Ectopic expression of lipid homeostasis genes in rare cells of the teleost pituitary gland

Eirill Ager-Wick*, Gersende Maugars*, Kristine von Krogh, Romain Fontaine, Khadeeja Siddique, Rasoul Nourizadeh-Lillabadi, Finn-Arne Weltzien & Christiaan Henkel**

Faculty of Veterinary Medicine, Norwegian University of Life Sciences, Ås, Norway

* Contributed equally

** Corresponding author: *christiaan.henkel@nmbu.no*

Abstract

Directing both organismal homeostasis and physiological adaptation, the pituitary is a key endocrine gland in all vertebrates. It communicates the needs of the organism to different organs by secreting hormones into the bloodstream. Here, we have used the model fish medaka to investigate the developmental dynamics in the pituitary using a comprehensive RNA-seq time series. Using expression trend analyses we show that one of the most prominent changes during sexual maturation is the strong decrease of transcripts encoding secreted lipid transport proteins, which are typically only produced by the liver. By integrating developmental trends with single-cell transcriptomics, we demonstrate that rare, previously uncharacterized cells are responsible for this expression pattern. With levels as high as for established peptide hormone-encoding genes, this ectopic expression exposes a major new mechanism in the juvenile teleost pituitary. In addition, it implies the existence of unexpected connections between endocrine communication, sexual maturation, and lipid homeostasis.

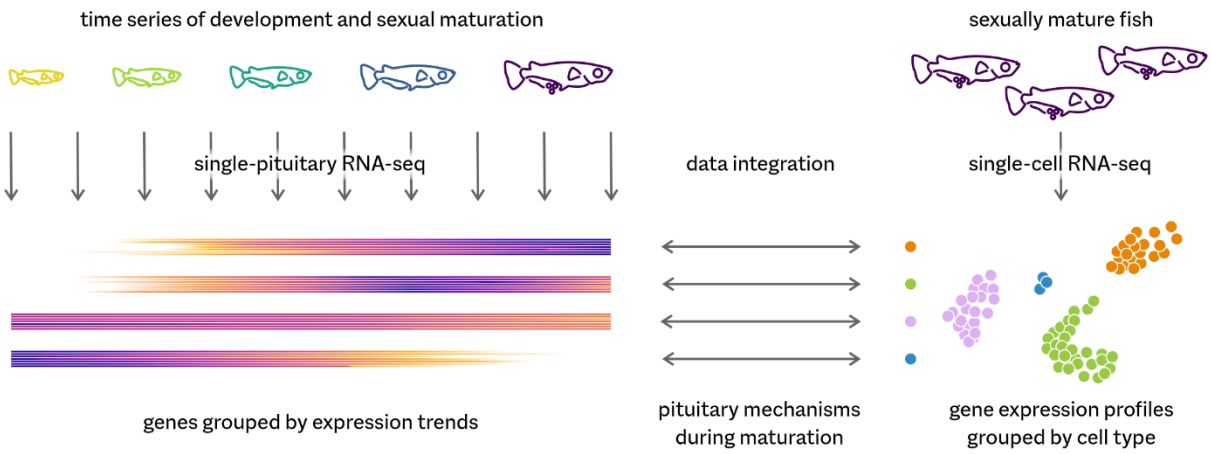

Figure 1. Study design. We infer gene expression trends during development and maturation from RNA-seq on pituitary glands of individual medaka fish. Using a previously described single-cell RNA-seq dataset on the pituitary of mature medaka, we can identify which cell types are associated with specific expression trends. In this simplified example, increase in gene expression in orange and green cells suggests they either proliferate or become more transcriptionally active during maturation; the expression trend associated with rare blue cells, on the other hand, implies they are more important during earlier developmental stages.

Introduction

All animals need to grow, reproduce, and maintain homeostasis – but all at the right time. Coordinating the present physiological needs of the organism and communicating these to different organs is the task of several endocrine glands. The pituitary is a key endocrine gland common to all vertebrates, involved in the regulation of many important physiological processes, which it modulates by releasing several peptide hormones into the bloodstream. By integrating signals derived from the brain, as well as feedback signals from downstream peripheral organs, the pituitary holds the focal position in several endocrine regulatory systems, including the brain-pituitary-gonadal (BPG), brain-pituitary-thyroid (BPT) and brain-pituitary-interrenal (BPI) axes[1].

Over the lifespan of an animal, the pituitary gland exhibits a high degree of developmental plasticity, allowing the proliferation or reduction of specific cell types to meet changing demands[2]. Distinct cell types are responsible for the production and secretion of at least eight different peptide hormones[1,3]. Sexual maturation, for instance, is controlled by pituitary gonadotrope cells producing follicle-stimulating and luteinizing hormones (Fsh and Lh, respectively). These cells are themselves under the control of several brain-derived factors (e.g. the hypothalamic gonadotropin-releasing hormone, Gnrh) and subject to feedback through sex steroids produced by the gonads and other organs[4]. Understanding pituitary mechanisms is, therefore, a central goal of endocrinological research, as it provides a key handle on the experimental management of an animal's physiology. In the case of teleost fish, this knowledge has important applications in aquaculture, ecology, and conservation.

One of the most important experimental perspectives on pituitary mechanisms is the transcriptome. Supported by high-throughput sequencing methods, it allows the sensitive inventory of a tissue's investment in specific functions. As an early example of the potential of transcriptomics, the last major new peptide hormone discovered in the pituitary – the teleost-specific somatolactin – was identified based on cDNA only[5]. A current extension of transcriptomics is single-cell (sc) RNA-seq, in which transcriptome profiling of individual cells within a heterogeneous tissue allows the inference of cell type-specific mechanisms and developmental trajectories[6].

To study the development of pituitary mechanisms over the course of sexual maturation, we here combine regular (bulk) RNA-seq and scRNA-seq on the pituitary gland of the model fish medaka, *Oryzias latipes* (figure 1). Using expression trend analyses on 54 individuals, ranging from juvenile to fully mature, we show two of the most prominent changes over development to be the expected increase of maturation-related hormone production, and the surprising expression and subsequent decrease of genes encoding proteins related to lipid homeostasis. As the latter (including apolipoprotein A, the structural component of high-density lipoprotein particles) are typically only secreted by the liver, their unexpected and strong expression in the juvenile pituitary implies a major new role in endocrine or paracrine communication.

Results

A time series of pituitary development

In order to capture major changes in the pituitary during sexual maturation, we designed a time series experiment in which we bred medaka siblings (offspring from three couples) under controlled conditions. We sampled the pituitaries and gonads of female fish after 31–178 days post-fertilization (dpf). The youngest fish are juveniles, and the smallest from which the pituitary gland can be reliably dissected; fish are fully sexually mature well before the final age. In addition, we sampled three older fish (249–277 dpf) from the same facility and the three parent fish (375 dpf). We used the gonads for a histological assessment of maturation, and profiled gene expression in individual pituitary glands using RNA-seq (9.5–16.2 million reads per sample).

We subsequently selected a group of 54 fish for which growth correlates well with age (figure 2a, table 1), and focused on genes that show medium to high expression in the pituitary. As a selection criterion, we took the appearance in any sample in the top 1000 by gene expression level. Over all samples, this results in a set of 2686 medium to high expression genes, which suggests samples differ considerably by expression profile.

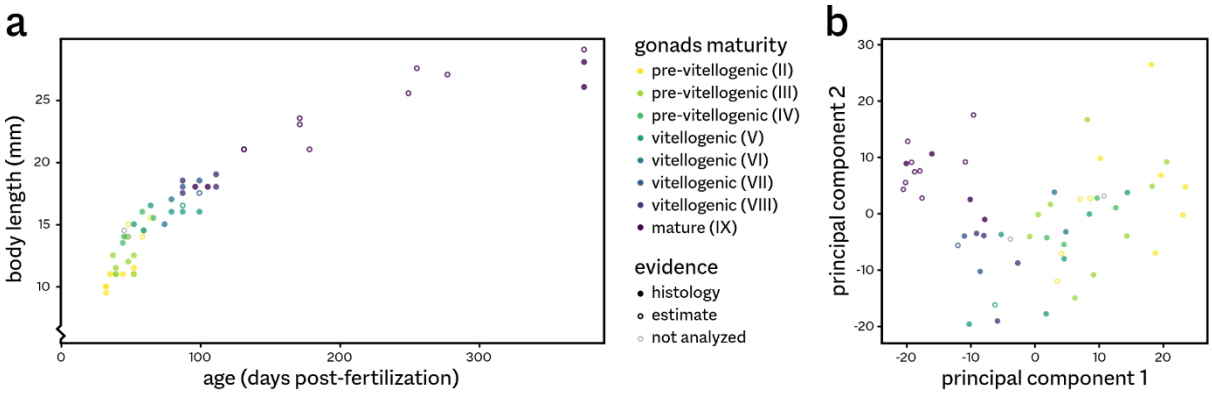

Figure 2. Growth and maturation in female medaka. a The fish used in this study range from young, small and immature (pre-vitellogenic gonads) to old, large and fully mature. Gonad maturity was assessed using histology for most samples; for a few fish, we only obtained a macroscopic estimate. We did not perform histological staging for the gonads of most of the mature, spawning fish. **b** Principal component analysis of the pituitary transcriptomes shows a strong correlation of the first principal component (PC1) with maturation. PC1 explains 21.9% of between-sample variance, PC2 11.9%.

