scientific reports

Check for updates

OPEN

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

Chanhee Kim[1], Avner Cnaani[2] & Dietmar Kültz[1]✉

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 and an optimized vector-based CRISPR/Cas9 system to functionally disrupt MYC in the tilapia genome and to establish causal links between MYC and cell functions, including cellular osmoregulation. A cell isolation and dilution strategy yielded polyclonal *myca* (a gene encoding MYC) knockout (ko) cell pools with low genetic variability and high gene editing efficiencies (as high as 98.2%). Subsequent isolation and dilution of cells from these pools produced a *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 wild-type cells in morphology. As such, they represent a new tool for causally linking *myca* to cellular osmoregulation and other cell functions.

MYC family genes encode class III basic helix-turn-helix (bHLH) transcription factors (MYC TF) that have been evolutionarily conserved in many animal species for at least 400 million years[1,2]. They are known to 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[1,3]. 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 TF also harbors a less common motif, a transcriptional regulatory domain, that is less evolutionarily conserved across species than the DNA-binding and dimerization domains[4]. This transcriptional regulatory domain was shown to strongly control transformation of rat embryo fibroblasts, as evidenced by using a domain deletion mutant[5]. Moreover, it has been suggested that deletion of any of the conserved domains outlined above can significantly impair MYC function[6]. Several lines of indirect evidence suggest that MYC TF is involved in governing cellular osmoregulation[7]. 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 an RNAi-mediated knockdown approach[8]. MYC has also been reported to directly modulate responses to abiotic stressors, including salinity stress, in plants (e.g., *Arabidopsis thaliana*[9,10]). This modulation is mediated via key hormonal signaling pathways important for plant salinity tolerance[11]. Such salinity stress-induced, non-canonical roles of MYC TF in plants were also indicated by an *in-silico* prediction study using bread wheat (*Triticum aestivum*)[12]. Moreover, microarray analyses of MYC ko rat cell lines revealed inositol monophosphatase (IMPA, A2 isoform), a key enzyme in the *myo*-inositol

[1]Department of Animal Sciences, University of California, Davis, CA 95616, USA. [2]Department of Poultry and Aquaculture, Institute of Animal Sciences, Agricultural Research Organization, Volcani Center, 7528809 Rishon LeZion, Israel. ✉email: dkueltz@ucdavis.edu

biosynthesis (MIB) compatible osmolyte pathway, as a target gene of MYC[13]. These studies suggest that MYC TF is important for controlling osmoregulatory mechanisms in eukaryotes.

The recent revolution of gene targeting approaches by implementing CRISPR/Cas-based approaches has enabled highly accurate and efficient genome editing that is superior to older gene targeting methods such as TALENs or ZFNs, which require covalent linkage of a specific DNA binding domain to a nuclease[14–16]. This innovative system was initially adopted from bacteria and archaea, in which it had evolved as a pathogen nucleic acid-targeting defense mechanism that conferred resistance to viral infection[17,18]. 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[19]. Most studies using this system to date have focused on model organisms such as human cell lines[20] and mouse[21] both in vivo and in vitro. They have demonstrated the great power of knockout (ko) models for functional studies aimed at causally linking genotypes and phenotypes[22–24]. 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[25–28]. For instance, double-allelic ko mutations were introduced in Atlantic salmon (*Salmo salar*) to alter pigmentation[25], Insulin-like Growth Factor Binding Protein-2b (IGF-BP2b) was targeted for ko in rainbow trout (*Oncorhynchus mykiss*)[26], alligator cathelicidin gene was targeted in a non-coding region of channel catfish (*Ictalurus punctatus*) genome (Simora et al. 2020) and myostatin ko mutations were performed in Channel catfish (Ictalurus punctatus) to improve growth and disease resistance (Coogan et al., 2022). Examples for in vivo gene targeting in fish also include zebrafish (*Danio rerio*) where all-in-one CRISPR/Cas9 components were injected into fertilized one-cell stage embryos to generate ko mutants[28].

Compared to a considerable number of CRISPR/Cas9 studies in whole fish in vivo, gene editing of fish cell lines in vitro has been rarely used and remains to be explored as a promising tool for high-throughput and low-cost establishment of causality between cellular phenotypes and genomic loci of interest. The first fish cell line that was genetically modified by CRISPR/Cas9 technology was reported for Chinook salmon (*Oncorhynchus tshawytscha*)[29]. Chinook salmon cell lines were also used to demonstrate the functionality of a vector-based expression system[30], as well as to optimize lentivirus-mediated infection for efficient delivery of recombinant DNA into host cells[31]. Our lab has recently optimized and established a vector-based CRISPR/Cas9 platform for tilapia cell lines[32]. This study was the first to enable in vitro gene targeting in euryhaline tilapia cells. Our 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[33].

Recent CRISPR/Cas9 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[34–36]. 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 wild-type or variable mutant proteins remains a concern[37]. Therefore, recent efforts have been geared towards generating homozygous clonal ko cell models, to ensure that the protein function is completely disrupted and that the resulting mutant 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[38]. 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*[39]. 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[40,41].

In this study, genetically engineered polyclonal and monoclonal tilapia cell lines were generated to facilitate studies of cellular functions of 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 has been 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[32] 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[42]; 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) 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).

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 gRNA design by providing an integrated reference genome of *O. niloticus* (Ensembl 76-Ore-nil1.0) for calculating off-target effect of sgRNAs by comparison to the 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 cloned into an optimized tilapia sgRNA expression vector, as described previously[32]. Complementary oligonucleotides (Eurofins Genomics) comprising each sgRNA's forward and reverse sequences (Table S1) were annealed to generate a ClaI restriction site at the 5' end an 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 S1) to confirm successful insertion of sgRNA target sequence.

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 for 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[32]. Briefly, medium was removed and cells surviving hygromycin B treatment 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 °C 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-gRNAscaff-HygroR 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[34] 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 and Fig. S3) 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 100 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 PureLink Genomic DNA mini Kit (Invitrogen) following manufacturer instructions. The test amplicon spanning the targeted region of *myca* was PCR-amplified using appropriate primers (Primer pair: myca_TideF1 and myca_TideR2, Table S1) and purified. Sanger sequencing was carried out at the DNA Sequencing Facility of UC Davis 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 truncated protein the structure of *myca* ko truncated protein was compared to wild-type 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 wild-type) MYC sequences were annotated with functional domains using InterPro version 90.0 (EMBL-EBI)[44]. Moreover, the 3D protein structures were predicted for both (mutant and wild-type) MYC proteins and visualized using AlphaFold[45], Mol*3D Viewer[46], and RCSB Protein Data Bank[47]. Combined with the above prediction tools, FunFOLDQA[48], 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 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[6]. 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* (Supplementary Fig. 1). To identify the domains that are important for tilapia MYC function and ensure that all functionally important domains are inactivated by CRISPR/Cas9 gene targeting, the tilapia *myca* gene (Gene ID: 100689989) was identified in the NCBI reference genome sequence for *Oreochromis 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[44] (EMBL-EBI). This approach 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 (Fig. 1a,d).

