Supplementary Information

The transcription factor *escargot* maintains stemness and suppresses differentiation in *Drosophila* intestinal stem cells

Jerome Korzelius[1], Svenja K. Naumann[1], Mariano A. Loza-Coll[2,3], Jessica S. K. Chan[1], Devanjali Dutta[1], Jessica Oberheim[1], Christine Gläßer[1], Tony D. Southall, Andrea H. Brand, D. Leanne Jones[2,3] and Bruce A. Edgar[1*]

Jerome Korzelius[1], Svenja K. Naumann[1], Mariano A. Loza-Coll[2,3], Jessica S. K. Chan[1], Devanjali Dutta[1], Jessica Oberheim[1], Christine Gläßer[1], Tony D. Southall[4,5], Andrea H. Brand[4], D. Leanne Jones[2,3] and Bruce A. Edgar[1*]

1. DKFZ/ZMBH Alliance, University of Heidelberg, Heidelberg, Germany

2. Laboratory of Genetics, The Salk Institute for Biological Studies, La Jolla, CA U.S.A.

3. Department of Molecular, Cell, and Developmental Biology, University of California- Los Angeles, Los Angeles, CA, U.S.A.

4. The Gurdon Institute and Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK

5. Present address: Department of Life Sciences, Imperial College London, London, UK

Table of contents

- Supplemental Materials and Methods

- Supplemental References

Figure S1. Loss of *esg* causes progenitor cell loss. A-B. Progenitor cell distribution in

esg-Gal4^{ts};UAS-GFP in the posterior midgut in control (*white* RNAi) animals. **B-B''**.

RNAi knockdown of *esg* causes loss of progenitor cell nests and the appearance of

both large, weakly GFP-positive cells with a large nucleus (arrows) and GFP$^+$/Pros$^+$

cells with a small nucleus (arrowhead). **C-D**. DE-Cadherin staining of *w*^{RNAi} control

and *esg^{RNAi}* midguts. **C**. Control GFP$^+$-progenitor cell nests have increased DE-

Cadherin staining at the membrane, especially at the ISC-EB border (arrows). **D**. In

esg^{RNAi} animals, GFP$^+$-cells with increased DE-cadherin are largely absent, except in

rare, highly GFP-positive doublet cells (arrowhead). Most weakly GFP$^+$-cells have

decreased DE-Cadherin staining but increased cellular and nuclear size (arrows).

Figure S2. Esg loss causes EB cells to differentiate into EC cells at an increased

speed. A-C. Loss of *esg* leads to an increase in GFP$^+$/Pdm1$^+$-double positive cells. **A**.

Su(H)GBE-Gal4^{ts}, UAS-GFP marks EB cells that differentiate into EC cells.

Approximately 1/3 of control EB cells expressing UAS-GFP are also positive for the

EC-marker Pdm1 (**A-A''**, arrowhead), whereas the majority is Pdm1-negative (**A-A''**,

arrows, quantification in **C**). **B**. *esg^{RNAi}* leads to a 2-fold increase in the number of

Pdm1-positive EB cells (**B-B''**, arrowheads, quantification in **C**). **C**. Quantification of

the fraction of Pdm1-positive EB cells from *Su(H)GBE-Gal4^{ts}, UAS-GFP* animals with

or without *esg^{RNAi}* (*n* =230 for control. *n* = 169 for *esg^{RNAi}*, *p*< 0,05%, Students t-test).

D-E. Delta staining of *Su(H)GBE-Gal4^{ts}, UAS-GFP* animals with or without *esg^{RNAi}*.

Control EBs appear as small, spindle-shaped cells located next to Delta-positive ISCs

(**D-D''**, arrows). Upon *esg^{RNAi}*, most of the EBs appear rounded and have a larger

nucleus as control EBs (**E-E''**, arrows). **F-I**. FACS profiles of *w*^{1118}, *Su(H)GBE-Gal4^{ts}* >

UAS-GFP, Su(H)GBE-Gal4^{ts}> UAS-GFP, UAS- esg^{RNAi} and *Su(H)GBE-Gal4^{ts}> UAS-GFP,*

UAS- esg animals. **H**. Expression of *esg^{RNAi}* results in a 3-fold decrease of the overall

number of GFP-positive EB cells (compare **G** and **H**). **I**. Overexpression of Esg results

in a 3-fold increase in the number of GFP-positive EB cells.

