SA04 - Subclustering of cells from the interstitial lineage

Stefan Siebert

November 1, 2018, updated on May 22, 2019

Summary

We subcluster cells from the interstitial cell lineage. The resulting data object is the starting point for URD trajectory reconstruction.

Preliminaries

```r
library(Seurat)
library(dplyr)
library(Matrix)
library(gtable)
library(grid)
library(gridExtra)
library(rlang)

# Function to find the full ID for gene of interest
hFind <- function(x) {
    return(ds.s1@data@Dimnames[[1]][grep(x, ds.s1@data@Dimnames[[1]], ignore.case = T)])
}

# We assume a folder 'objects' in the markdown directory that contains
# our raw count object and all Seurat objects
```

Subsetting - cells of the interstitial lineage

We load the full data set and extract interstitial clusters (Fig. 1). This subset was used to perform NMF to identify gene expression modules expressed in cells of the interstitial lineage (NMF analysis ic_K75).

```r
# Read Seurat object for the whole dataset
ds.s1 <- readRDS("objects/Hydra_Seurat_Whole_Transcriptome.rds")

# Suspected doublet cluster db (68 cells) was excluded from downstream
# analyses
ds.s1 <- SubsetData(object = ds.s1, ident.remove = c("db"), subset.raw = TRUE)

# Run this to restore original cluster numbering
ds.s1 <- SetAllIdent(object = ds.s1, id = "cluster_numbering")

# Clusters of the interstitial lineage
ds.ic <- SubsetData(object = ds.s1, ident.use = c("1", "5", "9", "12",
    "13", "15", "16", "17", "18", "19", "20", "21", "22", "24", "25", "28",
    "29", "30", "31", "32", "34", "35", "36", "38", "40"), subset.raw = TRUE)

p1 <- TSNEPlot(object = ds.s1, group.by = "res.1.5", do.return = T, do.label = T,
    no.legend = TRUE)
p2 <- TSNEPlot(object = ds.ic, group.by = "res.1.5", do.return = T, do.label = T,
    no.legend = TRUE)

plot_grid(p1, p2, ncol = 2, labels = "AUTO", label_size = 20, align = "h")
```

Clustering of cells

We cluster the cells (Fig. 2).

```r
# Identify variable genes
ds.ic <- FindVariableGenes(object = ds.ic, mean.function = ExpMean, dispersion.function = LogVMR,
    x.low.cutoff = 0.05, x.high.cutoff = 4, y.cutoff = 0.5)
# Scale
ds.ic <- ScaleData(object = ds.ic)
```

