

Transcriptomics Unravels Molecular Players Shaping Dorsal Lip Hypertrophy in the Vacuum Cleaner Cichlid, *Gnathochromis Permaxillaris*

Laurène Alicia Lecaudey

University of Graz

Pooja Singh

University of Graz

Christian Sturmbauer (✉ christian.sturmbauer@uni-graz.at)

University of Graz

Anna Duenser

University of Graz

Wolfgang Gessl

University of Graz

Ehsan Pashay Ahi

University of Graz

Research Article

Keywords: lip hypertrophy, gene expression, Cichlidae, *Gnathochromis permaxillaris*, RNA-Seq, Lake Tanganyika

Posted Date: March 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-283299/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Genomics on July 5th, 2021. See the published version at <https://doi.org/10.1186/s12864-021-07775-z>.

Abstract

Background: Teleosts display a spectacular diversity of craniofacial adaptations that often mediate ecological specializations. A considerable amount of research has revealed molecular players underlying skeletal craniofacial morphologies, but less is known about soft craniofacial phenotypes. Here we focus on a bizarre example of lip hypertrophy in the Lake Tanganyika cichlid, *Gnathochromis permaxillaris*, that is considered to be a dietary adaptation to suck invertebrates out of narrow crevices. In this study, we investigate the molecular and regulatory basis of lip development in *G. permaxillaris*, using a comparative transcriptomic approach.

Results: We identified a gene regulatory network involved in tissue overgrowth and cellular hypertrophy, potentially associated with the formation of a locally restricted hypertrophic lip in a teleost fish species. Of particular interest were the increased expression level of *apoda* and *fhl2*, as well as reduced expression of *cyp1a*, *gimap8*, *lama5* and *rasal3*, in the hypertrophic lip region which have been implicated in formation of lip structure in other vertebrates. Among the predicted upstream transcription factors, we found reduced expression of *foxp1* in the hypertrophic lip region, which is known to act as repressor of cell growth and proliferation and its function has been associated with hypertrophy of upper lip in human.

Conclusion: Our results provide a genetic foundation for future studies of molecular players shaping soft and exaggerated, but locally restricted, craniofacial morphological changes in fish and perhaps across vertebrates.

Introduction

Teleost fishes show striking adaptive diversity in their craniofacial anatomy, reflecting an equally striking variety of ecological and trophic specializations. It is therefore not surprising that this phenomenon has attracted the attention of developmental, molecular and evolutionary biologists, beyond model species such as zebrafish and medaka [1, 2]. In the last two decades, an impressive set of molecular players participating in the development and morphogenesis of craniofacial skeletal structures, including their interconnecting signalling pathways, have been described in teleost fish [2–4]. Much less is known about the morphogenic molecular players and underlying signals of craniofacial soft tissues, that also exhibit varied adaptive peculiarities. In cichlid fishes, the molecular mechanisms underlying exaggerated soft-tissue morphologies in lip, nose and nuchal hump phenotypes have been addressed only recently [5–10].

One such exaggerated craniofacial soft-tissue phenotype is the overgrowth of lip tissues, the so-called thick-lipped phenotype, which has been observed in cichlid species in lakes spanning different continents [6, 7, 11]. The repeated evolution of the hypertrophic lip phenotype in both African and Central American cichlids has been associated with the dietary adaptation to suck elusive invertebrates out of narrow crevices in rocky habitats [7, 12, 13]. Genetic studies so far, suggest numerous loci across the genome contribute small additive effects that can be linked to a variety of thick-lipped phenotypes [6, 7, 11]. These findings indicate that the thick-lipped phenotype is a complex trait [11]. The exaggerated thickening of the

lip might not follow a uniform pattern across the entire upper and lower lip and the extent of lip hypertrophy is significantly reduced in captivity under flake food diet, when the fish are not using this foraging strategy, so that lip hypertrophy seems highly dependent on foraging performance, making it a phenotypically plastic trait [11]. Both genetic and phenotypic plasticity take part in the adaptive morphogenesis of this phenotype [5, 14].

Another bizarre example of lip hypertrophy is *Gnathochromis permaxillaris*, a cichlid species with a shovel-like snout with a peculiarly thickened part of the upper lip [15]. *G. permaxillaris* is a benthivorous deepwater cichlid of the tribe Limnochromini from Lake Tanganyika [15] that develops a unique hypertrophy in the most anterior part of its upper lip (Fig. 1). Interestingly, when fish are raised in captivity, individuals develop this phenotype to a lesser extent (Peter Henninger, personal communication), suggesting some degree of phenotypic plasticity. Nothing is known about the molecular basis of lip development in *G. permaxillaris*. In this study, we aimed to find genes with that play a role in the formation of the dorsal part of the upper lip using RNA-sequencing. To this end we profiled gene expression differences between the hypertrophic part of the upper lip, versus the posterior part of the upper lip and the most anterior part of the lower lip, which both do not exhibit hypertrophic overgrowth in wild-caught young male adults of *G. permaxillaris* (Fig. 1). As a control, we used a closely related species in the tribe Limnochromini, *Greenwoodochromis bellcrossi* [15–17], which displays hypertrophy in the entire parts of both the upper and lower lips (Fig. 1). Using the regulatory sequences of differentially expressed genes in the hypertrophic lip region, we predicted their potential upstream transcriptional regulators, and applied qPCR expression analysis on the most interesting candidate genes to validate our results from the RNA-seq analysis. We identified a gene regulatory network, potentially associated with the formation of a locally restricted hypertrophic lip in the Limnochromini. Our results provide a foundation for future investigations of molecular players shaping exaggerated but locally restricted morphological changes in facial elements of fish and vertebrates.

Results

RNA-seq, differential gene expression and downstream analyses

The total RNA sequencing resulted between 3.84 and 8.13 million reads per sample and after removal of low quality reads, each sample had between 3.82 and 8.10 million reads (Table S1). The raw sequence reads have been submitted to the Sequencing Read Archive (SRA) of NCBI (accession number: PRJNA694145). The comparisons of the lip regions of *G. permaxillaris* resulted in 106 differentially expressed (DE) genes between the hypertrophic part of the upper lip (UL-1) versus both the non-hypertrophic part of the upper lip (UL-2) and the lower lip (LL) (Fig. 2A). Among these, 56 genes showed increased expression, whereas 47 genes had reduced expression in UL-1 (Fig. 2B and C). The heatmap clustering of the DE genes revealed at least two major branches in each group of up- or down-regulated gene sets indicating potential differences in transcriptional regulation within each group (Fig. 2B and C).

Using the list of 106 DE genes, we conducted gene ontology enrichment analysis and amongst the significantly enriched GO terms, biological processes related to cell movement, adhesion and growth had the highest enrichment ratios (Fig. 3A). The enrichment for genes related to locomotion raises the possibility of functional specificity for the hypertrophy of the most anterior part of the upper lip of *G. permaxillaris*. We also found enrichments of molecular processes related to GTPase and Ras protein signalling for the list of DE genes. Subjecting the same gene list to an interactome analysis revealed only a single large interconnected network of genes. However, the network contained several interesting genes, and particularly transcription factors (TFs), with potential role in tissue overgrowth and cellular hypertrophy (Fig. 3B). We also performed overrepresentation analysis of TF binding motifs on the regulatory sequences of the DE genes using MEME algorithm [18]. This resulted in five enriched motifs present in the regulatory sequences of at least 30 out of 106 DE genes (Table 2). In search for similarities of the identified motifs with known TF binding sites in vertebrates, we found at least two interesting TF candidates, *foxp1* and *heb/tcf12*, with overrepresented binding sites. Furthermore, both of these TFs showed regulatory connections with some of the DE genes in the identified interactome network, with *foxp1* having more connections than *tcf12* (Fig. 3B).

Expression validation using qPCR

The gene expression analysis by qPCR requires the identification of stably expressed reference genes [19], and our previous studies on East African cichlids confirmed that the validation of suitable reference gene(s) is an essential step that needs to be taken in every species, every tissue and experimental condition [20–24]. In order to choose adequate reference gene candidates, we conducted a ranking for the genes with no expression difference (FDR = 1 in the the RNA-seq comparisons) based on their coefficient variation (CV) throughout all the samples. The ten genes with lowest CV were selected as candidate reference genes for validation of their expression stability by qPCR. None of the validated reference genes in previous gene expression studies of East African cichlids have appeared among the ten candidates, indicating the necessity of reference gene validation for each experimental setup [20–24]. Overall, the reference gene rankings by the three algorithms, BestKeeper, geNorm and NormFinder software showed that two of the candidates, *pip4k2b* and *ss18*, were always among the top three most stable reference genes (Table 1). Therefore, we used geometric mean of Cq values for *pip4k2b* and *ss18* in each sample, as a normalization factor, in order to calculate the relative gene expression levels of our target genes.

