ensembl promoterome.

4.3 A case study: the TBP-independent *notail* promoter

The TBP knockdown microarray could confirm the initial observation that indeed not all genes expressed in early zebrafish development and other vertebrate model organisms require TBP-function (Muller *et al*, 2001; Veenstra *et al*, 2000). In this initial analysis, *notail*, a member of the T-box family of transcription factors, which is expressed during early development in all vertebrates and plays a major role in mesoderm specification, has been demonstrated to be expressed under TBP knockdown conditions in comparable levels to the c MO injected control (Muller *et al*, 2001). The bioinformatic analysis of promoters of genes significantly regulated in TBP morphants, suggests that the class of genes showing low variable changes in steady state RNA levels, the class that could be regulated by alternative transcription initiation mechanisms, tend to harbour less initiator motifs compared to the upregulated- and downregulated classes and to the Ensembl promoterome. However this observation is based on non-functional data. Therefore, the *notail* promoter is an ideal model to address the question of the sequence requirements of alternative transcription initiation mechanisms, which most likely target the promoter of *notail*, in functional and biological relevant assays.

4.3.1 Identification and *in silico* characterization of the *notail* promoter

The majority of core promoter elements interacting with the pre-initiation complex show a strong positional bias (reviewed in Butler & Kadonaga, 2002). However only a fraction of cDNAs in public databases have a complete 5' end marking the transcriptional start-site (see 4.2.1). To characterise the *notail* promoter *in silico* and *in vivo* made it therefore necessary to ascertain the TSS of *notail*. The carried out 5' RACE using primers based on the 5' end of the RefSeq sequence NM_131162.1, resulted in the generation of one single PCR product and, after sequencing, determined the TSS 11 nucleotides upstream of the 5' end of NM_131162.1 (Fig. 16A), suggesting that the *notail* promoter is located upstream of this position.

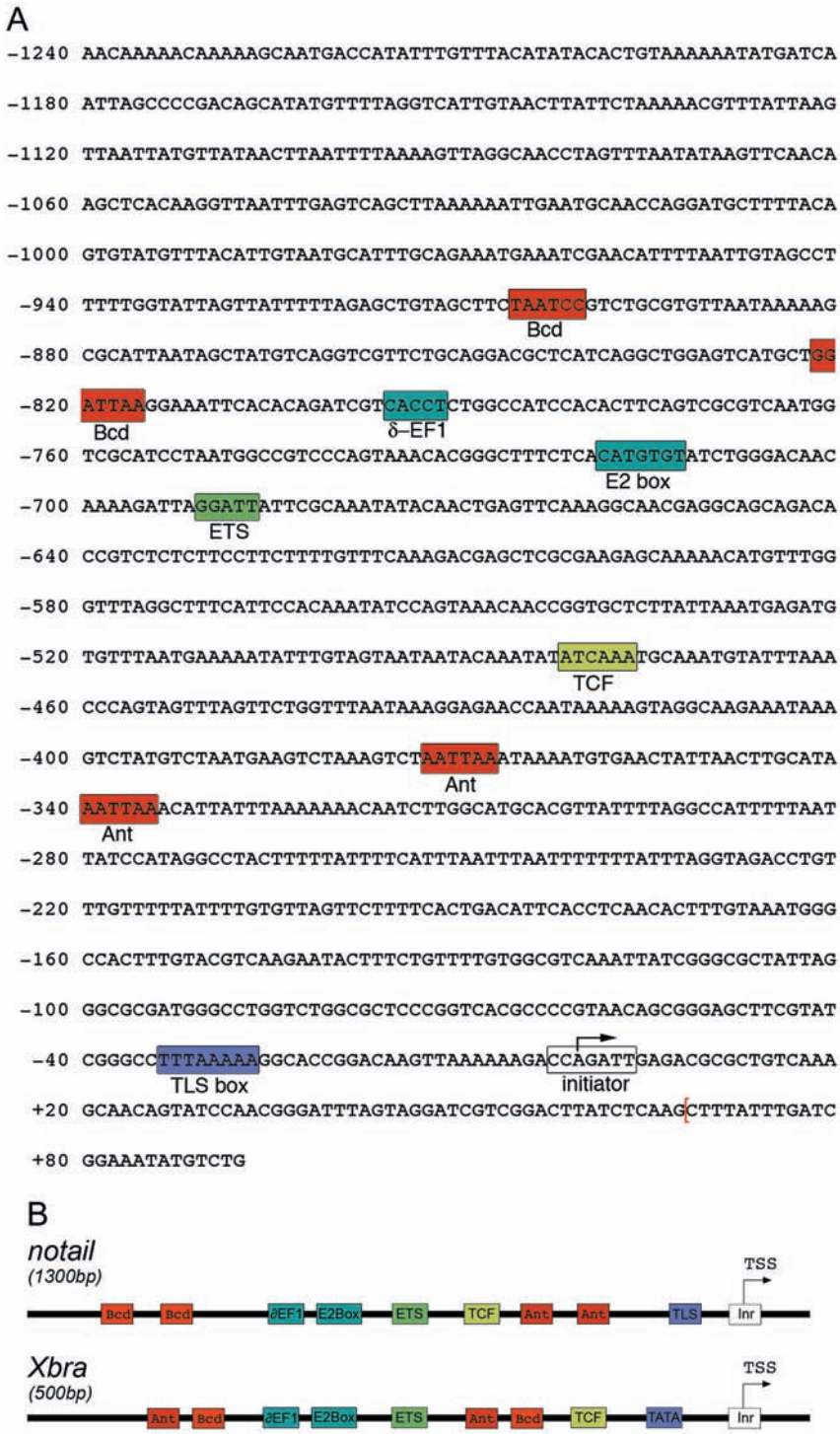

Figure 16. Transcription factor binding sites in the *notail* and *xbra* promoter sequences. (A) Sequence of the cloned *notail* promoter fragment and selected transcription factor binding sites in the *notail* promoter determined bioinformatically. The arrow marks the startsite of transcription as determined by 5'Race. The red bracked marks the 3' end of the cloned promoter fragment. (B) Selected transcription factor binding sites in the *notail* promoter and *xbra* promoter, modified from Latinkic *et. a*l, 1997.

The detailed analysis of *xbra*, the *notail* homolog in *Xenopus*, revealed a number of transcription factor binding sites that are important for the correct spatial expression of *xbra* (Latinkic *et al*, 1997; Lerchner *et al*, 2000). To elucidate if the *notail* core- and proximal promoter are composed of similar motifs, the *notail* promoter was bioinformatically analysed for putative transcription factor binding sites and revealed that the *notail* core promoter is comprised of an Initiator motif of the human consensus sequence which marks the TSS and a TATA-box showing one mismatch to the consensus sequence (TTTAAA instead of TATAAA), termed 'TLS box' (TATA like sequence) (Fig. 16B). The proximal promoter contains binding sites that have been described to be crucial for the activation and correct spatial expression in the *Xenopus xbra* promoter (Fig. 16B). *xbra* expression requires activity from canonical Wnt signalling through TCF binding sites (Vonica & Gumbiner, 2002) and FGF-signalling through the transcription factor ETS-2 for activity in the margin of the embryo (Kawachi *et al*, 2003) as well as nodal signalling for notochord expression in late/post gastrulation stages (Chen & Schier, 2001). Antennapedia- and Bicoid binding sites that are bound by Goosecoid, Mix.1 and other homoedomain transcription factors restrict the expression to ventral and lateral mesoderm (Latinkic *et al*, 1997) whereas the confinement of *xbra* to the margin is established by an E2-Box in conjuction with a deltaEF2 binding site (Latinkic *et al*, 1997; Lerchner *et al*, 2000). The *in silico* characterisation of the *notail* proximal promoter shows a comparable composition of transcription factor binding sites, which could imply that similar mechanisms control *notail* expression.

4.3.2 1.3 kb sequence upstream of the *notail* locus drives specific expression in the margin and the shield in zebrafish embryos

As a means of identifying DNA motifs conferring TBP-independence to the expression of a gene in functional assays, the promoter of the TBP-independent gene *notail* (Muller *et al*, 2001) was cloned. 1.3 kb 5' flanking region of *notail* was amplified by PCR using genomic DNA as a template. After sub-cloning the PCR-product into pCR2.1-Topo, the *notail* flanking region was excised with the restriction enzymes BamHI and XhoI and cloned into the *yfp* containing plasmid pBSII:*yfp* to monitor the expression resulting from the putative *notail* promoter.

To ascertain if the cloned promoter fragment would recapitulate the endogenous expression pattern of *notail*, 20 ng/µl of the circular plasmid containing 1.3 kb *notail* 5' flanking region (hereinafter referred to as 1.3*ntl*) was microinjected into wildtype one-cell stage embryos and transiently assayed for reporter activity at various stages during development

using UV microscopy. In 30% epiboly stages, endogenous *notail* mRNA, visualised by *in situ* hybridisation, is localised in the margin where the blastodisc meets the yolk (Fig. 17A). 1.3*ntl* is also active at this stage of development, however not in a ring like fashion, but at a distinct region near the margin of the blastodisc (Fig. 17E). In early gastrulation stages injection of 1.3*ntl* is, like the endogenous gene, leading to expression in the margin of the zebrafish embryo as well as reporter activity in the shield (Fig. 17B, F). In segmentation stages endogenous *notail* is expressed in the notochord and the budding tail (Fig. 17C). This expression pattern is recapitulated by 1.3*ntl*, but ectopic expression of the reporter was also observed (Fig. 17G). However, in late segmentation stages, 24 hours post fertilisation, when endogenous *notail* is expressed in the notochord and the tail (Fig. 17D), 1.3*ntl* shows no enrichment for YFP reporter signal in this tissues but strong YFP localisation is observed in muscle cells as well as other tissues (Fig. 17H). In contrast to this, YFP signal originating from the promoter less reporter construct was not detected in the analysed stages of development (data not shown, Fig. 18E).

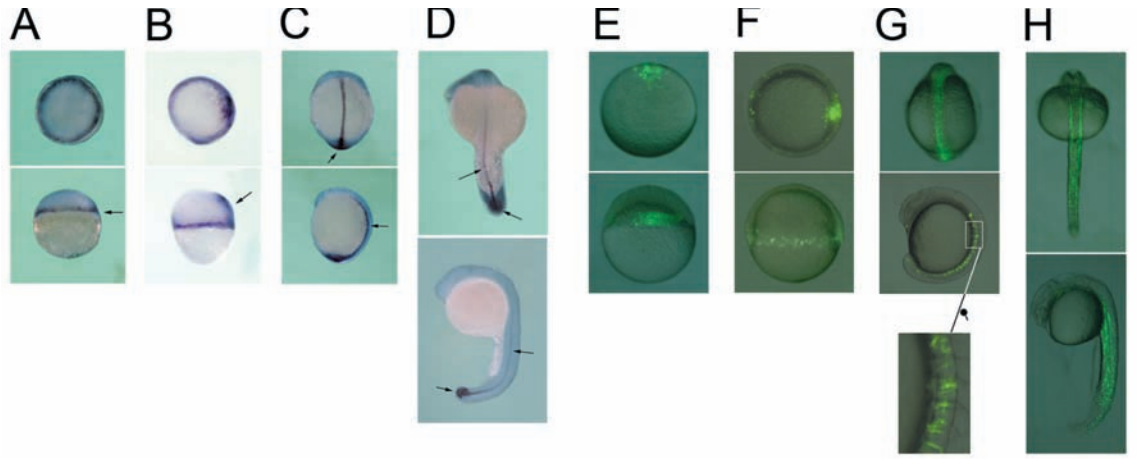

Figure 17. 1.3*ntl* recapitulates the endogenous *notail* expression pattern in early stages of development. *In situ* hybridisation of endogenous *notail* expression (A-D) and YFP expression under the control of the 1.3*ntl* reporter construct detected by UV-microscopy (E-H) in selected stages of development. (A) 30 % epiboly. The arrow points to the margin where the blastodisc meets the yolk. (B) Shield-stage. The arrow points to the margin and in the shield. (C) 6-somite stage. The arrows point to the notochord (bottom image) and the budding tail (top image). (D) 24 hpf. The arrows point to the notochord and the tail. (E) Expression of 1.3ntl at 30 % epiboly. (F) Shield stage. (G) 6-somite stage. Magnification of the white square shows specific expression in the notochord. (H) 24 hpf. Top row images show views from top, lateral views in bottom row images. Abbreviations, hpf, hours post fertilisation.

developmental stage	Analysed embryos	% yfp expressing	% spec. expression
30% epiboly	176	63,6	32,9
shield	281	96,8	68,1
6 somites	124	83,9	53
24 hpf	83	100	0

Table4. Summary of analysed embryos of 1.3*ntl* injections. Specific expression is regarded as equivalent to endogenous *notail* expression. Abbreviations, spec, specific.

These results indicate that a 1.3 kb 5' flanking region of *notail* is sufficient to recapitulate the endogenous expression pattern in early stages of zebrafish development, but not throughout development.

4.3.3 Deletion analysis of the *notail* promoter reveals a mesoderm specific region 600bp upstream of the TSS

To address the question where the regulatory region in the TBP promoter is located that is sufficient to restrict the expression to the margin and the shield in early gastrulation stages of zebrafish, a deletion series was generated by shortening the 5' end of the promoter using restriction enzymes followed by a re-ligation of the linearised plasmid (Fig. 18F). The resulting construct was then microinjected and transiently assayed for its ability to drive *notail* specific expression in the margin and the shield. Injection of a 120 bp sequence upstream of the *notail* TSS (cut with BamHI/XmnI), resulted in very weak and ectopic expression of the reporter, suggesting that this fragment is able to confer only basal activity without specificity (Fig. 18D). A reporter construct containing 300 bp upstream sequence of *notail* (cut with BamHI/SphI) leads to strong activation of the reporter in comparison to the 120 bp promoter fragment, but lacks the restriction to the margin and the notochord precursors (Fig. 18C). Injection of a 600 bp promoter fragment (cut with BamHI/SacI) was leading to the same observation, albeit stronger (Fig. 18B). As described above, a 1.3 kb promoter fragment is sufficient to lead to an expression pattern similar to endogenous *notail* in early gastrulation stages (Fig. 18A). To control that the observed expression pattern is attributable to the promoter fragment, a promoter-less control plasmid was injected. However, this construct was not able to activate the *yfp* reporter (Fig. 18E).

Taken together, these findings suggest that a region 600 bp upstream of the *notail* locus is assigning mesoderm specific activity to the *notail* promoter. These findings are in line with the *in silico* promoter characterization that shows that the majority of the elements required for the correct spatial expression in *Xenopus*, are found in a region from -600 bp to -1.3 kb in the *notail* promoter.

