Conserved molecular players involved in human nose morphogenesis underlie evolution of the exaggerated snout phenotype in cichlids

Anna Duenser[1], Pooja Singh[1,2,3], Laurène Alicia Lecaudey[1,4], Christian Sturmbauer[1]*, R. Craig Albertson[5], Wolfgang Gessl[1], Ehsan Pashay Ahi[1,6]

1. Institute of Biology, University of Graz, Universitätsplatz 2, A-8010 Graz, Austria
2. Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland
3. EAWAG, Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum, Switzerland
4. Department of Natural History, NTNU University Museum, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway
5. Department of Biology, University of Massachusetts, Amherst MA 01003, USA
6. Organismal and Evolutionary Biology Research Programme, University of Helsinki, Viikinkaari 9, 00014, Helsinki, Finland

*Author of Correspondence:

Christian Sturmbauer[1], Email: christian.sturmbauer@uni-graz.at

Abstract

Instances of repeated evolution of novel phenotypes can shed light on the conserved molecular mechanisms underlying morphological diversity. A rare example of an exaggerated soft tissue phenotype is the formation of a snout flap in fishes. This tissue flap develops from the upper lip and has evolved in one cichlid genus from Lake Malawi and one genus from Lake Tanganyika. To investigate the molecular basis of snout flap convergence, we used mRNA sequencing to compare two species with snout flap to their close relatives without snout flaps from each lake. Our analysis identified 201 genes that were repeatedly differentially expressed between species with and without snout flap in both lakes, suggesting shared pathways, even though the flaps serve different functions. Shared expressed genes are involved in proline and hydroxyproline metabolism, which have been linked to human skin and facial deformities. Additionally, we found enrichment for transcription factor binding sites at upstream regulatory sequences of differentially expressed genes. Among the enriched transcription factors were members of the FOX transcription factor family, especially *foxf1* and *foxa2*, which showed an increased expression in the flapped snout. Both of these factors are linked to nose morphogenesis in mammals. We also found *ap4* (tfap4), a transcription factor showing reduced expression in the flapped snout with an unknown role in craniofacial soft tissue development. As genes involved in cichlid snout flap development are associated with human mid-line facial dysmorphologies, our findings could hint at the conservation of genes involved in mid-line patterning across distant evolutionary lineages of vertebrates, although further functional studies are required to confirm this.

Key words: RNA-seq, Lake Malawi, Lake Tanganyika, snout flap, cichlids, functional conservation

Significance statement

The study of the evolution of similar physical traits across taxa can give insight into the molecular architecture underlying shared phenotypes. This has mostly been studied in bony structures, while soft tissue traits have been less intensely covered. We investigated the exaggerated snout in cichlid species from Lake Malawi and Lake Tanganyika and found that many genes involved in the development of the snout flap and are also associated with mid-line dysmorphologies in humans, implying a conservation across distant vertebrate lineages.

Introduction

The repeated evolution of phenotypes, reflecting particular ecological specializations, is a ubiquitous characteristic of adaptive radiations (Schluter & Nagel 1995; Losos et al. 1998; Rundle et al. 2000; Rüber et al. 1999). Cichlid adaptive radiations from the East African Great lakes display an impressive array of repeated morphological traits (Kocher et al. 1993), including a few dramatic examples of exaggerated phenotypes like the overgrowth of craniofacial soft tissues in various anatomical regions such as lips (Machado-Schiaffino et al. 2014; Manousaki et al. 2013; Colombo et al. 2013; Baumgarten et al. 2015; Lecaudey et al. 2019; Henning et al. 2017), the frontal head (nuchal hump) (Lecaudey et al. 2021) and the nose snout (or nose flap) (Concannon & Albertson 2015; Conith et al. 2018). Although there is increasing insight into the evolution of such phenotypic novelties, especially regarding hypertrophid lips, exaggerated soft tissue traits are less well studied than bony traits and the genetic mechanisms underlying these traits are not entirely understood. Comparative approaches can shed light on the genetic mechanisms that reconfigure the body plan and give rise to such complex traits. With examples of both parallel and non-parallel mechanism underlying cases of repeated evolution (e.g. Manousaki et al. 2013; Colombo et al. 2013) of

phenotypic novelties such comparisons can thus also help us to understand the molecular mechanisms that shape morphological diversity.

One of these repeated exaggerated phenotypes in cichlids is the snout flap, a pronounced projection that emanates from a flap of fibrous tissue just above the upper lip. It is a rare morphological innovation that has only evolved in two tribes of cichlid fishes from East Africa, the modern Haplochromines in Lake Malawi and the Ectodini in Lake Tanganyika (Concannon & Albertson 2015). When this snout is sexually monomorphic, it is thought to be a trophic adaptation that improves feeding efficiency (Konings 2007). When the snout is sexually dimorphic, it is hypothesised to be involved in sexual selection (Konings 2007; Concannon & Albertson 2015). The cichlid snout flap has been studied at the molecular level only in the genus *Labeotropheus* from Lake Malawi where it is sexually monomorphic and functions as a trophic adaptation to efficiently leverage algae from rocks (Concannon & Albertson 2015; Conith et al. 2018). A similar snout structure has also been described in two species from the Ectodini tribe (*Ophthalmotilapia nasuta* and *Asprotilapia leptura*) from Lake Tanganyika. In *A. leptura* it is sexually monomorphic and likely involved in increased foraging efficiency (similar to *Labeotropheus*), whereas in *O. nasuta* it is only found in mature males and is likely a secondary sexual character (Hanssens et al. 1999; Conith et al. 2019). Thus, the exaggerated snout is a convergent phenotype that evolved independently in two cichlid lineages that diverged > 9 MYA (Irisarri et al. 2018; Conith et al. 2019).

In *Labeotropheus*, the snout is evident histologically by the time the yolk is absorbed and exogenous feeding occurs (~1 month post-fertilization) (Conith et al. 2018; Concannon & Albertson 2015), and the early formation and growth of the snout is linked to the transforming growth factor beta (TGFβ) signalling pathway (Conith et al. 2018). However, it remains unclear which (1) genes and pathways contribute to the maintenance of this complex trait and if (2) these candidate genes and pathways can be linked to more conserved patterning of craniofacial

features. Furthermore, while previous research focused on the TGFβ signalling pathway, a more extensive molecular interaction map of the formation and maintenance of this exaggerated phenotype remains to be unravelled. A transcriptome-wide overview is particularly important since it is well-known that there is molecular cross-talk between the TGFβ signalling pathway and several other pathways which all play a pivotal role in craniofacial morphogenesis and adaptive evolutionary divergence in teleost fishes (Ahi 2016).

In this study, we set out to investigate the molecular mechanisms that underlie the formation and evolution of the exaggerated snout phenotype, in two non-sister cichlid lineages from lakes Tanganyika and Malawi (Figure 1) and link it to conserved molecular players in mid-line patterning. We compared two species that develop a snout; (1) *Labeotropheus trewavasae* (tribe Haplochromini) from Lake Malawi and (2) *Ophthalmotilapia nasuta* (tribe Ectodini) from Lake Tanganyika (Figure 1). As controls, we used two closely related species within each tribe that do not develop such a structure; (1) the Lake Malawi mbuna species *Tropheops tropheops* (Haplochromini) and (2) the Lake Tanganyika featherfin cichlid *Ophthalmotilapia ventralis* (Ectodini) (Figure 1). We used mRNA-sequencing to quantify gene expression differences between the exaggerated snout and non-snout tissues for each lake. Altogether, we identified parallel and non-parallel molecular mechanisms that underlie the evolution of the snout flap in Lake Malawi and Lake Tanganyika cichlids. Our study design provides valuable information on conserved regulatory mechanisms underlying the morphogenesis of a unique hypertrophic facial soft tissue in cichlids, which also exhibit striking similarity to those mechanisms driving craniofacial development and mid-line patterning in other vertebrates including humans. Notably, cichlids are already introduced as excellent models to study craniofacial skeletal deformities in humans (Powder & Albertson, 2016), and our study can be one of the first indications that cichlids can be used as models to study deformities in facial soft tissues as well.

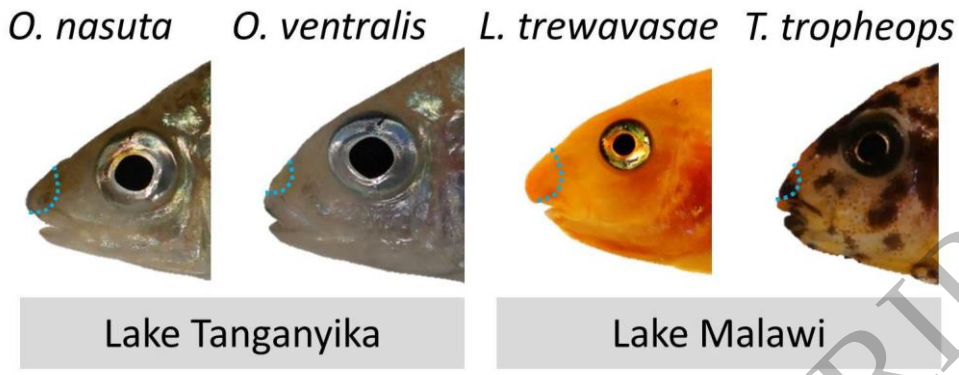

Fig.1. Convergent cases of snout flap evolution. East African cichlid species used in this study. The area of the soft tissue that was dissected is depicted by blue dashed lines. (*O. nasuta*) *Ophthalmotilapia nasuta*, (*O. ventralis*) *Ophthalmotilapia ventralis*, (*L. trewavasae*) *Labeotropheus trewavasae*, (*T. tropheops*) *Tropheops tropheops.*

Results

To investigate molecular mechanisms underlying the formation of a snout flap in two distant lineages of cichlids, we dissected the snout tissue of five biological replicates per species, which entailed the area above the upper lip including the nostrils. These tissue samples consisted of epidermis, dermis and the underlying connective tissue (Figure 1). Subsequently we extracted RNA of these five samples per species to quantify gene expression differences.

RNA-sequencing, gene expression and downstream analyses

The RNA-sequencing resulted in between 6.7 and 15.8 million reads per sample and after filtering of low-quality reads, between 4.6 and 11.1 million reads were retained for each sample (Supplementary Table S1). The raw data of sequence reads have been deposited in the Sequencing Read Archive (SRA; Supplementary Table 1) of NCBI (accession number: PRJNA770252). The final annotation of all merged species included 33,251 genes. Through pairwise comparisons between species of each lake radiation we identified 832 of the 33,251

genes (2.4%) with significant differential expression (FDR cut-off at $P < 0.05$) for the comparison of *O. nasuta* versus *O. ventralis*, while the comparison between *L. trawavasae* versus *T. tropheops* yielded 4,292 (12.7%) significant differentially expressed genes (FDR cut-off at $P < 0.05$).

GO enrichment analysis conducted for differentially expressed genes within each species pair comparison for Lake Tanganyika and Lake Malawi respectively revealed the involvement in biological processes like 'peptidyl-proline modification', 'tendon development' and 'cell adhesion' for the comparison of the Lake Tanganyika species (*O. nasuta* versus *O. ventralis*), while the Lake Malawi comparison (*L. trewavasae* versus *T. tropheops*) showed terms like 'cell matrix adhesion', 'apoptotic process involved in morphogenesis' as well as 'regulation of brown fat cell differentiation' amongst more cell specific processes (Supplementary Table S2).