Table 1

Ranking and statistical analyses of reference genes in the lip samples using three different algorithms.

| BestKeeper | | geNorm | | NormFinder | | | |
|----------------|-------|----------------|-------|----------------|-------|----------------|-------|
| Ranking | SD | Ranking | r | Ranking | M | Ranking | SV |
| <i>luc7l</i> | 0.260 | <i>ubxn4</i> | 0.908 | <i>pip4k2b</i> | 0.328 | <i>ss18</i> | 0.193 |
| <i>ss18</i> | 0.289 | <i>ss18</i> | 0.905 | <i>trir</i> | 0.333 | <i>pip4k2b</i> | 0.242 |
| <i>pip4k2b</i> | 0.293 | <i>pip4k2b</i> | 0.902 | <i>ss18</i> | 0.335 | <i>trir</i> | 0.248 |
| <i>trir</i> | 0.296 | <i>setd3</i> | 0.878 | <i>prkci</i> | 0.341 | <i>setd3</i> | 0.258 |
| <i>h3f3</i> | 0.328 | <i>trir</i> | 0.878 | <i>h3f3</i> | 0.352 | <i>h3f3</i> | 0.274 |
| <i>prkci</i> | 0.366 | <i>prkci</i> | 0.873 | <i>setd3</i> | 0.356 | <i>prkci</i> | 0.303 |
| <i>snrnp70</i> | 0.391 | <i>h3f3</i> | 0.828 | <i>ubxn4</i> | 0.364 | <i>ubxn4</i> | 0.329 |
| <i>vasp</i> | 0.402 | <i>vasp</i> | 0.806 | <i>vasp</i> | 0.421 | <i>luc7l</i> | 0.333 |
| <i>ubxn4</i> | 0.402 | <i>luc7l</i> | 0.611 | <i>luc7l</i> | 0.425 | <i>vasp</i> | 0.411 |
| <i>setd3</i> | 0.414 | <i>snrnp70</i> | 0.389 | <i>snrnp70</i> | 0.590 | <i>snrnp70</i> | 0.445 |

Abbreviations: SD = Standard deviation, r = Pearson product-moment correlation coefficient, SV = stability value, M = M value of stability.

Table 2

Predicted binding sites for potential upstream regulators of the differentially expressed genes identified by RNA-seq. PWD ID indicates positional weight matrix ID of a predicted binding site and E-values refer to matching similarity between the predicted motif sequences and the PWD IDs. The count implies on number of genes containing the predicted motif sequence on their regulatory region.

| TF binding site | PWM ID | Count | Predicted motif sequence | E-value |
|-----------------|--------|----------|-------------------------------|-----------|
| HEB | M00698 | 97 / 106 | CCTGCTG | 1.119e-09 |
| FOXP1 | M00987 | 81 / 106 | AAATAAANAACAAAAAAAWA | 4.441e-16 |
| SMAD3 | M00701 | 72 / 106 | ACASASASACASACA | 2.28E-07 |
| HEB | M00698 | 41 / 106 | KCCMRGCTGVCTGS | 3.51E-07 |
| FOXP1 | M00987 | 31 / 106 | TWTWYDTATWWRTWTATTTATWTATWWAT | 1.05E-09 |
| FOX | M00809 | | | 3.061E-08 |

Within the identified DE genes in the RNA-Seq result, we selected nine genes with a known role in facial morphogenesis and other related functions in vertebrates (Table 3), as well as the two predicted upstream TFs (*foxp1* and *tcf12*) to be tested by qPCR (Fig. 4). Based on the RNA-seq results, four of the selected candidate genes have shown reduced expression in the hypertrophic UL-1 (*cyp1a*, *gimap8*, *lama5* and

rasal3) whereas the other six genes (*apoda*, *foxf1*, *fhl2*, *igf2b* and *trim37*) had increased expression level in UL-1 of *G. permaxillaris*. The qPCR results showed that all of the genes followed similar expression patterns as found by RNA-seq, except for *trim37* which showed no significant difference between the lip regions in *G. permaxillaris*. However, when the qPCR expression results for *G. permaxillaris* were compared to the two hypertrophic lip regions of *G. bellcrossi*, we found only *apoda* and *fhl2* showing increased expression in all the hypertrophic lip regions across both species. The four genes with reduced expression in UL-1 for *G. permaxillaris*, also showed reduced expression in the hypertrophic lip regions of *G. bellcrossi*. Among the two predicted TFs, only *foxp1* showed consistent difference between hypertrophic and non-hypertrophic lip regions across both species, with reduced expression in the hypertrophic lip regions. This indicates potential transcriptional repressor effects of *foxp1* on the downstream genes in the hypertrophic lip region. Altogether, the qPCR results show consistency between RNA-seq and qPCR results confirming the validity of our transcriptome data analysis in this study.

Table 3

Differentially expressed genes in the hypertrophic region of dorsal lip in *Gnathochromis permaxillaris* with related functions in vertebrates.

| Gene | Related function | Organism | References |
|---------------|--|---------------------------|---|
| <i>apoda</i> | A multi-ligand transporter involved in neural cell survival found to be repressed in lip tissue of thick-lipped Midas cichlid | Human Mouse Cichlid | (Dassati, Waldner & Schweigreiter, 2014) (Manousaki et al., 2013) |
| <i>cyp1a</i> | Involved in xenobiotic and steroid metabolism and associated with lip and palate cleft and cancers | Human | (Linnenkamp et al., 2020) (Stuppia et al., 2011) |
| <i>fh12</i> | A transcriptional modulator of cell proliferation and apoptosis, found to be repressed in lip tissue of thick-lipped Midas cichlid | Human Cichlid | (Ng et al., 2011) (Labalette et al., 2008) (Manousaki et al., 2013) |
| <i>foxf1</i> | Micro-deletion causes lip and palate deformities | Human Mouse | (Shaw-Smith, 2010)(Xu et al., 2016) |
| <i>foxp1</i> | Non-functional mutation causes pronounced vermilion border of upper lip | Human | (Meerschaut et al., 2017) |
| <i>gimap8</i> | A GTP-binding protein with role in immunity and apoptosis, found to be repressed in lip tissue of thick-lipped Midas cichlid | Cichlid | (Manousaki et al., 2013) |
| <i>igf2</i> | Expression changes in Russell–Silver syndrome leads to thin vermilion border of upper lip | Human | (Peñaherrera et al., 2010) |
| <i>lama5</i> | An extracellular matrix glycoprotein mediating the attachment, migration and organization of cells into tissues, and implicated in lip inflammation and carcinogenesis | Human | (Peixoto da-Silva et al., 2012) |
| <i>rasal3</i> | A negative regulator of Ras pathway and its duplication and deletion are both linked to defective lip morphogenesis | Human | (Draaken et al., 2013) (Kosaki et al., 2011) |
| <i>Trim37</i> | A tripartite motif family member involved in developmental patterning and its mutation causes lip and palate cleft | Human | (Kumpf et al., 2013) |
| <i>tcf12</i> | Non-functional mutation causes thin upper lip and craniofacial deformities | Human | (Piard et al., 2015) |

Discussion

Delineating the molecular basis of adaptive morphologies is essential to understand their evolution and as well as their potential contribution to speciation. In this study, we investigated the soft-tissue craniofacial trait of the hypertrophic lip in *G. permaxillaris*, in a comparative transcriptomic framework with *G. bellcrossi*, which are both from the Limnochromini tribe belonging to the cichlid adaptive radiation of Lake Tanganyika. *Gnathochromis permaxillaris* is a deepwater cichlid living over mud bottom as far as oxygenated water reaches down and has a unique mode of foraging. The species slowly swims directly over the mud surface with its protruding snout hovering at the surface and at the same time opening the protruding mouth underneath, while ingesting mud that is then filtered through the gill arches. Its common name Vacuum cleaner cichlid is an adequate description of this mode of feeding. Thereby, the protruding upper lip with its small hypertrophy at its tip seems to be adaptive by boosting the efficiency of filtration (unpublished behavioural observations).

Using differential gene expression analysis in the hypertrophic lip region as input for gene ontology analysis, we found multiple biological processes to be enriched. These processes involve cell motility, adhesion and developmental growth, as well as regulation of GTPase mediated signals, particularly, the RAS signalling pathway. A recent integrated genomic and transcriptomic study in humans has revealed that cell adhesion, cell junction and extracellular structure organizations are among the major biological processes involving lip and cleft development and morphogenesis [25]. Furthermore, as studies on human and mouse have shown, biological processes involving developmental growth and cell proliferation are known to be the key mechanisms in upper lip development and morphogenesis [26]. On the other hand, RAS/MAPK signalling is among the well-known molecular pathways involved in development and morphogenesis of craniofacial structures in vertebrates [4, 27]. Activation of Ras signalling promotes cell proliferation, growth and survival in various tissues [28, 29], and because of these roles, several components of the Ras pathway are considered as therapeutic targets in different types of cancer [30]. In mammals, Ras signaling plays a pivotal role in skin development, dermal thickening and skin carcinogenesis [31]. Defective activity of the Ras pathway can cause a wide range of skin anomalies, such as thickened palms and soles, redundant skin, papilloma formation, excessive proliferation of keratinocytes and increased skin folds [31]. Therefore, the hypertrophic lip region in the anterior upper lip of *G. permaxillaris* can be a result of increased activity of Ras signaling in this region. However, further functional studies (such as a protein manipulation method recently used in cichlids to investigate regional activity of a growth signal [9]) are required to find out the molecular reason for the anatomically limited activation of this signal only in the anterior part of the upper lip in *G. permaxillaris*.

