Single-cell transcriptome analysis reveals cell-cell communication and thyrocyte diversity in the zebrafish thyroid gland.

Pierre Gillotay[1], Meghna Shankar[1], Benoit Haerlingen[1], Sema Elif Eski[1], Macarena Pozo-Morales[1], Inés Garteizgogeascoa Suñer[1], Susanne Reinhardt[2], Annekathrin Kränkel[2], Juliane Bläsche[2], Andreas Petzold[2], Nikolay Ninov[3], Gokul Kesavan[3], Christian Lange[3], Michael Brand[3], Vincent Detours[1], Sabine Costagliola[1]*, Sumeet Pal Singh[1]*

[1] IRIBHM, Université Libre de Bruxelles (ULB), Route de Lennik 808, 1070 Brussels, Belgium.
[2] DRESDEN-concept Genome Center, DFG NGS Competence Center, c/o Center for Molecular and Cellular Bioengineering, TU Dresden, Dresden, 01307, Germany
[3] Center for Molecular and Cellular Bioengineering, TU Dresden, Dresden, 01307, Germany.
* Corresponding authors: scostag@ulb.ac.be; sumeet.pal.singh@ulb.ac.be

Abstract

The thyroid gland regulates growth and metabolism via production of thyroid hormone in follicles composed of thyrocytes. So far, thyrocytes have been assumed to be a homogenous population. To uncover genetic heterogeneity in the thyrocyte population, and molecularly characterize the non-thyrocyte cells surrounding the follicle, we developed a single-cell transcriptome atlas of the zebrafish thyroid gland. The 6249-cell atlas includes profiles of thyrocytes, blood vessels, lymphatic vessels, immune cells and fibroblasts. Further, the thyrocytes could be split into two sub-populations with unique transcriptional signature, including differential expression of the transcription factor *pax2a*. To validate thyrocyte heterogeneity, we generated a CRISPR/Cas9-based *pax2a* knock-in line, which demonstrated specific *pax2a* expression in the thyrocytes. However, a population of *pax2a*-low mature thyrocytes interspersed within individual follicles could be distinguished, corroborating heterogeneity within the thyrocyte population. Our results identify and validate transcriptional differences within the nominally homogenous thyrocyte population.

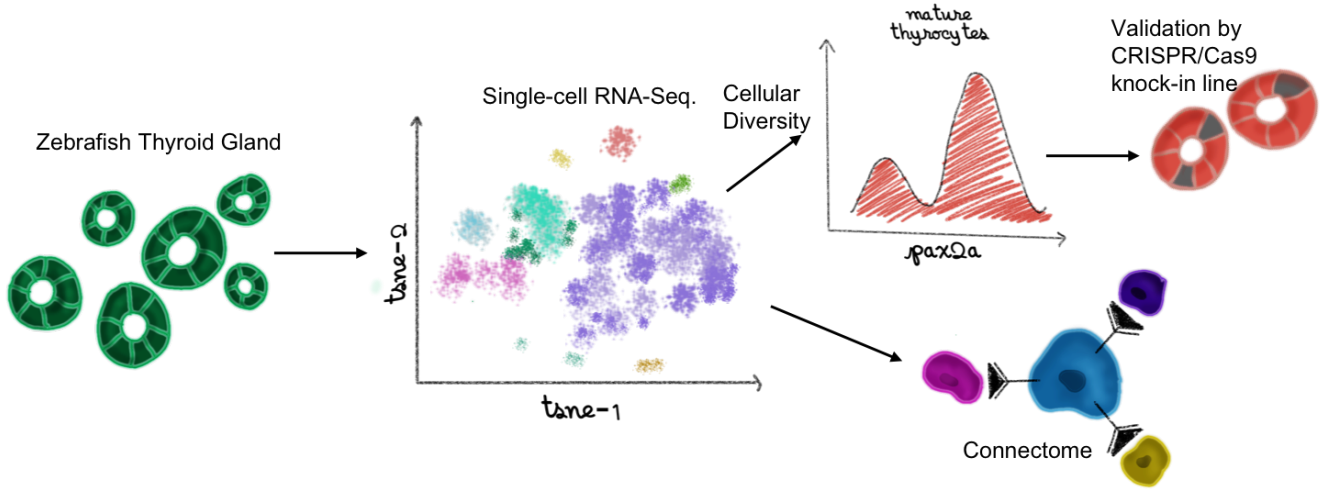
Keywords

Single-cell, transcriptomics, zebrafish, heterogeneity, thyroid gland, thyroid follicular cells, connective septa, CRISPR/Cas9, knock-in

One-line summary

Single-cell analysis uncovers latent heterogeneity in thyroid follicular cells.

Graphical Abstract



38 **Introduction**

39 The thyroid gland produces hormones thyroxine (T4) and triiodothyronine (T3)
40 that regulate body metabolism, growth, and development. Thyroid dysfunction, a
41 disease afflicting almost 100 million people worldwide (1), is common and treatable by
42 hormone replacement. If left untreated, however, it may result in profound adverse
43 effects on the human body, including mental retardation, goiter or dwarfism.

44 The thyroid gland is an endocrine organ with an intricate structure enabling
45 production, storage and release of the thyroid hormones. It contains numerous variable-
46 sized spherical follicles composed of thyroid follicular epithelial cells, or thyrocytes. The
47 thyrocytes generate the thyroid hormones in a multi-step process. They secrete and
48 store thyroglobulin (TG) in the lumen of the follicles. Additionally, they intake iodide from
49 the blood via sodium-iodide symporter (NIS / Slc5a5). At the interface between
50 thyrocytes and the lumen, thyroid peroxidase (TPO) expressed by the cells catalyzes
51 the coupling of iodide to tyrosyl residues of TG. Iodinated TG is absorbed back into the
52 thyrocyte and cleaved by cysteine proteases in lysosomes to form T4 and T3 (2).
53 Though the machinery responsible for the production of thyroid hormones by thyrocytes
54 is well established, it remains unknown if all the thyrocytes resident in the thyroid gland
55 are equally capable of generating thyroid hormones. In other words, the extent of
56 molecular homogeneity between individual thyrocytes has not yet been investigated.

57 Additionally, the thyroid gland contains many cell-types with potential roles in
58 modulating thyrocyte functionality. The gland contains an extensive distribution of blood
59 vessels, which carry iodide to the thyrocytes and carry thyroid hormones away from
60 them. The thyroid follicles are separated by a mesenchymal cell population, called

61 connective tissue septa, which also divides the gland into lobules. The mammalian
62 thyroid gland also contains parafollicular epithelial cells, or C-cells, that synthesize and
63 secrete the hormone calcitonin. These parafollicular epithelial cells are, however,
64 located outside the thyroid gland in fish and amphibians (3). Further, the presence of
65 immune cells and innervation has been demonstrated within the thyroid gland (4, 5).
66 Though we have a considerable understanding of these cell-types on a histological
67 level, we still lack the molecular characterization of the thyroid gland cell ensemble. This
68 extends to an incomplete appreciation of the impact of the diverse cell-populations on
69 thyroid follicular cell physiology.

70 To uncover the diversity within the thyrocyte population, and further characterize
71 the surrounding tissue at cellular resolution, we develop the first atlas of the thyroid
72 gland at single-cell resolution. For this, we build on the progress in single-cell
73 transcriptomics (6) to transcriptionally profile thousands of individual cells isolated from
74 the thyroid gland of adolescent and adult zebrafish. We demonstrate that these profiles
75 comprehensively represent the cells present in the zebrafish thyroid gland. Further, we
76 demonstrate the segregation of thyrocytes into two transcriptionally distinct sub-
77 populations. Utilizing the expression profiles of discrete cell populations, we build an
78 intercellular signaling network to uncover communication between thyrocytes and the
79 surrounding tissue. Finally, to enable easy access to the data, we have made the
80 zebrafish thyroid gland atlas available for online browsing.

81 **Results**

82 **Single-cell transcriptomics of the zebrafish thyroid gland.**

83 The zebrafish thyroid gland is composed of follicles scattered in the soft tissue
84 surrounding the ventral aorta (Fig. 1 A, B). Ventral aorta extends from the outflow tract
85 of the zebrafish heart and carries blood from the ventricle to the gills. Dissection of the
86 ventral aorta associated region (detailed in Methods section) provided us with tissue
87 that included the thyroid follicles and parts of zebrafish gills (Fig. 1C). Using *Tg(tg:nls-*
88 *EGFP)* transgenic line, which labels thyrocytes with nuclear green fluorescence (Fig.
89 1D), we estimated presence of 5.9 ± 1.9 % thyrocytes within the dissociated region (Fig.
90 1E).

91 To generate the molecular catalogue of the thyroid gland at cellular resolution,
92 we sampled the organ from two ages of zebrafish: 2 month post-fertilization (mpf) and 8
93 mpf (Supp. Figure 1). The time points span adolescent to adult transition in zebrafish,
94 with animals containing fully differentiated functional organs at both stages. By 2 mpf,
95 the adolescent animals have completed morphogenesis, but are yet to reach sexual
96 maturity. The animals sampled at 2 mpf were on average 2.6 cm in length and 123.8 mg
97 in weight. In contrast, fish at 8 mpf are sexually mature adults, with an average length of
98 3.5 cm and an average weight of 294.4 mg (Supp. Figure 1). To characterize the organ
99 cell-types in an unbiased manner, we dissected out the entire thyroid gland (Fig. 1B, C)
100 from six animals at each stage, and prepared the single-cell suspension for cDNA
101 library preparation. To guide thyroid gland dissection, we utilized the *Tg(tg:nls-mVenus-*
102 *T2A-NTR)* zebrafish reporter line (7) that labels thyrocytes with bright yellow fluorescent
103 protein (Fig. 1B'). The micro-dissected tissue was dissociated using enzymatic

104 digestion. The single-cell suspension was stained with calcein, which specifically labels
105 live cells with blue fluorescence. The live cells were then enriched using FACS (Fig. 1F
106 - G) to limit false positive signals from dead and/or ruptured cells (8). Thyrocytes
107 consisted of around 4 % of the alive cells at both stages, comparable to the percentage
108 obtained by immunofluorescence analysis (Fig. 1E). Twelve thousand live-cells, pooled
109 from six animals, were collected in separate tubes according to age and profiled using
110 droplet-based high-throughput single-cell RNA-sequencing provided by 10X Genomics
111 (9, 10). Droplet-based methods encapsulate cells with single-Poisson distribution (10).
112 This leads to approximately 50% cell capture rate, which is the ratio of the number of
113 cells detected by sequencing and the number of cells loaded. The 10X Genomics
114 pipeline uses molecule and cell-specific barcoding allowing transcript quantification
115 without amplification bias (11, 12). Using the Cell Ranger bioinformatics pipelines, the
116 resulting Next-Generation Sequencing libraries were mapped to the zebrafish genome,
117 de-multiplexed according to their cellular barcodes and quantified to generate gene/cell
118 UMI (unique molecular identifier) count tables. The Cell Ranger pipeline provided us
119 with 13,106 sequenced cells from 24,000 input cells (54.6 % cell capture rate). Quality-
120 based exclusion of single-cell transcriptomes was implemented based on mean library
121 size, percentage of mitochondrial reads and number of genes detected per cell. On
122 average, we detected 6,012 UMIs and 1,303 genes per cell (Supp. Figure 2). The
123 process recovered in total 6,249 cells out of 13,106 sequenced cells (47.7 % retention
124 rate), providing single-cell transcriptomic profiles for 2986 and 3263 individual cells for 2
125 mpf and 8 mpf, respectively.

126 **Identification of cell-types present in the zebrafish thyroid gland.**

127 To aid with visualization of the zebrafish thyroid gland single-cell RNA-Seq
128 (scRNA-Seq) data, we projected the cellular profiles onto t-distributed stochastic
129 neighbor embedding (t-SNE) plots, a non-linear dimensionality reduction technique (13)
130 (Fig. 2A). Using unsupervised graph-based clustering, we identified seven clusters for
131 the thyroid gland. Using the expression of genes involved in thyroid hormone
132 production, we could identify one of the clusters as thyroid follicular cells (Fig. 2B – D).
133 Specifically, the cluster displayed high relative expression of *tg* gene, which was further
134 enriched by background correction (Supp. Fig. 3); thereby demonstrating that the cells
135 represented differentiated thyroid follicular cells. The cluster, labeled as thyrocytes,
136 contains 267 cells. This represents 4.2 % of the total cells recovered after quality
137 control, similar to the proportion of thyrocytes quantified in the dissociated tissue by
138 imaging and FACS (Fig. 1E – G), suggesting lack of thyrocyte loss during the
139 sequencing procedure.

140 To define the identity of the remaining cell clusters, we generated cluster-specific
141 marker genes by performing differential gene expression analysis (Fig. 3A) (Supp.
142 Table 1). For four clusters, the marker genes included one or more known cell type–
143 specific identifiers. This included *gpr182* for endothelial cells; *acta2* for musculature;
144 *fcer1gl* for immune cells; and *ponzr3* for cells from zebrafish gills (Fig. 3B – E). Based
145 on these cell identifiers, the atlas includes 233 endothelial cells, 135 muscle lineage
146 cells, 914 immune cells and 199 cells from zebrafish gills. Notably, the endothelial cell
147 cluster includes blood vessels (*flt1* and *kdrl*) and lymphatic vessels (*mrc1a*, *prox1a*, *flt4*
148 and *lyve1b*) (Supp. Fig. 4); while the immune cell cluster includes macrophages

149 (*mpeg1.1* and *mfap4*), neutrophils (*lyz*) and lymphocytes (*il4*, *il13* and *il11b*) (Supp. Fig.
150 5).

151 For the remaining two clusters (number six and seven), we identified marker
152 genes that hinted towards identity of the cell-type. Specifically, *col1a2* and *tp63*
153 enriched in cluster number six and seven respectively (Fig. 3 F – G), are known
154 markers of fibroblasts (14, 15) and epithelial tissue (16–18). We performed gene-
155 ontology (GO) enrichment analysis of the marker genes to aid with classification (Supp.
156 Fig. 6). Cluster six demonstrated an enrichment of 'extracellular matrix structural
157 constituent', 'connective tissue development' and 'extracellular space', confirming the
158 presence of tissue fibroblasts in this cluster. Thus, we labelled cluster six as 'Stromal'
159 cells. Cluster seven displayed an enrichment of 'cell motility', 'cell migration' and
160 'epithelium development', suggestive of epithelial cells. Hence, we labelled cluster
161 seven as 'Non-Follicular Epithelium (NFE)', to distinguish them from the thyroid follicular
162 epithelial cells. Our data contains 3670 stromal cells and 831 non-follicular epithelial
163 cells.

