FUNCTION OF CORE PROMOTERS IN DIFFERENTIAL GENE REGULATION DURING EMBRYOGENESIS

By

JOCHEN GEHRIG

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Department of Medical and Molecular Genetics
School of Clinical and Experimental Medicine
College of Medical and Dental Sciences
The University of Birmingham
January 2010

Abstract

The core promoter is the ultimate target of all transcriptional regulatory processes. The recently discovered diversity of core promoters and basal transcription factors suggests a regulatory role in differential gene expression. However, the direct contribution of the core promoter remains poorly understood. I investigated core promoters and their putative role in differential gene regulation using the zebrafish embryo as an *in vivo* model system. To analyse the functional requirement for the general transcription factor TATA-box binding protein (TBP), a diverse set of promoters was tested for their TBP dependence. This analysis revealed a differential requirement of TBP for promoter activity. To further explore the roles of core promoters the ability of various core promoters to interact with tissue-specific enhancers was investigated. A high-throughput pipeline combining automated imaging with custom-designed software for registration of spatial reporter gene activity in thousands of zebrafish embryos was developed. The technology was applied in a large-scale screen analysing the tissue specific activities of 202 enhancer - core promoter combinations. A variety of interaction specificities observed suggests an important role of the core promoter in combinatorial gene regulation. Overall, these findings indicate that the core promoter significantly contributes to differential transcriptional regulation in the vertebrate embryo.

Acknowledgements

First I would like to thank my supervisor Dr Ferenc Müller for accepting me as a PhD student and for his continuous support, advice, encouragement and enthusiasm throughout my research. Thanks also go to all past and current members of the laboratory in Germany and in the UK for helpful discussions and the great working atmosphere: Marco Ferg, Yavor Hadzhiev, Eva Kalmar, Nan Li, Agnes Lovas, Matthew Rawlings, Jenny Roberts, Simone Schindler, Chengyi Song and Andreas Zaucker.

I am especially grateful to all the people involved in the screening project: Dr Urban Liebel (Institute of Toxicology and Genetics (ITG), Karlsruhe Institute of Technology (KIT), Germany) for providing the screening platform and for his support and enthusiasm, Dr Markus Reischl (Institute for Applied Computer Sciences, KIT, Germany) for the development of the custom design software and the great collaboration despite the geographical distance, Eva Kalmar for her initial work and the good teamwork, and all the helping hands involved in the handling of thousands of zebrafish embryos: Marco Ferg, Andreas Zaucker, Simone Schindler, Yavor Hadzhiev, Chengyi Song and Nadine Gröbner (ITG, KIT). I would like to thank Dr Agnes Lovas for preliminary experiments and sharing 5'-RACE data.

Finally, I would like to thank my parents and my partner Silke for their continued emotional support and encouragement during this project.

Table of contents

List of figures

List of tables

List of abbreviations

bp	base pair(s)
Br	brain domain
CAGE	cap analysis of gene expression
cDNA	complementary DNA
CFP	cyan fluorescent protein
CGI	CpG island
ChIP	chromatin immunoprecipitation
ChIP Seq	Chip sequencing
c-MO	control morpholino
CPE	core promoter element
CRE	cis-regulatory element
DBD	DNA binding domain
EST	expressed sequence tag
Ey	Eye domain
GFP	green fluorescent protein
GRB	genomic regulatory block
He	heart domain
hpf	hours post fertilisation
HRP	horse radish peroxidase
HTS	high throughput screening
kb	kilo base pair(s)
Mb	mega base pair(s)
MBT	mid-blastula transition
Mh	MHB domain
MHB	midbrain-hindbrain boundary
MZT	maternal-to-zygotic transition
NFR	nucleosome free region
No	notochord domain
nt	nucleotide(s)
PBS	phosphate buffered saline
PIC	pre-initiation complex
PKI	dominant negative protein kinase A
PTU	Phenyl-2-thiourea
PWM	positional weight matrix
RACE	rapid amplification of cDNA ends
RNAP II	RNA Polymerase II
SAGE	serial analysis of gene expression
Sk	skin domain
SNP	single nucleotide polymorphism
Sp	spinal cord domain

TBP-MO	TBP morpholino
TF	transcription factor
TFBS	transcription factor binding site
TLS	TATA-like sequence
TSS	transcription start site
UTR	untranslated region
wt	wildtype
YFP	yellow fluorescent protein
Yo	Yolk domain

1. GENERAL INTRODUCTION

1.1. Regulation of RNA polymerase II dependent transcription

Eukaryotic DNA dependent RNA Polymerase II (RNAP II) is responsible for the transcription of the genetic information encoded in the DNA sequence of tens of thousands of protein coding genes. The correct spatial and temporal transcription of genes needs to be tightly controlled as it is the first step in differential gene expression, which is a prerequisite for the development, growth and survival of all eukaryotic organisms. Thus, the analysis of gene regulatory processes is a key to the understanding how dysregulation of gene expression contributes to pathological conditions (Lee and Young 2000).

Initiation of transcription is controlled by a wide variety of mechanisms that integrate cellular and environmental signals into a transcriptional response. In metazoans, transcription initiation involves a multitude of transacting factors that interact with a complementing variety of cis-regulatory elements (CREs) which together form a gene regulatory network that controls the expression of most if not all protein coding genes. The information that dictates the spatiotemporal transcription of genes is generally encoded in the DNA sequence of three types of CREs (**Figure 1.1**): distal-acting enhancers or silencers, proximal promoters and core promoters. Promoter proximal and distal regions are the targets of sequence specific transcription factors, whereas the core promoter is the target of the pre-initiation complex (PIC) composed of the basal transcription factors and RNAP II. The distinct sequence regions and associated proteins interact with each other either directly or

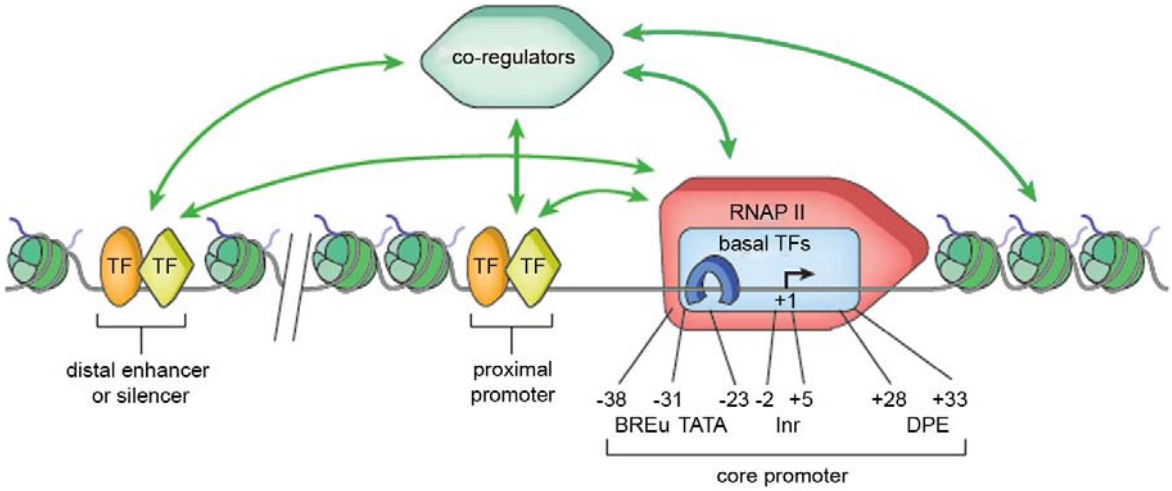


Figure 1.1| Cis-regulatory DNA elements involved in transcriptional regulation. A schematic representation of some CREs and associated factors involved in transcriptional regulation is shown. The core promoter encompasses the transcription start site (TSS) (black arrow) of a gene. The core promoter is bound by the PIC composed of the basal transcription factors (blue rectangle) and RNAP II (red 'rocket'). The blue 'horseshoe' depicts the TATA-Box binding protein (TBP). Basal transcription factors can interact with several core promoter elements (CPE) exemplified by the BREu, the TATA-Box, the Inr and the DPE. Numbers indicate the position of the CPEs relative to the TSS at +1. Proximal promoters are sequence regions generally extending up to several hundred bp upstream of the core promoter. Distal enhancers or silencers are CREs usually located distant to the core promoter. Transcription factors (TF, orange ovals and yellow diamonds) bind to regulatory elements in the proximal promoter region or in distal enhancers and interact with the PIC and core promoter chromatin (nucleosomes are indicated as green 'globes') either directly or via co-regulator complexes that mediate enhancer-promoter communication. For further details see **1.2.** - **1.4.** (adapted from Fuda et al. 2009).

indirectly, and this interplay of *cis*-regulatory elements and bound transcription factors directs the differential transcriptional activity of genes (Levine and Tjian 2003).

Ultimately, all these regulatory processes control the initiation of transcription at the core promoter. Thus, the core promoter lies at the centre of all gene regulatory processes governing transcriptional regulation (**Figure 1.1**). But despite this central role, the contribution of the core promoter to differential gene regulation is often overlooked (Juven-Gershon and Kadonaga 2009).

1.1.1. A classical model of transcription initiation

The term core promoter refers to the sequence encompassing the transcription start site (TSS) of a gene. The core promoter has been functionally defined as the minimal sequence required to direct accurate initiation of transcription *in-vitro* by the basal RNAP II machinery (Smale and Kadonaga 2003). Initially, core promoters have been considered highly similar in sequence composition and structure with an AT-rich sequence (TATA-box) frequently positioned at around -30 bp in relation to a clearly defined TSS located within a pyrimidine rich initiator sequence (Breathnach and Chambon 1981).

In line with the notion of a universal core promoter the processes leading to PIC assembly and initiation of transcription have also appeared to be a common mechanism. The PIC is composed of RNAP II and the basal transcription factors TFIID, -A, -B, -F, -E, and -H and is sufficient to direct basal transcription from a DNA template *in-vitro* (Orphanides et al. 1996; Roeder 1996). The PIC is believed to be assembled by the stepwise recruitment of factors or as a holoenzyme. The core promoter is mostly recognized by the multiprotein complex TFIID mainly via interaction of its subunit TBP (TATA-box binding protein) with the TATA-box, an event facilitated by TFIIA. In the stepwise model TFIID binding is followed by the entry of TFIIB, which stabilizes the TFIID-DNA complex and aids in TSS selection. Subsequently, the RNAP II - TFIIF complex, TFIIE and TFIIH are recruited. The nucleation of the PIC is followed by the melting of the core promoter DNA and initiation of transcription around 30 bp downstream of the TATA-box. Then TFIIH phosphorylates the C-terminal domain of RNAP II

leading to promoter clearance and transition into the elongation phase (Orphanides et al. 1996; Roeder 1996).

1.2. Diversity of core promoters

The above outlined model describes the core promoter as a rather static component in gene regulation serving mainly as a launching pad for transcription. In the last two decades, this view of the core promoter has been challenged by many observations. There are many different core promoter types with varying promoter architectures suggesting that gene regulatory processes at the core promoter are much more complex than initially thought.

1.2.1. Promoter discovery in eukaryotic genomes

One of the main obstacles in the bioinformatic and functional characterization of core promoters has been the limited availability of correct information of the TSS and thus the promoter of a gene. Bioinformatic predictions of TSSs have proven difficult due to the lack of common sequence features of core promoters (Ohler 2006). There are several methods to experimentally infer TSSs such as 5'-RACE, primer extension or nuclease protection assays. However, these methods target specific genomic regions or genes, respectively, and are therefore generally of low throughput. Thus, analyses of core promoters are often restricted to relatively small experimentally verified sets of promoters. Alternatively, promoter studies are based on TSS predictions using the mapping of 5'-ends of sequenced cDNAs or expressed sequence tags (EST), while the depth of coverage often remains limited (Sandelin et al. 2007).

In recent years, large scale genomic approaches have been developed to increase the information about the precise location of TSSs and promoters. Many of these methods are based on the high throughput sequencing of 5'-ends of cDNAs that encompass the full 5'-ends of transcripts using next generation sequencing. Methods successfully used include the sequencing of 5'-ends of full length cDNA (Suzuki et al. 2004), 5'-SAGE (Serial analysis of gene expression) (Hashimoto et al. 2004) and CAGE (Cap analysis of gene expression) (Shiraki et al. 2003). The large scale sequencing of 5´-ends of full length cDNAs has led to a constantly growing database of millions of eukaryotic TSSs (http://dbtss.hgc.jp/) covering tens of thousands of promoters for several model species including human, mouse and zebrafish (Wakaguri et al. 2008). However, the most comprehensive experimental promoter study to date is the analysis of the mouse and human genome using CAGE by Carninci *et al*. 2006 using massive parallel sequencing of millions of 20- or 21 nt-long sequence tags representing the 5'-ends of cap-trapped RNAs, i.e. RNAs isolated by a protocol that enriches for RNA species containing a 7-methylguanosine cap-structure at their 5' terminal end (Shiraki et al. 2003). As a result, they identified around 730,000 mouse and 665,000 human TSSs that clustered to 177,000 mouse and 184,000 human promoters, respectively (Carninci et al. 2006).

In addition to the transcript based sequencing approaches, methods using chromatin immunoprecipitation (ChIP) in combination with tiling microarrays (ChIP-Chip) or high throughput sequencing (Chip-Seq) to analyze the genome-wide distribution of promoter associated features like the binding of promoter recognition factors or chromatin signatures have been successfully used (Kim et al. 2005; Birney et al. 2007). While these methods are

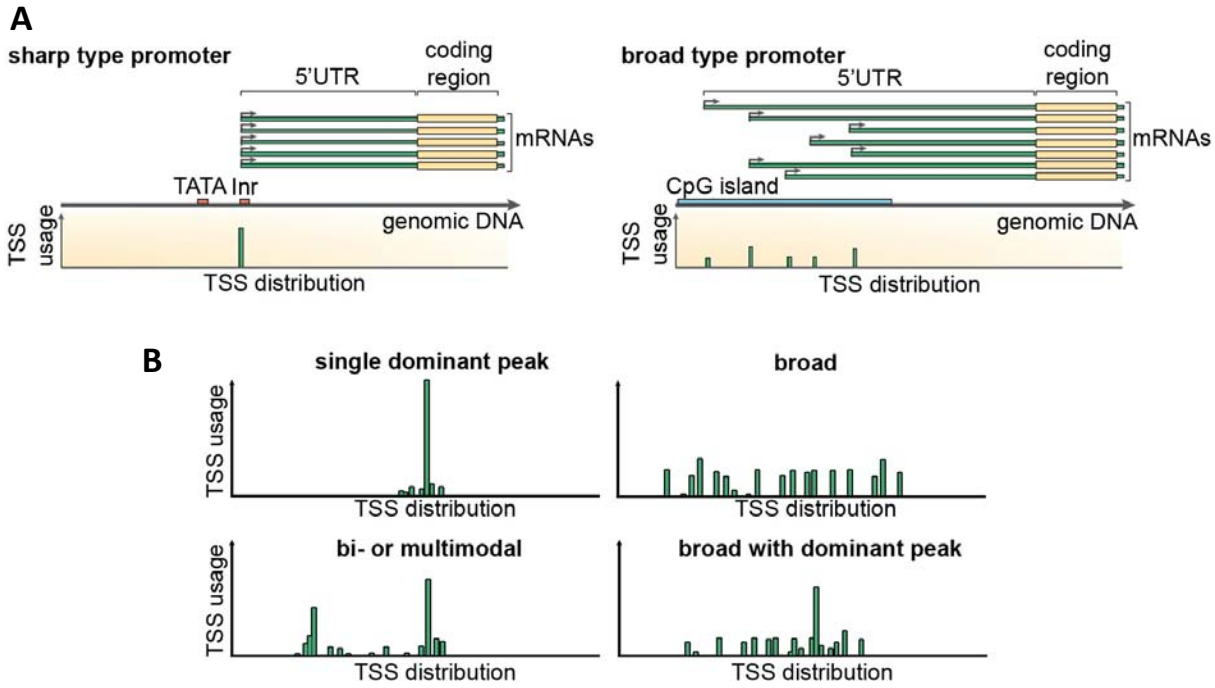


Figure 1.2|TSS classes of mammalian core promoters. (**A**) TSS distributions allow the classification of core promoters into two main categories: sharp and broad type promoters. Sharp type promoters have a clearly defined TSS and give rise to mRNAs with almost identical 5'UTRs. Broad type promoters initiate transcription at various positions leading to mRNA populations with distinct 5'UTRs, whereas the coding region is invariant. Promoter features often associated with the two distinct classes (TATA-Box, Initiator or CpG islands, respectively) are indicated. The different TSS distributions and usage frequencies within core promoter regions are depicted as green bars. (**B**) Subclassification of mammalian core promoters. Example TSS usages of core promoter classes are shown: Single dominant peak, general broad distribution, bi- or multimodal distribution and broad with dominant peak. TSS distributions are depicted as in **A** (adapted from Kawaji et al. 2006; Sandelin et al. 2007).

very powerful in the identification of promoter regions and promoter occupancy, the resolution remains limited and is often not sufficient to identify the precise location of the core promoter (Sandelin et al. 2007).

1.2.2. Transcription start site distributions define promoter classes

One of the most intriguing results of the large scale mapping of TSS was the discovery of several classes of core promoters that display differential distribution of TSSs within the core

promoter region suggesting differential regulatory mechanisms controlling transcription initiation (**Figure 1.2**) (Suzuki et al. 2001a; Carninci et al. 2006). Especially the CAGE based approach has revealed that the classical core promoter (see **1.1.1**) with a clearly defined TSS position is a rather rare feature in the mouse and human genomes. Instead, the majority of promoters have multiple start sites within the promoter region. Furthermore, these multiple TSSs cluster differentially in distinct promoters revealing various TSS distribution shapes indicating different modes of transcription initiation (**Figure 1.2**) (Carninci et al. 2005; Carninci et al. 2006). Based on >8,000 promoters supported by more than 100 CAGE tags two main promoter classes were generated: sharp and broad type promoters (**Figure 1.2A**) (Carninci et al. 2006; Kawaji et al. 2006). The sharp type class represents only 23% of the promoters examined and displays a narrowly defined TSS location often associated with tissue-specific expression and the presence of a TATA-box. The majority of promoters (67%) are represented in the broad type class. These promoters harbour multiple TSSs spanning on average a region of 71bp, but no more than 150 bp. Broad class promoters are mostly TATA-less and often associated with CpG islands and ubiquitously expressed genes (Carninci et al. 2005; Carninci et al. 2006). The broad type promoters can be further subdivided into three categories (**Figure 1.2B**): a general broad distribution, a bi- or multimodal distribution and a broad distribution with a dominant peak. Importantly, the TSSs distributions of promoters are conserved between orthologous genes of mouse and human adding further support to this classification (Carninci et al. 2006).

The sharp type class is the only category that is in agreement with a model of transcription initiation proposing that the recruitment of a PIC leads to transcriptional initiation at a

clearly defined TSS. Broad peak promoters do not match this model suggesting distinct mechanisms in transcriptional initiation (Muller et al. 2007; Sandelin et al. 2007; Juven-Gershon and Kadonaga 2009). It has been suggested that the broad distribution of the TSSs, within these usually core promoter element free promoters, might reflect a less tightly controlled initiation of transcription caused by less stringent binding of a PIC. Therefore, the multiple TSSs are the result of the consecutive or simultaneous binding of multiple PICs. Another explanation could be that the single TSSs in broad peak promoters reflect the TSS usage in distinct cell types and the broadness is the result of differential TSS usage in a heterogeneous cell population (Carninci et al. 2006). This is in line with the finding that TSS selection within broad class promoters can be regulated tissue-specifically (Kawaji et al. 2006).

1.2.3. Alternative promoters

Alternative promoters represent independent, mostly non-overlapping sites of transcriptional initiation often separated by up to several hundred bp (Carninci et al. 2006; Kimura et al. 2006) and are not equivalent to the differential usage of TSSs in distinct tissues (Kawaji et al. 2006). Analyses of vertebrate genomes have revealed that a large proportion of protein-coding genes are associated with more than one promoter (Carninci et al. 2006; Birney et al. 2007; Valen et al. 2009). For example, identification and functional analysis of 642 promoters, representing 1% of the human genome, in 16 cell lines has revealed that >20% of genes have functional alternative promoters (Cooper et al. 2006; Birney et al. 2007). Another study analysing the distribution of >1.7 million 5’ ends of full-length cDNAs suggests

that 52% of human genes are associated with on average 3 alternative promoters per gene (Kimura et al. 2006).

The selective usage of alternative promoters can contribute to differential gene regulation in various cell lineages, tissue types or developmental stages potentially via targeting by cell-type specifically expressed components of the basal transcription machinery (see **1.3**) (Davuluri et al. 2008). For example, it could be shown that the *EIF1AX* and *HBG1* genes are associated with a TATA-containing and a TATA-less promoter, which are active at different stages of mouse development (Davis and Schultz 2000; Duan et al. 2002). Additionally, usage of alternative promoters can give rise to protein isoforms due to the production of alternative transcripts that lack or contain additional protein coding information (Kimura et al. 2006; Davuluri et al. 2008). An example is the zebrafish *tp53* gene, which contains an intronic promoter that gives rise to a truncated isoform of tp53 that is activated upon apoptotic stimuli and modulates p53 activity (Chen et al. 2009).

Besides alternative promoters associated with the 5' ends of genes cap-trapping approaches have identified unconventional sites of putative transcriptional initiation (Carninci et al. 2005; Carninci et al. 2006). Many genes harbour tag clusters within their 3'UTR and functional analysis has shown that these sites indeed can serve as promoters in transgenic reporter assays. While the function of the transcripts originating from these sites remains elusive, it has been hypothesized that they are the promoters of genes generating non-coding RNAs that could play a role in the regulation of downstream genes (Carninci et al. 2006). In addition, many 5'ends of capped transcripts map into exons of genes and the origin

of these sites is mostly unknown (Carninci et al. 2006). Recently, it has been suggested that they do not display sites of transcriptional initiation, but indicate the capped products of a currently unknown mRNA degradation pathway (Affymetrix consortium 2009).

1.2.4. Bidirectional promoters

Despite the large genomic space in eukaryotic genomes many genes share a promoter region (Koyanagi et al. 2005). These bidirectional promoters are usually separated by less than 1000 bp and flanked by the TSSs of two divergent genes. In the human genome around 11% of protein-coding genes are controlled by bidirectional promoters (Trinklein et al. 2004). Many genes regulated by bidirectional promoters are co-expressed, while mutually exclusive expression is also observed. The mechanisms of transcriptional regulation at bidirectional promoters are unknown; however they usually lack the classical core promoter motifs and are often associated with CpG islands (Trinklein et al. 2004). Besides true bidirectionality with no overlap of transcripts there are cases with sense-antisense pairs overlapping at their 5' ends with the potential of transcriptional interference (Carninci et al. 2006).

1.2.5. Chromatin modifications at the core promoter

Transcriptional regulation at the core promoter occurs in the context of chromatin and modulation of nucleosome positioning and architecture at the core promoter can influence the interaction of regulatory proteins with the DNA. There is a plethora of literature regarding epigenetic mechanisms influencing gene activity and a comprehensive discussion

would be beyond the scope of this introduction, thus only a few major findings are highlighted.

The core promoters of constitutively active eukaryotic genes often overlap with nucleosome free regions (NFR) of around 150 bp in size; a configuration favouring the binding of basal transcription factors (Cairns 2009). Additionally, NFRs are often flanked by nucleosomes containing the histone variant H2A.Z implicating a role of H2A.Z in the maintenance or formation of NFRs. The promoters of regulated genes are usually covered with nucleosomes in their inactive state. These nucleosomes compete with transcription factors for DNA binding making the regulated promoters more reliant on chromatin remodelling factors (Barski et al. 2007; Heintzman and Ren 2007; Heintzman et al. 2007; Cairns 2009).

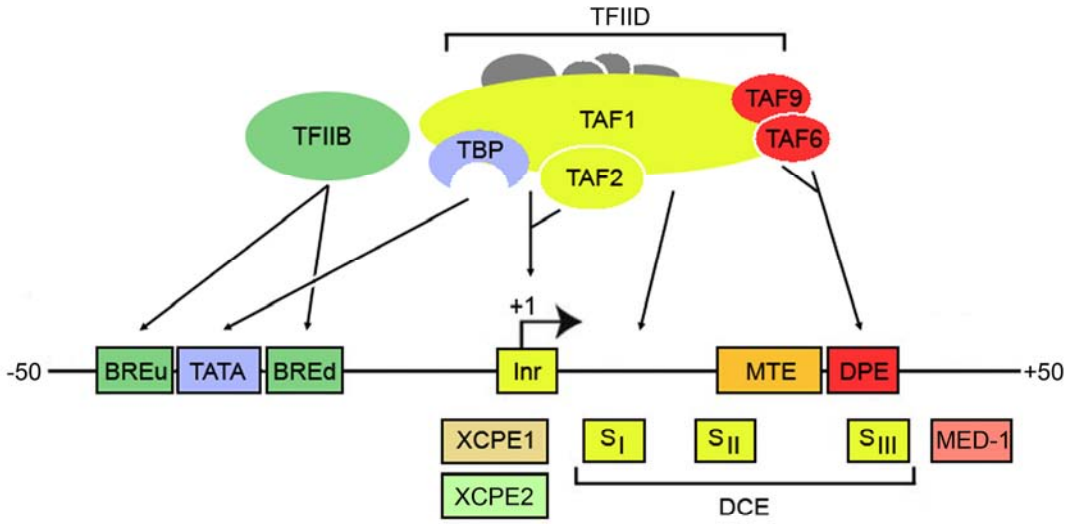
The N-terminal tails of histones within nucleosomes can be covalently modified and global analyses of the distribution of modified histones have revealed that several modifications are associated with promoters. Acetylation of histones can occur on many lysines in the various histone tails and acetylated histone 3 and 4 have been found to be highly associated with the promoter regions of active genes in several eukaryotic species (Schubeler et al. 2004; Kim et al. 2005). However, also active enhancer regions appear to be associated with this feature (Heintzman et al. 2007). Methylation of lysines, another common modification of histone tails can be associated with active or inactive promoter regions depending on the methylated residues within the N-terminal tail. For example, it is established in many eukaryotic organisms that the H3K4Me3 modification colocalizes with active promoters, and to a lesser extent with inactive promoters (Pokholok et al. 2005; Wardle et al. 2006; Barski et

al. 2007; Heintzman et al. 2007; Hon et al. 2009). Globally, the H3K4Me3 peaks around the TSS and is often followed by a gradient of H3K4Me2 and H3K4Me1 (Heintzman et al. 2007). Interestingly, it has been shown that H3K4Me3 is directly bound by TFIID via its subunit TAF3 providing a direct link between histone modifications and promoter recognition by the basal transcription machinery (Vermeulen et al. 2007). Inactive promoters generally lack the above mentioned modifications; instead they are associated with e.g. the H3K27Me3 modification (Akkers et al. 2009). Moreover, inactive promoters tend to be located in heterochromatic regions inaccessible to transcription factors (Heintzman and Ren 2007).

1.2.6. Core promoter elements

Core promoter elements (CPEs) are DNA sequence motifs of the core promoter that are important for the interaction with the basal transcription machinery and therefore promoter activity (Smale and Kadonaga 2003). Notably, there is no universal CPE present in all eukaryotic core promoters and the majority of core promoters do not carry any known CPE suggesting that many additional CPEs may remain to be discovered (**Figure 1.3**).

Although only the minority of core promoters are represented in the sharp class category, mechanistic studies regarding the function and structure of core promoters have mainly examined promoters of this class due to the biological significance of these mostly tissue-specific and highly regulated genes. Moreover, the identification of CPEs, which often display positional constraint in relation to the TSS, is facilitated in sharp class promoters. Therefore, most of the current knowledge regarding the function of CPEs is biased towards mechanisms acting on this class of promoters (Juven-Gershon and Kadonaga 2009).



BREu upstream of TATA-Box SSRCGCC	TATA-Box upstream T at -34 to -28 TATAWAAR	BREd -23 to -17 RTDKKKK	Inr -2 to +4/+5 YYANWYY (human) TCAKTY (*Drosophila*)	MTE +18 to +27 CSARCSSAAC
DPE +28 to +33 RGWYVT	XCPE1 -8 to +2 DSGYGGRASM	XCPE2 -10 to +2 VCYCRTTRCMY	DCE SI +6 to +11 CTTC SII +16 to +21 CTGT SIII +30 to +34 AGC	MED-1 +20 to +45 GCTCCS

Figure 1.3| Eukaryotic core promoter elements and interacting basal transcription factors. The core promoter diagram (upper panel), which is roughly to scale, shows the location of known CPEs. Above the diagram a schematic illustration of PIC components (TFIID and TFIIB) known to be involved in core promoter recognition is shown. TFIID components known to interact directly with CPEs are highlighted. Described protein-CPE interactions are indicated by an arrow and identical colouring. The BREu and BREd elements are recognized by TFIIB, the TATA-Box by TBP (and TBP2), the Inr by TAF1 and TAF2, the DPE by TAF6 and TAF9 and the DCE by TAF1. The MTE is known to interact with TFIID in general. XCPE1 and XCPE2 appear to act in a TAF independent manner. In the lower panel the location of each CPE in relation to the TSS (black arrow) at +1 and the CPE consensus sequences are given (modified after Thomas and Chiang 2006).

1.2.6.1. The TATA-box

The TATA-box is the first CPE that was identified. It was described as the key CPE indispensable for the recruitment of the basal transcription machinery and selection of the transcription start site in most eukaryotic promoters (Lifton et al. 1978; Breathnach and Chambon 1981; Orphanides et al. 1996). The key rate-limiting step in PIC assembly on TATA-containing promoters is the recognition of the TATA-box by the TFIID subunit TBP (Hernandez 1993). However, also TRF1 and TBP2, metazoan specific members of the TBP family, appear to be able to bind to TATA-box sequences (Hansen et al. 1997; Bartfai et al. 2004; Jallow et al. 2004).

The consensus sequence of the TATA-box is TATAWAAR (Bucher 1990), but also minor variants of this consensus can serve as TBP binding sites (Patikoglou et al. 1999). In metazoans the 5'-T is usually located between -34 to -28 (**Figure 1.3**), while its optimal position in relation to the TSS is -31 to -30 (Ponjavic et al. 2006). In *Saccharomyces cerevisiae* its position is more variable ranging from -50 to -200 in relation to the ATG start codon (Basehoar et al. 2004). Initially, the TATA-box was found in most eukaryotic core promoters examined, leading to the concept that it is a rather universal CPE. However, bioinformatic analyses of larger promoter sets have shown that the TATA-box is not a general feature of core promoters: In yeast the TATA-box was found in ~20% of all promoters (Basehoar et al. 2004). In *Drosophila* only between 28.3% to 43% of core promoters contain a TATA-box (Kutach and Kadonaga 2000; Ohler et al. 2002) and in the human genome, where core promoter architecture appears to be more diverse (Juven-Gershon and Kadonaga 2009), the

detected prevalence of the TATA-box is more variable ranging from 2.6%-27% (Trinklein et al. 2003; FitzGerald et al. 2004) depending on the analysed promoter set and the computational approach used. In the zebrafish genome the TATA-box is estimated to be present in 4.1% of the core promoters (Ferg 2008). Genome-wide CAGE studies in mouse and human have revealed that the presence of the TATA-box is associated with sharp class promoters of highly regulated, tissue-specific genes (Carninci et al. 2006; Ponjavic et al. 2006). The prevalence of single peak promoters is estimated to be approximately 25%, among which only 17% contain a consensus TATA-box (approx. 4.3% in total) (Carninci et al. 2006; Sandelin et al. 2007).

1.2.6.2. The initiator

The initiator (Inr) is a pyrimidine rich sequence motif encompassing the TSS of eukaryotic genes. The Inr is one of the first CPEs that was identified (Breathnach and Chambon 1981) and has been defined to be important for TSS selection and promoter strength *in-vitro* and *in-vivo*. It can function alone or in conjunction with other CPEs, often in a cooperative or synergistic manner (Smale and Baltimore 1989; Burke and Kadonaga 1996; Kutach and Kadonaga 2000). There is evidence that the Inr can interact with the basal transcription machinery via the TFIID subunits TAF1 and TAF2 (Chalkley and Verrijzer 1999). Its consensus sequence is $YYA_{+1}NWYY$ (Bucher 1990), while in *Drosophila* a less degenerate consensus $TCA_{+1}KTY$ is commonly used (**Figure 1.3**) (Arkhipova 1995; Kutach and Kadonaga 2000). The A_{+1} nucleotide is usually designated as the +1 start site, but transcription can also start at positions in the close vicinity. However, other CPEs (e.g. DPE, MTE) show a strict spacing

requirement to the A_{+1} position (Kutach and Kadonaga 2000; Lim et al. 2004). Several bioinformatic studies suggest that the Inr is a rather abundant CPE with estimated frequencies of 63-67% in *Drosophila* (Ohler et al. 2002) and 10.3% in the zebrafish genome (Ferg 2008). In the human genome estimates are much more diverse ranging from no apparent clustering at the TSS (FitzGerald et al. 2004) to 46% (Yang et al. 2007). In contrast, newer studies involving the genome wide mapping of TSSs by CAGE in the mouse and human genomes suggest that the Inr is a rare CPE as most promoters regardless of class lack the classical Inr. Many TSSs are rather associated with a YR[-1,+1]-dinucleotide, whereas the classical initiator, often in conjunction with the TATA-box, is a feature of a subset of sharp class promoters needed for very accurate initiation of transcription (Carninci et al. 2006; Sandelin et al. 2007; Juven-Gershon and Kadonaga 2009).

1.2.6.3. The Downstream promoter element

The DPE has been identified as a sequence motif present in Inr-containing TATA-less promoters located exactly +28 to +33 downstream of the A_{+1} position of the Inr and the spacing between the Inr and the DPE is critical for the cooperative function of the two elements (Burke and Kadonaga 1996; Kutach and Kadonaga 2000). Usually the DPE is present in TATA-less promoters, but it can also be found in some TATA containing promoters. Its mutation in TATA-less promoters leads to a significant reduction in basal transcription activity (Burke and Kadonaga 1996). The consensus sequence of the DPE is RGWYVT and appears to be conserved from *Drosophila* to human (**Figure 1.3**) (Burke and Kadonaga 1997), but it is currently unclear whether the DPE is important for mammalian

promoter function (Sandelin et al. 2007). In *Drosophila* the DPE seems to be as prevalent as the TATA-box and could be found in 8-40% of promoters analysed (Kutach and Kadonaga 2000; Ohler et al. 2002). It could be shown that the TFIID subunits TAF6 and TAF9 interact directly with the DPE (Burke and Kadonaga 1997). Interestingly, DPE and TATA dependent promoters can form a regulatory circuit involving TBP, NC2 and Mot1. While TBP activates transcription from TATA dependent promoters, it can repress activity of DPE dependent promoters. NC2 and Mot1 block the activating and repressing functions of TBP, thus repressing TATA-dependent promoters and elevating DPE dependent transcription (Willy et al. 2000; Hsu et al. 2008).

1.2.6.4. The TFIIB recognition elements

Two CPEs have been described that can interact with the basal transcription factor TFIIB and can be found in a subset of TATA containing promoters. The BREu element is located immediately upstream of TATA-boxes in a subset of TATA-containing promoters and its consensus sequence is SSRCGCC (where the 3'-C is immediately followed by the 5'-T of the TATA-box) (**Figure 1.3**) (Lagrange et al. 1998). The TFIIB-BREu interaction appears to be involved in stabilizing TBP-TATA binding and in orienting the PIC on the core promoter (Lagrange et al. 1998). The second TFIIB interacting element BREd is located immediately downstream of the TATA-box and its consensus sequence is RTDKKKK (**Figure 1.3**)(Deng and Roberts 2005). Both BRE elements seem to have a context-dependent activating or suppressing function, respectively (Lagrange et al. 1998; Evans et al. 2001; Deng and Roberts 2005).

1.2.6.5. Other core promoter elements

The downstream core element (DCE) was initially identified in the downstream region of promoters of the human β-globin locus (Lewis et al. 2000), but also other promoters harbour this CPE (Lee et al. 2005). The DCE has a three-partite structure consisting of three sub-elements S_I (+6 to +10: CTTC), S_{II} (+16 to +20: CTGT) and S_{III} (+30 to +33: AGC). It could be shown that the TFIID subunit TAF1 can interact with the DCE *in-vitro* (**Figure 1.3**) (Lee et al. 2005).

The motif 10 element (MTE) is an overrepresented motif in the downstream region of *Drosophila* core promoters (Ohler et al. 2002). The MTE with its consensus sequence CSARCSSAAC is located at +18 to +27 in a strict spacing requirement to the A+1 of the Inr and it was suggested that the MTE is recognized by TFIID (**Figure 1.3**)(Lim et al. 2004). The MTE acts cooperatively with the Inr and can act synergistically with the TATA-box and the DPE (Juven-Gershon et al. 2006).

The Multiple start site Element Downstream (MED-1) was found to be overrepresented in 14 out of 15 mouse and human promoters with multiple start sites. It is located between 20-45bp downstream of the most downstream TSS and its consensus is GCTCCS (**Figure 1.3**). MED-1 mutation led to decreased transcriptional activity in some of these core promoters (Ince and Scotto 1995).

The X core promoter elements (XCPE1 and XCPE2) are CPEs that were identified in the TATA-less HBV X gene promoter. The XCPE1 element encompasses the first TSS of the HBV X gene from -8 to +2 and has the consensus DSGYGGRASM (**Figure 1.3**). It can be found in

approximately 1% of human promoters and appears to be able to function in a TFIID independent manner (Tokusumi et al. 2007). The XCPE2 element harbours the second TSS of the HBV X gene and appears to be enriched in human broad class promoters (**Figure 1.3**) (Anish et al. 2009).

1.2.7. CpG islands

CpG islands are not a classical *cis*-regulatory motif, but rather a DNA sequence feature associated with promoter regions, but the direct consequence of its presence for transcriptional initiation at the core promoter is unknown (Gardiner-Garden and Frommer 1987; Heintzman and Ren 2007).

The majority of CpG (i.e. CG) dinucleotides in vertebrate genomes are methylated at the 5 carbon position of the cytosine ring. Methyl-cytosine is a hot-spot for mutations as it spontaneously deaminates to thymine, which leads to an underrepresentation of CpGs in the genome. However, there are unmethylated regions with an elevated CG content and little CpG suppression called CpG islands (Gardiner-Garden and Frommer 1987). These CpG islands are variable in length (around 0.5 to 2 kb) and overlap with the promoter regions of many genes. In the human genome it is estimated that around 50-70% of all promoters are located within CpG islands (Gardiner-Garden and Frommer 1987; Carninci et al. 2006; Saxonov et al. 2006). CpG island promoters are mainly associated with ubiquitously expressed, so-called housekeeping genes, although many exceptions exist. CpG islands can be the target of DNA methylation which plays a role in transcriptional silencing and it is

believed that methylation of CpG dinucleotides blocks transcription factors from interacting with the regulatory DNA by establishing a repressive chromatin structure (Jones et al. 1998).

In general, CpG island promoters lack any of the well known CPEs (e.g. TATA-box) and display a broad distribution of TSSs (Carninci et al. 2006). However, CpG islands often contain multiple target sites for the transcription factor Sp1, which are often located in vicinity to the TSS, suggesting a potential role of Sp1 in PIC recruitment (Gardiner-Garden and Frommer 1987).

1.3. Diversity of core promoter recognition factors

In addition to the huge diversity of core promoters, metazoan organisms have evolved a complementing variety of components of the basal transcription machinery. These sometimes tissue-specifically expressed variants suggest cell-type and gene-specific functions of core promoter binding proteins, thus adding a further layer of complexity to gene regulation at the core promoter level (**Figure 1.4**).

1.3.1. Variability in TAF composition of TFIID

The TFIID multi-subunit protein complex is thought to be the main component of the pre-initiation complex responsible for promoter recognition and its binding is considered as being the key rate limiting step in controlling core promoter activity. The TFIID complex is composed of TBP and 12 to 15 TBP associated factors (TAFs) (Albright and Tjian 2000; Tora 2002; Muller and Tora 2004). As already mentioned TBP is not the only TFIID component that is able to interact with the core promoter. Also certain TAFs are involved in core

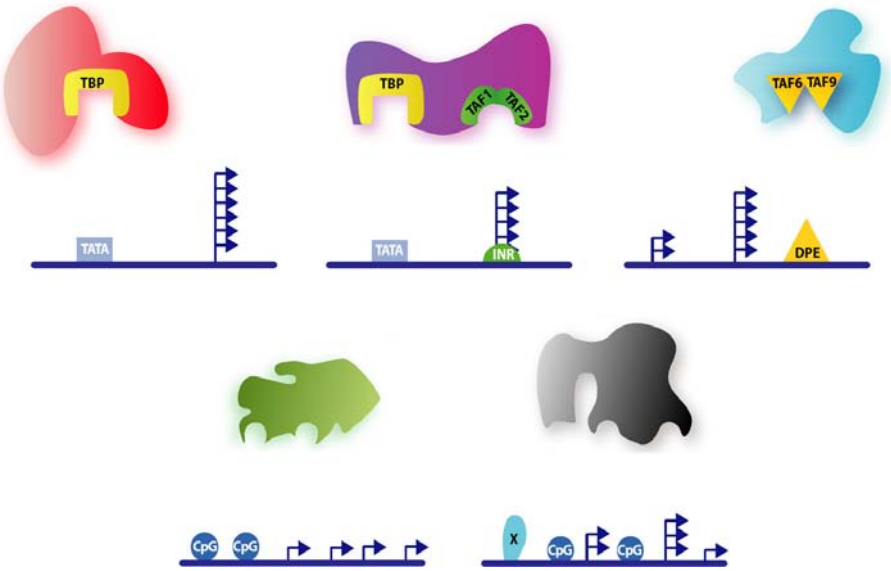


Figure 1.4|A model for core promoter specific protein complex nucleation. Shown are hypothetical core promoters with differential TSS distributions (blue arrows) and occurrence of CPEs or associated features. The distinct core promoters may be bound by a variety of hypothetical alternative protein complexes (as schematized above the promoter diagrams). In the upper panel known CPE-protein interactions are highlighted. Adapted from Muller et al. 2007.

promoter recognition (**Figure 1.3**). Thus, the same core promoter recognition factor can interact differentially with distinct core promoters potentially providing a mechanism of how core promoter architecture can contribute to differential gene activity (Smale and Kadonaga 2003).

The TFIID complex has been considered to have a rather invariant protein composition. However, it could be shown that only a subset of TAFs (TAF4, TAF5, TAF6, TAF8, TAF9, TAF12) is required for the formation of a stable core TFIID complex, while other components (e.g. TAF1, TAF2, TAF11 and TBP) have less contribution to TFIID complex stability. This suggests that the combinatorial addition of non-core components could give rise to a variety of TFIID complex variants (Wright et al. 2006). Indeed, TFIID complexes of different TAF composition have been identified and not all standard TAFs seem to be required equally in

different cell types suggesting tissue-specific functions (Bell and Tora 1999). An example is the murine TAF10, which is required during early mouse embryogenesis, but is dispensable in adult mouse keratinocytes (Mohan et al. 2003; Indra et al. 2005).

Furthermore, several TAF paralogous genes have been identified that are tissue-specifically expressed in several organisms (Hochheimer and Tjian 2003). For example, the paralogs of *Drosophila* TAF4, TAF5, TAF6, TAF8 and TAF12 are expressed specifically in developing spermatocytes within the testis and are required for tissue-specific target gene expression during male gametogenesis (Hiller et al. 2001; Hiller et al. 2004). Examples in mouse include the tissue-specifically expressed TAF4b, which is required for correct oocyte development and function, as well as for maintenance of spermatogenesis (Freiman et al. 2001; Falender et al. 2005a; Falender et al. 2005b).

In addition to the canonical TFIID complex, alternative TAF containing complexes have been identified including the TBP-free TAF containing complex TFTC or yeast SAGA and its human counterpart STAGA. In general, they are composed of several TAFs or TAF paralogs and other co-activator proteins (e.g. Histone acetyl transferases) and are also implicated in transcriptional activation (Thomas and Chiang 2006). For example, TFTC is able to mediate basal and activator dependent transcription on TATA-less and TATA-containing promoters in *in-vitro* assays (Wieczorek et al. 1998; Brand et al. 1999). A further example is the SAGA complex that appears to be involved in the recruitment of TBP to TFIID-independent promoters in yeast. Moreover, it has been shown that 10% of yeast genes, which are

enriched for stress induced genes and associated with TATA-box promoters, are dominantly dependent on SAGA function (Basehoar et al. 2004; Huisinga and Pugh 2004).

In summary, the existence of TAF paralogous genes, tissue specific differences of TAF expression and the existence of TFIID alternative complexes strongly suggests that core promoter recognition contributes to cell-type specific and gene-selective regulation of transcription (**Figure 1.4**). This is further supported by the existence of metazoan specific protein-coding genes encoding homologues to the TATA-box binding protein TBP.

1.3.2. The TATA-box binding protein

TBP is the best studied eukaryotic basal transcription factor. It was first identified as the main component of the TFIID complex mediating contact with the core promoter via interaction with the TATA-box (Kao et al. 1990). Additionally, it was identified to mediate DNA contact in the SL1 complex and the TFIIIB complex involved in transcription of ribosomal RNAs by RNA Polymerase I or small, stable RNAs (e.g. tRNAs) by RNAP III, respectively. Its requirement for transcription by all three eukaryotic RNA Polymerases from yeast to human has led to the notion that TBP is a universal eukaryotic transcription factor (Hernandez 1993; Orphanides et al. 1996).

There is a wealth of knowledge regarding the biochemistry of TBP. In short summary, TBP is a protein of 38 kD (*Drosophila* and human) and can be subdivided into two functional domains. The C-terminus forms a bipartite, saddle like structure consisting of two imperfect repeats and is highly conserved in yeast and mammals with 80% identity on the amino acid

sequence level (**Figure 1.5A**). In contrast, the N-terminal domain of TBP is more variable, while still conserved among vertebrates (Davidson 2003). The concave underside of the saddle shaped core domain is responsible for DNA binding and the C-terminal domain is sufficient to initiate transcription *in-vitro* when combined with the remaining basal transcription factors and RNAP II. TBP binding to the TATA-box of RNAP II dependent promoters leads to severe distortion of the DNA by about 90° and is the first step in PIC assembly followed by the recruitment of the other components of the basal transcription machinery. The convex upside of the core domain interacts with other basal transcription factors and a multitude of regulatory proteins (Nikolov et al. 1992; Burley and Roeder 1996; Nikolov et al. 1996; Orphanides et al. 1996; Davidson 2003; Thomas and Chiang 2006). Despite the extensive literature regarding the biochemistry of TBP less is known about its function and requirement in RNAP II transcription in the context of the living cell and in multicellular organisms.

Several studies have shown that TBP is not universally required for RNAP II dependent transcription in the vertebrate embryo. In *Xenopus* and zebrafish embryos knock-down of TBP by antisense oligonucleotides leads to an early embryonic arrest and an impairment of zygotic transcription concordant with the suggested role of TBP in activating the zygotic genome (see **1.6.2**). However, the majority of transcript levels remain unaffected after loss of TBP and ChIP experiments in *Xenopus* showed that TBP is not recruited to all core promoters of TBP dependent genes that have been examined (Veenstra et al. 2000; Muller et al. 2001; Ferg et al. 2007; Jacobi et al. 2007). In mice, knock-out of TBP is embryonically lethal and embryos arrest and undergo extensive apoptosis at the 30-40 cell stage. However,

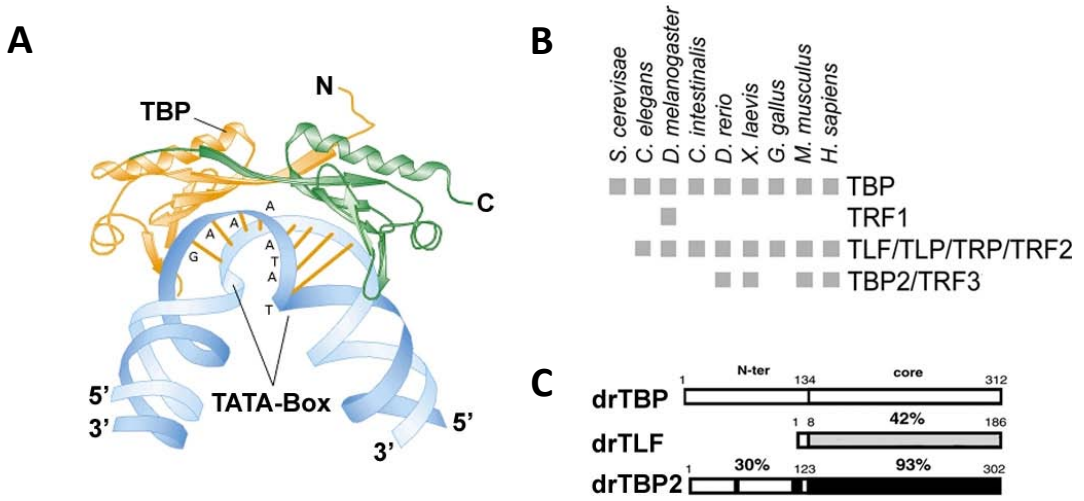


Figure 1.5| TBP and TBP family members. (**A**) Ribbon representation of the conserved C-terminal core domain of TBP bound to TATA-Box DNA. The imperfect repeat structure with the approximate twofold symmetry of the saddle like core domain is indicated (yellow and green halves). Binding of TBP leads to bending of the TATA-Box DNA. Taken from Lodish 2004. (**B**) TBP related proteins in model organisms. Grey boxes indicate genes encoding for TBP family members that can be found in the genomes of yeast and several metazoan species. (**C**) Schematic domain comparison of the three TBP family members described in zebrafish (dr). The percentages indicate amino acid conservation of the N-terminal (N-ter) and core domains of drTBP, drTLF and drTBP2. Numbers indicate amino acid positions. **B** and **C** taken from Bartfai et al. 2004.

cells explanted from these embryos survive for 2-4 days in culture and abundant RNAP II transcription can be detected in these $TBP^{-/-}$ cells, while RNA Polymerase I and III transcription is abolished (Martianov et al. 2002).

Overall, the non-universality of TBP suggests that there must be redundant or complementary mechanisms that can compensate for the loss of TBP. For example, TFIID binding to the core promoter can be mediated independently of TBP function via the interaction of TAFs with CPEs (**Fig. 1.3**). Additionally, it could be shown that the sequence specific transcription factor Sp1 can stabilize TFIID binding in a TBP independent mechanism (Kaufmann and Smale 1994). Furthermore, the existence of alternative promoter recognition

complexes like e.g. TFTC and of several TBP homologues proteins in metazoan animals provides another explanation for the redundancy of TBP (**Figure 1.5B**).

1.3.3. The TBP related factor 1

The TBP related factor 1 (TRF1) has only been described in *Drosophila* and *Anopheles* (**Figure 1.5B**) and there is no evidence for homology in vertebrates (Crowley et al. 1993; Reina and Hernandez 2007). In *Drosophila* it is ubiquitously expressed, while showing higher expression in the nervous system and in the gonads during early embryonic development (Hansen et al. 1997). TRF1 shares 63% homology with TBP within the C-terminal core domain and it can interact with TFIIA and TFIIB, bind to TATA-boxes, and direct RNAP II transcription *in-vitro* (Hansen et al. 1997). *In-vivo*, TRF1 binds together with TFIIA and TFIIB to the upstream alternative promoter of the *tudor* gene, while the downstream promoter is regulated by TBP (Holmes and Tjian 2000). However, TRF1 also forms a stable complex with *Drosophila* BRF1 suggesting a function in RNAP III transcription (Takada et al. 2000). Indeed, genome-wide ChIP-Chip experiments showed that TRF1 does bind to the promoters of few RNAP II dependent promoters, but can be found on RNAP III dependent genes in co-localization with BRF1. Thus, its main function seems to be to replace TBP in RNAP III transcription forming a TRF1-BRF1 complex (Isogai et al. 2007b).

1.3.4. The TBP like factor

The TBP like factor (TLF) can be found in many metazoan species from *C.elegans* to man (Dantonel et al. 1999). It has been described independently by several groups and various

names for this factor can be found in the literature (**Figure 1.5B**): TBP related factor 2 (TRF2) (Rabenstein et al. 1999), TBP related factor (TRF) (Maldonado 1999), TBP like factor (TLF) (Kaltenbach et al. 2000; Veenstra et al. 2000; Muller et al. 2001) and TBP-related protein (TRP) (Moore et al. 1999). The C-terminal core domain of TLF shows about 40-60% homology to TBP; however key amino acids required for TATA-box recognition are not conserved (**Figure 1.5C**). Unlike the highly conserved C-terminal domain of TBP, comparison of TLF core domains among metazoans shows only ~45% homology, suggesting species specific functions of the protein (Dantonel et al. 1999; Muller and Tora 2004).

TLF cannot bind to TATA-boxes but associates with TFIIA and TFIIB *in-vitro*. TLF cannot substitute TBP in *in-vitro* transcription reaction, it rather inhibits TBP dependent transcription potentially by sequestering TFIIA (Moore et al. 1999; Teichmann et al. 1999). In line with these observations, overexpression of TLF in human cells leads to reciprocal regulation of the *NF1* and *c-fos* genes. While the TATA-less *NF1* gene is directly upregulated, the TATA-containing *c-fos* gene is repressed by the overexpression of TLF (Chong et al. 2005). TLF was shown to be part of a complex with the DNA replication-related element binding factor (DREF). Indeed, the DREF/TLF complex was shown to control the DRE-element containing upstream promoter within the tandem promoter of the *Drosophila PCNA* gene (Hochheimer et al. 2002). Another target is the histone H1 promoter that is specifically regulated by TLF, while the core histones (H2A, H2B, H3 and H4) within the tandemly arrayed *Drosophila* histone gene cluster are regulated in a TBP dependent manner. DREF does not colocalize with TLF at the histone cluster suggesting the existence of distinct TLF-containing promoter recognition complexes (Isogai et al. 2007a). Furthermore, genome-wide

ChIP-Chip experiments in *Drosophila* have revealed >1000 binding sites of TLF among which 60% overlap with the 5'-regions of genes. These binding sites are DRE-motif enriched and mostly TATA-less, and the majority (80%) do not overlap with TBP-binding sites indicating that TLF regulates an alternative subset of promoters than TBP (Isogai et al. 2007a).

In *Drosophila* mutation of TLF is embryonically lethal during late embryogenesis and larval and pupal stages (Kopytova et al. 2006). RNAi knockdown of TLF in *C.elegans* revealed that TLF is required for early embryogenesis and zygotic transcription (Dantonel et al. 2000). In *Xenopus* knockdown of TLF by antisense-oligonucleotides leads to an early embryonic arrest and the expression of many genes is affected; embryos develop normally until the onset of zygotic transcription at the midblastula transition (see **1.6.1**) (Veenstra et al. 2000; Jacobi et al. 2007). This is in line with observations in zebrafish where the expression of a mutant isoform causes a severe phenotype accompanied with the lack of expression of several zygotic genes (Muller et al. 2001). In contrast to the severe phenotype in fish and frog TLF is not required during early embryogenesis in mouse as $TLF^{-/-}$ knockout (KO) mice are viable. However, male mice are sterile due to a late arrest in spermiogenesis, concordant with the differential expression of TLF in human and mice with high expression levels in the testis (Martianov et al. 2001; Zhang et al. 2001).

1.3.5. The TATA-box binding protein 2

The TBP related factor 3 (TRF3), or TATA-box binding protein 2 (TBP2), is a vertebrate specific member of the TBP-family (**Figure 1.5B**) and has been initially characterised in human (Persengiev et al. 2003), mouse (Deato and Tjian 2007), *Xenopus* (Jallow et al. 2004)

and zebrafish (Bartfai et al. 2004). TBP2 is more closely related to TBP than TRF1 and TLF with >90% identity between the C-terminal core domains, however, like for all TBP-proteins, the N-terminal domain is not conserved when compared to other members of the TBP-family (**Figure 1.5C**). Within the TBP2 family the C-terminal core domain is highly conserved, while the N-terminal domain is more diverse among vertebrates (Bartfai et al. 2004).

Like TBP and TRF1, TBP2 is able to bind to TATA-boxes. It can interact with TFIIA and TFIIB and it has been shown that TBP2 is able to mediate transcription *in-vitro* (Bartfai et al. 2004; Jallow et al. 2004). Additionally, TBP2 has also been found to be part of a TBP2-TAF3 complex and studies in a mouse muscle cell line have revealed that this complex appears to replace the canonical TFIID complex upon differentiation of myoblasts into myotubes suggesting that terminally differentiated cells employ an alternative core transcription machinery (Deato and Tjian 2007; Deato et al. 2008).