To understand if similar genes were involved in the formation of a snout across the two lakes, we investigated the intersection set of the two pairwise comparisons and could identify an overlapping list of 201 differentially expressed (DE) genes which were distinct between the flapped snout versus the non-flapped snout regions in both lakes (24.2% of the differentially expressed genes (DEG) for the Lake Tanganyika comparison and 4.8% for the Lake Malawi comparison) (Figure 2A) (Supplementary Table S3). Among the shared DE genes, 84.6% showed the same direction of expression with 74 genes being upregulated and 96 genes being downregulated in the flapped snout tissues in both comparisons which is a higher number of shared expression direction than one would expect by chance (Hypergeometric test, $P < 0.05$), whereas 31 genes showed expression differences in opposite directions across the comparisons for each Lake (Figure 2B-D). The heatmap clustering of the DE genes showed that there are at least two major branches in each group of up- or down-regulated gene sets, while the clustering of the DE genes with opposite expression pattern also revealed the presence of two major branches (Figure 2B-D). These clustering structures indicate distinct transcriptional regulations

within each group which potentially originated from the effects of different upstream regulators.

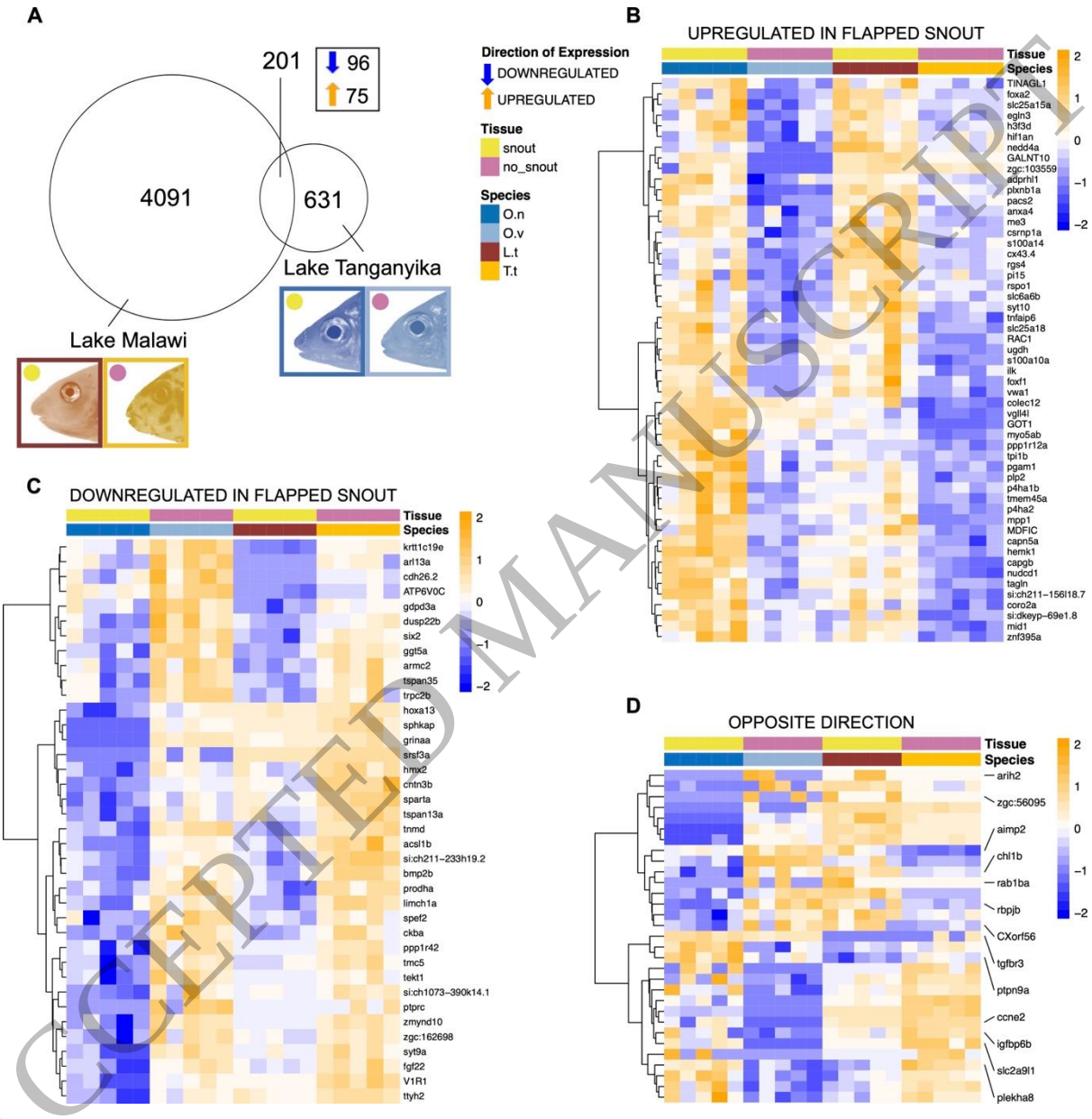

Fig. 2. Differentially expressed genes in the snout regions. (A) Venn diagram of genes with differential expression between the snout regions ("snout" and "no snout") for Lake Malawi and Lake Tanganyika and the overlap of 201 genes between the two comparisons of which 96 are downregulated and 75 are upregulated in the flapped snout of both comparisons. Dendrogram clusters of the overlapping annotated genes showing upregulation (B), and downregulation (C) in expression in the flapped snout tissue, as well as those showing differential expression in both comparisons but in

opposing directions (including not annotated genes) (D). Orange and blue shadings indicate higher and lower relative expression respectively. Lake Tanganyika: *Ophthalmotilapia nasuta* (O.n, dark blue), *Ophthalmotilapia ventralis* (O.v; light blue); Lake Malawi: *Labeotropheus trewavasae* (L.t; red), *Tropheops tropheops* (T.t; orange).

We performed gene ontology enrichment analysis using the list of the shared 201 DE genes as the input, and the result showed significant enrichment of GO terms for several biological processes such as amino acid metabolism (particularly proline related metabolic processes), 'tendon development', 'positive regulation of BMP signaling pathway' and cell adhesion and cell fate (Supplementary Table S2). When dividing the genes in their direction of expression in the snout flap, GO enrichment for upregulated genes was associated with 'peptidyl-proline hydroxylation', 'tendon development', 'muscle attachment', 'endothelial cell development', 'negative regulation of Notch signaling pathway' and although not significantly 'positive regulation of Wnt signaling pathway'. The downregulated genes were involved in a lot of terms related to cell fate commitment and negative regulation of cell fate as well as 'proline catabolic process' and 'positive regulation of BMP signaling pathway' (Supplementary Table S2).

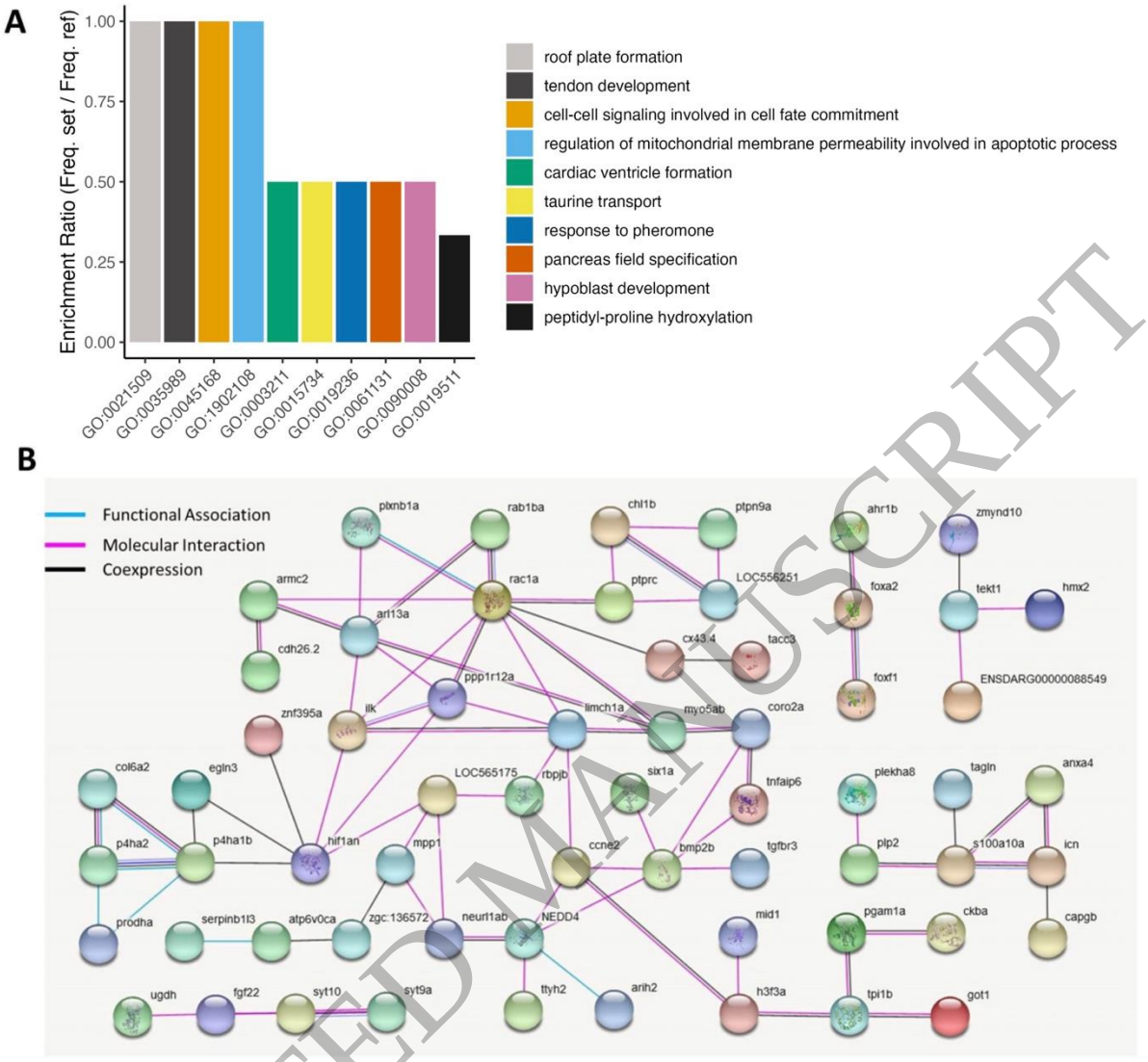

Fig. 3. Functional analyses of the overlapping differentially expressed genes in the flapped snout. (A) Enrichment for gene ontologies of biological processes using the shared 201 differentially expressed genes. (B) Functional interactions between the differentially expressed genes predicted based on zebrafish databases in STRING v10 (http://string-db.org/).

