Gene transcription in the zebrafish embryo: regulators and networks

Marco Ferg, Olivier Armant, Lixin Yang, Thomas Dickmeis, Sepand Rastegar and Uwe Strähle
Advance Access publication date 22 October 2013

Abstract
The precise spatial and temporal control of gene expression is a key process in the development, maintenance and regeneration of the vertebrate body. A substantial proportion of vertebrate genomes encode genes that control the transcription of the genetic information into mRNA. The zebrafish is particularly well suited to investigate gene regulatory networks underlying the control of gene expression during development due to the external development of its transparent embryos and the increasingly sophisticated tools for genetic manipulation available for this model system. We review here recent data on the analysis of cis-regulatory modules, transcriptional regulators and their integration into gene regulatory networks in the zebrafish, using the developing spinal cord as example.

Keywords: zebrafish; transcription regulator; gene regulatory network; cis-regulatory element; genomics

INTRODUCTION
The first step of gene expression is transcription. Differential gene transcription plays a crucial role in the development of the vertebrate body plan. Also, many homeostatic processes and responses in the adult organism are regulated by transcription. Thus, misexpression of genes due to deregulated transcription can lead to a variety of diseases including cancer and type II diabetes [1–3]. The importance of differential transcriptional regulation as a means to correctly express the information encoded in the genome is further underscored by the fact that a substantial proportion of genes in the genomes code for proteins that are involved in the control of transcription, such as transcription factors (TFs), chromatin-remodelling factors and proteins of the basic transcription machinery [4, 5]. Complex gene regulatory networks (GRNs) form the basis of cell differentiation during development of multicellular organisms and can control entire batteries of genes coordinately. In GRNs, specific combinations of transcriptional regulators (TRs) interact with cis-regulatory modules (CRMs) of responsive genes [6–8].

CRMs can be functionally subdivided into two groups: activating modules (promoter, enhancer, locus control region [LCR]) and repressive modules (silencer, insulator). Silencers inhibit gene expression through the interaction with repressor proteins that in turn initiate heterochromatin formation by recruiting histone-modifying enzyme complexes. Insulators confine the activity of an enhancer or...
silencers to its cognate gene promoter, and interaction of CCCTC-binding factor (CTCF) with insulators has been demonstrated to be crucial to shield distal enhancers/silencers [9]. This classical view of CTCF being a major marker for insulator elements has been questioned by 4C experiments showing that many of the CTCF-mediated long range interacting sites correlate with enhancer–promoter interactions. Further experiments showed that CTCF facilitates enhancer–promoter interactions directly through cohesin-mediated DNA looping [10–12].

Enhancers have been traditionally defined as a class of cis-regulatory elements that are able to activate and maintain gene expression in a location- and orientation-independent manner [13, 14]. They are found in the up- and downstream regions as well as in introns of the regulated gene. However, enhancers can also overlap with coding exons [15] or the 5′-untranslated region (UTR) [16]. They may be located as far as 1 megabase from the target gene, as has been shown, for example, for a limb-specific enhancer of the sonic hedgehog gene [17]. Enhancers are thought to represent islands of open chromatin that act as docking stations for TFs [18]. By a combination of distinct TFs that bind to the same enhancer, various signalling inputs can be integrated and interpreted, thereby leading to specific gene expression outputs. In the prevailing model, the physical interaction of enhancers with their target promoters is facilitated by a decondensation of the chromatin structure enabling the DNA to loop out from the nuclear periphery to the centre to form enhancer/promoter interactions [19–21]. A recent publication suggests CTCF/cohesin-mediated DNA looping is a determinant for promoter choice [22]. However, not all loci require looping-out for transcriptional activity [23].

Promoters can be functionally subdivided into the proximal promoter and the core promoter. Proximal promoters are located in the immediate vicinity of the transcription start site (TSS) and contain, like enhancers or silencers, recognition sites for DNA-binding factors that are involved in transcriptional regulation. The core promoter is usually defined as a 100–150 bp stretch of DNA encompassing the TSS that is sufficient to direct initiation of transcription by the RNA polymerase II machinery [24]. Core promoters can differ considerably in motif composition (reviewed in [25–27]) and can be bound by a multitude of promoter recognition complexes [28, 29]. Not all enhancers are able to interact with every type of core promoter [18, 30, 31], providing another level of regulatory control.

