**Investigating transcriptional regulatory mechanisms of euryhaline fish in response to salinity stress using a tilapia cell line**

By

Chanhee Kim

Dissertation

Submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Animal Biology

in the

Office of Graduate Studies

of the

University of California

Davis

Approved:

---

Dietmar Kültz, Ph.D., Chair

---

Richard E. Connon, Ph.D.

---

Elizabeth A. Maga, Ph.D.

Committee in Charge

2022

**ACKNOWLEDGMENT**

I would like to dedicate this dissertation to my wife Seunghee and my son Siwoo. The journey of pursuing Ph.D. has been very exciting with my wife all the time, and we had Siwoo almost at the end of this journey, which is a huge gift for us. I am really thrilled to explore the next step of our journey. I love you guys more than anything.

I want to thank my mom and dad, who both have been a highly renowned scientist in their field, serving as a member of National Academy of Science, Republic of Korea. They have given me a precious advice on what things I have to keep in mind throughout Ph.D. training for being a 'good' scientist. Also, my parents in law always support me to focus on my research, which I really appreciate all the time.

I will miss my wonderful lab mates who have been great friends, mentors, and colleagues. I have been really enjoyed discussing issues or concerns especially with Jens at bars. We have been able to solve the confronting hurdles many times, which was really exciting when looking back those moments. I am also extremely thankful to my committee members Dr. Connon and Dr. Maga for reviewing this dissertation and providing invaluable guidance. Last but not least, I feel very lucky to have Dr. Kültz as a major professor who allowed me to pursue my Ph.D. in UC Davis, a wonderful institute for biological research. I appreciate his mentorship and support. I wish to have him as my role model as a scientist.

**ABSTRACT**

Salinity in aquatic environments limits the abundance and distribution of fish in a particular ecological niche. Mozambique tilapia (*Oreochromis mossambicus*) can tolerate a wide range of salinity stress (they are euryhaline). Such stress is currently greatly intensified in many aquatic environments due to anthropogenically accelerated climate change. Effective osmoregulation is critical for fish and other aquatic organisms to adapt to salinity stress. Physiological and biochemical approaches to understand osmoregulatory mechanisms of fish have previously identified many genes, proteins, and biochemical pathways associated with compensatory responses to salinity stress. However, the underlying molecular mechanisms that control the osmotic regulation of these genes and pathways are still largely elusive. To address this knowledge gap, the overall objective of my thesis was to identify molecular underpinnings of cellular osmoregulatory mechanisms that are utilized by euryhaline fish. We used a cell line model (tilapia cell line) to study how osmoregulated genes are activated by hyperosmotic stress. The emphasis was on identifying and characterizing cis-elements and trans-factors that control osmotic regulation of gene expression by utilizing the advantages of this cell line as a genetically tractable experimental system. Salinity-responsive *cis*-regulatory elements (CREs) and their role in the hyperosmotic induction of the tilapia glutamine synthetase gene were identified and characterized using a targeted approach (Chapter 2). A systematic non-targeted approach that utilized bioinformatics and experimental tools was used to discover new osmoregulated CREs (Chapter 3). This non-targeted approach was based on enrichment of DNA sequence motifs in promoters of hyperosmotically upregulated genes. STREME1 was identified and experimentally validated as a new salinity-responsive CRE (Chapter 3). Lastly, CRISPR/Cas9 technology was used to engineer mono- and polyclonal tilapia cell lines that harbor a functionally inactive MYC transcription factor

(TF) to enable future loss-of-function studies (Chapter 4). Specifically, using a modified limiting dilution strategy several polyclonal knockout (ko) cell lines (heterogeneous cell pools) and a monoclonal *myca* ko cell line were generated. Most of the evolutionarily conserved functional domains of MYC were removed in the monoclonal ko cell line. The knowledge and tools generated in this dissertation research advance our understanding of the molecular mechanisms that link changes in environmental salinity and extracellular osmolality with the transcriptional regulation of specific sets of genes that underlie compensatory responses to salinity stress.

**TABLE OF CONTENTS**

# CHAPTER 1

## Overall Introduction

Euryhaline teleost fishes have a unique ability to tolerate changes in habitat salinity[1], which requires a complex suite of physiological and biochemical mechanisms that detect and regulate an imposed salinity stress in their environment. Approximately 10 % of all fish species are euryhaline[2], and these species possess the genomic diversity (molecular phenotypic traits) that has allowed their adaptation to occur on at both local and global scales over a long period of evolution[3,4]. During the evolutionary adaptive process some of these fish have acquired traits that provided them with a fitness advantage in inhabiting salinity stress environments. Such salinity stress is currently greatly intensified in many aquatic environments of the world mainly due to accelerated climate change[5,6]. Among many euryhaline fish, Mozambique tilapia, *Oreochromis mossambicus*, has been known to have an extremely high tolerance to especially high salinity, up to 3x seawater (SW)[7]. Thus, this fish is a valuable model to study physiological, cellular, and molecular mechanisms that enable effective osmoregulation of fish in response to highly fluctuating salinity environments, in particular hyperosmotic stress. To maintain their internal body fluids and plasma osmolality at approximately 300 milliosmoles/kg (mOsmol/kg) as teleost, a complex regulatory network that involves the initial sensing of changes in salinity and subsequent cellular and physiological compensatory mechanisms is required[8]. A number of previous studies have attempted to investigate physiological responses of Mozambique tilapia upon salinity stress with various focuses ranging from a specific type of cells (chloride cells) found in gill tissue to hormonal changes (e.g., cortisol, prolactin) in fish plasma[9–11] to deal with hyperosmotic stress. Also, morphological changes (e.g., gill epithelium) adapting to altered salinity have been reported[12]. However, the corresponding detailed biochemical/molecular

mechanisms fish cells employ to confer these functional/physiological changes have not yet been well studied in euryhaline fish. Given that previous studies have revealed key genes (e.g., Na+/K+-ATPase, carbonic anhydrase, and osmotic stress transcription factor 1(Ostf1)) and associated pathways (e.g., *myo*-inositol biosynthesis pathway) that are responsible for osmoregulation in Mozambique tilapia[13–16], there have been knowledge gaps that can be filled by studying more in-depth the mechanisms by which identified genes and pathways are being regulated upon salinity stress. In this dissertation, thus, we primarily focused on transcriptional regulatory mechanisms of salinity response of euryhaline fish using a tilapia cell line.

Transcriptional regulation is a fundamental biological process that can allow the cell or an organism to actively respond to environmental stimuli and it can be fine-tuned depending on changes in ambient conditions[17–19]. Mozambique tilapia, freshwater (FW)-originated fish, have throughout their evolutionary history acquired special osmoregulatory mechanisms that allow for effective adaptive responses to salinity stress, resulting from having to survive in high salinity environments. With a focus on transcriptional regulation to elucidate the unique salinity stress response of these fish, a targeted approach was first addressed (Chapter 2) in order to identify *cis*-regulatory elements (CREs) in a hyperosmotically-upregulated gene, which was identified via proteomics coupled with transcription inhibitor treatment using a tilapia brain(OmB) cell line. CREs and transcription factors (TFs) are main components comprising transcriptional regulatory networks (TRNs) that orchestrate spatial and temporal gene expression responses [20]. CREs could be representative of stress response DNA sequences, as they serve as binding sites for specific TFs that dynamically reprogram TRNs in response to fluctuating environmental conditions[21]. Thus, in Chapter 2, functional CREs responsive to salinity stress were empirically tested by employing an enhancer-trapping assay that exploited a dual-luciferase system. As a result, one copy of

osmolality/salinity-responsive enhancer 1 (OSRE1) was found in intron 1 of the glutamine synthetase gene (*GS*). Functionality of OSRE1 identified in *GS* (*GS*-OSRE1) was experimentally validated for its capability to induce transcription of the gene during hyperosmolality, confirming 1) an increase in transcriptional activity in a copy number dependent manner and 2) a disappearance of transcriptional induction by deleting the *GS*-OSRE1 sequence from the expression construct. This finding is comparable with a previous study[22] identifying several OSREs in other osmotically-regulated genes (inositol monophosphatase (*IMPA*) and *myo*-inositol phosphate synthase (*MIPS*), two key enzymes in the myo-inositol biosynthesis pathway). Because OSRE1 (found in hyperosmotically up-regulated genes: *IMPA*, *MIPS*, and *GS*) is validated to function as a driver for transcriptional induction of these genes, an untargeted and systematic approach was performed to identify additional functional CREs responsive to salinity stress in Chapter 3. Using a particular suite of bioinformatics tools (MEME)[23], allowing for the discovery of common DNA sequence motif(s) among co-regulated genes or proteins, five candidate DNA sequence motifs were identified as potential osmoresponsive CREs. Of these motifs, the top-ranked motif (STREME1) was shown to represent a binding site for the Forkhead box TF L1 (FoxL1), which rationalized FoxL1 genetic manipulation as a target for future functional study (e.g., CRISPR/Cas9-mediated gene targeting) to test its involvement in gene expression regulation upon salinity stress. To evaluate the functional role of the STREME1 in salinity-responsive transcriptional induction, the enhancer-trapping assay combined with targeted motif mutagenesis was conducted to demonstrate its critical role in hyperosmotic transcriptional induction. Transcriptional regulation of euryhaline fish during salinity stress was, therefore, studied using both targeted (Chapter 2) and untargeted (Chapter 3) approaches with a primary focus on CREs

and potential binding partner TF, which interplay harmoniously to modulate transcription during environmental stress[24,25].

A reverse genetics approach can be used to identify the altered phenotype from a genetically-manipulated mutant genotype, so as to understand the function of a gene while forward genetics associates a mutant phenotype to its genetic information[26]. Gene targeting has been an important tool of reverse genetics by generating numerous *in vivo* and *in vitro* knockout(ko) models, but conventional homologs recombination (HR) methods have limitations including low efficiency and[27]**.** Recent development of genome engineering tools, such as ZFNs (Zinc Finger Nucleases), TALENs (Transcription Activator-Like Effector Nucleases) and CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9), have made a paradigm shift on reverse genetics approaches[28]. Among these tools, the CRISPR/Cas9 system has become the most widely used system due to its versatility, simplicity, and cost-effectiveness[29,30]. Rapid advances in CRISPR/Cas9 technology have accelerated the expansion of our knowledge of the function and regulation of specific genes. To further explore salinity stress-induced TFs that have been suggested from previous studies and chapters in this dissertation, Chapter 4 aimed to establish a specific gene ko cell line model using the CRISPR/Cas9 system. This provided not only useful resources for future functional analysis, but also an optimized and efficient platform to target TFs of interest, leveraging the opportunity of performing loss-of-function studies. Prior efforts using a tilapia-optimized vector-based CRISPR/Cas9 system have enabled the production of a stable Cas9-expressing tilapia cell line[31]. This cell line allowed for the generation of a complete ko fish *in vitro* model(Chapter 4) whereas in the past, RNA interference was the only tool for addressing loss-of-function assessments with incomplete gene knockdown[32–34].

The overall theme of this dissertation thesis covered by Chapters 2, 3, and 4 aims to investigate osmoregulatory molecular mechanisms of euryhaline fish with a focus on transcriptional regulation largely mediated by CRE and TF.

## CHAPTER 2

## An osmolality/salinity-responsive enhancer 1 (OSRE1) in intron 1 promotes salinity induction of tilapia glutamine synthetase

*First-author published article, Scientific Reports, 2020, DOI: 10.1038/s41598-020-69090-z*

### Abstract

Euryhaline tilapia (*Oreochromis mossambicus*) are fish that tolerate a wide salinity range from fresh water to > 3x seawater. Even though the physiological effector mechanisms of osmoregulation that maintain plasma homeostasis in fresh water and seawater fish are well known, the corresponding molecular mechanisms that control switching between hyper- (fresh water) and hypo-osmoregulation (seawater) remain mostly elusive. In this study we show that hyperosmotic induction of *glutamine synthetase* represents a prominent part of this switch. Proteomics analysis of the *O. mossambicus* OmB cell line revealed that glutamine synthetase is transcriptionally regulated by hyperosmolality. Therefore, the 5' regulatory sequence of *O. mossambicus glutamine synthetase* was investigated. Using an enhancer trapping assay, we discovered a novel osmosensitive mechanism by which intron 1 positively mediates glutamine synthetase transcription. Intron 1 includes a single, functional copy of an osmoresponsive element, osmolality/salinity-responsive enhancer 1 (OSRE1). Unlike for conventional enhancers, the hyperosmotic induction of *glutamine synthetase* by intron 1 is position dependent. But irrespective

of intron 1 position, OSRE1 deletion from intron 1 abolishes hyperosmotic enhancer activity. These findings indicate that proper intron 1 positioning and the presence of an OSRE1 in intron 1 are required for precise enhancement of hyperosmotic *glutamine synthetase* expression.

**Keywords**

Osmoregulation, euryhalinity, transcriptional regulation, glutamine synthetase, enhancer

## Introduction

Euryhaline fish have evolved the capacity to utilize a suite of osmoresponsive genes for rapidly switching between hypo- and hyper-osmoregulation in response to salinity stress to maintain plasma ionic and osmotic homeostasis[35]. Mozambique tilapia (*O. mossambicus*) are representative euryhaline fish belonging to the family of cichlidae, which consists of many species that are uniquely adapted to specific environments[3,4]. A remarkable adaptive trait of *O. mossambicus* is its ability to tolerate large and rapid salinity fluctuations, ranging from 0 to 120 g/kg even though their osmoregulatory balance starts being compromised beyond 60-65 ppt[9,36]. The corresponding changes in plasma osmolality are normally low and within in the range of 305 to 330 mOsmol/kg. However, when salinity increases chronically to values greater than 85 g/kg or acutely by more than 30 g/kg then plasma osmolality increases between 450 and 550 mOsmol/kg have been reported[37–39]. Even more moderate but acute salinity stress occurring during transfer of tilapia from freshwater to 25 g/kg results in plasma osmolality increasing up to 460 mOsmol/kg at 15 h[40]. This species has evolved molecular mechanisms for rapidly turning on and off genes that encode enzymes and transporters involved in hypo- and hyper-osmoregulation[13,14]. However, the regulatory and evolutionary mechanisms controlling environmental (e.g., salinity, temperature, and hypoxia) regulation of gene expression in fish are still largely elusive. Many of

the genes and proteins involved in transepithelial ion transport and osmoregulation of euryhaline fish have been identified using candidate gene approaches such as qPCR and Western blotting or large-scale discovery approaches such as transcriptomics and proteomics. However, the regulatory mechanisms deciphering how abundances of the corresponding mRNAs and proteins are regulated are still largely elusive. For example, for many of the regulated genes it is not known whether their abundance change is due to transcriptional regulation or posttranscriptional RNA processing and whether *cis*- and *trans*-elements that regulate gene expression are involved. This lack of knowledge contrasts with the evolutionary diversity of the fish species, which have radiated into virtually any aquatic ecological niche. Previous studies investigating which parts of the genome have a functional role in the evolution of organisms have stressed *cis*-regulatory elements (CREs) as major targets of evolutionary adaptation[41]. Therefore, alterations of CREs are considered potent drivers of evolutionary adaptation[42].

CREs typically contain binding sites for transcriptional regulators that orchestrate gene expression in response to altered environmental and developmental contexts[43,44]. Many studies have focused on characterizing enhancers, the most studied type of CREs, involved in diseases, development, and cell- and tissue-type specificity, especially in mammalian models[43,45]. For example, in human renal cells the hyperosmotic induction of the sodium/myo-inositol cotransporter (*SMIT*) is mediated via several enhancers found in its 5'-untranslated region (UTR)[46]. In contrast to these findings in mammalian models, a comprehensive understanding of enhancer functions in fish exposed to salinity stress is still very limited. We have recently identified several copies of a CRE, the osmolality/salinity-responsive enhancer 1 (OSRE1) in the inositol monophosphatase (*IMPA1.1*) and *myo*-inositol phosphate synthase (*MIPS*) genes of *O. mossambicus*[22]. Enhancers such as OSRE1 are generally considered to function independent of

whether they occur in the 5' or 3' regulatory regions or in introns[47]. Although most enhancers, including OSRE1 in *O. mossambicus IMPA1.1* and *MIPS* genes[22], are found in the 5' regulatory region, intronic enhancers have been previously reported. For example, using human cell lines, Harris et al. have identified a tissue-specific enhancer in intron 1 of the cystic fibrosis transmembrane conductance regulator gene (*CFTR*)[48]. Another study has reported that enhancers located in intron 4 are responsible for differential expression of the Bone Morphogenetic Protein 6 gene (*Bmp6*), which underlies phenotypic differences between fresh water and seawater populations of threespine sticklebacks (*Gasterosteus aculeatus*)[49].

The *glutamine synthetase* gene (*GS)* encodes an evolutionarily highly conserved enzyme that catalyzes the conversion of ammonia to glutamine. It is thought to be crucial for detoxification of ammonia as a part of nitrogen metabolism in diverse organisms including vertebrates[50]. Most studies on glutamine synthetase in fish, including euryhaline *O. mossambicus, O. niloticus* and *Oncorhynchus mykiss*, have focused on abundance or activity of glutamine synthetase in different organs such as intestine, muscle, liver and gills[51–53]. In addition to its function for nitrogenous waste detoxification in fish, glutamine synthetase also has an important function to maintain osmotic homeostasis. Glutamine synthetase produces glutamine, which can be accumulated in cells as a compatible organic osmolyte to offset the perturbing effects of hyperosmotic stress[14,54]. For example, in gills of the swamp eel (*Monopterus albus*) the induction of *GS* has been shown to promote accumulation of the compatible osmolyte glutamine during hyperosmotic stress[55]. However, little is known about transcriptional regulation of *GS* during salinity stress to adjust osmoregulation in euryhaline fish adapting to altered salinity.

Salinity stress has been predicted to intensify in the future because of climate change-induced sea level rise that causes intrusion of salty water into freshwater habitat[56]. Better

knowledge of environmentally altered salinity effects on transcriptional regulation in fish is necessary to properly assess how global climate change that is predicted to accelerate salinization of many aquatic environments will impact on the biodiversity and the future evolution of fish. Tidally induced salinity changes could potentially lead to acute salinity stress by rapidly flooding freshwater ponds filled by rainfall or river water. In addition, large estuaries in tropical areas are prone to extreme, longer-term salinity increases that are predicted to intensify in the future. For example, the Saloum estuary in West Africa (Senegal) harboring tilapia (e.g., *Sarotherodon melanotheron*) species has already been reported to reach salinities up to 130 g/kg[57]. Moreover, chronic salinity increases to such extreme levels are also predicted to result from global warming in arid regions such as desert ponds or lakes[8,58,59]. In addition, mechanistic insight into fish salinity (hyperosmotic) stress adaptation of euryhaline fish helps elucidate mechanisms that can be targeted and will contribute to improve aquaculture practices in brackish and increasingly saline environments in arid and coastal areas impacted by climate change[60].

To contribute to better understanding adaptive mechanisms controlling fish osmoregulation, we investigated the transcriptional regulatory mechanism by which osmotic responsiveness is conferred to the *O. mossambicus GS*. First, we analyzed whether the salinity-induced abundance increase of glutamine synthetase protein is based on transcriptional regulation. Then, an enhancer trapping reporter assay was used to identify the specific genomic regions that are responsible for transcriptional induction of *GS* during hyperosmolality.

## Materials and Methods

### Cell culture

The tilapia OmB cell line was used for all experiments and luciferase reporter assays. OmB cells were maintained in L-15 medium containing 10 % (vol/vol) fetal bovine serum (FBS) and 1 % (vol/vol) penicillin-streptomycin at 26 ℃ and 2 % CO2. The purpose of FBS supplement is to support sufficient and reproducible OmB cell growth and potential variability issues derived from FBS (e.g., unknown components in FBS can interact with OmB cells or treatments) were minimized/resolved by employing proper controls in parallel with all treatments to isolate osmolality as the only variable factor. Using a large supply of OmB cell superstock (passage 15; P15), all experiments were carried out on OmB cells between P18 to P26. Cells were passaged every 3-4 d using a 1:5 splitting ratio. For applying hyperosmotic stress to OmB cells, hyperosmotic (650 mOsmol/kg) medium was prepared using hypersaline stock solution (osmolality: 2820 mOsmol/kg). This stock solution was made by adding an appropriate amount of NaCl to regular isosmotic (315 mOsmol/kg) L-15 medium. The hypersaline stock solution was then diluted with isosmotic medium to obtain hyperosmotic medium of 650 mOsmol/kg. Medium osmolality was always confirmed using a freezing point micro-osmometer (Advanced Instruments).

### Proteomics

Sample preparation by tryptic in solution digestion, data-independent acquisition (DIA) and targeted proteomics were performed as previously described using a nanoAcquity UPLC (Waters), an ImpactHD mass spectrometer (Bruker), and Skyline[61] targeted proteomics software[22]. Three peptides of GS that are identical in sequence in *O. mossambicus* and *O.*

*niloticus* (NCB Accession # XP_003444352.1) were used for quantitation (Supplementary Fig. S1). Three proteins, represented by at least three peptides each, were used for normalization (fatty acid-binding protein, NCB Accession # XP_003444095.3, beta-tubulin, NCB Accession # XP_003455078.1, and actin 2, NCB Accession # XP_003455997.3).

**Cloning**

Total genomic DNA was extracted from spleen tissue of Mozambique tilapia (*O. mossambicus*) using the PureLink Genomic DNA mini Kit (Invitrogen). Fish were maintained and euthanized before obtaining spleen tissue according to UC Davis approved Institutional Animal Care and Use Committee (IACUC) protocol # 19992. PCR primers were designed with Geneious 11.0.3 (Biomatters) using the *O. niloticus* glutamine synthetase (NCB Accession # XM_003444304.4 and XP_003444352.1) genomic sequence as a template. A CCCCC spacer followed by a restriction enzyme recognition site was added to the 5' end of each primer. The restriction enzymes KpnI, SacI, HindIII, and NcoI (New England BioLabs) were used to clone PCR amplicons representing genomic regions of the GS gene into pGL4.23 vector. Platinum PCR SuperMix (Thermo Fisher Scientific) and/or Q5® High-Fidelity DNA Polymerase (New England BioLabs) were used to amplify DNA fragments longer than 2 kb. For fragments < 2 kb, PCR Master Mix 2x (Promega) was used. PCR was conducted as follows: initial denaturation at 94°C for 3 min followed by 35 cycles of 94°C for 30 s, annealing: 48-60° for 30 s, elongation: 72°C for 0.5-2 min, and 72°C for 15 min. Annealing temperature and extension time were set according to the chemical features of the primers and the lengths of amplicons. PCR products were confirmed by agarose gel electrophoresis and sequentially either purified using the PureLink PCR Purification Kit (Thermo Fisher Scientific) or gel-extracted using the QIAquick® Gel Extraction Kit (Qiagen). Specific primers were designed for the translation start site (start codon, SC, +499)

and the 3' end of exon 1 (Ex1_3', +131). The SC and Ex1_3' primers included a NcoI restriction site that was already present in the wildtype GS donor sequence and in the pGL4.23 acceptor reporter plasmid. Therefore, genomic regions of interest that terminate at the SC and Ex1_3' sites could be cloned without changing any wildtype sequence. All amplified GS gene fragments were double-digested with two enzymes (combinations of KpnI, SacI, HindIII, and NcoI). Restriction enzyme digestion was conducted in 10 µL reaction buffer (CutSmart®Buffer and NEBuffer™ 1.1) containing 2 µL (10 U/µL) of each restriction enzyme, 0.5-2 µg of purified PCR product, and nuclease-free H2O ad 100 µL. After overnight incubation at 37°C, reactions were stopped by 20 min incubation at 80°C. Digested inserts and vectors were purified using the PureLink™ Quick PCR Purification Kit (Thermo Fisher Scientific) and ligated to produce recombinant constructs using T4 DNA ligase (Thermo Fisher Scientific). Ligation reactions contained 50 ng of vector, 10-20 ng of insert (depending on its size to yield a 1:3 or 1:5 molar ratio), 2 µL of ligase buffer, 1 µL of T4 ligase (1 U/µL) and nuclease-free H2O to 20 µL. Ligation proceeded at 25°C for 6h. The ligation products were transformed into 10-beta-competent *Escherichia coli* (New England Biolabs) as follows: First, a 50 µL aliquot of bacteria was thawed on ice for 5 min, then 10 µL of bacterial suspension was added to 1.5 µL of a single ligation reaction. Second, the mixture was kept on ice for 30 min, exposed to heat shock (42°C) for exactly 30 s, and placed back on ice for 5 min. Third, 190 µL of super optimal broth with catabolite repression medium (SOC, Thermo Fisher Scientific) was added and transformed bacteria were incubated at 250 rpm and 37°C for 60 min. After transformation, 30 µL of the bacterial solution was spread onto a pre-warmed (37°C) LB-ampicillin plate, which was used for single colony picking and colony PCR on the next day to confirm the presence of intended inserts. For colony PCR, tubes containing a bacterial clone were first quick-vortexed, then heated at 95°C for 15 min and quick-spun to remove debris. Three µL

of the supernatant were mixed with forward and reverse primers that flank the corresponding insert. Colony PCR conditions were the same as described above and amplicons were checked by agarose gel electrophoresis. Colonies that contained an insert of the expected size were chosen for plasmid purification. Each bacterial colony was inoculated into liquid LB medium and grown for 16-18 h to maximize plasmid yield. Liquid cultures were harvested and purified according to manufacturer's protocol using endotoxin-free PureLinkTM Quick Plasmid Miniprep Kit (Thermo Fisher Scientific). Insert sequences in purified DNA constructs were verified by Sanger sequencing at the University of California, Davis DNA Sequencing Facility before using the corresponding constructs for transient transfection into tilapia OmB cells.

**Enhancer trap reporter assays**

Enhancer trapping assays were performed according to the protocol previously reported by our laboratory[22]. To produce a backbone luciferase vector harboring the endogenous functional promoter of the *GS*, the functional *GS* core promoter region (GS-CP, -257 to +131, Fig. 3a) was cloned upstream of the firefly (*Photinus pyrails*) luciferase gene in the pGL4.23 vector (GenBank Accession Number DQ904455.1, Promega) and verified that it has constitutive activity but is not hyperosmotically inducible. The resulting reporter plasmid was named GS-CP+4.23. The GS-CP region was amplified using a forward primer that included a HindIII restriction site and a reverse primer that included a NcoI restriction site. The GS-CP region and pGL4.23 plasmid were digested with the same pair of restriction enzymes and followed by ligation. Cloning, purification, and sequence-validation were conducted as described in the cloning procedure.

The GS-CP+4.23 plasmid was used in combination with *hRluc (Renilla reniformis*) luciferase control plasmid pGL4.73 (GenBank Accession Number AY738229.1, Promega). Co-transfection of tilapia OmB cells with this control plasmid was used to normalize for variability of

transfection efficiency and cell number. One day prior to co-transfection OmB cells were seeded in 96-well plates (Thermo Fisher Scientific) at a density of 2 x $10^4$ cells per well. Co-transfection was performed when cells reached 80% to 90% confluency. Co-transfection was performed with ViaFect (Promega) reagent using previously optimized conditions[22]. Cells were allowed to recover for 24h after transfection before being dosed in either isosmotic (315 mOsmol/kg) or hyperosmotic (650 mOsmol/kg) media for 72h. Dual luciferase activity was measured in 96-well plates using a GloMax Navigator microplate luminometer (Promega). Four biological replicates were used for each experimental condition. All luciferase raw measurements were adjusted for transfection efficiency by normalizing the firefly luciferase activity to *Renilla* luciferase activity. They were expressed as fold-change in hyperosmotic media relative to isosmotic controls. One-way ANOVA was performed to assess statistical significance of the data and calculate p values using R package software (http://www.R-project.org/).

**Bioinformatics sequence analysis**

Intron 1 was searched for the occurrence of an OSRE1 consensus motif using a bioinformatics approach. For this purpose, Geneious 11.0.3 (Biomatters) was used. Both strands, sense and antisense, were searched. Sequence similarity searches were conducted by using the overall OSRE1-consensus sequence (DDKGGAAWWDWWYDNRB) as well as several experimentally validated and previously identified variants of OSRE1 sequences, including the 17 bp sequence AGTGGAAAAATACTAAG (*IMPA1.1*-OSRE1), as templates[22].

**Synthetic oligonucleotide annealing and GeneStrands synthesis**

The effect of *GS*-OSRE1 copy number variation and *GS*-OSRE1 deletion on hyperosmotic reporter activity was analyzed. Synthetic oligonucleotides containing different copy numbers of

*GS*-OSRE1 were produced by oligonucleotide annealing (Eurofins Genomics). *GS*-OSRE1 constructs containing one, two, three, four and five copies were generated. Forward and reverse PCR primers for amplifying each synthetic oligonucleotide were designed to contain SacI and HindIII restriction sites to enable subsequent cloning into GS-CP+4.23 vector (Supplementary Table S1). Synthetic oligonucleotides harboring more than three copies of *GS*-OSRE1 or mutated intron 1 (Intron 1▲*GS*-OSRE1) were longer than 100 bp. These longer inserts were synthesized using the GeneStrands method (Eurofins Genomics). Subsequently, each insert was separately cloned into GS-CP+4.23 luciferase reporter vector. After cloning into the reporter plasmid, the proper sequences of all inserts were verified by Sanger sequencing. These constructs were used to assess the effect of *GS*-OSRE1 copy number and deletion of *GS*-OSRE1 from intron 1 on reporter activity under hyperosmotic (650 mOsmol/kg) conditions relative to isosmotic controls (315 mOsmol/kg).

**Data Availability**

All data generated or analyzed during this study are included in this manuscript and its Supplementary information files. Sequence data for the 5' RS of the *O. mossambicus GS* gene investigated in this study can be found in GenBank with Accession Number: MN631059. The DIA assay library, results, and metadata for glutamine synthetase quantitation are publicly accessible in the targeted proteomics database Panorama Public[62] at the following link: https://panoramaweb.org/lUknq6.url.

## Results

### Hyperosmotic induction of glutamine synthetase is transcriptional and mediated by intron 1

Actinomycin D applied to OmB cells during exposure to hyperosmotic stress prevented glutamine synthetase production, which confirms that glutamine synthetase upregulation is mediated by transcriptional induction (Fig. 2-1 and Supplementary Fig. 2-S1). Quantitation of glutamine synthetase abundance revealed a 4.65 ± 0.18-fold increase during hyperosmotic stress (mean ± s.e.m, $p<0.0015$, Fig. 2-1). This increase in glutamine synthetase abundance was completely abolished by including 10 mM actinomycin D in the media to yield a slight 0.85 ± 0.09-fold reduction during hyperosmotic stress (mean ± s.e.m, $p=0.2726$, Fig. 2-1).

The 3.4-kb 5' regulatory sequence (RS), including the 5'-UTR, of the *O. mossambicus GS* was cloned, sequenced, and submitted to GenBank (GenBank Accession Number: MN631059). The start codon (SC) for translation was located in exon 2 (Fig. 2-2a). The first region tested for hyperosmotic enhancer activity was very long and spanned base pairs -2825 to the SC (+499). The corresponding plasmid construct with the first region inserted for a luciferase reporter assay is shown in Supplementary Fig. 2-S2. This region conferred a 3.2±0.09 (s.e.m)-fold ($p<0.001$) increase in luciferase reporter gene activity under hyperosmotic conditions relative to isosmotic controls (Fig. 2-2b). Iteratively narrowing this large region into successively shorter regions that had an identical 3' end but differed at the 5' end did not result in any loss of hyperosmotic induction of the reporter. These shortened constructs yielded 3.5± 0.18(s.e.m)-fold ($p<0.001$, -718 to SC), 3.4± 0.35(s.e.m)-fold ($p<0.001$, -257 to SC), 3.4± 0.24(s.e.m)-fold ($p<0.001$, -108 to SC), and 3.7±0.30(s.e.m)-fold ($p<0.001$, -60 to SC) reporter gene transcriptional induction, respectively (Fig. 2-2c). The shortest of these regions (559 bp) that is contained in all five constructs is

composed of the core promoter, exon 1, and intron 1. These results suggest that the core promoter, exon 1, and/ or intron 1 are responsible for induction of the *GS* gene during hyperosmolality.

Because introns have recently been shown to confer transcriptional enhancement of several eukaryotic genes[63], the role of intron 1 for the hyperosmotic *GS* induction was investigated further. First, intron 1 was excluded from all four shortened constructs to yield constructs that contain fragments whose 3' end coincided with the end of exon 1 (+131 bp downstream of the transcription start site, TSS) (Fig. 2-3a). Removal of intron 1 completely abolished the hyperosmotic induction of reporter activity for all four of these constructs (-718 to +131, -257 to +131, -108 to +131, and -60 to +131) (Fig. 2-3b). This result demonstrates that intron 1 of *GS* is required for its hyperosmotic transcriptional induction. To test the hyperosmotic induction of intron 1 in a more physiological context using the endogenous rather than a heterologous core promoter we isolated the *GS* core promoter (GS-CP). For this purpose, a reporter plasmid containing the GS-CP (-257 to +131) was constructed (Fig. 2-3a). The functional GS-CP region (-257 to +131) was selected from four putative GS-CP regions because previous studies have shown that for many genes the functional promoter spans from approximately 250 bp upstream of the TSS to 100 bp downstream[64]. Deleting the region spanning -257 to -108 bp from the GS-CP abolishes GS-CP activity. In addition, we have identified three downstream promoter elements (DPEs) in the GS-CP by motif searching for the 'RGWYVT' consensus motif (Fig. 2-3a).

**Intron 1 contains a single, functional copy of OSRE1**

Systematic bioinformatics searches of the entire intron 1 sequence for the occurrence of a previously identified OSRE1 was performed by utilizing the OSRE1-consensus (DDKGGAAWWDWWYDNRB) and several specific OSRE1 sequences (incl. *O. mossambicus*

*IMPA1.1*-OSRE1: AGTGGAAAAATACTAAG) that yielded high hyperosmotic induction of reporter activity in a previous study[22]. This approach enabled us to identify a single copy of OSRE1-like sequence (AGTGGAAAAATACAAC) in intron 1 of *GS*. This *GS*-OSRE1 was 16 bp long and almost identical (88%) to *IMPA1.1*-OSRE1, harboring only one gap and a single mismatch. *GS*-OSRE1 was localized on the reverse strand (Fig. 2-4a).

To verify whether *GS*-OSRE1 has functional activity as an enhancer element during salinity stress, a series of luciferase reporter plasmids driven by the endogenous GS-CP were constructed. Synthetic oligonucleotides harboring different numbers of copies of *GS*-OSRE1 were used to validate its function as an osmoresponsive enhancer. Constructs containing either a single copy or up to five copies of *GS*-OSRE1 were tested using dual luciferase reporter assays (Fig. 2-4b). Each of these constructs conferred hyperosmotic induction of reporter activity. Moreover, the extent of induction was proportional to the number of *GS*-OSRE1 copies. However, a single copy yielded only a very small albeit significant degree of hyperosmotic induction 1.2±0.11(s.e.m)-fold ($p<0.01$). In contrast, two copies (2.2±0.23(s.e.m)-fold, $p<0.01$), three copies (4.6±0.86(s.e.m)-fold, $p<0.01$), four copies (6.6±0.55(s.e.m)-fold, $p<0.001$), and five copies (7.6±0.40(s.e.m)-fold, $p<0.001$) of *GS*-OSRE1 yielded much greater hyperosmotic induction (Fig. 2-4c). These data demonstrate that *GS*-OSRE1 functions as an osmoresponsive CRE during hyperosmotic stress. However, they also show that a single copy of *GS*-OSRE1 is insufficient to explain the 3.4 to 3.7-fold hyperosmotic GS induction mediated by intron 1 (Fig. 2-2c).

After confirming the enhancer function of *GS*-OSRE1 we refined the consensus sequence for OSRE1 by including the *GS*-OSRE1 sequence in the consensus. This inclusion resulted in a change of the overall OSRE1 motif from DDKGGAAWWDWWYDNRB to DDKGGAAWWDWWYNNRB (Fig. 2-5).

**Hyperosmotic induction of *GS* depends on the location of intron 1 and requires OSRE1**

The dependence of hyperosmotic induction of *GS* on the location of intron 1 was investigated to address whether OSRE1-containing intron 1 behaves as a conventional position-independent enhancer. Unexpectedly, when intron 1 was positioned downstream of the GS-CP (which represents its native genomic context) the hyperosmotic induction of reporter activity was much lower than when it was trans-positioned upstream of the GS-CP (3.4-fold vs. 9.9-fold, Fig. 2-6a and 2-6b). This result shows that intron 1-mediated transcriptional regulation of *GS* during salinity stress depends on the location of intron 1, which is atypical for conventional enhancers[65]. This atypical but pronounced position-dependency of intron 1 mediated enhancement represents a potential mechanism for evolutionary tuning of enhancer responsiveness via trans-positioning regulatory elements.

In addition to establishing the position-dependency of intron 1 enhancement (Fig. 2-6a and 2-6b) and functionally validating OSRE1 (Fig. 2-4), we investigated whether *GS*-OSRE1 is necessary for the enhancer function of intron 1. To test whether the presence of *GS*-OSRE1 is essential for intron 1-mediated hyperosmotic transcriptional induction of *GS* the 16 bp OSRE1 sequence was deleted from intron 1. The rationale for this experiment was that, although a single copy of *GS*-OSRE1 was insufficient to account for the hyperosmotic induction of the *GS* gene (Fig. 2-4c), it may still be required as an essential component triggering the formation an inducible transcription factor complex. Two luciferase reporter plasmids with deletions of *GS*-OSRE1 were constructed and tested for luciferase activity in OmB cells exposed to iso- and hyperosmotic media (Supplementary Fig. 2-S3). One of these constructs harbored intron 1 downstream of the TSS in its native context and the other contained intron 1 trans-positioned upstream of the TSS (Fig. 2-6c). Selectively deleting *GS*-OSRE1 (16 bp) from intron 1 completely abolished the hyperosmotic

transcriptional induction conferred by intron 1, independent of the location of intron 1 (Fig. 2-6d). This result demonstrates that *GS*-OSRE1 is necessary for the enhancer activity of *GS* intron 1.

**Discussion**

In this study, we discovered a novel molecular mechanism where intron 1 harboring an osmoresponsive CRE (*GS*-OSRE1) positively regulates transcription of the tilapia *GS* under hyperosmotic stress. Furthermore, we identified that intron 1-mediated hyperosmotic *GS* induction requires the OSRE1 element and that its enhancer activity depends on the location of intron 1. The molecular mechanism of intron 1 enhancement of *GS* transcription during hyperosmotic stress differs from that of conventional context-inducible transcriptional enhancers, many of which have been previously shown to function independent of orientation or distance (relative position) to the TSS[66,67].

Euryhaline fish embrace a tolerance stage before osmoregulatory mechanisms have fully adjusted to altered salinity, which makes them temporarily vulnerable to dysregulation of osmotic homeostasis before adaptive mechanisms have been remodeled and can take effect[68]. Our group and others have documented that tilapia induce organic osmolyte synthesis in multiple tissues during salinity stress, which implies that plasma osmolality increases significantly under those conditions[16,69,70]. OmB cells were exposed to 650 mOsmol/kg, which exceeds the plasma osmolality increase documented in intact tilapia (450-550 mOsmo/kg, see introduction section). However, 650 mOsmol/kg is still below the maximal osmotolerance of this cell line (735 mOsmol/kg) and mechanistic dissection of physiological mechanisms is best performed when the corresponding mechanisms are robustly induced, even under slightly exaggerated conditions[70]. Moreover, cells and organisms have evolved a safety margin of physiological capacity in response to demand that exceeds physiological conditions actually experienced by as much as ten-fold[71].

Previously, we analyzed the 5' RS of *O. mossambicus IMPA1.1* and *MIPS* genes and identified six osmotically responsive CREs with a common 17 bp consensus motif, which we named OSRE1[22]. All of these OSRE1 elements were located between -232 and +56 bp relative to the TSS in both genes. The first osmotically responsive enhancers were identified in mammalian cell lines and named Tonicity responsive element (TonE)[72] and, alternatively, osmotic response element (ORE)[73]. Bai et al. have also characterized a distinct osmotic-responsive element (OsmoE) in a mouse kidney cell line and revealed its genomic locus to be further upstream (-808 to -791 bp relative to the TSS) in the *NHE-2* gene encoding the $Na^+/H^+$ exchanger-2[74]. In addition to the discovery of OsmoE, this study also identified a TonE-like element far upstream (-1201 to -1189 bp) in the same gene[74]. Therefore, our initial attempt to identify an osmotically responsive CRE in the *GS* utilized the 3.4-kb region of 5' RS spanning from –2825 to +526 bp. Our tilapia study and these previous observations in mammalian model systems suggest (with rare exceptions) that osmoresponsive CREs are located preferentially very close (within a few hundred bp) to the TSS. This knowledge informs comparative studies and future searches for osmoresponsive CREs in other genes and/or species.