We found that many of the genes with increased expression in the hypertrophic region of dorsal lip were already demonstrated to play a role in lip morphogenesis and pathobiology in other vertebrates, including *alx1* [32], *alx3* [33], *angptl2* [34], *arid3a* [35], *crip2* [36], *ddr2* [37], *dpysl2* [38], *itga5* [39], *lmna* [40], *mgp* [41], *rgs5* [42], *rhob* [42, 43], *rxfp2* [44], *sostdc1* [42, 45], *thbs3* [46], *vcan* [47, 48], and *vtn* [49]. Among the transcriptionally repressed genes, we also found candidates with functions which were previously implicated in defective lip morphogenesis in other vertebrates such as *cd96* [50], *ep300* [51][52], *fgfr11* [53], *frzb* [54], *hectd1* [55, 56], *mmp13* [49], *slc16a6* [57], and *syne1* [58]. Taken together, these results suggest that a similar set of genes might be involved in lip morphogenesis, not just across teleosts, but

across vertebrates, and further functional studies are required to investigate their specific role in morphological divergence of soft tissues in fish.

Among the genes with reduced expression in the hypertrophic lip regions, we validated four genes with qPCR, *cyp1a*, *gimap8*, *lama5* and *rasal3*, and found all of them showing a similar expression reduction pattern in all of the hypertrophic lip regions of both species. Cytochrome P450 (CYP) 1 alpha, *cyp1a*, encodes an enzyme with an important role in the cytochrome p450 xenobiotic metabolism and the synthesis of steroids and other lipids. *Cyp1a* is a downstream target for AHR, RAS and Wnt/ β -Catenin signaling pathways [59, 60], and all of these signals are demonstrated to play important roles in morphogenesis and adaptive radiations of craniofacial elements in teleost fish [4]. Differential regulation of *cyp1a* is implicated in craniofacial morphological divergence in fish [61]. Differential regulation of AHR, RAS and Wnt/ β -Catenin signals, as well as *cyp1a*, is also involved in the defective formation of the lip and palate in vertebrates [62–66]. We found reduced expression of *cyp1a* in the hypertrophic lip regions in both cichlid species, which can be explained by its inhibitory role on cell proliferation [67, 68].

The second gene, GTPase immunity-associated protein family member 8 or *gimap8*, encodes a nucleotide-binding protein that plays a role in the maintenance and survival of lymphocytes in mammals [69]. Consistent with our result, the same gene was found to be repressed in lip tissue of thick-lipped Midas cichlid [6]; however, its exact function during development and morphogenesis of soft tissues in vertebrates remained unclear. In humans, increased expression of *gimap8* orthologue has been reported during adipocyte differentiation, indicating its potential role in cell differentiation [70]. The third gene, *lama5*, encodes one of the vertebrate laminin alpha chain proteins, a family of extracellular matrix glycoproteins, which are the major noncollagenous constituent of basement membranes. In humans, *lama5* has been implicated in pathogenesis of lip inflammation and carcinogenesis [71]. In mouse, loss of *lama5* causes hyper-proliferation of basal keratinocytes, an increase in the number of immune cells and thickening of epidermis; thus, reduced expression of *lama5* in the hypertrophic lip regions might result in an increased number of keratinocytes and subsequently thicker epidermis in these regions [72]. The last gene, *rasal3*, RAS protein activator like-3, encodes a negative regulator of Ras signalling pathway and its duplication and deletion are both linked to defective lip morphogenesis in humans [73, 74]. The reduced expression of *rasal3* in the hypertrophic lip regions suggests higher activity of RAS signaling in this region, which is consistent with the enriched RAS related gene ontology for the differentially expressed genes.

Two of the genes with increased expression in hypertrophic lip regions, *apoda* and *fh12*, were particularly interesting, since both genes have been already reported as potential molecular players in the formation of thick-lipped phenotype of central American cichlids [6]. However, in contrast to our results, the expression of both genes have been shown to be repressed in hypertrophic lips of Midas cichlid. Four-and-a-half LIM domains 2, *fh12*, encodes a transcriptional modulator of cell proliferation and [75, 76], while apolipoprotein Da, *apoda*, encodes a multi-ligand transporter involved in neural cell survival [77]. The molecular reason for this discrepancy is unclear, but it is likely that these genes have dual and opposite modulatory functions under different cellular conditions. Future functional investigations are required to

unravel this discrepancy. It should be noted that *fh2* is also reported as an important molecular player in formation of egg-spot in cichlids indicating its functional diversity in the adaptive morphological divergence of cichlid fishes [78].

We predicted binding sites for Forkhead Box (FOX) transcription factors and the basic helix-loop-helix (bHLH) for transcription factor 12 (*tcf12/heb*) to be enriched on upstream regulatory sequences of many of the differentially expressed genes by RNA-seq. Among the differentially expressed genes, we only found *foxf1* as potential candidate that could bind to the enriched FOX binding site. Interestingly, micro-deletion in mammalian orthologue of *foxf1* have been shown to be associated with lip and palate deformities in mouse and human [79, 80]. However, the qPCR analysis revealed that, although *foxf1* expression was increased in the hypertrophic lip region of *G. permaxillaris*, its expression was not increased in the hypertrophic lip regions of *G. bellcrossi*. This could indicate that another member of the FOX family might be involved, whose expression difference was not detected by RNA-seq. In addition to the core FOX binding site, we also found a binding site for *foxp1* (another member of FOX family) to be enriched multiple times on the regulatory sequences. The qPCR analysis revealed that *foxp1* expression has a significant expression reduction in all the hypertrophic lip regions of both species. The reduced expression of *foxp1* suggests a potential repressive effect on transcription of the genes induced in the hypertrophic regions. Interestingly, *foxp1* has been already demonstrated to have repressive effects on transcription of many of its downstream targets [81]. Moreover, a mutation affecting *foxp1* function has been associated with hypertrophy of vermilion borders (side edges) of upper lip in humans [82]. The role of *foxp1* in lip hypertrophy might be linked to its function in inhibition of cell proliferation by repressing the transcription of growth/cell cycle stimulating factors [83–85]. The second predicted TF, *tcf12*, has been recently shown to be involved in hypertrophy of frontal head soft tissues (nuchal hump) in another East African cichlid species [10], however, we did not find its consistent expression difference between the lip regions by neither the RNA-seq nor the qPCR method in this study.

Conclusions

Understanding the molecular basis of key innovations is crucial to understand their role in adaptive evolution and speciation. Using the hypertrophic lip in the Lake Tanganyika endemic *G. permaxillaris*, we lay the foundation for studying locally restricted soft tissue developmental evolution in vertebrates. In this study, we found an interconnected gene regulatory network underlying the formation of locally restricted hypertrophy of dorsal lip which is an uncommon phenotype observed in cichlid fish. In the future, we advocate the integration of gene regulatory network analyses from various craniofacial tissues to understand how they collectively govern trophic and behavioural adaptations during cichlid adaptive radiation.

Methods

Fish rearing and tissue sampling

Five captive bred males of *G. permaxillaris* and five males of *G. bellcrossi* were raised in a large tank (approximately 2000 litres), together with same numbers of females per species, in an environment enriched with various stony shelters to minimize competition stress. Both species had the same age (young adult) and displayed similar swimming behaviour, but different feeding behaviours, with little or no intra- and inter-species aggression. We fed both species with the same diet, which is adjusted for Tanganyika cichlids (Tropical Tanganyika multi-ingredient flakes suitable for omnivorous cichlids). We sampled the species at the same time, when the protrusion of the anterior part of upper lip in *G. permaxillaris* males had appeared, while both upper and lower lips of *G. bellcrossi* were thickened as well (Fig. 1). At this young adult stage both species were displaying sexual behaviour, such as chasing females and territorial defending. Before the dissection step, the fish were placed in a solution with 0.2 g MS-222 per 1L water, and after being sacrificed, the lip regions specified in Fig. 1, which include epidermis, dermis and the underlying soft connective tissues in those regions, were all dissected. The entire tissues for each lip region per fish were considered as one biological replicate and were placed into separate tubes containing RNAlater (Qiagen) and stored at -20 °C. The fish sacrifice was performed based on the guidelines issued by the Federal Ministry of Science, Research and Economy of Austria and according to regulations of BMFWF. The study was carried out in compliance with the ARRIVE guidelines.

RNA extraction and cDNA synthesis

Total RNA was extracted from 15 lip tissue samples (five replicates per lip region) from *G. permaxillaris* and ten lip tissue samples from *G. bellcrossi* using the ReliaPrep™ RNA Tissue Miniprep System Kit (Promega). Each sample consisted of epidermis, dermis and the underlying soft connective tissues in the specified lip regions (Fig. 1). Tissue samples were placed into tubes containing 250 µl of lysis buffer mixed with 1-Thioglycerol and 1.4 mm ceramic beads. The tissues were homogenized thoroughly by FastPrep-24 Instrument (MP Biomedicals, CA, USA) and RNA extraction was performed based on manufacturer's ReliaPrep™ protocol for fibrous tissues. The extraction protocol followed included a column-based genomic DNA removal step and several purification steps. The total RNAs were diluted in 50 µl nuclease-free water and were quantified by a Nanophotometer (IMPLEN GmbH, Munich, Germany). The qualities of total RNAs were measured by R6K ScreenTape System using an Agilent 2200 TapeStation (Agilent Technologies), and all samples had a RNA integrity number (RIN) minimum of eight. Around 200 ng of the extracted total RNA from each sample was used for cDNA synthesis based on the manufacturer's protocol of High Capacity cDNA Reverse Transcription kit (Applied Biosystems) and 1:4 times cDNA dilution was used as input to perform qPCR.