To identify the evolutionarily most highly conserved regions in these functional domains, the tilapia MYC protein sequence was aligned to MYC sequences from 14 other vertebrate species. The transcriptional regulatory domain was not found conserved as a whole but rather contained multiple regions that were more highly conserved and interspersed between more divergent stretches of amino acid sequence (Fig. 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) (Fig. 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 sgRNAs—**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 means that the truncated mutant protein produced by this sgRNA still contains two highly conserved sequence blocks of the transcriptional regulatory domain (Fig. 1c). PCR primers for amplifying a region that includes 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 (Supplementary Fig. 2).

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 of 16.9% (TIDE) and 6% (ICE) (Fig. 2a).

To improve the gene editing efficiency of selected cells, isolate cells with biallelic *myca* ko, and eliminate genetic heterogeneity of isolated mutant cell populations, 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 (Fig. 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. The highest gene editing efficiency was 98.4% (TIDE) and 92% (ICE) for sgRNA1 colony #4 (Table S2).
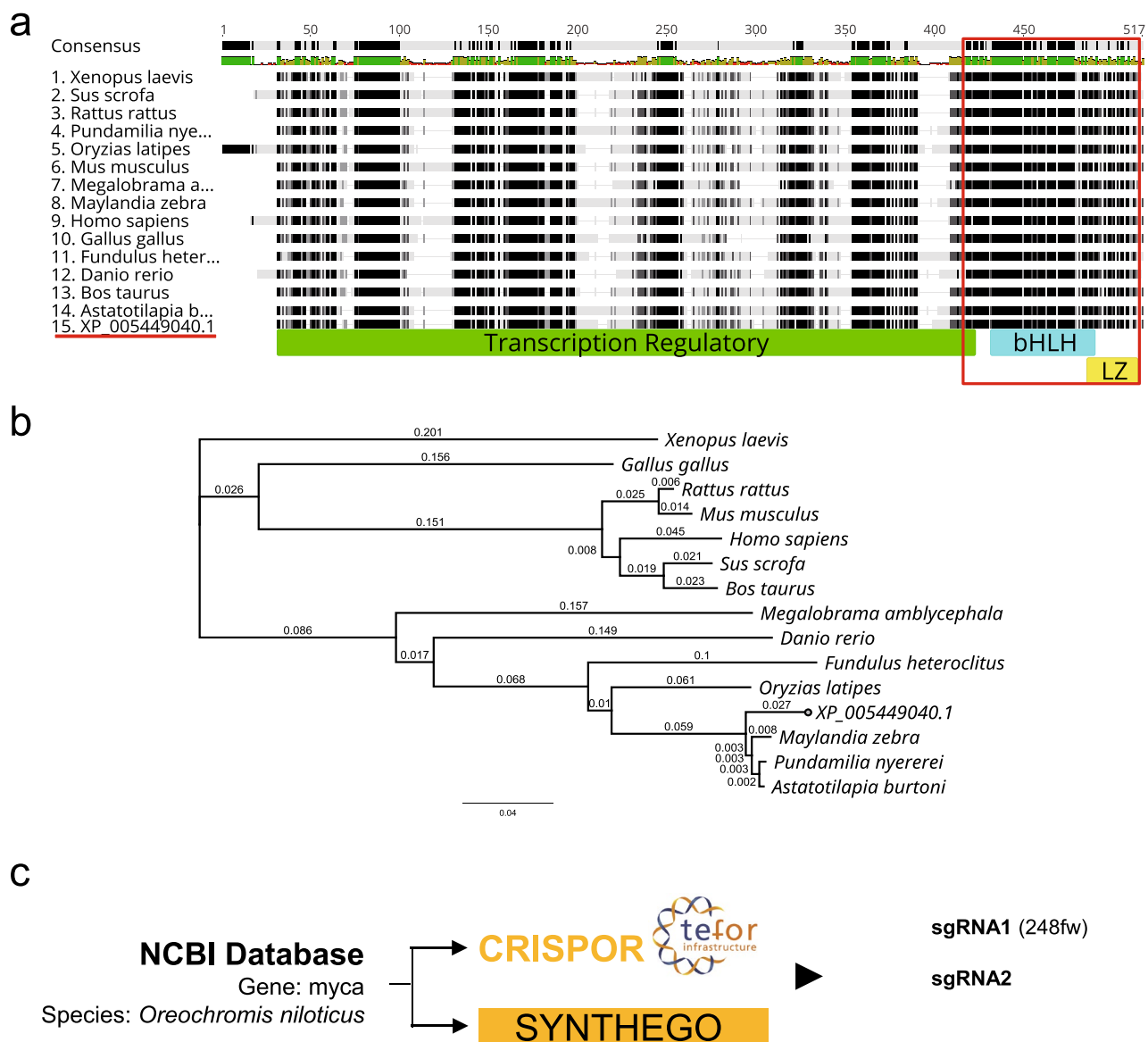
Figure 1. Evolutionary conservation of MYC protein domains and location of *myca* sgRNAs for CRIPSR/Cas9-mediated gene-editing. (**a**) Multiple alignment of MYC protein sequences from 15 vertebrate species including mammals, amphibians, and fishes generated with Geneious (Biomatters, cost matrix = Blosum62). The names of species included in the alignment are listed at the left side. The Oreochromis niloticus MYC protein sequence is labeled by its NCBI accession number (XP_005449040.1) 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.