LCRs are, like enhancers, activating CRMs. They possess chromatin domain-opening activity and are able to confer ‘copy number-dependent’ expression on a linked gene, i.e. the number of integrated transgenes determines the expression level and not the influence of the chromatin surrounding the transgenes (reviewed e.g. in [32, 33]).

Interestingly, enhancers and silencers share important properties, such as DNase I hypersensitivity, active chromatin marks and interaction with transcription factors and RNA polymerase II. It has, therefore, been proposed that it depends on the gene target and the developmental stage whether a given CRM activates or represses gene transcription [34].

CRMs contain the information required for precise transcriptional regulation. The implementation of this information is carried out by TRs through sequence-specific interactions with CRMs of their target genes. As described above, TFs activate or repress gene transcription by binding to enhancers or silencers, respectively. In addition, general TFs are involved in the basic processes of transcription, such as initiation, elongation and termination of transcription. However, they may be less ‘general’ than previously thought, as recent evidence suggests gene-specific employment of these factors [35, 36]. Finally, chromatin remodelling proteins alter the state of the chromatin and thereby affect transcription by regulating access to the CRMs.

In summary, GRNs can be defined as the implementation of the regulatory information encoded in the DNA of CRMs via their interactions with TRs. Modelling a GRN related to any development process requires knowledge of which TFs and signalling molecules are involved, when and where the genes are expressed and how these factors interact with each other and with CRMs [8]. The zebrafish is an ideal model to study the components and composition of GRNs. Its fast and external development and the high number of transparent embryos obtainable have proven very useful to address questions in transcriptional regulation. Zebrafish and other fish species have been utilized to analyse developmental GRNs [37, 38] and promoter/enhancer interactions [31], to validate and analyse enhancer function on a large scale [39–44], to address the regulatory potential of short sequences [45] and to identify regulatory regions driving tissue-specific expression.
in enhancer trap assays [46, 47]. Furthermore, recent progress in defining gene expression profiles on a global scale [4] demonstrates that zebrafish is a highly useful organism to comprehensively study GRNs in a vertebrate organism.

In this review, we will illustrate how zebrafish can be used to understand the regulatory logic of the vertebrate genome. We will summarize the current knowledge on specific aspects of CRMs regarding their distribution and the genomic features that have been associated with CRMs, examine the repertoire of TRs in the zebrafish genome and discuss one example of a transcriptional regulatory network that is well analysed in both zebrafish and mice.

**SEQUENCE CONSERVATION AND HISTONE MODIFICATIONS IDENTIFY CRMS**

CRMs are frequently characterized by conservation of sequence across species. DNA sequences with regulatory functions exhibit a lower mutation rate than non-functional DNA and therefore, like protein coding sequences, show significant sequence similarity across species. The *Fugu rubripes* genome was the first vertebrate genome to be assembled [48] after sequencing of the human genome [49, 50]. Comparative studies of the human and Fugu genomes established the presence of a large number of highly conserved non-coding elements (HCNEs) [51–54]. Since then, the growing number of whole-genome sequences from various organisms and the development of bioinformatic tools [55–59] have made comparative genomics a powerful means to identify putative CRMs [60–63].