With the discovery of functional OSRE1 in *GS*, *IMPA1.1* and *MIPS* (see above), further genes are deduced to be regulated via OSRE1. The production of compatible osmolytes represents a common functional role of all three genes (*GS*, *IMPA1.1*, *MIPS*). Therefore, it is possible that other genes with the same function are also regulated via OSRE1. One possible candidate is the aldose reductase gene (*AR*) which produces the organic osmolyte sorbitol and was shown to harbor a TonE/ORE less than 1 kb upstream of the transcription start site in a mammalian model[75]. Furthermore, the sodium- and chloride-dependent taurine transporter gene may be controlled by OSRE1 as its mRNA increases with salinity in tilapia[76,77]. The taurine transporter promotes

concentration of taurine, another compatible osmolyte, in cells exposed to hyperosmotic stress[78]. Another OSRE1 candidate gene is glycine synthase, which increases during hyperosmotic stress and promotes the production of glycine (a neutral amino acid) that can also function as a compatible osmolyte[9]. However, in all of these cases it is not sufficient to find OSRE1 consensus sequences near the promoter without experimentally validating them.

In addition to the role of hyperosmotically induced genes and cellular mechanisms of osmoregulation, a variety of other endpoints has been documented at higher levels of biological organization for euryhaline fish undergoing salinity adaptation. For example, physiological differences in organ function and phenotypic differences of tissues have been detected in osmoregulatory organs such as the gill, kidney, and intestine. Drinking rates and intestinal water absorption are increased in parallel to salinity and the number and size of ionocytes in gill epithelium of euryhaline fish increase during hyperosmotic stress[8]. Salinity adaptation at the whole organism level also includes significant integrative effects of hormones such as growth hormone (GH), insulin-like growth factor 1 (IGF-1), and cortisol to facilitate systemic integration of salinity adaptation to hyperosmolality[79].

We elucidated that intron 1 in combination with the endogenous GS-CP mediates transcriptional induction of the *GS* under salinity stress. Introns have been shown to boost gene expression in numerous ways including by providing binding sites for transcription factors, regulating the rate of transcription, promoting nuclear export, and stabilizing transcripts[63]. Several studies of plant species have identified a positive effect of introns on transcription or mRNA accumulation in a constitutive rather than context-dependent manner[80]. Moreover, most reported cases of intron-mediated transcriptional enhancement are stimulus-independent[81]. Only a small number of studies has thus far investigated the stimulus-responsiveness of introns.

However, some previous studies using different cell lines have shown that stimulus-dependent transcriptional regulation of a variety of genes is mediated by intron 1[82,83]. For example, in a human breast cancer cell model, intron 1 of the *ERBB2* proto-oncogene (*ERBB2*) contains a 409 bp sequence that mediates *ERBB2* transcriptional changes in response to oestrogens[82]. These previous studies reporting stimulus-dependent intron 1 mediated enhancement are consistent with our finding that intron 1 enhances *GS* transcription during hyperosmotic stress. Therefore, our study provides evidence that introns, which have often been regarded as "junk DNA" that is spliced out during mRNA processing, represent functional genomic targets for evolutionary adaptation to environmental changes.

Our observation that the degree of hyperosmotic transcriptional *GS* induction mediated by intron 1 is position-dependent suggests that the corresponding mechanism is distinct from typical CRE-mediated enhancement. A position-dependent effect of an intronic enhancer was also reported for intron 2 of the human beta-globin gene demonstrating that changes in the location of intron 2 relative to the promoter alters transcriptional activity three-fold[67]. This result is very similar to the three-fold change in transcriptional activity observed in our study when the location of *GS* intron 1 was altered. Moreover, the position dependence of intron-mediated enhancement (IME) has been well documented in plants[84].

In other cases reported for IME, however, intron 1 has been shown to act independent of its location[85]. It is likely that sequence rearrangements around the TSS result in conformational changes of the transcriptional machinery, which then affects transactivation efficiency[86]. Therefore, it is possible that intron 1 trans-positioning changes the structural conformation of the transcriptional machinery in a way that increases transactivation. The location of intron 1 in a position that does not maximally enhance transactivation suggests that evolution has favored

moderate over strong transcriptional induction of *GS* during hyperosmolality. Otherwise, transposition of the CRE elements included in intron 1 upstream of the TSS would have been evolutionarily favored. Possible reasons for limiting the extent to which *GS* is induction during hyperosmotic stress are as follows: Glutamine synthetase abundance during hyperosmotic stress may represent a compromise between its ability to produce a compatible organic osmolyte (glutamine) on the one hand and its consumption of energy (ATP) on the other hand. In most organisms including fish, glutamine synthetase is an essential enzyme that mediates bidirectional biochemical reactions, ammonia assimilation and glutamine biosynthesis[87]. Thus, a moderate increase of glutamine synthetase abundance during hyperosmotic stress may be evolutionarily favored as the most cost-effective strategy during salinity stress.

This study demonstrates that the *GS*-OSRE1 element in intron 1 is essential for transcriptional induction during hyperosmotic stress. We interpret these results as evidence that the OSRE1 element serves as a critical binding site for a hyperosmolality-inducible transcription factor. The prime candidate transcription factor for activating OSRE1 during hyperosmotic stress is nuclear factor of activated T cells (NFAT5). Mammalian NFAT5 is a fundamental regulator of the cellular response to osmotic stress in mammals. It binds to the TonE/ORE enhancer[88,89]. Since TonE/ORE and OSRE1 share a common core motif (TGGAAAA), tilapia NFAT5 has high potential for binding to OSRE1 and controlling its enhancer activity during hyperosmolality. NFAT5 also contributes to osmosensory or osmoregulatory mechanism in fish but its precise role and whether it binds to OSRE1 is still unclear[90,91]. Another candidate of an OSRE1 binding protein is the tilapia homolog of transcription factor II B (TFIIB). Tilapia TFIIB mRNA is induced rapidly and transiently within a short period of exposure of fish to salinity stress (4-fold within 2 h) whereas other stressors (oxidative stress and heat stress) did not trigger its induction[92].

Mammalian TFIIB is known to bind to a specific DNA sequence (B recognition element, BRE: SSRCGCC) to promote transcription of a gene by stabilizing the general transcriptional machinery. Thus, TFIIB is less likely to interact with OSRE1 directly but rather might be involved in stabilization of a multi-protein enhancer complex[93]. However, no sequence that resembles the mammalian BRE element is present in the proximal promoter region of *GS* suggesting that the homologous tilapia sequence diverges significantly from that of mammals, occurs in a region that is more distant from the *GS* core promoter or is not involved in the osmotic regulation of the *GS* gene. In accordance with the hypothetical existence of several candidate osmoresponsive transcription factors, our results suggest that a combination of inducible transcription factors is necessary for promoting transcriptional enhancement since a single copy of *GS*-OSRE1 outside its native intron 1 sequence context was inefficient for enhancing transactivation. We conclude that other, yet to be identified CREs, are present in intron 1 that interact with OSRE1 to result in transcriptional enhancement. Such combinatorial interactions between different CREs and corresponding transcription factors are common[94]. One important focus of future research will be to characterize such complexes and their interactions.

In conclusion, *GS* intron 1 was revealed to contain a single OSRE1 (*GS*-OSRE1) and to enhance transcriptional induction of *GS* in a tilapia (*O. mossambicus*) cell line exposed to hyperosmolality. The mechanism for this transcriptional enhancement of *GS* expression during hyperosmolality has two characteristics: 1. Its extent is dependent on the location of intron 1 relative to the TSS, 2. It requires *GS*-OSRE1 for intron 1 enhancer function. Furthermore, our data strongly suggest that the previously identified osmoresponsive CRE OSRE1 consensus sequence can be used for bioinformatics screening approaches that identify candidate OSRE1 sequences on a genome-wide bases[95]. Identification of the transcription factor(s) that bind to *GS*-OSRE1 and

potential other osmoresponsive CREs in intron 1 represents an intriguing future task to understand the process by which osmotic stress signals are perceived and transduced to regulate the expression of genes that compensate for salinity stress in euryhaline fish.

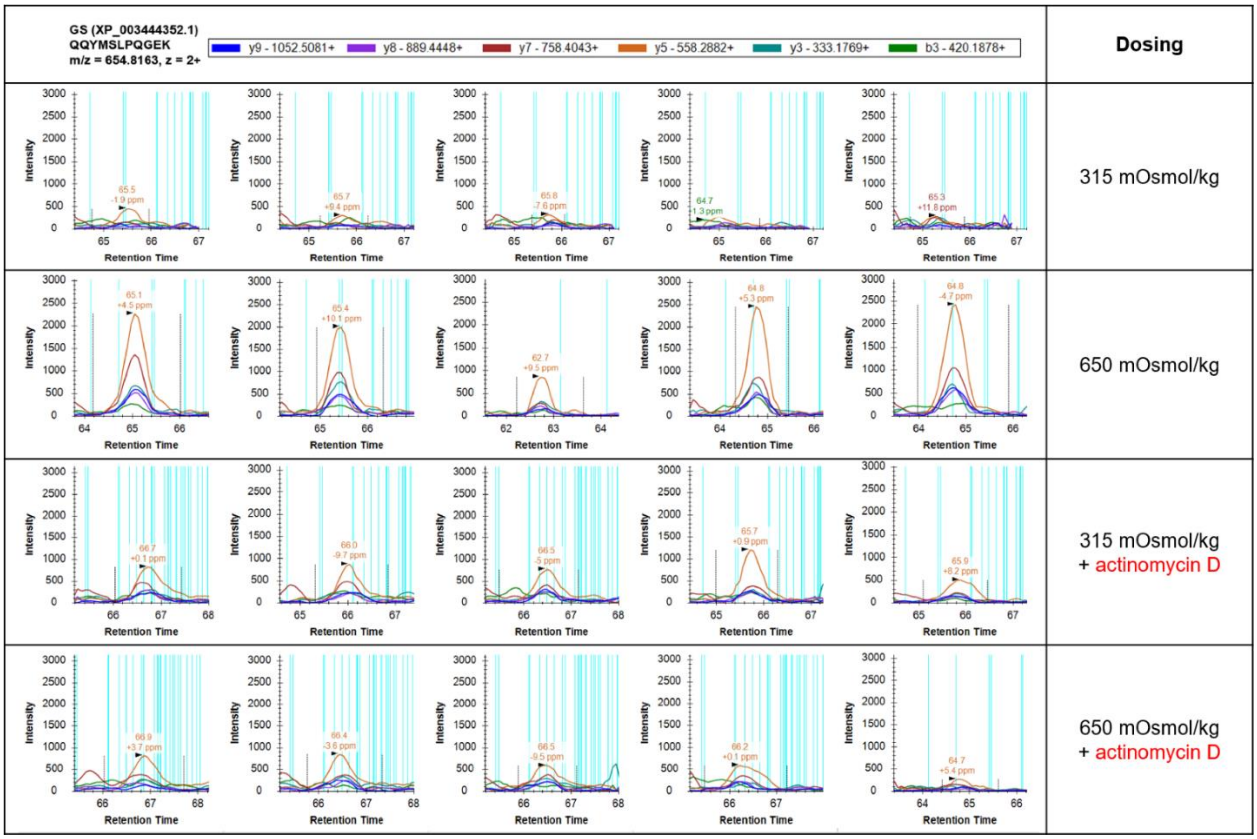



**Figure 2-1.** Targeted DIA-MS/Skyline protein quantitation of glutamine synthetase protein (GS, XP_003444352.1) in cells grown in four different medium conditions: isosmotic (315 mOsmol/kg), hyperosmotic (650 mOsmol/kg), isosmotic+10µM actinomycin D, hyperosmotic+10µM actinomycin D. Data for one of the four quantified peptides, QQYMSLPQGEK, is shown. Each treatment consisted of five biological replicates (from left to right).

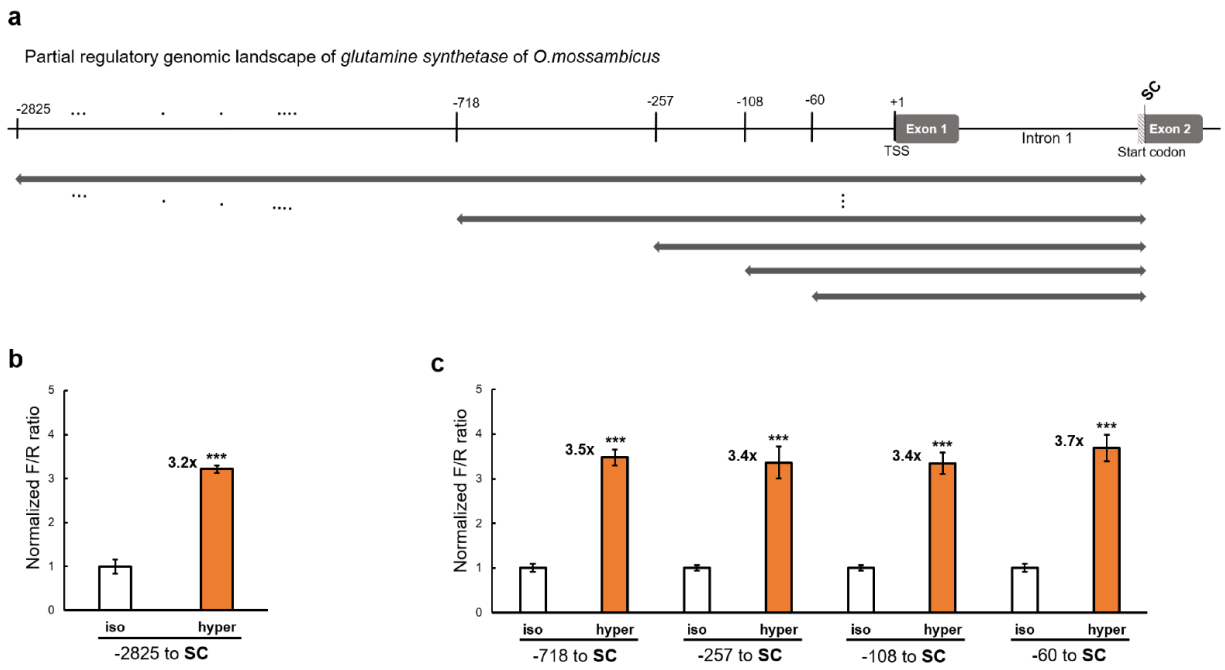



**Figure 2-2.** Narrowing of the osmotically regulated genomic region of the *GS* gene. (a) 3.4-kb long 5'-flanking genomic region and the 5'-UTR (including exon 1 and intron 1) of the *O. mossambicus GS* gene is illustrated. Numbers at the top indicate the genomic position relative to the transcription start site (TSS). The bars with arrows on both sides indicate the first set of genomic regions analyzed for hyperosmotic induction of reporter activity. The SC primer contains an NcoI restriction site at the translation start codon (SC); (b and c) Fold-change in luciferase reporter activity induced by hyperosmolality relative to isosmotic controls ((b) for the region -2825 to SC and (c) for successively shorter regions is shown). Normalized F/R ratio expresses inducible Firefly luciferase activity versus constitutive *Renilla* luciferase activity. This ratio was measured for both isosmotic (315 mOsmol/kg) and hyperosmotic (650 mOsmol/kg) conditions and normalized by setting isosmotic controls to one. The number of asterisks indicates the statistical significance of the hyperosmotic induction ($p<0.001$: ***, $p<0.01$: **, $p<0.05$: *, ns: not significantly different).

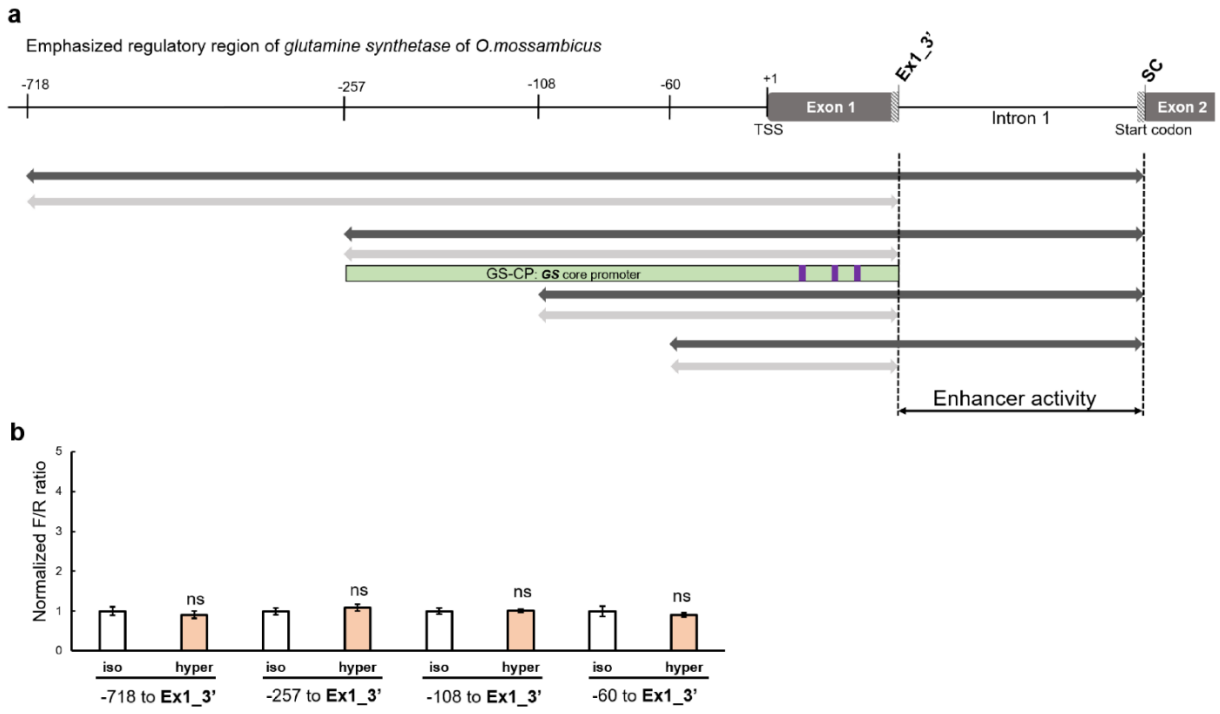



**Figure 2-3.** Identification of intron 1 as the genomic region necessary for hyperosmotic *GS* induction. (a) The genomic sequences used for reporter assays are shown. The Ex1_3' primer contains a NcoI restriction site at the 3' end of exon 1. The grey bars indicate constructs that exclude intron 1 from the original constructs tested in Fig. 2(black bars). The light green bar presents the functional *GS* core promoter that is used for constructing the backbone reporter vector GS-CP+4.23. Three purple bars in the light green bar indicate downstream promoter elements (DPEs) that match to 'RGWYVT' motif. (b) Hyperosmotic induction of reporter activity is completely abolished when intron 1 is excluded from the luciferase constructs. The normalized F/R ratio expresses inducible Firefly luciferase activity versus constitutive *Renilla* luciferase activity. This ratio was measured for both isosmotic (315 mOsmol/kg) and hyperosmotic (650 mOsmol/kg) conditions and normalized by setting isosmotic controls to one. The number of asterisks indicates the statistical significance of the hyperosmotic induction ($p<0.001$: ***, $p<0.01$: **, $p<0.05$: *, ns: not significantly different). Figure layout and abbreviations are as outlined in Fig. 2.2.

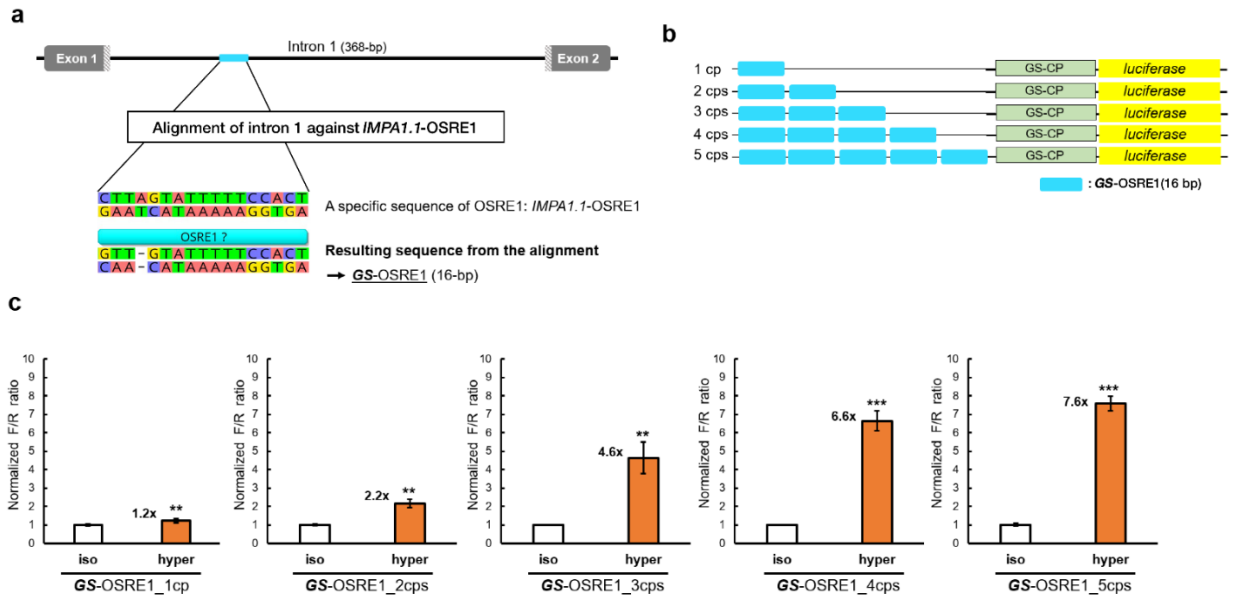



**Figure 2-4.** Identification of an osmoresponsive element (OSRE1) in intron 1 of *GS*. (a) Pairwise alignment of the *GS* intron 1 sequence against the 17 bp *IMPA1.1*-OSRE1 sequence yielded a match with 15 identical bases and 88.2 % of pairwise identity (one gap, one mismatch), which is referred to as *GS*-OSRE1. (b) Reporter constructs containing different copy number of *GS*-OSRE1 are depicted. (c) *GS*-OSRE1 represents an inducible enhancer whose transcriptional potency is proportional to copy number. (cp = copy, cps = copies). The normalized F/R ratio expresses inducible Firefly luciferase activity versus constitutive *Renilla* luciferase activity. This ratio was measured for both isosmotic (315 mOsmol/kg) and hyperosmotic (650 mOsmol/kg) conditions and normalized by setting isosmotic controls to one. The number of asterisks indicates the statistical significance of the hyperosmotic induction ($p<0.001$: ***, $p<0.01$: **, $p<0.05$: *, ns: not significantly different).

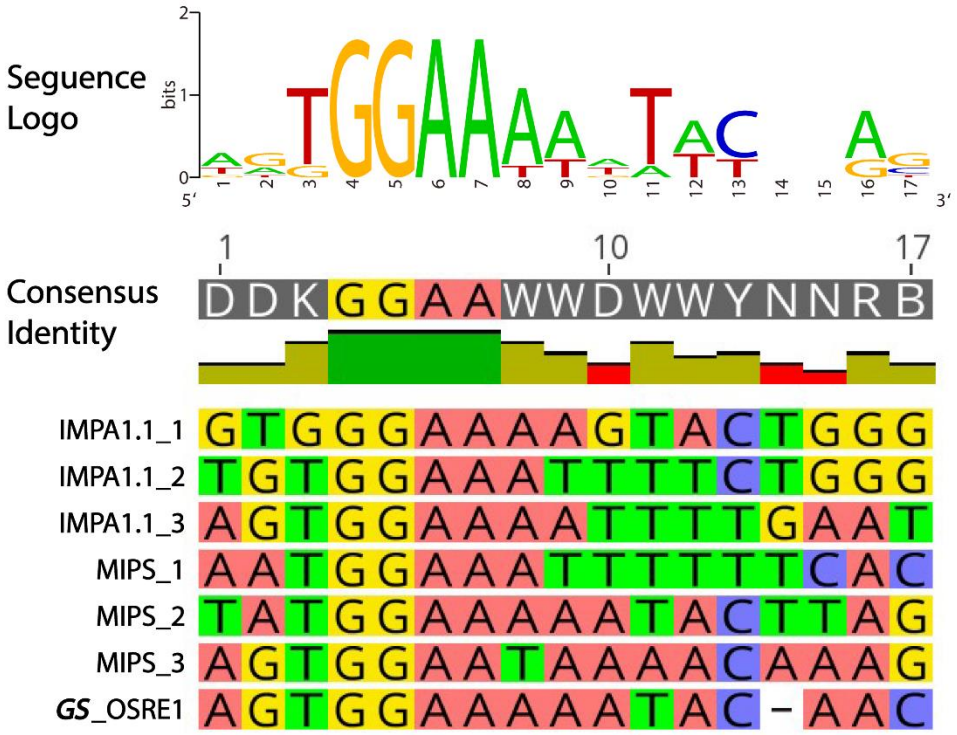



**Figure 2-5.** Refinement of osmolality/salinity-responsive enhancer 1 (OSRE1) consensus sequence. A multiple sequence alignment of *GS*-OSRE1 with previously identified OSRE1 motifs in *IMPA1.1* and *MIPS* genes[22] is shown. This alignment yields the refined consensus sequence DDKGGAAWWDWWYNNRB.

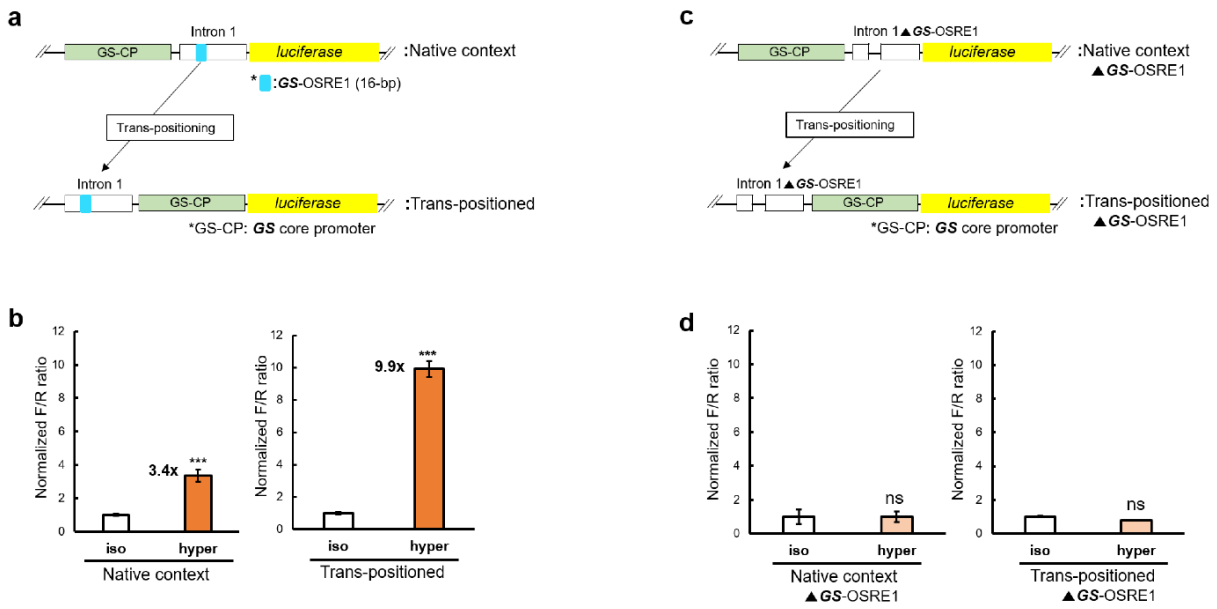



**Figure 2-6.** Two characteristics of the mechanism for hyperosmotic induction of *GS* by intron 1. (a) The genomic position of intron 1 was changed from downstream (native context) to upstream (trans-positioned) relative to the transcription start site (TSS). The light green bars indicate the *GS* core promoter (GS-CP). The light blue color indicates the *GS*-OSRE1 element. (b) The corresponding reporter gene activity results are illustrated in (a). (c, d) The effect of selective deletion of *GS*-OSRE1 from intron 1 (Intron 1▲*GS*-OSRE1) on reporter activity is shown. All reporter assays were carried out with reporter plasmid containing the GS-CP. Normalized F/R ratio expresses inducible Firefly luciferase activity versus constitutive *Renilla* luciferase activity. This ratio was measured for both isosmotic (315 mOsmol/kg) and hyperosmotic (650 mOsmol/kg) conditions and normalized by setting isosmotic controls to one. The number of asterisks indicates the statistical significance of the hyperosmotic induction ($p<0.001$: ***, $p<0.01$: **, $p<0.05$: *, ns: not significantly different).

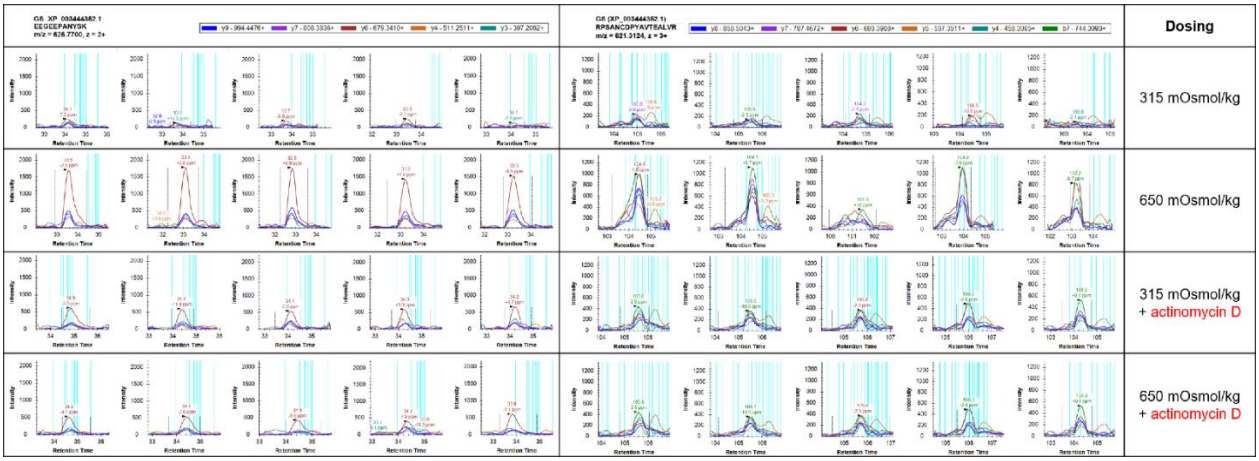



**Figure 2-S1.** Additional targeted SWATH-MS/Skyline protein quantitation data of GS protein in cells grown in four different medium conditions (Dosing: isosmotic (315 mOsmol/kg), hyperosmotic (650 mOsmol/kg), isosmotic+ 10µM actinomycin D, hyperosmotic+10µM actinomycin D). Two different peptides of glutamine synthetase (EEGEEPANYSK, RPSANCDPYAVTEALVR) are shown with five biological replicates for each dosing condition.

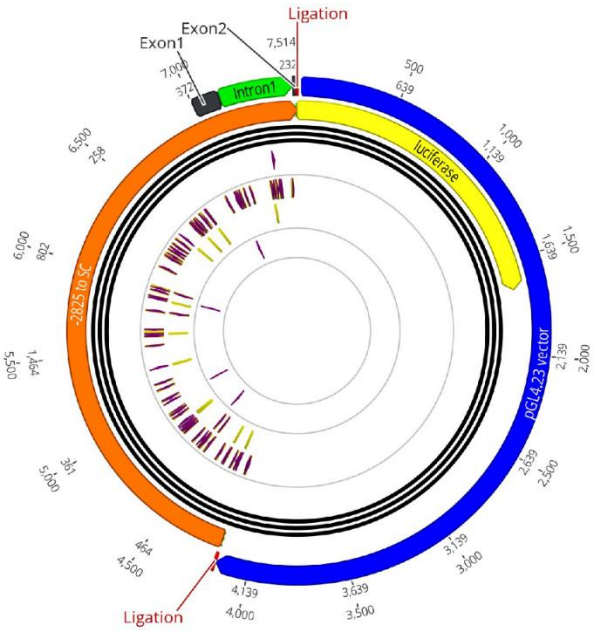



**Figure 2-S2.** Plasmid map of the initial luciferase reporter construct consisting of pGL4.23 vector (blue) and an insert representing the 3.4-kb 5' regulatory sequence (RS) of *GS* (orange, -2825 to +499). The 3' end of the insert (+499) represents the start codon (SC). Exon 1, a truncated Exon 2 (dark grey), intron 1 (green) of *GS* and the luciferase reporter gene (yellow) are indicated.

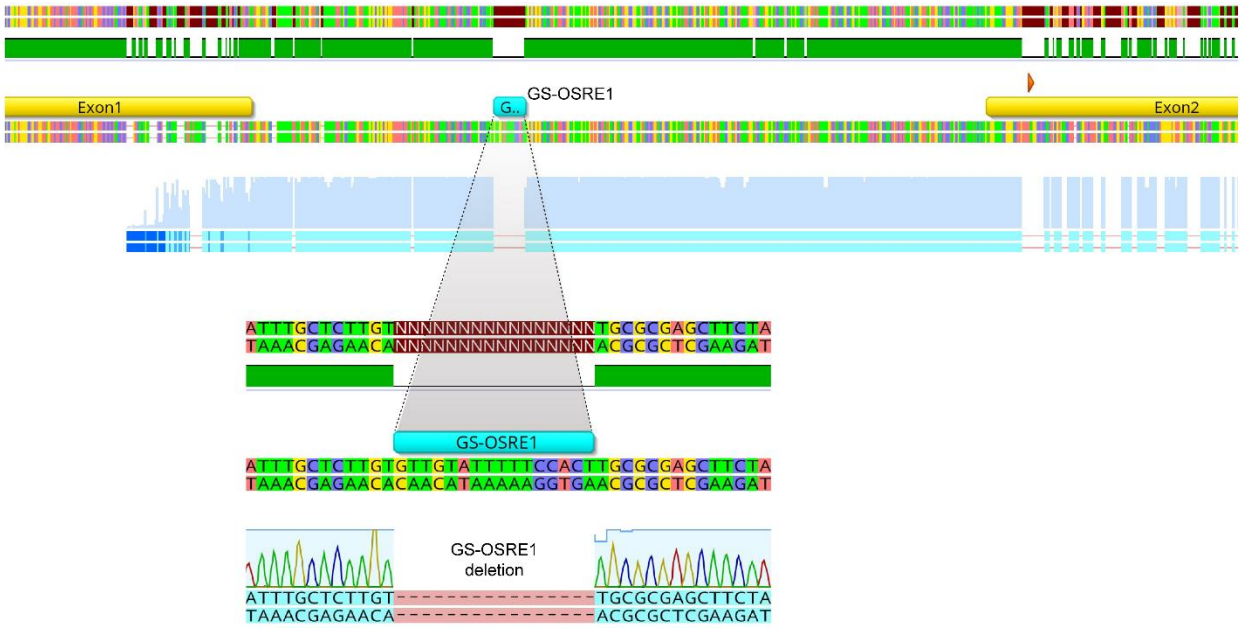



**Figure 2-S3.** The representative sequencing result from the reporter constructs showing a successful selective deletion of *GS*-OSRE1 from intron 1.

*GS*-OSRE1: GTTGTATTTTTCCACT

| Number of copy | Whole oligonucleotide sequence (bp length) | Forward primer sequence | Reverse primer sequence |
|---|---|---|---|
| 1 | CCCCCGAGCTCCATGCATGCATGCTGTGTTGTATTTTTCCACTTGCATGC ATGCATGCCGAGCTAAGCTTGGGGG (75 bp) | CCCCCGAGCTCCATGCATGCATG CTGTGTTGTATTTTTCCACTTGCA TGCATG | CCCCCAAGCTTAGCTCGGCATG CATGCATGC AAGTGGAA |
| 2 | CCCCCGAGCTCCGTTGTATTTTTCCACTACTGGTTGTATTTTTCCACTCGA GCTAAGCTTGGGGG (65 bp) | CCCCCGAGCTCCGTTGTATTTTT CCACTACTGGTTGTATTTTTCC | CCCCCAAGCTTAGCTCGAGTGG AAAAATACA ACCAGTAGTGG |
| 3 | CCCCCGAGCTCTGTGTTGTATTTTTCCACTTGCACTGTGTGTTGTATTTTT CCACTTGCACTGTGTGTTGTATTTTTCCACTTGCAAGCTTGGGGG (96 bp) | CCCCCGAGCTCTGTGTTGTATTT TTCCACTTGCACTGTGTGTTGTAT TTTTCCAC | CCCCCAAGCTTGCAAGTGGAAA AATACAACCACAGTGCAAGTGG AAAAATACAACACAC |
| 4 | CCCCCGAGCTCTGTGTTGTATTTTTCCACTTGCACTGTGTGTTGTATTTTT CCACTTGCCGAGCTTGTGTTGTATTTTTCCACTTGCACTGTGTGTTGTATT TTTCCACTTGCAAGCTTGGGGG (124 bp) | Genestrands synthesis (Eurofins Genomics) | |
| 5 | CCCCCGAGCTCTGTGTTGTATTTTTCCACTTGCACTGTGTGTTGTATTTTT CCACTTGCACTGTGTGTTGTATTTTTCCACTTGCACTGTGTGTTGTATTTTT CCACTTGCACTGTGTGTTGTATTTTTCCACTTGCAAGCTTGGGGG (148 bp) | Genestrands synthesis (Eurofins Genomics) | |

**Table 2-S1.** Synthetic oligonucleotide sequences containing different number of copies of *GS*-OSRE1 (SacI and HindIII restriction sites also included) and the corresponding Forward & Reverse primers sequences for their amplification. Red and black color variation separates multiple ***GS***-OSRE1 copies.

## CHAPTER 3

# Prediction and Experimental Validation of a New Salinity-Responsive Cis-Regulatory Element (CRE) in a Tilapia Cell Line

*First-author published article, Life, 2022, DOI: 10.3390/life12060787*


### Abstract

Transcriptional regulation is a major mechanism by which organisms integrate gene x environment interactions. It can be achieved by coordinated interplay between cis-regulatory elements (CREs) and transcription factors (TFs). Euryhaline tilapia (*Oreochromis mossambicus*) tolerate a wide range of salinity and thus are an appropriate model to examine transcriptional regulatory mechanisms during salinity stress in fish. Quantitative proteomics in combination with the transcription inhibitor actinomycin D revealed 19 proteins that are transcriptionally upregulated by hyperosmolality in tilapia brain (OmB) cells. We searched the extended proximal promoter up to intron1 of each corresponding gene for common motifs using motif discovery tools. The top-ranked motif identified (STREME1) represents a binding site for the Forkhead box TF L1 (FoxL1). STREME1 function during hyperosmolality was experimentally validated by choosing two of the 19 genes, chloride intracellular channel 2 (*clic2*) and uridine phosphorylase 1 (*upp1*), that are enriched in STREME1 in their extended promoters. Transcriptional induction of these genes during hyperosmolality requires STREME1, as evidenced by motif mutagenesis. We conclude that STREME1 represents a new functional CRE that contributes to gene x environment interactions during salinity stress in tilapia. Moreover, our results indicate that FoxL1 family TFs are contribute to hyperosmotic induction of genes in euryhaline fish.

**Keywords**

cellular osmoregulation; cis-regulatory element; fish; salinity; stress tolerance; transcription factor; transcriptional regulation.

## Introduction

A major challenge of biology is understanding the mechanisms that govern gene x environment interactions and the phenotypic diversity of organisms. Studies of physiological and biochemical ecology aimed at understanding and explaining how organisms adapt to environmental change and stress currently rely heavily on correlations of phenotypes with particular single nucleotide polymorphisms (SNPs) or other sequence variation and transcriptomics[96]. However, multiple levels of biological organization and regulation are interspersed between the genome and complex phenotypes with transcriptional regulation of gene expression being only one of many mechanisms by which changes in transcriptomes, proteomes, and complex cellular and organismal phenotypes are achieved[97].