To precisely localise the core promoter of *notail*, the region that is able to drive basal levels of activity and is able to interact with the enhancer region, three 60 bp deletions adjacent to each other were generated by a PCR based approach as described in 3.1.13. The deletions cover the region giving basal activity and start upstream of the initiator (Fig. 19D). While the third 60 bp deletion ranging from -180 to -120 nucleotides upstream of the initiator shows equal spatial expression as 1.3*ntl*, albeit weaker (Fig. 19A), deletion 2, ranging

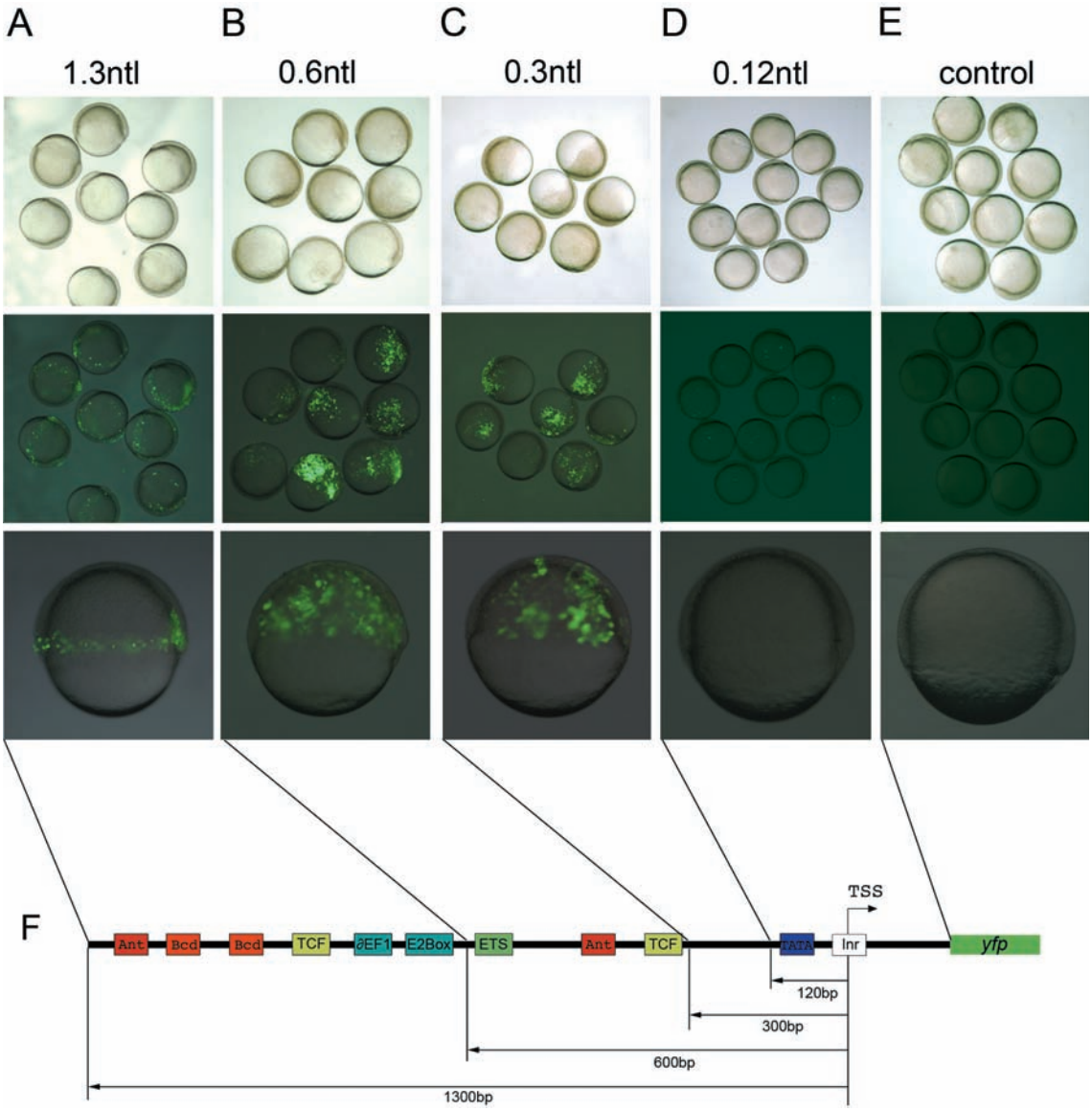

Figure 18. A shield and margin specific element is located 600 bp upstream of the *notail* locus. YFP expression of a deletion series of the 1.3*ntl* reporter construct using endogenous restriction enzyme recognition sites analysed by UV-microscopy at shield stage. (A) 1300 bp fragment (1.3*ntl*). (B) 600 bp promoter fragment. (C) 300 bp promoter fragment. (D). 120 bp promoter fragment. (E) Promoter less reporter construct. (F) Schematic representation of injected promoter fragments and included motifs in the respective promoter fragment. Group pictures of randomly oriented embryos imaged in bright field (top), fluorescence (middle) and UV-microscopy of single embryos in lateral view (bottom).

injected construct	Analysed embryos	% yfp expressing	% spec. expression
control	124	0	0
0.12ntl	108	15,7	0
0.3ntl	92	83,9	0
0.6ntl	128	100	0
1.3ntl	168	90,5	71,2

Table5. Summary of analysed embryos in the deletion series of the 1.3*ntl* reporter construct using endogenous restriction enzyme recognition sites. Specific expression is regarded as equivalent to endogenous *notail* expression. Abbreviations, spec, specific.

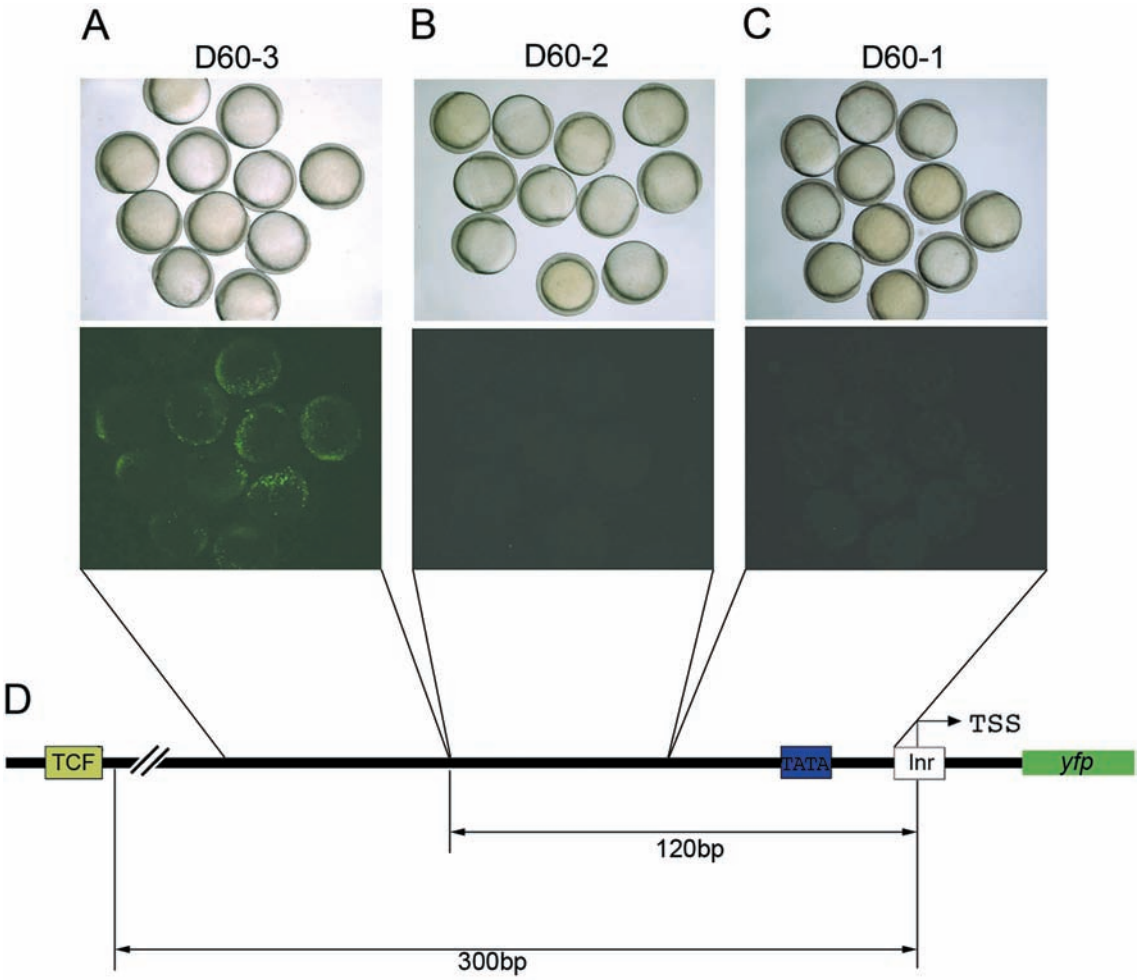

Figure 19. 60 bp deletions reveal the core promoter of *notail*. Deletion series of 3 adjacent 60 bp deletions in 1.3*ntl* analysed at shield stage by UV-microscopy. (A) Deletion ranging from -180 bp to -120 bp. (B) Deletion of a region ranging from -120 bp to -60 bp. (C) Deletion of a region ranging from -60 to +1. (D) Schematic representation of injected promoter fragments and included motifs in the respective promoter fragment. Group pictures of randomly oriented embryos imaged in bright field (top) and fluorescence (bottom).

injected construct	Analysed embryos	% yfp expressing	% spec. expression
D60-1	112	0	0
D60-2	174	15,7	5,3
D60-3	132	91,6	58,3

Table 6. Summary of analysed embryos in the 60 bp deletion series of the 1.3*ntl* reporter construct. Specific expression is regarded as equivalent to endogenous *notail* expression. Abbreviations, spec, specific.

from -120 to -60 and deletion 1 ranging from -60 to +1 are only able to activate the *yfp* reporter on very low levels (Fig. 19B, C). This indicates for a localisation of the *notail* minimal promoter in a region from +1 to -120 bp.

4.3.4 A DNA element conferring TBP-independence is located in the *notail* core promoter

To address the question if the 1.3*ntl* reporter construct shows the same independence of TBP-function as the endogenous *notail* promoter (Muller *et al*, 2001), triple-injection of 1.3*ntl* and the TBP-dependent *beta-actin*:*cfp* promoter construct together with TBP MO were performed. As a control, embryos were co-injected with c MO and analysed in transient expression assays in shield stage. Using this setup made it possible to control the injection efficiency as well as distinguish transcriptional activity of *beta-actin* and 1.3*ntl* by using different filters to detect the reporter signals (Fig. 20). To further control this experiment, alpha-amanitin, which blocks all Pol II transcription (Kedinger *et al*, 1970) was co-injected. In alpha-amanitin injected embryos no signal of the reporters CFP and YFP is detectable (Fig. 20B), suggesting that expression of the reporter genes requires transcriptional activity. In TBP MO injected embryos, the CFP signal originating from the TBP-dependent *beta-actin* promoter is lost, suggesting for efficient knockdown of TBP. In contrast to the loss of CFP signal, YFP signal originating from 1.3*ntl* is still expressed in comparable levels as the c MO injected embryos (Fig. 20A, C). This result suggests that a construct containing 1.3 kb of the *notail* promoter does not require TBP-function for transcriptional activity of the YFP reporter.

To confirm that expression of 1.3*ntl* in TBP MO co-injected embryos is not due to a lack of specificity of TBP MO, TBP MO2 was co-injected with 1.3*ntl* (Fig. 20D). Like in TBP MO co-injections, CFP signal is lost, whereas YFP signal-strength is comparable to co-injections of 1.3*ntl* with c MO, suggesting that expression of the 1.3ntl reporter construct under TBP knockdown conditions is not due to a lack of specificity of TBP MO.

To ascertain that the TBP-independence of the reporter construct is not based on the usage of an alternative TSS, diverged from the endogenous TSS, a 5' Race in c MO and TBP MO co-injected embryos carried out. The separated PCR-products by agarose gel electrophoresis (Fig. 20D) are equal in length, suggesting that in TBP morphants the same TSS can be utilised as in c MO injected embryos.

Taken together, these findings are leading to the conclusion that motifs causing the TBP-independence of the *notail* promoter are included in the isolated 1.3 kb *notail* promoter fragment.

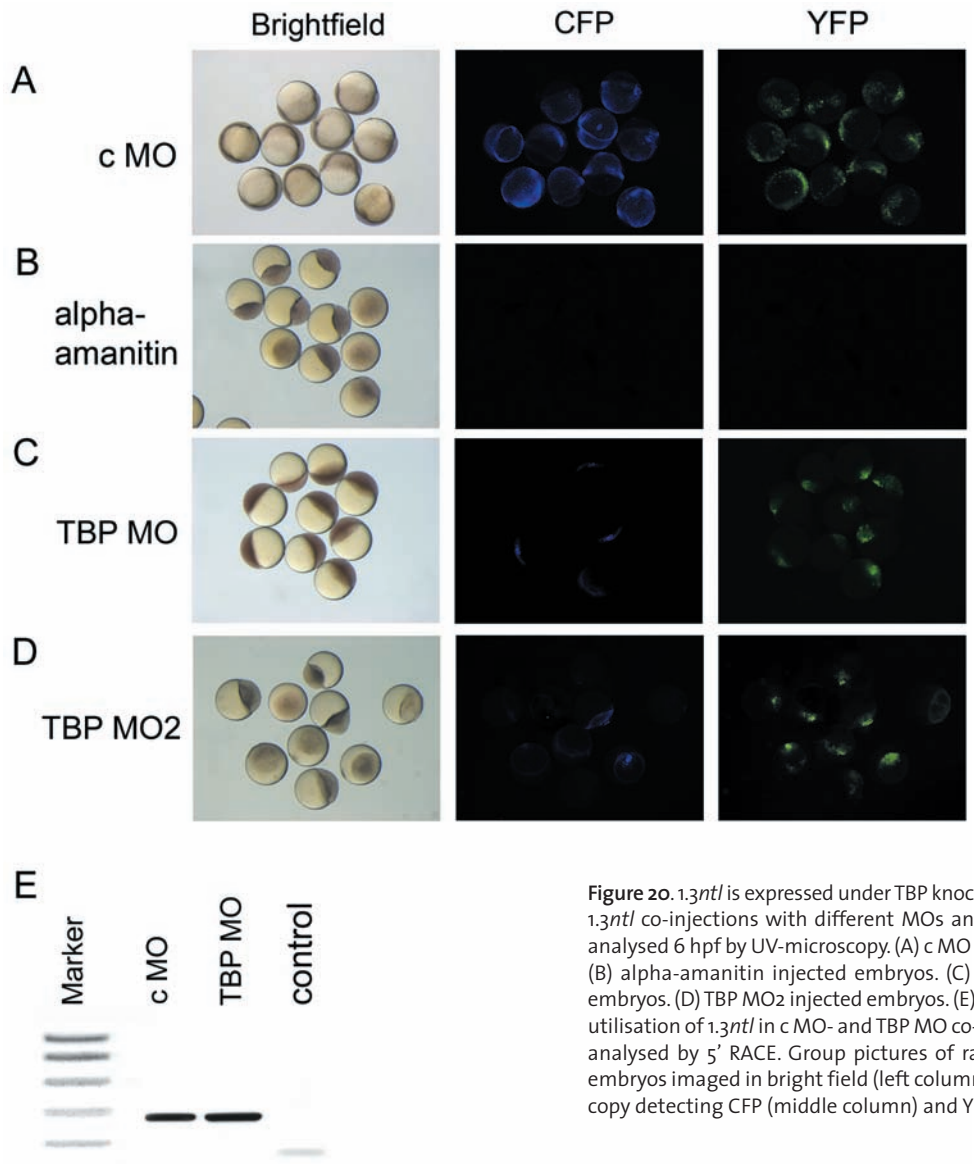

Figure 20. 1.3*ntl* is expressed under TBP knockdown conditions. 1.3*ntl* co-injections with different MOs and alpha-amanitin analysed 6 hpf by UV-microscopy. (A) c MO injected embryos. (B) alpha-amanitin injected embryos. (C) TBP MO injected embryos. (D) TBP MO2 injected embryos. (E) Gel picture of TSS utilisation of 1.3*ntl* in c MO- and TBP MO co-injected embryos analysed by 5' RACE. Group pictures of randomly oriented embryos imaged in bright field (left column) and UV-microscopy detecting CFP (middle column) and YFP (right column).

injected agent	Analysed embryos	% yfp expressing	% spec. expression	% cfp expression
c MO	204	92,6	65,3	96,1
alpha amanitin	186	0	0	0
TBP MO	167	85,6	0	3,7
TBP MO 2	124	82,9	0	2,8

Table 7. Summary of analysed embryos to elucidate the requirement for TBP-function of the 1.3*ntl* reporter construct. Specific expression is regarded as equivalent to endogenous *notail* expression. Abbreviations, spec, specific.