The binding of TRs to CRMs provides one mechanism of transcriptional regulation. Another level is introduced by the structure of the chromatin, which can control the accessibility of the CRMs. The fact that active regulatory CRMs are in an open chromatin state forms the basis of large-scale mapping of CRMs by DNAse I hypersensitivity assays [64]. DNAse I hypersensitivity provides a comprehensive picture of the active regulatory landscape of the genome in a particular cell, because it is a key feature of all classes of CRMs [65, 66]. Specific post-translational modifications of histones such as methylation and acetylation at particular positions are further landmarks of active and inactive chromatin structure of CRMs. For example, active enhancers have been shown to be marked by monomethylation of H3K4 (H3K4me1) [67–69]. However, this chromatin modification is not limited only to active enhancers. Several studies suggested the existence of both active and poised classes of enhancer elements. Active enhancers in a cell are engaged in the regulation of transcription, whereas poised enhancers are not actively regulating transcription, but can be rapidly converted to active enhancers, e.g. by external stimuli to the cell or in the course of differentiation [70]. The histone marks associated with poised enhancers contain both the H3K4me1 mark and trimethylation at K27 of histone H3 (H3K27me3). Active enhancers are distinguished from poised enhancers by the presence of acetylation of histone H3 at lysine 27 (H3K27ac) and by the absence of H3K27me3 [70–72]. Besides these post-translational modifications of histones, binding of chromatin remodelling enzymes has been shown to designate active enhancer modules. Binding of the acetyltransferase and transcriptional co-activator p300, for example, predicted with high accuracy where active enhancers are located in the genome in neural, limb and heart tissue [69, 73, 74]. Furthermore, the chromatin remodelling enzymes CHD7 and BRG1 appear to be present at active enhancers [71, 72].

**CRMS IN THE ZEBRAFISH GENOME**

Most genome-wide studies of the regulatory landscape have focused on mammalian cell line, while the dynamic changes of the histone code and TF occupancy of CRMs on a global scale during ontogenesis of a vertebrate are still unresolved. The zebrafish with its experimentally easily accessible early stages of development has a great potential to address these questions. ChIP (chromatin immunoprecipitation) on chip experiments confirmed enrichment of the histone mark H3K4me3 found in active promoters [75] also near the TSS in zebrafish [76]. Interestingly, many inactive developmental regulatory genes are marked by the repressive mark H3K27me3, while at the same time containing the activating chromatin modification H3K4me3, thus showing the characteristic signature of poised genes that is also found in developmental regulators in mouse embryonic stem cells [77, 78]. This ‘bivalent’ signature was not detected before midblastula transition, when the activation of the zygotic genome occurs [79]. It was suggested that such marks might
secure precise timing of the activation of developmental regulators [80]. Combining H3K4me1 and H3K4me3 profiles, signatures of active enhancers (H3K4me1) and active promoters (H3K4me3) in combination with the readily available expression data in zebrafish helped to computationally identify putative cis-regulatory sequences [39]. The generation of genomic tracks of H3K4me3, H3K4me1 and H3K27ac histone modifications at four developmental time-points of zebrafish embryogenesis lead to the identification of 50,000 potential cis-regulatory elements operating during the first 48 h of zebrafish development, a wealth of data that will help to deduce regulatory networks on the DNA sequence level [81].

Although many data are now available on the location of putative CRMs involved in embryonic development, a systematic analysis of the function of CRMs has not yet been carried out on a large scale in zebrafish. One crucial question is how conserved the functions of the CRMs encoded in HCNEs are. Case studies illustrated several possibilities. For example, the lateral stripe enhancer of the neurogenin1 gene retained its function for about 395 million years of evolution, as it drives expression in the dorsal telencephalon of both mouse and zebrafish [82]. In contrast, the function of CRMs in the shh gene that are conserved between mouse and zebrafish seems to have changed: the zebrafish sonic hedgehog (shh) enhancer ar-C directs mainly notochord expression in zebrafish embryos and also functions in the midline of mouse embryos. However, the mouse enhancer SFPE2, which exhibits sequence similarity with zebrafish ar-C, is floor plate-specific in the mouse and not functional in zebrafish.