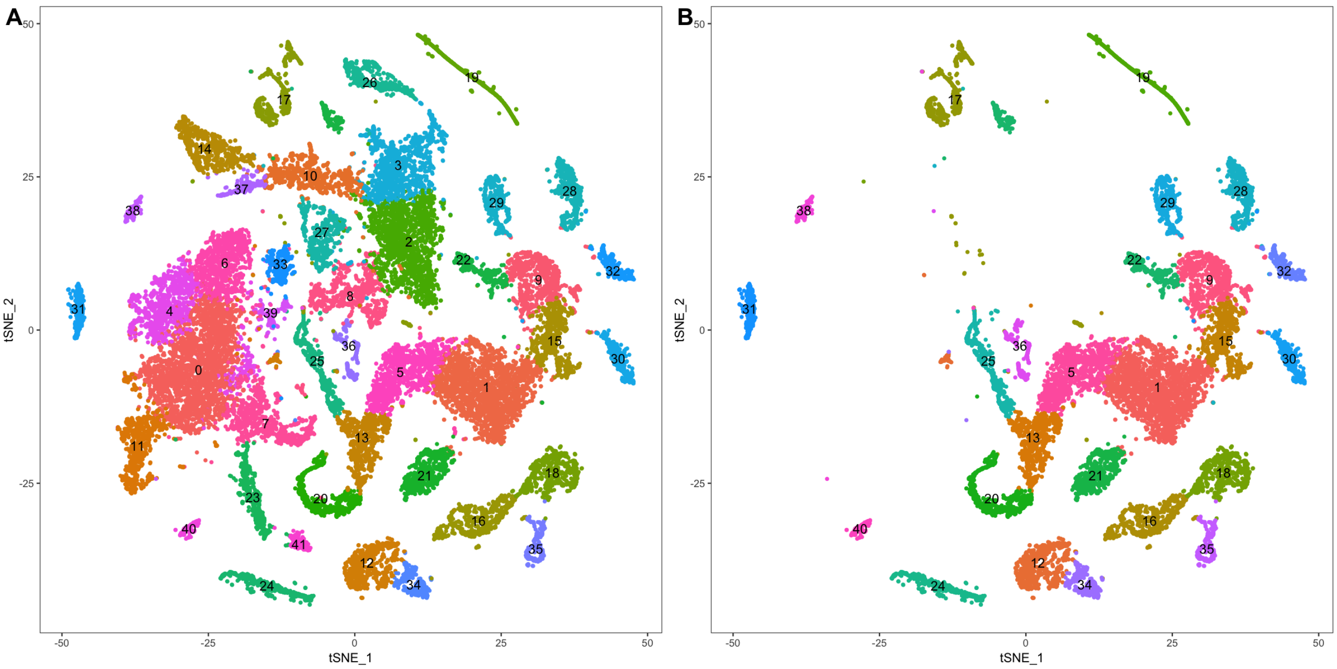

Figure 1: Subsetting cells of the interstitial lineage. A) Full t-SNE, B) Interstitial clusters.

```
# PCA on highly variable genes
ds.ic <- RunPCA(object = ds.ic, pc.genes = ds.ic@var.genes, pcs.compute = 40,
    do.print = TRUE, pcs.print = 1:5, genes.print = 20)

# Project PCA to find genes that weren't scored as highly variable, but
# should belong to a given PC and include them.
ds.ic <- ProjectPCA(object = ds.ic)

## Perform permutation test to directly calculate p-values ds.ic <-
## JackStraw(object = ds.ic, num.pc = 40, num.replicate = 100, do.print
## = FALSE) JackStrawPlot(object = ds.ic, PCs=1:40)

# Approximation of amount of variance encoded by each PC
PCElbowPlot(object = ds.ic, num.pc = 40)

# We run RunTSNE() for a range of principal components, seed and
# perplexities on a compute cluster: PCs 1:26 though 1:40; seed: 1,
# 300, 400; perplexity: 20, default (30), 40; The selected analysis
# considered PCs 1:31, perplexity 40, seed 300.

ds.ic <- FindClusters(object = ds.ic, reduction.type = "pca", dims.use = 1:31,
    force.recalc = TRUE, resolution = 1.5, print.output = 0)
ds.ic <- RunTSNE(object = ds.ic, dims.use = c(1:31), do.fast = T, perplexity = 40)

# saveRDS(ds.ic,'objects/ds.ic.s300_pc31_p40.rds')
```

```
# Since t-SNE is not deterministic we here load the object of our
# original analysis
ds.ic <- readRDS("objects/ds.ic.s300_pc31_p40.rds")

TSNEPlot(object = ds.ic, do.return = T, do.label = T, no.legend = TRUE,
    pt.size = 0.5)
```

Cluster annotation

We annotate the t-SNE using published marker genes (Fig. 3, 4). We apply neuron annotations from the neuronal subclustering (see analysis SA05_SubclustNeuronalCells).

