Supplementary Information Appendix

Cell Population	**Marker Gene**	**KHID (KH2012:KH.*)**
aATEN	GnRH2	C9.484
	GHSR	C1.315
BTN	ASIC1b	C1.215
	Synaphin	L164.21
PSC	βɣ-Crystallin	S605.3
	Sulfotransferase	C9.423
Epidermis (Control and PouIV-misexpressed)	FCL	C1.611
	Unknown	C8.844

Table S1. Gene names and matching KHIDs.

Gene	**KHID (KH2012:KH.*)**
PouIV	C2.42
Klf	C5.154
Neurogenin (Ng)	C6.129
β-Thymosine	C2.140

Table S2. Gene names and matching KHIDs for CESN ID

OE Subcluster	aATEN	BTN	CESN	PSC	Number of Cells	Percentage of Total Number
1	0.0474	0.2693	0.0198	0.0860	387	24.6
2	0.0601	0.4189	0.0805	0.0257	371	23.6
3	0.0925	0.4196	0.1440	0.1005	367	23.4
4	0.0577	0.4636	0.0169	0.0923	162	10.3
5	0.0375	0.2382	0.1770	0.0175	136	8.66
6	0.0200	0.1917	0.4324	0.0228	78	4.96
7	0.0594	0.3754	0.0280	0.2097	69	4.39

Table S3. Average of Cell-Type Specific Solved Coefficients for all cells in each subcluster population. Number of cells belonging to each subcluster and percentage of total cells is also given. Subcluster histograms are shown below table for visual aid.

Table S4. Primer Sequences

Gene Name	**Construct name**	**Forward primer**	**Reverse Primer**
POU IV	pSPCiPOU IVcDNA	aagcggccgctatgtttactaacatgct	ctacataatcacgtccccattaaa
CesA	pSPCesAPOUIVcDNA	atgcggatcctcgcggcgaagtaaagcgag	gatcggatcctgtcagaccaggtgttca
neurogenin	pSPneurogenin-1675/-1426fog220K	gatcctcgagccattactccgcaatgcgcg	gatcggatccaaattacccaaaatttgtcc
Foxg	pSPFoxg500bpK	gatcctcgagaatccgtcgtagagtgacaa	gatcgcggccgcctgtatgccgcgtttctc

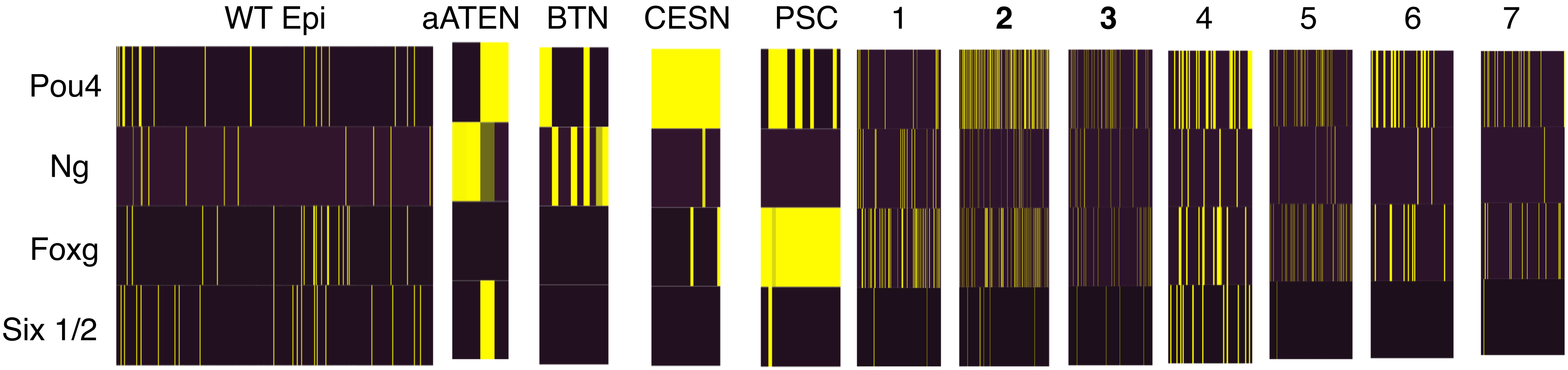


Fig. S1. Expression of Key Transcription Factors in WT Cell types and OE Clusters.

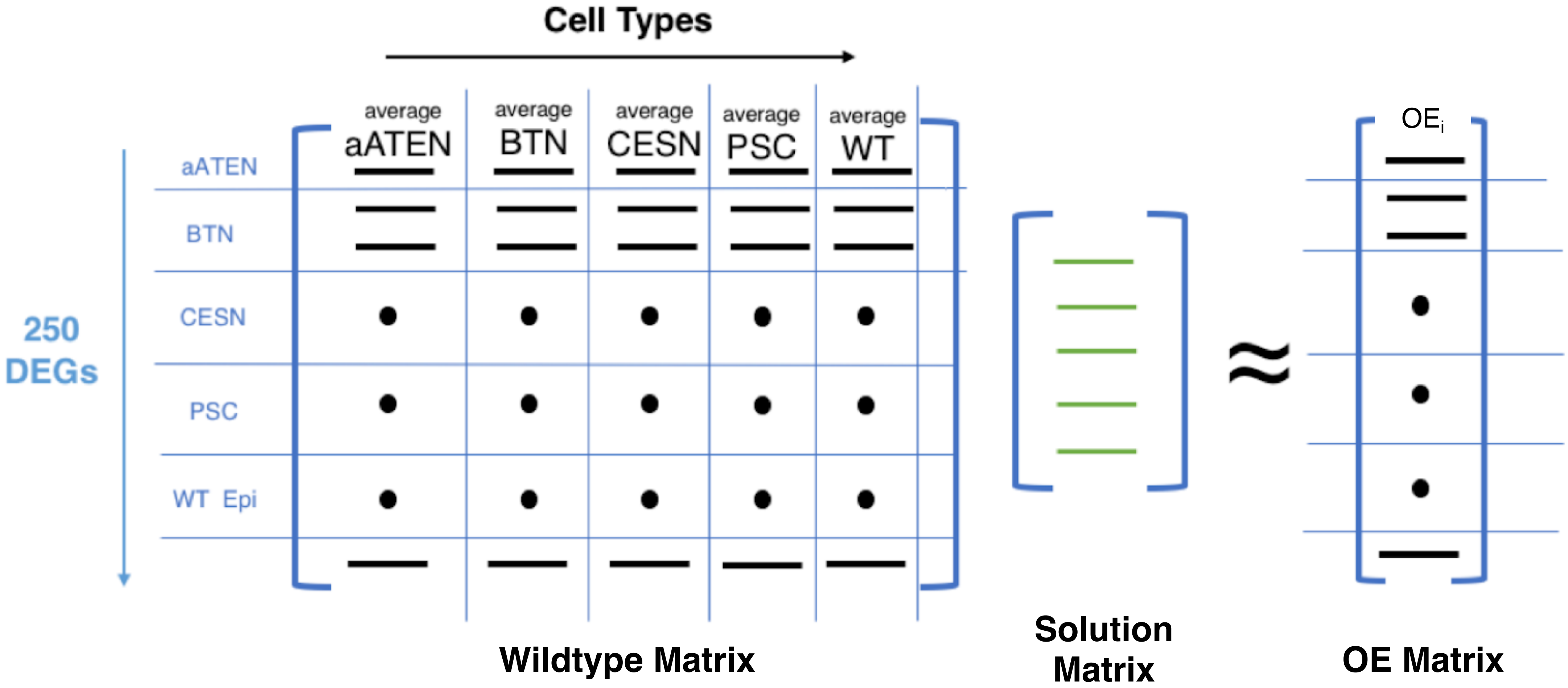


Fig. S2. Schematic of Linear Model. Each green entry in the Solution Matrix s a "Cell-type Specific Solved Coefficient"

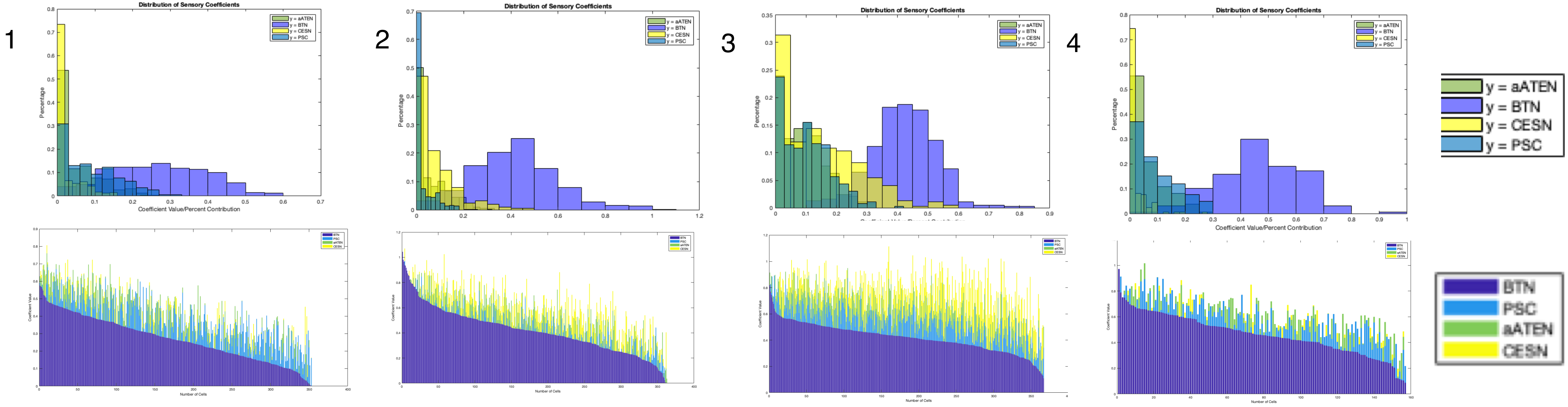


Fig. S3a. Visualization of Solved Coefficients for all cells (histogram, top) and Solved coefficients for single cells (stacked barplot, bottom) in subclusters 1-4. For the histograms, X-axis is coefficient value, Y-axis is the percentage of cells. For the stacked barplots, X-axis is a single cell, Y-axis is value of coefficient. Coefficients have bounds [0, 1].