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 suggested that, although virtually all cells in sgRNA1-colony#4 had been mutated, less than half yielded a functionally severely impaired protein since 58% did not harbor a frameshift mutation or a deletion of more than seven 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. For example, sgRNA1-colony#1 had a TIDE score of 48.6%, ICE score of 47%, and a KO-score of 47/100. sgRNA1-colony#2 had a TIDE score of 89.4%, ICE score of 91%, and a KO-score of 91/100. sgRNA2-colony#1 had a TIDE score of 84.9%, ICE score of 89%, and a KO-score of 88/100. sgRNA2-colony#4 had a TIDE score of 53%, ICE score of
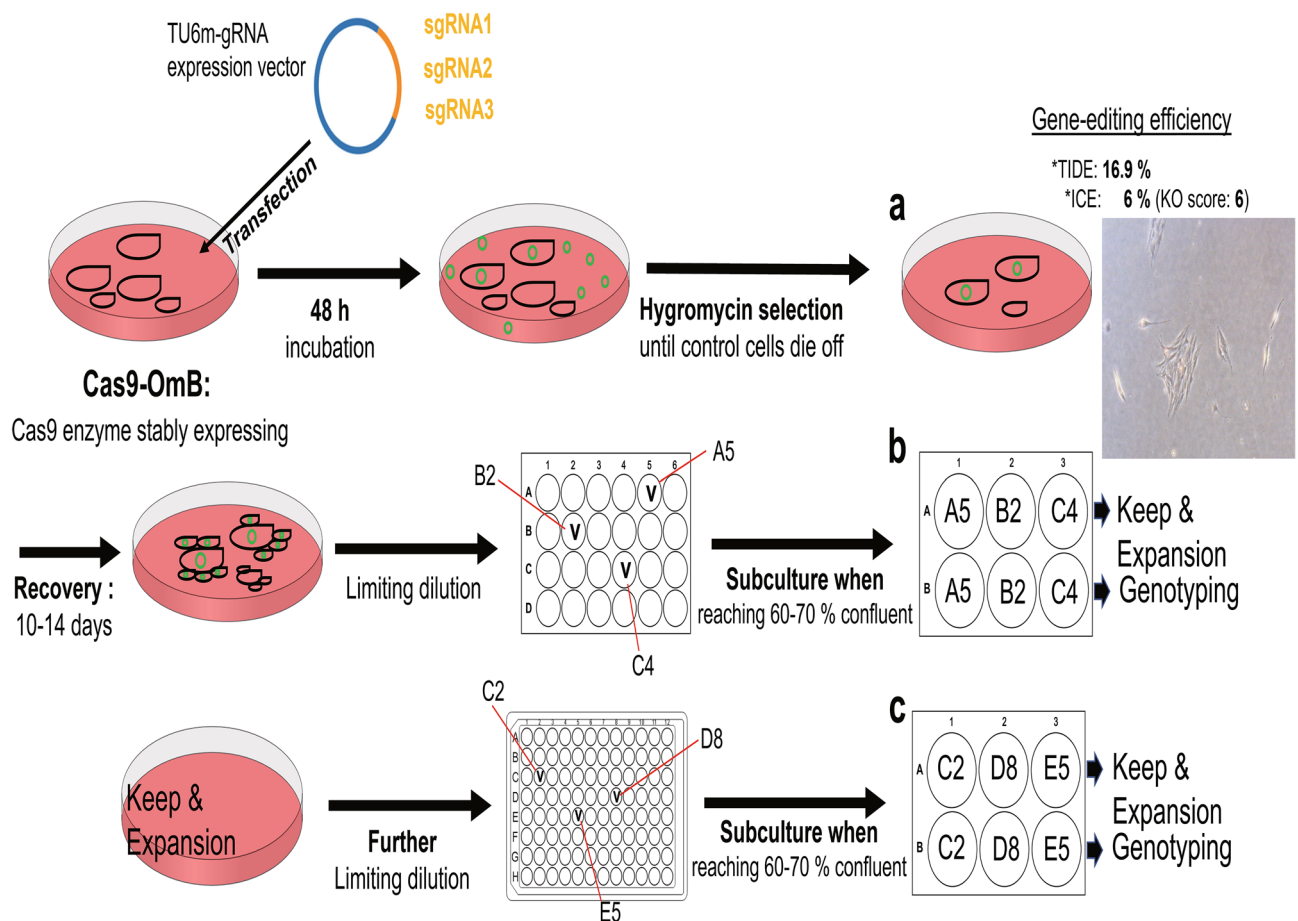
Figure 2. Schematic procedure of initial sgRNA-expression vector transfection (**a**) followed by antibiotic reagent selection and limiting dilution strategy to enrich and isolate poly-clonal cell pools (**b**) and monoclonal cell lines (**c**). Blue circle depicts sgRNA expression vector containing each sgRNA (orange part in the vector) 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 is 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.

50%, and a KO-score of 48/100 (Table S2). Thus, the majority of indel mutations that were isolated and enriched by the limiting dilution strategy resulted in effective frameshifts and potentially severe MYC truncation (Fig. 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 (Supplementary Fig. 3). This result suggested 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 those cells and passing them several times through a serological pipet (described in Fig. 2c).