To investigate whether these differences are linked to pituitary development over time, we performed a principal component analysis (figure 2b), which identifies the most important sources of expression variation among the samples. The main trend (principal component 1, explaining 21.9% of variance) indeed correlates well with sexual maturation. We therefore decided to use this first principal component (PC1) as a measure of developmental 'pseudotime', which can be used to unambiguously rank the pituitary transcriptomes. Of the 2686 medium-high expression genes, 52.7% are differentially expressed along this pseudotime axis, compared to 0.8% for the second principal component and 36.4% and 37.3% with age and length, respectively. When grouped by developmental stage, 47.6% of genes are differentially expressed between pre-vitellogenic (combined stages II–IV) and mature (IX) fish, 14.6% between pre-vitellogenic and vitellogenic (V–VIII), and 28.0% between vitellogenic and mature.

Table 1. Sampled fish per gonad maturation stage.

Stage	Description	Follicle/egg size (µm)	Nr. of fish	Nr. confirmed	Age (dpf)	Length (mm)
	No estimate		2	0	45–48	14.0–14.5
II	Early previtellogenic	60–90	10	6	32–64	10.0–15.5
III	Late previtellogenic	90–120	9	9	37–52	11.0–14.5
IV	Late previtellogenic	120–150	4	4	44–66	13.5–15.5
V	Early vitellogenic	150–250	7	6	52–99	14.5–16.5
VI	Early vitellogenic	250–400	2	2	74–79	15.0–17.0
VII	Late vitellogenic	400–500	3	2	87–99	17.5–18.5
VIII	Late vitellogenic	500–800	4	4	87–111	17.5–19.0
IX–X	Mature	800–1200	13	13	96–375	18.0–29.0

Gonadal maturation stage is based on initial macroscopic observations of the tissue, and was confirmed by histological examination (based on follicle size) for 39 fish. For 8 out of 13 mature fish, histological examination was not necessary, as externally visible eggs (approximate size 1200 µm) were present.

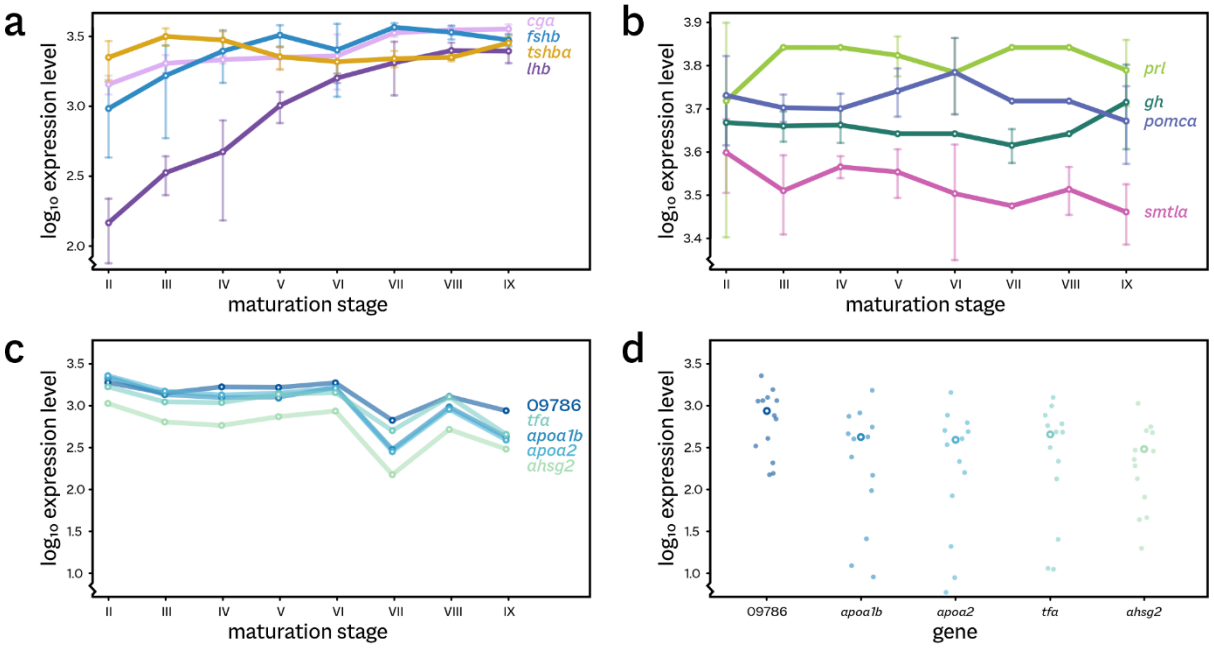

Figure 3. Gene expression patterns over development. a, b Expression levels of protein hormone-encoding genes. Shown are mean levels per maturation stage. Error bars represent standard deviations. Note the different y axis scales. **c** Mean expression levels of plasma transport genes. Expression decreases for adult animals, but is highly variable. **d** Individual expression values for the 13 mature animals (stage IX) are highly variable over up to two orders of magnitude, with the lowest expression measured in the fish furthest in development. Open circles: mean expression values (see panel c).

Divergent gene expression trends during maturation

The genes encoding the pituitary peptide hormones are amongst the most highly expressed throughout development (figure 3a, b). Both *fshb* and *lhb*, encoding the beta subunits of follicle-stimulating hormone and luteinizing hormone, respectively, show a significant increase in expression over time, as does the alpha subunit *cga*. *pomca* (encoding the hormone precursor pro-opiomelanocortin) and *prl* (prolactin) are consistently the most highly expressed genes in the entire pituitary. The levels of *gh* (growth hormone) and *tshba* (thyroid-stimulating hormone) increase slightly in mature animals, and *smtla* (somatolactin) expression decreases during development.

Surprisingly, the most highly expressed genes also include several encoding lipid homeostasis and transport proteins, such as *apoa1b* and *apoa2* (encoding the apolipoproteins A) and *tfa* (transferrin). Expression of these genes is usually associated with the liver, and has never been detected in the pituitary gland[7,8]. Their expression is as high as that of the established peptide hormones in younger fish (figure 3c), but declines precipitously in sexually mature animals (figure 3d), with the lowest expression observed in the most mature animals (by PC1 pseudotime).

These genes display highly similar expression trends over development, which can be an indication of common regulatory mechanisms or functions. In the same way, several hormone-encoding genes have comparable expression patterns. We therefore used weighted gene co-expression network analysis[9] (WGCNA) to categorize the 2686 medium-high expression genes into 24 expression modules, each representing a distinct expression trend (figure 4).

Nearly all of these expression modules are linked to specific biological processes (based on Gene Ontology overrepresentation tests), and many show a clear association with one or more maturation stages. For example, genes related to cell division are strongly downregulated in adult fish (module 19), while protein translation increases in these animals' pituitaries (module 7). The peptide hormone genes showing an increase over maturation are split between two large modules (4 and 6), while those decreasing are grouped in module 18. The consistently highly expressed *prl* is placed in a module (14) which shows no correlation with developmental pseudotime.