4.3.5 TBP-independence of *notail* is based on DNA motifs in the core promoter region

The above-described results (4.3.4) strongly suggest that the TBP-independence of *notail* is caused by sequence information located in the promoter itself. To analyse if the TBP-independence of *notail* can be transferred into a different enhancer/promoter background by replacing the *notail* core promoter with the core promoter of the TBP-dependent gene *sonic hedgehog* (Muller *et al*, 2001) and vice versa. The basis for this experiment was 1.3*ntl*:*yfp* and a *sonic hedgehog* reporter construct containing 2.4 kb 5' upstream sequence of the *sonic hedgehog* locus and three enhancer modules downstream of the *yfp* reporter (2.4*shh*)(Ertzer *et al*, 2007). To swap the core promoter regions of the two constructs, BamHI restriction sites were created in the promoter sequences of both constructs. The core promoters were then excised using BamHI/XhoI and swapped by cloning them into the respective background (Fig. 21). 184 bp upstream sequence of *notail*, including the initiator, was replaced by 129 bp 5' upstream sequence of *sonic hedgehog* and vice versa. The resulting reporter constructs were co-injected with either c MO or TBP MO into one cell stage embryos and analysed in transient expression assays by UV microscopy at shield stage. In c MO injected embryos, both constructs, the *notail* core promoter in *sonic hedgehog* background (*shh_ntl*:*yfp*) as well as the *sonic hedgehog* core promoter in *notail* background (*ntl_shh*:*yfp*) are able to activate transcription of the reporter in tissues resembling the wildtype expression pattern (Fig. 22A, E). *ntl_shh*:*yfp* expression is present in the margin and the shield (Fig. 22C), *shh_ntl*:*yfp* in the shield (Fig. 22G). The observation of the specificity of the expression pattern of both constructs, suggests the ability to communicate with the enhancer regions present in the respective construct. In TBP MO co-injections, the property of dependency on TBP-function of *sonic hedgehog* (Fig. 22F) has been transferred to the *notail* promoter. Signal originating from the reporter construct *ntl_shh*:*yfp* is no longer detectable in TBP Mo injected embryos (Fig. 22D). In line with this result, the *notail* core promoter enables *shh_ntl*:*yfp* to be active in TBP MO co-injected embryos (Fig. 22H). These findings are indicating that the TBP-independence of *notail* is founded in the core promoter, encoded on the DNA.

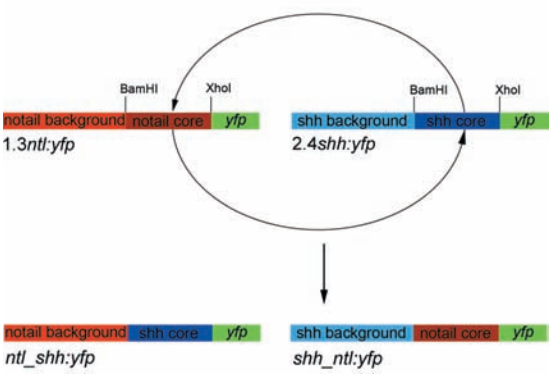

Figure 21. Schematic representation of the *notail*/*sonic hedgehog* core promoter swap.

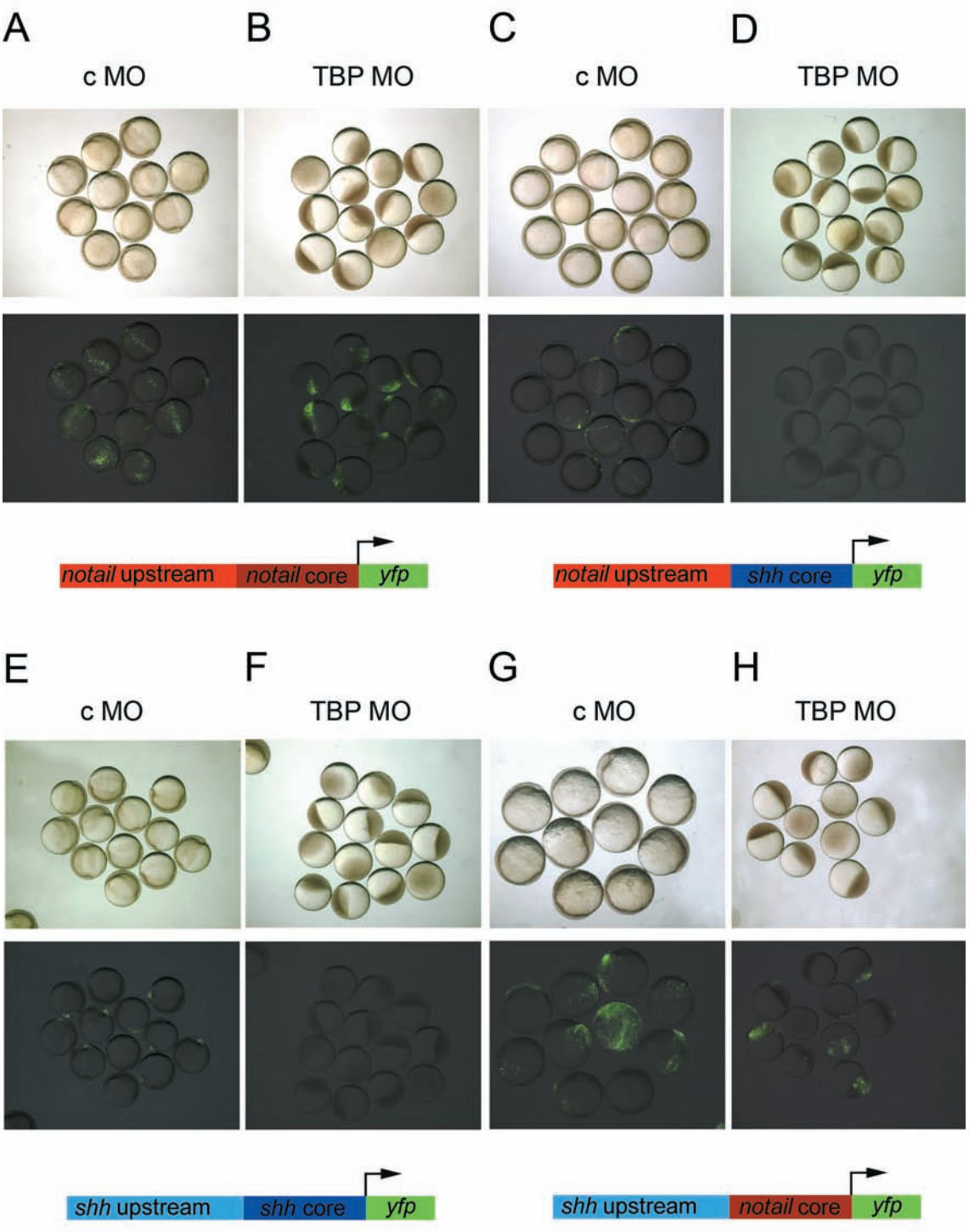

Figure 22. TBP-independence can be transferred into a different enhancer background. Co-injections of 1.3*ntl* and 2.4*shh* as well as the swapped constructs *shh_ntl:yfp* and *ntl_shh:yfp* analysed 6 hpf by UV-microscopy. (A) 1.3*ntl* co-injected with c MO. (B) 1.3*ntl* co-injected with TBP MO. (C) *ntl_shh:yfp* co-injected with c MO. (D) *ntl_shh:yfp* co-injected with TBP MO . (E) *2.4shh* co-injected with c MO. (F) *2.4shh* co-injected with TBP MO. (G) *shh_ntl:yfp* co-injected with c MO. (H) *shh_ntl:yfp* co-injected with TBP MO. Group pictures of randomly oriented embryos imaged in bright field (top) and fluorescence (bottom).

injected construct	Analysed embryos	% yfp expressing	% spec. expression
1.3ntl + c MO	184	97,2	63,1
1.3ntl + TBP MO	214	87,3	0
ntl_shh + c MO	186	75,8	47,1
ntl_shh + TBP MO	156	3,2	0
2.2shh + c MO	133	73,6	55,6
2.2shh + TBP MO	145	0	0
shh_ntl + c MO	164	90,2	46,3
shh_ntl + TBP MO	202	79,7	0

Table 8. Summary of analysed embryos in the *notail/sonic hedgehog* core promoter swap. Specific expression is regarded as equivalent to endogenous *notail* or *sonic hedgehog* expression. Abbreviations, spec, specific.

4.3.6 The TBP binding element TATA box is not sufficient to lead to TBP-dependence

Hitherto it could be shown that the TBP-independence of *notail* is attributable to the core promoter region and could be based on the interaction of DNA motifs with elements of an alternative transcription initiation complex. If this interaction also includes interaction with known core promoter elements, mutation of these elements might lead to a transcription initiation mechanism of *notail* requiring TBP-function. To test this hypothesis, point mutations in the TATA box like sequence (TTTAAA to ACGTGA) and the initiator (CCAGATT to GGCACGC) were created by a PCR based approach (see 3.1.13), turning the respective motif into an irrelevant sequence as well as mutating the TATA box like sequence into a canonical TATA box (TTTAAA to TATAAA). The resulting constructs were co-injected with either c MO or TBP MO and analysed at shield stage in transient expression assays. In c MO co-injected embryos neither the mutation of the initiator, nor the mutation of the TATA box like sequence to a random sequence or canonical TATA box led to a attenuation of the reporter signal or a change in specificity in comparison to the wildtype *notail* promoter construct (Fig. 23A-C). These results suggest that neither motif is functional in the context of the reporter construct in transient expression assays. TBP Mo co-injected embryos show no epibolic movements, but the YFP reporter signal of the above-mentioned constructs is still present (Fig. 23A-C). The presence of the reporter signal indicates that none of the mutated core promoter motifs in themselves are essential in the context of activating TBP-independent 1.3*ntl*:*yfp* expression and that the DNA motif TATA box alone is not sufficient to lead to TBP-independent initiation of transcription.

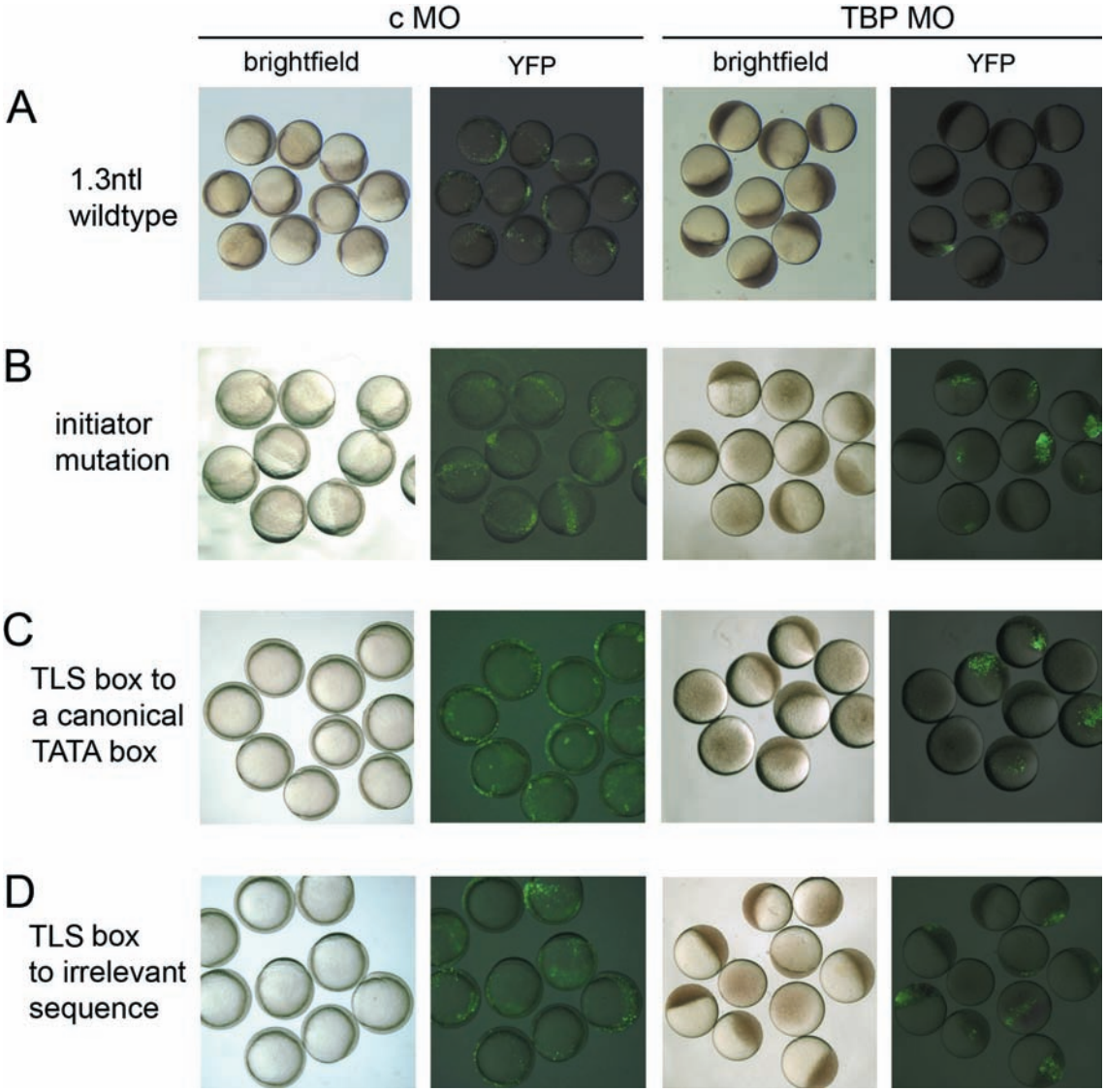

Figure 23. Mutational analysis of the core promoter elements bioinformatically determined in 1.3*ntl*. Constructs were co-injected with either c MO or TBP MO and analysed 6 hpf by UV-microscopy. (A) wildtype 1.3*ntl* promoter. (B) Mutation of the initiator to an irrelevant sequence. (C) Mutation of the TLS box to a canonical TATA box. (D) Mutation of the TLS box to an irrelevant sequence. Group pictures of randomly oriented embryos. Bright field and fluorescent YFP images as indicated in the Figure.

injected construct	Analysed embryos	% yfp expressing	% spec. expression
1.3ntl + c MO	211	92,1	64,2
1.3ntl + TBP MO	244	85,4	0
inr-mutation + c MO	316	82,8	64,8
inr-mutation + TBP MO	174	89,1	0
TLS-box to TATA box + c MO	336	95,2	62,8
TLS-box to TATA box + TBP MO	256	85,9	0
TLS box to irrelevant seq. + c MO	257	93,7	62,2
TLS box to irrelevant seq. + TBP MO	188	84,6	0

Table 9. Summary of analysed embryos in the mutational analysis of core promoter elements of 1.3*ntl*. Specific expression is regarded as equivalent to endogenous *notail* expression. Abbreviations, spec, specific.