We also applied the same list of the shared 201 DE genes for interactome analysis which demonstrated a large, interconnected network of genes with molecular and functional associations. Some of the genes in the network formed an interaction hub with the highest level of associations (based on the number of predicted interactions with other DE genes) with other genes such as *bmp2b*, *hif1an* and *rac1a*, suggesting their more pivotal role in the formation of

the flapped snout structure in cichlids (Figure 3B). Furthermore, we conducted TF binding motif overrepresentation analysis on the upstream regulatory sequences of the DE genes through MEME tool (Bailey et al. 2009). In total, seven motifs were enriched on the upstream regulatory sequences of at least 40 out of 201 DE genes (Table 1). Next, we checked the similarities of the enriched motifs with known TF binding sites in vertebrates and at least 11 TF candidates were identified to potentially bind to those motifs.

Table 1. Predicted motifs and upstream regulators potentially binding to them. Enriched motifs on upstream regulatory sequences of the DE genes are presented in degenerated sequence format. PWD IDs indicate positional weight matrix ID of predicted binding sites and E-values refer to matching similarity between the predicted motif sequences and the PWD IDs. The count implies the number of genes containing the predicted motif sequence on their regulatory region.

TF binding site	PWM ID	Count	Predicted motif sequence	E-value
FOXP1	M00987	71 / 201	AMAMACAMAMAMAMAMACACACAMAMACA	3.85E-12
FOXJ1	M00742			3.52E-08
RREB1	MA0073.1			1.87E-07
FOXJ1	M00742	47 / 201	AAAAASAAAMAAAMWMWCWKT	8.69E-10
FOX	M00809			9.15E-07
FOXD3	MA0041.1			3.95E-07
SP1	MA0079.2	41 / 201	CHCCYCCYCCYCCSCYCTCCY	1.12E-08
IRF9	M00258	61 / 201	KTTTTTYTTTYYCWK	2.90E-09
MEF2	M00405	72 / 201	RTTAAAAAAAA	4.28E-08
AP4	M00927	93 / 201	CWGCTGCWGCTKSTS	7.38E-08
HEB/tcf12	M00698	66 / 201	NYYCTGCTGD	1.03E-06

Expression analysis by qPCR

Validation of DE genes from RNA-seq was accomplished via quantitative RT-PCR (qPCR), normalised to stably expressed reference genes (Kubista et al. 2006). In our previous studies of East African cichlids, we found that validation of reference gene(s) is an essential step as genes only selected from literature are not necessarily the best choice and can vary a lot between different species and tissue types (Ahi, Singh, et al. 2019; EP. Ahi et al. 2020; E P.

Ahi et al. 2020; Pashay Ahi & Sefc 2018; Ahi, Richter, et al. 2019). We chose six candidate reference genes with a small log2 fold change and the lowest coefficient of variation (CV) throughout all the samples (Supplementary Table S4). Based on the rankings by the three software tools, BestKeeper, geNorm and NormFinder, only one of the candidate reference genes, *pak2b*, showed consistent stability, i.e. always ranked among top two most stable reference genes (Table 2). Thus, we used the Cq value of *pak2b* in each sample to normalize the relative gene expression levels of our target genes.

Table 2. Ranking of reference genes in the nose tissue samples using three different algorithms.

BestKeeper				geNorm		NormFinder	
Ranking	**SD**	**Ranking**	**r**	**Ranking**	**M**	**Ranking**	**SV**
sp3	0.461	*pak2b*	0.94	*pak2b*	0.369	*nup58*	0.310
pak2b	0.471	*pphln1*	0.931	*flot2a*	0.386	*pak2b*	0.408
flot2a	0.491	*flot2a*	0.926	*pphln1*	0.397	*pphln1*	0.443
nup58	0.509	*vps26a*	0.916	*sp3*	0.418	*flot2a*	0.498
vps26a	0.551	*nup58*	0.889	*nup58*	0.427	*sp3*	0.518
pphln1	0.587	*sp3*	0.887	*vps26a*	0.428	*vps26a*	0.646

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

Among the DE genes identified by RNA-seq, we chose 12 genes with a known role in nose morphogenesis and/or other related functions in craniofacial development mainly based on genetic studies in humans (Table 3), together with eight predicted upstream TFs (including *ap4*, *foxd3*, *foxj1*, *foxp1*, *irf9*, *mef2a*, *rreb1a* and *sp1*) for qPCR analysis (Figure 4).

Table 3. A selected set of differentially expressed genes in the flapped snout regions of studied cichlids with known related functions in nose morphogenesis in mammalian models.

Gene	Related function	Organism	References
adprhl1	Duplication of this gene is associated with prominent forehead, short and bulbous nose, and broad philtrum	Human	(De Pater et al. 2005)

bmp2	A ligand of the TGFβ signaling and its monoallelic deletion is associated with short upturned nose and long philtrum	Human	(Tan et al. 2017)
dusp22	Deletion at terminal end of this gene is associated with saddle shape nose morphogenesis	Human	(Hosono et al. 2020)
fgf22	Genomic rearrangement encompassing this gene is associated with elongation of nose with prominent nasal bridge	Human	(Quigley et al. 2004)
foxa2	Both deletion and missense variation in this gene causes hallow nasal bridge, short upturned nose and downturned nasolabial folds	Human	(Dines et al. 2019)
foxf1	Duplication and triplication causes bulbous nose and wide nasal bridge	Human	(Kucharczyk et al. 2014)
hmx2	Hemizygous deletion in this gene causes broad nasal bridge and prominent nose	Human	(Miller et al. 2009)
lyve1	Dysregulation of this gene is associated with cutaneous angiosarcoma on the nose	Human	(Mitteldorf et al. 2018)
prodh	Deletion and/or missense mutations in this gene causes frontal bossing, thin upper lip and short nose	Human	(Guilmatre et al. 2010)
rac1	Loss of function mutation in this gene causes failure in fusion of medial nasal processes and prominent nasal bridge	Human Mouse	(Thomas et al. 2010; Reijnders et al. 2017)
six2	Deletion in this gene causes frontonasal dysplasia syndrome in human with nasal clefting and broad nasal tip, and developmental deformities in nasal bridge in mouse	Human Mouse	(Hufnagel et al. 2016; Okello et al. 2017)
ugdh	Missense mutation in this gene causes bulbous nose and smooth philtrum	Human	(Alhamoudi et al. 2020)

Based on the RNA-seq results, six of these candidate genes displayed upregulation in expression in the flapped snout (*adprhl1*, *foxa2*, *foxf1*, *lyve1a*, *rac1* and *ugdh*), while the six other candidate genes (*bmp2b*, *dusp22b*, *fgf22*, *hmx2*, *prodha* and *six2a*) showed a downregulation in expression in the flapped snout. The results of qPCR analysis confirmed that almost all of the genes showed expression patterns similar to RNA-seq results, except for *bmp2b* and *hmx2* which showed no significant difference between the snout regions of L.t and T.t. Among the predicted TFs only *ap4* showed consistent differences across both comparisons displaying a slightly reduced expression in both species with protruded snouts (O.n and L.t). This indicates potential transcriptional repressor effects of *ap4* on the downstream genes in the hypertrophic snout region. Two predicted members of FOX transcription factors, *foxj1* and *foxp1*, showed expression differences but only in one of the comparisons (O.n vs O.v), which makes them unlikely candidates for upstream regulators of shared DEGs in both comparisons. Altogether, the qPCR results demonstrate consistency between RNA-seq and qPCR results confirming the validity of our transcriptome data analysis.

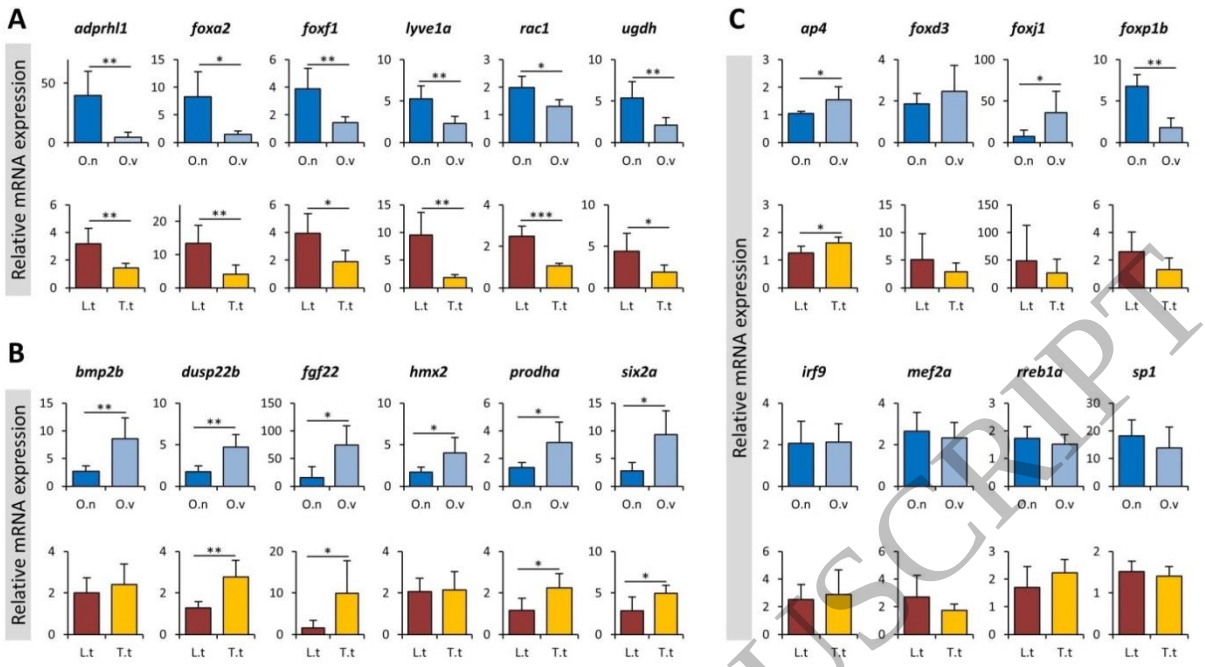

Fig. 4. qPCR expression analysis of a selected set of candidate genes. qPCR validation of expression differences for selected sets of genes showing upregulation (A) or downregulation (B) in snout tissues. (C) qPCR expression analysis of predicted transcription factors. The bars indicate mean and standard deviation of RQ expression values for five biological replicates per species. The asterisks above the bar represent significant expression differences (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). *Ophthalmotilapia nasuta* (O.n; dark blue), *Ophthalmotilapia ventralis* (O.v; light blue), *Labeotropheus trewavasae* (L.t; red), *Tropheops tropheops* (T.t; orange).

Discussion

Cases of repeated morphological evolution in recurrent diversivication events can contribute significantly to our understanding of the molecular architecture underlying shared phenotypes. The snout flap of *L. trewavasae* is thought to have evolved under natural selection (Concannon & Albertson 2015), as it plays a distinct role in the foraging efficiency for algal scraping (Konings 2007; Conith et al. 2019). Additionally, no difference in snout flap size has been detected between male and female of *Labeotropheus* and its formation has been shown to coincide with the time point of independent foraging, further supporting its function (Concannon & Albertson 2015). In contrast, only *O. nasuta* males show the distinct snout flap,

implying a role in mate choice (Concannon & Albertson 2015). Furthermore, both sexes of *O. nasuta* are planktivorous suction feeders, a feeding adaptation that is presumably not enhanced by a snout flap, although the snout of males continues to grow with increasing age (Hanssens et al. 1999). In a comparison of tissue types of the snout flap it has been found that the snout of *Labeotropheus fuelleborni* contains a lot more of intermaxillary ligament and loose connective tissue (80%) than the snout of *O. nasuta* (50%) (Conith et al. 2019). The morphological similarity of the snout flap across two cichlid radiations allows us to investigate if conserved molecular players are involved in the formation of snout, even if the morphologies possess different functions and differ in tissue composition and life-history.