164 We validated the presence of blood vessels, macrophages and stromal cells in
165 the thyroid gland using tissue specific transgenic lines (Fig. 4A – C).
166 Immunofluorescence (IF) analysis demonstrated physical proximity between thyrocytes
167 and blood vessels (Fig. 4A). Notably, we observed a subset of macrophages in direct
168 contact with thyroid follicles (Fig. 4B). In addition, we visualized NFE by immunostaining
169 against TP63 antibody (Fig. 4D), which revealed NFE scattered throughout the gills and
170 in the region adjacent to the follicles. Thus, the IF analysis successfully confirmed the
171 presence of different cell types identified in the single-cell atlas.

172 Our marker gene identification further established additional genes enriched in a
173 single cell-type in the thyroid gland (Fig. 3A) (Supp. Table 1). For instance, we identified
174 *cx30.3*, a connexin gene and *prdx1*, a gene involved in the antioxidant response, to be
175 specifically expressed in the thyrocytes. To enable further investigation of the clusters
176 and gene expression profiles, we have developed an interactive webtool for online
177 browsing (https://sumeet.shinyapps.io/zfthyroid/).

178 **Development of autocrine and paracrine signaling networks in the thyroid gland**
179 **using known ligand-receptor interactions.**

180 Having defined the cell types of the thyroid gland, we quantified potential cell-cell
181 interactions between thyrocytes and all cell types present in the organ (Fig. 5A) based
182 on a reference list of approximately 3,100 literature-supported interactions containing
183 receptors and ligands from receptor tyrosine kinase (RTK), extracellular matrix (ECM)-
184 integrin, chemokine and cytokine families (19). Although anatomical barriers between
185 cell types are not modeled in this analysis, we restricted the analysis to secreted ligands
186 for NFE, stroma and gills -- cell types that are physically separated from thyrocytes (Fig.
187 4C - D). For the remaining cell types, secreted and cell-membrane tethered ligands
188 were considered. The expression patterns of ligand-receptor pairs revealed a dense
189 intercellular communication network (Fig. 5B). The network consisted of 272 ligands
190 expressed on different cell-types with a corresponding receptor expressed on the
191 thyrocytes (Supp. Table 2). For instance, the stromal cells express the ligand *lpl*
192 (Lipoprotein Lipase) that signals through the *lrp2a* (zebrafish homologue of Megalin)
193 receptor (Fig. 5C). Stromal and smooth muscle cells express *dcn* (Decorin) whose
194 receptor *met* is expressed by thyrocytes. Further, the ligand *cyr61* is broadly expressed

195 in the thyroid gland, with one of its receptors, *itgb5*, an integrin isoform, expressed
196 specifically by the thyrocytes. The identified interactions also include autocrine
197 signaling. For example, the ligand *sema3b* and its receptor *nrp2a* are both present on
198 thyrocytes. GO-analysis for identified ligand-receptor pairs revealed genes involved in
199 'PI3K-Akt signaling pathway', 'MET signaling' and 'integrin binding' (Supp. Fig. 7).

200 **Thyrocytes are composed of transcriptionally distinct sub-populations.**

201 Next, we characterized the transcriptional differences within the thyrocyte
202 population. For this, we bioinformatically isolated the thyrocytes, and re-performed the
203 clustering pipeline on the isolated cell population. With this, we could segregate the
204 thyrocytes into two smaller clusters (Fig. 6A), labeled as 'Cluster_Blue' and
205 'Cluster_Red'. The two clusters displayed differences in the expression levels of 265
206 genes (Fig. 6B) (Supp. Table 3). Notably, Cathepsin B (*ctsba*) is significantly
207 downregulated in the blue cluster (Fold change = 1.6, p-value = 1.47×10^{-9}) (Fig. 6B –
208 C). Cathepsin B is a cysteine protease that is involved in the processing of iodinated
209 thyroglobulin to T4 and T3 in the thyrocyte lysosomes (2, 20). Moreover, fusion of
210 Cathepsin B and EGFP has been previously used to track thyroid hormone processing
211 lysosomes in rat thyroid epithelial cell lines (21).

212 Along with higher expression of Cathepsin B, the red cluster displayed significant
213 downregulation of *pax2a* expression (Fold change = 1.7, p-value = 8.24×10^{-9}) (Fig. 6B-
214 C). *pax2a* belongs to the PAX (paired box DNA-binding) domain containing family of
215 transcription factors. The loss of *pax2a* expression in the red cluster is notable, as
216 *pax2a* is an important regulator of thyrocyte development (22). Zebrafish thyroid
217 primordium expresses *pax2a* at 24 hpf (22), which is required for specification of the

218 thyroid follicles (23, 24). Consequently, zebrafish lacking *pax2a* fail to develop thyroid
219 follicles (22), which is similar to the Pax8 knock-out phenotype in mouse (25). The low
220 expression of *pax2a* in the red cluster, without a difference in *tg* expression (Supp. Fig.
221 8), suggests the presence of a thyrocyte sub-population with a distinct gene expression
222 signature.

223 Independent analysis of genetic entropy, a measure of the degree of uncertainty,
224 revealed transcriptional heterogeneity in 231 genes in the thyrocyte population (Fig. 6D)
225 (Supp. Table 4). Genes displaying statistically significant entropy (p-value < 0.05)
226 included *pax2a* and *ctsba*, corroborating their expression heterogeneity within
227 thyrocytes.

228 **Generation of *pax2a* knock-in reporter line.**

229 To validate the heterogeneity among the zebrafish thyrocytes, we focused on the
230 expression of *pax2a* transcription factor. We generated a knock-in line by inserting
231 monomeric Kusabira Orange 2 (mKO2) fluorescent protein to the 3' end of the
232 endogenous *pax2a* genomic location (Fig. 7A). The *pax2a*$^{pax2a\text{-}T2A\text{-}mKO2}$ (abbreviated as
233 *pax2a*^{mKO2}) reporter expression overlapped with PAX2A antibody staining in a majority of
234 regions at 9.5 hours post-fertilization (Fig. 7B). Moreover, the knock-in line displayed
235 mKO2 fluorescence in the otic vesicle, mid-hindbrain boundary, optic stalk, pronephros
236 and the thyroid gland (Fig. 7C – F, Supp. Movie 1), mimicking known expression of *pax2a*
237 during zebrafish development (26). Additionally, in order to assess whether the dynamics
238 of mKO2 expression would follow modifications in the expression of endogenous *pax2a,*
239 we used CRISPR/Cas9 technology to generate F0 knock-outs (also known as Crispant
240 (27)) of *pax2a* gene in our *pax2a*^{mKO2} line. The crispants displayed defects in thyroid

241 morphogenesis (Fig. 7G – H), mimicking the phenotype of *pax2a* loss-of-function
242 mutation (22). Live imaging of crispants at 55 hpf revealed strong decrease of mKO2
243 expression (Fig. 7G – H), thereby corroborating the faithful recapitulation of *pax2a*
244 expression by the newly generated reporter line.

245 **Segregation of thyrocyte sub-populations based on *pax2a* reporter expression.**

246 Upon investigating the fluorescence expression of the *pax2a* reporter in the
247 thyroid gland of adult zebrafish, we found strong and specific expression of *pax2a*
248 reporter in the thyrocytes lining the thyroid follicles (Fig. 8A – D). Although a majority of
249 thyrocytes displayed uniform expression of *pax2a* reporter, we could identify a small
250 population of *pax2a*^{mKO2}-Low thyrocytes (Fig. 8B – D). The *pax2a*^{mKO2}-Low thyrocytes
251 were not segregated, but scattered throughout the gland, thereby suggesting a mixing
252 of the two thyrocyte sub-populations.

253 To validate pax2a expression heterogeneity at a protein level, we performed
254 immunostaining against PAX2A in thyroid gland obtained from *Tg(tg:nls-EGFP)* animals
255 (Fig. 8E). For antibody staining, we utilized 8 µm thin sections of the thyroid gland to
256 ensure uniform antibody penetration to all cells. Confocal imaging of the stained
257 sections demonstrated the presence of PAX2A-Low and PAX2A-High thyrocytes (Fig.
258 8F). Notably, both PAX2A-Low and -High cells display *tg* promoter-driven EGFP
259 expression, thereby confirming their differentiated status.

260 Further, to quantify the proportions of *pax2a*^{mKO2}-Low and -High thyrocytes, we
261 performed FACS analysis on *pax2a*^{mKO2}; *Tg(tg:nls-EGFP)* double transgenic line (Fig.
262 9A – C). The *Tg(tg:nls-EGFP)* zebrafish line labels the thyrocyte population in green
263 fluorescence (27). We restricted our analysis to the thyrocyte population by gating for

264 GFP+ cells in the thyroid gland (Fig. 9A). Within the thyrocyte population, the cells
265 displayed a normal distribution of GFP fluorescence; however, thyrocytes could be split
266 into two sub-populations based on the levels of *pax2a* reporter expression (Fig. 9B – C).
267 Specifically, 75% of thyrocytes (202 out of 268 cells) displayed *pax2a*^{mKO2}-High
268 fluorescence, while 25% of thyrocytes (66 out of 268 cells) displayed *pax2a*^{mKO2}-Low
269 fluorescence levels.

270 In summary, the analysis of *pax2a* knock-in line, along with PAX2A
271 immunofluorescence imaging, validates the identification of thyrocyte sub-populations
272 within our single-cell RNA-Seq. data, and clearly demonstrates, for the first time, the
273 presence of transcriptionally diverse sub-populations of thyrocytes present in the thyroid
274 gland.

275 **Discussion**

276 We have applied for the first time unbiased single-cell gene expression analysis
277 to the thyroid gland. In contrast with the mainstream view that thyrocytes constitute a
278 molecularly uniform population, we identify two transcriptionally distinct sub-populations
279 of thyrocytes. The two sub-populations differed, among other genes (Supp. Table 3), in
280 the expression levels of a transcription factor *pax2a* and a cysteine protease Cathepsin
281 B (*ctsba*) (Fig. 6B – C). Cathepsin B is particularly notable as it enables the liberation of
282 thyroid hormone from thyrocytes by proteolytic processing of thyroglobulin (2, 20).

283 We validate the heterogeneity among the thyrocytes using a newly generated
284 knock-in reporter line for *pax2a* gene (Fig. 7). The knock-in reporter line was generated
285 using CRISPR/Cas9-based insertion of mKO2 fluorescent protein in the endogenous
286 *pax2a* genomic location. The *pax2a* knock-in line faithfully recapitulates the embryonic

287 expression of *pax2a* gene (Supp. Movie 1, Fig. 7B – F). Using the *pax2a* reporter line to
288 characterize the adult thyroid gland, we demonstrate the presence of *pax2a*^{mKO2}-Low
289 thyrocytes in the follicles (Fig. 8). Notably, *pax2a*^{mKO2}-Low and *pax2a*^{mKO2}-High
290 thyrocytes are present in the same follicle (Fig. 8C – D, F), raising the possibility of
291 contact-mediated interactions between the two sub-populations. It would be of interest
292 to build on this study and investigate the functional and replicative differences among
293 the two sub-populations of thyrocytes.

294 Our single-cell transcriptomics atlas provides a comprehensive genomics
295 resource to study the zebrafish thyroid gland in unprecedented detail. We performed
296 unbiased profiling of the thyroid gland, without enrichment for a specific cell-type. This
297 allowed us to capture yet poorly characterized cell-populations within the thyroid gland.
298 Specifically, we provide the molecular characteristics of the stromal tissue present in the
299 zebrafish thyroid gland. The stromal cells (Fig. 4C) display enrichment of extra-cellular
300 matrix (ECM) related genes (Supp. Table 1) and are possibly homologous to the
301 mesenchymal connective septa found in the mammalian thyroid gland. The connective
302 septa helps cluster the thyroid follicles into lobules. Notably, the expression of fgf
303 ligands from the mesenchymal septa cells has been implicated in lobe formation during
304 mouse thyroid gland development (28). It would be of interest to test if similar
305 morphological clustering of the thyroid follicles exists in zebrafish, and the role the
306 stromal cells play during development and growth of the thyrocytes.

307 Our atlas further identifies a non-follicular epithelial (NFE) cell-population present
308 near the zebrafish thyroid follicles. A subset of NFE are present in the gills (Fig. 4D) and
309 may potentially represent a progenitor population for the gills, similar to the TP63+ basal

310 layer in the zebrafish (29) and mammalian (30) epithelium. We also observe NFE
311 outside the gills (Fig. 4D), which may play a different role. It is interesting to note that
312 epithelial cells apart from follicular and parafollicular cells have been observed in the
313 mammalian thyroid gland. In a report from Dr. E. Baber published in 1876 (31),
314 histological examination of the dog thyroid gland displayed the presence of cells “beside
315 the stroma, lymphatics, blood vessels, & cells between the vesicles”. Dr. Baber labeled
316 the cells as ‘parenchyma’, and noted the existence of “numerous cells differing
317 markedly in size and shape from the epithelial cells amongst which they lie” (31). In
318 1907, Dr. Sophia Getzowa described an epithelial cells containing structure called the
319 Solid Cell Nests (SCN) of the thyroid (32). SCN are lumen containing irregular
320 structures located within the thyroid in mammals (33). SCN contain two types of
321 epithelial cells: main cells and C-cells, expressing TP63 and calcitonin respectively (34).
322 Notably, the NFE cells we identified in the zebrafish thyroid gland are marked with TP63
323 expression (Fig. 3G, 4D), raising the possibility of their homology with the main cells of
324 the SCN. C-cells, however, exist in the ultimobranchial bodies, which lies outside the
325 thyroid gland in zebrafish. The ultimobranchial bodies are the zebrafish homologues of
326 parafollicular cells and are located as a pair of follicles on top the sinus venous,
327 adjacent to the atrium and oesophagus (3). Cells adjacent to the atrium were removed
328 during our dissections (Fig. 1B). Additionally, NFE cells identified in our atlas do not
329 express the zebrafish homologue of calcitonin (*calca*) (Supp. Table 1), and thus it is
330 unlikely that the NFE cells we have identified would be related to cells of parafollicular
331 origin. Currently, the developmental origin of NFE cells and their role in thyroid gland is

332 unclear. To study the cell-population, transgenic zebrafish reagents driving expression
333 using the *tp63* regulatory region (35) could be utilized in future.
334 To survey the communication between thyrocytes, the functional unit of the
335 thyroid gland, and the other cell-types present in the thyroid gland, we constructed a
336 cellular interaction network (Fig. 5B). The network was built by matching the expression
337 of ligands in the diverse cell-types with the expression of receptor in the thyrocytes
338 (Supp. Table 2) (19). Based on literature survey, we manually identified multiple
339 interacting genes that have been implicated in thyroid diseases. For instance, the ligand
340 Decorin (*dcn*) is expressed by the stromal cells and musculature, with its receptor, MET,
341 present on thyrocytes (Fig. 5C). Decorin, a secreted proteoglycan, is considered a
342 "guardian from the matrix" (36), as it is an antagonist of growth factor signaling.
343 Importantly, Decorin expression has been reported to be downregulated in thyroid
344 cancer samples (37). Thus, stromal cells could modulate Decorin to control thyrocytes
345 growth. Further, interactions for CYR61 (associated with Graves' Disease (38)), LRP2 /
346 Megalin (involved in thyroglobulin processing (39)) and NRP2 (associated with thyroid
347 cancer metastasis (40)) were identified (Fig. 3C). The hypothesis generated by the
348 theoretical ligand-receptor interaction network can be tested *in vivo* in zebrafish or *in*
349 *vitro* by manipulation of thyrocytes in thyroid organoid models (41) to gain valuable
350 insight into thyroid gland homeostasis.
351 The current atlas is restricted to healthy adolescent and adult thyroid gland. The
352 two stages represent a period of slow growth in zebrafish. Thus, genes driving cellular
353 proliferation might be repressed at these stages. Additionally, the low number of cells
354 per cluster obtained at each stage restricts an in-depth analysis of the transcriptional

355 difference with age. In future, it would be of interest to extend the atlas by increasing
356 cell numbers and by including single-cell transcriptomics from embryonic and old fish,
357 providing a comprehensive resource for development, homeostasis and aging of the
358 thyroid gland. It would be of further interest to profile zebrafish models of thyroid
359 disorder (42, 43) to understand the cellular and molecular changes underlying organ
360 dysfunction. Combined with the power of CRISPR/Cas9 based screen that we have
361 established for the thyroid gland (27), this resource will provide a roadmap for the
362 functional elucidation of cell type specific programs during thyroid gland growth and
363 homeostasis.