RNA-Seq library, transcriptome assembly and gene expression

To obtain a list of gene transcripts from the lip tissues, we performed RNA-Seq library preparation, based on the protocol of Standard TruSeq Stranded mRNA Sample Prep Kit (Illumina) with 500ng of total RNA per tissue input. We assessed the library qualities by running them on a D1000 ScreenTape (Agilent) using TapeStation 2200 (Agilent). We diluted the libraries to an optimal quantity recommended for sequencing, and then pooled the libraries with equal molarity. To generate 125bp paired-end reads, the RNA-sequencing was performed in the NGS Facility at Vienna Biocenter Core Facilities (VBCF, Austria) on an Illumina HiSeq2500. Next, the raw reads were de-multiplexed based on the unique barcodes incorporated in each sample during the library preparation step. The quality control step was conducted on the raw reads from each sample using the FASTQC analysis tool [86]. For each sample, the low quality reads were discarded, following the standard quality trimming step of the Trimmomatic software [87]. To do this, the filtering criteria was scaled to only maintain the reads with phred +33 quality score of at least 34 for all bases for a minimum length of 50 bp. The *de novo* transcriptome assembly of the lip tissues was implemented using the quality trimmed paired-end reads of all samples through the Trinity package [88, 89].

The transcript abundances were calculated using the assembled transcriptome and Kallisto tool within the Trinity package, in order to attain the transcript expression levels in each sample [90]. The gene expression levels were compared between the lip regions within each species. For each comparison, the transcripts abundances of all samples included in the comparison were used to build a normalized expression matrix by Trinity software. Subsequently, transcripts showing differential expression were identified through edgeR package [91–94] and the R Bioconductor software (R version 3.4.4, R Development Core Team 2018). The differentially expressed genes were filtered by a false-discovery rate (FDR) cutoff of 0.05 [95] and those with minimum of 2 fold-expression change were used to create heatmaps. Finally, the enrichment for gene ontology (GO) terms for biological process was performed through Database for Annotation, Visualization and Integrated Discovery (DAVID) [96]. Furthermore, the knowledge-based interactions between the gene products (109 DE genes) were investigated by STRING v10 (<http://string-db.org/>), using zebrafish databases for protein interactomes [97].

Primer design and qPCR

The qPCR primers for candidate genes (selected based on the RNA-seq results), were designed after aligning their assembled sequences to their homologous sequences from other East African cichlid tribes from Lake Tanganyika [98–100], as well as *Oreochromis niloticus* (Tilapiini). Through these alignments, we identified the conserved sequence regions across East African cichlids at the exon junctions (using CLC Genomic Workbench, CLC Bio, Denmark, and annotated genome of *Astatotilapia burtoni* from the Ensembl database, <http://www.ensembl.org>). The primers were designed for short amplicon size (<200 bp) using Primer Express 3.0 (Applied Biosystems, CA, USA), and their secondary structures and dimerization were assessed through OligoAnalyzer 3.1 (Integrated DNA Technology) (Table S2). The qPCR steps provided by the protocol of Maxima SYBR Green/ROX qPCR Master Mix (2X) (Thermo Fisher Scientific, Germany) were performed following the guidelines for optimal experimental set-up for each

qPCR run [101]. The qPCR program was set for a 2 min at 50°C, a 10 min at 95°C, 40 cycles of 15 sec at 95°C and 1 min at 60°C, followed by an additional step of dissociation at 60°C – 95°C. The primer efficiency (E values) for each gene was calculated through standard curves generated by serial dilutions of pooled cDNA samples. The standard curves were run in triplicates and calculated using the following formula: $E = 10^{-1/\text{slope}}$ (Table S2).

In order to select candidate reference genes, we used the transcriptome data and followed an approach that we have already used in our previous studies [10, 102]. In brief, we first identified the genes showing no expression difference (FDR = 1) between the lip regions for each species and ranked them according to their level of expression to attain those with highest expression. Next, we ranked the genes based on their coefficient of variation (CV of expression levels) across the replicates and we selected the top ten genes shared between the transcriptome comparisons of both species as candidate reference genes. Finally, after qPCR expression analysis of the ten genes across all samples, we ranked them based on their expression stability by three different algorithms: BestKeeper [103], NormFinder [104] and geNorm [105]. Thus, we used the geometric means of the Cq values of the top two most stable reference genes to normalize Cq values of target genes in each sample ($\Delta Cq \text{ target} = Cq \text{ target} - Cq \text{ reference}$). The relative expression levels (RQ) were calculated by $2^{-\Delta\Delta Cq}$ approach [106] and the log-transformed RQ values were used for ANOVA and Tukey's HSD post hoc tests to calculate to statistically differences between the groups.

Declarations

Acknowledgements

The authors thank Holger Zimmermann and Stephan Koblmüller for sharing their precious knowledge on cichlid fishes of Lake Tanganyika, Sylvia Schäffer for sharing her experience on RNA-seq library preparation, and Martin Grube and his lab for technical assistance and access to their real-time PCR System. The authors acknowledge the financial support by the University of Graz.

Author Contributions

EPA, LL, CS and WG conceived the study. WG did the fish handling and photography, and EPA and AD conducted the sampling and tissue dissection. LL, EPA and AD conducted the RNA lab work and analyses of the gene expression data. EPA, PS, LL and CS wrote the manuscript with input from AD and WG. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

All experimental protocols related to the fishes used in this study were approved by the Federal Ministry of Science, Research and Economy of Austria, under permit BMFWF-66.007/0004-WF/V/3b/2016 in accordance with the ethical guidelines and regulations of the BMFWF.

Consent for publication

Not applicable

Funding

The project is funded by Austrian Science Fund granted to CS (Grant number: P29838).

Availability of data and materials

All the data represented in this study are provided within the main manuscript or in the supplementary materials.

References

1. Schartl M. Beyond the zebrafish: diverse fish species for modeling human disease. *Dis Model Mech.* 2014;7:181. doi:10.1242/DMM.012245.
2. Powder KE, Albertson RC. Cichlid fishes as a model to understand normal and clinical craniofacial variation. *Dev Biol.* 2016;415:338–46. doi:10.1016/J.YDBIO.2015.12.018.
3. Hulsey CD, Fraser GJ, Strelman JT. Evolution and Development of Complex Biomechanical Systems: 300 Million Years of Fish Jaws. *Zebrafish.* 2005;2:243–57. <http://www.ncbi.nlm.nih.gov/pubmed/18248183>.
4. Ahi EP. Signalling pathways in trophic skeletal development and morphogenesis: Insights from studies on teleost fish. *Dev Biol.* 2016;420:11–31. doi:10.1016/J.YDBIO.2016.10.003.
5. Machado-Schiaffino G, Henning F, Meyer A. Species-specific differences in adaptive phenotypic plasticity in an ecologically relevant trophic trait: hypertrophic lips in Midas cichlid fishes. *Evolution (N Y).* 2014;68:2086–91. doi:10.1111/evo.12367.
6. Manousaki T, Hull PM, Kusche H, Machado-Schiaffino G, Franchini P, Harrod C, et al. Parsing parallel evolution: ecological divergence and differential gene expression in the adaptive radiations of thick-lipped Midas cichlid fishes from Nicaragua. *Mol Ecol.* 2013;22:650–69. <http://www.ncbi.nlm.nih.gov/pubmed/23057963>. Accessed 11 Aug 2015.

7. Colombo M, Diepeveen ET, Muschick M, Santos ME, Indermaur A, Boileau N, et al. The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. *Mol Ecol*. 2013;22:670–84. doi:10.1111/mec.12029.
8. Concannon MR, Albertson RC. The genetic and developmental basis of an exaggerated craniofacial trait in East African cichlids. *J Exp Zool B Mol Dev Evol*. 2015;324:662–70. doi:10.1002/jez.b.22641.
9. Conith MR, Hu Y, Conith AJ, Maginnis MA, Webb JF, Albertson RC. Genetic and developmental origins of a unique foraging adaptation in a Lake Malawi cichlid genus. *Proc Natl Acad Sci*. 2018;115:7063–8. doi:10.1073/PNAS.1719798115.
10. Lecaudey LA, Sturmbauer C, Singh P, Ahi EP. Molecular mechanisms underlying nuchal hump formation in dolphin cichlid, *Cyrtocara moorii*. *Sci Rep*. 2019;9:1–13.
11. Henning F, Machado-Schiaffino G, Baumgarten L, Meyer A. Genetic dissection of adaptive form and function in rapidly speciating cichlid fishes. *Evolution (N Y)*. 2017;71:1297–312. doi:10.1111/evo.13206.
12. Baumgarten L, Machado-Schiaffino G, Henning F, Meyer A. What big lips are good for: On the adaptive function of repeatedly evolved hypertrophied lips of cichlid fishes. *Biol J Linn Soc*. 2015;115:448–55. doi:10.1111/bij.12502.
13. Vranken N, Van Steenberge M, Kayenbergh A, Snoeks J. The lobed-lipped species of *Haplochromis* (Teleostei, Cichlidae) from Lake Edward, two instead of one. *J Great Lakes Res*. 2020;46:1079–89.
14. Machado-Schiaffino G, Kautt AF, Torres-Dowdall J, Baumgarten L, Henning F, Meyer A. Incipient speciation driven by hypertrophied lips in Midas cichlid fishes? *Mol Ecol*. 2017;26:2348–62. doi:10.1111/mec.14029.
15. Duftner N, Koblmüller S, Sturmbauer C. Evolutionary Relationships of the Limnochromini, a Tribe of Benthic Deepwater Cichlid Fish Endemic to Lake Tanganyika, East Africa. *J Mol Evol*. 2005;60:277–89. doi:10.1007/s00239-004-0017-8.
16. Kirchberger PC, Sefc KM, Sturmbauer C, Koblmüller S. Outgroup effects on root position and tree topology in the AFLP phylogeny of a rapidly radiating lineage of cichlid fish. *Mol Phylogenet Evol*. 2014;70:57–62. doi:10.1016/J.YMPEV.2013.09.005.
17. Ronco F, Matschiner M, Böhne A, Boila A, Büscher HH, El Taher A, et al. Drivers and dynamics of a massive adaptive radiation in cichlid fishes. *Nature*. 2021;589:76–81.
18. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37 Web Server issue:W202-8. doi:10.1093/nar/gkp335.
19. Kubista M, Andrade JM, Bengtsson M, Forootan A, Jonák J, Lind K, et al. The real-time polymerase chain reaction. *Mol Aspects Med*. 2006;27:95–125. doi:10.1016/j.mam.2005.12.007.
20. Ahi EP, Richter F, Sefc KM. A gene expression study of ornamental fin shape in *Neolamprologus brichardi*, an African cichlid species. *Sci Rep*. 2017;7:17398. doi:10.1038/s41598-017-17778-0.
21. Ahi EP, Sefc KM. Anterior-posterior gene expression differences in three Lake Malawi cichlid fishes with variation in body stripe orientation. *PeerJ*. 2017;5:e4080. doi:10.7717/peerj.4080.