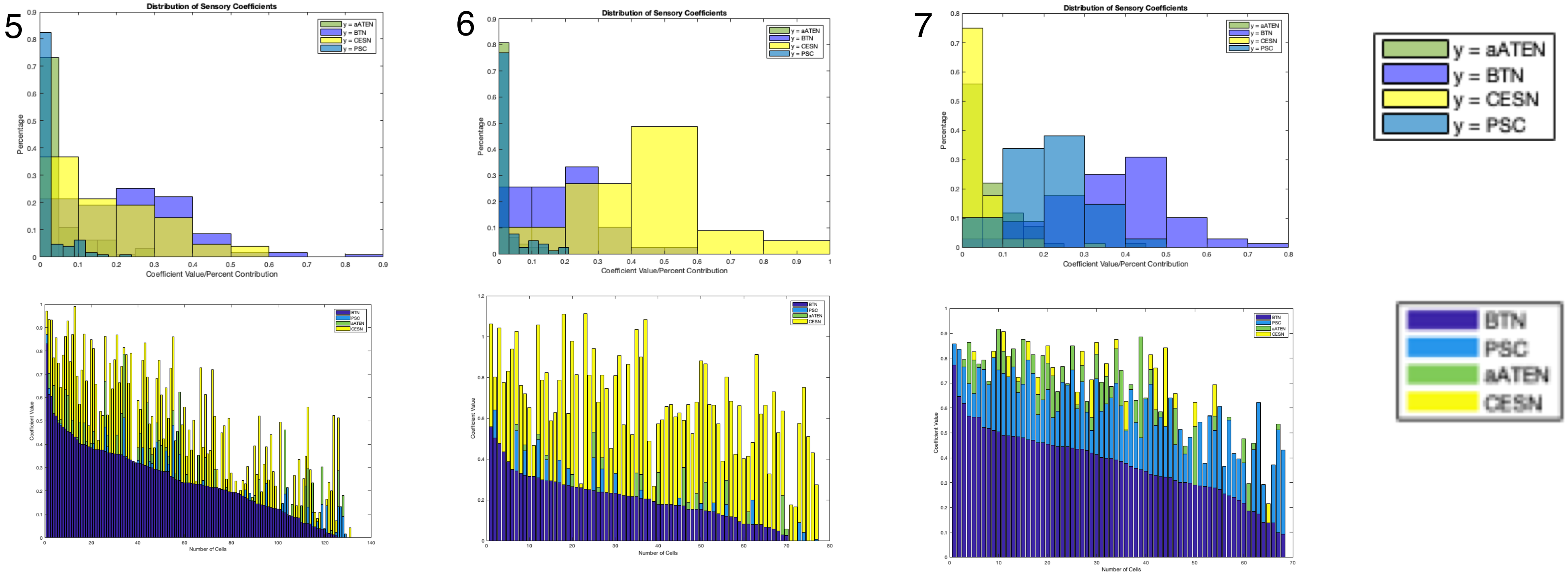
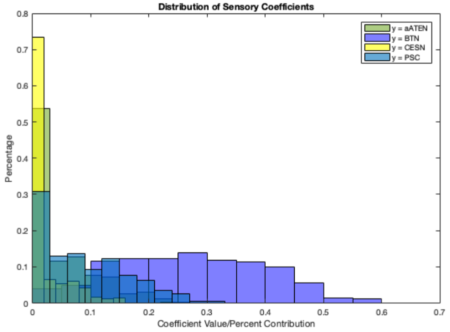


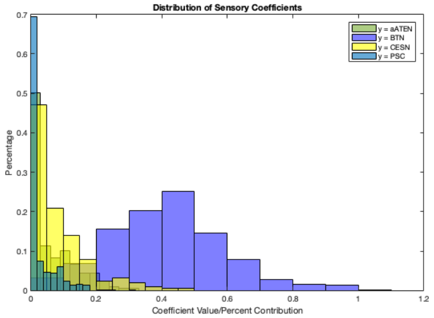
Fig. S3b. Visualization of Solved Coefficients for all cells (histogram, top) and Solved coefficients for single cells (stacked barplot, bottom) in subclusters 5-7. For the histograms, X-axis is coefficient value, Y-axis is the percentage of cells. For the stacked barplots, X-axis is a single cell, Y-axis is value of coefficient. Coefficients have bounds [0, 1].

OE Subcluster	aATEN	BTN	CESN	PSC	Number of Cells	Percentage of Total Number
1	0.0474	0.2693	0.0198	0.0860	387	24.6
2	0.0601	0.4189	0.0805	0.0257	371	23.6
3	0.0925	0.4196	0.1440	0.1005	367	23.4
4	0.0577	0.4636	0.0169	0.0923	162	10.3
5	0.0375	0.2382	0.1770	0.0175	136	8.66
6	0.0200	0.1917	0.4324	0.0228	78	4.96
7	0.0594	0.3754	0.0280	0.2097	69	4.39

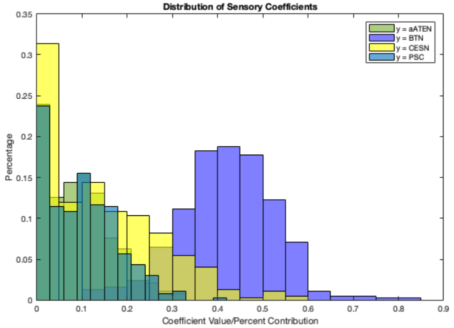
1



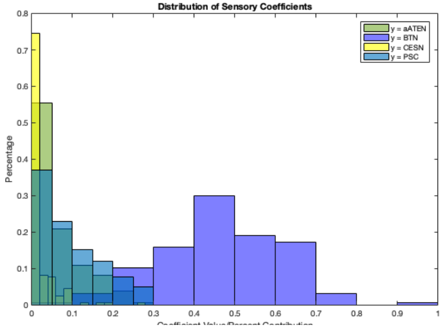
2



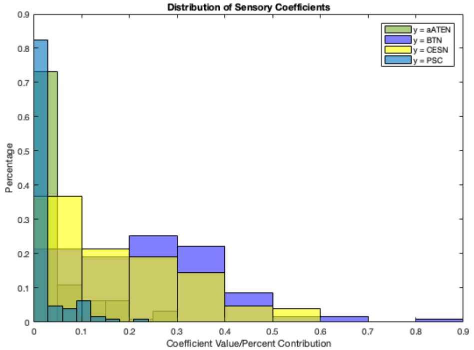
3



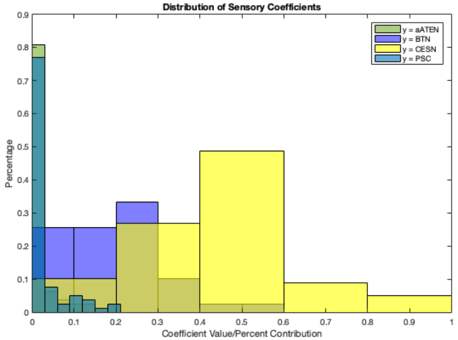
4



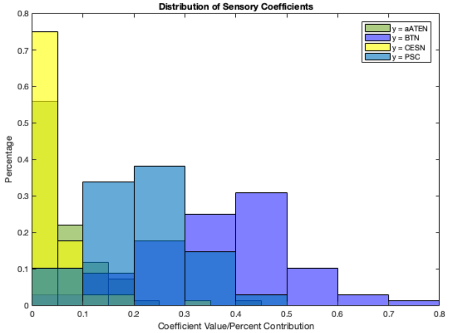
5



6



7



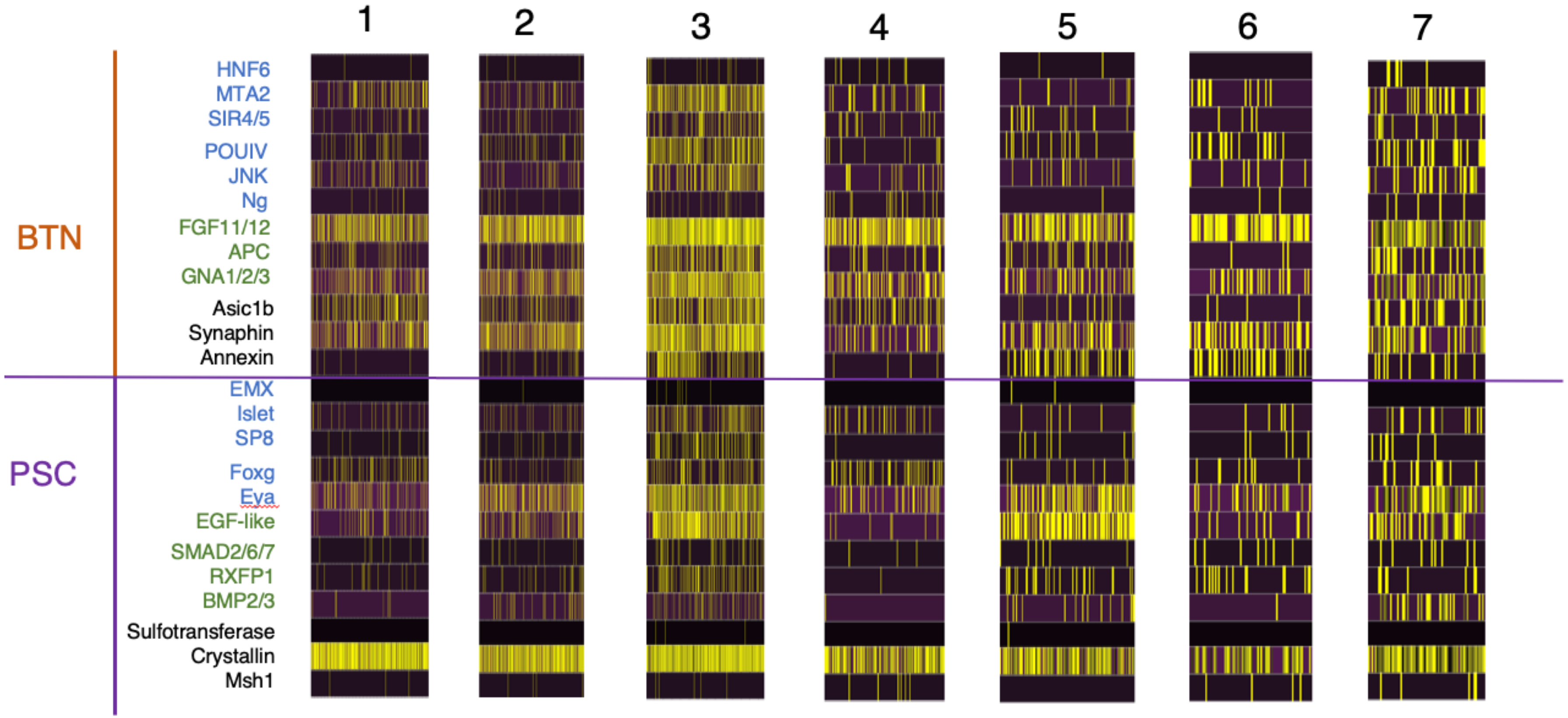


Fig. S4. Expression Levels of Various BTN and PSC Markers among Cells in OE Subclusters, Wildtype BTNs, and Wildtype PSCs. The above heatmap visualizes expression levels of various BTN and PSC transcription factor (blue), signaling molecule (green), and effector (black) genes.