364 In summary, our work provides the first molecular map of the zebrafish thyroid
365 gland at cellular resolution. The atlas contains the molecular characterization of the
366 thyroid gland stromal population, identification of non-follicular epithelial cells, and
367 demonstrate the transcriptional heterogeneity among zebrafish thyrocytes. Further, by
368 constructing cell-cell communication network, the atlas provides clues into tissue
369 dynamics present within the organ. Finally, the dataset has been made available for
370 browsing via an interactive webtool (https://sumeet.shinyapps.io/zfthyroid/). We hope
371 that our efforts will expand the understanding of thyrocytes beyond a nominally
372 homogenous endocrine cell population; providing a complex picture of the diversity in
373 thyrocyte identity and function.
374

375 **Methods**

376 **Zebrafish strains and husbandry**

377 Wild-type or transgenic zebrafish of the outbred AB, WIK, or a hybrid WIK/AB strain
378 were used in all experiments. Zebrafish were raised under standard conditions at 28 °C.
379 Animals were chosen at random for all experiments. Published transgenic strains used
380 in this study were *Tg(tg:nls-mVenus-T2A-NTR)* (7), *Tg(tg:nls-EGFP)* (27),
381 *Tg(kdrl:GFP)*^{la116} (44), *Tg(mpeg1.1:mCherry)*^{gl23} (45) and *Tg(col1a2:LOXP-mCherry-*
382 *NTR)*^{cn11} (referred as *Tg(Col1a2:mCherry)*) (14). Experiments with *Tg(tg:nls-mVenus-*
383 *T2A-NTR)* were conducted in accordance with the Animal Welfare Act and with
384 permission of the Landesdirektion Sachsen, Germany (DD24-5131/346/11, DD24-
385 5131/346/12, DD24.1-5131/476/2, TVV21/2018 and all corresponding amendments).
386 Zebrafish husbandry and experiments with all other transgenic lines was performed
387 under standard conditions in accordance with institutional (Université Libre de Bruxelles
388 (ULB)) and national ethical and animal welfare guidelines and regulation, which were
389 approved by the ethical committee for animal welfare (CEBEA) from the Université Libre
390 de Bruxelles (protocols 578N-579N).

391 **Dissection of the zebrafish thyroid gland**

392 The dissection of thyroid gland in zebrafish was performed by using the ventral aorta as
393 a reference (Fig. 1A, B). In zebrafish, the thyroid follicles sit loosely in soft tissue around
394 the ventral aorta. Ventral aorta connects to the outflow tract that further joins with the
395 heart ventricle. During dissociation, cells connected to the ventral aorta, including parts

396 of zebrafish gills (Fig. 1C) were kept intact to avoid injuring the organ during
397 dissociation.
398 In detail, zebrafish were euthanized in 0.2% Tricaine (MS-222, Sigma E10521)
399 solution. Using fine forceps, the lower jaw was separated from the upper jaw and
400 disconnected from the gut by pinching near the gills. The dissected tissue was carefully
401 cleaned by removing muscle, skin, pectoral fin and lateral cartilages of the lower jaw.
402 The cleaned tissue when observed from the ventral side under brightfield clearly shows
403 the ventral aorta as a thick pink blood vessel extending from the heart towards the
404 basibranchial cartilage (Fig. 1A). Next, the surrounding gills are pinched off using fine
405 forceps, taking care to keep the ventral aorta intact (Fig. 1B). This leaves small parts of
406 gills connected to the ventral aorta (Fig. 1C). Lastly, the ventral aorta is disconnected
407 from the outflow tract by pinching with fine forceps (dashed line in Fig. 1B).

408 **Single cell suspension of zebrafish thyroid gland**

409 Single cell suspension of zebrafish thyroid gland was performed by adapting the cell
410 dissociation protocol outlined in Singh et al., Scientific Reports, 2018 (46). In brief, the
411 thyroid gland was collected and dissociated into single cells by incubation in TrypLE
412 (ThermoFisher, 12563029) with 0.1% Pluronic F-68 (ThermoFisher, 24040032) at 37 °C
413 in a benchtop shaker set at 450 rpm for 45 min. Following dissociation, TrypLE was
414 inactivated with 10% FBS, and the cells pelleted by centrifugation at 500g for 10min at 4
415 °C. The supernatant was carefully discarded and the pellet re-suspended in 500 uL of
416 HBSS (without Ca, Mg) + 0.1% Pluronic F-68. To remove debris, the solution was
417 passed over a 30 µm cell filter (Miltenyi Biotec, 130-041-407). To remove dead cells,
418 calcein violet (ThermoFisher, C34858) was added at a final concentration of 1 µM and

419 the cell suspension incubated at room temperature for 20 minutes. The single cell
420 preparation was sorted with the appropriate gates, including excitation with UV (405 nm)
421 laser for identification of alive cells (calcein+) (Fig. 1F – G). FACS was performed
422 through 100 µm nozzle.

423 **Single-cell profiling of the zebrafish thyoid gland**

424 For single-cell RNA-seq of the zebrafish thyroid gland using the 10X Genomics
425 platform, cell suspension was prepared as mentioned above from the thyroid glands of
426 six 2 month post fertilization and six 8 month post-fertilization *Tg(tg:nls-mVenus-T2A-*
427 *NTR)* animals. The cell suspension was adjusted with Hanks' Balanced Salt Solution
428 (without calcium and magnesium) to a density of 800cells/µl, and diluted with nuclease-
429 free water according to the manufacturer's instructions to yield 12,000 cells.
430 Subsequently, the cells were carefully mixed with reverse transcription mix before
431 loading the cells on the 10X Genomics Chromium system (10). After the gel emulsion
432 bead suspension underwent the reverse transcription reaction, emulsion was broken
433 and DNA purified using Silane beads. The complementary DNA was amplified with 10
434 cycles, following the guidelines of the 10x Genomics user manual. The 10X Genomics
435 single cell RNA-seq library preparation—involving fragmentation, dA tailing, adapter
436 ligation and indexing PCR—was performed based on the manufacturer's protocol. After
437 quantification, the libraries were sequenced on an Illumina NextSeq 550 machine using
438 a HighOutput flowcell in paired-end mode (R1: 26 cycles; I1: 8 cycles; R2: 57 cycles),
439 thus generating ~45mio fragments. The raw sequencing data were then processed with
440 the 'count' command of the Cell Ranger software (v.2.1.0) provided by 10x Genomics
441 with the option '–expect-cells' set to 10,000 (all other options were used as per default).

442 To build the reference for Cell Ranger, zebrafish genome (GRCz10) as well as gene
443 annotation (Ensembl 87) were downloaded from Ensembl and the annotation was
444 filtered with the 'mkgtf ' command of Cell Ranger (options: '–
445 attribute=gene_biotype:protein_coding– attribute=gene_biotype:lincRNA –
446 attribute=gene_biotype:antisense'). Genome sequence and filtered annotation were
447 then used as input to the 'mkref ' command of Cell Ranger to build the appropriate Cell
448 Ranger Reference.

449 **Analysis of single-cell RNA-Seq. of the zebrafish thyroid gland**

450 The raw data generated from 10x Chromium pipeline was clustered using Seurat 2.3.4
451 (47) using the recommended analysis pipeline. Briefly, the raw data as UMI-counts was
452 log-normalized, regressed to remove the effect of library size and mitochondrial counts,
453 and scaled. Highly variable genes were identified for PCA analysis and graph-based
454 clustering using shared nearest neighbour (SNN). For clustering, the first five principal
455 components (PCs) were utilized as they displayed significant deviation from uniform
456 distribution as accessed by JackStraw analysis. Further, a resolution of 0.3 for SNN was
457 used for clustering. Marker genes identified for each cluster were used to classify the
458 cell-type. The thyrocyte cluster was isolated and sub-clustered using the first three PCs
459 and 0.3 resolution to identify and label sub-populations.

460 **Development of intercellular signaling network**

461 Development of intercellular signaling network for zebrafish was performed as
462 described in Cosacak et al. (48). Briefly, ligands expressed in 20% of a cell-population
463 were identified. A connection between cell-type and thyrocyte was made if the

464 expression of the corresponding receptor was identified in 20% of thyrocytes. The
465 connectome contains secreted and membrane-tethered ligands. For cell-types that do
466 not physically contact the thyrocytes (gills, NFE and stroma), membrane-tethered
467 ligands were manually removed from the connectome.

468 **Background correction for thyrocyte gene expression**

469 Supervised background correction for the thyrocyte population was performed using
470 DecontX (49). As input, normalized data and clustering information from Seurat was
471 used. The method using Bayesian approach to model gene expression as a mixture of
472 expression in the expected cell-population plus background expression accessed from
473 remaining cell-types. Background noise is removed, which likely resembles free mRNA
474 released from injured and ruptured cells. As background correction required clustering
475 information, the corrected data was not utilized for re-clustering to avoid circular use of
476 the data.

477 **Genetic entropy analysis for thyrocyte population**

478 Quantification of genetic entropy was performed using ROGUE (Ratio of Global
479 Unshifted Entropy) (50). As input, raw counts of thyrocytes that passed quality control
480 were used. Default parameters were used for analysis. The algorithm provided a
481 measure of entropy (degree of uncertainty / heterogeneity), along with p-value, within
482 the population.

483 **Data Availability**

484 The raw 10X data, along with tabulated count data are available publicly from GEO
485 under accession number GSE133466. The atlas for online browsing is available at
486 https://sumeet.shinyapps.io/zfthyroid/.

487 **Generation of knock-in *pax2a*$^{pax2a\text{-}T2A\text{-}mKO2}$ zebrafish line**

488 For generation of *pax2a* reporter line, we designed a single-guide RNA (sgRNA) targeting
489 the STOP codon of the *pax2a* coding sequence (GCTGCGATGGTAACTAGTGG). We
490 then generated a donor construct in which the sequence encoding for the monomeric
491 Kusabira orange (mKO2) protein was fused to a viral 2A peptide linker. This reporter
492 cassette was flanked by left (1000bp) and right (2000bp) homology arms of the *pax2a*
493 genomic DNA region around the stop codon therefore preventing the sgRNA from cutting
494 the donor construct. sgRNA design, production and validation were done as previously
495 described (27, 51). Wild-type embryos were injected with 3 nL of the injection mix
496 containing the sgRNA (final concentration 80 ng/μL), the donor construct (final
497 concentration 7.5 ng/μL), the protein Cas9 (recombinant cas protein from *S. pyogenes*
498 PNA Bio CP01, final concentration 100 ng/μL) and KCL (final concentration 200 mM).
499 Upon homologous recombination of this reporter construct in the endogenous locus,
500 *pax2a*-expressing cells were fluorescently labelled by mKO2. This *pax2a*$^{pax2a\text{-}T2A\text{-}mKO2}$ line
501 is referenced as *pax2a*^{mKO2} in the text.

502 **Generation of *pax2a* crispants**

503 Somatic mutagenesis of pax2a gene was carried out exactly as mentioned in Trubiroha
504 et al., Scientific Reports, 2018 (27). Briefly, sgRNA targeting the exon 2 of *pax2a* was

505 generated as described in the publication. Following the strategy described in the
506 publication, Cas9 protein along with sgRNA was injected in one-cell stage of zebrafish
507 embryos for disruption of *pax2a* gene. Non-injected animals were used as controls.

508 **Tissue collection**

509 To facilitate confocal imaging of the thyroid gland, the organ was manually dissected
510 from fish as previously described and fixed. Fish were killed in Tricaine followed by
511 dissection of the gland, which was fixed by immersion in 4% paraformaldehyde (PFA) +
512 1% Triton-X overnight at 4 °C. The gland was washed 2 – 3 times in PBS to remove
513 PFA before proceeding.

514 **Quantification of proportion of thyrocytes within the dissected tissue**

515 To quantify the proportion of thyrocytes within the dissected tissue, the gland was
516 dissected from *Tg(tg:nls-EGFP)* animals and fixed as described above. The fixed tissue
517 was permeabilized by three washes with 1% PBT (1x PBS + 1 % Triton-X-100). Nuclei
518 were stained by immersing the tissue in 1 µg / ml Hoechst prepared in 1x PBS for two
519 hours at room temperature. The tissue was immersed in 30% sucrose solution overnight
520 at 4 °C, embedded in Tissue Freezing Medium (Leica 14020108926) and frozen at -80
521 °C. Thin sections (8 µm) were obtained using cryostat (Leica CM3050 S), collected on
522 frosted glass slides (Thermo Scientific 12362098) and covered with glass coverslip of
523 #1 thickness (Carl Roth GmbH NK79.1) using mounting media (Dako S3023). The
524 sections were imaged on Zeiss LSM 780 confocal microscope. Confocal images were
525 analyzed in Fiji using the following step: threshold using 'IsoData' to distinguish signal
526 from background, 'watershed' transformation to separate joined nuclei and 'measure'

527 function to obtain nuclei count. With this, the green channel (number of thyrocyte nuclei)
528 and blue channel (total number of nuclei) was measured for seven transverse sections
529 obtained from three animals. Percentage was calculated by taking the ratio of thyrocyte
530 nuclei to total nuclei.