22. Ahi EP, Sefc KM. A gene expression study of dorso-ventrally restricted pigment pattern in adult fins of *Neolamprologus meeli*, an African cichlid species. PeerJ. 2017;5:e2843. doi:10.7717/peerj.2843.
23. Ahi EP, Singh P, Lecaudey LA, Gessl W, Sturmbauer C. Maternal mRNA input of growth and stress-response-related genes in cichlids in relation to egg size and trophic specialization. Evodevo. 2018;9:23. doi:10.1186/s13227-018-0112-3.
24. Yang CG, Wang XL, Tian J, Liu W, Wu F, Jiang M, et al. Evaluation of reference genes for quantitative real-time RT-PCR analysis of gene expression in Nile tilapia (*Oreochromis niloticus*). Gene. 2013;527:183–92.
25. Yan F, Dai Y, Iwata J, Zhao Z, Jia P. An integrative, genomic, transcriptomic and network-assisted study to identify genes associated with human cleft lip with or without cleft palate. BMC Med Genomics. 2020;13:39. doi:10.1186/s12920-020-0675-4.
26. Jiang R, Bush JO, Lidral AC. Development of the upper lip: Morphogenetic and molecular mechanisms. Developmental Dynamics. 2006;235:1152–66.
27. Spears R, Svoboda KKH. Growth factors and signaling proteins in craniofacial development. Semin Orthod. 2005;11:184–98.
28. Drosten M, Dhawahir A, Sum EYM, Urosevic J, Lechuga CG, Esteban LM, et al. Genetic analysis of Ras signalling pathways in cell proliferation, migration and survival. EMBO J. 2010;29:1091–104. doi:10.1038/emboj.2010.7.
29. Papaioannou G, Mirzamohammadi F, Kobayashi T. Ras signaling regulates osteoprogenitor cell proliferation and bone formation. Cell Death Dis. 2016;7:e2405–e2405.
30. Takashima A, Faller D V. Targeting the RAS oncogene. Expert Opinion on Therapeutic Targets. 2013;17:507–31.
31. Doma E, Rupp C, Baccarini M. EGFR-Ras-Raf Signaling in Epidermal Stem Cells: Roles in Hair Follicle Development, Regeneration, Tissue Remodeling and Epidermal Cancers. Int J Mol Sci. 2013;14:19361–84. doi:10.3390/ijms141019361.
32. Uz E, Alanay Y, Aktas D, Vargel I, Gucer S, Tuncbilek G, et al. Disruption of ALX1 Causes Extreme Microphthalmia and Severe Facial Clefing: Expanding the Spectrum of Autosomal-Recessive ALX-Related Frontonasal Dysplasia. Am J Hum Genet. 2010;86:789–96.
33. Twigg SRF, Versnel SL, Nürnberg G, Lees MM, Bhat M, Hammond P, et al. Frontorhiny, a Distinctive Presentation of Frontonasal Dysplasia Caused by Recessive Mutations in the ALX3 Homeobox Gene. Am J Hum Genet. 2009;84:698–705.
34. Ehret JK, Engels H, Cremer K, Becker J, Zimmermann JP, Wohlleber E, et al. Microdeletions in 9q33.3-q34.11 in five patients with intellectual disability, microcephaly, and seizures of incomplete penetrance: Is STXBP1 not the only causative gene? Mol Cytogenet. 2015;8:1–14.
35. Kuroda Y, Saito T, Nagai J-I, Ida K, Naruto T, Masuno M, et al. Microdeletion of 19p13.3 in a girl with Peutz-Jeghers syndrome, intellectual disability, hypotonia, and distinctive features. Am J Med Genet Part A. 2015;167:389–93. doi:10.1002/ajmg.a.36813.

36. Engels H, Schüler HM, Zink AM, Wohlleber E, Brockschmidt A, Hoischen A, et al. A phenotype map for 14q32.3 terminal deletions. *Am J Med Genet Part A*. 2012;158A:695–706. doi:10.1002/ajmg.a.35256.
37. Al-Kindi A, Kizhakkedath P, Xu H, John A, Sayegh AA, Ganesh A, et al. A novel mutation in *DDR2* causing spondylo-meta-epiphyseal dysplasia with short limbs and abnormal calcifications (SMED-SL) results in defective intra-cellular trafficking. *BMC Med Genet*. 2014;15:42. doi:10.1186/1471-2350-15-42.
38. Ozgen HM, Staal WG, Barber JC, De Jonge M V., Eleveld MJ, Beemer FA, et al. A novel 6.14 Mb duplication of chromosome 8p21 in a patient with autism and self mutilation. *J Autism Dev Disord*. 2009;39:322–9.
39. Chang HW, Yen CY, Chen CH, Tsai JH, Tang JY, Chang YT, et al. Evaluation of the mRNA expression levels of integrins $\alpha 3$, $\alpha 5$, $\beta 1$ and $\beta 6$ as tumor biomarkers of oral squamous cell carcinoma. *Oncol Lett*. 2018;16:4773–81.
40. Doubaj Y, De Sandre-Giovannoli A, Vera E-V, Navarro CL, Elalaoui SC, Tajir M, et al. An inherited *LMNA* gene mutation in atypical Progeria syndrome. *Am J Med Genet Part A*. 2012;158A:2881–7. doi:10.1002/ajmg.a.35557.
41. Marulanda J, Murshed M. Role of Matrix Gla protein in midface development: Recent advances. *Oral Dis*. 2018;24:78–83. doi:10.1111/odi.12758.
42. Li H, Jones KL, Hooper JE, Williams T. The molecular anatomy of mammalian upper lip and primary palate fusion at single cell resolution. *Dev*. 2019;146.
43. Tsuda M, Yamada T, Mikoya T, Sogabe I, Nakashima M, Minakami H, et al. A type of familial cleft of the soft palate maps to 2p24.2-p24.1 or 2p21-p12. *J Hum Genet*. 2010;55:124–6.
44. Haldeman-Englert CR, Naeem T, Geiger EA, Warnock A, Feret H, Ciano M, et al. A 781-kb deletion of 13q12.3 in a patient with Peters plus syndrome. *American Journal of Medical Genetics, Part A*. 2009;149:1842–5.
45. Song L, Li Y, Wang K, Wang YZ, Molotkov A, Gao L, et al. Lrp6-mediated canonical Wnt signaling is required for lip formation and fusion. *Development*. 2009;136:3161–71.
46. Brugmann SA, Powder KE, Young NM, Goodnough LH, Hahn SM, James AW, et al. Comparative gene expression analysis of avian embryonic facial structures reveals new candidates for human craniofacial disorders. *Hum Mol Genet*. 2010;19:920–30.
47. Gritli-Linde A. The Mouse as a Developmental Model for Cleft Lip and Palate Research. In: *Frontiers of Oral Biology*. Karger Publishers; 2012. p. 32–51. doi:10.1159/000337523.
48. Yang J, Yu X, Zhu G, Wang R, Lou S, Zhu W, et al. Integrating GWAS and eQTL to predict genes and pathways for non-syndromic cleft lip with or without palate. *Oral Dis*. 2020;:odi.13699. doi:10.1111/odi.13699.
49. Marinucci L, Balloni S, Bodo M, Carinci F, Pezzetti F, Stabellini G, et al. Patterns of some extracellular matrix gene expression are similar in cells from cleft lip-palate patients and in human palatal fibroblasts exposed to diazepam in culture. *Toxicology*. 2009;257:10–6.