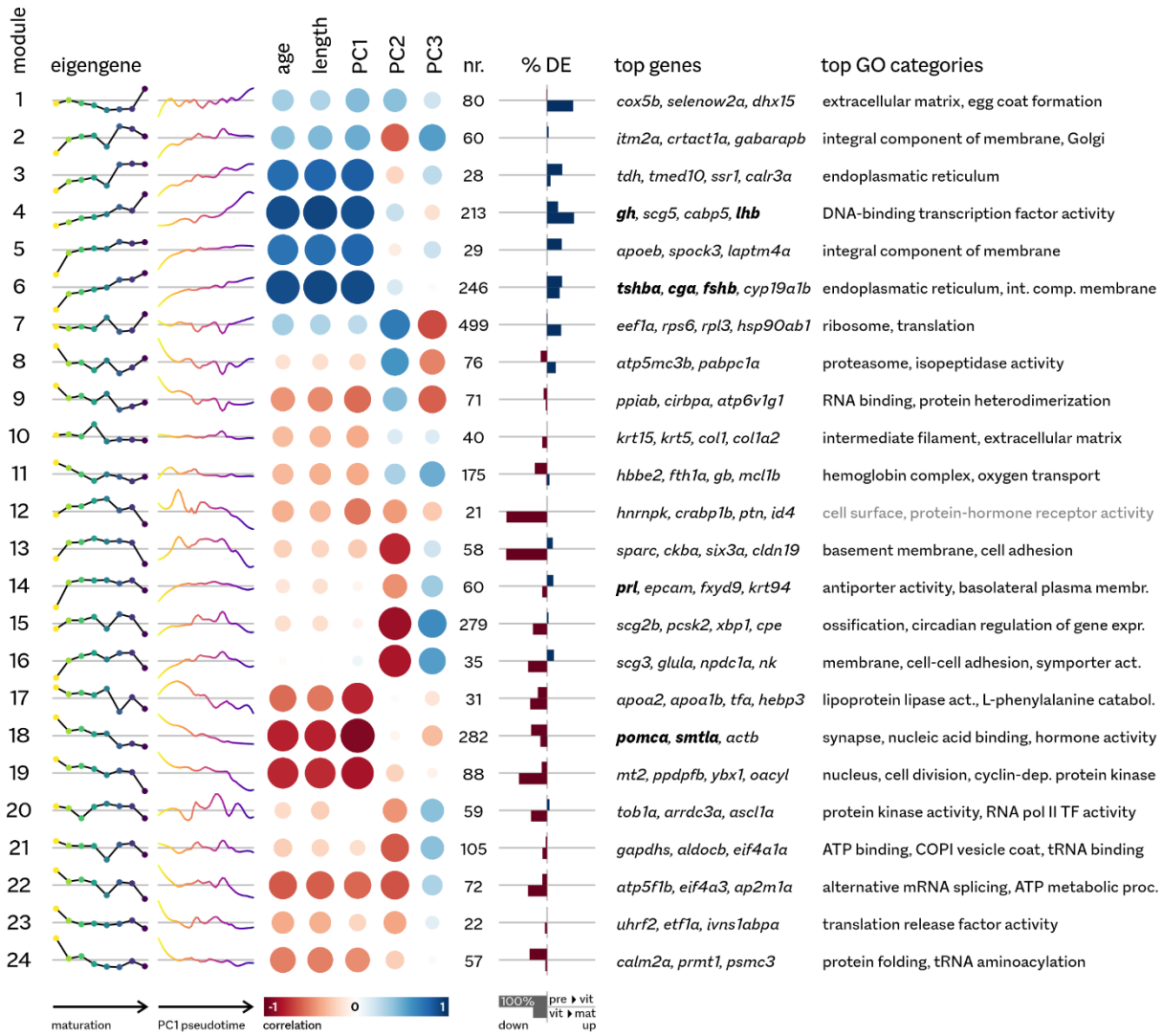

Figure 4. Expression modules. For each module, the typical expression pattern (eigengene) is shown both across gonad maturation stages and along principal component 1 (see figure 1b). These trends are based on variation around the mean expression over all samples (horizontal grey lines). Also shown are correlation with fish age and length, and with additional principal components; the number of genes in the module; the fractions of genes differentially expressed in either of two contrasts (previtellogenic-vitellogenic and vitellogenic-mature); top genes (by mean expression) in the module; and the most significantly overrepresented Gene Ontology (GO) categories (5% false discovery rate). All genes encoding the pituitary peptide hormones are amongst the highest expressed, and shown in bold. GO overrepresentation was not statistically significant for module 12 (grey text).

Juvenile expression of lipid transport and homeostasis genes

The co-expression analysis also groups the highly expressed lipid transport genes into a single, relatively small module (17, figure 4). This entire module comprises 31 genes (figure 5a), all of which decline strongly in expression over the course of development and sexual maturation (figure 5a–c). Most of these genes are involved in either amino acid biosynthesis and catabolism, specifically of phenylalanine (*pah*, *gatm*, *ces3*, *hpdb*, *hgd*), or in lipophilic compound transport and lipid homeostasis (*rbp4*, apolipoprotein C-I, *lipg*, *lpl*, *ahsg2*, *apoa2*, *apoa1b*). In addition, two uncharacterized genes (ENSORLG00000009771 and ENSORLG00000009786) are homologues of orosomucoid (alpha-1-acid glycoprotein), which in mammals is a plasma carrier protein of lipophilic compounds. Likewise, transferrin (*tfa*) is a plasma carrier of iron ions, and *hebp3* is a member of the heme-binding protein family.

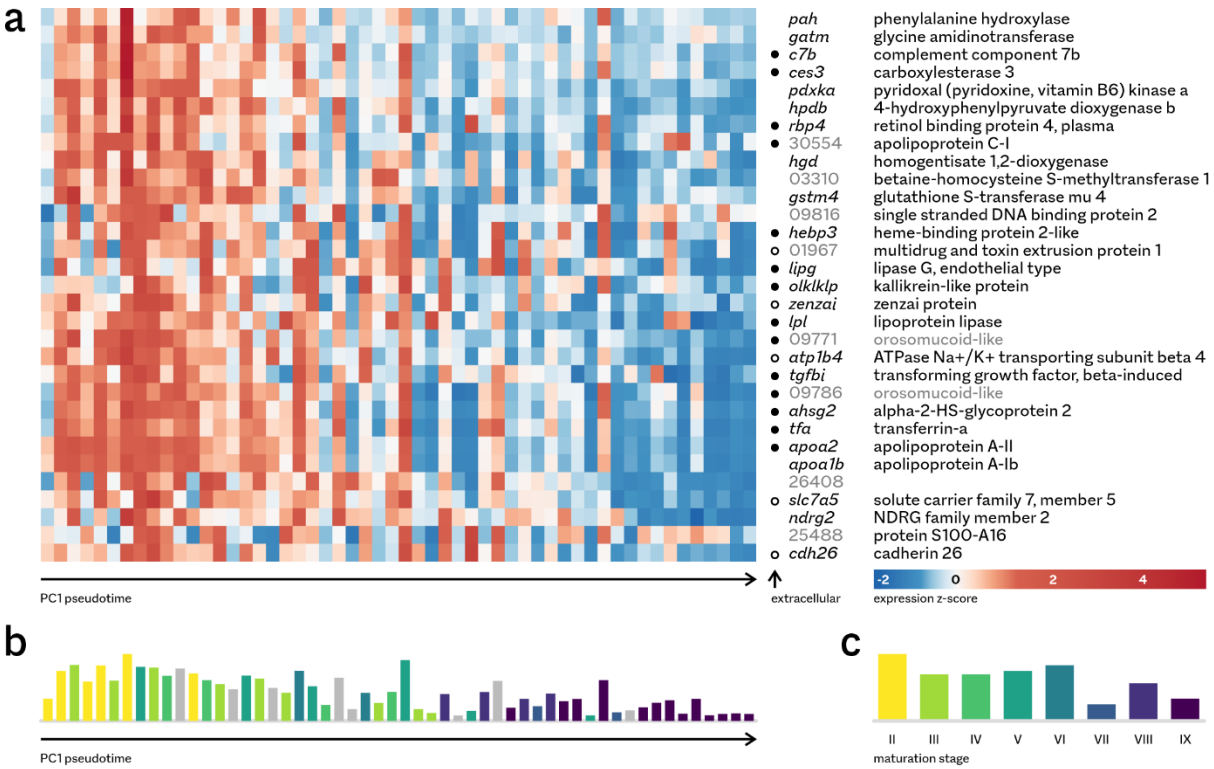

Figure 5. Genes in module 17 decrease in expression over development. a Heatmap of module 17 gene expression patterns, showing expression level variation in standard deviations around the mean for each gene (z-scores). Predicted extracellular protein products are marked by black circles, predicted cell membrane proteins by open circles. Unnamed genes are listed as numbers, which refer to Ensembl gene identifiers (e.g. 30554 is short for ENSORLG00000030554). Grey text for descriptions is based on our own analyses (no description in Ensembl). Expression of ENSORLG00000009816 (single-stranded DNA binding protein 2) appears to be caused by an annotation artifact (see figure 6b). **b** Total expression of module 17 genes over developmental pseudotime, scaled relative to the sample with the highest expression for this module. Colours indicate confirmed gonadal maturation stage (see panel c), with grey indicating samples lacking detailed histological confirmation. **c** Idem, mean expression per maturation stage.

These highly expressed carrier genes are typically specific to the liver, with their protein products secreted into the bloodstream. This homology-based functional annotation is corroborated by computational prediction of the subcellular location for the encoded proteins, which for nearly all of these genes strongly supports a secreted product (figure 5a).

In order to exclude the remote possibility of contamination with liver tissue, we compared total gene expression in the adult pituitary to that in liver[10]. Although *apoa1b* and *tfa* are indeed the most abundant transcripts in the medaka liver, overall the liver and pituitary transcriptomes are incompatible (figure 6a). For example, the highly expressed liver genes *fabp10a* (fatty acid binding protein 10a) and *c3-1* (complement component 3) are not detectable at substantial levels in the pituitary transcriptomes, while the most highly expressed gene in module 17, orosomucoid-like ENSORLG00000009786, is barely expressed in the liver at all. Interestingly, this gene is one of three (divergent) local duplicates, of which the other two *are* expressed in both the liver and the pituitary (figure 6b).

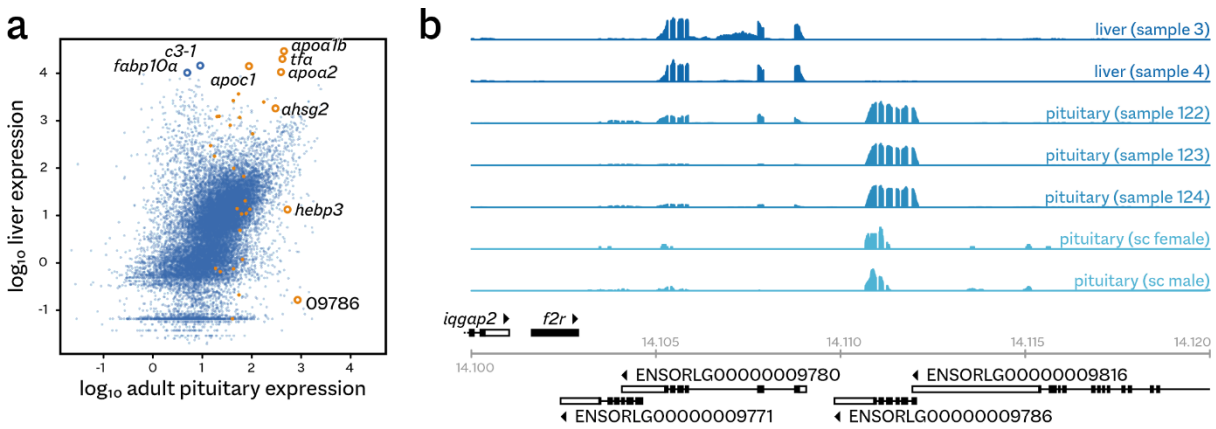

Figure 6. Expression of liver-specific genes in the pituitary is not caused by liver tissue. a Scatterplot showing the overall expression patterns in medaka pituitary and liver. Module 17 genes are highlighted in orange, with the most highly expressed annotated. Also annotated are the most highly expressed genes in the liver. The horizontal band of genes showing very low expression in liver is an artifact caused by the much higher sequencing depth for the liver samples. **b** Local gene expression for the orosomucoid-like locus on medaka chromosome 9. The locus consists of three local duplicates (ENSORLG00000009771/9780/9786), which are expressed at different levels in the pituitary and liver. A fourth gene, ENSORLG00000009816 (encoding single-stranded DNA binding protein 2) has a long 3' UTR annotation that overlaps the first exon of orosomucoid-like ENSORLG00000009786. Expression profiles for each sample are scaled to the maximum local expression value. Note that different sequencing library preparation protocols affect expression patterns over a gene: for example, the 10x single-cell protocol (sc) captures 3' expression only.