531 **Immunofluorescence and image acquisition**

532 Whole-mount immunofluorescence was performed on thyroid gland collected as
533 described above. The collected samples were permeabilized in 1% PBT (Triton-X-100)
534 and blocked in 4% PBTB (BSA). Primary and secondary antibody stainings were
535 performed overnight at 4 °C. Primary antibodies used in this study were anti-PAX2A
536 (rabbit, Genetex GTX128127) at 1:250, anti-EGFP (chicken, Abcam ab13970) at 1:1000,
537 anti-E-Cadherin (mouse, BD bioscience cat 610181) at 1:200, anti-monomeric Kusabira-
538 Orange 2 (mouse, MBL amalgaam M-168-3M) at 1:200, anti-monomeric Kusabira-
539 Orange 2 (rabbit, MBL amalgam PM051M) at 1:250 and anti-p63 (mouse, Santa Cruz
540 Biotechnology 4A4) at 1:200. Secondary antibodies at 1:250 dilutions used in this study
541 were Alexa Fluor 488 anti-chicken (Jackson ImmunoResearch laboratories 703-545-
542 155), Alexa Fluor 647 anti-rabbit (Jackson ImmunoResearch laboratories 711-605-152),
543 Alexa Fluor 647 anti-mouse (Jackson ImmunoResearch laboratories 715-605-150),
544 Cy™3-conjugated anti-rabbit (Jackson ImmunoResearch laboratories 711-165-152) and
545 Cy™3-conjugated anti-mouse (Jackson ImmunoResearch laboratories 715-165-150).
546 When needed nuclei were staining using DAPI at a 1:1000 dilution. Samples were
547 mounted in NuSieve™ GTG™ Agarose (Lonza cat50080) and imaged on a glass bottom
548 FluoroDish™ (WPI FD3510-100) using a Zeiss LSM 780 confocal microscope or Leica

549 DMI 6000b microscope. ImageJ was used to add scale bars and PowerPoint was used
550 for adding arrows and labels.

551 **FACS-based reporter analysis**

552 For analysing the levels of *pax2a*^{mKO2} by FACS, single-cell suspension from the thyroid
553 gland of 5 mpf *Tg(tg:nls-GFP)*; *pax2a*^{mKO2} animals was prepared as described earlier
554 and stained with 1 μM calcein violet (ThermoFisher, C34858). Cells were sorted and
555 analyzed using FACS-Aria II (BD Bioscience). Thyrocytes were selected by gating for
556 calcein+ GFP+ population, and mKO2 expression level recorded for analysis.

557 **Gene Ontology (GO) Analysis**

558 Gene ontology (GO) analysis was performed using DAVID (52). The list of genes was
559 uploaded on the web browser of DAVID and statistically significant (p-value < 0.05) GO
560 terms were identified using default parameters.

561 **Statistical analysis**

562 Statistical analysis was performed using R. No animals were excluded from analysis.
563 Blinding was not performed during analysis. Analysis of normal distribution was not
564 performed.
565

566 **References**

567 1. Taylor PN, Albrecht D, Scholz A, Gutierrez-Buey G, Lazarus JH, Dayan CM,
568 Okosieme OE 2018 Global epidemiology of hyperthyroidism and hypothyroidism.
569 Nat Rev Endocrinol **14**:301–316.
570 2. Brix K, Lemansky P, Herzog V 1996 Evidence for extracellularly acting cathepsins
571 mediating thyroid hormone liberation in thyroid epithelial cells. Endocrinology
572 **137**:1963–74.
573 3. Alt B, Reibe S, Feitosa NM, Elsalini OA, Wendl T, Rohr KB 2006 Analysis of origin
574 and growth of the thyroid gland in zebrafish. Dev Dyn **235**:1872–1883.
575 4. Linehan SA, Martínez-Pomares L, da Silva RP, Gordon S 2001 Endogenous
576 ligands of carbohydrate recognition domains of the mannose receptor in murine
577 macrophages, endothelial cells and secretory cells; potential relevance to
578 inflammation and immunity. Eur J Immunol **31**:1857–66.
579 5. Nonidez JF 1931 Innervation of the thyroid gland. II. Origin and course of the
580 thyroid nerves in the dog. Am J Anat **48**:299–329.
581 6. Svensson V, Vento-Tormo R, Teichmann SA 2018 Exponential scaling of single-
582 cell RNA-seq in the past decade. Nat Protoc **13**:599–604.
583 7. McMenamin SK, Bain EJ, McCann AE, Patterson LB, Eom DS, Waller ZP, Hamill
584 JC, Kuhlman JA, Eisen JS, Parichy DM 2014 Thyroid hormone-dependent adult
585 pigment cell lineage and pattern in zebrafish. Science (80-) **345**:1358–1361.
586 8. AlJanahi AA, Danielsen M, Dunbar CE 2018 An Introduction to the Analysis of
587 Single-Cell RNA-Sequencing Data. Mol Ther Methods Clin Dev **10**:189–196.
588 9. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I,

589 Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR,
590 Shalek AK, Regev A, McCarroll SA 2015 Highly Parallel Genome-wide
591 Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell **161**:1202–
592 1214.
593 10. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB,
594 Wheeler TD, McDermott GP, Zhu J, Gregory MT, Shuga J, Montesclaros L,
595 Underwood JG, Masquelier DA, Nishimura SY, Schnall-Levin M, Wyatt PW,
596 Hindson CM, Bharadwaj R, Wong A, Ness KD, Beppu LW, Deeg HJ, McFarland
597 C, Loeb KR, Valente WJ, Ericson NG, Stevens EA, Radich JP, Mikkelsen TS,
598 Hindson BJ, Bielas JH 2017 Massively parallel digital transcriptional profiling of
599 single cells. Nat Commun **8**:14049.
600 11. Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, Taipale J
601 2011 Counting absolute numbers of molecules using unique molecular identifiers.
602 Nat Methods **9**:72–4.
603 12. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, Lönnerberg P,
604 Linnarsson S 2014 Quantitative single-cell RNA-seq with unique molecular
605 identifiers. Nat Methods **11**:163–6.
606 13. Maaten L van der, Hinton G 2008 Visualizing data using t-SNE. J Mach Learn
607 Res **9**:2579–2605.
608 14. Sánchez-Iranzo H, Galardi-Castilla M, Sanz-Morejón A, González-Rosa JM,
609 Costa R, Ernst A, Sainz de Aja J, Langa X, Mercader N 2018 Transient fibrosis
610 resolves via fibroblast inactivation in the regenerating zebrafish heart. Proc Natl
611 Acad Sci U S A **115**:4188–4193.

612 15. Denton CP, Zheng B, Shiwen X, Zhang Z, Bou-Gharios G, Eberspaecher H, Black
613 CM, de Crombrugghe B 2001 Activation of a fibroblast-specific enhancer of the
614 proalpha2(I) collagen gene in tight-skin mice. Arthritis Rheum **44**:712–22.
615 16. Lisse TS, Middleton LJ, Pellegrini AD, Martin PB, Spaulding EL, Lopes O, Brochu
616 EA, Carter E V, Waldron A, Rieger S 2016 Paclitaxel-induced epithelial damage
617 and ectopic MMP-13 expression promotes neurotoxicity in zebrafish. Proc Natl
618 Acad Sci U S A **113**:E2189-98.
619 17. Reischauer S, Levesque MP, Nüsslein-Volhard C, Sonawane M 2009 Lgl2
620 executes its function as a tumor suppressor by regulating ErbB signaling in the
621 zebrafish epidermis. PLoS Genet **5**:e1000720.
622 18. Barbieri CE, Pietenpol JA 2006 p63 and epithelial biology. Exp Cell Res **312**:695–
623 706.
624 19. Ramilowski JA, Goldberg T, Harshbarger J, Kloppman E, Lizio M, Satagopam VP,
625 Itoh M, Kawaji H, Carninci P, Rost B, Forrest ARR 2015 A draft network of ligand-
626 receptor-mediated multicellular signalling in human. Nat Commun **6**:7866.
627 20. Brix K, Linke M, Tepel C, Herzog V 2001 Cysteine proteinases mediate
628 extracellular prohormone processing in the thyroid. Biol Chem **382**:717–25.
629 21. Linke M 2002 Trafficking of lysosomal cathepsin B--green fluorescent protein to
630 the surface of thyroid epithelial cells involves the endosomal/lysosomal
631 compartment. J Cell Sci **115**:4877–4889.
632 22. Wendl T, Lun K, Mione M, Favor J, Brand M, Wilson SW, Rohr KB 2002 Pax2.1 is
633 required for the development of thyroid follicles in zebrafish. Development
634 **129**:3751–60.

635 23. Porazzi P, Calebiro D, Benato F, Tiso N, Persani L 2009 Thyroid gland
636 development and function in the zebrafish model. Mol Cell Endocrinol **312**:14–23.
637 24. Pfeffer PL, Gerster T, Lun K, Brand M, Busslinger M 1998 Characterization of
638 three novel members of the zebrafish Pax2/5/8 family: dependency of Pax5 and
639 Pax8 expression on the Pax2.1 (noi) function. Development **125**:3063–74.
640 25. Mansouri A, Chowdhury K, Gruss P 1998 Follicular cells of the thyroid gland
641 require Pax8 gene function. Nat Genet **19**:87–90.
642 26. Kesavan G, Chekuru A, Machate A, Brand M 2017 CRISPR/Cas9-Mediated
643 Zebrafish Knock-in as a Novel Strategy to Study Midbrain-Hindbrain Boundary
644 Development. Front Neuroanat **11**:52.
645 27. Trubiroha A, Gillotay P, Giusti N, Gacquer D, Libert F, Lefort A, Haerlingen B, De
646 Deken X, Opitz R, Costagliola S 2018 A Rapid CRISPR/Cas-based Mutagenesis
647 Assay in Zebrafish for Identification of Genes Involved in Thyroid Morphogenesis
648 and Function. Sci Rep **8**:5647.
649 28. Liang S, Johansson E, Barila G, Altschuler DL, Fagman H, Nilsson M 2018 A
650 branching morphogenesis program governs embryonic growth of the thyroid
651 gland. Development **145**:dev146829.
652 29. Guzman A, Ramos-Balderas JL, Carrillo-Rosas S, Maldonado E 2013 A stem cell
653 proliferation burst forms new layers of P63 expressing suprabasal cells during
654 zebrafish postembryonic epidermal development. Biol Open **2**:1179–1186.
655 30. Yang A, Schweitzer R, Sun D, Kaghad M, Walker N, Bronson RT, Tabin C,
656 Sharpe A, Caput D, Crum C, McKeon F 1999 p63 is essential for regenerative
657 proliferation in limb, craniofacial and epithelial development. Nature **398**:714–8.

658 31. Baber EC 1876 XXI. Contributions to the minute anatomy of the thyroid gland of
659 the dog. Philos Trans R Soc London **166**:557–568.
660 32. Getzowa S 1907 Über die Glandula parathyreodeaa, intrathyreoideale Zellhaufen
661 derselben und Reste des postbranchialen Körpers. Virchows Arch Pathol Anat
662 Physiol Klin Med **188**:181–235.
663 33. Harach HR 1988 Solid cell nests of the thyroid. J Pathol **155**:191–200.
664 34. Ríos Moreno MJ, Galera-Ruiz H, De Miguel M, López MIC, Illanes M, Galera-
665 Davidson H 2011 Inmunohistochemical profile of solid cell nest of thyroid gland.
666 Endocr Pathol **22**:35–9.
667 35. Rasmussen JP, Sack GS, Martin SM, Sagasti A 2015 Vertebrate epidermal cells
668 are broad-specificity phagocytes that clear sensory axon debris. J Neurosci
669 **35**:559–70.
670 36. Neill T, Schaefer L, Iozzo R V 2012 Decorin: a guardian from the matrix. Am J
671 Pathol **181**:380–7.
672 37. Martínez-Aguilar J, Clifton-Bligh R, Molloy MP 2016 Proteomics of thyroid
673 tumours provides new insights into their molecular composition and changes
674 associated with malignancy. Sci Rep **6**:23660.
675 38. Planck T, Shahida B, Sjögren M, Groop L, Hallengren B, Lantz M 2014
676 Association of BTG2, CYR61, ZFP36, and SCD gene polymorphisms with
677 Graves' disease and ophthalmopathy. Thyroid **24**:1156–61.
678 39. Marinò M, McCluskey RT 2000 Megalin-mediated transcytosis of thyroglobulin by
679 thyroid cells is a calmodulin-dependent process. Thyroid **10**:461–9.
680 40. Tu D-G, Chang W-W, Jan M-S, Tu C-W, Lu Y-C, Tai C-K 2016 Promotion of

681 metastasis of thyroid cancer cells via NRP-2-mediated induction. Oncol Lett
682 **12**:4224–4230.
683 41. Antonica F, Kasprzyk DF, Opitz R, Iacovino M, Liao X-H, Dumitrescu AM, Refetoff
684 S, Peremans K, Manto M, Kyba M, Costagliola S 2012 Generation of functional
685 thyroid from embryonic stem cells. Nature **491**:66–71.
686 42. Chopra K, Ishibashi S, Amaya E 2019 Zebrafish duox mutations provide a model
687 for human congenital hypothyroidism. Biol Open **8**.
688 43. Anelli V, Villefranc JA, Chhangawala S, Martinez-McFaline R, Riva E, Nguyen A,
689 Verma A, Bareja R, Chen Z, Scognamiglio T, Elemento O, Houvras Y 2017
690 Oncogenic BRAF disrupts thyroid morphogenesis and function via twist
691 expression. Elife **6**.
692 44. Choi J, Dong L, Ahn J, Dao D, Hammerschmidt M, Chen J-N 2007 FoxH1
693 negatively modulates flk1 gene expression and vascular formation in zebrafish.
694 Dev Biol **304**:735–744.
695 45. Ellett F, Pase L, Hayman JW, Andrianopoulos A, Lieschke GJ 2011 mpeg1
696 promoter transgenes direct macrophage-lineage expression in zebrafish. Blood
697 **117**:e49-56.
698 46. Singh SP, Janjuha S, Chaudhuri S, Reinhardt S, Kränkel A, Dietz S, Eugster A,
699 Bilgin H, Korkmaz S, Zararsız G, Ninov N, Reid JE 2018 Machine learning based
700 classification of cells into chronological stages using single-cell transcriptomics.
701 Sci Rep **8**:17156.
702 47. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R 2018 Integrating single-cell
703 transcriptomic data across different conditions, technologies, and species. Nat

704 Biotechnol doi: 10.1101/164889.

705 48. Cosacak MI, Bhattarai P, Reinhardt S, Petzold A, Dahl A, Zhang Y, Kizil C 2019
706 Single-Cell Transcriptomics Analyses of Neural Stem Cell Heterogeneity and
707 Contextual Plasticity in a Zebrafish Brain Model of Amyloid Toxicity. Cell Rep
708 **27**:1307-1318.e3.