Figure S3. *UAS-esg*-expressing cells in the progenitor compartment express both

the ISC marker *Dl-lacZ* and the EB marker *Su(H)-GBE-lacZ*. A-B. The *esg*-F/O system

was combined with either *Dl^{05151}-lacZ* or *Su(H)-GBE-lacZ* in the absence or presence

of *UAS-esg*. **A.** Control animals have 1 or more *Dl-lacZ*-positive cells within the clone

(**A-A'**, arrows). **B.** *UAS-esg*-expressing F/O-clones consist of small diploid cells that

were both *Dl-lacZ*-positive (**B-B'**, arrows) and –negative (**B-B'**, arrowheads). **C.**

Control cells contain 1 or more *Su(H)-GBE*-positive cells (**C-C'**, arrows). **D**. Similarly,

UAS-esg-expressing F/O-clones have both *Su(H)-GBE*-positive (**D-D'**, arrows) and –

negative cells (**D-D'**, arrowheads). **E**. Quantifications of *Dl-lacZ^{+}* cells/clone from

clones in (**A-B**). *n*= 327 for control, *n*=542 for *UAS-esg*. *p < 0,05*, Mann-Whitney U-

test. **F**. Quantifications of *Su(H)-GBE-lacZ^{+}* cells/clone from clones in (**C-D**). *n*= 153 for

control, *n*=408 for *UAS-esg*. *p < 0,0001*, Mann-Whitney U-test.

Figure S4. Overexpression of Esg in ECs leads to induction of cytokine production

and triggers a regenerative response in stem cells. **A-D**. Cells expressing a control

(*white*) RNAi have low levels of cell renewal (A, arrow) and a regularly structured

epithelium (**B**). **C-D**. EC-specific expression of *UAS-esg* results in a striking increase in

mitotic cells (**C**) and massive loss of GFP-positive EC cells (**D**). **E-F**. Ectopic Esg

expression with *MyoIA^{ts}* induces strong *Upd3-lacZ* reporter expression 24 hours after

induction. Control animals have a very low baseline expression of this cytokine

reporter under homeostatic conditions, whereas *UAS-esg* induction results in widespread induction of *Upd3-lacZ* in EC cells and an increase in the mitotic index of these midguts (**F'**, arrows, quantification in **G**). **G.** Quantification of mitoses/midgut after 2 days of transgene induction. (*n* > 10 for each genotype). **H-J.** Prolonged expression of *UAS-esg* results in a loss of midgut tissue integrity resulting in lethality (**I**, quantification in **K**). This is partially rescued by co-expression of *UAS-dIAP*. **K.** Kaplan-Meier analysis of survival over time for flies of the appropriate genotype. *n* = 3 X 20 animals for each genotype.

Figure S5. Loss of Esg in *Notch^RNAi* tumors leads to precocious differentiation towards the EE fate. **A-A'.** *Notch^RNAi* midguts have tumorous growths that take up almost the entire posterior midgut and are highly pH3-positive (**A'**). **B-B'.** *esg^G66B Notch^RNAi* midguts have only few small overgrowths and are less pH3-positive than *Notch^RNAi* midguts. **C.** *Notch^RNAi* tumors form large sheets with Pros-positive cells that reside mainly along the periphery of the clone (A-A''). **D.** *esg^G66B Notch^RNAi* clones are smaller and consist mainly of Pros-positive cells (B-B''). **E.** Quantification of the percentage of Pros-positive cells/clone for *Notch^RNAi* and *esg^G66B Notch^RNAi* clones. *n* = 25 clones for *Notch^RNAi* clones and *n* = 27 for *esg^G66B Notch^RNAi* clones. *p*<0,0001 Student's t-test

Table S1. Table showing normalized gene expression values (log2 scale, mean RPKM values) of genes across midgut cell types. At least 2-3 biological replicates for each cell type were used for this analysis. The RPKM values are color-coded row-wise for each gene and are comparable across cell types. Color coding- Red: low gene expression; Yellow: medium gene expression; Green: High gene expression.

Table S2. List of genes differentially regulated upon *esg^ts^* > *esg^RNAi^* or *UAS-esg*

expression. Full list of genes that change significantly (>2-fold, at 95% confidence) in

sorted *esg^ts^, UAS-GFP* cells expressing either *UAS-esg* or *esg^RNAi^* for 24 hours.

Differential expression data was generated using edgeR. List data includes gene

symbol, CG number, $\log_2$-fold-change and fold-change values for each gene as well

as significance (p-value with Benjamini-Hochberg correction), GO-terms and Flybase

description for each gene.

Table S3. List of enriched GO-terms in the set of differentially regulated genes

from *esg^ts^* > *esg^RNAi^* animals. GO-term enrichment was calculated using amiGO (see

Supplemental Experimental Procedures). This list includes all GO-terms that are

enriched in the list of genes that change significantly upon *esg^RNAi^* in sorted *esg^ts^*,

UAS-GFP cells (see Table S2).