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 (Fig. 4b), had altered growth patterns resulting in more densely clustered patches (Fig. 4e,f), and had lower proliferation rates than wildtype cells, i.e., they need more time to reach confluency (e.g., Fig. 4a,d).
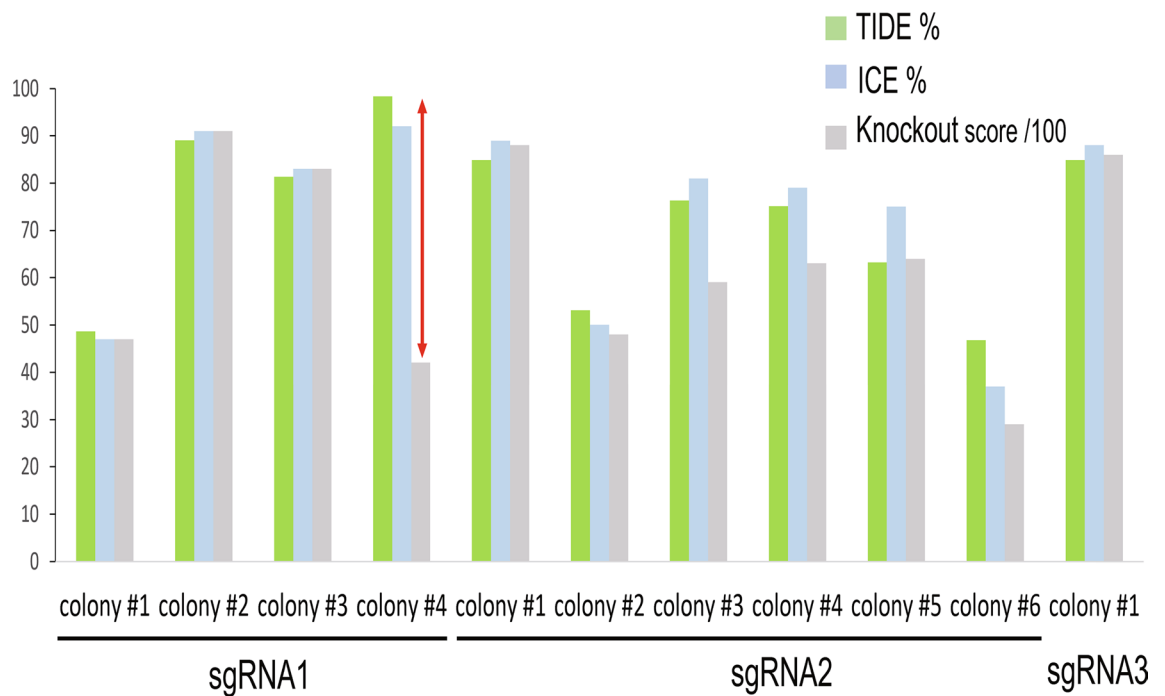
Figure 3. Quantitative indel mutation efficiencies of cell colonies (pools) produced by limiting dilution of cells transfected with sgRNA1, sgRNA2, and sgRNA3 plasmids. Individual cell colonies isolated by the limiting dilution method into individual wells of a 24-well plate at an average density of one cell per well 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. The resulting bar-graph visually indicates high consistency of all scores except for sgRNA1-colony#4, which has a lower KO-score.
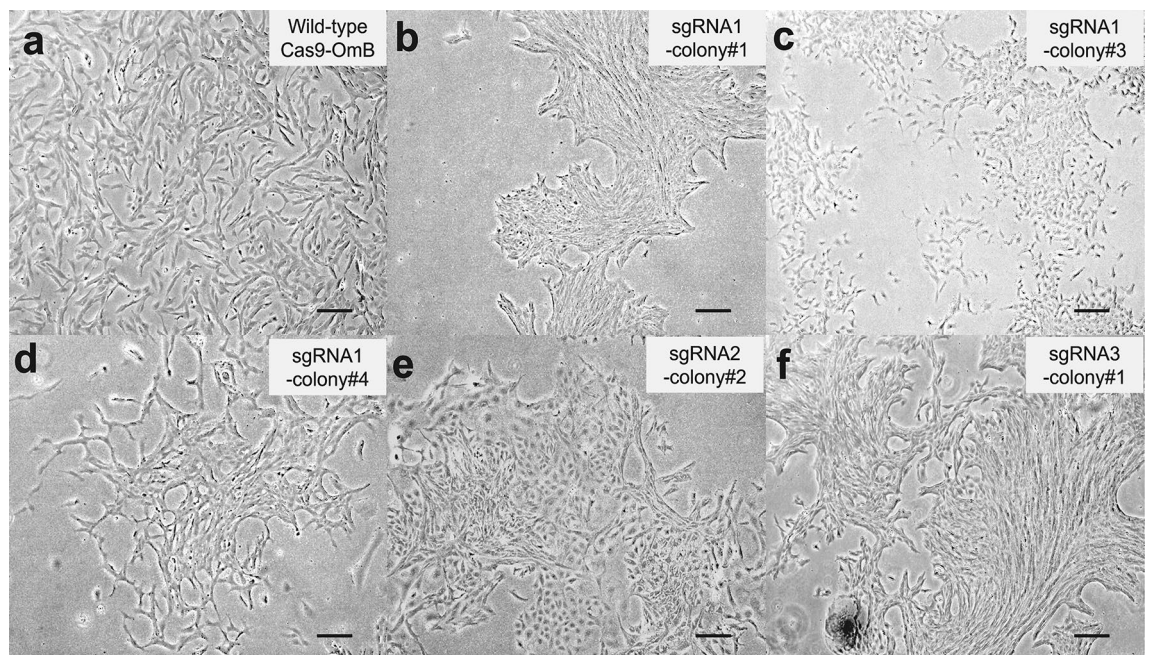

Figure 4. Representative images of cell morphology of wild-type Cas9-OmB cells and *myca* polyclonal ko cell pools (Scale bar, 200 μm). All micrographs were taken on an inverted microscope (Leica Dmi8) and imaged 5–7 days after transferring the cell colonies grown and tracked in a 24-well plate to a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

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 (Supplementary Fig. 3).

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 $R^2 = 0.99$ (Fig. 5a), and ICE indel efficiency was 100% with $R^2 = 1$ (Fig. 5d). Moreover, the KO-score was 100/100, which indicated that the 1-bp deletion mutation in the *myca* gene might result in functional disruption of MYC TF. ICE analysis confirmed a homogeneous monoclonal genotype consisting to 100% of the 1-bp deletion mutant (Fig. 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 (Fig. 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 was predicted to be truncated to only contain the first 120 amino acids due to premature translation termination via an early stop codon (TGA) (Fig. 6a). The majority of transcription regulatory domain and all of the DNA-binding domain (bHLH + LZ) were missing in the mutant MYC protein, indicating elimination of its function as a TF.

Protein structures of both wild-type and mutant MYC as generated using AlphaFold[45], RCSB Protein Data Bank[47], and Mol* 3D Viewer[46] were profoundly different between wild-type and mutant MYC (Fig. 6b,c). The removal of protein domains necessary for TF function from the truncated mutant protein was confirmed using FunFOLDQA[48], 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 to much shorter cell extensions (Fig. 7, white arrows) compared to in the morphology of wildtype cells showing elongated cell extensions even though the main cell body of mutant cells was comparable in size to wild-type cells (Fig. 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[49–51]. MYC has been studied extensively in the context of cancer biology, because of its overexpression in malignant tumors and activation of many hallmarks of cancer[52]. However, other functions of MYC that are not directly relevant for cancer biology have not received much attention. Since MYC TF is evolutionarily highly conserved and have many transcriptional targets[5], they are likely central regulators of the CSR and other diverse cellular functions that are important for normal non-pathological physiology[53].