709 49. Yang S, Corbett SE, Koga Y, Wang Z, Johnson WE, Yajima M, Campbell JD
710 2020 Decontamination of ambient RNA in single-cell RNA-seq with DecontX.
711 Genome Biol **21**:57.

712 50. Liu B, Li C, Li Z, Ren X, Zhang Z 2019 ROGUE: an entropy-based universal
713 metric for assessing the purity of single cell population. bioRxiv 819581.

714 51. Varshney GK, Pei W, LaFave MC, Idol J, Xu L, Gallardo V, Carrington B, Bishop
715 K, Jones M, Li M, Harper U, Huang SC, Prakash A, Chen W, Sood R, Ledin J,
716 Burgess SM 2015 High-throughput gene targeting and phenotyping in zebrafish
717 using CRISPR/Cas9. Genome Res **25**:1030–42.

718 52. Huang D, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R,
719 Baseler MW, Lane HC, Lempicki RA 2007 The DAVID Gene Functional
720 Classification Tool: a novel biological module-centric algorithm to functionally
721 analyze large gene lists. Genome Biol **8**:R183.

722

723

724 **Acknowledgements**

725 We thank members of the Costagliola and Singh lab for comments on the manuscript,
726 members of Center for Regenerative Therapies Dresden (CRTD) fish, FACS and
727 sequencing facility, and members of IRIBHM fish facility for technical assistance. We
728 thank J.-M. Vanderwinden from the Light Microscopy Facility and Christine Dubois from
729 the FACS facility for technical assistance at ULB. We are grateful to Priyanka Oberoi for
730 illustrations. P.G is Fund for Research in the Industry and the Agriculture (FRIA)
731 Research fellow; M.S. is FNRS Research Fellow (34985615 - THYSCEFA); S.C. is
732 FNRS Senior Research Associate. V.D. acknowledges grants from the Fond Naets
733 (J1813300), the Fondation Contre le Cancer (2016-093) and FNRS (EQP/OL
734 U.N019.19, J006120F). Work by M.B., C.L. and G.K. was supported by grants to M.B.
735 from the Deutsche Forschungsgemeinschaft and European Union (European Research
736 Council AdG Zf-BrainReg). Work by S.P.S. was supported by MISU funding from the
737 FNRS (34772792 – SCHISM). This work was supported by grants from the Belgian
738 National Fund for Scientific Research (FNRS) (FRSM 3-4598-12; CDR-J.0145.16, GEQ
739 U.G030.19), the Fonds d'Encouragement à la Recherche de l'Université Libre de
740 Bruxelles (FER-ULB).

741 **Author contribution**

742 S.P.S. conceptualized the project. N.N., G.K., C.L. and M.B. provided reagents and
743 animals for single-cell RNA-Sequencing. S.P.S., S.R., A.K., J.B., and A.P. performed
744 the single-cell RNA-Sequencing. S.P.S., S.E.E., V.D., S.C. analysed and interpreted the
745 data. S.P.S. developed the online browser. P.G. and B.H. generated the *pax2a* knock-in

line. P.G. and M.S. analysed the *pax2a* reporter line, M.P.M. and I.G.S. collected immunofluorescence images. S.P.S. wrote the first draft and S.C., P.G., S.E.E edited the manuscript. S.P.S. and S.C. acquired funding for the project. All authors read and approved the final manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

752 **Figure Legends**

753 **Figure 1: Isolation of zebrafish thyroid gland**

754 **(A – B)** A brightfield image showing the zebrafish thyroid gland along with surrounding
755 tissue. The thyroid follicles reside in the soft tissue surrounding the ventral aorta, which
756 extends from the outflow tract of the heart into the gills towards the basibranchial
757 cartilage in the lower jaw. The thyroid follicular cells, or thyrocytes, are labeled in green
758 in the *Tg(tg:nls-mVenus-NTR)* transgenic line (B'). **(C)** Maximum intensity projection of
759 3D confocal stack obtained from the dissected thyroid gland labeled with DAPI. **(D)**
760 Confocal scan of a transverse section across the dissected thyroid gland from *Tg(tg:nls-*
761 *EGFP)* animal at 3 mpf. Sections were stained with DAPI to visualize cells surrounding
762 thyroid follicles. **(E)** Boxplot depicting the proportion of thyrocytes present in transverse
763 sections obtained from three *Tg(tg:nls-EGFP)* animals at 3 mpf. Each dot represents a
764 transverse section. **(F – G)** Representative FACS plot of single cells from *Tg(tg:nls-*
765 *mVenus-T2A-NTR)* animals at 2 mpf (F) and 8 mpf (G). Calcein (Pacific Blue) labels live
766 cells, while green fluorescence (FITC) labels thyrocytes. Percentage values represent
767 proportion of calcein+ thyrocytes within total calcein+ cells.

768 **Figure 2: Single-cell RNA-Seq. of the zebrafish thyroid gland**

769 **(A)** A t-SNE plot displaying the 6249 single-cells profiled in the zebrafish thyroid gland
770 atlas. The colors represent cell clusters denoting a specific cell-type. **(D – F)** Cluster #1
771 represents the thyrocytes that express *tg*, *slc5a5* (NIS) and *tpo*.

772 **Figure 3: Gene expression signature of the different cell-types in the zebrafish**
773 **thyroid gland**

774 **(A)** Heatmap depicting five genes specifically expressed in each one of the seven
775 clusters of the zebrafish thyroid gland atlas. **(B – G)** t-SNE plots overlaid with the
776 expression of a gene specific to each of the cluster. The endothelium cluster (cluster #2)
777 is a mix of blood vessels and lymphatic vessels (see Supp. Fig. 4), while the immune
778 cell cluster (cluster #4) is a mix of macrophages, neutrophils and lymphocytes (see
779 Supp. Fig. 5).

780 **Figure 4: Immunofluorescence-based visualization of cell-types surrounding**
781 **zebrafish thyroid follicles**

782 Images show immunofluorescence labeling of thyroid gland from adult zebrafish.
783 Transverse sections were utilized for imaging. The organ was isolated from tissue-
784 specific transgenic lines to allow marking of a particular cell-type adjacent to the thyroid
785 follicle. Blood vessels were marked using *Tg(kdrl:EGFP)* **(A)**, macrophages using
786 *Tg(mpeg1.1:mCherry)* **(B)** and stroma using *Tg(col1a2:mCherry)* **(C)**. Thyrocytes were
787 labeled with *pax2a*^{mKO2} expression in (A) (described in Fig. 7) and *Tg(tg:nls-EGFP)*
788 expression in (B – C). **(D)** NFE was labeled using antibody against TP63 in sections of
789 the thyroid gland isolated from *Tg(tg:nls-EGFP)* animals. Gills are marked based on
790 their morphological appearance. DAPI labels nuclei. Scale bars: 10 μm (A - B), 50 μm
791 (C – D).

792 **Figure 5: Connectome of the zebrafish thyroid gland identifies a dense**
793 **intercellular signaling network**

794 **(A)** To build a connectome for the atlas, the ligands expressed specifically in each cell-
795 type were matched with their corresponding receptors in the thyrocytes. **(B)** A highly
796 connected intercellular interaction network is identified by the connectome. The number
797 of ligand-receptor pairs identified between two cell-types is denoted alongside the
798 arrows. For NFE, Gills and Stromal cells, the connectome was restricted to secreted
799 ligands. **(C)** A dotplot depicting examples of paracrine and autocrine signaling in the
800 thyroid gland. The dots represent expression level in the different cell-types of the atlas.

801 **Figure 6: Thyrocytes can be subdivided into two transcriptionally distinct sub-**
802 **populations**

803 **(A**) Unsupervised clustering of the thyrocyte population identifies two sub-populations.
804 **(B)** Heatmap depicting the top ten most differentially expressed genes between the two
805 sub-populations. **(C)** Violin plots depicting the expression levels of *pax2a*, *ctsba* and *tg*
806 in each sub-population. Y-axis represents scaled data. **(D)** Dot plot depicting expression
807 entropy on Y-axis against average gene expression on X-axis for the thyrocyte
808 population. Each dot depicts a gene, with red dots depicting genes that show
809 statistically significant (p-value < 0.05) difference in entropy from expected value.
810 Expected value is represented by black regression line. *pax2a* and *ctsba* are marked on
811 the graph.

812 **Figure 7: *pax2a^{mKO2}* knock-in line faithfully reports *pax2a* expression and knock-**
813 **down**

814 **(A)** Schematic of the knock-in strategy used to generate the *pax2a^{mKO2}* line. Double
815 strand break was induced between the penultimate codon and the STOP codon of *pax2a*
816 gene using CRISPR/Cas9. DNA repair integrates the donor construct at the site of double
817 strand break, resulting in a *pax2a* reporter line. The donor construct contains T2A-mKO2
818 reporter cassette flanked by left homology (LH) and right homology (RH) arms. **(B)** Whole
819 mount immunofluorescence of 9.5hpf *pax2a*^{mKO2} embryos stained with anti-mKO2
820 antibody (red) and anti-PAX2A antibody (green). Anterior is to the left, and dorsal side is
821 to the top. **(C)** Whole mount immunofluorescence of 55 hpf *pax2a^{mKO2}*; *Tg(tg:nls-EGFP)*
822 stained with PAX2A antibody (PAX2A-Ab) displays an overlap of mKO2 and PAX2A-Ab
823 signal. The otic vesicle (OV), mid-hindbrain barrier (MHB), interneurons (IN) and thyroid
824 gland (THY) is labelled. **(D – F)** Confocal microscopy imaging of a sagittal section of a 55
825 hpf *pax2a^{mKO2}*; *Tg(tg:nls-EGFP)* embryos showing co-localization of mKO2 and *pax2a* in
826 the pronephros (D), thyroid gland (E) and mid-hindbrain barrier (F). In the thyroid gland,
827 mKO2, PAX2A-Ab and thyrocyte-specific GFP (green) show co-localization. Scale bars:
828 100µm (C) and 50µm (D – F). Anterior to the right, white dashed line represents the
829 outline of the embryo. **(G – H)** Snapshots from live imaging of 55 hpf *pax2a*^{mKO2}; *Tg(tg:nls-*
830 *EGFP)* embryos injected with sgRNA targeting *pax2a* coding sequence. The anterior part
831 of a representative control embryo (G) is shown alongside a representative crispant (H).
832 Crispants display a strong reduction of mKO2 fluorescence, as well as an absence of
833 GFP signal suggesting absence of thyroid (THY) tissue.

834 **Figure 8: Immunofluorescence-based validation of thyrocyte heterogeneity**

835 **(A-D)** Analysis of 3 mpf thyroid gland from *pax2a*^{mKO2} zebrafish shows heterogeneity in
836 *pax2a* reporter expression. **(A)** Whole mount confocal imaging of a 3 mpf *pax2a*^{mKO2}
837 thyroid labelled with mKO2 (red), E-cadherin (cyan, not shown in 'A' for clarity reasons)
838 and DAPI (dark blue) for nuclear localisation. **(B – D)** Optical sections of three follicles,
839 with mKO2-Low cells labelled with arrows. E-cadherin (B' – D') and DAPI (B'' – D'')
840 staining shows that absence of mKO2 signal does not correspond to an absence of cells.
841 Anterior to the bottom of the pictures. **(E)** Confocal image of thyroid gland section from
842 *Tg(tg:nls-EGFP)* at 4 mpf stained with PAX2A antibody and DAPI. The dotted region is
843 displayed at high magnification in **(F)**. Arrows marks thyrocytes displaying low PAX2A
844 staining. Notably, PAX2A-Low thyrocytes display *tg*-driven EGFP expression,
845 demonstrating their differentiated status. Scale bars: 250 µm (A), 50 µm (B – E), 10 µm
846 (F).

847 **Figure 9: Flow cytometry-based validation of thyrocyte heterogeneity**

848 Cells from the thyroid gland of 5 mpf *Tg(tg:nls-GFP); pax2a*^{mKO2} animals were labelled
849 with calcein (live cell marker) and analysed using FACS. **(A)** A FACS plot showing calcein
850 on X-axis and GFP on Y-axis. The box encompassing the GFP+ cells represents the
851 thyrocyte population, which was gated for further analysis. **(B)** Histogram showing the
852 distribution of GFP intensity in thyrocytes. **(C)** Histogram showing the distribution of
853 mKO2 intensity in thyrocytes. Thyrocytes were selected by gating for GFP+ population.
854 Horizontal lines indicate the mKO2-Low and mKO2-High expression level, with

855 percentage values representing proportion of thyrocytes with mKO2-Low and mKO2-High

856 expression.