5.0 Discussion

Summary of the results

In the most conventional model, the first step in Polymerase II dependent transcription initiation is the binding of TFII-D through its subunit TBP to the TATA box in the core promoter region (reviewed in Roeder, 1991) and TBP was argued to be essential for transcription by Pol II ((Orphanides *et al*, 1996). However, recent publications suggest that TBP is not universally required to initiate transcription (Muller *et al*, 2001; Veenstra *et al*, 2000) and that the TATA box is present in less than 10 % of eukaryotic promoters (reviewed in Sandelin *et al*, 2007). These findings raised the question of the extent and the nature of genes requiring TBP function in the activation of the zygotic genome and the characteristics of gene promoters transcribed by TBP-dependent and -independent mechanisms.

The results presented demonstrate that only a subset of genes expressed in the early zebrafish embryo are dependent on TBP function. Furthermore it could be shown that TBP is specifically required for the transcription dependent degradation of a subclass of maternally deposited mRNAs. The study presented here suggests that the maternal mRNAs, which require TBP for their degradation during and after the MBT involve a miR-430 micro RNA dependent mechanism. Therefore, TBP plays a major role in the transition from a transcriptionally inactive state to a transcriptionally active zygote. The majority of genes show no significant change in their steady state RNA levels in embryos in which TBP function is blocked. This finding could indicate alternative pre-initiation complexes to be recruited to the core promoter region of these genes. The analysis of core promoter elements of zebrafish genes that were analysed for TBP dependence indicated the binding site of TBP the TATA box is underrepresented in the class showing low variable changes in gene expression, the class holding genes that are likely transcribed by an alternative transcription initiation mechanism. In contrast he initiator element is underrepresented in the class of upregulated genes of TBP morphants. However there is no preference for a core promoter element in the class of genes downregulated in TBP morphants. To get better insight into the sequence requirements of TBP independent transcription mechanisms, the TBP independent *notail* gene promoter was analysed. The results suggest that there are specific sequences probably reflecting transcription factor binding sites that confer TBP independence in the core promoter region. The functional analysis of core promoter elements in the *notail* promoter also

demonstrates that the TATA box alone is not sufficient to recruit TBP and make the promoter TBP dependent for its activation.

5.1 TBP function is only required for a subset of genes in early zebrafish development

The use of antisense morpholino oligonucleotides to knock down TBP in conjunction with large-scale gene expression profiling techniques indicates that the expression of a large proportion of genes is not significantly affected in TBP depleted embryos. By using a second morpholino that targets a sequence more upstream in the 5'UTR of tbp and the application of rescue experiments using recombinant *xtbp* RNA that does not include the interaction site of the TBP-MOs, strongly suggest that the observed changes in expression levels on the microarray are specific to the loss of TBP-function. Based on the foldchange of RNA levels, three distinct response groups where established: genes which show reduced RNA levels by knockdown of TBP, genes showing increased RNA levels and genes that do not significantly change steady state RNA levels in TBP MO injected embryos. This result could indicate that alternative transcription initiation mechanisms, which are not affected in TBP depleted embryos, are involved in transcriptional processes of these genes. Protein complexes like TFTC and SAGA are known to activate transcriptional processes in a TBP-free manner (Huisinga & Pugh, 2004; Wieczorek *et al*, 1998). Furthermore, members of the TBP family of proteins (such as TLF /TRF2 (Rabenstein *et al*, 1999) and TBP2/TRF3 (Persengiev *et al*, 2003) could be involved in these alternative transcription initiation mechanisms. Although the DNA interaction site of TLF has so far not been identified in vertebrates (Thomas & Chiang, 2006), a TLF/DREF complex has been identified that targets a subset of promoters through binding of DREF to DRE-elements on the DNA (Hochheimer *et al*, 2002). Beside this enrichment for DRE motifs of TLF target genes, the promoter analysis of more than hundred bona-fide TLF target genes revealed that TLF binds to promoters, which are devoid of the TATA motif (Chong *et al*, 2005). TLF could be demonstrated to be indispensable for the early development of *C. elegans*, zebrafish and *Xenopus* (Dantonel *et al*, 2000)(Muller *et al*, 2001; Veenstra *et al*, 2000). Recent studies suggest that TLF-function is linked to genes that are preferentially expressed in the embryo (Jacobi *et al*, 2007). The vertebrate specific protein TBP2 is also required for early developmental processes (Bartfai *et al*, 2004; Jallow *et al*, 2004) and shares a highly homologous core domain with TBP (92% identity) and is able to bind to the TATA box motif on the DNA in the core promoter region. TBP2-function has been linked to ventral specification (Jacobi *et al*, 2007) myogenesis (Deato & Tjian, 2007) and specification of the haematopoietic linage (Hart *et al*, 2007). Taken together, these results establish that TBP fam-

ily proteins have specific, none overlapping functions in developmental processes, a finding which makes it likely that a subclass of genes showing low variable changes in RNA levels on the TBP knockdown microarray are regulated by alternative core promoter recognition complexes that contain TLF or TBP2. The question for specific target genes of TBP family proteins could be addressed by ChIP-on-chip (chromatin immunoprecipitation with microarray technology) followed by a gene ontology analysis to identify specific classes of genes regulated by TBP, TLF and TBP2.

In addition, TLF and TBP2 form complexes with TFIIA and TFIIB (Bartfai *et al*, 2004), proteins that directly interact with TBP. As both proteins seem to be able to function in similar complexes, TLF and TBP2 could also partially restore transcription of genes that, in the wild-type embryo, are transcribed by a TBP-dependent mechanism. In *Xenopus*, moderate overexpression of TBP2 has been demonstrated to partially rescue an antisense knockdown of TBP (Jallow *et al*, 2004). To address the question to what extend TBP2 or TLF could restore TBP dependent transcription mechanisms, would make it necessary to individually knock down each member of the TBP family and analyse the expression profile on a genome wide scale. In zebrafish however, TLF- and TBP2 protein are deposited maternally, making it impossible to target these factors by morpholino oligonucleotides in the early embryo. In *Xenopus* however, where none of the TBP related factors is present in the egg, the knockdown of TLF or TBP2 indicates for a requirement of each factor in specific developmental processes. TLF has been linked to genes that are preferentially expressed in the embryo, TBP2 to ventral specification (Jacobi *et al*, 2007).

The comparison of the TBP knockdown microarray with maternal datasets demonstrates that the low variable (or TBP independent) class of genes also contains genes, which are maternally deposited. This low variable increase or decrease in RNA levels in comparison to the control could have two reasons: (i) as not all maternally deposited transcripts are degraded soon after MBT (Mathavan *et al*, 2005), these maternal transcripts could be degraded later in development. (ii) The maternal component is degraded by a TBP-dependent mechanism. If transcription of the zygotic component of a transcript is also TBP dependent, the steady state of the RNA in TBP depleted embryos is similar to the control. If maternal genes are present in the class of low variable genes, what is the exact proportion of genes transcribed by alternative promoter recognition complexes in the dome stage embryo? A comparison of the TBP knockdown microarray to genes downregulated in alpha-amanitin injected embryos could elucidate if genes transcribed by Pol II also require TBP. However, this comparison would only concern a subclass of transcripts, which are not deposited maternally

or the maternal component is already degraded (Genes with maternal component degraded by a transcription dependent mechanism or degraded later in development are present in upregulated- and low variable set of the alpha-amanitin microarray). To ascertain if genes present in the class of genes showing low variable changes in RNA levels are actually transcribed by alternative transcription initiation mechanisms and its proportion compared to late degrading maternally deposited transcripts, would make it necessary to carry out biochemical studies that target the promoter occupancy of alternative transcription initiation complexes as well as the transcriptional activity of genes present in this class.

The class of genes upregulated by TBP knockdown contains a high number of maternally deposited transcripts. This observation suggested that TBP is specifically required for the transcription dependent degradation of a subclass of maternally deposited mRNAs. Furthermore the involvement of the miR-430 pathway in the degradation of these transcripts was suggested by the results presented here. Upregulation of genes in TBP morphants could however also occur if TBP is a direct or indirect repressor. The analysis of promoter constructs to directly assess the transcriptional activity of a reporter gene, thus excluding a possible interference with maternal transcripts, demonstrates that 4 out of the 12 analysed promoters show significant increase of the reporter signal in TBP morphants compared to the control (carried out by Jochen Gehrig, published in Ferg *et al*, 2007). This result suggests that TBP can act as a repressor. To check if this is a direct or indirect effect needs to be addressed by chromatin immunoprecipitation to show if TBP is bound in the promoter region of these gene promoters and a detailed promoter occupancy analysis before and after activation of these promoters. Furthermore, genes, which are transcribed by a TLF- or TBP2 dependent mechanism, could be normally repressed by TBP. It was described that TLF can indirectly inhibit transcription from TATA-containing promoters by sequestering essential factors (Moore *et al*, 1999; Teichmann *et al*, 1999) and TBP overexpression has been shown to inhibit transcription of TLF target genes (Chong *et al*, 2005). It is therefore likely that TBP and TLF regulate respective promoters in a reciprocal manner. A recent publication suggests that reduced levels of TBP leads to increased levels of TLF- and TBP2 mRNA levels (Bush *et al*, 2008), which in return could lead to an activation of their target genes. In comparison to the c MO injected control, these genes would most likely be found in the upregulated class of genes.

The class of genes downregulated by TBP knockdown are most likely genes, which require TBP for transcriptional activity. Although the analysed stage in development, shortly after the activation of the zygotic genome, minimises the chance of secondary effects, it can-

not be excluded that components of the transcriptional regulation of a gene indirectly require TBP function, and result in the downregulation of its activity in TBP morphants.

The comparison of the TBP knockdown microarray with transcripts showing the highest accumulation in the unfertilised egg (Mathavan *et al*, 2005) and genes upregulated on the alpha-amanitin could demonstrate that maternal genes are present in the upregulated- and downregulated class of genes.

As TBP is required for only a subset of genes in the early development of zebrafish, it raises the question for the functional specification of genes downregulated in TBP depleted embryos. Do genes that are significantly regulated in TBP morphants belong to a specific ontology class? The analysis of ontology groups in the TBP knockdown dataset did not lead to clear results (data not shown). In comparison to other vertebrate model organisms, the functional annotation of genes by Gene ontology (GO) in zebrafish is still underdeveloped, a circumstance that could explain the lack of statistically analysable data in GO analysis. However, the recent analysis of gene expression profiles in various stages during zebrafish ontogeny (Konantz M, Otto G-W, Weller C, Saric M. Geisler R. Microarray analysis of gene expression in zebrafish development, manuscript in preparation) rendered it possible to address if genes significantly regulated in TBP morphants are specifically expressed in a certain stage of development or constitutively active during development (analysed by Remo Sanges, published in Ferg *et al*, 2007). The Meta-analysis could show that genes downregulated by the knockdown of TBP tend to be stage-specific, whereas the upregulated class of genes tends to be expressed throughout development. The class of genes showing low variable changes in steady state RNA levels is composed of both, stage specific- and constitutively active genes. These findings indicate that genes requiring TBP-function have a specific role during development and suggests for stringent controlled expression, which has been associated with single peak promoters and a high frequency of TATA boxes (Carninci *et al*, 2006). Genes that are upregulated in TBP morphants were shown to be more likely constitutively expressed genes. Interestingly it was shown previously in mammals that constitutively expressed genes tend to be associated with broad peak promoters and CpG islands but lack of TATA Box (Carninci *et al*, 2006). In contrast tissue specific genes (expected to have dynamic activity in different stages of ontogeny were shown to be enriched for TATA boxes, It will be interesting to address the promoter features of zebrafish promoters to be able to associate TBP function with promoter characteristics.

5.2 Trancriptional startsite selection in TBP morphants

To address the question, if the same promoter is used when TBP is lost, or if alternative promoters are utilised in TBO morphants, the transcriptional startsite utilisation was analysed by 5' RACE. The analysis of TSS utilisation of zygotic genes in TBP MO injected embryos could show that in the majority of analysed genes, initiation of transcription preferentially starts in the same nucleotide position as compared to the c MO injected control. However, the analysis also demonstrates that in case of certain genes in TBP depleted embryos transcriptional start sites are utilised that are not detectable in c MO injected embryos and vice versa. In one case, transcription of the gene *tppp3* preferentially utilises a startsite that is not detectable in c Mo injected embryos. The analysed genes emanate from a zygotic activity shortly after MBT and do not have a maternal component. This demonstrates that transcription initiation in TBP depleted embryos takes place and suggests for the nucleation of an alternative pre-initiation complex in the class of genes showing low variable changes in RNA levels on the TBP knockdown microarray. It, however, also raises the question, if the nucleated pre-initiation complex in TBP morphants is identical to the pre-initiation complex in wild type embryos. Recent studies suggest that both TLF and TBP2 are overexpressed in TBP heterozygous cells (Bush *et al*, 2008) and can compensate to some extend for the loss of TBP-function (Bush *et al*, 2008; Jacobi *et al*, 2007; Jallow *et al*, 2004). Furthermore, the core-domain of TBP and TBP2 are nearly identical (Bartfai *et al*, 2004), suggesting, that the sequence requirements to interact with the DNA in the core promoter region could be similar and could replace TBP in pre-initiation complexes nucleating a similar transcriptional startsite.

The utilisation of an independent start site of transcription in the 5' UTR of *tppp3* in TBP MO injected embryos that is not detectable in c Mo injected embryos, suggests an alternative pre-initiation complex which form on the promoter of of *tppp3* and could indicate that different pre-initiation complexes can initiate transcription from the same gene. Though specific roles for TBP2 and TLF in early development have been described (Jacobi *et al*, 2007)(Gazdag *et al*, 2007), this finding raises the question if multiple pre-initiation mechanisms are involved in the regulation of a developmental gene. Alterations of intracellular TBP protein levels can affect the transcription of specific promoters (Bendjennat & Weil, 2008; Bush *et al*, 2008). Therefore, dependent on cell-type or stage of development, initiation of transcription of a gene could be controlled by different pre-initiation complexes. To elucidate if multiple pre-initiation complexes are involved in the regulation of a gene, would

make it necessary to analyse the promoter occupancy by individual components of pre-initiation complexes during various stages in development and/or cell-types.