We found quite differing numbers of DE genes between the comparisons within Lake Malawi and Lake Tanganyika, with roughly five times more differentially expressed genes between the chosen species pair from Lake Malawi over the species pair from Lake Tanganyika. This, most likely, can not be explained by the use of differing genera for the Lake Malawi comparison as the species flock shows a low sequence divergence of 0.1% - 0.25% probably due to the young age of the radiation (Malinsky et al. 2018) (Supplementary Fig. S1), but could be an indicator for the aforementioned difference in tissue composition of *L. fuelleborni* (Conith et al. 2019). Additionally, the GO enrichment analyses for the two within lake comparisons showed quite distinct enrichment terms. GO enrichment analysis for DE genes within the Lake Malawi comparison between *L. trewavasae* and *T. tropheops* yielded terms like ‘cell-matrix adhesion’, ‘regulation of brown fat cell differentiation’ and ‘apoptotic process involved in morphogenesis’ among terms involved in nerve development and different terms not readily connected to snout morphology (Supplementary Table S2). This could hint at a stunning difference in organization of connective tissue in *Labeotropheus* compared to *Tropheops*. Conith et al. 2018a found that the connective tissue (identified as the intermaxillary ligament) of *Labeotropheus*, which is high in collagen, invades the sourrounding loose

connective tissue and anchors to the epithelium potentially helping with the stiffness of the snout to improve foraging. The GO enrichment for DE genes between *O. nasuta* and *O. ventralis* revealed terms linked to cell fate and cell shape regulation, 'peptidyl-proline modification' and 'tendon development'. This suggests a difference in collagen/tendon development and cell adhesion and fate between the two Lake Tanganyika species, where the structure is not as unique as in *Labeotrpheus*, and shows an overall increase in the proportion of skin and other tissue, much greater than in *Labeotrpheus* (Conith et al. 2018a).

To understand if similar molecular mechanisms underly these seemingly similar phenotypes (yet different histologically) across both lakes, we looked at the intersection set of both comparisons and found many of DE genes, both upregulated and downregulated, that are associated with craniofacial development and involved in human dysmorphologies, many with mid-line facial defects including those that effect the nose in literature. Among the upregulated genes with related functions were *adprhl1* (De Pater et al. 2005), *angptl2* (Ehret et al. 2015), *colec12* (Zlotina et al. 2016), *cx43* (McLachlan et al. 2005), *foxa2* (Dines et al. 2019), *foxf1* (Kucharczyk et al. 2014), *galnt10* (Starkovich et al. 2016), *got1* (Tomkins et al. 1983), *lyve1* (Mitteldorf et al. 2018), *mdfic* (Kosho et al. 2008), *mid1* (Preiksaitiene et al. 2015)(Hüning et al. 2013), *nudcd1* (Selenti et al. 2015), *pacs2* (Holder et al. 2012), *plxnb1* (Haldeman-Englert et al. 2009), *rac1* (Thomas et al. 2010)(Reijnders et al. 2017), *rspo1* (Wieacker & Volleth 2007), *s100a10* (Sawyer et al. 2007), *slc25a18* (Chen et al. 2013), *slc6a6* (Kariminejad et al. 2015), *ugdh* (Alhamoudi et al. 2020), *vgll4* (Czeschik et al. 2014)(Barrionuevo et al. 2014), and *vwa1* (Giannikou et al. 2012). Among the downregulated genes we also found the following candidates to have such roles; *acsl1* (Yakut et al. 2015), *adgb* (Alazami et al. 2016), *arl13* (Brugmann et al. 2010), *ATP6v0c* (Mucha et al. 2019; Tinker et al. 2021), *bmp2* (Tan et al. 2017), *cntn3* (Ţuţulan-Cuniţă et al. 2012), *dusp22* (Hosono et al. 2020)(Martinez-Glez et al. 2007), *fgf22* (Quigley et al. 2004), *gdpd3* (Dell'Edera et al. 2018), *grina* (Bonaglia et al. 2005),

hmx2 (Miller et al. 2009), *hoxa13* (Fryssira et al. 2011), *il23r* (Rivera-Pedroza et al. 2017), *ppp1r42* (Mordaunt et al. 2015), *prodh* (Guilmatre et al. 2010), *six2* (Hufnagel et al. 2016)(Okello et al. 2017), *srsf3* (Pillai et al. 2019), *syt9* (Sofos et al. 2012), and *trpc2* (Sansone et al. 2014)(Zhang et al. 2010). Interestingly, one of the downregulated genes, *pi15*, is known as an important molecular player in beak formation in birds (Nimmagadda et al. 2015). Even among the overlapping DE genes which showed opposing expression patterns between the two comparisons, we still found at least four genes to have been associated with craniofacial mid-line defects in other vertebrates, including *ccne2* (Jain et al. 2010), *plekha8* (Schulz et al. 2008), *rab1b* (Alwadei et al. 2016), *RBPJ* (Nakayama et al. 2014) and *tgfbr3* (Lopes et al. 2019). The most likely explanation for opposing expression of these genes can be the existence of bimodality in their expression pattern. Bimodality of gene expression is a mechanism contributing to phenotypic diversity (Ochab-Marcinek & Tabaka 2010), and it can be reflected by up- or down-regulation of a gene during the same biological process. This regulatory bimodality can have various causes such as differential/opposing action of transcription factors (e.g. negative feedback loop), post-transcriptional factors (e.g. microRNA and circular RNA) and stochastic events. Interestingly, a highly conserved negative feedback loop in Notch signaling has already been shown to be mediated by oppsoing roles of RBPJ (Tanigaki & Honjo 2010). This indicates a potential bimodality of *rbpj* expression through a negative feedback loop in regulation of Notch signaling, which plays a crucial role in the formation of the mid line structures including nasal structures (Tanigaki & Honjo 2010). Including developmental time series for expression profiling in future studies can help to identify whether the shared DEGs with opposing expressions also show bimodality in their expression.

These findings demonstrate that similar sets of genes are involved in mid-line patterning and growth across evoulutionary distant vertebrates. Thus, functional studies investigating their specific role in divergent morphogenesis of snout structures in fish can provide valuable

information about the conserved molecular mechanisms underlying the formation of facial soft tissues (Powder & Albertson 2016). Moreover, future studies with developmental time series, hisotological analyses, species crossing (particulalry for the species of Lake Malawi) and female *O. nasuta* are required to explore undelying mechanims of potential convergent evolution, and to tease apart genes involved in snout development from those that only play a role in exaggeration of the snout.

Conducting GO enrichment analysis on the list of upregulated DE genes also revealed the involvement of several biological processes such as proline metabolism, 'tendon development', as well as Notch and Wnt signalling pathways (although Wnt signalling not significantly). Interestingly, a defective proline and hydroxyproline metabolisms has been already associated with a range of skin and facial deformities including abnormal nose morphogenesis in humans (Kiratli & Satilmiş 1998; Kretz et al. 2011; Zaki et al. 2016; Baumgartner et al. 2016). A defective proline metabolism is known to severely affect collagen formation and extracellular matrix integrity, and subsequently cell adhesion (Velez et al. 2019; Karna et al. 2020; Xinjie et al. 2001; Javitt et al. 2019; Noguchi et al. 2020). We found genes involved in 'peptidyl-proline hydroxylation' enriched in the upregulated genes as well as 'proline catabolic process' in the enrichment analysis of downregulated genes. In addition, it has been recently shown that the biosynthesis of proline is tightly regulated by transforming growth factor-beta (Tgfβ) (Schwörer et al. 2020), a TF that also plays role in the early development of the flapped snout in cichlids (Conith et al. 2018). We also found components of this pathway (e.g., *tgfbr3*) to be differentially expressed, and both of the enriched pathways, Wnt and Notch, have conserved crosstalk with Tgfβ signal in regulation of various molecular, cellular and developmental events (Attisano & Labbé 2004; Chesnutt et al. 2004; Arnold et al. 2019; Klüppel & Wrana 2005; Ahi 2016). In addition, both Wnt and Notch signalling pathways are known to play a pivotal role in craniofacial development and morphogenesis including the

formation of middle structures including nasal structures (Penton et al. 2012; Pakvasa et al. 2020; Brugmann et al. 2007; Wang et al. 2011; Singh et al. 2021).

The induction of Tgfβ signalling is required for the establishment of cell-cell contacts in different tissues, whereas later induction of Notch signal stabilizes the Tgfβ mediated effects (Klüppel & Wrana 2005). In the context of the snout, it is possible that activation of Tgfβ is required for early snout induction (Conith et al. 2018) and that continued snout growth is maintained via Notch signalling. This potential time dependent crosstalk may be mediated through the downstream targets of Notch and Tgfβ signals, since it is shown that both signals can regulate similar target genes (Klüppel & Wrana 2005; De Jong et al. 2004), including *foxa2*, a member of the FOX family of transcription factors (both signals suppress *foxa2* expression) (Kondratyeva et al. 2016; Liu et al. 2012). In our study, we found upregulation of *foxa2* in the flapped snout region, and interestingly, a recent study in human shows that a deletion in *Foxa2* can cause a variety of nasal deformities (Dines et al. 2019). Moreover, we found *rbpjb*, a major transcription factor mediating canonical Notch signal (Tanigaki et al. 2002), to be differentially expressed in the flapped snout of both species. In mice, *Rbpj* is shown to regulate a receptor of Tgfβ signal (*Tgfbr1*), thus making a reciprocal positive regulatory loop between the two pathways (Valdez et al. 2012). We also found another receptor of Tgfβ signal (*tgfbr3*) to show a similar expression pattern as *rbpjb* raising the possibility of the existence of such a reciprocal regulatory loop in flapped snout cichlids. In human, a deletion in *RBPJ* gene has been linked to abnormal thickening of the nose and lip (Nakayama et al. 2014). On the other hand, Bmp2 signal which is regarded as another molecular cross point between Tgfβ and Notch pathways (De Jong et al. 2004), mediates its signal through *Tgfbr3* (Hill et al. 2012). Previous studies in cichlids had proposed variations in Bmp expression as a molecular player in adaptive morphological divergence in different skeletal structures (Gunter et al. 2013; Albertson et al. 2005; Ahi et al. 2017; Hulsey et al. 2016). We found downregulation of *bmp2b* expression

suggesting that a key regulator linking both pathways is affected in the flapped snout region. Furthermore, deletion of *Bmp2* in human has been reported to cause a range of nose and lip deformities (Tan et al. 2017). Taken together, these findings suggest complex interactions between Notch and Tgfβ signals in the formation and possibly the maintenance of the flapped snout structure in cichlids.