Figure 1

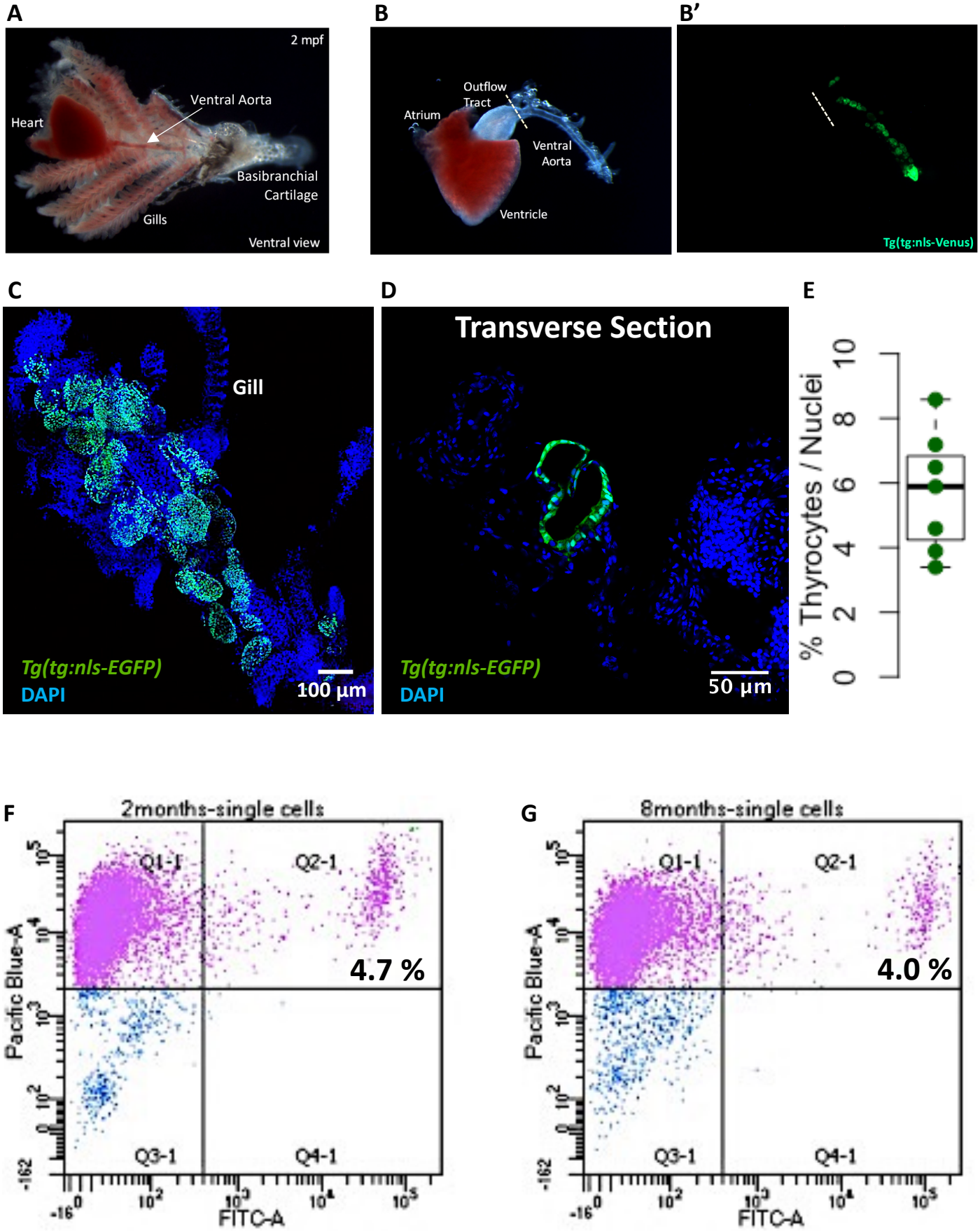

A
2 mpf
Ventral Aorta
Heart
Basibranchial
Cartilage
Gills
Ventral view
B
Outflow
Tract
Atrium
Ventral
Aorta
Ventricle
B'
Tg(tg:nls-Venus)
C
Gill
Tg(tg:nls-EGFP)
DAPI
100 µm
D
Transverse Section
Tg(tg:nls-EGFP)
DAPI
50 µm
E
% Thyrocytes / Nuclei
0
2
4
6
8
10
F
2months-single cells
Q1-1
Q2-1
4.7 %
Q3-1
Q4-1
Pacific Blue-A
FITC-A
G
8months-single cells
Q1-1
Q2-1
4.0 %
Q3-1
Q4-1
Pacific Blue-A
FITC-A

Figure 2

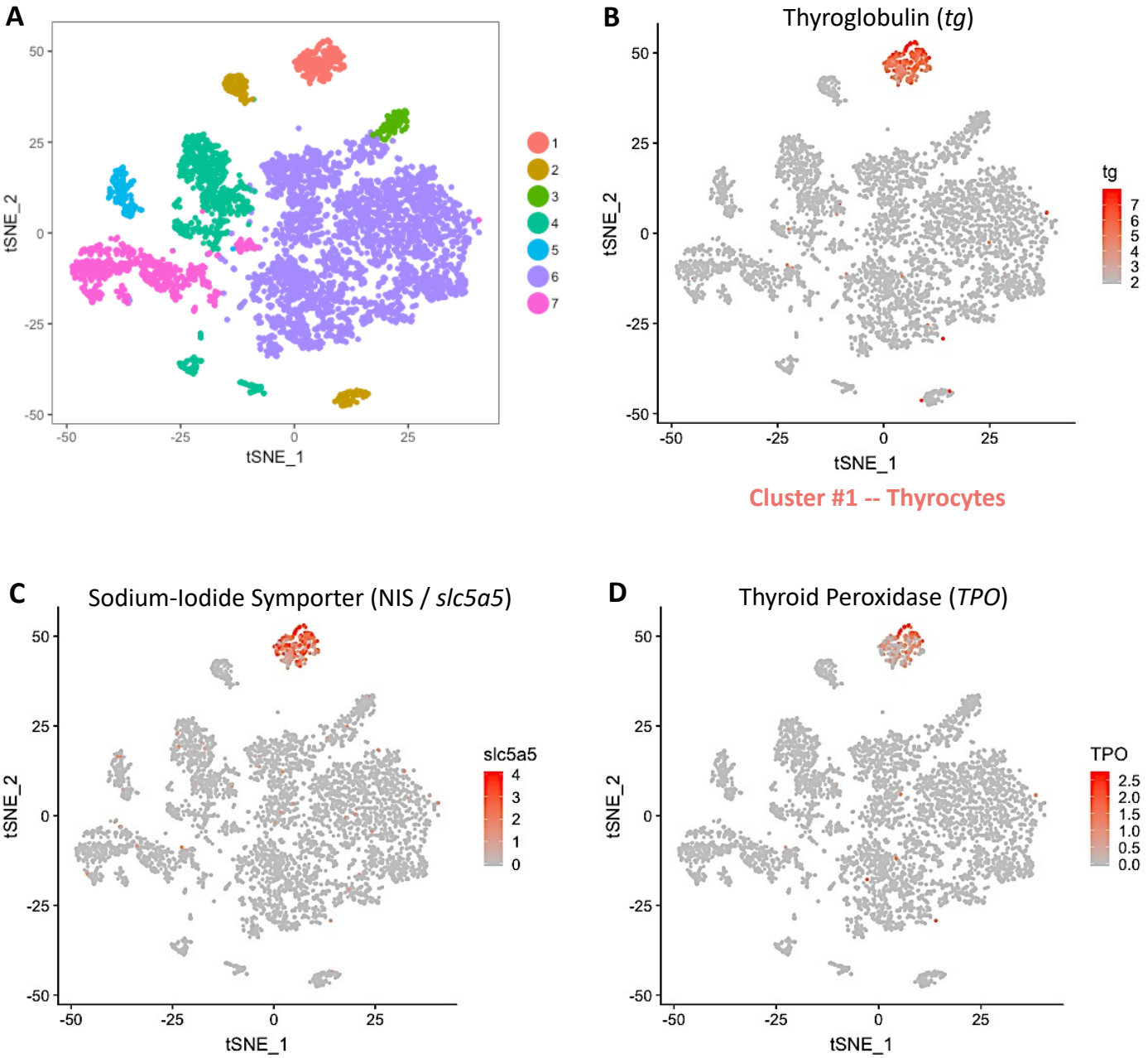
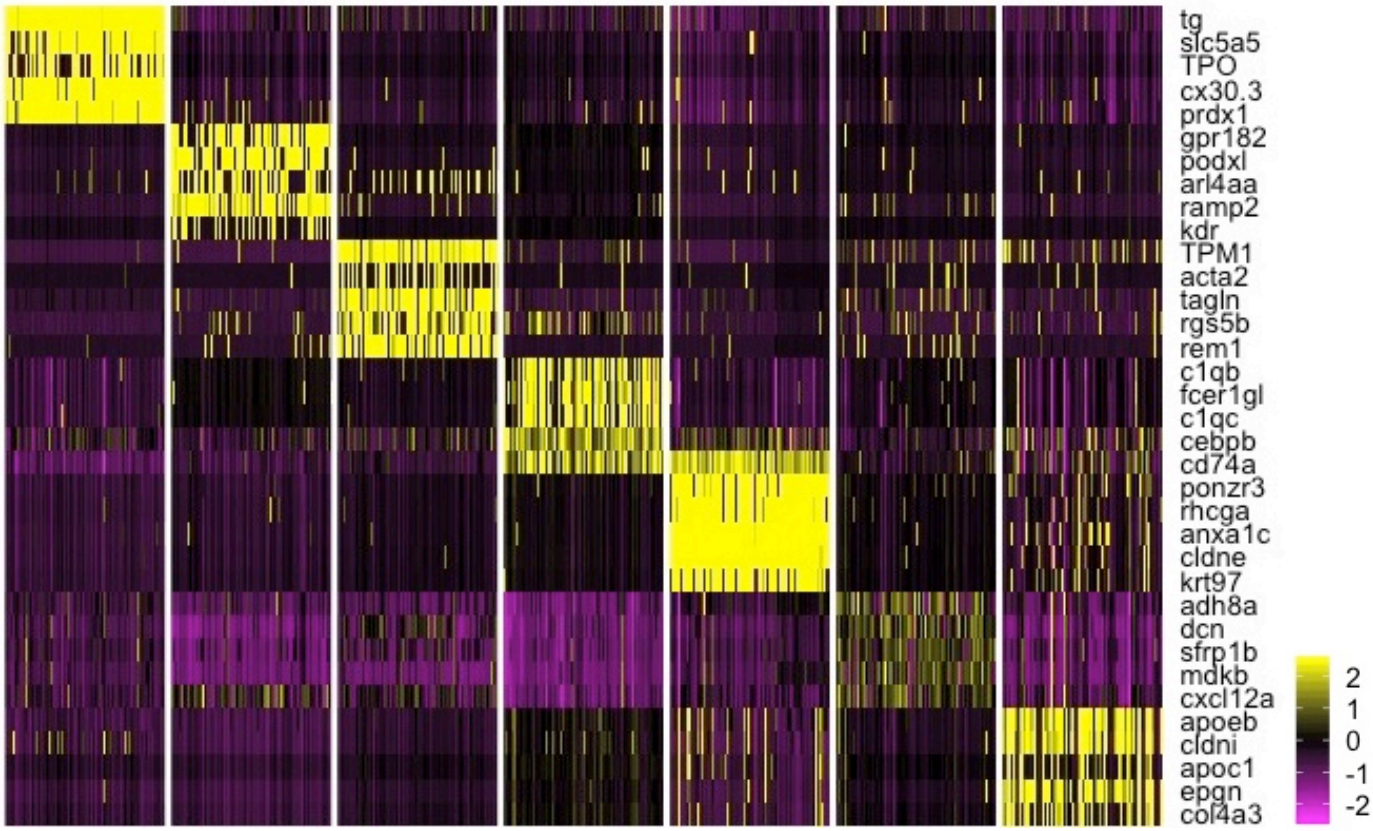
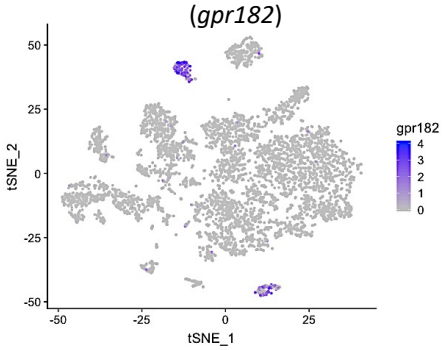

A
tSNE_2
tSNE_1
50
25
0
-25
-50
1
2
3
4
5
6
7
B
Thyroglobulin (*tg*)
tg
7
6
5
4
3
2
Cluster #1 -- Thyrocytes
C
Sodium-Iodide Symporter (NIS / *slc5a5*)
slc5a5
4
3
2
1
0
D
Thyroid Peroxidase (*TPO*)
TPO
2.5
2.0
1.5
1.0
0.5
0.0

Figure 3

A

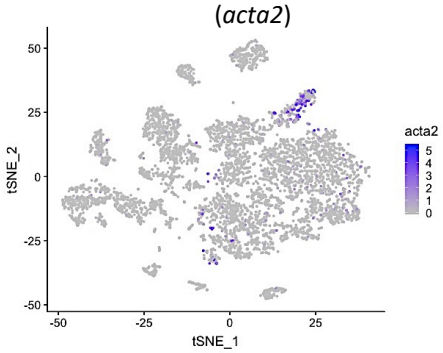


B Adrenomedullin Receptor (*gpr182*)



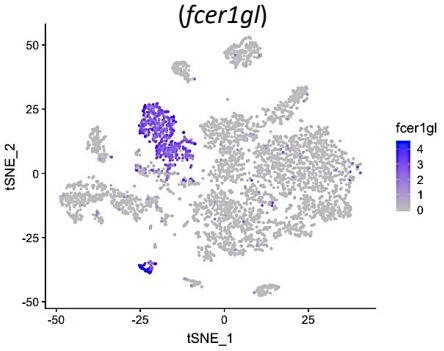
Cluster #2
Endothelium

C Actin Alpha 2, Smooth Muscle (*acta2*)



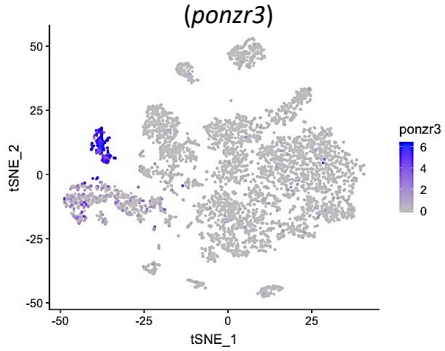
Cluster #3
Musculature

D Fc Fragment Of IgE Receptor Ig (*fcer1gl*)



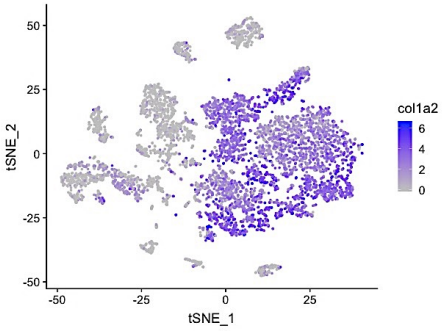
Cluster #4
Immune Cells

E Plac8 onzin related protein 3 (*ponzr3*)



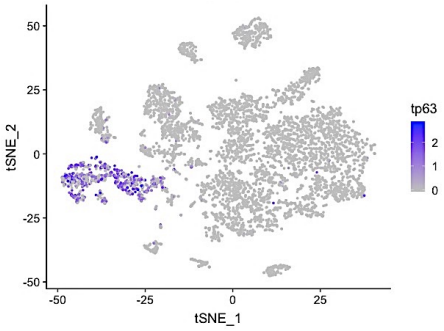
Cluster #5
Gills

F Collagen1A2 (*col1a2*)