Gene expression trends are associated with specific cell types

As it needs to meet changing demands during development, the medaka pituitary is known to undergo compositional changes over time[2]. The dynamic landscape of transcriptional trends seen in figure 3 may therefore be a consequence of changes in cell type composition, in addition to changes in the regulation of gene expression within cells.

We have recently described a single-cell transcriptomics (scRNA-seq) dataset[3] on the medaka pituitary gland (figure 7a). These data comprise shallow transcriptomic profiles for 6396 cells (2592 female, 3804 male) derived from mature animals. Clustering of cells by similarity in expression (figure 7a) reveals distinct cell types linked to the production of specific peptide hormones. However, several large and small clusters of cells remain functionally uncharacterized (labelled A, B and C in figure 7a).

Integrating expression modules (figure 4) with scRNA-seq data can potentially uncover which cell types are responsible for specific expression trends. We therefore mapped the total expression of genes in each WGCNA module to the scRNA-seq clustering. This often presents a distinct expression pattern over cell types (figure 7b). For example, genes in module 4 are predominantly expressed in somatotropes and Lh-gonadotropes, which is consistent with the inclusion of *gh* and *lhb* in this module (together with 211 other genes).

Figure 7c summarizes this approach for all cell clusters and all WGCNA modules. Here, it is immediately apparent that many modules are indeed associated with specific cell types. This is also true for several of the modules showing the strongest change over development and maturation, suggesting these cells change number, overall transcriptional activity, or transcriptional investment in these genes.

These links between cell types and expression modules could potentially help expose the functions of the three uncharacterized cell types (A, B and C in figure 7a), which show strong associations with the relatively small modules 17 (A), 5 (B), and 12 (C). Of these, especially the first one is intriguing, as it

contains the highly expressed lipid homeostasis genes. In the adult pituitary glands sampled for scRNA-seq, only a small fraction of previously uncharacterized cell cluster A expresses these genes at high levels.

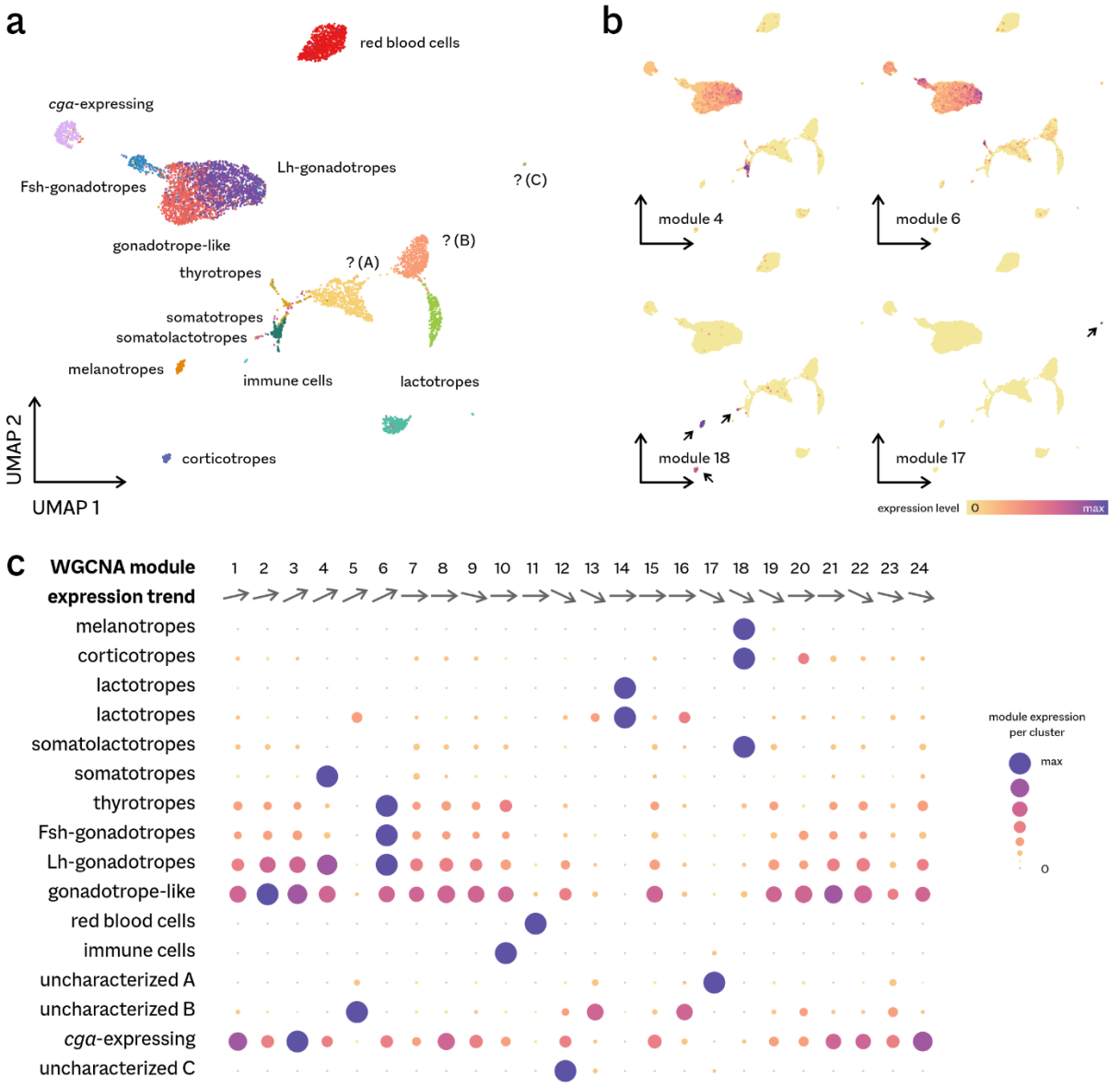

Figure 7. Integration of expression modules with single-cell RNA-seq data. a In this UMAP (uniform manifold approximation and projection) visualization of medaka pituitary scRNA-seq, cells cluster together if their transcriptomic profiles are similar. Based on this, we previously recognized 16 distinct cell types[3]. No function was defined for three of these cell clusters. **b** Examples of module expression in single cells. Module 4 (increasing over development) appears mostly linked to somatotropes and Lh-gonadotropes; module 6 (also increasing) to Lh- and Fsh-gonadotropes and thyrotropes. Module 18 is an example of a closer association with specific cell types (in this case, *pomca*-expressing melanotropes and corticotropes, as well as somatolactotropes). Module 17 is only expressed by a few cells in uncharacterized cluster A (arrow, co-localizing in this projection with uncharacterized cluster C). **c** Expression modules are often associated with specific cell types. Shown is the total expression of each module in each cell cluster, corrected for cell cluster size (total expression in all cells within a cluster) and gene module size (total expression in the module). Values are scaled to show the most important module per cell type (rather than the most important cell type per module). Expression trends over development are based on the eigengenes shown in figure 4.

Rare cell populations are responsible for the liver-like expression signature

The highly similar variation in lipid homeostasis gene expression (figure 5) indicates coordinated expression, and the association with a single (uncharacterized) cell type (figure 7b, c) implies this is caused by co-expression in specific cells. Of the 6396 scRNA-seq cells, only two show very high expression of these genes, and at the same relative expression levels as in the compound pituitary (figure 8a).

Conversely, nearly all genes highly expressed in these cells belong to module 17, and are restricted to a few cells (figure 8b, f). The only exception is the highly expressed gene *ptgdsb.1* (encoding a lipocalin family protein, possibly prostaglandin D_2 synthase), which is broadly expressed in all three of the uncharacterized cell clusters (figure 8c).

In total, five cells present a clear module 17 expression profile. All of these belong to the previously uncharacterized cell cluster A, but they also appear similar to another small cluster of 21 cells (C) that we found in male scRNA-seq data only (figure 8d–f). These specialized, rare cells are responsible for most of the overall expression of module 17 genes, suggesting they are either much more numerous or transcriptionally active in the juvenile pituitary gland.