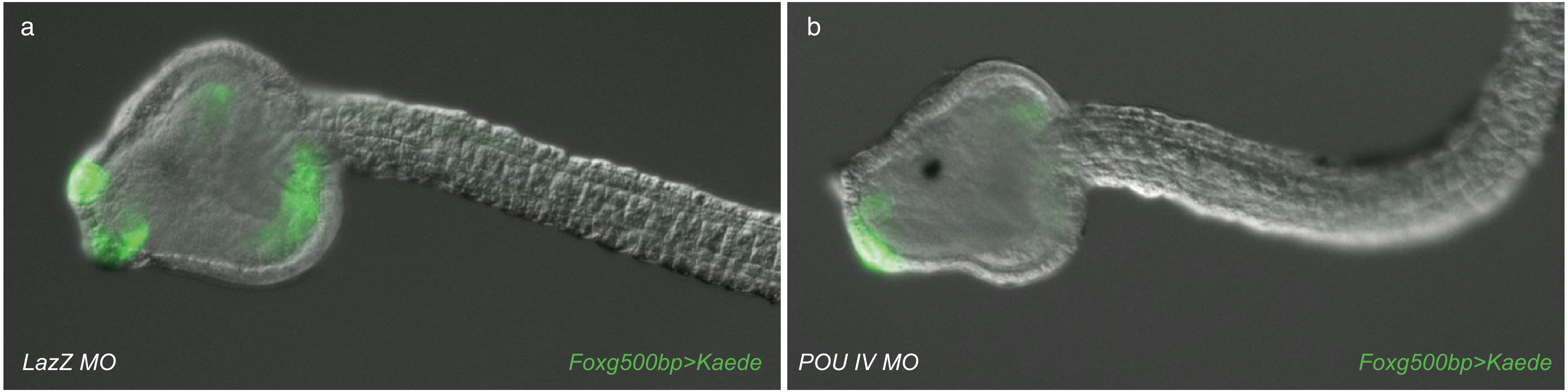


Fig. S5. Expression of Foxg is diminished in larvae injected with POU IV morpholino. (a) Kaede fluorescence in *Foxg>Kaede* injected larva **(b)** Foxg expression is diminished in PSCs upon loss of POU IV expression.

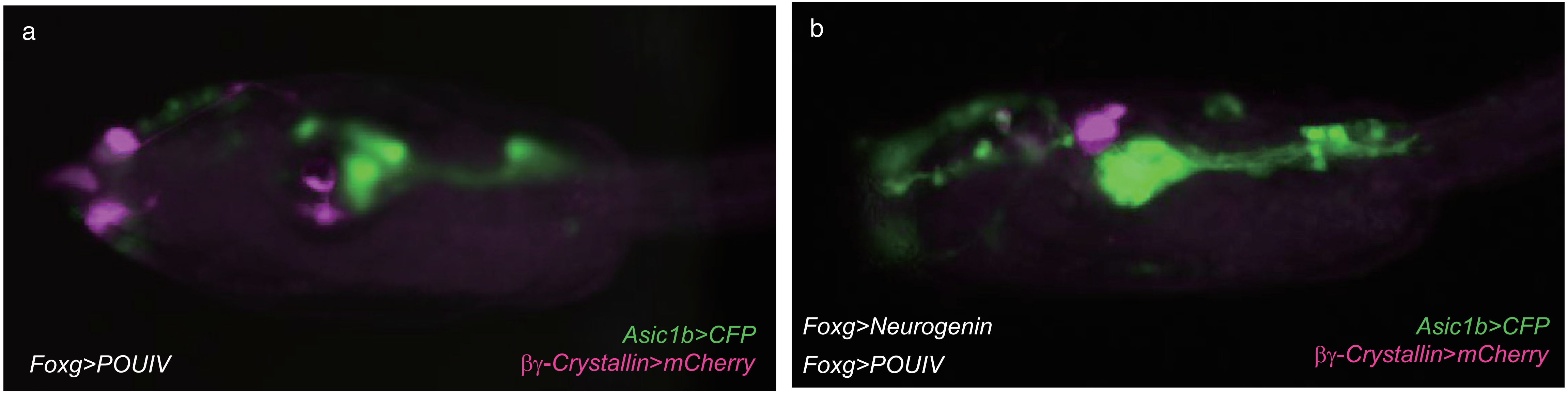


Fig. S6. Misexpression of POUIV and Neurogenin induces more complete transformation of PSCs to BTNs.
(a) *Foxg>POU IV*, like *Foxg>Neurogenin* (**Fig. 4 g-h**) induces the expression of Asic1b in PSCs and Bγ-Crystallin expression is maintained, demonstrating the hybrid phenotype.
(b) *Foxg>POUIV+Neurogenin* injected larva express Asci1b and lose the expression of Bγ-Crystallin.

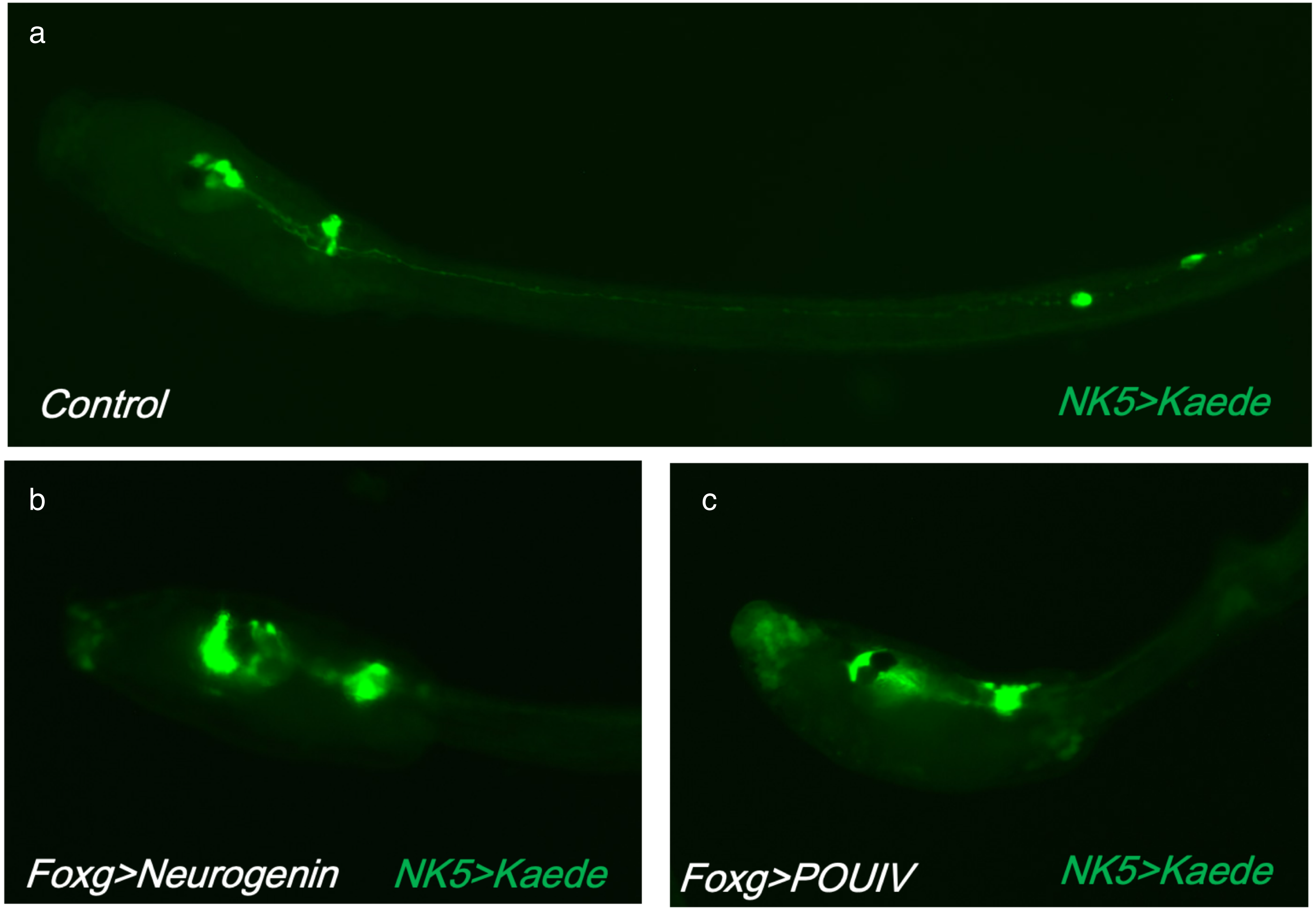
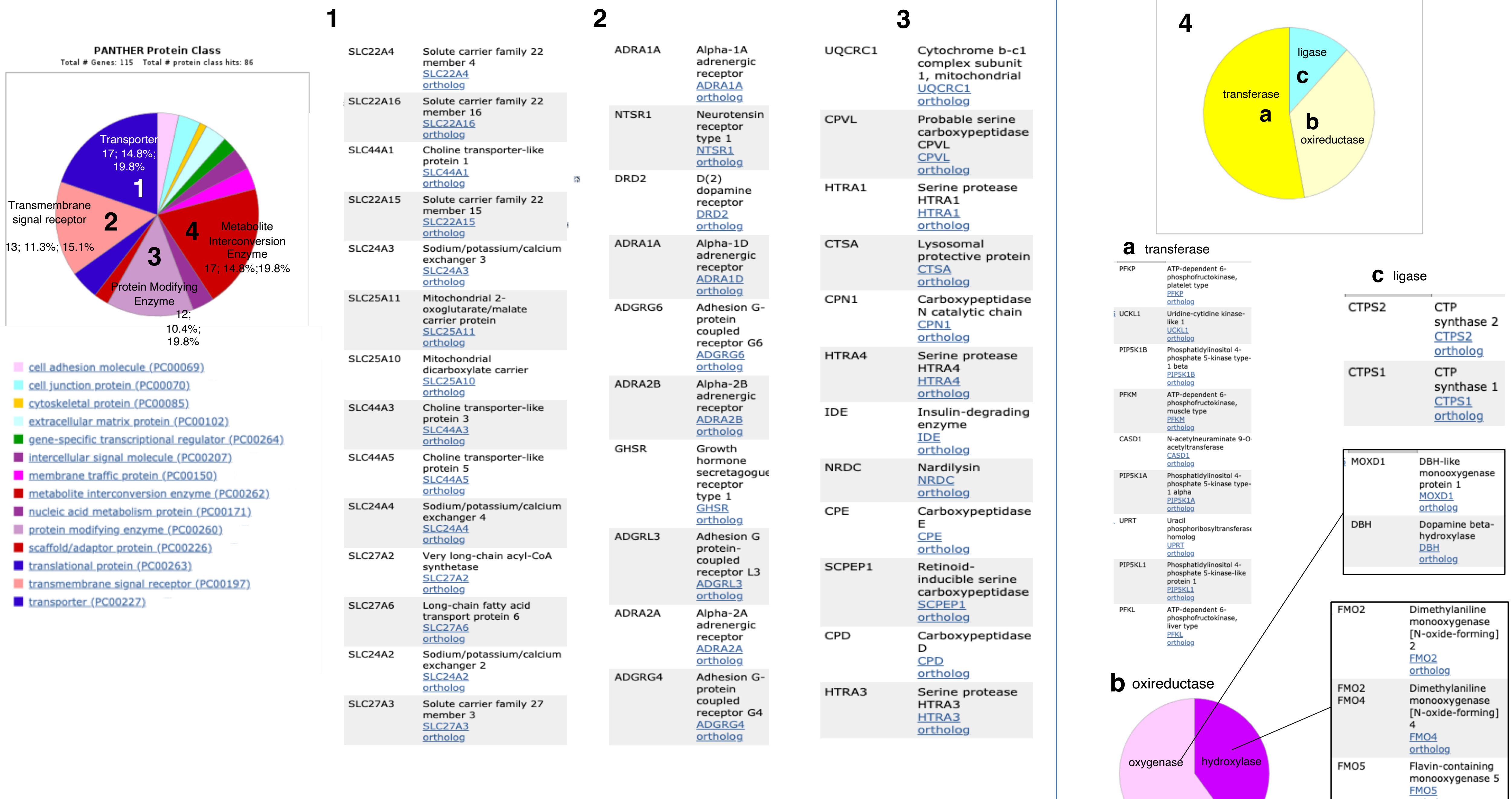


Fig. S7. Misexpression of POU IV and Neurogenin in PSCs induce the expression of *NK5* in PSCs. (a) Kaede fluorescence in *NK5>Kaede* injected larva. **(b)** Kaede fluorescence in *Foxg>Neurogenin* and *NK5>Kaede* injected larva. **(c)** Kaede fluorescence in *Foxg>POUIV* and *NK5>Kaede* injected larva. Misexpression of POU IV and Neurogenin in PSCs induce the expression of NK5, another BTN marker, in PSCs.