Cluster #6
Stromal

G Tumor Protein P63 (*tp63*)



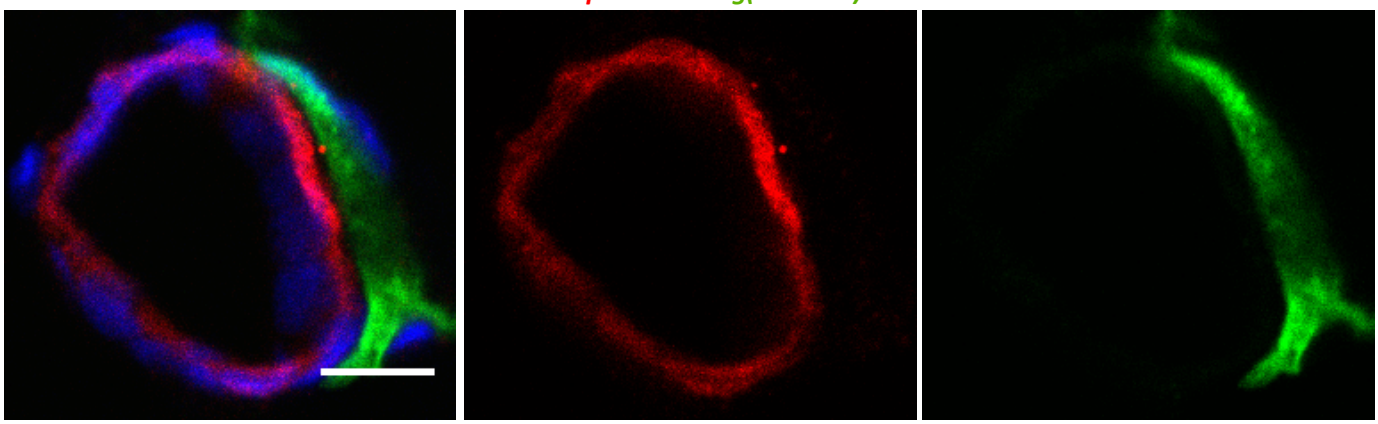
Cluster #7
Non-Follicular Epithelium (NFE)

Figure 4

A

Cluster #2 Endothelium

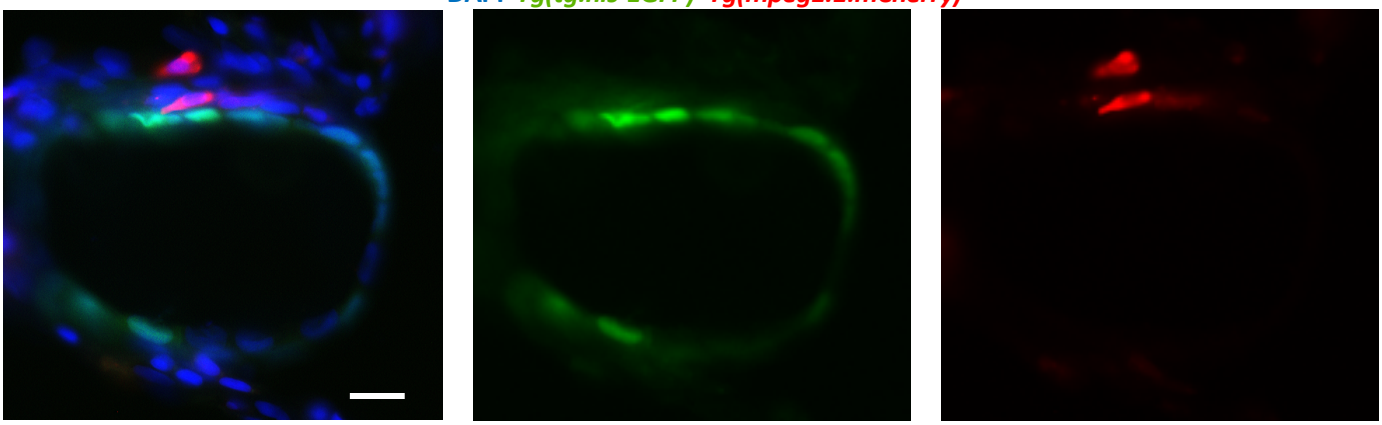
DAPI *pax2a^{mKO2}* *Tg(kdrl:GFP)*



B

Cluster #4 Immune Cells

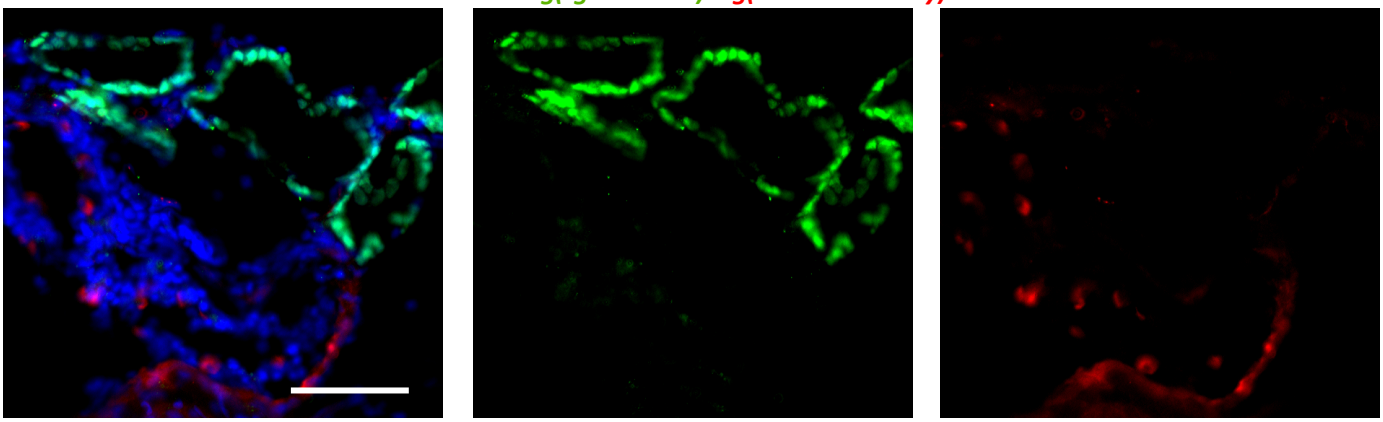
DAPI *Tg(tg:nls-EGFP)* *Tg(mpeg1.1:mCherry)*



C

Cluster #6 Stromal

DAPI *Tg(tg:nls-EGFP)* *Tg(col1a2:mCherry)*



D

Cluster #7 NFE

DAPI *Tg(tg:nls-EGFP)* TP63-Ab

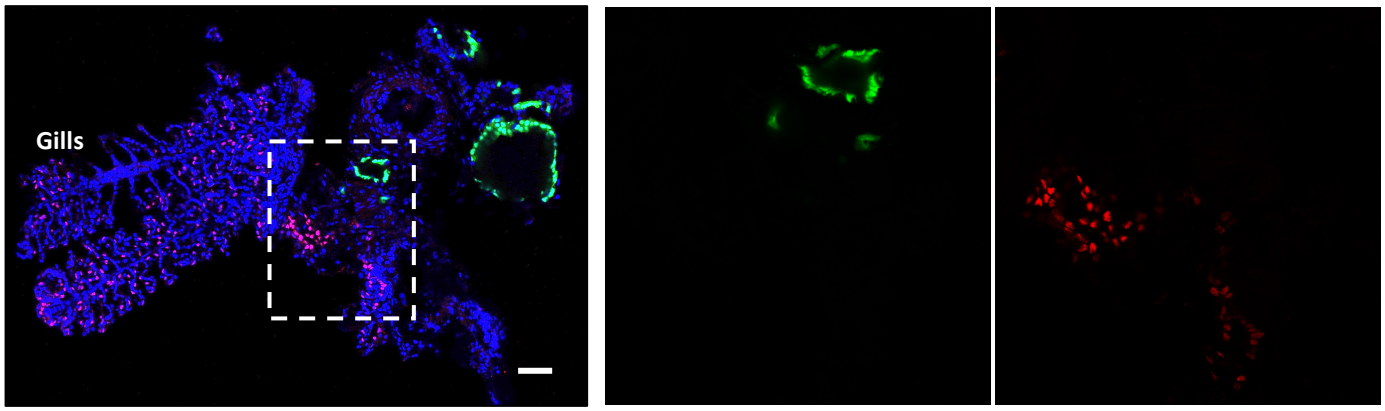


Figure 5

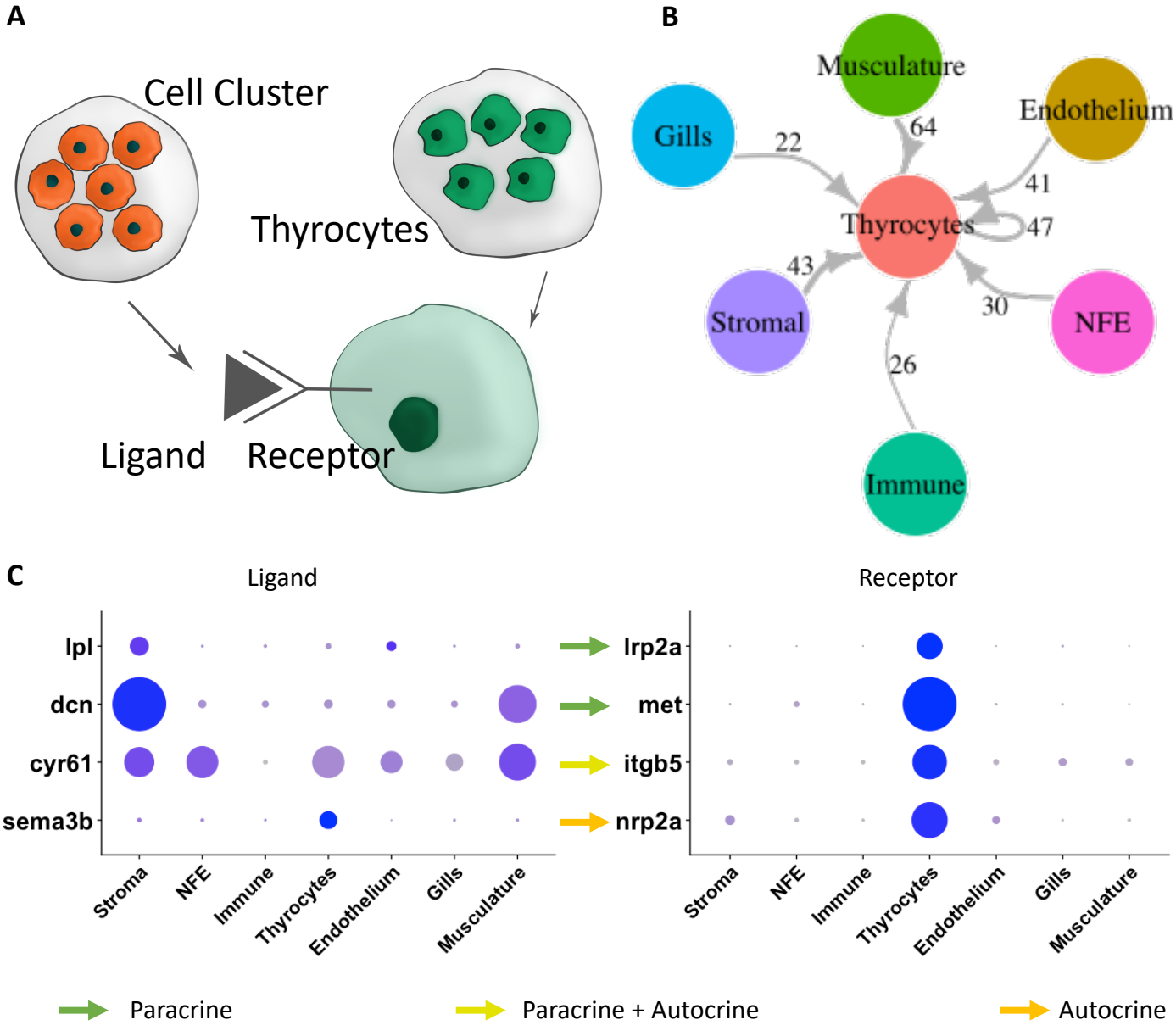

A
Cell Cluster
Thyrocytes
Ligand
Receptor
B
Musculature
Gills
Endothelium
Thyrocytes
Stromal
NFE
Immune
22
64
41
47
43
30
26
C
Ligand
Receptor
lpl
dcn
cyr61
sema3b
lrp2a
met
itgb5
nrp2a
Stroma
NFE
Immune
Thyrocytes
Endothelium
Gills
Musculature
Paracrine
Paracrine + Autocrine
Autocrine