50. Kaname T, Yanagi K, Chinen Y, Makita Y, Okamoto N, Maehara H, et al. Mutations in CD96, a member of the immunoglobulin superfamily, cause a form of the C (Opitz trigonocephaly) syndrome. *Am J Hum Genet.* 2007;81:835–41.
51. Bartholdi D, Roelfsema JH, Papadia F, Breuning MH, Niedrist D, Hennekam RC, et al. Genetic heterogeneity in Rubinstein-Taybi syndrome: Delineation of the phenotype of the first patients carrying mutations in EP300. *J Med Genet.* 2007;44:327–33.
52. Woods SA, Robinson HB, Kohler LJ, Agamanolis D, Sterbenz G, Khalifa M. Exome sequencing identifies a novel *EP300* frame shift mutation in a patient with features that overlap cornelia de lange syndrome. *Am J Med Genet Part A.* 2014;164:251–8. doi:10.1002/ajmg.a.36237.
53. Chen CP, Chen CY, Chern SR, Wu PS, Chen SW, Lai ST, et al. Prenatal diagnosis of a 1.6-Mb 4p16.3 interstitial microdeletion encompassing FGFRL1 and TACC3 associated with bilateral cleft lip and palate of Wolf-Hirschhorn syndrome facial dysmorphism and short long bones. *Taiwan J Obstet Gynecol.* 2017;56:821–6.
54. Kurosaka H, Iulianella A, Williams T, Trainor PA. Disrupting hedgehog and WNT signaling interactions promotes cleft lip pathogenesis. *J Clin Invest.* 2014;124:1660–71.
55. Mohamad Shah NS, Salahshourifar I, Sulong S, Wan Sulaiman WA, Halim AS. Discovery of candidate genes for nonsyndromic cleft lip palate through genome-wide linkage analysis of large extended families in the Malay population. *BMC Genet.* 2016;17:1–9.
56. Gajera M, Desai N, Suzuki A, Li A, Zhang M, Jun G, et al. MicroRNA-655-3p and microRNA-497-5p inhibit cell proliferation in cultured human lip cells through the regulation of genes related to human cleft lip. *BMC Med Genomics.* 2019;12:1–18.
57. Vergult S, Dauber A, Chiaie BD, Van Oudenhove E, Simon M, Rihani A, et al. 17q24.2 microdeletions: A new syndromal entity with intellectual disability, truncal obesity, mood swings and hallucinations. *Eur J Hum Genet.* 2012;20:534–9.
58. Osoegawa K, Vessere GM, Utami KH, Mansilla MA, Johnson MK, Riley BM, et al. Identification of novel candidate genes associated with cleft lip and palate using array comparative genomic hybridisation. *J Med Genet.* 2008;45:81–6.
59. Ma Q. Induction of CYP1A1. The AhR / DRE Paradigm Transcription, Receptor Regulation, and Expanding Biological Roles. *Curr Drug Metab.* 2005;2:149–64.
60. Braeuning A. Regulation of Cytochrome P450 Expression by Ras- and β -Catenin-Dependent Signaling. *Curr Drug Metab.* 2009;10:138–58.
61. Ahi EP, Steinhäuser SS, Pálsson A, Franzdóttir SR, Snorrason SS, Maier VH, et al. Differential expression of the Aryl hydrocarbon receptor pathway associates with craniofacial polymorphism in sympatric Arctic charr. *Evodevo.* 2015;6:27.
62. Barrow LL, Wines ME, Romitti PA, Holdener BC, Murray JC. Aryl hydrocarbon receptor nuclear translocator 2 (ARNT2): Structure, gene mapping, polymorphisms, and candidate evaluation for human orofacial clefts. *Teratology.* 2002;66:85–90. doi:10.1002/tera.10062.

63. Linnenkamp BDW, Raskin S, Esposito SE, Herai RH. A comprehensive analysis of AHRR gene as a candidate for cleft lip with or without cleft palate. *Mutation Research - Reviews in Mutation Research*. 2020;785:108319.
64. Stuppia L, Capogreco M, Marzo G, La Rovere D, Antonucci I, Gatta V, et al. Genetics of Syndromic and Nonsyndromic Cleft Lip and Palate. *J Craniofac Surg*. 2011;22:1722–6. doi:10.1097/SCS.0b013e31822e5e4d.
65. Vijayan V, Ummer R, Weber R, Silva R, Letra A. Association of WNT pathway genes with nonsyndromic cleft lip with or without cleft palate. *Cleft Palate-Craniofacial J*. 2018;55:335–41. doi:10.1177/1055665617732782.
66. Pereira T dos SF, Amorim LSD, Pereira NB, Vitória JG, Duarte-Andrade FF, Guimarães LM, et al. Oral pyogenic granulomas show MAPK/ERK signaling pathway activation, which occurs independently of *BRAF*, *KRAS*, *HRAS*, *NRAS*, *GNA11*, and *GNA14* mutations. *J Oral Pathol Med*. 2019;48:906–10. doi:10.1111/jop.12922.
67. Androutsopoulos VP, Tsatsakis AM, Spandidos DA. Cytochrome P450 CYP1A1: Wider roles in cancer progression and prevention. *BMC Cancer*. 2009;9:1–17.
68. Winslow S, Scholz A, Rappl P, Brauß TF, Mertens C, Jung M, et al. Macrophages attenuate the transcription of CYP1A1 in breast tumor cells and enhance their proliferation. *PLoS One*. 2019;14:e0209694. doi:10.1371/journal.pone.0209694.
69. Webb LMC, Pascall JC, Hepburn L, Carter C, Turner M, Butcher GW. Generation and Characterisation of Mice Deficient in the Multi-GTPase Domain Containing Protein, GIMAP8. *PLoS One*. 2014;9:e110294. doi:10.1371/journal.pone.0110294.
70. Menssen A, Häupl T, Sittlinger M, Delorme B, Charbord P, Ringe J. Differential gene expression profiling of human bone marrow-derived mesenchymal stem cells during adipogenic development. *BMC Genomics*. 2011;12:1–17.
71. Peixoto da-Silva J, Lourenço S, Nico M, Silva FH, Martins MT, Costa-Neves A. Expression of laminin-5 and integrins in actinic cheilitis and superficially invasive squamous cell carcinomas of the lip. *Pathol Res Pract*. 2012;208:598–603.
72. Wegner J, Loser K, Apsite G, Nischt R, Eckes B, Krieg T, et al. Laminin α 5 in the keratinocyte basement membrane is required for epidermal–dermal intercommunication. *Matrix Biol*. 2016;56:24–41.
73. Draaken M, Mughal SS, Pennimpede T, Wolter S, Wittler L, Ebert A-K, et al. Isolated bladder exstrophy associated with a de novo 0.9 Mb microduplication on chromosome 19p13.12. *Birth Defects Res Part A Clin Mol Teratol*. 2013;97:133–9. doi:10.1002/bdra.23112.
74. Kosaki K, Saito H, Kosaki R, Torii C, Kishi K, Takahashi T. Branchial arch defects and 19p13.12 microdeletion: Defining the critical region into a 0.8 M base interval. *Am J Med Genet Part A*. 2011;155:2212–4. doi:10.1002/ajmg.a.33908.
75. Ng CF, Ng PKS, Lui VWY, Li J, Chan JYW, Fung KP, et al. FHL2 exhibits anti-proliferative and anti-apoptotic activities in liver cancer cells. *Cancer Lett*. 2011;304:97–106.

76. Labalette C, Nouët Y, Sobczak-Thepot J, Armengol C, Levillayer F, Gendron MC, et al. The LIM-only protein FHL2 regulates cyclin D1 expression and cell proliferation. *J Biol Chem*. 2008;283:15201–8.
77. Dassati S, Waldner A, Schweigreiter R. Apolipoprotein D takes center stage in the stress response of the aging and degenerative brain. *Neurobiology of Aging*. 2014;35:1632–42.
78. Santos ME, Braasch I, Boileau N, Meyer BS, Sauter L, Böhne A, et al. The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. *Nat Commun*. 2014;5:1–11.
79. Shaw-Smith C. Genetic factors in esophageal atresia, tracheo-esophageal fistula and the VACTERL association: Roles for FOXF1 and the 16q24.1 FOX transcription factor gene cluster, and review of the literature. *European Journal of Medical Genetics*. 2010;53:6–13.
80. Xu J, Liu H, Lan Y, Aronow BJ, Kalinichenko V V., Jiang R. A Shh-Foxf-Fgf18-Shh Molecular Circuit Regulating Palate Development. *PLOS Genet*. 2016;12:e1005769. doi:10.1371/journal.pgen.1005769.
81. Wang B, Lin D, Li C, Tucker P. Multiple Domains Define the Expression and Regulatory Properties of Foxp1 Forkhead Transcriptional Repressors. *J Biol Chem*. 2003;278:24259–68.
82. Meerschaut I, Rochefort D, Revençu N, Pètre J, Corsello C, Rouleau GA, et al. FOXP1-related intellectual disability syndrome: A recognisable entity. *J Med Genet*. 2017;54:613–23.
83. Leishman E, Howard JM, Garcia GE, Miao Q, Ku AT, Dekker JD, et al. Foxp1 maintains hair follicle stem cell quiescence through regulation of Fgf18. *Dev*. 2013;140:3809–18.
84. Stephen TL, Rutkowski MR, Allegrezza MJ, Perales-Puchalt A, Tesone AJ, Svoronos N, et al. Transforming growth factor β -mediated suppression of antitumor T cells requires Foxp1 transcription factor expression. *Immunity*. 2014;41:427–39.
85. Cheng L, Shi X, Huo D, Zhao Y, Zhang H. MiR-449b-5p regulates cell proliferation, migration and radioresistance in cervical cancer by interacting with the transcription suppressor FOXP1. *Eur J Pharmacol*. 2019;856:172399.
86. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2012. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
87. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
88. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 2013;8:1494–512.
89. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29:644–52.
90. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34:525–7. doi:10.1038/nbt.3519.
91. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–40.