In *Xenopus* and zebrafish, TBP2 is differentially expressed with higher levels of expression in the ventral side of the embryo (Bartfai et al. 2004; Jallow et al. 2004). TBP2 is required for early development in fish and frog as demonstrated by depletion experiments (Bartfai et al. 2004; Jallow et al. 2004; Hart et al. 2007) and it appears to be required for the expression of many zygotic genes in frog (see chapter **1.6.2**) (Jacobi et al. 2007). Additionally, it has been shown that in zebrafish embryos the TBP2-TAF3 complex is required for the initiation of hematopoiesis by regulating a single master regulator gene *mespa* (Hart et al. 2007; Hart et al. 2009).

In *Xenopus* and zebrafish, TBP2 is highly expressed in the gonads, specifically in the ovary, while TBP is undetectable at the protein level (Bartfai et al. 2004; Jallow et al. 2004). Indeed, in *Xenopus* TBP2 replaces TBP in oocyte specific transcription (Akhtar and Veenstra 2009). In apparent contradiction to the uncovered role of TBP2 in induced myogenesis (Deato and Tjian 2007; Deato et al. 2008), other studies have shown that in the mouse TBP2 is exclusively expressed in the ovary, specifically in oocytes (Xiao et al. 2006; Gazdag et al. 2007). Furthermore, KO-mice deficient of TBP2 are viable with no apparent phenotype, however, female TBP2$^{-/-}$ mice are infertile due to defective folliculogenesis (Gazdag et al. 2009).

1.4. Core promoters in combinatorial gene regulation

The high variability of core promoter architecture coupled with tissue specific activity and function of core promoter binding proteins suggest that the core promoter can actively contribute to differential gene expression due to the differential recruitment of components of the basal transcription machinery in various cell types. In addition, another important functional consequence of core promoter diversity is the contribution to combinatorial gene regulation. Especially the selective interaction with other CREs, in particular with distally located enhancers, appears to be influenced by the core promoter identity.

1.4.1. *Cis*-regulatory elements besides the core promoter

A typical metazoan gene is usually associated with a highly structured regulatory DNA that directs the often complex expression patterns. Besides the already discussed core promoter,

these additional cis-regulatory regions include the promoter-proximal regions and distal CREs (**Figure 1.1**). Moreover, many genes are under the control of several autonomous elements that regulate their expression in a modular fashion. Each of these elements can be responsible for the modulation of the core promoter activity in different cell types or at different times, or in response to different stimuli. Thus, the overall expression pattern of a gene is often the cumulative result of the action of several proximal and distal CREs acting independently (Levine and Tjian 2003; Maston et al. 2006).

1.4.1.1. The proximal promoter

The proximal promoter includes the DNA regions upstream of the core promoter, usually extending several hundred bp from the TSS (Butler and Kadonaga 2002). This region often harbours sequence features important for general activation or elevation of transcription. Examples include the GC-Box or CCAAT-Box, which are recognition sites for the general activators Sp1 or CTF (CCAAT-binding transcription factor) and CBF (CCAAT-binding factor), respectively (Bucher 1990; Suzuki et al. 2001b). However, the proximal region may contain regulatory elements targeted by the same set of transcription factors that also acts on distal elements. In fact, proximal regions can harbour enhancer elements; therefore the distinction from other regulatory elements is mainly based on distance and orientation (Heintzman and Ren 2007). Analysis of extended human promoter regions in transgenesis assays has revealed that the regions upstream of the core promoter can often be subdivided into regions with activating and repressing function. While sequences located -100 to -350 to the TSS often contribute positively to core promoter activity, regions from -500 to -1000 are

frequently associated with regulatory information that negatively affects transcription initiation (Cooper et al. 2006). Additionally, proximal promoter regions may contain proximal promoter-tethering elements or promoter targeting sequences, which both are involved in the specific recruitment of cognate enhancers by protein binding that mediates enhancer contact (Zhou and Levine 1999; Calhoun et al. 2002).

1.4.1.2. Enhancers and Silencers

Transcriptional enhancers have been initially described as *cis*-regulatory sequences that can modulate the expression of their target gene in an orientation and distance independent manner (Banerji et al. 1981). In metazoans, a typical enhancer is up to several hundred bp long and usually consists of a cluster of transcription factor binding sites (TFBS), which are small sequence motifs of 6-12 bps length. TFBSs serve as recognition sites for sequence specific transcription factors that work cooperatively to enhance target gene transcription (Maston et al. 2006). Within an enhancer the organization of TFBS can be critical, thus the orientation and distance independence only applies for an enhancer as a whole (Remenyi et al. 2004). The regulatory regions harbouring the enhancer elements can be located up- or downstream or even within introns of their target genes (Levine and Tjian 2003). Therefore, the distance to the target core promoter is highly variable and can range from a few hundred bp to many kb, and in extreme cases the enhancer can be located up to a 1 Mb away (Lettice et al. 2003). Enhancers, are generally located in *cis* to their target, although *trans* regulation has been described in *Drosophila* (Morris et al. 1998).

Silencers are CREs that are targeted by transcription factors with a negative effect on transcription, thereby blocking or dampening the target gene activity in certain cell-types. Generally, they share many of the features of enhancers like the positioning in relation to the target gene and the modular organization. Silencers can occur as independent CREs or they can be part of a distal enhancer (Ogbourne and Antalis 1998). In fact, enhancer bound activator proteins can switch to a repressor by cell-type specific recruitment of co-regulators (Maston et al. 2006).

1.4.1.3. Insulators

Insulators or boundary elements are *cis*-regulatory regions that block genes from the transcriptional activity of neighbouring loci. Insulators have two main functions: First, they can block enhancer-promoter communication, thus shielding promoters from the influence of heterologous CREs, and second they can block the spreading of repressive chromatin conformations, i.e. heterochromatin (Heintzman et al. 2009). The exact mechanism how enhancer blocking activity is mediated is unknown. Although several proteins are described which mediate insulator function in *Drosophila* (West et al. 2002), in vertebrates only the binding of CTCF (CCCTC binding factor) could be associated with insulator function. It is suggested that CTCF functions in establishing chromatin conformations that disfavour certain enhancer-promoter interactions (West and Fraser 2005). Interestingly, CTCF binding sites are occupied in various cell types, suggesting a cell-type independent function of insulators (Kim et al. 2005). Insulator function has also been attributed to matrix-associated regions (MARs), which are non-coding DNA elements that mediate the attachment of

chromatin loops to the nuclear matrix generating separated structural domains (Glazko et al. 2003).

1.4.2. Transcription factors and enhancer activity

Transcription factors (TFs) are sequence-specific DNA binding proteins that mediate the correct spatiotemporal gene regulatory activity of proximal and distal CREs. TFs usually have a modular structure consisting of a DNA-binding domain (DBD), domains important for protein-protein interaction and transactivation (activators) or repression domains (repressors), which are required to stimulate or inhibit transcription upon DNA binding, respectively (Kadonaga 2004). TFs physically interact with TFBS within CREs through the DBD and this binding modulates target gene transcription by controlling the transcription initiation rate at the core promoter. There are thousands of TF genes in metazoan genomes and it is estimated that 5-10% of all protein-coding genes encode for TFs highlighting the importance of these regulatory proteins that control the transcription of most if not all genes (Wilson et al. 2008; Martinez and Walhout 2009). Numerous DBDs with specific DNA binding properties have been identified that allow grouping TFs with similar DBDs into families. While there are many distinct TF families, there is an overrepresentation of certain DBD classes in the human genome: cysteine rich zinc-finger (C2H2 ZF), homeo box, helix-loop-helix (HLH), basic leucine zipper (bZIP), and forkhead (Tupler et al. 2001; Wilson et al. 2008; Vaquerizas et al. 2009). Usually, TFs can bind to a range of similar TFBS and the exact sequence can influence binding affinity; therefore the consensus sequences of TFBS are often highly degenerate making the prediction of functional TFBS difficult as the short motifs

can occur by chance at high frequency. Moreover, distinct TFs of certain families can bind to similar TFBS suggesting overlapping activities (Martinez and Walhout 2009; Portales-Casamar et al. 2009).

TFs are usually pleiotropic as a single TF can control many target genes. Moreover, certain transcription factors can serve as master regulators (e.g. PAX6) controlling directly or indirectly a multitude of genes and therefore the expression of a single TF can change the entire transcriptional program of a cell or cell lineage. However, the vast majority of genes are controlled not only by a single, but by a combination of TFs that bind to the target CREs and control the processes leading to transcriptional modulation in an often cooperative manner (Hobert 2008). At the posttranslational level the activity of TFs can be modulated by several mechanism including covalent modifications (e.g. phosphorylation) of TFs, ligand/co-factor binding or selective interaction with other TFs. Also TFBS accessibility, controlled by chromatin structure, plays an important role. The cell type specific activity of TFs is often dependent on the cell type specific expression of TFs themselves, which is again controlled by more upstream regulatory levels. Furthermore, TFs can regulate each other in so-called feedback and feedforward loops, which together with the hierarchical organisation of TFs and targets (and thus CREs and promoters) form a transcriptional regulatory network that directs the correct spatiotemporal transcription of all genes (Hobert 2008; Martinez and Walhout 2009).

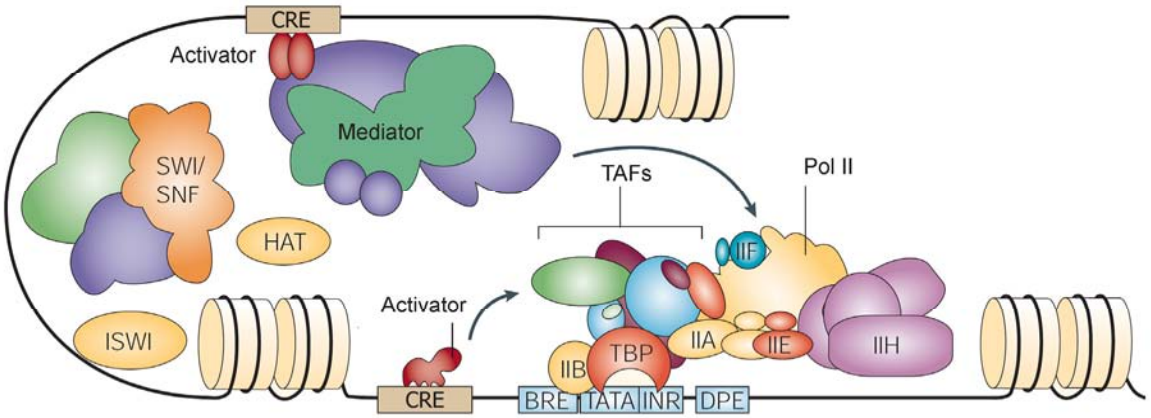


Figure 1.6| A model of activator mediated initiation of transcription. The core promoter is bound by the PIC consisting of the basal transcription factors TFIIA, -IIB, -IIE, -IIF, IIH and TFIID. TFIID is composed of several TAFs and TBP. The PIC receives stimulatory input from CRE bound activators either via direct or indirect protein-protein interactions. Indirect interactions are exemplified by the mediator complex. Additionally, the chromatin-remodelling complexes ISWI and SWI/SNF as well as a HAT (histone acetyl transferase) are shown. Taken from Taatjes et al. 2004.

1.4.2.1. Activator mediated initiation of transcription

Transcription factors modulate target gene activity upon response to cellular or environmental signals by interpreting and transmitting the encoded regulatory information to the basal transcription machinery. As a result, the binding of activators to CREs stimulates PIC assembly and initiation of transcription at the core promoter (**Figure 1.6**).

In a simplified model for transcriptional activation, activator binding leads to the displacement and modification of nucleosomes at the core promoter by the recruitment of ATP-dependent chromatin remodelling complexes (e.g. SWI/SNF, ISWI) and histone-modifying factors (e.g. histone acetyltransferases CBP/p300, SAGA or PCAF), which together render the chromatin at the core promoter accessible for the binding of the basal transcription factors (Cosma 2002; Kadonaga 2004). Promoter-enhancer communication is

mediated in part by direct protein-protein interactions of activators with the basal transcription machinery (e.g. TFIID via TAFs) (Thomas and Chiang 2006). Additionally, activators can recruit co-activators that indirectly transmit signals from the enhancer to the basal transcription factors. One of the best known co-activators is the Mediator, a multi-protein complex without intrinsic sequence-specific DNA binding capability that is conserved from yeast to man (**Figure 1.5**) (Taatjes et al. 2004; Conaway et al. 2005).

It must be noted, that the outlined model is an oversimplification as the order of events and the exact involvement of factors can be highly variable at different gene loci (Cosma 2002; Taatjes et al. 2004). However, it illustrates that it is necessary to establish contacts between activator proteins and the basal transcription machinery at the core promoter. As the enhancer-promoter distance can be considerably large such interactions must be established over long chromosomal distances. Several models have been proposed to explain how this is achieved. The scanning model proposes that activators bind to an enhancer and then they move along the DNA until they have found their cognate promoter (Blackwood and Kadonaga 1998). Another model suggests that the enhancer directly contacts the core promoter, involving mechanisms that loop out the intervening non-participating DNA. Indeed, chromosome capture conformation (3C) techniques could show that enhancers co-localize and thus loop with the core promoter DNA (Carter et al. 2002). The establishment of a loop conformation is achieved by either free diffusion of the enhancer through the nucleoplasm or the enhancer is guided along the intervening DNA, a mechanism named facilitated tracking (Blackwood and Kadonaga 1998; Zhu et al. 2007). It has been shown that also the target gene locus can loop out from its chromosome territory facilitating enhancer-

promoter interaction. Interestingly, in the case of the ZRS enhancer, which drives *Shh* (*sonic hedgehog*) expression in the zone of polarizing activity (ZPA) of the developing limb bud (Lettice et al. 2003), the looping out of the target locus appears to be required for transcriptional activation, whereas enhancer-promoter interaction establishes transcriptional competence (Amano et al. 2009). This is in line with models proposing that actively transcribed genes often co-localize with focal concentration of RNAP II within the nucleus, so-called 'transcription factories' (Osborne et al. 2004). Additionally, it has been shown that enhancers can also recruit components of the basal transcription machinery (Szutorisz et al. 2005). In the case of the human ε-globin gene, the HS2 enhancer-complex contains TBP and RNAP II which are then guided to the core promoter by facilitated tracking (Zhu et al. 2007).

1.4.3. Enhancer-promoter interaction specificity

Enhancers are not only located far away from their cognate target promoter, they can also be located in proximity or even within introns of unrelated 'bystander' genes (Lettice et al. 2003; Kikuta et al. 2007). An example is the *ZRS* enhancer of the *Shh* gene. In the mouse genome, this enhancer is located around 1 Mb away within intron 5 of the unrelated lmbr1 gene with additional unrelated genes in the intervening DNA (Lettice et al. 2003). In addition to long distance regulation, there could be scenarios were an enhancer must activate only one of several core promoters in its immediate vicinity (e.g. alternative promoters, adjacent loci). Therefore, an enhancer can be exposed to several core promoters indicating that mechanisms directing an enhancer to its target core promoter must exist. In part, this is

mediated by the influence of other CREs like insulators or proximal promoter-tethering elements, which provide directionality or specificity, respectively (Burgess-Beusse et al. 2002; Calhoun et al. 2002). However, also the core promoter identity plays an important role in restricting the stimulatory capacity of enhancers.

1.4.3.1. Core promoter architecture and enhancer-promoter specificity

Early evidence for enhancer-promoter interaction specificity comes from experiments in *Drosophila* showing that certain enhancers cannot interact with heterologous promoters of nearby genes (Li and Noll 1994; Merli et al. 1996). These observations are complemented by experiments in *Drosophila* analysing the ability of enhancers to activate either TATA- or DPE-dependent core promoters. The AE1 or IAB5 enhancers exhibit a preference for the TATA-containing *even-skipped* promoter relative to the TATA-less DPE-containing *white* promoter when linked to both promoters in a reporter construct. However, both enhancers are fully competent to activate the DPE-dependent *white* promoter when blocked from interacting with the *even-skipped* promoter by an insulator indicating core promoter competition (Ohtsuki et al. 1998). In another study, the production of *Drosophila* sister lines that carry either a TATA-containing or a DPE-containing promoter-reporter transgene at exactly the same chromosomal position resulted in 18 promoter insertions which trapped enhancers. The analysis of the lines yielded 3 DPE specific, 1 TATA-specific and 14 non-specific enhancers, thus confirming enhancer-promoter specificity in a chromosomal context (Butler and Kadonaga 2001). In line with these result is the analysis of the *Drosophila Hox* gene cluster. The core promoters of nearly all *Drosophila Hox* genes contain a conserved and

functionally important DPE element. The TF Caudal is a key regulator of the *Hox* gene cluster and it could be shown that activation of *Hox* core promoters by a Caudal-dependent enhancer requires the DPE element. In addition, also other DPE dependent promoters show that functional link. While Caudal also mediates weak activation from TATA-containing promoters, the presence of an additional BREu element completely blocks transcriptional activation. In DPE dependent promoters the insertion of a BREu element has no significant effect on transcriptional activation; therefore the BREu element can specifically suppress Caudal-dependent activation in TATA-containing promoters (Juven-Gershon et al. 2008a). Together, these result underline that the diversity of core promoter motif compositions can play an important role in combinatorial gene regulation.

1.4.3.2. Promoter features of targets of long-range regulation

Genomic regulatory blocks (GRBs) are large chromosomal regions found in *Drosophila* and vertebrate genomes with preserved synteny across insects or vertebrates, respectively, indicating an evolutionary constraint on the architecture of these regions (Engstrom et al. 2007; Kikuta et al. 2007). GRBs are characterized by the presence of highly-conserved non-coding elements (HCNE), which frequently function as long-range enhancers, unrelated bystander genes and likely target genes, which are often developmental genes with complex spatiotemporal expression patterns. The analysis of GRBs has revealed some further features of potential target promoters of distal enhancers that distinguish them from heterologous promoters. In *Drosophila*, the bioinformatic examination of core promoters of developmental transcriptional regulators within GRBs has shown that they are highly

enriched for the Inr. This enrichment of the Inr (with or without DPE) is also seen when analyzing all *Drosophila* core promoters that fall into the developmental gene classification, suggesting a role of this CPE in enhancer-promoter specificity as this gene class is often targeted by long-range elements (Engstrom et al. 2007). In vertebrate genomes, the distinction of bystander and target promoters is not as apparent as both are associated with CpG islands and display a broad distribution of TSSs (Carninci et al. 2006; Akalin et al. 2009). However, likely target genes are often located within multiple large CpG islands, as compared to bystander genes. Additionally, target promoters have a complex architecture with multiple CAGE-tag clusters (potential alternative promoters) within a 4kb window around the most prominent TSS, suggesting the existence of several enhancer-promoter pairings (Akalin et al. 2009). While these studies do not functionally address the contribution of the core promoter to enhancer-promoter specificity, they strongly suggest that core promoter architecture plays an important role in long-range gene regulation.

1.5. Cis-regulatory elements and human disease

Transcription factors, CREs and the components of the basal transcription machinery control nearly every aspect of transcriptional regulatory processes. Dysregulation of transcription, caused by aberrant function of regulatory elements and factors involved is often associated with human disease and cancer. While there is considerable knowledge about the pathological role of coding mutations, less is known about the contribution of CRE mutations to human disease (Kleinjan and van Heyningen 2005; Maston et al. 2006).

Human core promoters can be substantially polymorphic as shown by the comparison of a set of human promoter sequences derived from several individuals (Hoogendoorn et al. 2003). This analysis revealed single nucleotide polymorphisms (SNP) in approximately 35% of the analysed promoters. Functional testing of a subset of these promoters in transgenic cell culture assays has shown that one third of the variants give rise to altered reporter gene activity (Hoogendoorn et al. 2003; Buckland et al. 2005). Interestingly, the location of functional SNPs is biased to positions within 100 bp distance to the TSS, while non-functional SNPs are distributed evenly over the promoter region. This bias suggests that polymorphism of the core and to a lesser extent the proximal promoter can substantially contribute to altered gene expression *in-vivo* contributing to phenotypic variation and disease susceptibility (Buckland et al. 2005). One of the relatively rare well understood examples of how core promoter variation contributes to human disease are mutations causing nucleotide variants within the DCE (**Figure 1.3**). This CPE has been identified in individuals with β-thalassemia, a disease in which the downregulation of β-globin leads to precipitation of the excess α-globin in erythrocytes inhibiting their maturation. Mutations at either +22 or +33 in relation to the TSS lead to a decreased expression of β-globin by potentially reducing the interaction of the DCE with TAF1 (Lewis et al. 2000; Lee et al. 2005).

Several mostly congenital diseases are associated with aberrant function of more distal CREs (Maston et al. 2006). These disorders are caused by disruption or impairment of the enhancer-core promoter communication due to mutation, physical dissociation or changes in chromatin structure (Kleinjan and van Heyningen 2005). There are many known examples where the mutation of TFBSs in the proximal promoter leads to aberrant gene expression

(Maston et al. 2006; Laurila and Lähdesmäki 2009). For example the analysis of the amyloid precursor protein gene (*APP*), which plays a role in Alzheimers disease (AD), has identified 3 mutations in the proximal promoter that are associated with early onset of AD. The mutations abolished (AP-2 and HES-1) or created (Oct-1) putative TFBS, respectively. The promoter variants showed increased activity when tested in transgenic reporter gene assays in line with the higher *APP* mRNA levels in AD brains of some patients (Theuns et al. 2006). Another example is the mutation of a GATA-1 binding site in the proximal promoter of the δ-globin gene, which reduces GATA-1 binding affinity and is associated with δ-thalassemia (Matsuda et al. 1992).

The pathological role of mutations of more distal or long-range elements is much less explored, mainly due to difficulties in identifying cognate enhancers and to some extent in identifying enhancers themselves (Visel et al. 2009). Physical dissociation of a long-range enhancer and promoter can be seen in a number of aniridia cases, a disease characterized by the absence of the iris and other related eye anomalies, which is normally caused by haploinsufficency of PAX6, a paired-box/homeodomain transcription factor and master regulator of oculogenesis. In several patients, coding mutations could not be detected, instead chromosomal rearrangements downstream of PAX6 have been observed (Fantes et al. 1995; Kleinjan et al. 2001). Functional analysis of the rearranged chromosomal regions identified a long-range regulatory element that controls PAX6 expression and is normally located 200 kb downstream within the last intron of the unrelated bystander gene *ELP4* (Kleinjan et al. 2001). Another example is the *SHH* gene, which is controlled by a complex array of proximal, intronic and long-range enhancers (Epstein et al. 1999; Muller et al. 1999;

Lettice et al. 2003; Jeong et al. 2006) and encodes for a secreted signalling protein involved in many developmental processes. Haploinsufficency of *SHH* is associated with holoprosencephaly (HPE) (Belloni et al. 1996). More recently, it has been shown that also a rare nucleotide variant within the *SBE2* enhancer, which drives *SHH* expression in the hypothalamus and is located around 460 kb upstream, is associated with HPE. Interestingly, the TF Six3, known to be associated with HPE, displays reduced binding affinity to the mutant enhancer (Jeong et al. 2008). The SBE2 is not the only SHH enhancer associated with disease. Mutations of the SHH ZRS enhancer (described in **1.4.3**) lead to dysregulation of SHH in the developing limb bud, which causes e.g. polydactyly or triphalangeal thumb (Lettice et al. 2003; Sun et al. 2008).

1.6. The maternal to zygotic transition

The early development of most animal model systems including nematodes, insects, fishes and amphibians is characterized by rapid and mostly synchronous cell cycles or cleavages that subdivide the zygote into a population of blastomeres or a syncytial blastoderm. In mammals initial cell cycles are asynchronous and considerably longer (O'Farrell et al. 2004). Common to most animals is that this early period is mainly controlled by maternal gene products that have been deposited in the egg during oogenesis (Tadros and Lipshitz 2009). In contrast, zygotic genes are almost exclusively transcriptionally silent during this early phase. Extensive transcriptional activation of the zygotic genome occurs mostly in later cell cycles, while the exact cycle number varies considerably between species (O'Farrell et al. 2004). Activation of zygotic transcription is often paralleled by loss of synchrony and elongation of

cell cycles and the onset of differentiation processes (Newport and Kirschner 1982a; Newport and Kirschner 1982b; Kane and Kimmel 1993). Furthermore, zygotic genome activation coincides with the controlled degradation of maternally inherited mRNAs (**Figure 1.7**) (Mathavan et al. 2005; Giraldez et al. 2006; De Renzis et al. 2007). The transition from maternally controlled to zygotically controlled embryonic development is often synonymously termed mid-blastula transition (MBT) or maternal-to-zygotic transition (MZT).

1.6.1. Regulation of transcription and mRNA levels at the MZT

Several mechanisms have been proposed that could control the delayed activation of zygotic transcription during the MZT. During early cell cycles or cleavages the number of nuclei and thus the amount of DNA/chromatin is continuously increasing, while the volume of cytoplasm remains constant. Experiments with polyspermic *Xenopus* embryos which have a modified nucleocytoplasmic ratio have an altered onset of MBT (Newport and Kirschner 1982b; Newport and Kirschner 1982a). Similarly, the zebrafish maternal mutant *futile cycle*, which undergoes anucleate cleavages due to defective karyogenesis, shows an earlier onset of MBT in the giant nuclei carrying cells (Dekens et al. 2003). These examples demonstrate further the validity of the classic observations by Newport and Kirschner et al 1982 suggesting that the nucleocytoplasmic ratio plays a key role in the delayed onset of zygotic transcription by potentially titrating chromatin components (Prioleau et al. 1994) or maternally inherited transcriptional repressors present in the cytoplasm. The identity of such repressors remains elusive in most animal models (Schier 2007). However, experiments in *Xenopus* have shown that depletion of the maternally provided cytosine methyltransferase

xDnmt1 in pre-MBT embryos leads to premature transcription. The repressive function appears to be independent from its catalytic activity indicating methylation independent repression of target genes (Dunican et al. 2008). Another model suggests that the rapid cell cycles during pre-MBT prohibit zygotic transcription. Indeed, it was observed that mitosis can abort nascent transcripts (Shermoen and O'Farrell 1991). Additionally, *Xenopus* embryos (Kimelman et al. 1987) and *Drosophila* embryos (Edgar and Schubiger 1986) with experimentally induced cell cycle lengthening display earlier onset of transcription. These observations are in concordance with the early onset of transcription in mouse, where the first cell cycles are comparably long (O'Farrell et al. 2004; Schier 2007). The timing of the onset of zygotic transcription could also be controlled by the activity or availability of components of the basal transcription machinery. Indeed, several studies have addressed the role of general transcription factors in activating zygotic transcription, especially the role of TBP family members during MBT has been analysed (see **1.6.2**).

As mentioned above the activation of zygotic transcription at the MZT is paralleled by the degradation of maternal mRNAs that potentially interfere with later development. In *Drosophila* the RNA-binding protein Smaug has been identified to bind to *cis*-elements located within the ORF of the maternally deposited *hsp38* mRNA leading to its deadenylation and subsequent degradation (Semotok et al. 2005; Semotok et al. 2008). In zebrafish the zygotically expressed microRNA *miR-430* has been found to be responsible for the clearance of several hundred maternally inherited transcripts around MBT. *miR-430* binds to complementary sequence motifs in the 3'UTR of target mRNAs promoting their deadenylation and degradation. The expression of *miR-430* is initiated at the MBT, thus

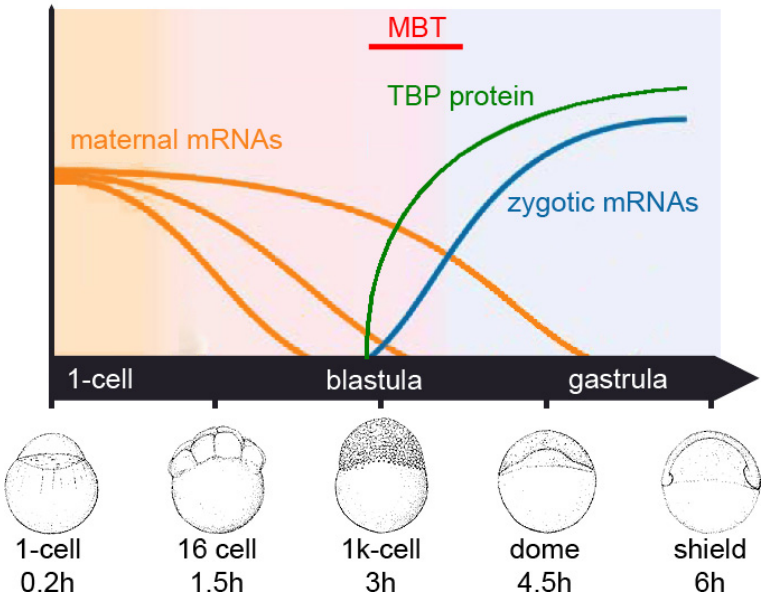


Figure 1.7| Transcript dynamics during early zebrafish development. Schematic illustration of the differential degradation of maternal mRNAs (orange) and the activation of zygotic transcription (blue) during zebrafish MBT (red bar). Rising TBP protein levels coinciding with MBT are depicted in green. On the x-axis the time post fertilization and the corresponding developmental stages are indicated. Modified after Schier 2007.

directly linking zygotic genome activation to clearance of maternal mRNAs (Giraldez et al. 2006).

1.6.2. The role of TBP-family members in the mid-blastula transition

The TBP protein is not expressed in pre-MBT *Xenopus* and zebrafish embryos, but TBP mRNA is deposited in the egg maternally. At MBT the maternal TBP mRNA is translated into protein, thus rising TBP protein levels coincide with the activation of zygotic transcription (**Figure 1.7**) (Veenstra et al. 1999; Muller et al. 2001). Similarly to frog and fish, TBP levels increase in the mouse embryo around the onset of zygotic transcription (Gazdag et al. 2007).

First evidence for the role of TBP-family members in MZT came from studies analyzing the requirement and function of TBP during MBT in *Xenopus*. When TBP is pre-bound to a

promoter construct prior to injection in *Xenopus* embryos precocious, but transient, pre-MBT transcription can be observed (Prioleau et al. 1994). The co-injection of relatively large amounts of non-specific DNA leads to stable transcription potentially due to titration of repressive components in the pre-MBT embryo (Prioleau et al. 1994). Moreover, *Xenopus* egg extract substituted with TBP directs transcription from a reporter construct *in-vitro* and precocious translation of TBP-mRNA in pre-MBT embryos can activate transcription (Veenstra et al. 1999). Overall, these findings suggest that TBP protein levels are rate-limiting for transcription during early embryogenesis. Indeed, the inhibition of maternal TBP-mRNA translation in *Xenopus* and zebrafish embryos leads to a severe phenotype characterized by an early embryonic arrest and impaired zygotic transcription. However, this requirement for TBP is not universal as not all zygotic genes are affected and it could be shown that the TBP-family members TLF and TBP2 are also involved in the activation of zygotic transcription in fish and frog (Dantonel et al. 2000; Veenstra et al. 2000; Muller et al. 2001; Bartfai et al. 2004; Jallow et al. 2004). Similarly, TBP function is not universally required for zygotic transcription in mouse (Martianov et al. 2002).

Microarray analyses of transcript levels of TBP, TLF and TBP2 depleted *Xenopus* embryos have revealed that 69% of transcript levels remain unaffected after TBP-knockdown, whereas the knockdown of TLF or TBP2 has a major impact on zygotic transcription, as 74% or 60% of transcript levels are downregulated, respectively (Jacobi et al. 2007). Metaanalysis of the microarray data using gene orthology has revealed that the TBP-family members are required for different subsets of genes. While TBP is mainly required for the regulation of frog genes that are often orthologs of yeast genes, TLF and TBP2 targets are enriched for

metazoan- or vertebrate specific genes. Moreover, gene ontology analysis has shown that TBP targets are enriched for transcripts expressed maternally and in the adult. In contrast, TLF targets are enriched for embryonically expressed transcripts linked to catabolism and TBP2 is linked to vertebrate-specific embryonic genes (Jacobi et al. 2007). Similarly, microarray expression profiling after TBP depletion in zebrafish embryos has revealed that the majority (65.3%) of transcript levels remain unaffected (Ferg et al. 2007). Only 17.5% of the analysed transcripts show a significant reduction of their expression levels. These downregulated genes are significantly enriched for developmental stage specific genes potentially in line with the enrichment of TATA-containing promoters among highly regulated tissue specific genes (Carninci et al. 2006; Ferg et al. 2007). However, no statistically significant correlation between TATA-box occurrence and TBP-dependence could be established due to the low abundance of the motif in the zebrafish genome (Ferg et al. 2007; Ferg 2008). A surprising result of the zebrafish study is the large proportion of transcripts that are upregulated after TBP knockdown (17.1%), which are enriched for constitutively expressed genes. Comparison of the upregulated gene set reveals that there is substantial overlap (52%) with maternally inherited transcripts providing a link between TBP dependent zygotic transcription and degradation of maternal mRNAs. Strikingly, 63% of these TBP-dependent degrading transcripts are also targets of the miR-430 dependent pathway responsible for the clearance of a large number of maternally inherited mRNAs (Giraldez et al. 2006). The fact that miR-430 is still expressed in TBP depleted zebrafish embryos indicates that TBP dependent transcription is required for processes acting downstream of miR-430 transcription in the degradation pathway (Ferg et al. 2007).

Taken together, the above mentioned studies show that TBP-family members have acquired specific functions during early vertebrate embryogenesis concordant with the metazoan or vertebrate specific diversification of the TBP-family proteins. Moreover, it shows that core promoter recognition proteins, long being considered as static components in the regulation of gene expression, are involved in the specific regulation of subsets of genes and play specialised roles in differential gene expression during vertebrate embryogenesis.

1.7. Zebrafish as a model organism

The zebrafish *Danio rerio* is a small subtropical freshwater fish that was initially selected as a model to study developmental biology and genetic processes underlying embryogenesis due to several key embryological and experimental advantages (Dahm and Geisler 2006). Zebrafish are easy to maintain, have a short generation time (3-4 months) and under laboratory conditions they breed throughout the year. They are highly fertile and a pair of zebrafish can produce 100-300 eggs weekly allowing the generation of thousands of embryos in a relatively short period of time. The embryos develop *ex-utero* making them readily accessible to experimental manipulation (e.g. microinjection) right after fertilization. Additionally, early embryonic and larval stages are optically transparent allowing the observation of developmental processes inside the living embryo. Moreover, the transparency facilitates the usage of fluorescent dyes or proteins allowing e.g. lineage tracing experiments or the tracking of reporter gene expression. Zebrafish embryos develop very rapidly in comparison to other vertebrate model organisms: e.g. 24 hours post fertilization (hpf) many organ rudiments are in place and 120 hpf the larvae start feeding

indicating that embryogenesis is complete (Kimmel et al. 1995; Westerfield 2000) Importantly, zebrafish embryos are vertebrates therefore they offer developmental processes and a body plan very similar to other vertebrates including mouse and human. The zebrafish genome consists of 25 chromosomes and has a haploid size of $1.5*10^9$ bp and encodes for approximately 20.000 genes (Dahm 2005). It is sequenced and well annotated, facilitating gene discovery and identification of CREs. This is complemented by a large database of EST and mRNA sequences (http://www.ensembl.org/Danio_rerio) and a comprehensive database of expression patterns at different developmental stages (http://zfin.org). Additionally, a database of transcript dynamics (e.g. http://zf-espresso.tuebingen.mpg.de) at different developmental time points derived from microarray expression profiling experiments is available. Importantly, coding and frequently also non-coding sequences are conserved between zebrafish and other vertebrate species allowing comparative genomics and underlining the relevance of the zebrafish as a model organism (Dahm and Geisler 2006; Yang et al. 2009).

The zebrafish embryo is amenable to several genetic manipulations. For example, forward genetic screens using chemical or insertional mutagenesis are widely employed and have led to the identification of thousands of characterized mutations that affect development, physiology and behaviour as well as several mutants of components of the basal transcription machinery (Driever et al. 1996; Haffter and Nusslein-Volhard 1996; Amsterdam et al. 2004). The random insertion of mutations is complemented by methods to disrupt the function of a gene of interest, e.g. gene-knockdown by morpholino antisense-oligonucleotides (Ekker and Larson 2001). The screening of randomly generated mutant

libraries for a mutation of interest (TILLING) is another reverse genetics approach successfully used (Wienholds et al. 2003; Dahm and Geisler 2006). Until recently, targeted knock-out in zebrafish has been unsuccessful, but the development of specifically engineered zinc finger nucleases that introduce genomic lesions at specific sites has made the targeted generation of mutants feasible (Meng et al. 2008).

1.7.1. Studying gene regulatory function by transgenesis in zebrafish

The zebrafish embryo provides an optimal model system for the analysis of CREs. Especially the transparency and the external development of the embryos in combination with a range of transgenic methods available allow the functional characterization of CREs in the context of a living vertebrate embryo.

Testing of reporter constructs, in which a reporter gene is under the influence of CREs, can be done transiently. Transient transgenesis is accomplished by injecting plasmids or BACs into the cytoplasm of fertilized eggs (Westerfield et al. 1992; Yang et al. 2006). Alternatively, putative CREs can be co-injected in conjunction with a minimal promoter-reporter construct, which then concatemerize in the embryo (Muller et al. 1997). The injected DNA initially replicates during early cleavage stages. During gastrulation the foreign DNA gets mostly degraded and the remainder is randomly integrated in the host genome leading to transient transgenic embryos that express the reporter gene in a highly mosaic manner. Mosaicism is potentially caused by the uneven distribution of reporter constructs during early cleavages combined with cell-type specific activation of transgenes (Stuart et al. 1988; Westerfield et al. 1992). The mosaic reporter expression is a drawback of this technique, but it can be

overcome by compiling the reporter gene expression from several embryos leading to a cumulative expression pattern that reflects the overall activity of the regulatory elements tested (Westerfield et al. 1992; Muller et al. 1997; Muller et al. 1999). Also the usage of Tol2 transposon mediated transgenesis has been reported to reduce mosaicism in the F0 generation (Fisher et al. 2006). Transient analysis in zebrafish is fast and efficient, and has been extensively used in the functional analysis of predicted and known zebrafish and non-zebrafish CREs. Moreover, the ease of microinjection and the large number of eggs available allow the generation of hundreds of transient transgenic embryos in parallel, facilitating the analysis of CRE function on a larger scale (Barton et al. 2001; Dickmeis et al. 2004; Woolfe et al. 2005; Sanges et al. 2006; Jarinova et al. 2008).

In addition to the transient approach, CREs can be assayed after stable integration of DNA into the germline leading to stable transgenic fish with non-mosaic reporter gene expression in the F1 generation (Stuart et al. 1988; Stuart et al. 1990). The method most frequently used to generate stable transgenic lines is the injection of plasmids or BACs into zebrafish embryos, which randomly integrate into the genome. However, the frequency of stable integration into the germline is relatively low; approximately 5-20% of fishes show germline transmission of the transgene (Amsterdam and Becker 2005). The integration efficiency can be elevated by the co-injection of the meganuclease I-Sce I (Grabher et al. 2004) or by retrovirus or transposon assisted transgenesis (Gaiano et al. 1996; Kawakami et al. 2000). Due to the lack of mosaicism, stable transgenics recapitulate the regulatory activity of the tested regulatory elements more faithfully than transient transgenics, but the approach is

generally more laborious and of lower throughput as the transgene activity is usually assessed in daughter generations.

Another transgenic application in zebrafish is the random insertion of gene trap or enhancer trap vectors to generate stable transgenic lines expressing a reporter gene under the regulatory control acting at the integration site (Amsterdam and Becker 2005). These trap approaches have been successfully used to screen the zebrafish genome for CREs and tissue-specifically regulated genes as well as to dissect the *cis*-regulatory logic of genomic loci (Kawakami et al. 2004; Ellingsen et al. 2005; Kikuta et al. 2007). Additionally, these trap lines can serve as a starting point for other studies (e.g. phenotyping, drug screening) as they often represent tissue-markers allowing the visualization of subpopulation of cells in the living zebrafish embryo (Amsterdam and Becker 2005).

1.7.2. Large-scale screening approaches

The zebrafish embryo is an ideal model system to be used in large scale screening approaches as it combines the complexity of a vertebrate embryo with features needed for high-throughput screening (HTS). Especially, the large number of embryos available combined with their small size (approximately 1mm diameter) facilitates HTS approaches as the embryos are small enough to be arrayed in multi-well plates (Kari et al. 2007). Such large scale screens do not only include the already discussed mutagenesis and transgenesis screens but also approaches like toxicological screens (Yang et al. 2009), chemical or drug discovery screens (Zon and Peterson 2005) and behavioural screens (Rihel et al. 2010). The HTS potential of the zebrafish embryo is further enhanced when combined with the plethora

of mutants and transgenic lines available. For instance, zebrafish *gridlock* mutant embryos display an abnormal formation of the dorsal aorta that prevents blood circulation to the trunk resembling human aortic coarctation (Zhong et al. 2000). This mutant has been successfully used to screen a library of 5,000 chemicals leading to the identification of two compounds that can rescue the phenotype, potentially by activating the VEGF pathway (Peterson et al. 2004; Zon and Peterson 2005). Another example is the screening of a compound library for chemicals with antiangiogenic activity. Using the stable transgenic line *tg(VEGFR2:GRCFP),* in which the reporter gene is exclusively expressed in blood vessels of the developing embryo, it was possible to identify 1 compound with antiangiogenic activity, which could be verified in human endothelial cell assays (Tran et al. 2007).

Despite the apparent HTS potential of the zebrafish embryo, only a few technologies that exploit these advantages in an automated manner have been reported. A key challenge for automation is the scoring of expression patterns or phenotypes within the complexity of a vertebrate embryo. Successful examples include the study of Tran *et al*. 2007 who developed an algorithm that can automatically quantify alterations of reporter signal activity in the trunk of embryos of the *tg(VEGFR2:GRCFP)* transgenic line. A further example is the study of Vogt *et al*. 2009 who utilized the Cognitive Network Technology (CNT, Definiens, Munich Germany) to quantify fluorescently labelled blood vessels in the *Tg(fli1:EGFP)*^{y1} transgenic line in an orientation independent manner. However, both approaches are restricted to a pre-defined stereotypical pattern, thus limiting their application.

2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Chemicals

All chemicals used were of analytical or molecular biological quality grade. Standard laboratory chemicals were purchased from Sigma-Aldrich, Poole, UK or Fisher Scientific, Loughborough, UK unless otherwise stated. Salts (e.g. NaCl or $CaCl_2$) or solvents (e.g. ethanol or chloroform) commonly used in most molecular biology laboratories are not listed.

Ampicillin	Sigma-Aldrich, Poole, UK
Bovine serum albumin (BSA)	Serva, Heidelberg, Germany
DNA ladder (100bp, 1kb, 2-log)	Promega, Southampton, UK
DNA ladder (100bp, 1kb, 2-log)	New England Biolabs, Hitchin, UK
dNTPs	5Prime, Nottingham, UK
Gentamycin	Fisher, Loughborough, UK
Nuclease free water	Ambion, Warrington, UK
Paraformaldehyde	Sigma-Aldrich, Poole, UK
D-PBS (with $CaCl_2$ and $MgCl_2$)	Invitrogen, Paisley, UK
Phenol red solution	Sigma-Aldrich, Poole, UK
N-Phenylthiourea	Sigma-Aldrich, Poole, UK
molecular biology grade glycogen	Invitrogen, Paisley, UK
Kanamycin	Sigma-Aldrich, Poole, UK

2.1.2. Enzymes

Restriction enzymes	Invitrogen, Paisley, UK
	New England Biolabs, Hitchin, UK
	Promega, Southampton, UK
RNase A	Sigma-Aldrich, Poole, UK
Protease from *Streptomyces griseus* Type XIV	Sigma-Aldrich, Poole, UK
Proteinase K	Sigma-Aldrich, Poole, UK
GoTaq DNA Polymerase	Promega, Southampton, UK
PCR extender system	5Prime, Nottingham, UK
T4 DNA Ligase	Promega, Southampton, UK

2.1.3. Equipment and Consumables

Standard molecular biology laboratory equipment such as water baths, centrifuges, bacterial incubators, heat blocks, plastic ware, pipettes and similar items are not listed.

Analogue gas microinjector	Tritech Research, Los Angeles, USA
borosilicate glass capillaries (OD 1mm, ID 0.78mm)	Harvard Apparatus, Kent, UK
Digital microscope camera DFC300 FX	Nikon, Kingston, UK
Flaming-Brown needle puller	Sutter Instruments, Novato, USA
Fluorescence stereo microscope MZ FLI-II	Leica, Bensheim, Germany
Incubator for fish embryos	Heraeus, Hanau, Germany
Microfiltration columns	Pall, Ann Harbor, USA
Microloader tips	Eppendorf, Hamburg, Germany
NanoDrop ND-1000	Peqlab, Erlangen, Germany
Scan^R high content screening station	Olympus, Hamburg, Germany

Stereomicroscope SMZ645	Nikon, Kingston, UK
Vacuum manifold	Promega, Southampton, UK

2.1.4. Antibodies

Primary antibodies:

- Rabbit anti-GFP polyclonal antibody (Invitrogen, Karlsruhe, Germany; Cat.-No. R970-01); working dilution for whole-mount antibody staining 1:200.
- Rabbit anti-GFP polyclonal antibody (Torrey Pines Biolabs, Houston, USA; Cat.-No. TP401); working dilution for whole-mount antibody staining 1:500.

Secondary antibody:

- Goat anti-rabbit polyclonal HRP conjugated (Dako, Glostrup, Denmark, Cat.-No. P0448); working dilution for whole-mount antibody staining 1:500.

2.1.5. Morpholinos

All morpholino oligonucleotides were purchased from GeneTools, Philomath, USA.

Standard control morpholino:
5'-CCTCTTACCTCAGTTACAATTTATA-3'

TBP-MO (targeting ATG region of zebrafish *tbp* mRNA; working concentration: 1 mM):
5'-GAGGTAGGCTGTTGTTATGTTCCAT-3' (Muller et al. 2001)

TBP-MO 2 (targeting 5'UTR of *tbp* mRNA; working concentration: 0.75 mM):
5'-CAAAAGACGTAAACGATAATTCGCA-3'

2.1.6. Kits

mMESSAGE mMACHINE Kit	Ambion, Warrington, UK
Peroxidase substrate kit	Vectorlabs, Burlingame, USA
pGlow-TOPO reporter kit	Invitrogen, Paisley, UK
Plasmid Maxi Kit	Qiagen, Crawley, UK
PureYield Plasmid Midiprep System	Promega, Southampton, UK
QIAprep Spin Miniprep Kit	Qiagen, Crawley, UK
Wizard SV Gel and PCR Clean-Up System	Promega, Southampton, UK

2.1.7. Oligonucleotides

All oligonucleotides were designed using the Primer3 software and purchased from Metabion, Martinsried, Germany. PCR – Primers are listed in **Appendix 1** and **3**.

2.1.8. Bacterial strains

For general cloning and TOPO cloning the TOP10 strain was used (Invitrogen, Paisley, UK).

2.1.9. Zebrafish lines

All wildtype embryos used derived from the AB or AB* strains.

Stable transgenic embryos derived from the *Tg(2.2shh:ABC:GFP)* stable transgenic line*;* the wildtype background is unknown. The transgene is stably expressed for >10 generations and is composed of head to tail concatamers of the transgene in an unknown number of copies (Shkumatava et al. 2004).

2.1.10. Buffers and solutions

If not otherwise stated buffers and solutions were prepared using deionised water.

- Hank's Medium: 0.137 M NaCl, 5.4 mM KCl, 0.25 mM Na_2HPO_4, 0.44 mM KH_2PO_4, 1.3 mM $CaCl_2$, 1.0 mM Mg SO_4, 4.2 mM $NaHCO_3$
- Lysis buffer for genomic DNA extraction: 10 mM Tris-HCl, 0.1 M EDTA, 0.5% SDS; adjust to pH 8
- Pronase stock (10mg/ml): dissolve Protease from *Streptomyces griseus* Type XIV in Hank's Medium and incubate at 37°C for 1 h.
- Proteinase K stock solution (10mg/ml): dissolve in 100 mM Tris-HCl pH 8, 6 mM $CaCl_2$ and add 1 volume glycerol; store @ -20C
- RNase A stock solution (10-20mg/ml): dissolve in 100 mM Tris-HCl (pH 7.4) and boil for 15 min; cool down to room temperature and add 1 volume glycerol; store @ -20C
- TE-buffer: 10 mM Tris-HCl (stock: 1M pH 7.5), 1 mM EDTA (0.5M pH 8.0)
- 4% Paraformaldehyde (PFA): stir 4% PFA in D-PBS at 60°C until completely dissolved; store at +4°C for maximally 1 week
- PTW: supplement D-PBS (with $CaCl_2$ and $MgCl_2$) with 0.1% Tween-20
- PBT: supplement PTW with 1% BSA
- PTU stock solution (0.06%; 20x): stir 0.06% PTU in Hank's or fish water at 60°C until completely dissolved; store at room temperature

2.1.11. Plasmids

Plasmids for mRNA synthesis

pCS2+xTBPc	Fragment encoding for c-terminal core domain of Xenopus TBP (xTBPc)	(Veenstra et al. 2000)
pFOL876	Fragment encoding for bacterial transposase IS30	(Szabo et al. 2003)
pCS+CFP	Fragment encoding for cyan fluorescent protein (CFP)	-
pCSdnReg	Fragment encoding for dominant negative variant of PKA (PKI)	(Muller et al. 2000)
pCS+Tol2	Fragment encoding for Tol2 transposase	(Parinov et al. 2004)

Reporter constructs

β-actin:yfp	4.5 promoter of carp *β-actin*	F. Mueller (unpublished)
β-actin:cfp	4.5 promoter of carp *β-actin*	F. Mueller (unpublished)
Tol2-*shha arC::shha*-Tol2	*shha arC* enhancer linked to *shha* core promoter in front of *venus* flanked by Tol2 sites	Y. Hadzhiev (unpublished)

Gateway constructs

pDONR-221-P1/P2	Promoter entry vector	(Roure et al. 2007)
pDONR-221-P3/P5	Enhancer entry vector	(Roure et al. 2007)
pSP72-R3-ccdB/cmR-R5::RfA-venus	Destination vector	(Roure et al. 2007)

Promoter TOPO cloning

pGlow – TOPO	TOPO cloning site in front of *cycle3-gfp*	Invitrogen, Paisley, UK

2.2. Molecular biological methods

2.2.1. Phenol – chloroform extraction of nucleic acids

To remove proteins and lipids from nucleic acid containing solutions an equal volume of Tris-buffer saturated phenol-chloroform (ratio 1:1) was added. The solution was vortexed briefly and centrifuged at 10,000 rpm for 10 min to separate the aqueous and organic phases. Then, the upper aqueous phase was transferred to a fresh microcentrifuge tube and an equal volume of chloroform was added. The mixture was vortexed briefly and centrifuged at 10,000 rpm for 10 min. Again, the upper aqueous layer was transferred to a fresh microcentrifuge tube. The purified nucleic acid solutions were stored at -20°C or precipitated immediately.

2.2.2. Precipitation of nucleic acids

DNA was precipitated by mixing DNA containing aqueous solutions with either 1/25 volume of 5M sodium chloride (final concentration 0.2M) or 1/10 volume of 3M sodium acetate pH5.2 (final concentration 0.3M) and 2.5 volumes of 100% ethanol. If the DNA concentration was expected to be low (<1µg total DNA) 20 µg of molecular biology grade glycogen were added additionally. The solutions were incubated at -20°C for at least 30 min and subsequently centrifuged at 13,000 rpm for 20 min at 4°C. The supernatant was discarded and the pellet was washed with 70% ethanol. After an additional centrifugation step at 13,000 rpm for 5 min at 4°C the supernatant was discarded, the pellet was air dried and

resuspended in the appropriate volume of TE buffer or nuclease free water and stored at -20°C.

RNA was precipitated accordingly, except that the precipitation was carried out using 1/10 volume of 5M ammonium acetate (final concentration 0.5M) and 1 volume of 100% isopropanol. RNA solutions were stored at -80°C.

2.2.3. Isolation of plasmid DNA

Plasmid DNA preparations were carried out using commercially available Mini-, Midi- and Maxiprep kits (Promega or Qiagen). The kits from both manufacturers are based on the alkaline lysis protocol for plasmid DNA isolation (Sambrook and Russell 2001). Bacterial cells were harvested by centrifugation from bacterial liquid cultures. The bacterial cell pellets were lysed and the plasmid DNA was separated from genomic DNA and bacterial debris by centrifugation or filtration. The plasmid DNA containing solutions were applied to an anion exchange or silica membrane column that binds the plasmid DNA. Remaining proteins, RNAs and bacterial endotoxins were removed by different washing steps. Finally the column bound plasmid DNA was eluted in an appropriate volume of nuclease free water or TE buffer and stored at -20°C.

2.2.3.1. Plasmid DNA Miniprep

For plasmid DNA minipreparation the QIAprep Spin Miniprep Kit was used. 2 ml of a bacterial liquid culture were transferred to a microcentrifuge tube and the cells were harvested by centrifugation at 8,000 rpm for 3 minutes. The supernatant was discarded and

the bacterial cell pellet was resuspended in 250 µl P1 buffer. To lyse the cells 250 µl P2 buffer were added and mixed thoroughly by inverting the tube 4-6 times. 350 µl N3 buffer were added and mixed immediately by inverting the tube 4-6 times. The solution was centrifuged at 13,000 rpm for 10 min. The supernatant was applied to a spin column and spun at 13,000 rpm for 30s. The flow-through was discarded. The column was washed by adding 0.5 ml of buffer PB and centrifugation at 13,000 rpm for 30s. The column was washed again by adding 0.75 ml of buffer PE and centrifugation at 13,000 rpm for 30s. After each washing step, the flow through was discarded. The column was centrifuged at 13,000 rpm for 1 min to remove residual buffers and ethanol. The column was transferred to a clean microcentrifuge tube and the DNA was eluted by adding 50 µl of water. After 1min incubation the DNA was collected by centrifugation at 13,000 rpm for 1 min.

2.2.3.2. Plasmid DNA Midiprep

For plasmid DNA midipreparation the PureYield Plasmid Midiprep System was used. Cells were harvested from 50 ml bacterial liquid culture by centrifugation at 5,000 rpm for 5 min. The bacterial cell pellet was resuspended in 3 ml cell resuspension solution. To lyse the cells 3ml of cell lysis solution were added and mixed by inverting the tube 3-5 times. After 5 min incubation the lysis was stopped by adding 5 ml of neutralization solution and mixed by inverting 5 times. The lysate was centrifuged at 15,000 x g for 15 min. Meanwhile a column stack was assembled by placing a clearing column on a binding column and placed onto a vacuum manifold. The supernatant was poured into the column stack and the vacuum was applied until all the supernatant had passed the columns. The column membranes were

washed by adding 5 ml endotoxin removal wash and 20 ml column wash solution. The vacuum was applied until the membrane was dry. The binding column was transferred into a 50 ml tube and the DNA was eluted by adding 600 µl nuclease free water and collected by centrifugation at 1,500 x g for 5 min.

2.2.3.3. Plasmid DNA Maxiprep

For plasmid DNA maxipreparation the Qiagen Plasmid Maxi Kit was used. Cells were harvested from 100 ml bacterial liquid culture by centrifugation at 6,000 x g for 15 min at 4°C. The bacterial cell pellet was resuspended in 10 ml buffer P1. Cells were lysed by adding 10 ml P2 buffer. After thorough mixing the lysate was incubated for 5 min at room temperature. The lysis was stopped by adding 10 ml of chilled P3 buffer and incubation for 20 min on ice. Then, the solution was centrifuged at 20,000 x g for 30 min at 4°C. Meanwhile, a binding column was equilibrated by applying 10 ml QBT buffer. After centrifugation, the supernatant was transferred to the equilibrated column. The column was washed twice with 30 ml QC buffer and once with 15 ml QF buffer. The DNA was eluted in 15 ml QF buffer. At each step, columns were allowed to empty by gravity flow. The DNA was precipitated by adding 10.5 ml isopropanol and centrifuged at 15,000 x g for 30 min at 4°C. The pellet was washed with 5 ml 70% ethanol, spun again at 15,000 x g for 10 min at 4°C and resuspended in 200 µl of nuclease free water.

2.2.4. Isolation of genomic DNA

After removal of excess liquid, tissue samples or embryos were frozen in liquid nitrogen and subsequently homogenized while kept frozen. The tissue powder (approx. 500 mg) was spread on the surface of 10 ml lysis buffer containing 20 µg/ml RNAse A and submerged by gentle shaking. The suspension was incubated for 30 min at 37°C. After addition of Proteinase K (final concentration 100 µg/ml) the solution was incubated for at least 3 hours at 55°C. The lysate was purified by two extractions with Phenol and one extraction with phenol-chloroform. The DNA was precipitated using 0.2 volumes of 10 M Ammonium acetate and 2 volumes of ethanol. The DNA precipitate was taken out using a U-shaped Pasteur pipette, washed with 70% ethanol and air-dried. Then the DNA was dissolved in an appropriate volume of TE – buffer.