5.3 TBP has a specific function in the pathway of maternal transcript degradation

The comparison of the TBP knockdown microarray with datasets containing transcripts that are maternally deposited indicated an enrichment of these mRNA species in the class of genes upregulated by the knock down of TBP and genes showing no significant change in RNA levels. This finding suggests for a requirement of TBP function in the program of maternal mRNA degradation. The transcription dependent degradation of maternal mRNA was reported to be linked to the microRNA miR-430 (Giraldez *et al*, 2006). The authors could show that binding of miR-430 to its target sequence in the 3' UTR of transcripts leads to deadenylation and subsequent degradation. The analysis of TBP function in this process indicates that TBP is specifically required in a subset of maternal mRNAs, which are degraded by a miR-430 dependent mechanism in post MBT stages, before gastrulation starts.

The primary microRNA transcripts were shown to be transcribed by polymerase II (Lee *et al*, 2004), a finding which could suggest that the pre-initiation complex formed at the microRNA promoter requires TBP to initiate transcription. However, the mature form of miR-430 can still be detected by northern blot under TBP knockdown conditions (carried out by Tina Rathjen in the laboratory of Dr. Tamas Dalmay, Norwich England, published in Ferg *et al*, 2007), suggesting that TBP is neither required for the transcription of the primary microRNA, nor for the maturation process of miRNA. These findings make it likely that TBP-function is required downstream of the maturation of miR-430. The exact mechanisms how the binding of miR-430 to its targets leads to deadenylation are not well understood yet. However, as TBP depletion leads to the perturbed degradation of only a subgroup of miR-430 target mRNAs, could suggest that not the general degradation machinery itself is affected, but an activity which is part of the recognition process in the 3'UTR of these targets.

miR-430 is one of the first microRNAs expressed in the developing embryo and accumulates after the maternal-to-zygotic transition. Initially, miR-430 was identified to regulate brain morphogenesis in zebrafish (Giraldez *et al*, 2005). Apparently the microRNA has two distinct functions, which raises the question if TBP would be required in this process as well (however this cannot be studied without conditional knock down of TBP). MicroRNAs are thought to have the potential to regulate many differentiation processes and have been shown to be involved in neurogenesis (Giraldez *et al*, 2005; Krichevsky *et al*, 2003), cardiogenesis (Zhao *et al*, 2007) and limb development (Harfe *et al*, 2005). It would be interesting

to address if the requirement for TBP-function in mRNA degradation through microRNAs is limited to the clearance of maternal mRNAs or if TBP is also needed in microRNA mechanisms involved in regulating developmental processes by post-transcriptional regulation of gene expression.

Another important question is the generality of the TBP requirement in the miR-430 dependent degradation mechanism. In *Drosophila*, two pathways for maternal RNA degradation have been identified. The first pathway is driven by maternally encoded factors, the second, zygotic pathway, becomes active 2 hours after fertilization (Bashirullah *et al*, 1999). Again using *Drosophila* as a model organism, a recent study could demonstrate that zygotically expressed microRNAs, like in zebrafish, target maternal mRNAs for turnover and are part of the zygotic degradation pathway. However, the microRNAs identified in this study are unrelated to miR-430, suggesting for a convergent evolution of the pathways (Bushati *et al*, 2008). The study presented here could provide the basis to analyse if TBP-function is required for the regulated degradation of maternal transcripts by this evolutionary conserved pathway in other vertebrates.

The comparison of the TBP knockdown microarray dataset with miR-430 targets (Giraldez *et al*, 2006) implies that not all maternally deposited RNAs showing increased steady state RNA levels in TBP depleted embryos are degraded by a miR-430 dependent mechanisms. On the other hand, the degradation dynamics of miR-430 dependent RNA degradation and TBP dependent RNA degradation are very similar. These findings could suggest for a mRNA degradation mechanism acting parallel to the miR-430 pathway. The mechanisms leading to the regulated degradation of maternally deposited transcripts in the developing embryo are largely unknown. The indication of a second pathway that, like miR-430, acts zygotically, could prove helpful for studies engaged in the elucidation of maternal mRNA degradation mechanisms

5.4 The Ensembl promoterome is characterised by a low frequency of core promoter elements

The characterisation of promoters in respect of DNA elements occurring in the core promoter region relies on the available information of the 5' end of transcripts, which marks the TSS. As too few verified 5' ends are available in zebrafish, there was a need for a comparison of bioinformatically predicted- and experimentally verified 5' ends to see if the Ensembl data could be sued. The variance analysis comparing a dataset of experimentally verified 5' ends of genes (Suzuki *et al*, 2001) to a dataset based on predicted TSS information in

the Ensembl promoterome demonstrates the feasibility of characterising promoters in a large scale, as the 5' ends of 65% of the sequences overlapping both datasets are located within a ±20 bp window. The assessment of TSS information using cage tag libraries in mouse and human (Carninci *et al*, 2006) demonstrated that in only a fraction of genes transcription starts from a single start site. The majority of genes utilise multiple startsites or multiple promoters. The approach used here for identification of promoter regions however utilised the sequence information of only one startsite per gene. Together with the above-mentioned limitations of predicted TSS information, the analysis of the Ensembl promoterome could only present a qualitative approach to study the frequency of core promoter elements in zebrafish.

To elucidate if the well-described core promoter elements initiator, TATA box and DPE form cluster in the core promoter region, a finding which would indicate for an increase of occurrences at a certain position, a clustering analysis of the above mentioned core promoter elements was carried out. This analysis demonstrates that the initiator and TATA box, but not the DPE form clear peaks at a distinct position and suggests for a local overrepresentation of the two sequence motifs in the core promoter region. In previous studies, neither the initiator, nor the DPE showed an accumulation of occurrences at a certain position and was reasoned to be caused by high degree of degeneracy the DNA-motifs hold (FitzGerald *et al*, 2004). The initiator consensus sequence contains 128 sub sequences, the DPE 24 possible combinations. In conjunction with the nucleotide frequency in the core promoter region, this degeneracy makes it unlikely that these motifs show an enrichment of occurrences at a certain location. A possible explanation why the initiator sequence clusters nonetheless, though it is far more degenerate than the DPE, is the preference for sub sequences that contain the central dinucleotide motif CA. The database of eukaryotic transcription factor binding profiles Japsar (Sandelin *et al*, 2004) assumes that this dinucleotide is the most prominent feature of the initiator consensus sequence and was associated with more active transcriptional startsites (Carninci *et al*, 2006). Another possible explanation for the inability to detect clustering of the DPE beside the degeneracy of the motif is the application of consensus sequences in this study. Consensus sequence based approaches use regular expressions to identify strings- or patterns of characters on DNA sequences and represent only the most abundant nucleotide of a motif. The application of positional weight matrices could lead to the detection of the DPE to cluster in the core promoter region, as positional weight matrices provide a quantitative measure of the likelihood of each nucleotide being present

at each position in the motif. However, such a weight matrix has not been identified for the vertebrate DPE.

To determine the frequency of core promoter elements in the Ensembl PPR dataset at the consensus position, a position-restricted search for the core promoter elements initiator (human and *Drosophila* consensus sequence), TATA box, DPE and CpG islands was carried out and resulted in the identification of a low frequency of the TATA box and the initiator. The TATA box was found to be present in 4.1%, the initiator of the human consensus sequence in 10.3% and the initiator of the *Drosophila* consensus sequence in 4.0% of analysed promoters in the Ensembl PPR dataset. Recently published studies in human using CAGE data from DBTSS (Wakaguri *et al*, 2007) demonstrated that the TATA box is present in ~10%, the initiator in ~48% of analysed promoters. In comparison to these studies, both motifs, initiator and TATA box, are less abundant in the Ensembl zebrafish PPR dataset than in human. Are there fundamental differences in the core promoter architecture comparing human to zebrafish? The analysis of zebrafish promoters with an experimentally verified transcriptional startsite (Wakaguri *et al*, 2007), suggests very similar proportion of promoters to contain a TATA box (4.1%) or an initiator (11.7%) (data not shown) as the analysis of the Ensembl promoterome. However, DBTSS holds experimentally gained TSS information of 3417 zebrafish genes supported by an average of 5 tags per gene. Hence, DBTSS and Ensembl lay the foundation to bioinformatically characterise core promoters in zebrafish, but do not contain sufficiently large number of TSSs to directly compare to the CAGE tag analysis carried out in mouse and human (Carninci *et al*, 2006). The CAGE tag analysis gave new insights into the promoter composition on a genome wide scale and allowed to identify the compositional properties of the core promoter region at a very high resolution (Bajic *et al*, 2006) as well as large-scale cross-species comparisons (Bajic *et al*, 2006; Jin *et al*, 2006). These comparative genomic studies on the promoter level could demonstrate that a high number of core promoter elements in orthologous human and mouse genes are conserved and improves the signal-to-noise ratio. Similar approaches using a more distantly related organism like zebrafish, could further improve this ratio and would allow to address the biological relevance of the in silico identified promoter composition in functional assays using large scale screening techniques. However, these studies would have to be aware of the fundamentally differences in the nucleotide frequencies in different species. The analysis of the nucleotide frequency of the Ensembl PPR dataset demonstrates an increase of the GC-content around the putative transcriptional startsite in a specific shape. Similar approaches in human, mouse and Fugu demonstrate that this shape in the variation of the GC content is organism spe-

cific and was found to correlate with gene expression (Aerts *et al*, 2004). Variations in the base composition around the transcriptional startsite between distinct classes could lead to a change in the probability of a nucleotide to occur. As a consequence, this also leads to a change in the probability of a core promoter element to occur. Therefore, the base composition in the core promoter region has to be taken into account in comparative genomic studies on the promoter level.

5.5 Requirement for TBP function in the context of promoter elements

The bioinformatic analysis of core promoter elements present on promoters of significantly regulated genes on the TBP knockdown microarray indicates an under representation of the TATA box in the low variable class of genes and an under representation of the initiator in the class of upregulated genes. The class of downregulated genes displays a distribution of core promoter elements similar to the Ensembl promoterome. Both motifs, TATA box and initiator, have been suggested to be indicative for single peak promoters (see 1.3.1) and linked to tissue specific expression. Both motifs are under represented in broad peak promoters (see 1.3.1), associated with ubiquitous expression and a high frequency of CpG islands (Carninci *et al*, 2006; Martinez *et al*, 1994; Schug *et al*, 2005; Suzuki *et al*, 2001). The meta-analysis demonstrated that genes upregulated in TBP morphants are linked to constitutive activity, whereas genes downregulated by TBP knockdown tend to be stage specific (see 4.1).

These findings could suggest, that genes upregulated by TBP knockdown contain less frequently an initiator because this class contains constitutively active genes with broad peak promoters. The class of low variable genes less frequently contain a TATA box, because a proportion of these genes are transcribed by alternative mechanisms. However, it has to be emphasised that the number of analysed promoters and the number of incidences of core promoter elements is very low. Cross species comparisons of the core promoter region demonstrated similar motif composition in orthologous genes (Bajic *et al*, 2006; Jin *et al*, 2006). As the annotation of the human and mouse genome is nearing completion, the analysis of orthologous genes present in the TBP knockdown microarray could increase the number of analysable promoters.

The functional analysis of the *notail* promoter elements by mutational analysis of the bioinformatically determined core promoter elements (discussed in 4.7) could demonstrate that neither the initiator, nor the TLS box is required for the specificity of the reporter construct. Furthermore, mutation of the TLS box to a consensus TATA box demonstrated that the

TATA box alone is insufficient to lead to TBP-dependent transcription initiation mechanisms. Accordingly, these results suggest that the initiator does not determine an alternative pre-initiation complex and the mere appearance of a TATA box is not sufficient for TBP dependent initiation of transcription.

Furthermore, the sequence requirements for TBP-dependent and –independent transcription initiation mechanisms can only be addressed on genes, which are actively transcribed at the analysed stage of development. As discussed in previous sections, the datasets representing the three classes of RNA changes on the TBP knockdown microarray are very heterogeneous: the upregulated class contains a large proportion of maternal transcripts. Although maternally deposited genes have been removed, because a requirement for TBP-function in regulating transcription of maternal genes cannot be addressed in the experimental system presented here, the removal might not have been efficient enough to lead to a dataset containing only zygotically active genes. The "contamination" of maternal transcripts also likely in the class of genes showing no significant changes in RNA levels. This dataset is also likely to contain genes, which are transcribed by alternative transcription initiation mechanisms, and each mechanism could have its own sequence requirements. The class of genes downregulated in TBP morphants is also likely to contain genes, which are downregulated through secondary effects. Accordingly, the heterogeneity of the TBP knockdown microarray dataset and the low frequency of occurrences of core promoter motifs does not allow to draw a final conclusion on the sequence requirements of TBP-dependent and -independent transcription initiation mechanism and emphasises the need for genome wide promoter occupancy studies of individual components of pre-initiation complexes by ChIP-on-chip followed by a bioinformatically aided sequence analysis to address the question of the sequence requirements of TBP-dependent and -independent transcription initiation mechanisms.

5.6 1.3 kb *notail* promoter region (1.3*ntl*) contains DNA elements conferring specific expression in the notochord

Microinjection of a reporter construct containing 1.3kb 5' flanking region of *notail* is capable to confer an expression pattern of the reporter gene that recapitulates the endogenous expression pattern of *notail* in early stages of zebrafish development. During gastrulation, the reporter gene is expressed in the margin where the blastodisc meets the yolk and in the shield. In segmentation stages, 1.3*ntl* is expressed in the notochord and in the budding tail. A deletion series utilising endogenous restriction enzyme recognition sites in the

notail promoter, could aid in locating a regulatory sequence block in a region -600bp to -1300bp upstream of the TSS with enhancer effect in the margin as well as in the shield and differentiating notochord. Presumably, this region contains transcription factor binding sites that mediate *notail* specific expression. In contrast to this finding, the analysis of *brachyury* expression in mouse by transgenes comprising up to 8.3 kb 5' sequence and 5 kb 3' sequence did not lead to expression in axial mesoderm (Clements *et al*, 1996). Similar studies in *Xenopus* to elucidate the regulatory mechanisms of the *brachyury* homolog *xbra* led to the same conclusion (Latinkic *et al*, 1997). It was suggested that activation of *brachyury* expression has different requirements in different populations of mesodermal cells. The signal responsible for activation of *brachyury* expression in the notochord might be different from the signal requirements in mesoderm formation (Clements *et al*, 1996). Accordingly, studies in zebrafish revealed that in maternal zygotic one eyed pinhead mutants (MZoep), *notail* is still expressed in the ventral and lateral margin during gastrulation and in the tailbud, but not in the dorsal margin or, in later stages, in the notochord (Chen & Schier, 2001). This suggests for two distinct expression domains of *notail*: a nodal signalling independent expression domain in mesoderm formation and a nodal dependent expression domain in axial mesoderm specification Taken together, these finding suggest that 1.3*ntl* includes transcription factor binding sites required for the interaction with Smad signalling molecules, which are not present in the proximal promoter of *brachyury* and *xbra*. Has the regulatory sequence block that is present in zebrafish been separated during mouse- and *Xenopus* evolution, physically dividing the two expression domains? This question could be addressed by a comparative analysis of genomic sequences using the zebrafish shield enhancer as a reference.