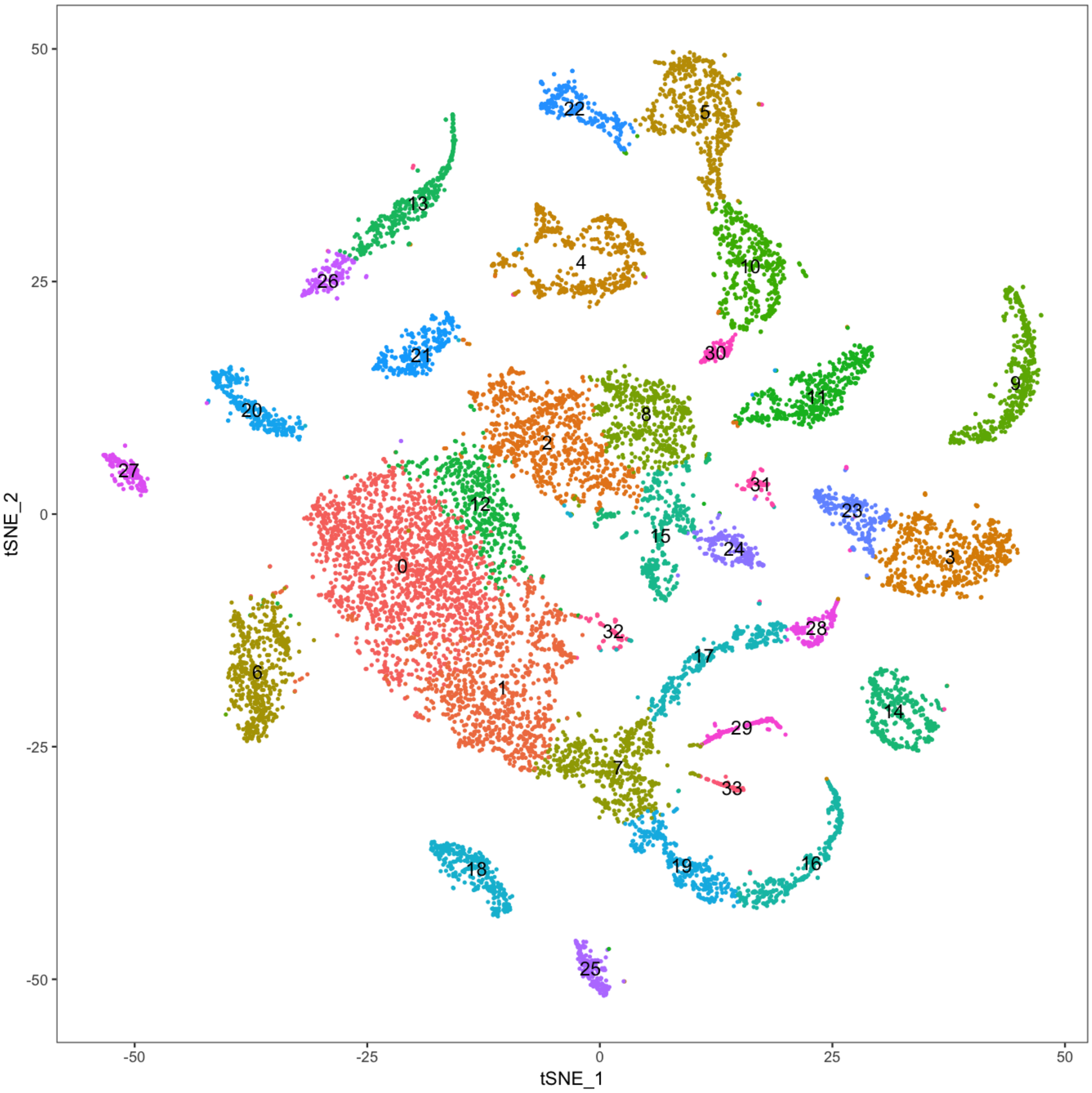

Figure 2: t-SNE plot for cells from the interstitial cell lineage.

```
# Genes to be plotted
gene.names <- c(hFind("t34367aep"), hFind("t11407aep"), hFind("t15393aep"),
    hFind("t3974aep"), hFind("t27424aep"), hFind("t12642aep"), hFind("t15237aep"),
    hFind("t22117aep"), hFind("t23176aep"), hFind("t11117aep"), hFind("t38683aep"),
    hFind("t7059aep"), hFind("t35863aep"), hFind("t14102aep"), hFind("t8678aep"),
    hFind("t18356aep"))

# Updated gene names for readability

new.names <- c("Hywi", "Cnnos1", "HvSoxC", "ELAV2 (t3974)", "Myb (t27424)",
    "FOXL1 (t12642)", "nowa-1", "HyDkk-3", "nematocilin A", "periculin1a",
    "H10A (t38683)", "MUC2 (t7059)", "HyTSR1", "HyDkk1/2/4 C", "HyDkk1/2/4 A",
    "CHIA (t18356)")

# Annotate
update.names(gene.names, new.names)

# Plot with tsne
p <- FeaturePlot(ds.ic, c("Hywi", "Cnnos1", "HvSoxC", "ELAV2 (t3974)",
    "Myb (t27424)", "FOXL1 (t12642)", "nowa-1", "HyDkk-3", "nematocilin A",
    "periculin1a", "H10A (t38683)", "MUC2 (t7059)", "HyTSR1", "HyDkk1/2/4 C",
    "HyDkk1/2/4 A", "CHIA (t18356)"), cols.use = c("grey", "blue"), do.return = TRUE)

plot_grid(plotlist = p, labels = "AUTO", label_size = 30, align = "h",
    ncol = 4)
```

```
## Annotate cluster
```

```
# Multiple cluster annotations were used in the course of this study.
# We store them in the Seurat object.  They can be restored using
# Seurat function SetAllIdent().

# Load whole transcriptome cluster numbering
ds.ic <- SetAllIdent(object = ds.ic, id = "cluster_numbering")

# Stash cluster ids from the whole transcriptome clustering
ds.ic <- StashIdent(object = ds.ic, save.name = "wt_clusters")
ds.ic@meta.data$cluster_numbering <- NULL

# Run this to restore original interstitial cluster numbering
ds.ic <- SetAllIdent(object = ds.ic, id = "res.1.5")

# Current ids
current.cluster.ids <- as.character(0:33)

# Cluster labeling scheme used in the bioRxiv preprint (Fig. 3A)
cluster.names <- c("SC/nb", "nb1", "nc_gc_prog", "smgc2", "nem", "zmg1",
    "fmgl1", "nb2", "nc_prog", "fmgl2_nurse", "gmgc_head", "nc1", "SC/prog",
    "mgl2", "nc2", "nc_prog/nc9", "nb5", "nb4", "nc3", "nb3", "nc4", "nc5",
    "zmg2", "smgc1", "nc8", "nc6", "mgl1", "nc7", "nb6", "nb7", "gmgc_hyp",
    "db1", "db2", "nb8")

# Update names in Seurat object
ds.ic@ident <- plyr::mapvalues(x = ds.ic@ident, from = current.cluster.ids,
    to = cluster.names)

# Stash labels
ds.ic <- StashIdent(object = ds.ic, save.name = "cluster.preprint")

# Run this to restore original cluster numbering
ds.ic <- SetAllIdent(object = ds.ic, id = "res.1.5")

# Modified cluster labeling with consistent neuron labels across
# clusterings, short labels (useful when working in R)
cluster.names <- c("SC/nb", "nb1", "n_gc_prog", "smgc2", "nematocyte",
    "zmg1", "fmgl1", "nb2", "n_prog", "fmgl2", "gmgc_head", "n_ec2", "ISC/prog",
    "mgl2", "n_ec3", "n_prog/n_ec1A", "nb5", "nb4", "n_en1", "nb3", "n_en2",
    "n_ec4", "zmg2", "smgc1", "n_ec1B", "n_ec5", "mgl1", "n_en3", "nb6",
    "nb7", "gmgc_hyp", "db1", "db2", "nb8")

# Update names in Seurat object
```