1

SLC22A4	Solute carrier family 22 member 4 SLC22A4 ortholog
SLC22A16	Solute carrier family 22 member 16 SLC22A16 ortholog
SLC44A1	Choline transporter-like protein 1 SLC44A1 ortholog
SLC22A15	Solute carrier family 22 member 15 SLC22A15 ortholog
SLC24A3	Sodium/potassium/calcium exchanger 3 SLC24A3 ortholog
SLC25A11	Mitochondrial 2-oxoglutarate/malate carrier protein SLC25A11 ortholog
SLC25A10	Mitochondrial dicarboxylate carrier SLC25A10 ortholog
SLC44A3	Choline transporter-like protein 3 SLC44A3 ortholog
SLC44A5	Choline transporter-like protein 5 SLC44A5 ortholog
SLC24A4	Sodium/potassium/calcium exchanger 4 SLC24A4 ortholog
SLC27A2	Very long-chain acyl-CoA synthetase SLC27A2 ortholog
SLC27A6	Long-chain fatty acid transport protein 6 SLC27A6 ortholog
SLC24A2	Sodium/potassium/calcium exchanger 2 SLC24A2 ortholog
SLC27A3	Solute carrier family 27 member 3 SLC27A3 ortholog

2

ADRA1A	Alpha-1A adrenergic receptor ADRA1A ortholog
NTSR1	Neurotensin receptor type 1 NTSR1 ortholog
DRD2	D(2) dopamine receptor DRD2 ortholog
ADRA1A	Alpha-1D adrenergic receptor ADRA1D ortholog
ADGRG6	Adhesion G-protein coupled receptor G6 ADGRG6 ortholog
ADRA2B	Alpha-2B adrenergic receptor ADRA2B ortholog
GHSR	Growth hormone secretagogue receptor type 1 GHSR ortholog
ADGRL3	Adhesion G protein-coupled receptor L3 ADGRL3 ortholog
ADRA2A	Alpha-2A adrenergic receptor ADRA2A ortholog
ADGRG4	Adhesion G-protein coupled receptor G4 ADGRG4 ortholog

3

UQCRC1	Cytochrome b-c1 complex subunit 1, mitochondrial UQCRC1 ortholog
CPVL	Probable serine carboxypeptidase CPVL CPVL ortholog
HTRA1	Serine protease HTRA1 HTRA1 ortholog
CTSA	Lysosomal protective protein CTSA ortholog
CPN1	Carboxypeptidase N catalytic chain CPN1 ortholog
HTRA4	Serine protease HTRA4 HTRA4 ortholog
IDE	Insulin-degrading enzyme IDE ortholog
NRDC	Nardilysin NRDC ortholog
CPE	Carboxypeptidase E CPE ortholog
SCPEP1	Retinoid-inducible serine carboxypeptidase SCPEP1 ortholog
CPD	Carboxypeptidase D CPD ortholog
HTRA3	Serine protease HTRA3 HTRA3 ortholog

a transferase

PFKP	ATP-dependent 6-phosphofructokinase, platelet type PFKP ortholog
UCKL1	Uridine-cytidine kinase-like 1 UCKL1 ortholog
PIP5K1B	Phosphatidylinositol 4-phosphate 5-kinase type-1 beta PIP5K1B ortholog
PFKM	ATP-dependent 6-phosphofructokinase, muscle type PFKM ortholog
CASD1	N-acetylneuraminate 9-O-acetyltransferase CASD1 ortholog
PIP5K1A	Phosphatidylinositol 4-phosphate 5-kinase type-1 alpha PIP5K1A ortholog
UPRT	Uracil phosphoribosyltransferase homolog UPRT ortholog
PIP5KL1	Phosphatidylinositol 4-phosphate 5-kinase-like protein 1 PIP5KL1 ortholog
PFKL	ATP-dependent 6-phosphofructokinase, liver type PFKL ortholog

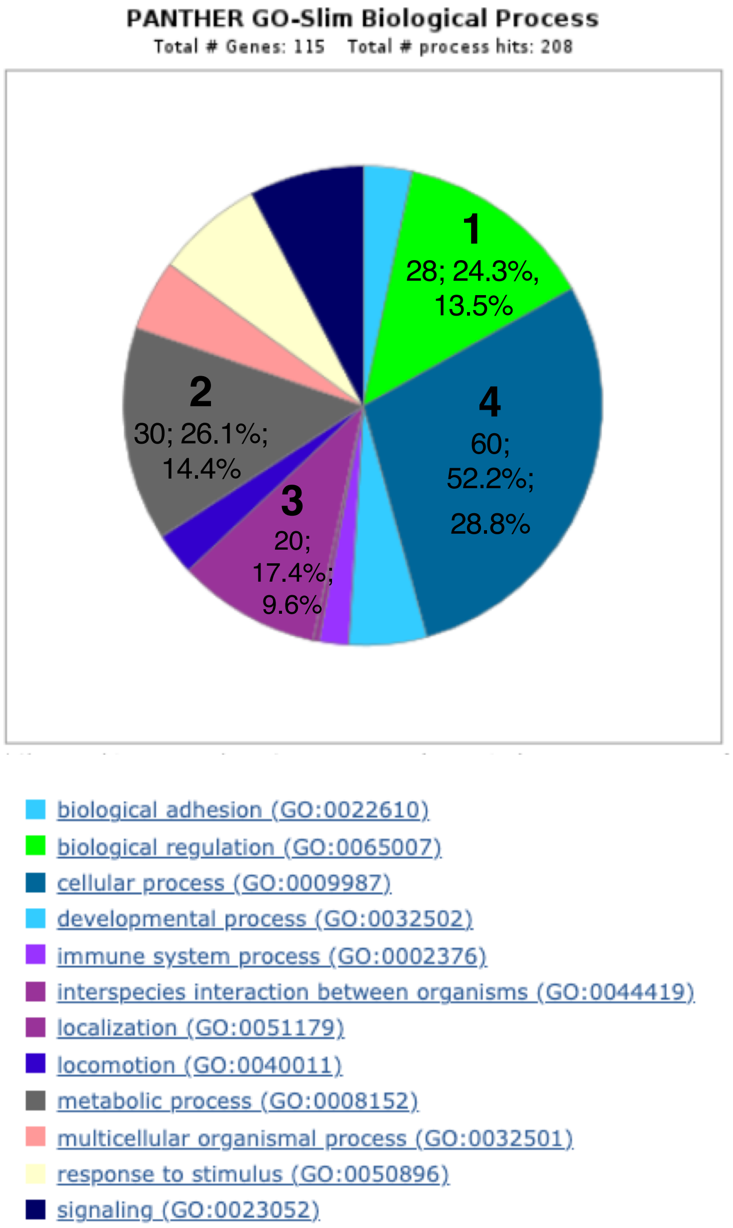
c ligase

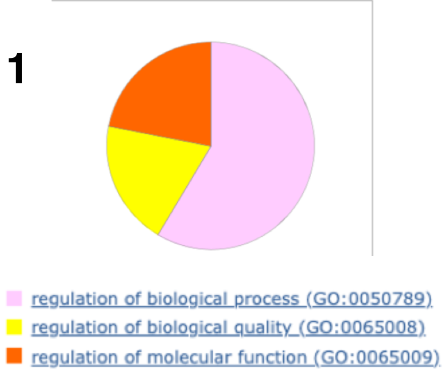
CTPS2	CTP synthase 2 CTPS2 ortholog
CTPS1	CTP synthase 1 CTPS1 ortholog

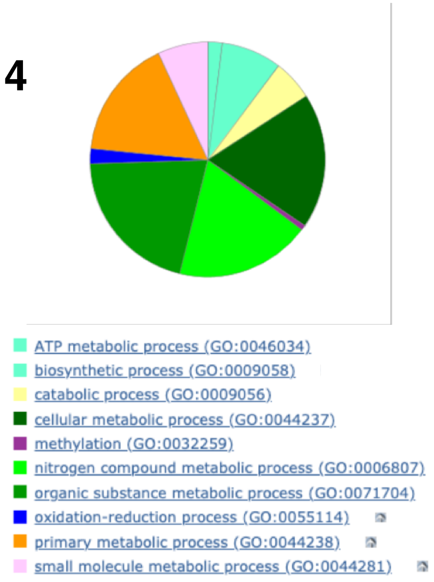
MOXD1	DBH-like monooxygenase protein 1 MOXD1 ortholog
DBH	Dopamine beta-hydroxylase DBH ortholog

FMO2	Dimethylaniline monooxygenase [N-oxide-forming] 2 FMO2 ortholog
FMO2 FMO4	Dimethylaniline monooxygenase [N-oxide-forming] 4 FMO4 ortholog
FMO5	Flavin-containing monooxygenase 5 FMO5 ortholog

Fig. S8. PANTHER Protein Class classification of human orthologs of OE DEGs.
genes; Percent of gene hit against total # genes; Percent of gene hit against total # Protein Class hits







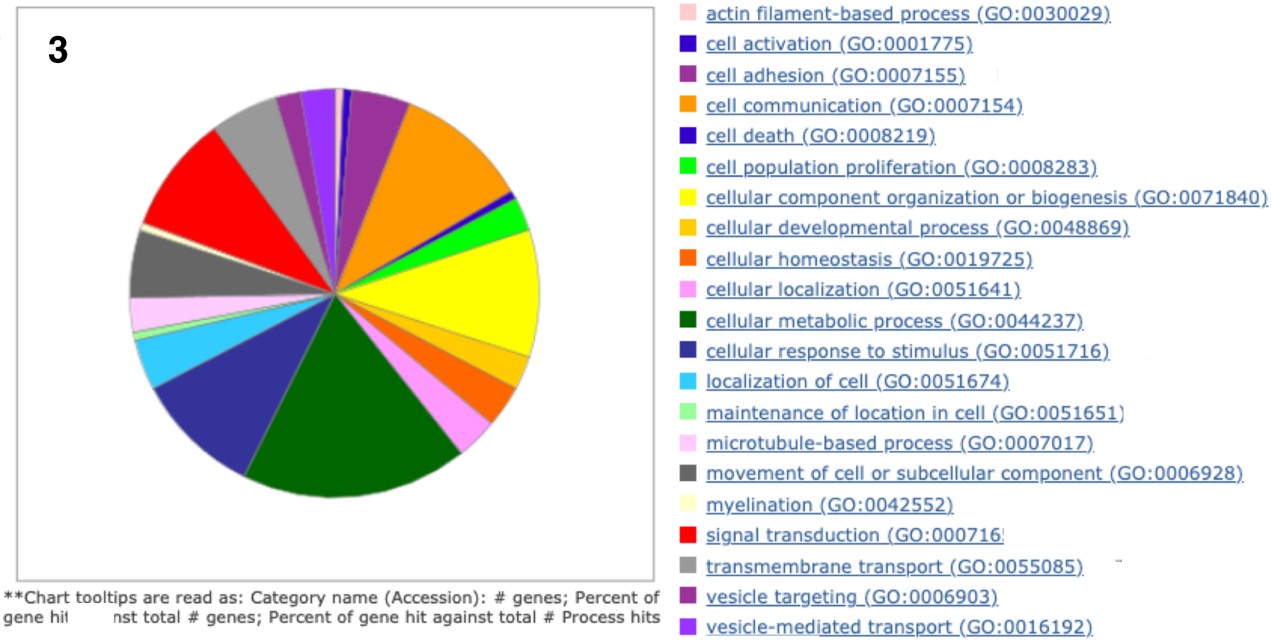


**Chart tooltips are read as: Category name (Accession): # genes; Percent of gene hit nst total # genes; Percent of gene hit against total # Process hits