5.7 Mutational analysis of core promoter elements in 1.3*ntl*

The initial observation that *notail* is still expressed in TBP depleted embryos (Muller *et al*, 2001) raised the question for the sequence requirements of TBP-independent transcription initiation mechanism. To address the question if these requirements are based on DNA motifs present in the *notail* core promoter, the bioinformatically determined initiator and TLS box were mutated to random sequences and the TSL box to a canonical TATA box. The functional assays could demonstrate that neither the mutation of the initiator, nor the mutation of the TSL box leads to an attenuation of the reporter signal or a loss in specificity in the context of the reporter construct. Does these findings suggest that none of the bioinformatically predicted core promoter elements are functional? It was suggested that

the initiator and the TATA box are the only known core promoter elements that, alone, are able to recruit the pre-initiation complex and initiate transcription (Sandelin *et al*, 2007). The TLS box shows one mismatch compared to the consensus sequence (TTTAAA instead of (TATAA[TA]A (Bucher, 1990)). However the positional weight matrix established by Bucher shows that in 8.9% of analysed promoters, the second nucleotide A is replaced by the nucleotide T and resembles the TLS box. Therefore, the TSL box could represent a "weak" TATA box that, as the initiator, is in itself strong enough to direct the pre-initiation complex to the startsite of transcription in constructs were only one motif is mutated. To address the question if the TLS box or the initiator is sufficient to recruit the pre-initiation complex would make in necessary to mutate both motifs in the same construct and analyse the signal strength of the YFP reporter compared to 1.3*ntl*.

By mutating the TLS box to a consensus TATA box, in combination with the initiator, a core promoter was created that resembles the classical promoter TFIID containing pre-initiation complexes bind to and raises the question if in c MO co-injected embryos, the pre-initiation complex bound to the promoter of this construct is identical to the pre-initiation complex bound to the wildtype promoter construct. As this construct is able to activate transcription of the reporter in TBP depleted embryos, suggests that this promoter construct includes the sequences requirements for different pre-initiation complexes to be formed in the core promoter region. Studies in *Drosophila* could demonstrate that a subclass of gene promoters are regulated through the utilisation of tandem core promoters composed of two distinct transcriptional startsites, which are regulated either by TLF or TBP (Hochheimer *et al*, 2002; Isogai *et al*, 2007). Future studies will have to address the promoter occupancy by individual components of pre-initiation complexes of this promoter construct and the transcriptional startsite utilisation that would indicate for a tandem core promoter. If only one TSS is utilised, It would be very interesting to address the regulatory mechanism that lead to differential recruitment of pre-initiation complexes to a core promoter and if similar promoter structures are present in vertebrate genomes.

5.8 Is TPB-function required for promoter/enhancer interaction?

Co-injection of 1.3*ntl* and TBP MO could demonstrate that the reporter construct is, like the endogenous gene, still active under TBP knockdown conditions. However, expression of the reporter gene in TBP depleted embryos is not restricted to the margin, but displays a random distribution of YFP expressing cells in the blastodisc. The replacement of the ntl core promoter by the *notail* core promoter region (*shh_ntl:yfp*) leads to the same conclusion. Co-

injection with c MO leads to specific expression in the shield and recapitulates the endogenous expression pattern of ntl. The specific expression of the chimeric promoter demonstrates the ability of *sonic hedgehog* specific enhancers to communicate with the pre-initiation complex formed on the *notail* core promoter, whereas in TBP MO co-injected embryos, this specificity is lost. This loss of specificity of 1.3*ntl* and *shh_ntl*:*yfp* in TBP morphants raises several questions: (i) is the loss of specificity caused by a miss migration of cells? Although it could be demonstrated that only the minority of genes require TBP function in early zebrafish development, the general transcription factor still plays a key role in transcription initiation mechanisms of many genes. The phenotype of TBP-depleted embryos is complex and is characterised by a block of epiboly. This complex phenotype could also include miss migration of cells, which in the wildtype embryo are located at the margin, but in TBP morphants are scattered all over the blastodisc. However, this is unlikely as the endogenous *notail* is still expressed in a ring in TBP MO injected embryos (Muller *et al*, 2001), suggesting that this loss of specificity is a property of the promoter constructs, not of the endogenous gene. Accordingly, this suggests that the interaction of the enhancer with the core promoter is perturbed in the transgene situation, possibly due to additional enhancers complementing or the PIC may form correctly due to alternative mechanism activated by the genomic context. (ii) Is TBP required for the correct temporal and spatial expression of *notail* through mediating the interaction of cis-regulatory elements with the pre-initiation complex? This mediation could be indirect or indirect. TBP-dependent transcription initiation mechanisms of a transcription factor or member of the mediator complex would point to an indirect requirement for TBP-function in controlling *notail* expression. However, TBP could also be required to directly establish the bridge between activators binding to motifs in the proximal promoter and the pre-initiation complex assembled in the core promoter of *notail*. This possibility again raises the question, if the nucleated pre-initiation complex in TBP morphants is identical to the pre-initiation complex in wildtype embryos (see 4.2). The analysis of enhancer/promoter specificity, the ability of enhancers to activate transcription from different core promoter classes could demonstrate that enhancer/promoter interaction depends on the promoter architecture (Butler & Kadonaga, 2001). This finding could suggest that the pre-initiation complex formed in the core promoter region is involved in the mediation of promoter/enhancer interaction. Accordingly, if the nucleated pre-initiation complex in TBP morphants is not identical to the pre-initiation complex in wildtype embryos, specificity of the reporter construct could be lost. To elucidate if TBP function is required for promoter/enhancer interaction or if an alternative pre-initiation complex is utilised in TBP de-

pleted embryos that is able to restore expression, but not to interact with the cognate enhancer, will have to be addressed by biochemical assays that allow to analyse the nature of the pre-initiation complex bound to the core promoter in c MO and TBP MO injected embryos. Furthermore, assays would have to be established that are able to address the question if TBP interacts with the enhancer/promoter complex in genes regulated by alternative transcription initiation mechanisms.

6.0 Supplement

All supplementary information of the work presented here, is stored on the supplemental CD, which is attached to the last page of the dissertation and has the following content:

6.1 Supplementary tables

Supplementary table I
Significantly regulated genes on the TBP knockdown microarray

Supplementary table II
Primers used in this study

Supplementary table III
Intersection of the TBP microarray with transcripts showing highest abundance in the unfertilised egg (Mathavan et all 2005)

Supplementary table IV
Significantly regulated genes on the alpha-amanitin microarray

Supplementary table V
Intersection of the TBP microarray with upregulated genes in alpha Amanitin injected zebrafish embryos

Supplementary table VI
Intersection of the TBP microarray with miR-430 targets (Giraldez et. al 2005)

6.2 Supplementary files

Supplementary file I
Ensembl PPR dataset

Supplementary file II
Dataset of bootstrapped control sequences

Supplementary file III
Gene promoters downregulated on the TBP knockdown microarray

Supplementary file IV
Gene promoters showing low variable changes in gene expression on the TBP knockdown microarray

6.3 perl scripts used in this study

determine_nucfreq
findPattern_restricted
findPattern_unrestricted
promoter_extraction
promoter_shuffle
TSS_validation
UniGene_parser

7.0 References

Aerts S, Thijs G, Dabrowski M, Moreau Y, De Moor B (2004) Comprehensive analysis of the base composition around the transcription start site in Metazoa. *BMC genomics* 5: 34

Alizadeh Z, Kageyama S, Aoki F (2005) Degradation of maternal mRNA in mouse embryos: selective degradation of specific mRNAs after fertilization. *Molecular reproduction and development* 72: 281-290

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of molecular biology* 215: 403-410

Amsterdam A (2006) Insertional mutagenesis in zebrafish: genes for development, genes for disease. *Briefings in functional genomics & proteomics* 5: 19-23

Amsterdam A, Becker TS (2005) Transgenes as screening tools to probe and manipulate the zebrafish genome. *Dev Dyn* 234: 255-268

Amsterdam A, Burgess S, Golling G, Chen W, Sun Z, Townsend K, Farrington S, Haldi M, Hopkins N (1999) A large-scale insertional mutagenesis screen in zebrafish. *Genes & development* 13: 2713-2724

Antequera F, Bird A (1993) Number of CpG islands and genes in human and mouse. *Proceedings of the National Academy of Sciences of the United States of America* 90: 11995-11999

Arnold SJ, Stappert J, Bauer A, Kispert A, Herrmann BG, Kemler R (2000) Brachyury is a target gene of the Wnt/beta-catenin signaling pathway. *Mechanisms of development* 91: 249-258

Bajic VB, Seah SH, Chong A, Zhang G, Koh JL, Brusic V (2002) Dragon Promoter Finder: recognition of vertebrate RNA polymerase II promoters. *Bioinformatics (Oxford, England)* 18: 198-199

Bajic VB, Tan SL, Christoffels A, Schonbach C, Lipovich L, Yang L, Hofmann O, Kruger A, Hide W, Kai C, Kawai J, Hume DA, Carninci P, Hayashizaki Y (2006) Mice and men: their promoter properties. *PLoS genetics* 2: e54

Bajic VB, Tan SL, Suzuki Y, Sugano S (2004) Promoter prediction analysis on the whole human genome. *Nature biotechnology* 22: 1467-1473

Bally-Cuif L, Schatz WJ, Ho RK (1998) Characterization of the zebrafish Orb/CPEB-related RNA binding protein and localization of maternal components in the zebrafish oocyte. *Mechanisms of development* 77: 31-47

Bartfai R, Balduf C, Hilton T, Rathmann Y, Hadzhiev Y, Tora L, Orban L, Muller F (2004) TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish. *Curr Biol* 14: 593-598

Basehoar AD, Zanton SJ, Pugh BF (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell* 116: 699-709

Bashirullah A, Halsell SR, Cooperstock RL, Kloc M, Karaiskakis A, Fisher WW, Fu W, Hamilton JK, Etkin LD, Lipshitz HD (1999) Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in Drosophila melanogaster. *The EMBO journal* 18: 2610-2620

Bauer M, Katzenberger JD, Hamm AC, Bonaus M, Zinke I, Jaekel J, Pankratz MJ (2006) Purine and folate metabolism as a potential target of sex-specific nutrient allocation in Drosophila and its implication for lifespan-reproduction tradeoff. *Physiological genomics* 25: 393-404

Baumann M, Sander K (1984) Bipartite axiation follows incomplete epiboly in zebrafish embryos treated with chemical teratogens. *The Journal of experimental zoology* 230: 363-376

Bell B, Scheer E, Tora L (2001) Identification of hTAF(II)80 delta links apoptotic signaling pathways to transcription factor TFIID function. *Molecular cell* 8: 591-600

Bell B, Tora L (1999) Regulation of gene expression by multiple forms of TFIID and other novel TAFII-containing complexes. *Experimental cell research* 246: 11-19

Bendjennat M, Weil PA (2008) The transcriptional repressor activator protein Rap1p is a direct regulator of TATA-binding protein. *The Journal of biological chemistry* 283: 8699-8710

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57: 289-300

Bird AP (1986) CpG-rich islands and the function of DNA methylation. *Nature* 321: 209-213

Brand M, Leurent C, Mallouh V, Tora L, Schultz P (1999) Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC. *Science (New York, NY* 286: 2151-2153

Breathnach R, Chambon P (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annual review of biochemistry* 50: 349-383

Brou C, Chaudhary S, Davidson I, Lutz Y, Wu J, Egly JM, Tora L, Chambon P (1993) Distinct TFIID complexes mediate the effect of different transcriptional activators. *The EMBO journal* 12: 489-499

Bucher P (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *Journal of molecular biology* 212: 563-578

Burke TW, Kadonaga JT (1996) Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes & development* 10: 711-724

Burke TW, Kadonaga JT (1997) The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes & development* 11: 3020-3031

Burley SK, Roeder RG (1996) Biochemistry and structural biology of transcription factor IID (TFIID). *Annual review of biochemistry* 65: 769-799

Bush SD, Richard P, Manley JL (2008) Variations in intracellular levels of TATA binding protein can affect specific genes by different mechanisms. *Molecular and cellular biology* 28: 83-92

Bushati N, Stark A, Brennecke J, Cohen SM (2008) Temporal Reciprocity of miRNAs and Their Targets during the Maternal-to-Zygotic Transition in Drosophila. *Curr Biol* 18: 501-506

Butler JE, Kadonaga JT (2001) Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes & development* 15: 2515-2519

Butler JE, Kadonaga JT (2002) The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes & development* 16: 2583-2592

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond SM, Wells CA, Orlando V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA, Hayashizaki Y (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nature genetics* 38: 626-635

Caron C, Mengus G, Dubrowskaya V, Roisin A, Davidson I, Jalinot P (1997) Human TAF(II)28 interacts with the human T cell leukemia virus type I Tax transactivator and promotes its transcriptional activity. *Proceedings of the National Academy of Sciences of the United States of America* 94: 3662-3667

Casey ES, O'Reilly MA, Conlon FL, Smith JC (1998) The T-box transcription factor Brachyury regulates expression of eFGF through binding to a non-palindromic response element. *Development (Cambridge, England)* 125: 3887-3894

Casey ES, Tada M, Fairclough L, Wylie CC, Heasman J, Smith JC (1999) Bix4 is activated directly by VegT and mediates endoderm formation in Xenopus development. *Development (Cambridge, England)* 126: 4193-4200

Cavallini B, Faus I, Matthes H, Chipoulet JM, Winsor B, Egly JM, Chambon P (1989) Cloning of the gene encoding the yeast protein BTF1Y, which can substitute for the human TATA box-binding factor. *Proceedings of the National Academy of Sciences of the United States of America* 86: 9803-9807

Chalkley GE, Verrijzer CP (1999) DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *The EMBO journal* 18: 4835-4845

Chen Y, Schier AF (2001) The zebrafish Nodal signal Squint functions as a morphogen. *Nature* 411: 607-610

Chiang CM, Roeder RG (1995) Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. *Science (New York, NY* 267: 531-536

Chong JA, Moran MM, Teichmann M, Kaczmarek JS, Roeder R, Clapham DE (2005) TATA-binding protein (TBP)-like factor (TLF) is a functional regulator of transcription: reciprocal regu-

lation of the neurofibromatosis type 1 and c-fos genes by TLF/TRF2 and TBP. *Molecular and cellular biology* 25: 2632-2643

Clements D, Taylor HC, Herrmann BG, Stott D (1996) Distinct regulatory control of the Brachyury gene in axial and non-axial mesoderm suggests separation of mesoderm lineages early in mouse gastrulation. *Mechanisms of development* 56: 139-149

Cooper SJ, Trinklein ND, Anton ED, Nguyen L, Myers RM (2006) Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome research* 16: 1-10

Corden J, Wasylyk B, Buchwalder A, Sassone-Corsi P, Kedinger C, Chambon P (1980) Promoter sequences of eukaryotic protein-coding genes. *Science (New York, NY* 209: 1406-1414

Crowley TE, Hoey T, Liu JK, Jan YN, Jan LY, Tjian R (1993) A new factor related to TATA-binding protein has highly restricted expression patterns in Drosophila. *Nature* 361: 557-561

Dantonel JC, Quintin S, Lakatos L, Labouesse M, Tora L (2000) TBP-like factor is required for embryonic RNA polymerase II transcription in C. elegans. *Molecular cell* 6: 715-722

Dantonel JC, Wurtz JM, Poch O, Moras D, Tora L (1999) The TBP-like factor: an alternative transcription factor in metazoa? *Trends in biochemical sciences* 24: 335-339

