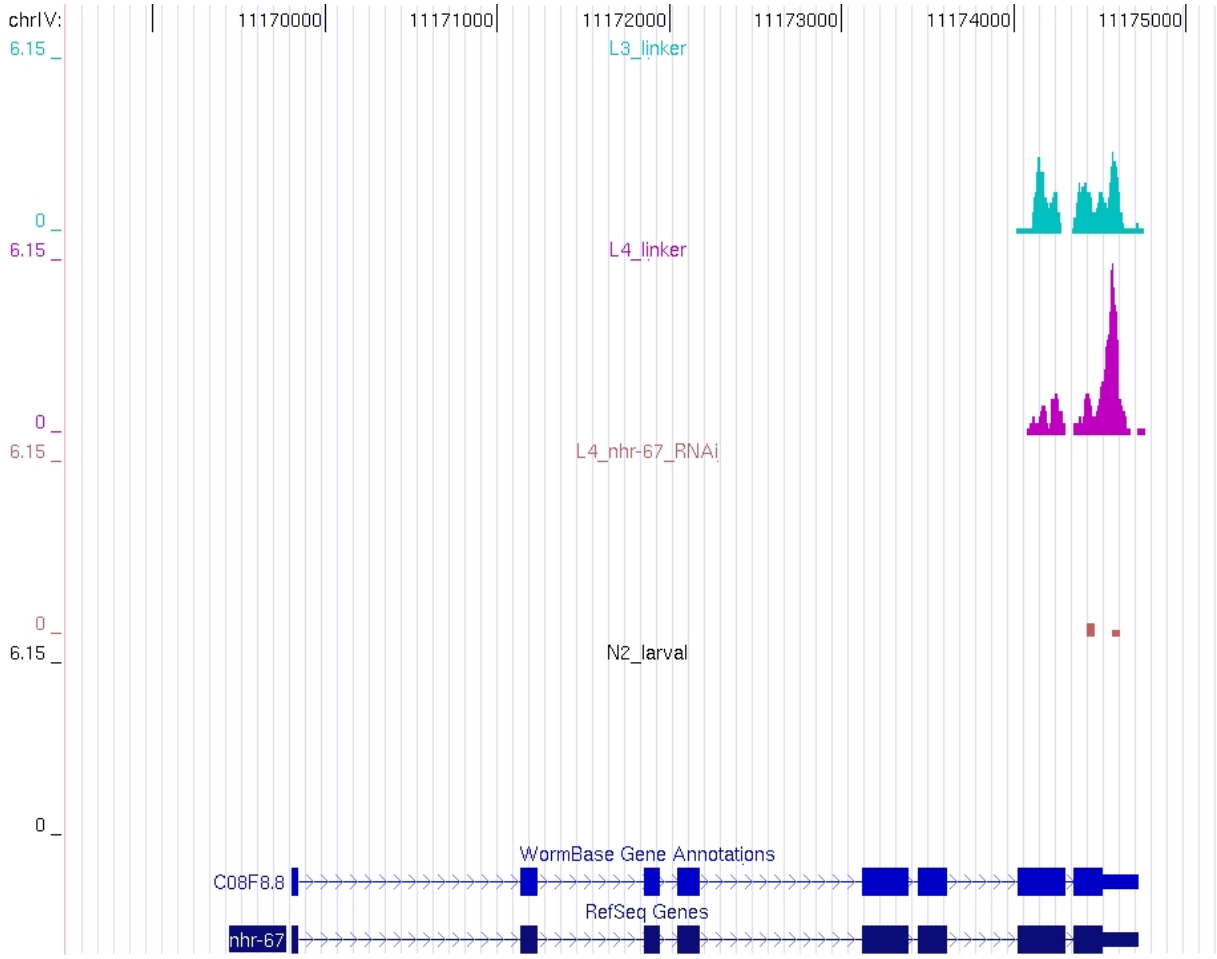
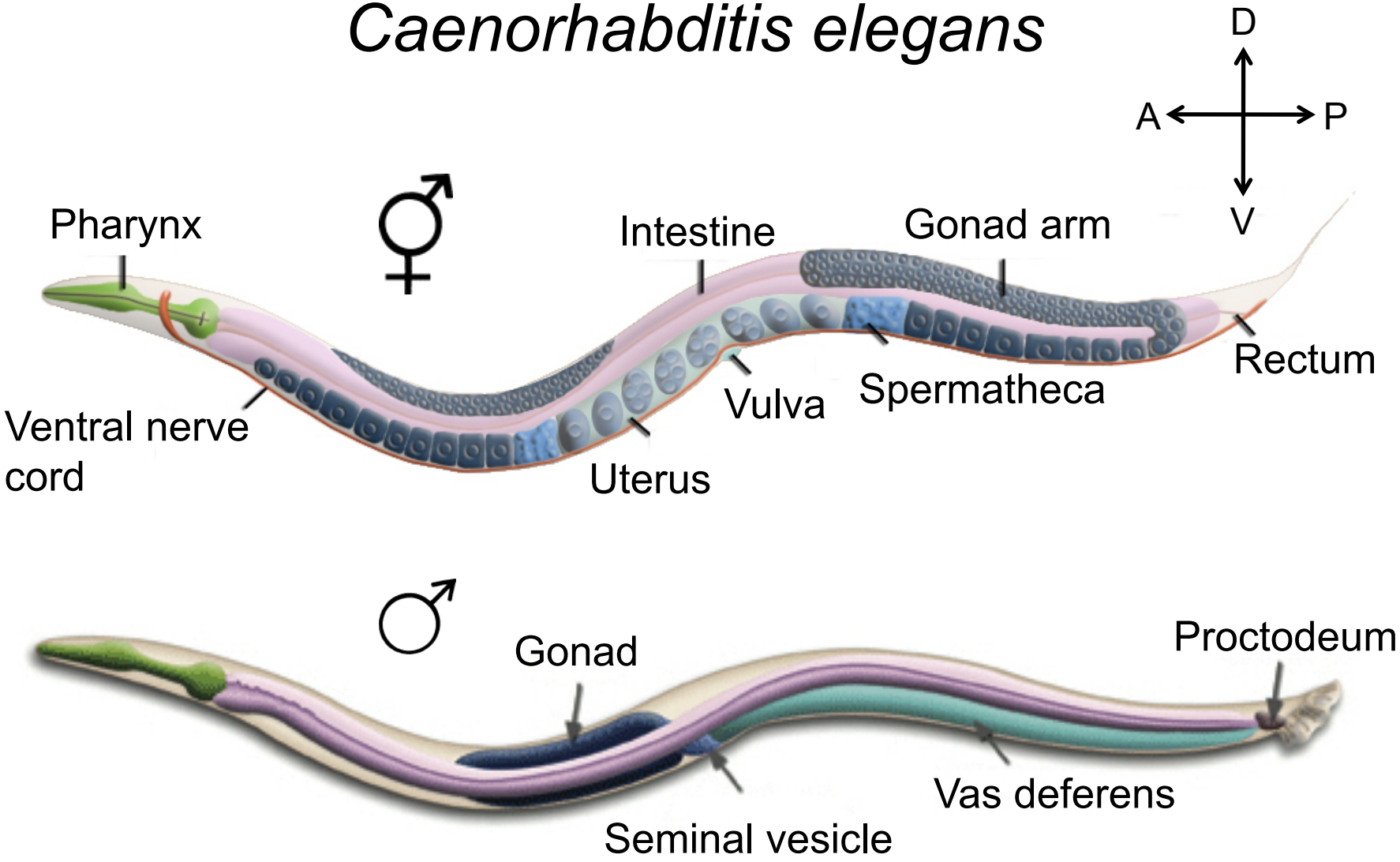
Single-cell RNA-seq in *C. elegans*