Finally, we also conducted enrichment for TF binding sites on regulatory sequences of DEGs and found several potential binding sites for TFs that may play a role in the formation of a flapped snout. The most represented TF binding sites belonged to members of FOX transcription factor family, e.g. *foxd3*, *foxj1* and *foxp1,* as well as a consensus binding site for the FOX family. In East African cichlids, both *foxd3* and *foxp1* were predicted to regulate a gene network involved in exaggerated fin elongation (Pashay Ahi & Sefc 2018; Ahi, Richter, et al. 2019). Additionally, *foxp1* was recently suggested as an upstream regulator of genes involved in the formation of the hypertrophic lip in another East African cichlid species (Lecaudey et al. 2021). None of the predicted FOX members (*foxd3*, *foxj1* and *foxp1*) displayed consistent differential expression across both comparisons. It is, therefore, possible that the two other FOX members identified by RNA-seq and qPCR, *foxf1* and *foxa2*, are the key regulators of the entire list of DEGs, since they might bind to the predicted consensus FOX binding site. In addition, both *foxf1* and *foxa2* displayed consistently increased expression in the flapped snout in both comparisons, and are also implicated in the nose morphogenesis in mammals (Dines et al. 2019; Kucharczyk et al. 2014). We have recently found *foxf1* among the differentially expressed genes in hypertrophied lips of an East African cichlid species as well (Lecaudey et al. 2021), suggesting a potential role of *foxf1* in soft tissues exaggeration in cichlids.

Among the other predicted TF binding site we found overrepresentation of binding motif for *tcf12*, a transcription factor with known roles in development and morphogenesis of the

frontal bone and cranial vault thickening in mammals (Piard et al. 2015; Sharma et al. 2013). Moreover, we have previously identified *tcf12* as a potential key player in the formation of a nuchal hump in an East African cichlid (Lecaudey et al. 2019). In this study, we did not detect its differential expression in the snout region. However, there might be other types of potential variations in these TFs (for example alternative splicing (Singh & Ahi 2022)), which are not necessarily reflected in their overall expression differences, but still lead to changes in their regulatory effects. Interestingly, mutations causing missense, frame shift and splicing changes are already reported for *tcf12*, which could lead to craniofacial deformities in humans (Sharma et al. 2013).

The only predicted TF with consistent expression difference in both comparisons was *ap4* (or *tfap4*), i.e. showing slight but significant reduced expression in the flapped snout. *ap4* encodes a member of the basic helix-loop-helix-zipper (bHLH-ZIP) family, and can act as a transcriptional activator or repressor on a variety of downstream target genes mediating cell fate decisions (Wong et al. 2021). We also found both up- and down-regulated genes among the predicted downstream targest of *ap4* (i.e. 93 genes contained ap4 binding site), which confirms its potential transcriptional activating or repressing roles. The exact role of *ap4* in craniofacial morphogenesis of soft tissues is unclear and although deletions in a genomic region containing this gene appeared to cause facial dysmorphisms in humans such as prominent beaked nose and micrognathia (Gervasini et al. 2007), these phenotypes are mainly thought to be linked to mutations in a neighboring gene (*CBP* or *CREBBP*) in this region. Future functional studies are required to verify the potential role of *ap4* in formation and morphogenesis of craniofacial soft tissues in fish.

Conclusions

The snout flap in *Labeotropheus trewavasae* and *Ophthalmotilapia nasuta* is a striking and rare example of an exaggerated soft tissue trait that has evolved recurrently in the cichlid radiations of Lake Malawi and Lake Tanganyika, albeit for different functions. Comparing the transcriptional landscape of the snout flap tissue of these two species with the snout of close relatives that do not develop such a structure, we identified 201 genes that were repeatedly recruited to give rise to the snout flap phenotype even after > 9 MYA of divergence. Our study provides support for a change in proline hydroxylation, a mechanism also linked to human facial deformations, to be a mechanism for metabolic changes involved in the formation of the snout flap in fish. Additionally, we found indications of complex interactions between the transforming growth factor-beta (Tgfβ), regulating the biosynthesis of proline, and Notch signalling, associated with morphogenesis and craniofacial development, in the formation and maintenance of the snout flap. Upstream of the differentially expressed genes we identified transcription factors belonging to the FOX family (especially *foxf1* and *foxa2*) which are both linked to the morphogenesis of the nose in mammals and *ap4* a transcription factor that showed reduced expression in the species with snout flap, but with an unknown role in craniofacial soft tissue formation. We want to emphasise that the identification of genes involved in snout morphogenesis in fish can shed light on the conserved molecular mechanisms crucial for the development and shaping of facial soft tissue. In the future it would be important to build on these findings and confirm the reuse of these genes and pathways across more distant teleost groups.

Methods

Fish rearing and tissue sampling

Five captive bred males of each *O. nasuta, O. ventralis*, and five captive bred females of *L. trewavasae* and *T. tropheops* were raised and kept in a large tank (approximately 450 litres)

containing multiple stony shelters to decrease competition stress. All specimens were at the young adult stage and have been fed with the same diet, Tropical multi-ingredient flakes suitable for omnivorous cichlids. The two species in each comparison were sampled at the same time when the protrusion of the flapped snout had already appeared (Figure 1). To perform the dissections, we used a solution with 0.3 g MS222 per 1L water to euthanize the fish, and similar snout regions, an area above the upper lip encompassing the nostrils which includes epidermis, dermis and the underlying soft connective tissues, were sampled for each fish (Figure 1). The sampled snout tissues for each individual were placed into separate tubes containing RNAlater (Qiagen) and stored at -20 C°. The sacrificing of fish followed the guidelines of the Federal Ministry of Science, Research and Economy of Austria according to the regulations of the BMWFW.

RNA extraction and cDNA synthesis

Total RNA was extracted from 20 dissected snout tissue samples (5 biological replicates per species) following the TRIzol method (Thermo Fischer Scientific). Each dissected sample included epidermis, dermis, and the underlying fibrous/connective tissues of the specified nose regions (Figure 1). Tissue samples were placed into tubes containing 1 ml of TRIzol with a ceramic bead (1.4mm) and homogenized using a FastPrep-24 Instrument (MP Biomedicals, CA, USA). RNA extraction followed the protocol of TRIzol RNA extraction from Thermo Fischer Scientific. A DNA removal step with DNase followed the extraction (invitrogen). The total RNAs were dissolved in 50 µl nuclease-free water and their concentrations were quantified through a Nanophotometer (IMPLEN GmbH, Munich, Germany). We measured the quality of RNAs with the R6K ScreenTape System using an Agilent 2200 TapeStation (Agilent Technologies) and RNA integrity numbers (RIN) above 7 were aimed at for all samples. To synthesize cDNA for qPCR analysis, we used 500 ng of the total RNA per sample and followed

the manufacturer's protocol of the High Capacity cDNA Reverse Transcription kit (Applied Biosystems), and the resulted cDNAs were diluted 1:4 to be used for the qPCR reaction.

RNA-seq library preparation and gene expression quantification

To attain transcriptome data of the snout tissues, we conducted RNA-seq library preparation with 1000 ng of total RNA per tissue sample as input and following the protocol of the Standard TruSeq Stranded mRNA Sample Prep Kit (Illumina) with indexing adapters. The library qualities were assessed using D1000 ScreenTape and reagents (Agilent) on a TapeStation 2200 machine (Agilent). In order to reach an optimal quantity recommended for sequencing, we diluted the libraries and pooled them with equal molar concentration for each library. The RNA-sequencing was conducted in the NGS Facility at Vienna Biocenter Core Facilities (VBCF, Austria) on an Illumina HiSeq2500 and generated between 6.7 and 15.8 million paired-end reads with 125bp length per sample (Supplementary Table S1). Raw reads were de-multiplexed based on unique barcodes by the same facility. The quality of the reads was assessed with FastQC (v0.11.8) (Andrews 2012), and reads were filtered for a quality > 28 and a minimum length of 70 bp with Trimmomatic (v0.3.9) (Bolger et al. 2014). Reads were aligned to the *O. niloticus* reference genome (Conte et al. 2017) of the University of Maryland using RNAstar (v2.7.3.a) (Dobin et al. 2013). To check the mapping statistics, we used samtools idxstats (v1.9) (Danecek et al. 2021) (Supplementary Table S1) and further merged the single files for species and Lake with picard (v2.21.7) (Picard Toolkit. 2019. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/). We used StringTie (v.2.0.6) (Pertea et al. 2015) to assemble the alignments into potential transcripts without a reference. This step was conducted separately for single files (per biological replicate) and the merged files (per species and per Lake). The single files per biological replicate were further merged into species. This process of repeated merging steps was implemented to reduce the probability of false positives.

To assess the accuracy of the mapping we used gffcompare (v0.11.2) (Pertea & Pertea 2020) to compare our annotations to the reference annotation. Subsequently we filtered for monoexonic transcripts that were not contained in our reference and the transcripts assigned the class code 'possible polymerase run-on' by gffcompare. As the maximum intron length of the *O. niloticus* reference is 200000 bp, we also filtered for that in the produced annotation. The expression estimates for each transcript were based on these annotations and generated with StringTie (v.2.0.6) with no multimapping allowed. The final raw count matrices were produced from the expression estimates with a Perl script from the griffith lab (https://github.com/griffithlab/rnaseq_tutorial/blob/master/scripts/stringtie_expression_matrix.pl) and the code used in this analysis is available at this github repository (https://github.com/annaduenser/snout_flap_RNAseq).

Differential expression analysis was conducted using DESeq2 (Love et al. 2014) in R (R Core Team 2017) running comparisons for each Lake separately. DESeq2 estimates variance-mean dependence based on a model using negative binomial distribution using the raw counts (Love et al. 2014). A FDR of $P < 0.05$ was chosen as the cut-off for the adjusted *P*-value to determine differentially expressed genes.

For the downstream analysis, an enrichment step for gene ontology (GO) terms of biological processes was conducted in R using topGO (v2.48.0) (Alexa & Rahnenfuhrer 2019) with the method *weight* and using Fisher's exact tests for the enrichment analysis, while GO terms for *Oreochromis niloticus* were aquired via the biomaRt package (v2.46.1) (Durinck et al. 2005, 2009) from the Ensemble database. To further group functionally similar GO terms, we also used REVIGO (Supek et al. 2011) using simRel scores (Schlicker et al. 2006). To predict the potential upstream regulators of DE genes, we conducted motif overrepresentation analysis on 4 kb upstream sequences (promoter and 5′-UTR) of these genes using the annotated genome of

the Nile tilapia (Zerbino et al. 2018) and MEME tool (Bailey et al. 2009). The motifs that were present in the promoters of at least one fifth of the total 201 DEGs were compared to position weight matrices (PWMs) in the TRANSFAC database (Matys et al. 2003) via STAMP (Mahony & Benos 2007) in order to identify matching TF binding sites. In addition, we investigated the functional interactions between the products of DE genes through STRING v10 (http://string-db.org/), a knowledge based interaction prediction tool, and zebrafish databases for protein interactomes (Szklarczyk et al. 2017).