One mechanism by which organisms respond to environmental signals (e.g., temperature changes, salinity fluctuations, etc.) is by regulating gene expression[98,99]. Transcriptional regulation of specific gene(s) is a fundamental regulatory process for controlling gene expression[18]. Understanding transcriptional regulation is thus critical for elucidating how molecular mechanisms shape the phenotypic changes of organisms in response to environmental stress[100]. Transcriptional regulation is based on the interaction of cis-regulatory elements (CREs) that control the transcription of associated genes and transcription factors (TFs) that recognize and bind to CREs to influence transcription of those genes[101]. Harmonious interactions (binding events) between those two components in response to environmental stimuli are known to govern

gene expression in an organized manner[25,102,103]. Despite much attention and interest in environmental control of gene expression and many studies documenting elaborate changes of transcriptomes in response to environmental stresses, little is known about the molecular mechanisms that control transcriptional regulation in stress tolerant (eurytopic) organisms exposed to environmental stress.

Mozambique tilapia (*Oreochromis mossambicus*) are eurytopic fish that are highly tolerant to many environmental stresses, including large salinity changes. Gene expression patterns of tilapia have been correlated with various phenotypic characteristics that are important for aquaculture, e.g., muscle growth and meat quality[104,105]. Another important trait for tilapia aquaculture is environmental resilience. Several tilapia species, including *O. mossambicus*, have undergone a remarkable adaptive evolution to cope with large salinity fluctuations in their environment. *O. mossambicus* is able to tolerate salinities from 0 to 120 g/kg and plasma osmolality changes ranging from 305 to 800 mOsmol/kg[106]. This astonishing phenotypic plasticity renders Mozambique tilapia an excellent model for investigating the underlying molecular mechanisms that orchestrate the control of gene expression during hyperosmotic salinity stress.

The influence of salinity on gene expression patterns in tilapia has been investigated, complemented by studies of other systems-level, holistic molecular phenotypes, notably metabolomes and proteomes[107–109]. These systems-level studies have revealed that salinity stress has very pronounced effects on transcriptomes and proteomes, causing significant changes in hundreds of gene products. Although these studies have correlated many gene products with salinity stress in tilapia and other euryhaline fish, the regulatory mechanisms that are causal for such changes are mostly elusive. Few studies have identified the mechanism of regulation of

transcripts and proteins, i.e., whether regulation takes place at the level of transcription (gene expression), posttranscriptional mRNA abundance regulation, translational regulation, or protein turnover. We have previously demonstrated that gene expression control by a specific novel CRE, the osmotic/salinity response element 1 (OSRE1), is largely responsible for the hyperosmotic upregulation of several osmoregulated proteins[22,110].

Approaches for identifying and experimentally validating regulatory sequences, such as CREs of a particular gene that mediate a response to environmental stress (e.g., temperature and salinity), have been mostly used for relatively few model species[46,111]. They require robust genomic resources and are laborious and technically challenging. However, as more genomic sequence information and effective computational tools have become available for a greater diversity and number of species, genome-wide comparative approaches for identifying potential regulatory sequences such as CREs have become more powerful and are now commonly used for yeast, certain plants, and mammalian models[103,112–114]. The combination of computational prediction and experimental validation represents a powerful tool for elucidating the mechanisms that underlie changes in gene expression in response to environmental or developmental cues and for establishing causality between changes in certain transcript and protein abundances and environmentally controlled signaling networks[44,115]. Recently, such approaches have been used to delineate transcriptional networks in zebrafish (*Danio rerio*) and understand genetic networks that control the physiological adaptation of fish[116,117]. In one study, differentially expressed genes of zebrafish exposed to low temperature were analyzed for enriched CREs using a motif discovery program, and subsequent experimental validation of the identified motifs revealed cis- and trans- elements (CREs and TFs) that control gene expression during cold stress[116]. In another study, tilapia (*Oreochromis niloticus*) and zebrafish were compared to

decipher divergent aspects of cold stress responses by identifying TF binding sites in extended promoter region of genes with species-specific regulation during cold stress. This approach was complemented by experimental validation and yielded a genetic network of cold stress responses in different fish species[117].

In this study, a similar comparative bioinformatics approach was used to identify a novel CRE and corresponding TF candidate, and then experimentally validate the functionality of the candidate CRE during hyperosmotic stress.

## Materials and Methods

### Hyperosmotic Stress Challenge and Actinomycin D Treatment

The *O. mossambicus* brain (OmB) cell line was subjected to all hyperosmotic stress challenges. L-15 medium containing 5% (vol/vol) fetal bovine serum (FBS) and 1% (vol/vol) penicillin-streptomycin at 26 °C was used to grow OmB cells at 2% CO2 as previously described[22,110]. Using a large supply of OmB cell superstock (passage 15; P15), all experiments were conducted on OmB cells between P20 and P27. OmB cells were passaged every 3–4 days using a 1:5 splitting ratio and exposed to hyperosmotic medium (osmolality: 650 mOsmol/kg) during hyperosmotic stress challenge. The hyperosmotic medium was made by adding an appropriate volume of hyperosmotic stock solution (osmolality: 2820 mOsmol/kg) to isosmotic L15 medium (osmolality: of 315 mOsmol/kg). An appropriate amount of NaCl was added to isosmotic L-15 medium to prepare the hyperosmotic stock solution. Medium osmolality was measured by freezing point micro-osmometer (Advanced Instruments). All exposures were performed by acutely increasing medium osmolality from 315 to 650 mOsmol/kg for 24 h. Parallel handling controls were subjected to medium change without increasing the medium

osmolality. Actinomycin D, a widely-used transcription initiation inhibitor[118,119], was added at a concentration of 10 µM to a subset of hyperosmotically challenged OmB cells and isosmotic controls to analyze the contribution of transcriptional regulation in the hyperosmotic upregulation of protein.

**Quantitative Proteomics**

Sample preparation and in-solution trypsin digestion were performed as previously described[120]. A DIA-LCMS2 approach was used to ensure highly accurate relative quantitation of many proteins. DIA was invented in 2012[121,122] and avoids undersampling of peaks and inconsistent peak picking. DIA-LCMS2 is also known under the acronym sequentially windowed acquisition of all theoretically possible MS spectra (SWATH)-MS[123,124] and represents a merger of targeted MS approaches such as selected reaction monitoring (SRM) and non-targeted MS2 spectra acquisition[125]. DIA targeting of specific transitions, precursors, peptides, and proteins is performed post-acquisition by interrogating all MS2 spectra present in a sample against a previously validated DIA assay library. Using a previously published procedure[120], we have generated a high quality DIA assay (MS2 spectral) library for *O. mossambicus* OmB cells which includes 3043 unique proteins meeting stringent quality control (QC) criteria and consisting of non-redundant diagnostic peptides (Figure 3-S1). DIA data were acquired as previously described[126] and analyzed with Skyline[61], mProphet[127], MSstats[128]. They were deposited and are publicly accessible at PanoramaPublic[62] and ProteomeXchange[129] (see Data Availability Statement). The following parameters were used for MSstats analysis of quantitative DIA data: normalization method = equalize medians, confidence interval = 95%, scope = protein, summary method = Tukey's median polish, mProphet Q value cutoff = 0.05.

**Motif Discovery and Refinement**

Motif-based sequence analyses were performed using the MEME bioinformatics suite[23]. Specifically, three MEME suite analysis tools were used: STREME[130], TOMTOM[131], and FIMO[132]. Common motifs in a set of the regulatory sequences were searched for using STREME, a motif discovery tool that identifies motifs, which are enriched in the input sequences (regulatory sequences from 19 transcriptionally osmoregulated tilapia genes). STREME compares the input sequences to a control dataset that is generated by shuffling each of the input sequences. Approximately 5 kb representing the extended promoters up to intron 1 were extracted in FASTA format for each of the 19 genes using the genome database of *O. niloticus* (isolate F11D_XX linkage groupS, O_niloticus_UMD_NMBU, whole genome shotgun sequence, NCBI). STREME was carried out on these 19 regulatory sequences using default parameters, except the range was set to between 8 and 18 bp to capture pertinent potential transcription factor binding sites according to the typical length range of binding sites for TFs[133,134]. To estimate false discovery rate, STREME processes both the input sequences and an equally large decoy set consisting of their reverse complements. This approach permits the use of Fisher's exact test as a statistical test for assessing statistical significance of motif enrichment. Significant STREME motifs identified with this approach were then evaluated with the TOMTOM motif comparison tool to compare these motifs with known TF binding sites (CREs). In the TOMTOM approach, to sequences were aligned to curated eukaryotic DNA-JASPAR, vertebrate, and UniPROBE mouse databases of known CREs with a *p*-value cut-off of 5e-3 and sequence divergence cutoff of fewer than 2 bases. FIMO was subsequently run to scan and annotate all occurrences of TOMTOM-annotated motifs in each regulatory region of the 19 hyperosmotically induced genes. The FIMO tool converts each input motif into a log-odds position specific scoring matrix (PSSM) and uses each PSSM to

independently scan each input sequence. All positions in each sequence that match a motif with a statistically significant log-odds score are then reported. The q-value is similar to a *p*-value but corrected for multiple testing, and a q-value of 0.01 or less was used as the threshold for statistical significance using FIMO.

**Cloning**

Genomic DNA used for PCR amplification was extracted from OmB cells using the PureLink Genomic DNA mini kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA). PCR primers were designed using Geneious Prime 2022.0.1 (Biomatters, https://www.geneious.com) with the *O. niloticus* genomic sequences of chloride intracellular channel 2 (NCBI Gene ID # 100694858) and uridine phosphorylase 1 (NCBI Gene ID # 100690403) as templates. A CCCCC spacer followed by a restriction enzyme recognition site was added to the 5' end of each PCR primer. The restriction enzymes KpnI and NcoI (New England Biolabs, Ipswich, Massachusetts, USA) were used to clone PCR amplicons representing extended proximal promoters up to intron 1 of each gene into pBS_EGFP expression vector. Q5 high-fidelity DNA polymerase (New England Biolabs, Ipswich, Massachusetts, USA) and GoTaq Green Master Mix (Promega, Madison, Wisconsin, USA) were used to amplify DNA fragments after confirming single-band PCR amplicons on regular DNA agarose gels. PCR was carried out as follows: initial denaturation at 95 ℃ for 3 min followed by 35 cycles of 95 ℃ for 30 s, annealing: 50–60° for 30 s, elongation: 72 ℃ for 0.5–1 min and 72 ℃ for 15 min. Annealing temperature and extension time were adjusted according to the chemical features of the primers (e.g., Tm) and the lengths of amplicons. PCR amplicons were checked by agarose gel electrophoresis and subsequently purified using the PureLink PCR Purification Kit (Thermo Fisher Scientific) if a single band was detected. In cases where multiple bands were visible on an agarose gel, a specific band with the expected

size was gel-extracted using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). Specific primers were designed to clone the parts harboring predicted motifs within proximal regulatory sequences of each gene. Primer pairs were generated to be compatible with KpnI and NcoI sites in the acceptor expression vector (clic2_5'KpnI, clic2_3'NcoI, upp1_5'KpnI, upp1_3'NcoI, the corresponding primer sequences are listed in table 3-S1).

PCR amplicons and pBS_EGFP vector were double digested with KpnI and NcoI restriction enzymes. Restriction enzyme digestion reactions were prepared as follows: 10 µL reaction buffer (rCutSmartBuffer™ and NEBuffer™ r1.1), 2 µL (10 U/µL) of each restriction enzyme, 0.5–2 µg of purified PCR amplicons (or pBS_EGFP vector), and nuclease-free $H_2O$ were added to 100 µL. After overnight incubation at 37 °C to ensure complete digestion, reactions were stopped by heating at 80 °C for 20 min. Digested PCR amplicon inserts and vectors were purified using the PureLink Quick PCR Purification Kit (Thermo Fisher Scientific) and subsequently ligated to produce desired recombinant constructs for experimental validation using T4 DNA ligase (Thermo Fisher Scientific). Ligation reactions were prepared as follows: 50 ng of vector, 10–20 ng of insert (1:5 molar ratio), 2 µL of ligase buffer, 1 µL of T4 ligase (1 U/µL) and nuclease-free $H_2O$ added to 20 µL. Ligation reactions were incubated in a thermocycler (Mastercycler, Eppendorf) at 25 °C for 5 h. The ligation products were transformed into 10-beta-competent *E. coli* (New England Biolabs) as previously described[110]. After transformation, an appropriate amount of the bacterial solution was spread onto a prewarmed (37 °C) LB-ampicillin plate. The plate was used for single colony picking and subsequent colony PCR to check for the presence of intended PCR amplicon inserts.

Colony PCR was performed by heating tubes containing a single bacterial clone picked from the plates at 95 °C for 15 min and quick centrifugation, and resulting supernatants were used

as a template. The supernatant (3 µL) was mixed with forward (M13_Forward) and reverse primers (GFP_R5) that flank the corresponding PCR amplicon insert (Table 3-S1). Colony PCR thermocycler conditions were the same as described above and amplicons were confirmed by agarose gel electrophoresis. Colonies that harbored an insert of the expected size were chosen for bacterial cell cultures followed by plasmid purification. Each bacterial colony was inoculated into liquid LB medium and grown for 18–20 h to obtain a sufficient amount of plasmid. Liquid cultures were harvested and purified according to manufacturer's protocol using endotoxin-free PureLink Quick Plasmid Miniprep Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Insert sequences in purified DNA constructs were verified by Sanger sequencing with M13_Forward and GFP_R5 primers at the University of California, Davis DNA Sequencing Facility.

**Site-Directed Motif Mutagenesis**

"Overlap Extension PCR" was used to mutate candidate motifs identified in regulatory regions of the tilapia *clic2* and *upp1* genes. Two or three independent PCR amplifications were performed using the extended PCR primers containing nucleotide replacements for introducing nonfunctional motifs and complementary sequences for stable binding into sequence fragments. PCR amplicons representing fragments of the overall sequence were then then used as PCR templates (1 µL of each PCR amplicon) and subsequently stitched together using PCR with the outermost primers to obtain a single intermediate PCR amplicon. The final amplifications of the entire 1 kb long regulatory regions containing the mutated motifs of *clic2* and *upp1* were then performed using the same PCR primers as those used for amplification of the corresponding wildtype sequences (Table 3-S1). The sequences for all mutagenesis constructs for *clic2* and *upp1* were confirmed by Sanger sequencing after each plasmid was miniprepped as described above (Figure 3-S1).

### Quantitative Fluorescent Reporter Assay

To perform quantitative fluorescent reporter assays, tilapia OmB cells were seeded in six-well plates (Corning, Glendale, Arizona, USA) and transiently transfected at 80 % confluency with four different enhanced green fluorescent protein (GFP) expression vectors containing regulatory regions of *clic2* (wildtype and mutant) and *upp1* (wildtype and mutant). After 24 h, the transfected cells were dosed either with hyperosmotic (650 mOsmol/kg) or isosmotic (315 mOsmol/kg) medium for 24 h. For GFP quantification, a Dmi8 fluorescence microscope (Leica) was used to capture fluorescence micrographs of OmB cells cotransfected with one of the GFP-expression vectors and a control vector containing red fluorescent protein (RFP) that was used for normalization. Instead of capturing a random single fluorescence image of part of the each well, a complete tile scan of the well was performed to quantify fluorescence in the entire well for all conditions using the Dmi8 automatic stage microscope (Leica, Wetzlar, Germany) and Leica Application Suite X (LAS X) software. Tile scanning of each well was carried out to detect GFP and RFP signals from the designated part of each well. Intensity sum values were used to calculate the relative GFP/RFP intensity ratio. Five independent biological replicates (individual wells) were used to enable testing for statistical significance of treatment effects on GFP/RFP intensity ratio.

## Results

### Transcriptional Regulation Is Required for Upregulation of Proteins in OmB Cells Exposed to Hyperosmotic Stress

Increases in protein abundances of Mozambique tilapia OmB cells exposed for 24 h to hyperosmolality (650 mOsmol/kg) compared to isosmotic controls (315 mOsmol/kg) were calculated based on DIA data using Skyline and MSstats and visualized using volcano plots.

Remarkably, the upregulation of all statistically significant proteins (multiple testing correct $p < 0.05$ and fold-change >2) was abolished when transcription was inhibited by the inclusion of 10 μM actinomycin D in the medium (Figure 3-1). This result confirms transcriptional regulation as a predominant mechanism underlying the upregulation of proteins during hyperosmotic stress. Nineteen hyperosmotically upregulated proteins whose upregulation was completely abolished by transcription inhibition were chosen to serve as a basis for identifying common motifs in their regulatory sequences. These proteins are indicated as red triangles in Figure 3-1. Inositol-monophosphatase (XP_005449080.1) was not included in this set even though it showed the same pattern of regulation (blue diamond in Figure 3-1) because the extent of upregulation was more than an order of magnitude greater than for the other proteins and we had previously analyzed the regulation of this protein and its corresponding gene in depth[22]. Interestingly, one of the 19 proteins selected for motif searching (ferritin, heavy subunit, XP_003445743.1) was significantly downregulated in the presence of actinomycin D, suggesting that it may be subject to very rapid turnover in OmB cells exposed to hyperosmolality.

**Discovery of Putative CRE Motifs That Mediate Hyperosmotic Induction of Tilapia Genes**

Regulatory sequences (extended promoter up to intron 1) of the 19 genes that showed transcriptional upregulation of corresponding proteins during hyperosmotic stress were obtained by searching the NCBI genome database for *O. niloticus*. Geneious prime 2022.0.1 (Biomatters) was used to extract and visualize approximately 5 kb of each of these 19 regulatory sequences (Figure 3-2). The criteria by which the regulatory sequences of each gene were selected from the downloaded NCBI sequence database were as follows: 1. Trim up to 5 kb long upstream regulatory region relative to transcription start site (TSS); 2. If the upstream regulatory region is overlapped with another gene body nearby, trim up to that point of the overlap; 3. 5’ untranslated regions

(UTRs) were included (such as exon 1 and intron 1). The last criterion we adopted in this study was to rationalize according to our previous publications elucidating where CREs (in particular, osmotically responsive CREs) are located. We previously identified seven osmotically responsive CREs (OSRE1) that were localized between −232 and +56 relative to the TSS and intron 1 in several osmotically regulated genes [22,110]. Thus, emphasis was placed on including this region for each of the 19 genes.

STREME analysis was performed on the 19 regulatory sequences to find putative hyperosmolality-responsive motifs that are enriched in the set of the regulatory sequences (Figure 3-3). STREME analysis yielded five motifs. For each of the resulting STREME motifs (STREME1 to STREME5), detailed information (e.g., logo, motif sequence, score) is shown in Figure 3-3. These five discovered motifs were then subjected to TOMTOM motif comparison analysis to see if any motif discovered by STREME resembles a previously known TF binding site (Figure 3-3). TOMTOM compares motifs against publicly known TF binding motif databases (e.g., JASPAR) and ranks the motifs in the database to produce an alignment for each significant match. This analysis revealed that STREME1 and STREME2 best match to the Forkhead box protein L1 secondary motif (FoxL1_2nd) and metal response element binding transcription factor 1 secondary motif (Mtf1_2nd), respectively. The other three motifs (STREME3, STREME4, and STREME5), however, yielded no statistically significant match with the cutoff values of $p$-value < 0.001 and q-value < 0.05 (Figure 3-3), indicating that these motifs are perhaps distinct in tilapia compared to the organisms included in the databases used by TOMTOM. Nevertheless, the TOMTOM-driven refinement process allowed prediction of putative transcription factors for two of the five discovered motifs (STREME1 and STREME2).

**Annotating STREM1 Hit Localization on the 19 Regulatory Sequences and Selecting for Candidate Gene Regulatory Regions to be Experimentally Tested**

We chose to focus on the most highly significant motif, STREME1, for further analyses based on the results generated by STREME and TOMTOM. Next, we investigated STREME1 by performing FIMO analysis to scan all 19 regulatory sequences for occurrences of the STREME1 motif. This analysis revealed multiple occurrences in each sequence in total (342), 51 of which were statistically significant at *p*-value < 0.0001 and q-value < 0.01 (Figure 3-4). A complete list of the location of all motifs in each sequence is provided in Supplementary Materials (Table 3-S2).

Due to the highest probability of STREME1 being a functional motif predicted by motif screening, significant occurrences of this motif detected by FIMO were annotated on each of the 19 regulatory sequences to visualize their genomic localization using Geneious Prime software (Figure 3-5). Then, we examined the regions including 1 kb upstream relative to TSS and 5' UTR regions (including exon 1 and intron 1) to determine any enrichment pattern of the STREME1 motif in this region. The rationale for first focusing on this region was that proximal promoters, noncoding exon 1, and intron 1 were previously shown to harbor osmoresponsive CREs, which facilitate transcriptional induction during hyperosmolality[22,110]. Chloride intracellular channel 2 (*clic2*) and uridine phosphorylase 1 (*upp1*) each had three significant occurrences of the STREME1 motif in this region and were selected for experimental validation of the functionality of this motif during hyperosmotic stress. Since we used the genomic sequence of *O. niloticus*, but OmB cells were derived from *O. mossambicus*, these sequences were cloned and re-sequenced from *O. mossambicus* genomic DNA. As expected (the two species are very similar, forming fully functional and fertile hybrids in nature[135,136]), the pairwise identity between *O. mossambicus* and *O. niloticus* sequences for these regulatory sequences was very high—95.4% of 1037 bp for

*clic2* and 96.2% of 1216 bp for *upp1*, respectively—and all STREME1 motifs were conserved (Figure 3-S2).

**Experimental Validation of the Selected Candidate Gene STREME1 Motifs**

The proximal extended promoter sequences of two candidate genes, *clic2* and *upp1*, were PCR amplified and cloned into EGFP-reporter vector (Figure 3-S3) to test their transcriptional activity during hyperosmolality. The comparative transcriptional activities were measured by GFP signals and tile scan images (RFP signal was used to normalize GFP signal) using the fluorescence microscope and subsequently quantified using the processing software installed. The approximate 1 kb proximal extended promoter sequences isolated from *clic2* and *upp1*, were shown to drive transcriptional induction in response to hyperosmolality (Figure 3-6A-left panel, Figure 3-6B-left panel, Figure 3-6D, and Figure 3-6E).

To determine whether STREME1 is responsible for hyperosmotic inducibility of osmoregulated tilapia genes and to what extent it contributes to their regulation in response to hyperosmolality, the STREME1 wildtype and STREME1 mutant forms of each proximal extended promoter region (boxes surrounded by red lines in Figure 3-5) were analyzed using GFP/RFP reporter assay. Hyperosmotic induction of both *clic2* and *upp1* was confirmed using 1kb of their wildtype proximal extended promoter regions. Moreover, when all three STREME motifs were mutated to render them nonfunctional, the hyperosmotic inducibility of both genes was significantly reduced (Figure 3-6). Interestingly, the reduction was about two- to three-fold in both cases. However, since *upp1* hyperosmotic induction was much greater than *clic2* hyperosmotic induction, only mutation of *clic2* STREME1 motifs completely abolished hyperosmotic inducibility of the reporter. In contrast, reporter activity was still significantly higher in hyperosmotic medium for the upp1 mutant construct suggesting that this 1 kb regulatory sequence

contains other osmotically responsive CREs in addition to STREME1. Overall, however, these results represent experimental validation of STREME1 as a novel salinity-responsive CRE in euryhaline tilapia.

## Discussion

### The Role of CREs in Environmental Acclimation of Fish

In the present study, a sequential approach consisting of a multi-step bioinformatics methodology followed by experimental validation of the function of candidate sequences was used to identify a novel CRE (STREME1). Moreover, we predict the corresponding TF required for transcriptional activation of salinity-induced genes via STREME1 in euryhaline tilapia. We hypothesized that transcriptionally coregulated genes encoding hyperosmotically induced proteins have common regulatory elements that control their expression during hyperosmolality. Hyperosmolality-induced proteins in tilapia OmB cells were identified by quantitative proteomics and their transcriptional activation was verified using actinomycin D treatment. Transcriptional regulation in response to environmental cues such as hyperosmotic stress is largely governed by CREs and TFs[137,138]. For example, stress response CREs and stress-induced TFs that respond to a variety of stressors, such as heat shock, oxidative stress, and osmotic stress, have been dissected using the yeast model *Saccharomyces cerevisiae*[139]. Many studies have identified environmentally regulated genes, transcripts, and proteins, but many fewer have focused on the mechanisms by which CREs and TFs regulate mRNA and protein abundances. Nonetheless, some studies have attempted to elucidate environmentally induced transcriptional regulation of fish, including two studies from our lab that have identified osmolality/salinity responsive element 1 (OSRE1) as a CRE necessary for hyperosmotic induction of tilapia inositol monophosphatase 1

(IMPA1), *myo*-inositol phosphate synthase (MIPS), and glutamate synthetase (GS) genes[22,110]. In addition to previous studies, some CREs have been identified in other fish species. *PelB* enhancer (CRE) was identified as a major driver of *Pitx1* gene expression in the developing hind limb in sticklebacks. *Pitx1* encodes a homeodomain TF that controls hind limb development of the fish[140]. In zebrafish, a number of p63 TF binding sites (CREs) are located upstream of epidermal genes (e.g., *dlx3b*, *grhl1*, and *myh9a*) that are regulated as a p63-TF-controlled gene regulatory network[141]. Osmotic stress transcription factor 1 (OSTF1/TSC22D3) and TFIIB as salinity-induced TFs in tilapia whose induction precedes that of osmoregulatory effector genes were previously identified[15,142,143]. OSTF1 has since been confirmed as a rapidly osmoregulated gene in several other species of euryhaline fish[144]. In medaka intestine, OSTF1 mRNA is upregulated along with serum/glucocorticoid regulated kinase (SGK1) [145]. The importance of cis-regulatory elements for adaptive divergence of marine vs. freshwater sticklebacks was emphasized without specifying the specific cis-elements that are involved[146]. Our identification of a functional CRE (STREME1) and its putative transacting factorFoxL1 provides a new specific target for dissecting mechanisms of osmosensory signal transduction in euryhaline fishes.

**Transcriptional Regulation of Genes that Penetrates to Proteins and Phenotypes**

In the present study, we have focused on hyperosmotically regulated proteins to emphasize corresponding genes whose regulation penetrates to phenotypes and take into account frequently observed lack of correlation of inducible mRNA versus protein abundance regulation[126,147–150]. This approach contrasts with many studies on fish that have been performed at the transcriptome level in response to different environmental signals, including changes in osmolality. One study examined the transcriptome profile of gill tissue of euryhaline estuarine goby, Gillichthys mirabilis, exposed to osmotic stress to identify osmotically responsive mRNAs. This

study revealed many effector genes that encode putative osmosensory signaling proteins, including insulin receptor substrate-2 (IRS-2) and insulin-like growth factor binding protein 1 (IGFBP-1)[151]. Another study investigating the liver of spotted sea bass, Lateolabrax maculatus, challenged with salinity stress found 455 differentially expressed genes (DEGs) by RNA-seq, including many involved in cell signaling[152]. Deep sequencing of the gill transcriptome of hybrid tilapia exposed to salinity stress revealed many DEGs with signaling functions, e.g., carbonic anhydrase (CA), aquapoin-1 (AQP-1), and calcium/calmodulin-dependent protein kinase (CaM kinase) II[153]. However, these and many other transcriptomics studies do not identify the mechanism of mRNA abundance regulation, i.e., whether it is transcriptional or posttranscriptional, and they do not demonstrate that mRNA regulation penetrates to the level of proteins to affect phenotype. Our study demonstrates osmoregulatory transcriptional regulation both by analyzing osmotic effects on protein abundance and by utilizing the specific transcription inhibitor actinomycin D[119]. This inhibitor has been used extensively for confirming transcriptional regulation of mRNA abundances, including in fish exposed to environmental stress. For instance, we previously utilized actinomycin D to investigate the mechanism of mRNA induction for OSTF1 in gills of tilapia exposed to hyperosmotic stress[15]. Another study used actinomycin D to show that hyperosmotic OSTF1 induction in gill cells of Japanese eels (Anguilla japonica) is in part due to transcriptional regulation[154].

Because phenotypes of cell lines can change with passage number, we have consistently used a narrow range of passages (20 to 27) of the OmB cell line. Nevertheless, we have previously documented that hyperosmotic stress response phenotypes of the OmB and OmL cell lines do not differ in their response to hyperosmolality and corresponding phenotype when low (passage 10 and 11) and high (passage 63) passages were compared [155].

**STREME1 as a Novel Hyperosmotically Inducible CRE of Euryhaline Tilapia**

Whether the predicted STREME1 motif is necessary for transcriptional regulation of *clic2* and *upp1* genes during hyperosmotic stress was experimentally tested. For both genes, mutagenesis of STREME1 significantly reduced the hyperosmotic inducibility. Intriguingly, STREME1 motif mutagenesis almost completely abolished hyperosmotic inducibility of the 1 kb proximal extended promoter region of *clic2* while that of *upp1* was only partially abolished after mutagenesis despite both regulatory regions being approximately equal in length and containing 3 STREME1 sites each. These data indicate that other CRE/TF binding sites are involved in hyperosmotic induction of *upp1*. For *clic2*, however, STREME1 plays a dominant role for the hyperosmotic activation. Combinatorial transcriptional regulation of *upp1* during hyperosmolality by multiple TFs is consistent with combinatorial regulation of many other genes in a diverse array of contexts as demonstrated in fruit flies, yeast, and mammals[94,156]. This cooperativity of multiple TFs with corresponding binding sites (CREs) has gained much attention as it can explain highly complex spatiotemporal transcriptional regulation[94]. Although combinatorial transcriptional regulation has been mostly studied in model organisms, there are some reports of combinatorial functions of TFs in fish. A study on the molecular mechanism of arterial formation investigating arterial-specific gene regulation in zebrafish has demonstrated that arterial specification is regulated by combinatorial binding of both the Notch and SOXF TFs[157]. Another zebrafish study on the involvement of ETS family TFs in early endothelial specification and differentiation elucidated that four members of this TF family (*fli1, fli1b, ets1, and etsrp*) function in combination with each other to achieve full vascular development, which was confirmed by introducing defective mutants of each gene[158].

Based on previous studies and those of others investigating CREs in several osmoresponsive genes in fish and mammals, respectively, a majority of CREs are localized within proximal promoter regions (within 1 kb upstream relative to TSS) or even intron 1 (5' UTR) [17,18,74,75]. Consequently, we have focused on the approximately 1 kb extended promoter regions of *clic2* and *upp1* for experimental validation. However, other CREs contributing to the overall transcriptional regulation of hyperosmotically inducible genes are likely also involved in a combinatorial manner. For example, in mammals, salinity-responsive enhancers are scattered over a 50 kb region relative to the TSS[46]. Long-range inducible CREs have also been revealed for other contexts in diverse model species [10,21,76]. Thus, although the reporter studies utilized can unambiguously demonstrate that a particular CRE is necessary and contributes to the hyperosmotic regulation, it is not possible to conclude whether it is sufficient even if hyperosmotic induction is completely abolished by mutagenesis as is the case for the *clic2* 1 kb extended promoter.

**Roles of CLIC2 and UPP1 during Hyperosmolality**

Sequences of chloride intracellular channel (CLIC) proteins are highly conserved among vertebrates but individual CLIC family members have multiple distinct cellular functions[159]. CLIC2 is the least studied CLIC family member. A mechanistic study of CLIC2 functions in human cancer tissues demonstrated that, apart from chloride transport, CLIC2 is involved in tight junction formation [160]. Tight junctions are known to be critical for osmoregulation, including in Mozambique tilapia gill epithelium[161]. Therefore, transcriptional regulation of CLIC2 upon hyperosmolality may be a physiological response that contributes not only to cellular osmoregulation but also to integrative osmoregulation at higher levels of organization[162].

The *upp1* gene encodes an enzyme that catalyzes uridine to produce ribose phosphate and uracil[163]. The produced molecule is then utilized as a carbon and energy source or in the process

of nucleotide synthesis[164]. Both uses can facilitate cellular osmoregulation and salinity stress responses of tilapia because substantial amounts of energy are required to cope with stressful conditions[165]. Moreover, it is necessary to produce more nucleotides including uracil (for *de novo* generating RNA molecules) to compensate for reduced nucleotide pools caused by stress-induced DNA and RNA damage[166,167]. The nonspecific nature of such effects of environmental stress on macromolecular damage and the induction of *upp1* during acute heat stress in black rockfish (*Sebastes schlegelii*) supports its role for replenishing building blocks for RNA pools during stress[168]. Moreover, the *upp2* gene of Javanese medaka (*Oryzias javanicus*), was shown to be induced by yet another type of stress, bisphenol A (BPA), which is a potent environmental toxicant, implicating *upp2* in the compensation of BPA chemical toxicant stress[169].

**Other Candidate Binding Sites for Hyperosmolality Inducible TFs**

Five other candidate motifs for hyperosmotically inducible CREs have been identified (STREME2-5) although, unlike STREME1, they have not been experimentally validated in this study. STREME2 was predicted to serve as a putative binding site for Mtf1. Known functional roles of Mtf1 include the activation of metal-induced expression of metallothionein (MT) genes[170,171]. Recently, it has been demonstrated that Mtf1 is involved in stress signaling and iron homeostasis in zebrafish[170], which supports our finding of Mtf1 as a potential TF involved in hyperosmotic stress responses. The other candidate motifs (STREME3, STREME4, and STREME5) identified did not meet the statistical significance threshold for any known TF. It is possible that these motifs are sufficiently different in fish from mammals and other organisms for which comprehensive TF binding motif databases are available. A common binding site shared by multiple STREME motifs is that for Sox TFs although none of the corresponding matches meets

the significance threshold (Table 3-S3). Sox TFs control development, cell survival, and physiological homeostasis[172].

TFs regulate the expression of genes having roles in a variety of environmental contexts through sequence-specific interactions with DNA and their DNA recognition specificity has been regarded as a crucial factor of transcriptional regulatory networks[173,174]. TF binding site databases document the binding preferences of TFs based on curated data from model organisms. Binding preferences of TFs from select model organisms to specific sequences have been extensively examined using protein binding microarray (PBM) technology, which assesses in vitro DNA binding preferences of TFs from yeast, mice, and humans[175–177]. These studies have demonstrated that distinct modes of DNA binding exist for many TFs and different (primary and secondary) motifs can be bound with potentially distinct regulatory functions that depend on the cellular environment. Our results show that the STREME1 motif matches the FoxL1 secondary binding site while it differs from the FoxL1 primary binding site, which is shared with other Fox TFs (GTAAACA). It has been suggested that the secondary binding specificity of FoxL1 has been acquired to permit usage of this TF in multiple contexts for controlling a variety of cellular processes throughout the evolution of transcriptional regulatory networks[178]. Thus, the secondary FoxL1 binding motif may have been favored during the evolution of transcriptional regulatory networks that control hyperosmotic stress responses[179].

**Fox L1 as a Putative Hyperosmotically Inducible TF Binding to STREME1**

An intraspecific comparative genomics approach has allowed for identification of STREME1 and its putative TF, FoxL1, as a CRE/TF duo necessary for the hyperosmotic induction of tilapia genes. The STREME1 motif (AAAACAAAACAMWAAA) contains the core sequence (CAAAACAA) of FoxL1 binding sites in mammals. In mammals, Fox family TFs, including

FoxL1, have been described as important regulators of carcinogenesis[180] and stem cell differentiation[181]. Studying the effect of FoxL1 on the Wnt/β-catenin signaling pathway, Perreault et al. established that FoxL1 inhibits this pathway to deplete β-catenin in the nucleus, which in turn decreases cell proliferation in a *FoxL1*-null mouse model[182]. In contrast, another group demonstrated that FoxL1 can activate the same pathway by promoting the induction of tumor necrosis factor (TNF) related apoptosis-inducing ligand (TRAIL) in cancer cells[183]. Thus, in a mammalian system, FoxL1 TF has been shown to act as either activator or repressor depending on the specific combinatorial context, presumably defined by which other sets of TFs it interacts with. Interestingly, the Wnt/β-catenin pathway has been implicated in osmoregulation in tilapia[126]. In zebrafish, one study suggests that FoxL1 acts as transcriptional repressor of the sonic hedgehog (*shh*) gene, regulating central nervous system development[184]. This finding contrasts to our proposed role of FoxL1 as a transcriptional activator. However, little is known about physiological roles of FoxL1 in environmental stress responses and nothing about its function in the hyperosmotic stress response in fish. Moreover, as outlined above, depending on context, FoxL1 can also act as transcriptional activator. Furthermore, it is possible that in fish, other Fox family TFs bind to STREME 1 even though the STREME1 motif is most similar to the mammalian FoxL1 binding site. The TOMTOM-generated TF candidates identified in our study (Table 3-S3) included not only FoxL1 for the STREME1 and STREME2 motifs, but also FoxK1 for the STREME2 and STREME5 motifs. FoxL1 and FoxK1 binding sites are highly similar, which renders both of these TFs strong candidates for hyperosmotically activated TFs in euryhaline fish.

## Conclusions

Using a bioinformatics approach based on intraspecific comparative genomics, we identified a novel hyperosmotically inducible CRE of euryhaline tilapia, STREME1. STREME1 function during hyperosmotic stress was experimentally validated using reporter assays in combination with site-directed mutagenesis of two different genes (*clic2* and *upp1*). Furthermore, FoxL1 and potentially its close ortholog FoxK1 were identified as candidate TFs that bind to STREME1 and possibly additional CREs (STREME2 and STREME5) in hyperosmotically regulated tilapia genes. This systematic approach consisting of intraspecific comparative genomics and experimental validation represents a powerful complement to widespread RNA-seq studies to identify the mechanisms by which stress-induced genes are regulated during specific environmental contexts.

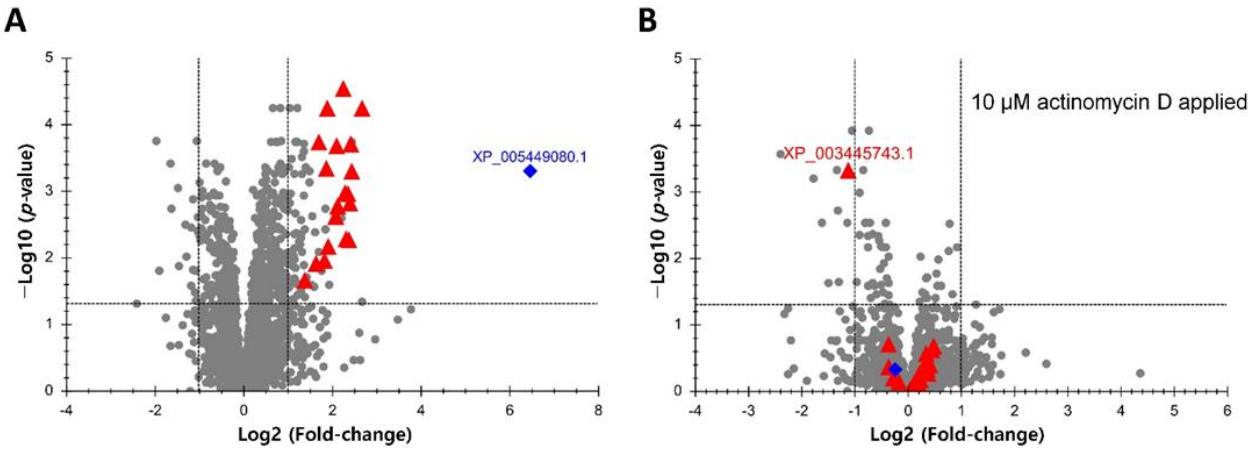


**Figure 3-1.** Relative abundances of 3043 *Oreochromis mossambicus* proteins in OmB cells exposed for 24 h to hyperosmotic stress (650 mOsmol/kg) versus isosmotic media (315 mOsmol/kg). (A) Volcano plot indicating the 19 significantly up-regulated proteins that were selected for comparative sequence analyses and motif searches (red triangles). The x axis displays the fold change of protein abundance in hyperosmotic versus isosmotic medium on a log 2 scale. The y axis displays the negative decadic logarithm of the MSstats-adjusted (multiple testing corrected) *p* value. Inositol monophosphatase (blue diamond) was not included in this set because its FC was much greater than that of the other proteins' and it had been analyzed previously in depth[22]. (B) Volcano plot for the same proteins as shown in panel A except that OmB cells were exposed to hyperosmolality in the presence of 10 µM of the transcriptional inhibitor actinomycin D. Data are based on five replicates for each treatment and control group. For accession numbers of all 19 proteins indicated by red triangles and used for further analyses by motif searching please refer to Figure 3-2.