Erich Schwarz, Cornell



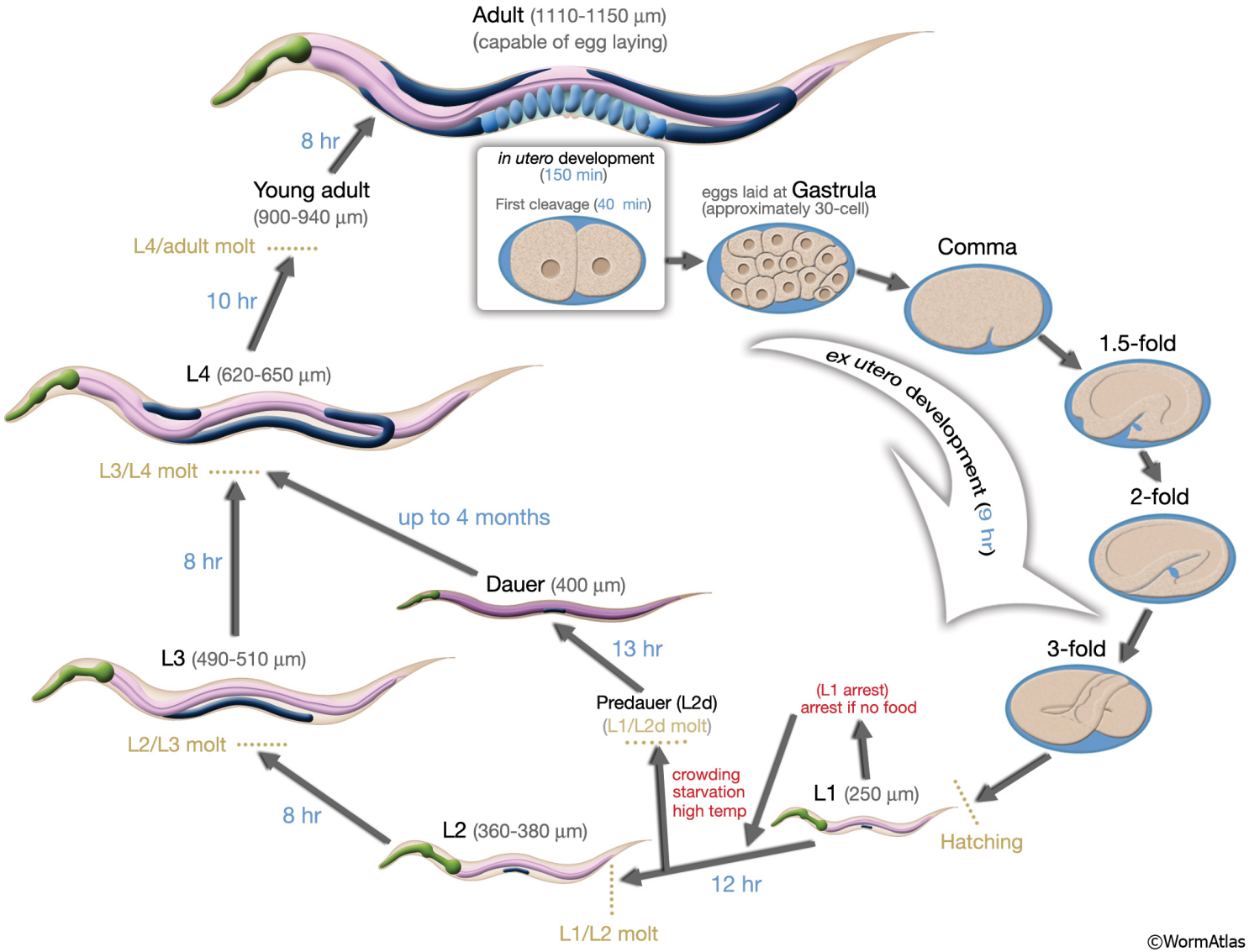
MSU NGS Course, June 2013