Table S4. List of genes differentially regulated upon upon *MyoIA^ts^* > *UAS-esg*

expression acquired from edgeR analysis. Full list of genes that change significantly

(>2-fold, at 95% confidence) in whole midguts upon expression of either Esg alone or

Esg and dIAP in ECs with *MyoIA^ts^*. Control animals are *MyoIA^ts^* animals crossed to

empty RNAi vector (VDRC 60100). Differential expression data was generated using

edgeR. List data includes gene symbol, CG number, $\log_2$-fold-change and fold-change

values for each gene as well as significance (p-value with Benjamini-Hochberg

correction), GO-terms and Flybase description for each gene.

Table S5. List of enriched GO-terms in the set of differentially regulated genes

from *MyoIA^ts^* > *UAS-esg* whole midguts. GO-term enrichment was calculated using

amiGO. This list includes all GO-terms that are enriched in the list of genes that

change significantly in whole midguts upon expression of *UAS-esg* in ECs with

MyoIA^{ts} (see Table S4).

Table S6. List of high-confidence Esg-target genes. Comparison of genes that went

up upon *esg*^{ts} > *esg*^{RNAi} (see Table S2), went down upon *MyoIA*^{ts} > *UAS-esg*

expression (see Table S4) and had 1 or more Esg-binding regions within a 5 kb range

from the transcription start site. See Figure 5H for a Venn-diagram graphical

representation.

Supplemental Materials and Methods

Fly stocks used in this study

Driver lines: **esg^{ts}**: *y,w;esg-GAL4/CyO;tub-GAL80^{ts},UAS-GFP/Tm6B*, **MyoIA^{ts}**: *w; Myo1A-Gal4^{NP0001}/CyO; tub-Gal80^{ts}, UAS-GFP/TM6B*, **Su(H)-GBE^{ts}**: *Su(H)GBE-Gal4,UAS-CD8-GFP/CyO;tubGal80^{ts}/Tm6B,Tb*, **esg-F/O**: *w;esg-Gal4,UAS-GFP,tub-Gal80^{ts}/CyO;UAS-flp,act>CD2>Gal4/Tm6B* , **MyoIA^{ts};Upd3.1-lacZ**: *w; Myo1A-Gal4/CyO,tub-Gal80^{ts}, UAS-GFP/CyO; Upd3.1-lacZ/Tm6B*. **UAS-transgenes**: *y,w;UASt-esg/CyO, UAS-N^{intra}, UAS-dIAP, UAS-Pdm1-RD*. **Reporter genes:** *y,w;esg-lacZ$^{[k00606]/}$CyO, Su(H)-GBE-lacZ/Tm3, Sb, Dl^{05151}-lacZ, mira-GFP*. **MARCM stocks**: **MARCM 40A**: *y,w,hsflp$^{1.22}$,tub-Gal4,UAS-GFP;tub-Gal80,FRT40A/CyO,act-GFP* **MARCM 82B**: *y,w,hs-flp$^{1.22}$,tub-Gal4,UAS-GFP;tub-Gal80,FRT82B/Tm3,Sb*, **MARCM 80B**: *y,w,hs-flp$^{1.22}$,tub-Gal4,UAS-GFP;tub-Gal80,FRT80B/Tm6B, w;FRT40A, w;FRT82B w;FRT80B* **Mutants**: *y,w;esg^{G66B},FRT40A/In2(LR), Gla, Bc,Elp*. **RNAi stocks**: *UAS-esg^{RNAi}* (TRiP-line JF03134, BL28514) *UAS-esg^{RNAi}* (TRiP-line HMS00025, BL34063), *UAS-Notch^{RNAi}*.

Clonal analysis and quantification

Clonal analyses of ISCs and their progeny were achieved using the *esg*-FlipOut system (Jiang et al., 2009) or the MARCM system (Lee and Luo, 1999) as indicated. Animals were allowed to age for a minimum of 3 days after eclosion before clonal induction. MARCM clones were induced by heat-shock for 45 minutes in a 37°C waterbath. *esg*-F/O clones were induced by shifting animals to 29°C for the indicated times. Clonal composition was counted by taking stacks from ROIs from a minimum of 5 posterior midguts. The percentage of Delta/Pdm1-positive cells was determined

in these ROIs by using the CellCounter plugin for Fiji. Percentages were plotted as the percentage of GFP-positive cell type/total number of GFP-positive clonal cells. For survival assays, animals were shifted and/or infected for the indicated time and survival was assayed every 1-2 days. For *P.e.* survival assays, animals were transferred to vials with fresh P.e. every 3-4 days. Results are presented as pooled data from 3 biological repeats of 15-20 animals each. Kaplan-Meier survival plots were generated in Graphpad Prism. DAPI intensity quantification (Figure 6H and Figure 2E) was done using Imaris v7.3 and values (DAPI intensity X nuclear volume) for each nucleus were plotted as a frequency distribution histogram (Figure 6H) or box-plot (Figure 2E, whiskers: 2.5-97.5 percentile) in Graphpad Prism. Mitotic indices were determined by manually counting pH3-positive cells in > 10 female adult intestines per genotype. For graphs, error bars indicate standard deviation, unless indicated otherwise.