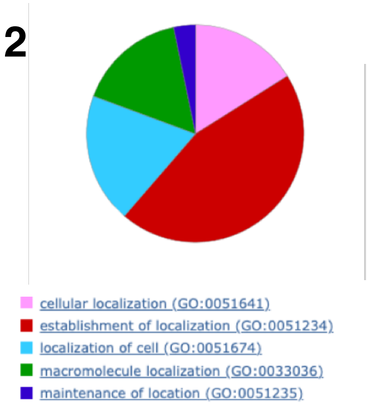


Fig. S9. PANTHER Biological Process classification of human orthologs of OE DEGs.
genes; Percent of gene hit against total # genes; Percent of gene hit against total # Protein Class hits

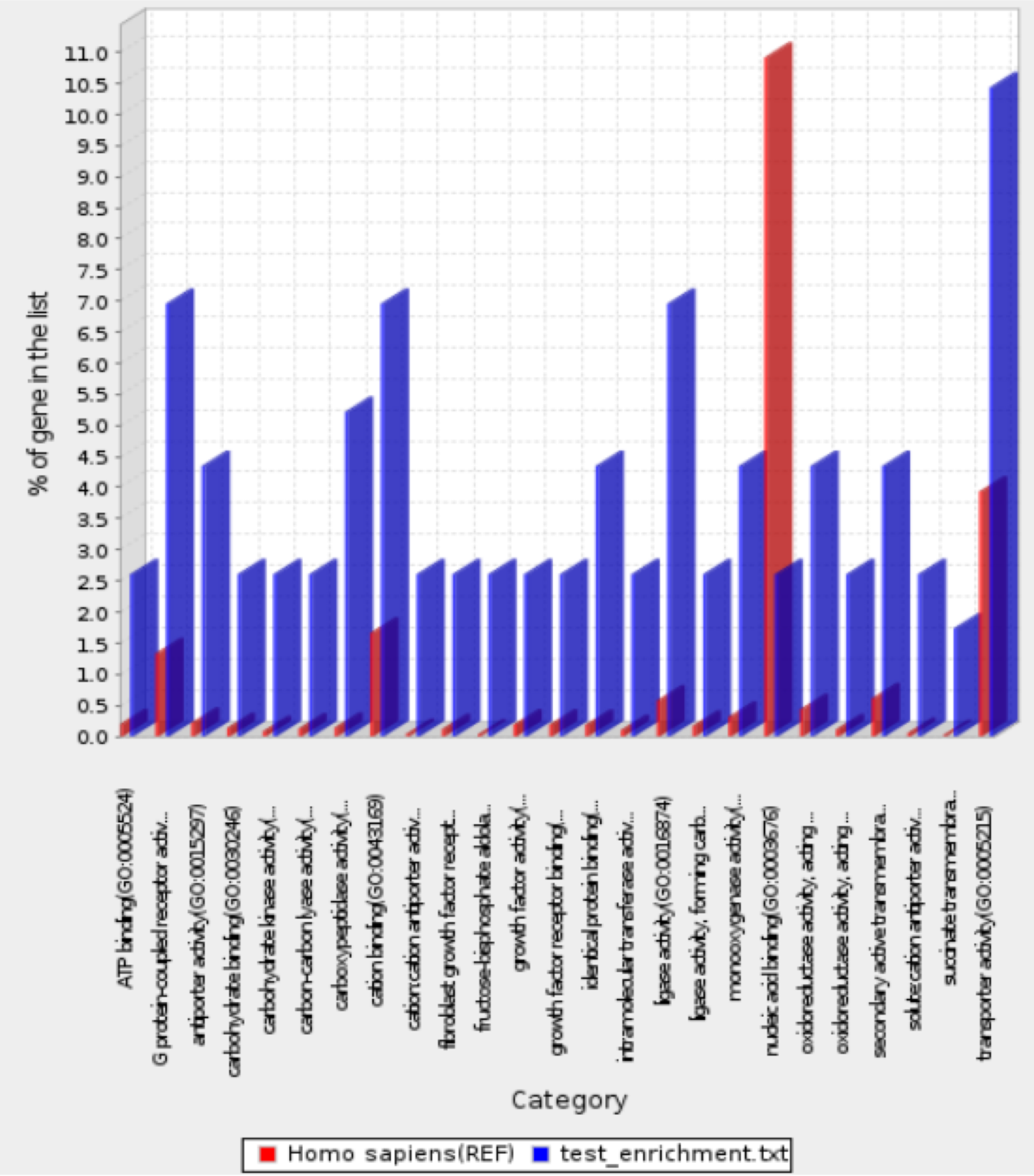


Fig. S10. Results of Statistical Overrepresentation of **PANTHER GO-Slim Molecular Function terms.**

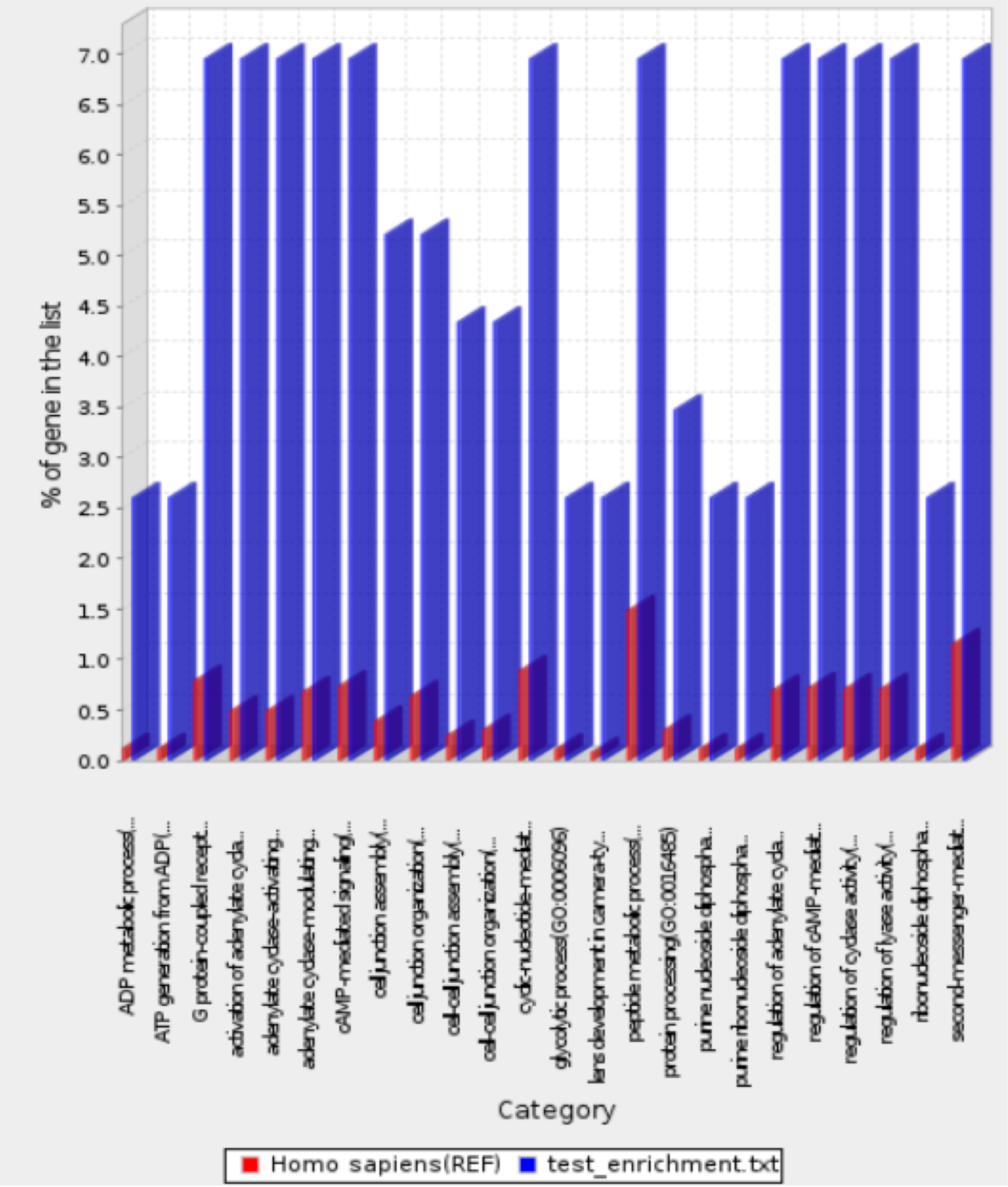


Fig. S11. Results of Statistical Overrepresentation of **PANTHER GO-Slim Biological Process terms.**