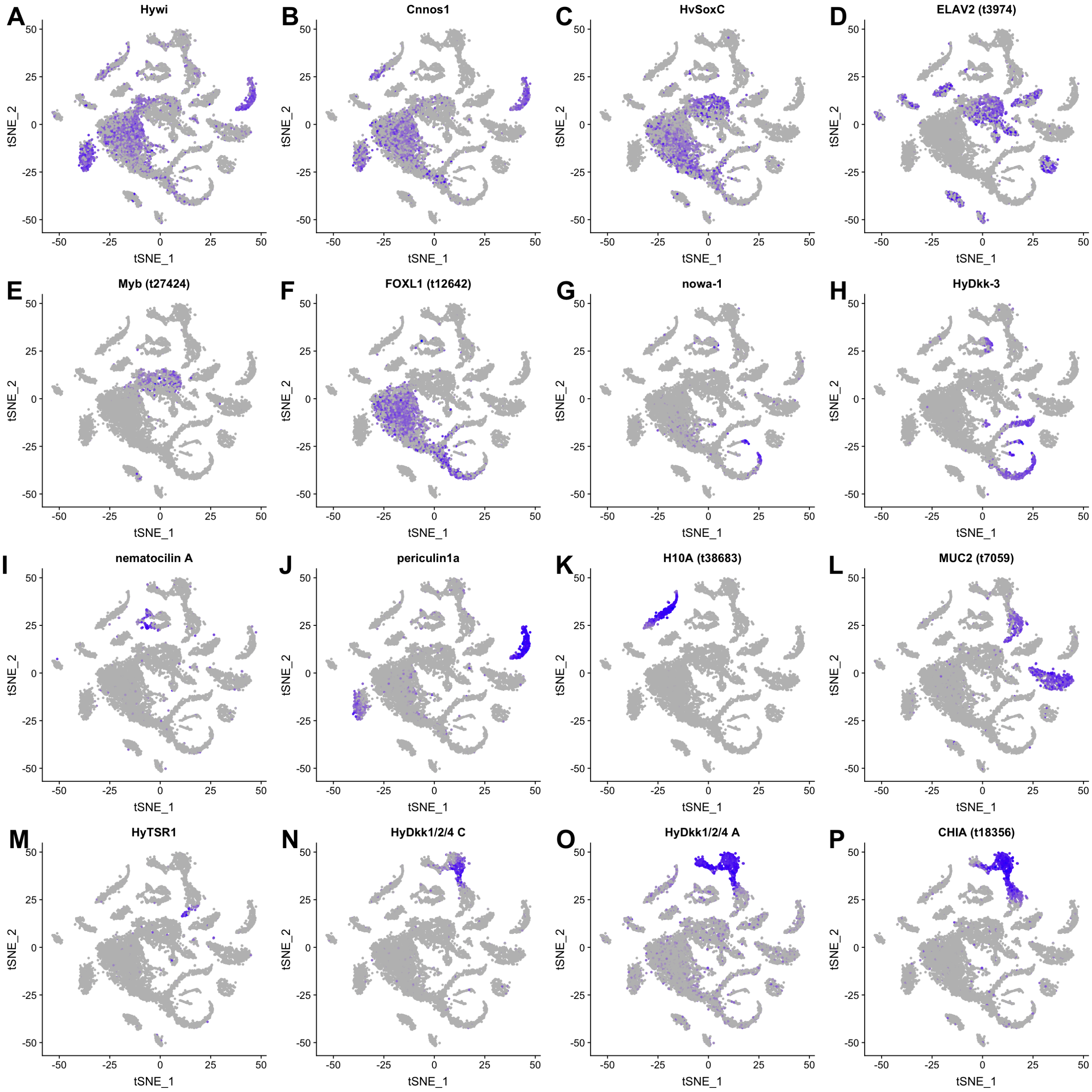

Figure 3: Selected markers used for cluster annotation. A) Interstitial stem cells, progenitor, germline - Hywi (*1*). B) Interstitial stem cells, progenitor, germline - Cnnos1 (*2*). C) Differentiating progenitors - HvSoxC (this study)(*3*). D) Neuronal cells - ELAV2 (t3974) (this study). E) Neurogenesis, gland cell differentiation - Myb (t27424) (this study). F) Interstitial stem cells, nematoblasts - FOXL1 (t12642) (this study). G) Nematoblasts - Nowa-1 (*4*). H) Nematoblasts - HyDkk-3 (*5*). I) Differentiated nematocyte - nematocillin A (*6*). J) Female germline - periculin1a (*7*). K) Male germline - histone H10A (t3863) (this study). L) Granular and spumous mucous gland cells - MUC2 (t7059). M) Granular mucous gland cells - HyTSR1 (*8*). N) Zymogen gland cell - Hydkk1/2/4 C (*9*). O) Zymogen gland cell - Hydkk1/2/4 A (*9*). P) Zymogen gland cell - CHIA (t18356) (this study).