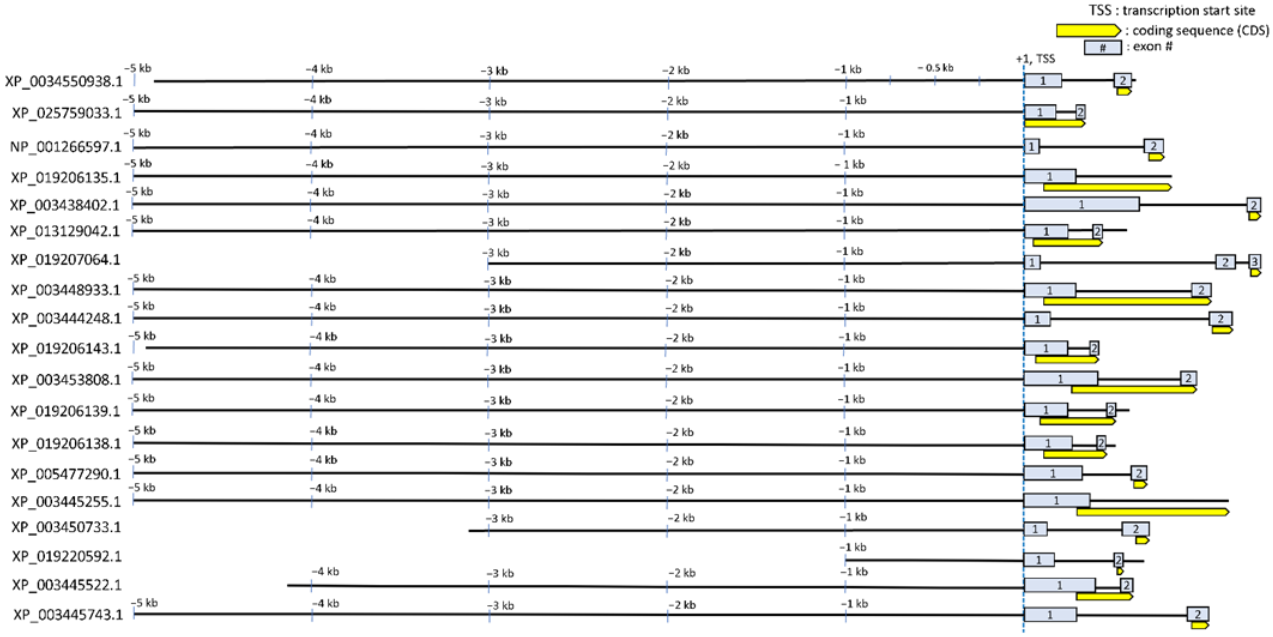



**Figure 3-2.** Schematic genomic landscape of the regulatory sequences of 19 hyperosmotically induced tilapia genes. Each of the regulatory sequences up to 5 kb upstream (some genes have upstream regulatory regions less than 5 kb due to overlapping with gene body of other genes) and 5' UTR (up to CDS) are depicted as black lines. Light grey boxes indicate exons and the yellow boxes with one-sided arrow indicate the coding sequence (CDS) of each gene including intron 1 if

applicable. Each sequence is labeled with the NCBI accession number of the corresponding protein. Relative genomic positions (e.g., −3 kb) from transcription start site (TSS, +1) are presented.

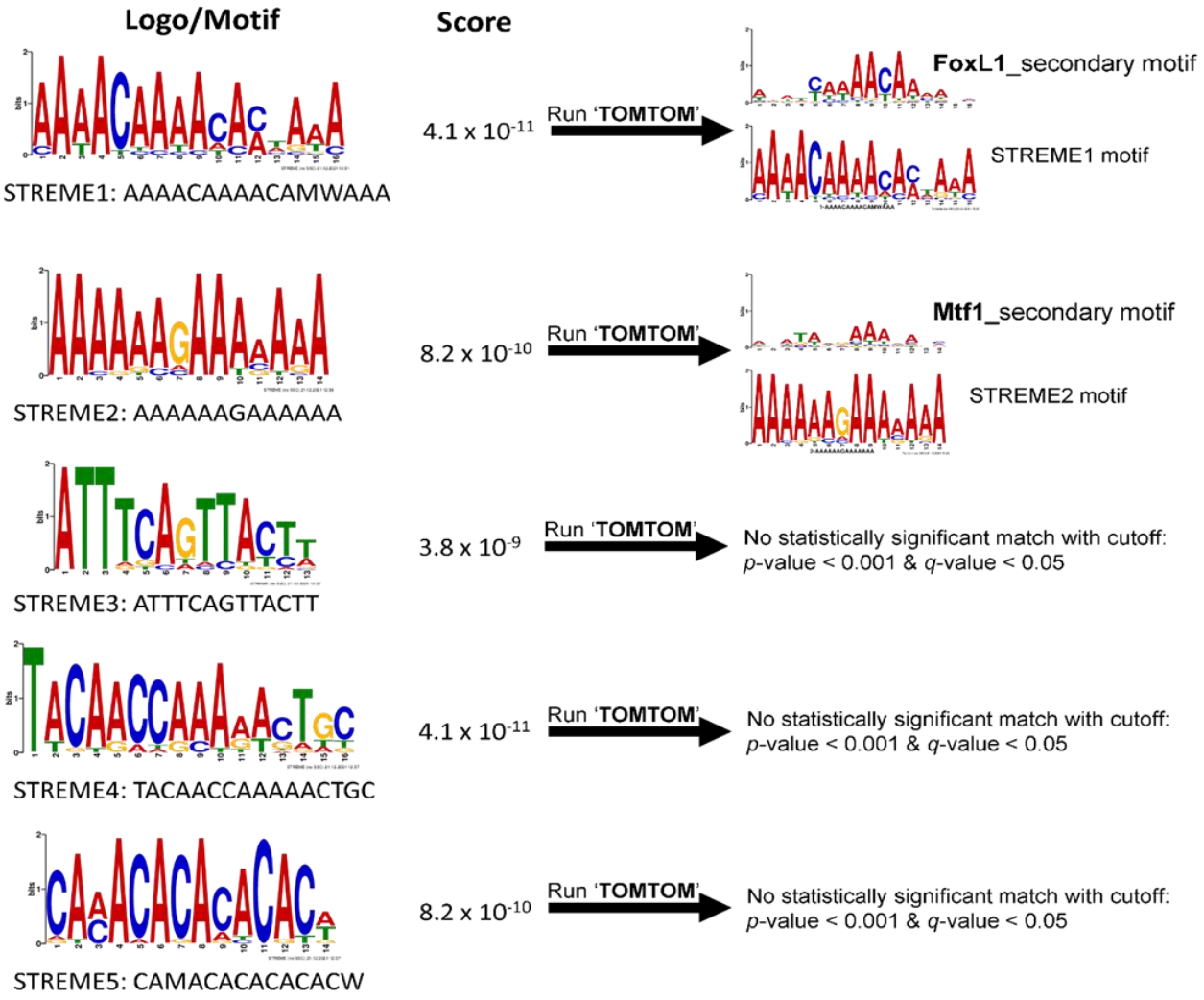
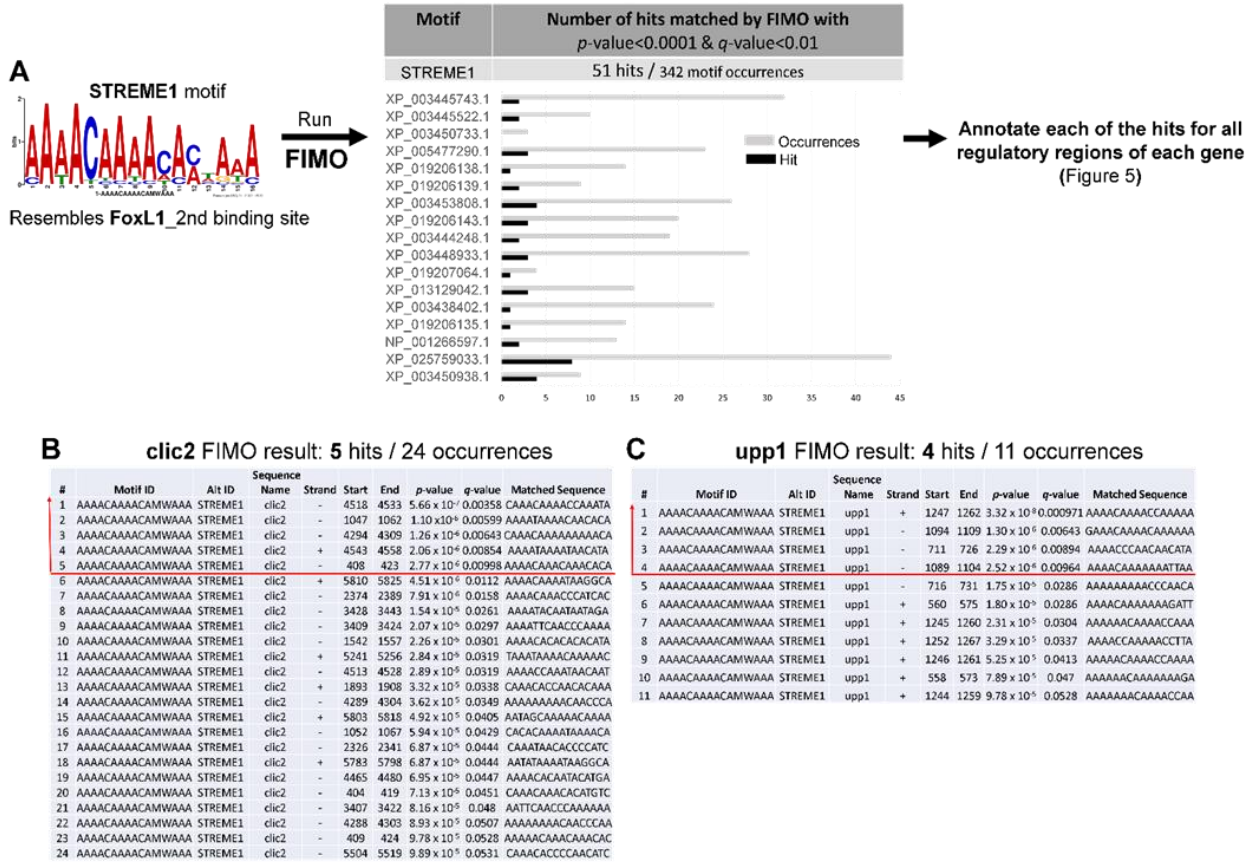



**Figure 3-3.** Five motifs (STREME1 to STREME5) identified by STREME motif discovery tool using 19 hyperosmotically induced tilapia sequences as input. The logo and sequence for each motif is indicated on the left and corresponding STREME score in the center. The result of TOMTOM prediction of known TF binding sites is indicated on the right. On the right are STREME1 motif results indicating a match to the FoxL1 secondary binding motif and STREME2 motif results indicating a match to the Mtf1 secondary motif from TOMTOM search. The other three STREME motifs did not match to any known TF binding motif with the cutoff criteria.




**B** **clic2** FIMO result: **5** hits / 24 occurrences

| # | Motif ID | Alt ID | Sequence Name | Strand | Start | End | *p*-value | *q*-value | Matched Sequence |
|---|---|---|---|---|---|---|---|---|---|
| 1 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 4518 | 4533 | $5.66 \times 10^{-7}$ | 0.00358 | CAAACAAAACCAAATA |
| 2 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 1047 | 1062 | $1.10 \times 10^{-6}$ | 0.00599 | AAAATAAAACAACACA |
| 3 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 4294 | 4309 | $1.26 \times 10^{-6}$ | 0.00643 | CAAACAAAAAAAAACA |
| 4 | AAAACAAAACAMWAAA | STREME1 | clic2 | + | 4543 | 4558 | $2.06 \times 10^{-6}$ | 0.00854 | AAAATAAAATAACATA |
| 5 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 408 | 423 | $2.77 \times 10^{-6}$ | 0.00998 | AAAACAAACAAACACA |
| 6 | AAAACAAAACAMWAAA | STREME1 | clic2 | + | 5810 | 5825 | $4.51 \times 10^{-6}$ | 0.0112 | AAAACAAAATAAGGCA |
| 7 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 2374 | 2389 | $7.91 \times 10^{-6}$ | 0.0158 | AAAACAAACCCATCAC |
| 8 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 3428 | 3443 | $1.54 \times 10^{-5}$ | 0.0261 | AAAATACAATAATAGA |
| 9 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 3409 | 3424 | $2.07 \times 10^{-5}$ | 0.0297 | AAAATTCAACCCAAAA |
| 10 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 1542 | 1557 | $2.26 \times 10^{-5}$ | 0.0301 | AAAACACACACACATA |
| 11 | AAAACAAAACAMWAAA | STREME1 | clic2 | + | 5241 | 5256 | $2.84 \times 10^{-5}$ | 0.0319 | TAAATAAAACAAAAAC |
| 12 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 4513 | 4528 | $2.89 \times 10^{-5}$ | 0.0319 | AAAACCAAATAACAAT |
| 13 | AAAACAAAACAMWAAA | STREME1 | clic2 | + | 1893 | 1908 | $3.32 \times 10^{-5}$ | 0.0338 | CAAACACCAACACAAA |
| 14 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 4289 | 4304 | $3.62 \times 10^{-5}$ | 0.0349 | AAAAAAAAACAACCCA |
| 15 | AAAACAAAACAMWAAA | STREME1 | clic2 | + | 5803 | 5818 | $4.92 \times 10^{-5}$ | 0.0405 | AATAGCAAAAACAAAA |
| 16 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 1052 | 1067 | $5.94 \times 10^{-5}$ | 0.0429 | CACACAAAATAAAACA |
| 17 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 2326 | 2341 | $6.87 \times 10^{-5}$ | 0.0444 | CAAATAACACCCATC |
| 18 | AAAACAAAACAMWAAA | STREME1 | clic2 | + | 5783 | 5798 | $6.87 \times 10^{-5}$ | 0.0444 | AATATAAAATAAGGCA |
| 19 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 4465 | 4480 | $6.95 \times 10^{-5}$ | 0.0447 | AAAACACAATACATGA |
| 20 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 404 | 419 | $7.13 \times 10^{-5}$ | 0.0451 | CAAACAAACACATGTC |
| 21 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 3407 | 3422 | $8.16 \times 10^{-5}$ | 0.048 | AATTCAACCCAAAAAA |
| 22 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 4288 | 4303 | $8.93 \times 10^{-5}$ | 0.0507 | AAAAAAAACAACCCAA |
| 23 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 409 | 424 | $9.78 \times 10^{-5}$ | 0.0528 | AAAAACAAACAAACAC |
| 24 | AAAACAAAACAMWAAA | STREME1 | clic2 | - | 5504 | 5519 | $9.89 \times 10^{-5}$ | 0.0531 | CAAACACCCCAACATC |

**C** **upp1** FIMO result: **4** hits / 11 occurrences

| # | Motif ID | Alt ID | Sequence Name | Strand | Start | End | *p*-value | *q*-value | Matched Sequence |
|---|---|---|---|---|---|---|---|---|---|
| 1 | AAAACAAAACAMWAAA | STREME1 | upp1 | + | 1247 | 1262 | $3.32 \times 10^{-8}$ | 0.000971 | AAAACAAAACCAAAAA |
| 2 | AAAACAAAACAMWAAA | STREME1 | upp1 | - | 1094 | 1109 | $1.30 \times 10^{-6}$ | 0.00643 | GAAACAAAACAAAAAA |
| 3 | AAAACAAAACAMWAAA | STREME1 | upp1 | - | 711 | 726 | $2.29 \times 10^{-6}$ | 0.00894 | AAAACCCAACAACATA |
| 4 | AAAACAAAACAMWAAA | STREME1 | upp1 | - | 1089 | 1104 | $2.52 \times 10^{-6}$ | 0.00964 | AAAACAAAAAAATTAA |
| 5 | AAAACAAAACAMWAAA | STREME1 | upp1 | - | 716 | 731 | $1.75 \times 10^{-5}$ | 0.0286 | AAAAAAAAACCCAACA |
| 6 | AAAACAAAACAMWAAA | STREME1 | upp1 | + | 560 | 575 | $1.80 \times 10^{-5}$ | 0.0286 | AAAACAAAAAAAGATT |
| 7 | AAAACAAAACAMWAAA | STREME1 | upp1 | + | 1245 | 1260 | $2.31 \times 10^{-5}$ | 0.0304 | AAAAAACAAAACCAAA |
| 8 | AAAACAAAACAMWAAA | STREME1 | upp1 | + | 1252 | 1267 | $3.29 \times 10^{-5}$ | 0.0337 | AAAACCAAAAACCTTA |
| 9 | AAAACAAAACAMWAAA | STREME1 | upp1 | + | 1246 | 1261 | $5.25 \times 10^{-5}$ | 0.0413 | AAAAACAAAACCAAAA |
| 10 | AAAACAAAACAMWAAA | STREME1 | upp1 | + | 558 | 573 | $7.89 \times 10^{-5}$ | 0.047 | AAAAAACAAAAAAAGA |
| 11 | AAAACAAAACAMWAAA | STREME1 | upp1 | + | 1244 | 1259 | $9.78 \times 10^{-5}$ | 0.0528 | AAAAAAACAAAACCAA |

**Figure 3-4.** STREME1 motif scanning for occurrences in the regulatory sequences of 19 hyperosmotically induced tilapia genes. (A) FIMO was performed using the STREME1 motif to find all occurrences in the regulatory sequences. The sequence information for all identified STREME1 motifs in the regulatory sequences resulting from FIMO, such as whether STREME1 is found in sense or antisense strand; start and end position; *p*-value, and q-value are provided in supplementary Table 3-S2. Fifty-one significant hits out of 342 total STREME1 occurrences were identified throughout all 19 regulatory sequences and were screened by two statistical cut off values using FIMO default *p*-value (0.0001) and q-value (0.01) thresholds. The number of occurrences of significant hits (black bars) and total STREME1 occurrences (grey bars) is illustrated for each of gene except for *clic2* and *upp1*, which were chosen for further experimental validation. Some occurrences of these 51 hits represent overlapping sequences, which were consolidated into a single motif in Figure 3-5. Detailed information about STREME1 motifs identified in *clic2* (B) and *upp1* (C) regulatory sequences is shown in panels B and C with significance thresholds indicated by a red line.

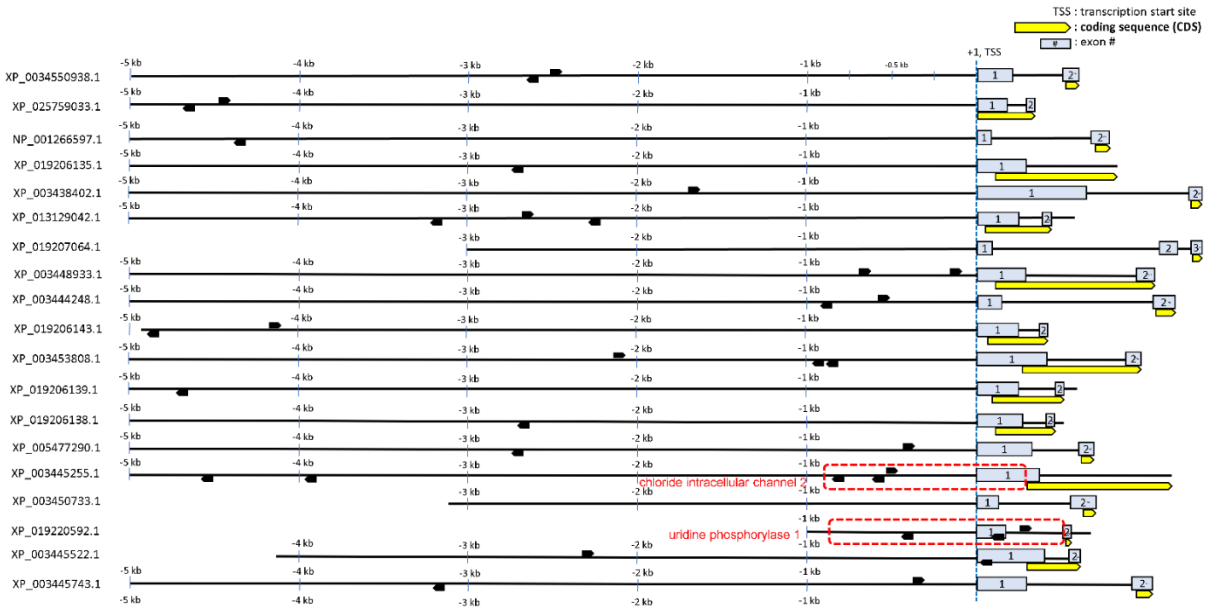



**Figure 3-5.** Annotation of significant STREME1 motif occurrences in regulatory sequences of 19 hyperosmotically induced tilapia genes. Each of the regulatory sequences analyzed by motif discovery is depicted as a black line. Light grey boxes indicate exons and the yellow boxes with one-sided arrow indicate coding sequence (CDS) including intron 1 if applicable. Significant hit STREME1 motifs (analyzed in figure 3-4) are displayed using black bars with one-sided arrow indicating orientation. The proximal extended promoter sequences (approximately 1 kb upstream relative to TSS and 5'-UTR) for *clic2* and *upp1* that were used for experimental validation of STREME1 motifs are boxed by a dashed red line.

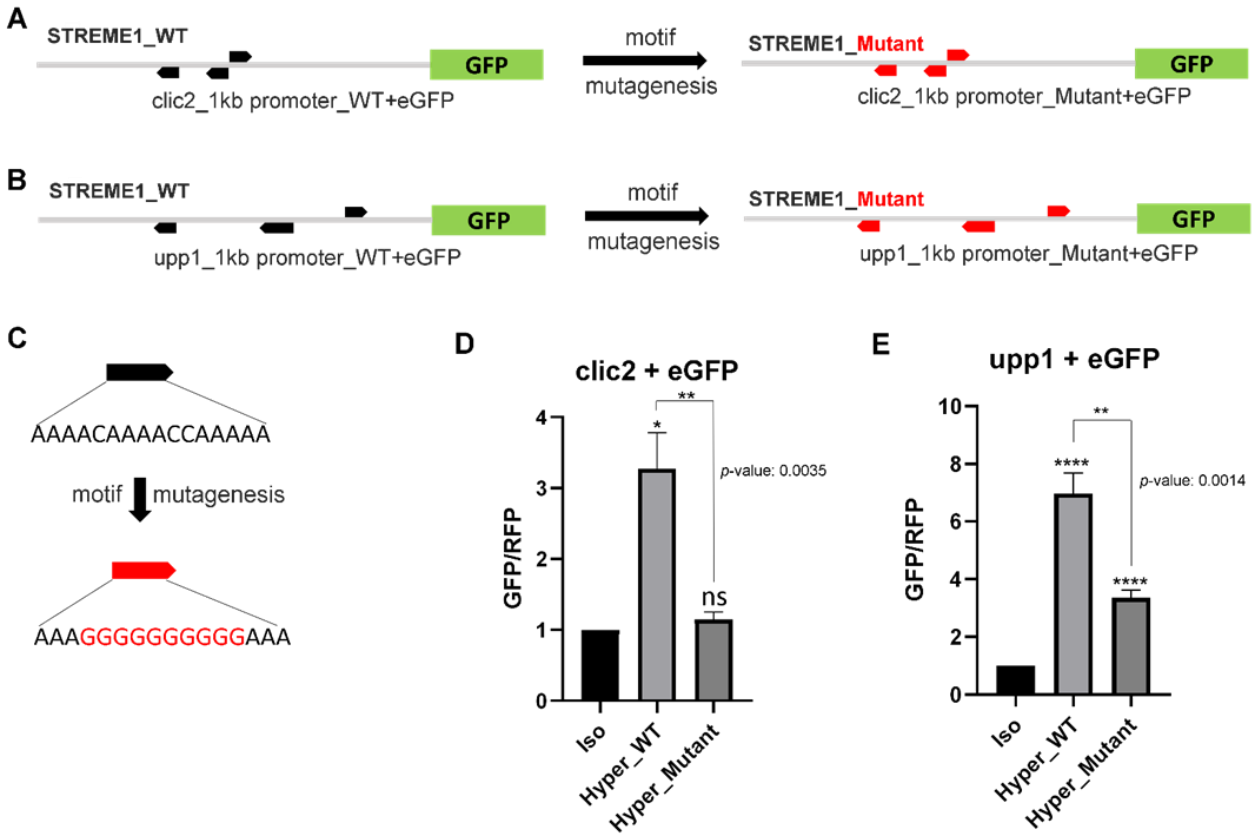



**Figure 3-6.** Experimental validation of STREME1 using GFP/RFP reporter assay and motif mutagenesis. Motif mutagenesis was used to replace the original STREME1 (black arrows) with a nonfunctional sequence (red arrows) by changing the core nucleotides with nucleotides not contained in that core region for chloride intracellular channel 2 (*clic2*, A) and uridine phosphorylase (*upp1*, B). (C) STREME1 motif mutagenesis strategy indicates the sequence difference between wild type (WT, black arrow) and mutant (red arrow) motifs. The transcriptional activities of the proximal regulatory regions from *clic2* (D) and *upp1* (E) during hyperosmolality (compared to isosmotic control medium) were measured by GFP signal (normalized with RFP control) using expression vector systems. t-test was used to calculate statistical significance yielding *p*-values. Iso: Isosmotic control, Hyper_WT: wild type regulatory sequence treated with hyperosmotic treatment, Hyper_Mutant: mutant regulatory sequence treated with hyperosmotic treatment, n = 5 (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$)

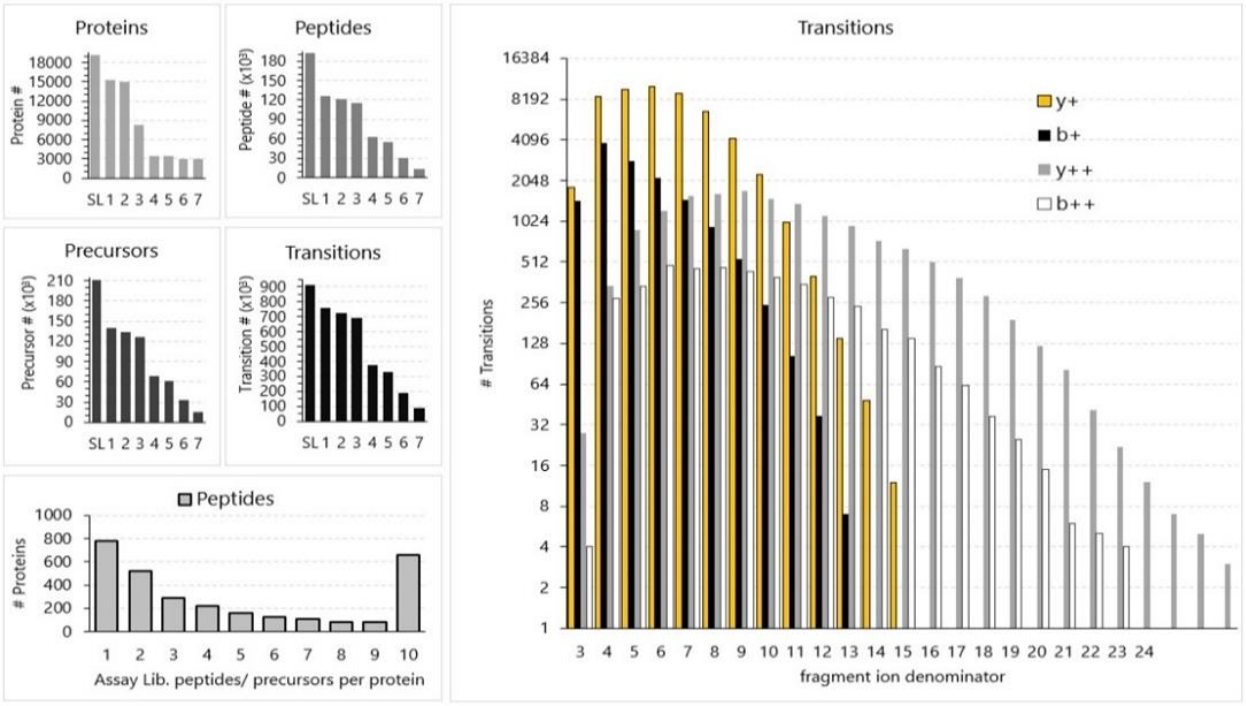
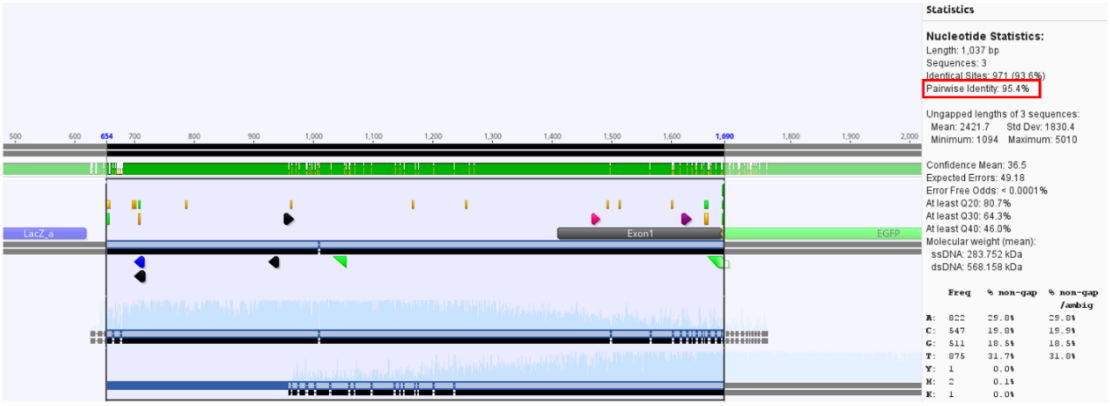
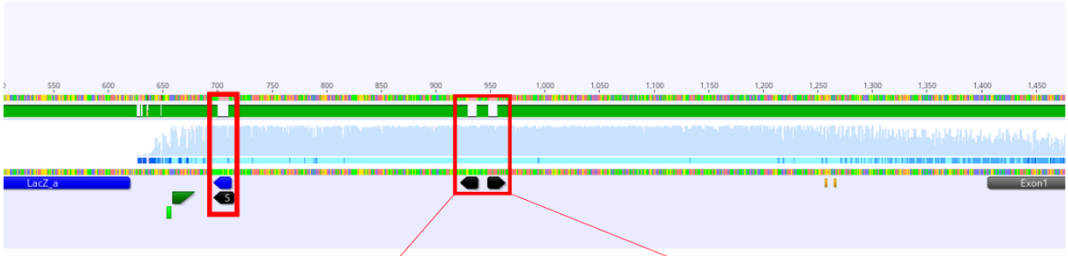



**Figure 3-S1.** DIA assay library of *Oreochromis mossambicus* OmB cells. The initial spectral library (SL) contains over 18,000 proteins, 200,000 precursors, and almost 1 million transitions. Seven quality control (QC) filters[185] were applied to create the DIA assay library containing 3043 unique proteins, 15,211 precursors (peptides), and 87,184 diagnostic transitions. Most proteins are represented by at least 2 diagnostic peptides. The remainder (25%) was identified by at least 2 unique peptides but only 1 remains after applying all DIA QC filters.

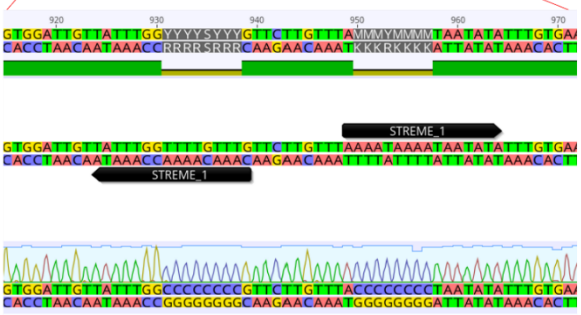
A. chloride intracellular channel 2 (clic2)_Wild Type



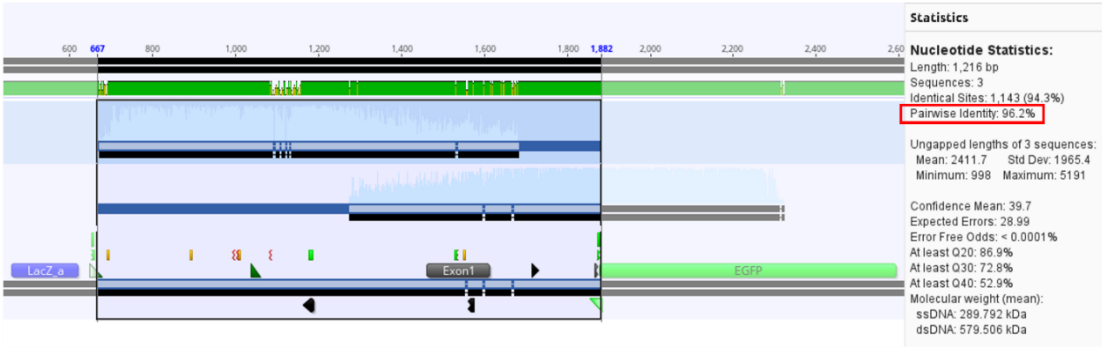

B. chloride intracellular channel 2 (clic2) _STREME1 motif mutant



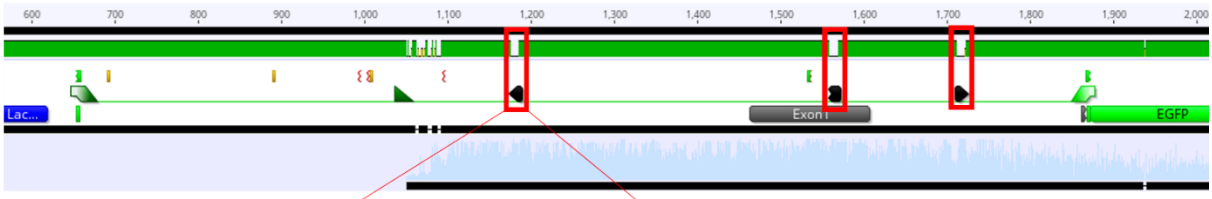

C. clic2_STREME1 motif mutagenesis representative result



D. uridine phosphorylase 1 (upp1)_Wild Type




E. uridine phosphorylase 1 (upp1)_STREME1 motif mutant




F. upp1_STREME1 motif mutagenesis representative result


**Figure 3-S2.** Sequence information of wild-type and mutant versions of approximately 1kb long proximal extended promoter regions of *clic2* and *upp1* that were used for quantitative GFP reporter assays. (A) The pairwise identity of *clic2* sequences between *Oreochromis mossambicus* and *Oreochromis niloticus* (reference genome) is 95.4 % as marked by red box. (B) STREME1 motifs found in the 1 kb promoter region of *clic2* were mutated using site-directed mutagenesis and confirmed by Sanger sequencing as indicated by red boxes. (C) A representative result of mutagenesis was shown with distinct chromatograms between wild-type reference (upper, STREME1 annotated) and mutant form (bottom). (D) The pairwise identity of *upp1* sequences between *Oreochromis mossambicus* and *Oreochromis niloticus* (reference genome) is 96.2 % as marked by red box. (B) STREME1 motifs found in the 1 kb promoter region of *upp1* were mutated using site-directed mutagenesis and confirmed by Sanger sequencing as indicated by red boxes.

(C) A representative result of mutagenesis was shown with distinct chromatograms between wild-type reference (upper, STREME1 annotated) and mutant form (bottom).

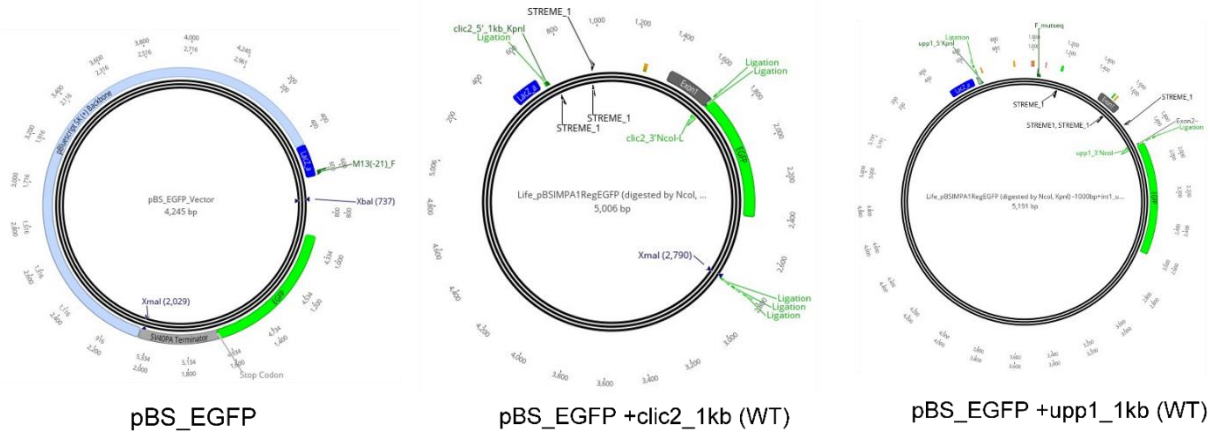



**Figure 3-S3.** The schematic maps of the plasmids constructed for testing two candidate regions of *clic2* and *upp1*. pBS_EGFP (basic backbone vector) is displayed at the left side, clic2_1kb_proximal extended promoter cloned pBS_EGFP (pBS_EGFP+clic2_1kb (WT)) at center, and upp1_1kb_proximal extended promoter cloned pBS_EGFP (pBS_EGFP+upp1_1kb (WT)) at the right side.

| Primer Purpose | Primer Name | Primer Sequence |
|---|---|---|
| Wildtype PCR and Cloning | clic2_5'KpnI | CCCCCGGTACCCCTAATCAGAAGGATGCAGA |
| | clic2_3'NcoI | CCCCCCCATGGTTTATCCCCCAAAAGTTTTTGTTT |
| | upp1_5'KpnI | CCCCCGGTACCGTCCTGAACTGAAGTGATG |
| | upp1_3'NcoI | CCCCCCCATGGTTCCTGCGAACAGCCGAA |
| | clic2_5'KpnI | CCCCCGGTACCCCTAATCAGAAGGATGCAGA |
| Mutant PCR and Cloning | clic2_1p_3'overlapped | GCATTAGCAAAGGGGGGGGGGAAAAAACCAAGCTGAAGTTCATTA |
| | clic2_2p_5'overlapped | GCTTGGTTTTTTCCCCCCCCCCTTTGCTAATGCTATTGTTTTCGAA |
| | clic2_2p_3'overlapped | GCATTAGCAAAGGGGGGGGGGAAAAAACCAAGCTGAAGTTCATTA |
| | clic2_3p_5'overlapped | GCTTGGTTTTTTCCCCCCCCCCTTTGCTAATGCTATTGTTTTCGAA |
| | upp1_1p_3'overlapped | AGTGCCGGAATAAAAAAAAGGGGGGGGGGATATAGCTGGC |
| | upp1_2p_5'overlapped | CCAAGGCCAGCTATATCCCCCCCCCCTTTTTTTTATTCCG |
| | upp1_2p_3'overlapped | CACGAAACCAAGAAGGGGGGGGGGAAAATTAAAAAGAAGATACAGAT |
| | upp1_3p_5'overlapped | TATCTTCTTTTTAATTTTCCCCCCCCCCTTCTTGGTTTCGTGAGTC |
| | upp1_3p_3'overlapped | AGATCTGTACAATTAAGGTTTCCCCCCCCCCTTTTTTAAAATTTAAAAC |
| | upp1_4p_5'overlapped | TTAAATTTTAAAAAAGGGGGGGGGGAAACCTTAATTGTACAGATCTGAT |
| Sequencing | M13_Forward | TGTAAAACGACGGCCAGT |
| | GFP_R5 | TTCTTCTGCTTGTCGGCCAT |

**Table 3-S1.** Primer sequence information used in this study

Additional Supplementary Materials (Table 3-S2: STREME1_FIMO search result of the 19 regulatory regions for both occurrences and hits; Table 3-S3: Lists of TFs for STREME motifs searched by TOMTOM) for this chapter can be found online at: https://www.mdpi.com/2075-1729/12/6/787#supplementary

## CHAPTER 4

## Removal of evolutionarily conserved functional MYC domains in a tilapia cell line using a vector-based CRISPR/Cas9 system

*First author, planned to submit to Transgenic Research, December 2022*


### Abstract

MYC transcription factors have critical roles in facilitating a variety of cellular functions that have been highly conserved among species during evolution. However, despite circumstantial evidence for an involvement of MYC in animal osmoregulation, mechanistic links between MYC function and osmoregulation are missing. Mozambique tilapia (*Oreochromis mossambicus*) represents an excellent model system to study these links because it is highly euryhaline and highly tolerant to osmotic (salinity) stress at both the whole organism and cellular levels of biological organization. Here, we utilize an *O. mossambicus* brain cell line (OmB cells) and an optimized vector-based CRISPR/Cas9 system to functionally knockout (ko) MYC in the tilapia genome and to establish causal links between MYC and cell function, including cellular osmoregulation. A cell isolation and dilution strategy yielded polyclonal MYC (*myca*) ko cell pools with low genetic variability and high gene editing efficiencies (as high as 98.2 %). Further isolation and dilution of cells from these pools yielded a monoclonal myca ko cell line harboring a 1-bp deletion that caused a frameshift mutation. This frameshift functionally inactivated the transcriptional regulatory and DNA-binding domains predicted by bioinformatics and structural analyses. Both the polyclonal and monoclonal myca ko cell lines were viable, propagated well in standard medium, and differed from wildtype cells in morphology. As such, they represent a new tool for causally linking myca to cellular osmoregulation and other cell functions.