- doi:10.1093/bioinformatics/btp616.
92. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010;11:R25. doi:10.1186/gb-2010-11-3-r25.
 93. Lun ATL, Chen Y, Smyth GK. It's DE-licious: A Recipe for Differential Expression Analyses of RNA-seq Experiments Using Quasi-Likelihood Methods in edgeR. Humana Press, New York, NY; 2016. p. 391–416. doi:10.1007/978-1-4939-3578-9_19.
 94. Chen Y, Lun ATL, Smyth GK. Differential Expression Analysis of Complex RNA-seq Experiments Using edgeR. In: *Statistical Analysis of Next Generation Sequencing Data*. Cham: Springer International Publishing; 2014. p. 51–74. doi:10.1007/978-3-319-07212-8_3.
 95. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A Practical and powerful approach to multiple testing. *JRoyStatistSoc.* 1995;57:289–300. <http://inspirehep.net/record/1625829/>. Accessed 18 Oct 2018.
 96. Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, et al. The DAVID Gene Functional Classification Tool: A novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* 2007;8:1–16.
 97. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* 2017;45:D362–8. doi:10.1093/nar/gkw937.
 98. Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, et al. The genomic substrate for adaptive radiation in African cichlid fish. *Nature.* 2014;513:375–81. doi:10.1038/nature13726.
 99. Santos ME, Baldo L, Gu L, Boileau N, Musilova Z, Salzburger W. Comparative transcriptomics of anal fin pigmentation patterns in cichlid fishes. *BMC Genomics.* 2016;17:712. doi:10.1186/s12864-016-3046-y.
 100. Singh P, Börger C, More H, Sturmbauer C. The Role of Alternative Splicing and Differential Gene Expression in Cichlid Adaptive Radiation. *Genome Biol Evol.* 2017;9:2764–81. doi:10.1093/gbe/evx204.
 101. Hellems J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* 2007;8:R19. doi:10.1186/gb-2007-8-2-r19.
 102. Ahi EP, Singh P, Duenser A, Gessl W, Sturmbauer C. Divergence in larval jaw gene expression reflects differential trophic adaptation in haplochromine cichlids prior to foraging. *BMC Evol Biol.* 2019;19:150. doi:10.1186/s12862-019-1483-3.
 103. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pairwise correlations. *Biotechnol Lett.* 2004;26:509–15. <http://www.ncbi.nlm.nih.gov/pubmed/15127793>. Accessed 14 Jul 2016.
 104. Andersen CL, Jensen JL, Ørntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization,

applied to bladder and colon cancer data sets. *Cancer Res.* 2004;64:5245–50. doi:10.1158/0008-5472.CAN-04-0496.

105. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 2002;3:RESEARCH0034. <http://www.ncbi.nlm.nih.gov/pubmed/12184808>. Accessed 14 Jul 2016.
106. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 2001;29:e45. <http://www.ncbi.nlm.nih.gov/pubmed/11328886>. Accessed 14 Jul 2016.

Figures

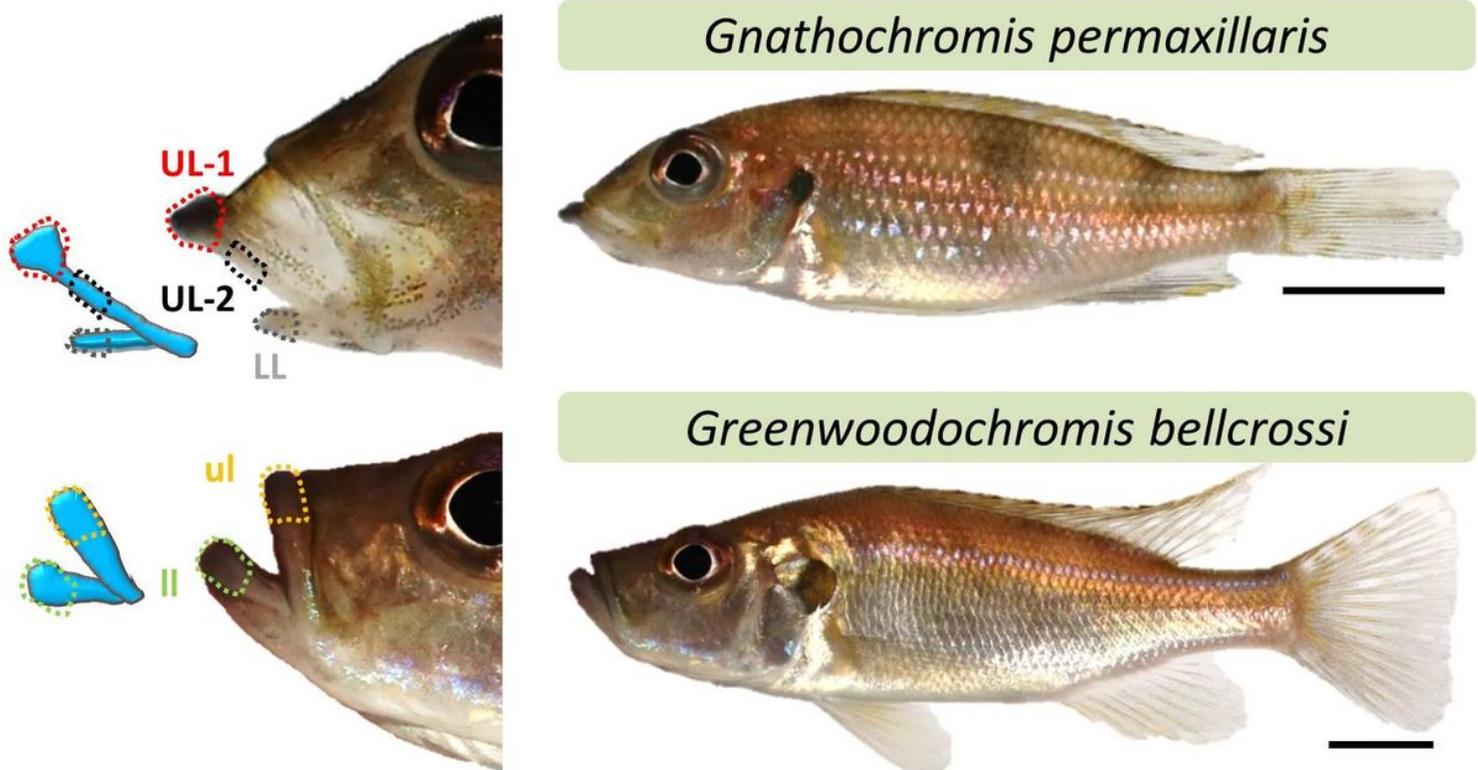
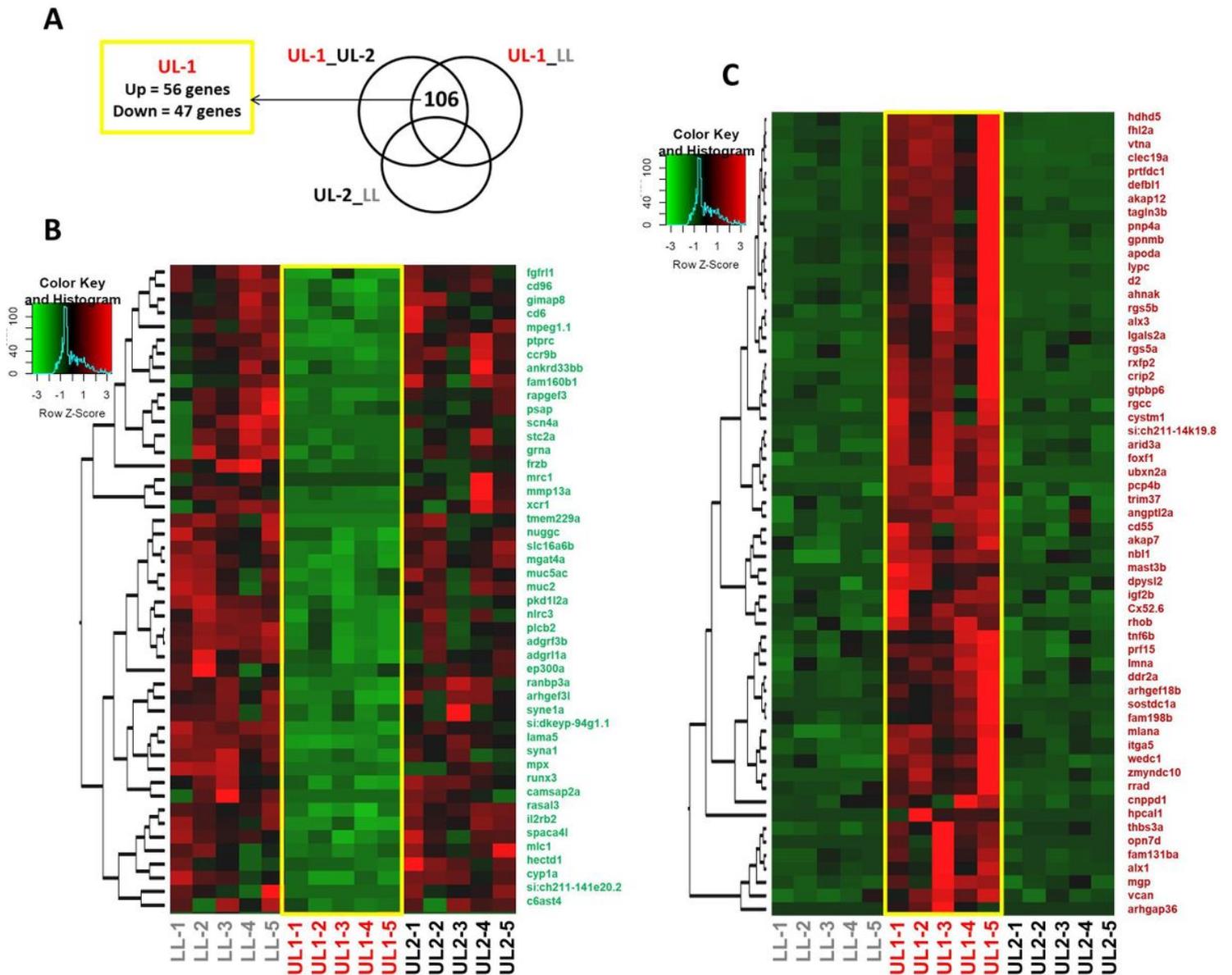


Figure 1

Two East African cichlid species of the tribe Limnochromini from Lake Tanganyika used in this study. The areas from which the soft tissue samples are taken in lips are delineated by coloured dash lines; UL/ul and LL/ll refer to upper and lower lip, respectively.