2.2.5. Determination of nucleic acid concentration

Concentrations of aqueous nucleic acid solutions were determined using a NanoDrop spectrophotometer (Peqlab, Erlangen). The optical density at 260 nm wavelength was measured, which corresponds to the absorption maximum of nucleic acids. Additionally, the absorption at 280 nm was measured to check for potential contamination with proteins. Optimally the ratio OD260/OD280 should be 2:1.

2.2.6. Restriction digest of DNA

Digestion of DNA using restriction endonucleases was carried out according the instructions of the enzyme supplier. The digests were carried out in the appropriate buffer and a two- to

tenfold excess of enzyme was used. Usually 2-3 Units of enzyme per microgram of DNA were used. If not otherwise stated by the manufacturer the reaction mix was incubated 1-4 hours depending on the amount of DNA and the excess of enzyme used.

2.2.7. Ligation of DNA fragments

Ligation of DNA fragments generated by restriction digest with cohesive ends was carried out using T4 DNA Ligase. When cloning a fragment into a plasmid vector a 3:1 molar ratio of vector:insert was used. For a 20 µl reaction 100 ng of vector DNA were mixed with 2 µl of 10x ligation buffer, the corresponding amount of insert DNA, the appropriate amount of water and 1-2 units of T4 DNA Ligase. The reaction mix was incubated at room temperature for 3 hours.

2.2.8. Polymerase Chain Reaction

The selective amplification of DNA fragments from DNA templates was carried out by PCR (= polymerase chain reaction). Depending on the purpose of the PCR two different PCR systems were utilised. If no proofreading activity was required an ordinary Taq DNA Polymerase was used. If proofreading activity was required, e.g. for the generation of fragments for molecular cloning, the PCR extender system was used. This system consists of a blend of Taq DNA Polymerase and proofreading enzymes that allows the efficient amplification of relatively long fragments with a very low error rate. The PCRs were carried out according to the manufacturer's instructions. Annealing temperatures and extension times were adjusted according to the specific set of primers and expected size of the PCR product.

For amplification of promoter regions of larger promoter fragments (1-2 kb) a nested PCR was carried out. A nested PCR is a two step PCR reaction that increases specificity and efficiency of amplifications. In a first round of PCR an outer primer pair is used to amplify a target region. In a second round of PCR an inner primer pair is used to amplify a region located within the first PCR product. Due to the enrichment of target DNA in the first PCR round, the amplification of non-specific side products is drastically reduced in the second round.

2.2.9. Agarose Gel electrophoresis

Agarose gel electrophoresis was carried out to check the size and quality of nucleic acid samples, as well as to separate fragments of different sizes. Depending on the expected size of the analyzed nucleic acid samples the concentration of prepared agarose gels ranged between 0.8% and 2% agarose in 1x TBE-buffer supplemented with 0.5µg/ml Ethidiumbromide. Before loading the samples were mixed with an appropriate volume of loading buffer. In addition to the samples a DNA ladder was loaded to determine the size and approximate quantity of the nucleic acid samples. The electrophoresis was carried out using 3-5V/cm gel length. Nucleic acids were visualized using an UV-Transilluminator.

2.2.10. Isolation of DNA from Agarose Gels

To isolate DNA fragments from agarose gels the Promega Wizard SV Gel and PCR Clean-Up System was used. The desired DNA fragments were cut out from the gels on a UV transilluminator and the gel slice was dissolved in 10µl Membrane Binding Solution per

10mg of gel and incubated at 50–65°C until the gel slice was completely dissolved. Then the solution was transferred to a binding column, incubated for 1 min and spun for 1 min at 16,000 x g. The flowthrough was discarded. The column was washed twice with 700 µl and 500 µl Membrane wash solution, respectively. Each time the column was spun for 5 min at 16,000 x g and the flowthrough was discarded. The column was dried by centrifugation and evaporation and transferred to a fresh microcentrifuge tube. Then the DNA was eluted with 50µl nuclease-free water and collected by centrifugation for 1 min at 16,000 x g.

2.2.11. Adenylation of DNA fragments

When gel-purified PCR products were used for TOPO-cloning they were adenylated by incubating them with 0.2 mM dATP, 1.5U Taq DNA Polymerase and an appropriate volume of PCR reaction buffer for 15 min at 72°C.

2.2.12. TOPO cloning

For TOPO cloning the pGlow TOPO reporter kit from Invitrogen was utilized using slightly modified conditions. The ligation reaction consisted of 1 µl PCR-product (with 3´-A over-hangs), 0.5 µl salt solution, 1 µl nuclease-free water and 0.5 µl of the linearized pGlow-TOPO vector covalently linked to Topoisomerase I. The reaction mix was incubated for 7 min at room temperature. Then 2 µl of the reaction mix were used to transform chemically competent *E.coli* and the remainder was stored at -20°C.

2.2.13. Sequencing

All sequencing relevant to this thesis was carried out by GATC Biotech, Konstanz, Germany.

2.2.14. *In-vitro* transcription

Preparation of mRNA for microinjections was carried out using the mMESSAGE mMACHINE Kit from Ambion. The reaction mix was assembled by combining 10 µl 2X NTP/CAP, 2 µl 10x reaction buffer, an appropriate volume of water, 1µg of template DNA and 2 µl enzyme mix, so the final volume was 20 µl. The reaction mix was incubated at 37°C for 2h. Template DNA was digested by adding 1 µl of TURBO DNase and incubation at 37°C for 15 min. *In-vitro* transcribed mRNA was recovered by phenol-chloroform extraction and subsequent isopropanol precipitation.

2.3. Microbiological methods

2.3.1. Bacterial cell culture

Bacterial culture plates were generated by spreading bacterial liquid culture onto LB-Agar-plates. If a selection for resistance was carried out the LB-Agar was supplemented with the relevant antibiotic at an appropriate concentration (50-100 µg/ml ampicillin or 50 µg/ml kanamycin). Bacterial culture plates were incubated overnight at 37°C.

Bacterial liquid cultures were generated by inoculating LB-Media with a colony from a bacterial culture plate or with a small volume of liquid culture. Relevant antibiotics were added if required. Bacterial liquid cultures were grown at 37°C until the desired OD was reached.

2.3.2. Transformation of chemically competent E.coli

Chemically competent *E.coli* were thawed on ice and mixed with 1-2 ng plasmid DNA or an appropriate volume of ligation reaction. The cells were incubated with the DNA for 10 min on ice, heat-shocked for 30 s at 42°C and placed again on ice for 2 min. After addition of 250 µl SOC- or 500 µl LB-media the cells were incubated for 1 h at 37°C. Finally 100-200 µl of the bacterial cell culture was spread onto antibiotic containing LB-Agar-plates.

2.3.3. Colony PCR

To identify positive clones from bacterial culture plates a colony PCR was carried out. A single colony was picked using a pipette tip and the cells were transferred to a PCR tube containing a ready PCR reaction mix. The cells that remained on the pipette tip were used to inoculate a bacterial liquid culture.

2.4. Immunhistochemical methods

2.4.1. Fixation of zebrafish embryos

Gastrula stage zebrafish embryos were fixed by transferring them into ice-cold 4% PFA. After overnight incubation at 4°C the embryos were dehydrated by replacement of the fixative with 100% methanol and stored at -20°C.

2.4.2. Preparation of fish powder

Fish powder was used to pre-block antibodies. Zebrafish embryos were ground while kept frozen using liquid nitrogen. The excess water and fat was removed by acetone extractions.

The fish powder was dried on a whatman paper under the hood. Fish powder was stored at 4°C.

2.4.3. Whole-mount immunostaining

Embryos stored in methanol were rehydrated by successive washes with 75%, 50%, 25% methanol in PTW. Subsequently, the embryos were washed 3x with PTW and 1x with distilled water. Then the embryos were incubated in pre-cooled acetone for exactly 7 min at -20°C to increase the permeability of cellular membranes. After that, the embryos were washed 1 x 1 min in distilled water, 1 x 5 min in 1:1 PTW/water and 2 x 5 min in PTW. Then they were rinsed once in PBT and blocked for 3-5 h in PBT. In parallel the primary antibody was pre-blocked in PBT containing fish powder. After blocking the fish powder was removed by centrifugation and the embryos were incubated with the primary antibody overnight at 4°C. The next day unbound antibody was removed by 6x5 min washes with PBT. Again the embryos were blocked for 3-5h in PBT and in parallel the secondary antibody was blocked in PBT containing fish powder. After removal of the fish powder the embryos were incubated overnight with the secondary antibody. Excess secondary antibody was removed by 6x5 min washes with PTW.

For the staining reaction the embryos were transferred into 24 well plates and incubated for 2-10 min in DAB staining solution according to the manufacturer's instructions. When sufficient signal strength was reached the enzymatic staining reaction was stopped by replacing the solution with 4% PFA in PBS. Stained embryos were stored in fixative at 4°C and are stable for several weeks.

2.4.4. Image analysis using ImageScope

Images of immunostainings were analysed using Aperio ImageScope software (version 7.1.32.1024; Aperio Technologies, Vista, CA, USA; freely available at www.aperio.com).

To quantify immunostainings, the stained gastrula stage embryos were deyolked and oriented with animal pole up on 3% Agar in PBS. Images were acquired using a DFC300 FX digital camera (Leica, Bensheim, Germany) connected to a MZ FLIII stereomicroscope (Leica, Bensheim, Germany). Colour quantification was performed applying the Aperio positive pixel count algorithm using default settings for brown colour quantification. To quantify fluorescent images groups of randomly oriented embryos were imaged using the same microscopic setup. The settings of the Aperio positive pixel count algorithm were adjusted to count pixels of the respective colour of the fluorescent signal.

For each image algorithm threshold settings were specified to assure correct detection of stained tissue. The correct detection of the desired colour and intensities was confirmed on the basis of the positive "mark-up" image (**Figure 3.5B,C**). The number of positive pixels (N = Nwp+Np+Nsp) of GFP positive embryos was measured by defining an annotation layer for each embryo. Non-expressing embryos were considered as N = 0.

2.5. Fish husbandry and embryological methods

2.5.1. Keeping of adult zebrafish

Adult zebrafish stocks are kept in groups of 7-15 pairs depending on tank size. The water conditions are: conductivity 400-500 µS, hardness 5° dH, pH 7.0-7.5 and temperature between 26°C and 28°C. As the zebrafish breeding behaviour is affected by the day/night cycle the photo period in the fish facility is adjusted to a 14 hours light/10 hours dark cycle. The zebrafish are fed twice a day and water quality (ammonium, nitrate, etc) is controlled on a weekly basis.

2.5.2. Embryo production

Crosses of adult zebrafish for embryo production were carried out in 1 litre crossing cages. The crossing cages contain an inlay with a removable divider that allows separating the sexes. Additionally, the bottom of the inlay contains a mesh that allows the separation of the eggs from the parents after egg laying to avoid parental cannibalism. In the evening 1 or 2 pairs of zebrafish were set up in the crossing cages, while sexes were kept separated. The next morning with the beginning of the light period sexes were combined by removing the divider. Eggs were collected 5-15 min after laying by pouring the water through a small net and were transferred to a petri dish.

2.5.3. Raising embryos and larvae

Dechorionated zebrafish embryos were raised in Hank's Medium at 28-30°C until they had reached the desired developmental stage. When chorionated embryos were raised fish water was optionally used. Regularly, the embryos were cleaned and the media was exchanged. To avoid growth of bacteria and fungi Gentamycin (final concentration 50 µg/ml) was added to the Hank's Medium. Optionally, the medium was supplemented with 0.003% PTU (Phenylthiourea) to avoid pigment formation.

2.5.4. Dechorionation

Enzyme assisted removal of embryonic egg shells was carried out using Pronase (10 mg/ml). Embryos were dechorionated either directly after egg collection or around 24 hours post fertilization. Embryos were transferred in petri dishes and most of the media was removed. Then 0.5 ml of Pronase stock solution was added. The embryos were observed under a stereomicroscope while gently swirling them until the first embryos started to come out of the chorion. Then they were transferred to a beaker and washed three times with approximately 500 ml of system water. After dechorionation embryos were transferred back to petri dishes containing Hank's medium plus Gentamycin and raised as described above. If embryos were dechorionated right after fertilization agar coated petri dishes were used to avoid damaging of the very sensitive early embryos. For dechorionation of older embryos, i.e. >24 hpf (prim-5), only one wash step was carried out after most of the embryos were out of the chorion and embryos can be optionally transferred back to system water.

Alternatively, older embryos (>24hpf) were dechorionated manually using forceps.

2.5.5. Microinjection of zebrafish embryos

Microinjection of zebrafish embryos was carried out using gas microinjectors and stereomicroscopes. The needles for the microinjection were prepared from borosilicate capillaries (inner diameter: 0.78 mm, outer diameter: 1.0 mm) on a Flaming-Brown needle puller. The needles were filled with 1-3 µl injection solution using Eppendorf microloader pipette tips. For the microinjections embryos were either dechorionated right after egg collection and transferred into agar-coated petri dishes or they were injected through the chorion after transfer to a petri dish and removal of all liquid. The injection solutions contained phenol red that allowed controlling injection efficiency qualitatively and quantitatively. For double injections dechorionated embryos were exposed to a second round of injection.

2.5.6. Preparation of injection solutions

Injections solutions were prepared by combining plasmid DNA, mRNA or Morpholinos with nuclease free water and phenol red (final concentration 0.1 %). The concentrations of experimental compounds varied depending on the assay: plasmid DNA 5-100 ng/µl, mRNA 10-100 ng/µl and Morpholino 0.5-1 mM. Injection solutions were filtered prior to injections using spin filter columns and stored at -20°C/-80°C until needed.

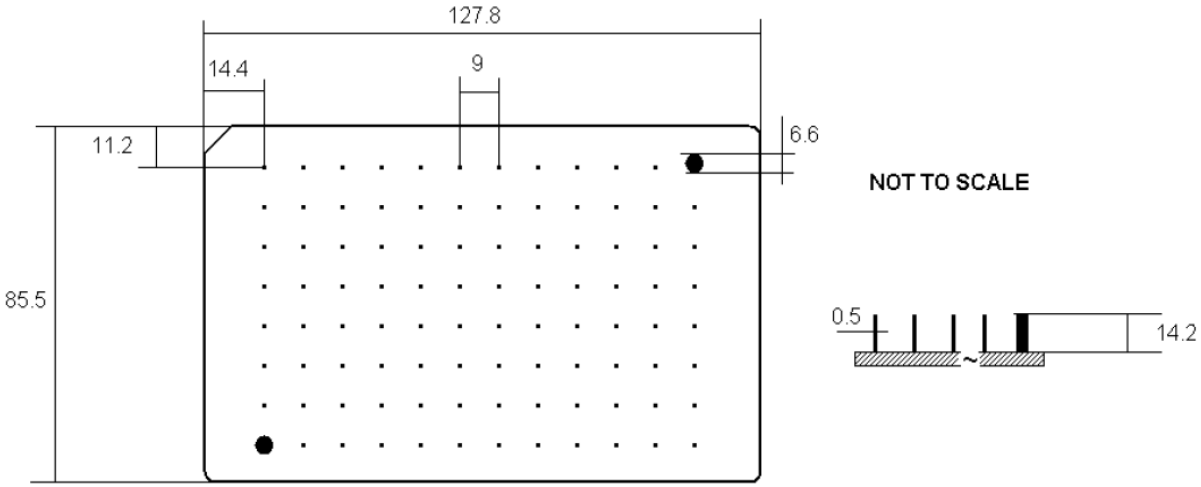


Figure 2.1| Depiction of template used for the generation of agarose embedded 96-well plates. The template consists of a PVC plate with 94 metal pins of 0.5 mm diameter that match the position of the centre of each well. At the position of 2 corner wells the template has 2 guiding pins that facilitate the insertion into the 96 well plates. The template was custom built by POLYGON Szerszám Készito, Budapest, Hungary.

2.6. High throughput screening related protocols

2.6.1. Preparation of 96 well plates for high throughput imaging

Agarose embedded 96-well plates (Greiner Cat.-No. 655180) were generated by pipetting 45 µl of hot 1% agarose in fish water into each well using a multichannel pipette while the four corner wells were omitted, thus 92 wells were filled. To avoid precocious solidification of the agarose the plate was kept on a heat block at 70-80°C. To avoid meniscus formation, a droplet of 100% ethanol was added on top of the hot agarose. To generate concentric depressions within the agarose, a template with 94 pins (**Figure 2.1**) was inserted into the 96-well plate and the plate was kept at room temperature for at least 15 min. After solidification of the agarose the template was removed. Before usage plates were pre-warmed and washed to remove remaining traces of ethanol. Plates were stored in a sealed

plastic bag at 4°C, while for long term storage or after usage gentamycin containing water was added to each well to avoid growth of fungi and drying up of the agarose.

2.6.2. High throughput microscopy and data storage

Prior to microscopy, embryos were anaesthetized with 0.03% tricaine and transferred into agarose embedded plates in a volume of 100 µl using a cut 200 µl tip. Embryos were oriented using a bent subcutaneous injection needle. Plates were sealed using transparent PCR foil and loaded into a stacking system.

The plates were imaged using a Olympus Scan^R high content screening microscope (Liebel et al. 2003) (Olympus Biosystems) with a 2.5x objective (Plan-Apo) and an Olympus Biosystems DB-1 (1,300 × 1,024 pixels) camera in bright field and with CFP, YFP filter cubes. The plates were automatically transferred from the stacking system to the microscope stage by a SWAP plate gripper (Hamilton). Imaging occurred through the bottom of the plate and image integration times were fixed with 180 ms for CFP and 1,000 ms for YFP. The light source was an ultra stable MT-20 (Olympus Biosystems) with a 150 W xenon lamp. The focal plane of the embryos was detected by an object detection autofocus algorithm. We acquired images of each embryo in four z-dimension slices with 55 µm distance. Data management, thumbnail gallery generation and data compression was carried out via an assembly of LabView software modules (National Instruments) generated by Dr Urban Liebel, Institute of Toxicology and Genetics, Karlsruhe Institute of Technology, Germany.

2.6.3. Generation of multisite Gateway entry clones and expression vectors

The isolation of promoters and enhancers and the generation of constructs using a modified Multisite Gateway system, which is a recombination based approach that utilizes enzymes of the λ-phage recombination system (Roure et al. 2007), were done by Eva Kalmar. Therefore, I will just give a brief summary. A detailed description can be found in (Roure et al. 2007) and (Kalmár 2009).

Enhancer and promoter fragments were amplified from genomic DNA using the PCR extender system in a nested PCR. In the first round of PCR primers were used that flank the loci and additionally contain 12 bp of adaptor DNA that forms part of the attB recombination sites. In the second round of PCR primers were used that form the complete attB sites (29 bp) and thus partially overlap with the adaptor sites and extend the PCR product with full attB recombination sites. In a next step entry clones (ECs) were generated by carrying out PB recombination reactions between PCR amplified enhancer or promoter fragments and pDONR 221-P1/P2 or pDONR-221-P3/P5 donor vectors using Gateway BP Clonase enzyme mix (Invitrogen, Paisley, UK) (Roure et al. 2007). For the generation of pSP72-B3-enhancer-B5::B1:core promoter:B2-Venus expression vectors (ExV), LR recombination reactions between promoter and enhancer containing ECs and the destination vector (pSP72-R3-ccdB/cmR-R5::RfA-venus) were carried out using Gateway LR Clonase Enzyme Mix (Invitrogen, Paisley, UK) (Roure et al. 2007). Plasmids were verified by sequencing (ECs) or by restriction digestion (ExVs).

2.6.4. Merging of experimental repeats

The merging routine algorithm to evaluate pattern reproducibility was designed by Dr Markus Reischl, Institute of Applied Computer Sciences, Karlsruhe Institute of Technology, Germany.

Experimental data from distinct injection experiments was merged after homogeneity check. To this end, discrete domain expression values were introduced describing low, medium and high expressions (-1, 0 and 1). Domain specific expressions in all embryos of the repeats were ranked and assigned a -1 if they belong to the 33% lowest, a 1 if they belong to the 33% highest expressions or 0 for the remainder. Thus, the mean value of the discrete domain expression values for all repeats equals zero for each domain expression. For individual repeats, the mean value x was unequal zero. The differences of the individual to the overall mean were calculated and a value between 0 and 1 was obtained. 0 stands for an identical distribution, 1 means that only one of the discrete domain expression values occurs. This procedure was repeated for each domain. Each variation (for each domain expression in each repeat) was assigned a performance value (see Gehrig et al. 2009) for further details). The two domain expressions with the highest variations per repeat were cut off and all others of the same repeat were multiplied to obtain a performance index. If this performance index was above the heuristic found threshold 0.1, the repeats were merged.

2.7. Bioinformatic methods and resources

2.7.1. Local software

For local sequence analyses either BioEdit 7.0.9 for Windows 95/98/NT/2000/XP or DNA-Strider 1.3f16 for Mac OS IX/X were used.

Additionally, bioinformatic analyses were carried out using custom designed scripts written in the Perl 5.8 programming language.

2.7.2. Web based sequence databases

For sequence extractions and analyses (e.g. BLAST/BLAT analyses, primer design, sequence motif search) the following web databases/resources were used:

Bioinformatic Harvester IV	http://harvester.fzk.de/
BioMart	http://www.biomart.org/
Ensembl	http://www.ensembl.org/
Galaxy	http://galaxy.psu.edu/
MotifViz	http://biowulf.bu.edu/MotifViz/
NCBI	http://www.ncbi.nlm.nih.gov/
Primer3	http://frodo.wi.mit.edu/primer3/
UCSC genome browser	http://genome.ucsc.edu/
ZFIN	http://zfin.org/

2.7.3. EST distribution analysis

The chromosomal position of 5‘-ends of ESTs were extracted from the all_est table downloaded from the UCSC genome browser database, which contained information about the alignment between around 820.000 zebrafish ESTs and the zebrafish genome (January 2009). The zebrafish genome used was zv#7 (danRer5) (Rhead et al. 2009). Using a set of self written Perl scripts (available on request) the table was parsed for strand, chromosomal window or other features associated with EST tracks.

In principle, for the determination of EST numbers at certain chromosomal positions or regions the number of 5’ ends of ESTs at this position or within a window was counted. The 5’ ends were determined by the orientation of the ESTs. However, the strand information does not reflect the transcription direction of ESTs because they can be sequenced from either side (Rhead et al. 2009). Therefore, only chromosomal regions close to 5’ ends of genes were analysed to minimize the contribution of false-positive 5’ends derived from ESTs mapped to the opposite strand of their transcription direction, i.e. actual 3’ ends of ESTs.

To determine windows with maximal numbers of 5’ ends a sliding window approach was chosen. Therefore, a chromosomal window was scanned by “moving” a 10 bp window over the region in 1 bp steps, e.g. for a 100 bp window the occurrence of 5’ ends of ESTs was counted in 91 possible 10 bp windows.

3. ANALYSIS FOR TBP-DEPENDENCE OF PROMOTERS OF EARLY EXPRESSED ZYGOTIC GENES

Foreword

The work presented in this chapter is a continuation of my diploma project carried out in the laboratory of Dr Ferenc Müller at the Institute of Toxicology and Genetics at the Forschungszentrum Karlsruhe (Jochen Gehrig, *Analyse der TBP-Abhängigkeit von Promotoren früh exprimierter zygotischer Gene im Embryo des Zebrafisches Danio rerio,* submitted to the University of Karlsruhe, Germany, May 2006). During the course of this work I carried out the prediction, cloning and preliminary functional analysis of some of the promoters described in the results section. As a PhD student, I substantially expanded the analysis by including more promoters, improving promoter predictions, more control experiments and a more detailed functional analysis. Therefore, the majority of the presented data is unique to this thesis.

The work presented in this chapter has been partially published in Ferg et al. 2007.

3.1. Introduction and overview

3.1.1. The zebrafish midblastula transition

Like in most animal models the early development of zebrafish is characterized by rapid cell cycles and transcriptional quiescence followed by the activation of the zygotic genome and the commencement of differentiation processes; a period referred to as midblastula transition (MBT) (Kane and Kimmel 1993). This transition from a transcriptionally almost silent to a transcriptionally highly active state combined with the experimental accessibility of the early zebrafish embryo provides an interesting model system to study the regulation of transcription *in-vivo*.

The zebrafish MBT starts during the tenth cell cycle at the 512-cell stage. It shares many of the common features of the MZT/MBT, like the cell cycle lengthening, loss of cell cycle synchrony and loss of pluripotency. It coincides with broad and sudden activation of zygotic transcription of a large number of genes representing all major gene ontologies (Kane and Kimmel 1993; Mathavan et al. 2005; O'Boyle et al. 2007). Furthermore, the MBT is paralleled by the differential degradation of maternal mRNAs, which control pre-MBT and influence later embryonic development (Mathavan et al. 2005; Giraldez et al. 2006). In zebrafish, there is evidence that the delayed onset of transcription is controlled by at least two mechanisms: (i) the nucleocytoplasmic ratio (NCR) as indicated by the earlier onset of zygotic transcription in the polyploid nuclei carrying cells of *fue* mutants (Dekens et al. 2003) and (ii) deficiencies of the basal transcription machinery in pre-MBT embryos as suggested by the absence of the

TBP protein in pre-MBT embryos and the functional requirement of TBP family members in activation of zygotic transcription (Muller et al. 2001; Bartfai et al. 2004).

3.1.2. The role of TBP in zebrafish MBT

Knockdown of TBP by antisense morpholino oligonucleotides in zebrafish embryos leads to a severe phenotype characterized by strongly impaired epiboly, i.e. the almost complete lack of gastrulation movements, which is accompanied by reduced zygotic transcription. However, the analysis of transcript levels in TBP depleted embryos has shown that it is not universally required (Muller et al. 2001). In fact, only the minority of transcript levels are significantly downregulated as revealed by transcriptome analysis (Ferg et al. 2007). The majority of transcript levels remain unaffected and a surprisingly large fraction of transcripts is upregulated after TBP knockdown. As transcription dependent mRNA degradation pathways are perturbed in TBP depleted embryos, the elevated transcript levels are in part due to the loss of maternal mRNA degradation (Giraldez et al. 2006; Ferg et al. 2007). Besides TBP, also the other TBP family members TLF and TBP2 appear to play an important role during MBT and subsequent events, and their function could contribute to the unexpectedly moderate changes in transcript levels seen after TBP depletion (Muller et al. 2001; Bartfai et al. 2004).

3.1.3. Aims

There is strong evidence that the basal transcription factor TBP plays an important role in the activation of zygotic transcription during the zebrafish MBT. However, the observed non-

generality of gene responses to loss of TBP raises the question of what are the specific functions of TBP during early zebrafish development (Muller et al. 2001; Ferg et al. 2007). The analysis of the functional requirement for TBP at the transcript level allows only indirect observations due to the abundance of dynamically regulated maternal mRNAs and the major contribution of post-transcriptional regulatory processes. Hence, its requirement for initiation of transcription during MBT remains largely elusive. Therefore, the objective of this study was to investigate the functional requirement for TBP at the promoter level in zebrafish embryos, thus excluding posttranscriptional regulatory mechanisms. With this aim, I investigated the TBP dependence of promoters of early expressed zygotic genes in a transient transgenic reporter gene assay, in which I compared reporter gene activities in control and TBP depleted early zebrafish embryos. By investigating the direct functional requirement of TBP function for promoter activity I expected to expand the knowledge of the role of TBP in activating transcription during early zebrafish development.

3.2. Results

3.2.1. Promoters analysed in this study

The promoters chosen to analyse for TBP-dependence (**Table 3.1**) derived from two sets of genes. The first set consisted of 19 promoters of genes, which were detected in a SSH (suppressive subtractive hybridization) screen, which aimed to identify zygotically activated or upregulated transcripts, respectively (O'Boyle et al. 2007). Additionally, 5' - RACE (rapid amplification of cDNA ends) data was available for these 19 genes, which were generated by Dr Agnes Lovas in a preparatory study (Agnes Lovas, Ferenc Müller unpublished data). Since

TBP is known to play an important role in activating transcription during MBT, genes activated or upregulated at the MBT are an interesting group to study the direct transcriptional requirement of TBP. The second set contained 14 promoters of genes known to be early expressed (Muller et al. 2001; Ferg et al. 2007) including genes of general interest for their TBP dependence like the TBP family member genes *tbp* and *tbpl1* (zebrafish gene name for *tlf*).

3.2.2. Prediction of promoter regions

To predict the core promoter regions of the 33 selected genes, I mapped putative full-length cDNAs extracted from GenBank to the zebrafish assembly zv#4 (danRer2) or zv#5 (danRer3) using BLAST. I determined the region harbouring the core promoter as the sequence encompassing the 5'-end of the cDNA. For 19 promoters 5' - RACE data was available and these sequences map in 18 cases to the predicted core promoter regions, thus confirming the cDNA based predictions. The 5' - RACE of *sox3* maps to the end of the coding sequence of *sox3*, therefore it was considered as an erroneously truncated sequence (**Table 3.1**).

I mapped the promoter regions to the most recent assembly available (2009) in the UCSC genome browser (zv#7) using *in-silico* PCR (Kent et al. 2002). Hence, chromosomal coordinates presented in **Table 3.1** are based on assembly zv#7. An exception is the sequence verified promoter region of *tram1*, which could only be mapped to the assembly zv#4 due to loss of this region in subsequent assemblies.

3.2.3. Preparation of promoter-reporter constructs

To physically extract the promoters, I amplified them from genomic DNA by nested PCR using the PCR extender system that contains an enzyme mix with proofreading activity. I designed the inner primers (see **Appendix 1**) such that the amplified region spans a region of 1-2 kb upstream and up to 30-300 bp downstream of the putative TSS (**Table 3.1**). A relatively large promoter region was chosen, because I assumed that the core promoter, i.e. the actual target of TBP, would not be sufficiently active to give rise to detectable reporter gene expression. Consequently, I accepted that additional regulatory information might be present in these promoters. I chose the inner primer pair to exclude any ATG triplet in the predicted downstream region, which is not followed by an in-frame stop codon to avoid frameshift mutations in the subsequently generated reporter constructs. This approach resulted in the relatively large variations in the chosen downstream region (**Table 3.1**).

To generate reporter constructs, I cloned the 33 amplified promoter fragments in front of a *gfp* reporter gene using the pGlow-TOPO vector (see **Appendix 1**). Additionally, I cloned a 1.3 kb promoter fragment of the zebrafish *ntl* (*no tail*) gene as a positive control. The 1.3 *ntl* promoter is known to drive specific expression in gastrula stage embryos (Ferg 2008). In the following I will refer to pGlow promoter constructs as *promoter:gfp*. All generated promoter constructs were verified by sequencing using the GFP reverse priming site (see **Appendix 1**) which allowed sequencing of a region of around 700 to 850 bp covering the core and in parts the proximal promoter region.

Table 3.1| Promoters analysed in this study. For legend see next page

gene symbol	chromosomal positon	strand	size (bp)	Most 5' GenBank mRNA	Position of most upstream TSS of mRNA	Position of 5'RACE	position of EST 10 bp window	number of EST in 10 bp window	TSS distribution	CPE/CGI occurence (position)	functional at 50%/shield
anxa1a	chr5:19963203-19965235	+	2033	AY178793	1963	n/a	1972	97	dominant	TATA (1933)	+
apoeb	chr16:24387752-24389211	+	1460	BC154034	1071	1073	1080	483	dominant	TATA (1041)	+
(apoeb2)	chr16:24387752-24389210	+	1460	Y13652	1452	n/a	1441	645	dominant	-	+
atp6v1g1	chr5:50837844-50839220	+	1377	BC152208	1301	1333	1308	82	dominant	-	+
c20orf45	chr6:58934223-58935972	+	1750	BC045872	1682	1688	1726	39	broad	-	+
ccnb2	chr7:32959980-32961435	-	1456	BC066507	1370	n/a	1366	353	broad	-	+
ccne	chr7:43205866-43207273	-	1408	BC045842	1361	n/a	1367	13	dominant	-	+
cdk7	chr5:65928414-65929833	+	1420	DQ294347	1360	1395	1391	13	broad	-	+
clstn2	chr2:31827433-31828887	+	1454	AM422117	1449	1376	n/a	n/a	no conclusion	-	+
cxadr	chr10:21100280-21102079	+	1800	BC045286	1736	n/a	1741	6	no conclusion	-	-
foxa	chr14:38474941-38476443	-	1503	BC080222	1402	n/a	1430	6	no conclusion	-	+
gngt2	chr3:22392977-22394992	+	2016	BC059612	1952	n/a	1932	42	dominant	-	+
gtf2a1	chr20:12835427-12836911	+	1485	BC048894	1425	1425	1402	29	broad	TATA (1394)	+
her7	chr5:61590699-61592733	-	2035	BC092709	1963	n/a	1990	3	no conclusion	TATA (1933)	-
klf4	chr2:27759296-27761141	+	1846	AM422104	1670	1674	1688	10	broad	-	+
kpna2	chr3:21229734-21231116	-	1383	BC075790	1339	1335	1339	102	dominant	-	+
krt4	chr6:26288942-26290833	-	1892	BC155774	1803	n/a	1808	344	dominant	TATA (1774)	+
marcksl1	chr13:17455392-17457243	-	1852	BC058848	1725	1704	1707	67	broad	-	+
mkks	Zv7_scaffold2635:5049-6475	-	1427	BC045401	1387	1387	1410	2	no conclusion	-	(+)
ndr1	chr21:11382503-11383825	-	1323	AF002218	1145	n/a	1240	3	no conclusion	-	+
otx1	chr17:20278393-20279732	-	1340	BC045290	1282	1279	1283	10	dominant	-	+
pcbp2	chr9:614692-615936	-	1245	BC045508	1215	1213	1238	50	broad	CGI	+

Table 3.1 continued| Promoters analysed in this study.

gene symbol	chromosomal positon	strand	size (bp)	Most 5' GenBank mRNA	Position of most upstream TSS of mRNA	Position of 5'RACE	position of EST 10 bp window	number of EST in 10 bp window	TSS distribution	CPE/CGI occurence (position)	functional at 50%/shield
raver1	chr3:44606447-44608398	+	1952	BC095698	1781	n/a	1788	4	no conclusion	-	+
rdh10	chr2:24615019-24616394	+	1376	BC054596	1300	1279	1286	148	dominant	-	+
sft2d3	chr9:905461-906480	-	1020	BC114252	944	976	968	4	no conclusion	CGI	+
sox3	chr14:36980556-36981855	-	1300	BC092845	1216	2220	1237	17	broad	-	+
tbp	chr13:24700979-24702404	+	1426	BC055549	1310	n/a	1313	57	broad	TLS (1276)	+
tbpl1	chr23:30363454-30364858	+	1405	BC085661	1329	n/a	1332	8	broad	-	+
tbx16	chr8:46497139-46499143	+	2005	BC095003	1843	n/a	1851	10	dominant	-	+
thy1	chr5:21575747-21577247	+	1501	BC152206	1302	1305	1311	123	dominant	-	+
tppp3	chr7:28263831-28265738	-	1908	BC062530	1863	n/a	1839	9	broad	-	-
tram1	chr24:14311577-14313084**	+	1508	BC052143	1378	1405	n/a	n/a	no conclusion	-	+
vox	chr13:52346787-52348195	+	1409	BC116578	1372	1372	1378	7	no conclusion	TATA (1341)	+
zgc:66242	chr5:49461531-49463363	+	1833	BC055560	1694	1687	1673	3	no conclusion	-	-

Legend to Table 3.1| Column 1: Gene symbols based on zv#7. *apoeb2* stands for the second putative TSS region within the *apoeb* promoter fragment. Column 2-4: Chromosomal position in zv#7 (** in zv#4), strand/orientation and size of the cloned promoter fragment. Column 5-6: Accession number of the most 5' located GenBank mRNA and its position within the promoter. Column 7: Position of the 5'-RACE verified TSS within the fragment; n/a: 5'-RACE not available. Column 8-9: Position of the 10 bp window containing maxima of 5'-ends of EST within the promoter and the number of 5' – ends of ESTs within the window. Column 10-11: Putative TSS distribution based on 5' ends of ESTs mapping in 150 bp window encompassing the EST based TSS position (column 8) and the occurrence of CPEs; CGI indicates overlap of the cloned promoter fragment with a CGI. Column 12: Functionality of generated promoter constructs at the 50%/shield stage. For further details see text. (Abbreviations: n/a, not applicable; TLS, TATA-like sequence).

3.2.4. *In-silico* verification of promoters

Since not all promoter locations were experimentally verified by 5'-RACE, I carried out an *in-silico* analysis of the TSS regions using core promoter associated annotations for assembly zv#7 present in the UCSC genome browser database (Rhead et al. 2009). This included the 5' ends of zebrafish mRNAs and the position of 5'ends of cognate ESTs.

A commonly used approach to determine a representative TSS of a putative promoter region is the identification of the most upstream located 5' end of associated cDNAs (e.g. Suzuki et al. 2001b). Therefore, I mapped the most upstream 5' ends of GenBank mRNAs available in the UCSC genome browser database. All mRNAs for which 5'RACE validation was available mapped to the promoter region with on average 15 bp distance to the 5' - RACE verified TSSs, thus confirming the core promoter predictions in these cases (**Table 3.1**). It must be noted that the GenBank mRNA set was in part redundant (8 promoters) to the previously used putative full length cDNAs, thus the confirmation of TSS predictions was only partially independent. Moreover, the genomic position of the most 5' located nt of mRNAs is not necessarily the most representative TSS (Kawaji et al. 2006). Therefore, I carried out an analysis of the distribution of 5'ends of ESTs available in the UCSC genome browser database within the predicted promoter region. To this end, I chose a sliding window approach to identify the 10 bp window within the promoter fragment containing the most 5'ends of ESTs with identical directionality to the gene. In all analysed cases the centre of the 10 bp window is located in close proximity to the predicted core promoter region, i.e. with in average 13 bp or 17 bp distance to the 5' - RACE verified or GenBank mRNA predicted TSS, respectively,

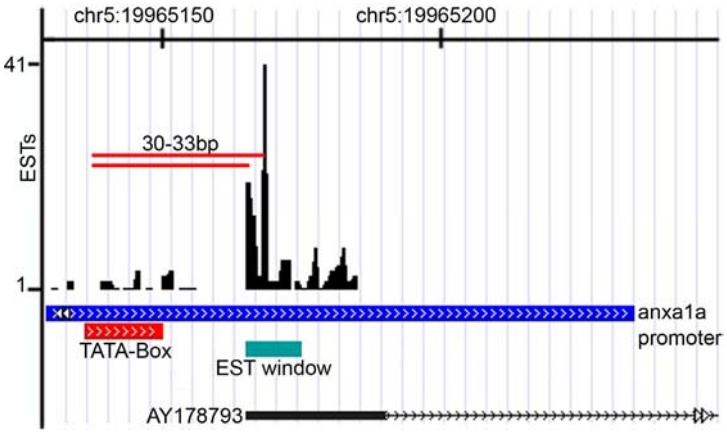


Figure 3.1| Core promoter features of analysed promoters. Shown is a schematic depiction of the *anxa1a* promoter. The blue horizontal bar depicts the 3'-end of the cloned promoter region. Black vertical bars indicate the distribution of ESTs. Additionally shown are the location of the 10bp-EST window (turquoise), the GenBank mRNA AY178793 (black) and the location of the TATA-Box (red). TATA-TSS-spacing is indicated.

thus providing further support to the promoter definitions. Moreover, in the case of *apoeb*, this approach identified a putative alternative core promoter (named *apoeb2*) within the cloned promoter region located approximately 370 bp downstream of the RACE verified TSS, which was additionally supported by a GenBank mRNA (**Table 3.1**). Together, the 5' end mapping approaches suggest that the cloned promoter fragments encompass downstream promoter regions that range from 23 to 385 bp (in average 94 bp). Finally, both TSS prediction methods strongly supported that the chosen upstream regions harbour a core promoter of the corresponding gene.

To investigate if additional TSS clusters were missed, I expanded the EST sliding window approach to include regions beyond the predicted promoter fragment (+/- 500 bp window around GenBank mRNA TSS). The approach verified the core promoter predictions in 97% of the analysed cases. In the case of *cxadr* it revealed that a second potential TSS cluster located directly downstream of the chosen fragment was omitted.

3.2.5. Analysis of core promoter features

To characterize the promoter regions, I carried out an in-silico analysis of promoter properties. The differential TSS distribution of core promoters suggests differential mechanisms of transcription initiation and can be used for core promoter classification (Carninci et al. 2006). Therefore, I investigated the distribution of TSSs within the core promoter region by analysing the occurrence of 5'-ends of all GenBank mRNAs and ESTs in a +-75bp window encompassing the EST based TSS prediction. I considered the TSS distribution as dominant if more than 20% of 5'-ends are located at a single nt and as broad if the putative TSSs were distributed more widely. If less than 20 5'-ends were available no conclusion was made (**Table 3.1**).

Core promoters can be further classified by the occurrence of CPEs within the core promoter region and their presence may be indicative for differential modes of transcription initiation. Therefore, I investigated the occurrence of known CPEs. To this end, I searched a 400 bp window encompassing the mRNA based core promoter prediction for the presence of CPEs. The region chosen contains downstream sequences not included in the cloned fragments to check if downstream core promoter elements were missed. Positional weight matrices (PWM) for almost all known CPEs (**Figure 1.3**) were extracted from the JASPAR database (Portales-Casamar et al. 2009). The sequences were scanned for the occurrence of CPEs using the PWM search program Possum using default settings (Fu et al. 2004). No PWM was available for XCPE2, therefore I performed a pattern match approach using the consensus sequence presented in **Figure 1.3**. I only considered a CPE hit as valid if it was located in

correct distance to any predicted or verified TSS within the core promoter region (**Figure 1.3** and **3.1**). Based on this analysis the TATA-box was the only CPE detected in correct TSS spacing. It was found between -33 to -29 in relation to either the EST, 5' - RACE or mRNA supported TSSs in the *anxa1a*, *apoeb*, *gtf2a1*, *her7*, *krt4* and *vox* promoters. Additionally, if less stringent PWM search conditions are used the *tbp* promoter contains a TATA-like sequence (TLS) located at -34 potentially displaying a non-canonical TATA-box. No further correctly positioned CPEs could be found in correct TSS spacing using lower stringency settings.

To check whether the promoter regions are associated with CpG islands (CGI), I searched a +-2000 bp region around the GenBank based TSS for the occurrence of CGIs using the CpG island explorer (Wang and Leung 2004) and CpG island definitions proposed by Takai and Jones 2002; Han and Zhao 2008: minimum length 500 bp, GC content greater than 55% and observed CpG/expected CpG of 0.65. Using these definitions only the *pcbp2*, *sft2d3*, *sox3* and *thy1* promoter regions were associated with CGIs. However, only the cloned promoter fragments of *pcbp2* and *sft2d3* partially overlap with these CGIs.

3.2.6. Functional analysis of promoter constructs

In order to functionally analyse the generated promoter constructs, I carried out a transient transgenic reporter gene assay. To this end, I injected the promoter constructs into zygote stage zebrafish embryos. The reporter gene expression was analysed at the 50% epiboly/shield stage, i.e. around 6 hours post fertilization (hpf) (Kimmel et al. 1995). I elected the 50%/shield stage to allow for sufficient time for the expression of the reporter gene and

GFP maturation as it is 3 hours post MBT and thus 3 hours after the start of zygotic transcription (Kane and Kimmel 1993). Moreover, I assumed that, it is still early enough to minimize indirect effects as the protein repertoire is still limited in comparison to older embryos as differentiation processes have just commenced.

At the 50%/shield stage, embryos were fixed and reporter gene activity was analysed by whole-mount immunostaining. I chose to carry out the detection of reporter gene expression immunohistochemically as the pGlow-TOPO vector contains a *gfp* mutant called *cycle3-gfp*. This *gfp* mutant exhibits high fluorescence in mammalian cells, but the fluorescent signal was barely detectable in zebrafish, thus hampering a fast analysis of transient transgenic zebrafish embryos at the fluorescence level. More importantly, the comparison of promoter activities required analysing embryos of exactly the same stage. However, early zebrafish embryos develop very rapidly, therefore the time window to analyse reporter gene activities at a given stage was relatively short. This would have made it relatively difficult to compare reporter gene activities of control and treatment groups efficiently, but it was easily accomplished by the parallel fixation and subsequent immunostaining of embryos.

3.2.6.1. Evaluation of immunohistochemical detection of reporter activity

The pGlow vector had not been previously used in zebrafish embryos. Therefore, I carried out control experiments to test for background activity caused by the presence of cryptic promoters in the vector backbone. Additionally, I wanted to control for non-specific binding of the used antibodies to embryonic proteins. To this end, I injected a negative control

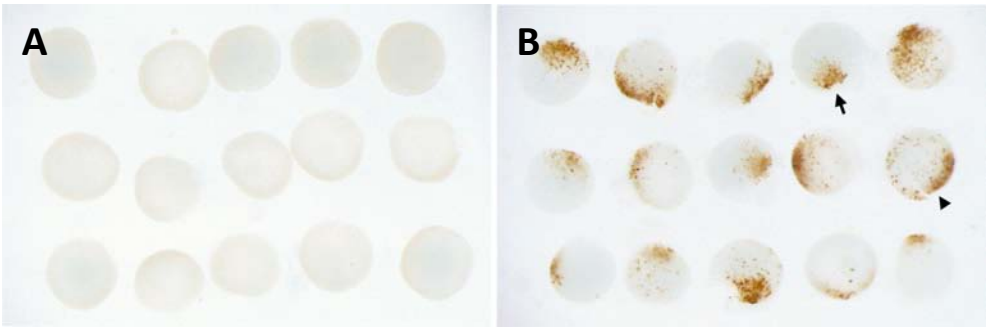


Figure 3.2| Control immunostainings. (**A,B**) Whole-mount immunostainings of 50%/shield stage embryos injected with (**A**) the negative control construct *apoeb_rev:gfp* (0%, n=72) or (**B**) the positive control construct 1.3 *ntl:gfp* (100%, n=42). The brown colour indicates mosaic reporter gene expression. *1.3 ntl:gfp* drives reporter gene expression in the embryonic shield (arrow) and margin (arrowhead).

construct containing the *apoeb* promoter in reverse orientation. No staining was observed in embryos injected with the *apoeb_rev:gfp* construct (**Figure 3.2A**).

To test if the chosen reporter vector and the detection method allow the analysis of zebrafish promoters in a transient transgenic assay, I used the *1.3 ntl* promoter as a positive control. The strong expression in the margin and shield observed after immunostaining (**Figure 3.2B**) of *1.3ntl:gfp* injected embryos is identical to the previously seen pattern using a different reporter vector (Ferg 2008). In summary, I concluded that the chosen assay is suitable to analyse promoter function in early zebrafish embryos.

3.2.6.2. The majority of promoter constructs are functional in 50%/shield stage embryos

In order to identify promoters which are functional at the 50%/shield stage, I carried out a pre-screen experiment in which I tested the generated promoter constructs and analysed reporter gene activities by immunostainings. Embryos injected with the *cxadr:gfp*, *tppp3:gfp*, *her7:gfp* and *zgc:66242:gfp* constructs showed no reporter signal at the analysed stage. Additionally, injection of the *mkks:gfp* construct led to very weak, spurious activity in a very

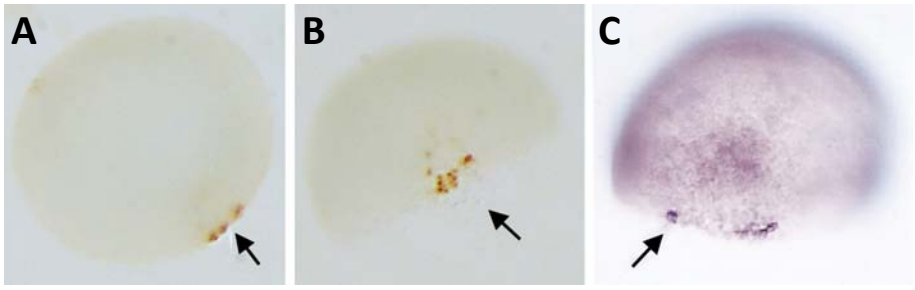


Figure 3.3 | Expression pattern driven by *ndr1:gfp*. (**A-B**) Whole-mount immunostaining of *ndr1:gfp* injected shield stage embryos: (**A**) animal pole view and (**B**) lateral view. Arrows point to the reporter gene expression in the putative shield region. (**C**) *in-situ* hybridization of shield stage embryos indicating *ndr1* mRNA expression in a small population of cells in the shield region (taken from Rebagliati et al. 1998). Arrow points to *ndr1* mRNA expression in the embryonic shield. In all shown examples the yolk ball was removed.

small number of embryos. Since these promoter fragments did not drive sufficiently detectable reporter gene expression, I excluded these five constructs from further analysis. The remaining 28 promoters were active in 50%/shield stage embryos (**Table 3.2**), whereas strong activity differences in the level of reporter gene expression driven by distinct promoters could be observed (**Table 3.2** and **Figure 3.5D**).

With the exception of *ndr1:gfp* the analysed promoter fragments did not show a restricted expression pattern. The mosaic expression was either ubiquitous, ectopic or the endogenous expression pattern is unknown. In shield stage embryos endogenous *ndr1* is expressed in a small population of superficial cells in the shield region (Feldman et al. 1998). The *ndr1:gfp* construct gave rise to a very specific *gfp* expression pattern recapitulating the endogenous expression pattern of *ndr1* (**Figure 3.3**), suggesting that the *1.3 kb ndr1* promoter contains sufficient regulatory information to control the tissue specific expression of *ndr1* in shield stage embryos. If *ndr1:gfp* recapitulates the endogenous expression pattern at different

developmental stages or if the GFP expression does indeed co-localize with endogenous *ndr1*-mRNA was not analyzed as it was beyond the scope of this study.

3.2.7. Analysis for TBP dependence of isolated zebrafish promoters

In zebrafish pre-MBT embryos TBP protein levels are barely detectable, but maternally inherited *tbp* mRNA is abundant. The maternal *tbp* mRNA is translated during MBT and the protein expression peaks at early gastrula stages, thus rising TBP protein levels coincide with the activation of zygotic transcription during MBT (**Figure 1.7**) (Muller et al. 2001; Bartfai et al. 2004). To block TBP function, I carried out a knock-down of TBP using a TBP-antisense morpholino oligonucleotide (TBP-MO) to inhibit the translation of maternal and newly synthesized zygotic *tbp* mRNA. The TBP-MO is complementary to the sequence region at the ATG start codon of the *tbp*-mRNA, thus inhibiting the initiation of translation. As control I used a standard control morpholino oligonucleotide (c-MO), which exhibits no sequence similarities to the zebrafish genome. The reliability and efficiency of the knockdown conditions were confirmed in Muller et al. 2001.

To test promoters for TBP dependence, the promoter constructs were either co-injected with c-MO or TBP-MO. Therefore, I prepared injection solutions containing an equal amount of a particular *promoter:gfp* construct (50-100 ng/µl) plus either c-MO or TBP-MO and injected them into zygote stage zebrafish embryos. The embryos were fixed when c-MO injected embryos reached the 50%/shield stage. Since the TBP depletion led to a severe embryonic phenotype it was impossible to phenotypically infer the embryonic stage of TBP morphants, therefore embryos injected with a particular promoter construct derived from

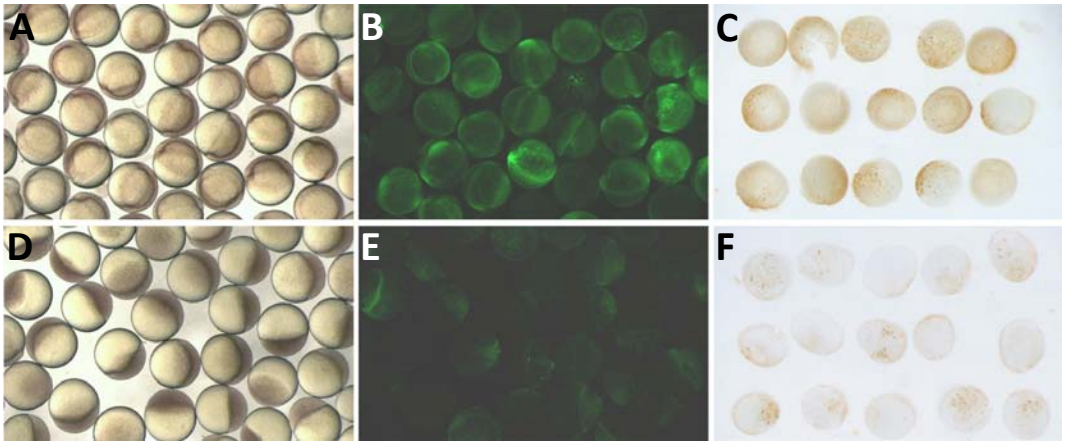


Figure 3.4| Comparison of different detection methods. (**A-F**) Representative groups of shield stage embryos injected with *β-actin:yfp* and (**A-C**) c-MO or (**D-F**) TBP-MO, respectively: (**A,D**) Brightfield view, (**B,E**) YFP channel view of **A,D** and (**C,F**) antibody staining. Embryos shown in **C,F** derived from the same injection experiment as the embryos shown in **A,B,D,E**.

the same batch of eggs to ensure that comparable stages were used in the subsequent comparison of promoter activities. Staining reactions were observed under the stereomicroscope and control and treatment staining reactions were stopped in parallel.

3.2.7.1. Immunodetection of GFP allows comparison of promoter activities

To control if the immunohistochemical detection of GFP allows the qualitative and semi-quantitative detection of differential reporter gene expression, I carried out an antibody staining of embryos injected with a *β-actin:yfp* construct. This constructs contains a 4.5 kb promoter region of the carp *β-actin* gene, which is known to be a TBP dependent promoter (F. Mueller, unpublished observation). I analysed the injected embryos for reporter gene expression by fluorescence microscopy. Subsequently, I carried out an immunostaining using the very same set of embryos. As expected, the reporter gene activity was drastically reduced after TBP knockdown, which could be detected using either method (**Figure 3.4**).

Therefore, I concluded that the immunohistochemical staining provides a sufficient readout for the comparison of reporter gene activities in early zebrafish embryos.

3.2.7.2. The majority of promoter activities are not significantly changed

To investigate the TBP dependence of promoters of early expressed zygotic genes during early zebrafish development, I tested the 28 functional promoters in TBP morphants as described above. Strikingly, the vast majority of promoters were still active after TBP knockdown. However, they exhibited differential responses to loss of TBP as judged by visual inspection: downregulation, not apparently changed and upregulated (**Table 3.2**).

To obtain a semiquantitative numerical readout of the reporter gene activity, I quantified the staining by counting the number of brown pixels above a certain threshold (positive pixels) within acquired images using the Aperio Imagescope software (**Figure 3.5B,C**). To contrast promoter activities of c-MO and TBP-MO injected embryos, the average number of positive pixels per embryo of control and treatment group was compared. The semi-quantitative image analysis allowed contrasting absolute and relative changes of promoter activities (**Figure 3.5D** and **3.6**), which matched the visual impression in all cases. To check whether there is a statistically significant difference between the two populations, I have applied a Mann-Whitney-U test. I used rather stringent criteria (p-value cut-off of 0.01), since the reporter gene expression varies considerably between single embryos due to the mosaic expression and variations caused by inaccuracy of microinjections such as slight differences of the injected volume or in the targeting of the eggs.

To verify the obtained results, I repeated each injection experiment at least once. The injection of five promoter constructs (*tbx16*, *gtf2a1*, *cdk7*, *ccne* and *ndr1*) gave varying results in different repeats. Since the results were inconclusive, i.e. it was impossible to clearly infer the promoter response to TBP knockdown, I excluded these constructs from further analysis.

In summary, a relatively small fraction of promoters 30.4% (7/23) was downregulated, i.e. they showed significantly reduced reporter activity after TBP knockdown (**Figure 3.6**). This result is consistent with the suggested role of TBP in activating transcription of many genes during early development. An example is the *otx1* promoter shown in **Figure 3.5A**. In contrast, the majority 52.2% (12/23) of promoter activities, including the *apoeb* promoter (**Figure 3.5A**), were not significantly changed after TBP knockdown, indicating alternative mechanisms for initiation of transcription from these promoters. Interestingly, 5' RACE experiments carried out to analyse TSS usage of the *apoeb* promoter could show that the loss of TBP does not lead to the utilization of the downstream alternative promoter (Ferg et al. 2007). A surprisingly large fraction of 17.4% (4/23) showed upregulation after TBP knockdown (e.g. *atp6v1g1*), suggesting a negative regulatory role of the TBP protein (**Figure 3.6**). Interestingly, this set includes the promoter of *tbp* itself (**Figure 3.7**), suggesting a negative autoregulatory role of TBP on its own gene promoter. However, overexpression of TBP does not repress promoter activity (experiment discussed in **3.2.8**).