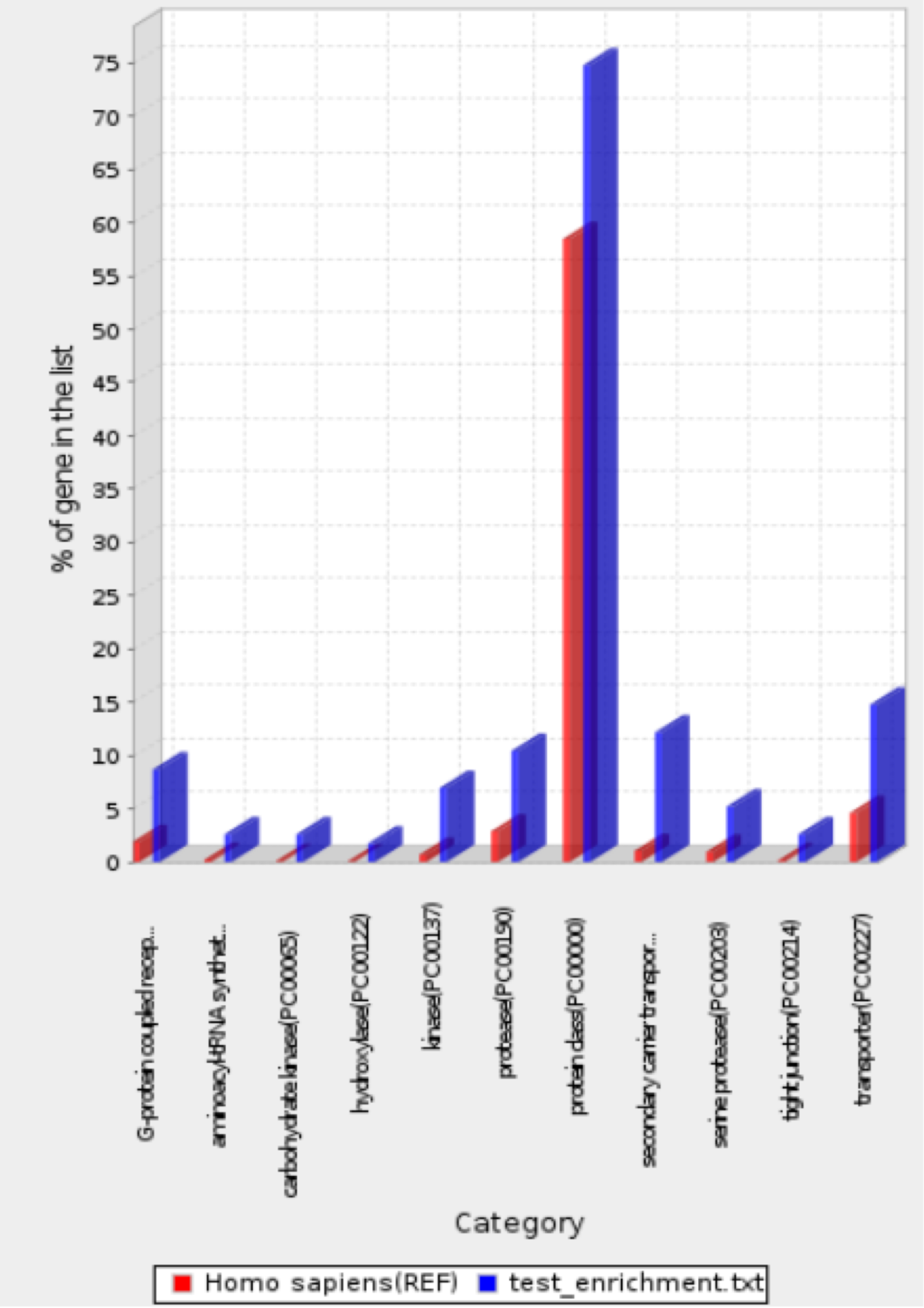
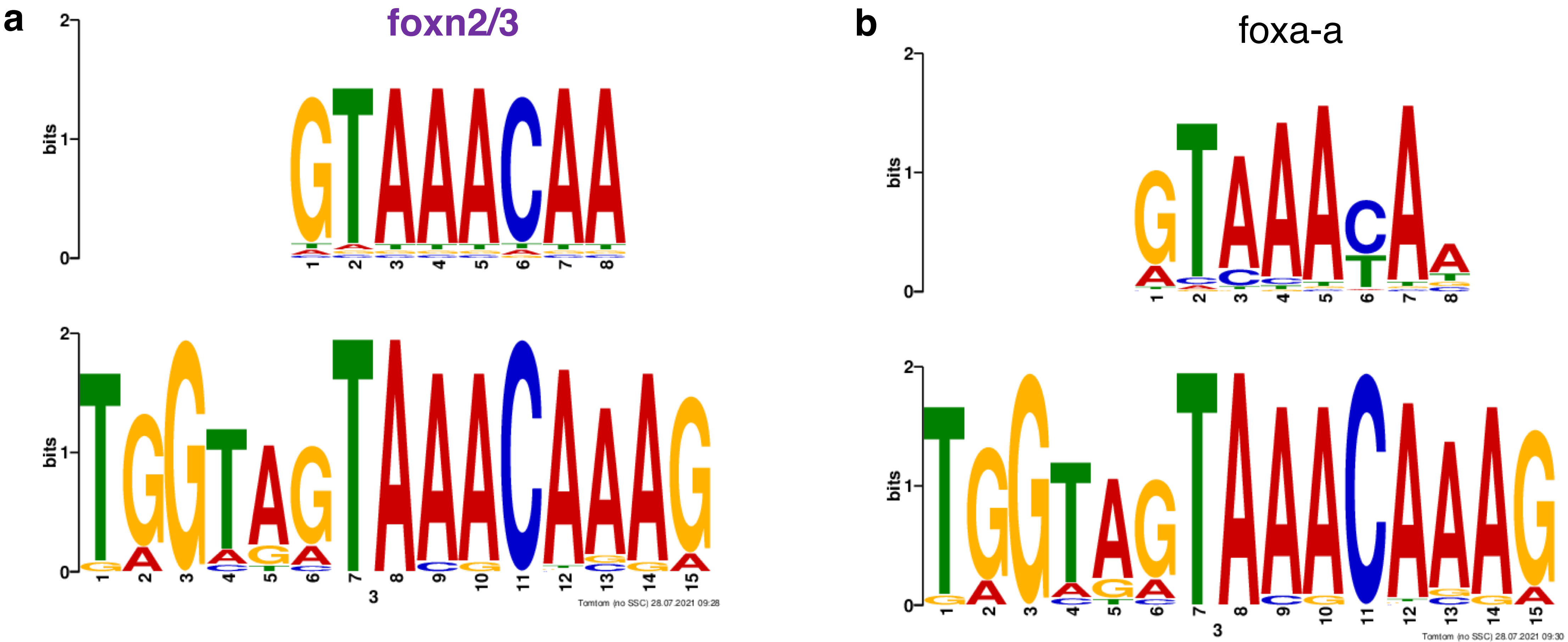


Fig. S12. Results of Statistical Overrepresentation **PANTHER GO- Protein Class terms.**

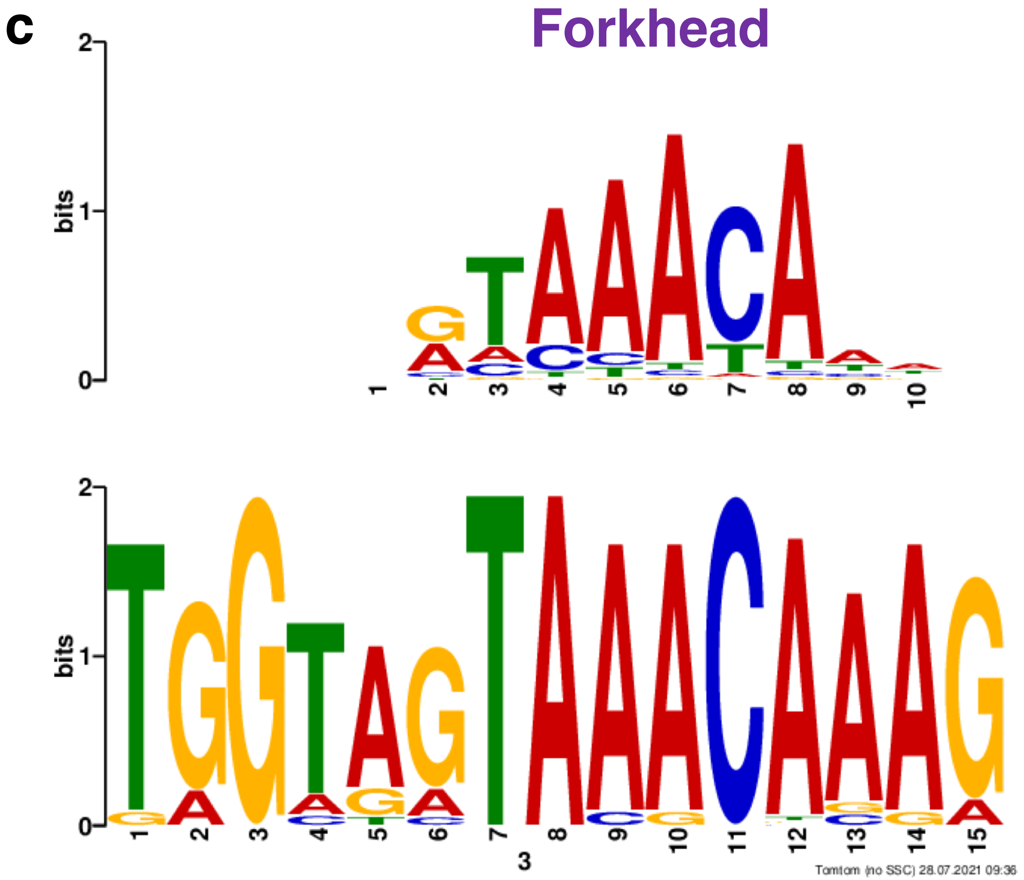
Fig. S13. Discovered motif from putative cis-regulatory regions of Novel Genes.



p-value	5.93e-07
E-value	9.30e-05
q-value	1.81e-04

p-value	1.19e-05
E-value	1.87e-03
q-value	1.69e-03

Fig. S14a, b. Matches (Transcription Factors in Ciona) to discovered motif and respective statistical significance.