**Keywords**

evolutionarily conserved domain, MYC, transcription factor, osmoregulation, CRISPR/Cas9, clonal cell line, tilapia

## Introduction

MYC family genes encode class III basic helix-turn-helix (bHLH) transcription factors (TFs) that have been evolutionarily conserved in many animal species for at least 400 million years [186,187]. They have canonical roles in cell proliferation, differentiation, and apoptosis and are characterized by the presence of a leucine zipper (LZ) adjacent to the bHLH domain[186,188]. The bHLH and LZ domains are common functional motifs for DNA binding and dimerization that are often also found in other TF families, which indicates that these domains arose from an evolutionarily ancient ancestral TF. MYC TFs also harbor a less common motif, a transcriptional regulatory domain, that is less evolutionarily conserved across species than the DNA-binding and dimerization domains[189]. This transcriptional regulatory domain was shown to strongly control transformation of rat embryo fibroblasts as evidenced by using a domain deletion mutant[190]. Moreover, it has been suggested that deletion of any of the conserved domains outlined above can significantly impair MYC function[191]. Several lines of indirect evidence suggest that MYC TFs are involved in governing cellular osmoregulation[192]. For example, a recent study has revealed that the *myo*-inositol biosynthesis (MIB) pathway of euryhaline turbot (*Scophthalmus maximus*) is positively regulated by MYC, which was demonstrated using a RNAi-mediated knockdown approach[193]. MYC has also been reported to directly modulate responses to abiotic stressors, including salinity stress, in plants (e.g., *Arabidopsis thaliana*[194,195]). This modulation is mediated via key hormonal signaling pathways important for plant salinity tolerance[196]. Such salinity stress-induced, non-canonical roles of MYC TFs in plants were also indicated by an *in-*

*silico* prediction study using bread wheat (*Triticum aestivum*)[197]. Moreover, microarray analyses of MYC ko rat cell lines has revealed inositol monophosphatase (IMPA, A2 isoform), a key enzyme in the *myo*-inositol biosynthesis (MIB) compatible osmolyte pathway, as a target gene of MYC[198]. These studies suggest that MYC TFs are important for controlling osmoregulatory mechanisms in eukaryotes.

The recent revolution of gene targeting approaches by implementing CRISPR/Cas based methodologies has enabled highly accurate and efficient genome editing that is more easily employed than older gene targeting methods such as TALENs or ZFNs, which require covalent linkage of a specific DNA binding domain to a nuclease[28,29,199]. This innovative system was initially adopted from bacteria and archaea, in which it has evolved as a pathogen nucleic acid-targeting defense mechanism that confers resistance to viral infection[200,201]. The simplicity and high efficiency of the CRISPR/Cas9 system renders it convenient, cost-effective, and multimodal tool for gene editing in a variety of organisms[202]. Most studies using this system to date have focused on mammalian models such as human[203] and mouse[204] both *in vivo* and *in vitro.* They have demonstrated the power of ko models for functional studies aimed at causally linking genotypes and phenotypes[205–207]. In contrast to the well-established mammalian models, CRISPR/Cas9 approaches have been used to a lesser extent with lower vertebrates such as fish, even though numerous studies have shown that this gene targeting system can be successfully utilized to genetically modify aquaculture fish species *in vivo*[208–211]. For instance, double-allelic ko mutations were introduced in Atlantic salmon (*Salmo salar*) to alter pigmentation [208], and Insulin-like Growth Factor Binding Protein-2b (IGF-BP2b) was targeted for ko in rainbow trout (*Oncorhynchus mykiss*)[209]. Examples for *in vivo* gene targeting in fish also

include zebrafish (*Danio rerio*) where all-in-on CRISPR/Cas9 components were injected into fertilized one-cell stage embryos to generate ko mutants[211].

The first fish cell line that was genetically modified by CRISPR/Cas9 technology was reported for Chinook salmon (*Oncorhynchus tshawytscha*)[212]. Chinook salmon cell lines were also used to demonstrate the functionality of a vector-based expression system[213], as well as to optimize lentivirus-mediated infection for efficient delivery of recombinant DNA into host cells [214]. A vector-based CRISPR/Cas9 platform for tilapia cell lines was recently optimized and established[31], being the first study to enable *in vitro* gene targeting in euryhaline tilapia cells. This vector-based *in vitro* approach differs from the *in vivo* approach used for whole tilapia, which is based on microinjection of gRNA and Cas9 mRNA or protein into Nile tilapia (*Oreochromis niloticus*) fertilized eggs[215].

Recent CRISPR/Cas approaches aim to obtain highly precise and consistent ko models that are characterized by very high gene editing efficiency and/or clonality to exclude potentially confounding factors arising from heterogeneity of ko cells[216–218]. Although highly heterogeneous, pooled ko cell populations with high mutational efficiency (about 80 % or above) are routinely used for short-term loss-of-function studies, interference arising from expression of wildtype or variable mutant proteins remains a concern[219]. Therefore, recent efforts have been geared towards generating homozygous clonal ko cell models to ensure that protein function is completely disrupted and resulting mutant (fragment) proteins that cause effects, are consistent. Several previous studies have reported successful production of clonal ko cell lines for some fish species. For example, Liu et al. have generated a Japanese medaka (*Oryzias latipes*) ko cell line using RNP transfection for CRISPR/Cas9 gene targeting, followed by isolating a ko clonal cell line having a 9-nt deletion in the *sytl5* gene starting with an initial ko cell pool showing 50 % gene

editing efficiency[220]. Furthermore, a modified Chinook Salmon (*Oncorhynchus tshawytscha*) cell line that stably expresses Cas9 protein has been used to generate a monoclonal *stat2* ko cell line harboring a 2-nt frame-shift deletion in *stat2*[221]. Such recent efforts to either generate edited pools (polyclonal) or isolate clonal lines (monoclonal) of CRIPSR/Cas9 gene edited cells can be expanded beyond canonical model species of fish to enable broad comparative and evolutionary studies[222,223].

In this study, genetically engineered polyclonal and monoclonal tilapia cell lines were generated to facilitate studies of the cellular functions of the MYC TF, specifically its role in osmoregulation. A strategy utilizing a DNA vector based CRISPR/Cas9 system followed by serial dilution of mutant cells for efficiently isolating clonal ko cell lines was applied. We present the first successful report of applying targeted gene-editing in combination with serial dilution of a heterogeneous cell population to generate low genetic variability polyclonal and monoclonal tilapia cell lines to enable future functional analyses for assessment of causal genotype-phenotype links.

## Materials and Methods

### Cell culture

A tilapia Cas9-OmB cell line[31] previously generated in our lab was used in this study. The genomic presence and expression of Cas9 transgene was verified by an array of PCRs targeting transgene amplicons using both genomic DNA (gDNA) and complementary DNA (cDNA). Cas9-OmB cell working stock (passage 40 of the original OmB cell line[224]; P40) was thawed and maintained at ambient $CO_2$ and 26 ˚C in L-15 medium (Hyclone, SH30910.03) containing 10% (vol/vol) fetal bovine serum (FBS, Gibco, 11415-064), 1 % (vol/vol) Penicillin-

Streptomycin (Sigma-Aldrich, P4333). When culture plates reached a confluency of 80-90 %, cells were passaged (at 3–4-day intervals) using a 1:5 splitting ratio. For applying hyperosmotic stress to cells, hyperosmotic (650 mOsmol/ kg) media was prepared using hypersaline stock solution (osmolality 2,820 mOsmol/kg). This stock solution was made by adding an appropriate amount of sodium chloride (NaCl) to regular isosmotic (310 mOsmol/kg) L-15 medium. The hypersaline stock solution was then diluted with isosmotic medium to obtain hyperosmotic medium of 650 mOsmol/kg. Medium osmolality was always confirmed using a freezing point micro-osmometer (Advanced Instruments).

***MYC functional domain annotation***

To identify the domains that are important for tilapia MYC function and ensure that all functionally important domains were inactivated by CRISPR/Cas9 gene targeting, the tilapia *myca* gene (Gene ID: 100689989) was identified in the NCBI reference genome sequence for *O. niloticus* (NC_031974.2) along with the corresponding mRNA (XM_005448983.4) and protein (XP_005449040.1) sequences. The 432 amino acid sequence of tilapia MYC TF was then used to annotate functional domains using InterPro version 90.0[225] (EMBL-EBI). To further identify the evolutionarily most highly conserved regions in these functional domains, the annotated tilapia MYC protein sequence was aligned to MYC sequences from 14 other vertebrate species using Geneious bioinformatics software (Biomatters, https://www.geneious.com).

***Generation of sgRNA vectors***

The *O. niloticus* reference genome deposited at NCBI was used to derive the coding sequence (CDS) for tilapia *myca*, the MYC proto-oncogene bHLH transcription factor a (Gene ID: 100689989). A *myca* CDS region spanning exons 2 and 3 was submitted to CRISPOR[43] version

5.01 to design small guide RNAs (sgRNAs) for efficient gene targeting of *myca*. In addition to using CRISPOR, the tilapia *myca* gene was also analyzed with the CRISPR Knockout Guide Design tool provided by SYNTHEGO (https://design.synthego.com/#/) to design sgRNAs using different algorithms. CRISPOR and SYNTHEGO tools both support convenient tilapia sgRNA design by providing an integrated reference genome of *O. niloticus* (Ensembl 76-Orenil1.0) for calculating off-target effects of sgRNAs by comparison to whole genome sequences. This aspect of sgRNA design is critical for optimizing specificity. To further validate the selection of the best possible *myca* sgRNAs, their potential off-targets effects were also manually evaluated using the NCBI reference genome for *O. niloticus*. Sequences for sgRNA1, sgRNA2, and sgRNA3 were searched against nucleotide sequences using Blastn limited to highly similar sequences (megablast) and restricted to entries associated with the organism "*Oreochromis niloticus*" (taxid:8128). No off-target genes were identified that matched any of the three top scoring *myca* sgRNAs suggested by CRISPOR and SYNTHEGO tools. These top three sgRNAs were then cloned into an optimized tilapia sgRNA expression vector as describe previously[32]. Complementary oligonucleotides (Eurofins Genomics) comprising each sgRNAs forward and reverse sequences (Table 4-S1) were annealed to generate a ClaI restriction site at the 5' end a XbaI restriction site at the 3' end. The annealed oligonucleotide was then ligated into ClaI (New England BioLabs) and XbaI (New England BioLabs) double-digested TU6m-gRNAscaffHygroR vector[32] using T4 DNA ligase (Promega). The resulting sgRNA expression vectors for *myca* sgRNAs 1 – 3 were sequenced (sgRNA_seqP1, Table 4-S1) to confirm successful insertion of sgRNA target sequences.

**Transfection and antibiotic-resistance selection of tilapia Cas9-OmB cells**

For each well of a six-well cell culture plate, two micrograms of TU6m-gRNAscaffHygroR vector containing either *myca* sgRNA1, sgRNA2, or sgRNA3 were added to 200 µL Opti-MEM I

Reduced Serum Medium (Gibco) and 6 µL ViaFect reagent (Promega) to initiate the formation of transfection complexes. Stabilized transfection complexes yielded after 15 min incubation were then applied to 80 % confluent Cas9-OmB cells (P43) by adding the transfection complex solution evenly without any medium change. After 48 h, all medium was removed from transfected cells and a non-transfected control and replaced with 2 mL of L-15 medium containing selection media containing 500 µg/ml hygromycin B (Invitrogen, 10687-010). Transfected wells were maintained in selection medium until one day after all cells were detached from the surface of the non-transfected control well. Half of the wells that were transfected with each *myca* sgRNA were then used for analyzing *myca* ko efficiency. This was done as previously described[31]. Briefly, medium was removed and cells surviving hygromycin B selection were rinsed with Dulbecco's phosphate buffered saline (DPBS, Gibco, 14190-144). They were then scraped from the surface of the well into fresh 0.5 mL DPBS, transferred to a 1.5 mL microcentrifuge tube, and centrifuged for 5 min at 18,000 g. After removal of supernatant, cell pellets were lysed in 20 µl of 25 mM NaOH by incubation at 95℃ for 15 min followed by addition of 50 µl of 40 mM Tris–HCl. The resulting solution containing extracted template DNA was used directly for PCR to generate *myca* test amplicons for analysis of mutational efficiency.

**Limiting dilution strategy**

After selection of hygromycin B-resistant cells containing TU6m-gRNAscaffHygroR vector expressing either *myca* sgRNA1, sgRNA2, or sgRNA3, the genetic heterogeneity of mutated, selected cells was serially reduced using a limiting dilution strategy to generate more homogeneous ko cell lines. The protocol for this strategy was adapted from a previous publication [216] and public protocols (https://www.synthego.com/resources/Limiting-Dilution-&-Clonal-Expansion-Protocol, https://www.addgene.org/protocols/limiting-dilution/) and modified as

follows. Selected ko cells were allowed to recover for 14 days in culture to reach 20-30 % confluency. They were then harvested as a single cell suspension, counted with a hemocytometer (Hausser Scientific), and diluted to an average concentration of one cell per well before plating into a 24-well cell culture plate. The wells were visually screened periodically to track cells forming colonies using an inverted microscope (DMi1, Leica) for 14 days. When colonies reached 60-70 % confluency, they were harvested and split evenly into two new wells of a six-well plate. One well was used for continuous expansion and the other for genotyping the *myca* mutation in the corresponding cell population.

Another round of serial dilution was performed after genotyping the cell populations resulting from the first round of dilution. Further limiting dilution for isolating a single clonal *myca* ko cell line was performed by splitting the most promising cell population from dilution round 1 (sgRNA1-colony#3, see results) into a 96-well plate at an average density of one cell per well. Cell density was determined by counting cells in a single cell suspension after harvest with hemocytometer (Hausser Scientific). The cell suspension was then diluted to 5 cells per mL medium. Each well received 500 µL of this cell suspension such that the average seeding density was 0.5 cells/well. Seeding an average of 0.5 cells/well ensures that some wells receive a single cell, while minimizing the likelihood that any well receives more than one cell. Cells were maintained for 14 days to track the wells containing a single clonal cell by regular inspection with a microscope (DMi1, Leica).

***Genotyping***

Genomic DNA was extracted using the PureLink Genomic DNA mini Kit (Invitrogen) following manufacturer instructions. The test amplicon spanning the targeted region of *myca* was PCR-amplified and purified using appropriate primers (Primer pair: myca_TideF1 and

myca_TideR2, Table 4-S1). Sanger sequencing was carried out at the UC Davis DNA Sequencing Facility using the same primers as those used for PCR. DNA sequences and chromatograms were then analyzed with TIDE (Tracking of Indels by Decomposition; shinyapps.datacurators.nl/tide/) and ICE (inference of CRISPR Edits; SYNTHEGO - CRISPR Performance Analysis) to obtain quantitative overall target gene editing efficiency and indel mutation frequencies from each mono- or poly-clonal *myca* ko cell line. The PCR amplicon using genomic DNA extracted from wild-type Cas9-OmB cells was used as the control sample.

***Cellular phenotyping***

Cells were seeded in 6-well plates (Corning, Tewksbury, MA, USA) containing 2 mL complete L-15 medium. After 72 h, attached and proliferating cells were visualized by phase contrast microscopy using an inverted microscope (DMi8, Leica)

***Prediction of 3D protein structure***

To confirm functional ko of the resulting protein, the structure of *myca* ko truncated protein was compared to wildtype protein. Geneious 2022.0.1 (Biomatters, https://www.geneious.com) bioinformatics software was used to predict pre-mature translation termination (early stop codon) resulting from deletion of a single nucleotide from wild-type *myca* in a monoclonal ko mutant (*myca* ko clonal(-1), see results). The resulting mutant MYC protein sequence was generated by translating the cDNA sequence and compared to the wild-type MYC protein sequence. Both (mutant and wildtype) MYC sequences were annotated with functional domains using InterPro version 90.0 (EMBL-EBI)[225]. Moreover, the 3D protein structures were predicted for both (mutant and wildtype) MYC proteins and visualized using AlphaFold[226], Mol*3D Viewer[227], and RCSB Protein Data Bank[228]. Combined with the above prediction tools, FunFOLDQA

[229], a protein ligand binding site residue prediction tool, was also employed to reveal whether any DNA binding capacity is preserved in the truncated mutant MYC protein.

## Results

### *Identification of evolutionarily conserved tilapia MYC domains for functional inactivation*

MYC TF orthologs are highly conserved across many species although the N-terminal part is often more variable than the C-terminal part of MYC, indicating that the latter has been functionally more highly conserved during evolution [191]. In addition to its canonical cellular functions, MYC (encoded by *myca*) may contribute to tilapia osmoregulation as *myca* mRNA is induced during hyperosmolality and multiple MYC binding sites (E boxes) have been identified in the promoter region of the highly hyperosmotically induced tilapia gene *IMPA1.1* (Figure 4-S1). The MYC functional domain annotation followed by evolutionarily conserved region search identified the long transcriptional regulatory domain at N-terminus and the short DNA-binding domain, consisting of basic helix-loop-helix (bHLH) and leucine zipper (LZ) motifs, at C-terminus (Figure 4-1A and 4-1D).

It was also identified the transcriptional regulatory domain was conserved as a whole but rather contained multiple regions that were more highly conserved and interspersed between more divergent stretches of amino acid sequence (Figure 4-1A). In contrast, the entire DNA-binding domain, especially the bHLH part of this domain, was highly evolutionarily conserved with a pairwise identity of 90.8 % among all 15 species. A phylogenetic tree generated with Geneious (Biomatters) based on the multiple sequence alignment illustrates that tilapia MYC is most similar to MYC of other African cichlids followed by other euryhaline teleosts (medaka and killifish) (Figure 4-1B).

Because highly conserved clusters of sequences were found in the N-terminal transcriptional regulatory domain of tilapia MYC, we aimed to design sgRNAs in the very beginning of the coding sequence to avoid retention of any potentially functional domain in a truncated mutant protein resulting from frame-shift mutation. Two sgRNAs - sgRNA2 (rk#3) and sgRNA3 (rk#4) - designed with CRIPSPOR and SYNTHEGO both met this criterion in addition to having the lowest predicted off-target effects. A third sgRNA - sgRNA1 (248fw) – had the highest scores in both CRISPOR and SYNTHEGO but was located in the earlier part of the transcriptional regulatory domain (227 bp downstream of the start codon), which meant that the truncated mutant protein produced by this sgRNA still contained two highly conserved sequence blocks of the transcriptional regulatory domain (Figure 4-1C). PCR primers for amplifying a region that included target loci of all three sgRNAs (test amplicon) were designed and subsequently used to sequence mutated genomic loci. This test amplicon was sequenced for the wild-type *O. mossambicus* Cas9-OmB cells and compared to the *O. niloticus* reference sequence. The pairwise sequence identity of this test amplicon between *O. mossambicus* and *O. niloticus* was 96.9 % overall and 100% for all three sgRNA target sequences (Figure 4-S2).

***Isolation and serial dilution of myca knockout cell lines with low genetic variability***

The first *myca* gene editing experiment was performed with sgRNA1. In this pilot experiment, the sequence of the target locus in Cas9-OmB cells transfected with TU6m-sgRNA1-expression vector and selected with hygromycin was compared to that of wild-type Cas9-OmB control cells at the same locus immediately after hygromycin selection. The Sanger sequencing chromatograms of wild-type and sgRNA1 mutant *myca* test amplicons as analyzed by TIDE and ICE tools indicated relatively poor gene editing efficiency(monoallelic) of 16.9 % (TIDE) and 6 % (ICE) (Figure 4-2A).

To improve the gene editing efficiency of selected cells, enrich for cells with biallelic *myca* ko, and reduce genetic heterogeneity of the mutant cell population, a limiting dilution strategy was employed in a series of experiments that utilized all three *myca* sgRNAs. In these experiments, cells were allowed to recover from selection by incubation in complete media for 14 days. Cell recovery restored proliferation rate and provided sufficient cell numbers for applying a strategy of limiting dilution. Dilution of the initial *myca* ko cell mixture into a 24-well plate at an average density of one cell per well (Figure 4-2B), dramatically improved the overall gene editing efficiency scores in the resulting cell lines ranging from 46.8 % to 98.4 % in TIDE efficiency and from 37 % to 92 % in ICE efficiency (Figure 4-3A). The highest gene editing efficiency (biallelic) was 98.4% (TIDE) and 92% (ICE) for sgRNA1 colony #4.

In addition to indel percentage, ICE analysis also provides a KO-score that is derived from calculating the proportion of indels that cause a frameshift or are longer than 21 bp. For instance, the KO-score of sgRNA1-colony#4 was 42/100. This score suggests that, although virtually all cells in sgRNA1-colony#4 were mutated, less than half yield a functionally or severely impaired protein because 58% do not harbor a frameshift mutation or a deletion of more than 7 amino acids. Nevertheless, such discrepancy between gene editing efficiency and KO-score was the exception and the two scores were virtually identical for most mutant cell populations isolated after limiting dilution (see Figure 4-3A). Thus, the majority of indel mutations that were isolated and enriched by the limiting dilution strategy resulted in effective frameshifts and severe MYC truncation (Figure 4-3).

The 1-bp deletion genotype present in the knockout pool of sgRNA1-colony#3 showed the highest contribution (57 %) of any single genotype in any of the cell populations isolated by the first limited dilution series (Figure 4-S3). This result suggests that sgRNA1-colony#3 was the most

promising for isolating a monoclonal mutant cell line harboring a *myca* frameshift ko. Thus, we decided to perform a second series of limiting dilution for this population of cells after generating a single cell suspension of sgRNA1-colony#3 by trypsinization of these cells and passing them several times through a serological pipet. Interestingly, altered morphologies and growth patterns were observed in most *myca* ko cell pools resulting from the first series of limiting dilution. For example, *myca* ko cells appeared more adherent to each other (e.g., Figure 4-4B), had altered growth patterns resulting in more densely clustered patches (e.g., Figure 4-4F), and had lower proliferation rates than wildtype cells, i.e., they needed more time to reach confluency (e.g., Figure 4-4D).

***Isolation of a myca-knockout monoclonal tilapia cell line***

The first series of limiting dilution produced polyclonal KO cell pools with high gene editing efficiency that can be directly used for functional analyses. However, to unambiguously rule out any off-target effects on the phenotype of interest it is preferable to use multiple homogenous monoclonal KO lines for functional analyses. The statistical likelihood that two different monoclonal lines harbor the same off-target mutation is infinitesimally small. Thus, if a consistent phenotype is observed it cannot be due to off-target effects. Therefore, the possibility of generating a monoclonal KO line by another series of limiting dilution of KO cells was explored. To demonstrate proof of principle the sgRNA1-colony#3 KO cell pool resulting from the first limiting dilution series was chosen because it included a 1-bp deletion genotype that accounted for 57 % of the total cell population (Figure 4-S3). This cell pool was diluted and seeded into a 96-well plate at an average concentration of 0.5 cells per well. Two cell colonies were detected after 14 days of seeding and isolated into separate wells. One of these colonies was confirmed to be monoclonal. It consisted of the 1-bp deletion genotype that was most abundant in the starting

population of cells. TIDE analysis showed 98.5 % gene editing efficiency with $R^2 = 0.99$ (Figure 4-5A), and ICE indel efficiency was 100 % with $R^2 = 1$ (Figure 4-5D). Moreover, the KO-score was 100 /100 (biallelic mutation), which indicates that the 1-bp deletion mutation in the *myca* gene results in functional disruption of MYC TF. ICE analysis confirmed a homogeneous monoclonal genotype consisting to 100 % of the 1-bp deletion mutant (Figure 4-5B). The original Sanger sequence trace also confirmed cleanly that the 1-bp deletion of a cytosine was present in all copies of the target test amplicon (Figure 4-5C). These data provide evidence that a two-step serial limiting strategy can be applied for isolating monoclonal KO mutant cell colonies from an initially highly heterogeneous mixture of genotypes.

***Structure of the predicted loss-of-function monoclonal mutant MYC protein***

Compared to wild-type MYC protein made up of 432 amino acids, the mutant MYC protein expressed in the monoclonal OmB-*myca*KO1 mutant cell line is predicted to be truncated to only contain the first 120 amino acids due to premature translation termination via an early stop codon (TGA) (Figure 4-6A). The majority of transcription regulatory domain and all of the DNA-binding domain (bHLH + LZ) are missing in the mutant MYC protein, which abolishes its function as a TF. Predicted protein structures of both wild-type and mutant MYC were profoundly different between wild-type and mutant MYC (Figure 4-6B and 4-6C). The removal of protein domains necessary for TF function from the truncated mutant protein was confirmed using FunFOLDQA[229], a protein ligand binding site residue prediction tool, which indicated ‘No DNA binding capacity’ of the truncated mutant MYC protein.

### *Morphological differences between Cas9-OmB-mycaKO1 and wildtype Cas9-OmB cells*

In addition to the general tendency for forming tighter cell clusters and decreased proliferation rate (time to confluency) outlined for heterogeneous mutant cell populations above, a conspicuous morphological difference was observed for the Cas9-OmB-*myca*KO1 cell line compared to *myca* wild-type Cas9-OmB cells. The morphology of both wild-type and mutant cell lines was compared after two additional passages to expand the mutant line for cryopreservation. The mutant cells appeared smaller due much shorter elongated cell extensions that are characteristic of wildtype cells even though the main cell body of mutant cells was comparable in size to wildtype cells (Figure 4-7).

## Discussion

### *Potential non-canonical role of MYC as an osmoregulatory transcription factor*

MYC is a well-studied transcription factor having numerous cellular functions such as the regulation of cell growth, cell cycle, cell differentiation, global mRNA translation, and cellular stress response (CSR) in a wide variety of organisms[230–232]. MYC has been studied extensively in the context of cancer biology because of it is overexpression in malignant tumors and its activation of many hallmarks of cancer[233]. However, other functions of MYC that are not directly relevant for cancer biology have not received much attention. Since MYC TFs are evolutionarily highly conserved and have many transcriptional targets[190], they are likely central regulators of the CSR and other diverse cellular functions that are important for normal non-pathological physiology[234].

To enable causal links between MYC function and the CSR in tilapia, in particular during salinity (osmotic) stress, this study generated tilapia *myca* ko cell lines. Previous studies suggest

that *myca*, which is the gene encoding for tilapia MYC, may be involved in governing mechanisms of teleost osmoregulation[193,198]. Furthermore, our results show that, unlike *myc2*, *myca* is elevated at the mRNA level during hyperosmotic stress in tilapia OmB cells (**Figure 4-S2A**). MYC regulates its downstream target genes by binding to the E-box (CACGTG or CATGTG), a MYC-specific cis-regulatory element (CRE). Intriguingly, three E-box sequences were found in the proximal promoter region (within 1.1 kb of the transcription start site) of the tilapia *IMPA1.1* gene encoding the most osmoresponsive enzyme in tilapia (Figure 4-S2B). We have previously shown that hyperosmotic transcriptional induction of *IMPA1.1* is at least partly mediated by several osmolality/salinity-responsive element 1 (OSRE1) CREs[22]. The OSRE1 core consensus sequence resembles that of the mammalian tonicity response element (TonE), which is the binding site for nuclear factor of activated T-cells 5 (NFAT5)[235]. This similarity suggests that tilapia NFAT5 contributes to the osmotic regulation of the *IMPA1.1* gene.

However, gene expression is often regulated in a combinatorial manner and dependent on multiple different CREs and TFs. Our findings of hyperosmotic MYC TF mRNA elevation and the presence of multiple MYC CREs (E-boxes) in *IMPA1.1* suggests that combinatorial transcriptional regulation involving NFAT5 and MYC controls the hyperosmotic induction of tilapia *IMPA1.1*. To enable functional analyses of the role of MYC and its interaction with NFAT5 for osmoregulatory target gene transactivation, this study generated multiple poly- and mono-clonal *myca* ko tilapia cell lines that can be used in combination with reporter assays[236,237], molecular and cellular phenotyping[238–240], and other approaches[241,242] to comprehensively characterize the role of MYC for teleost osmoregulation.

***Limiting dilution is an effective strategy for cloning specific cell genotypes***

Using gene targeting by CRIPSR/Cas9 results in a heterogenous pool of edited cells with varying indel populations[216]. Such highly heterogeneous polyclonal cell pools can have poor overall gene editing efficiency due to inefficient delivery or low sgRNA-dependent mutational efficiency[243,244]. Moreover, some of the cells in this heterogeneous pool may harbor non-specific mutations that can cause off-target effects. One way to eliminate such potential off-targets is to isolate genetically distinct cell pools and check for consistent phenotypes when multiple of these distinct pools are used for experiments that evaluate functional consequences of gene ko. In this study, we applied a serial limiting dilution strategy to increase gene editing efficiency and decrease the genetic heterogeneity of polyclonal and monoclonal mutant (gene-edited ko) cells. Both polyclonal and monoclonal cell lines are useful depending on the purpose of the subsequent experiment[218,221,245].

We successfully produced ten polyclonal cell pools with reduced heterogeneity in the indel genotypes when compared to the original mixture of gene edited cells. Five of these polyclonal lines had much higher gene editing efficiencies (>80%) than the starting population of gene edited cells. The remaining five cell pools with <80% gene editing efficiency also showed improved mutational efficiency relative to the starting population. The high gene editing efficiencies achieved during the initial series of limiting dilution illustrate that this strategy represents a rapid way for enrichment of desirable genotypes. However, wildtype mRNA and protein expression from unedited or heterozygous cells harboring a haploid mutant genotype represents a potential pitfall of studying polyclonal ko cells. This would be particularly problematic for long-term studies if unedited or haploid cells have a growth rate that exceeds those of diploid mutants.

To eliminate concerns about possible phenotype masking effects due to the potential presence of unedited or haploid genotypes in a polyclonal cell pool, we performed another series of limiting dilution. For this second dilution series a polyclonal cell line from the first series of dilution was chosen, which had the highest frequency of frameshift (functionally inactivating) mutations as determined by ICE analysis. We demonstrate that it is possible to isolate a monoclonal tilapia cell line harboring a 1-bp deletion in the *myca* gene after only two rounds of limiting serial dilution. In theory, it is possible to decrease the cell concentration during the first round of dilution even further and seed more aliquots (e.g., into a 1536-well plate) to obtain monoclonal lines after a single limiting dilution step. However, in praxis, our approach of subsequent serial dilutions at an average seeding density of 1 cell per well followed by an average seeding density of 0.5 cells per well yielded the best results with cells retaining their ability to form colonies within a short time (within 4-5 weeks). This approach also reduced the likelihood of seeding cell clusters present in cell suspensions during the second round of limiting dilution[217]. Thus, limiting dilution is an effective strategy to overcome (1) the genetic heterogeneity of mutant cells, (2) the potential presence of unedited or haploid mutant cells in mixed populations, and (3) the low delivery efficiency of vector-based CRISPR/Cas9 in cultured fish cells[246–248]. The limiting dilution strategy is not only fast but also very cost-effective.

Our approach overcomes difficulties often encountered when attempting to isolate clonal mutant cell lines from non-canonical model organisms[249]. For example, workflows for isolating gene-edited clonal cell lines by fluorescent-aided sorting system are currently only feasible for mammalian cells[250]. Moreover, even if expensive cell sorting devices and corresponding labeling approaches are available, cells would be potentially exposed to the non-sterile conditions and in danger of being contaminated. Cell viability is also negatively affected by the cell sorting

process. In conclusion, the serial limiting dilution strategy used in this study provides a robust, rapid, and cost-effective platform for generating ko clonal cell lines for studies of mutagenesis effects on cellular phenotypes of teleost fishes.

***Interpretation of gene editing results by TIDE and ICE analyses***

To analyze results from CRISPR/Cas9 gene editing experiments, the test amplicon sequence spanning the target site must be analyzed with bioinformatics tools that deconvolute the indel heterogeneity into interpretable scores. We used TIDE and ICE analyses for this purpose. Sanger sequence chromatograms of test amplicons from genomic DNA of gene edited and wildtype cells and the corresponding sgRNA sequence are used as the input data. TIDE[251] has been widely adopted for mutation detection since its development in 2014 as an accurate, versatile, and time-saving alternative to restriction enzyme-based assays[243,252]. The ICE analysis pipeline (SYNTHEGO)[253] has been developed more recently and examined rigorously by comparing to next generation sequencing (NGS)-based amplicon sequencing data. This evaluation revealed that the accuracy of ICE analysis is comparable to that of NGS-based approaches such as CRISPResso2, which aligns deep sequencing reads to a reference sequence[254].

In addition to scoring the overall gene editing efficiency, ICE also provides a Knockout score (KO-score) which is a useful measure to determine how many of the contributing indels are likely to result in a functional ko of the targeted gene. The main advantages of using both TIDE and ICE in combination are being able to: (1) Compare the consistency of two independent measures for gene editing efficiency, which are derived from different analytical algorithms, (2) Obtain detailed information about distribution and frequency of different types of indels, and (3) Estimate the frequency of indels resulting in a functional ko. The advantages and complementary

features for the combined use of TIDE and ICE for indel heterogeneity analyses has been reported previously[255].

One important finding of our study was a substantial difference between KO-score and indel efficiency observed in sgRNA1-colony#4. While TIDE and ICE indel efficiency scores for this cell population were 98.4 % and 92 %, respectively, the KO-score was only 42 % (Figure 4-3B). This information represents a critical factor for choosing particular cell populations for functional studies. In this case, sgRNA1-colony#4 cells would be eliminated from consideration for functional phenotype analyses because most of the mutants generated produce a MYC protein that is very similar to the wildtype protein.

The scores provided by TIDE and ICE analyses provide effective and unbiased selection criteria of choosing a particular population of cells from the first round of limiting dilution for subsequent serial dilution of cells. We selected the sgRNA1-colony#3 cell population resulting from the first limiting dilution for the second round of limiting dilution because of the high abundance of a specific genotype (a 1-bp deletion accounting for 51.4 % of the indels) and because the frameshift mutation introduced a premature translation termination codon shortly after the target site, Figure 4-6). Because of this premature stop codon, the mutant protein that is generated is very short, which minimizes potential non-specific side effects of expressing an abnormal protein in cells that are subjected to complex phenotype analyses.

***myca clonal KO tilapia cell line and its potential use and implications***

The limiting dilution strategy had already been successfully used in a few cases to generate clonal ko cells in different fish species including carp (*Cyprinus carpio*)[221,256], medaka (*Oryzias latipes*)[257], and chinook salmon *(Oncorhynchus tshawytscha)*[220]. These studies

were focused on investigating mechanisms of resistance to viral infection. However, generation of polyclonal or monoclonal ko cell lines from heterogeneous indel genotypes generated by CRISPR/Cas9 technology, whether by limiting dilution or other approaches, has not been reported prior to this study for any tilapia species.

Here we established proof of principle that monoclonal tilapia ko cells can be generated by limiting dilution of heterogeneous indel genotypes resulting from using a vector-based CRISPR/Cas9 system with cell lines. Such monoclonal ko lines have a defined genotype that facilitates the interpretation of functional studies aimed at evaluating the effects of gene inactivation on cellular phenotypes, for example osmoregulatory, disease resistance, proliferation, or other phenotypes that are informative for understanding basic physiological mechanisms but also of great interest from an applied perspective, e.g., for improving aquaculture[60,258]. Tilapia are a widely used aquaculture species, second only to carp regarding global production yields. However, like many other organisms, tilapia are subjected to climate change and pollution, which negatively affects their performance and their natural habitat, and impacts aquaculture efforts. Mechanistic insight derived from gene targeting studies helps to understand, properly interpret, and compensate for such impacts to facilitate mitigating these negative effects.

In the current study, the *myca* gene encoding MYC TF was chosen as to demonstrate the effectiveness of the limiting dilution strategy for generating a clonal ko cell line. This cell line and the polyclonal *myca* ko lines generated in the present study enable testing the role that MYC TF plays for tilapia cellular osmoregulation, its contribution to the activation of the myo-inositol biosynthesis (MIB) pathway, and its contribution to other cell functions. This approach can be extended to other targets and species of interest, for example genes important for aquaculture traits other than salinity tolerance and candidate genes other than *myca*, including those identified by

previous GWAS and SNP analysis in multiple aquaculture fish species[259–261]. These ko cell lines allow for deep functional analyses that associate specific genotypes with complex phenotypes, including systems level molecular phenotypes revealed by transcriptomics[262] and proteomics[263].

Surprisingly, the cell viability and growth rate of the *myca* clonal ko cell line did not differ from that of wild-type Cas9-OmB cells. This contrasts with several polyclonal ko cell pools obtained after the first round of limiting dilution, which display reduced proliferation even after three passages. This result indicates that, not only is *myca* ko not lethal, but it also does not hamper cell propagation *in vitro*. This phenomenon is interesting as we had expected that all ko cell lines might show similar reduced cell viability and slower cell growth compared to the wild-type cells because of the known essential roles of MYC in numerous cellular processes including cell growth, proliferation, and differentiation[264]. Nevertheless, the aberrant morphologies of tilapia *myca* ko cell lines (Figures 4-4C and 4-4D) support the idea that cell differentiation, one key cellular phenotype controlled by MYC TF, is notably altered relative to wild-type cells.

Although aberrant cell morphology is reported for *myca* ko cells in this study, more in-depth molecular phenotyping will help understand the physiological consequences of *myca* ko in future studies. Specifically, osmotolerance phenotypes of *myca* ko cells can be analyzed[224] and quantitative proteomics approaches can be employed to provide functional insights into biochemical and genetic networks that are controlled by MYC TF[265,266]. The limiting dilution approach and the *myca* ko tilapia cell lines generated in this study will empower such future functional analyses. In conclusion, this study successfully used a vector-based CRISPR/Cas9 approach in combination with a serial limiting dilution strategy to generate mono- and poly-clonal

tilapia *myca* ko cell lines for in depth cellular phenotyping studies directed at investigating functions of MYC TF in tilapia.