Caenorhabditis elegans



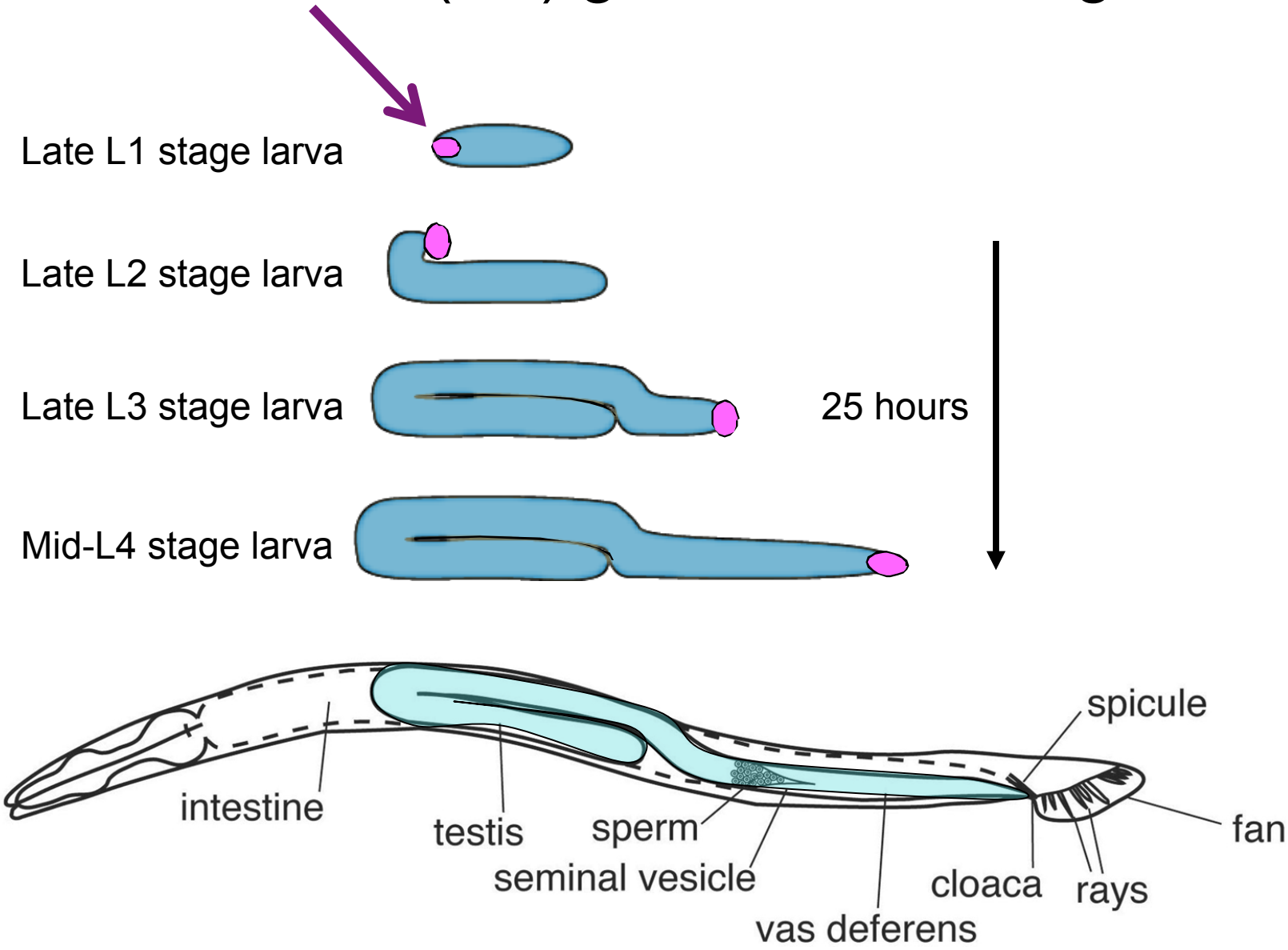
Ref.: Hall (2013), *www.wormatlas.org*.