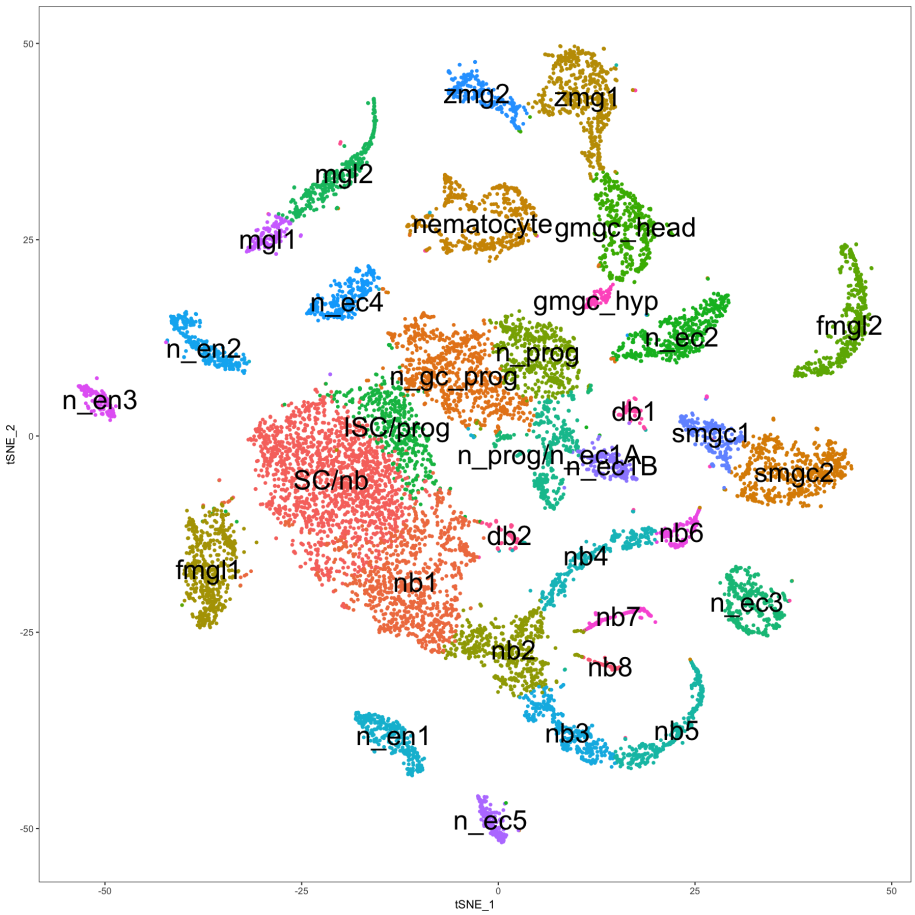

Figure 4: Annotated t-SNE representation for cells of the interstitial cell lineage. db: doublet, fmgl: female germline, gc: gland cell, gmgc: granular mucous cell, mgl: male germline, n: neuron, nb: nematoblast, smgc: spumous mucous cell, prog: progenitor, SC: stem cell, zmg: zymogen gland cell.

```
ds.ic@ident <- plyr::mapvalues(x = ds.ic@ident, from = current.cluster.ids,
    to = cluster.names)

# Stash labels
ds.ic <- StashIdent(object = ds.ic, save.name = "cluster.short")

TSNEPlot(object = ds.ic, do.return = T, do.label = T, label.size = 9, no.legend = TRUE)
```

```
# Run this to restore original cluster numbering
ds.ic <- SetAllIdent(object = ds.ic, id = "res.1.5")

# Modified cluster labeling with consistent neuron labels across
# clusterings and long labels for Broad portal presentation
cluster.names <- c("ISC/nematoblast", "nematoblast_1", "neuron/gland_cell_progenitor",
    "spumous_mucous_gland_cell_2", "nematocyte", "zymogen_gland_cell_1",
    "female_germline_1", "nematoblast_2", "neuron_progenitor", "female_germline_2_nurse_cell",
    "granular_mucous_gland_cell(head)", "neuron_ec2", "ISC/progenitor",
    "male_germline_2", "neuron_ec3", "neuron_prog/neuron_ec1A", "nematoblast_5",
    "nematoblast_4", "neuron_en1", "nematoblast_3", "neuron_en2", "neuron_ec4",
    "zymogen_gland_cell_2", "spumous_mucous_gland_cell_1", "neuron_ec1B",
    "neuron_ec5", "male_germline_1", "neuron_en3", "nematoblast_6", "nematoblast_7",
    "granular_mucous_gland_cell(hyp)", "db1", "db2", "nematoblast_8")

# Update names in Seurat object
ds.ic@ident <- plyr::mapvalues(x = ds.ic@ident, from = current.cluster.ids,
    to = cluster.names)

# Stash labels
```

```
ds.ic <- StashIdent(object = ds.ic, save.name = "cluster.long.portal")

# Run this to restore original cluster numbering
ds.ic <- SetAllIdent(object = ds.ic, id = "res.1.5")

# Modified cluster labeling used in manuscript (Fig. 3A)
cluster.names <- c("ISC/nb", "nb1", "n/gc_prog", "smgc2", "nematocyte",
    "zymogen_gc_1", "female_germline_1", "nb_2", "n_progenitor", "female_germline_2_nurse_cell",
    "granular_mgc(head)", "neuron_ec2", "ISC/prog", "male_germline_2",
    "neuron_ec3", "n_prog/neuron_ec1A", "nb5", "nb4", "neuron_en1", "nb3",
    "neuron_en2", "neuron_ec4", "zymogen_gc_2", "smgc1", "neuron_ec1B",
    "neuron_ec5", "male_germline_1", "neuron_en3", "nb_6", "nb_7", "granular_mgc(hyp)",
    "db1", "db2", "nb_8")

# Update names in Seurat object
ds.ic@ident <- plyr::mapvalues(x = ds.ic@ident, from = current.cluster.ids,
    to = cluster.names)

ds.ic <- StashIdent(object = ds.ic, save.name = "cluster.manuscript")

# restore a particular labeling scheme ds.ic <- SetAllIdent(object =
# ds.ic, id = 'res.1.5') ds.ic <- SetAllIdent(object = ds.ic, id =
# 'cluster.preprint') ds.ic <- SetAllIdent(object = ds.ic, id =
# 'cluster.short') ds.ic <- SetAllIdent(object = ds.ic, id =
# 'cluster.long.portal') ds.ic <- SetAllIdent(object = ds.ic, id =
# 'cluster.manuscript')
```