Another important question is why CRMs of developmental regulators show higher sequence conservation than CRMs of other genes; why do not all conserved sequences have an obvious function? Why are not all functional CRMs highly conserved? Moreover, it is more than questionable whether up to 90% conservation is necessary to maintain the binding of the same TFs, given the degeneracy of the recognition sequences of many TFs. In fact, functional dissection of five notochord enhancers with evolutionarily conserved sequences revealed two short DNA elements mediating notochord expression. However, these elements were embedded within the conserved sequences of the five enhancers in an almost random fashion, suggesting that the conservation of the sequence is only indirectly linked to binding of TFs [83]. The findings that non-coding RNAs bi-directionally transcribed on enhancers (eRNAs) [84] significantly increase promoter–enhancer interactions [85] and that activating ncRNAs (ncRNA-a) play an important role in establishing and maintaining chromatin structure [86, 87] opened up a new chapter in cis-regulation of gene expression and might help to answer some of these questions.

Taken together, analysis of CRMs of the zebrafish genome now provides a framework to understand precisely how classes of CRMs are composed, how they interact with epigenetic marks and why some of them are conserved across large evolutionary distances. More functional analyses will be required, and one important element of these is knowledge about which transcriptional regulators interact with the CRMs.

**CHARACTERIZATION OF THE TRANSCRIPTIONAL REGULATORY REPertoire OF THE ZEBRAFISH GENOME**

The availability of the entire zebrafish genome sequence [88] and the elucidation of its gene structures by computational means to identify or predict TSSs, exon/intron structure, UTRs, protein domains and regulatory elements, provides a rich resource to mine genomes for genes involved in transcriptional regulation. The InterPro database provides a resource of functional annotation on proteins by classifying them into families. It compiles protein signatures from a number of databases, thereby integrating all data into a unique searchable resource [89]. A thorough search of the InterPro protein domains led to the selection of 483 domains associated with transcriptional regulation activities (Figure 1A). Mining the zebrafish genome for genes encoding at least one protein domain linked to transcriptional regulation revealed 3302 genes, including TFs with specific DNA binding domains, genes involved in chromatin modification and factors of the general transcriptional machinery (Figure 1B). Of these, 2488 TR genes can be detected reliably by deep sequencing (349 million 76 bp long paired-end reads) of mRNA isolated from 24 h post-fertilization (hpf) embryos. This suggests that nearly 75% of the TR gene repertoire of the zebrafish is expressed at a significant level at 24 hpf [4], a stage of particular
importance as it represents the evolutionarily con-
served phylotypic stage of this model organism [90]. The remaining TR genes may be expressed at
very low levels, below the detection limit, or at later
stages or only under certain physiological conditions.

Of the 3302 TR genes detected in the zebrafish
genome, 2677 genes can be assigned to the TF sub-
class of TRs. Precedent analysis of human and mouse
genomes suggested that 1500–2000 genomic loci
encode DNA-binding TFs [5], roughly 10 times

Figure 1: Zebrafish transcriptional regulators. (A) InterPro protein domains specific to transcriptional regulation. The number of protein domains specific for each category of TR (transcription factor, chromatin remodelling, basal transcriptional machinery) is indicated (InterPro release 19). (B) The number of genomic loci encoding genes with at least one domain functionally linked to transcriptional control are indicated for each category (Zv9 assembly, Ensembl version 60 annotation). (C) Expression patterns of chromatin remodellers of the BTB/POZ family in 24 hpf embryos. *bach2* and *zbtb46* are expressed in the telencephalon, the *zbtb16* homologue is expressed in the whole spinal cord and the posterior tuberculum in the diencephalon, *zbtb43* is expressed in the retina and the tectum and *btbd6a* is expressed in the epiphysis, the telencephalon, the hindbrain and the spinal cord, while its para-
ologue *btbd6b* is expressed in the somites, the telencephalon and other parts of the fore- and midbrain. *hic1* and *bcl6b* are expressed in the vascular system of the trunk and the head, whereas the *bcl6b* homologue is expressed only in the vascular system of the head.
the size of the TF repertoire of the yeast Saccharomyces cerevisiae (200–350 loci) [91]. However, the fraction of the total number of genes that encode TFs does not increase that much, as only a 2-fold increase is observed from yeast (4% TF genes) to human (8%). In zebrafish, 9% of all genes encode TFs, slightly more compared with other vertebrates. This might be due to the additional genome duplication during the evolution of teleost fishes [92]. The increased number of encoded TF genes alone cannot explain the enormous complexity of the vertebrate body, with many distinct cell types and complex body functions, compared to unicellular organisms such as yeast. The differences seem to lie mainly in the combinatorial action of TRs as well as in specific post-transcriptional and post-translational modifications that will give alternative outputs in different cellular contexts [93–95].