Primer design and qPCR

We designed the qPCR primers on conserved regions of the selected candidate genes by aligning their assembled sequences to their already available homologous mRNA sequences from *Ophthalmotilapia ventralis* (Böhne et al. 2014), *Metriaclima zebra*, *Pundamilia nyererei*, *Neolamprologus brichardi* and *Astatotilapia burtoni* (Brawand et al. 2014), as well as *Oreochromis niloticus*. After aligning the conserved sequence regions across the abovementioned East African cichlids, we identified the exon junctions (using CLC Genomic Workbench, CLC Bio, Denmark, and annotated genome of *Astatotilapia burtoni* in the Ensembl database, http://www.ensembl.org). The primer designing steps were conducted as described previously (Pashay Ahi & Sefc 2018; Ahi, Richter, et al. 2019) using Primer Express 3.0 (Applied Biosystems, CA, USA) (Supplementary Table S5). The qPCR was performed based on the protocol provided by Maxima SYBR Green/ROX qPCR Master Mix (2X) (Thermo Fisher Scientific, Germany) following the guidelines for optimal experimental set-up for each qPCR run (Hellemans et al. 2007). The qPCR program was set for 2 min at 50°C, 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/slope] (Supplementary Table S5).

In order to select stably expressed candidate reference genes, we filtered for genes with a low log2 fold change and subsequently ranked the remaining genes according to low coefficient of variation. The top six most stable genes shared across the transcriptome comparisons were selected as candidate reference genes (Supplementary Table S4). After qPCR expression analysis of the six genes across all samples, we ranked them based on their expression stability by three different algorithms: BestKeeper (Pfaffl et al. 2004), NormFinder (Andersen et al. 2004) and geNorm (Vandesompele et al. 2002). We used the Cq values of the top most stable reference genes to normalize Cq values of target genes in each sample (ΔCq target = Cq target – Cq reference). The relative expression levels (RQ) were calculated by $2^{-\Delta\Delta Cq}$ method (Pfaffl 2001) and the log-transformed RQ values were used for independent-samples t-tests to calculate the statistical differences.

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, CA and CS conceived the project. WG contributed to fish husbandry and photography, and EPA and AD conducted the sampling and tissue dissection. AD, EPA and LL conducted the RNA lab work. AD, PS, LL, EPA contributed to the analyses and all authors to manuscript writing. CS and EPA contributed to funding. This work was supported by the Austrian Science

Fund [project number P29838] awarded to CS. All authors approved the final version of the manuscript.

Competing financial interests

The authors declare no competing interests.

Ethical approval

No experiments were conducted on the fish prior to sampling, so an ethics approval is not required according to the Austrian animal welfare law. Fish keeping and sacrifice was carried out in our certified aquarium facility in accordance with the Austrian animal welfare law.

Data availability

The data underlying this article are available in the Sequencing Read Archive (SRA) of NCBI at https://www.ncbi.nlm.nih.gov/ and can be accessed with PRJNA770252.

References

Ahi E. 2016. Signalling pathways in trophic skeletal development and morphogenesis: Insights from studies on teleost fish. Dev. Biol. 420:11–31. doi: http://dx.doi.org/10.1016/j.ydbio.2016.10.003.

Ahi EP, Duenser A, Singh P, Gessl W, Sturmbauer C. 2020. Appetite regulating genes may contribute to herbivory versus carnivory trophic divergence in haplochromine cichlids. PeerJ. 8:e8375. doi: 10.7717/peerj.8375.

Ahi EP et al. 2020. Expression levels of the tetratricopeptide repeat protein gene ttc39b covary with carotenoid-based skin colour in cichlid fish. Biol. Lett. 16:20200629. doi: 10.1098/rsbl.2020.0629.

Ahi EP, Richter F, Lecaudey LA, Sefc KM. 2019. Gene expression profiling suggests differences in molecular mechanisms of fin elongation between cichlid species. Sci. Rep. 9:9052. doi: 10.1038/s41598-019-45599-w.

Ahi EP, Richter F, Sefc KM. 2017. A gene expression study of ornamental fin shape in Neolamprologus brichardi, an African cichlid species. Sci. Rep. 7:17398. doi: 10.1038/s41598-017-17778-0.

Ahi EP, Singh P, Duenser A, Gessl W, Sturmbauer C. 2019. Divergence in larval jaw gene

expression reflects differential trophic adaptation in haplochromine cichlids prior to foraging. BMC Evol. Biol. 19:150. doi: 10.1186/s12862-019-1483-3.

Alazami AM et al. 2016. Expanding the clinical and genetic heterogeneity of hereditary disorders of connective tissue. Hum. Genet. 135:525–540. doi: 10.1007/s00439-016-1660-z.

Albertson RC, Streelman JT, Kocher TD, Yelick PC. 2005. Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies. Proc. Natl. Acad. Sci. U. S. A. 102:16287–92. doi: 10.1073/pnas.0506649102.

Alexa A, Rahnenfuhrer J. 2019. topGO: Enrichment Analysis for Gene Ontology version 2.42.0 from Bioconductor. R Packag. https://rdrr.io/bioc/topGO/ (Accessed August 31, 2022).

Alhamoudi KM et al. 2020. A Missense Mutation in the UGDH Gene Is Associated With Developmental Delay and Axial Hypotonia. Front. Pediatr. 8. doi: 10.3389/fped.2020.00071.

Alwadei AH et al. 2016. Loss-of-function mutation in *RUSC2* causes intellectual disability and secondary microcephaly. Dev. Med. Child Neurol. 58:1317–1322. doi: 10.1111/dmcn.13250.

Andersen CL, Jensen JL, Ørntoft TF. 2004. 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. 64:5245–50. doi: 10.1158/0008-5472.CAN-04-0496.

Andrews S. 2012. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Arnold CP, Benham-Pyle BW, Lange JJ, Wood CJ, Sánchez Alvarado A. 2019. Wnt and TGFβ coordinate growth and patterning to regulate size-dependent behaviour. Nature. 572:655–659. doi: 10.1038/s41586-019-1478-7.

Attisano L, Labbé E. 2004. TGFβ and Wnt pathway cross-talk. Cancer Metastasis Rev. 23:53–61. doi: 10.1023/A:1025811012690.

Bailey TL et al. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37:W202-8. doi: 10.1093/nar/gkp335.

Barrionuevo MG, Aybar MJ, Tríbulo C. 2014. Two different vestigial like 4 genes are differentially expressed during Xenopus laevis development. Int. J. Dev. Biol. 58:369–377. doi: 10.1387/ijdb.130353ct.

Baumgarten L, Machado-Schiaffino G, Henning F, Meyer A. 2015. What big lips are good for: On the adaptive function of repeatedly evolved hypertrophied lips of cichlid fishes. Biol. J. Linn. Soc. 115:448–455. doi: 10.1111/bij.12502.

Baumgartner MR, Valle D, Dionisi-Vici C. 2016. Disorders of Ornithine and Proline Metabolism. In: Inborn Metabolic Diseases. Springer Berlin Heidelberg pp. 321–331. doi: 10.1007/978-3-662-49771-5_21.

Böhne A, Sengstag T, Salzburger W. 2014. Comparative transcriptomics in East African cichlids reveals sex- and species-specific expression and new candidates for sex differentiation in fishes. Genome Biol. Evol. 6:2567–2585. doi: 10.1093/gbe/evu200.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120. doi: 10.1093/bioinformatics/btu170.

Bonaglia MC et al. 2005. A 2.3 Mb duplication of chromosome 8q24.3 associated with

severe mental retardation and epilepsy detected by standard karyotype. Eur. J. Hum. Genet. 13:586–591. doi: 10.1038/sj.ejhg.5201369.

Brawand D et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature. 513:375–381. doi: 10.1038/nature13726.

Brugmann SA et al. 2007. Wnt signaling mediates regional specification in the vertebrate face. Development. 134:3283–95. doi: 10.1242/dev.005132.

Brugmann SA, Cordero DR, Helms JA. 2010. Craniofacial ciliopathies: A new classification for craniofacial disorders. Am. J. Med. Genet. Part A. 152A:2995–3006. doi: 10.1002/ajmg.a.33727.

Chen CP, Ko TM, Chen YY, Su JW, Wang W. 2013. Prenatal diagnosis and molecular cytogenetic characterization of mosaicism for a small supernumerary marker chromosome derived from chromosome 22 associated with cat eye syndrome. Gene. 527:384–388. doi: 10.1016/j.gene.2013.05.061.

Chesnutt C, Burrus LW, Brown AMC, Niswander L. 2004. Coordinate regulation of neural tube patterning and proliferation by TGFβ and WNT activity. Dev. Biol. 274:334–347. doi: 10.1016/j.ydbio.2004.07.019.

Colombo M et al. 2013. The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. Mol. Ecol. 22:670–684. doi: 10.1111/mec.12029.

Concannon MR, Albertson RC. 2015. The genetic and developmental basis of an exaggerated craniofacial trait in East African cichlids. J. Exp. Zool. B. Mol. Dev. Evol. 324:662–70. doi: 10.1002/jez.b.22641.

Conith MR et al. 2018. Genetic and developmental origins of a unique foraging adaptation in a Lake Malawi cichlid genus. Proc. Natl. Acad. Sci. 115:7063–7068. doi: 10.1073/PNAS.1719798115.

Conith MR, Conith AJ, Albertson RC. 2019. Evolution of a soft-tissue foraging adaptation in African cichlids: Roles for novelty, convergence, and constraint. Evolution (N. Y). 73:2072–2084. doi: 10.1111/evo.13824.

Conte MA, Gammerdinger WJ, Bartie KL, Penman DJ, Kocher TD. 2017. A high quality assembly of the Nile Tilapia (Oreochromis niloticus) genome reveals the structure of two sex determination regions. BMC Genomics. 18:341. doi: 10.1186/s12864-017-3723-5.

Czeschik JC et al. 2014. A patient with a de-novo deletion 3p25.3 and features overlapping with Rubinstein–Taybi syndrome. Clin. Dysmorphol. 23:67–70. doi: 10.1097/MCD.0000000000000035.

Danecek P et al. 2021. Twelve years of SAMtools and BCFtools. Gigascience. 10. doi: 10.1093/gigascience/giab008.

Dell'Edera D et al. 2018. 16p11.2 microdeletion syndrome: A case report. J. Med. Case Rep. 12:1–6. doi: 10.1186/s13256-018-1587-1.

Dines JN et al. 2019. Expanding phenotype with severe midline brain anomalies and missense variant supports a causal role for *FOXA2* in 20p11.2 deletion syndrome. Am. J. Med. Genet. Part A. 179:ajmg.a.61281. doi: 10.1002/ajmg.a.61281.

Dobin A et al. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29:15–21. doi: 10.1093/bioinformatics/bts635.

Durinck S et al. 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 21:3439–3440. doi: 10.1093/bioinformatics/bti525.

Durinck S, Spellman PT, Birney E, Huber W. 2009. Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. Nat. Protoc. 4:1184–1191. doi: 10.1038/nprot.2009.97.

Ehret JK et al. 2015. 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. 8:1–14. doi: 10.1186/s13039-015-0178-8.

Fryssira H et al. 2011. Severe Developmental Delay in a Patient with 7p21.1–p14.3 Microdeletion Spanning the <i>TWIST</i> Gene and the <i>HOXA</i> Gene Cluster. Mol. Syndromol. 2:45–49. doi: 10.1159/000334313.

Gervasini C et al. 2007. High frequency of mosaic CREBBP deletions in Rubinstein-Taybi syndrome patients and mapping of somatic and germ-line breakpoints. Genomics. 90:567–573. doi: 10.1016/j.ygeno.2007.07.012.