Caenorhabditis elegans



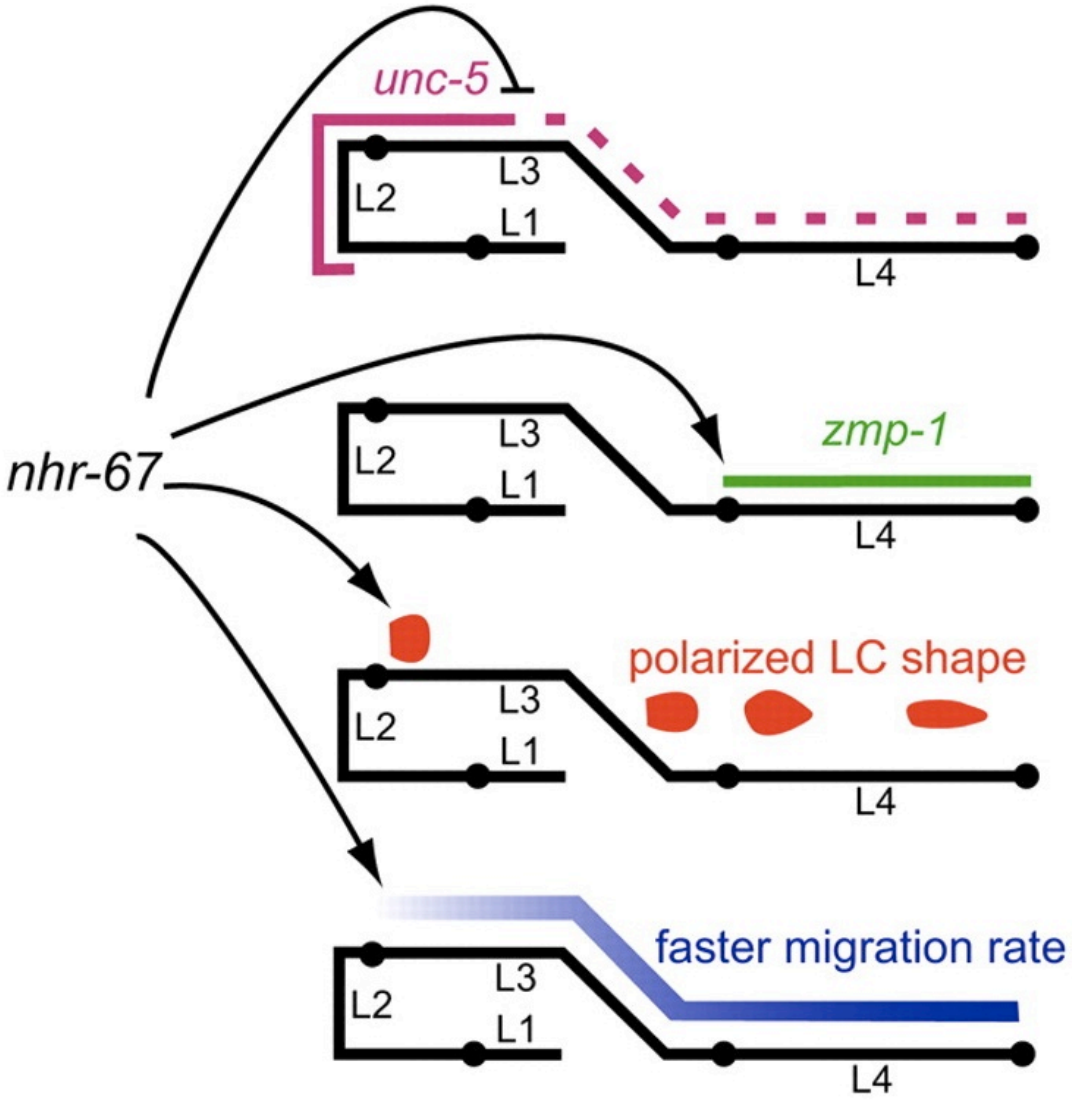
Ref.: Hall (2013), *www.wormatlas.org*.

The linker cell (LC) guides the male gonad



Ref.: Kato and Sternberg (2009), Development, *136*, 3907-3915.

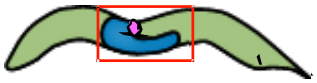
Linker cells partially require *nhr-67* (*tailless*/Tlx)

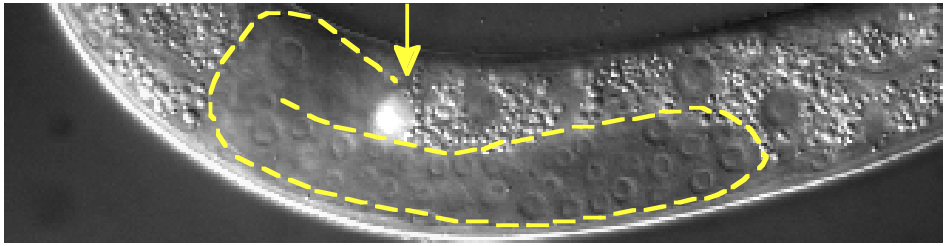


Ref.: Kato and Sternberg (2009), Development, *136*, 3907-3915.