The precise spatiotemporal expression of many TR genes is essential for normal vertebrate development. In a systematic analysis of expression of 1711 TR genes in the 24 hpf zebrafish embryo, around 200 TR genes were found to be expressed in a single tissue, whereas the vast majority of TRs (n = 1504) were detected either in multiple tissues or ubiquitously. The highly restricted expression pattern of these 200 TR genes makes them prime candidates for functional studies. This group of tissue-specific genes is highly enriched for TFs (82%; n = 207) [4].

The central nervous system and specifically the spinal cord and telencephalon express a majority of the TR genes (~60% of the TRs analysed by in situ hybridization). A significant correlation of genes co-expressed in different tissues was observed in a few brain regions, including the ectodermally derived cranial sensory ganglia (otic vesicle and olfactory bulb) and several forebrain structures as well as the retina and the tectum [4]. Interestingly, retina and tectum are functionally coupled by topographical projections of retinal axons into the tectum. However, at a global level, there was no extensive co-expression of specific TRs in multiple tissues at 24 hpf. This is in contrast to the components of cell signalling pathways, which are frequently organized into synexpression groups [96, 97]. Together, these results show that there is no extensive overlap between the TR expression patterns. This suggests a high flexibility in the combination of different factors in order to generate alternative regulatory outputs.

Genes encoding general TFs and chromatin remodelling proteins are usually ubiquitously expressed at 24 hpf with the exception of the BTB–POZ family (named after Broad complex, Tramtrack and Bric à brac/poxviruses and zinc finger). A significant proportion of these genes (38 out of the 118 mapped genes) have expression patterns restricted to the somites and the central nervous system. The BTB/POZ domain is an evolutionary conserved domain involved in protein–protein interactions leading to dimerization. Proteins containing both the BTB/POZ domain and C2H2 zinc finger DNA-binding domain have been shown to promote transcriptional repression through the recruitment of co-repressor proteins such as histone deacetylase (HDAC), N-CoR and SMRT [98–100]. Six genes of the BTB/POZ family are expressed in parts of the nervous system of the 24 hpf zebrafish embryos (Figure 1C). The tissue-restricted expression of these TR genes suggests that some members of the BTB/POZ—zinc finger family may control cellular differentiation during embryogenesis, possibly by the recruitment of transcriptional repressors such as HDACs.

Functions of ubiquitously expressed genes can still vary dramatically from one tissue to another. For example, bgl1 (alias smarca4), a protein associated with the Swi/Snf-like Brg1/Brm-associated factors (BAF) chromatin remodelling complex, is expressed in a broad range of tissues during development and in the adult zebrafish. However, mutation of bgl1 leads to specific defects in the zebrafish heart [101] and in retinal neurogenesis [102, 103]. Furthermore, disrupting the dosage balance between Brg1 and the cardiac TF Tbx5 led to impaired heart development in the mouse, indicating that normal formation of this tissue relies on precise levels of functional BAF complexes. These results highlight the importance of ubiquitous chromatin remodelers in the acquisition of specific cellular fates. The fact that most TRs are expressed ubiquitously emphasizes the importance of combinatorial action with tissue-restricted factors as well as that of post-translational and post-transcriptional regulation.

Thus, data on the global repertoire of TRs and on their expression during development are now complementing the CRM and epigenetic modification data. In order to be able to model transcriptional regulation during development, these data will have to be assembled into a global GRN. This will require large-scale functional analysis by perturbing normal networks, e.g. by creating mutants or other loss of function phenotypes for particular regulators or even for CRMs. Such large-scale efforts might be
built on smaller scale studies that were aimed at understanding particular subprocesses of development. In the next paragraph, we will discuss an example of such a study that illustrates the experimental advantages of the zebrafish model.