Table 3.2| Overview of TBP dependence analysis. The corresponding gene symbols of the analysed 28 promoter regions are shown together with the percentage of GFP-positive embryos as revealed by antibody staining and the number of analysed c-MO and TBP-MO injected embryos.

promoter region	c-MO		TBP-MO	
	GFP positive	n	GFP positive	n
anxa1a	93.26%	178	81.90%	116
apoeb	93.92%	148	93.29%	149
atp6v1g1	38.46%	130	35.71%	112
c20orf45	74.07%	81	24.66%	73
ccnb2	64.60%	113	78.18%	110
ccne	41.67%	84	49.33%	75
cdk7	67.69%	130	73.19%	138
clstn2	61.42%	127	45.60%	125
foxa	51.85%	108	25.42%	59
gngt2	29.06%	117	23.81%	126
gtf2a1	47.18%	142	46.62%	148
klf4	65.71%	105	12.21%	131
kpna2	55.07%	138	42.11%	76
krt4	93.80%	129	95.69%	116
marcksl1	97.47%	158	99.33%	149
ndr1	40.00%	125	60.83%	120
otx1	39.80%	98	3.00%	100
pcbp2	51.18%	170	35.07%	134
raver1	57.59%	191	36.90%	168
rdh10b	53.54%	99	69.86%	73
sft2d3	46.75%	154	42.96%	142
sox3	42.11%	114	8.93%	112
tbp	52.27%	220	92.68%	205
tbpl1	95.49%	133	96.49%	171
tbx16	51.39%	144	50.63%	160
thy1	53.06%	98	9.38%	96
tram1	58.59%	99	56.78%	118
vox	94.74%	95	91.72%	145

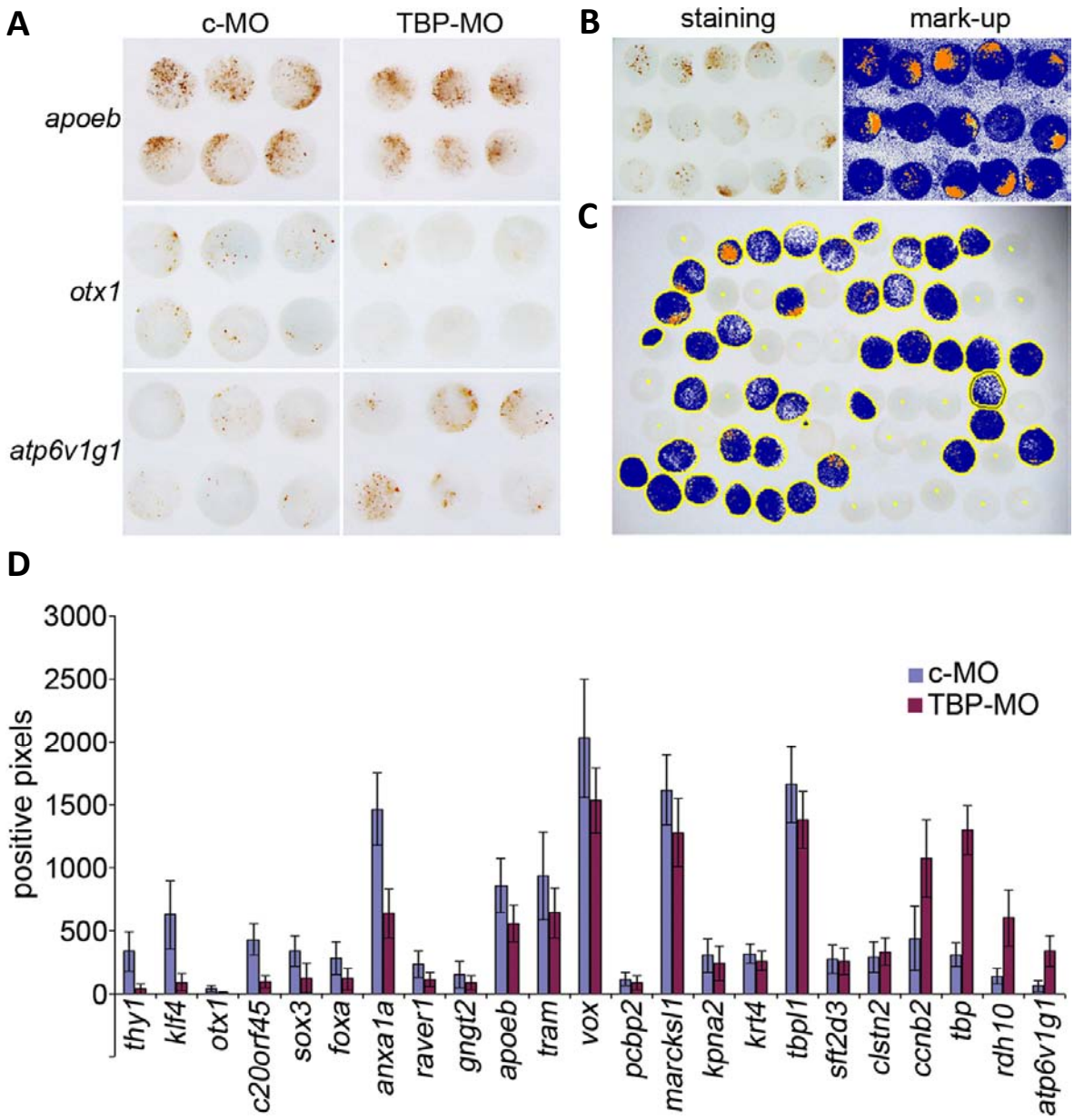


Figure 3.5| Semiquantitative analysis for TBP dependence of zebrafish promoters. (**A**) Representative examples of immunostained embryos injected with *promoter:gfp* constructs as indicated on the left and c-MO or TBP-MO, respectively. (**B,C**) Illustration of the image analysis carried out. (**B**) Immunostaining image (left) and positive "mark-up" image (right) illustrating the working principle of the Aperio Imagescope software. Yellow and orange pixels in the right panel indicate quantified pixels (positive pixels) in the left panel. (**C**) Annotation layers (yellow circles) around each expressing embryo with mark-up image superimposed, non-expressing embryos were considered as 0. (**D**) Comparison of promoter activities. The bar chart shows the average number of positive pixels per embryo indicating GFP activity in immunohistochemically stained c-MO (blue) and TBP-MO (purple) injected zebrafish embryos. Error bars indicate the 99% confidence interval of the mean.

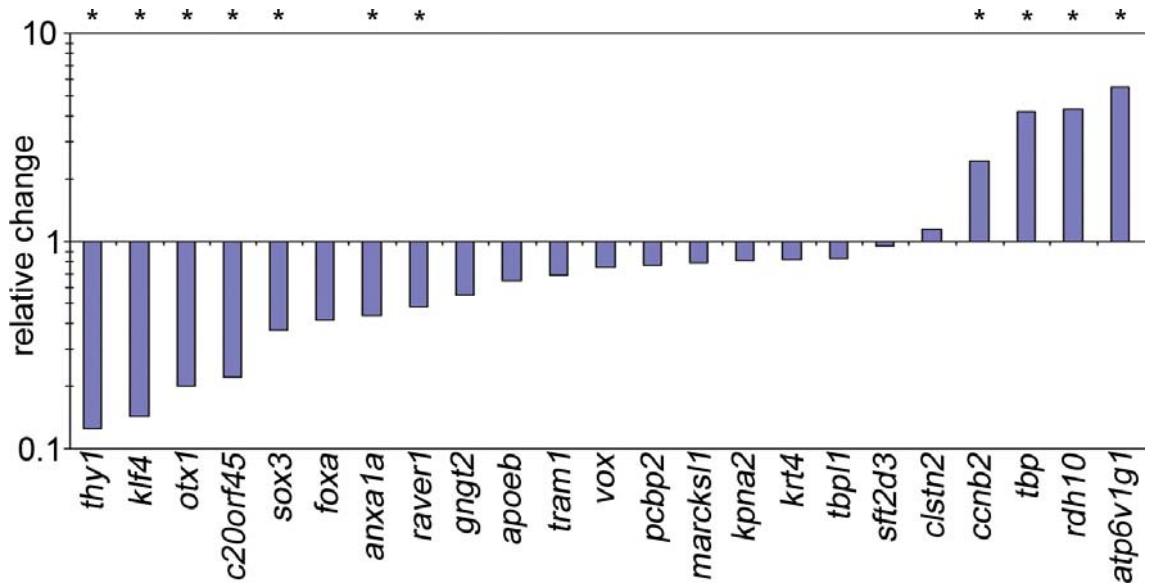


Figure 3.6| Identification of TBP response groups. Relative changes of promoter activities after loss of TBP given as the ratio of number of positive pixels between c-MO and TBP-MO injected embryos presented on a log scale. Asterisks indicate statistical significance by Mann-Whitney-U test at a p-value cut-off of 0.01.

3.2.7.3. No correlation of TBP dependence with promoter features

The presence of CPEs could indicate the functional requirement of certain components of the basal transcription machinery (Smale and Kadonaga 2003). Therefore, I aimed to investigate whether the occurrence of CPEs correlates with the TBP dependence of promoters. The TATA-box is the only CPE found and it is present in four promoters analysed for TBP dependence. It was found in the TBP dependent promoter of *anxa1a*, which is consistent with the function of the TATA-box as TBP recognition site. However, it was also present in the *vox, krt4* and *apoeb* promoters, which exhibited no significant change in activity after TBP knockdown. Additionally, the *tbp* promoter, which is upregulated after loss of TBP, contains a TATA-like sequence (TLS) element. Judging from this small set of TATA- or TLS containing promoters the presence of a TATA-box appeared not to be an indicator for TBP-dependence. Altogether, the low abundance of CPEs hampered a statistical evaluation

of correlation. For the same reason, no correlation with CGI occurrence could be established, as an overlap with CpG island was observed for only 2 promoter fragments.

It has been reported that sharp peak promoters are associated with tissue specifically expressed genes and the presence of a TATA-box, suggesting that TBP dependent mechanism control initiation of transcription from these promoters (Carninci et al. 2006). In this analysis no correlation could be established, because promoters for which a TSS distribution judgment could be made were evenly distributed within TBP response groups.

3.2.8. Promoter responses are specific to the loss of TBP

Morpholinos are commonly used to study gene function in zebrafish and other model organisms. However, they can cause non-specific effects due to the binding to unpredicted sequences in non-target transcripts or the non-specific activation of apoptotic pathways causing neural cell death, while the latter is most prominent in older embryos (>1 day post fertilization) (Ekker and Larson 2001; Robu et al. 2007). To check for potential off-target effects, I carried out control experiments to investigate if the observed promoter responses are specific to the loss of TBP protein function.

I subcloned the *tbp* promoter from the *tbp:gfp* construct using primers linked to Sal-I and Nco-I restriction sites into the *pCS+yfp* vector. The hitherto generated *1.4tbp:yfp* construct exhibits upregulation after loss of TBP as seen with the *tbp:gfp* construct (**Table 3.3 and Figure 3.7B,C**). This construct was used in a double injection experiment, in which I initially co-injected the *1.4tbp:yfp* plasmid (50 ng/µl) with TBP-MO and a *β-actin:cfp* construct (20

ng/µl) that contains the same regulatory elements as discussed in **1.2.7.1.**. Then, I split the injected embryos into three batches and injected each group a second time with different reagents: (I) water, (II) *is30*-mRNA (100 ng/µl), which encodes for an *Escherichia coli* transposase that causes no observable phenotype in zebrafish and therefore served as control mRNA (Szabo et al. 2003) and (III) *xtbpc*-mRNA (100 ng/µl), which is a TBP-MO resistant *Xenopus tbp* mRNA (*xtbpc*), which encodes for the biologically active C-terminal core domain (aa 104-297) and is sufficient to fulfil TBP function in eukaryotic cells (Veenstra et al. 2000; Schmidt et al. 2003). Since the embryos had received the same dosage of reporter constructs and TBP-MO, differences in promoter activity between treatment groups could only be caused by the molecules injected in the second round, thus providing an internally controlled experimental setup. As expected, I observed a decrease in *1.4tbp:yfp* activity and an increase in *β-actin:cfp* after *xtbpc* mRNA injection in comparison to *is30* mRNA or water injected embryos. Additionally, the *xtbpc* mRNA injected embryos exhibited a less severe phenotype than the water or *is30* mRNA injected embryos suggesting rescue of the TBP morphant phenotype (**Table 3.3** and **Figure 3.7C-F**).This indicates that only the *xtbpc* mRNA can rescue the effects caused by the TBP Morpholino. Therefore, at least for the *tbp* promoter and the *carp β-actin* promoter the responses are specific to the loss of TBP function. When I carried out a similar experiment using c-MO instead of TBP-MO, no significant change in reporter gene expression could be observed after water, *is30* mRNA or x*tbpc* mRNA injection, suggesting that overexpression of xTBPc or IS30 does not have an effect on promoter activity. Moreover, it shows that despite the potential negative

regulatory role of TBP in regulating its own gene promoter, the overexpression of xTBPc does not repress *tbp* promoter activity (**Table 3.3**).

As an additional specificity control I co-injected the *1.4tbp:yfp* construct with an alternative TBP-MO. This TBP-MO 2 targets sequence regions in the 5'UTR of the *tbp* mRNA, therefore inhibiting translation. TBP-MO 2 injected embryos exhibited a very similar TBP-knockdown phenotype. Moreover, TBP-MO 2 injection led to a similar upregulation of the *tbp* promoter as seen with the previous TBP-MO (**Table 3.3** and **Figure 3.7G**). Thus, the TBP MO2 injections further support that the observed promoter responses are specific to the loss of TBP.

Table 3.3| Overview of specificity control experiments. Shown are the percentages of expressing embryos together with the number of embryos analysed. For further details see text and **Figure 3.7**.

single injections	c-MO		TBP-MO	
	expressing embryos	n	expressing embryos	n
1.4tbp:yfp	40.06%	285	70.09%	254

single injections	c-MO		TBP-MO 2	
	expressing embryos	n	expressing embryos	n
1.4tbp:yfp	35.19%	216	74.88%	207

double injections	water		*is30*		*xtbpc*	
	expressing embryos	n	expressing embryos	n	expressing embryos	n
1.4tbp:yfp **+ TBP-MO**	75.45%	234	81.50%	241	62.38%	226
β-actin:cfp **+ TBP-MO**	44.02%		39.00%		46.02%	
1.4tbp:yfp **+ c-MO**	63.63%	161	62.60%	176	61.33%	187
β-actin:cfp **+ c-MO**	52.76%		53.98%		49.20%	

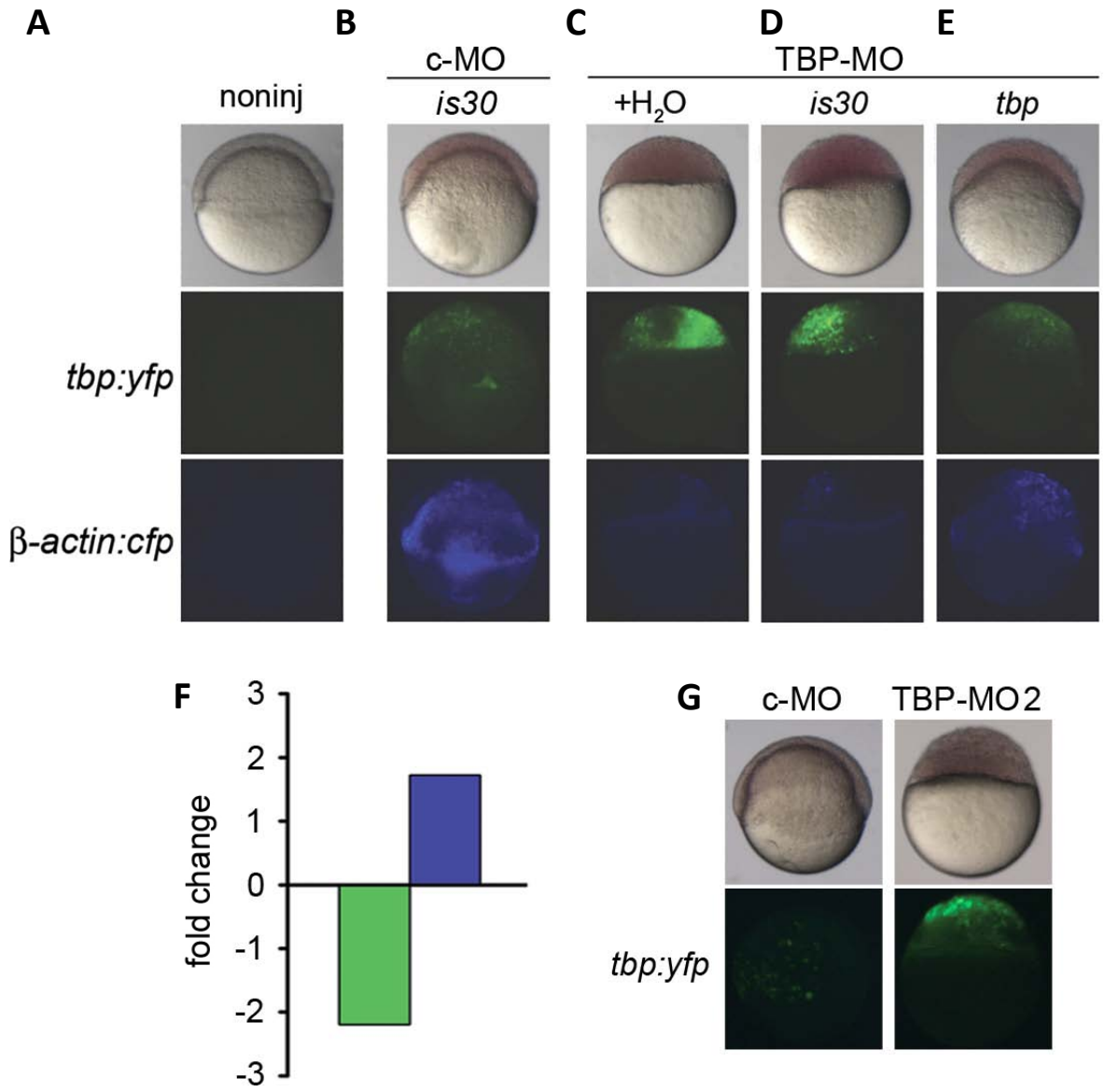


Figure 3.7|Specificity control experiments. (**A-E**) Rescue of promoter responses and TBP-MO phenotype. Shown are representative embryos at 6 hpf: (**A**) non-injected embryo, (**B-E**) embryos injected with *1.4tbp:yfp* and *β–actin:cfp* and c-MO or TBP-MO, respectively. The embryos were split into batches and exposed to a second round of injection as indicated beneath the horizontal line. (**F**) The bar chart shows the fold changes of *1.4tbp:yfp* (green) and *β-actin:cfp* (blue) activity in TBP depleted embryos upon double injection involving either *is30* or *xtbpc* mRNA (as indicated in **D** and **E**). Shown is the ratio of blue and green pixels measured using the Aperio Imagescope software. (**G**) Alternative TBP-MO injection. Representative embryos injected with *1.4tbp*:yfp and either c-MO or TBP-MO 2.

3.3. Discussion

3.3.1. Prediction of TSSs and core promoters

The isolation and functional characterization of promoters requires the correct identification of the TSSs of genes. TSSs can be inferred experimentally by analysing the 5' ends of transcripts on a gene-by-gene basis (e.g. 5'-RACE, primer extension or nuclease protection assays) or by hybridization and sequencing based genome scale approaches (e.g. tiling arrays, 5' SAGE or CAGE) (Sandelin et al. 2007). Alternatively, TSSs can be predicted with reasonable accuracy using available transcript associated sequences like cDNAs or ESTs (Suzuki et al. 2001b; Ohler et al. 2002).

In this study, the TSSs were mapped by a combination of experimental and bioinformatic methods. For a subset of genes 5' RACE sequences were available, thus providing experimentally verified, high quality TSS data. For the remaining promoters I identified the TSS regions by mapping the most upstream 5' ends of cDNAs or ESTs. Since the positions of 5'RACE and predicted TSSs for a subset of 18 genes are in good agreement, I concluded that the bioinformatic approach can predict core promoter regions with high confidence. However, all three methods have disadvantages, because they may reveal a single and not necessarily representative TSS (5'-RACE and cDNA) or lead to potentially noisy TSS predictions due to the prevalence of 5' truncated sequences (ESTs). TSS distributions of core promoters can be rather complex, thus both methods can potentially reveal the real TSS distribution only partially (Carninci et al. 2006). Therefore, the classification of the analysed promoters into broad and dominant promoters is an approximation and should be taken

with caution, especially when low numbers of TSSs were available. A more accurate TSS definition can be achieved by the utilization of large scale TSS mapping approaches. In fact, the TSS distribution and promoter utilisation at different developmental stages of the zebrafish embryo is currently explored using CAGE, revealing that the zebrafish promoterome is highly complex, comparable to the situation in mouse or human cells (Carninci, P., Lenhard, B., Strähle, U., Müller, F., unpublished observation; Carninci et al. 2006).

3.3.2. Functional verification of promoters

The majority of the predicted promoter fragments are functional in the analysed stage as revealed by the transient transgenic reporter assay, therefore confirming my promoter predictions. Thus, I have generated a dataset of functionally verified zebrafish promoters, which could serve as starting point for other transgenic applications or gene regulatory studies (e.g. **chapter 5**). However, five promoters were inactive, indicating that important regulatory information is missing or that the reporter gene activity was below the detection limit of the reporter gene assay chosen. In the case of the *cxadr* promoter the lack of activity was potentially due to the lack of downstream core promoter sequences as a second prominent TSS cluster was omitted, underlining the importance of accurate TSS prediction in promoter analysis.

3.3.3. Analysis of CPE occurrence

In this work, I have analysed the occurrence of all known CPEs described in **chapter 1** by performing a PWM search, which is a commonly used approach to search for the occurrence of sequence motifs (Gomez-Skarmeta et al. 2006). Many of the known CPEs have only been functionally verified in *Drosophila* (Sandelin et al. 2007; Juven-Gershon and Kadonaga 2009) and, furthermore, the function of CPEs and their consensus sequences are poorly characterized in zebrafish in comparison to other vertebrate species. However, due to the lack of alternatives I utilized the experimentally verified sequence motifs of CPEs of multicellular organisms available in the JASPAR database (Portales-Casamar et al. 2009).

Except the TATA-box none of the known CPEs was found, potentially due to the low prevalence of CPEs combined with the small number of promoters analysed or due to alternative CPE composition in zebrafish. However, the TATA-box was found in 18% of the analysed promoter, which is higher than the 4% detected in approximately 4000 predicted zebrafish promoters (Ferg 2008). Potentially, the difference reflects the different approach and core promoter definitions used, which is frequently the cause of the varying frequency estimates of CPEs given in the literature (Juven-Gershon et al. 2008b). TATA-boxes were found in all three response groups of TBP morphants indicating that the bioinformatic prediction of a TATA-box might not be an indicator of TBP dependence or that the majority of the found TATA-boxes are non-functional or redundant, respectively. However, the establishment of a clear correlation between the occurrence of CPEs and promoter function would require the analysis of a much larger set of promoters. Nevertheless, the presence of

a TATA-box in 3 promoters that do not require TBP for activity is surprising as TBP binding to the TATA-box is one of the best studied regulatory steps of transcription initiation (Orphanides et al. 1996). However, also TBP2 was shown to be able to bind to TATA-box DNA providing a potential TBP-independent mechanism that acts on TATA depending core promoters (Bartfai et al. 2004). Alternatively, redundant or compensatory mechanisms could initiate transcription from these promoters in TBP morphants (see **3.3.4**).

3.3.4. TBP dependence analysis

In this study, I utilized the efficient block of TBP function by TBP-MO to study the functional requirement of TBP for promoter activity during early zebrafish development. By analysing 28 distinct promoters in TBP morphants, I could show that the TBP protein is differentially required for the activation of promoters during early zebrafish development, therefore complementing the transcript based analysis presented in Ferg et al. 2007. Importantly, the specificity control experiments carried out provide experimental verification of my main observations as they strongly suggest that the observed promoter responses are specific to the loss of TBP function.

The overall proportion of TBP response groups was very similar to the overall changes of transcript levels seen in a microarray study (Ferg et al. 2007). However, from 10 promoters of genes represented on the microarray only 3 promoter responses matched the transcript response (data not shown). This low agreement is potentially due to the posttranscriptional regulatory events influencing steady state transcript levels. Additionally, the analysis of

promoters in an artificial transgenic context can cause differences as the isolated fragments may lack regulatory information influencing endogenous gene response.

The analysis for TBP dependence revealed that around one third of the analysed promoters are significantly downregulated in TBP morphants, indicating that TBP is required for transcriptional activation of these promoters, which is consistent with the suggested role of TBP as a crucial core promoter recognition protein and as a key player in activating zygotic transcription of many genes during zebrafish MBT.

The majority of promoter activities were weakly or not affected after loss of TBP, thus transcription at these promoters must initiate independently of TBP in TBP-morphants. This could indicate that these promoters are indeed TBP independent, i.e. alternative mechanism are utilized also under wildtype conditions, or that redundant or compensatory mechanisms can substitute for TBP function. A potential explanation is that other members of the TBP family have complementing roles in the early zebrafish embryo. Indeed, there is evidence that TLF and TBP2 play important roles in controlling zygotic genome activation in fish and frog (Veenstra et al. 2000; Muller et al. 2001; Bartfai et al. 2004; Jallow et al. 2004; Jacobi et al. 2007). Especially, the work in *Xenopus* could show that the different TBP family members regulate distinct subsets of genes during MBT (Jacobi et al. 2007). To investigate if promoters depend on TLF and TBP2 during zebrafish genome activation would require blocking the protein function. However, the knockdown of both proteins in early embryos is hampered due to the presence of maternally inherited protein (Bartfai et al. 2004, F. Mueller, unpublished data) and would therefore require the generation of maternal-zygotic

mutants, the development of alternative technologies to interfere with gene expression in the oocyte or technologies that specifically interfere with protein function in the embryo. Another hypothetical scenario explaining the redundancy of TBP is that the canonical TFIID complex can recognize core promoters in the absence of TBP, e.g. via interaction of TAFs with the core promoter DNA. Indeed, it has been shown that TFIID can bind to promoters independently of the DNA binding activity of TBP (Martinez et al. 1994). Furthermore, the TFIID core subcomplex is stable in the absence of TBP, suggesting that TBP is only a peripheral unit of the complex (Wright et al. 2006). Hence, one could speculate that TBP function is redundant for TFIID binding to many RNAP II dependent promoters. Finally, TBP-free TAF containing complexes have been described which could also account for TBP independent transcription (Brand et al. 1999).

It has been previously shown that the transcript levels of a relatively large subset of genes are elevated in early zebrafish embryos after loss of TBP, partially due to impaired transcription dependent mRNA degradation processes (Ferg 2008). In this study I could show that the loss of TBP also leads to increased transcriptional activity of a subset of zebrafish promoters suggesting a negative regulatory role on gene transcription. Therefore, it is tempting to speculate that TBP directly represses core promoters during early zebrafish development. The observation of the negative regulatory role on transcription is in concordance with observations in *Drosophila,* which revealed that promoter occupancy of the *hsp70* promoter by TBP and certain TAFs is diminished upon activation suggesting that TBP occupancy and transcriptional activity can be inverse correlated (Lebedeva et al. 2005). Additionally, studies in Drosophila S2 cells could show that the overexpression of TBP

decreases transcription from TATA-less DPE dependent promoters, and in human cells, it was observed that TBP overexpression represses the TATA-less *NF1* promoter (Chong et al. 2005; Hsu et al. 2008).

The loss of TBP leads to increased transcriptional activity of the *tbp* promoter, indicating a negative autoregulatory feedback loop controlling *tbp* expression. This is in line with observations that TBP protein and *tbp* mRNA levels are inversely correlated at late blastula and early gastrula stages (Bartfai et al. 2004). However, the overexpression of xTBPc did not lead to repression of the *tbp* promoter, potentially indicating that *tbp* promoter activity is increased when TBP levels fall below a certain threshold, whereas steady-state protein levels are controlled by posttranscriptional regulatory mechanisms.

Finally, the usage of an alternative TBP-MO and the internally controlled rescue experiments could show that upregulation and downregulation of promoter activities are specific to the loss of TBP. The rescue is partial as the embryos display a milder, but still severely retarded phenotype. Additionally, also the promoter activities are not fully restored when compared to c-MO injected embryos. Potentially, a simple microinjection experiment cannot accurately generate TBP protein levels that would re-establish gene regulatory programs acting during MBT, especially for a protein that is dynamically regulated during early zebrafish development (Bartfai et al. 2004). Furthermore, the truncated *Xenopus* protein might not be fully capable to compensate for all zebrafish TBP functions.

Despite the apparent non-universality of TBP, it remains a crucial protein involved in transcriptional initiation of a multitude of genes and its loss severely impairs gene regulatory

programs. Thus, the loss of TBP might affect the expression of many genes that are involved in transcriptional regulation. Moreover, the chosen promoter fragments are relatively large, so they can harbour additional CREs besides the core promoter that serve as targets for regulatory proteins whose expression might be affected in TBP morphants. For example, the loss of TBP could block the expression of transcriptional activators or repressors acting at the additional regulatory sites present in the promoter fragments, therefore leading to indirect downregulation or upregulation of the promoters, respectively. Thus, the analysis carried out cannot reveal whether TBP directly regulates the core promoters of the analysed genes. The contribution of indirect effects could be minimized by analysing core promoters only, but this would require more sensitive reporter gene readouts (e.g. luciferase, qPCR), because the basal activity of core promoters is frequently very weak (see e.g. **Chapter 5**). Alternatively, the core promoter output could be boosted by linking it to a transcriptional enhancer whose activity is known to be unaffected by the loss of TBP, providing a system in which the change of activity can be caused by the change of regulatory processes acting at the core promoter only. Overall, a clear identification of direct targets of TBP could only be generated by the establishment of a link between TBP binding and functional requirement of TBP function. This would require analysing the promoter occupancy by TBP *in-vivo* and comparing it to the TBP dependence data generated. To this end, I have started to investigate TBP occupancy by chromatin immunoprecipitation (ChIP) experiments. While I have been able to establish a working ChIP protocol using an antibody known to work in zebrafish embryo extracts (anti-H3K4Me3) (Wardle et al. 2006), the ChIP experiments carried out using antibodies against human or mouse TBP have been unsuccessful (data not

shown). Thus, the direct functional requirement of TBP for core promoter activity during zebrafish MBT remains to be explored.

The approach presented here allowed the functional analysis of a few dozen promoters in a relatively short period of time and has the potential to be scaled up to the analysis of hundreds of promoters if combined with high-throughput screening protocols available (see **Chapter 4** and **5**). However, to get global insights into the direct role of TBP in activating zygotic transcription it must be complemented with the utilization of genome scale approaches. For example, the combination of ChIP with next generation sequencing (ChIP-Seq) or tiling arrays (ChIP-Chip) would allow the identification of the occupancy of promoters at different developmental timepoints around MBT. The intersection of global ChIP data with transcriptome based approaches like microarrays, CAGE or RNA-Seq would provide genome-scale insights into promoter dynamics and the role of TBP in regulating transcription during MBT. Large scale genomic data in combination with high-throughput functional analysis of promoters could have the potential to decipher the role of TBP in regulating core promoters and thus its role in activation of zygotic transcription.

4. DEVELOPMENT OF A HIGH THROUGHPUT ASSAY FOR THE AUTOMATED ANALYSIS OF CORE PROMOTER - ENHANCER INTERACTION SPECIFICITY

Foreword

The work presented in **chapter 4** and **5** is part of a highly collaborative, interdisciplinary project, which was carried out in cooperation with Dr Eva Kalmar, a former PhD student in the laboratory of Dr Ferenc Müller, as well as Dr Markus Reischl, Institute for Applied Computer Sciences (IAI) and Dr Urban Liebel, Institute of Toxicology and Genetics (ITG), both at the Karlsruhe Institute of Technology (KIT), Germany. The contributions of Eva Kalmar have already been presented in her PhD thesis (Kalmár 2009). In my thesis I will put emphasis on aspects of the project I have carried out.

My contributions to the results presented in **chapter 4** are as follows:

In collaboration with Eva Kalmar I designed and led the screening of enhancer – core promoter interactions which involved the microinjection, pre-screening, plating and imaging of embryos. The project started in summer 2007. After 6 month Eva Kalmar left the laboratory and I took over the organization, conduction and supervision of experiments. We received help in the handling of embryos and crossing of zebrafish from Marco Ferg, Andreas Zaucker, Simone Schindler, Yavor Hadzhiev, Chengyi Song and Nadine Gröbner (**4.2.1.** and **4.2.2.**). The screening microscope and associated software modules were designed and

constructed by Urban Liebel (**4.2.2.**). The custom designed automated image processing software for detection and referencing of zebrafish embryos and quantification of fluorescent reporter signal presented was developed and executed by Markus Reischl. Although, I have not contributed to the actual programming, I provided conceptual ideas. I will outline working principles and discuss applications as it is crucial for understanding of **chapter 4** and **5** (details of the algorithms can be found in Gehrig et al. 2009). Moreover, I significantly contributed to the testing and quality control of the software (**4.2.3.** and **4.2.8.**). In collaboration with Markus Reischl, I designed and carried out the segmentation of the reference embryo shape and the validation experiments. In detail, I carried out the manual segmentation of individual embryos (**4.2.4.** and **4.2.8.**). I designed and carried out the control experiments to evaluate the domain specific quantification of fluorescent signals (**4.2.5.**) and the transient transgenesis approach (**4.2.6.** and **4.2.7.**). I also designed and carried out the model experiments to test the further applicability of the developed system (**4.2.8** and **4.2.9.**).

The work presented in **chapter 4** has been partially published in Gehrig et al. 2009.

4.1. Introduction and overview

The analysis of eukaryotic genomes combined with comparative genomics approaches has led to the identification of a multitude of putative non-coding functional elements including enhancers (Visel et al. 2009). Moreover, genome-wide studies have revealed an unexpected diversity of core promoters (Sandelin et al. 2007). Thus, a main challenge of the post-genomic era is to functionally verify and characterize these elements, in particular within the complexity of a living organism (Maston et al. 2006; Visel et al. 2009). However, in order to complement large-scale genomic approaches similarly high capacity assays need to be developed that provide sufficient high-throughput screening (HTS) potential to functionally analyse large numbers of elements (Pepperkok and Ellenberg 2006).

4.1.1. High-throughput screening analysis of *cis*-regulatory elements using transgenic zebrafish embryos

The zebrafish embryo is amenable to HTS applications due to its various experimental advantages such as the large number of eggs available and its optical transparency (Zon and Peterson 2005). Therefore, it provides an attractive model system to address biological questions in the context of a developing vertebrate embryo in a large scale. The analysis of reporter gene expression in transient and stable transgenic zebrafish embryos has the potential to provide an *in-vivo* system for the efficient testing of CREs in a high throughput manner. However, while several pioneering zebrafish studies have been reported, they have remained limited in capacity due to the lack of genuine high throughput protocols that allow

spatial registration of reporter gene activities in transgenic zebrafish embryos (Ellingsen et al. 2005; Woolfe et al. 2005).

While promising advance has been reported for the automated detection of fluorescent reporter gene expression (Tran et al. 2007; Vogt et al. 2009), methods for the automated processing of tissue-specificity information in thousands of fish embryos that express fluorescent reporter genes in distinct tissues are not yet available. Since CREs can drive expression in various tissues, the exploitation of the zebrafish in the high throughput analysis of CREs remains limited. Therefore, the potential of the zebrafish embryo as a large-scale test system for CRE function would be greatly enhanced by the development of tools that allow the simultaneous and automated registration of reporter gene expression in multiple tissues or domains, respectively.

4.1.2. Challenges in zebrafish high throughput screening

The handling and experimental manipulation of thousands of embryos combined with the demand for acquisition and analysis of large amounts of data can render the HTS of zebrafish embryos to a rather challenging task (Yang et al. 2009). For that reason, zebrafish HTS approaches would largely benefit from the development of new technologies that allow the semi- or fully automated screening and analysis of embryos. Although significant advances have been reported that partially overcome the challenges by utilizing semi-automated image acquisition approaches (Tran et al. 2007), there is further demand in the establishment of HTS protocols that increase the degree of automation. Additionally, methods for automated image analysis have been developed, that can detect morphological

changes of embryonic structures or fluorescently labelled tissues (Liu et al. 2006; Tran et al. 2007; Vogt et al. 2009). However, these methods are currently restricted to the automated analysis of pre-defined embryonic features or reporter gene expression patterns, thus restricting their applicability to more specific questions.

4.1.3. Aims

The uncovered diversity of core promoters and complementing variety of core promoter binding proteins has raised the question to what extent does the core promoter contribute to differential gene regulation (Muller and Tora 2004; Sandelin et al. 2007). A suggested regulatory function of core promoters is to contribute to combinatorial gene regulation by limiting the activating effect of enhancers (see **1.4**). Indeed, there is evidence for enhancer - core promoter interaction specificity in the *Drosophila* system, but in the context of the vertebrate embryo its contribution to interaction specificities is unknown. In this context, we aimed to explore enhancer - core promoter interaction specificity in the context of the zebrafish embryo. To this end, we designed a large scale experiment in which we mapped the interaction specificities of 19 core promoters with 11 enhancers in all possible combinations in a transient transgenic reporter gene assay (see **chapter 5**). The investigation of more than 200 interaction specificities necessitated the development of a high-throughput protocol that allowed the screening of tens of thousands of live arrayed zebrafish embryos. Therefore, we developed a HTS pipeline by combining fully automated image acquisition with custom designed image analysis software. The development and

evaluation of the developed method is described in this chapter. A detailed discussion of the enhancer-core promoter interaction screen is provided separately in **chapter 5**.

In the context of this highly collaborative and interdisciplinary study, we aimed to establish a versatile HTS assay that allows detecting spatially and quantitatively various fluorescent reporter activities in zebrafish embryos without the demand of prior knowledge of reporter gene expression patterns. To this end, we have developed a flexible system that is applicable in various transgenic zebrafish screening approaches like e.g. the large scale functional analysis of CREs or transgenic phenotyping.

4.2. Results

To address interaction specificities between core promoters and enhancers in a transient transgenic reporter gene assay in a large scale, we developed and utilized a pipeline for the automated analysis of transgenic embryos which consisted of three main experimental steps: (I) the microinjection and manual preparation of zebrafish embryos for imaging, (II) the automated image acquisition and (III) the automated analysis of fluorescent reporter gene expression patterns.

4.2.1. Microinjection and embryo handling

To test enhancer-core promoter interactions and to prepare embryos for automated imaging and image analysis, we injected enhancer-core promoter constructs (see **5.2.3**) into zygote stage zebrafish embryos through the chorion together with *in-vitro* synthesized *cfp* mRNA (15ng/µl). When we assayed embryos of stable transgenic lines or wild-type embryos, we

only injected *cfp* mRNA. The CFP signal deriving from the injected *cfp* mRNA served as an ubiquitous tissue marker (**Figure 4.1**) that was important for the subsequent image processing as it provided an outline definition and was used to define the location of certain embryonic structures (e.g. the dorsal border of the hindbrain). Furthermore, the *cfp* mRNA served as injection quality control in transient transgenic assays as only successfully injected embryos expressed the CFP protein. Additionally, we chose the CFP concentration such that the generated signal was within the linear range allowing a rough estimation of injected volumes in later stages (**Appendix 2**). The embryos were raised in PTU containing fish water to prevent pigmentation, because this would have interfered with subsequent image processing and fluorescent signal analysis. Around the prim-5 stage (Kimmel et al. 1995) we enzymatically dechorionated the embryos and carried out a manual pre-screen in which we discarded CFP negative embryos from further analysis. At the prim-20 stage we anesthetized and plated CFP positive embryos in 96-well plates, which were embedded with agarose that contained a concentric depression of 0.5 mm diameter. We manually oriented the plated embryos by inserting the yolk ball into the 0.5 mm hole, which ensured the lateral orientation of the embryos and therefore a consistent lateral view in the acquired images (**Figure 4.1A**). The prim-20 stage was chosen as at this stage most vertebrate specific features and organs are recognizable allowing a reasonably accurate judgement of tissue specific activity of tested regulatory elements (Kimmel et al. 1995). Moreover, it is still early enough to observe sufficient reporter gene activity in transient transgenic embryos. Additionally, it is a commonly used developmental stage for the functional analysis of CREs (Woolfe et al. 2005).

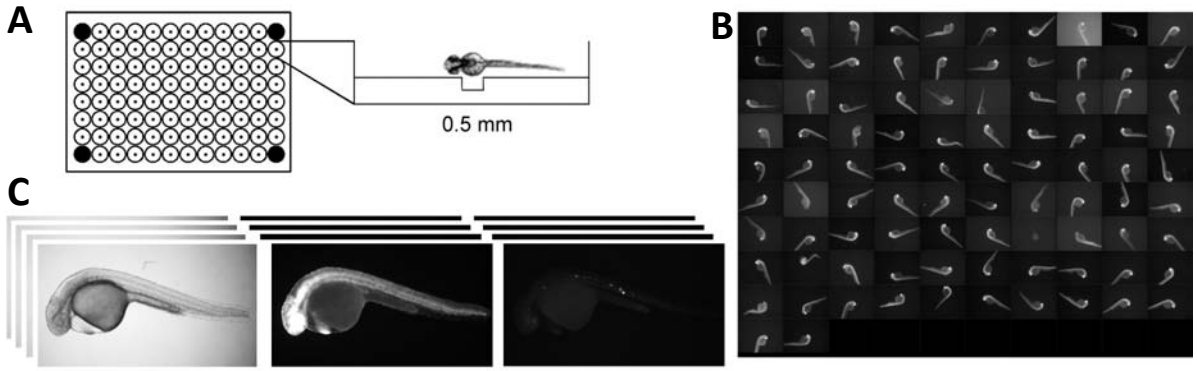


Figure 4.1| A pipeline for automated image acquisition. (**A**) Dechorionated and CFP positive embryos were sorted into agar embedded 96-well plates. The lateral orientation of embryos was facilitated by a 0.5 mm hole holding the yolk ball (see also **Figure 2.1**). (**B**) Example data showing the automatically acquired CFP images of one focal plane from one experimental plate. (**C**) Each embryo was imaged in 4 z-slices in three different channels: brightfield (left panel), CFP (middle panel) and YFP (right panel).

4.2.2. Automated image acquisition

To image embryos in high-throughput we used an automated fluorescence screening microscope (Liebel et al. 2003). After injecting, plating and orienting the embryos (see above), the 96-well plates were loaded into the stacking system of the Scan^R screening station (Olympus, Hamburg, Germany). The microscope was linked to a robotic arm that automatically transferred the 96-well plates to the microscope stage. The programmable motorised microscope stage automatically placed each well/embryo in front of the objective and sample focussing was achieved using a software autofocus. Altogether, these features allowed the consecutive and automated imaging of experimental plates with the potential of imaging thousands of live arrayed embryos within a few hours (**Figure 4.1B**).

We imaged each embryo in three different channels: the brightfield channel which provided an overall overview and was important for the image processing, the CFP channel (see **4.2.1**) and the YFP channel which provided the actual experimental data, i.e. the reporter gene

expression patterns driven by the regulatory elements used in the respective transient or stable transgenic embryos (**Figure 4.1C**). Additionally, in every channel each embryo was imaged in 4 different z-slices (i.e. different focal planes) with 55 µm distance. This captured different layers of the embryo and ensured that most of the embryonic structures were in focus on at least one image. Thus, in total we acquired 12 images per embryo (**Figure 4.1C**).

The microscopy system allowed us to assay transient transgenic embryos with significantly increased speed in comparison to manual imaging protocols. To utilize this potential efficiently, we set up a work schedule that divided the screening into 2 day intervals involving the teamwork of three people: On the first day, embryos were injected by 2 persons while a third person was controlling the mating of fishes and collecting the fertilised eggs. After injections embryos were split into batches of 100 and cleaned. On the second day two persons dechorionated the embryos, carried out the manual pre-screen for CFP signal, the plating, the manual orientation and imaging of embryos, while the third person was setting up fishes for next day's injections. Using this work plan, we were able to image embryos deriving from up to 3 x 20 different injections per week depending on egg production. On average we crossed 3 x 100-120 pairs of adult fish per week, thus the screening demands a medium sized fish facility. In total, within a period of approximately 10 weeks a team of 3 persons was able to image >30,000 embryos leading to the generation of >360,000 images, illustrating the throughput that can be achieved using this imaging pipeline (for results of the test application see **chapter 5**).

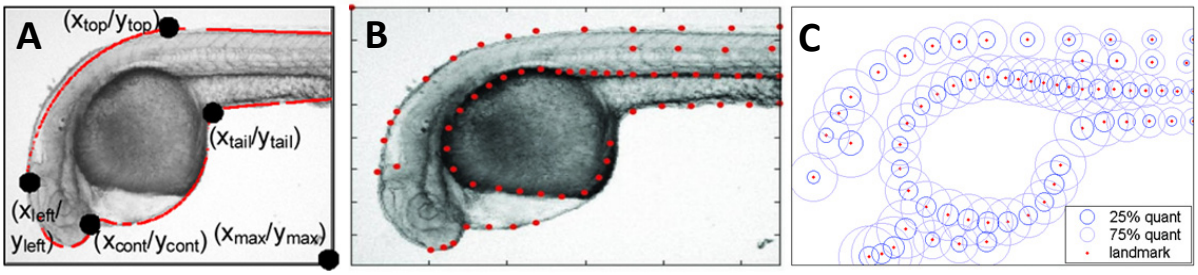


Figure 4.2| Simplified overview of embryo detection and referencing. (**A**) Illustration of some of the detected embryonic features. Black dots on embryo image indicate four characteristic points in the head and yolk region and an image coordinate. The modelled embryo outline is shown as a red line. (**B**) Embryo image with the 64 landmark points used in subsequent analysis highlighted (red dots). (**C**) The mean position of landmarks from 12,582 embryos generating a reference shape (red dots) with quantile circles (25%, 75%). Software developed by Markus Reischl, IAI, KIT.

4.2.3. Embryo detection and referencing

To overcome the obstacle of manually analysing tens or hundreds of thousands of acquired images we collaborated with Dr Markus Reischl, IAI, KIT, Germany who developed custom designed image analysis software for the automated detection, orientation and referencing of zebrafish embryos within images (details can be found in Gehrig et al. 2009).

In brief, the CFP and brightfield images were used to define the embryonic outline and the position of certain embryonic reference points. Individual images were adjusted by rotation and reflection, so all embryos were oriented the same way with anterior to the left and dorsal to the top. Erroneous images (e.g. CFP negative) were automatically discarded. The 4 z-slices acquired for each embryo were combined to obtain an image with maximum sharpness (from now on referred to as extended focus image). Then a mathematical model was created that led to the generation of landmarks (anchor points) which describe the ventral and dorsal body curvature as well as the location of certain embryonic features (**Figure 4.2A,B**). In summary, landmarks were generated for the embryonic outline, the

outline of the yolk which describes gross embryo morphology and key morphological points like the location of the notochord or cerebellum, which allow detecting the body axis and circumferential variation in morphology during development, respectively (**Figure 4.2C**). A set of 51 well registered, manually selected images of prim-20 stage embryos were used to train a learning classifier which allowed removing badly registered images (e.g. malformed embryos, wrong developmental stage) from the entire dataset.

4.2.3.1. Reference embryo shape and image warping

As a result of the referencing, each embryo was represented by 64 landmarks which describe its morphology (**Figure 4.2B**). The landmarks were used to obtain the average XY-position of the landmarks of 12,582 well registered prim-20 embryos. These average positions describe a reference embryo shape which is representative for all embryos analysed (**Figure 4.2C**). The reference embryo shape was used in all subsequent analysis steps as each experimental embryo was warped onto it, i.e. the images were deformed such that the position of the 64 landmarks of an individual embryo matched the landmarks of the reference embryo shape. As a result, the landmark positions of each experimental embryo were identical and consequently also the location of tissues was highly similar, making individual embryos directly comparable to each other.

4.2.3.2. Embryo projection to generate compound expression patterns

Our goal was to automate the analysis of the spatial reporter gene activity through visualization and segmentation of the expression pattern of a single transgene in transient and stable transgenic embryos. To reveal the tissue distribution of a particular reporter gene

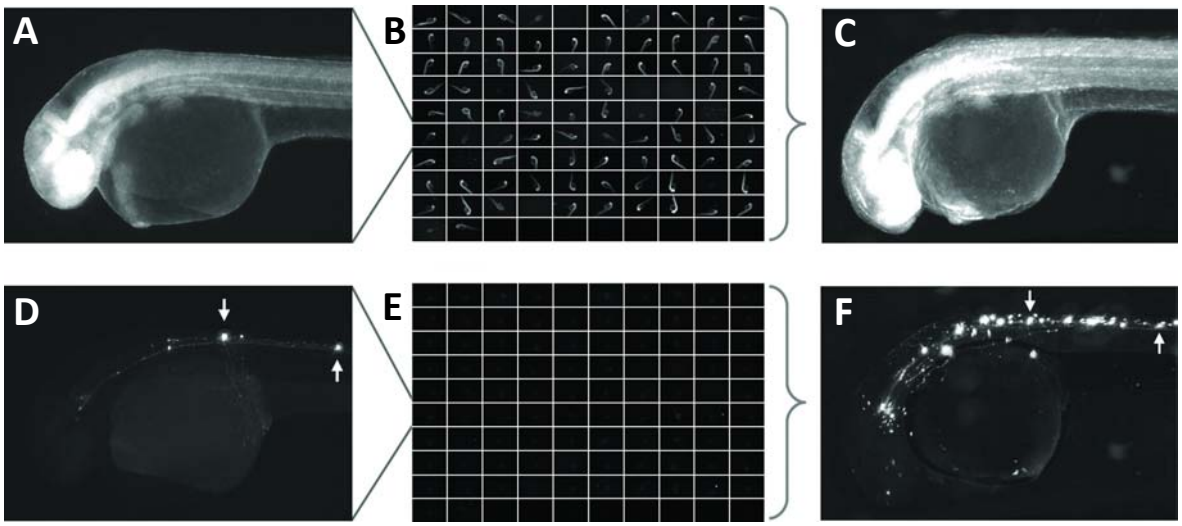


Figure 4.3| Embryo image warping tool reveals overall reporter gene expression pattern. (**A-F**) CFP (top row) and YFP (bottom row) filter views of embryos injected with *cfp* mRNA and an *isl1 zCREST2::eng2b::venus* reporter construct. Embryos shown in (**A-C**) are identical to the embryos shown in (**D-F**). (**A,D**) single transient transgenic embryo; (**B,E**) thumbnail images of entire experimental plate. (**C,F**) Maximum projection overlay of 59 individual embryos. Arrows in **F** point to the same neurons as in **D**.

activity, individual embryos were warped onto the reference embryo shape and overlaid. For visualisation purposes, the pixel brightness values at any position were derived and the expression pattern was visualized by taking the mean value (mean projections) or the maximum value (maximum projections) of each pixel (**Figure 4.3, 4.5** and **4.8**).

To evaluate the two types of overlay projection techniques, we compared the outcome of mean and maximum projections of transient transgenic and stable transgenic embryos. The projection tool proved especially useful for the analysis of transient transgenic embryos, as the injection of reporter constructs into zebrafish embryos leads to mosaic expression patterns in the F0 generation, thus the overall expression pattern driven by the regulatory elements used can be difficult to infer. This was overcome by the generation of maximum projections which generate a compound expression pattern that allows the analysis of the

overall spatial distribution of the fluorescent reporter signal. An example is shown in **Figure 4.3**, which shows embryos injected with an *isl1 zCREST2* enhancer (Uemura et al. 2005) and an *eng2b* promoter linked to a *venus* reporter gene. While single embryos showed reporter gene expression in only a small number of cells with limited information about overall specificity of CRE activity, the maximum projection generated from the full experimental batch of embryos revealed a comprehensive expression pattern with Venus positive neuronal cells in the spinal cord region. The quality of the warping and overlay approach was demonstrated by the projections carried out on the CFP images (**Figure 4.3A-C**). Maximum projections are particularly helpful for the analysis of mosaic patterns, because they highlight even weak expression in single or small numbers of cells and provide cellular resolution. In contrast, for the visualization of stable patterns usually seen in transgenic lines the generation of mean projection overlays proved superior as it can resolve intensity differences between tissues, that is less well visualized using maximum projection (**Figure 4.5A-C** and **4.8**).

4.2.4. Segmentation of reference embryo shape

Besides the visualization of tissue-specific reporter gene activity, we aimed to quantify the spatial distributions of reporter gene expression representing the tissue-specific activities of distinct enhancer-promoter combinations. To this end, we assigned each pixel in the reference embryo shape 2-dimensional position information. Therefore, we subdivided the reference embryo shape into 8 arbitrarily defined segmentation domains (for domain

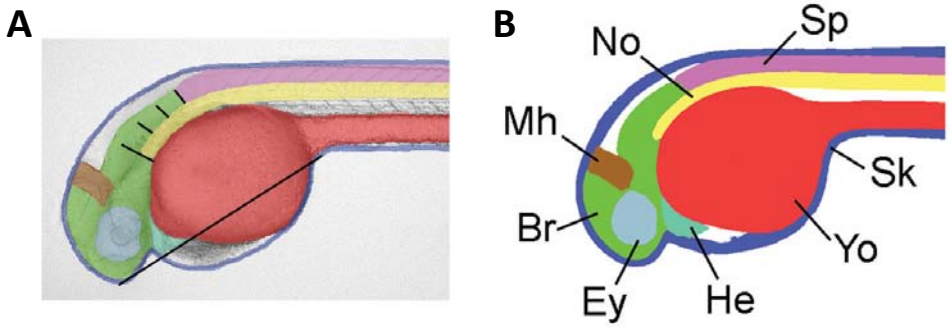


Figure 4.4| Generation of reference embryo shape. (**A**) Single embryo in which segmentation domains are marked by distinct colours. Colours are as in B. (**B**) Segmented reference embryo shape made by the reference landmarks as shown in **Figure 4.2C** and assignment of segmentation domains based on the manual referencing of 26 prim-20 embryos as shown in **A**. Abbreviations: Yo, yolk domain; Sp, spinal cord domain; No, notochord domain; He, heart domain; Mh, MHB domain; Br, brain domain; Sk, skin domain Ey, eye domain. For domain definitions see **4.2.4.1**.

definitions see **4.2.4.1** and **Figure 4.4**), each of them representing several embryonic tissues due to the 2-dimensional nature of the approach.

To segment and assign position information to the reference embryo shape, we averaged the position of eight domains from a set of 26 randomly picked well registered single prim-20 stage embryos. Therefore, we segmented these individual embryos by assigning each pixel to one of the eight domains described in **4.2.4.1** and highlighted in **Figure 4.4**. To this end, we manually marked each segmentation domain with a different colour using Adobe Photoshop (**Figure 4.4A**). Then the manually segmented embryos were warped onto the reference embryo shape, leading to a reference embryo shape in which each pixel was associated with 26 domain assignments (ideally 26 identical). A pixel in the reference shape was assigned to a domain when >30% (in the case of the skin 10%) of the 26 domain assignments were identical. As a result we obtained a segmented reference embryo shape with domain information, which allowed the assignment of position information to each pixel of experimental embryos after warping (**Figure 4.4B**).

4.2.4.1. Definition of segmentation domains

We defined the spinal cord (Sp) domain as the region which is bordered anteriorly by a vertical bar at 2 otic vesicle length posterior to the otic vesicle and includes posteriorly the region between the notochord and the skin domain of the embryo, thus including overlapping somites. The Sp - domain is bordered ventrally by the notochord (No) domain, which encompasses the notochord and extends anteriorly to the anterior end of the otic vesicle. The midbrain-hindbrain boundary (MHB) domain (Mh) contains the two prominent vertical furrows anterior and posterior to the MHB, therefore it includes parts of the posterior tectum and tegmentum as well as cerebellar tissue and tissue of rhombomere 1. Ventrally, the Mh - domain extends to the floorplate. The eye domain (Ey) contains the retina region including the lens placode and brain tissues between the eyes. We defined the brain domain (Br) as the region anterior to the Sp - domain excluding the Nc-, Ey- and Mh-domain. The skin domain (Sk) contains an 8 pixel wide stripe at the outline of the embryo, including mainly skin cells of the midsection and parts of the median fin fold. The yolk domain (Yo) contains the yolk and the yolk extension. Finally, we defined the heart domain (He) as the curved keel shaped territory bordered dorsally by the embryo proper and ventrally by the Yo – domain. The anterior part of the He – domain was arbitrarily set as the line between the anterior tip of the brain and the ventral joint of the yolk ball and the yolk extension (transverse line in **Figure 4.4A**).