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[8,13]. Furthermore, our results show that, unlike *myc2, myca* is elevated at the mRNA level during hyperosmotic stress in tilapia OmB cells (Supplementary Fig. 1a). 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 (Supplementary Fig. 1b). We had previously shown that hyperosmotic transcriptional induction of *IMPA1.1* is at least partly mediated by several osmolality/salinity-responsive element 1 (OSRE1) CREs[54]. 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)[55]. 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 depends 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
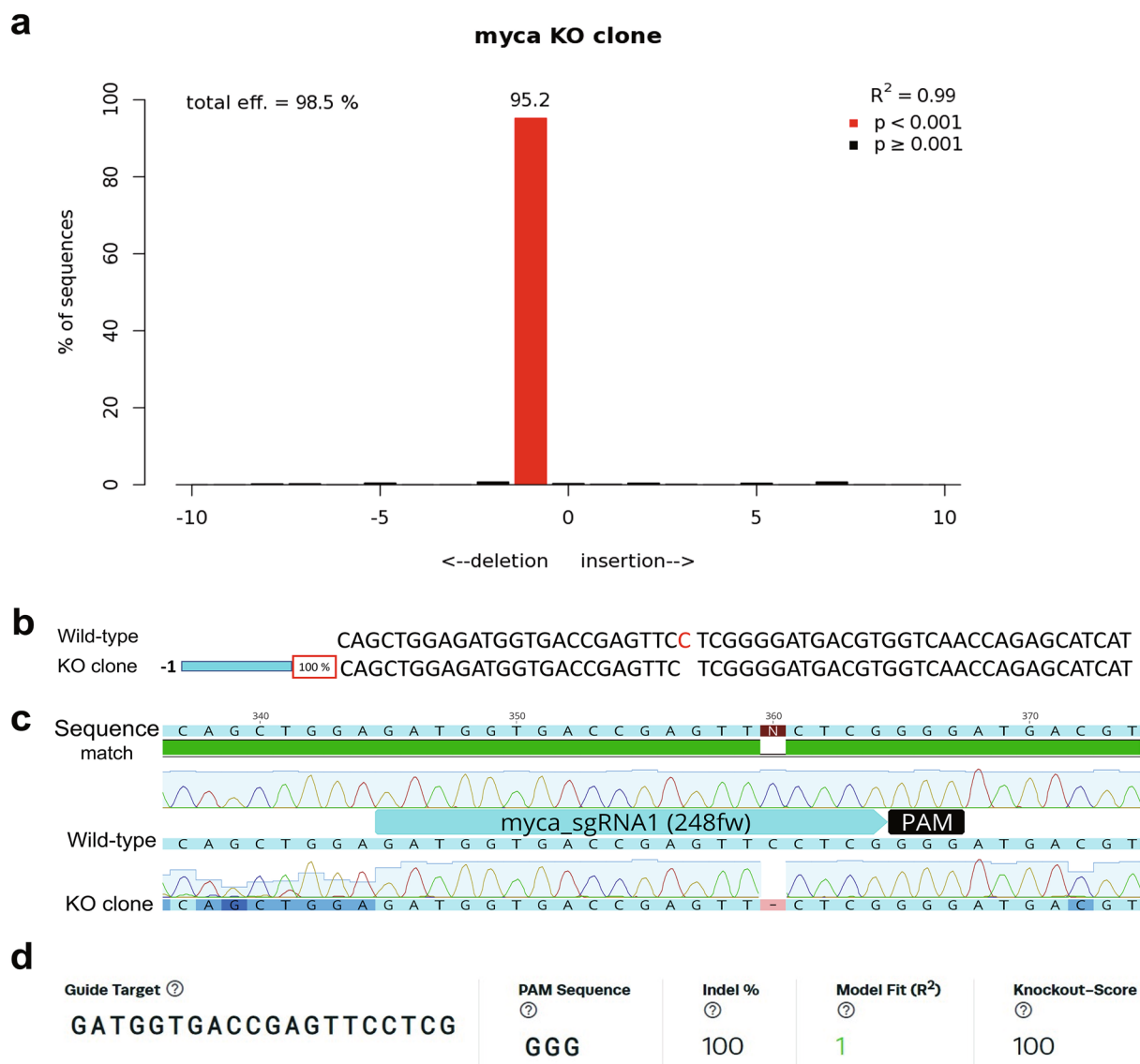
Figure 5. Confirmation of the homogeneity of mutant genotype in a *myca* monoclonal ko cell line. (**a**) TIDE analysis of genomic DNA extracted from a cell colony after the second series of limiting dilution. The X-axis indicates the nature of indels while the Y-axis depicts the percentages of the corresponding 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 (R^2), and Knockout-Score.

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[56,57], molecular and cellular phenotyping[58–60], and other approaches[61,62] to comprehensively characterize the role of MYC for teleost osmoregulation.