Identification of doublet clusters

We load gene modules that were identified in NMF analyses for the whole dataset (wt_K96). We plot scores for selected epithelial metagenes on the t-SNE representation for the whole dataset and on the t-SNE representation for the cells of the interstitial lineage (Fig. 5). High scores for epithelial gene modules suggest an epithelial component in the transcriptomes of cells in interstitial clusters db1 and db2. These cells were excluded from downstream analyses. We update cluster labels to improve clarity and accessibility (Fig. 6).

```
# Ic metagene scores
cellScores <- read.csv("nmf/ic_K75/GoodMeta_CellScores.csv", row.names = 1,
    check.names = F)
cellScores <- as.data.frame(cellScores)

# Make metagenes columns
cellScores <- t(cellScores)
# Fix cell ids
rownames(cellScores) <- sub("X", "", rownames(cellScores))
rownames(cellScores) <- sub("\\.", "-", rownames(cellScores))

# Match and load
cellScores <- cellScores[match(rownames(ds.ic@meta.data), rownames(cellScores)),
    ]
ds.ic@meta.data <- cbind(ds.ic@meta.data, cellScores)
head(ds.ic@meta.data)

# Whole metagene scores
cellScores <- read.csv("nmf/wt_K96/GoodMeta_CellScores.csv", row.names = 1,
    check.names = F)
cellScores <- as.data.frame(cellScores)

# Make metagenes columns
cellScores <- t(cellScores)
# Fix cell ids
rownames(cellScores) <- sub("X", "", rownames(cellScores))
rownames(cellScores) <- sub("\\.", "-", rownames(cellScores))

# Match and load
cellScores <- cellScores[match(rownames(ds.ic@meta.data), rownames(cellScores)),
    ]
ds.ic@meta.data <- cbind(ds.ic@meta.data, cellScores)
head(ds.ic@meta.data)
```

```r
# saveRDS('objects/Hydra_Seurat_IC.rds')
```

```r
# Plot epithelial metagene scores on full t-SNE and interstitial t-SNE
p <- FeaturePlot(ds.s1, c("wt11", "wt36", "wt40", "wt42", "wt12", "wt28",
    "wt30", "wt38"), do.return = TRUE, cols.use = c("grey", "blue"))
p1 <- FeaturePlot(ds.ic, c("wt11", "wt36", "wt40", "wt42", "wt12", "wt28",
    "wt30", "wt38"), do.return = TRUE, cols.use = c("grey", "blue"))

plot_grid(p[[1]], p[[2]], p[[3]], p[[4]], p1[[1]], p1[[2]], p1[[3]], p1[[4]],
    p[[5]], p[[6]], p[[7]], p[[8]], p1[[5]], p1[[6]], p1[[7]], p1[[8]],
    ncol = 4, labels = "AUTO", label_size = 20, align = "h")
```

Reclustering of cells prior to URD trajectory reconstruction

We recalculate variable genes, PCs and recluster after excluding clusters db1 and db2. PCs and clustering from this analysis are used in the URD trajectory reconstruction. We add them to our interstitial cell Seurat object (Hydra_Seurat_IC.rds) as ic.original.pca and clustering2.

```r
# Remove doublet cluster
ds.ic.s1 <- SubsetData(object = ds.ic, ident.remove = c("db1", "db2"),
    subset.raw = TRUE)

ds.ic.s1 <- FindVariableGenes(object = ds.ic.s1, mean.function = ExpMean,
    dispersion.function = LogVMR, x.low.cutoff = 0.05, x.high.cutoff = 4,
    y.cutoff = 0.5)
ds.ic.s1 <- ScaleData(object = ds.ic.s1)
ds.ic.s1 <- RunPCA(object = ds.ic.s1, pc.genes = ds.ic.s1@var.genes, pcs.compute = 40,
    do.print = TRUE, pcs.print = 1:5, genes.print = 20)
ds.ic.s1 <- ProjectPCA(object = ds.ic.s1)
ds.ic.s1 <- FindClusters(object = ds.ic.s1, reduction.type = "pca", dims.use = 1:31,
    force.recalc = TRUE, resolution = 1.5, print.output = 0)
ds.ic.s1 <- RunTSNE(object = ds.ic.s1, dims.use = c(1:31), do.fast = T,
    perplexity = 40)

# We add the updated PCA and the updated clustering to our interstitial
# cell object

# Extract the clustering
ds.ic.s1.meta <- ds.ic.s1@meta.data

ds.ic.s1.meta <- ds.ic.s1.meta[c(5)]

ds.ic.s1.meta <- as.data.frame(ds.ic.s1.meta)
# Rename clustering
names(ds.ic.s1.meta) <- c("clustering2")

# add gene id column
ds.ic@meta.data$id <- 1:nrow(ds.ic@meta.data)

# Add the clustering to metadata
ds.ic@meta.data <- merge(ds.ic@meta.data, ds.ic.s1.meta, by = "row.names",
    all.x = TRUE)
# Add row names
rownames(ds.ic@meta.data) <- ds.ic@meta.data$Row.names
ds.ic@meta.data[order(ds.ic@meta.data$id), ]

# Drop gene id column, row.name columns
ds.ic@meta.data$Row.names <- NULL
ds.ic@meta.data$id <- NULL

# Add updated PCA to Hydra_Seurat_IC.rds. This is the 2nd PCA in this
# document but the one that was originally used in the URD analysis.
ds.ic@dr$ic.original.pca <- ds.ic.s1@dr$pca

# saveRDS(ds.ic,'/objects/Hydra_Seurat_IC.rds')
```