Davidson I (2003) The genetics of TBP and TBP-related factors. *Trends in biochemical sciences* 28: 391-398

Davuluri RV, Suzuki Y, Sugano S, Zhang MQ (2000) CART classification of human 5' UTR sequences. *Genome research* 10: 1807-1816

De Renzis S, Elemento O, Tavazoie S, Wieschaus EF (2007) Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. *PLoS biology* 5: e117

Deato MD, Tjian R (2007) Switching of the core transcription machinery during myogenesis. *Genes & development* 21: 2137-2149

Deng W, Roberts SG (2005) A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes & development* 19: 2418-2423

Dobrovolskaia-Zavadskaia N (1927) Sur la mortification spontanee de la queue chez la souris nouveau-nee et sur l'existence d'un caractere heriditaire 'non-viable'. *Crit Rev Soc Biol* 97: 114–116

Driever W, Solnica-Krezel L, Schier AF, Neuhauss SC, Malicki J, Stemple DL, Stainier DY, Zwartkruis F, Abdelilah S, Rangini Z, Belak J, Boggs C (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. *Development (Cambridge, England)* 123: 37-46

Dynlacht BD, Hoey T, Tjian R (1991) Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. *Cell* 66: 563-576

Ertzer R, Muller F, Hadzhiev Y, Rathnam S, Fischer N, Rastegar S, Strahle U (2007) Cooperation of sonic hedgehog enhancers in midline expression. *Developmental biology* 301: 578-589

Falender AE, Shimada M, Lo YK, Richards JS (2005) TAF4b, a TBP associated factor, is required for oocyte development and function. *Developmental biology* 288: 405-419

Feaver WJ, Gileadi O, Li Y, Kornberg RD (1991) CTD kinase associated with yeast RNA polymerase II initiation factor b. *Cell* 67: 1223-1230

Ferg M, Sanges R, Gehrig J, Kiss J, Bauer M, Lovas A, Szabo M, Yang L, Straehle U, Pankratz MJ, Olasz F, Stupka E, Muller F (2007) The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish. *The EMBO journal* 26: 3945-3956

Fickett JW, Hatzigeorgiou AG (1997) Eukaryotic promoter recognition. *Genome research* 7: 861-878

Fisher RA, Yates F (1938) Statistical tables for biological, agricultural and medical research. *London: Oliver & Boyd* ISBN 0-02-844720-4

FitzGerald PC, Shlyakhtenko A, Mir AA, Vinson C (2004) Clustering of DNA sequences in human promoters. *Genome research* 14: 1562-1574

Flicek P, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, Down T, Dyer SC, Eyre T, Fitzgerald S, Fernandez-Banet J, Graf S, Haider S, Hammond M, Holland R, Howe KL, Howe K, Johnson N, Jenkinson A, Kahari A, Keefe D, Kokocinski F, Kulesha E, Lawson D, Longden I, Megy K, Meidl P, Overduin B, Parker A, Pritchard B, Prlic A, Rice S, Rios D, Schuster M, Sealy I, Slater G, Smedley D, Spudich G, Trevanion S, Vilella AJ, Vogel J, White S, Wood M, Birney E, Cox T, Curwen V, Durbin R, Fernandez-Suarez XM, Herrero J, Hubbard TJ, Kasprzyk A, Proctor G, Smith J, Ureta-Vidal A, Searle S (2008) Ensembl 2008. *Nucleic acids research* 36: D707-714

Freiman RN, Albright SR, Zheng S, Sha WC, Hammer RE, Tjian R (2001) Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development. *Science (New York, NY* 293: 2084-2087

Frith MC, Li MC, Weng Z (2003) Cluster-Buster: Finding dense clusters of motifs in DNA sequences. *Nucleic acids research* 31: 3666-3668

Gaiano N, Allende M, Amsterdam A, Kawakami K, Hopkins N (1996) Highly efficient germ-line transmission of proviral insertions in zebrafish. *Proceedings of the National Academy of Sciences of the United States of America* 93: 7777-7782

Gazdag E, Rajkovic A, Torres-Padilla ME, Tora L (2007) 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 (Cambridge, England)* 134: 51-62

Gershenzon NI, Ioshikhes IP (2005) Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics (Oxford, England)* 21: 1295-1300

Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, Baskerville S, Hammond SM, Bartel DP, Schier AF (2005) MicroRNAs regulate brain morphogenesis in zebrafish. *Science (New York, NY* 308: 833-838

Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF (2006) Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science (New York, NY* 312: 75-79

Goldberg ML (1979) Sequence analysis of Drosophila histone genes. *PhD dissertation, Stanford University, California*

Grant PA, Schieltz D, Pray-Grant MG, Steger DJ, Reese JC, Yates JR, 3rd, Workman JL (1998) A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. *Cell* 94: 45-53

Green JB, Smith JC, Gerhart JC (1994) Slow emergence of a multithreshold response to activin requires cell-contact-dependent sharpening but not prepattern. *Development (Cambridge, England)* 120: 2271-2278

Griffin K, Patient R, Holder N (1995) Analysis of FGF function in normal and no tail zebrafish embryos reveals separate mechanisms for formation of the trunk and the tail. *Development (Cambridge, England)* 121: 2983-2994

Griffin KJ, Amacher SL, Kimmel CB, Kimelman D (1998) Molecular identification of spadetail: regulation of zebrafish trunk and tail mesoderm formation by T-box genes. *Development (Cambridge, England)* 125: 3379-3388

Guermah M, Ge K, Chiang CM, Roeder RG (2003) The TBN protein, which is essential for early embryonic mouse development, is an inducible TAFII implicated in adipogenesis. *Molecular cell* 12: 991-1001

Haffter P, Granato M, Brand M, Mullins MC, Hammerschmidt M, Kane DA, Odenthal J, van Eeden FJ, Jiang YJ, Heisenberg CP, Kelsh RN, Furutani-Seiki M, Vogelsang E, Beuchle D, Schach U, Fabian C, Nusslein-Volhard C (1996) The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. *Development (Cambridge, England)* 123: 1-36

Halpern ME, Ho RK, Walker C, Kimmel CB (1993) Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation. *Cell* 75: 99-111

Hansen SK, Takada S, Jacobson RH, Lis JT, Tjian R (1997) Transcription properties of a cell type-specific TATA-binding protein, TRF. *Cell* 91: 71-83

Hardy S, Brand M, Mittler G, Yanagisawa J, Kato S, Meisterernst M, Tora L (2002) TATA-binding protein-free TAF-containing complex (TFTC) and p300 are both required for efficient transcriptional activation. *The Journal of biological chemistry* 277: 32875-32882

Harfe BD, McManus MT, Mansfield JH, Hornstein E, Tabin CJ (2005) The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. *Proceedings of the National Academy of Sciences of the United States of America* 102: 10898-10903

Hart DO, Raha T, Lawson ND, Green MR (2007) Initiation of zebrafish haematopoiesis by the TATA-box-binding protein-related factor Trf3. *Nature* 450: 1082-1085

Helmlinger D, Hardy S, Sasorith S, Klein F, Robert F, Weber C, Miguet L, Potier N, Van-Dorsselaer A, Wurtz JM, Mandel JL, Tora L, Devys D (2004) Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. *Human molecular genetics* 13: 1257-1265

Hernandez N (1993) TBP, a universal eukaryotic transcription factor? *Genes & development* 7: 1291-1308

Herrmann BG, Labeit S, Poustka A, King TR, Lehrach H (1990) Cloning of the T gene required in mesoderm formation in the mouse. *Nature* 343: 617-622

Higuchi R, Krummel B, Saiki RK (1988) A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic acids research* 16: 7351-7367

Hiller MA, Lin TY, Wood C, Fuller MT (2001) Developmental regulation of transcription by a tissue-specific TAF homolog. *Genes & development* 15: 1021-1030

Hochheimer A, Zhou S, Zheng S, Holmes MC, Tjian R (2002) TRF2 associates with DREF and directs promoter-selective gene expression in Drosophila. *Nature* 420: 439-445

Hoey T, Dynlacht BD, Peterson MG, Pugh BF, Tjian R (1990) Isolation and characterization of the Drosophila gene encoding the TATA box binding protein, TFIID. *Cell* 61: 1179-1186

Holmes MC, Tjian R (2000) Promoter-selective properties of the TBP-related factor TRF1. *Science (New York, NY* 288: 867-870

Huisinga KL, Pugh BF (2004) A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. *Molecular cell* 13: 573-585

Hultmark D, Klemenz R, Gehring WJ (1986) Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. *Cell* 44: 429-438

Imbalzano AN, Zaret KS, Kingston RE (1994) Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. *The Journal of biological chemistry* 269: 8280-8286

Isaacs HV, Pownall ME, Slack JM (1994) eFGF regulates Xbra expression during Xenopus gastrulation. *The EMBO journal* 13: 4469-4481

Isogai Y, Keles S, Prestel M, Hochheimer A, Tjian R (2007) Transcription of histone gene cluster by differential core-promoter factors. *Genes & development* 21: 2936-2949

Jacobi UG, Akkers RC, Pierson ES, Weeks DL, Dagle JM, Veenstra GJ (2007) TBP paralogs accommodate metazoan- and vertebrate-specific developmental gene regulation. *The EMBO journal* 26: 3900-3909

Jallow Z, Jacobi UG, Weeks DL, Dawid IB, Veenstra GJ (2004) Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in Xenopus. *Proceedings of the National Academy of Sciences of the United States of America* 101: 13525-13530

Jin VX, Singer GA, Agosto-Perez FJ, Liyanarachchi S, Davuluri RV (2006) Genome-wide analy-

sis of core promoter elements from conserved human and mouse orthologous pairs. *BMC bioinformatics* 7: 114

Kaltenbach L, Horner MA, Rothman JH, Mango SE (2000) The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during C. elegans embryogenesis. *Molecular cell* 6: 705-713

Kane DA, Hammerschmidt M, Mullins MC, Maischein HM, Brand M, van Eeden FJ, Furutani-Seiki M, Granato M, Haffter P, Heisenberg CP, Jiang YJ, Kelsh RN, Odenthal J, Warga RM, Nusslein-Volhard C (1996) The zebrafish epiboly mutants. *Development (Cambridge, England)* 123: 47-55

Kane DA, Kimmel CB (1993) The zebrafish midblastula transition. *Development (Cambridge, England)* 119: 447-456

Kawachi K, Masuyama N, Nishida E (2003) Essential role of the transcription factor Ets-2 in Xenopus early development. *The Journal of biological chemistry* 278: 5473-5477

Kawakami K (2004) Transgenesis and gene trap methods in zebrafish by using the Tol2 transposable element. *Methods in cell biology* 77: 201-222

Kawakami K, Koga A, Hori H, Shima A (1998) Excision of the tol2 transposable element of the medaka fish, Oryzias latipes, in zebrafish, Danio rerio. *Gene* 225: 17-22

Kedinger C, Gniazdowski M, Mandel JL, Jr., Gissinger F, Chambon P (1970) Alpha-amanitin: a specific inhibitor of one of two DNA-pendent RNA polymerase activities from calf thymus. *Biochemical and biophysical research communications* 38: 165-171

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B (2005) A high-resolution map of active promoters in the human genome. *Nature* 436: 876-880

Kimmel CB (1989) Genetics and early development of zebrafish. *Trends Genet* 5: 283-288

Kispert A, Hermann BG (1993) The Brachyury gene encodes a novel DNA binding protein. *The EMBO journal* 12: 4898-4899

Kispert A, Koschorz B, Herrmann BG (1995a) The T protein encoded by Brachyury is a tissue-specific transcription factor. *The EMBO journal* 14: 4763-4772

Kispert A, Ortner H, Cooke J, Herrmann BG (1995b) The chick Brachyury gene: developmental expression pattern and response to axial induction by localized activin. *Developmental biology* 168: 406-415

Kopytova DV, Krasnov AN, Kopantceva MR, Nabirochkina EN, Nikolenko JV, Maksimenko O, Kurshakova MM, Lebedeva LA, Yerokhin MM, Simonova OB, Korochkin LI, Tora L, Georgiev PG, Georgieva SG (2006) Two isoforms of Drosophila TRF2 are involved in embryonic development, premeiotic chromatin condensation, and proper differentiation of germ cells of both sexes. *Molecular and cellular biology* 26: 7492-7505

Krichevsky AM, King KS, Donahue CP, Khrapko K, Kosik KS (2003) A microRNA array reveals extensive regulation of microRNAs during brain development. *RNA (New York, NY* 9: 1274-1281

Kutach AK, Kadonaga JT (2000) The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Molecular and cellular biology* 20: 4754-4764

Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH (1998) New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes & development* 12: 34-44

Latinkic BV, Umbhauer M, Neal KA, Lerchner W, Smith JC, Cunliffe V (1997) The Xenopus Brachyury promoter is activated by FGF and low concentrations of activin and suppressed by high concentrations of activin and by paired-type homeodomain proteins. *Genes & development* 11: 3265-3276

Lee DH, Gershenzon N, Gupta M, Ioshikhes IP, Reinberg D, Lewis BA (2005) Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. *Molecular and cellular biology* 25: 9674-9686

Lee TI, Causton HC, Holstege FC, Shen WC, Hannett N, Jennings EG, Winston F, Green MR, Young RA (2000) Redundant roles for the TFIID and SAGA complexes in global transcription. *Nature* 405: 701-704

Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN (2004) MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal* 23: 4051-4060

Lerchner W, Latinkic BV, Remacle JE, Huylebroeck D, Smith JC (2000) Region-specific activation of the Xenopus brachyury promoter involves active repression in ectoderm and endoderm: a study using transgenic frog embryos. *Development (Cambridge, England)* 127: 2729-2739

Lewis BA, Kim TK, Orkin SH (2000) A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. *Proceedings of the National Academy of Sciences of the United States of America* 97: 7172-7177

Lim CY, Santoso B, Boulay T, Dong E, Ohler U, Kadonaga JT (2004) The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes & development* 18: 1606-1617

Martianov I, Fimia GM, Dierich A, Parvinen M, Sassone-Corsi P, Davidson I (2001) Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. *Molecular cell* 7: 509-515

Martinez E, Chiang CM, Ge H, Roeder RG (1994) TATA-binding protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. *The EMBO journal* 13: 3115-3126

Martinez E, Kundu TK, Fu J, Roeder RG (1998) A human SPT3-TAFII31-GCN5-L acetylase complex distinct from transcription factor IID. *The Journal of biological chemistry* 273: 23781-23785

Mathavan S, Lee SG, Mak A, Miller LD, Murthy KR, Govindarajan KR, Tong Y, Wu YL, Lam SH, Yang H, Ruan Y, Korzh V, Gong Z, Liu ET, Lufkin T (2005) Transcriptome analysis of zebrafish

embryogenesis using microarrays. *PLoS genetics* 1: 260-276

Matsui T, Segall J, Weil PA, Roeder RG (1980) Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *The Journal of biological chemistry* 255: 11992-11996

Mengus G, May M, Carre L, Chambon P, Davidson I (1997) Human TAF(II)135 potentiates transcriptional activation by the AF-2s of the retinoic acid, vitamin D3, and thyroid hormone receptors in mammalian cells. *Genes & development* 11: 1381-1395

Molina C, Grotewold E (2005) Genome wide analysis of Arabidopsis core promoters. *BMC genomics* 6: 25