**GENE REGULATORY NETWORKS SPECIFYING ZEBRAFISH VENTRAL SPINAL CORD NEURONS**

Many TFs act in specific cellular contexts, thereby addressing distinct downstream genes or GRNs. Frequently, TFs are organized in pathways, which form a hierarchy of TF gene interactions. The challenge is to understand these cascades and to link the action of TFs to specific cellular outcomes. The vertebrate spinal cord offers well-studied examples of how different cell types are specified by combinations of TFs.

The spinal cord is patterned along its dorsoventral axis by antagonistic signalling interactions of the Shh, the bone morphogenetic protein and the Wnt pathways [104]. In the ventral spinal cord, the morphogen Shh, which is secreted from the underlying notochord and the floor plate of the spinal cord, induces five distinct ventral neuronal subtypes via a concentration gradient [105]. Distinct concentrations of Shh induce or repress the spatial expression of TFs, which belong predominantly to the homeodomain protein (HD) or basic helix–loop–helix (bHLH) families (Figure 2; [105]). Class II TFs are induced, while class I TFs are repressed by Shh (Figure 2C). As a consequence of this Shh activity, the five ventral neural progenitor domains express each a specific combination of TF genes (Figure 2A and B). Further cross-repressive interactions between these TF proteins sharply delineate the boundaries between the ventral neuronal domains [106]. These progenitor domains will give rise to distinct populations of neuronal subtypes (Figure 2B). Close to the ventral source of Shh, V3 interneurons form, followed at a more dorsal position by motoneurons (MNs) and then by V2, V1 and V0 interneurons further dorsally (Figure 2B).

The Gli family of zinc finger transcription factors mediates Hedgehog (Hh) signalling in all vertebrates by activating or repressing the expression of downstream target genes [108]. As is the case for mammals, zebrafish Shh is expressed in the notochord and floor plate and specifies ventral spinal cord fates [109]. Most of the TFs expressed in the spinal cord of the mouse that mark its distinct domains are expressed in the zebrafish spinal cord in a similar dorsoventral pattern (Figure 2A). It was therefore proposed that gene regulatory networks underlying spinal cord patterning are highly conserved during vertebrate evolution [109].

One of the advantages of the zebrafish model organism is the existence of a huge collection of mutants for all major signalling pathways. For instance, more than 10 zebrafish mutations have been shown to affect different components of the Shh signalling cascade [110]. Analysis of zebrafish mutants in Hh pathway components as well as over-expression of Shh highlighted differences in the requirement of Shh signalling in spinal cord development in mouse and zebrafish embryos: in contrast to mouse Shh signalling, zebrafish Hh signalling appears to be partially dispensable for the specification of some of the ventral neuronal subtypes [106, 111]. In the total absence of Hh signalling in maternal and zygotic smu mutants, motoneuron progenitors (pMN) are strongly reduced in numbers and p3 progenitors are totally absent. In contrast, p2, p1 and p0 progenitors as well as the post-mitotic V2, V1 and V0 interneurons that develop from these progenitor domains are present [111]. These results suggest that the activity of this signalling pathway in zebrafish is primarily required for the specification of the p3 progenitor domains. As in mouse, the expression of nkx2.2 and nkx2.9 genes in the p3 domain is dependent on Hh signals secreted from the floor plate and the notochord [111–115]. The expression of mouse Nkx2.2 and 2.9, and zebrafish nkx2.2a and 2.9 is driven by conserved Shh responsive CRMs, which are bound by Gli factors [115–118], suggesting that Hh directly regulates the expression of these transcription factors. In line with this hypothesis, in gli1 (dtr) mutants the expression of the nkx2 genes is greatly reduced, while misexpression of Gli1 results in ectopic induction of nkx2.9. Furthermore, a construct containing the Gli consensus binding site of the nkx2.9 enhancer drives green fluorescent protein (GFP) expression in the ventral neural tube, and mutation of this Gli-binding site abolishes GFP expression [115]. In addition, morpholino-mediated knockdown of nkx2a/b and nkx2.9 genes leads to an expansion of olig2 expression, which is a marker of pMN, into the p3 domain and abolishes the expression of the V3 marker sim1 (leucine zipper/PAS domain) [38]. Thus, zebrafish nkx2a/b and nkx2.9...
genes are required downstream of the Hh pathway for the specification of the p3 domain and the production of V3 interneurons and they contribute to the correct dorsoventral positioning of the pMN domain by repressing $\text{olig2}$ expression in a negative cross-regulatory interaction similar as in the mouse [105].