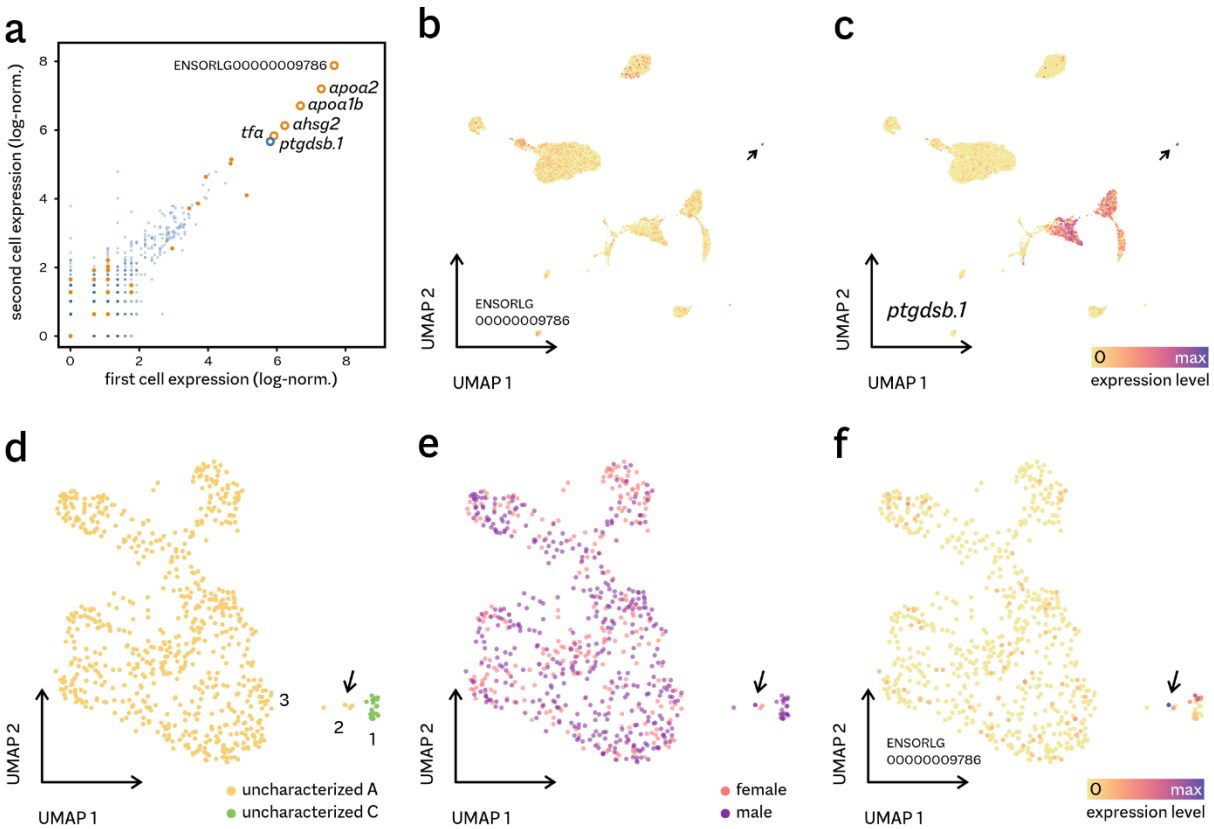

Figure 8. Rare cells display liver-like expression profiles. a Two cells (in the male scRNA-seq data) show high expression of module 17 genes (orange). **b** These cells belong to a small group of five (male and female) cells expressing module 17 genes (arrow). **c** These cells also highly express *ptgdsb.1*, which has a wider distribution of expression and does not follow the temporal trend specified by module 17. **d** In an UMAP projection focusing on the two uncharacterized clusters A (yellow) and C (green) only, the five cells (here marked as group 2) appear similar to a group (1) of 21 cells. **e** This group 1 was only detected in the male pituitary, group 2 in both females and males. **f** Both small groups express lipid transport genes.

These two small groups of cells (together accounting for 0.4% of scRNA-seq expression profiles) are characterized by the differential overexpression of a total of 134 genes. 74 of these genes are highly expressed in the compound pituitary at some developmental timepoint, and therefore included in the WGCNA analysis. Figure 9a shows how expression of these genes is specific to either or both of the rare cell groups, and sometimes present in the related cells as well (group 3 in figure 8d). In the developmental RNA-seq series, expression of most of these genes declines over pseudotime, but with different dynamics per WGCNA module (figure 9b).

Genes from three modules are strongly overrepresented amongst the 74: module 12 (43% of module genes), module 13 (26%) and module 17 (45%). All three modules are characterized by decreased expression in adults (figure 4) and associated with previously uncharacterized cell types (figure 6c). Of these, modules 12 and 17 are highly specific to the rare cells, while module 13 genes are more widely expressed across all three uncharacterized clusters (figure 7c).

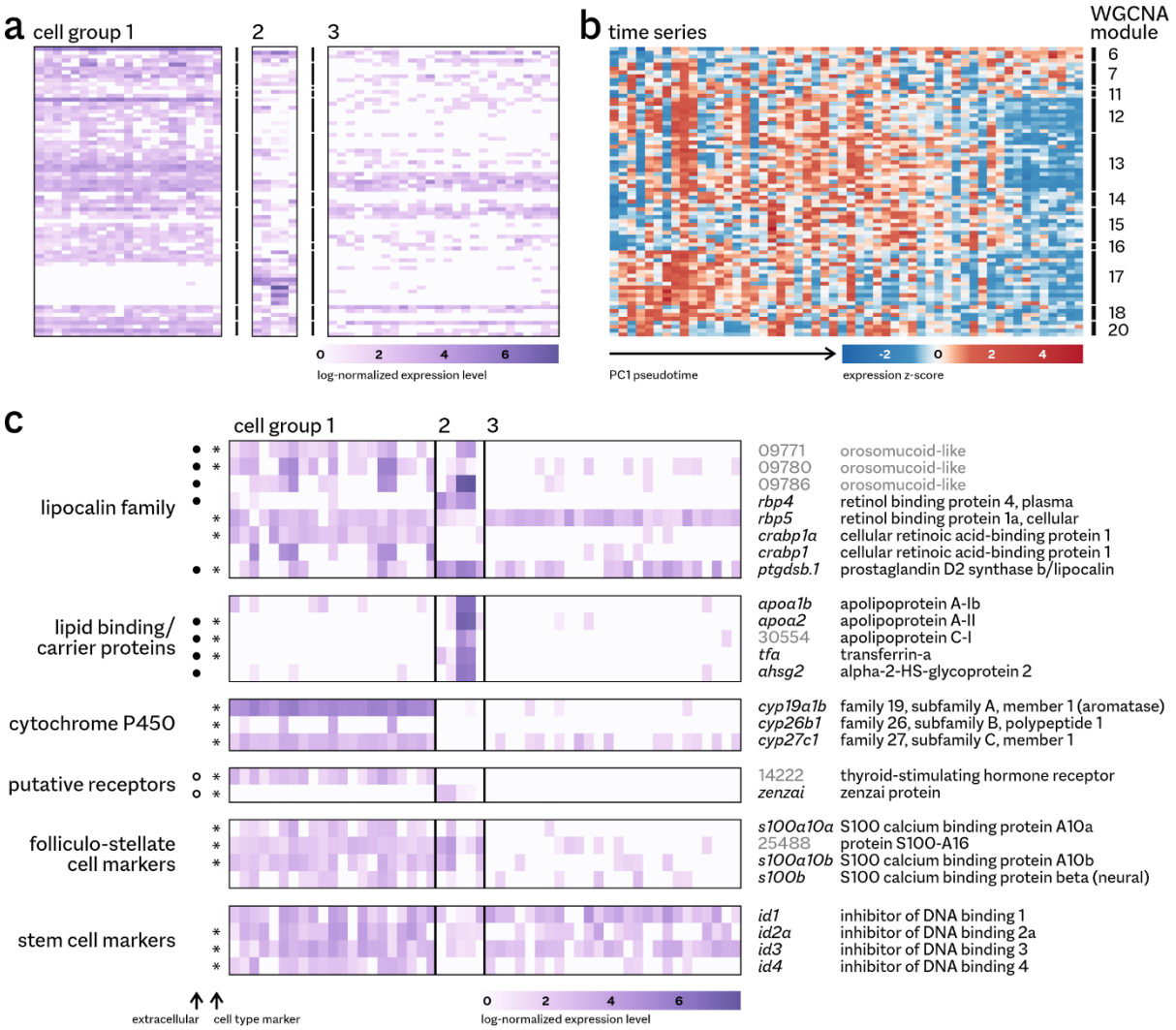

Figure 9. Expression profiles of two rare cell populations. a 134 genes are differentially expressed between either or both rare cells groups and the general pituitary (all cells excluding uncharacterized clusters A and C). See figure 8d for the definition of the cell groups. Group 3 are the 26 cells of cluster A that are most similar to the two rare cell populations (by cellular pseudotime analysis). 74 of 134 genes are shown in this heatmap, which corresponds to the genes included in time series WGCNA modules (**b**). **c** Selected highly and/or differentially (*) expressed genes suggest functional roles for the rare cell populations. Predicted extracellular protein products are marked by black circles, cell membrane proteins by open circles.

The marker genes and the genes generally highly expressed in these small groups of cells contain clear functional patterns (figure 9c). The five cells in the smaller group 2 (less than 0.1% of cells in the adult pituitary) express the liver-specific lipid transport genes, with two cells showing very high expression levels (figures 8a and 9c). Both groups express the orosomucoid-like genes, as well as *ptgdsb.1* and several genes encoding retinol/retinoic acid-binding proteins. These proteins belong to the same superfamily of lipocalins/calycins: small, usually extracellular proteins with lipophilic molecule binding functions[11,12]. Consistent with this observation, the 21 cells in the larger group 1 (0.3% of pituitary cells) specifically express cytochrome P450 enzymes involved in the synthesis of lipophilic compounds. *cyp19a1b* encodes aromatase, the enzyme converting androgens to estrogens and involved in steroid feedback along the BPG axis[4]; *cyp26b1* and *cyp27c1* encode enzymes typically involved in processing retinoids.