Flow cytometry and RNA-Seq data analysis

For midgut cell dissociation for RNA-Seq, 1 mg/ml of Elastase were added per sample and guts incubated for 1 h at 27 °C, shaking at 600 rpm. To help dissociation, samples were pipetted vigorously every 15 min. For FACS profiling, tissue dissociation was performed by treatment with 7.5 mg/ml of Collagenase for 45 min followed by addition of 50 µl of 10X Trypsin/EDTA (Sigma) for 15-20 min. For RNA-Seq, a minimum of 20,000 GFP[+] cells/condition was sorted on a FACS Aria II Flow Cytometer (BD Biosciences) using a 70 µm nozzle and FACS Diva software. Intestinal cells obtained from w^{1118} animals where used to control for auto-fluorescence and set sorting gates. Dead cells were excluded from the sorting procedure by staining

with 3 μl of Propidium Iodide (1 mg/ml, Life Technologies/Molecular Probes). RNA was isolated from sorted cells using the Arcturus PicoPure RNA Isolation Kit (Life Technologies). mRNA was amplified using the Arcturus RiboAmp HS Plus Amplification Kit (Life Technologies). Amplified RNA integrity was determined with an Agilent 2100 Bioanalyzer before library generation. cDNA libraries were generated using the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB), and NEBNext Multiplex Oligos for Illumina (NEB) were used for multiplexing. RNA sequencing was performed on a Hi-Seq2000 (Illumina) with 50 bp single-end read length. Differential expression analysis was carried out using a pipeline described in (Anders et al., 2013). First, raw reads were aligned against the *Drosophila* genome (version 5.73, BDGP) using tophat with default settings. Reads being aligned were counted using HTSeqcount with a minimum alignment score of 10, while unmapped reads were trimmed using Trimmomatic cutting adapter and other illumina-specific sequences from the read and performing a 5-mer sliding window clipping with quality of 30. Surviving reads were then remapped using tophat with default settings and subsequently counted. After that, the counts of both mapped and previously unmapped reads were added and used as basis for differential expression analysis. edgeR was used for differential expression analysis (Robinson et al., 2010). Genes with at least 1 read per million in at least n samples with n= lowest sample size were kept, others discarded; p-value adjustment method: Benjamini-Hochberg (Benjamini and Hochberg, 1995). GO-term enrichment analysis was done with the amiGO software (Boyle et al., 2004; Carbon et al., 2009). P-value cut-off was p < 0,01, minimum number of gene products: 2.

qRT-PCR Primer sequences

Upd3 F: 5'-cccagccaacgatttttatg-3' **Upd3** R: 5'-tgttaccgctccggctac-3', Universal Probe

Library probe: #18 **Pdm1 F**: 5'-cgggataaatcgaaggaagc-3' **Pdm1 R**: 5'-

agtatttgatgtgtttgcgacttt-3', Universal Probe Library probe: #62, **GAPDH1 F**: 5'-

gctccgggaaaaggaaaa-3', **GAPDH1 R**: 5'-tccgttaattccgatcttcg-3', Universal Probe

Library probe: #102

Supplemental References

Anders, S., McCarthy, D.J., Chen, Y., Okoniewski, M., Smyth, G.K., Huber, W., and Robinson, M.D. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc *8*, 1765-1786.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological *57*, 289-300.

Boyle, E.I., Weng, S.A., Gollub, J., Jin, H., Botstein, D., Cherry, J.M., and Sherlock, G. (2004). GO::TermFinder - open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics *20*, 3710-3715.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., Hub, A., and Grp, W.P.W. (2009). AmiGO: online access to ontology and annotation data. Bioinformatics *25*, 288-289.

Jiang, H., Patel, P.H., Kohlmaier, A., Grenley, M.O., McEwen, D.G., and Edgar, B.A. (2009). Cytokine/Jak/Stat signaling mediates regeneration and homeostasis in the Drosophila midgut. Cell *137*, 1343-1355.

Lee, T., and Luo, L. (1999). Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. Neuron *22*, 451-461.

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