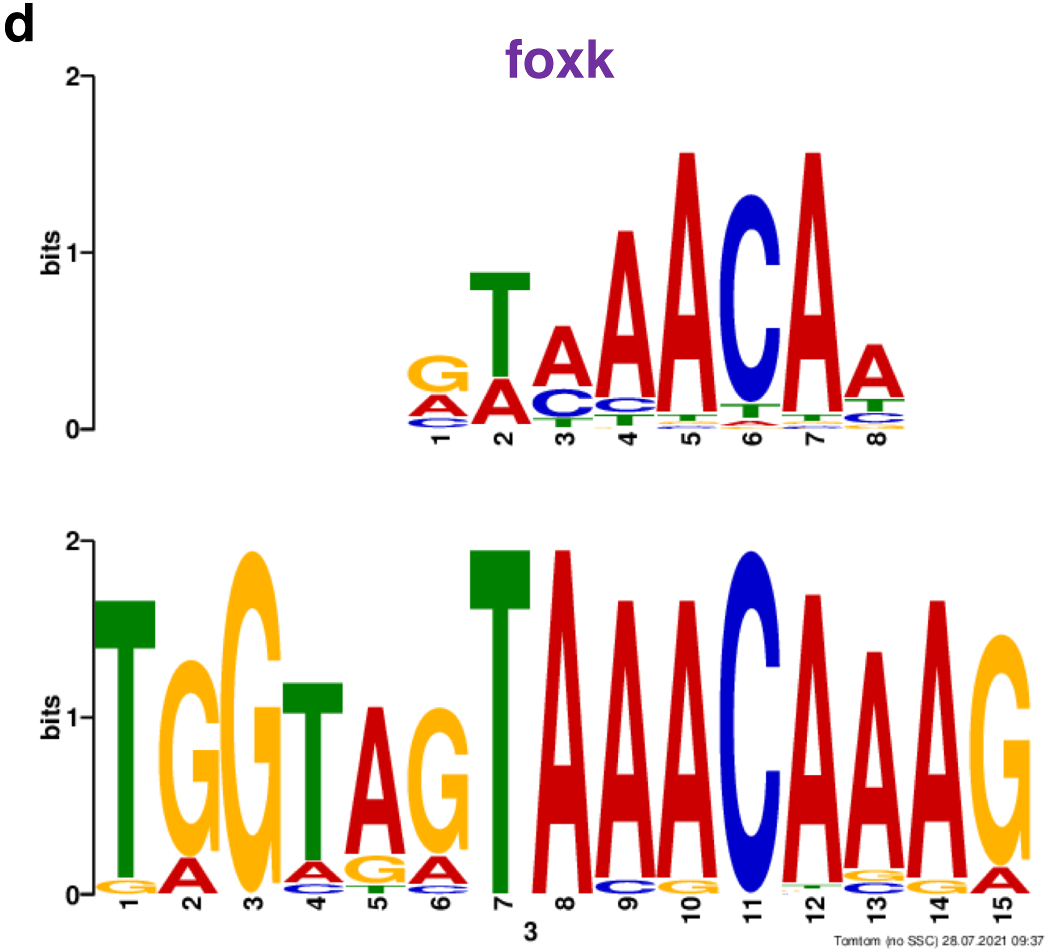


p-value 1.66e-05

E-value 2.60e-03

q-value 1.69e-03

<u>Motifs from Related TFs in *H. sapiens*</u>
FOXJ2/3, FOXK1, FOXI1

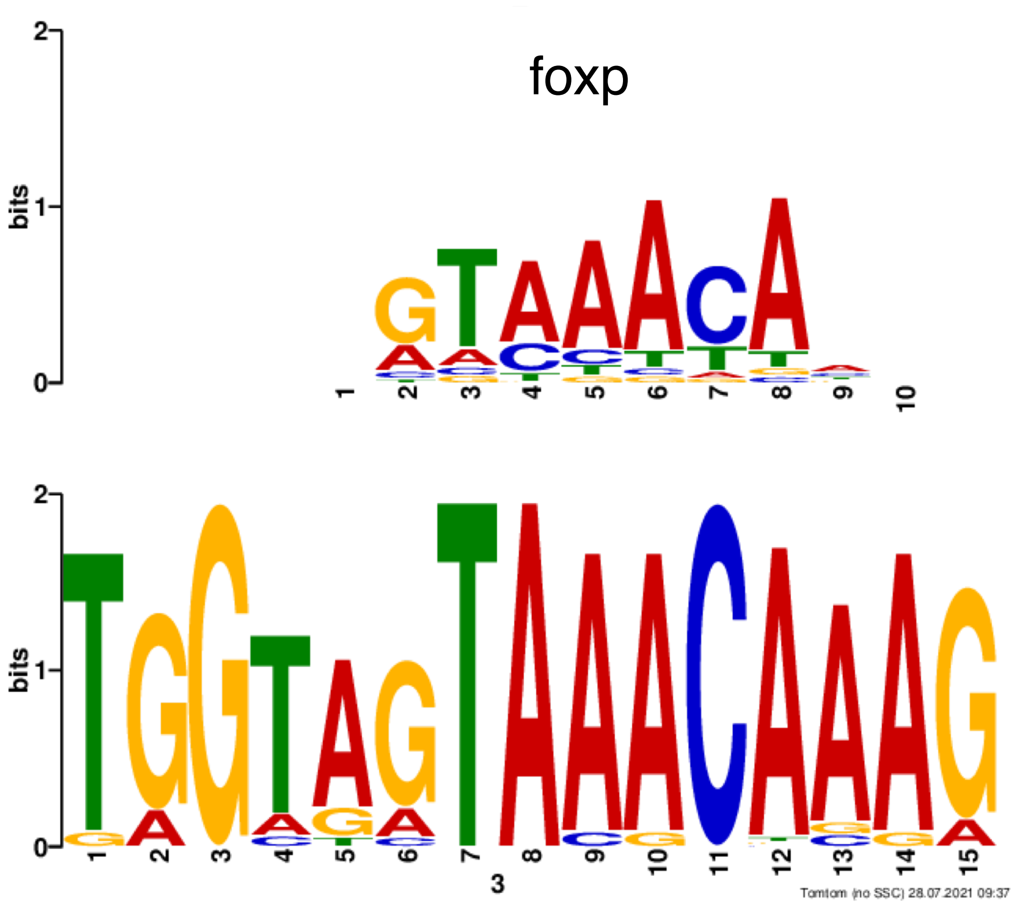


p-value 3.66e-05

E-value 5.75e-03

q-value 2.80e-03

Fig. S14c, d. Match to discovered motif and respective statistical significance. Since there wasn't an exact match to a Transcription Factor in Ciona, similar matches in Human are reported.



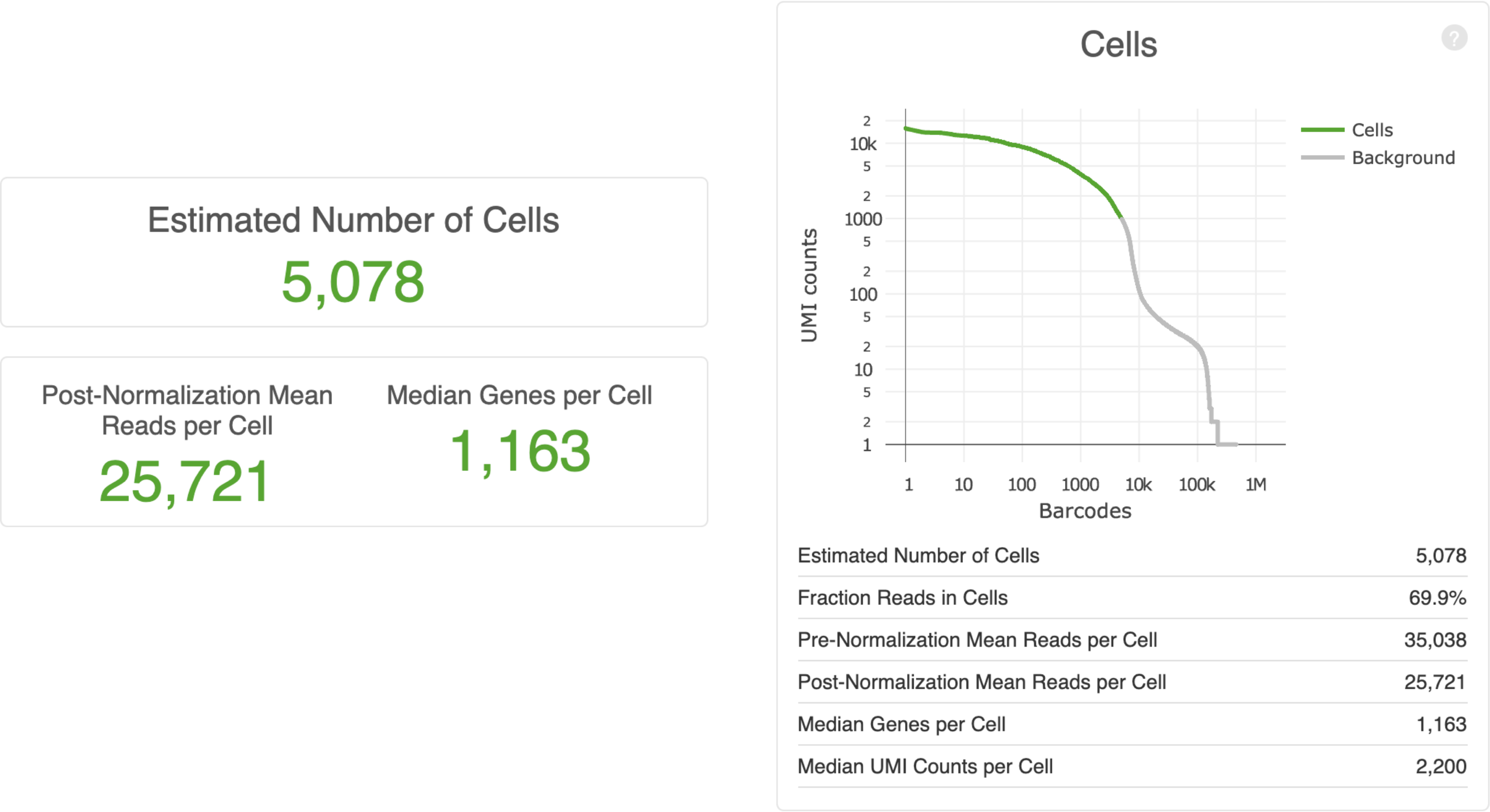
p-value	6.85e-04
E-value	1.08e-01
q-value	4.19e-02

Fig. S14e. Match (Transcription Factor in Ciona) to discovered motif and respective statistical significance.

List S1. Forkhead TFs that are significantly more highly expressed in ME vs WT Epidermis

They are in purple and bold font in Figs. S6a-e.

"FOXJ1" "FOXK1" "FOXK2" "FOXE3" "FOXI2" "FOXI3" "FOXF1" "FOXF2" "FOXL1" "FOXA2" "FOXB1" "FOXB2" "FOXK1" "FOXK2" "FOXD1" "FOXD3" "FOXD4" "FOXI1" "FOXI2" "FOXI3" "FOXG1" "FOXI1" "FOXI2" "FOXF1" "FOXH1" "FOXQ1" "FOXI3" "FOXJ2" "FOXJ3" "FOXN1" "FOXN2" "FOXN3" "FOXO1" "FOXO3" "FOXO4" "FOXF1" "FOXH1" "FOXQ1" "FOXF1" "FOXF2" "FOXQ1"



Estimated Number of Cells

5,078

Post-Normalization Mean Reads per Cell

25,721

Median Genes per Cell

1,163

Estimated Number of Cells	5,078
Fraction Reads in Cells	69.9%
Pre-Normalization Mean Reads per Cell	35,038
Post-Normalization Mean Reads per Cell	25,721
Median Genes per Cell	1,163
Median UMI Counts per Cell	2,200

Fig. S15. 10X Cell Ranger Summary Statistics of Single-cell Data Acquisition. The expected number of cells of the Ciona tadpole at MTB stage is ~1600. These statistics demonstrate coverage of ~3x of the embryo.

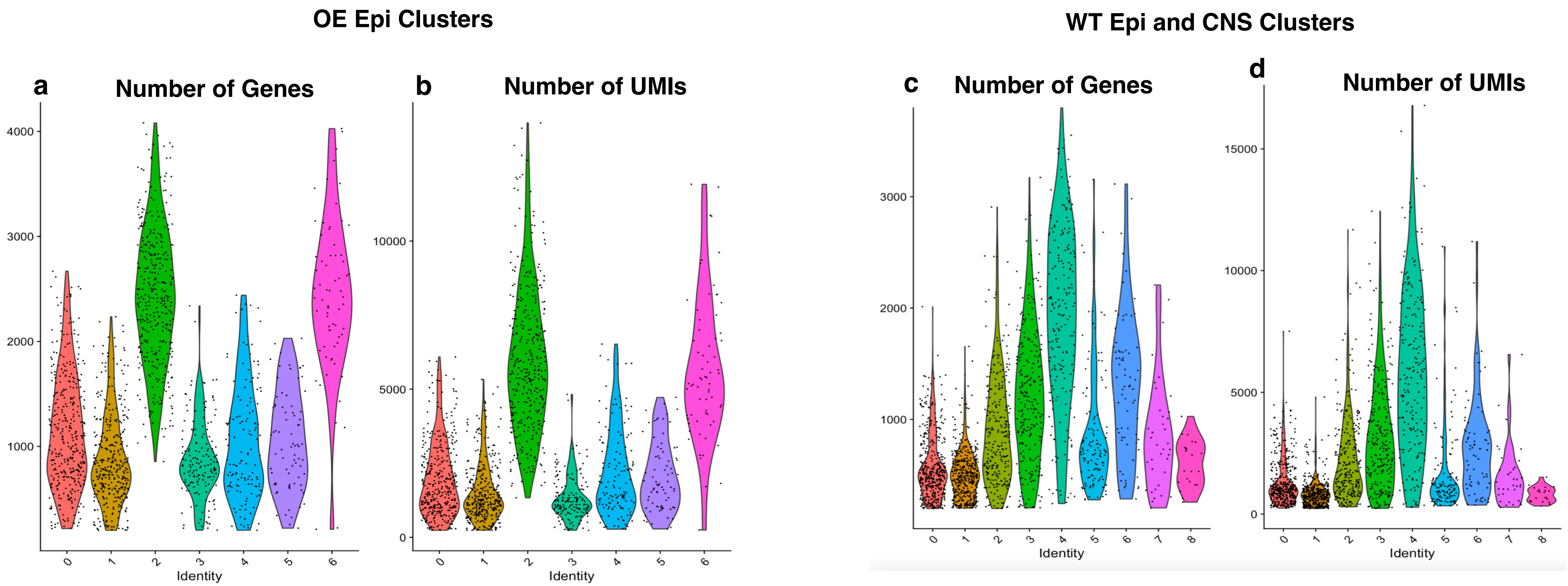


Fig. S16. Number of Genes and UMIs in various OE Epi (a, b) and WT Epi and CNS (c, d) clusters.
Note: All Epidermal and Sensory Neuron types were identified from various Epi and CNS clusters.