In addition, both cell groups specifically express putative receptor genes for the lipid transport genes involved. The receptor for thyroid stimulating hormone has been shown to respond directly to human orosomucoid[13], and is expressed exclusively in the 21-cell group. *zenzai*, specifically expressed in the five cells, is one of five medaka orthologues of scavenging receptor B, which binds to apolipoprotein A[14].

Finally, the differentially expressed genes contain an indication of cell type identity. All three uncharacterized clusters (figure 7a) express several S100 genes, which encode calcium binding proteins. Four paralogues are specific to the smaller groups (figure 9c). In mammals, this protein family is the established marker for the folliculo-stellate cells of the pituitary gland[15,16], a heterogeneous population presumed to be involved in paracrine communication[17] and exhibiting some of the characteristics of stem cells[18] (e.g. expression of *sox2*). Interestingly, this possible stem cell-like nature of the rare, juvenile-specific, lipid carrier-producing cells is supported by the high expression of all four medaka paralogues of the *inhibitor of DNA binding* developmental/stemness transcription factors.

Discussion

In this study, we have probed the development of the medaka pituitary gland by combining longitudinal RNA-seq data with in-depth scRNA-seq for a single timepoint (figure 1). The association between many developmental expression trends and cell type clusters (figure 7c) allows us to partially extrapolate cell type activity over time. This is analogous to explicit deconvolution of RNA-seq samples, but without the need for assumptions on the transcriptomic states of the cells in the younger fish. Here, this approach reveals the existence of novel populations of cells that are transcriptionally prominent in juveniles, yet physically all but absent in adults, without the need to perform scRNA-seq on challenging juvenile samples.

A disadvantage of this methodology is that different cell types need to exhibit distinct developmental dynamics and unique expression profiles, and that it requires cell types to be detectable at the single scRNA-seq timepoint. We have therefore combined scRNA-seq data from female and male fish to increase cell numbers (to 6396, still less than the approximately 15000 making up an adult medaka pituitary gland[19]). As our time series RNA-seq is based on females only, this potentially confounds the analysis. Both female and male scRNA-seq data, however, appear highly similar overall. One population that is not present in the female scRNA-seq data is the group of 21 cells (figure 8e). Their unequivocal association with a declining expression trend (figure 7c) provides strong evidence that this cell type is actually present in younger females, but not detectable in adults.

The two rare cell populations ectopically express genes not expected in the pituitary, most notably those encoding apolipoprotein A-I and A-II. These proteins are the structural components of high-density lipoprotein (HDL) particles produced and secreted by the liver[20,21]. We therefore ruled out the possibility these expression patterns can be explained by (highly implausible) contamination of our RNA-seq samples with medaka liver tissue (figure 6). In further support of the legitimacy of the phenomenon, we examined recently published zebrafish scRNA-seq data[22], which confirm the expression of *apoa1*, *apoa2* and *tfa* in the pituitary gland (figure 10).

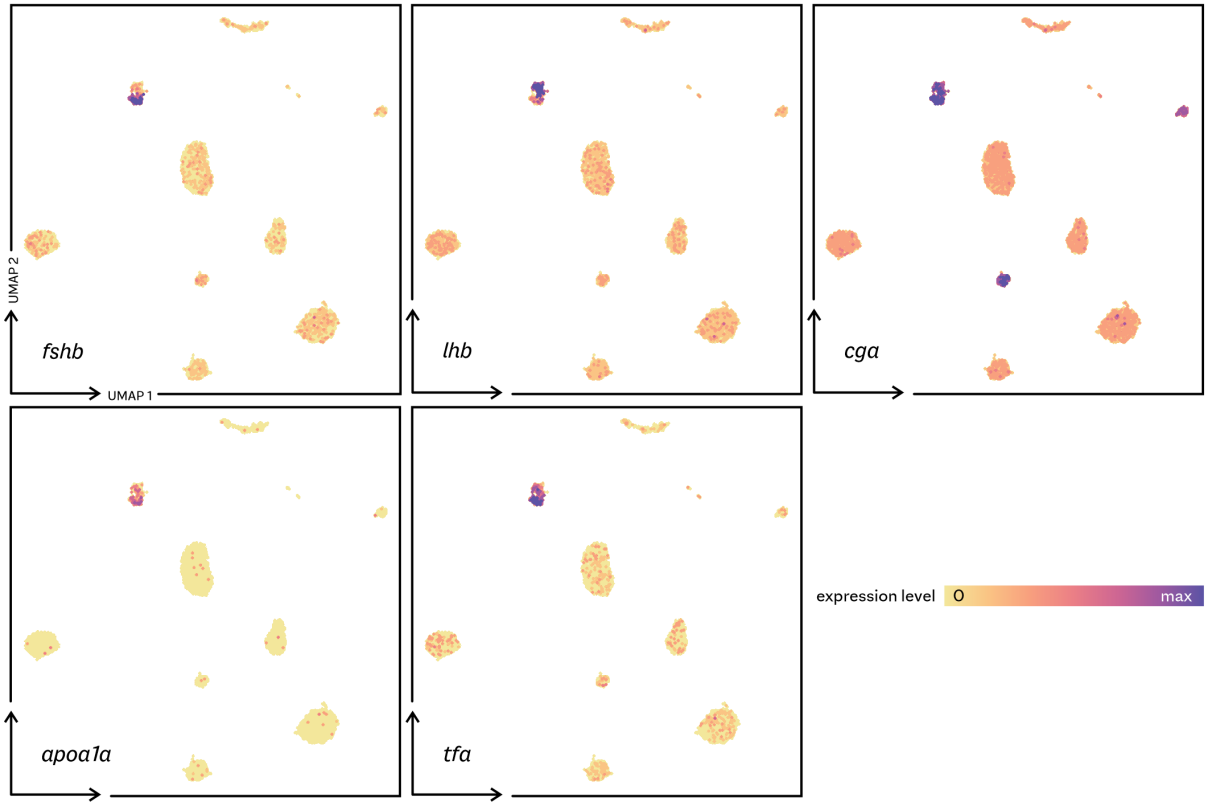

Figure 10. In the zebrafish pituitary gland[22], *apoa1a* and *tfa* are co-expressed in *fshb-/lhb*-expressing gonadotrope cells.

Interestingly, in zebrafish, these genes are expressed in gonadotropes instead of in folliculo-stellate cells (figure 10). Medaka and zebrafish are only very distantly related, and represent the two largest radiations near the base of the teleost fish clade. The gene regulatory programme(s) responsible are therefore likely common to all teleosts, but they may have been substantially rewired after the divergence of these model species. Together with the expression of different *apoa1* paralogues (*apoa1b* in medaka, *apoa1a* in zebrafish), this opens up the possibility that this programme is an evolutionary novelty ultimately caused by the whole-genome duplication event which shaped teleost evolution and development[23].

Further evidence on the expression patterns of *apoa1/2* in teleosts is inconclusive. In adult carp (a close relative of zebrafish), both show a tissue distribution that closely resembles the mammalian pattern[8], with no expression at all in the pituitary gland. In several teleosts, however, considerable expression is sometimes detected in the skin, brain, and testis, amongst other tissues[24,25]. In these cases, apolipoprotein A is often assumed to have a local anti-microbial function, a role that is difficult to reconcile with high expression in the pituitary.

Mammals show no clear expression of *apoa1/2* in the pituitary[7,26–29], although given the developmental dynamics in medaka, as well as the variation amongst teleosts, this current absence of evidence should not be considered conclusive. Intriguingly, when transfected with *apoa* expression constructs, murine pituicytes secrete HDL particles that are indistinguishable from liver-derived HDL[30]. Even in the absence of a mammalian counterpart for the teleost phenomenon, the involvement of key components of cholesterol transport in fish endocrinological systems suggests the regulation of lipid homeostasis in general contains hidden complexities.

The established physiological role of HDL is reverse cholesterol transport[21]. HDL particles sequester cholesterol from the bloodstream and deliver it back to the liver. In addition, the gonads and adrenal glands/interrenal tissue interact with HDL to obtain cholesterol, which they need as the raw material for the synthesis of steroid hormones. It is unlikely the pituitary-derived apolipoproteins primarily engage in this scavenging function, as even high juvenile production will be fully eclipsed by the much larger liver. However, many lipoproteins are thought to be involved in delivering messenger molecules in the brain[31]. This suggests the existence of a novel endocrine communication channel along the BPG and BPI axes, in which pituitary-derived HDL-like particles could deliver steroids or other lipophilic compounds to peripheral organs expressing the HDL receptor. Curiously, one medaka homologue of the HDL receptor (*zenzai*) is co-expressed specifically in the apolipoprotein A-producing cells (figure 9c). In medaka, *zenzai* has been shown to be essential for germ cell survival and gonad development[32].

A similar situation – co-expression of a lipid transport gene and its putative receptor – occurs in the other novel cell type. These express several orosomucoid-like genes[33], which in mammals affect thyroid hormone homeostasis through a direct interaction with the thyroid stimulating hormone (Tsh) receptor[13]. In our data, Tsh itself is produced by pituitary thyrotropes (see figure 7a), and the Tsh receptor is exclusively expressed in the rare cells (figure 9c). In the human pituitary, the Tsh receptor is expressed in a subset of folliculo-stellate cells[34].