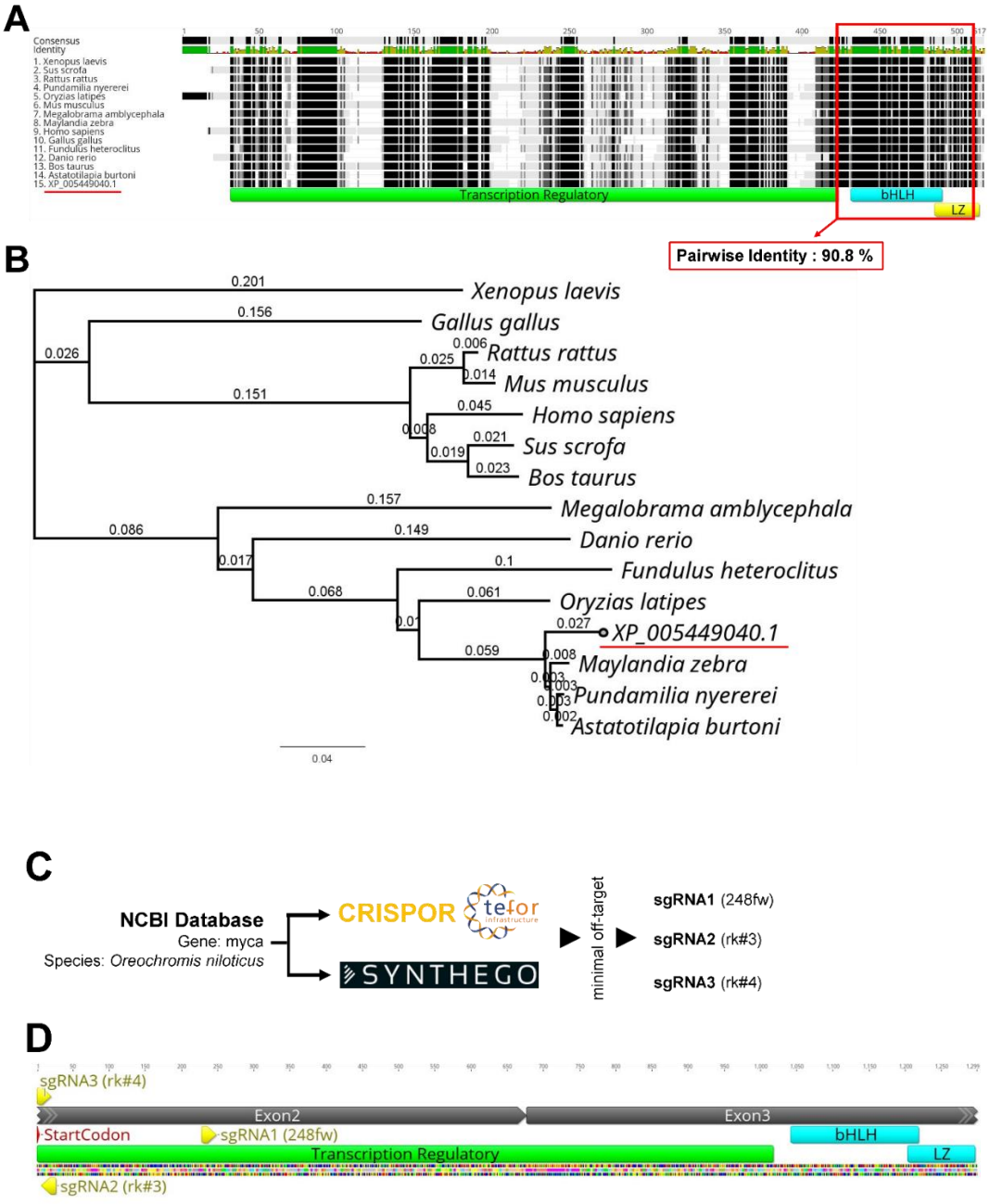



**Figure 4-1.** Evolutionary conservation of MYC protein domains and location of myca sgRNAs for CRIPSR/Cas9-mediated gene-editing. **(A)** Multiple alignment of MYC protein sequences from

15 vertebrate species including mammals, amphibians, and fishes generated with Geneious (Biomatters, cost matrix = Blosum62). The names of species included in the alignment are listed on the left side. The Oreochromis niloticus MYC protein sequence is labeled by its NCBI accession number (XP_003439292.1) and underlined red. Black blocks indicate highly conserved regions while lighter ones (grey blocks) represent less conserved regions. The transcriptional regulatory domain is depicted by a green-colored bar and the DNA binding domain (bHLH plus LZ) is depicted by cyan-colored bars. The pairwise identity of the DNA binding domain is outlined in red. **(B)** Phylogenetic tree corresponding to the alignment shown in panel A generated using Geneious (Biomatters, genetic distance model = Jukes-Cantor, method = Neighbor-Joining). **(C)** Workflow of sgRNA design using the O. niloticus NCBI reference genome to screen for off-target effects identified the three best sgRNAs for tilapia myca. **(D)** Tilapia myca CDS (1299 bp) annotated with corresponding functional protein domains (green, cyan bars) and the location of sgRNA targets (yellow arrows). Exons are indicated by grey bars at the top.

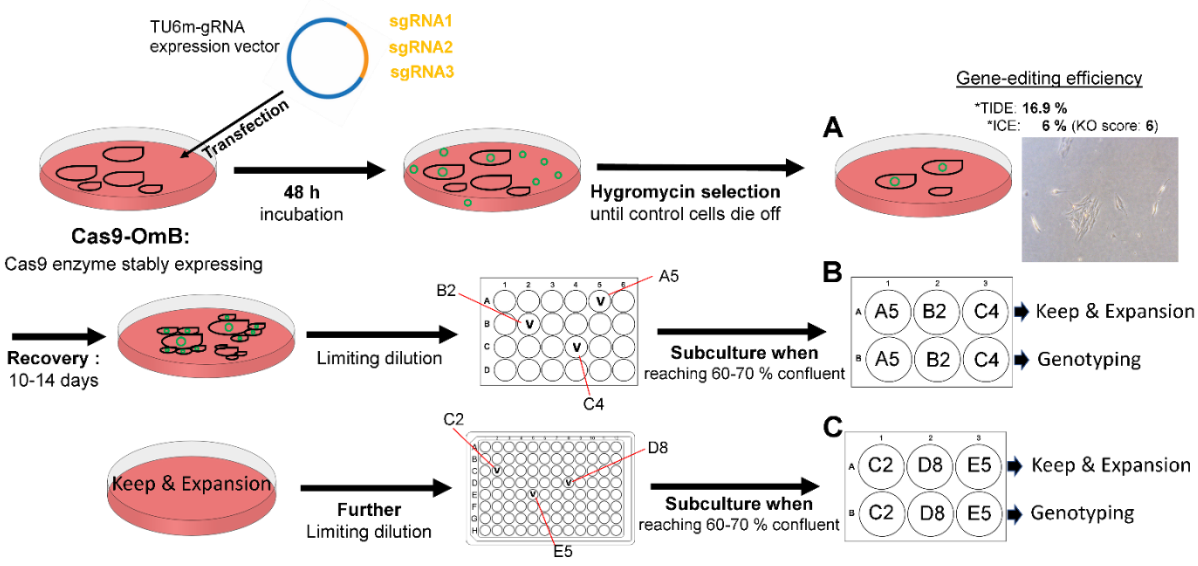



**Figure 4-2.** Schematic of selection and dilution strategy. Blue circle depicts sgRNA expression vector containing each sgRNA (orange part) targeting the myca gene. Cas9-OmB cells are depicted

as black cell-shaped components in dishes. Green small circles in cell culture dishes represent plasmid vector used for transfection. **(A)** Representation of the initial Hygromycin-selection of highly heterogeneous myca ko cells. The micrograph at the right illustrates the very low confluency of cells surviving the selection process. CRISPR/Cas9-mediated gene editing efficiency of the initial batch of selected cells was low as indicated by TIDE (Tracking of Indels by Decomposition; shinyapps.datacurators.nl/tide/) and ICE (Inference of CRISPR Edits; SYNTHEGO - CRISPR Performance Analysis) scores above the cell micrograph. The Knockout (KO)-score is even lower (6%) indicating that the proportion of cells with a functional ko is very low. **(B)** Initial series of limiting dilution after recovery of the initially selected cells shown in (A). The wells marked with 'v' (A5, B2, and C4) represent wells containing cell colonies. When these cell colonies reached 70% confluency, they were split 1:2 and transferred to two new wells, one of which was propagated for expansion and the other harvested for genotyping. **(C)** Second limiting dilution series using a single polyclonal KO cell pool generated in (B). Each well of the 6-well plate is labeled with the well locations (C2, D8, and E5) of the previous step. A 'v' marks wells that contained cell colonies (C2, D8, and D5). These were grown to 70% confluency, split 1:2, and transferred to two new wells, one of which was propagated for expansion and the other harvested for genotyping.

**A**

| Cell colony # | TIDE efficiency % | ICE Indel % | ICE-KO score /100 |
|---|---|---|---|
| **sgRNA1**-colony#1 | 48.6 | 47 | 47 |
| **sgRNA1**-colony#2 | 89.1 | 91 | 91 |
| **sgRNA1**-colony#3 | 81.3 | 83 | 83 |
| **sgRNA1**-colony#4 | 98.4 | 92 | 42 |
| **sgRNA2**-colony#1 | 84.9 | 89 | 88 |
| **sgRNA2**-colony#2 | 53.1 | 50 | 48 |
| **sgRNA2**-colony#3 | 76.3 | 81 | 59 |
| **sgRNA2**-colony#4 | 75.1 | 79 | 63 |
| **sgRNA2**-colony#5 | 63.2 | 75 | 64 |
| **sgRNA2**-colony#6 | 46.8 | 37 | 29 |
| **sgRNA3**-colony#1 | 84.9 | 88 | 86 |

**B**


**Figure 4-3.** Quantitative indel mutation efficiencies of cell colonies (pools) produced by limiting dilution of cells transfected with sgRNA1, sgRNA2, and sgRNA3 plasmids. **(A)** Individual cell colonies isolated by the limiting dilution method into individual wells of a 24-well plate at an average density of one cell per well (Figure 4-2B) were genotyped and quantitatively analyzed using both TIDE and ICE tools. In addition to obtaining indel mutational efficiencies, ICE provides a KO-score which is the proportion of indels with a frameshift or exceeding 21 bp in length. **(B)** Bar-graph visually indicates high consistency of all scores except for sgRNA1-colony#4, which has a lower KO-score.

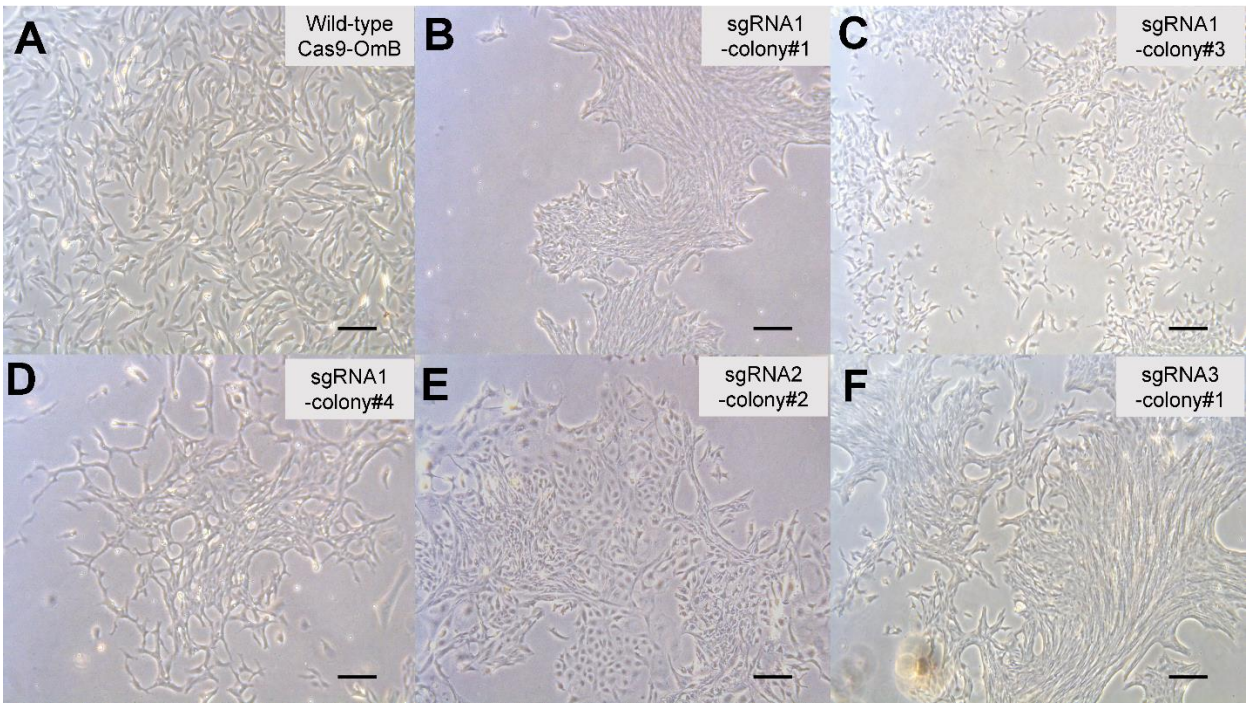



**Figure 4-4.** Representative images of cell morphology of wild-type Cas9-OmB cells and myca polyclonal ko cell pools (Scale bar, 200 μm). All micrographs were taken on an inverted microscope (Leica Dmi8) and imaged 5-7 days after transferring the cell colonies grown and tracked in a 24-well plate to a 6-well plate.

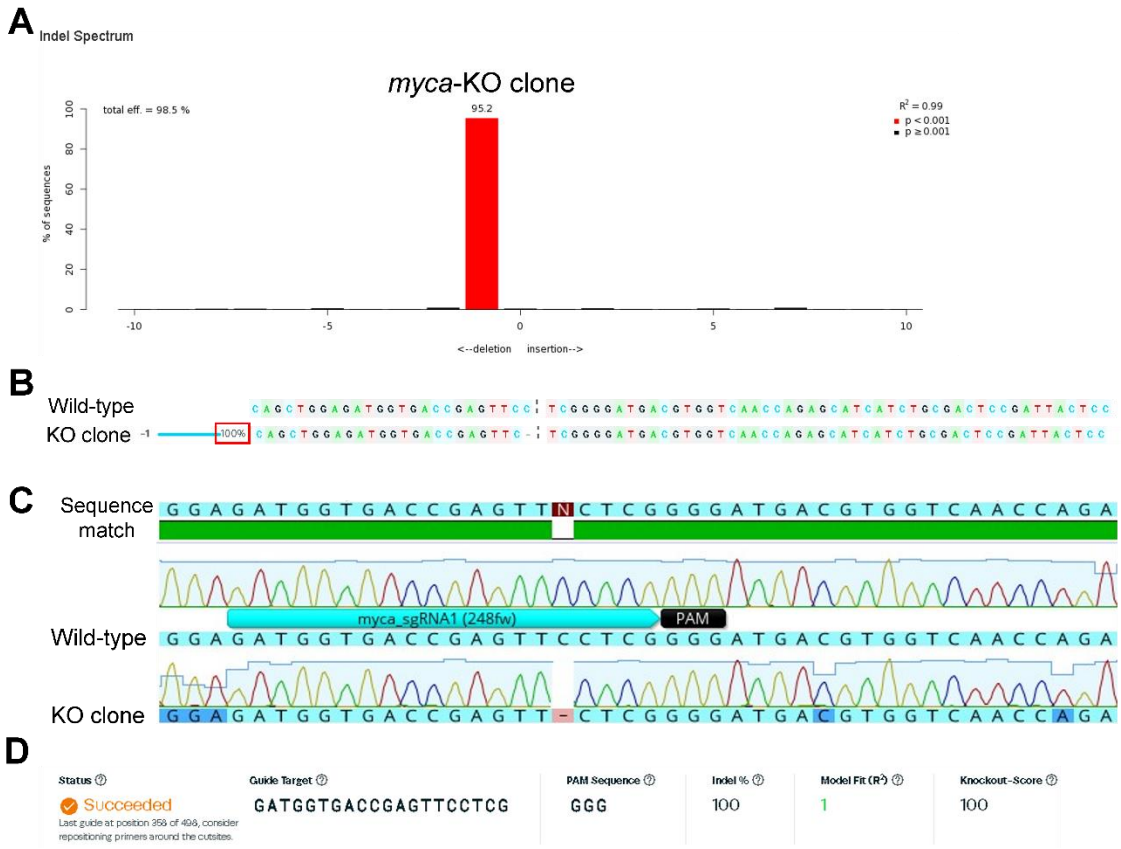



**Figure 4-5.** Confirmation of the homogeneity of mutant genotype in a myca monoclonal ko cell line. **(A)** TIDE analysis of genomic DNA extracted from a cell colony after the second series of limiting dilution. The X-axis indicates the nature of indels while the Y-axis depicts the percentages of the corresponding sequences. R-square refers to quality of the sequence reads from Sanger sequencing chromatograms with a value above 0.9 considered acceptable. The significance cutoff was set at a default $p < 0.001$ threshold. **(B)** The ICE analysis result of sequence distribution and frequency (%) of myca monoclonal ko cell line. The top row of nucleotides shows the wild-type sequence for the region surrounding target site. The bottom row indicates the gene-edited mutant sequence (ko clone). The dashed vertical black line indicates the cut site. Both sequences are aligned perfectly except for the 1-bp deletion in the mutant. **(C)** Alignment of wild-type and gene-edited (ko clone) sequencing chromatograms produced with Geneious 2022.0.1 (Biomatters) The

cyan-colored bar depicts sgRNA1 target sequence and the black bar shows PAM sequence. **(D)** ICE results include sgRNA target sequence, PAM sequence, indel efficiency, model fitness (R2), and Knockout-Score.

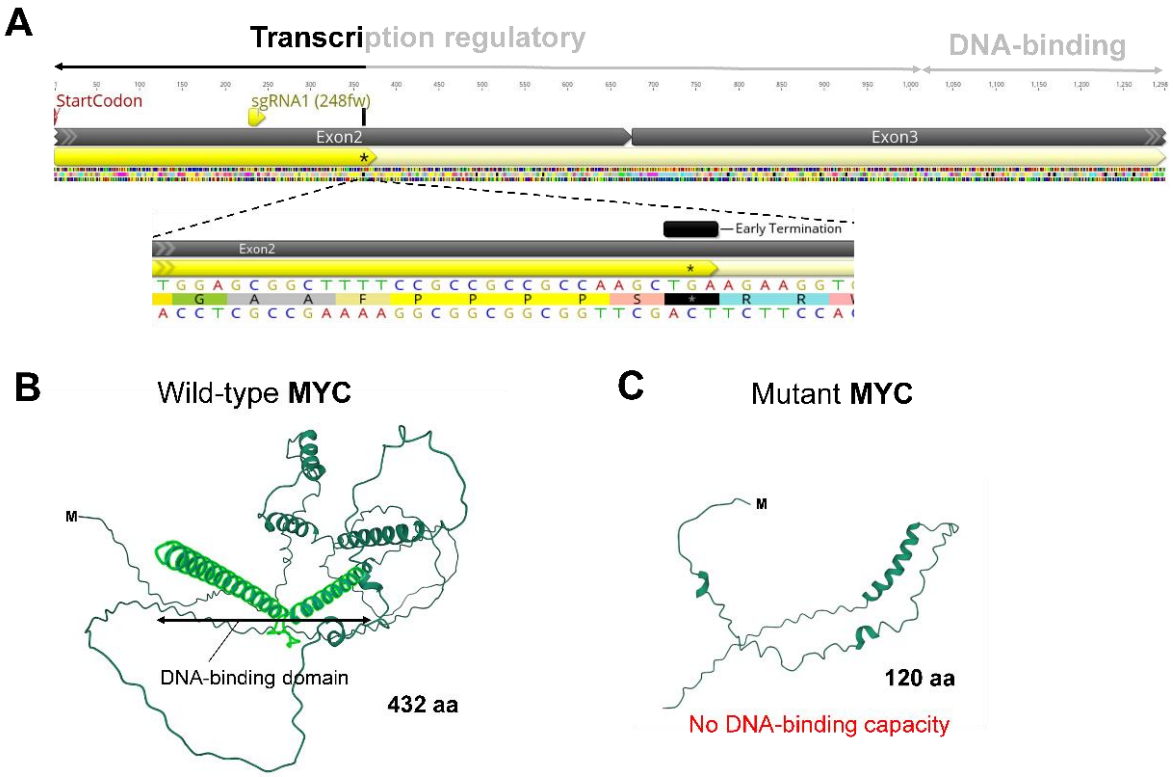



**Figure 4-6.** Prediction of loss of function due to truncation of a mutant MYC protein produced by the monoclonal Cas9-OmB-mycaKO1 cell line. **(A)** Wild-type and mutant MYC amino acid sequences were annotated on the coding sequence (CDS, shown yellow highlighted in the upper row). The region of the mutant MYC sequence that is delimited by a premature stop codon (TGA) caused by a 1 bp deletion in the target site is enlarged in the bottom row. **(B)** 3D structure of wild-type MYC protein (432 amino acids) as modeled using AlphaFold[226] (https://alphafold.ebi.ac.uk/entry/Q45RH2), Mol*3D Viewer[227], and RCSB Protein Data Bank

[228]. The scissor-like structure composed of two alpha-helixes (light green) represents the main DNA-binding domain. **(C)** 3D structure of mutant MYC protein (120 amino acids) as predicted using the same tools and approach as in panel B.

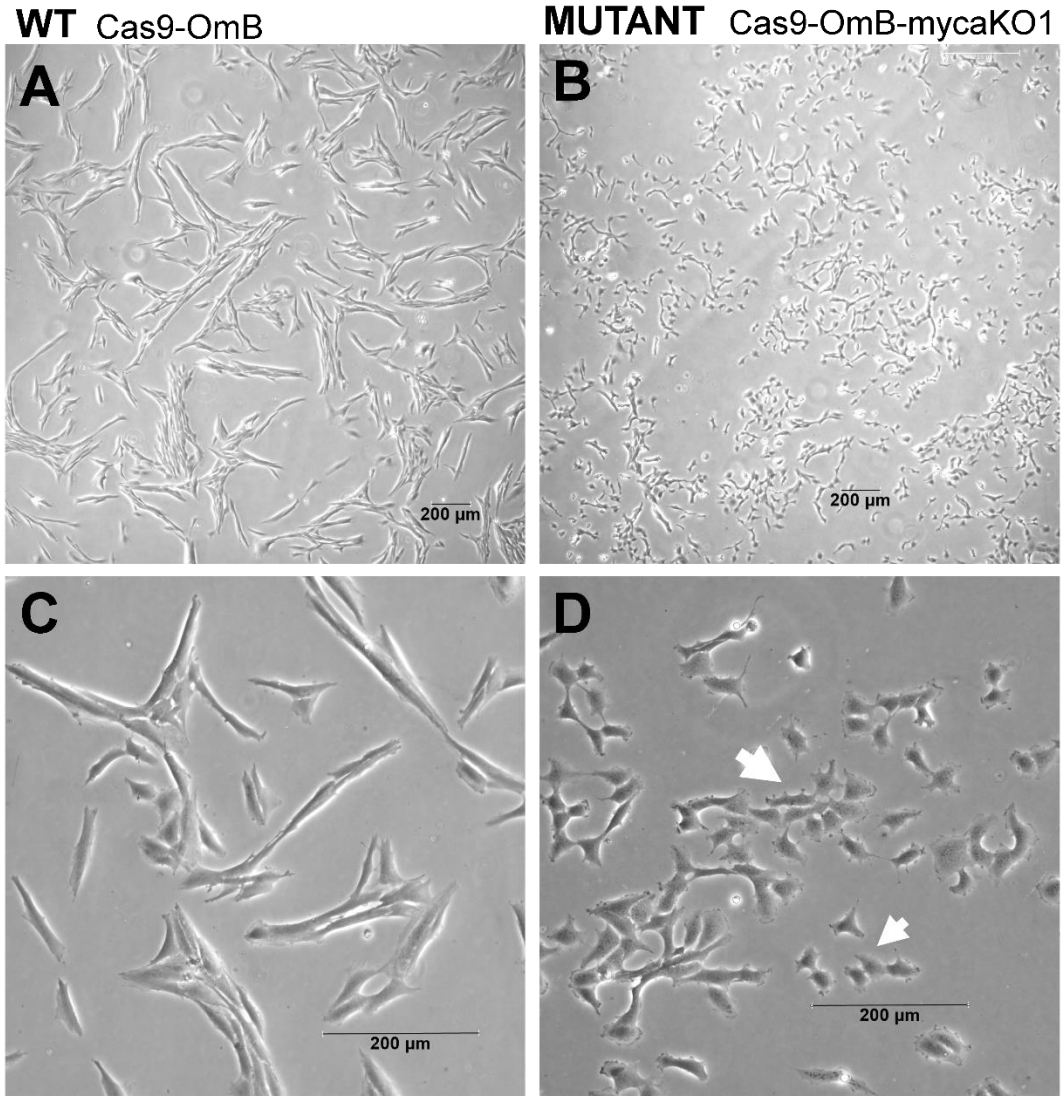



**Figure 4-7.** Cell morphology of tilapia cell lines. Micrographs showing differences in morphology of Cas9-OmB wildtype cells **(A, C)** versus Cas9-OmB-mycaKO1 mutant cells **(B, D)**. The lack of elongated cell extensions in selected mutant cells is illustrated by white arrows. Images were taken

using 5x and 20x phase objective on an inverted microscope (Leica Dmi8). The bar at the bottom of each image indicates a distance of 200 µm.

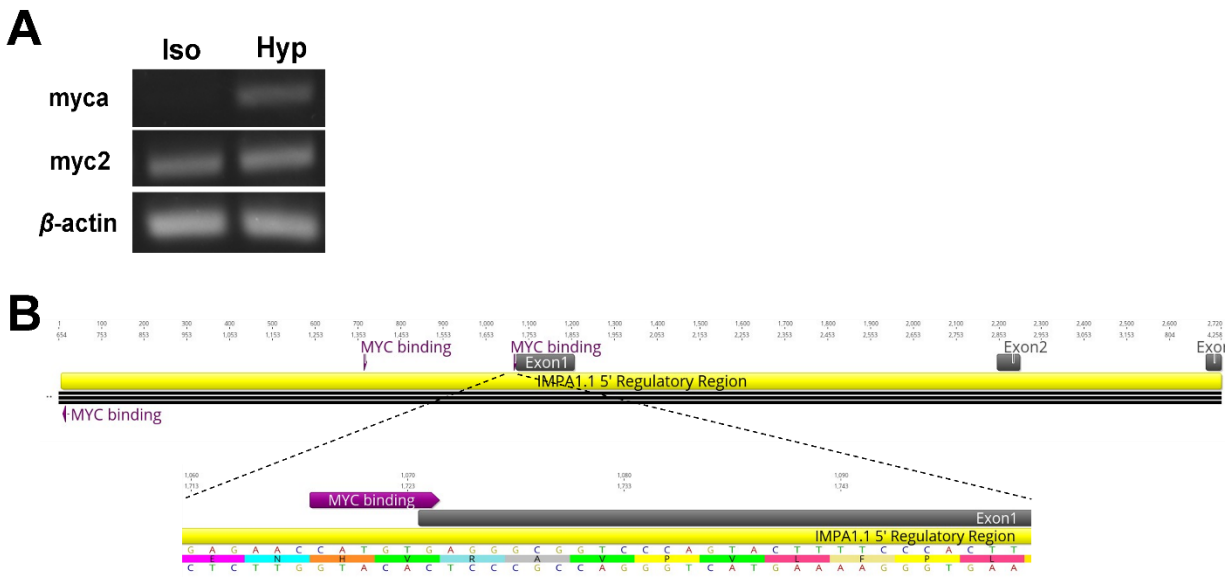



**Figure 4-S1.** MYC mRNA abundance in tilapia OmB cells. **(A)** Semi-quantitative cDNA agarose gel indicating increased myca mRNA abundance after exposure of OmB cells to 650 mOsmol/kg hyperosmolality for 6h. mRNA was isolated from both iso- (Iso) and hyperosmotically (Hyp) grown OmB cells using PureLink RNA mini kit (Invitrogen, 12183020 ), reverse transcribed into cDNA using Verso cDNA Synthesis Kit (ThermoFisher, AB1453A), and PCR amplified for 35 cycles using Mastercycler (Eppendorf, 6333000022). **(B)** MYC binding motifs (purple-colored) identified with the Geneious "Find Motifs" function (Geneious 2022.0.1, Biomatters) annotated to the O. mossambicus IMPA1.1 5' regulatory region (proximal promoter).


**Figure 4-S2.** Sequence confirmation of sgRNAs targeting Oreochromis mossambicus MYC (myca). **(A)** The myca targeted sequence (test amplicon region) was compared between O. mossambicus and O. niloticus by aligning the PCR-amplified sequence covering all sgRNA target sequences with the O. niloticus reference sequence. The pairwise identity between the two sequences is 96.9 %. **(B-D)** sgRNA coding sequence validation after cloning into TU6 gRNA expression vector to confirm 100% match of sgRNAs to corresponding O. mossambicus myca locus. **(B)** sgRNA1 (248fw), **(C)** sgRNA2 (rk#3), and **(D)** sgRNA3 (rk#4).

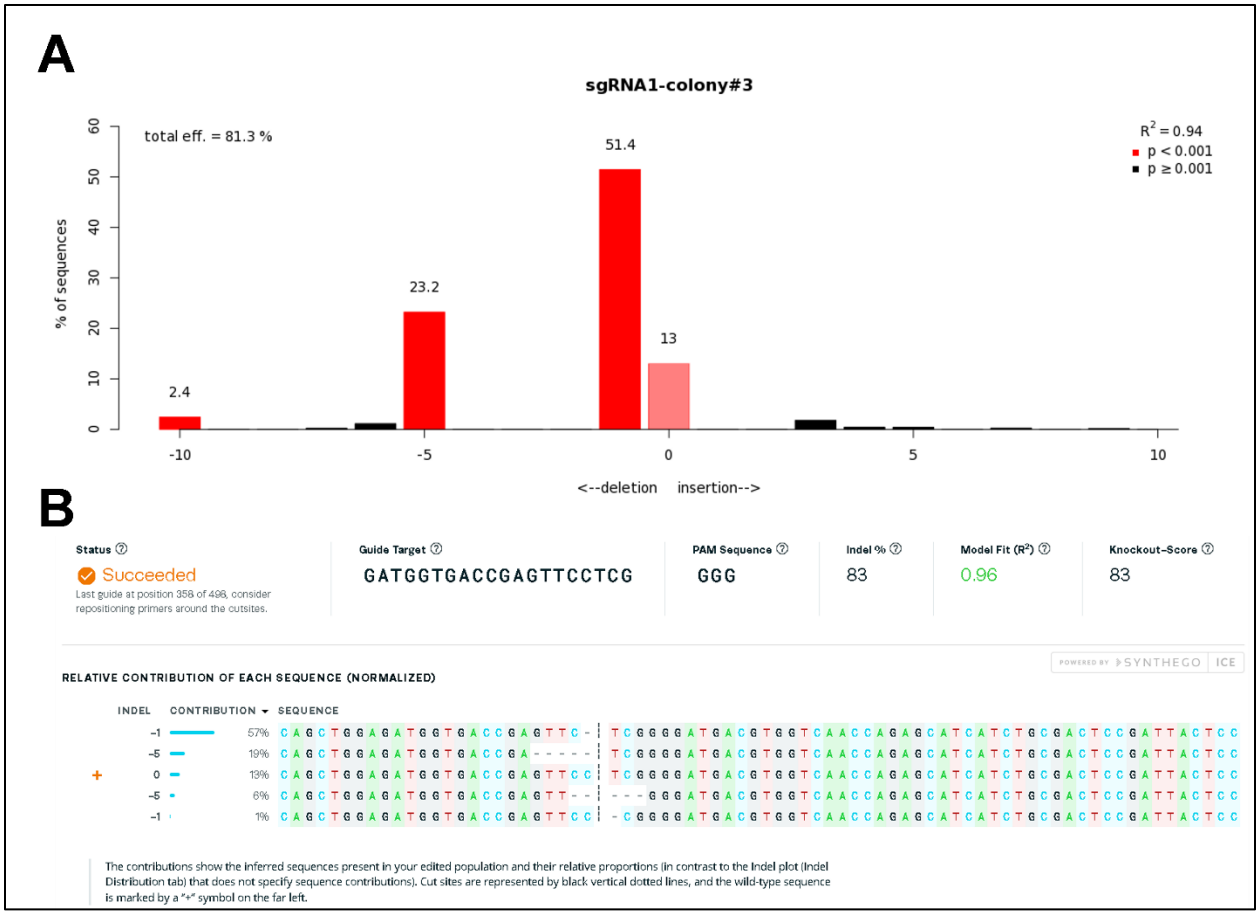



**Figure 4-S3.** Relative contribution of each genotype present in the sgRNA1-colony#3 ko cell pool after the first round of limiting dilution. **(A)** TIDE analysis result with significant indels shown as red bars (p-value < 0.001). **(B)** Corresponding ICE analysis result showing high agreement with the TIDE data. In addition, ICE analysis also provides the knockout score, which is the same as the Indel frequency (83%). Matching scores for indel frequency and knockout score indicate that all three types of indels present in this cell population (-1del, -5del.1, -5del.2, and -10del) all cause frameshift mutations. The R-square value refers to quality of the sequence reads from Sanger sequencing with a value above 0.9 considered acceptable.

| Primer Purpose | Primer Name | Primer Sequence |
|---|---|---|
| sgRNA oligo annealing | myca_sgRNA1(248fw)Target | CGATGGTGACCGAGTTCCTCGGTTTTAGAG |
| | myca_sgRNA1(248fw)Compl | CTAGCTCTAAAACCGAGGAACTCGGTCACCAT |
| | myca_sgRNA2(rk#3)Target | CGCTCGCCAAACTTGGATTCCGGTTTTAGAG |
| | myca_sgRNA2(rk#3)Compl | CTAGCTCTAAAACCGGAATCCAAGTTTGGCGAG |
| | myca_sgRNA3(rk#4)Target | CGCATGCCGCGGAATCCAAGTTGTTTTAGAG |
| | myca_sgRNA3(rk#4)Compl | CTAGCTCTAAAACAACTTGGATTCCGCGGCATG |
| Genotyping | myca_TideF1 | TGGAGGGAGTTGACCATGAAAG |
| | myca_TideR2 | CTCGGACACCACCTTCTTCA |
| | sgRNA_segP1 (for sequencing) | GTATACTATGTGCCGAATTTCC |
| myca mRNA detection | myca_qPCR_F | TGTCACTGCCGCACTGGAAT |
| | myca_qPCR_R | CAAACTTGGATTCCGCGGCA |
| myc2 mRNA detection | myc2_qPCR_F | GTGTTCCTGGGTGAGAAGCA |
| | myc2_qPCR_R | TGTCCACTGTCACCACATCG |

**Table 4-S1.** Sequences of primers used in this study

# CHAPTER 5

## Summary and Future Directions

### Summary

The research addressed in this dissertation aims to understand molecular mechanisms underlying cellular osmoregulation of euryhaline fish during salinity stress using a cell line model with a particular emphasis on transcriptional regulation. We identified multiple mechanisms by which transcriptional regulation of tilapia genes is controlled during hyperosmotic stress. These mechanisms contribute to the unique physiological capacity of euryhaline fish to tolerate salinity stress. Transcriptional regulation *per se* is a fundamental molecular mechanism governing cellular stress responses under sub-optimal conditions to maintain physiological homeostasis. Therefore and because of the large knowledge gap regarding the link between perception of an osmotic signal and induction of gene expression, Chapters 2 and 3 focus on the transcriptional regulatory mechanisms and networks controlling hyperosmolality-responsive genes and the corresponding proteins. These chapters have identified novel osmoregulatory CREs in multiple hyperosmotically induced genes. They have also identified potential candidate binding partner TFs that activate those genes and provide a focus for future research in this field. A targeted gene-specific (glutamine synthetase, GS) approach (Chapter 2) has revealed unexpected localization of OSRE1 (a salinity-responsive CRE) in intron 1 of the tilapia *GS* gene. This interesting finding assigns a novel functional role to an intronic non-coding region of *GS* and illustrates that introns can no longer be regarded as  trash across the entire genome[267]. In contrast to Chapter 2, an untargeted systematic approach was pursued to discover novel CREs by aligning the 5' regulatory regions of many hyperosmotically, transcriptionally induced genes that lead to increases in corresponding

protein levels during salinity stress. This approach leveraged bioinformatics tools to enable *de novo* discovery of a new salinity-responsive CRE which we named STREME1 (Chapter 3). The STREME1 CRE was experimentally validated in order to confirm its functional role for the transcriptional induction of salinity-responsive genes. The approach taken in Chapter 3 is broadly applicable to many other biological questions that require a systematic approach that takes advantage of systems level data on gene expression without prior bias towards a candidate CRE or TF. This approach lends itself very well to take advantage of massive 'omics' datasets that are now commonly produced. Combined with gene targeting approaches they enable establishment of causal links between environmental signals and gene expression on the one hand and gene expression changes and phenotypes on the other hand. This dissertation has identified several CREs and candidate TFs that control transcriptional regulation in hyperosmotically stressed tilapia cells. Moreover, the research reported in Chapter 4 has provided the tools to causally link the function of one of these TFs, MYC, to the hyperosmotic induction of osmoregulated genes and other cellular phenotypes associated with cellular stress responses and other aspects of cell physiology. To facilitate robust and efficient gene targeting to manipulate the function of TFs during osmotic stress, specific *myca* ko mono- and poly-clonal tilapia cell lines were produced using a CRISPR/Cas9 system optimized for tilapia cells. Chapter 4 reports the successful establishment of *myca* ko clonal cell lines (polyclonal cell pools and a monoclonal cell line) by providing proof-of-principle that a serial limiting dilution strategy represents a rapid and cost-effective means for isolating specific mutant genotypes from complex heterogeneous cell populations. This approach can now be extended to other TFs beyond MYC to test their involvement in tilapia transcription regulatory mechanisms and/ or osmosensory signaling during salinity stress.

In summary, this dissertation contributes to advancing our knowledge about salinity-responsive molecular mechanisms with an emphasis on how transcriptional regulation is controlled and establishing a platform for future functional analyses by providing genetically engineered clonal cell line models to address the role of specific TFs.

**Future directions**

Findings and discoveries obtained from efforts described above suggest several follow-up studies to further advance our understanding of molecular mechanisms of osmoregulation in euryhaline fish. First, a pull-down approach using a bead-immobilized DNA sequence can be used to directly isolate TFs that bind to the salinity-responsive CREs identified in this dissertation (OSRE1 and STREME1). Such an unbiased approach, where all putative TFs could be candidates, would reveal gene-regulatory networks in euryhaline fish that operate during osmotic stress and generate novel hypotheses regarding osmosensory signaling mechanisms. Several candidate TFs have already been identified in our lab, including MYC, NFAT5, and OSTF1 but it is likely that combinatorial signaling involves even more TFs[268]. Therefore, the above approach will either assess the role of known TFs in osmoregulation or expand our knowledge of additional transcriptional regulatory mechanisms by identifying novel TFs that contribute to osmoregulatory responses.

Moreover, an efficient system to produce gene ko cell lines using CRISPR/Cas9 technology has been developed in Chapter 4, which can now be leveraged to elucidate osmoregulatory function of the candidate STREME1 binding partner, FoxL1, identified in Chapter 3. By producing a monoclonal *foxl1*-ko cell line according to the limiting dilution strategy described in Chapter 4, it would be possible to compare hyperosmotic transcriptional activity of proximal regulatory regions of a reporter gene harboring STREME1 motifs using different cell

lines (a normal OmB cell line and a *foxl1*-ko cell line). This approach can highlight crucial information about whether the putative TF, identified *in silico*, is functionally involved in controlling osmoresponsive gene-regulatory networks during salinity stress. Such studies focused on causally linking the function of osmoregulated TFs to the osmotic regulation of gene expression and gene regulatory networks will strengthen our mechanistic understanding of how euryhaline fish respond to salinity stress.

**BIBLIOGRAPHY**


1. Schultz, E.T.; McCormick, S.D. 10 - Euryhalinity in An Evolutionary Context. In *Fish Physiology*; McCormick, S.D., Farrell, A.P., Brauner, C.J., Eds.; Euryhaline Fishes; Academic Press, 2012; Vol. 32, pp. 477–533.
2. Jawad, L.A. FISH PHYSIOLOGY: EURYHALINE FISHES - Edited by S. D. McCormick, A. P. Farrell & C. J. Brauner. *J. Fish Biol.* **2013**, *83*, 1485–1486, doi:10.1111/jfb.12249.
3. Turner, G.F. Adaptive Radiation of Cichlid Fish. *Curr. Biol.* **2007**, *17*, R827–R831, doi:10.1016/j.cub.2007.07.026.
4. Brawand, D.; Wagner, C.E.; Li, Y.I.; Malinsky, M.; Keller, I.; Fan, S.; Simakov, O.; Ng, A.Y.; Lim, Z.W.; Bezault, E.; et al. The Genomic Substrate for Adaptive Radiation in African Cichlid Fish. *Nature* **2014**, *513*, 375–381, doi:10.1038/nature13726.
5. Walther, G.-R.; Post, E.; Convey, P.; Menzel, A.; Parmesan, C.; Beebee, T.J.C.; Fromentin, J.-M.; Hoegh-Guldberg, O.; Bairlein, F. Ecological Responses to Recent Climate Change. *Nature* **2002**, *416*, 389–395, doi:10.1038/416389a.
6. Malhi, Y.; Franklin, J.; Seddon, N.; Solan, M.; Turner, M.G.; Field, C.B.; Knowlton, N. Climate Change and Ecosystems: Threats, Opportunities and Solutions. *Philos. Trans. R. Soc. B Biol. Sci.* **2020**, *375*, 20190104, doi:10.1098/rstb.2019.0104.
7. Uchida, K.; Kaneko, T.; Miyazaki, H.; Hasegawa, S.; Hirano, T. Excellent Salinity Tolerance of Mozambique Tilapia (Oreochromis Mossambicus): Elevated Chloride Cell Activity in the Branchial and Opercular Epithelia of the Fish Adapted to Concentrated Seawater. *Zoolog. Sci.* **2000**, *17*, 149–160, doi:10.2108/zsj.17.149.
8. Kültz, D. Physiological Mechanisms Used by Fish to Cope with Salinity Stress. *J. Exp. Biol.* **2015**, *218*, 1907–1914, doi:10.1242/jeb.118695.
9. Fiess, J.C.; Kunkel-Patterson, A.; Mathias, L.; Riley, L.G.; Yancey, P.H.; Hirano, T.; Grau, E.G. Effects of Environmental Salinity and Temperature on Osmoregulatory Ability, Organic Osmolytes, and Plasma Hormone Profiles in the Mozambique Tilapia (Oreochromis Mossambicus). *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **2007**, *146*, 252–264, doi:10.1016/j.cbpa.2006.10.027.