Taken together, it appears that the transcription factors and their role in the specification of different progenitor domains and of post-mitotic interneurons are largely conserved between mammals and zebrafish, but that the signals required for the correct spatial and temporal expression of these factors are only partially maintained. It has still to be determined which other signals are required for the early expression of pMN, p2, p1 and p0 progenitor domain genes in the zebrafish [111].

Thus, the patterns of expression of the TFs in the mouse and zebrafish spinal cord are very similar and the mechanisms underlying neuronal differentiation appear to be conserved, a conclusion also supported by other studies [112, 119, 120]. However, expression of TFs is not necessarily always an indicator of maintained function. This is exemplified by the GRNs controlling differentiation of two closely related types of GABAergic interneurons, the Kolmer–Agdhur neurons KA’ and KA’’ [38]. Homologues of these cells have so far not been described in the mouse spinal cord. In contrast to other neurons in the spinal cord, the cell bodies of both KA interneuron types are positioned close to the central canal, into which they extend cilia [121]. KA’ cells are required for spontaneous free swimming, while the function of KA’’ neurons is unknown [122]. The KA’’ interneurons occupy a ventral position and develop from the lateral floor plate next to the progenitors of V3 interneurons. The KA’ cells reside at a more dorsal position and arise from the motoneuron progenitor domain [38, 123–125]. The specification and formation of KA interneurons involve the Notch and Hh signalling pathways [38, 123–125]. KA’’ interneurons in the lateral floor express the TFs $\text{nrx2.2a, nrx2.2b, nrx2.9, tal2, gata2}$ and $\text{gata3}$ (Figure 3). The expression of $\text{tal2, gata2}$ and $\text{gata3}$ as well as of the GABA-synthesizing enzyme $\text{glutamic acid decarboxylase 67 (gad67)}$ is shared between KA’ and KA’’ cells (Figure 3; [38, 119]). The functional relationships

---

**Figure 2:** Schematic representation of the ventral spinal cord of mouse and zebrafish. (A) TF genes expressed in the progenitor domains (p) of mouse (left) and zebrafish (right). The expression patterns of the TF genes in the progenitors and post-mitotic interneurons of mouse and zebrafish are almost identical. (B) Ventral progenitor domains and their corresponding post-mitotic neurons. (C) Distinct concentrations of Shh induce the spatial expression of class II genes or repress class I gene expression (left). Progenitor domain borders are refined and maintained by negative cross-regulatory interactions among proteins, which share common boundaries. Negative cross-regulatory interaction can take place between class I and class II proteins or between proteins of the same class (centre). Specific combinations of homoeodomain transcription factor proteins in each progenitor domain determine the neuronal subtype (right). Figure modified from a previous study [107].
of these factors and their role in the specification of the two distinct KA interneuron subtypes were investigated in a systematic knockdown approach using morpholinos designed against each one of these TFs [38]. Differentiation of the correct number of KA cells depends on the activity of Nkx2.9 that acts in cooperation with Nkx2.2a and b. All three of these nkx2 genes are necessary for the expression of the zinc finger transcription factor gata2 and gata3 in KA" cells. Gata2 but not Gata3 is necessary for expression of the bHLH transcription factor tal2 that acts upstream of gad67 in KA" cells (Figure 3). The four genes, tal2, gata2, gata3, gad67 are also expressed in KA' cells, and they depend on the bHLH transcription factor Olig2 rather than on Nkx2 genes in these cells. Curiously, knock-down of gata2 does not affect tal2 or gad67 expression in KA' cells (Figure 3) suggesting different functional connections. Expression of tal2 and gad67 is rather dependent on Gata3 in these cells. Moreover, tal2, although like gata2 it is expressed in KA' interneurons, is not required for gad67 expression in these cells (Figure 3). Hence, the functional connections between the different regulatory genes differ in the two KA cell types: although gata2 and tal2 are also expressed in KA' cells, they are dispensable for gad67 expression in these cells. Instead, olig2 and gata3 are required for the differentiation of gad67-expressing KA' cells (Figure 3; [38]).