These observations suggest both rare cell types engage in paracrine (or autocrine) signalling: communication to local cells (or feedback to the producing cell itself) mediated by secreted proteins[17]. Such a mechanism could play a role in the coordinated disappearance or transcriptional inactivation (by unknown mechanisms) of these cells over the course of maturation. A paracrine role is also consistent with their apparent identity as subpopulations of the folliculo-stellate cells[15].

The very high expression levels in the juvenile pituitary, however, do not support an exclusive intra-pituitary paracrine role. *apoa1b*, *apoa2*, *tfa*, and one orosomucoid-like gene are amongst the most highly expressed in immature fish, attaining the same levels as the genes encoding the established peptide hormones (figure 3). Our combined evidence therefore implies teleosts co-opted these secreted liver proteins as endocrine messengers.

Methods

Fish husbandry and growth experiment

We studied the pituitary glands and gonads of Japanese medaka (*Oryzias latipes*) of the d-rR genetic background. In addition, these fish express green fluorescent protein under the control of the *lhb* promotor[35], facilitating pituitary dissection even in small animals. We sampled eggs from a tank containing three females and three males. After hatching in petri dishes, fish from a single egg collection date were transferred to 1-liter tanks in a re-circulation system. After one month, fish were transferred to 3-liter tanks containing exactly 10 fish of the same age. Fish were kept at 28 °C on a

14/10h light/dark cycle, and were fed three times a day with a mix of dry food and *Artemia*. The animal experiments performed for this study followed EU/Norwegian regulations for animal welfare and for the use of animals in experiments.

Fish sampling

We sampled 118 female fish from our time series experiment (aged 31–178 days after egg collection), as well as the three female parents (375 days) and three mature females of intermediate age (249–277 days), housed in the same facility. The fish were sacrificed between 09.00–11.00h in the morning by severing of the spinal cord, and immediate dissection of the pituitary. Pituitaries were carefully dissected using fine forceps and extra care was taken to prevent cross-contamination between samples by carefully cleaning all equipment with ethanol after dissecting each fish, and also before sampling the pituitary. Each pituitary gland was put in an 0.5 ml Eppendorf tube containing 10 µl ice-cold medaka-adjusted[36] PBS (pH 7.75 and mOsm 290 mOsm) and 5 ceramic beads of 1.4 mm (MP Biomedials), directly followed by the addition of 300 ul Trizol Reagent (Invitrogen) to each tube. The sample was vortexed at high speed for 1 minute and snap frozen in liquid nitrogen. Samples were stored at -80 °C until RNA isolation.

Subsequently, ovaries were collected and transferred to small glass bottles containing 4% glutaraldehyde (Merck Millipore) 0.1 M phosphate-buffered solution (pH 7.2) and left to incubate overnight at 4 °C. The samples were then stored in 70% EtOH at 4 °C until histological examination. Whenever possible, ovaries were dissected out and examined macroscopically to obtain a preliminary maturity estimate. Fin clips were collected for sex genotyping.

Sex genotyping

Genomic DNA was isolated from fin clips by incubation in 25 µl alkaline lysis buffer (25 mM NaOH, 0.2 mM EDTA) for 5 minutes at 95 °C and subsequent neutralization using 25 µl 40 mM Tris-HCl (pH 8.0). Genotypic sex was determined by genomic PCR for the presence or absence of the *dmy* gene (primers 5'-CCGGGTGCCCAAGTGCTCCCGCTG and 5'-GATCGTCCCTCCACAGAGAAGAGA) and analysis by agarose gel electrophoresis[37,38].

Transcriptome sequencing

RNA was isolated using a standard Trizol protocol, with minor modifications. Samples were thawed and vortexed for 10 seconds with 120 µl chloroform, incubated at room temperature for 3 minutes, and centrifuged at 14000g for 15 minutes at 4 °C. Glycoblue (Invitrogen) (17.5 µg/ml) was added to each tube to enhance precipitation and visualize the pellet. The aqueous phase (ca. 160 µl) was transferred to a new tube and RNA was precipitated in 160 µl 100% isopropanol by incubation at -20 °C for 1 hour, followed by centrifugation at 14000 g for 15 minutes at 4 °C. The pellet was then washed in 300 µl 75 % ice-cold ethanol and the sample centrifuged at 10000g for 10 minutes at 4 °C. The supernatant was removed and the RNA pellet briefly dried (approximately 5 minutes) at room temperature. The pellet was dissolved in 10 µl of RNase-free water. To avoid unnecessary freeze-thawing of the samples, aliquots of 2 µl of each sample were made for Bioanalyzer analysis. Samples were stored at -80 °C until use.

RNA concentration and integrity of all samples were investigated using the Bioanalyzer Eucaryote ttal RNA pico kit. All samples included in the downstream RNA-seq analysis had RIN values between 7–10 (mean 9.3, median 9.6).

68 samples were selected for sequencing, of which 54 are included in the present analysis (13 additional sequenced pituitaries were from fish selected as growth curve outliers during the experiment; one sample was an outlier after sequencing). RNA was shipped on dry ice for library

preparation and sequencing at Future Genomics Technologies (Leiden, The Netherlands). cDNA was generated and amplified using the Smart-SEQ HT Kit for low RNA input according to the manufacturer's recommendations (Takara Bio), which was used for generating sequencing libraries with the Nextera XT kit. Libraries were sequenced at 151 nt paired-end on an Illumina NovaSeq 6000 system.

Histological analysis

The fixated tissue was dehydrated in a series of increasing concentrations of EtOH (70–100%), each step lasting at least >30 minutes. The last step (100%) was repeated trice and then replaced with approximately 5 ml of preparation solution (100 ml Technovit 7100 with 1 g of Hardener I (Heraeus Kulzer, Hanau, Germany)) and kept at slow shaking at room temperature overnight. After infiltration, tissue samples were embedded in cold Histoform S (Heraeus Kulzer) with approximately 1 ml preparation solution and 50 µl Hardener II (Heraeus Kulzer) and incubated at 37 °C. Cured samples were mounted on Histoblocs using Technovit 3040 (both from Heraeus Kulzer). The gonads were sectioned using a Leica RM2245 microtome (Leica Biosystems, Wetzlar, Germany). Sagittal sections (3 µm) were made from the periphery until the middle of the gonad and collected on microscope slides every 30–90 µm, depending on the gross maturity of the gonad. Dried sections were stained with Toluidine Blue O (Sigma-Aldrich) and mounted with Coverquick 4000 (VWR International, Radnor, PA, USA) prior to microscopy analysis. The developmental stage of the tissue was determined by the most advanced ovarian follicle present (see table 1)[39].

Additional data

Liver RNA-seq data for two female adult medaka were downloaded from the NCBI Sequence Read Archive (BioProject PRJEB37848, accessions ERS4513972 and ERS4513973). These fish (of the d-rR genetic background) were grown in the same facility as the fish of our main experiment[10]. The data are based on RNA-seq libraries prepared using the Illumina TruSeq Stranded mRNA HT Sample Prep Kit and sequenced at 125 nt paired-end on an Illumina HiSeq 2500 system.

Single-cell RNA-seq expression matrices were downloaded from NCBI Gene Expression Omnibus for medaka (GSE162787) and zebrafish (GSE148591), as well as for mouse (GSE120410, GSE146619), rat (GSE132224) and human (GSE142653). In the case of medaka, these fish are of the same transgenic line as those used in our main experiment.

Data analysis

All FASTQ data were aligned against the medaka HdrR reference genome (Ensembl version 94) using STAR[40] (version 2.7.6a). Alignments were inspected or processed using Samtools[41] (version 1.10) and quantified using htseq-count[42] (version 0.11.2) using the *intersection-nonempty* setting. The resulting data were analyzed in R (version 4.0.3) using the edgeR[43] (version 3.32.0) and cqn[44] (version 1.36.0) packages, using the Superheat (version 0.1.0) and Corrplot (version 0.84) packages for visualization. As our data are both amplified for low input material and derive from developmentally heterogeneous samples, they do not fit the assumptions for normalization by scaling commonly applied to RNA-seq (e.g. minimal changes in gene expression between samples). Counts for 23635 protein-encoding genes were therefore normalized between samples using quantile normalization as previously described[45].

Differential expression between maturity cohorts, and along age, length or principal components of the dataset were calculated using edgeR using the GLM functionality. Expression trend modules were identified using weighted gene correlation network analysis[9] (WGCNA package, version 1.69). Gene Ontology (GO) overrepresentation was analyzed using GOseq[46] (version 1.42.0) using GO annotations obtained from Ensembl BioMart. We only tested overrepresentation for categories containing more

than one gene and with at least one gene in the test group. Initial p-values for both differential expression and GO overrepresentation were corrected using the Benjamini-Hochberg procedure to control the false discovery rate at 5%.

Single-cell sequencing data were analyzed using Seurat[47] (version 3.2.3) for dataset integration and dimensionality reduction. Differential expression between groups of cells was determined using the Seurat *FindMarkers* function (using the default Wilcoxon rank-sum test), using an adjusted p-value threshold of 0.05 (Bonferroni correction). The reported differential expression results are the combination of four contrasts: overexpression of group 1, 2, or both, compared agains all pituitary cell excluding the uncharacterized clusters A and C; and differential expression between groups 1 and 2. Only genes expressed in at least two cells in the rare groups were considered. Monocle3 (version 0.2.3) was used to infer cellular pseudotime[48].

Subcellular protein sorting was predicted using SignalP[49] (version 5.0) and DeepLoc[50] (version 1.0), available at *www.cbs.dtu.dk/services*.