10. Watanabe, S.; Kaneko, T.; Aida, K. Aquaporin-3 Expressed in the Basolateral Membrane of Gill Chloride Cells in Mozambique Tilapia Oreochromis Mossambicus Adapted to Freshwater and Seawater. *J. Exp. Biol.* **2005**, *208*, 2673–2682, doi:10.1242/jeb.01684.
11. Choi, J.; Lee, K.M.; Inokuchi, M.; Kaneko, T. Acute Responses of Gill Mitochondria-Rich Cells in Mozambique Tilapia Oreochromis Mossambicus Following Transfer from Normal Freshwater to Deionized Freshwater. *Fish. Sci.* **2009**, doi:10.1007/s12562-009-0195-9.
12. Kültz, D.; Jürss, K.; Jonas, L. Cellular and Epithelial Adjustments to Altered Salinity in the Gill and Opercular Epithelium of a Cichlid Fish (Oreochromis Mossambicus). *Cell Tissue Res.* **1995**, *279*, 65–73, doi:10.1007/BF00300692.
13. Kültz, D.; Bastrop, R.; Jürss, K.; Siebers, D. Mitochondria-Rich (MR) Cells and the Activities of the Na+K+-ATPase and Carbonic Anhydrase in the Gill and Opercular Epithelium of Oreochromis Mossambicus Adapted to Various Salinities. *Comp. Biochem. Physiol. Part B Comp. Biochem.* **1992**, *102*, 293–301, doi:10.1016/0305-0491(92)90125-B.
14. Sacchi, R.; Gardell, A.M.; Chang, N.; Kültz, D. Osmotic Regulation and Tissue Localization of the Myo-Inositol Biosynthesis Pathway in Tilapia (Oreochromis Mossambicus) Larvae. *J. Exp. Zool. Part Ecol. Genet. Physiol.* **2014**, *321*, 457–466, doi:10.1002/jez.1878.
15. Fiol, D.F.; Chan, S.Y.; Kültz, D. Regulation of Osmotic Stress Transcription Factor 1 (Ostf1) in Tilapia (Oreochromis Mossambicus) Gill Epithelium during Salinity Stress. *J. Exp. Biol.* **2006**, *209*, 3257–3265, doi:10.1242/jeb.02352.
16. Sacchi, R.; Li, J.; Villarreal, F.; Gardell, A.M.; Kültz, D. Salinity-Induced Regulation of the Myo-Inositol Biosynthesis Pathway in Tilapia Gill Epithelium. *J. Exp. Biol.* **2013**, *216*, 4626–4638, doi:10.1242/jeb.093823.
17. López-Maury, L.; Marguerat, S.; Bähler, J. Tuning Gene Expression to Changing Environments: From Rapid Responses to Evolutionary Adaptation. *Nat. Rev. Genet.* **2008**, *9*, 583–593, doi:10.1038/nrg2398.
18. Cooper, G.M. Regulation of Transcription in Eukaryotes. *Cell Mol. Approach 2nd Ed.* **2000**.
19. Cases, I.; de Lorenzo, V.; Ouzounis, C.A. Transcription Regulation and Environmental Adaptation in Bacteria. *Trends Microbiol.* **2003**, *11*, 248–253, doi:10.1016/s0966-842x(03)00103-3.

20. Shibata, M.; Gulden, F.O.; Sestan, N. From Trans to Cis: Transcriptional Regulatory Networks in Neocortical Development. *Trends Genet. TIG* **2015**, *31*, 77–87, doi:10.1016/j.tig.2014.12.004.
21. Swift, J.; Coruzzi, G. A Matter of Time - How Transient Transcription Factor Interactions Create Dynamic Gene Regulatory Networks. *Biochim. Biophys. Acta* **2017**, *1860*, 75–83, doi:10.1016/j.bbagrm.2016.08.007.
22. Wang, X.; Kültz, D. Osmolality/Salinity-Responsive Enhancers (OSREs) Control Induction of Osmoprotective Genes in Euryhaline Fish. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E2729–E2738, doi:10.1073/pnas.1614712114.
23. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for Motif Discovery and Searching. *Nucleic Acids Res.* **2009**, *37*, W202-208, doi:10.1093/nar/gkp335.
24. Savarese, F.; Grosschedl, R. Blurring Cis and Trans in Gene Regulation. *Cell* **2006**, *126*, 248–250, doi:10.1016/j.cell.2006.07.008.
25. Reuveni, E.; Getselter, D.; Oron, O.; Elliott, E. Differential Contribution of Cis and Trans Gene Transcription Regulatory Mechanisms in Amygdala and Prefrontal Cortex and Modulation by Social Stress. *Sci. Rep.* **2018**, *8*, 6339, doi:10.1038/s41598-018-24544-3.
26. Gao, A.W.; uit de Bos, J.; Sterken, M.G.; Kammenga, J.E.; Smith, R.L.; Houtkooper, R.H. Forward and Reverse Genetics Approaches to Uncover Metabolic Aging Pathways in Caenorhabditis Elegans. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **2018**, *1864*, 2697–2706, doi:10.1016/j.bbadis.2017.09.006.
27. Gurumurthy, C.B.; Grati, M.; Ohtsuka, M.; Schilit, S.L.P.; Quadros, R.M.; Liu, X.Z. CRISPR: A Versatile Tool for Both Forward and Reverse Genetics Research. *Hum. Genet.* **2016**, *135*, 971–976, doi:10.1007/s00439-016-1704-4.
28. Gaj, T.; Gersbach, C.A.; Barbas, C.F. ZFN, TALEN, and CRISPR/Cas-Based Methods for Genome Engineering. *Trends Biotechnol.* **2013**, *31*, 397–405, doi:10.1016/j.tibtech.2013.04.004.
29. Hsu, P.D.; Lander, E.S.; Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **2014**, *157*, 1262–1278, doi:10.1016/j.cell.2014.05.010.

30. Zhang, C.; Quan, R.; Wang, J. Development and Application of CRISPR/Cas9 Technologies in Genomic Editing. *Hum. Mol. Genet.* **2018**, *27*, R79–R88, doi:10.1093/hmg/ddy120.
31. Hamar, J.; Kültz, D. An Efficient Vector-Based CRISPR/Cas9 System in an Oreochromis Mossambicus Cell Line Using Endogenous Promoters. *Sci. Rep.* **2021**, *11*, 7854, doi:10.1038/s41598-021-87068-3.
32. Zenke, K.; Okinaka, Y. Establishing an Effective Gene Knockdown System Using Cultured Cells of the Model Fish Medaka (Oryzias Latipes). *Biol. Methods Protoc.* **2022**, *7*, bpac011, doi:10.1093/biomethods/bpac011.
33. Menanteau-Ledouble, S.; Schachner, O.; Lawrence, M.L.; El-Matbouli, M. Effects of SiRNA Silencing on the Susceptibility of the Fish Cell Line CHSE-214 to Yersinia Ruckeri. *Vet. Res.* **2020**, *51*, 45, doi:10.1186/s13567-020-00760-6.
34. Collet, B.; Collins, C.; Cheyne, V.; Lester, K. Plasmid-Driven RNA Interference in Fish Cell Lines. *In Vitro Cell. Dev. Biol. Anim.* **2022**, *58*, 189–193, doi:10.1007/s11626-022-00645-2.
35. Fiol, D.F.; Kültz, D. Osmotic Stress Sensing and Signaling in Fishes. *FEBS J.* **2007**, *274*, 5790–5798, doi:10.1111/j.1742-4658.2007.06099.x.
36. Kültz, D.; Onken, H. Long-Term Acclimation of the Teleost Oreochromis Mossambicus to Various Salinities: Two Different Strategies in Mastering Hypertonic Stress. *Mar. Biol.* **1993**, *117*, 527–533, doi:10.1007/BF00349328.
37. Moorman, B.P.; Lerner, D.T.; Grau, E.G.; Seale, A.P. The Effects of Acute Salinity Challenges on Osmoregulation in Mozambique Tilapia Reared in a Tidally Changing Salinity. *J. Exp. Biol.* **2015**, *218*, 731–739, doi:10.1242/jeb.112664.
38. Breves, J.P.; Hasegawa, S.; Yoshioka, M.; Fox, B.K.; Davis, L.K.; Lerner, D.T.; Takei, Y.; Hirano, T.; Grau, E.G. Acute Salinity Challenges in Mozambique and Nile Tilapia: Differential Responses of Plasma Prolactin, Growth Hormone and Branchial Expression of Ion Transporters. *Gen. Comp. Endocrinol.* **2010**, *167*, 135–142, doi:10.1016/j.ygcen.2010.01.022.
39. Gardell, A.M.; Yang, J.; Sacchi, R.; Fangue, N.A.; Hammock, B.D.; Kültz, D. Tilapia (Oreochromis Mossambicus) Brain Cells Respond to Hyperosmotic Challenge by Inducing Myo-Inositol Biosynthesis. *J. Exp. Biol.* **2013**, *216*, 4615–4625, doi:10.1242/jeb.088906.

40. Kammerer, B.D.; Cech, J.J.; Kültz, D. Rapid Changes in Plasma Cortisol, Osmolality, and Respiration in Response to Salinity Stress in Tilapia (Oreochromis Mossambicus). *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **2010**, *157*, 260–265, doi:10.1016/j.cbpa.2010.07.009.
41. Wray, G.A. The Evolutionary Significance of Cis-Regulatory Mutations. *Nat. Rev. Genet.* **2007**, *8*, 206–216, doi:10.1038/nrg2063.
42. Wittkopp, P.J.; Kalay, G. Cis-Regulatory Elements: Molecular Mechanisms and Evolutionary Processes Underlying Divergence. *Nat. Rev. Genet.* **2011**, *13*, 59–69, doi:10.1038/nrg3095.
43. Ong, C.-T.; Corces, V.G. Enhancer Function: New Insights into the Regulation of Tissue-Specific Gene Expression. *Nat. Rev. Genet.* **2011**, *12*, 283–293, doi:10.1038/nrg2957.
44. Visel, A.; Blow, M.J.; Li, Z.; Zhang, T.; Akiyama, J.A.; Holt, A.; Plajzer-Frick, I.; Shoukry, M.; Wright, C.; Chen, F.; et al. ChIP-Seq Accurately Predicts Tissue-Specific Activity of Enhancers. *Nature* **2009**, *457*, 854–858, doi:10.1038/nature07730.
45. Dickel, D.E.; Visel, A.; Pennacchio, L.A. Functional Anatomy of Distant-Acting Mammalian Enhancers. *Philos. Trans. R. Soc. B Biol. Sci.* **2013**, *368*, doi:10.1098/rstb.2012.0359.
46. Rim, J.S.; Atta, M.G.; Dahl, S.C.; Berry, G.T.; Handler, J.S.; Kwon, H.M. Transcription of the Sodium/Myo-Inositol Cotransporter Gene Is Regulated by Multiple Tonicity-Responsive Enhancers Spread over 50 Kilobase Pairs in the 5'-Flanking Region. *J. Biol. Chem.* **1998**, *273*, 20615–20621, doi:10.1074/jbc.273.32.20615.
47. Pennacchio, L.A.; Bickmore, W.; Dean, A.; Nobrega, M.A.; Bejerano, G. Enhancers: Five Essential Questions. *Nat. Rev. Genet.* **2013**, *14*, 288–295, doi:10.1038/nrg3458.
48. Smith, A.N.; Barth, M.L.; McDowell, T.L.; Moulin, D.S.; Nuthall, H.N.; Hollingsworth, M.A.; Harris, A. A Regulatory Element in Intron 1 of the Cystic Fibrosis Transmembrane Conductance Regulator Gene. *J. Biol. Chem.* **1996**, *271*, 9947–9954, doi:10.1074/jbc.271.17.9947.
49. Cleves, P.A.; Hart, J.C.; Agoglia, R.M.; Jimenez, M.T.; Erickson, P.A.; Gai, L.; Miller, C.T. An Intronic Enhancer of Bmp6 Underlies Evolved Tooth Gain in Sticklebacks. *PLOS Genet.* **2018**, *14*, e1007449, doi:10.1371/journal.pgen.1007449.

50. Veauvy, C.M.; McDonald, M.D.; Van Audekerke, J.; Vanhoutte, G.; Van Camp, N.; Van der Linden, A.; Walsh, P.J. Ammonia Affects Brain Nitrogen Metabolism but Not Hydration Status in the Gulf Toadfish (Opsanus Beta). *Aquat. Toxicol. Amst. Neth.* **2005**, *74*, 32–46, doi:10.1016/j.aquatox.2005.05.003.
51. Essex-Fraser, P.A.; Steele, S.L.; Bernier, N.J.; Murray, B.W.; Stevens, E.D.; Wright, P.A. Expression of Four Glutamine Synthetase Genes in the Early Stages of Development of Rainbow Trout (Oncorhynchus Mykiss) in Relationship to Nitrogen Excretion. *J. Biol. Chem.* **2005**, *280*, 20268–20273, doi:10.1074/jbc.M412338200.
52. Webb, J.T.; Brown, G.W. Some Properties and Occurrence of Glutamine Synthetase in Fish. *Comp. Biochem. Physiol. B* **1976**, *54*, 171–175.
53. Chew, S.F.; Tng, Y.Y.M.; Wee, N.L.J.; Tok, C.Y.; Wilson, J.M.; Ip, Y.K. Intestinal Osmoregulatory Acclimation and Nitrogen Metabolism in Juveniles of the Freshwater Marble Goby Exposed to Seawater. *J. Comp. Physiol. [B]* **2010**, *180*, 511–520, doi:10.1007/s00360-009-0436-3.
54. Yancey, P.H. Organic Osmolytes as Compatible, Metabolic and Counteracting Cytoprotectants in High Osmolarity and Other Stresses. *J. Exp. Biol.* **2005**, *208*, 2819–2830, doi:10.1242/jeb.01730.
55. Tok, C.Y.; Chew, S.F.; Peh, W.Y.X.; Loong, A.M.; Wong, W.P.; Ip, Y.K. Glutamine Accumulation and Up-Regulation of Glutamine Synthetase Activity in the Swamp Eel, Monopterus Albus (Zuiew), Exposed to Brackish Water. *J. Exp. Biol.* **2009**, *212*, 1248–1258, doi:10.1242/jeb.025395.
56. Kopp, R.E.; Gilmore, E.A.; Little, C.M.; Lorenzo-Trueba, J.; Ramenzoni, V.C.; Sweet, W.V. Usable Science for Managing the Risks of Sea-Level Rise. *Earths Future* **2019**, *7*, 1235–1269, doi:10.1029/2018EF001145.
57. Salinity and Tides in Alluvial Estuaries - 1st Edition Available online: https://www.elsevier.com/books/salinity-and-tides-in-alluvial-estuaries/savenije/978-0-444-52107-1 (accessed on 28 May 2020).
58. Wedderburn, S.D.; Barnes, T.C.; Hillyard, K.A. Shifts in Fish Assemblages Indicate Failed Recovery of Threatened Species Following Prolonged Drought in Terminating Lakes of the Murray–Darling Basin, Australia. *Hydrobiologia* **2014**, *730*, 179–190, doi:10.1007/s10750-014-1836-2.

59. Cañedo-Argüelles, M.; Kefford, B.; Schäfer, R. Salt in Freshwaters: Causes, Effects and Prospects - Introduction to the Theme Issue. *Philos. Trans. R. Soc. B Biol. Sci.* **2019**, *374*, doi:10.1098/rstb.2018.0002.
60. Aquaculture Genomics, Genetics and Breeding Workshop; Abdelrahman, H.; ElHady, M.; Alcivar-Warren, A.; Allen, S.; Al-Tobasei, R.; Bao, L.; Beck, B.; Blackburn, H.; Bosworth, B.; et al. Aquaculture Genomics, Genetics and Breeding in the United States: Current Status, Challenges, and Priorities for Future Research. *BMC Genomics* **2017**, *18*, 191, doi:10.1186/s12864-017-3557-1.
61. Pino, L.K.; Searle, B.C.; Bollinger, J.G.; Nunn, B.; MacLean, B.; MacCoss, M.J. The Skyline Ecosystem: Informatics for Quantitative Mass Spectrometry Proteomics. *Mass Spectrom Rev* **2017**, *39*, 229–244, doi:10.1002/mas.21540.
62. Sharma, V.; Eckels, J.; Taylor, G.K.; Shulman, N.J.; Stergachis, A.B.; Joyner, S.A.; Yan, P.; Whiteaker, J.R.; Halusa, G.N.; Schilling, B.; et al. Panorama: A Targeted Proteomics Knowledge Base. *J Proteome Res* **2014**, *13*, 4205–4210, doi:10.1021/pr5006636.
63. Shaul, O. How Introns Enhance Gene Expression. *Int. J. Biochem. Cell Biol.* **2017**, *91*, 145–155, doi:10.1016/j.biocel.2017.06.016.
64. Kutach, A.K.; Kadonaga, J.T. The Downstream Promoter Element DPE Appears to Be as Widely Used as the TATA Box in Drosophila Core Promoters. *Mol. Cell. Biol.* **2000**, *20*, 4754–4764, doi:10.1128/mcb.20.13.4754-4764.2000.
65. Shlyueva, D.; Stampfel, G.; Stark, A. Transcriptional Enhancers: From Properties to Genome-Wide Predictions. *Nat. Rev. Genet.* **2014**, *15*, 272–286, doi:10.1038/nrg3682.
66. Visel, A.; Rubin, E.M.; Pennacchio, L.A. Genomic Views of Distant-Acting Enhancers. *Nature* **2009**, *461*, 199–205, doi:10.1038/nature08451.
67. Zomorodipour, A.; Jahromi, E.M.; Ataei, F.; Valimehr, S. Position Dependence of an Enhancer Activity of the Human Beta-Globin Intron-Ii, within a Heterologous Gene. *J. Mol. Med. Ther.* **2017**, *1*.
68. Marshall, W.S. 8 - Osmoregulation in Estuarine and Intertidal Fishes. In *Fish Physiology*; McCormick, S.D., Farrell, A.P., Brauner, C.J., Eds.; Euryhaline Fishes; Academic Press, 2012; Vol. 32, pp. 395–434.
69. Kalujnaia, S.; Gellatly, S.A.; Hazon, N.; Villasenor, A.; Yancey, P.H.; Cramb, G. Seawater Acclimation and Inositol Monophosphatase Isoform Expression in the European Eel

(Anguilla Anguilla) and Nile Tilapia (Orechromis Niloticus). *Am. J. Physiol.-Regul. Integr. Comp. Physiol.* **2013**, *305*, R369–R384, doi:10.1152/ajpregu.00044.2013.

70. Gardell, A.M.; Qin, Q.; Rice, R.H.; Li, J.; Kültz, D. Derivation and Osmotolerance Characterization of Three Immortalized Tilapia (Oreochromis Mossambicus) Cell Lines. *PLoS ONE* **2014**, *9*, doi:10.1371/journal.pone.0095919.
71. Diamond, J. Quantitative Evolutionary Design. *J. Physiol.* **2002**, *542*, 337–345, doi:10.1113/jphysiol.2002.018366.
72. Takenaka, M.; Preston, A.S.; Kwon, H.M.; Handler, J.S. The Tonicity-Sensitive Element That Mediates Increased Transcription of the Betaine Transporter Gene in Response to Hypertonic Stress. *J. Biol. Chem.* **1994**, *269*, 29379–29381.
73. Ferraris, J.D.; Williams, C.K.; Jung, K.-Y.; Bedford, J.J.; Burg, M.B.; García-Pérez, A. ORE, a Eukaryotic Minimal Essential Osmotic Response Element THE ALDOSE REDUCTASE GENE IN HYPEROSMOTIC STRESS. *J. Biol. Chem.* **1996**, *271*, 18318–18321, doi:10.1074/jbc.271.31.18318.
74. Bai, L.; Collins, J.F.; Muller, Y.L.; Xu, H.; Kiela, P.R.; Ghishan, F.K. Characterization of Cis-Elements Required for Osmotic Response of Rat Na(+)/H(+) Exchanger-2 (NHE-2) Gene. *Am. J. Physiol.* **1999**, *277*, R1112-1119, doi:10.1152/ajpregu.1999.277.4.R1112.
75. Ko, B.C.B.; Ruepp, B.; Bohren, K.M.; Gabbay, K.H.; Chung, S.S.M. Identification and Characterization of Multiple Osmotic Response Sequences in the Human Aldose Reductase Gene. *J. Biol. Chem.* **1997**, *272*, 16431–16437, doi:10.1074/jbc.272.26.16431.
76. Takeuchi, K.; Toyohara, H.; Kinoshita, M.; Sakaguchi, M. Role of Taurine in Hyperosmotic Stress Response of Fish Cells. *Fish. Sci.* **2002**, *68*, 1177–1180, doi:10.2331/fishsci.68.sup2_1177.
77. Takeuchi, K.; Toyohara, H.; Kinoshita, M.; Sakaguchi, M. Ubiquitous Increase in Taurine Transporter MRNA in Tissues of Tilapia (Oreochromis Mossambicus) during High-Salinity Adaptation. *Fish Physiol. Biochem.* **2000**, *23*, 173–182, doi:10.1023/A:1007889725718.
78. Ozasa, H.; Gould, K.G. Protective Effect of Taurine from Osmotic Stress on Chimpanzee Spermatozoa. *Arch. Androl.* **1982**, *9*, 121–126, doi:10.3109/01485018208990229.
79. Foskett, J.K.; Bern, H.A.; Machen, T.E.; Conner, M. Chloride Cells and the Hormonal Control of Teleost Fish Osmoregulation. *J. Exp. Biol.* **1983**, *106*, 255–281.

80. Rose, A.B. Requirements for Intron-Mediated Enhancement of Gene Expression in Arabidopsis. *RNA* **2002**, *8*, 1444–1453.
81. Wang, B.; Wang, H.; Gao, C.; Liu, Y.; Jin, C.; Sun, M.; Zhang, Q.; Qi, J. Functional Analysis of the Promoter Region of Japanese Flounder (Paralichthys Olivaceus) β-Actin Gene: A Useful Tool for Gene Research in Marine Fish. *Int. J. Mol. Sci.* **2018**, *19*, doi:10.3390/ijms19051401.
82. Bates, N.P.; Hurst, H.C. An Intron 1 Enhancer Element Mediates Oestrogen-Induced Suppression of ERBB2 Expression. *Oncogene* **1997**, *15*, 473–481, doi:10.1038/sj.onc.1201368.
83. Bruhat, A.; Tourmente, S.; Chapel, S.; Sobrier, M.L.; Couderc, J.L.; Dastugue, B. Regulatory Elements in the First Intron Contribute to Transcriptional Regulation of the Beta 3 Tubulin Gene by 20-Hydroxyecdysone in Drosophila Kc Cells. *Nucleic Acids Res.* **1990**, *18*, 2861–2867, doi:10.1093/nar/18.10.2861.
84. Gallegos, J.E.; Rose, A.B. The Enduring Mystery of Intron-Mediated Enhancement. *Plant Sci.* **2015**, *237*, 8–15, doi:10.1016/j.plantsci.2015.04.017.
85. Tourmente, S.; Chapel, S.; Dreau, D.; Drake, M.E.; Bruhat, A.; Couderc, J.L.; Dastugue, B. Enhancer and Silencer Elements within the First Intron Mediate the Transcriptional Regulation of the Beta 3 Tubulin Gene by 20-Hydroxyecdysone in Drosophila Kc Cells. *Insect Biochem. Mol. Biol.* **1993**, *23*, 137–143, doi:10.1016/0965-1748(93)90092-7.
86. Lis, M.; Walther, D. The Orientation of Transcription Factor Binding Site Motifs in Gene Promoter Regions: Does It Matter? *BMC Genomics* **2016**, *17*, 185, doi:10.1186/s12864-016-2549-x.
87. Kumada, Y.; Benson, D.R.; Hillemann, D.; Hosted, T.J.; Rochefort, D.A.; Thompson, C.J.; Wohlleben, W.; Tateno, Y. Evolution of the Glutamine Synthetase Gene, One of the Oldest Existing and Functioning Genes. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 3009–3013, doi:10.1073/pnas.90.7.3009.
88. Woo, S.K.; Dahl, S.C.; Handler, J.S.; Kwon, H.M. How Salt Regulates Genes: Function of a Rel-like Transcription Factor TonEBP. *Am. J. Physiol. Renal Physiol.* **2000**, *278*, F1006-1012, doi:10.1152/ajprenal.2000.278.6.F1006.
89. Cheung, C.Y.; Ko, B.C. NFAT5 in Cellular Adaptation to Hypertonic Stress – Regulations and Functional Significance. *J. Mol. Signal.* **2013**, *8*, 5, doi:10.1186/1750-2187-8-5.

90. Lorgen, M.; Jorgensen, E.H.; Jordan, W.C.; Martin, S.A.M.; Hazlerigg, D.G. NFAT5 Genes Are Part of the Osmotic Regulatory System in Atlantic Salmon (Salmo Salar). *Mar. Genomics* **2017**, *31*, 25–31, doi:10.1016/j.margen.2016.06.004.
91. López-Rodríguez, C.; Aramburu, J.; Jin, L.; Rakeman, A.S.; Michino, M.; Rao, A. Bridging the NFAT and NF-KB Families: NFAT5 Dimerization Regulates Cytokine Gene Transcription in Response to Osmotic Stress. *Immunity* **2001**, *15*, 47–58, doi:10.1016/S1074-7613(01)00165-0.
92. Fiol, D.F.; Kültz, D. Rapid Hyperosmotic Coinduction of Two Tilapia (Oreochromis Mossambicus) Transcription Factors in Gill Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 927–932, doi:10.1073/pnas.0408956102.
93. Gelev, V.; Zabolotny, J.M.; Lange, M.; Hiromura, M.; Yoo, S.W.; Orlando, J.S.; Kushnir, A.; Horikoshi, N.; Paquet, E.; Bachvarov, D.; et al. A New Paradigm for Transcription Factor TFIIB Functionality. *Sci. Rep.* **2014**, *4*, 3664, doi:10.1038/srep03664.
94. Kato, M.; Hata, N.; Banerjee, N.; Futcher, B.; Zhang, M.Q. Identifying Combinatorial Regulation of Transcription Factors and Binding Motifs. *Genome Biol.* **2004**, *5*, R56, doi:10.1186/gb-2004-5-8-r56.
95. Mandriani, B.; Castellana, S.; Rinaldi, C.; Manzoni, M.; Venuto, S.; Rodriguez-Aznar, E.; Galceran, J.; Nieto, M.A.; Borsani, G.; Monti, E.; et al. Identification of P53-Target Genes in *Danio Rerio*. *Sci. Rep.* **2016**, *6*, 32474, doi:10.1038/srep32474.
96. Wray, N.R.; Yang, J.; Hayes, B.J.; Price, A.L.; Goddard, M.E.; Visscher, P.M. Pitfalls of Predicting Complex Traits from SNPs. *Nat. Rev. Genet.* **2013**, *14*, 507–515, doi:10.1038/nrg3457.
97. Kültz, D. Physiological Mechanisms Used by Fish to Cope with Salinity Stress. *J. Exp. Biol.* **2015**, *218*, 1907–1914, doi:10.1242/jeb.118695.
98. Marie, A.D.; Smith, S.; Green, A.J.; Rico, C.; Lejeusne, C. *Transcriptomic Response to Thermal and Salinity Stress in Introduced and Native Sympatric Palaemon Caridean Shrimps. *Sci. Rep.* **2017**, *7*, 13980, doi:10.1038/s41598-017-13631-6.
99. Almeida-Dalmet, S.; Litchfield, C.D.; Gillevet, P.; Baxter, B.K. *Differential Gene Expression in Response to Salinity and Temperature in a Haloarcula Strain from Great Salt Lake, Utah. *Genes* **2018**, *9*, 52, doi:10.3390/genes9010052.

100. Wray, G.A.; Hahn, M.W.; Abouheif, E.; Balhoff, J.P.; Pizer, M.; Rockman, M.V.; Romano, L.A. *The Evolution of Transcriptional Regulation in Eukaryotes. *Mol. Biol. Evol.* **2003**, *20*, 1377–1419, doi:10.1093/molbev/msg140.

101. Zhang, B.; Wang, O.; Qin, J.; Liu, S.; Sun, S.; Liu, H.; Kuang, J.; Jiang, G.; Zhang, W. Cis-Acting Elements and Trans-Acting Factors in the Transcriptional Regulation of Raf Kinase Inhibitory Protein Expression. *PloS One* **2013**, *8*, e83097, doi:10.1371/journal.pone.0083097.

102. Mattioli, K.; Oliveros, W.; Gerhardinger, C.; Andergassen, D.; Maass, P.G.; Rinn, J.L.; Melé, M. Cis and Trans Effects Differentially Contribute to the Evolution of Promoters and Enhancers. *Genome Biol.* **2020**, *21*, 210, doi:10.1186/s13059-020-02110-3.

103. Berman, B.P.; Pfeiffer, B.D.; Laverty, T.R.; Salzberg, S.L.; Rubin, G.M.; Eisen, M.B.; Celniker, S.E. Computational Identification of Developmental Enhancers: Conservation and Function of Transcription Factor Binding-Site Clusters in Drosophila Melanogaster and Drosophila Pseudoobscura. *Genome Biol.* **2004**, *5*, R61, doi:10.1186/gb-2004-5-9-r61.

104. Huang, C.W.; Li, Y.H.; Hu, S.Y.; Chi, J.R.; Lin, G.H.; Lin, C.C.; Gong, H.Y.; Chen, J.Y.; Chen, R.H.; Chang, S.J.; et al. Differential Expression Patterns of Growth-Related MicroRNAs in the Skeletal Muscle of Nile Tilapia (Oreochromis Niloticus)1. *J. Anim. Sci.* **2012**, *90*, 4266–4279, doi:10.2527/jas.2012-5142.

105. Konstantinidis, I.; Sætrom, P.; Mjelle, R.; Nedoluzhko, A.V.; Robledo, D.; Fernandes, J.M.O. Major Gene Expression Changes and Epigenetic Remodelling in Nile Tilapia Muscle after Just One Generation of Domestication. *Epigenetics* **2020**, *15*, 1052–1067, doi:10.1080/15592294.2020.1748914.

106. Root, L.T.; Kültz, D. Gill Proteome Networks Explain Energy Homeostasis during Salinity Stress in Oreochromis Mossambicus. *Mol Ecol* **2022**, *in review*, Authorea: https://doi.org/10.22541/au.164433207.74361091/v1.

107. Qin, H.; Yu, Z.; Zhu, Z.; Lin, Y.; Xia, J.; Jia, Y. The Integrated Analyses of Metabolomics and Transcriptomics in Gill of GIFT Tilapia in Response to Long Term Salinity Challenge. *Aquac. Fish.* **2021**, doi:10.1016/j.aaf.2021.02.006.

108. Kültz, D.; Li, J.; Gardell, A.; Sacchi, R. Quantitative Molecular Phenotyping of Gill Remodeling in a Cichlid Fish Responding to Salinity Stress *. *Mol. Cell. Proteomics* **2013**, *12*, 3962–3975, doi:10.1074/mcp.M113.029827.

109. Moorman, B.P.; Yamaguchi, Y.; Lerner, D.T.; Grau, E.G.; Seale, A.P. Rearing Mozambique Tilapia in Tidally-Changing Salinities: Effects on Growth and the Growth Hormone/Insulin-like Growth Factor I Axis. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **2016**, *198*, 8–14, doi:10.1016/j.cbpa.2016.03.014.

110. Kim, C.; Kültz, D. An Osmolality/Salinity-Responsive Enhancer 1 (OSRE1) in Intron 1 Promotes Salinity Induction of Tilapia Glutamine Synthetase. *Sci. Rep.* **2020**, *10*, 12103, doi:10.1038/s41598-020-69090-z.

111. Zarka, D.G.; Vogel, J.T.; Cook, D.; Thomashow, M.F. Cold Induction of Arabidopsis CBF Genes Involves Multiple ICE (Inducer of CBF Expression) Promoter Elements and a Cold-Regulatory Circuit That Is Desensitized by Low Temperature. *Plant Physiol.* **2003**, *133*, 910–918, doi:10.1104/pp.103.027169.

112. Hughes, J.D.; Estep, P.W.; Tavazoie, S.; Church, G.M. Computational Identification of Cis-Regulatory Elements Associated with Groups of Functionally Related Genes in Saccharomyces Cerevisiae. *J. Mol. Biol.* **2000**, *296*, 1205–1214, doi:10.1006/jmbi.2000.3519.

113. Elkon, R.; Zeller, K.I.; Linhart, C.; Dang, C.V.; Shamir, R.; Shiloh, Y. In Silico Identification of Transcriptional Regulators Associated with C-Myc. *Nucleic Acids Res.* **2004**, *32*, 4955–4961, doi:10.1093/nar/gkh816.

114. Ma, S.; Bohnert, H.J. Integration of Arabidopsis Thaliana Stress-Related Transcript Profiles, Promoter Structures, and Cell-Specific Expression. *Genome Biol.* **2007**, *8*, R49, doi:10.1186/gb-2007-8-4-r49.

115. 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. Distinct and Predictive Chromatin Signatures of Transcriptional Promoters and Enhancers in the Human Genome. *Nat. Genet.* **2007**, *39*, 311–318, doi:10.1038/ng1966.

116. Hu, P.; Liu, M.; Zhang, D.; Wang, J.; Niu, H.; Liu, Y.; Wu, Z.; Han, B.; Zhai, W.; Shen, Y.; et al. Global Identification of the Genetic Networks and Cis-Regulatory Elements of the Cold Response in Zebrafish. *Nucleic Acids Res.* **2015**, *43*, 9198–9213, doi:10.1093/nar/gkv780.

117. Hu, P.; Liu, M.; Liu, Y.; Wang, J.; Zhang, D.; Niu, H.; Jiang, S.; Wang, J.; Zhang, D.; Han, B.; et al. Transcriptome Comparison Reveals a Genetic Network Regulating the Lower Temperature Limit in Fish. *Sci. Rep.* **2016**, *6*, doi:10.1038/srep28952.
118. Cassé, C.; Giannoni, F.; Nguyen, V.T.; Dubois, M.-F.; Bensaude, O. The Transcriptional Inhibitors, Actinomycin D and α-Amanitin, Activate the HIV-1 Promoter and Favor Phosphorylation of the RNA Polymerase II C-Terminal Domain*. *J. Biol. Chem.* **1999**, *274*, 16097–16106, doi:10.1074/jbc.274.23.16097.
119. Koba, M.; Konopa, J. [Actinomycin D and its mechanisms of action]. *Postepy Hig. Med. Doswiadczalnej Online* **2005**, *59*, 290–298.
120. Li, J.; Levitan, B.; Gomez-Jimenez, S.; Kültz, D. Development of a Gill Assay Library for Ecological Proteomics of Threespine Sticklebacks (Gasterosteus Aculeatus). *Mol. Cell. Proteomics MCP* **2018**, *17*, 2146–2163, doi:10.1074/mcp.RA118.000973.
121. Gillet, L.C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol Cell Proteomics* **2012**, *11*, O111 016717, doi:10.1074/mcp.O111.016717.
122. Collins, B.C.; Gillet, L.C.; Rosenberger, G.; Rost, H.L.; Vichalkovski, A.; Gstaiger, M.; Aebersold, R. Quantifying Protein Interaction Dynamics by SWATH Mass Spectrometry: Application to the 14-3-3 System. *Nat Methods* **2013**, *10*, 1246–1253, doi:10.1038/nmeth.2703.
123. Arnhard, K.; Gottschall, A.; Pitterl, F.; Oberacher, H. Applying "Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra" (SWATH) for Systematic Toxicological Analysis with Liquid Chromatography-High-Resolution Tandem Mass Spectrometry. *Anal Bioanal Chem* **2015**, *407*, 405–414, doi:10.1007/s00216-014-8262-1.
124. Huang, Q.; Yang, L.; Luo, J.; Guo, L.; Wang, Z.; Yang, X.; Jin, W.; Fang, Y.; Ye, J.; Shan, B.; et al. SWATH Enables Precise Label-Free Quantification on Proteome Scale. *Proteomics* **2015**, *15*, 1215–1223, doi:10.1002/pmic.201400270.
125. Koopmans, F.; Ho, J.T.C.; Smit, A.B.; Li, K.W. Comparative Analyses of Data Independent Acquisition Mass Spectrometric Approaches: DIA, WiSIM-DIA, and Untargeted DIA. *Proteomics* **2018**, *18*, 1–6, doi:10.1002/pmic.201700304.

126. Root, L.; Campo, A.; MacNiven, L.; Con, P.; Cnaani, A.; Kültz, D. Nonlinear Effects of Environmental Salinity on the Gill Transcriptome versus Proteome of Oreochromis Niloticus. *Genomics* **2021**, *113*, 3235–3249.
127. Reiter, L.; Rinner, O.; Picotti, P.; Huttenhain, R.; Beck, M.; Brusniak, M.Y.; Hengartner, M.O.; Aebersold, R. MProphet: Automated Data Processing and Statistical Validation for Large-Scale SRM Experiments. *Nat Methods* **2011**, *8*, 430–435, doi:10.1038/nmeth.1584.
128. Choi, M.; Chang, C.Y.; Clough, T.; Broudy, D.; Killeen, T.; MacLean, B.; Vitek, O. MSstats: An R Package for Statistical Analysis of Quantitative Mass Spectrometry-Based Proteomic Experiments. *Bioinformatics* **2014**, *30*, 2524–2526, doi:10.1093/bioinformatics/btu305.
129. Deutsch, E.W.; Bandeira, N.; Sharma, V.; Perez-Riverol, Y.; Carver, J.J.; Kundu, D.J.; García-Seisdedos, D.; Jarnuczak, A.F.; Hewapathirana, S.; Pullman, B.S.; et al. The ProteomeXchange Consortium in 2020: Enabling 'Big Data' Approaches in Proteomics. *Nucleic Acids Res.* **2020**, *48*, D1145–D1152, doi:10.1093/nar/gkz984.
130. Bailey, T.L. STREME: Accurate and Versatile Sequence Motif Discovery. *bioRxiv* **2020**, 2020.11.23.394619, doi:10.1101/2020.11.23.394619.
131. Gupta, S.; Stamatoyannopoulos, J.A.; Bailey, T.L.; Noble, W.S. Quantifying Similarity between Motifs. *Genome Biol.* **2007**, *8*, R24, doi:10.1186/gb-2007-8-2-r24.
132. Grant, C.E.; Bailey, T.L.; Noble, W.S. FIMO: Scanning for Occurrences of a given Motif. *Bioinformatics* **2011**, *27*, 1017–1018, doi:10.1093/bioinformatics/btr064.
133. Yang, L.; Zhou, T.; Dror, I.; Mathelier, A.; Wasserman, W.W.; Gordân, R.; Rohs, R. TFBSshape: A Motif Database for DNA Shape Features of Transcription Factor Binding Sites. *Nucleic Acids Res.* **2014**, *42*, D148–D155, doi:10.1093/nar/gkt1087.
134. Salekin, S.; Zhang, J.M.; Huang, Y. Base-Pair Resolution Detection of Transcription Factor Binding Site by Deep Deconvolutional Network. *Bioinformatics* **2018**, *34*, 3446–3453, doi:10.1093/bioinformatics/bty383.
135. Hena, A.; Kamal, M.; Mair, G.C. Salinity Tolerance in Superior Genotypes of Tilapia, Oreochromis Niloticus, Oreochromis Mossambicus and Their Hybrids. *Aquaculture* **2005**, *247*, 189–201, doi:10.1016/j.aquaculture.2005.02.008.
136. Firmat, C.; Alibert, P.; Losseau, M.; Baroiller, J.-F.; Schliewen, U.K. Successive Invasion-Mediated Interspecific Hybridizations and Population Structure in the Endangered Cichlid

Oreochromis Mossambicus. *PLOS ONE* **2013**, *8*, e63880, doi:10.1371/journal.pone.0063880.