Limiting dilution is an effective strategy for cloning specific cell genotypes. Targeting genes in vitro using a CRIPSR/Cas9 system results in a heterogenous pool of edited cells with varying indel populations[34]. Such highly heterogeneous polyclonal cell pools can have poor overall gene editing efficiency due to inefficient delivery or low sgRNA-dependent mutational efficiency[63,64]. 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
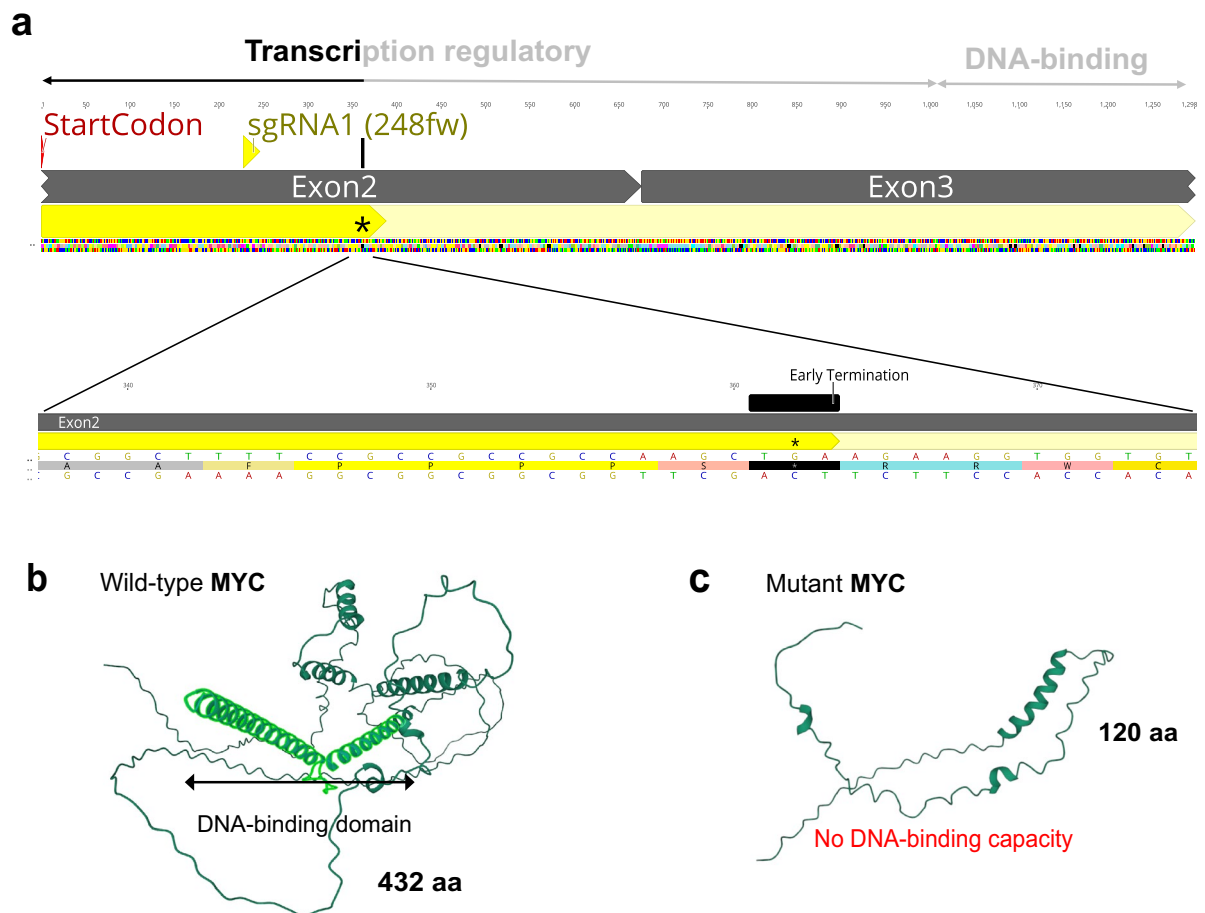
Figure 6. Prediction of loss of function due to truncation of a mutant MYC protein produced by the monoclonal Cas9-OmB-mycaKO1 cell line. (**a**) Wild-type and mutant MYC amino acid sequences were annotated on the coding sequence (CDS, shown yellow highlighted in the upper row). The region of the mutant 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 (https://alphafold.ebi.ac.uk/entry/Q45RH2), Mol*3D Viewer, and RCSB Protein Data Bank. 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**).

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[29,36,39].

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%). 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
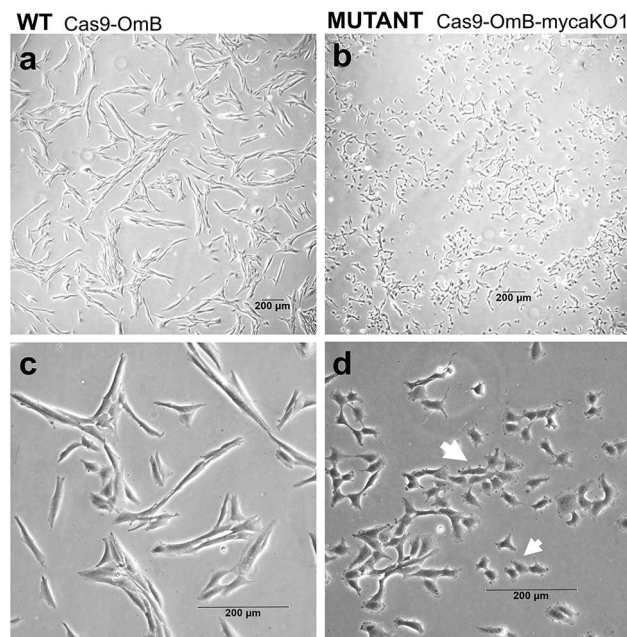
Figure 7. Cell morphology of tilapia cell lines. Micrographs showing differences in morphology of Cas9-OmB wild-type cells (**a,c**) versus Cas9-OmB-mycaKO1 mutant cells (**b,d**). The lack of elongated cell extensions in selected mutant cells is illustrated by white arrows. Images were taken using 5 × and 20 × phase objective on an inverted microscope (Leica Dmi8). The bar at the bottom of each image indicates a distance of 200 μm.

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[35]. 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[65–67]. 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[68]. For example, workflows for isolating gene-edited clonal cell lines by fluorescent-aided sorting system are currently only feasible for mammalian cells[69]. 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. 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 were used as the input data. TIDE[70] 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[63,71]. The ICE analysis pipeline (SYNTHEGO)[72] has been developed more recently and examined rigorously by comparing to 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[73].

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. Because of these advantages and their complementary features, the combined use of TIDE and ICE for indel heterogeneity analyses has been reported previously[74].

One important finding of our study was a substantial difference between KO-score and indel efficiency observed in the 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% (Fig. 3 and Table S2). This information represents a critical factor for choosing particular cell populations for functional studies. In this case, the sgRNA1-colony#4 cells

would be eliminated from consideration for functional phenotype analyses because almost half of the mutants generated produced wildtype MYC proteins that is functionally intact. The genetic heterogeneity was higher for other mutant populations including sgRNA2-colonies #2 and #3, which is why sgRNA1-colony#3 was chosen for serial dilution (see also supplementary Fig. 3b).

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 indel accounts for 51.4%) and because the frameshift mutation introduced a premature translation termination codon shortly after the target site, Fig. 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. Of note, other colonies (e.g., sgRNA2-colony#2 and #3) with intermedial difference between indel efficiencies (TIDE and ICE) and KO scores (with high indel frequency) were also ruled out for subsequent dilution mainly due to contributions of individual mutant genotypes were less than 51.4% of sgRNA1-colony#3 (Supplementary Fig. 3).

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[39,75], medaka[76], and Chinook salmon[38]. 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[77,78]. Tilapia are widely used aquaculture species, second only to carp regarding global production yields. However, like many other organisms, tilapia are subject to climate change and pollution, which negatively affects their performance and natural habitat and aquaculture. 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 analyses in multiple aquaculture fish species[79–81]. These ko cell lines allow for deep functional analyses that associate specific genotypes with complex phenotypes, including systems level molecular phenotypes revealed by transcriptomics[82] and proteomics[83].

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 displayed reduced proliferation even after three passages. This result indicates that *myca* ko is not lethal, and even does not hamper cell propagation in vitro. This phenomenon is interesting as we 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[84]. Nevertheless, the aberrant morphologies of tilapia *myca* ko cell lines (Figs. 4c,d) 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[42] and quantitative proteomics approaches can be employed to provide functional insights into biochemical and genetic networks that are controlled by MYC TF[85,86]. The limiting dilution approach and the *myca* ko tilapia cell lines generated in this study will empower such future functional analyses.