LC samples harvested for RNA-seq

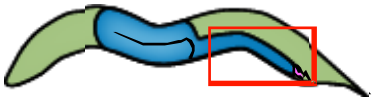
L3 stage

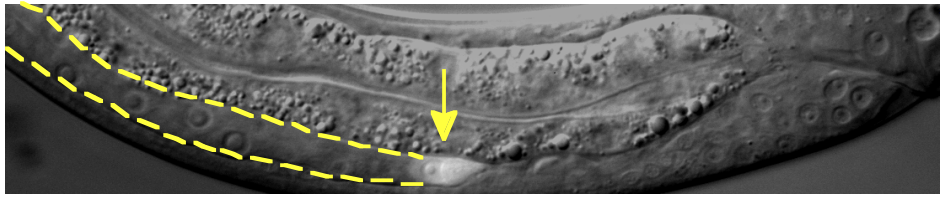




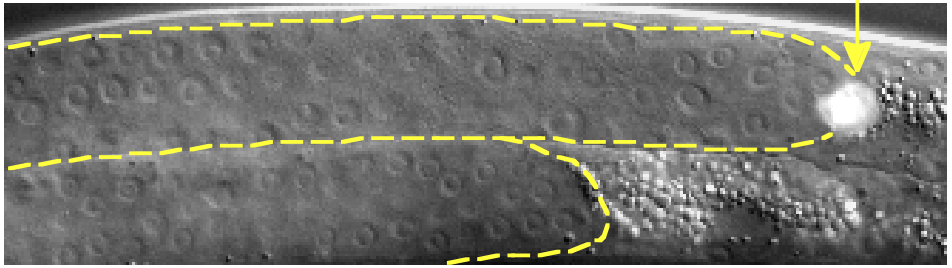
LC is labeled by YFP

L4 stage





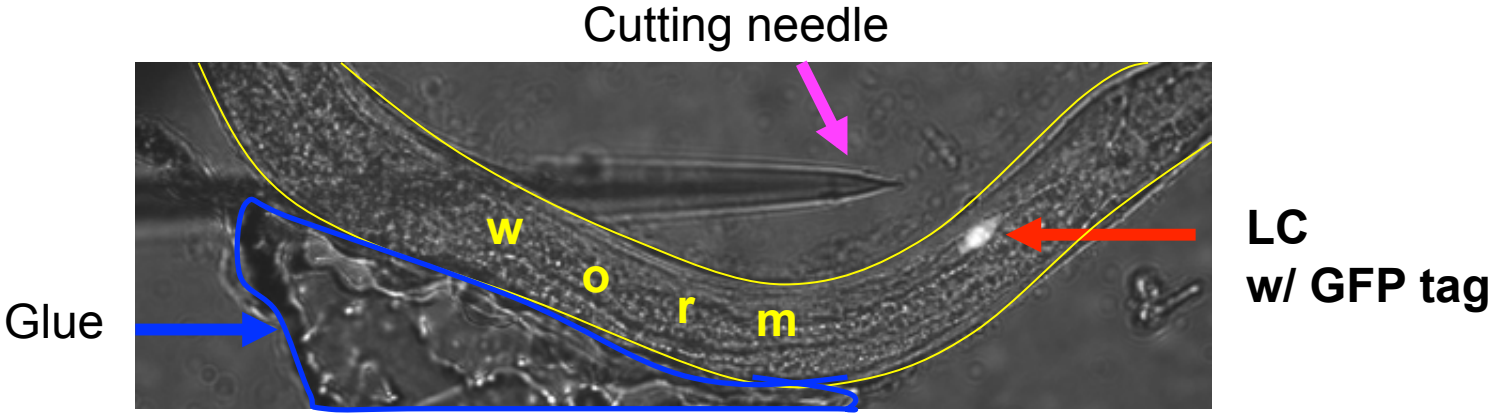
L4, *nhr-67*(-)