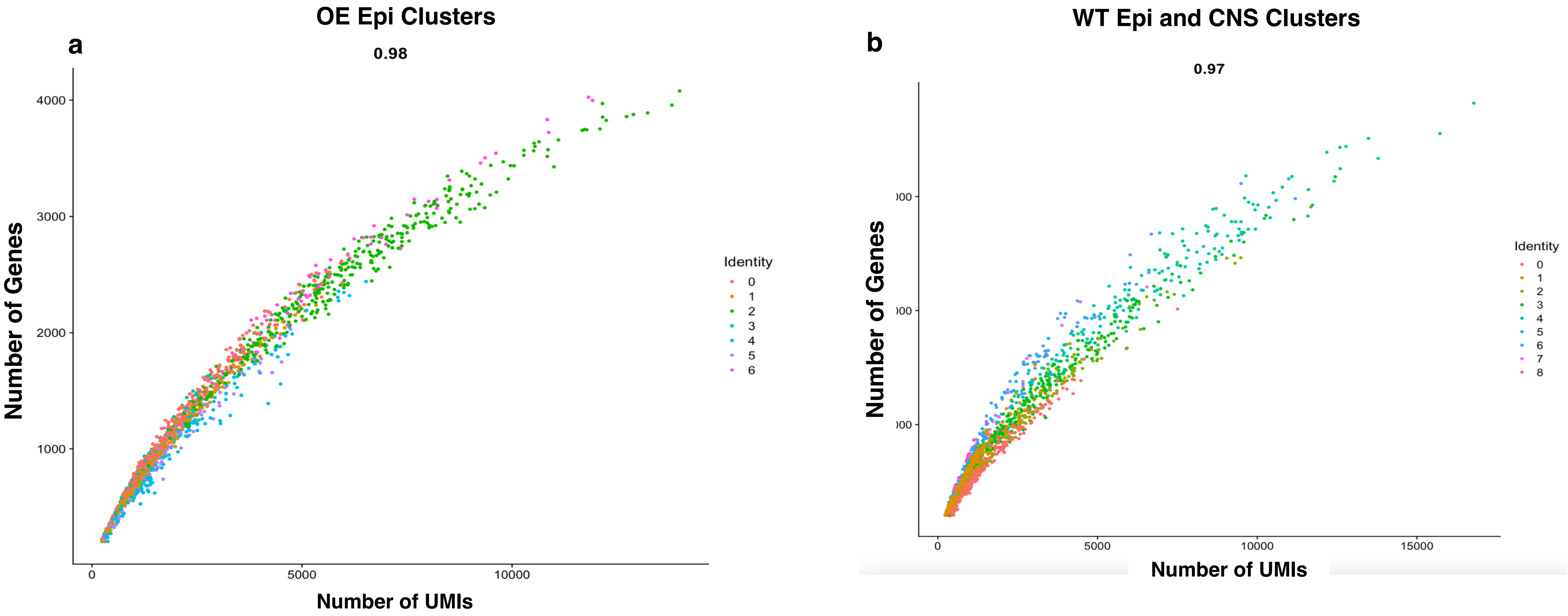


Fig. S17. Scatterplot and Correlations of Number of Genes and Number of UMIs in OE Epi (a) and WT Epi and CNS (b) clusters.

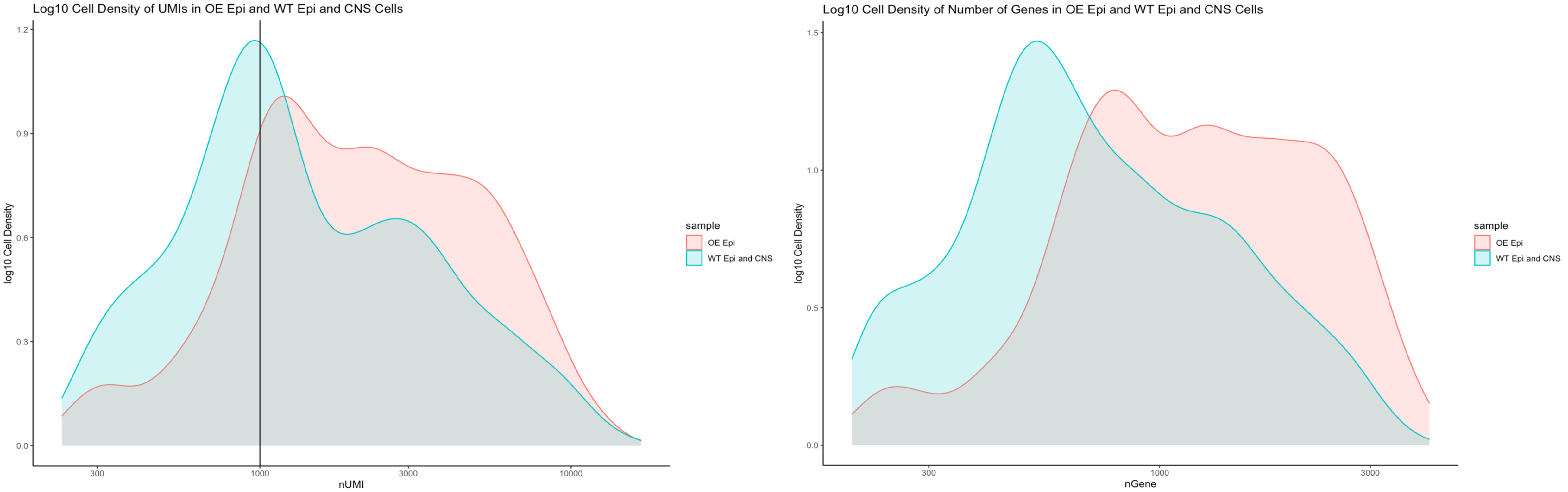


Fig. S18. Log10 Cell Density of UMIs (a) and Number of Genes (b) in OE Epi and WT Epi and CNS Cells. Vertical line is drawn at 1000 nUMIs to demonstrate high coverage of transcripts in both WT and OE cell populations.

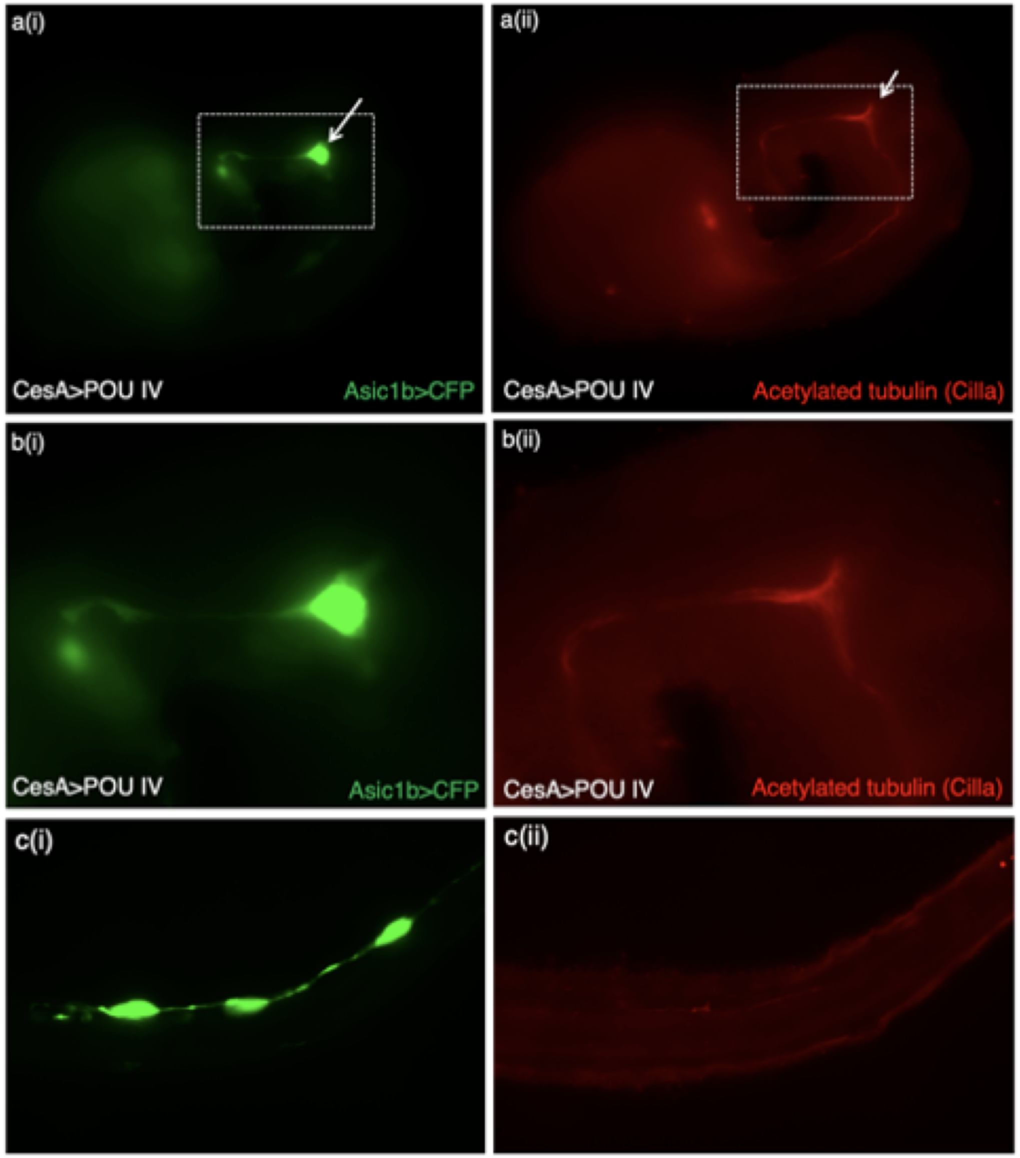


Fig. S19. Misexpression of POU IV induces the differentiation of synthetic cell types, which have both properties of BTNs and PSCs (ai-ii) CFP fluorescence (green) and acetylated tubulin immunofluorescence (red) in the *CesA>POU IV* and *Ascic1b>CFP* injected larvae. A BTN marker gene, Asic1b is expressed in the PSCs in *CesA>POU IV* misexpressed larvae. **(bi-ii)** are zoomed-in images of (ai-ii).This cell shows hybrid properties of BTNs (has long axons) and PSCs (has cilia). **(c)** BTNs do not have cilia.