4.2.4.2. Validation of segmented reference embryo shape

Since individual embryos can vary significantly in size and overall morphology, we evaluated the accuracy of domain assignment after warping. To this end, we manually segmented 29 randomly selected or 30 stage selected (i.e. matching the reference embryo shape morphology) embryos, respectively. The embryos were warped onto the segmented reference embryo shape and the previous and new domain assignments were compared. Despite differences in position, orientation and size of embryos in the randomly selected set, the validation experiment indicated an average accuracy of 81.4%, which indicates the percentage of all domains that are correctly assigned (**Appendix 2**). The accuracy for single domains ranged from 72% (He) to 97% (Yo). For the stage synchronous embryos, an average accuracy of 86% was obtained and single domain results ranged from 76% (He) to 98% (Yo). Moreover, the validation experiment revealed that certain domains were more likely to be misassigned to another domain, thus leading to over- or underrepresentation of certain domains, respectively (last row in the tables in **Appendix 2**). For example, pixels of several domains were misassigned to the brain domain indicating an overrepresentation (approx. 1.2 fold) of pixels recognized as brain.

4.2.5. Spatial quantification of reporter gene expression

To achieve automated registration of the various reporter gene activities, we utilized the segmented reference embryo shape to quantitatively analyse the spatial distribution of fluorescent signals in transgenic zebrafish embryos. Therefore, the extended focus images of single embryos were warped onto the reference shape and the intensity of pixels assigned to

certain domains was measured. The signal intensity within a domain was defined as the sum of pixel intensities that exceeded a dynamic threshold, which was determined for each individual image based on the image background. Each segmentation domain had a different size, therefore we normalised the quantification values by the size of the respective domain. Additionally, background yolk fluorescence and saturated areas were deduced. To obtain the overall expression pattern driven by a particular transgene all corresponding extended focus images of experimental embryos were analysed and the mean value for each segmentation domain was calculated. As a result, the expression pattern of each transgene was described by 8 values that indicate the area- and embryo number normalized sum of pixel intensities (expression grade, **Figure 4.5**).

4.2.5.1. Quantitative readout of spatial reporter gene activity

To test if the developed pixel count algorithm provides a quantitative readout of reporter gene activity, we designed a control experiment, in which we analysed the signal intensities of genotypes with defined reporter expression levels: non-transgenic, hemizygous (1 transgene allele per genome) and heterozygous (2 transgene alleles per genome) embryos of the *Tg(2.2shh:gfp:ABC)* stable transgenic line (Shkumatava et al. 2004). We expected that the higher number of allelic copies would lead to a higher fluorescent intensity of homozygous embryos. To determine the embryonic genotype, we identified the parental genotype of individual transgenic males and females after crosses with wild-type fishes by the mendelian frequency of transgenic embryos among the offspring. To minimize individual variation, we obtained heterozygous and homozygous embryos by two independent crosses

of a single identified homozygous female with either a wild type or a homozygous transgenic male, respectively. We microinjected non-transgenic and transgenic embryos with *cfp* mRNA and imaged them and processed the images as described above.

We compared overall reporter activities by taking the sum of pixel intensities of all segmentation domains measured. As depicted in **Figure 4.5D** the algorithm could successfully distinguish hemi- and homozygous batches of embryos**,** as indicated by the median pixel intensity count that was proportional to the expected allelic copy numbers of the GFP transgene and indicated on average double fluorescence intensity in homozygous as compared to hemizygous embryos.

The *2.2shha:gfp:ABC* transgene (**Figure 4.5A**) is highly active in the hypothalamus, zona limitans intrathalamica, the tegmentum, the floorplate and the notochord. This was correctly assigned to the domains of the Ey-, Br-, Mh- and Nc-domains (**Figure 4.5E**), as demonstrated by the comparison of mean projections and the quantification values of segmentation domains displayed in a radar plot (**Figure 4.5C,F**). Thus, the algorithm successfully quantified the spatial distribution of reporter signal in *Tg(2.2shh:gfp:ABC)* embryos within these segmentation domains.

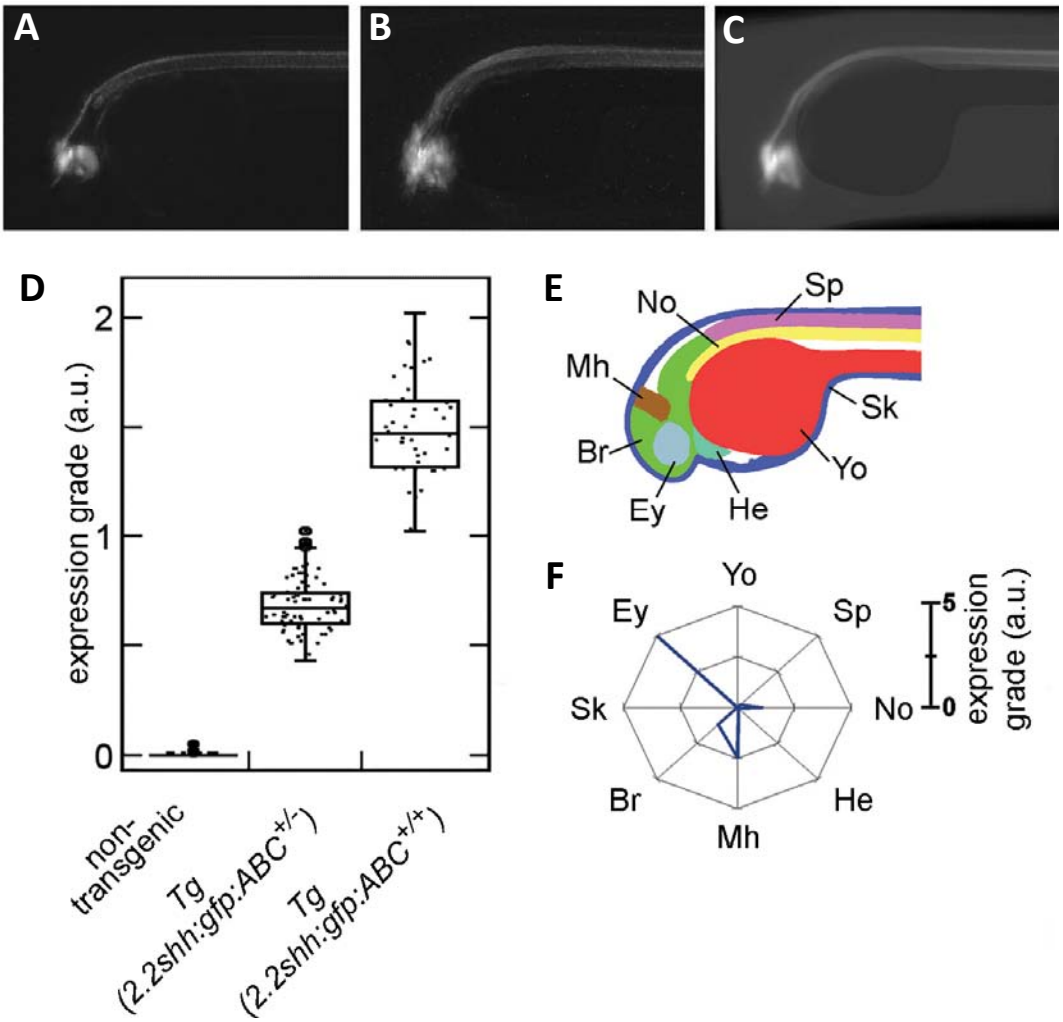


Figure 4.5| Signal quantification of embryos of the *Tg(2.2shh:gfp:ABC)* transgenic line. (**A-C**) YFP filter views of (**A**) single embryos, (**B**) maximum projection and (**C**) mean projection of the stable transgenic line. (**D**) Box plot indicating distributions of measured GFP signal intensities of non-transgenic (n = 76), hemizygous *Tg(2.2shh:gfp:ABC)*$^{+/-}$ (n = 80) and homozygous *Tg(2.2shh:gfp:ABC)*$^{+/+}$ (n = 40) embryos. Dots indicate expression grades of single embryos. (**E**) Reference embryo shape as in **Figure 4.4**. (**F**) Domain-specific distribution of fluorescent reporter signal intensities indicated as a radar plot of segmentation domains. Expression grade values are plotted along the axes and connected by a blue line. Abbreviations as in **Figure 4.4** and a.u.: arbitrary unit.

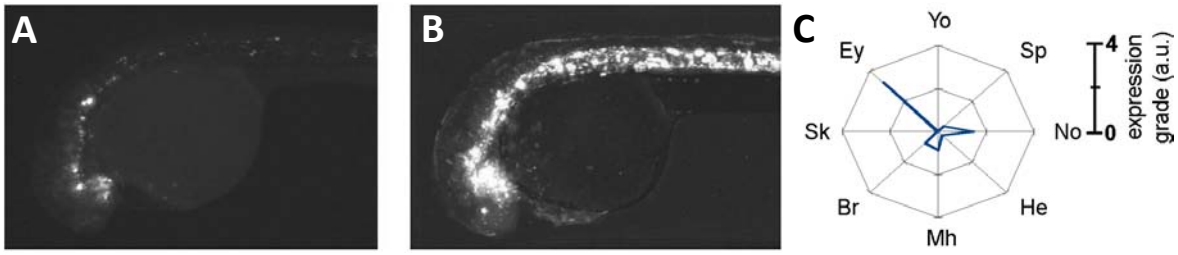


Figure 4.6| Domain specific quantification of mosaic reporter gene activity. (**A**) Single transient transgenic embryo injected with the *shha arC:krt4:venus* construct. (**B**) Maximum projection of 65 stage selected embryos injected with *shha arC:krt4:venus*. (**C**) Results of domain specific quantification of fluorescent reporter signal of embryos injected with *shha arC:krt4:venus* presented as a radar plot. a.u.: arbitrary units.

4.2.5.2. Quantification of mosaic expression patterns

The main goal of this study was to develop a HTS protocol for the automated analysis of spatial reporter signal distribution in transient transgenic zebrafish embryos. Therefore, we tested if the developed algorithm can also reveal the domain specific activity in embryos with mosaic distribution of signal. The domain-specific quantification indicated reporter gene signal in the same segmentation domains as seen in maximum projection images of embryos injected with a *shha arC:krt4:venus* enhancer-promoter reporter construct (**Figure 4.6,** for further examples see **Figure 5.4**). Furthermore, we successfully applied the developed tool in the quantification of mosaic patterns in an extensive analysis of enhancer-promoter interaction specificity investigating the spatial activity of 202 enhancer - core promoter combinations in transient transgenic zebrafish embryos. The results are discussed in detail in **chapter 5**.

4.2.6. Evaluation of the transient transgenesis approach

In this study, we chose to investigate the interaction specificity between enhancers and core promoters in a transient transgenic reporter gene assay. We generated transient transgenic embryos by the “classic” and widely used approach of simply microinjecting circular plasmids into zebrafish embryos, which results in strong mosaicism, thus demanding the generation of compound expression patterns by analysing several transient transgenic embryos (Westerfield et al. 1992).

To evaluate whether the analysis of mosaic activity can sufficiently and reproducibly reveal the overall activity of the regulatory elements used we carried out a set of control experiments.

4.2.6.1. Comparison of different transient transgenesis methods

Tol2 is an autonomously active transposon identified in the medaka fish that jumps in a cut-and-paste mechanism, i.e. it is excised from the donor site and integrates at the acceptor site (Kawakami et al. 2000). It consists of the *tol2* gene encoding for a transposase which is flanked by DNA regions important for transposition. When reporter constructs in which the transgene is flanked by Tol2 sequences are co-injected with *tol2* mRNA the transgenesis efficiency can be elevated, because the Tol2 transposase protein excises the transgene and stably integrates it as a single copy insertion. Therefore, this system is widely used for the generation of stable transgenic zebrafish lines as it increases germ line transmission rates (Parinov et al. 2004; Kawakami 2007). Additionally, it has been reported that its use can

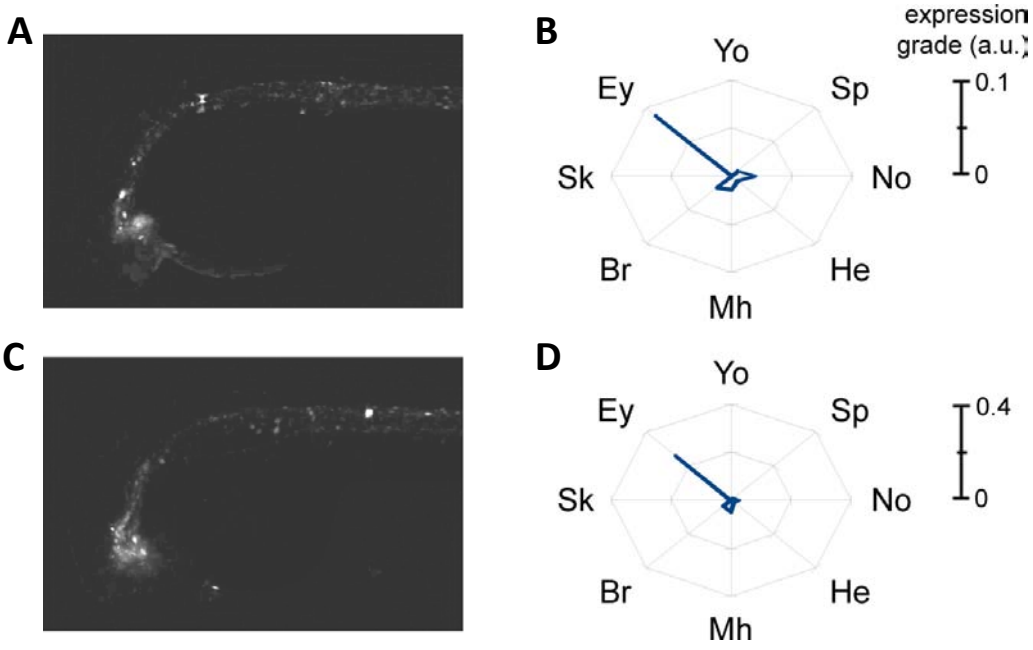


Figure 4.7| Comparison of transient transgenesis methods. (**A,C**) Maximum projections and (**B,D**) radar plots showing domain specific signal quantification indicating Venus activity in embryos injected with (**A,B**) Tol2-*shha arC::shha*-Tol2 without *tol2* mRNA (84.5%, n=84) and (**C,D**) Tol2-*shha arC::shha*-Tol2 plus *tol2* mRNA (85.4%, n=110). Numbers in brackets show the number of expressing embryos. Abbreviations as in **Figure 4.4** and for domain definitions see **4.2.4.1**.

reduce the reporter signal mosaicism observed in the F0 generation after microinjection, because early integration events increase the frequency of transgene activation among embryonic cells (Fisher et al. 2006).

To analyse if the usage of Tol2 would provide a benefit to our analysis we carried out an experiment in which we analysed the transient activity of a *shha arC::shha* construct with flanking Tol2 sites. The plasmid (10 ng/µl) was injected together with or without *tol2* mRNA (50 ng/µl), and the embryos were analysed as described above. In both cases the overall expression patterns were very similar as indicated in the maximum projections and in the radar plots shown in **Figure 4.7**. However, Tol2 co-injection resulted in an overall higher

activity of enhancer - core promoter combinations as indicated by the higher mean values of domain specific activity. This was due to the less mosaic transgene activity of single embryos.

Although the Tol2 system seemed to provide a slight advantage, we continued with the "classic" system as it provided sufficient detail to judge the overall expression pattern driven by the regulatory elements used. And, importantly, it has the advantage of technical simplicity, which was crucial for standardized microinjections in a large scale (for discussion see **4.3.6**).

4.2.6.2. Comparison to stable transgenesis

We used the overlay and projection technique to overcome the effect of reporter gene expression mosaicism in transient transgenics by generating compound expression patterns that revealed the overall spatial activity of the tested regulatory elements. Reporter signal mosaicism can also be overcome by the generation of stable transgenics, which have the major advantage that spatial CRE activity can be inferred from very low numbers of transgenic embryos. However, the generation of stable transgenic lines is rather laborious, making it technically difficult to utilize them in the analysis of CREs in a large scale.

To test whether the expression patterns reconstructed by the chosen transient transgenic assay are comparable to the continuous transgene activity revealed after stable integration, we compared embryos injected with the *2.2shh:gfp:ABC* construct to a corresponding stable transgenic line *Tg(2.2shh:gfp:ABC)*. The overlay expression patterns resulting from transient transgenic embryos were comparable to those in the stable transgenic line (**Figure 4.8**). This

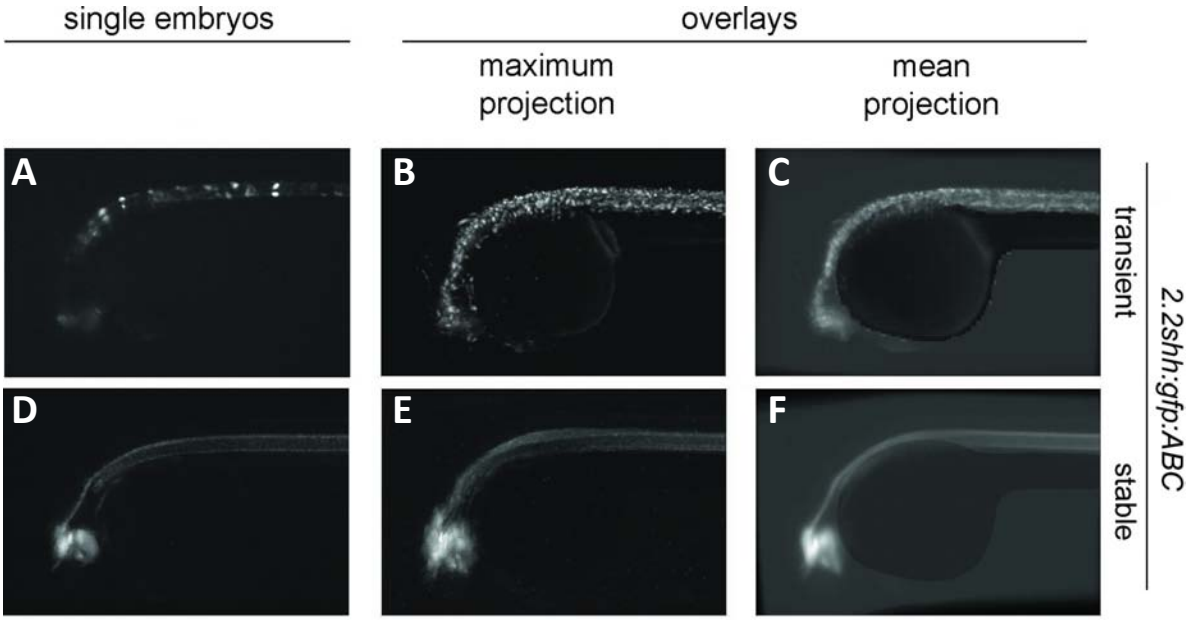


Figure 4.8| Comparison of projection and different transgenesis methods. (**A-C**) Transient transgenic embryos injected with the *2.2shh:gfp:ABC* construct. (**D-F**) Stable transgenic embryos of the *Tg(2.2shh:gfp:ABC)* stable transgenic line. (**A,D**) single embryos, (**B,E**) maximum projections and (**C,F**) mean projections are shown. All images are YFP filter views.

indicates that the reconstruction of patterns from individual embryos using the warping and projection tool can reliably reveal the overall spatial activity of the regulatory elements used.

4.2.7. Reproducibility of transient transgenesis experiments

The mosaic nature and the variable expression among embryos within an injection batch as well as variability introduced by microinjections raised the question if the transient transgenic approach leads to a reproducible reconstruction of overall expression patterns. To address that, we carried out 4 independent repeats of an injection experiment with the *shha arC::krt4* construct and processed the embryos using the image analysis tool as described. For each repeat we obtained 8 domain specific quantification values describing spatial signal distribution (**Table 4.1**).

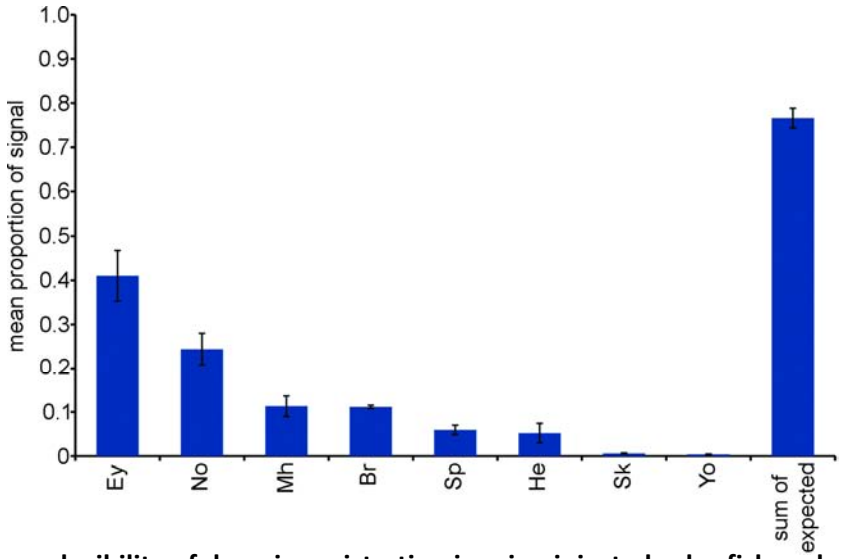


Figure 4.9| Reproducibility of domain registration in microinjected zebrafish embryos. Bar chart diagram of 4 repeats of an injection experiment with the construct *shha arC::krt4*. The mean percentages of Venus signal in the tissue domains detected in the repeats are indicated. Abbreviations as in **Fig 4.4.**. See also T**able 4.1**. Error bars indicate the standard error of the mean.

Table 4.1| Quantification of *shha arC::krt4* microinjection repeats. Column 1 and 2 show the repeat number and the number of embryos analysed. The remaining columns show the domain specific quantification results as arbitrary units (upper panel) or the percentage of signal within segmentation domains (lower panel). Expected domains of *arC* activity are highlighted in bold. The last column shows the total activity within expected domains. Abbreviations as in **Figure 4.4.**

Repeat no.	number of embryos	Yo	Ey	Sk	Br	Mh	He	No	Sp	expected domains
1	90	0.044	**5.247**	0.077	**1.196**	1.522	0.409	**2.623**	0.634	**9.07**
2	84	0.011	**1.559**	0.017	**0.443**	0.639	0.050	**0.843**	0.185	**2.85**
3	89	0.012	**2.887**	0.019	**0.628**	0.475	0.239	**0.990**	0.238	**4.51**
4	90	0.063	**2.613**	0.091	**1.153**	0.682	1.195	**3.589**	0.934	**7.36**
1	90	0.37%	**44.65%**	0.66%	**10.18%**	12.95%	3.48%	**22.32%**	5.39%	**77.14%**
2	84	0.29%	**41.61%**	0.45%	**11.82%**	17.05%	1.33%	**22.50%**	4.94%	**75.93%**
3	89	0.22%	**52.61%**	0.35%	**11.44%**	8.66%	4.35%	**18.04%**	4.34%	**82.09%**
4	90	0.61%	**25.32%**	0.88%	**11.17%**	6.61%	11.58%	**34.78%**	9.05%	**71.27%**

The *shha arC* enhancer is expected to drive activity in the hypothalamus, tegmentum and notochord of prim-20 stage embryos (**Table 5.3**). Therefore, we also calculated the expression within the expected domains of activity: notochord (No), brain (Br) and eye (Ey). While the overall signal intensity was different among repeats (**Table 4.1**), the distribution of

signal was reproducible as indicated by the mean proportion of signal within segmentation domains in **Figure 4.9**. This indicates that the transient transgenic approach can reproducibly reveal overall expression patterns.

4.2.8. Application of the image analysis tool to other developmental stages

We initially developed the screening tool for zebrafish embryos of the prim-20 stage, thus the application of the screening tool was restricted to a defined stage, although slight stage variations could be handled as demonstrated in the validation of the segmented reference embryo shape using randomly selected embryos (**see 4.2.4.2**). To test whether the developed system can be applied to other developmental stages with different overall morphology, we used the embryo detection and segmentation tool to analyse long-pec stage embryos (Kimmel et al. 1995). To this end, the algorithms were modified and we adjusted the domain definitions for embryos of the long-pec stage.

Based on 23 long-pec stage embryos a new reference embryo shape was generated (**Figure 4.10A,C**). To carry out spatial quantification of reporter signal, we segmented the long-pec reference embryo shape using 23 manually segmented single embryos (as described in **4.2.4.2**). The domain definitions were identical to the prim-20 stage with two exceptions: (I) The He - domain definition was adjusted such that the anterior border was now set by a line between the most ventral position of the eye and yolk and (II) the ventral border of the Mh – domain was newly set as the line which is tangent to the eye and parallel to the floorplate (**Figure 4.10B**). We tested the accuracy of domain assignments by using 19 stage selected, well registered embryos as described in **4.2.4.2**. The results are comparable to the results

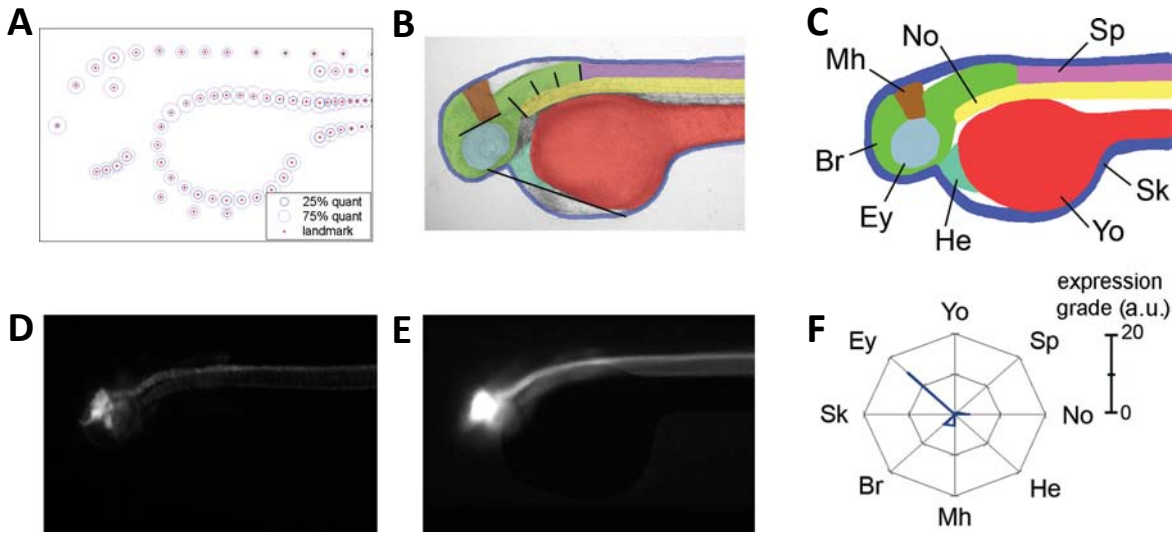


Figure 4.10| Analysis of long-pec stage embryos. (**A**) Mean positions of landmarks generated from 23 long-pec stage embryos (**B**) Manually marked segmentation domains of a single long-pec stage embryo. Colours are as in **C**. (**C**) Segmented reference embryo shape made by the reference landmarks as shown in **A**. The assignment of segmentation domains was based on the manual referencing of 23 long-pec embryos as shown in **B**. (**D**) Single long-pec embryo of the *Tg(2.2shh:gfp:ABC)* transgenic line. (**E**) Mean projection. (**F**) Domain-specific distribution of fluorescent reporter signal intensities indicated as a radar plot of segmentation domains. Abbreviations as in **Fig. 4.4**.

obtained with prim-20 stage embryos, with in average 82.5% accuracy, while results for single domains range from 74% (He) to 94% (Yo) (**Appendix 2**).

The newly generated reference embryo shape was successfully used to generate overlay projections of long-pec stage embryos of the *Tg(2.2shh:gfp:ABC)* transgenic line (**Figure 4.10D,E**). Moreover, using long-pec stage embryos of the *Tg(2.2shh:gfp:ABC)* transgenic line, we successfully quantified the fluorescent reporter gene activity in the overlapping Ey-, Br-, Mh- and Nc-domains (**Figure 4.10F**).

4.2.9. Detection of reporter gene expression phenotypes

Transgenic zebrafish embryos, in which certain tissues are fluorescently labelled, can provide a useful screening tool for the detection of phenotypes which are characterized by subtle morphological changes or changes in gene expression (e.g. Tran et al. 2007). Therefore, we wanted to investigate if the developed system has the potential to be used in fluorescent reporter gene expression based screening applications besides the analysis of CRE function. To this end, we utilized the observation that the hedgehog signalling pathway can be modulated by the overexpression of a dominant negative isoform of protein kinase A (PKI), which mimics constitutively active hedgehog signalling leading to an expansion of *shha* mRNA expression in the dorsal brain (Hammerschmidt et al. 1996; Muller et al. 2000). Accordingly, we observed that the microinjection of *pki* mRNA (10 ng/µl) leads to an expansion of GFP expression into the dorsal brain in embryos of the *Tg(2.2shh:gfp:ABC)* transgenic line as compared to control mRNA (*tol2* mRNA) injected embryos (**Figure 4.11A-D**). Importantly, we did not observe gross morphological changes at the mRNA concentration used, as this would have hampered the registration of embryos by the automated image analysis tool. When injected into hemizygous transgenic embryos and analysed with the developed system, we could successfully detect the increase in reporter gene expression and, moreover, we were able to resolve the spatial changes as indicated by the change of activity in the overlapping domains (**Figure 4.11D** and **Appendix 2**). This result suggests that the screening system is sufficiently sensitive to detect minor reporter gene expression changes that may emerge in screening situations and demonstrates its utility in gene expression phenotype analysis.

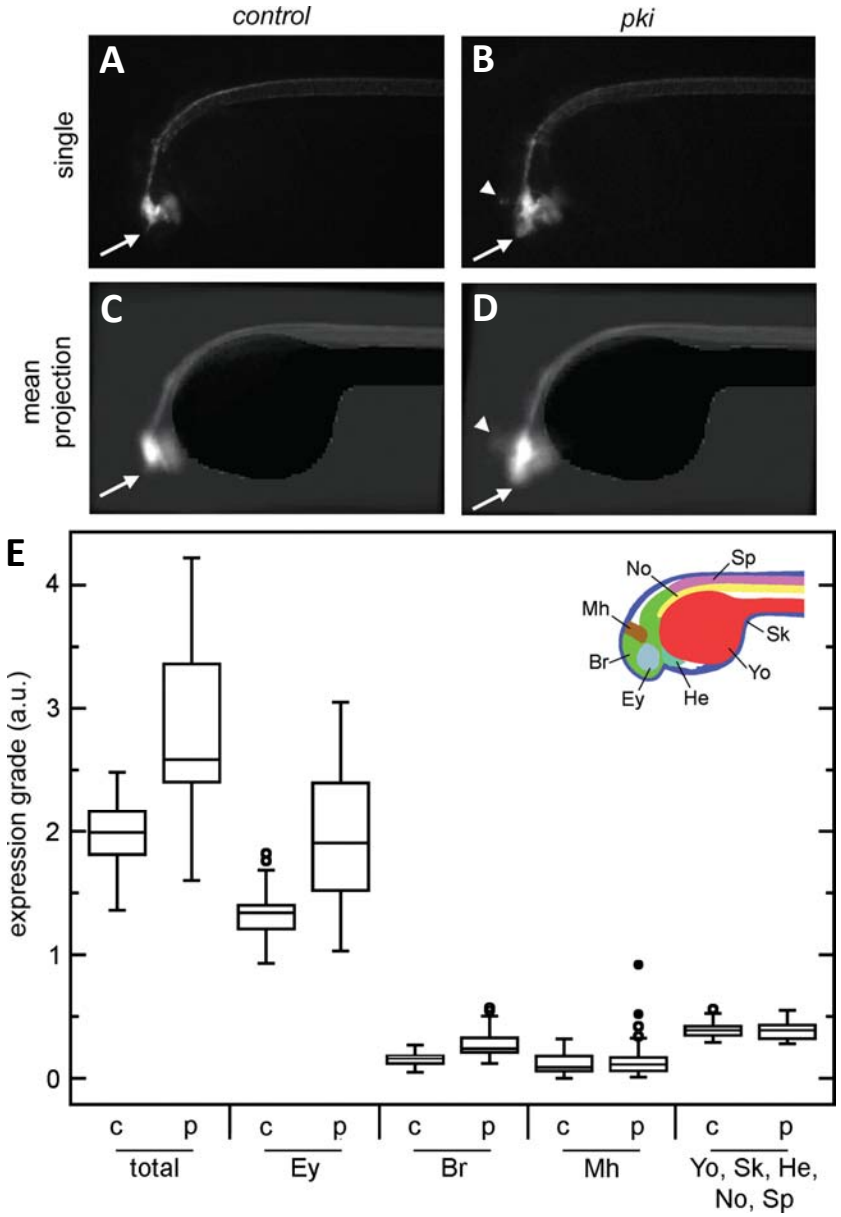


Figure 4.11| Detection of domain specific changes of reporter gene expression in a stable transgenic line. (**A,B**) GFP expression in single heterozygous embryos of the stable transgenic line *Tg(2.2shh:gfp:ABC)* injected with control mRNA (**A**) or *pki* mRNA (**B**). (**C,D**) Mean projections of YFP images of control mRNA injected embryos (n = 27) or *pki* mRNA injected embryos (n = 32). Arrows point at the zona limitans intrathalamica with expansion of reporter gene expression in *pki* mRNA injected embryos. Arrowheads indicate the additional reporter gene expression in the dorsal midbrain and hindbrain in *pki* mRNA injected embryos. (**E**) Box plots illustrating the domain-specific changes of reporter gene expression in *pki* mRNA injected embryos. The distributions of signal intensities of total embryos and of selected affected domains are shown. Abbreviations of domains are as described in text. Abbreviations: c, control mRNA–injected (n = 27); p, pki mRNA–injected (n = 32); a.u., arbitrary unit.

4.3. Discussion

4.3.1. High throughput imaging and sample preparation

In this project we have established an imaging pipeline that allows the automated 2D image acquisition of thousands of live-arrayed zebrafish embryos, thus eliminating a common bottleneck of zebrafish HTS screening approaches. We successfully adapted an automated screening microscope initially designed for high content cell culture screening approaches for the use of zebrafish embryos (Liebel et al. 2003). With the pipeline used it is possible to image up to 2,000 embryos within 4 h by a single microscope system. The throughput could be further enhanced by an increase in embryo production and the use of multiple systems running in parallel or extension of the period of imaging, respectively. However, to achieve such a throughput protocols need to be developed that overcome the manual sample preparation which remains a bottleneck of the approach.

While an experienced experimenter can inject up to 2,000 embryos per hour, the large scale generation of transient transgenic embryos would benefit from methods that allow the parallel transfection of embryos. Methods for the generation of transgenic zebrafish by electroporation have been reported (Buono and Linser 1992; Muller et al. 1993), but the relatively low efficiency and survival rate appear to prevent a wider use. The automation of microinjection of zebrafish embryos could be an alternative solution. In fact, the development of zebrafish injection robot prototypes, which can inject with high accuracy and reasonable speed, has been reported (Wang et al. 2007). Additionally, for the analysis of stable transgenic lines with the described image processing software, microinjection would

be rendered completely obsolete if experimental embryos were obtained by outcrosses with transgenic lines that ubiquitously express a distinct fluorescent reporter gene that can substitute for the ubiquitous CFP signal.

The manual sorting and plating of embryos is another labour intensive step in the pipeline, thus the screening efficiency would be greatly improved by automation tools. While devices to automatically sort and dispense zebrafish embryos have been reported (Makky et al. 2008), their applicability in research projects using live fluorescently labelled embryos needs to be demonstrated. The usage of agarose embedded plates provided a simple and efficient way to laterally orient embryos, which was a pre-requisite for the 2-dimensional approach chosen. However, it required the manual orientation of each experimental embryo, thus adding further to the work load prior to image acquisition. Therefore, orientation independent image processing (Vogt et al. 2009) or automated imaging that allows 3D-resolution (e.g. confocal) would be greatly beneficial.

4.3.2. Overlay projections allow visualization of expression patterns

The analysis of reporter gene expression patterns driven by regulatory elements in transient transgenic embryos requires the generation of compound expression patterns due to the mosaicism in the F0 generation. Classically, this is done by manually plotting positive cells of single embryos on schematic drawings of zebrafish embryos thus providing an expression map (Westerfield et al. 1992; Muller et al. 1997). The generation of maximum projections overcomes the reporter expression mosaicism in an analogous fashion with the added benefit of automation. The developed tool can project tens of thousands of embryos within

a few hours, thus providing significant advances over the labour-intensive manual protocols. Moreover, the visualization of data by mean and maximum projections provides intuitive access to the experimental data. This was crucial for the HTS screening of transient transgenic zebrafish embryos as it provided a fast visual overview of the multitude of individual experiments carried out (see also **Chapter 5**). Since maximum projection highlights expression in single cells of individual embryos they are particularly useful for weak and mosaic activity, but it also means that tissue-specific expression strength differences can be hidden. Therefore, the continuous patterns of transgenic lines are better visualized in mean projections. In stable transgenic lines the expression patterns can already be inferred from single stable transgenic embryos. Nevertheless, the generation of projections from transgenic lines is useful for the visualization of overall changes of expression patterns, which can vary on the individual level, like e.g. seen in the PKI experiment (**Figure 4.11**).

4.3.3. A 2-dimensional approach to analyse spatial reporter activity

To quantitatively analyse the spatial distribution of signal we chose a simplified 2-dimensional approach, which allowed short computing times and kept hardware requirements low. We arbitrarily subdivided the reference embryo shape into 8 segmentation domains, which we considered of sufficient detail to resolve the spatial activity of transgenes, as each domain represents key embryonic tissues (such as the retina, notochord or spinal cord). The segmentation domains are user-defined and thus flexible, i.e. they can be modified to cover different embryonic structures thus emphasizing different patterns. Furthermore, the number of domains can be increased thus elevating the spatial

resolution of the reporter signal read-out. As a consequence of the two-dimensional nature of the segmentation domains, signals assigned to a domain can emanate from tissues that underlie or overlie the embryonic structure that gives name to the domain. This could only be overcome by the 3-dimensional analysis of reporter gene expression. Potentially the separate analysis of several z-slices would partially allow 3-dimensional assignment of signal, but ultimately only 3-dimensional imaging could provide sufficient spatial resolution.

The shape and size of both embryos and domains can vary greatly among individuals owing to the natural shape variation and irregularities of developmental stages used. Therefore, we tested the accuracy of the domain assignment of individual embryos in a control experiment in which we analysed domain assignments after warping. This revealed that the algorithm is robust and performs well with randomly selected embryos; however, the accuracy increased when we selected embryos that matched the morphology of the reference embryo shape, thus emphasizing the importance of generating embryos at a well-defined stage. Moreover, we observed significant variation in the accuracy of detection of different domains. It was apparent that well referenced domains such as yolk, skin and notochord are warped more reliably, indicating that the domain assignments could be further improved by increasing the number of landmarks generated prior to the segmentation step.

4.3.4. The algorithm provides a quantitative read-out of reporter gene expression

We utilized the segmented reference embryo to quantify the spatial distribution of fluorescent signals providing a simplified 2-dimensional resolution of spatial transgene activity. The comparison of quantification results and the observed expression patterns seen

in maximum or mean projections of transient or stable transgenic embryos, respectively, demonstrated reliable spatial signal detection in the overlapping domains, despite the inaccuracies of domain detection revealed in the control experiments. Moreover, the reliable quantification of fluorescent signal intensities could be demonstrated in a control experiment, analysing the transgene activity in hemi- and heterozygous embryos of the *Tg(2.2shh:gfp:ABC)* transgenic line. We assumed that these embryos provided a test sample with defined signal intensities as embryos of a particular genotype had a defined copy number of transgenes and we expected that the transgene activity correlates with the distinct number of copies. On the population level the algorithm could successfully distinguish the genotypes, but on the individual level variations within genotype groups were observed. These are potentially due to minor differences in the orientation of embryos leading to differences in signal capturing. This demonstrates the importance of accurate orientation of embryos in 2-dimensional approaches as well as of using larger number of embryos. Additionally, the algorithm successfully quantified fluorescent signals as suggested by the visual correlation of domain specific quantification values and reporter gene expression patterns.

4.3.5. Transient transgenesis to analyse interaction specificities

The analysis of CREs in transient transgenic embryos is technically simple and fast, thus allowing the rapid scoring of tissue-specific reporter gene activities. Therefore, it has been extensively used in the characterization of CRE activities e.g. Woolfe et al. 2005. Although the approach is rapid a drawback is that expression is not seen in all cells that would activate

the transgene in a stable transgenic line demanding the reconstruction of expression patterns from several embryos. The Tol2 system has been reported to reduce mosaicism, thus reducing the number of embryos required (Fisher et al. 2006). Therefore, we tested if the Tol2 system would allow a more efficient testing of enhancer-promoter interactions in our assay. Indeed, the activation rate per embryo was higher, so potentially by using Tol2 the overall number of embryos needed could have been reduced. However, that would have required co-injecting *tol2* mRNA with the reporter constructs, which would have introduced another potential variable to the approach as mRNAs are prone to degradation during the injection procedure. Indeed, we have observed considerable variations of CFP signal from what we mainly attributed to the degradation of *cfp* mRNA during microinjections (**Appendix 2**).

Moreover, it has been reported that Tol2 mediated integration leads to preferential integration within or near transcriptionally active sites (Kondrychyn et al. 2009), that potentially increases the probability of position effects, i.e. the transgene comes under control of the regulatory information acting at the integration site. This is consistent with the wide and successful application of the Tol2 system in enhancer trap studies (Kawakami 2007). Position effects would especially have a large impact on reconstructed expression patterns deriving from low numbers of embryos. Indeed, we observed ectopic activation of reporter signal in the hatching gland and blood vessels in the F0 generation when testing several alternative enhancer-promoter combinations (Yavor Hadzhiev, unpublished observation). The 'classic' approach we chose can also be influenced by position effects, but we assume that they are less frequent and averaged out by analysing higher number of

embryos. Potential position effects could be further minimized by using constructs in which the transgene is flanked by insulators such as the mouse GAB or chicken 5'HS5, which have been reported to partially maintain their insulating activity in zebrafish (Bessa et al. 2009). In addition to Tol2, it was also shown that the co-injection of I-SceI homing endonuclease with transgenes flanked by I-SceI recognition sites can reduce reporter mosaicism in the F0 generation (Thermes et al. 2002), but we have not explored this option as in the laboratory we could not verify the described effect (Yavor Hadzhiev, unpublished observation).

In this study, we could demonstrate that the utilized overlay and projection technique used on transient transgenics revealed highly similar expression patterns to stable transgenic embryos. However, the direct comparison of transient and stable transgenics has to be treated with caution as both systems have advantages and disadvantages. In stable transgenic lines each embryonic cell harbours an equal copy number of the transgene, thus stable transgenics show higher resolution and are less noisy than transients. CRE analysis in stable transgenics is currently the only technique that can completely overcome reporter gene mosaicism in zebrafish. Moreover, the stable integration overcomes the loss of some regulatory control that requires the chromosomal context, which is a potential disadvantage of the transient approach. On the other hand, stable transgenic lines are more prone to position effects and examples where the expression pattern differs among transgenic lines carrying the very same transgene exist (Hadzhiev et al. 2007). Nevertheless, the increased accuracy in stable transgenics resulting in a more faithful recapitulation of CRE activity would have provided a major advantage. The main argument against the usage of transgenic lines was that the generation of more than 200 lines is time-consuming and very demanding in

terms of laboratory/animal facility space, making this approach almost inapplicable in the screening of enhancer-promoter interactions in a high-throughput manner. In summary, while keeping technical limitations in mind, the transient transgenic approach provided a rapid and technically simple screening assay particularly suited for the analysis of CRE activity in a high-throughput manner.

Due to the limitations of the transient approach, we carried out a control experiment to analyse the reproducibility of pattern reconstruction. This experiment revealed that independent experiments led to highly similar spatial reporter signal activities, illustrating the suitability of the approach in analysing spatial CRE activity. However, it also showed that the overall expression intensity can vary among repeats, potentially caused by variations introduced by microinjections and differential distribution of plasmid among embryonic cells. This suggested that the overall expression strength is not a precise indicator for the overall interaction strength of enhancers and promoters.

4.3.6. Analysis of long-pec stage embryos

One of the key advantages of the zebrafish embryo is its temporal resolution, i.e. the possibility to address biological questions at different time points during development. Since we initially designed the system for the analysis of the prim-20 developmental stage, the biological applications the tool could be applied to were restricted to a single developmental stage. To test the further applicability of the system, the tool was adapted for the use with long-pec stage embryos. We could demonstrate that the domain assignment had comparable accuracy to the prim-20 stage. Moreover, the analysis of long-pec stage

embryos of the *Tg(2.2shh:gfp:ABC)* transgenic line resulted in comparable mean projections and domain specific quantification of signal to those of the same transgenic line at the prim-20 stage. Together, these results demonstrate the flexibility of the approach in analysing fluorescent signals at different developmental stages.

4.3.7. Applications of the developed high-throughput screening tool

The system has proven useful in the automated quantitative analysis of spatial transgene activity, thus providing a tool for the high throughput analysis of CREs in transient transgenic assays (see also **Chapter 5**) and in stable transgenic embryos. Therefore, we suggest that the developed tool can be applied to the large-scale functional testing of bioinformatically predicted putative CREs using the zebrafish embryo as a model system (Ellingsen et al. 2005; Woolfe et al. 2005). Additionally, we could show that the developed tool is sufficiently sensitive to detect minor alterations of transgene activity as demonstrated by a modelling experiment in which we modified transgene activity of the *Tg(2.2shh:gfp:ABC)* stable transgenic line by modulating the Hedgehog signalling pathway. Since we could detect minor alterations of reporter gene expression patterns, the developed tool may be also utilized in transgenic phenotyping applications such as the detection of gene expression phenotypes after exposure to toxic or pharmacological compounds or the automated detection of gene expression phenotypes in transgenic genetic screen (Jin et al. 2007). However, it must be noted that the system does not perform well with embryos with severely altered morphology thus limiting its application to mild phenotypes. Malformed embryos are either sorted out, thus providing at least a very basic distinction of morphological phenotypes, or

lead to misassignment of domains. To utilize the developed system for scoring of morphological phenotypes would require major modifications of the software. Potentially, the 'training' of the algorithm using several defined and frequently occurring morphological changes would allow a categorisation of morphological phenotypes.

The current technology allows the investigation of spatial fluorescent signals in 2 dimensions. However, to fully exploit the potential of reporter gene based HTS approaches in a developing organism requires the development of technologies that allow the automated examination of expression patterns in the third dimension. Therefore, the establishment of screening systems that combine high throughput with imaging technologies (e.g. confocal microscopy, selective plane illumination microscopy) that allow a better resolution of transgene activity in the various cell- and tissue types present in the zebrafish embryo would provide a significant advance. Ultimately, only the investigation of reporter gene activity on a single cell resolution within the complexity of an organism allows the correct dissection of transgene activity.

5. MAPPING OF ENHANCER - CORE PROMOTER INTERACTIONS IN ZEBRAFISH EMBRYOS

Foreword

See foreword of **chapter 4** and below.

My contributions to the results presented in **chapter 5** are as follows:

I carried out the TSS prediction and *in-silico* characterization of core promoters (**5.2.1.** and **5.2.8.**). I carried out the expected domain analysis of enhancers (**5.2.2.** and **5.2.8.**). My contribution to the screening of embryos (**5.2.4.**) are explained in the foreword of **chapter 4**. In collaboration with Markus Reischl and Eva Kalmar, I designed and carried out the enhancer – core promoter interaction mapping and analyzed data (**5.2.5.** and **5.2.6.**). I designed and carried out the analysis of interaction specificities of core promoters including the interactivity analysis (**5.2.7.**) and the specificity analysis (**5.2.8**).

The work presented in **chapter 5** has been partially published in Gehrig et al. 2009.

5.1. Introduction and overview

Tissue-specific gene transcription has been thought to be mainly controlled by the interplay of transcription factors and proximal or distal CREs such as enhancers. In contrast, core promoters have been considered as rather passive components in differential gene regulation serving mainly as initiation site for transcription. However, the identification of many different types of core promoters and a complementing variety of core promoter binding proteins have indicated that regulatory processes at the core promoter level are rather complex suggesting that their function extends beyond transcription initiation (Juven-Gershon and Kadonaga 2009).

5.1.1. Core promoter diversity and enhancer-promoter specificity

A potential functional significance for core promoter diversity could be to define the interaction specificity with enhancers. Many metazoan genes are regulated by enhancers that can be exposed to potentially several genes within their striking distance, thus enhancer - promoter specificity could provide a mechanism controlling the cell-type or tissue specific activation of target genes. In fact, functional studies in *Drosophila* have revealed that the core promoter identity, in particular the CPE composition of core promoters, can define interaction specificity (Ohtsuki et al. 1998; Butler and Kadonaga 2001; Juven-Gershon et al. 2008a). In vertebrates the contribution of core promoters to enhancer-promoter interaction specificity remains largely elusive, but bioinformatic studies suggest that target promoters of long-range enhancers exhibit different features than non-target core promoters suggesting specific enhancer - core promoter pairings (Kikuta et al. 2007; Akalin et al. 2009).

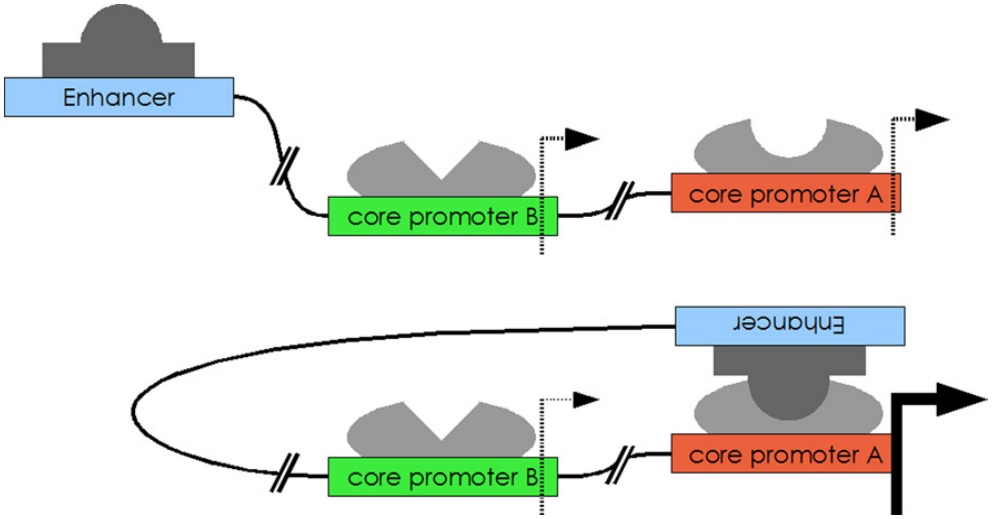


Figure 5.1| Simplified model of enhancer-promoter specificity. Enhancer bound proteins can interact with proteins bound at core promoter A, but not with proteins at core promoter B, thus leading to activation of core promoter A only.

A hypothetical explanation for enhancer-promoter specificity could be that PICs at different core promoters interact differentially with the core promoter DNA or display an alternative protein composition leading to topologically or structurally distinct protein complexes, respectively. The variant forms of PICs could have differential capabilities to interact with enhancer bound proteins, therefore restricting the activating effect of enhancers to a certain type of promoter (**Figure 5.1**) (Smale 2001).

5.1.2. Implications of enhancer-promoter specificity in transgenesis technologies

The interaction specificity between enhancers and core promoters has also important technical implications for transgenic applications in which the generation of reporter signal depends on the interaction of enhancers with core promoters, such as enhancer trap assays or the functional analysis of CREs. Since the core promoter can potentially restrict the interaction to certain enhancers or impose bias on tissue specific activity, the result of

studies may be skewed depending on the core promoter used (Amsterdam and Becker 2005; Gomez-Skarmeta et al. 2006). In fact, the outcome of reported enhancer trap screens carried out in zebrafish appears to be dependent on the core promoter used, i.e. the activation rate per genomic integration or the gene ontology of endogenous genes regulated by the trapped enhancers differs (Parinov et al. 2004; Ellingsen et al. 2005).

5.1.3. Aims

While there is ample evidence for enhancer - core promoter interaction specificity in *Drosophila*, its contribution to differential gene regulation in the context of the vertebrate embryo is unknown. Therefore, the objective of this study was to explore whether there is sequence determined enhancer-promoter specificity using the zebrafish embryo as a model system. To this end, we utilized a previously generated promoter resource (**Chapter 3**) to design a large scale experiment, in which we addressed the interaction specificities of 19 core promoters and 11 enhancers in all possible combinations. A total of 202 enhancer-promoter combinations were cloned in front of a *venus* reporter gene. We analysed the activity of the *enhancer::promoter* reporter constructs in a transient transgenic approach, using a HTS tool for the automated and quantitative analysis of spatial distribution of fluorescent signal in zebrafish embryos which was developed in the context of this study (the development of the HTS tool is discussed in **Chapter 4**).

By mapping enhancer-promoter interactions, we expected to gain insights into the role of core promoters in restricting enhancer activity in the developing vertebrate embryo. Additionally, by screening 19 promoters for their differential capability to interact with

various enhancers, we aimed to create a core promoter resource for other transgenic applications.

5.2. Results

5.2.1. Core promoters used in this study

To analyse enhancer-promoter interaction specificities, we chose a diverse set of 19 core promoters from genes that are known to be expressed in the early zebrafish embryo. We selected all core promoter regions from a pool of promoters which were functionally verified in reporter gene assays prior to this study either as promoter fragments including additional proximal sequences (see **chapter 3**) or in co-injection experiments with a known enhancer (see **Appendix 3**). Promoters were selected to represent either ubiquitously expressed (*e.g. pcbp2, tbp*) or spatially restricted genes (e.g. *apoeb, c20orf45*). Additionally, we chose promoters to represent different promoter classes; this includes the differential occurrence of CPEs or distinct distribution of TSS within the promoter region (**Table 5.1**).

We defined the exact location of core promoters on the basis of TSS distributions. We chose the core promoters when at least 1 verified TSS was annotated to the genome sequence (zv#7). To predict the main TSS region we used the DBTSS database and the ensembl genome browser (Wakaguri et al. 2008; Hubbard et al. 2009). To verify the TSS regions and define TSS distribution shapes, we analysed the distribution of 5' ends of ESTs and GenBank mRNAs in the predicted core promoter regions and in regions flanking the defined TSSs using the EST annotation present in the UCSC genome browser (Rhead et al. 2009). Analysis for

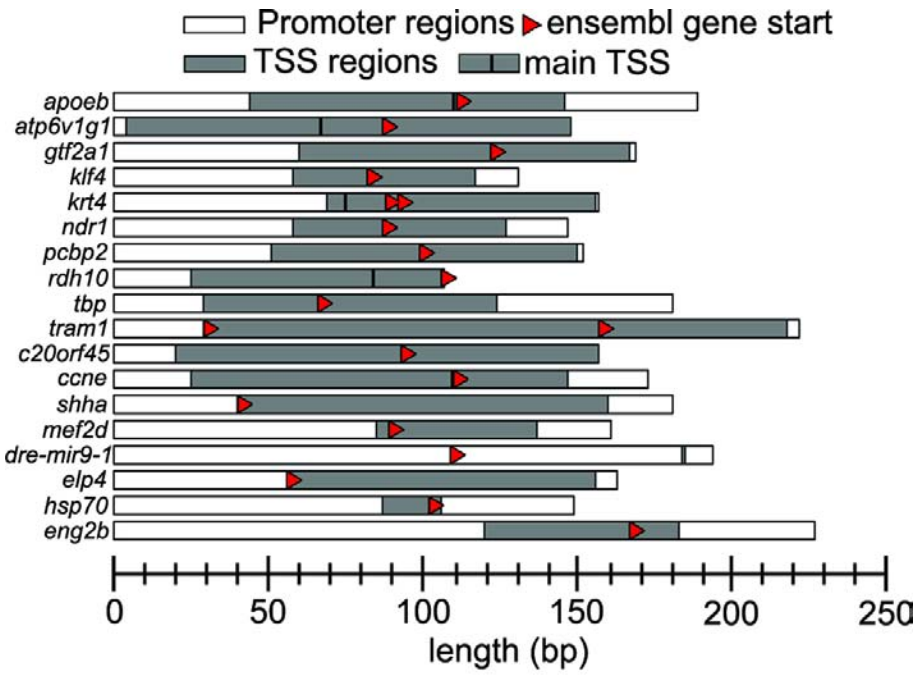


Figure 5.2| Core promoters used in this study. Core promoters are represented as horizontal bars indicating length in bp. Grey areas represent transcription start site regions covered by 5' ends of ESTs or GenBank mRNAs that map in the promoter region. Red triangles indicate the position of ensembl gene starts and black vertical bars indicate the position of dominant TSS peaks (main TSS) if applicable. See also **Table 5.1**.

occurrence of CPEs was carried out using Possum (Fu et al. 2004) and CGI analysis using CpG island explorer (Wang and Leung 2004). TSS, EST, CPE and CGI analyses were carried out according to **chapter 3.2.4.**

Primers were designed to flank a region of 107-227 bp incorporating the main TSS in cases where a dominant TSS peak was evident (**Figure 5.2**). However, for most promoters no TSS peak was evident. In these cases we designed the primers to include the highest possible number of potential TSSs within size constraints (**Appendix 3** and **Table 5.1**). As a negative control promoter we used the *pbFOG* promoter of *Ciona intestinalis* which is transcriptionally silent in zebrafish (Roure et al. 2007; Kalmár 2009).