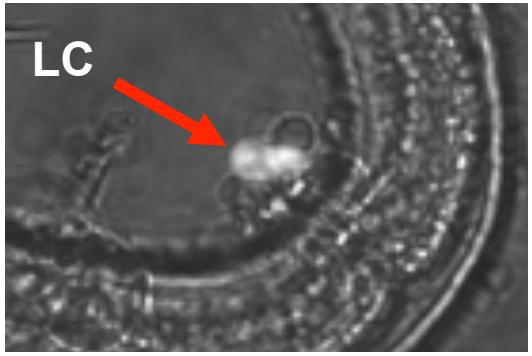


Whole L1+ larvae

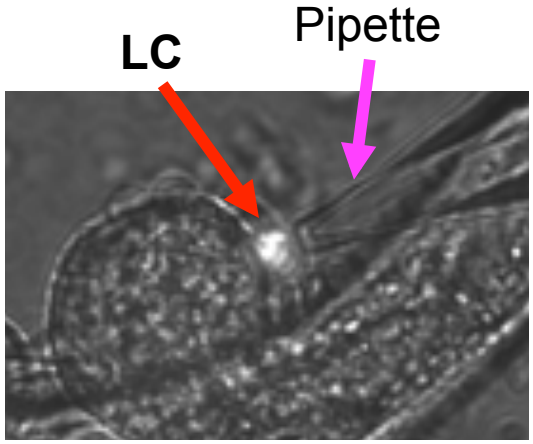
Microdissecting individual LCs

1. Glue the worm onto an agar pad and immerse it in buffer.





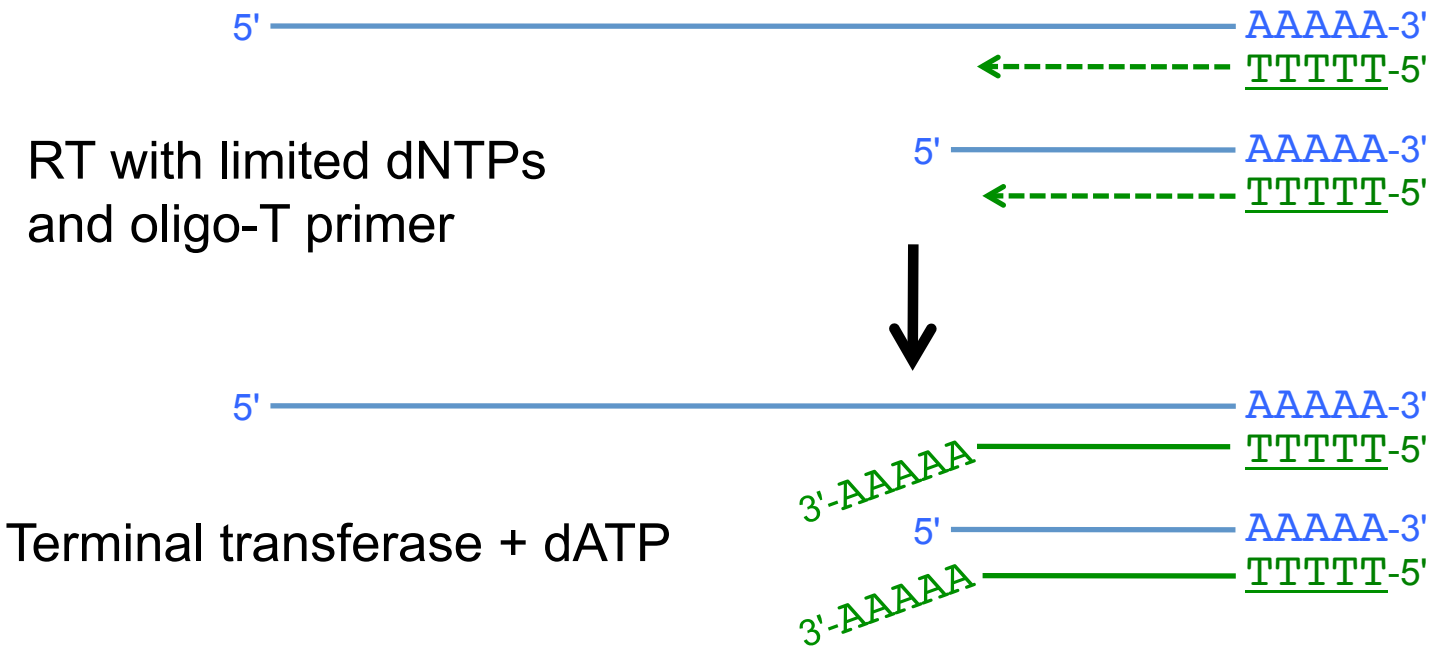
2. Cut into the worm to release the LC.



3. Suck the LC into a pipette and snap-freeze it in liquid nitrogen.

Ref.: Lockery and Goodman (1998), Meth. Enzym. *293*, 201-217.

RT-PCR of dissected cells



PCR with
~uniform templates

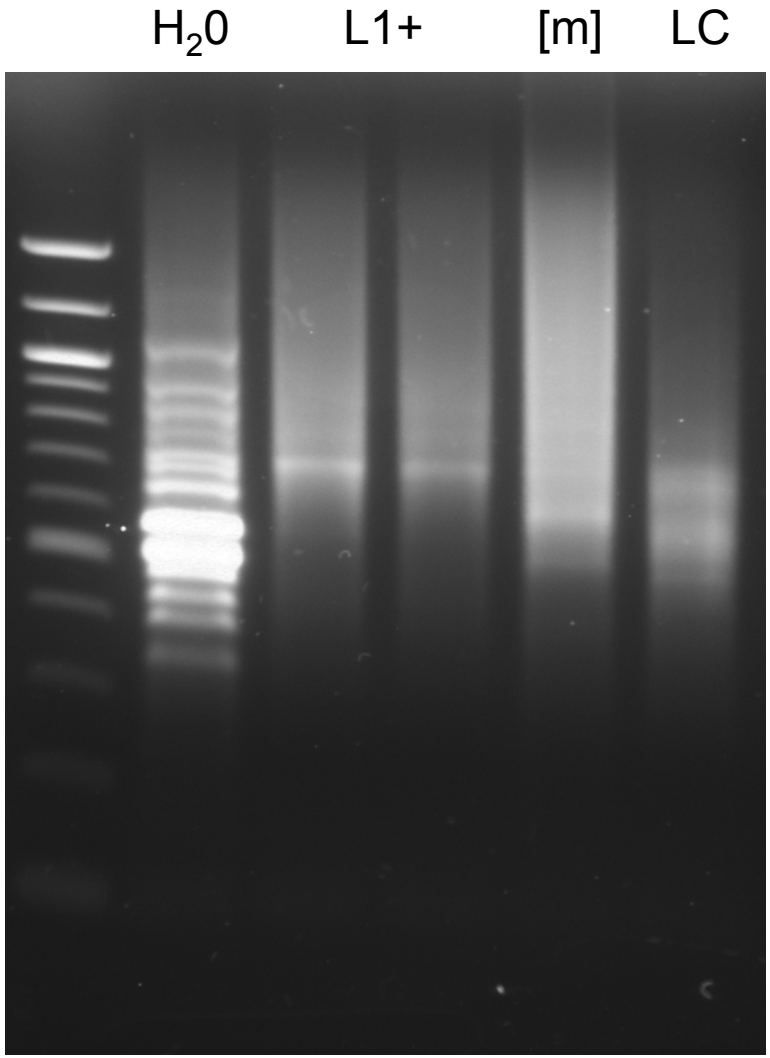
5'-[X]-TTTTT — AAAAA-[Y]-3'
3'-AAAAA — TTTTT-[X]-5'