In summary, this study has 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.

Data availability

The datasets generated and/or analysed during the current study are available as DNA sequence datafor gRNA-sand PCR primers, which are provided in the main manuscript or supplementary data file. The tilapia *myca* gene (Gene ID: 100689989) was identified in the NCBI reference genome sequence for *Oreochromis niloticus* (NC_031974.2), along with the corresponding sequences for mRNA (XM_005448983.4) and protein (XP_005449040.1).

References

1. Dang, C. V. *et al.* Function of the c-Myc oncogenic transcription factor. *Exp Cell Res.* **253**, 63–77 (1999).
2. Schmidt, E. V. The role of c-myc in cellular growth control. *Oncogene* **18**, 2988–2996 (1999).
3. 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.* **4**, R69 (2003).
4. 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. USA* **92**, 10217–10221 (1995).
5. 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. USA* **84**, 170–173 (1987).
6. Moura, M. T. *et al.* Evolutionary-driven C-MYC gene expression in mammalian fibroblasts. *Sci. Rep.* **10**, 11056 (2020).
7. Kültz, D. Evolution of cellular stress response mechanisms. *J. Exp. Zool. A Ecol. Integr. Physiol.* https://doi.org/10.1002/jez.2347 (2020).
8. Ma, A. *et al.* Osmoregulation by the myo-inositol biosynthesis pathway in turbot Scophthalmus maximus and its regulation by anabolite and c-Myc. *Comput. Biochem. Physiol. Part A Mol. Integr. Physiol.* **242**, 1136 (2020).
9. 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.* **287**, 2560–2576 (2020).
10. Valenzuela, C. E. *et al.* Salt stress response triggers activation of the jasmonate signaling pathway leading to inhibition of cell elongation in Arabidopsis primary root. *J. Exp. Bot.* **67**, 4209–4220 (2016).
11. Yoon, Y. *et al.* The role of stress-responsive transcription factors in modulating abiotic stress tolerance in plants. *Agronomy* **10**, 788 (2020).
12. 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* **14**, e0223962 (2019).
13. O'Connell, B. C. *et al.* A large scale genetic analysis of c-Myc-regulated gene expression patterns * 210. *J. Biol. Chem.* **278**, 12563–12573 (2003).
14. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
15. Barrangou, R. & Doudna, J. A. Applications of CRISPR technologies in research and beyond. *Nat. Biotechnol.* **34**, 933–941 (2016).
16. Gaj, T., Gersbach, C. A. & Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* **31**, 397–405 (2013).
17. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
18. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338 (2012).
19. Hilton, I. B. & Gersbach, C. A. Enabling functional genomics with genome engineering. *Genome Res.* **25**, 1442–1455 (2015).
20. Li, H. *et al.* Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Sig. Transduct. Target Ther.* **5**, 1–23 (2020).
21. Burgio, G. Redefining mouse transgenesis with CRISPR/Cas9 genome editing technology. *Genome Biol.* **19**, 27 (2018).
22. Campenhout, C. V. *et al.* Guidelines for optimized gene knockout using CRISPR/Cas9. *Biotechniques* **66**, 295–302 (2019).
23. Hana, S. *et al.* Highly efficient neuronal gene knockout in vivo by CRISPR-Cas9 via neonatal intracerebroventricular injection of AAV in mice. *Gene Ther.* **28**, 646–658 (2021).
24. Mou, H., Kennedy, Z., Anderson, D. G., Yin, H. & Xue, W. Precision cancer mouse models through genome editing with CRISPR-Cas9. *Genome Med.* **7**, 53 (2015).
25. 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* **9**, e108622 (2014).
26. 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.* **8**, 16054 (2018).
27. Irion, U., Krauss, J. & Nüsslein-Volhard, C. Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. *Development* **141**, 4827–4830 (2014).
28. Hruscha, A. *et al.* Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development* **140**, 4982–4987 (2013).
29. 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. (NY)* **18**, 449–452 (2016).
30. Escobar-Aguirre, S. *et al.* Development of a bicistronic vector for the expression of a CRISPR/Cas9-mCherry system in fish cell lines. *Cells* **8**, E75 (2019).
31. Gratacap, R. L. *et al.* Efficient CRISPR/Cas9 genome editing in a salmonid fish cell line using a lentivirus delivery system. *BMC Biotechnol.* **20**, 35 (2020).
32. Hamar, J. & Kültz, D. An efficient vector-based CRISPR/Cas9 system in an Oreochromis mossambicus cell line using endogenous promoters. *Sci. Rep.* **11**, 7854 (2021).
33. Li, M. *et al.* Efficient and heritable gene targeting in tilapia by CRISPR/Cas9. *Genetics* **197**, 591–599 (2014).
34. 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.* **128**, e100 (2019).
35. 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* **4**, 16 (2021).
36. Yang, R. *et al.* Monoclonal cell line generation and CRISPR/Cas9 manipulation via single-cell electroporation. *Small* **14**, e1702495 (2018).
37. Fenerty, K. E. *et al.* 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* **6**, 133 (2018).
38. 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* **7**, bio035170 (2018).
39. Dehler, C. E. *et al.* Viral resistance and IFN signaling in STAT2 knockout fish cells. *J. Immunol.* **203**, 465–475 (2019).
40. Gratacap, R. L., Wargelius, A., Edvardsen, R. B. & Houston, R. D. Potential of genome editing to improve aquaculture breeding and production. *Trends Genet* **35**, 672–684 (2019).
41. Yang, Z., Yu, Y., Tay, Y. X. & Yue, G. H. Genome editing and its applications in genetic improvement in aquaculture. *Rev. Aquac.* **14**, 178–191 (2022).
42. 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* **9**, e95919 (2014).