Table 5.1| Overview of core promoter features

Gene symbol	Chromosomal Position zv#7	Strand	Ensembl Transcript(s)	gene expression in embryogenesis	Total Number of TSS	TSS Distribution	Size (bp)	Ensembl TSS position	Main TSS position	Most 5' TSS position	Most 3' TSS position	CPE/CGI occurence
apoeb	chr16:24388715-24388903	+	ENSDART00000058965	spatially restricted	627	dom	189	111	110	44	146	TATA (78)
atp6v1g1	chr5:50839078-50839225	+	ENSDART00000036331	spatially restricted	197	dom	148	87	67	4	148	
gtf2a1	chr20:12836731-12836899	+	ENSDART00000006443	not spatially restricted	75	brd	169	122	n/a	60	167	TATA (90)
klf4	chr2:27760887-27761017	+	ENSDART00000056655	spatially restricted	24	brd	131	84	n/a	58	117	
krt4	chr6:29,881,458-29,881,623*	-	ENSDART00000012644	spatially restricted	365*	dom	166	n/a	82	76	152	TATA (50)
krt4	chr6:26,288,946-26,289,102**	-	ENSDART00000012644	spatially restricted	604	dom	157	88/92	75	69	156	TATA (43)
			ENSDART00000104080									
ndr1	chr21:11382506-11382652	-	ENSDART00000079692	spatially restricted	5	n/a	147	87	n/a	58	127	
pcbp2	chr9:614661-614812	-	ENSDART00000025885	not spatially restricted	110	brd	152	99	n/a	51	150	*CGI*
rdh10	chr2:24616214-24616320	+	ENSDART00000004903	spatially restricted	211	dom	107	106	84	25	106	
tbp	chr13:24702224-24702404	+	ENSDART00000016211	not spatially restricted	110	brd	181	66	n/a	29	124	TLS (31)
tram1	chr24:17550732-17550953	+	ENSDART00000004664	spatially restricted	105	brd	222	29/157	n/a	29	218	
			ENSDART00000105973									
c20orf45	chr6:58935813-58935969	+	ENSDART00000008243	spatially restricted	90	brd	157	93	n/a	20	157	
ccne	chr7:43205849-43206021	-	ENSDART00000007883	spatially restricted	23	dom	173	110	110	25	147	
shha	chr7:36664074-36664254	-	ENSDART00000099204	spatially restricted	5	n/a	181	40	n/a	40	159	*CGI*
mef2d	chr16:20571360-20571520	+	ENSDART00000009826	spatially restricted	21	brd	161	89	n/a	85	137	
dre-mir9-1	chr16:20549606-20549799	-	ENSDART00000070292	spatially restiricted	2	n/a	194	109	n/a	184	185	
elp4	chr7:8747729-8747891	-	ENSDART00000002498	spatially restricted	25	brd	163	57	n/a	57	156	MED-1 (92)
hsp70	chr3:23548288-23548436	+	ENSDART00000010477	spatially restricted	5	n/a	149	102	n/a	87	106	TATA (57)
eng2b	chr2:24409903-24410129	+	ENSDART00000056748	spatially restricted	10	n/a	227	167	n/a	120	183	

Legend to Table 5.1|Shown are the corresponding gene symbols, the chromosomal position and orientation/strand of the core promoters as well as the ensembl transcript ID (Column 1-4; * assembly zv#6, ** based on BLAT). Column 5 indicates endogenous gene expression in 2 categories (Rebagliati et al. 1998; Thisse et al. 2001; Thisse et al. 2005; Wienholds et al. 2005; Krone et al. 2005; Ertzer et al. 2007; Hinits et al. 2007). Column 6 indicates the total number of 5'ends of ESTs and GenBank mRNAs within the core promoter region. Column 7 shows TSS distribution classification: dominant (dom) if more than 20% of TSS map to a single nt, broad (brd) if less than 20% map to a single nt and n/a if in total less than 20 TSS map to the core promoter region. Column 8 shows the size of the core promoter fragments in bp. Column 9-12 gives TSS positions within the promoter fragments (see also **Figure 5.2**). Column 13 indicates the occurrence of CPEs or the location within CGIs, respectively.

5.2.2. Enhancers used in this study

The enhancers were chosen from the published literature or were identified by Eva Kalmar (*eng2b reg5*; (Kalmár 2009)) and represent various tissue-specific activities in the embryo with a bias for neural specific enhancers (**Table 5.2** and **Table 5.3**). The described spatial activity of all selected enhancers matched or overlapped with the expression pattern of the corresponding gene (**Table 5.3**). The following enhancers were chosen: *shha arC* (Muller et al. 1999), *bactin2 intron1* (Muller et al. 1997), *pax6b eye* (Woolfe et al. 2005), *eng2b CXE* (Song et al. 1996), *eng2b reg5* (Kalmár 2009), *dre-mir9-1* (Kikuta et al. 2007), *myl7* (Huang et al. 2003), *isl1 zCREST2* (Uemura et al. 2005), *dlx2b/6a ei* (Zerucha et al. 2000) and *mnx1 regB* (Nakano et al. 2005). As a negative control enhancer we used *VC_909* a non-conserved, noncoding fragment without enhancer activity from *Takifugu rubripes* (Sanges et al. 2006).

Table 5.2| Overview of enhancers. Shown are the symbol, the full name, the size in bp, the genome used to isolate the enhancer and the corresponding genomic position if applicable.

Symbol	Name	Size (bp)	Genome	Chromosomal position
control	*control (VC_909)*	576	Takifugu rubripes	ChrUn:54537362-54537937
shha arC	*sonic hedgehog a activation region C*	462	D. rerio Zv6	chr7: 49531072-49531514
bactin2 i1	*bactin2 intron1*	508	Cyprinus carpio	N/A
pax6b	*pax6b eye enhancer*	343	Takifugu rubripes	chrUn:292744986-292745329
eng2b CXE	*engrailed 2b CXE*	968	D. rerio Zv6	chr2: 27210540-27210955
eng2b reg5	*engrailed 2b conserved region 5*	416	D. rerio Zv6	chr2: 27200525-27199606
dre-mir9-1	*micro-RNA 9 brain enhancer*	365	D. rerio Zv6	chr16: 26936634-26936998
myl7	*myosin, light polypeptide 7, regulatory, cardiac enhancer*	285	D. rerio Zv6	chr8: 66577515-66577801
isl1	*islet1 zCREST2*	724	D. rerio Zv6	chr5: 53167655-53168379
dlx2b/6a ei	*distal-less homeobox gene 2b/6a, ei*	479	D. rerio Zv6	chr19: 41680276-41680755
mnx1 regB	*motor neuron and pancreas homeobox 1, regB*	215	D. rerio Zv6	chr7: 48712492-48712707

Table 5.3| Tissue specificity of enhancers. See legend below

	Reported enhancer activity			Endogenous target gene expression pattern	
Symbol	**Promoter used for identification**	**Enhancer specificity**	**Corresponding segmentation domains**	**Tissues (prim6-prim25)**	**Corresponding segmentation domains**
control	*mouse hsp68*	-	-	-	-
shha arC	*zebrafish shha*	hypothalamus, tegmentum, notochord	brain, eye,notochord, (mhb)	**hypothalamus, tegmentum**, zona limitans intrathalamica, floor plate; earlier: **Notochord**	brain, eye, notochord, (mhb)
bactin2 i1	*carp β-actin2*	ubiquitous	ubiquitous	**ubiquitous**	ubiquitous
pax6b	*mouse β-globin*	eye, brain, spinal cord	eye, (brain, spinal cord)	**retina, hindbrain**, pancreas, tegmentum, **spinal cord**	eye, brain, spinal cord, (yolk)
eng2b CXE	*mouse hsp68*	MHB	mhb	**midbrain hindbrain boundary** and in posterior midbrain	mhb
eng2b reg5	*zebrafish eng2b*	MHB, hindbrain	mhb, brain	**midbrain hindbrain boundary** and in posterior midbrain	mhb
dre-mir9-1	*zebrafish gata2*	dorsal telencephalon	brain	**dorsal telencephalon**, proliferating cells of brain, eyes	brain/eye
myl7	*zebrafish myl7*	heart	heart	**heart**	heart
isl1	*zebrafish hsp70*	trigeminal ganglion neurons, Rohon-Beard neurons, secondary motoneurons in spinal cord	spinal cord, (brain)	forebrain nuclei, branchial arches, **trigeminal ganglia**, epiphysis, hindbrain neurons, **dorsal and ventral spinal chord neurons**, pretectum, ganglion cell layer of retina, pancreas, liver primordium	spinal cord, brain, eye
dlx2b/6a ei	*human β-globin*	forebrain	brain	dlx6: mandibular arch, hyoid arch, pharyngeal arches, otic vesicle, **forebrain**; dlx5: ventral diencephalon and **telencephalon**	brain, eye
mnx1 regB	*mouse β-globin*	spinal motoneurons	spinal cord	**developing motor neurons**	spinal cord

Legend to Table 5.3| In the left half the reported enhancer activity is shown. Given are the promoters used for enhancer identification in the original publication, the tissues of enhancer activity and the corresponding overlapping segmentation domains. In the right half the endogenous target gene expression in prim-6 to prim-25 embryos is shown (Ellies et al. 1997; Yelon et al. 1999; Muller et al. 1999; Thisse et al. 2004; Wienholds et al. 2005; Guner et al. 2007; Ertzer et al. 2007). Tissues overlapping with enhancer activity are shown in bold. Overlapping segmentation domains are indicated. Domains in brackets indicate minor overlap.

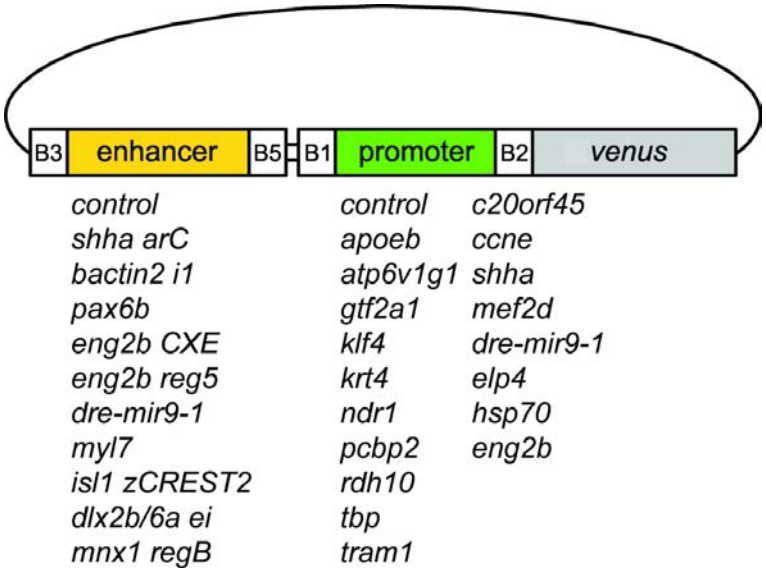


Figure 5.3| Schematic representation of reporter constructs. Shown are a diagram of the reporter construct used together with a list of 11 enhancers and 19 core promoter fragments that were used with a *venus* reporter in Multisite Gateway expression constructs. B3, B5, B1 and B2 indicate recombination sites.

5.2.3. Isolation of fragments and generation of reporter constructs

The *in-vitro* isolation of enhancers and promoters and the generation of constructs were carried out by Eva Kalmar. A detailed description of the cloning procedure can be found in Roure et al. 2007 and Kalmár 2009.

Enhancers and core promoter fragments were amplified from genomic DNA using an enzyme mix with proofreading activity. To generate reporter constructs a modified multisite Gateway system was used. In a first step 19 promoter and 11 enhancer entry clones were generated, respectively. The entry clones were used to simultaneously recombine enhancer and core promoter fragments into the expression vector, which carries two cloning/recombination sites in front of a *venus* reporter gene (**Figure 5.3**). This allowed the shuffling of enhancers and core promoters and thus the generation of more than 200

reporter constructs in a very efficient way. In total 202 enhancer-promoter-venus constructs were generated, which are from now on referred to as *enhancer::promoter* (**Figure 5.3**). Importantly, it was shown that the insertion of gateway vector–derived DNA sequences between regulatory sequences and the *venus* reporter gene does not affect enhancer-promoter interaction. Moreover, the activity of regulatory elements in the Gateway vector was comparable to other commonly used reporter vectors (Roure et al. 2007).

5.2.4. Microinjection and data acquisition

To investigate the activities of the 202 *enhancer::promoter* constructs, we carried out a transient transgenic reporter gene assay. Therefore, we injected the reporter constructs (5ng/µl) into the yolk of zygote stage embryos, i.e. before a visible blastodisc had formed, together with *in-vitro* synthesized *cfp* mRNA serving as injection quality control. To analyse the transient transgene activity we utilized the developed HTS tool as described in **chapter 4**.

In brief recapitulation, we manually selected CFP positive embryos at prim-5 and plated them at prim-20 into agar-embedded 96 well plates. Using a robotic screening microscope we automatically acquired images of the laterally oriented embryos in 4 z-slices in 3 channels: brightfield, CFP and YFP (**Figure 4.1**). As a result, we imaged a total of 17,793 embryos representing 202 enhancer-promoter combinations in a total of 213,516 images.

To analyse the spatial distribution of fluorescent signal, we used custom designed image analysis software. The four z-slices were combined to obtain a single extended focus image. The extended focus image was warped onto an arbitrarily segmented 2-dimensional

reference embryo shape and the Venus signal was visualized using maximum projections and quantified within segmentation domains. Thus, the activity of each of the 202 *enhancer::promoter* constructs was represented by 8 values indicating the spatial activity driven by the regulatory elements used (**Figure 4.2** to **4.6**). Comparative analysis of spatial distribution of reporter gene activities was carried out for experiments with minimum 30 embryos that passed quality controls. When the number of analyzable embryos was lower than 30, we carried out further repeats of experiments and the experimental data was merged after evaluation of pattern reproducibility. For each injection experiment, the 5% of embryos with the highest Venus signal (outliers) were excluded (**Appendix 3**).

5.2.5. Large scale mapping of enhancer-promoter interactions

To address the interaction specificities between the 19 core promoters and 11 enhancers, we carried out a comparative analysis of domain specific reporter gene activities as described above. In total 12,582 embryos were analysed with in average 62 embryos per construct. The quantification results for all 202 microinjection experiments as well as the number of experimental embryos are listed in **Appendix 3**. To visualize spatial reporter gene activities, we converted the quantification data of individual experiments into a colour code in which each of the 8 domains is represented by a colour, while the intensity of the colour represents reporter signal intensity. Thus, each experiment was represented by a colour coded square indicating the spatial activity of *enhancer::promoter* combinations (**Figure 5.4A to C**). The maximum colour intensity for a given domain was set as the maximum signal intensity for that domain within all experiments carried out (**Figure 5.4C**). We observed a

huge dynamic range of signal activities (several orders of magnitude), therefore we applied a root scale for visualization of domain specific quantification values as a colour code.

To create an overview of all experiments we generated an interaction matrix in which we ranked the core promoters and enhancers according their overall activity (**Figure 5.5**). Therefore, we generated an expressivity value for each element by averaging all eight domain activities for all corresponding constructs after subtraction of the activity of the control experiments with the same enhancer or promoter, respectively.

5.2.6. Functionality of core promoters and enhancers

Five core promoters showed substantial basal activity (*krt4*, *klf4*, *gtf2a1*, *c20orf45* and *apoeb*), i.e. their activity was considerably higher than the activity of the *control::control* construct, whereas the remaining promoters showed no significantly higher activity in conjunction with the control enhancer (last row **Figure 5.5**). However, each of the core promoters without apparent activity above the background level was active in conjunction with at least one of the enhancers (**Figure 5.5**). This indicates that all core promoters chosen are functional as they are capable to respond to enhancer input. Similarly, all enhancers activated at least one core promoter above the activity observed with the control constructs thus confirming enhancer functionality. None of the enhancers showed significant activity in conjunction with the control promoter (last column in **Figure 5.5**).

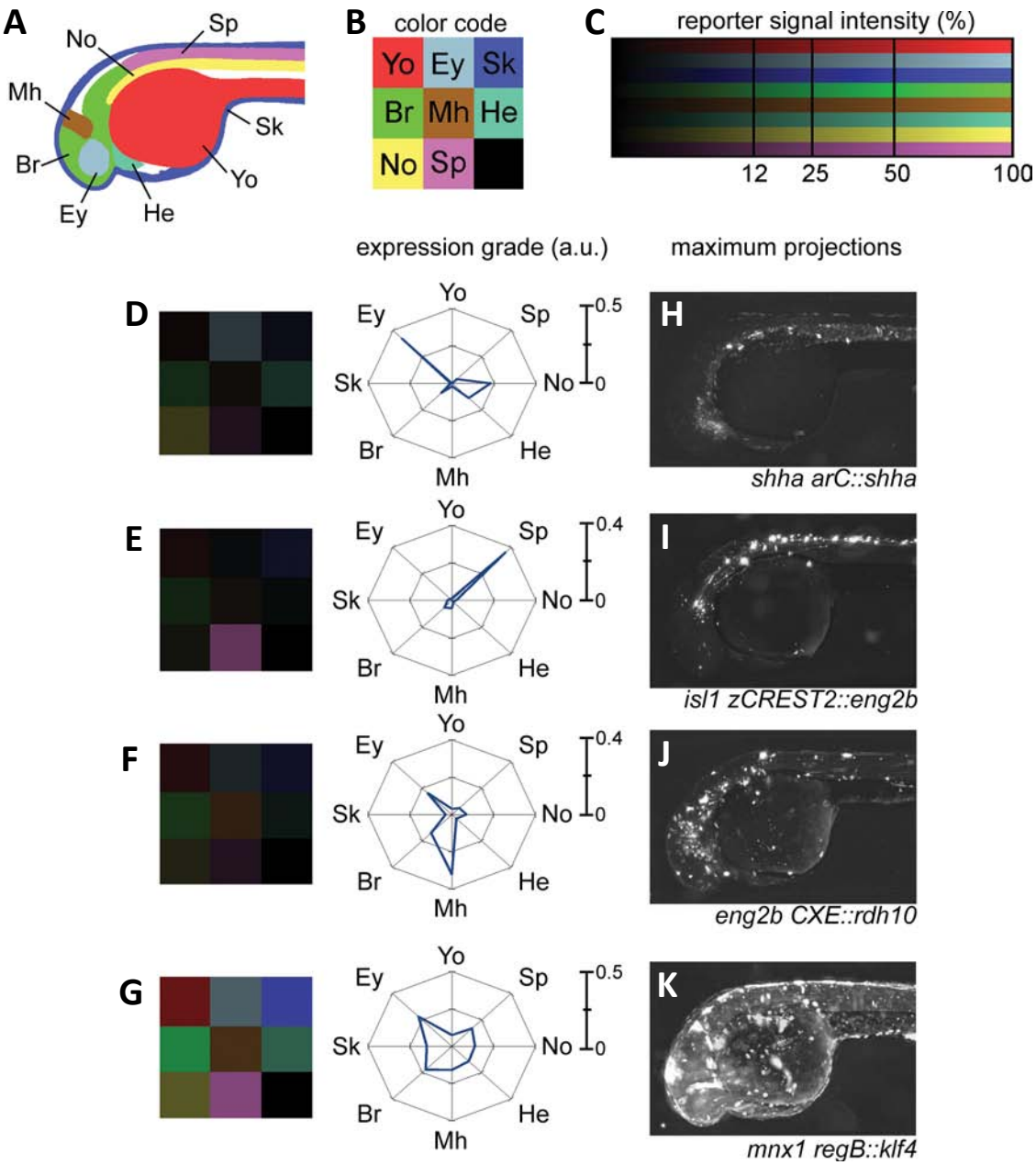


Figure 5.4| Domain specific analysis of enhancer - core promoter interaction activities. (**A**) Segmented reference embryo shape (see also **Chapter 4**). (**B**) Colour code to indicate domain specific reporter activities. (**C**) Colour scale illustrating colour values and corresponding reporter signal intensities. (**D-K**) Illustrative examples of distinct expression profiles generated by transient expression of *enhancer::promoter* constructs (**D-G**) Distribution of Venus signal intensities normalized to domain size indicated by colour codes (left) and radar plots (right) for *shha arC*::shha (**H**), *isl1 zCREST2::eng2b* (**I**), *eng2b CXE::rdh10* (**J**) and *mnx1 regB::klf4* (**K**). Abbreviations as in **Figure 4.4**.

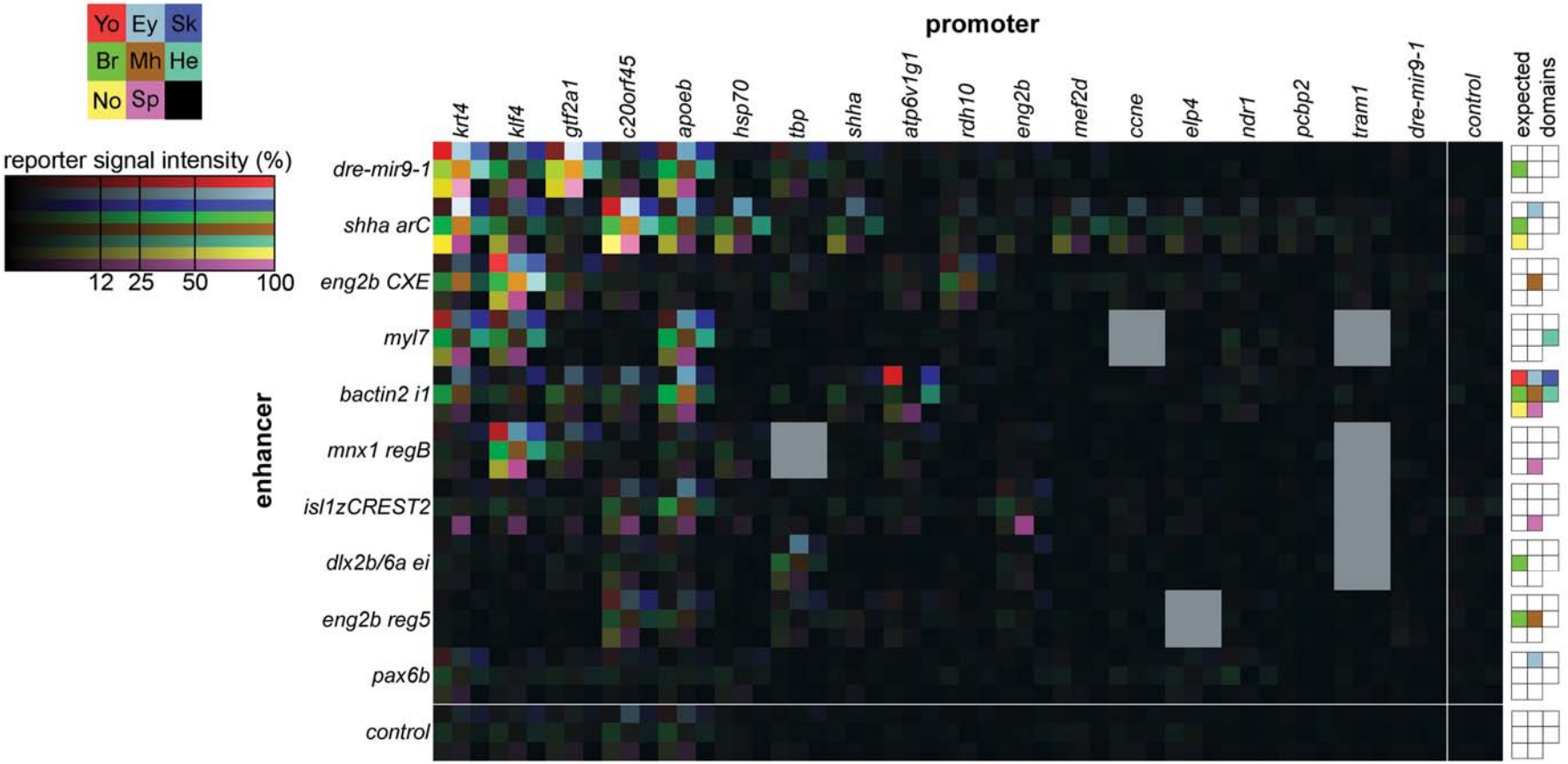


Figure 5.5| Activities of enhancer-promoter combinations. Shown is an interaction matrix of all 202 reporter constructs representing a total of 12,582 transient transgenic embryos. The colour intensity codes are arranged into rows (enhancers) and columns (promoters). Both are sorted in descending order according the average activity value of the enhancers or promoters (expressivity, see text). The white lines demarcate intensity values of negative control constructs. Grey boxes indicate reporter constructs not assayed. The last column on the right schematizes the expected activity domains of the enhancers used (see also **Table 5.3**).

5.2.7. Analysis of interaction specificities

The matrix in **Figure 5.5** illustrates the large variation of overall and domain specific Venus activity observed among the various combinations of core promoters and enhancers. For instance, we observed for several core promoters dramatic but varying upregulation after the addition of an enhancer (e.g. *krt4* or *klf4*), while other core promoters were only moderately upregulated (e.g. *shha* or *eng2b*). Moreover, enhancers did not necessarily interact with their cognate core promoters more strongly than with heterologous core promoters. For example, the *shha arC* enhancer was found to be more active with *krt4* or *hsp70* than with the *shha* core promoter, while tissue specificity was mostly retained (**Figure 5.5**). Besides these examples, the overall comparison of differential reporter activities allowed several observations to be made about interaction specificities of core promoters including general responsiveness to enhancers and recapitulation of tissue-specificity inherent to the linked enhancers.

5.2.7.1. General interaction abilities of core promoters

To investigate the general interaction abilities of core promoters, we analysed if a certain *enhancer::promoter* combination gave expression above the control level. We considered a combination as 'active', when the overall reporter activity exceeded the maximum of the negative promoter- and the negative enhancer control by the factor 2. This threshold was arbitrarily chosen, but the visual comparison of signal intensities of embryos injected with experimental or control constructs for a subset of experiments supported that this choice would reveal interactions with high confidence.

A notable observation of this interactivity analysis was the differential ability of core promoters to interact with enhancers. For example, the *ndr1* and *eng2b* core promoters interacted with most enhancers differentially (**Figure 5.6A**). While both were able to be activated by the *shha arC* enhancer in the notochord and ventral brain, only the *eng2b* core promoter was activated by the *isl1 zCREST2* enhancer in spinal cord neurons (**Figure 5.6B**). This indicated that the core promoter can limit the activating effect of enhancers. To visualize all individual experiments that exceed the control level and to reveal interaction patterns of core promoters and enhancers an interactivity matrix is shown in **Figure 5.6C**. The interactivity matrix revealed that the observation of differential interaction abilities of core promoters can be extended to the entire dataset. Strikingly, none of the core promoters was able to interact with all enhancers assayed and, furthermore, each core promoter interacted with a different subset of enhancers. Additionally, the number of interactions varied significantly among core promoters. Several core promoters were rather promiscuous (e.g. *klf4* or *eng2b*), i.e. they could interact with the majority of the enhancers assayed, while the degree of upregulation varied strongly. On the other hand, several core promoters were very specific as they could interact only with a small subset of enhancers (e.g. *c20orf45* or *ccne*) (**Figure 5.6C**). Accordingly, similar observations can be made from the enhancer point of view with strong, promiscuous enhancers such as *shha arC* or *eng2b CXE*, and very selective enhancers like *isl1 zCREST2* (**Figure 5.5** and **5.6C**). For a more detailed discussion of enhancer interaction specificities please see (Kalmár 2009).

Overall, these observations suggest that the establishment and strength of interactions of enhancers and core promoters were dependent on the identity of the core promoter used.

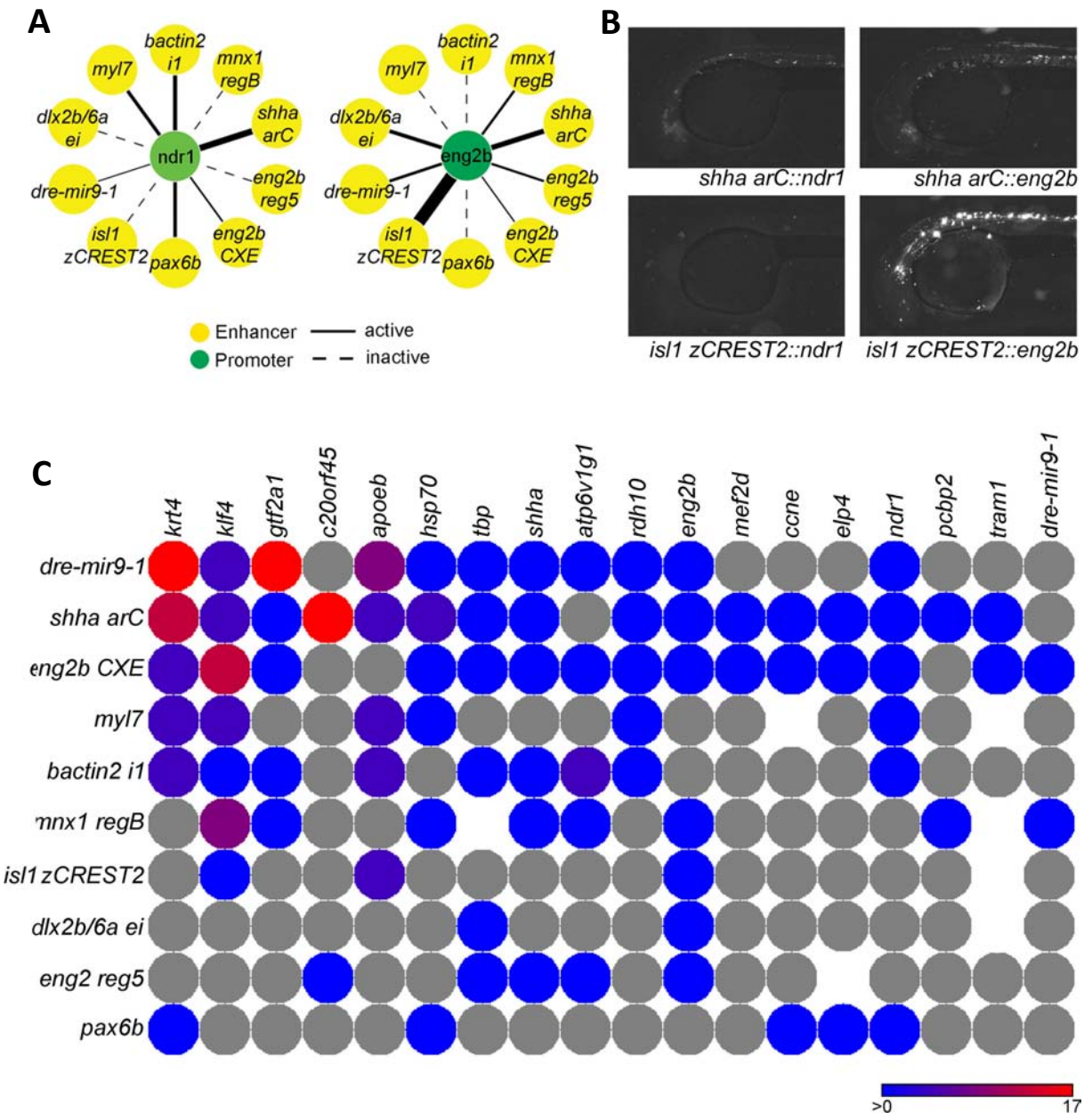


Figure 5.6| Differential interactivity of core promoters. (**A**) Interactions of the *ndr1* and *eng2b* promoters with enhancers. Thickness of black lines indicates relative strength of interaction proportional to the total sum of brightness of pixels. 'Active' indicates that signal intensity exceeds two times the maximum of the relevant controls. (**B**) Maximum projections of embryos injected with the indicated *enhancer::promoter* constructs. (**C**) Interactivity matrix showing 'active' experiments as coloured discs. Blue indicates weak and red indicates strong activity (as in **A**). Grey discs indicate 'inactive' experiments. Gaps indicate constructs not assayed.

5.2.7.2. Domain specificity of enhancer - core promoter interactions

The interactivity analysis discussed above disregards that the enhancer-promoter combinations resulted in reporter activity within expected domains of enhancer activity with varying efficiency. However, tissue specific activation of core promoters forms an important part of the overall interaction specificity. To aid in comparison of the observed activities with expected activities, an additional column is inserted in **Figure 5.5** that schematically depicts the expected activity of the linked enhancers based on published transgenic enhancer assays (**Table 5.3**).

For instance, the promiscuous *shha arC* enhancer activated all interacting core promoters within its expected activity domains eye, brain and notochord while additional ectopic activity was observed (e.g. with *c20orf45*). Also other enhancers gave tissue specific activity with heterologous and/or cognate core promoters (e.g. *eng2b CXE* or *isl1 zCREST2*) (**Figure 5.4H-J** and **5.5**), whereas some enhancers were not able to activate core promoters in the expected domains. For example the brain enhancer *dre-mir9-1* did not lead to brain specific activity in any combination including the *dre-mir9-1* core promoter; it rather acted as a general activator that largely activated a subset of core promoters (**Figure 5.5**). The injection of many *enhancer::promoter* constructs resulted in ectopic activity either in addition to or in replacement of the expected activity (e.g. **Figure 5.4G,K**), respectively, depending on the core promoter used (**Figure 5.5**). Ectopic activation was especially observed with core promoters that exhibited basal activity and that often had the tendency to be highly upregulated by linked enhancers (e.g. *klf4*).

The differential efficiencies of core promoters to respond to enhancers in a domain specific, and thus to some degree tissue specific manner, is an important consideration for promoter choice in other transgenic applications such as enhancer trap or the transgenic assay of predicted CREs, because the core promoter used in these assays could largely influence the interactions with either the endogenous enhancer near to the integration site or the attached putative enhancer, respectively. In order to evaluate the reliability of a core promoter to respond to enhancer input in a domain specific manner, we examined the activity of 19 core promoters in conjunction with three highly specific enhancers used in the screen: *shha arC*, *isl1 zCREST2* and *eng2b CXE.* For each enhancer - core promoter combination we calculated the percentage of tissue-specific reporter signal within the expected domains of enhancer activity (**Table 5.3** and **Figure 5.5**). Then we took the mean of these percentages as specificity value reflecting the proportion of signal matching enhancer activity. As a result we assigned each core promoter a specificity value, which allowed ranking the promoters according to their domain specific interactions.

A summary of the interaction abilities of core promoters is shown in **Figure 5.7** in which we plotted the specificity values of individual core promoters together with the proportion of experiments that exceed the maximum of the controls (interactivity). This interaction plot together with the matrices in **Figure 5.5** and **Figure 5.6C** allowed identifying sets of promoters (e.g. *krt4*, *hsp70* and *eng2b*) with the majority of Venus expression in expected domains and broad capability of interactions with the majority of enhancers. For that reason, we propose that these core promoters can serve as useful tools in other transgenic application.

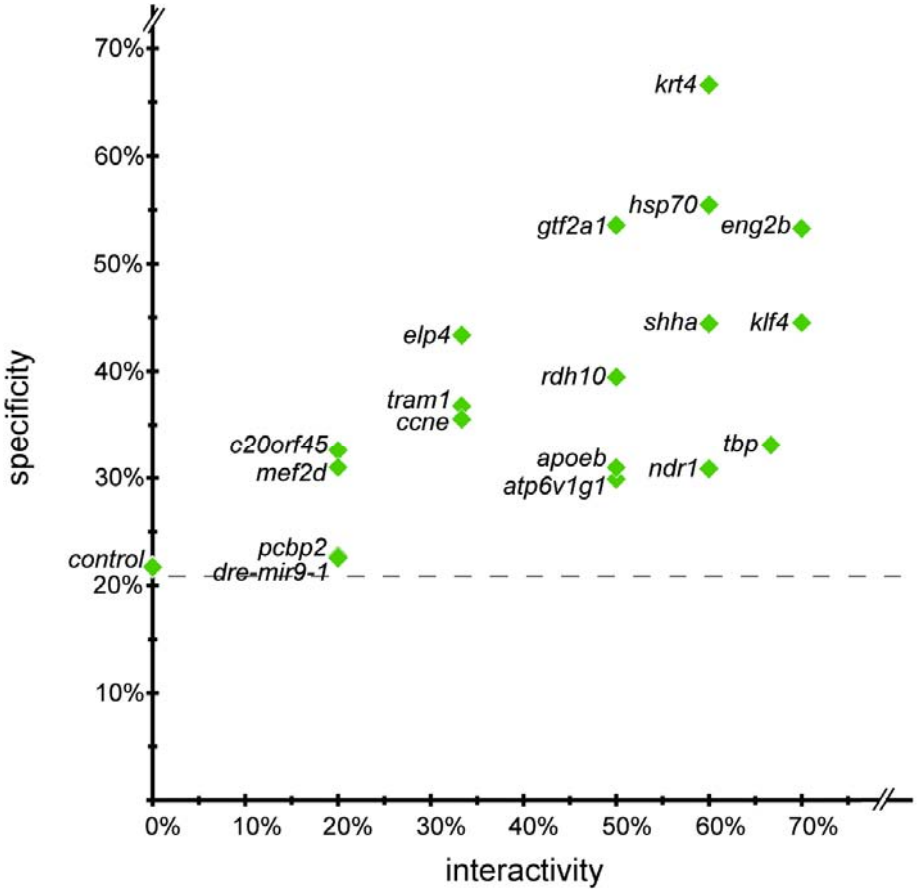


Figure 5.7| Interaction specificities of core promoters. Plot of domain specificity (mean percentage) of reporter signal in the expected domains in conjunction with the *shha arC*, *eng2b CXE* and *isl1 zCREST2* enhancers versus interactivity of core promoters (**see Figure 5.6C**). The expected domains are indicated in **Figure 5.5** and **Table 5.3**. The dashed line indicates the expected proportion of fluorescence signal by random mosaic expression. As in total 5 segmentation domains for 3 experiments were examined a random distribution of signal (12.5% signal per domain) leads to a specificity value of 20.8%.

5.2.8. Correlation with core promoter features

In this study we analysed the interaction specificity of a diverse set of core promoters, because we aimed to identify core promoter properties that correlate with promoter promiscuity or specificity, respectively. However, the relatively small set of core promoters combined with the rare occurrence of CPEs and CGIs as well as the lack of TSS distribution information for many core promoters hindered a meaningful statistical evaluation of

correlation (**Table 5.1**). Nevertheless, one observation could be made: 3 of 4 TATA-containing promoters were identified as 'good' interactors: *krt4*, *hsp70* and *gtf2a1*. The TATA containing *apoeb* and the TLS containing *tbp* core promoters were also interactive, while both led to a substantial amount of ectopic activity (**Figure 5.7**).

5.3. Discussion

5.3.1. Core promoter definitions and enhancer selection

A core promoter is defined as the minimal stretch of sequence that can nucleate the assembly of a functional pre-initiation complex and can integrate stimulatory signals from distally located sites (Smale and Kadonaga 2003). Besides this functional definition, a core promoter can be defined by its location in relation to verified TSSs (Sandelin et al. 2007). In fact, the most accurate method to date for identifying the location of a core promoter on a genomic scale is the CAGE analysis of TSSs (Carninci et al. 2006).

In this study one of the key issues was the accurate definition of core promoters. Therefore, we identified core promoters by the location of TSS using full-length cDNA and EST clustering analysis, which are commonly used methods to map TSS regions (Suzuki et al. 2001a; Ohler et al. 2002). As the 5'ends of transcript associated sequences always derive from biochemical assays this approach fulfils the criteria of experimental verification and can therefore be considered as sufficient to predict the location of core promoter regions. However, the TSS data utilized in this study derives from publically available sequence databases that do not always contain comprehensive TSS information for all promoter

regions (e.g. *dre-mir9-1*; **Table 5.1**). Therefore, we only chose core promoters from a pool of previously verified promoters. Importantly, all core promoters within the screen were activated by at least one enhancer. Thus, we had two independent lines of evidence confirming the core promoter definitions: TSS mapping and functional verification.

Enhancers were chosen based on available literature or on functional studies carried out by Eva Kalmar (Kalmár 2009). When the original studies were carried out in mouse, the activity of the conserved zebrafish enhancers was tested prior to this study and their tissue-specific activity was confirmed in transient transgenic assays (Kalmár 2009). Thus, all chosen enhancers were functionally verified. The bias towards neural enhancers reflects a bias towards studying genes specifically expressed in the developing nervous system.

5.3.2. Cloning strategy

The utilized Multisite Gateway system provided an efficient tool to relatively quickly generate a large number of constructs as the Gateway system allows the efficient shuffling of elements. The suitability of the vector to dissect CRE activity was extensively demonstrated prior to the screen in *Ciona intestinalis*, zebrafish and mouse cells (Roure et al. 2007). The generated enhancer and promoter entry clones can be used to easily recombine the cloned elements into various destination vectors which carry compatible recombination sites, including several distinct vectors for transgenesis and expression vectors (Roure et al. 2007). Therefore, each generated vector adds to a growing resource of plasmids which can be used for multiple purposes.

5.3.3. Differential interaction abilities of core promoters

In this study we carried out a HTS experiment in which we explored the interaction specificities between 19 core promoters and 11 enhancers in all possible combinations. We functionally assayed the interaction specificities of 202 enhancer-core promoter combinations using a platform for the automated registration of spatial reporter signal activity in transient transgenic embryos. This analysis revealed a high degree of interaction specificities of core promoters, indicating that in the context of the zebrafish embryo the core promoter largely contributes to interaction specificities. This finding is consistent with the proposed contribution of core promoter diversity to differential gene regulation and the observed core promoter selective action of enhancers in *Drosophila* (Smale 2001; Juven-Gershon and Kadonaga 2009). However, in the context of a vertebrate embryo the contribution of core promoter identity to enhancer-promoter interaction specificity has not been demonstrated before.

We investigated enhancer - core promoter interactions by analysing two aspects: the general ability of core promoters to respond to enhancer input (interactivity) and the activation of the core promoter in expected activity domains of the enhancer (specificity). Although we observed substantial variations in overall upregulation, we did not use overall expression strength (expressivity) to make general conclusion about the interaction abilities of core promoters as it was considerably influenced by several factors: the dynamic range observed, the tendency of strong promoters to be highly upregulated by enhancers as well as the variations introduced by the transient transgenesis approach (see reproducibility

control experiment in **chapter 4**). We only used upregulation to make a judgment whether a certain enhancer-promoter combination was active or inactive and to highlight trends.

The interactivity analysis revealed that all core promoters had differential abilities to be activated by enhancers, including core promoters that were rather promiscuous and core promoters that were highly selective. It must be noted that the threshold chosen to consider an interaction as 'active' (2-fold over control level) was arbitrarily chosen, as there is no measure to draw a line between a biologically meaningful and non-meaningful upregulation, especially in respect of the technical limitations of the microinjection approach. Therefore, this choice potentially introduces false negative and false positive interactions. Nevertheless, we assume that potential artefacts are restricted to the single construct level and the application of a common threshold to a large set of experiments reveals overall tendencies of interaction specificities of core promoters which was ultimately the goal of this study.

The automated spatial quantification of reporter signal allowed us to compare the domain specific activation of distinct enhancer-promoter combinations. Therefore, we investigated the distribution of reporter signal within segmentation domains which showed major overlap with previously experimentally verified enhancer activity. We did not compare reporter signal activity to the endogenous gene expression patterns as the enhancers do not always fully recapitulate the gene expression pattern as they frequently regulate target gene expression in subdomains only due to the modular nature of cis-regulation. This analysis revealed that the core promoters also largely differ in their ability to be activated in a tissue-specific manner. We identified core promoters that tended to give ectopic activity while

others were mainly activated in the expected domains. The specificity ranking was based on reporter signal activity within domains of expected enhancer activity, while we disregarded the overall activity of experiments for this analysis, i.e. that this measurement was based on pattern analysis and is therefore not influenced by signal intensity variations inherent to the approach. Nevertheless, the 2-dimensional approach and the relatively low resolution of 8 segmentation domains could hide subtle variations in reporter gene expression. While we could have increased the resolution by creating more segmentation domains, we believe that the chosen segmentation provides sufficient resolution to analyse spatial activity of CREs as they reflect well defined organs and tissues and thus allow a reasonably accurate judgement of overall expression patterns (see also **Chapter 4**). Additionally, the 8 domains provide similar resolution to the segmentation used in other studies investigating CRE activities in transient transgenic assays in zebrafish (Woolfe et al. 2005).

By choosing an assay that brought enhancers and core promoters into close proximity we excluded regulatory mechanism that contribute to interaction specificities in the chromosomal context, suggesting that the observed results mainly depend on core promoter identity. In our approach we isolated enhancers and core promoters from their endogenous genomic location and analysed the interaction specificity in an artificial transgenic context. Additionally, the cloning into the multisite gateway expression vector brought core promoters and enhancers into close vicinity. Therefore, we excluded endogenous regulatory mechanism controlling enhancer - promoter interaction such as promoter proximal tethering elements, insulating activities of CREs and chromosomal boundaries as well as the distance between enhancers and core promoters. Thus, the interaction abilities of core promoters

should mainly depend on properties encoded in the sequence of the respective DNA elements. However, in this study we could not establish a clear link between core promoter associated properties such as CPE occurrence, overlap with CGIs, TSS distribution or promoter strength (based on EST abundance) and the interactivity/specificity of a core promoter. Nevertheless, there was a tendency that core promoters with significant basal activity responded more strongly to enhancer input. Additionally, the TATA-box or TLS were enriched among core promoters that were interactive, thus it is tempting to speculate that the presence of this CPE is associated with enhancer responsiveness, especially as the TATA-box appears to correlate with tissue-specifically expressed genes (Carninci et al. 2006). However, the establishment of statistically significant correlations between core promoter properties and interactivity/specificity would require extending the current assay to a much larger and more diverse set of core promoters.

Since we did not address the mechanistic basis for the apparent enhancer-promoter specificity, models explaining these observations are speculative and hypothetical. Like already mentioned above, a hypothesis is that protein-complexes nucleating at distinct core promoters could have differential abilities to interact or communicate with enhancer bound proteins, respectively (**Figure 5.1**). For example, the core promoter specific composition of PICs could restrict the activating input of enhancers by permitting protein-protein interactions with some enhancer bound activators while inhibiting interactions with proteins bound at alternative enhancers. Moreover, the differential interaction of PIC components with core promoter DNA could induce conformational changes of the core promoter bound protein complexes, thus influencing protein-protein interaction abilities. This model is

supported by many observations including the cell type specific expression of TFIID components, the existence of TFIID alternative protein complexes (Hochheimer and Tjian 2003; Muller and Tora 2004), the existence and functional diversification of TBP-family members (Davidson 2003; Reina and Hernandez 2007) as well as the large variety of core promoter compositions and, importantly, their reported contribution to enhancer - promoter interaction specificity (Smale 2001; Sandelin et al. 2007; Juven-Gershon and Kadonaga 2009). In addition, enhancer - core promoter interaction specificities could be influenced by the differential recruitment of co-activator complexes, such as the mediator complexes, by enhancer bound activators. These co-activator complexes could exhibit differential subunit compositions, therefore restricting their function to certain subsets of core promoters. Furthermore, distinct activators could interact differentially with co-activators, thus inducing conformational changes of the complexes that might facilitate the interaction with certain core promoter bound proteins (Levine and Tjian 2003; Taatjes et al. 2004).

5.3.4. A core promoter resource for transgenic applications

The success of transgenic applications where reporter signal is generated by the interaction of enhancers and promoters is largely dependent on the core promoter used. Therefore, the 19 core promoters analysed can serve as a resource for core promoter choice in such applications. The interaction specificity ranking will aid in choosing promoters for the analysis of CRE function and in enhancer trapping assays, as well as in conventional transgenic reporter gene assays to generate cell-type specific transgenic lines. Notably,

strong promoters were efficiently activated by enhancers, while they also tended to generate more ectopic expression (e.g. *klf4* and *apoeb*). In contrast, some weaker promoters (such as *hsp70* and *eng2b*) were also likely to interact with a variety of enhancers and thus may be useful in enhancer studies. Interestingly, the *krt4* and *hsp70* core promoters which we identified as good 'interactors' have been successfully used in previous enhancer trap assays, although alternative promoter fragments containing additional upstream sequences were utilized (Parinov et al. 2004; Nagayoshi et al. 2008). For the *krt4* core promoter, it could be shown that it is efficiently activated by enhancers of genes belonging to various gene ontologies at a high frequency (Parinov et al. 2004; Amsterdam and Becker 2005). Thus, this provides coincidental support for the identification of *krt4* as one of the best enhancer responding core promoters.

6. SUMMARY AND CONCLUSIONS

6.1. Differential requirement of TBP for promoter function during zebrafish mid-blastula transition

In RNA Polymerase II dependent transcription, the binding of TBP to the TATA-box has been considered as the key rate-limiting step in promoter recognition and thus in initiation of transcription. Moreover, TBP was shown to be rate-limiting in zygotic genome activation. However, functional analyses of TBP requirement in several model organisms have revealed that mRNA levels and transcriptional activity of RNA Polymerase II are to a large extent not dependent on TBP function (Veenstra et al. 2000; Muller et al. 2001; Martianov et al. 2002; Ferg et al. 2007; Jacobi et al. 2007).

Since most of the evidence about the non-universality of TBP function derives from analyses of steady state transcript levels, which can be influenced by posttranscriptional regulatory mechanisms (Jacobi et al. 2007; Ferg et al. 2007), the main objective of the study presented in **chapter 3** was to directly address the transcriptional requirement of TBP using the activation of zygotic transcription during zebrafish mid-blastula transition as an *in-vivo* model system. Therefore, I isolated a set of zebrafish promoters of early expressed zygotic genes and functionally analysed them in a transient transgenic reporter gene assay. The efficient block of TBP function by TBP-MO in the early zebrafish embryo allowed the analysis of TBP dependence of 28 promoters in the context of a living vertebrate embryo providing further insights into the functional requirement of TBP in activating zygotic transcription.

The comparison of promoter activities in control and TBP-depleted embryos revealed that around 30% of the assayed promoter fragments required TBP for transcriptional activity supporting the crucial role of TBP in initiation of transcription. However, the majority of promoter activities (approx. 52%) were not significantly changed suggesting that TBP is not required or that compensatory or redundant mechanisms exist to initiate transcription from these promoters. Moreover, the loss of TBP led to upregulation of promoters indicating a repressive role of TBP in transcription. Finally, I carried out specificity control experiments including rescue experiments and the usage of an alternative TBP-Morpholino that showed that the observed promoter responses in TBP morphants were specific to the loss of TBP, thus providing experimental verification of my main observations. Together, the observed effects of loss of TBP on the promoter level complement transcript based studies that revealed a differential requirement of TBP function during activation of zygotic transcription (Ferg et al. 2007), thus expanding the understanding of the specific roles of TBP in regulating gene expression during early zebrafish development.

The analysis for TBP dependence of promoters could not establish a link between the occurrence of CPEs (e.g. TATA-Box) and the TBP dependence of a promoter, potentially due to the generally very low abundance of CPEs. Potentially, the generation of such correlations would require the analysis of a much larger set of promoters. Moreover, the study could not reveal if the observed promoter responses in TBP morphants are direct or indirect. Thus, suggested future work includes the analysis of promoter occupancy and intersection with functional data on the single promoter as well as on a genome-wide level.

6.2. Development of an high throughput pipeline for the mapping of fluorescent reporter signal in zebrafish embryos

The availability of simple and cheap experimental techniques and the ease of detection of organ/tissue and cellular phenotypes in the transparent *ex utero* developing zebrafish embryo combined with the large number of embryos available make the zebrafish an ideal model for high content applications (Zon and Peterson 2005), in particular for the rapid and efficient testing of *cis*-regulatory elements in transgenic embryos (Ellingsen et al. 2005; Woolfe et al. 2005). However, the application of genuine HTS using zebrafish embryos demands the development of protocols that allow the automated acquisition and analysis of the tremendous amount of data that can be generated in such approaches.

The analysis of enhancer – core promoter specificity (see **6.3**) in a large scale screen using transient transgenic zebrafish embryos necessitated a protocol that allowed the efficient scoring of spatial transgene activities. However, most of the currently available technologies do not provide the flexibility to analyse various reporter gene expression patterns in a high-throughput (Liu et al. 2006; Tran et al. 2007; Vogt et al. 2009). Therefore, the objective of the study discussed in **chapter 4** was to develop a flexible high throughput image acquisition and processing system for the quantitative and spatial registration of fluorescence signals in transient transgenic zebrafish embryos. In an collaborative and interdisciplinary effort, we established a protocol for the automated image acquisition of thousands of live-arrayed zebrafish embryos using an automated screening microscope. This was combined with custom-designed automated image analysis software for the detection of fluorescence

signals in zebrafish embryos in a quantitative and domain specific manner. The visualization of expression patterns by mean and maximum projections provided intuitive access to the experimental data and, moreover, the quantification of signal within user-defined segmentation domains provided a simplified 2-dimensional resolution of spatial transgene activity. We demonstrated the quality of the quantification algorithm by the successful distinction of fluorescence intensities of hemi- and heterozygous embryos of a stable transgenic line. Furthermore, we could verify the specificity of domain registration by validation experiments and visual correlation of domain specific quantification with expression patterns. We demonstrated the utility of the developed tool in the analysis of continuous and mosaic reporter gene expression patterns. Additionally, we could demonstrate that the transient transgenesis approach combined with automated reconstruction of expression patterns is suitable to analyse spatial activity of transgenes in a high-throughput manner. Besides, the direct application in our enhancer - core promoter interaction screen, in which we used transient transgenic prim-20 stage embryos, we successfully tested the applicability of the developed system in fluorescent reporter gene expression based phenotyping approaches and with different developmental stages.

Overall, we suggest that the system might be useful for the functional large-scale analysis of bioinformatically predicted CREs (Woolfe et al. 2005) and in fluorescent reporter gene based phenotypic screens (Jin et al. 2007). A drawback of the approach is the 2-dimensional read-out, which restricts the overall possible spatial resolution. Therefore, future protocols would be largely improved by the utilization of technologies that allow a three – dimensional registration of signals.

6.3. A high-throughput screen in zebrafish embryos reveals differential enhancer – core promoter interaction abilities

The core promoter is the ultimate target of all gene regulatory processes, yet its contribution to differential gene expression is often overlooked. The multitude of existing core promoter types together with the functional diversity of core promoter binding proteins suggests that it can contribute to combinatorial gene regulation (Sandelin et al. 2007). A suggested regulatory role of the core promoter is that it can contribute to enhancer - promoter specificity. In fact, in the *Drosophila* model system it was found that core promoter composition or identity can restrict the activating effect of enhancers (Juven-Gershon and Kadonaga 2009).

In the developing vertebrate embryo direct evidence for enhancer - core promoter interaction specificities is lacking. Therefore, the aim of the study presented in **Chapter 5** was to explore the interaction specificities between a diverse set of core promoters and tissue-specific enhancers in the context of the developing zebrafish embryo. To this end, we carried out a high-throughput screening experiment in which we addressed the interaction specificities between 11 enhancers and 19 core promoters in all possible combinations. The enhancer and core promoters were linked in front of a *venus gfp* reporter gene using a multisite gateway system. To analyse the reporter constructs we developed and utilized a high-throughput pipeline for the automated registration of spatial reporter signal activity in transient transgenic zebrafish embryos (see **6.2**). The analysis of 202 reporter constructs revealed a diversity of enhancer-promoter interactions and showed that core promoters

largely differ in their general ability to interact with enhancers as well as their ability to recapitulate tissue-specific activity inherent to the linked enhancer. By choosing an assay that brought enhancers and core promoters into close proximity we excluded regulatory mechanisms that contribute to interaction specificities in the chromosomal context, suggesting that the observed results mainly depend on the core promoter identity. However, we could not establish a clear correlation between interaction specificities and core promoter properties. This would require the screening of a much larger promoter set to generate sufficiently high numbers for statistical analyses. Overall, the results indicate that also in the context of the zebrafish embryo the core promoter largely contributes to interaction specificities.

The core promoter can have large influence on the outcome of transgenic applications that depend on the interaction of enhancers and promoters (e.g. enhancer trap) as it potentially imposes bias on enhancer activity. In this project, we have identified a set of promoters with broad interaction abilities and low tendency for ectopic activity. Therefore, the screening of 19 core promoters for interaction specificities has generated a resource for promoter choice in transgenic applications.