Giannikou K et al. 2012. Further delineation of novel 1p36 rearrangements by array-CGH analysis: Narrowing the breakpoints and clarifying the 'extended' phenotype. Gene. 506:360–368. doi: 10.1016/j.gene.2012.06.060.

Guilmatre A et al. 2010. Type I hyperprolinemia: genotype/phenotype correlations. Hum. Mutat. 31:961–965. doi: 10.1002/humu.21296.

Gunter HM et al. 2013. Shaping development through mechanical strain: the transcriptional basis of diet-induced phenotypic plasticity in a cichlid fish. Mol. Ecol. 22:4516–4531. doi: 10.1111/mec.12417.

Haldeman-Englert CR et al. 2009. A 3.1-Mb microdeletion of 3p21.31 associated with cortical blindness, cleft lip, CNS abnormalities, and developmental delay. Eur. J. Med. Genet. 52:265–268. doi: 10.1016/j.ejmg.2008.11.005.

Hanssens M, Snoeks J, Verheyen E. 1999. A morphometric revision of the genus Ophthalmotilapia (Teleostei, Cichlidae) from Lake Tanganyika (East Africa). Zool. J. Linn. Soc. 125:487–512. doi: 10.1111/j.1096-3642.1999.tb00602.x.

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. 8:R19. doi: 10.1186/gb-2007-8-2-r19.

Henning F, Machado-Schiaffino G, Baumgarten L, Meyer A. 2017. Genetic dissection of adaptive form and function in rapidly speciating cichlid fishes. Evolution (N. Y). 71:1297–1312. doi: 10.1111/evo.13206.

Hill CR et al. 2012. BMP2 signals loss of epithelial character in epicardial cells but requires the Type III TGFβ receptor to promote invasion. Cell. Signal. 24:1012–1022. doi: 10.1016/j.cellsig.2011.12.022.

Holder JL, Lotze TE, Bacino C, Cheung S-W. 2012. A child with an inherited 0.31 Mb microdeletion of chromosome 14q32.33: Further delineation of a critical region for the 14q32 deletion syndrome. Am. J. Med. Genet. Part A. 158A:1962–1966. doi: 10.1002/ajmg.a.35289.

Hosono K et al. 2020. A case of childhood glaucoma with a combined partial monosomy

6p25 and partial trisomy 18p11 due to an unbalanced translocation. Ophthalmic Genet. 41:175–182. doi: 10.1080/13816810.2020.1744019.

Hufnagel RB et al. 2016. A new frontonasal dysplasia syndrome associated with deletion of the *SIX2* gene. Am. J. Med. Genet. Part A. 170:487–491. doi: 10.1002/ajmg.a.37441.

Hulsey CD, Fraser GJ, Meyer A. 2016. Biting into the genome to phenome map: Developmental genetic modularity of cichlid fish dentitions. In: Integrative and Comparative Biology.Vol. 56 Oxford University Press pp. 373–388. doi: 10.1093/icb/icw059.

Hüning I et al. 2013. Exon 2 duplication of the MID1 gene in a patient with a mild phenotype of Opitz G/BBB syndrome. Eur. J. Med. Genet. 56:188–191. doi: 10.1016/j.ejmg.2013.01.004.

Irisarri I et al. 2018. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. Nat. Commun. 9:1–12. doi: 10.1038/s41467-018-05479-9.

Jain S, Yang P, Farrell SA. 2010. A case of 8q22.1 microdeletion without the Nablus mask-like facial syndrome phenotype. Eur. J. Med. Genet. 53:108–110. doi: 10.1016/j.ejmg.2009.12.006.

Javitt G et al. 2019. *cis* -Proline mutants of quiescin sulfhydryl oxidase 1 with altered redox properties undermine extracellular matrix integrity and cell adhesion in fibroblast cultures. Protein Sci. 28:228–238. doi: 10.1002/pro.3537.

De Jong DS et al. 2004. Regulation of Notch signaling genes during BMP2-induced differentiation of osteoblast precursor cells. Biochem. Biophys. Res. Commun. 320:100–107. doi: 10.1016/j.bbrc.2004.05.150.

Kariminejad A et al. 2015. Intellectual disability, muscle weakness and characteristic face in three siblings: A newly described recessive syndrome mapping to 3p24.3-p25.3. Am. J. Med. Genet. Part A. 167:2508–2515. doi: 10.1002/ajmg.a.37248.

Karna E, Szoka L, Huynh TYL, Palka JA. 2020. Proline-dependent regulation of collagen metabolism. Cell. Mol. Life Sci. 77:1911–1918. doi: 10.1007/s00018-019-03363-3.

Kiratli H, Satilmiş M. 1998. Prolidase deficiency associated with pathologic myopia. Ophthalmic Genet. 19:49–53. doi: 10.1076/opge.19.1.49.2180.

Klüppel M, Wrana JL. 2005. Turning it up a Notch: cross-talk between TGFβ and Notch signaling. BioEssays. 27:115–118. doi: 10.1002/bies.20187.

Kocher TD, Conroy JA, McKaye KR, Stauffer JR. 1993. Similar morphologies of cichlid fish in lakes tanganyika and malawi are due to convergence. Mol. Phylogenet. Evol. 2:158–165. doi: 10.1006/mpev.1993.1016.

Kondratyeva LG et al. 2016. Downregulation of expression of mater genes SOX9, FOXA2, and GATA4 in pancreatic cancer cells stimulated with TGFβ1 epithelial–mesenchymal transition. Dokl. Biochem. Biophys. 469:257–259. doi: 10.1134/S1607672916040062.

Konings A. 2007. *Malawi cichlids in their natural habitat 4th Edition*. 4th ed. Cichlid Press. El Paso, Texas, USA.: El Paso, Texas, USA.

Kosho T et al. 2008. De-novo balanced translocation between 7q31 and 10p14 in a girl with central precocious puberty, moderate mental retardation, and severe speech impairment. Clin. Dysmorphol. 17:31–34. doi: 10.1097/MCD.0b013e3282f17688.

Kretz R et al. 2011. Defect in proline synthesis: Pyrroline-5-carboxylate reductase 1 deficiency leads to a complex clinical phenotype with collagen and elastin abnormalities. J. Inherit. Metab. Dis. 34:731–739. doi: 10.1007/s10545-011-9319-3.

Kubista M et al. 2006. The real-time polymerase chain reaction. Mol. Aspects Med. 27:95–125. doi: 10.1016/j.mam.2005.12.007.

Kucharczyk M et al. 2014. The first case of a patient with de novo partial distal 16q tetrasomy and a data's review. Am. J. Med. Genet. Part A. 164:2541–2550. doi: 10.1002/ajmg.a.36686.

Lecaudey LA et al. 2021. Transcriptomics unravels molecular players shaping dorsal lip hypertrophy in the vacuum cleaner cichlid, Gnathochromis permaxillaris. BMC Genomics. 22:506. doi: 10.1186/s12864-021-07775-z.

Lecaudey LA, Sturmbauer C, Singh P, Ahi EP. 2019. Molecular mechanisms underlying nuchal hump formation in dolphin cichlid, Cyrtocara moorii. Sci. Rep. doi: 10.1038/s41598-019-56771-7.

Liu M et al. 2012. IKKα Activation of NOTCH Links Tumorigenesis via FOXA2 Suppression. Mol. Cell. 45:171–184. doi: 10.1016/j.molcel.2011.11.018.

Lopes F et al. 2019. Genomic imbalances defining novel intellectual disability associated loci. Orphanet J. Rare Dis. 14:1–13. doi: 10.1186/s13023-019-1135-0.

Losos JB, Jackman TR, Larson A, De Queiroz K, Rodríguez-Schettino L. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. Science (80-.). 279:2115–2118. doi: 10.1126/science.279.5359.2115.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550. doi: 10.1186/s13059-014-0550-8.

Machado-Schiaffino G, Henning F, Meyer A. 2014. Species-specific differences in adaptive phenotypic plasticity in an ecologically relevant trophic trait: hypertrophic lips in Midas cichlid fishes. Evolution (N. Y). 68:2086–2091. doi: 10.1111/evo.12367.

Mahony S, Benos P V. 2007. STAMP: a web tool for exploring DNA-binding motif similarities. Nucleic Acids Res. 35:W253-8. doi: 10.1093/nar/gkm272.

Malinsky M et al. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat. Ecol. Evol. 2:1940–1955. doi: 10.1038/s41559-018-0717-x.

Manousaki T et al. 2013. Parsing parallel evolution: ecological divergence and differential gene expression in the adaptive radiations of thick-lipped Midas cichlid fishes from Nicaragua. Mol. Ecol. 22:650–69. doi: 10.1111/mec.12034.

Martinez-Glez V et al. 2007. Clinical presentation of a variant of Axenfeld-Rieger syndrome associated with subtelomeric 6p deletion. Eur. J. Med. Genet. 50:120–127. doi: 10.1016/j.ejmg.2006.10.005.

Matys V et al. 2003. TRANSFAC®: Transcriptional regulation, from patterns to profiles. Nucleic Acids Res. 31:374–378. doi: 10.1093/nar/gkg108.

McLachlan E et al. 2005. Functional Characterization of Oculodentodigital Dysplasia-Associated Cx43 Mutants. Cell Commun. Adhes. 12:279–292. doi: 10.1080/15419060500514143.

Miller ND et al. 2009. Molecular (SNP) analyses of overlapping hemizygous deletions of 10q25.3 to 10qter in four patients: Evidence for HMX2 and HMX3 as candidate genes in hearing and vestibular function. Am. J. Med. Genet. Part A. 149A:669–680. doi: 10.1002/ajmg.a.32705.

Mitteldorf C et al. 2018. Deceptively bland cutaneous angiosarcoma on the nose mimicking hemangioma-A clinicopathologic and immunohistochemical analysis. J. Cutan. Pathol. 45:652–658. doi: 10.1111/cup.13275.

Mordaunt D et al. 2015. 8q13.1-q13.2 deletion associated with inferior cerebellar vermian hypoplasia and digital anomalies: A new syndrome? Pediatr. Neurol. 52:230-234.e1. doi: 10.1016/j.pediatrneurol.2014.09.002.

Mucha BE et al. 2019. A new microdeletion syndrome involving TBC1D24, ATP6V0C, and PDPK1 causes epilepsy, microcephaly, and developmental delay. Genet. Med. 21:1058–1064. doi: 10.1038/s41436-018-0290-3.

Nakayama T et al. 2014. RBPJ is disrupted in a case of proximal 4p deletion syndrome with epilepsy. Brain Dev. 36:532–536. doi: 10.1016/j.braindev.2013.07.009.

Nimmagadda S et al. 2015. Identification and functional analysis of novel facial patterning genes in the duplicated beak chicken embryo. Dev. Biol. 407:275–288. doi: 10.1016/j.ydbio.2015.09.007.

Noguchi Y, Iwasaki Y, Ueda M, Kakinoki S. 2020. Surfaces immobilized with oligo-prolines prevent protein adsorption and cell adhesion. J. Mater. Chem. B. 8:2233–2237. doi: 10.1039/d0tb00051e.

Ochab-Marcinek A, Tabaka M. 2010. Bimodal gene expression in noncooperative regulatory systems. Proc. Natl. Acad. Sci. U. S. A. 107:22096–22101. doi: 10.1073/pnas.1008965107.