Figure 6

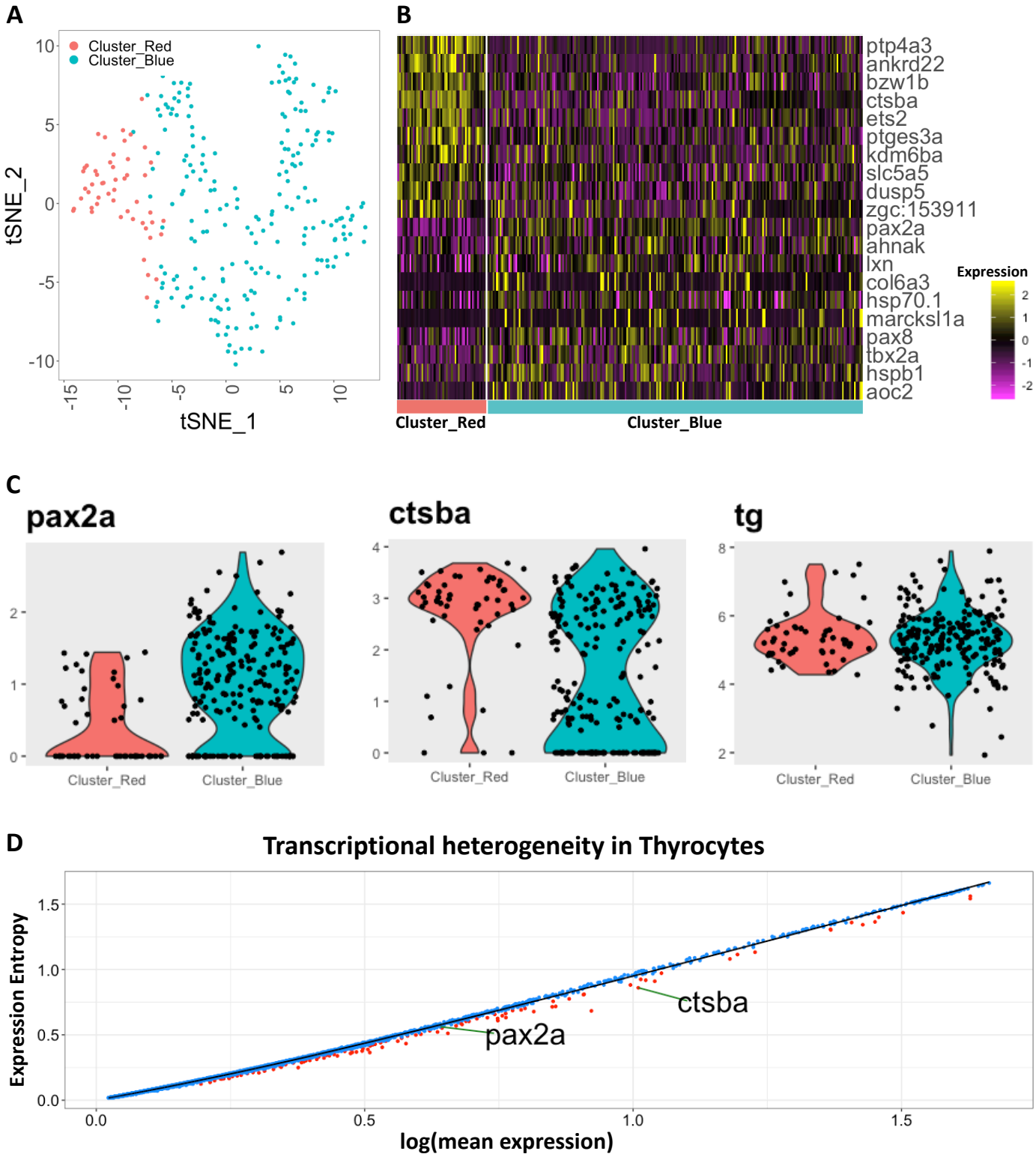

A
Cluster_Red
Cluster_Blue
tSNE_2
tSNE_1
B
ptp4a3
ankrd22
bzw1b
ctsba
ets2
ptges3a
kdm6ba
slc5a5
dusp5
zgc:153911
pax2a
ahnak
lxn
col6a3
hsp70.1
marcksl1a
pax8
tbx2a
hspb1
aoc2
Expression
Cluster_Red
Cluster_Blue
C
pax2a
ctsba
tg
Cluster_Red
Cluster_Blue
D
Transcriptional heterogeneity in Thyrocytes
Expression Entropy
log(mean expression)
ctsba
pax2a

Figure 7

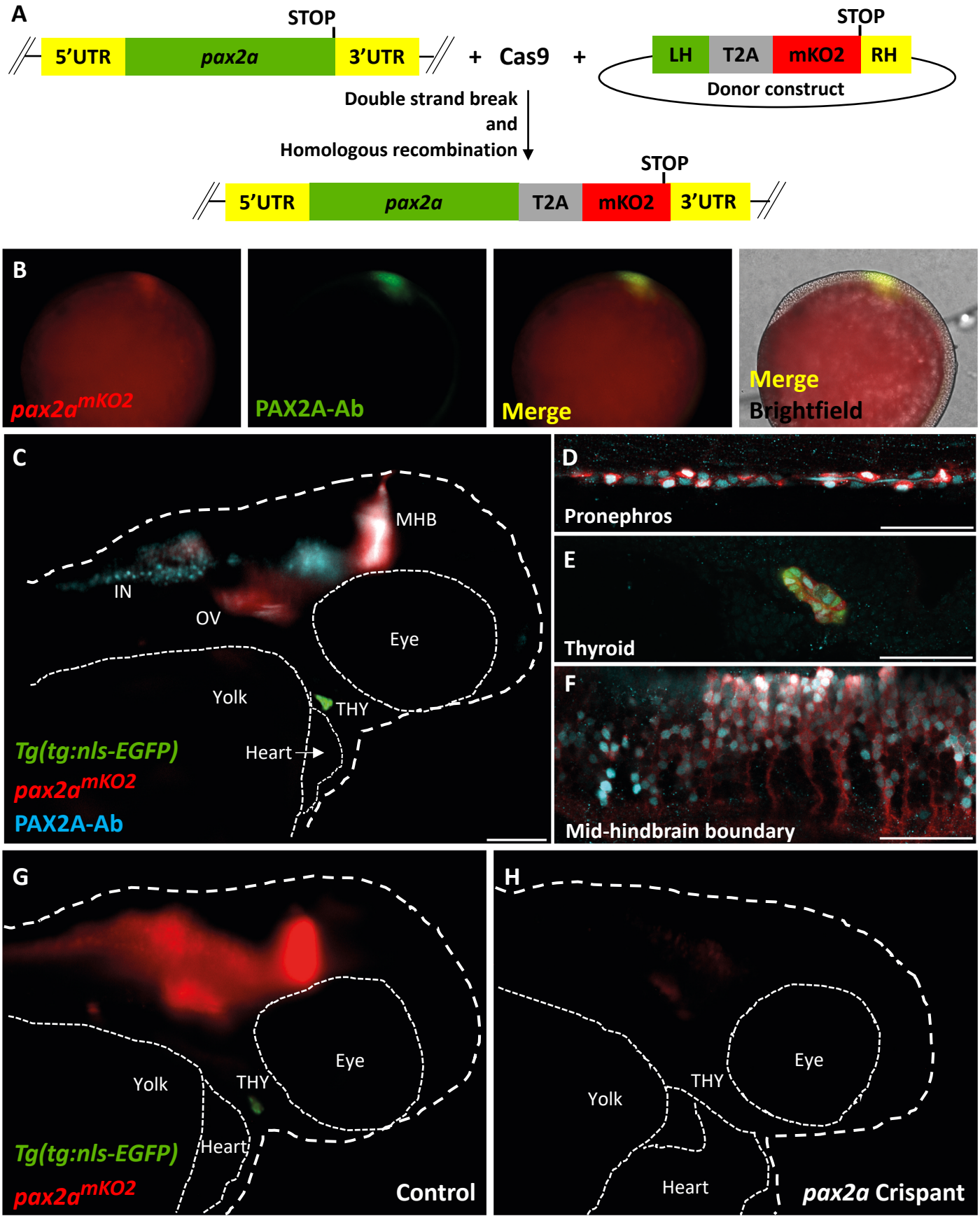

A
STOP
5'UTR
pax2a
3'UTR
+ Cas9 +
STOP
LH
T2A
mKO2
RH
Donor construct
Double strand break
and
Homologous recombination
STOP
5'UTR
pax2a
T2A
mKO2
3'UTR
B
pax2a mKO2
PAX2A-Ab
Merge
Merge
Brightfield
C
MHB
IN
OV
Eye
Yolk
THY
Heart
Tg(tg:nls-EGFP)
pax2a mKO2
PAX2A-Ab
D
Pronephros
E
Thyroid
F
Mid-hindbrain boundary
G
Eye
Yolk
THY
Heart
Tg(tg:nls-EGFP)
pax2a mKO2
Control
H
Eye
THY
Yolk
Heart
pax2a Crispant

Figure 8

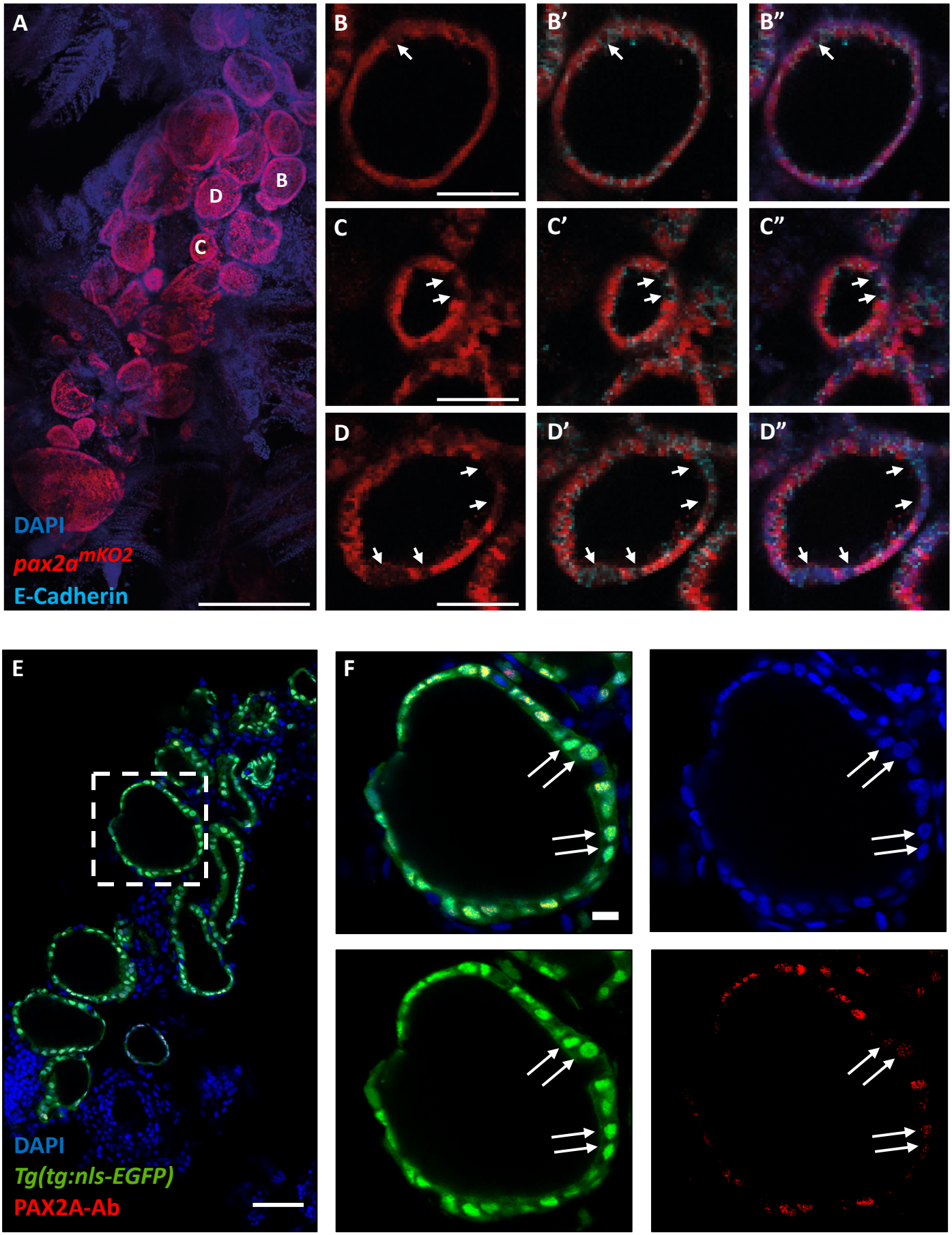
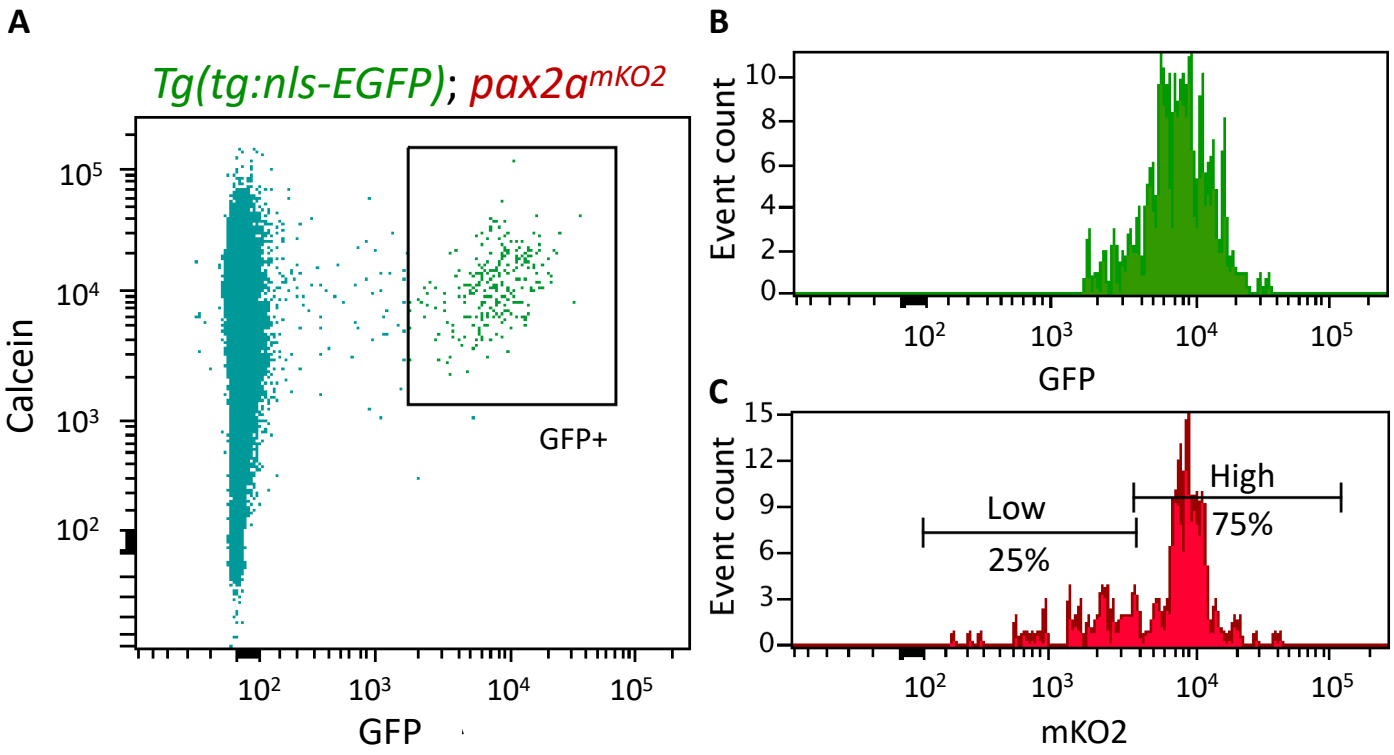

A
B
D
C
DAPI
pax2a mKO2
E-Cadherin
B
B'
B''
C
C'
C''
D
D'
D''
E
DAPI
Tg(tg:nls-EGFP)
PAX2A-Ab
F

Figure 9


A
Tg(tg:nls-EGFP); pax2a^mKO2
Calcein
GFP
GFP+
B
Event count
GFP
C
Event count
mKO2
Low
25%
High
75%