7. APPENDICES

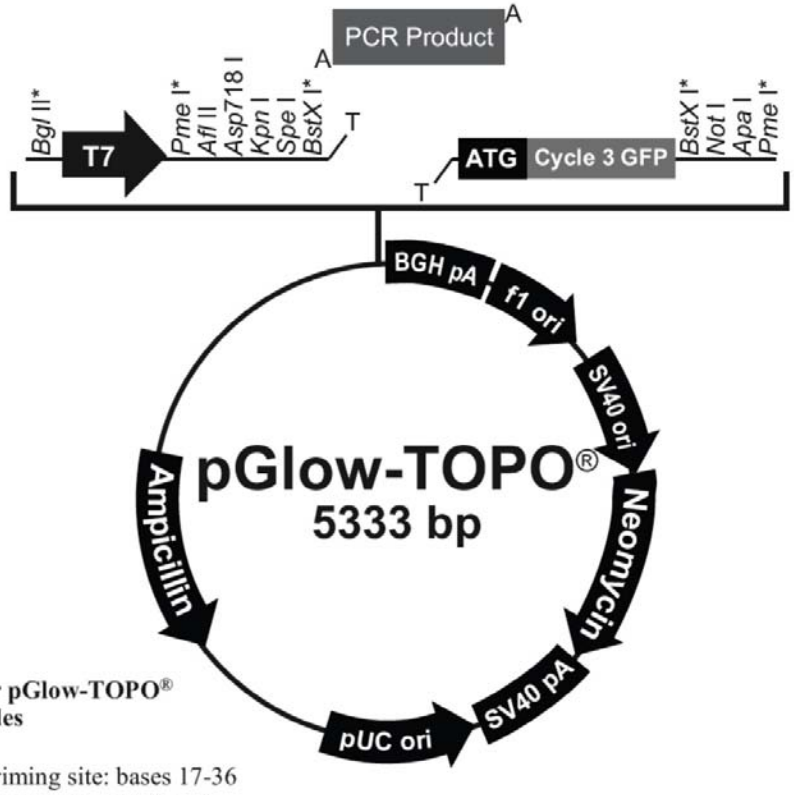
7.1. Appendix 1 - Additional information for chapter 3

Nested primer pairs used to amplify the promoter regions

promoter region		left primer	right primer
anxa1a	outer	GCATTGTGGTGGTCTGGATT	TTTTTAACAAACTGACAACACTGG
	inner	GTCAATCAATTGAGAATCACAAGA	TGGCTATTAAATGAAGGGTAAGAA
apoeb	outer	GTTACTTCAGTTTTCAGCATCACATT	TTAAGCAGATGACCTGATAAATATGC
	inner	TAGCAATTCTTATTAGTGTTTGATCATT	ATGTTGATGGACAGACAAGAACA
atp6v1g1	outer	TTACATTTAGCGCTTTTACACATAGG	TGACAAGCAAAACGATTACTTACTTT
	inner	CCAATTATGATATAGGGATGGTTAGG	GGTACGGATTTTATTTGAATCAATG
c20orf45	outer	ACTCATACATTACGGCCAATTTAGTT	GGTATAAGAAAAACCGTCACTCACTT
	inner	AAACTCCACACAGAAACACCAACT	TTTTTAGATGTTTAACGGGCGACT
ccnb2	outer	AAAAGTTTTTAGTGTATAGGGTGGACA	TTTCTTCTCCAAATACTCACCAGAG
	inner	CACATTTTCAATGAAGTGTAGCTTAT	GATGCCTCGGTTAAATTAAATAAAAA
ccne	outer	TGTGTGTTTAGCAGAAGAAAGAAACT	TATCCTCGACTGCAAAAGTTAGTATG
	inner	ACATTCTATAACAAAATGACCAGTGC	CATATGGCCGGGATGTGATGTGTTC
cdk7	outer	TCTGTTGCTCTTGGTGATTCTAAG	GAGTGAAATCGAATGTTGTACTGC
	inner	GTTTAAACATACACCCGCTCCTCT	ATGTTTTTCAACTGTAAACTTGTTCTTT
clstn2	outer	CAGACAGTTTGTGTTGCCTTAACT	CGATGACCTAATTCTTCCTCTGTT
	inner	TGTGTTCAATCCACATAAAGTTGTTA	GACGATGTACTGAAGTCTCGCTAGT
cxadr	outer	AACATGCAACTACGCCATCA	GCTGTTTCGCGTTCACTCTT
	inner	ATCCAGCAGGGGTCATTAGA	CAGAAAGTCCAGCCGCAGCTC
foxa	outer	CAAACACATTAAGTTGAACATTACCA	CTCTTTCTCGTTTAACTTCGTCAATC
	inner	TTTTCAGTGTAGATTCGATTTCAAAG	TCAACTAAATGCGAAATGTTTCAGT
gngt2	outer	CCACCGGTGTCGCTAATAAT	TATGACCAGAGCACTGTTATGTTT
	inner	TCCCCCTAAAATTGTTCCAA	CTATTGGTCAGTGGGTCAGC
gtf2a1	outer	AATTACTAAACGCTGTTACAACCAAA	GAAAAACACACCGCGAATATCAAG
	inner	TTCGCTAATTATACAGTACACAAGCA	ATAAAAACACAGCTCTTTACGGTCTT
her7	outer	TTTACACAGTTATCCTGTGAGCAA	TCAGTGAGGATATGATTCCAGAAA
	inner	GCGTCTCCATTATATGGCAAG	TGTCCAAGCTTTCTTCTGTTAGG
klf4	outer	TTAAACAAGTGCAGGCCGTTATC	AGTAGCTACAAACAAACACGCTGA
	inner	AACCCGGTGCCTACAAATAA	AGGTTGAACTCACCCGAACT

promoter region		left primer	right primer
kpna2	outer	CCATGCGCTATAGATCAATAAACTAA	CCTGGAAAATAAAATATATCCGTGAC
	inner	TAGCATCCATTTGATTGTAAAAGTGT	TAAATACGAGCAAAAGTAGGAACTCA
krt4	outer	CGATTGCGTAATGCTATTAAATG	CATGATGCCTGTGTCTTTGAG
	inner	GCTGTGACACCCAACTGCT	TGCTGAGAAGGAGGTACGAGA
marcksl1	outer	ACATGCAAACTCCTCACAGAAAC	CCTCTTTGTGTCTTCTTCACAG
	inner	CAGCAACCTTCTTGCTGTGA	GGGGGTCCTTCAATTCAGAC
mkks	outer	CCTGGTCTTTGTAGCTGACTATGTAT	ACCTAATAGCTTATAAAACGCGCATA
	inner	TGAGCTAACTATACCCTCTGTCTCAA	ACTAACTGAAGAAAGAACAGCCAAAG
ndr1	outer	GGCTCACTAGACTTAAAGTTGGTTTT	TTTATTCCTGAGGGTCTCTTGAAAT
	inner	TACTAATGACACAGATGCAACAAAAC	ATGTCAAATCAAGGTAATAATCCACA
otx1	outer	TTGATCTATTGAAGAACGTGTAAAGC	CTGGACTTGCTGTAGTCTATTCAGAG
	inner	GTTTGATTGGATGCCCCATATAGTA	TTCTCGTTCGTGTCTGTCACTAGTTT
pcbp2	outer	GTTGCTATGAAATCAATATTTACTCTTT	AGCTCCTCTCTTACAAAGACAACAAG
	inner	CAATATTCTACATCATTTGCTCTCAA	AGCTCCTCTCTTACAAAGACAACAAG
raver1	outer	CATCAATCTGTCACTAGAAGATTTCA	GTCGGTGAACACTCGTCCTC
	inner	CGACATCAGCTTTGCGTTTA	TCAGTGTTTTTGGCCGTTT
rdh10	outer	AGAGGAATCAGCATAGTCGATAAGAC	CAGATTATTTTAAAAGTGACCACGAA
	inner	CTTGTACTCCTTGTGGTAAGCTTCAG	ACTAGATAAGAAATCAAACCGCAAAC
sft2d3	outer	CATATAGGGTGGGGTGTTATTAAA	GTCTATGGTTGTGCTGGAGCTG
	inner	CTGTCACAATACTGTTACCCAAAAAC	GCTTTTAGAGTGAGCCAGATATTCCT
sox3	outer	GTAACAACACTTAAGAGCAGTTTGGA	TAATCTCGGTTTCCATCATGTTATAC
	inner	GTGTTCTCGATGTGCTAATACTGATT	TAAAGTCTTTCAAAAGTCTCCACAAC
tbp	outer	GGCCAAGCTTTACTAAAAACAAATAA	AACCCAAGCAGTTAACACTTGTAATA
	inner	AATAGCAGCATCACATTCAACTACTC	CTCCGTCTAGAAACAGTGTTAGATCA
tbpl1	outer	CATTAACTTGCCGTAAATTTTGAGTT	AGTCCGTTCCAGGTAATTTTATTATG
	inner	AATTTCATTGTTTATTTGCAGAATTG	GAAGATCAAGGAAAGTAAACAAAAGC
tbx16	outer	TCAGGTAGCGAATCTCAGCA	TCACCTCTGATAGCCTGCATT
	inner	CCCATCCACACTGAGACAGA	TGTCCTCTAAAAGAAAATGTCAGAAA
thy1	outer	GAGTGTAACCTGTGCAGATCTTTTTA	CGTTAGGAACATAAGTCAAGATAGCA
	inner	AGAGCCTCAAATATGATGACTTTCTT	CGTTAGGAACATAAGTCAAGATAGCA
tppp3	outer	TGTGGGTCATGTACGTTTGG	TTACCTGTGTTTCTGGCTTCC
	inner	TGCATAAAGCGCTGTGAATC	ACACACACTCCCAGGTCTGT
tram1	outer	TCTTTTCGACTTAGTCCCTAGTAATCA	ATAAGACCCAGGAGAAACACCATAG
	inner	AAGAACCTTTCTGTCTTACGCATC	CATCTTGGTAGATGTTGAGAGAACTT
vox	outer	AATTGTGTTGCTACTGTAATCCCTTT	TAAATAAACCTTGTCGTATGATGTCG
	inner	GTTGCTACTGTAATCCCTTTGATCC	GAAGTGAATCTATCCGAGTACAATCC
zgc:66242	outer	CGCAGCCTCGGTTATATTTTATTA	GCAGATGCTACTGCAAGTAAACAA
	inner	TCGTTTAACATTCTGATTCCAA	AACGTACTAACACACGCAGGT

pGlow-TOPO plasmid map (Invitrogen, Paisley, UK)



Comments for pGlow-TOPO®
5333 nucleotides

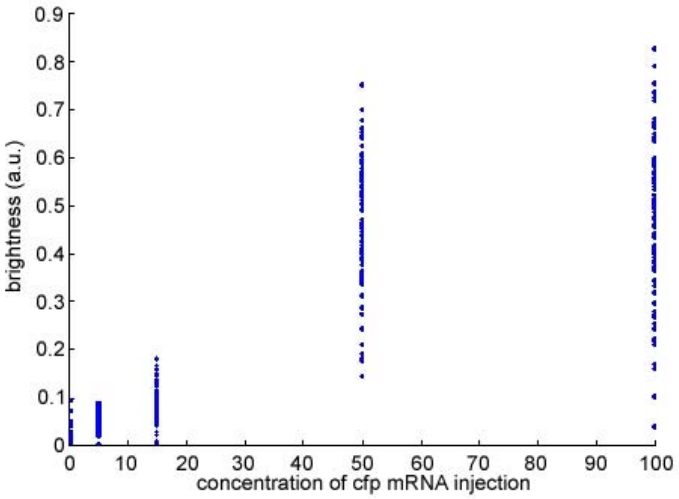
T7 promoter/priming site: bases 17-36
TOPO® Cloning site: bases 116-117
Initiation ATG: bases 140-142
Cycle 3 GFP ORF: bases 140-859
GFP Reverse priming site: bases 251-272
BGH polyadenylation sequence: bases 926-1153
f1 origin: bases 1199-1627
SV40 promoter and origin: bases 1681-1963
Neomycin resistance gene: bases 2038-2832
SV40 polyadenylation sequence: bases 3006-3136
pUC origin: bases 3519-4192 (complementary strand)
Ampicillin resistance gene: bases 4337-5197 (complementary strand)

*pGlow-TOPO™ has only two *Pme* I sites as shown.

```
                  Bgl II*     T7 promoter/priming site                  Pme I*  Afl II     Asp718 I  Kpn I
1   GACGGATCGG GAGATCTAAT ACGACTCACT ATAGGGAGAC CCAAGCTGGC TAGCGTTTAA ACTTAAGCTT GGTACCGAGC

          Spe I              BstX I*                                  Bgl II* Pst I* Xba I*      GFP ORF
81  TCGGATCCAC TAGTCCAGTG TGGTGGAATT GCCCTT PCR AAGG GCAATTCTGC AGATCTAGA ATG GCT AGC
                              ... CGGGAA Product TTCC ...                    Met Ala Ser

149 AAA GGA GAA GAA CTT TTC ACT GGA GTT GTC CCA ATT CTT GTT GAA TTA GAT GGT GAT GTT AAT GGG
    Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly
                                                        GFP Reverse priming site
215 CAC AAA TTT TCT GTC AGT GGA GAG GGT GAA GGT GAT GCT ACA TAC GGA AAG CTT ACC CTT AAA TTT
    His Lys Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe

281 ATT TGC ACT ACT GGA AAA ...
    Ile Cys Thr Thr Gly Lys ...
```

7.2. Appendix 2 - Additional information for chapter 4

CFP concentration curve: CFP signal quantification results of embryos injected with 0.5 ng, 5 ng, 15 ng, 50 ng and 100 ng of cfp mRNA.



Validation of domain registration. The tables show the results of domain assignment after warping. Yellow highlighted numbers indicate the percentage of correct assignments.

1.) Validation experiment using 29 randomly chosen prim-20 embryos

%	Yo	Ey	Sk	Br	Mh	He	No	Sp	Sum
Yo	**97**	0	2	0	0	0	0	0	99
Ey	0	**77**	0	21	1	0	0	0	99
Sk	2	0	**83**	2	0	1	0	4	92
Br	2	6	5	**76**	4	3	1	1	98
Mh	0	1	4	19	**74**	0	0	0	98
He	8	0	8	4	0	**72**	0	0	92
No	6	0	0	4	0	0	**84**	4	98
Sp	0	0	5	2	0	0	6	**86**	99
Sum	115	84	107	128	79	76	91	95	

Average accuracy: 81.4%

2.) Validation experiment using 30 stage synchronous prim-20 embryos from random experiments

%	Yo	Ey	Sk	Br	Mh	He	No	Sp	Sum
Yo	**98**	0	1	0	0	0	0	0	99
Ey	0	**77**	0	21	2	0	0	0	100
Sk	3	0	**86**	2	0	1	0	4	96
Br	2	5	3	**81**	5	1	1	1	99
Mh	0	0	4	5	**91**	0	0	0	100
He	13	0	2	9	0	**76**	0	0	100
No	3	0	0	2	0	0	**87**	8	100
Sp	0	0	4	1	0	0	1	**93**	99
Sum	119	82	100	121	98	78	89	106	

Average accuracy: 86%

3.) Validation experiment using 19 stage synchronous long-pec embryos from one experiment

%	Yo	Ey	Sk	Br	Mh	He	No	Sp	Sum
Yo	**94**	0	4	0	0	0	0	0	98
Ey	0	**84**	0	15	1	0	0	0	100
Sk	2	0	**77**	7	1	1	0	9	97
Br	0	6	3	**87**	2	0	1	0	99
Mh	0	7	3	14	**75**	0	0	0	99
He	10	0	3	3	0	**74**	0	0	90
No	4	0	0	4	0	0	**85**	3	96
Sp	0	0	1	9	0	0	9	**81**	100
Sum	110	97	91	139	79	75	95	93	

Average accuracy: 82.5%

Domain specific GFP quantification for reporter expression phenotyping. Tables show the domain specific quantification results of single embryos.

a) *Tg(2.2shh:gfp:ABC)* embryos injected with *control (tol2)* mRNA

Embryo	Yo	Ey	Sk	Br	Mh	He	No	Sp	Sum of Yo, Sk, He, No, Sp	Sum total
1	0.003	1.100	0.000	0.267	0.235	0.006	0.316	0.103	0.42861	2.031
2	0.000	1.214	0.000	0.159	0.079	0.005	0.223	0.130	0.3584	1.81
3	0.000	1.376	0.000	0.215	0.245	0.005	0.231	0.135	0.37124	2.207
4	0.003	1.341	0.000	0.264	0.322	0.006	0.392	0.022	0.4236	2.35
5	0.000	1.404	0.000	0.182	0.086	0.006	0.229	0.157	0.3936	2.066
6	0.002	1.221	0.000	0.096	0.036	0.010	0.338	0.014	0.36432	1.717
7	0.002	1.824	0.000	0.138	0.112	0.010	0.332	0.059	0.40243	2.476
8	0.000	1.213	0.000	0.106	0.085	0.006	0.309	0.095	0.41039	1.815
9	0.000	1.336	0.000	0.167	0.105	0.005	0.247	0.095	0.34793	1.956
10	0.000	1.262	0.000	0.127	0.058	0.007	0.294	0.098	0.40053	1.847
11	0.017	1.368	0.000	0.131	0.081	0.019	0.239	0.029	0.30352	1.884
12	0.000	1.342	0.000	0.163	0.151	0.006	0.284	0.148	0.43786	2.094
13	0.001	1.385	0.000	0.145	0.052	0.061	0.315	0.048	0.42519	2.007
14	0.000	1.308	0.000	0.168	0.196	0.006	0.263	0.080	0.34933	2.022
15	0.000	1.402	0.000	0.181	0.063	0.007	0.289	0.110	0.40603	2.052
16	0.001	0.974	0.000	0.047	0.000	0.012	0.246	0.084	0.34313	1.364
17	0.000	1.180	0.000	0.082	0.063	0.005	0.208	0.080	0.29338	1.618
18	0.003	1.384	0.000	0.125	0.052	0.007	0.401	0.009	0.42008	1.981
19	0.000	1.487	0.000	0.227	0.181	0.018	0.254	0.164	0.43667	2.332
20	0.000	0.933	0.003	0.099	0.000	0.208	0.236	0.112	0.55883	1.591
21	0.000	1.118	0.000	0.108	0.059	0.021	0.327	0.071	0.41922	1.704
22	0.000	1.142	0.000	0.123	0.108	0.006	0.250	0.102	0.35809	1.731
23	0.001	1.505	0.000	0.233	0.184	0.007	0.288	0.160	0.45622	2.378
24	0.002	1.450	0.001	0.194	0.197	0.006	0.266	0.042	0.31574	2.157
25	0.000	1.229	0.000	0.166	0.136	0.006	0.271	0.086	0.36384	1.896
26	0.011	1.760	0.000	0.156	0.058	0.028	0.228	0.025	0.29286	2.267
27	0.006	1.399	0.000	0.134	0.089	0.007	0.348	0.008	0.3686	1.991

b) *Tg(2.2shh:gfp:ABC)* embryos injected with *pki* mRNA

Embryo	Yo	Ey	Sk	Br	Mh	He	No	Sp	Sum of Yo, Sk, He, No, Sp	Sum total
1	0.000	2.380	0.000	0.321	0.164	0.025	0.220	0.176	0.4211	3.287
2	0.000	1.358	0.000	0.121	0.013	0.091	0.225	0.102	0.41844	1.911
3	0.000	1.051	0.000	0.159	0.081	0.005	0.203	0.098	0.30686	1.597
4	0.000	2.308	0.000	0.308	0.107	0.006	0.235	0.177	0.41877	3.142
5	0.000	1.890	0.000	0.240	0.147	0.006	0.230	0.057	0.29398	2.571
6	0.000	2.396	0.000	0.210	0.065	0.015	0.256	0.099	0.37012	3.041
7	0.000	3.050	0.000	0.357	0.262	0.040	0.188	0.173	0.40063	4.069
8	0.000	2.754	0.000	0.236	0.105	0.006	0.273	0.088	0.36762	3.462
9	0.000	2.412	0.000	0.330	0.110	0.043	0.285	0.226	0.55472	3.406
10	0.000	1.196	0.000	0.148	0.047	0.019	0.228	0.128	0.37533	1.767
11	0.000	1.770	0.001	0.316	0.081	0.006	0.312	0.062	0.38027	2.546
12	0.000	1.923	0.000	0.234	0.069	0.043	0.195	0.139	0.37739	2.603
13	0.000	1.032	0.000	0.550	0.516	0.007	0.249	0.167	0.42327	2.522
14	0.000	1.578	0.000	0.226	0.166	0.007	0.346	0.085	0.43726	2.407
15	0.000	1.923	0.001	0.204	0.058	0.061	0.196	0.103	0.36079	2.546
16	0.015	2.127	0.000	0.244	0.043	0.099	0.341	0.011	0.46681	2.88
17	0.000	1.435	0.000	0.244	0.339	0.006	0.211	0.160	0.37741	2.396
18	0.001	1.837	0.002	0.296	0.109	0.024	0.272	0.014	0.31271	2.554
19	0.000	1.501	0.000	0.269	0.116	0.032	0.243	0.155	0.43103	2.316
20	0.000	2.027	0.000	0.226	0.106	0.037	0.240	0.147	0.42353	2.783
21	0.000	2.441	0.001	0.567	0.920	0.005	0.153	0.133	0.29212	4.22
22	0.000	1.804	0.000	0.209	0.145	0.017	0.209	0.070	0.29655	2.455
23	0.000	2.150	0.000	0.309	0.174	0.009	0.210	0.212	0.43086	3.064
24	0.000	2.799	0.000	0.386	0.256	0.006	0.207	0.066	0.27911	3.72
25	0.000	1.627	0.001	0.252	0.021	0.112	0.236	0.172	0.52099	2.421
26	0.000	1.266	0.001	0.214	0.039	0.013	0.227	0.162	0.40368	1.922
27	0.000	2.859	0.007	0.350	0.106	0.084	0.208	0.225	0.52387	3.838
28	0.000	3.034	0.001	0.373	0.157	0.007	0.236	0.192	0.43492	3.998
29	0.000	1.786	0.000	0.174	0.027	0.007	0.262	0.040	0.30871	2.296
30	0.001	2.228	0.000	0.370	0.425	0.007	0.280	0.074	0.36187	3.384
31	0.000	1.206	0.018	0.214	0.136	0.006	0.145	0.111	0.27968	1.835
32	0.001	1.862	0.000	0.205	0.063	0.079	0.283	0.137	0.50058	2.632

7.3. Appendix 3 – Additional information for chapter 5

Primers used for the generation or *enhancer::promoter* constructs

Promoters	**Forward primer**	**Reverse Primer**
control	AAGCTTCGTGTATTGTACGG	TATGTGTGTTATTTTTGTATAG
apoeb	TGGGATGACAAAAGACGA	CCCTTCTGTAATAAGAGGATGA
atp6v1g1	CTGTGAGTCTCGTGCAGTC	GCTTTGGTACGGATTTTATTT
gtf2a1	CAGCTGACTGCACGGTAAGA	CTCTTTACGGTCTTATTCACAGTCC
klf4	ACTACATCCCAAGCGTCAT	AGGTGTTTACTCTCATTCAGT
krt4	CAAGTGTGTGTGTGTGTGAGAG	CTGAGAAGGAGGTACGAGAGTG
ndr1	CTGACCATCAAAAGACTGCAAG	TCAAATCAAGGTAATAACCACACG
pcpb2	CAGTGTGCAGTGTGGAGTACG	GGGGAAGAGGGAAGACACG
rdh10	CATAACAGGCGGACACAC	CCACGAAATCTGCCCAAA
tbp	AGTATGCGAGCCAATAGTGC	CTCCGTCTAGAAACAGTGTTAGATCA
tram1	GCTCTCTCGTCTCCTTGC	GTTCTGGATCACAAACTCATGG
c20orf45	GGGAGATTTTCCATTTAGATTGC	TTTAGAGTTTAACGGGCGACT
ccne	GTGCTTCGTTGTCATTCTAGGAG	AGTCTGTAAGCAGGCAGCAT
shha	GTTTTGTGGGATAACATCAGAAGTG	CGGAGGTTTGCGGCGGGGA
mef2d	CTTCCACACAGCAGTATCCATTCTA	GGTTATTATTTAGCCCGTACAGTCA
dre-mir9-1	TTGATCTAAATACAGTTGACTTTCTAA	GGATTCTTGTTACTTTCGGTTA
elp4	TCTCTTTCTGATTGGCTGAGATTAC	GCTGCGGGTTTTCTTCTGA
hsp70	TTGATTGGTCGAACATGCTG	CAGTCCGCTCGCTGTCTC
eng2b	TGAGAATAAGGCGAGGTTGG	TTCAGAATCAAAGCAGTAGACCTG
Enhancers	**Forward primer**	**Reverse Primer**
control	GTGTGTCATCCTCATCCACG	CATTCCATGATGGTGCTCTG
shha arc	AGCTTGACAACGGAGAGCAT	GAAACGCGCACATAAGGAAT
βactin2-i1	GCAGCCCTTCAAGTCTTTCATTT	GACAAAGGAAGTCCCTCTGCATT
pax6-eye	GCTGGCAAACACACTAACTTCACTT	TCATGTTTCTGTGTTTTGTTGCAGT
eng2b CXE	TATCTTGTCCCCATTCCAACAGAG	ATGTCAGCCAGAATGGTCAAAAAC
eng2b reg5	CGATACACTTTGATGATACGCATTG	GCTCACATGACATTTCTCATTTTCC
dre-mir9-1 enh	ATTCCTTTCCTTGGCATCAA	GGGACACCGTTGTTCCTCTA
myl7	CCATCCTTTTCATCCCTCAA	AGCTTTGTCTACTCACCATGTTC
isl1 zCREST2	TCCAGCACCATAATTCACCA	CCAGTATCGTGCAGCCCTA
dlx2b/6a ei	AATCAGAAAGCAAGGCAAAATTAG	TGTCATATAAACACACTGGCTGAA
mnx1 regB	ATGTGGAGGATCGGTGTCAT	CCGGTGACTTGTTGATTTCC
Gateway primers	**attB-primers (adaptors are in *italic*)**	
attB1	GGGGACAAGTTTGTACA*AAAAAGCAGGCT*	
attB2	GGGGACCACTTTGTACA*AGAAAGCTGGGT*	
attB3	GGGGACAAGTTTGTATA*ATAAAGTAGGCT*	
attB5	GGGGACCACTTTGTATA*CAAAAGTTGGGT*	

Functionality of core promoters

Gene symbol	Functional with co-injected *ar-C* enhancer prior to screen	Functional in the screen with *ar-C* enhancer above maximum of control	Functional with other enhancers in the screen above maximum of control	5' RACE verification	Functional with proximal sequences
apoeb	n/a	+	+	+	chapter 3
atp6v1g1	n/a	-	+	+	chapter 3
gtf2a1	n/a	+	+	+	chapter 3
klf4	n/a	+	+	n/a	chapter 3
krt4	n/a	+	+	n/a	chapter 3
ndr1	n/a	+	+	n/a	chapter 3
pcbp2	n/a	+	-	+	chapter 3
rdh10	n/a	+	+	+	chapter 3
tbp	n/a	+	+	n/a	chapter 3
tram	n/a	+	+	n/a	chapter 3
c20orf45	n/a	+	+	+	chapter 3
ccne	n/a	+	+	n/a	chapter 3
shha	n/a	+	+	n/a	(Müller et al. 1999)
mef2d	43.3%(n=141)	+	(+)	n/a	n/a
dre-mir9-1	37.3%(n=118)	-	+	n/a	n/a
elp4	54.1% (n=98)	+	+	n/a	n/a
hsp70	79.8%(n=129)	+	+	n/a	n/a
eng2b	53.4%(n=163)	+	+	n/a	n/a

Number of embryos and domain specific Venus signal intensity data of enhancer-promoter combinations assayed. IA: images acquired (total number of embryos; n = 17793, mean = 88), QC: quality control (number of embryos passing QC; n = 13253, mean = 66), OR: outlier removal (number of embryos after OR, thus number of embryos analysed; n = 12582, mean = 62), PE: positive embryos (percentage of embryos exceeding maximum of *control::control*).

Enhancer	Promoter	IA	QC	OR	PE	Yo	Ey	Sk	Br	Mh	He	No	Sp
bactin2 i1	*apoeb*	84	66	63	93.7%	0.002	1.493	0.038	0.866	0.789	0.121	0.129	0.208
control	*apoeb*	73	60	57	77.2%	0.007	0.158	0.008	0.045	0.017	0.017	0.026	0.034
dlx2b/6a ei	*apoeb*	58	54	51	45.1%	0.008	0.018	0.000	0.010	0.004	0.000	0.001	0.009
dre-mir9-1	*apoeb*	82	63	60	96.7%	0.125	1.648	0.392	0.862	0.421	0.731	0.548	0.545
eng2b CXE	*apoeb*	77	60	57	45.6%	0.002	0.025	0.003	0.008	0.004	0.001	0.012	0.012
eng2b reg5	*apoeb*	87	74	70	77.1%	0.002	0.165	0.018	0.056	0.086	0.008	0.009	0.015
isl1	*apoeb*	165	117	111	80.2%	0.004	0.553	0.024	0.279	0.232	0.098	0.052	0.121
mnx1 regB	*apoeb*	88	82	78	75.6%	0.006	0.093	0.009	0.033	0.027	0.002	0.012	0.031
myl7	*apoeb*	70	64	61	90.2%	0.052	0.959	0.371	0.660	0.401	0.709	0.221	0.302
pax6b	*apoeb*	48	30	28	64.3%	0.000	0.010	0.002	0.003	0.016	0.008	0.005	0.000
shha arC	*apoeb*	86	75	71	81.7%	0.022	1.345	0.031	0.367	0.367	0.076	0.593	0.150
bactin2 i1	*atp6v1g1*	123	83	79	62.0%	0.517	0.002	0.295	0.029	0.003	0.300	0.094	0.123
control	*atp6v1g1*	82	61	58	15.5%	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000
dlx2b/6a ei	*atp6v1g1*	91	77	73	4.1%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
dre-mir9-1	*atp6v1g1*	83	55	52	25.0%	0.002	0.002	0.004	0.001	0.000	0.002	0.002	0.000
eng2b CXE	*atp6v1g1*	52	49	47	51.1%	0.011	0.013	0.002	0.007	0.018	0.003	0.007	0.034
eng2b reg5	*atp6v1g1*	92	81	77	20.8%	0.006	0.002	0.003	0.001	0.001	0.002	0.006	0.001
isl1	*atp6v1g1*	68	52	49	79.6%	0.001	0.011	0.001	0.004	0.005	0.002	0.003	0.012
mnx1 regB	*atp6v1g1*	80	55	52	51.9%	0.004	0.014	0.011	0.016	0.000	0.002	0.014	0.014
myl7	*atp6v1g1*	77	43	41	36.6%	0.000	0.001	0.000	0.000	0.001	0.002	0.000	0.000
pax6b	*atp6v1g1*	85	78	74	14.9%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
shha arC	*atp6v1g1*	75	45	43	23.3%	0.001	0.001	0.001	0.001	0.000	0.001	0.000	0.000
shha arC	*c20orf45*	81	56	53	100.0%	0.471	3.704	0.450	1.768	1.720	1.596	4.879	1.527
isl1	*c20orf45*	87	78	74	91.9%	0.016	0.230	0.009	0.108	0.056	0.033	0.086	0.150
dre-mir9-1	*c20orf45*	87	63	60	70.0%	0.010	0.069	0.060	0.057	0.018	0.100	0.085	0.076
eng2b reg5	*c20orf45*	87	78	74	85.1%	0.065	0.236	0.055	0.084	0.099	0.073	0.137	0.063
eng2b CXE	*c20orf45*	74	57	54	53.7%	0.014	0.022	0.002	0.005	0.016	0.007	0.031	0.039
dlx2b/6a ei	*c20orf45*	83	72	68	67.6%	0.004	0.053	0.011	0.018	0.008	0.007	0.057	0.030
bactin2 i1	*c20orf45*	91	76	72	88.9%	0.007	0.379	0.013	0.075	0.065	0.027	0.030	0.026
control	*c20orf45*	68	54	51	68.6%	0.002	0.231	0.006	0.035	0.016	0.008	0.025	0.026
mnx1 regB	*c20orf45*	101	78	74	39.2%	0.001	0.033	0.000	0.003	0.001	0.001	0.003	0.004
pax6b	*c20orf45*	37	31	29	65.5%	0.009	0.019	0.003	0.007	0.006	0.003	0.001	0.000
myl7	*c20orf45*	85	68	65	24.6%	0.000	0.002	0.002	0.001	0.001	0.000	0.001	0.000

Enhancer	Promoter	IA	QC	OR	PE	Yo	Ey	Sk	Br	Mh	He	No	Sp
bactin2 i1	*ccne*	71	49	47	27.7%	0.000	0.000	0.001	0.001	0.001	0.000	0.000	0.000
control	*ccne*	90	66	63	17.5%	0.000	0.000	0.001	0.000	0.001	0.001	0.001	0.001
dlx2b/6a ei	*ccne*	81	41	39	33.3%	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.000
dre-mir9-1	*ccne*	78	67	64	9.4%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
eng2b CXE	*ccne*	72	55	52	46.2%	0.000	0.003	0.001	0.004	0.001	0.004	0.006	0.000
eng2b reg5	*ccne*	86	69	66	18.2%	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.000
isl1	*ccne*	215	139	132	29.5%	0.000	0.005	0.000	0.002	0.002	0.000	0.001	0.002
mnx1 regB	*ccne*	83	66	63	12.7%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pax6b	*ccne*	47	30	28	50.0%	0.000	0.001	0.000	0.004	0.002	0.003	0.000	0.000
shha arC	*ccne*	63	37	35	80.0%	0.000	0.184	0.001	0.036	0.001	0.030	0.075	0.021
bactin2 i1	*control*	66	48	46	23.9%	0.000	0.001	0.001	0.003	0.000	0.000	0.000	0.000
control	*control*	76	67	64	0.0%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
dlx2b/6a ei	*control*	84	75	71	22.5%	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000
dre-mir9-1	*control*	87	69	66	10.6%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
eng2b CXE	*control*	92	66	63	14.3%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
eng2b reg5	*control*	86	64	61	11.5%	0.001	0.001	0.000	0.001	0.000	0.000	0.000	0.000
isl1	*control*	89	34	32	21.9%	0.001	0.001	0.002	0.006	0.004	0.012	0.002	0.004
mnx1 regB	*control*	117	83	79	11.4%	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
myl7	*control*	77	67	64	20.3%	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.000
pax6b	*control*	113	73	70	21.4%	0.000	0.001	0.000	0.001	0.000	0.002	0.000	0.000
shha arC	*control*	161	78	74	21.6%	0.005	0.000	0.002	0.000	0.000	0.001	0.010	0.001
bactin2 i1	*dre-mir9-1*	88	74	70	10.0%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
control	*dre-mir9-1*	141	106	101	4.0%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
dlx2b/6a ei	*dre-mir9-1*	88	66	63	6.3%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
dre-mir9-1	*dre-mir9-1*	75	53	50	22.0%	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
eng2b CXE	*dre-mir9-1*	73	56	53	43.4%	0.002	0.002	0.001	0.002	0.003	0.005	0.001	0.001
eng2b reg5	*dre-mir9-1*	73	49	47	21.3%	0.000	0.000	0.000	0.001	0.000	0.000	0.003	0.001
isl1	*dre-mir9-1*	53	47	45	11.1%	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.001
mnx1 regB	*dre-mir9-1*	87	70	66	10.6%	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.001
myl7	*dre-mir9-1*	132	66	63	17.5%	0.001	0.000	0.001	0.001	0.000	0.000	0.000	0.000
pax6b	*dre-mir9-1*	87	70	66	0.0%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
shha arC	*dre-mir9-1*	87	84	80	11.3%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
bactin2 i1	*elp4*	69	63	60	15.0%	0.000	0.000	0.001	0.000	0.000	0.003	0.002	0.001
control	*elp4*	80	67	64	20.3%	0.000	0.001	0.000	0.002	0.002	0.000	0.001	0.001
dlx2b/6a ei	*elp4*	90	65	62	6.5%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
dre-mir9-1	*elp4*	87	59	56	12.5%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
eng2b CXE	*elp4*	47	37	35	37.1%	0.000	0.000	0.002	0.006	0.004	0.003	0.003	0.005
isl1	*elp4*	85	74	70	25.7%	0.001	0.000	0.000	0.000	0.002	0.000	0.002	0.002
mnx1 regB	*elp4*	88	71	67	17.9%	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
myl7	*elp4*	51	40	38	15.8%	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000
pax6b	*elp4*	127	98	93	33.3%	0.001	0.014	0.002	0.003	0.002	0.005	0.004	0.002
shha arC	*elp4*	110	86	82	61.0%	0.003	0.162	0.002	0.020	0.014	0.017	0.062	0.010

Enhancer	Promoter	IA	QC	OR	PE	Yo	Ey	Sk	Br	Mh	He	No	Sp
bactin2 i1	*eng2b*	150	111	105	12.4%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
control	*eng2b*	79	65	62	9.7%	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000
dlx2b/6a ei	*eng2b*	92	52	49	34.7%	0.001	0.001	0.014	0.005	0.008	0.000	0.003	0.013
dre-mir9-1	*eng2b*	80	68	65	38.5%	0.003	0.002	0.006	0.006	0.000	0.003	0.002	0.003
eng2b CXE	*eng2b*	84	65	62	27.4%	0.001	0.000	0.000	0.001	0.000	0.001	0.001	0.001
eng2b reg5	*eng2b*	67	51	48	35.4%	0.000	0.001	0.004	0.005	0.003	0.000	0.001	0.002
isl1	*eng2b*	77	62	59	67.8%	0.012	0.011	0.024	0.051	0.035	0.005	0.016	0.362
mnx1 regB	*eng2b*	90	76	72	27.8%	0.001	0.008	0.004	0.001	0.001	0.001	0.001	0.002
myl7	*eng2b*	95	56	53	17.0%	0.000	0.001	0.000	0.001	0.000	0.000	0.001	0.000
pax6b	*eng2b*	80	59	56	23.2%	0.000	0.000	0.001	0.000	0.000	0.004	0.000	0.000
shha arC	*eng2b*	85	71	67	55.2%	0.002	0.034	0.001	0.002	0.003	0.001	0.042	0.002
bactin2 i1	*gtf2a1*	87	59	56	73.2%	0.002	0.325	0.016	0.128	0.109	0.020	0.035	0.034
control	*gtf2a1*	79	70	66	62.1%	0.002	0.027	0.003	0.009	0.006	0.001	0.011	0.013
dlx2b/6a ei	*gtf2a1*	85	57	54	46.3%	0.000	0.020	0.001	0.004	0.001	0.001	0.003	0.002
dre-mir9-1	*gtf2a1*	69	42	40	95.0%	0.171	4.795	0.920	3.362	3.497	1.599	1.367	1.817
eng2b CXE	*gtf2a1*	78	55	52	59.6%	0.017	0.033	0.048	0.086	0.132	0.032	0.133	0.060
eng2b reg5	*gtf2a1*	56	44	42	7.1%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
isl1	*gtf2a1*	49	41	39	79.5%	0.000	0.001	0.003	0.004	0.004	0.001	0.001	0.026
mnx1 regB	*gtf2a1*	124	97	92	68.5%	0.005	0.272	0.046	0.099	0.081	0.011	0.042	0.032
myl7	*gtf2a1*	54	47	45	26.7%	0.001	0.012	0.002	0.002	0.000	0.000	0.000	0.002
pax6b	*gtf2a1*	78	36	34	73.5%	0.000	0.009	0.002	0.006	0.014	0.014	0.001	0.000
shha arC	*gtf2a1*	250	155	146	82.2%	0.003	0.167	0.003	0.026	0.039	0.048	0.061	0.009
bactin2 i1	*hsp70*	90	77	73	27.4%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
control	*hsp70*	153	133	126	20.6%	0.001	0.000	0.000	0.001	0.002	0.000	0.002	0.001
dlx2b/6a ei	*hsp70*	69	33	31	19.4%	0.000	0.001	0.000	0.000	0.000	0.000	0.001	0.000
dre-mir9-1	*hsp70*	60	42	40	42.5%	0.002	0.009	0.002	0.002	0.000	0.012	0.008	0.000
eng2b CXE	*hsp70*	74	54	51	39.2%	0.005	0.002	0.001	0.003	0.010	0.001	0.000	0.003
eng2b reg5	*hsp70*	73	60	57	10.5%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
isl1	*hsp70*	232	166	158	31.6%	0.000	0.003	0.001	0.004	0.004	0.001	0.001	0.017
mnx1 regB	*hsp70*	81	60	57	49.1%	0.005	0.004	0.007	0.037	0.005	0.003	0.042	0.032
myl7	*hsp70*	64	31	29	44.8%	0.001	0.007	0.001	0.002	0.001	0.006	0.000	0.000
pax6b	*hsp70*	87	74	70	32.9%	0.000	0.012	0.007	0.007	0.002	0.005	0.006	0.010
shha arC	*hsp70*	38	30	28	89.3%	0.006	1.178	0.005	0.166	0.144	0.417	0.348	0.102
bactin2 i1	*klf4*	80	43	41	58.5%	0.003	0.032	0.066	0.043	0.099	0.001	0.003	0.014
control	*klf4*	131	112	107	29.9%	0.002	0.019	0.015	0.014	0.003	0.021	0.005	0.015
dlx2b/6a ei	*klf4*	70	55	52	26.9%	0.002	0.001	0.002	0.001	0.000	0.000	0.000	0.001
dre-mir9-1	*klf4*	90	69	66	74.2%	0.022	0.529	0.318	0.362	0.135	0.149	0.128	0.206
eng2b CXE	*klf4*	61	50	47	89.4%	1.910	1.362	1.076	1.464	2.821	2.767	0.739	0.688
eng2b reg5	*klf4*	74	60	57	10.5%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
isl1	*klf4*	123	86	81	55.6%	0.006	0.011	0.014	0.034	0.033	0.004	0.008	0.108
mnx1 regB	*klf4*	89	70	66	86.4%	0.302	1.118	0.599	0.880	0.640	0.572	0.549	0.676
myl7	*klf4*	54	45	43	88.4%	0.110	0.372	0.241	0.253	0.163	0.399	0.190	0.234
pax6b	*klf4*	82	32	30	50.0%	0.000	0.002	0.005	0.004	0.008	0.007	0.002	0.006
shha arC	*klf4*	82	59	56	85.7%	0.047	0.301	0.179	0.196	0.096	0.148	0.507	0.145

Enhancer	Promoter	IA	QC	OR	PE	Yo	Ey	Sk	Br	Mh	He	No	Sp
bactin2 i1	*krt4*	175	138	131	87.0%	0.003	0.427	0.019	0.357	0.381	0.034	0.049	0.076
control	*krt4*	84	67	64	53.1%	0.002	0.062	0.009	0.023	0.020	0.005	0.008	0.012
dlx2b/6a ei	*krt4*	82	66	63	49.2%	0.003	0.019	0.001	0.005	0.004	0.001	0.004	0.005
dre-mir9-1	*krt4*	72	61	58	93.1%	0.765	3.154	1.283	2.969	2.079	2.149	1.823	1.873
eng2b CXE	*krt4*	84	42	40	82.5%	0.018	0.296	0.011	0.236	0.904	0.124	0.050	0.022
eng2b reg5	*krt4*	90	48	46	30.4%	0.001	0.000	0.001	0.001	0.002	0.000	0.002	0.000
isl1	*krt4*	46	37	35	85.7%	0.000	0.013	0.006	0.015	0.035	0.009	0.002	0.173
mnx1 regB	*krt4*	130	95	90	58.9%	0.013	0.033	0.020	0.014	0.017	0.003	0.021	0.019
myl7	*krt4*	91	79	75	89.3%	0.230	0.402	0.377	0.406	0.212	0.509	0.340	0.342
pax6b	*krt4*	54	33	31	74.2%	0.025	0.118	0.032	0.064	0.039	0.032	0.040	0.030
shha arC	*krt4*	90	71	67	97.0%	0.044	5.247	0.077	1.196	1.522	0.409	2.623	0.634
bactin2 i1	*mef2d*	87	80	76	18.4%	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
control	*mef2d*	84	64	61	9.8%	0.000	0.000	0.000	0.001	0.005	0.000	0.000	0.001
dlx2b/6a ei	*mef2d*	90	65	62	12.9%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
dre-mir9-1	*mef2d*	81	52	49	38.8%	0.000	0.000	0.003	0.002	0.000	0.003	0.001	0.001
eng2b CXE	*mef2d*	67	44	42	23.8%	0.000	0.000	0.000	0.001	0.004	0.007	0.003	0.000
eng2b reg5	*mef2d*	76	58	55	10.9%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
isl1	*mef2d*	49	34	32	15.6%	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000
mnx1 regB	*mef2d*	56	48	46	15.2%	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
myl7	*mef2d*	56	35	33	9.1%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pax6b	*mef2d*	66	35	33	36.4%	0.001	0.000	0.000	0.002	0.001	0.000	0.001	0.001
shha arC	*mef2d*	60	33	31	77.4%	0.008	0.182	0.002	0.040	0.013	0.115	0.170	0.061
bactin2 i1	*ndr1*	40	30	28	85.7%	0.001	0.026	0.003	0.002	0.001	0.002	0.004	0.011
control	*ndr1*	136	114	108	3.7%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
dlx2b/6a ei	*ndr1*	90	73	69	7.2%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
dre-mir9-1	*ndr1*	87	53	50	18.0%	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001
eng2b CXE	*ndr1*	73	50	47	29.8%	0.000	0.000	0.002	0.003	0.001	0.001	0.000	0.002
eng2b reg5	*ndr1*	87	76	72	19.4%	0.000	0.001	0.002	0.001	0.000	0.000	0.000	0.000
isl1	*ndr1*	89	78	74	8.1%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
mnx1 regB	*ndr1*	89	75	71	5.6%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
myl7	*ndr1*	62	53	50	40.0%	0.000	0.001	0.001	0.015	0.000	0.007	0.010	0.003
pax6b	*ndr1*	62	50	47	29.8%	0.002	0.010	0.000	0.018	0.000	0.008	0.000	0.001
shha arC	*ndr1*	80	66	63	74.6%	0.000	0.077	0.000	0.006	0.013	0.002	0.023	0.009
bactin2 i1	*pcbp2*	92	85	81	18.5%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
control	*pcbp2*	79	48	46	8.7%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
dlx2b/6a ei	*pcbp2*	56	45	43	14.0%	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
dre-mir9-1	*pcbp2*	90	81	77	5.2%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
eng2b CXE	*pcbp2*	82	67	64	10.9%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
eng2b reg5	*pcbp2*	89	78	74	4.1%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
isl1	*pcbp2*	170	124	118	14.4%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
mnx1 regB	*pcbp2*	74	62	59	5.1%	0.000	0.001	0.000	0.000	0.000	0.003	0.000	0.000
myl7	*pcbp2*	84	44	42	19.0%	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002
pax6b	*pcbp2*	87	71	67	3.0%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
shha arC	*pcbp2*	90	62	59	74.6%	0.004	0.050	0.000	0.007	0.004	0.022	0.014	0.003

Enhancer	Promoter	IA	QC	OR	PE	Yo	Ey	Sk	Br	Mh	He	No	Sp
bactin2 i1	*rdh10*	73	48	46	52.2%	0.000	0.005	0.003	0.002	0.000	0.000	0.001	0.003
control	*rdh10*	77	62	59	13.6%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
dlx2b/6a ei	*rdh10*	85	51	48	4.2%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
dre-mir9-1	*rdh10*	113	83	78	47.4%	0.005	0.008	0.006	0.003	0.000	0.004	0.003	0.008
eng2b CXE	*rdh10*	85	71	67	80.6%	0.034	0.158	0.031	0.142	0.318	0.029	0.072	0.047
eng2b reg5	*rdh10*	91	83	79	19.0%	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
isl1	*rdh10*	59	35	33	30.3%	0.000	0.000	0.000	0.000	0.001	0.013	0.000	0.000
mnx1 regB	*rdh10*	91	85	81	9.9%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
myl7	*rdh10*	84	66	63	36.5%	0.004	0.000	0.001	0.001	0.001	0.003	0.003	0.000
pax6b	*rdh10*	55	36	34	20.6%	0.000	0.002	0.000	0.002	0.000	0.001	0.000	0.000
shha arC	*rdh10*	62	56	53	66.0%	0.005	0.082	0.000	0.015	0.008	0.012	0.036	0.008
bactin2 i1	*shha*	59	45	43	37.2%	0.003	0.022	0.023	0.044	0.003	0.004	0.012	0.006
control	*shha*	80	67	64	18.8%	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
dlx2b/6a ei	*shha*	86	48	46	4.3%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
dre-mir9-1	*shha*	79	64	61	45.9%	0.004	0.032	0.019	0.016	0.001	0.010	0.012	0.024
eng2b CXE	*shha*	137	73	69	21.7%	0.000	0.000	0.000	0.000	0.001	0.002	0.001	0.000
eng2b reg5	*shha*	91	63	60	30.0%	0.000	0.000	0.004	0.006	0.024	0.001	0.001	0.006
isl1	*shha*	245	164	155	28.4%	0.000	0.001	0.000	0.007	0.000	0.004	0.002	0.006
mnx1 regB	*shha*	152	103	98	17.3%	0.000	0.001	0.001	0.000	0.001	0.000	0.002	0.000
myl7	*shha*	79	64	61	23.0%	0.000	0.000	0.001	0.000	0.000	0.002	0.000	0.000
pax6b	*shha*	88	68	65	15.4%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
shha arC	*shha*	80	61	58	89.7%	0.003	0.415	0.007	0.086	0.020	0.136	0.234	0.038
bactin2 i1	*tbp*	96	80	76	13.2%	0.000	0.000	0.002	0.011	0.000	0.000	0.000	0.000
control	*tbp*	77	38	36	38.9%	0.000	0.001	0.002	0.000	0.000	0.000	0.002	0.001
dlx2b/6a ei	*tbp*	83	55	52	75.0%	0.003	0.562	0.011	0.135	0.114	0.033	0.065	0.034
dre-mir9-1	*tbp*	77	44	42	85.7%	0.026	0.073	0.060	0.033	0.035	0.065	0.082	0.052
eng2b CXE	*tbp*	84	55	52	69.2%	0.000	0.063	0.003	0.015	0.013	0.006	0.016	0.014
eng2b reg5	*tbp*	87	62	59	59.3%	0.004	0.068	0.017	0.007	0.002	0.009	0.019	0.009
isl1	*tbp*	90	63	60	46.7%	0.000	0.006	0.001	0.001	0.001	0.000	0.004	0.003
myl7	*tbp*	111	92	87	8.0%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pax6b	*tbp*	85	74	70	4.3%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
shha arC	*tbp*	174	140	133	54.1%	0.005	0.019	0.002	0.005	0.002	0.004	0.012	0.002
bactin2 i1	*tram1*	172	113	107	13.1%	0.000	0.000	0.001	0.004	0.001	0.000	0.001	0.000
control	*tram1*	119	93	89	3.4%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
dre-mir9-1	*tram1*	89	80	76	17.1%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
eng2b CXE	*tram1*	52	40	38	36.8%	0.001	0.007	0.001	0.003	0.006	0.000	0.002	0.007
eng2b reg5	*tram1*	80	54	51	11.8%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pax6b	*tram1*	109	83	79	16.5%	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
shha arC	*tram1*	95	62	59	49.2%	0.002	0.008	0.001	0.002	0.004	0.009	0.009	0.002

7.4. Appendix 4 – General information

Full names of corresponding gene products of promoters used in this thesis

gene symbol	full name
anxa1a	*annexin A1a*
apoeb	*apolipoprotein Eb precursor*
atp6v1g1	*vacuolar H+ ATPase G1*
c20orf45	*chromosome 20 open reading frame 45*
ccnb2	*cyclin B2*
ccne	*cyclin E*
cdk7	*cyclin-dependent kinase 7*
clstn2	*calsyntenin 2*
dre-mir9-1	*Danio rerio miR-9-1*
elp4	*elongator complex protein 4*
eng2b	*homeobox protein engrailed-2b*
foxa	*forkhead box A*
gngt2	*gamma transducing activity polypeptide 2*
gtf2a1	*general transcription factor IIA*
hsp70	*heat shock cognate 70-kd protein*
klf4	*Kruppel-like factor 4*
kpna2	*karyopherin alpha 2*
krt4	*keratin 4*
marcksl1	*MARCKS-like 1*
mef2d	*myocyte enhancer factor 2d*
ndr1	*nodal-related 1*
otx1	*homeobox protein OTX1*
pcbp2	*poly(rC) binding protein 2*
raver1	*ribonucleoprotein, PTB-binding 1*
rdh10	*retinol dehydrogenase 10*
sft2d3	*SFT2 domain containing 3*
shha	*sonic hedgehog protein A precursor*
sox3	*transcription factor Sox-3*
tbp	*TATA-box-binding protein*
tbpl1	*TATA box-binding protein-like protein 1*
tbx16	*T-box gene 16*
thy1	*Thy-1 cell surface antigen*
tram1	*translocating chain-associating membrane protein*
vox	*ventral homeobox*
cxadr	*coxsackievirus and adenovirus receptor homolog*
tppp3	*tubulin polymerization-promoting protein*
her7	*hairy and enhancer of split related-7*
zgc:66242	*zgc:66242*
mkks	McKusick-Kaufman syndrome

8. REFERENCES

Affymetrix / Cold Spring Harbor Laboratory ENCODE Transcriptome Project (2009). "Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs." *Nature*, 457(7232), 1028-32.

Akalin, A., Fredman, D., Arner, E., Dong, X., Bryne, J. C., Suzuki, H., Daub, C. O., Hayashizaki, Y., and Lenhard, B. (2009). "Transcriptional features of genomic regulatory blocks." *Genome Biol*, 10(4), R38.

Akhtar, W., and Veenstra, G. J. (2009). "TBP2 is a substitute for TBP in Xenopus oocyte transcription." *BMC Biol*, 7, 45.

Akkers, R. C., van Heeringen, S. J., Jacobi, U. G., Janssen-Megens, E. M., Francoijs, K. J., Stunnenberg, H. G., and Veenstra, G. J. (2009). "A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos." *Dev Cell*, 17(3), 425-34.

Albright, S. R., and Tjian, R. (2000). "TAFs revisited: more data reveal new twists and confirm old ideas." *Gene*, 242(1-2), 1-13.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). "Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription." *Dev Cell*, 16(1), 47-57.

Amsterdam, A., and Becker, T. S. (2005). "Transgenes as screening tools to probe and manipulate the zebrafish genome." *Dev Dyn*, 234(2), 255-68.

Amsterdam, A., Nissen, R. M., Sun, Z., Swindell, E. C., Farrington, S., and Hopkins, N. (2004). "Identification of 315 genes essential for early zebrafish development." *Proc Natl Acad Sci U S A*, 101(35), 12792-7.

Anish, R., Hossain, M. B., Jacobson, R. H., and Takada, S. (2009). "Characterization of transcription from TATA-less promoters: identification of a new core promoter element XCPE2 and analysis of factor requirements." *PLoS One*, 4(4), e5103.

Arkhipova, I. R. (1995). "Promoter elements in Drosophila melanogaster revealed by sequence analysis." *Genetics*, 139(3), 1359-69.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). "Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences." *Cell*, 27(2 Pt 1), 299-308.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). "High-resolution profiling of histone methylations in the human genome." *Cell*, 129(4), 823-37.

Bartfai, R., Balduf, C., Hilton, T., Rathmann, Y., Hadzhiev, Y., Tora, L., Orban, L., and Muller, F. (2004). "TBP2, a vertebrate-specific member of the TBP family, is required in embryonic development of zebrafish." *Curr Biol*, 14(7), 593-8.

Barton, L. M., Gottgens, B., Gering, M., Gilbert, J. G., Grafham, D., Rogers, J., Bentley, D., Patient, R., and Green, A. R. (2001). "Regulation of the stem cell leukemia (SCL) gene: a tale of two fishes." *Proc Natl Acad Sci U S A*, 98(12), 6747-52.

Basehoar, A. D., Zanton, S. J., and Pugh, B. F. (2004). "Identification and distinct regulation of yeast TATA box-containing genes." *Cell*, 116(5), 699-709.

Bell, B., and Tora, L. (1999). "Regulation of gene expression by multiple forms of TFIID and other novel TAFII-containing complexes." *Exp Cell Res*, 246(1), 11-9.

Belloni, E., Muenke, M., Roessler, E., Traverso, G., Siegel-Bartelt, J., Frumkin, A., Mitchell, H. F., Donis-Keller, H., Helms, C., Hing, A. V., et al. (1996). "Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly." *Nat Genet*, 14(3), 353-6.

Bessa, J., Tena, J. J., de la Calle-Mustienes, E., Fernandez-Minan, A., Naranjo, S., Fernandez, A., Montoliu, L., Akalin, A., Lenhard, B., Casares, F., et al. (2009). "Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish." *Dev Dyn*, 238(9), 2409-17.

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., et al. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." *Nature*, 447(7146), 799-816.

Blackwood, E. M., and Kadonaga, J. T. (1998). "Going the distance: a current view of enhancer action." *Science*, 281(5373), 60-3.