137. Sheshadri, S.A.; Nishanth, M.J.; Simon, B. Stress-Mediated Cis-Element Transcription Factor Interactions Interconnecting Primary and Specialized Metabolism in Planta. *Front. Plant Sci.* **2016**, *7*, 1725, doi:10.3389/fpls.2016.01725.
138. Lin, L.-H.; Lee, H.-C.; Li, W.-H.; Chen, B.-S. A Systematic Approach to Detecting Transcription Factors in Response to Environmental Stresses. *BMC Bioinformatics* **2007**, *8*, 473, doi:10.1186/1471-2105-8-473.
139. Estruch, F. Stress-Controlled Transcription Factors, Stress-Induced Genes and Stress Tolerance in Budding Yeast. *FEMS Microbiol. Rev.* **2000**, *24*, 469–486, doi:10.1111/j.1574-6976.2000.tb00551.x.
140. Thompson, A.C.; Capellini, T.D.; Guenther, C.A.; Chan, Y.F.; Infante, C.R.; Menke, D.B.; Kingsley, D.M. A Novel Enhancer near the Pitx1 Gene Influences Development and Evolution of Pelvic Appendages in Vertebrates. *eLife* **2018**, *7*, e38555, doi:10.7554/eLife.38555.
141. Gallardo-Fuentes, L.; Santos-Pereira, J.M.; Tena, J.J. Functional Conservation of Divergent P63-Bound Cis-Regulatory Elements. *Front. Genet.* **2020**, *11*.
142. Fiol, D.F.; Kültz, D. Rapid Hyperosmotic Coinduction of Two Tilapia (Oreochromis Mossambicus) Transcription Factors in Gill Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 927–932, doi:10.1073/pnas.0408956102.
143. Fiol, D.F.; Mak, S.K.; Kültz, D. Specific TSC22 Domain Transcripts Are Hypertonically Induced and Alternatively Spliced to Protect Mouse Kidney Cells during Osmotic Stress. *FEBS J.* **2007**, *274*, 109–124, doi:10.1111/j.1742-4658.2006.05569.x.
144. Tse, W.K.F. The Role of Osmotic Stress Transcription Factor 1 in Fishes. *Front. Zool.* **2014**, *11*, doi:10.1186/s12983-014-0086-5.
145. Wong, M.K.-S.; Ozaki, H.; Suzuki, Y.; Iwasaki, W.; Takei, Y. Discovery of Osmotic Sensitive Transcription Factors in Fish Intestine via a Transcriptomic Approach. *BMC Genomics* **2014**, *15*, 1134, doi:10.1186/1471-2164-15-1134.
146. Verta, J.-P.; Jones, F.C. Predominance of Cis-Regulatory Changes in Parallel Expression Divergence of Sticklebacks. *eLife* **2019**, *8*, e43785, doi:10.7554/eLife.43785.

147. Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global Quantification of Mammalian Gene Expression Control. *Nature* **2011**, *473*, 337–342, doi:10.1038/nature10098.

148. Suhre, K.; McCarthy, M.I.; Schwenk, J.M. Genetics Meets Proteomics: Perspectives for Large Population-Based Studies. *Nat. Rev. Genet.* **2021**, *22*, 19–37, doi:10.1038/s41576-020-0268-2.

149. Buccitelli, C.; Selbach, M. MRNAs, Proteins and the Emerging Principles of Gene Expression Control. *Nat. Rev. Genet.* **2020**, 1–15, doi:10.1038/s41576-020-0258-4.

150. Pascal, L.E.; True, L.D.; Campbell, D.S.; Deutsch, E.W.; Risk, M.; Coleman, I.M.; Eichner, L.J.; Nelson, P.S.; Liu, A.Y. Correlation of MRNA and Protein Levels: Cell Type-Specific Gene Expression of Cluster Designation Antigens in the Prostate. *BMC Genomics* **2008**, *9*, 246, doi:10.1186/1471-2164-9-246.

151. Evans, T.G.; Somero, G.N. A Microarray-Based Transcriptomic Time-Course of Hyper- and Hypo-Osmotic Stress Signaling Events in the Euryhaline Fish Gillichthys Mirabilis: Osmosensors to Effectors. *J. Exp. Biol.* **2008**, *211*, 3636–3649, doi:10.1242/jeb.022160.

152. Zhang, X.; Wen, H.; Wang, H.; Ren, Y.; Zhao, J.; Li, Y. RNA-Seq Analysis of Salinity Stress–Responsive Transcriptome in the Liver of Spotted Sea Bass (Lateolabrax Maculatus). *PLoS ONE* **2017**, *12*, e0173238, doi:10.1371/journal.pone.0173238.

153. Su, H.; Ma, D.; Zhu, H.; Liu, Z.; Gao, F. Transcriptomic Response to Three Osmotic Stresses in Gills of Hybrid Tilapia (Oreochromis Mossambicus Female × O. Urolepis Hornorum Male). *BMC Genomics* **2020**, *21*, 110, doi:10.1186/s12864-020-6512-5.

154. Tse, W.K.F.; Chow, S.C.; Wong, C.K.C. The Cloning of Eel Osmotic Stress Transcription Factor and the Regulation of Its Expression in Primary Gill Cell Culture. *J. Exp. Biol.* **2008**, *211*, 1964–1968, doi:10.1242/jeb.017368.

155. Gardell, A.M.; Qin, Q.; Rice, R.H.; Li, J.; Kültz, D. Derivation and Osmotolerance Characterization of Three Immortalized Tilapia (Oreochromis Mossambicus) Cell Lines. *PloS One* **2014**, *9*, e95919, doi:10.1371/journal.pone.0095919.

156. Reiter, F.; Wienerroither, S.; Stark, A. Combinatorial Function of Transcription Factors and Cofactors. *Curr. Opin. Genet. Dev.* **2017**, *43*, 73–81, doi:10.1016/j.gde.2016.12.007.

157. Sacilotto, N.; Monteiro, R.; Fritzsche, M.; Becker, P.W.; Sanchez-del-Campo, L.; Liu, K.; Pinheiro, P.; Ratnayaka, I.; Davies, B.; Goding, C.R.; et al. Analysis of Dll4 Regulation

Reveals a Combinatorial Role for Sox and Notch in Arterial Development. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 11893–11898, doi:10.1073/pnas.1300805110.

158. Pham, V.N.; Lawson, N.D.; Mugford, J.W.; Dye, L.; Castranova, D.; Lo, B.; Weinstein, B.M. Combinatorial Function of ETS Transcription Factors in the Developing Vasculature. *Dev. Biol.* **2007**, *303*, 772–783, doi:10.1016/j.ydbio.2006.10.030.
159. Khamici, H.A.; Brown, L.J.; Hossain, K.R.; Hudson, A.L.; Sinclair-Burton, A.A.; Ng, J.P.M.; Daniel, E.L.; Hare, J.E.; Cornell, B.A.; Curmi, P.M.G.; et al. Members of the Chloride Intracellular Ion Channel Protein Family Demonstrate Glutaredoxin-Like Enzymatic Activity. *PLOS ONE* **2015**, *10*, e115699, doi:10.1371/journal.pone.0115699.
160. Ueno, Y.; Ozaki, S.; Umakoshi, A.; Yano, H.; Choudhury, M.E.; Abe, N.; Sumida, Y.; Kuwabara, J.; Uchida, R.; Islam, A.; et al. Chloride Intracellular Channel Protein 2 in Cancer and Non-Cancer Human Tissues: Relationship with Tight Junctions. *Tissue Barriers* **2019**, *7*, 1593775, doi:10.1080/21688370.2019.1593775.
161. Tipsmark, C.K.; Baltzegar, D.A.; Ozden, O.; Grubb, B.J.; Borski, R.J. Salinity Regulates Claudin MRNA and Protein Expression in the Teleost Gill. *Am. J. Physiol.-Regul. Integr. Comp. Physiol.* **2008**, *294*, R1004–R1014, doi:10.1152/ajpregu.00112.2007.
162. Tang, V.W.; Goodenough, D.A. Paracellular Ion Channel at the Tight Junction. *Biophys. J.* **2003**, *84*, 1660–1673, doi:10.1016/S0006-3495(03)74975-3.
163. Pizzorno, G.; Cao, D.; Leffert, J.J.; Russell, R.L.; Zhang, D.; Handschumacher, R.E. Homeostatic Control of Uridine and the Role of Uridine Phosphorylase: A Biological and Clinical Update. *Biochim. Biophys. Acta* **2002**, *1587*, 133–144, doi:10.1016/s0925-4439(02)00076-5.
164. Watanabe, S.; Uchida, T. Cloning and Expression of Human Uridine Phosphorylase. *Biochem. Biophys. Res. Commun.* **1995**, *216*, 265–272, doi:10.1006/bbrc.1995.2619.
165. Kültz, D. Molecular and Evolutionary Basis of the Cellular Stress Response. *Annu. Rev. Physiol.* **2005**, *67*, 225–257, doi:10.1146/annurev.physiol.67.040403.103635.
166. Wozniak, K.J.; Simmons, L.A. Hydroxyurea Induces a Stress Response That Alters DNA Replication and Nucleotide Metabolism in Bacillus Subtilis. *J. Bacteriol.* **2021**, *203*, e0017121, doi:10.1128/JB.00171-21.

167. Kilstrup, M.; Hammer, K.; Ruhdal Jensen, P.; Martinussen, J. Nucleotide Metabolism and Its Control in Lactic Acid Bacteria. *FEMS Microbiol. Rev.* **2005**, *29*, 555–590, doi:10.1016/j.fmrre.2005.04.006.
168. Lyu, L.; Wen, H.; Li, Y.; Li, J.; Zhao, J.; Zhang, S.; Song, M.; Wang, X. Deep Transcriptomic Analysis of Black Rockfish (Sebastes Schlegelii) Provides New Insights on Responses to Acute Temperature Stress. *Sci. Rep.* **2018**, *8*, 9113, doi:10.1038/s41598-018-27013-z.
169. Woo, S.; Denis, V.; Yum, S. Transcriptional Changes Caused by Bisphenol A in Oryzias Javanicus, a Fish Species Highly Adaptable to Environmental Salinity. *Mar. Drugs* **2014**, *12*, 983–998, doi:10.3390/md12020983.
170. O'Shields, B.; McArthur, A.G.; Holowiecki, A.; Kamper, M.; Tapley, J.; Jenny, M.J. Inhibition of Endogenous MTF-1 Signaling in Zebrafish Embryos Identifies Novel Roles for MTF-1 in Development. *Biochim. Biophys. Acta* **2014**, *1843*, 1818–1833, doi:10.1016/j.bbamcr.2014.04.015.
171. van Loo, K.M.J.; Schaub, C.; Pitsch, J.; Kulbida, R.; Opitz, T.; Ekstein, D.; Dalal, A.; Urbach, H.; Beck, H.; Yaari, Y.; et al. Zinc Regulates a Key Transcriptional Pathway for Epileptogenesis via Metal-Regulatory Transcription Factor 1. *Nat. Commun.* **2015**, *6*, 8688, doi:10.1038/ncomms9688.
172. Stevanovic, M.; Drakulic, D.; Lazic, A.; Ninkovic, D.S.; Schwirtlich, M.; Mojsin, M. SOX Transcription Factors as Important Regulators of Neuronal and Glial Differentiation During Nervous System Development and Adult Neurogenesis. *Front. Mol. Neurosci.* **2021**, *14*.
173. Inukai, S.; Kock, K.H.; Bulyk, M.L. Transcription Factor-DNA Binding: Beyond Binding Site Motifs. *Curr. Opin. Genet. Dev.* **2017**, *43*, 110–119, doi:10.1016/j.gde.2017.02.007.
174. Geertz, M.; Maerkl, S.J. Experimental Strategies for Studying Transcription Factor-DNA Binding Specificities. *Brief. Funct. Genomics* **2010**, *9*, 362–373, doi:10.1093/bfgp/elq023.
175. Berger, M.F.; Philippakis, A.A.; Qureshi, A.M.; He, F.S.; Estep, P.W.; Bulyk, M.L. Compact, Universal DNA Microarrays to Comprehensively Determine Transcription-Factor Binding Site Specificities. *Nat. Biotechnol.* **2006**, *24*, 1429–1435, doi:10.1038/nbt1246.

176. Andrilenas, K.K.; Penvose, A.; Siggers, T. Using Protein-Binding Microarrays to Study Transcription Factor Specificity: Homologs, Isoforms and Complexes. *Brief. Funct. Genomics* **2015**, *14*, 17–29, doi:10.1093/bfgp/elu046.

177. Gordân, R.; Murphy, K.F.; McCord, R.P.; Zhu, C.; Vedenko, A.; Bulyk, M.L. Curated Collection of Yeast Transcription Factor DNA Binding Specificity Data Reveals Novel Structural and Gene Regulatory Insights. *Genome Biol.* **2011**, *12*, R125, doi:10.1186/gb-2011-12-12-r125.

178. Nakagawa, S.; Gisselbrecht, S.S.; Rogers, J.M.; Hartl, D.L.; Bulyk, M.L. DNA-Binding Specificity Changes in the Evolution of Forkhead Transcription Factors. *Proc. Natl. Acad. Sci.* **2013**, *110*, 12349–12354, doi:10.1073/pnas.1310430110.

179. Badis, G.; Berger, M.F.; Philippakis, A.A.; Talukder, S.; Gehrke, A.R.; Jaeger, S.A.; Chan, E.T.; Metzler, G.; Vedenko, A.; Chen, X.; et al. Diversity and Complexity in DNA Recognition by Transcription Factors. *Science* **2009**, *324*, 1720–1723, doi:10.1126/science.1162327.

180. Chen, A.; Zhong, L.; Lv, J. FOXL1 Overexpression Is Associated with Poor Outcome in Patients with Glioma. *Oncol. Lett.* **2019**, *18*, 751–757, doi:10.3892/ol.2019.10351.

181. Aoki, R.; Shoshkes-Carmel, M.; Gao, N.; Shin, S.; May, C.L.; Golson, M.L.; Zahm, A.M.; Ray, M.; Wiser, C.L.; Wright, C.V.E.; et al. Foxl1-Expressing Mesenchymal Cells Constitute the Intestinal Stem Cell Niche. *Cell. Mol. Gastroenterol. Hepatol.* **2016**, *2*, 175–188, doi:10.1016/j.jcmgh.2015.12.004.

182. Perreault, N.; Katz, J.P.; Sackett, S.D.; Kaestner, K.H. Foxl1 Controls the Wnt/β-Catenin Pathway by Modulating the Expression of Proteoglycans in the Gut *. *J. Biol. Chem.* **2001**, *276*, 43328–43333, doi:10.1074/jbc.M104366200.

183. Zhang, G.; He, P.; Gaedcke, J.; Ghadimi, B.M.; Ried, T.; Yfantis, H.G.; Lee, D.H.; Hanna, N.; Alexander, H.R.; Hussain, S.P. FOXL1, a Novel Candidate Tumor Suppressor, Inhibits Tumor Aggressiveness and Predicts Outcome in Human Pancreatic Cancer. *Cancer Res.* **2013**, *73*, 5416–5425, doi:10.1158/0008-5472.CAN-13-0362.

184. Nakada, C.; Satoh, S.; Tabata, Y.; Arai, K.; Watanabe, S. Transcriptional Repressor Foxl1 Regulates Central Nervous System Development by Suppressing Shh Expression in Zebra Fish. *Mol. Cell. Biol.* **2006**, *26*, 7246–7257, doi:10.1128/MCB.00429-06.

185. Li, J.; Levitan, B.B.; Gomez-Jimenez, S.; Kültz, D. Ecological Proteomics of Three-Spine Sticklebacks (Gasterosteus Aculeatus) with a Standardized Gill DIA Assay. *Integr. Comp. Biol.* **2018**, *58*, e133.

186. Dang, C.V.; Resar, L.M.; Emison, E.; Kim, S.; Li, Q.; Prescott, J.E.; Wonsey, D.; Zeller, K. Function of the C-Myc Oncogenic Transcription Factor. *Exp. Cell Res.* **1999**, *253*, 63–77, doi:10.1006/excr.1999.4686.

187. Schmidt, E.V. The Role of C-Myc in Cellular Growth Control. *Oncogene* **1999**, *18*, 2988–2996, doi:10.1038/sj.onc.1202751.

188. Zeller, K.I.; Jegga, A.G.; Aronow, B.J.; O'Donnell, K.A.; Dang, C.V. An Integrated Database of Genes Responsive to the Myc Oncogenic Transcription Factor: Identification of Direct Genomic Targets. *Genome Biol.* **2003**, *4*, R69, doi:10.1186/gb-2003-4-10-r69.

189. Atchley, W.R.; Fitch, W.M. Myc and Max: Molecular Evolution of a Family of Proto-Oncogene Products and Their Dimerization Partner. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 10217–10221.

190. Sarid, J.; Halazonetis, T.D.; Murphy, W.; Leder, P. Evolutionarily Conserved Regions of the Human C-Myc Protein Can Be Uncoupled from Transforming Activity. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 170–173.

191. Moura, M.T.; Silva, R.L.O.; Cantanhêde, L.F.; Ferreira-Silva, J.C.; Nascimento, P.S.; Benko-Iseppon, A.M.; Oliveira, M.A.L. Evolutionary-Driven C-MYC Gene Expression in Mammalian Fibroblasts. *Sci. Rep.* **2020**, *10*, 11056, doi:10.1038/s41598-020-67391-x.

192. Kültz, D. Evolution of Cellular Stress Response Mechanisms. *J. Exp. Zool. Part Ecol. Integr. Physiol.* **2020**, doi:10.1002/jez.2347.

193. Ma, A.; Cui, W.; Wang, X.; Zhang, W.; Liu, Z.; Zhang, J.; Zhao, T. Osmoregulation by the Myo-Inositol Biosynthesis Pathway in Turbot Scophthalmus Maximus and Its Regulation by Anabolite and c-Myc. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **2020**, *242*, 110636, doi:10.1016/j.cbpa.2019.110636.

194. Verma, D.; Jalmi, S.K.; Bhagat, P.K.; Verma, N.; Sinha, A.K. A BHLH Transcription Factor, MYC2, Imparts Salt Intolerance by Regulating Proline Biosynthesis in Arabidopsis. *FEBS J.* **2020**, *287*, 2560–2576, doi:10.1111/febs.15157.

195. Valenzuela, C.E.; Acevedo-Acevedo, O.; Miranda, G.S.; Vergara-Barros, P.; Holuigue, L.; Figueroa, C.R.; Figueroa, P.M. Salt Stress Response Triggers Activation of the Jasmonate

Signaling Pathway Leading to Inhibition of Cell Elongation in Arabidopsis Primary Root. *J. Exp. Bot.* **2016**, *67*, 4209–4220, doi:10.1093/jxb/erw202.

196. Yoon, Y.; Seo, D.H.; Shin, H.; Kim, H.J.; Kim, C.M.; Jang, G. The Role of Stress-Responsive Transcription Factors in Modulating Abiotic Stress Tolerance in Plants. *Agronomy* **2020**, *10*, 788, doi:10.3390/agronomy10060788.
197. Nahas, L.D.; Al-Husein, N.; Lababidi, G.; Hamwieh, A. In-Silico Prediction of Novel Genes Responsive to Drought and Salinity Stress Tolerance in Bread Wheat (Triticum Aestivum). *PLOS ONE* **2019**, *14*, e0223962, doi:10.1371/journal.pone.0223962.
198. O'Connell, B.C.; Cheung, A.F.; Simkevich, C.P.; Tam, W.; Ren, X.; Mateyak, M.K.; Sedivy, J.M. A Large Scale Genetic Analysis of C-Myc-Regulated Gene Expression Patterns * 210. *J. Biol. Chem.* **2003**, *278*, 12563–12573, doi:10.1074/jbc.M210462200.
199. Barrangou, R.; Doudna, J.A. Applications of CRISPR Technologies in Research and Beyond. *Nat. Biotechnol.* **2016**, *34*, 933–941, doi:10.1038/nbt.3659.
200. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A.; Charpentier, E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337*, 816–821, doi:10.1126/science.1225829.
201. Wiedenheft, B.; Sternberg, S.H.; Doudna, J.A. RNA-Guided Genetic Silencing Systems in Bacteria and Archaea. *Nature* **2012**, *482*, 331–338, doi:10.1038/nature10886.
202. Hilton, I.B.; Gersbach, C.A. Enabling Functional Genomics with Genome Engineering. *Genome Res.* **2015**, *25*, 1442–1455, doi:10.1101/gr.190124.115.
203. Li, H.; Yang, Y.; Hong, W.; Huang, M.; Wu, M.; Zhao, X. Applications of Genome Editing Technology in the Targeted Therapy of Human Diseases: Mechanisms, Advances and Prospects. *Signal Transduct. Target. Ther.* **2020**, *5*, 1–23, doi:10.1038/s41392-019-0089-y.
204. Burgio, G. Redefining Mouse Transgenesis with CRISPR/Cas9 Genome Editing Technology. *Genome Biol.* **2018**, *19*, 27, doi:10.1186/s13059-018-1409-1.
205. Campenhout, C.V.; Cabochette, P.; Veillard, A.-C.; Laczik, M.; Zelisko-Schmidt, A.; Sabatel, C.; Dhainaut, M.; Vanhollebeke, B.; Gueydan, C.; Kruys, V. Guidelines for Optimized Gene Knockout Using CRISPR/Cas9. *BioTechniques* **2019**, *66*, 295–302, doi:10.2144/btn-2018-0187.
206. Hana, S.; Peterson, M.; McLaughlin, H.; Marshall, E.; Fabian, A.J.; McKissick, O.; Koszka, K.; Marsh, G.; Craft, M.; Xu, S.; et al. Highly Efficient Neuronal Gene Knockout in Vivo

by CRISPR-Cas9 via Neonatal Intracerebroventricular Injection of AAV in Mice. *Gene Ther.* **2021**, *28*, 646–658, doi:10.1038/s41434-021-00224-2.

207. Mou, H.; Kennedy, Z.; Anderson, D.G.; Yin, H.; Xue, W. Precision Cancer Mouse Models through Genome Editing with CRISPR-Cas9. *Genome Med.* **2015**, *7*, 53, doi:10.1186/s13073-015-0178-7.
208. Edvardsen, R.B.; Leininger, S.; Kleppe, L.; Skaftnesmo, K.O.; Wargelius, A. Targeted Mutagenesis in Atlantic Salmon (Salmo Salar L.) Using the CRISPR/Cas9 System Induces Complete Knockout Individuals in the F0 Generation. *PloS One* **2014**, *9*, e108622, doi:10.1371/journal.pone.0108622.
209. Cleveland, B.M.; Yamaguchi, G.; Radler, L.M.; Shimizu, M. Editing the Duplicated Insulin-like Growth Factor Binding Protein-2b Gene in Rainbow Trout (Oncorhynchus Mykiss). *Sci. Rep.* **2018**, *8*, 16054, doi:10.1038/s41598-018-34326-6.
210. Irion, U.; Krauss, J.; Nüsslein-Volhard, C. Precise and Efficient Genome Editing in Zebrafish Using the CRISPR/Cas9 System. *Dev. Camb. Engl.* **2014**, *141*, 4827–4830, doi:10.1242/dev.115584.
211. Hruscha, A.; Krawitz, P.; Rechenberg, A.; Heinrich, V.; Hecht, J.; Haass, C.; Schmid, B. Efficient CRISPR/Cas9 Genome Editing with Low off-Target Effects in Zebrafish. *Dev. Camb. Engl.* **2013**, *140*, 4982–4987, doi:10.1242/dev.099085.
212. Dehler, C.E.; Boudinot, P.; Martin, S.A.M.; Collet, B. Development of an Efficient Genome Editing Method by CRISPR/Cas9 in a Fish Cell Line. *Mar. Biotechnol. N. Y. N* **2016**, *18*, 449–452, doi:10.1007/s10126-016-9708-6.
213. Escobar-Aguirre, S.; Arancibia, D.; Escorza, A.; Bravo, C.; Andrés, M.E.; Zamorano, P.; Martínez, V. Development of a Bicistronic Vector for the Expression of a CRISPR/Cas9-MCherry System in Fish Cell Lines. *Cells* **2019**, *8*, E75, doi:10.3390/cells8010075.
214. Gratacap, R.L.; Regan, T.; Dehler, C.E.; Martin, S.A.M.; Boudinot, P.; Collet, B.; Houston, R.D. Efficient CRISPR/Cas9 Genome Editing in a Salmonid Fish Cell Line Using a Lentivirus Delivery System. *BMC Biotechnol.* **2020**, *20*, 35, doi:10.1186/s12896-020-00626-x.
215. Li, M.; Yang, H.; Zhao, J.; Fang, L.; Shi, H.; Li, M.; Sun, Y.; Zhang, X.; Jiang, D.; Zhou, L.; et al. Efficient and Heritable Gene Targeting in Tilapia by CRISPR/Cas9. *Genetics* **2014**, *197*, 591–599, doi:10.1534/genetics.114.163667.

216. Giuliano, C.J.; Lin, A.; Girish, V.; Sheltzer, J.M. Generating Single Cell-Derived Knockout Clones in Mammalian Cells with CRISPR/Cas9. *Curr. Protoc. Mol. Biol.* **2019**, *128*, e100, doi:10.1002/cpmb.100.
217. Ye, M.; Wilhelm, M.; Gentschev, I.; Szalay, A. A Modified Limiting Dilution Method for Monoclonal Stable Cell Line Selection Using a Real-Time Fluorescence Imaging System: A Practical Workflow and Advanced Applications. *Methods Protoc.* **2021**, *4*, 16, doi:10.3390/mps4010016.
218. Yang, R.; Lemaître, V.; Huang, C.; Haddadi, A.; McNaughton, R.; Espinosa, H.D. Monoclonal Cell Line Generation and CRISPR/Cas9 Manipulation via Single-Cell Electroporation. *Small Weinh. Bergstr. Ger.* **2018**, *14*, e1702495, doi:10.1002/smll.201702495.
219. Fenerty, K.E.; Padget, M.; Wolfson, B.; Gameiro, S.R.; Su, Z.; Lee, J.H.; Rabizadeh, S.; Soon-Shiong, P.; Hodge, J.W. Immunotherapy Utilizing the Combination of Natural Killer- and Antibody Dependent Cellular Cytotoxicity (ADCC)-Mediating Agents with Poly (ADP-Ribose) Polymerase (PARP) Inhibition. *J. Immunother. Cancer* **2018**, *6*, 133, doi:10.1186/s40425-018-0445-4.
220. Liu, Q.; Yuan, Y.; Zhu, F.; Hong, Y.; Ge, R. Efficient Genome Editing Using CRISPR/Cas9 Ribonucleoprotein Approach in Cultured Medaka Fish Cells. *Biol. Open* **2018**, *7*, bio035170, doi:10.1242/bio.035170.
221. Dehler, C.E.; Lester, K.; Pelle, G.D.; Jouneau, L.; Houel, A.; Collins, C.; Dovgan, T.; Machat, R.; Zou, J.; Boudinot, P.; et al. Viral Resistance and IFN Signaling in STAT2 Knockout Fish Cells. *J. Immunol.* **2019**, *203*, 465–475, doi:10.4049/jimmunol.1801376.
222. Gratacap, R.L.; Wargelius, A.; Edvardsen, R.B.; Houston, R.D. Potential of Genome Editing to Improve Aquaculture Breeding and Production. *Trends Genet. TIG* **2019**, *35*, 672–684, doi:10.1016/j.tig.2019.06.006.
223. Yang, Z.; Yu, Y.; Tay, Y.X.; Yue, G.H. Genome Editing and Its Applications in Genetic Improvement in Aquaculture. *Rev. Aquac.* **2022**, *14*, 178–191, doi:10.1111/raq.12591.
224. Gardell, A.M.; Qin, Q.; Rice, R.H.; Li, J.; Kültz, D. Derivation and Osmotolerance Characterization of Three Immortalized Tilapia (Oreochromis Mossambicus) Cell Lines. *PLOS ONE* **2014**, *9*, e95919, doi:10.1371/journal.pone.0095919.

225. Blum, M.; Chang, H.-Y.; Chuguransky, S.; Grego, T.; Kandasaamy, S.; Mitchell, A.; Nuka, G.; Paysan-Lafosse, T.; Qureshi, M.; Raj, S.; et al. The InterPro Protein Families and Domains Database: 20 Years On. *Nucleic Acids Res.* **2021**, *49*, D344–D354, doi:10.1093/nar/gkaa977.

226. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Res.* **2022**, *50*, D439–D444, doi:10.1093/nar/gkab1061.

227. Sehnal, D.; Bittrich, S.; Deshpande, M.; Svobodová, R.; Berka, K.; Bazgier, V.; Velankar, S.; Burley, S.K.; Koča, J.; Rose, A.S. Mol* Viewer: Modern Web App for 3D Visualization and Analysis of Large Biomolecular Structures. *Nucleic Acids Res.* **2021**, *49*, W431–W437, doi:10.1093/nar/gkab314.

228. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242, doi:10.1093/nar/28.1.235.

229. Roche, D.B.; Buenavista, M.T.; McGuffin, L.J. FunFOLDQA: A Quality Assessment Tool for Protein-Ligand Binding Site Residue Predictions. *PloS One* **2012**, *7*, e38219, doi:10.1371/journal.pone.0038219.

230. de la Cova, C.; Johnston, L.A. Myc in Model Organisms: A View from the Flyroom. *Semin. Cancer Biol.* **2006**, *16*, 303–312, doi:10.1016/j.semcancer.2006.07.010.

231. Dang, C.V. C-Myc Target Genes Involved in Cell Growth, Apoptosis, and Metabolism. *Mol. Cell. Biol.* **1999**, *19*, 1–11.

232. Ahmadi, S.E.; Rahimi, S.; Zarandi, B.; Chegeni, R.; Safa, M. MYC: A Multipurpose Oncogene with Prognostic and Therapeutic Implications in Blood Malignancies. *J. Hematol. Oncol.J Hematol Oncol* **2021**, *14*, 121, doi:10.1186/s13045-021-01111-4.

233. Dang, C.V. MYC on the Path to Cancer. *Cell* **2012**, *149*, 22–35, doi:10.1016/j.cell.2012.03.003.

234. Chou, S.-J.; Tole, S. Lhx2, an Evolutionarily Conserved, Multifunctional Regulator of Forebrain Development. *Brain Res.* **2019**, *1705*, 1–14, doi:10.1016/j.brainres.2018.02.046.

235. Burg, M.B.; Kwon, E.D.; Kültz, D. Regulation of Gene Expression by Hypertonicity. *Annu. Rev. Physiol.* **1997**, *59*, 437–455, doi:10.1146/annurev.physiol.59.1.437.

236. Kim, C.; Kültz, D. An Osmolality/Salinity-Responsive Enhancer 1 (OSRE1) in Intron 1 Promotes Salinity Induction of Tilapia Glutamine Synthetase. *Sci. Rep.* **2020**, *10*, 12103, doi:10.1038/s41598-020-69090-z.
237. Kim, C.; Wang, X.; Kültz, D. Prediction and Experimental Validation of a New Salinity-Responsive Cis-Regulatory Element (CRE) in a Tilapia Cell Line. *Life Basel Switz.* **2022**, *12*, 787, doi:10.3390/life12060787.
238. Kültz, D.; Fiol, D.; Valkova, N.; Gomez-Jimenez, S.; Chan, S.Y.; Lee, J. Functional Genomics and Proteomics of the Cellular Osmotic Stress Response in "non-Model" Organisms. *J. Exp. Biol.* **2007**, *210*, 1593–1601, doi:10.1242/jeb.000141.
239. Kültz, D.; Li, J.; Gardell, A.; Sacchi, R. Quantitative Molecular Phenotyping of Gill Remodeling in a Cichlid Fish Responding to Salinity Stress. *Mol. Cell. Proteomics* **2013**, *12*, 3962–3975, doi:10.1074/mcp.M113.029827.
240. Root, L.; Campo, A.; MacNiven, L.; Cnaani, A.; Kültz, D. A Data-Independent Acquisition (DIA) Assay Library for Quantitation of Environmental Effects on the Kidney Proteome of *Oreochromis Niloticus*. *Mol Ecol Approaches* **2021**, *21*, 2486–2503.
241. Jayaram, N.; Usvyat, D.; R. Martin, A.C. Evaluating Tools for Transcription Factor Binding Site Prediction. *BMC Bioinformatics* **2016**, *17*, 547, doi:10.1186/s12859-016-1298-9.
242. Yadav, S.; Yadava, Y.K.; Kohli, D.; Meena, S.; Kalwan, G.; Bharadwaj, C.; Gaikwad, K.; Arora, A.; Jain, P.K. Genome-Wide Identification, in Silico Characterization and Expression Analysis of the RNA Helicase Gene Family in Chickpea (C. Arietinum L.). *Sci. Rep.* **2022**, *12*, 9778, doi:10.1038/s41598-022-13823-9.
243. Arndell, T.; Sharma, N.; Langridge, P.; Baumann, U.; Watson-Haigh, N.S.; Whitford, R. GRNA Validation for Wheat Genome Editing with the CRISPR-Cas9 System. *BMC Biotechnol.* **2019**, *19*, 71, doi:10.1186/s12896-019-0565-z.
244. Etard, C.; Joshi, S.; Stegmaier, J.; Mikut, R.; Strähle, U. Tracking of Indels by DEcomposition Is a Simple and Effective Method to Assess Efficiency of Guide RNAs in Zebrafish. *Zebrafish* **2017**, *14*, 586–588, doi:10.1089/zeb.2017.1454.
245. Development of an Efficient Genome Editing Method by CRISPR/Cas9 in a Fish Cell Line | SpringerLink Available online: https://link.springer.com/article/10.1007/s10126-016-9708-6 (accessed on 22 September 2022).

246. Lopez, A.; Fernandez-Alonso, M.; Rocha, A.; Estepa, A.; Coll, J.M. Transfection of Epithelioma Papulosum Cyprini (EPC) Carp Cells. *Biotechnol. Lett.* **2001**, *23*, 481–487, doi:10.1023/A:1010393723002.

247. Collet, B.; Collins, C.; Lester, K. Engineered Cell Lines for Fish Health Research. *Dev. Comp. Immunol.* **2018**, *80*, 34–40, doi:10.1016/j.dci.2017.01.013.

248. Gratacap, R.L.; Jin, Y.H.; Mantsopoulou, M.; Houston, R.D. Efficient Genome Editing in Multiple Salmonid Cell Lines Using Ribonucleoprotein Complexes. *Mar. Biotechnol.* **2020**, *22*, 717–724, doi:10.1007/s10126-020-09995-y.

249. Weinguny, M.; Klanert, G.; Eisenhut, P.; Jonsson, A.; Ivansson, D.; Lövgren, A.; Borth, N. Directed Evolution Approach to Enhance Efficiency and Speed of Outgrowth during Single Cell Subcloning of Chinese Hamster Ovary Cells. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1320–1329, doi:10.1016/j.csbj.2020.05.020.

250. Munoz, A.; Morachis, J.M. High Efficiency Sorting and Outgrowth for Single-Cell Cloning of Mammalian Cell Lines. *Biotechnol. Lett.* **2022**, doi:10.1007/s10529-022-03300-8.

251. Brinkman, E.K.; Chen, T.; Amendola, M.; van Steensel, B. Easy Quantitative Assessment of Genome Editing by Sequence Trace Decomposition. *Nucleic Acids Res.* **2014**, *42*, e168, doi:10.1093/nar/gku936.

252. Sentmanat, M.F.; Peters, S.T.; Florian, C.P.; Connelly, J.P.; Pruett-Miller, S.M. A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Sci. Rep.* **2018**, *8*, 888, doi:10.1038/s41598-018-19441-8.

253. Conant, D.; Hsiau, T.; Rossi, N.; Oki, J.; Maures, T.; Waite, K.; Yang, J.; Joshi, S.; Kelso, R.; Holden, K.; et al. Inference of CRISPR Edits from Sanger Trace Data. *CRISPR J.* **2022**, *5*, 123–130, doi:10.1089/crispr.2021.0113.

254. Clement, K.; Rees, H.; Canver, M.C.; Gehrke, J.M.; Farouni, R.; Hsu, J.Y.; Cole, M.A.; Liu, D.R.; Joung, J.K.; Bauer, D.E.; et al. CRISPResso2 Provides Accurate and Rapid Genome Editing Sequence Analysis. *Nat. Biotechnol.* **2019**, *37*, 224–226, doi:10.1038/s41587-019-0032-3.

255. Jin, J.; Xu, Y.; Huo, L.; Ma, L.; Scott, A.W.; Pizzi, M.P.; Li, Y.; Wang, Y.; Yao, X.; Song, S.; et al. An Improved Strategy for CRISPR/Cas9 Gene Knockout and Subsequent Wildtype and Mutant Gene Rescue. *PLOS ONE* **2020**, *15*, e0228910, doi:10.1371/journal.pone.0228910.

256. Ma, J.; Fan, Y.; Zhou, Y.; Liu, W.; Jiang, N.; Zhang, J.; Zeng, L. Efficient Resistance to Grass Carp Reovirus Infection in JAM-A Knockout Cells Using CRISPR/Cas9. *Fish Shellfish Immunol.* **2018**, *76*, 206–215, doi:10.1016/j.fsi.2018.02.039.

257. Kim, M.S.; Shin, M.J.; Kim, K.H. Increase of Viral Hemorrhagic Septicemia Virus Growth by Knockout of IRF9 Gene in Epithelioma Papulosum Cyprini Cells. *Fish Shellfish Immunol.* **2018**, *83*, 443–448, doi:10.1016/j.fsi.2018.09.025.

258. Houston, R.D.; Bean, T.P.; Macqueen, D.J.; Gundappa, M.K.; Jin, Y.H.; Jenkins, T.L.; Selly, S.L.C.; Martin, S.A.M.; Stevens, J.R.; Santos, E.M.; et al. Harnessing Genomics to Fast-Track Genetic Improvement in Aquaculture. *Nat. Rev. Genet.* **2020**, *21*, 389–409, doi:10.1038/s41576-020-0227-y.

259. Omeka, W.K.M.; Liyanage, D.S.; Lee, S.; Lim, C.; Yang, H.; Sandamalika, W.M.G.; Udayantha, H.M.V.; Kim, G.; Ganeshalingam, S.; Jeong, T.; et al. Genome-Wide Association Study (GWAS) of Growth Traits in Olive Flounder (Paralichthys Olivaceus). *Aquaculture* **2022**, *555*, 738257, doi:10.1016/j.aquaculture.2022.738257.

260. Zhou, Y.; Fu, H.-C.; Wang, Y.-Y.; Huang, H.-Z. Genome-Wide Association Study Reveals Growth-Related SNPs and Candidate Genes in Mandarin Fish (Siniperca Chuatsi). *Aquaculture* **2022**, *550*, 737879, doi:10.1016/j.aquaculture.2021.737879.

261. Tsai, H.-Y.; Hamilton, A.; Tinch, A.E.; Guy, D.R.; Gharbi, K.; Stear, M.J.; Matika, O.; Bishop, S.C.; Houston, R.D. Genome Wide Association and Genomic Prediction for Growth Traits in Juvenile Farmed Atlantic Salmon Using a High Density SNP Array. *BMC Genomics* **2015**, *16*, 969, doi:10.1186/s12864-015-2117-9.

262. Xie, C.; Bekpen, C.; Künzel, S.; Keshavarz, M.; Krebs-Wheaton, R.; Skrabar, N.; Ullrich, K.K.; Zhang, W.; Tautz, D. Dedicated Transcriptomics Combined with Power Analysis Lead to Functional Understanding of Genes with Weak Phenotypic Changes in Knockout Lines. *PLoS Comput. Biol.* **2020**, *16*, e1008354, doi:10.1371/journal.pcbi.1008354.

263. Dolgalev, G.; Poverennaya, E. Applications of CRISPR-Cas Technologies to Proteomics. *Genes* **2021**, *12*, 1790, doi:10.3390/genes12111790.

264. Das, S.K.; Lewis, B.A.; Levens, D. MYC: A Complex Problem. *Trends Cell Biol.* **2022**, *0*, doi:10.1016/j.tcb.2022.07.006.

265. Wang, Y.-Y.; Hsu, S.-H.; Tsai, H.-Y.; Cheng, F.-Y.; Cheng, M.-C. Transcriptomic and Proteomic Analysis of CRISPR/Cas9-Mediated ARC-Knockout HEK293 Cells. *Int. J. Mol. Sci.* **2022**, *23*, 4498, doi:10.3390/ijms23094498.
266. Zhang, J.; Kim, S.; Li, L.; Kemp, C.J.; Jiang, C.; Lü, J. Proteomic and Transcriptomic Profiling of Pten Gene-Knockout Mouse Model of Prostate Cancer. *The Prostate* **2020**, *80*, 588–605, doi:10.1002/pros.23972.
267. Barrett, L.W.; Fletcher, S.; Wilton, S.D. Regulation of Eukaryotic Gene Expression by the Untranslated Gene Regions and Other Non-Coding Elements. *Cell. Mol. Life Sci.* **2012**, *69*, 3613–3634, doi:10.1007/s00018-012-0990-9.
268. Kültz, D. The Combinatorial Nature of Osmosensing in Fishes. *Physiol. Bethesda Md* **2012**, *27*, 259–275, doi:10.1152/physiol.00014.2012.