Moore PA, Ozer J, Salunek M, Jan G, Zerby D, Campbell S, Lieberman PM (1999) A human TATA binding protein-related protein with altered DNA binding specificity inhibits transcription from multiple promoters and activators. *Molecular and cellular biology* 19: 7610-7620

Muller F, Blader P, Rastegar S, Fischer N, Knochel W, Strahle U (1999) Characterization of zebrafish smad1, smad2 and smad5: the amino-terminus of smad1 and smad5 is required for specific function in the embryo. *Mechanisms of development* 88: 73-88

Muller F, Demeny MA, Tora L (2007) New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors. *The Journal of biological chemistry* 282: 14685-14689

Muller F, Lakatos L, Dantonel J, Strahle U, Tora L (2001) TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish. *Curr Biol* 11: 282-287

Murphey RD, Zon LI (2006) Small molecule screening in the zebrafish. *Methods (San Diego, Calif* 39: 255-261

Nasevicius A, Ekker SC (2000) Effective targeted gene 'knockdown' in zebrafish. *Nature genetics* 26: 216-220

Newport J, Kirschner M (1982a) A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell* 30: 675-686

Newport J, Kirschner M (1982b) A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. *Cell* 30: 687-696

Nikolov DB, Chen H, Halay ED, Usheva AA, Hisatake K, Lee DK, Roeder RG, Burley SK (1995) Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* 377: 119-128

O'Farrell PH, Stumpff J, Su TT (2004) Embryonic cleavage cycles: how is a mouse like a fly? *Curr Biol* 14: R35-45

Ogryzko VV, Kotani T, Zhang X, Schiltz RL, Howard T, Yang XJ, Howard BH, Qin J, Nakatani Y (1998) Histone-like TAFs within the PCAF histone acetylase complex. *Cell* 94: 35-44

Ohbayashi T, Kishimoto T, Makino Y, Shimada M, Nakadai T, Aoki T, Kawata T, Niwa S, Tamura T (1999) Isolation of cDNA, chromosome mapping, and expression of the human TBP-like

protein. *Biochemical and biophysical research communications* 255: 137-142

Ohbayashi T, Shimada M, Nakadai T, Wada T, Handa H, Tamura T (2003) Vertebrate TBP-like protein (TLP/TRF2/TLF) stimulates TATA-less terminal deoxynucleotidyl transferase promoters in a transient reporter assay, and TFIIA-binding capacity of TLP is required for this function. *Nucleic acids research* 31: 2127-2133

Ohler U (2006) Identification of core promoter modules in Drosophila and their application in accurate transcription start site prediction. *Nucleic acids research* 34: 5943-5950

Ohler U, Liao GC, Niemann H, Rubin GM (2002) Computational analysis of core promoters in the Drosophila genome. *Genome biology* 3: RESEARCH0087

Orphanides G, Lagrange T, Reinberg D (1996) The general transcription factors of RNA polymerase II. *Genes & development* 10: 2657-2683

Palhan VB, Chen S, Peng GH, Tjernberg A, Gamper AM, Fan Y, Chait BT, La Spada AR, Roeder RG (2005) Polyglutamine-expanded ataxin-7 inhibits STAGA histone acetyltransferase activity to produce retinal degeneration. *Proceedings of the National Academy of Sciences of the United States of America* 102: 8472-8477

Pan G, Greenblatt J (1994) Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region immediately upstream of the initiation site. *The Journal of biological chemistry* 269: 30101-30104

Parinov S, Kondrichin I, Korzh V, Emelyanov A (2004) Tol2 transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes in vivo. *Dev Dyn* 231: 449-459

Pelegri F (2003) Maternal factors in zebrafish development. *Dev Dyn* 228: 535-554

Persengiev SP, Zhu X, Dixit BL, Maston GA, Kittler EL, Green MR (2003) TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed. *Proceedings of the National Academy of Sciences of the United States of America* 100: 14887-14891

Prioleau MN, Huet J, Sentenac A, Mechali M (1994) Competition between chromatin and transcription complex assembly regulates gene expression during early development. *Cell* 77: 439-449

Pugh BF, Tjian R (1990) Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell* 61: 1187-1197

Pugh BF, Tjian R (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes & development* 5: 1935-1945

Rabenstein MD, Zhou S, Lis JT, Tjian R (1999) TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. *Proceedings of the National Academy of Sciences of the United States of America* 96: 4791-4796

Robbins LG (1984) Developmental Use of Gene Products in Drosophila: The Maternal-Zygotic Transition. *Genetics* 108: 361-375

Rodaway A, Takeda H, Koshida S, Broadbent J, Price B, Smith JC, Patient R, Holder N (1999) Induction of the mesendoderm in the zebrafish germ ring by yolk cell-derived TGF-beta family signals and discrimination of mesoderm and endoderm by FGF. *Development (Cambridge, England)* 126: 3067-3078

Roeder RG (1991) The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends in biochemical sciences* 16: 402-408

Roeder RG (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends in biochemical sciences* 21: 327-335

Saka Y, Tada M, Smith JC (2000) A screen for targets of the Xenopus T-box gene Xbra. *Mechanisms of development* 93: 27-39

Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic acids research* 32: D91-94

Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume DA (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature reviews* 8: 424-436

Saxonov S, Berg P, Brutlag DL (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America* 103: 1412-1417

Scherf M, Klingenhoff A, Werner T (2000) Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. *Journal of molecular biology* 297: 599-606

Schier AF (2007) The maternal-zygotic transition: death and birth of RNAs. *Science (New York, NY* 316: 406-407

Schmidt EE, Bondareva AA, Radke JR, Capecchi MR (2003) Fundamental cellular processes do not require vertebrate-specific sequences within the TATA-binding protein. *The Journal of biological chemistry* 278: 6168-6174

Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ, Jr. (2005) Promoter features related to tissue specificity as measured by Shannon entropy. *Genome biology* 6: R33

Schulte-Merker S, Ho RK, Herrmann BG, Nusslein-Volhard C (1992) The protein product of the zebrafish homologue of the mouse T gene is expressed in nuclei of the germ ring and the notochord of the early embryo. *Development (Cambridge, England)* 116: 1021-1032

Schulte-Merker S, Smith JC (1995) Mesoderm formation in response to Brachyury requires FGF signalling. *Curr Biol* 5: 62-67

Smale ST (2001) Core promoters: active contributors to combinatorial gene regulation. *Genes & development* 15: 2503-2508

Smale ST, Baltimore D (1989) The "initiator" as a transcription control element. *Cell* 57: 103-113

Smale ST, Jain A, Kaufmann J, Emami KH, Lo K, Garraway IP (1998) The initiator element: a paradigm for core promoter heterogeneity within metazoan protein-coding genes. *Cold Spring Harbor symposia on quantitative biology* 63: 21-31

Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. *Annual review of biochemistry* 72: 449-479

Smale ST, Schmidt MC, Berk AJ, Baltimore D (1990) Transcriptional activation by Sp1 as directed through TATA or initiator: specific requirement for mammalian transcription factor IID. *Proceedings of the National Academy of Sciences of the United States of America* 87: 4509-4513

Smith J (1999) T-box genes: what they do and how they do it. *Trends Genet* 15: 154-158

Smith JC, Price BM, Green JB, Weigel D, Herrmann BG (1991) Expression of a Xenopus homolog of Brachyury (T) is an immediate-early response to mesoderm induction. *Cell* 67: 79-87

Smith JC, Price BM, Van Nimmen K, Huylebroeck D (1990) Identification of a potent Xenopus mesoderm-inducing factor as a homologue of activin A. *Nature* 345: 729-731

Stancheva I, El-Maarri O, Walter J, Niveleau A, Meehan RR (2002) DNA methylation at promoter regions regulates the timing of gene activation in Xenopus laevis embryos. *Developmental biology* 243: 155-165

Suzuki Y, Tsunoda T, Sese J, Taira H, Mizushima-Sugano J, Hata H, Ota T, Isogai T, Tanaka T, Nakamura Y, Suyama A, Sakaki Y, Morishita S, Okubo K, Sugano S (2001) Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome research* 11: 677-684

Tada M, Casey ES, Fairclough L, Smith JC (1998) Bix1, a direct target of Xenopus T-box genes, causes formation of ventral mesoderm and endoderm. *Development (Cambridge, England)* 125: 3997-4006

Tada M, Smith JC (2000) Xwnt11 is a target of Xenopus Brachyury: regulation of gastrulation movements via Dishevelled, but not through the canonical Wnt pathway. *Development (Cambridge, England)* 127: 2227-2238

Takada S, Lis JT, Zhou S, Tjian R (2000) A TRF1:BRF complex directs Drosophila RNA polymerase III transcription. *Cell* 101: 459-469

Teichmann M, Wang Z, Martinez E, Tjernberg A, Zhang D, Vollmer F, Chait BT, Roeder RG (1999) Human TATA-binding protein-related factor-2 (hTRF2) stably associates with hTFIIA in HeLa cells. *Proceedings of the National Academy of Sciences of the United States of America* 96: 13720-13725

Thomas MC, Chiang CM (2006) The general transcription machinery and general cofactors. *Critical reviews in biochemistry and molecular biology* 41: 105-178

Thompson EM, Legouy E, Renard JP (1998) Mouse embryos do not wait for the MBT: chro-

matin and RNA polymerase remodeling in genome activation at the onset of development. *Developmental genetics* 22: 31-42

Tokusumi Y, Ma Y, Song X, Jacobson RH, Takada S (2007) The new core promoter element XCPE1 (X Core Promoter Element 1) directs activator-, mediator-, and TATA-binding protein-dependent but TFIID-independent RNA polymerase II transcription from TATA-less promoters. *Molecular and cellular biology* 27: 1844-1858

Tran TC, Sneed B, Haider J, Blavo D, White A, Aiyejorun T, Baranowski TC, Rubinstein AL, Doan TN, Dingledine R, Sandberg EM (2007) Automated, quantitative screening assay for antiangiogenic compounds using transgenic zebrafish. *Cancer research* 67: 11386-11392

Veenstra GJ, Weeks DL, Wolffe AP (2000) Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in Xenopus. *Science (New York, NY* 290: 2312-2315

Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M, Timmers HT (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* 131: 58-69

Vonica A, Gumbiner BM (2002) Zygotic Wnt activity is required for Brachyury expression in the early Xenopus laevis embryo. *Developmental biology* 250: 112-127

Wakaguri H, Yamashita R, Suzuki Y, Sugano S, Nakai K (2007) DBTSS: database of transcription start sites, progress report 2008. *Nucleic acids research*

Wang Y, Barbacioru C, Hyland F, Xiao W, Hunkapiller KL, Blake J, Chan F, Gonzalez C, Zhang L, Samaha RR (2006) Large scale real-time PCR validation on gene expression measurements from two commercial long-oligonucleotide microarrays. *BMC genomics* 7: 59

Warga RM, Kimmel CB (1990) Cell movements during epiboly and gastrulation in zebrafish. *Development (Cambridge, England)* 108: 569-580

Wasylyk B, Hagman J, Gutierrez-Hartmann A (1998) Ets transcription factors: nuclear effectors of the Ras-MAP-kinase signaling pathway. *Trends in biochemical sciences* 23: 213-216

Weinzierl RO, Dynlacht BD, Tjian R (1993) Largest subunit of Drosophila transcription factor IID directs assembly of a complex containing TBP and a coactivator. *Nature* 362: 511-517

Weis L, Reinberg D (1997) Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. *Molecular and cellular biology* 17: 2973-2984

Wieczorek E, Brand M, Jacq X, Tora L (1998) Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II. *Nature* 393: 187-191

Willy PJ, Kobayashi R, Kadonaga JT (2000) A basal transcription factor that activates or represses transcription. *Science (New York, NY* 290: 982-985

Wu PY, Ruhlmann C, Winston F, Schultz P (2004) Molecular architecture of the S. cerevisiae SAGA complex. *Molecular cell* 15: 199-208

Wu S, Xie X, Liew AW, Yan H (2007) Eukaryotic promoter prediction based on relative entropy and positional information. *Physical review* 75: 041908

Yamakoshi K, Shimoda N (2003) De novo DNA methylation at the CpG island of the zebrafish no tail gene. *Genesis* 37: 195-202

Zawel L, Reinberg D (1993) Initiation of transcription by RNA polymerase II: a multi-step process. *Progress in nucleic acid research and molecular biology* 44: 67-108

Zawel L, Reinberg D (1995) Common themes in assembly and function of eukaryotic transcription complexes. *Annual review of biochemistry* 64: 533-561

Zhang D, Penttila TL, Morris PL, Roeder RG (2001) Cell- and stage-specific high-level expression of TBP-related factor 2 (TRF2) during mouse spermatogenesis. *Mechanisms of development* 106: 203-205

Zhao X, Herr W (2002) A regulated two-step mechanism of TBP binding to DNA: a solvent-exposed surface of TBP inhibits TATA box recognition. *Cell* 108: 615-627

Zhao Y, Ransom JF, Li A, Vedantham V, von Drehle M, Muth AN, Tsuchihashi T, McManus MT, Schwartz RJ, Srivastava D (2007) Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell* 129: 303-317

Acknowledgements

First I would like to thank my supervisor Dr. Ferenc Müller for accepting me as Ph.D student in his lab, for his supervision help and support during my work and the useful suggestions concerning the composition of my thesis.

I thank Prof. Dr. Uwe Strähle for his interest in this work and for accepting to be the first reviewer of my thesis. For the same reason I would like to thank Prof. Dr. Jochen Wittbrodt and for accepting to be the second reviewer of the thesis.

I would like to thank all people involved in this work. Our collaborators Dr. Elia Stupka and Dr. Remo Sanges from the Bioinformatics-CBM in Trieste for the bioinformatic analysis of the microarray data, Dr. Janos Kiss, Dr. Monika Szabo and Ferenc Olasz from the Agricultural Biotechnology centre Gödöllö, Dr. Michael Pankratz, Dr. Lixin Yang and Dr. Matthias Bauer from the Institut for Toxicology and Genetics, Forschungszentrum Karlsruhe, for the microarray data.

I would also like to thank Yavor Hadzhiev for providing the 2.4*shh* construct, Seesandra V. Rajagopala for his initial work on the *notail* promoter, the lab of Tamas Dalmay for detecting miR-430 in zebrafish embryos, Alexander Schier and Antonio Giraldez for sharing miR-430 target site containing GFP constructs, Robert Geisler for sharing unpublished data, Dr. Laszlo Tora for more than one helpfull discussion, Simone Schindler and Carolin Balduf for help and technical assistance. The fish facility Staff: Nadine Borel, Christina Villani, Mandy Thomauske, Richard Braun and Stephanie Wiessner for taking care and keeping the fish in good conditions.

I thank all current and previous lab members: Yavor Hadzhiev, Jochen Gehrig, Andreas Zaucker, Eva Kalmar and Agnes Lovas for their help, discussion and nice working atmosphere.

How can I acknowledge all the people behind the scenes? The people who fed me and took care that I go to work (in time)?
I would like to thank:

My wife Esther, for always being supportive in not only the above mentioned issues.
Inga, who still thinks that I might be a physician soon.
Maria, for various infections that reminded me where real life happens.
My sister Daniela for last second lay outing.

And all the other guys who still think that I train zebrafish.