Brand, M., Leurent, C., Mallouh, V., Tora, L., and Schultz, P. (1999). "Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC." *Science*, 286(5447), 2151-3.

Breathnach, R., and Chambon, P. (1981). "Organization and expression of eucaryotic split genes coding for proteins." *Annu Rev Biochem*, 50, 349-83.

Bucher, P. (1990). "Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences." *J Mol Biol*, 212(4), 563-78.

Buckland, P. R., Hoogendoorn, B., Coleman, S. L., Guy, C. A., Smith, S. K., and O'Donovan, M. C. (2005). "Strong bias in the location of functional promoter polymorphisms." *Hum Mutat*, 26(3), 214-23.

Buono, R. J., and Linser, P. J. (1992). "Transient expression of RSVCAT in transgenic zebrafish made by electroporation." *Mol Mar Biol Biotechnol*, 1(4-5), 271-5.

Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A., and Felsenfeld, G. (2002). "The insulation of genes from external enhancers and silencing chromatin." *Proc Natl Acad Sci U S A*, 99 Suppl 4, 16433-7.

Burke, T. W., and Kadonaga, J. T. (1996). "Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters." *Genes Dev*, 10(6), 711-24.

Burke, T. W., and Kadonaga, J. T. (1997). "The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila." *Genes Dev*, 11(22), 3020-31.

Burley, S. K., and Roeder, R. G. (1996). "Biochemistry and structural biology of transcription factor IID (TFIID)." *Annu Rev Biochem*, 65, 769-99.

Butler, J. E., and Kadonaga, J. T. (2001). "Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs." *Genes Dev*, 15(19), 2515-9.

Butler, J. E., and Kadonaga, J. T. (2002). "The RNA polymerase II core promoter: a key component in the regulation of gene expression." *Genes Dev*, 16(20), 2583-92.

Cairns, B. R. (2009). "The logic of chromatin architecture and remodelling at promoters." *Nature*, 461(7261), 193-8.

Calhoun, V. C., Stathopoulos, A., and Levine, M. (2002). "Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex." *Proc Natl Acad Sci U S A*, 99(14), 9243-7.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). "The transcriptional landscape of the mammalian genome." *Science*, 309(5740), 1559-63.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A., Taylor, M. S., Engstrom, P. G., Frith, M. C., et al. (2006). "Genome-wide analysis of mammalian promoter architecture and evolution." *Nat Genet*, 38(6), 626-35.

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F., and Fraser, P. (2002). "Long-range chromatin regulatory interactions in vivo." *Nat Genet*, 32(4), 623-6.

Chalkley, G. E., and Verrijzer, C. P. (1999). "DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator." *EMBO J*, 18(17), 4835-45.

Chen, J., Ng, S. M., Chang, C., Zhang, Z., Bourdon, J.-C., Lane, D. P., and Peng, J. (2009). "p53 isoform Δ113p53 is a p53 target gene that antagonizes p53 apoptotic activity via BclxL activation in zebrafish." *Genes & Development*, 23(3), 278-290.

Chong, J. A., Moran, M. M., Teichmann, M., Kaczmarek, J. S., Roeder, R., and Clapham, D. E. (2005). "TATA-binding protein (TBP)-like factor (TLF) is a functional regulator of transcription: reciprocal regulation of the neurofibromatosis type 1 and c-fos genes by TLF/TRF2 and TBP." *Mol Cell Biol*, 25(7), 2632-43.

Conaway, J. W., Florens, L., Sato, S., Tomomori-Sato, C., Parmely, T. J., Yao, T., Swanson, S. K., Banks, C. A., Washburn, M. P., and Conaway, R. C. (2005). "The mammalian Mediator complex." *FEBS Lett*, 579(4), 904-8.

Cooper, S. J., Trinklein, N. D., Anton, E. D., Nguyen, L., and Myers, R. M. (2006). "Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome." *Genome Res*, 16(1), 1-10.

Cosma, M. P. (2002). "Ordered recruitment: gene-specific mechanism of transcription activation." *Mol Cell*, 10(2), 227-36.

Crowley, T. E., Hoey, T., Liu, J. K., Jan, Y. N., Jan, L. Y., and Tjian, R. (1993). "A new factor related to TATA-binding protein has highly restricted expression patterns in Drosophila." *Nature*, 361(6412), 557-61.

Dahm, R., and Geisler, R. (2006). "Learning from small fry: the zebrafish as a genetic model organism for aquaculture fish species." *Mar Biotechnol (NY)*, 8(4), 329-45.

Dahm, R., Geisler, R., Nuesslein-Volhard, C. (2005). "Zebrafish (Danio rerio) Genome and Genetics." *Encyclopedia of Molecular Cell Biology and Molecular Medicine*, Volume 15, 2nd Edition.

Dantonel, J. C., Quintin, S., Lakatos, L., Labouesse, M., and Tora, L. (2000). "TBP-like factor is required for embryonic RNA polymerase II transcription in C. elegans." *Mol Cell*, 6(3), 715-22.

Dantonel, J. C., Wurtz, J. M., Poch, O., Moras, D., and Tora, L. (1999). "The TBP-like factor: an alternative transcription factor in metazoa?" *Trends in Biochemical Sciences*, 24(9), 335-9.

Davidson, I. (2003). "The genetics of TBP and TBP-related factors." *Trends in Biochemical Sciences*, 28(7), 391-8.

Davis, W., Jr., and Schultz, R. M. (2000). "Developmental change in TATA-box utilization during preimplantation mouse development." *Dev Biol*, 218(2), 275-83.

Davuluri, R. V., Suzuki, Y., Sugano, S., Plass, C., and Huang, T. H. (2008). "The functional consequences of alternative promoter use in mammalian genomes." *Trends Genet*, 24(4), 167-77.

De Renzis, S., Elemento, O., Tavazoie, S., and Wieschaus, E. F. (2007). "Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo." *PLoS Biol*, 5(5), e117.

Deato, M. D., Marr, M. T., Sottero, T., Inouye, C., Hu, P., and Tjian, R. (2008). "MyoD targets TAF3/TRF3 to activate myogenin transcription." *Mol Cell*, 32(1), 96-105.

Deato, M. D., and Tjian, R. (2007). "Switching of the core transcription machinery during myogenesis." *Genes Dev*, 21(17), 2137-49.

Dekens, M. P., Pelegri, F. J., Maischein, H. M., and Nusslein-Volhard, C. (2003). "The maternal-effect gene futile cycle is essential for pronuclear congression and mitotic spindle assembly in the zebrafish zygote." *Development*, 130(17), 3907-16.

Deng, W., and Roberts, S. G. (2005). "A core promoter element downstream of the TATA box that is recognized by TFIIB." *Genes Dev*, 19(20), 2418-23.

Dickmeis, T., Plessy, C., Rastegar, S., Aanstad, P., Herwig, R., Chalmel, F., Fischer, N., and Strahle, U. (2004). "Expression profiling and comparative genomics identify a conserved regulatory region controlling midline expression in the zebrafish embryo." *Genome Res*, 14(2), 228-38.

Driever, W., Solnica-Krezel, L., Schier, A. F., Neuhauss, S. C., Malicki, J., Stemple, D. L., Stainier, D. Y., Zwartkruis, F., Abdelilah, S., Rangini, Z., et al. (1996). "A genetic screen for mutations affecting embryogenesis in zebrafish." *Development*, 123, 37-46.

Duan, Z. J., Fang, X., Rohde, A., Han, H., Stamatoyannopoulos, G., and Li, Q. (2002). "Developmental specificity of recruitment of TBP to the TATA box of the human gamma-globin gene." *Proc Natl Acad Sci U S A*, 99(8), 5509-14.

Dunican, D. S., Ruzov, A., Hackett, J. A., and Meehan, R. R. (2008). "xDnmt1 regulates transcriptional silencing in pre-MBT Xenopus embryos independently of its catalytic function." *Development*, 135(7), 1295-302.

Edgar, B. A., and Schubiger, G. (1986). "Parameters controlling transcriptional activation during early Drosophila development." *Cell*, 44(6), 871-7.

Ekker, S. C., and Larson, J. D. (2001). "Morphant technology in model developmental systems." *Genesis*, 30(3), 89-93.

Ellies, D. L., Stock, D. W., Hatch, G., Giroux, G., Weiss, K. M., and Ekker, M. (1997). "Relationship between the genomic organization and the overlapping embryonic expression patterns of the zebrafish dlx genes." *Genomics*, 45(3), 580-90.

Ellingsen, S., Laplante, M. A., Konig, M., Kikuta, H., Furmanek, T., Hoivik, E. A., and Becker, T. S. (2005). "Large-scale enhancer detection in the zebrafish genome." *Development*, 132(17), 3799-811.

Engstrom, P. G., Ho Sui, S. J., Drivenes, O., Becker, T. S., and Lenhard, B. (2007). "Genomic regulatory blocks underlie extensive microsynteny conservation in insects." *Genome Res*, 17(12), 1898-908.

Epstein, D. J., McMahon, A. P., and Joyner, A. L. (1999). "Regionalization of Sonic hedgehog transcription along the anteroposterior axis of the mouse central nervous system is regulated by Hnf3-dependent and -independent mechanisms." *Development*, 126(2), 281-92.

Ertzer, R., Muller, F., Hadzhiev, Y., Rathnam, S., Fischer, N., Rastegar, S., and Strahle, U. (2007). "Cooperation of sonic hedgehog enhancers in midline expression." *Dev Biol*, 301(2), 578-89.

Evans, R., Fairley, J. A., and Roberts, S. G. (2001). "Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB." *Genes Dev*, 15(22), 2945-9.

Falender, A. E., Freiman, R. N., Geles, K. G., Lo, K. C., Hwang, K., Lamb, D. J., Morris, P. L., Tjian, R., and Richards, J. S. (2005a). "Maintenance of spermatogenesis requires TAF4b, a gonad-specific subunit of TFIID." *Genes Dev*, 19(7), 794-803.

Falender, A. E., Shimada, M., Lo, Y. K., and Richards, J. S. (2005b). "TAF4b, a TBP associated factor, is required for oocyte development and function." *Dev Biol*, 288(2), 405-19.

Fantes, J., Redeker, B., Breen, M., Boyle, S., Brown, J., Fletcher, J., Jones, S., Bickmore, W., Fukushima, Y., Mannens, M., et al. (1995). "Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype." *Hum Mol Genet*, 4(3), 415-22.

Feldman, B., Gates, M. A., Egan, E. S., Dougan, S. T., Rennebeck, G., Sirotkin, H. I., Schier, A. F., and Talbot, W. S. (1998). "Zebrafish organizer development and germ-layer formation require nodal-related signals." *Nature*, 395(6698), 181-5.

Ferg, M. (2008). "Large scale- and functional analysis for the requirement of TBP-function in early zebrafish development". University Heidelberg. http://www.ub.uni-heidelberg.de/archiv/8920/

Ferg, M., Sanges, R., Gehrig, J., Kiss, J., Bauer, M., Lovas, A., Szabo, M., Yang, L., Straehle, U., Pankratz, M. J., et al. (2007). "The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish." *EMBO J*, 26(17), 3945-56.

Fisher, S., Grice, E. A., Vinton, R. M., Bessling, S. L., and McCallion, A. S. (2006). "Conservation of RET regulatory function from human to zebrafish without sequence similarity." *Science*, 312(5771), 276-9.

FitzGerald, P. C., Shlyakhtenko, A., Mir, A. A., and Vinson, C. (2004). "Clustering of DNA sequences in human promoters." *Genome Res*, 14(8), 1562-74.

Freiman, R. N., Albright, S. R., Zheng, S., Sha, W. C., Hammer, R. E., and Tjian, R. (2001). "Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development." *Science*, 293(5537), 2084-7.

Fu, Y., Frith, M. C., Haverty, P. M., and Weng, Z. (2004). "MotifViz: an analysis and visualization tool for motif discovery." *Nucl. Acids Res.*, 32(suppl_2), W420-423.

Fuda, N. J., Ardehali, M. B., and Lis, J. T. (2009). "Defining mechanisms that regulate RNA polymerase II transcription in vivo." *Nature*, 461(7261), 186-92.

Gaiano, N., Allende, M., Amsterdam, A., Kawakami, K., and Hopkins, N. (1996). "Highly efficient germ-line transmission of proviral insertions in zebrafish." *Proc Natl Acad Sci U S A*, 93(15), 7777-82.

Gardiner-Garden, M., and Frommer, M. (1987). "CpG islands in vertebrate genomes." *J Mol Biol*, 196(2), 261-82.

Gazdag, E., Rajkovic, A., Torres-Padilla, M. E., and 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*, 134(1), 51-62.

Gazdag, E., Santenard, A., Ziegler-Birling, C., Altobelli, G., Poch, O., Tora, L., and Torres-Padilla, M. E. (2009). "TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte." *Genes Dev*, 23(18), 2210-23.

Gehrig, J., Reischl, M., Kalmar, E., Ferg, M., Hadzhiev, Y., Zaucker, A., Song, C., Schindler, S., Liebel, U., and Muller, F. (2009). "Automated high-throughput mapping of promoter-enhancer interactions in zebrafish embryos." *Nat Meth*, 6(12), 911-916.

Giraldez, A. J., Mishima, Y., Rihel, J., Grocock, R. J., Van Dongen, S., Inoue, K., Enright, A. J., and Schier, A. F. (2006). "Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs." *Science*, 312(5770), 75-9.

Glazko, G. V., Koonin, E. V., Rogozin, I. B., and Shabalina, S. A. (2003). "A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions." *Trends in Genetics*, 19(3), 119-124.

Gomez-Skarmeta, J. L., Lenhard, B., and Becker, T. S. (2006). "New technologies, new findings, and new concepts in the study of vertebrate cis-regulatory sequences." *Dev Dyn*, 235(4), 870-85.

Grabher, C., Joly, J. S., and Wittbrodt, J. (2004). "Highly efficient zebrafish transgenesis mediated by the meganuclease I-SceI." *Methods Cell Biol*, 77, 381-401.

Guner, B., and Karlstrom, R. O. (2007). "Cloning of zebrafish nkx6.2 and a comprehensive analysis of the conserved transcriptional response to Hedgehog/Gli signaling in the zebrafish neural tube." *Gene Expr Patterns*, 7(5), 596-605.

Hadzhiev, Y., Lele, Z., Schindler, S., Wilson, S., Ahlberg, P., Strahle, U., and Muller, F. (2007). "Hedgehog signaling patterns the outgrowth of unpaired skeletal appendages in zebrafish." *BMC Developmental Biology*, 7(1), 75.

Haffter, P., and Nusslein-Volhard, C. (1996). "Large scale genetics in a small vertebrate, the zebrafish." *Int J Dev Biol*, 40(1), 221-7.

Hammerschmidt, M., Bitgood, M. J., and McMahon, A. P. (1996). "Protein kinase A is a common negative regulator of Hedgehog signaling in the vertebrate embryo." *Genes Dev*, 10(6), 647-58.

Han, L., and Zhao, Z. (2008). "Comparative analysis of CpG islands in four fish genomes." *Comp Funct Genomics*, 565631.

Hansen, S. K., Takada, S., Jacobson, R. H., Lis, J. T., and Tjian, R. (1997). "Transcription properties of a cell type-specific TATA-binding protein, TRF." *Cell*, 91(1), 71-83.

Hart, D. O., Raha, T., Lawson, N. D., and Green, M. R. (2007). "Initiation of zebrafish haematopoiesis by the TATA-box-binding protein-related factor Trf3." *Nature*, 450(7172), 1082-5.

Hart, D. O., Santra, M. K., Raha, T., and Green, M. R. (2009). "Selective interaction between Trf3 and Taf3 required for early development and hematopoiesis." *Dev Dyn*, 238(10), 2540-9.

Hashimoto, S., Suzuki, Y., Kasai, Y., Morohoshi, K., Yamada, T., Sese, J., Morishita, S., Sugano, S., and Matsushima, K. (2004). "5'-end SAGE for the analysis of transcriptional start sites." *Nat Biotechnol*, 22(9), 1146-9.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., et al. (2009). "Histone modifications at human enhancers reflect global cell-type-specific gene expression." *Nature*, 459(7243), 108-12.

Heintzman, N. D., and Ren, B. (2007). "The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome." *Cell Mol Life Sci*, 64(4), 386-400.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., et al. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." *Nat Genet*, 39(3), 311-8.

Hernandez, N. (1993). "TBP, a universal eukaryotic transcription factor?" *Genes Dev*, 7(7B), 1291-308.

Hiller, M., Chen, X., Pringle, M. J., Suchorolski, M., Sancak, Y., Viswanathan, S., Bolival, B., Lin, T. Y., Marino, S., and Fuller, M. T. (2004). "Testis-specific TAF homologs collaborate to control a tissue-specific transcription program." *Development*, 131(21), 5297-308.

Hiller, M. A., Lin, T. Y., Wood, C., and Fuller, M. T. (2001). "Developmental regulation of transcription by a tissue-specific TAF homolog." *Genes Dev*, 15(8), 1021-30.

Hinits, Y., and Hughes, S. M. (2007). "Mef2s are required for thick filament formation in nascent muscle fibres." *Development*, 134(13), 2511-9.

Hobert, O. (2008). "Gene regulation by transcription factors and microRNAs." *Science*, 319(5871), 1785-6.

Hochheimer, A., and Tjian, R. (2003). "Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression." *Genes Dev*, 17(11), 1309-20.

Hochheimer, A., Zhou, S., Zheng, S., Holmes, M. C., and Tjian, R. (2002). "TRF2 associates with DREF and directs promoter-selective gene expression in Drosophila." *Nature*, 420(6914), 439-45.

Holmes, M. C., and Tjian, R. (2000). "Promoter-selective properties of the TBP-related factor TRF1." *Science*, 288(5467), 867-70.

Hon, G. C., Hawkins, R. D., and Ren, B. (2009). "Predictive chromatin signatures in the mammalian genome." *Hum Mol Genet*, 18(R2), R195-201.

Hoogendoorn, B., Coleman, S. L., Guy, C. A., Smith, K., Bowen, T., Buckland, P. R., and O'Donovan, M. C. (2003). "Functional analysis of human promoter polymorphisms." *Hum Mol Genet*, 12(18), 2249-54.

Hsu, J. Y., Juven-Gershon, T., Marr, M. T., 2nd, Wright, K. J., Tjian, R., and Kadonaga, J. T. (2008). "TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription." *Genes Dev*, 22(17), 2353-8.

Huang, C. J., Tu, C. T., Hsiao, C. D., Hsieh, F. J., and Tsai, H. J. (2003). "Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish." *Dev Dyn*, 228(1), 30-40.

Hubbard, T. J. P., Aken, B. L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Clarke, L., et al. (2009). "Ensembl 2009." *Nucl. Acids Res.*, 37(suppl_1), D690-697.

Huisinga, K. L., and Pugh, B. F. (2004). "A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae." *Mol Cell*, 13(4), 573-85.

Ince, T. A., and Scotto, K. W. (1995). "A conserved downstream element defines a new class of RNA polymerase II promoters." *J Biol Chem*, 270(51), 30249-52.

Indra, A. K., Mohan, W. S., 2nd, Frontini, M., Scheer, E., Messaddeq, N., Metzger, D., and Tora, L. (2005). "TAF10 is required for the establishment of skin barrier function in foetal, but not in adult mouse epidermis." *Dev Biol*, 285(1), 28-37.

Isogai, Y., Keles, S., Prestel, M., Hochheimer, A., and Tjian, R. (2007a). "Transcription of histone gene cluster by differential core-promoter factors." *Genes Dev*, 21(22), 2936-49.

Isogai, Y., Takada, S., Tjian, R., and Keles, S. (2007b). "Novel TRF1/BRF target genes revealed by genome-wide analysis of Drosophila Pol III transcription." *EMBO J*, 26(1), 79-89.

Jacobi, U. G., Akkers, R. C., Pierson, E. S., Weeks, D. L., Dagle, J. M., and Veenstra, G. J. (2007). "TBP paralogs accommodate metazoan- and vertebrate-specific developmental gene regulation." *EMBO J*, 26(17), 3900-9.

Jallow, Z., Jacobi, U. G., Weeks, D. L., Dawid, I. B., and Veenstra, G. J. (2004). "Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in Xenopus." *Proc Natl Acad Sci U S A*, 101(37), 13525-30.

Jarinova, O., Hatch, G., Poitras, L., Prudhomme, C., Grzyb, M., Aubin, J., Berube-Simard, F. A., Jeannotte, L., and Ekker, M. (2008). "Functional resolution of duplicated hoxb5 genes in teleosts." *Development*, 135(21), 3543-53.

Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M., and Epstein, D. J. (2006). "A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers." *Development*, 133(4), 761-72.

Jeong, Y., Leskow, F. C., El-Jaick, K., Roessler, E., Muenke, M., Yocum, A., Dubourg, C., Li, X., Geng, X., Oliver, G., et al. (2008). "Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein." *Nat Genet*, 40(11), 1348-53.

Jin, S. W., Herzog, W., Santoro, M. M., Mitchell, T. S., Frantsve, J., Jungblut, B., Beis, D., Scott, I. C., D'Amico, L. A., Ober, E. A., et al. (2007). "A transgene-assisted genetic screen identifies essential regulators of vascular development in vertebrate embryos." *Dev Biol*, 307(1), 29-42.

Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., and Wolffe, A. P. (1998). "Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription." *Nat Genet*, 19(2), 187-91.

Juven-Gershon, T., Cheng, S., and Kadonaga, J. T. (2006). "Rational design of a super core promoter that enhances gene expression." *Nat Methods*, 3(11), 917-22.

Juven-Gershon, T., Hsu, J. Y., and Kadonaga, J. T. (2008a). "Caudal, a key developmental regulator, is a DPE-specific transcriptional factor." *Genes Dev*, 22(20), 2823-30.

Juven-Gershon, T., Hsu, J. Y., Theisen, J. W., and Kadonaga, J. T. (2008b). "The RNA polymerase II core promoter - the gateway to transcription." *Curr Opin Cell Biol*, 20(3), 253-9.

Juven-Gershon, T., and Kadonaga, J. T. (2009). "Regulation of gene expression via the core promoter and the basal transcriptional machinery." *Dev Biol*.

Kadonaga, J. T. (2004). "Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors." *Cell*, 116(2), 247-57.

Kalmár, É. (2009). "Analysis of the cis-regulatory structure of developmentally regulated genes in zebrafish embryo". Universitätsbibliothek der Universität Heidelberg: Heidelberg. http://archiv.ub.uni-heidelberg.de/volltextserver/volltexte/2009/9443

Kaltenbach, L., Horner, M. A., Rothman, J. H., and Mango, S. E. (2000). "The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during C. elegans embryogenesis." *Mol Cell*, 6(3), 705-13.

Kane, D., and Kimmel, C. (1993). "The zebrafish midblastula transition." *Development*, 119(2), 447-456.

Kao, C. C., Lieberman, P. M., Schmidt, M. C., Zhou, Q., Pei, R., and Berk, A. J. (1990). "Cloning of a transcriptionally active human TATA binding factor." *Science*, 248(4963), 1646-50.

Kari, G., Rodeck, U., and Dicker, A. P. (2007). "Zebrafish: an emerging model system for human disease and drug discovery." *Clin Pharmacol Ther*, 82(1), 70-80.

Kaufmann, J., and Smale, S. T. (1994). "Direct recognition of initiator elements by a component of the transcription factor IID complex." *Genes Dev*, 8(7), 821-9.

Kawaji, H., Frith, M. C., Katayama, S., Sandelin, A., Kai, C., Kawai, J., Carninci, P., and Hayashizaki, Y. (2006). "Dynamic usage of transcription start sites within core promoters." *Genome Biol*, 7(12), R118.

Kawakami, K. (2007). "Tol2: a versatile gene transfer vector in vertebrates." *Genome Biol*, 8 Suppl 1, S7.

Kawakami, K., Shima, A., and Kawakami, N. (2000). "Identification of a functional transposase of the Tol2 element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage." *Proc Natl Acad Sci U S A*, 97(21), 11403-8.

Kawakami, K., Takeda, H., Kawakami, N., Kobayashi, M., Matsuda, N., and Mishina, M. (2004). "A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish." *Dev Cell*, 7(1), 133-44.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, D. (2002). "The human genome browser at UCSC." *Genome Res*, 12(6), 996-1006.

Kikuta, H., Laplante, M., Navratilova, P., Komisarczuk, A. Z., Engstrom, P. G., Fredman, D., Akalin, A., Caccamo, M., Sealy, I., Howe, K., et al. (2007). "Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates." *Genome Res*, 17(5), 545-55.

Kim, T. H., Barrera, L. O., Zheng, M., Qu, C., Singer, M. A., Richmond, T. A., Wu, Y., Green, R. D., and Ren, B. (2005). "A high-resolution map of active promoters in the human genome." *Nature*, 436(7052), 876-80.

Kimelman, D., Kirschner, M., and Scherson, T. (1987). "The events of the midblastula transition in Xenopus are regulated by changes in the cell cycle." *Cell*, 48(3), 399-407.

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., and Schilling, T. F. (1995). "Stages of embryonic development of the zebrafish." *Dev Dyn*, 203(3), 253-310.

Kimura, K., Wakamatsu, A., Suzuki, Y., Ota, T., Nishikawa, T., Yamashita, R., Yamamoto, J., Sekine, M., Tsuritani, K., Wakaguri, H., et al. (2006). "Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes." *Genome Res*, 16(1), 55-65.

Kleinjan, D. A., Seawright, A., Schedl, A., Quinlan, R. A., Danes, S., and van Heyningen, V. (2001). "Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6." *Hum Mol Genet*, 10(19), 2049-59.

Kleinjan, D. A., and van Heyningen, V. (2005). "Long-range control of gene expression: emerging mechanisms and disruption in disease." *Am J Hum Genet*, 76(1), 8-32.

Kondrychyn, I., Garcia-Lecea, M., Emelyanov, A., Parinov, S., and Korzh, V. (2009). "Genome-wide analysis of Tol2 transposon reintegration in zebrafish." *BMC Genomics*, 10, 418.

Kopytova, D. V., Krasnov, A. N., Kopantceva, M. R., Nabirochkina, E. N., Nikolenko, J. V., Maksimenko, O., Kurshakova, M. M., Lebedeva, L. A., Yerokhin, M. M., Simonova, O. B., et al. (2006). "Two isoforms of Drosophila TRF2 are involved in embryonic development, premeiotic chromatin condensation, and proper differentiation of germ cells of both sexes." *Mol Cell Biol*, 26(20), 7492-505.

Koyanagi, K. O., Hagiwara, M., Itoh, T., Gojobori, T., and Imanishi, T. (2005). "Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system." *Gene*, 353(2), 169-76.

Krone, P. H., Blechinger, S. R., Evans, T. G., Ryan, J. A., Noonan, E. J., and Hightower, L. E. (2005). "Use of fish liver PLHC-1 cells and zebrafish embryos in cytotoxicity assays." *Methods*, 35(2), 176-87.

Kutach, A. K., and Kadonaga, J. T. (2000). "The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters." *Mol Cell Biol*, 20(13), 4754-64.

Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D., and Ebright, R. H. (1998). "New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB." *Genes Dev*, 12(1), 34-44.

Laurila, K., and Lähdesmäki, H. (2009). "Systematic analysis of disease-related regulatory mutation classes reveals distinct effects on transcription factor binding." *In Silico Biology*, 9(0018).

Lebedeva, L. A., Nabirochkina, E. N., Kurshakova, M. M., Robert, F., Krasnov, A. N., Evgen'ev, M. B., Kadonaga, J. T., Georgieva, S. G., and Tora, L. (2005). "Occupancy of the Drosophila hsp70 promoter by a subset of basal transcription factors diminishes upon transcriptional activation." *Proc Natl Acad Sci U S A*, 102(50), 18087-92.

Lee, D. H., Gershenzon, N., Gupta, M., Ioshikhes, I. P., Reinberg, D., and Lewis, B. A. (2005). "Functional characterization of core promoter elements: the downstream core element is recognized by TAF1." *Mol Cell Biol*, 25(21), 9674-86.

Lee, T. I., and Young, R. A. (2000). "Transcription of eukaryotic protein-coding genes." *Annu Rev Genet*, 34, 77-137.

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E., and de Graaff, E. (2003). "A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly." *Hum Mol Genet*, 12(14), 1725-35.

Levine, M., and Tjian, R. (2003). "Transcription regulation and animal diversity." *Nature*, 424(6945), 147-51.

Lewis, B. A., Kim, T. K., and Orkin, S. H. (2000). "A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts." *Proc Natl Acad Sci U S A*, 97(13), 7172-7.

Li, X., and Noll, M. (1994). "Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the Drosophila embryo." *EMBO J*, 13(2), 400-6.

Liebel, U., Starkuviene, V., Erfle, H., Simpson, J. C., Poustka, A., Wiemann, S., and Pepperkok, R. (2003). "A microscope-based screening platform for large-scale functional protein analysis in intact cells." *FEBS Lett*, 554(3), 394-8.

Lifton, R. P., Goldberg, M. L., Karp, R. W., and Hogness, D. S. (1978). "The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications." *Cold Spring Harb Symp Quant Biol*, 42 Pt 2, 1047-51.

Lim, C. Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., and Kadonaga, J. T. (2004). "The MTE, a new core promoter element for transcription by RNA polymerase II." *Genes Dev*, 18(13), 1606-17.

Liu, T., Lu, J., Wang, Y., Campbell, W. A., Huang, L., Zhu, J., Xia, W., and Wong, S. T. (2006). "Computerized image analysis for quantitative neuronal phenotyping in zebrafish." *J Neurosci Methods*, 153(2), 190-202.

Lodish, H. F. (2004). *Molecular cell biology*, New York: W. H. Freeman.

Makky, K., Duvnjak, P., Pramanik, K., Ramchandran, R., and Mayer, A. N. (2008). "A whole-animal microplate assay for metabolic rate using zebrafish." *J Biomol Screen*, 13(10), 960-7.

Maldonado, E. (1999). "Transcriptional functions of a new mammalian TATA-binding protein-related factor." *J Biol Chem*, 274(19), 12963-6.

Martianov, I., Fimia, G. M., Dierich, A., Parvinen, M., Sassone-Corsi, P., and Davidson, I. (2001). "Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene." *Mol Cell*, 7(3), 509-15.

Martianov, I., Viville, S., and Davidson, I. (2002). "RNA polymerase II transcription in murine cells lacking the TATA binding protein." *Science*, 298(5595), 1036-9.

Martinez, E., Chiang, C. M., Ge, H., and Roeder, R. G. (1994). "TATA-binding protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter." *EMBO J*, 13(13), 3115-26.

Martinez, N. J., and Walhout, A. J. (2009). "The interplay between transcription factors and microRNAs in genome-scale regulatory networks." *Bioessays*, 31(4), 435-45.

Maston, G. A., Evans, S. K., and Green, M. R. (2006). "Transcriptional regulatory elements in the human genome." *Annu Rev Genomics Hum Genet*, 7, 29-59.

Mathavan, S., Lee, S. G., Mak, A., Miller, L. D., Murthy, K. R., Govindarajan, K. R., Tong, Y., Wu, Y. L., Lam, S. H., Yang, H., et al. (2005). "Transcriptome analysis of zebrafish embryogenesis using microarrays." *PLoS Genet*, 1(2), 260-76.

Matsuda, M., Sakamoto, N., and Fukumaki, Y. (1992). "Delta-thalassemia caused by disruption of the site for an erythroid-specific transcription factor, GATA-1, in the delta-globin gene promoter." *Blood*, 80(5), 1347-51.

Meng, X., Noyes, M. B., Zhu, L. J., Lawson, N. D., and Wolfe, S. A. (2008). "Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases." *Nat Biotechnol*, 26(6), 695-701.

Merli, C., Bergstrom, D. E., Cygan, J. A., and Blackman, R. K. (1996). "Promoter specificity mediates the independent regulation of neighboring genes." *Genes Dev*, 10(10), 1260-70.

Mohan, W. S., Jr., Scheer, E., Wendling, O., Metzger, D., and Tora, L. (2003). "TAF10 (TAF(II)30) is necessary for TFIID stability and early embryogenesis in mice." *Mol Cell Biol*, 23(12), 4307-18.

Moore, P. A., Ozer, J., Salunek, M., Jan, G., Zerby, D., Campbell, S., and Lieberman, P. M. (1999). "A human TATA binding protein-related protein with altered DNA binding specificity inhibits transcription from multiple promoters and activators." *Mol Cell Biol*, 19(11), 7610-20.

Morris, J. R., Chen, J. L., Geyer, P. K., and Wu, C. T. (1998). "Two modes of transvection: enhancer action in trans and bypass of a chromatin insulator in cis." *Proc Natl Acad Sci U S A*, 95(18), 10740-5.

Muller, F., Albert, S., Blader, P., Fischer, N., Hallonet, M., and Strahle, U. (2000). "Direct action of the nodal-related signal cyclops in induction of sonic hedgehog in the ventral midline of the CNS." *Development*, 127(18), 3889-97.

Muller, F., Chang, B., Albert, S., Fischer, N., Tora, L., and Strahle, U. (1999). "Intronic enhancers control expression of zebrafish sonic hedgehog in floor plate and notochord." *Development*, 126(10), 2103-16.

Muller, F., Demeny, M. A., and Tora, L. (2007). "New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors." *J Biol Chem*, 282(20), 14685-9.

Muller, F., Lakatos, L., Dantonel, J., Strahle, U., and Tora, L. (2001). "TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish." *Curr Biol*, 11(4), 282-7.

Muller, F., Lele, Z., Varadi, L., Menczel, L., and Orban, L. (1993). "Efficient transient expression system based on square pulse electroporation and in vivo luciferase assay of fertilized fish eggs." *FEBS Lett*, 324(1), 27-32.

Muller, F., and Tora, L. (2004). "The multicoloured world of promoter recognition complexes." *EMBO J*, 23(1), 2-8.

Muller, F., Williams, D. W., Kobolak, J., Gauvry, L., Goldspink, G., Orban, L., and Maclean, N. (1997). "Activator effect of coinjected enhancers on the muscle-specific expression of promoters in zebrafish embryos." *Mol Reprod Dev*, 47(4), 404-12.

Nagayoshi, S., Hayashi, E., Abe, G., Osato, N., Asakawa, K., Urasaki, A., Horikawa, K., Ikeo, K., Takeda, H., and Kawakami, K. (2008). "Insertional mutagenesis by the Tol2 transposon-mediated enhancer trap approach generated mutations in two developmental genes: tcf7 and synembryn-like." *Development*, 135(1), 159-169.

Nakano, T., Windrem, M., Zappavigna, V., and Goldman, S. A. (2005). "Identification of a conserved 125 base-pair Hb9 enhancer that specifies gene expression to spinal motor neurons." *Dev Biol*, 283(2), 474-85.

Newport, J., and Kirschner, M. (1982a). "A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage." *Cell*, 30(3), 675-86.

Newport, J., and Kirschner, M. (1982b). "A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription." *Cell*, 30(3), 687-96.

Nikolov, D. B., Chen, H., Halay, E. D., Hoffman, A., Roeder, R. G., and Burley, S. K. (1996). "Crystal structure of a human TATA box-binding protein/TATA element complex." *Proc Natl Acad Sci U S A*, 93(10), 4862-7.

Nikolov, D. B., Hu, S. H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N. H., Roeder, R. G., and Burley, S. K. (1992). "Crystal structure of TFIID TATA-box binding protein." *Nature*, 360(6399), 40-6.

O'Boyle, S., Bree, R. T., McLoughlin, S., Grealy, M., and Byrnes, L. (2007). "Identification of zygotic genes expressed at the midblastula transition in zebrafish." *Biochem Biophys Res Commun*, 358(2), 462-8.

O'Farrell, P. H., Stumpff, J., and Su, T. T. (2004). "Embryonic cleavage cycles: how is a mouse like a fly?" *Curr Biol*, 14(1), R35-45.

Ogbourne, S., and Antalis, T. M. (1998). "Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes." *Biochem J*, 331 (Pt 1), 1-14.

Ohler, U. (2006). "Identification of core promoter modules in Drosophila and their application in accurate transcription start site prediction." *Nucleic Acids Res*, 34(20), 5943-50.

Ohler, U., Liao, G. C., Niemann, H., and Rubin, G. M. (2002). "Computational analysis of core promoters in the Drosophila genome." *Genome Biol*, 3(12), RESEARCH0087.

Ohtsuki, S., Levine, M., and Cai, H. N. (1998). "Different core promoters possess distinct regulatory activities in the Drosophila embryo." *Genes Dev*, 12(4), 547-56.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). "The general transcription factors of RNA polymerase II." *Genes Dev*, 10(21), 2657-83.

Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J. A., Lopes, S., Reik, W., et al. (2004). "Active genes dynamically colocalize to shared sites of ongoing transcription." *Nat Genet*, 36(10), 1065-71.

Parinov, S., Kondrichin, I., Korzh, V., and Emelyanov, A. (2004). "Tol2 transposon-mediated enhancer trap to identify developmentally regulated zebrafish genes in vivo." *Dev Dyn*, 231(2), 449-59.

Patikoglou, G. A., Kim, J. L., Sun, L., Yang, S. H., Kodadek, T., and Burley, S. K. (1999). "TATA element recognition by the TATA box-binding protein has been conserved throughout evolution." *Genes Dev*, 13(24), 3217-30.

Pepperkok, R., and Ellenberg, J. (2006). "High-throughput fluorescence microscopy for systems biology." *Nat Rev Mol Cell Biol*, 7(9), 690-6.

Persengiev, S. P., Zhu, X., Dixit, B. L., Maston, G. A., Kittler, E. L., and Green, M. R. (2003). "TRF3, a TATA-box-binding protein-related factor, is vertebrate-specific and widely expressed." *Proc Natl Acad Sci U S A*, 100(25), 14887-91.

Peterson, R. T., Shaw, S. Y., Peterson, T. A., Milan, D. J., Zhong, T. P., Schreiber, S. L., MacRae, C. A., and Fishman, M. C. (2004). "Chemical suppression of a genetic mutation in a zebrafish model of aortic coarctation." *Nat Biotechnol*, 22(5), 595-9.

Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E., et al. (2005). "Genome-wide map of nucleosome acetylation and methylation in yeast." *Cell*, 122(4), 517-27.

Ponjavic, J., Lenhard, B., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., and Sandelin, A. (2006). "Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters." *Genome Biol*, 7(8), R78.

Portales-Casamar, E., Thongjuea, S., Kwon, A. T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W. W., and Sandelin, A. (2009). "JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles." *Nucleic Acids Res*.

Prioleau, M. N., Huet, J., Sentenac, A., and Mechali, M. (1994). "Competition between chromatin and transcription complex assembly regulates gene expression during early development." *Cell*, 77(3), 439-49.

Rabenstein, M. D., Zhou, S., Lis, J. T., and Tjian, R. (1999). "TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family." *Proc Natl Acad Sci U S A*, 96(9), 4791-6.

Rebagliati, M. R., Toyama, R., Fricke, C., Haffter, P., and Dawid, I. B. (1998). "Zebrafish nodal-related genes are implicated in axial patterning and establishing left-right asymmetry." *Dev Biol*, 199(2), 261-72.

Reina, J. H., and Hernandez, N. (2007). "On a roll for new TRF targets." *Genes Dev*, 21(22), 2855-60.

Remenyi, A., Scholer, H. R., and Wilmanns, M. (2004). "Combinatorial control of gene expression." *Nat Struct Mol Biol*, 11(9), 812-5.

Rhead, B., Karolchik, D., Kuhn, R. M., Hinrichs, A. S., Zweig, A. S., Fujita, P. A., Diekhans, M., Smith, K. E., Rosenbloom, K. R., Raney, B. J., et al. (2009). "The UCSC genome browser database: update 2010." *Nucleic Acids Res*.

Rihel, J., Prober, D. A., Arvanites, A., Lam, K., Zimmerman, S., Jang, S., Haggarty, S. J., Kokel, D., Rubin, L. L., Peterson, R. T., et al. (2010). "Zebrafish behavioral profiling links drugs to biological targets and rest/wake regulation." *Science*, 327(5963), 348-51.

Robu, M. E., Larson, J. D., Nasevicius, A., Beiraghi, S., Brenner, C., Farber, S. A., and Ekker, S. C. (2007). "p53 activation by knockdown technologies." *PLoS Genet*, 3(5), e78.

Roeder, R. G. (1996). "The role of general initiation factors in transcription by RNA polymerase II." *Trends in Biochemical Sciences*, 21(9), 327-35.

Roure, A., Rothbacher, U., Robin, F., Kalmar, E., Ferone, G., Lamy, C., Missero, C., Mueller, F., and Lemaire, P. (2007). "A multicassette Gateway vector set for high throughput and comparative analyses in ciona and vertebrate embryos." *PLoS One*, 2(9), e916.

Sambrook, J., and Russell, D. W. (2001). *Molecular cloning : a laboratory manual*, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D. A. (2007). "Mammalian RNA polymerase II core promoters: insights from genome-wide studies." *Nat Rev Genet*, 8(6), 424-36.

Sanges, R., Kalmar, E., Claudiani, P., D'Amato, M., Muller, F., and Stupka, E. (2006). "Shuffling of cis-regulatory elements is a pervasive feature of the vertebrate lineage." *Genome Biol*, 7(7), R56.

Saxonov, S., Berg, P., and Brutlag, D. L. (2006). "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters." *Proc Natl Acad Sci U S A*, 103(5), 1412-7.

Schier, A. F. (2007). "The maternal-zygotic transition: death and birth of RNAs." *Science*, 316(5823), 406-7.

Schmidt, E. E., Bondareva, A. A., Radke, J. R., and Capecchi, M. R. (2003). "Fundamental cellular processes do not require vertebrate-specific sequences within the TATA-binding protein." *J Biol Chem*, 278(8), 6168-74.

Schubeler, D., MacAlpine, D. M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D. E., O'Neill, L. P., Turner, B. M., Delrow, J., et al. (2004). "The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote." *Genes Dev*, 18(11), 1263-71.

Semotok, J. L., Cooperstock, R. L., Pinder, B. D., Vari, H. K., Lipshitz, H. D., and Smibert, C. A. (2005). "Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early Drosophila embryo." *Curr Biol*, 15(4), 284-94.

Semotok, J. L., Luo, H., Cooperstock, R. L., Karaiskakis, A., Vari, H. K., Smibert, C. A., and Lipshitz, H. D. (2008). "Drosophila maternal Hsp83 mRNA destabilization is directed by multiple SMAUG recognition elements in the open reading frame." *Mol Cell Biol*, 28(22), 6757-72.

Shermoen, A. W., and O'Farrell, P. H. (1991). "Progression of the cell cycle through mitosis leads to abortion of nascent transcripts." *Cell*, 67(2), 303-10.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., et al. (2003). "Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage." *Proc Natl Acad Sci U S A*, 100(26), 15776-81.

Shkumatava, A., Fischer, S., Muller, F., Strahle, U., and Neumann, C. J. (2004). "Sonic hedgehog, secreted by amacrine cells, acts as a short-range signal to direct differentiation and lamination in the zebrafish retina." *Development*, 131(16), 3849-58.

Smale, S. T. (2001). "Core promoters: active contributors to combinatorial gene regulation." *Genes Dev*, 15(19), 2503-8.

Smale, S. T., and Baltimore, D. (1989). "The "initiator" as a transcription control element." *Cell*, 57(1), 103-13.

Smale, S. T., and Kadonaga, J. T. (2003). "The RNA polymerase II core promoter." *Annu Rev Biochem*, 72, 449-79.

Song, D. L., Chalepakis, G., Gruss, P., and Joyner, A. L. (1996). "Two Pax-binding sites are required for early embryonic brain expression of an Engrailed-2 transgene." *Development*, 122(2), 627-35.

Stuart, G. W., McMurray, J. V., and Westerfield, M. (1988). "Replication, integration and stable germ-line transmission of foreign sequences injected into early zebrafish embryos." *Development*, 103(2), 403-12.

Stuart, G. W., Vielkind, J. R., McMurray, J. V., and Westerfield, M. (1990). "Stable lines of transgenic zebrafish exhibit reproducible patterns of transgene expression." *Development*, 109(3), 577-84.

Sun, M., Ma, F., Zeng, X., Liu, Q., Zhao, X. L., Wu, F. X., Wu, G. P., Zhang, Z. F., Gu, B., Zhao, Y. F., et al. (2008). "Triphalangeal thumb-polysyndactyly syndrome and syndactyly type IV are caused by genomic duplications involving the long range, limb-specific SHH enhancer." *J Med Genet*, 45(9), 589-95.

Suzuki, Y., Taira, H., Tsunoda, T., Mizushima-Sugano, J., Sese, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Morishita, S., et al. (2001a). "Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites." *EMBO Rep*, 2(5), 388-93.

Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., et al. (2001b). "Identification and characterization of the

potential promoter regions of 1031 kinds of human genes." *Genome Res*, 11(5), 677-84.

Suzuki, Y., Yamashita, R., Shirota, M., Sakakibara, Y., Chiba, J., Mizushima-Sugano, J., Kel, A. E., Arakawa, T., Carninci, P., Kawai, J., et al. (2004). "Large-scale collection and characterization of promoters of human and mouse genes." *In Silico Biol*, 4(4), 429-44.

Szabo, M., Muller, F., Kiss, J., Balduf, C., Strahle, U., and Olasz, F. (2003). "Transposition and targeting of the prokaryotic mobile element IS30 in zebrafish." *FEBS Lett*, 550(1-3), 46-50.

Szutorisz, H., Dillon, N., and Tora, L. (2005). "The role of enhancers as centres for general transcription factor recruitment." *Trends in Biochemical Sciences*, 30(11), 593-9.

Taatjes, D. J., Marr, M. T., and Tjian, R. (2004). "Regulatory diversity among metazoan co-activator complexes." *Nat Rev Mol Cell Biol*, 5(5), 403-10.

Tadros, W., and Lipshitz, H. D. (2009). "The maternal-to-zygotic transition: a play in two acts." *Development*, 136(18), 3033-42.

Takada, S., Lis, J. T., Zhou, S., and Tjian, R. (2000). "A TRF1:BRF complex directs Drosophila RNA polymerase III transcription." *Cell*, 101(5), 459-69.

Takai, D., and Jones, P. A. (2002). "Comprehensive analysis of CpG islands in human chromosomes 21 and 22." *Proc Natl Acad Sci U S A*, 99(6), 3740-5.

Teichmann, M., Wang, Z., Martinez, E., Tjernberg, A., Zhang, D., Vollmer, F., Chait, B. T., and Roeder, R. G. (1999). "Human TATA-binding protein-related factor-2 (hTRF2) stably associates with hTFIIA in HeLa cells." *Proc Natl Acad Sci U S A*, 96(24), 13720-5.

Thermes, V., Grabher, C., Ristoratore, F., Bourrat, F., Choulika, A., Wittbrodt, J., and Joly, J. S. (2002). "I-SceI meganuclease mediates highly efficient transgenesis in fish." *Mech Dev*, 118(1-2), 91-8.

Theuns, J., Brouwers, N., Engelborghs, S., Sleegers, K., Bogaerts, V., Corsmit, E., De Pooter, T., van Duijn, C. M., De Deyn, P. P., and Van Broeckhoven, C. (2006). "Promoter mutations that increase amyloid precursor-protein expression are associated with Alzheimer disease." *Am J Hum Genet*, 78(6), 936-46.

Thisse, B., Pflumio, S., Fürthauer, M., Loppin, B., Heyer, V., Degrave, A., Woehl, R., Lux, A., Steffan, T., Charbonnier, X.Q. and Thisse, C. (2001). "Expression of the zebrafish genome during embryogenesis (NIH R01 RR15402)." *ZFIN Direct Data Submission (http://zfin.org)*. .

Thisse, B., Thisse, C. (2004). "Fast Release Clones: A High Throughput Expression Analysis." *ZFIN Direct Data Submission (http://zfin.org)*. .

Thomas, M. C., and Chiang, C. M. (2006). "The general transcription machinery and general cofactors." *Crit Rev Biochem Mol Biol*, 41(3), 105-78.

Tokusumi, Y., Ma, Y., Song, X., Jacobson, R. H., and 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." *Mol Cell Biol*, 27(5), 1844-58.

Tora, L. (2002). "A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription." *Genes Dev*, 16(6), 673-5.

Tran, T. C., Sneed, B., Haider, J., Blavo, D., White, A., Aiyejorun, T., Baranowski, T. C., Rubinstein, A. L., Doan, T. N., Dingledine, R., et al. (2007). "Automated, quantitative screening assay for antiangiogenic compounds using transgenic zebrafish." *Cancer Res*, 67(23), 11386-92.

Trinklein, N. D., Aldred, S. F., Hartman, S. J., Schroeder, D. I., Otillar, R. P., and Myers, R. M. (2004). "An abundance of bidirectional promoters in the human genome." *Genome Res*, 14(1), 62-6.

Trinklein, N. D., Aldred, S. J., Saldanha, A. J., and Myers, R. M. (2003). "Identification and functional analysis of human transcriptional promoters." *Genome Res*, 13(2), 308-12.

Tupler, R., Perini, G., and Green, M. R. (2001). "Expressing the human genome." *Nature*, 409(6822), 832-3.

Uemura, O., Okada, Y., Ando, H., Guedj, M., Higashijima, S., Shimazaki, T., Chino, N., Okano, H., and Okamoto, H. (2005). "Comparative functional genomics revealed conservation and diversification of three enhancers of the isl1 gene for motor and sensory neuron-specific expression." *Dev Biol*, 278(2), 587-606.

Valen, E., Pascarella, G., Chalk, A., Maeda, N., Kojima, M., Kawazu, C., Murata, M., Nishiyori, H., Lazarevic, D., Motti, D., et al. (2009). "Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE." *Genome Res*, 19(2), 255-65.

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., and Luscombe, N. M. (2009). "A census of human transcription factors: function, expression and evolution." *Nat Rev Genet*, 10(4), 252-63.

Veenstra, G. J., Destree, O. H., and Wolffe, A. P. (1999). "Translation of maternal TATA-binding protein mRNA potentiates basal but not activated transcription in Xenopus embryos at the midblastula transition." *Mol Cell Biol*, 19(12), 7972-82.

Veenstra, G. J., Weeks, D. L., and Wolffe, A. P. (2000). "Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in Xenopus." *Science*, 290(5500), 2312-5.

Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W., van Schaik, F. M., Varier, R. A., Baltissen, M. P., Stunnenberg, H. G., Mann, M., and Timmers, H. T. (2007). "Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4." *Cell*, 131(1), 58-69.

Visel, A., Rubin, E. M., and Pennacchio, L. A. (2009). "Genomic views of distant-acting enhancers." *Nature*, 461(7261), 199-205.

Vogt, A., Cholewinski, A., Shen, X., Nelson, S. G., Lazo, J. S., Tsang, M., and Hukriede, N. A. (2009). "Automated image-based phenotypic analysis in zebrafish embryos." *Dev Dyn*, 238(3), 656-63.

Wakaguri, H., Yamashita, R., Suzuki, Y., Sugano, S., and Nakai, K. (2008). "DBTSS: database of transcription start sites, progress report 2008." *Nucleic Acids Res*, 36(Database issue), D97-101.

Wang, W., Liu, X., Gelinas, D., Ciruna, B., and Sun, Y. (2007). "A fully automated robotic system for microinjection of zebrafish embryos." *PLoS One*, 2(9), e862.

Wang, Y., and Leung, F. C. (2004). "An evaluation of new criteria for CpG islands in the human genome as gene markers." *Bioinformatics*, 20(7), 1170-7.

Wardle, F. C., Odom, D. T., Bell, G. W., Yuan, B., Danford, T. W., Wiellette, E. L., Herbolsheimer, E., Sive, H. L., Young, R. A., and Smith, J. C. (2006). "Zebrafish promoter microarrays identify actively transcribed embryonic genes." *Genome Biol*, 7(8), R71.

West, A. G., and Fraser, P. (2005). "Remote control of gene transcription." *Hum Mol Genet*, 14 Spec No 1, R101-11.

West, A. G., Gaszner, M., and Felsenfeld, G. (2002). "Insulators: many functions, many mechanisms." *Genes Dev*, 16(3), 271-88.

Westerfield, M. (2000). "The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio)." *University of Oregon Press*(4th).

Westerfield, M., Wegner, J., Jegalian, B. G., DeRobertis, E. M., and Puschel, A. W. (1992). "Specific activation of mammalian Hox promoters in mosaic transgenic zebrafish." *Genes Dev*, 6(4), 591-8.

Wieczorek, E., Brand, M., Jacq, X., and Tora, L. (1998). "Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II." *Nature*, 393(6681), 187-91.

Wienholds, E., Kloosterman, W. P., Miska, E., Alvarez-Saavedra, E., Berezikov, E., de Bruijn, E., Horvitz, H. R., Kauppinen, S., and Plasterk, R. H. (2005). "MicroRNA expression in zebrafish embryonic development." *Science*, 309(5732), 310-1.

Wienholds, E., van Eeden, F., Kosters, M., Mudde, J., Plasterk, R. H., and Cuppen, E. (2003). "Efficient target-selected mutagenesis in zebrafish." *Genome Res*, 13(12), 2700-7.

Willy, P. J., Kobayashi, R., and Kadonaga, J. T. (2000). "A basal transcription factor that activates or represses transcription." *Science*, 290(5493), 982-5.

Wilson, D., Charoensawan, V., Kummerfeld, S. K., and Teichmann, S. A. (2008). "DBD--taxonomically broad transcription factor predictions: new content and functionality." *Nucleic Acids Res*, 36(Database issue), D88-92.

Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., Smith, S. F., North, P., Callaway, H., Kelly, K., et al. (2005). "Highly conserved non-coding sequences are associated with vertebrate development." *PLoS Biol*, 3(1), e7.

Wright, K. J., Marr, M. T., 2nd, and Tjian, R. (2006). "TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter." *Proc Natl Acad Sci U S A*, 103(33), 12347-52.

Xiao, L., Kim, M., and DeJong, J. (2006). "Developmental and cell type-specific regulation of core promoter transcription factors in germ cells of frogs and mice." *Gene Expr Patterns*, 6(4), 409-19.

Yang, C., Bolotin, E., Jiang, T., Sladek, F. M., and Martinez, E. (2007). "Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters." *Gene*, 389(1), 52-65.

Yang, L., Ho, N. Y., Alshut, R., Legradi, J., Weiss, C., Reischl, M., Mikut, R., Liebel, U., Muller, F., and Strahle, U. (2009). "Zebrafish embryos as models for embryotoxic and teratological effects of chemicals." *Reprod Toxicol*, 28(2), 245-53.

Yang, Z., Jiang, H., Chachainasakul, T., Gong, S., Yang, X. W., Heintz, N., and Lin, S. (2006). "Modified bacterial artificial chromosomes for zebrafish transgenesis." *Methods*, 39(3), 183-8.

Yelon, D., Horne, S. A., and Stainier, D. Y. (1999). "Restricted expression of cardiac myosin genes reveals regulated aspects of heart tube assembly in zebrafish." *Dev Biol*, 214(1), 23-37.

Zerucha, T., Stuhmer, T., Hatch, G., Park, B. K., Long, Q., Yu, G., Gambarotta, A., Schultz, J. R., Rubenstein, J. L., and Ekker, M. (2000). "A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain." *J Neurosci*, 20(2), 709-21.

Zhang, D., Penttila, T. L., Morris, P. L., Teichmann, M., and Roeder, R. G. (2001). "Spermiogenesis deficiency in mice lacking the Trf2 gene." *Science*, 292(5519), 1153-5.

Zhong, T. P., Rosenberg, M., Mohideen, M. A., Weinstein, B., and Fishman, M. C. (2000). "gridlock, an HLH gene required for assembly of the aorta in zebrafish." *Science*, 287(5459), 1820-4.

Zhou, J., and Levine, M. (1999). "A novel cis-regulatory element, the PTS, mediates an anti-insulator activity in the Drosophila embryo." *Cell*, 99(6), 567-75.

Zhu, X., Ling, J., Zhang, L., Pi, W., Wu, M., and Tuan, D. (2007). "A facilitated tracking and transcription mechanism of long-range enhancer function." *Nucleic Acids Res*, 35(16), 5532-44.

Zon, L. I., and Peterson, R. T. (2005). "In vivo drug discovery in the zebrafish." *Nat Rev Drug Discov*, 4(1), 35-44.

9. PUBLICATIONS

Ferg, M.*, Sanges, R.*, **Gehrig, J**., Kiss, J., Bauer, M., Lovas, A., Szabo, M., Yang, L., Strähle, U., Pankratz, M.J., Olasz, F., Stupka, E., and Müller, F. (2007). "The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish." *EMBO Journal* 26(17): 3945-3956. (*contributed equally)

Gehrig, J.*, Reischl, M.*, Kalmar, E.*, Ferg, M., Hadzhiev, Y., Zaucker, A., Song, C., Schindler, S., Liebel, U. and Müller, F. (2009). "Automated high throughput mapping of promoter-enhancer interactions in zebrafish embryos." *Nature Methods* 6(12): 911-916. (*contributed equally)