43. Concordet, J.-P. & Haeussler, M. CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucl. Acids Res.* **46**, W242–W245 (2018).
44. Blum, M. *et al.* The InterPro protein families and domains database: 20 years on. *Nucl. Acids Res.* **49**, D344–D354 (2021).
45. Varadi, M. *et al.* AlphaFold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucl. Acids Res.* **50**, D439–D444 (2022).
46. Sehnal, D. *et al.* Mol* Viewer: Modern web app for 3D visualization and analysis of large biomolecular structures. *Nucl. Acids Res.* **49**, W431–W437 (2021).
47. Berman, H. M. *et al.* The protein data bank. *Nucl. Acids Res.* **28**, 235–242 (2000).
48. Roche, D. B., Buenavista, M. T. & McGuffin, L. J. FunFOLDQA: a quality assessment tool for protein-ligand binding site residue predictions. *PLoS ONE* **7**, e38219 (2012).
49. de la Cova, C. & Johnston, L. A. Myc in model organisms: A view from the flyroom. *Semin. Cancer Biol.* **16**, 303–312 (2006).
50. Dang, C. V. c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Mol. Cell Biol.* **19**, 1–11 (1999).
51. 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.* **14**, 121 (2021).
52. Dang, C. V. MYC on the path to cancer. *Cell* **149**, 22–35 (2012).
53. Chou, S.-J. & Tole, S. Lhx2, an evolutionarily conserved, multifunctional regulator of forebrain development. *Brain Res.* **1705**, 1–14 (2019).
54. Wang, X. & Kültz, D. Osmolality/salinity-responsive enhancers (OSREs) control induction of osmoprotective genes in euryhaline fish. *Proc. Natl. Acad. Sci. USA* **114**, E2729–E2738 (2017).
55. Burg, M. B., Kwon, E. D. & Kültz, D. Regulation of gene expression by hypertonicity. *Annu. Rev. Physiol.* **59**, 437–455 (1997).
56. Kim, C. & Kültz, D. An osmolality/salinity-responsive enhancer 1 (OSRE1) in intron 1 promotes salinity induction of tilapia glutamine synthetase. *Sci. Rep.* **10**, 12103 (2020).
57. 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)* **12**, 787 (2022).
58. Kültz, D. *et al.* Functional genomics and proteomics of the cellular osmotic stress response in 'non-model' organisms. *J. Exp. Biol.* **210**, 1593–1601 (2007).
59. Kültz, D., Li, J., Gardell, A. & Sacchi, R. Quantitative molecular phenotyping of gill remodeling in a cichlid fish responding to salinity stress. *Mol. Cell. Proteom.* **12**, 3962–3975 (2013).
60. 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. Approach.* **21**, 2486–2503 (2021).
61. Jayaram, N., Usvyat, D. & Martin, A. C. Evaluating tools for transcription factor binding site prediction. *BMC Bioinf.* **17**, 547 (2016).
62. Yadav, S. *et al.* Genome-wide identification, in silico characterization and expression analysis of the RNA helicase gene family in chickpea (C. arietinum L.). *Sci. Rep.* **12**, 9778 (2022).
63. Arndell, T. *et al.* gRNA validation for wheat genome editing with the CRISPR-Cas9 system. *BMC Biotechnol.* **19**, 71 (2019).
64. 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* **14**, 586–588 (2017).
65. Lopez, A., Fernandez-Alonso, M., Rocha, A., Estepa, A. & Coll, J. M. Transfection of epithelioma papulosum cyprini (EPC) carp cells. *Biotech. Lett.* **23**, 481–487 (2001).
66. Collet, B., Collins, C. & Lester, K. Engineered cell lines for fish health research. *Dev Comp Immunol* **80**, 34–40 (2018).
67. Gratacap, R. L., Jin, Y. H., Mantsopoulou, M. & Houston, R. D. Efficient genome editing in multiple salmonid cell lines using ribonucleoprotein complexes. *Mar Biotechnol* **22**, 717–724 (2020).
68. Weinguny, M. *et al.* Directed evolution approach to enhance efficiency and speed of outgrowth during single cell subcloning of Chinese Hamster Ovary cells. *Comput. Struct. Biotechnol. J.* **18**, 1320–1329 (2020).
69. Munoz, A. & Morachis, J. M. High efficiency sorting and outgrowth for single-cell cloning of mammalian cell lines. *Biotechnol. Lett.* https://doi.org/10.1007/s10529-022-03300-8 (2022).
70. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucl. Acids Res.* **42**, e168 (2014).
71. 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.* **8**, 888 (2018).
72. Conant, D. *et al.* Inference of CRISPR Edits from Sanger Trace Data. *CRISPR J* **5**, 123–130 (2022).
73. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
74. Jin, J. *et al.* An improved strategy for CRISPR/Cas9 gene knockout and subsequent wildtype and mutant gene rescue. *PLoS ONE* **15**, e0228910 (2020).
75. Ma, J. *et al.* Efficient resistance to grass carp reovirus infection in JAM-A knockout cells using CRISPR/Cas9. *Fish. Shellfish Immunol.* **76**, 206–215 (2018).
76. 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.* **83**, 443–448 (2018).
77. Aquaculture Genomics, Genetics and Breeding Workshop *et al.* Aquaculture genomics, genetics and breeding in the United States: current status, challenges, and priorities for future research. *BMC Genom.* **18**, 191 (2017).
78. Houston, R. D. *et al.* Harnessing genomics to fast-track genetic improvement in aquaculture. *Nat. Rev. Genet.* **21**, 389–409 (2020).
79. Omeka, W. K. M. *et al.* Genome-wide association study (GWAS) of growth traits in olive flounder (Paralichthys olivaceus). *Aquaculture* **555**, 738257 (2022).
80. 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* **550**, 737879 (2022).
81. Tsai, H.-Y. *et al.* Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genom.* **16**, 969 (2015).
82. Xie, C. *et al.* Dedicated transcriptomics combined with power analysis lead to functional understanding of genes with weak phenotypic changes in knockout lines. *PLoS Comput. Biol.* **16**, e1008354 (2020).
83. Dolgalev, G. & Poverennaya, E. Applications of CRISPR-cas technologies to proteomics. *Genes* **12**, 1790 (2021).
84. Das, S. K., Lewis, B. A. & Levens, D. MYC: A complex problem. *Trends Cell Biol.* (2022).
85. 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.* **23**, 4498 (2022).
86. Zhang, J. *et al.* Proteomic and transcriptomic profiling of Pten gene-knockout mouse model of prostate cancer. *Prostate* **80**, 588–605 (2020).

Acknowledgements

Author contributions

Conceptualization, C.K. and D.K.; methodology, C.K.; software, C.K.; validation, C.K.; formal analysis, C.K.; investigation, C.K.; resources, D.K.; data curation, C.K.; writing—original draft preparation, C.K.; writing—review and editing, C.K., A.C., and D.K.; visualization, C.K.; supervision, D.K.; project administration, D.K.; funding acquisition, A.C. and D.K. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-37928-x.

Correspondence and requests for materials should be addressed to D.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.