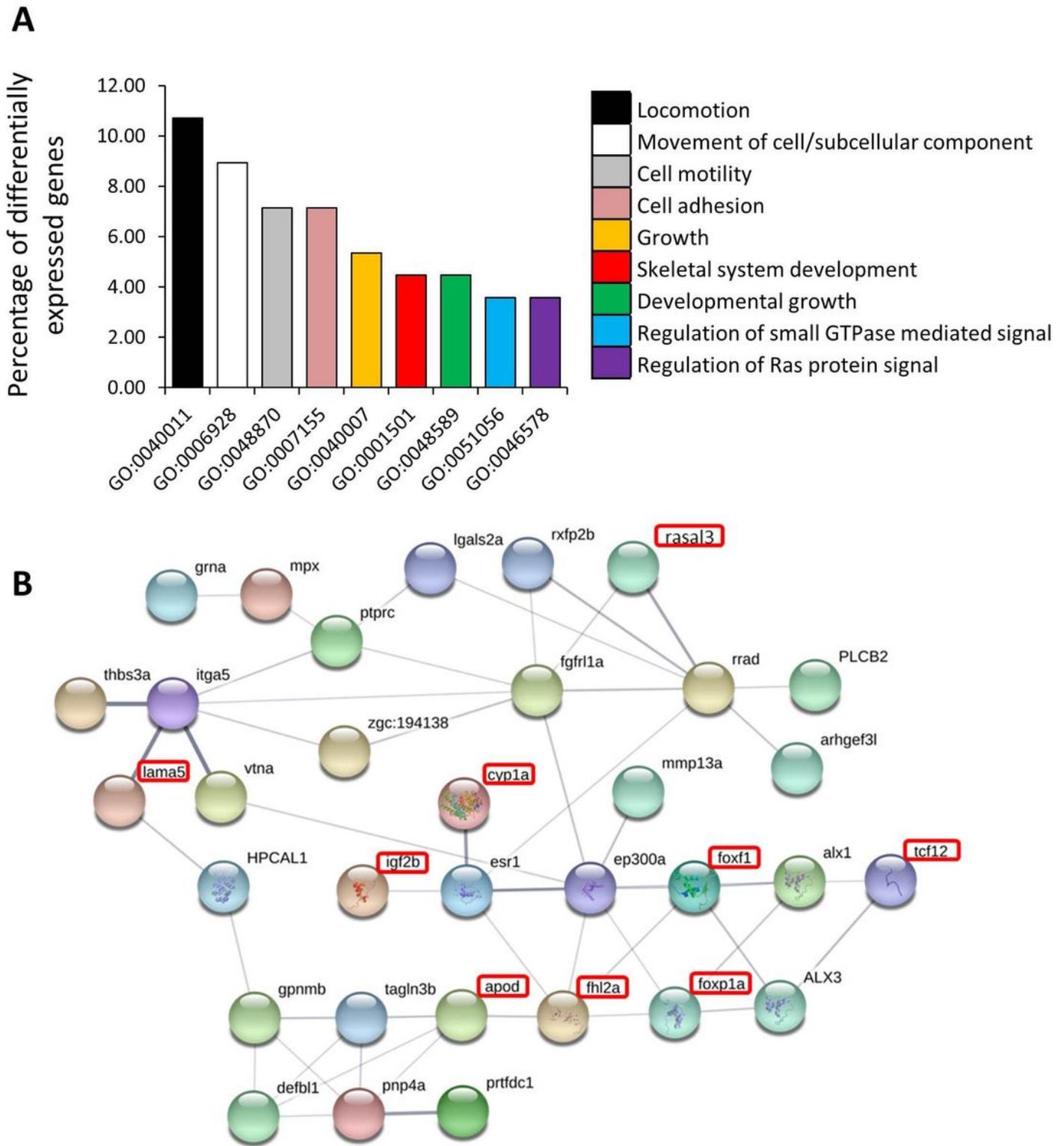


Figure 3

Downstream functional analyses of differentially expressed genes. (A) Gene ontology enrichment analysis for biological processes using the shared 106 differentially expressed genes conducted with Manteia tool. (B) Predicted functional associations/interactions between the differentially expressed genes based on zebrafish databases in STRING v10 (<http://string-db.org/>). The differential expression of the genes specified with red squares are confirmed by both qPCR and RNA-seq.

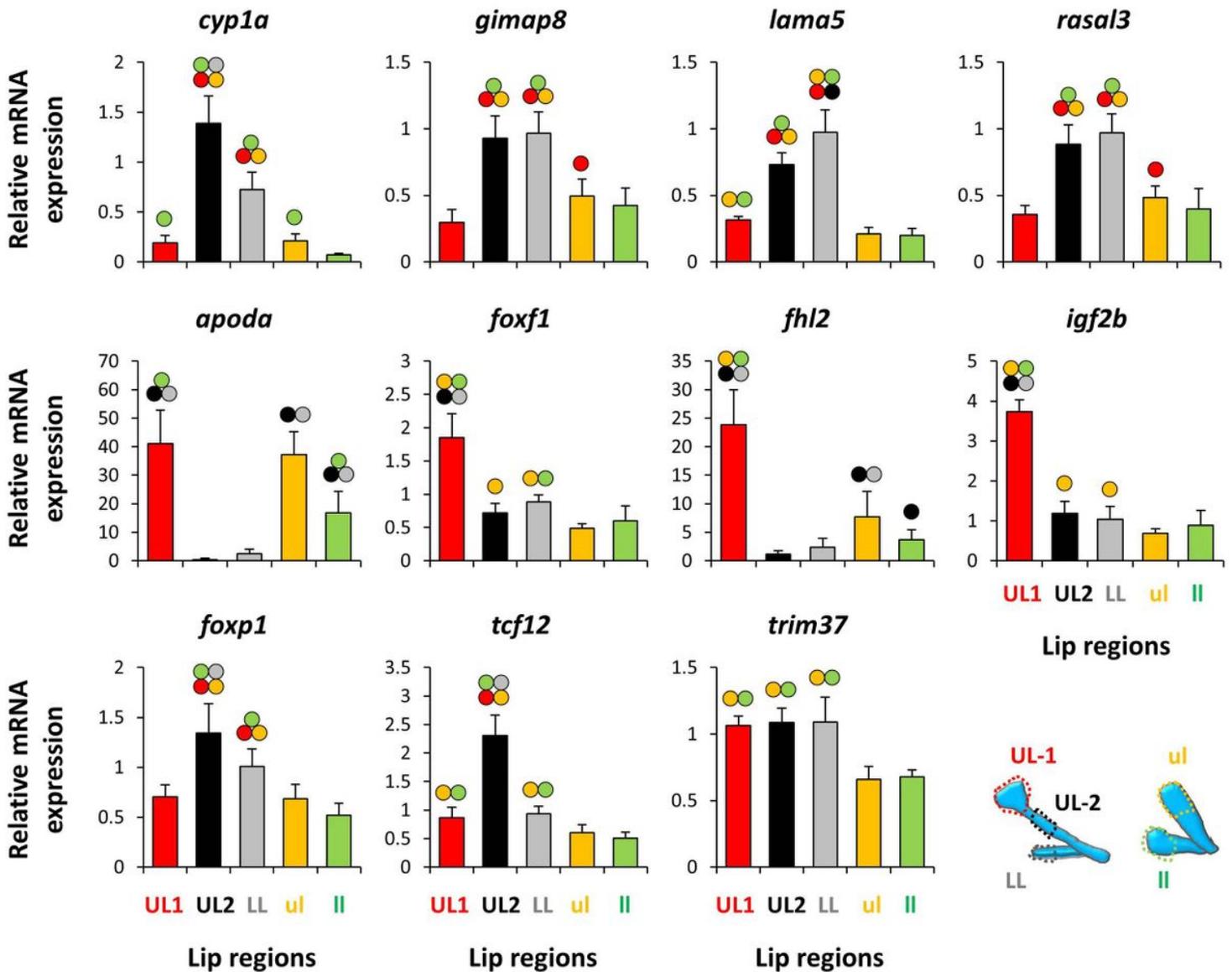


Figure 4

Expression analysis of a selection of candidate genes using qPCR. The bars represent means and standard deviations of RQ values for five biological replicates in each lip region. Circles above bars indicate significantly elevated expression ($P < 0.05$) in comparisons between the lip regions (i.e., compared to the bar matching the colour code of the circle).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.xlsx](#)
- [Additionalfile2.xlsx](#)