Okello DO et al. 2017. Six2 Plays an Intrinsic Role in Regulating Proliferation of Mesenchymal Cells in the Developing Palate. Front. Physiol. 8:955. doi: 10.3389/fphys.2017.00955.

Pakvasa M et al. 2020. Notch signaling: Its essential roles in bone and craniofacial development. Genes Dis. doi: 10.1016/j.gendis.2020.04.006.

Pashay Ahi E, Sefc KM. 2018. Towards a gene regulatory network shaping the fins of the Princess cichlid. Sci. Rep. 8:9602. doi: 10.1038/s41598-018-27977-y.

De Pater JM et al. 2005. Striking Facial Dysmorphisms and Restricted Thymic Development in a Fetus with a 6-Megabase Deletion of Chromosome 14q. Pediatr. Dev. Pathol. 8:497–503. doi: 10.1007/s10024-005-0041-8.

Penton AL, Leonard LD, Spinner NB. 2012. Notch signaling in human development and disease. Semin. Cell Dev. Biol. 23:450–457. doi: 10.1016/j.semcdb.2012.01.010.

Pertea G, Pertea M. 2020. GFF Utilities: GffRead and GffCompare. F1000Research. 9. doi: 10.12688/f1000research.23297.2.

Pertea M et al. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33:290–295. doi: 10.1038/nbt.3122.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29:e45.

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. 2004. Determination of stable

housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. Biotechnol. Lett. 26:509–15.

Piard J et al. 2015. *TCF12* microdeletion in a 72-year-old woman with intellectual disability. Am. J. Med. Genet. Part A. 167:1897–1901. doi: 10.1002/ajmg.a.37083.

Pillai NR et al. 2019. Novel deletion of 6p21.31p21.1 associated with laryngeal cleft, developmental delay, dysmorphic features and vascular anomaly. Eur. J. Med. Genet. 62:103531. doi: 10.1016/j.ejmg.2018.08.012.

Powder KE, Albertson RC. 2016. Cichlid fishes as a model to understand normal and clinical craniofacial variation. Dev. Biol. 415:338–346. doi: 10.1016/J.YDBIO.2015.12.018.

Preiksaitiene E et al. 2015. R368X mutation in MID1 among recurrent mutations in patients with X-linked Opitz G/BBB syndrome. Clin. Dysmorphol. 24:7–12. doi: 10.1097/MCD.0000000000000059.

Quigley DI, Kaiser-Rogers K, Aylsworth AS, Rao KW. 2004. Submicroscopic deletion 9(q34.4) and duplication 19(p13.3): Identified by subtelomere specific FISH probes. Am. J. Med. Genet. 125A:67–72. doi: 10.1002/ajmg.a.20457.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. doi: 10.1007/978-3-540-74686-7.

Reijnders MRF et al. 2017. RAC1 Missense Mutations in Developmental Disorders with Diverse Phenotypes. Am. J. Hum. Genet. 101:466–477. doi: 10.1016/j.ajhg.2017.08.007.

Rivera-Pedroza CI et al. 2017. Chromosome 1p31.1p31.3 Deletion in a Patient with Craniosynostosis, Central Nervous System and Renal Malformation: Case Report and Review of the Literature. Mol. Syndromol. 8:30–35. doi: 10.1159/000452609.

Rüber L, Verheyen E, Meyer A. 1999. Replicated evolution of trophic specializations in an endemic cichlid fish lineage from Lake Tanganyika. Proc. Natl. Acad. Sci. U. S. A. 96:10230–10235. doi: 10.1073/pnas.96.18.10230.

Rundle HD, Nagel L, Boughman JW, Schluter D. 2000. Natural selection and parallel speciation in sympatric sticklebacks. Science (80-.). 287:306–308. doi: 10.1126/science.287.5451.306.

Sansone A, Syed AS, Tantalaki E, Korsching SI, Manzini I. 2014. Trpc2 is expressed in two olfactory subsystems, the main and the vomeronasal system of larval Xenopus laevis. J. Exp. Biol. 217:2235–2238. doi: 10.1242/jeb.103465.

Sawyer JR, Binz RL, Swanson CM, Lim C. 2007. De novo proximal duplication of 1(q12q22) in a female infant with multiple congenital anomalies. Am. J. Med. Genet. Part A. 143A:338–342. doi: 10.1002/ajmg.a.31604.

Schlicker A, Domingues FS, Rahnenführer J, Lengauer T. 2006. A new measure for functional similarity of gene products based on gene ontology. BMC Bioinformatics. 7:302. doi: 10.1186/1471-2105-7-302.

Schluter D, Nagel LM. 1995. Parallel speciation by natural selection. Am. Nat. 146:292–301. doi: 10.1086/285799.

Schulz S, Volleth M, Muschke P, Wieland I, Wieacker P. 2008. Greig cephalopolysyndactyly (GCPS) contiguous gene syndrome in a boy with a 14 Mb deletion in region 7p13-14 caused by a paternal balanced insertion (5; 7). Appl. Clin. Genet. 1:19–22. doi: 10.2147/tacg.s4401.

Schwörer S et al. 2020. Proline biosynthesis is a vent for TGFβ-induced mitochondrial redox stress. EMBO J. 39. doi: 10.15252/embj.2019103334.

Selenti N et al. 2015. An interstitial deletion at 8q23.1-q24.12 associated with Langer-Giedion syndrome/Trichorhinophalangeal syndrome (TRPS) type II and Cornelia de Lange syndrome 4. Mol. Cytogenet. 8:64. doi: 10.1186/s13039-015-0169-9.

Sharma VP et al. 2013. Mutations in TCF12, encoding a basic helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis. Nat. Genet. 45:304–307. doi: 10.1038/ng.2531.

Singh P, Ahi EP. 2022. The importance of alternative splicing in adaptive evolution. Mol. Ecol. 31:1928–1938. doi: 10.1111/mec.16377.

Singh P, Ahi EP, Sturmbauer C. 2021. Gene coexpression networks reveal molecular interactions underlying cichlid jaw modularity. BMC Ecol. Evol. 21:1–17. doi: 10.1186/s12862-021-01787-9.

Sofos E et al. 2012. A novel familial 11p15.4 microduplication associated with intellectual disability, dysmorphic features, and obesity with involvement of the ZNF214 gene. Am. J. Med. Genet. Part A. 158A:50–58. doi: 10.1002/ajmg.a.34290.

Starkovich M, Lalani SR, Mercer CL, Scott DA. 2016. Chromosome 5q33 deletions associated with congenital heart defects. Am. J. Med. Genet. Part A. 170:3338–3342. doi: 10.1002/ajmg.a.37957.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms Gibas, C, editor. PLoS One. 6:e21800. doi: 10.1371/journal.pone.0021800.

Szklarczyk D et al. 2017. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res. 45:D362–D368. doi: 10.1093/nar/gkw937.

Tan TY et al. 2017. Monoallelic BMP2 Variants Predicted to Result in Haploinsufficiency Cause Craniofacial, Skeletal, and Cardiac Features Overlapping Those of 20p12 Deletions. Am. J. Hum. Genet. 101:985–994. doi: 10.1016/j.ajhg.2017.10.006.

Tanigaki K et al. 2002. Notch-RBP-J signaling is involved in cell fate determination of marginal zone B cells. Nat. Immunol. 3:443–450. doi: 10.1038/ni793.

Tanigaki K, Honjo T. 2010. *Two Opposing roles of RBP-J in Notch signaling*. Academic Press doi: 10.1016/S0070-2153(10)92007-3.

Thomas PS, Kim J, Nunez S, Glogauer M, Kaartinen V. 2010. Neural crest cell-specific deletion of Rac1 results in defective cell-matrix interactions and severe craniofacial and cardiovascular malformations. Dev. Biol. 340:613–625. doi: 10.1016/j.ydbio.2010.02.021.

Tinker RJ et al. 2021. Haploinsufficiency of <scp> *ATP6V0C* </scp> possibly underlies 16p13.3 deletions that cause microcephaly, seizures, and neurodevelopmental disorder. Am. J. Med. Genet. Part A. 185:196–202. doi: 10.1002/ajmg.a.61905.

Tomkins DJ, Gitelman BJ, Roberts MH. 1983. Confirmation of a de novo duplication, dup(10) (q24→q26), by GOT1 gene dosage studies. Hum. Genet. 63:369–373. doi: 10.1007/BF00274764.

Ţuţulan-Cuniţă AC et al. 2012. 3p interstitial deletion: Novel case report and review. J. Child

Neurol. 27:1062–1066. doi: 10.1177/0883073811431016.

Valdez JM et al. 2012. Notch and TGFβ form a reciprocal positive regulatory loop that suppresses murine prostate basal stem/progenitor cell activity. Cell Stem Cell. 11:676–688. doi: 10.1016/j.stem.2012.07.003.

Vandesompele J et al. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3:RESEARCH0034.

Velez DO et al. 2019. 3D collagen architecture regulates cell adhesion through degradability, thereby controlling metabolic and oxidative stress. Integr. Biol. 11:221–234. doi: 10.1093/intbio/zyz019.

Wang Y, Song L, Zhou CJ. 2011. The canonical Wnt/ß-catenin signaling pathway regulates Fgf signaling for early facial development. Dev. Biol. 349:250–260. doi: 10.1016/j.ydbio.2010.11.004.

Wieacker P, Volleth M. 2007. <i>WNT4</i> and <i>RSPO1</i> Are Not Involved in a Case of Male-to-Female Sex Reversal with Partial Duplication of 1p. Sex. Dev. 1:111–113. doi: 10.1159/000100032.

Wong MM-K, Joyson SM, Hermeking H, Chiu SK. 2021. Transcription Factor AP4 Mediates Cell Fate Decisions: To Divide, Age, or Die. Cancers (Basel). 13:676. doi: 10.3390/cancers13040676.

Xinjie L et al. 2001. Evaluation of the role of proline residues flanking the RGD motif of dendroaspin, an inhibitior of platelet aggregation and cell adhesion. Biochem. J. 355:633–638. doi: 10.1042/bj3550633.

Yakut S et al. 2015. A familial interstitial 4q35 deletion with no discernible clinical effects. Am. J. Med. Genet. Part A. 167:1836–1841. doi: 10.1002/ajmg.a.37097.

Zaki MS et al. 2016. *PYCR2* Mutations cause a lethal syndrome of microcephaly and failure to thrive. Ann. Neurol. 80:59–70. doi: 10.1002/ana.24678.

Zerbino DR et al. 2018. Ensembl 2018. Nucleic Acids Res. 46:D754–D761. doi: 10.1093/nar/gkx1098.

Zhang P, Yang C, Delay RJ. 2010. Odors activate dual pathways, a TRPC2 and a AA-dependent pathway, in mouse vomeronasal neurons. Am. J. Physiol. Physiol. 298:C1253–C1264. doi: 10.1152/ajpcell.00271.2009.

Zlotina A et al. 2016. Ring chromosome 18 in combination with 18q12.1 (DTNA) interstitial microdeletion in a patient with multiple congenital defects. Mol. Cytogenet. 9:1–7. doi: 10.1186/s13039-016-0229-9.