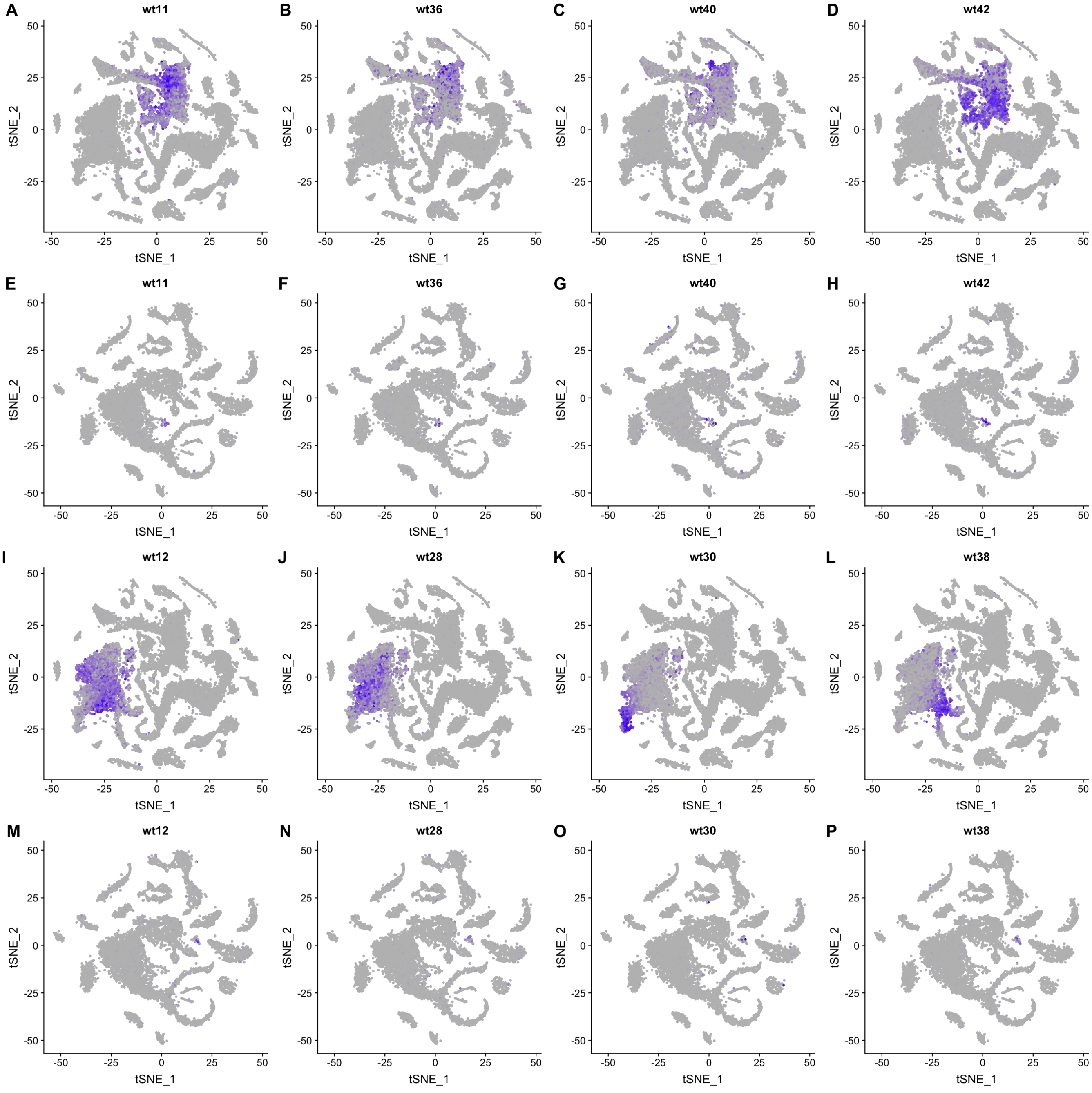

Figure 5: Expression of epithelial metagenes in cells of the interstitial lineage indicate hybrid transcriptomes in cells of clusters db1 and db2. A-D) Expression of four ectodermal gene modules visualized on the whole dataset t-SNE plot. E-H) Expression of ectodermal gene modules plotted on the t-SNE representation for the interstitial cell subset. I-L) Expression of four endodermal gene modules on the whole dataset t-SNE plot. M-P) Expression of endodermal gene modules plotted on the t-SNE representation for the interstitial cell subset.

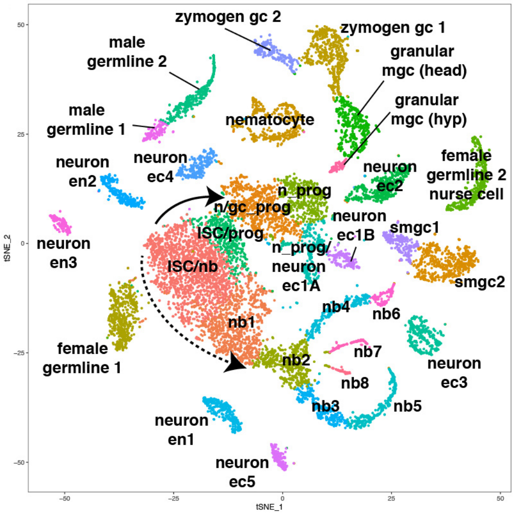

Figure 6: Annotated t-SNE plot with labels edited for clarity (manuscript Fig. 3A). ec: ectoderm, en: endoderm, gc: gland cell, hyp: hypostome, ISC: interstitial multipotent stem cell, mgc: mucous gland cell, n: neuron, nb: nematoblast, smgc: spumous mucous gland cell, prog: progenitor.