These data suggest that the expression state of TFs per se is not sufficient to determine the cell fate or differentiation status of cells, and that one would need to know the complete expression repertoire of TRs to predict the fate of a cell. Also, the post-transcriptional and post-translational modifications of the TRs might play a role as, for instance, the phosphorylation state of the Olig2 protein determines the choice between the MN and the oligodendrocyte cell fate [126, 127]. Even this information might not be enough, as the epigenetic history of the cell might also play a role by determining which CRMs are accessible to the TR combination present in it.

CONCLUSIONS AND PERSPECTIVES
Over the last 20 years, zebrafish has become a valuable model organism to address questions of gene regulation and regulatory networks. This model has been shown to be advantageous in reverse and forward genetic screens, in reporter assays to study enhancer/promoter function as well as in ChIP studies targeting histone modifications and TRs, thereby delivering the resources for an in-depth analysis of GRNs. The above described analysis of Kolmer–Agdhur interneuron differentiation demonstrated that although TFs can be expressed in two specific cell types, they do not necessarily have the same regulatory function in both cell types. It simultaneously underlines the major advantage of zebrafish: the potential to easily and rapidly address complex questions where the questions were actually phrased—in the organism. The examples of cell specification in a single organ, which have been presented here, also demonstrate what would be required to investigate the mechanisms underlying TR on a global scale throughout development. A systematic knock out of TRs by the transcription activator-like effector (TALE) [128] or CRISPR/CAS9 [129] systems would help to understand their integration in GRNs. TALEs could also be utilized to identify and characterize CRMs. Depending on the fused effector, nucleases can excise CRMs, while repressors and activators could temporarily change the activity of the targeted CRM(s) [130]. Experiments like this could help to
resolve how multiple CRM modules work together to define the expression domains of a gene and would demonstrate if a CRM identified in reporter assays is actually required for accurate expression in a given context. The ENCODE project demonstrated that functional enhancers can be identified by binding of TRs at distinct loci in multiple cell lines. A similarly systematic ChIP-seq study in zebrafish might shed light on the changes of CRM TR occupancy over time in development. Provided the availability of good antibodies, TF-binding sites could be discovered rapidly in fish. In summary, the time is ready for a zebrafish ENCODE project could be discovered rapidly in fish. In summary, the time is ready for a zebrafish ENCODE project to systematically mine the embryo of this animal system for regulatory interactions and to combine these with expression states and systematic functional studies of TRs.

**Key Points**

- CRMs are DNA sequences that regulate gene expression.
- TRs modulate gene expression on the protein level by interaction with CRMs.
- Gene regulatory networks are formed by a multi-layered interaction of TRs with CRMs resulting in the precise regulation of gene expression.
- The midline composed of notochord and floor plate is an organizing centre for the specification of neuronal fate in the spinal cord controlling complex gene regulatory networks.

**FUNDING**

The work in our laboratory was supported by the EU IP ZF-Health FP7-Health-2009-242048; NeuroXsys Health-F4-2009 No. 223262, the Interreg network for synthetic biology in the Upper Rhine valley (NSB–Upper Rhine); the BMBF funded network Erasys Bio BMBF KZ: 0315716 and the BioInterfaces programme of the Helmholtz association.

**References**


23. Morey C, Da Silva NR, Perry P, et al. Nuclear reorganisation and chromatin decondensation are conserved, but