Acknowledgements

We would like to thank Lourdes Carreon G Tan for fish facility maintenance, Simen Sandve for information on the liver data, and Ron Dirks, Kjetil Hodne, and Erik Burgerhout for discussion. This work was supported by the Norwegian University of Life Sciences and the Norwegian Research Council grants no 251307, 255601, and 248828.

Author contributions

GM, EA-W, F-AW and CH conceived and designed the study. GM and EA-W performed the experiments. KvK analyzed gonad samples. RN-L performed genotyping. KS and RF contributed data analysis. CH analyzed the data and wrote the paper, with contributions from the other authors.

References

1. Weltzien, F.-A., Andersson, E., Andersen, Ø., Shalchian-Tabrizi, K. & Norberg, B. The brain–pituitary–gonad axis in male teleosts, with special emphasis on flatfish (Pleuronectiformes). *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **137**, 447–477 (2004).

2. Fontaine, R. *et al.* Gonadotrope plasticity at cellular, population and structural levels: A comparison between fishes and mammals. *Gen. Comp. Endocrinol.* **287**, 113344 (2020).

3. Siddique, K., Ager-Wick, E., Fontaine, R., Weltzien, F.-A. & Henkel, C. V. Characterization of hormone-producing cell types in the teleost pituitary gland using single-cell RNA-seq. *bioRxiv* 2020.12.14.422690 (2020) doi:10.1101/2020.12.14.422690.

4. Fontaine, R., Royan, M. R., von Krogh, K., Weltzien, F.-A. & Baker, D. M. Direct and Indirect Effects of Sex Steroids on Gonadotrope Cell Plasticity in the Teleost Fish Pituitary. *Front. Endocrinol.* **11**, (2020).

5. Ono, M. *et al.* cDNA cloning of somatolactin, a pituitary protein related to growth hormone and prolactin. *Proc. Natl. Acad. Sci.* **87**, 4330–4334 (1990).

6. Cheung, L. Y. M. & Rizzoti, K. Cell population characterization and discovery using single-cell technologies in endocrine systems. *J. Mol. Endocrinol.* **65**, R35–R51 (2020).

7. Sjöstedt, E. *et al.* An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science* **367**, eaay5947 (2020).

8. Kolder, I. C. R. M. *et al.* A full-body transcriptome and proteome resource for the European common carp. *BMC Genomics* **17**, 701 (2016).

9. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

10. Gillard, G. B. *et al.* Comparative regulomics reveals pervasive selection on gene dosage following whole genome duplication. *bioRxiv* 2020.07.20.212316 (2020) doi:10.1101/2020.07.20.212316.

11. Flower, D. R. The lipocalin protein family: structure and function. *Biochem. J.* **318**, 1–14 (1996).

12. Breustedt, D. A., Schönfeld, D. L. & Skerra, A. Comparative ligand-binding analysis of ten human lipocalins. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1764**, 161–173 (2006).

13. Zimmermann-Belsing, T., Rasmussen, Å. K., Feldt-Rasmussen, U. & Bøg-Hansen, T. C. The influence of alpha1-acid glycoprotein (orosomucoid) and its glycoforms on the function of human thyrocytes and CHO cells transfected with the human TSH receptor. *Mol. Cell. Endocrinol.* **188**, 241–251 (2002).

14. Shen, W.-J., Azhar, S. & Kraemer, F. B. SR-B1: A Unique Multifunctional Receptor for Cholesterol Influx and Efflux. *Annu. Rev. Physiol.* **80**, 95–116 (2018).

15. Allaerts, W. & Vankelecom, H. History and perspectives of pituitary folliculo-stellate cell research. *Eur. J. Endocrinol.* **153**, 1–12 (2005).

16. Devnath, S. & Inoue, K. An Insight to Pituitary Folliculo-Stellate Cells. *J. Neuroendocrinol.* **20**, 687–691 (2008).

17. Denef, C. Paracrinicity: The Story of 30 Years of Cellular Pituitary Crosstalk. *J. Neuroendocrinol.* **20**, 1–70 (2008).

18. Fauquier, T., Rizzoti, K., Dattani, M., Lovell-Badge, R. & Robinson, I. C. A. F. SOX2-expressing progenitor cells generate all of the major cell types in the adult mouse pituitary gland. *Proc. Natl. Acad. Sci.* **105**, 2907–2912 (2008).

19. Fontaine, R., Ager-Wick, E., Hodne, K. & Weltzien, F.-A. Plasticity in medaka gonadotropes via cell proliferation and phenotypic conversion. *J. Endocrinol.* **245**, 21–37 (2020).

20. Van der Vorst, E. P. C. High-Density Lipoproteins and Apolipoprotein A1. in *Vertebrate and Invertebrate Respiratory Proteins, Lipoproteins and other Body Fluid Proteins* (eds. Hoeger, U. & Harris, J. R.) 399–420 (Springer International Publishing, 2020). doi:10.1007/978-3-030-41769-7_16.

21. Luo, J., Yang, H. & Song, B.-L. Mechanisms and regulation of cholesterol homeostasis. *Nat. Rev. Mol. Cell Biol.* **21**, 225–245 (2020).

22. Fabian, P. *et al.* Lineage analysis reveals an endodermal contribution to the vertebrate pituitary. *Science* **370**, 463–467 (2020).

23. Glasauer, S. M. K. & Neuhauss, S. C. F. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genomics* **289**, 1045–1060 (2014).

24. Román-Padilla, J., Rodríguez-Rúa, A., Manchado, M. & Hachero-Cruzado, I. Molecular characterization and developmental expression patterns of apolipoprotein A-I in Senegalese sole (*Solea senegalensis* Kaup). *Gene Expr. Patterns GEP* **21**, 7–18 (2016).

25. Dietrich, M. A. *et al.* Expression of apolipoprotein A-I and A-II in rainbow trout reproductive tract and their possible role in antibacterial defence. *Fish Shellfish Immunol.* **45**, 750–756 (2015).

26. Cheung, L. Y. M. *et al.* Single-Cell RNA Sequencing Reveals Novel Markers of Male Pituitary Stem Cells and Hormone-Producing Cell Types. *Endocrinology* **159**, 3910–3924 (2018).

27. Fletcher, P. A. *et al.* Cell Type- and Sex-Dependent Transcriptome Profiles of Rat Anterior Pituitary Cells. *Front. Endocrinol.* **10**, (2019).

28. Ho, Y. *et al.* Single-cell transcriptomic analysis of adult mouse pituitary reveals sexual dimorphism and physiologic demand-induced cellular plasticity. *Protein Cell* **11**, 565–583 (2020).

29. Zhang, S. *et al.* Single-cell transcriptomics identifies divergent developmental lineage trajectories during human pituitary development. *Nat. Commun.* **11**, 5275 (2020).

30. Fennewald, S. M., Hamilton, R. L. & Gordon, J. I. Expression of human preproapo AI and pre(delta pro)apoAI in a murine pituitary cell line (AtT-20). A comparison of their intracellular compartmentalization and lipid affiliation. *J. Biol. Chem.* **263**, 15568–15577 (1988).

31. Wang, H. & Eckel, R. H. What are lipoproteins doing in the brain? *Trends Endocrinol. Metab.* **25**, 8–14 (2014).

32. Saito, D. *et al.* Proliferation of germ cells during gonadal sex differentiation in medaka: Insights from germ cell-depleted mutant zenzai. *Dev. Biol.* **310**, 280–290 (2007).

33. Luo, Z., Lei, H., Sun, Y., Liu, X. & Su, D.-F. Orosomucoid, an acute response protein with multiple modulating activities. *J. Physiol. Biochem.* **71**, 329–340 (2015).

34. Prummel, M. F. *et al.* Expression of the Thyroid-Stimulating Hormone Receptor in the Folliculo-Stellate Cells of the Human Anterior Pituitary. *J. Clin. Endocrinol. Metab.* **85**, 4347–4353 (2000).

35. Hildahl, J. *et al.* Developmental tracing of luteinizing hormone β-subunit gene expression using green fluorescent protein transgenic medaka (*Oryzias latipes*) reveals a putative novel developmental function. *Dev. Dyn.* **241**, 1665–1677 (2012).

36. Ager-Wick, E. *et al.* Preparation of a High-quality Primary Cell Culture from Fish Pituitaries. *JoVE J. Vis. Exp.* e58159 (2018) doi:10.3791/58159.

37. Shinomiya, A., Otake, H., Togashi, K., Hamaguchi, S. & Sakaizumi, M. Field Survey of Sex-Reversals in the Medaka, *Oryzias latipes*: Genotypic Sexing of Wild Populations. *Zoolog. Sci.* **21**, 613–619 (2004).

38. Ansai, S. & Kinoshita, M. Targeted mutagenesis using CRISPR/Cas system in medaka. *Biol. Open* **3**, 362–371 (2014).

39. Iwamatsu, T., Ohta, T., Oshima, E. & Sakai, N. Oogenesis in the medaka *Oryzias latipes*–stages of oocyte development. *Zoolog. Sci.* **5**, p353-373 (1988).

40. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

41. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

42. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

43. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

44. Hansen, K. D., Irizarry, R. A. & Wu, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostat. Oxf. Engl.* **13**, 204–216 (2012).

45. Ager-Wick, E., Henkel, C. V., Haug, T. M. & Weltzien, F.-A. Using normalization to resolve RNA-Seq biases caused by amplification from minimal input. *Physiol. Genomics* **46**, 808–820 (2014).

46. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).

47. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

48. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

49. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).

50. Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H. & Winther, O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33**, 3387–3395 (2017).