Software versions

This document was computed on Sun May 26 16:50:31 2019 with the following R package versions.

```
R version 3.4.4 (2018-03-15)
Platform: x86_64-apple-darwin15.6.0 (64-bit)
Running under: macOS  10.14.2

Matrix products: default
BLAS: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRlapack.dylib

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] grid      stats     graphics  grDevices utils     datasets  methods  
[8] base     

other attached packages:
 [1] bindrcpp_0.2.2 rlang_0.3.0.1  gridExtra_2.3  gtable_0.2.0  
 [5] dplyr_0.7.7    Seurat_2.3.4   Matrix_1.2-15  cowplot_0.9.2 
 [9] ggplot2_3.1.0  knitr_1.20    

loaded via a namespace (and not attached):
  [1] Rtsne_0.13          colorspace_1.3-2    class_7.3-14       
  [4] modeltools_0.2-21   ggridges_0.4.1      mclust_5.4         
  [7] rprojroot_1.3-2     htmlTable_1.11.2    base64enc_0.1-3    
 [10] rstudioapi_0.7      proxy_0.4-21        npsurv_0.4-0       
 [13] flexmix_2.3-14      bit64_0.9-7         mvtnorm_1.0-7      
 [16] codetools_0.2-15    splines_3.4.4       R.methodsS3_1.7.1  
 [19] lsei_1.2-0          robustbase_0.92-8   Formula_1.2-2      
 [22] jsonlite_1.5        ica_1.0-1           cluster_2.0.6      
 [25] kernlab_0.9-25      png_0.1-7           R.oo_1.21.0        
 [28] compiler_3.4.4      httr_1.3.1          backports_1.1.2    
 [31] assertthat_0.2.0    lazyeval_0.2.1      formatR_1.5        
 [34] lars_1.2            acepack_1.4.1       htmltools_0.3.6    
 [37] tools_3.4.4         igraph_1.2.2        glue_1.3.0         
 [40] RANN_2.6            reshape2_1.4.3      Rcpp_0.12.19       
 [43] trimcluster_0.1-2   gdata_2.18.0        ape_5.1            
 [46] nlme_3.1-131.1      iterators_1.0.9     fpc_2.1-11         
 [49] lmtest_0.9-35       xfun_0.1            stringr_1.3.0      
 [52] irlba_2.3.2          gtools_3.5.0         DEoptimR_1.0-8
 [55] MASS_7.3-49          zoo_1.8-1            scales_1.0.0
 [58] doSNOW_1.0.16        parallel_3.4.4       RColorBrewer_1.1-2
 [61] yaml_2.1.18          reticulate_1.10      pbapply_1.3-4
 [64] rpart_4.1-13         segmented_0.5-3.0    latticeExtra_0.6-28
 [67] stringi_1.1.6        highr_0.6            foreach_1.4.4
 [70] checkmate_1.8.5      caTools_1.17.1.1     SDMTools_1.1-221
 [73] pkgconfig_2.0.2      dtw_1.18-1           prabclus_2.2-6
 [76] bitops_1.0-6         evaluate_0.10.1      lattice_0.20-35
 [79] ROCR_1.0-7           purrr_0.2.5          bindr_0.1.1
 [82] labeling_0.3         htmlwidgets_1.0      bit_1.1-12
 [85] tidyselect_0.2.5     plyr_1.8.4           magrittr_1.5
 [88] bookdown_0.7         R6_2.3.0             snow_0.4-2
 [91] gplots_3.0.1         Hmisc_4.1-1          pillar_1.2.1
 [94] foreign_0.8-69       withr_2.1.2          fitdistrplus_1.0-11
 [97] mixtools_1.1.0       survival_2.41-3      nnet_7.3-12
[100] tibble_1.4.2         tsne_0.1-3           crayon_1.3.4
[103] hdf5r_1.0.1          KernSmooth_2.23-15   rmarkdown_1.9
[106] data.table_1.11.8    metap_0.8            digest_0.6.18
[109] diptest_0.75-7       tidyr_0.8.0          R.utils_2.6.0
[112] stats4_3.4.4         munsell_0.5.0
```

References

1. C. E. Juliano *et al.*, PIWI proteins and PIWI-interacting RNAs function in Hydra somatic stem cells. *Proceedings of the National Academy of Sciences of the United States of America.* **111**, 337–342 (2014).

2. K. Mochizuki, H. Sano, S. Kobayashi, C. Nishimiya-Fujisawa, T. Fujisawa, Expression and evolutionary conservation of nanos-related genes in Hydra. *Wilhelm Roux' Archiv für Entwicklungsmechanik der Organismen.* **210**, 591–602 (2000).

3. Hemmrich, Georg *et al.*, Molecular signatures of the three stem cell lineages in hydra and the emergence of stem cell function at the base of multicellularity. *Molecular biology and evolution.* **29**, 3267–3280 (2012).

4. Engel, Ulrike *et al.*, Nowa, a novel protein with minicollagen Cys-rich domains, is involved in nematocyst formation in Hydra. *Journal of cell science.* **115**, 3923–3934 (2002).

5. Fedders, Henning, Augustin, René, Bosch, Thomas C G, A Dickkopf- 3-related gene is expressed in differentiating nematocytes in the basal metazoan Hydra. *Wilhelm Roux' Archiv für Entwicklungsmechanik der Organismen.* **214**, 72–80 (2004).

6. J. S. Hwang *et al.*, Cilium evolution: identification of a novel protein, nematocilin, in the mechanosensory cilium of Hydra nematocytes. *Molecular biology and evolution.* **25**, 2009–2017 (2008).

7. S. Fraune *et al.*, In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides. *Proceedings of the National Academy of Sciences of the United States of America.* **107**, 18067–18072 (2010).

8. S. Siebert, F. Anton-Erxleben, T. C. G. Bosch, Cell type complexity in the basal metazoan Hydra is maintained by both stem cell based mechanisms and transdifferentiation. *Developmental biology.* **313**, 13–24 (2008).

9. R. Augustin *et al.*, Dickkopf related genes are components of the positional value gradient in Hydra. **296**, 62–70 (2006).