5'-[X]-TTTTT — AAAAA-[Y]-3'
3'-AAAAA — TTTTT-[X]-5'

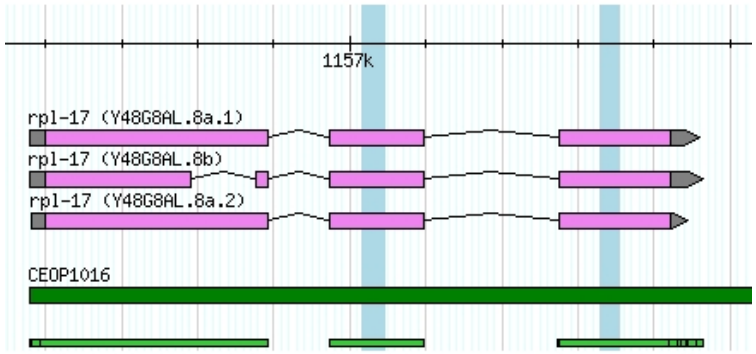
Ref.: Dulac and Axel (1995), Cell *20*, 195-206.

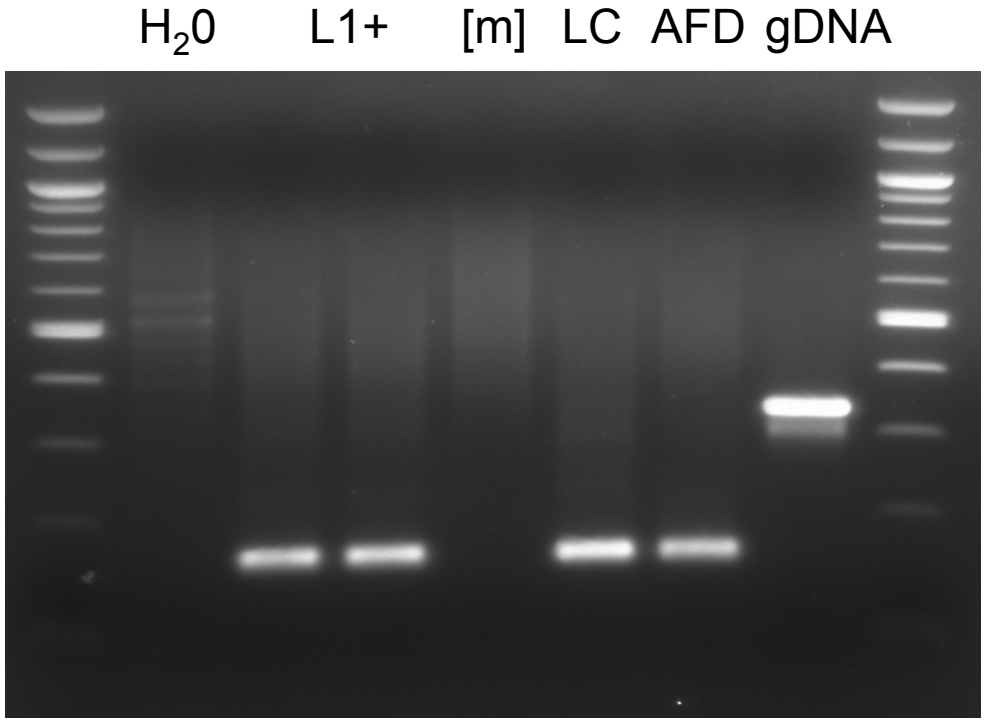
RT-PCR products from a single linker cell

Primary RT-PCRs:

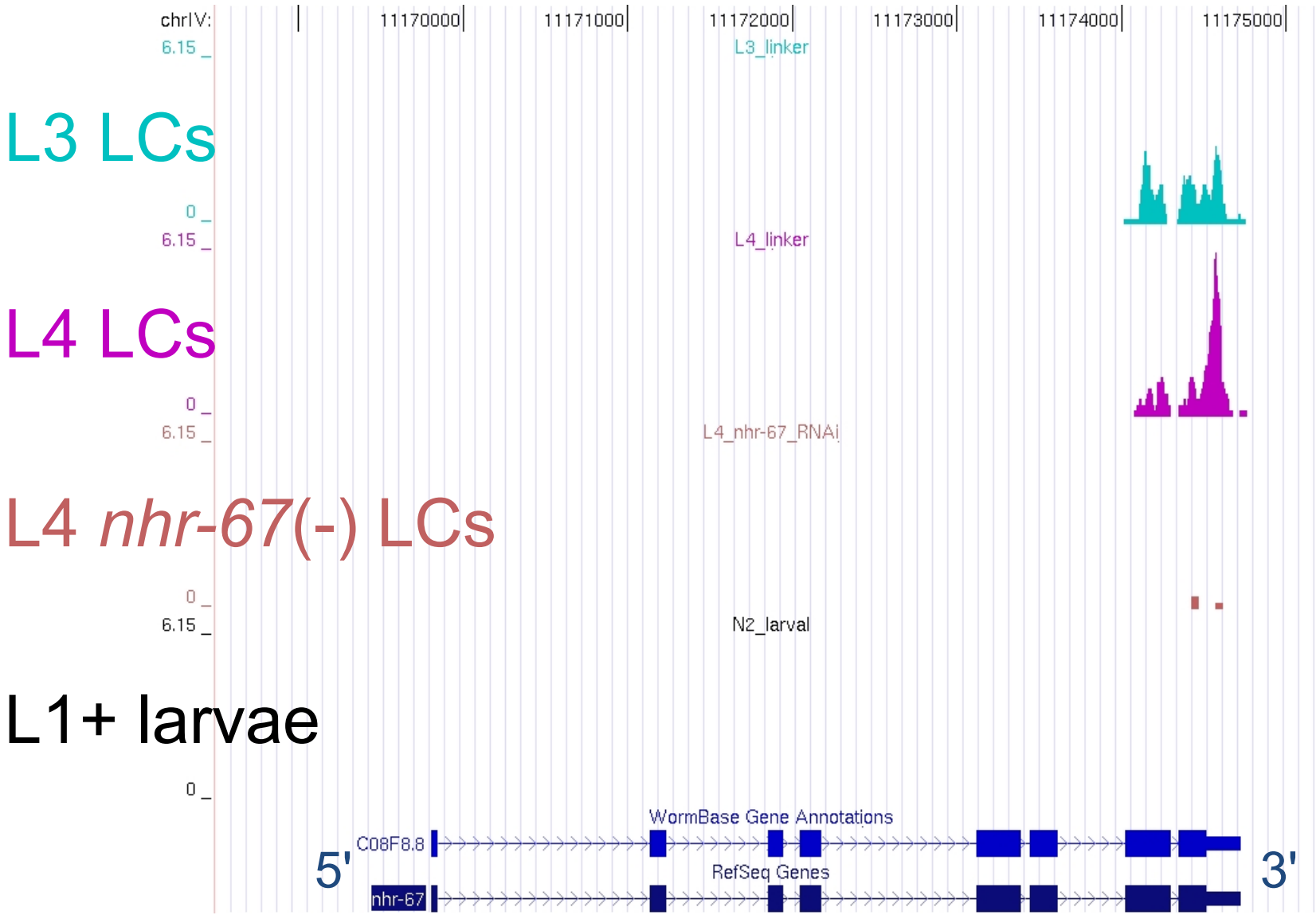


rpl-17 test PCRs:





RNA-seq of pools (five LCs or samples apiece)

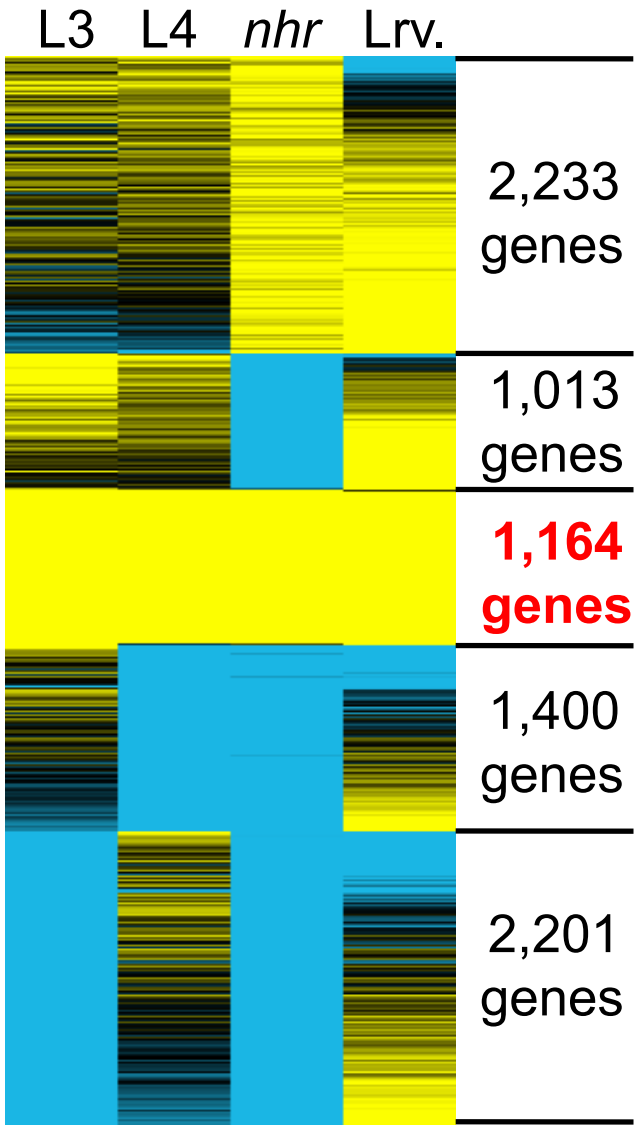


nhr-67

Expressed genes

Maximum detectable	20,252
L3-stage LC	5,740
L4-stage LC	6,603
nhr-67(RNAi) L4 LC	5,290
L3 or L4 w.t. LC	8,011
Wild-type L1+ larvae	13,152

A simple k=5 split defines ubiquitous genes



NHR-67-inhibited, broad expression

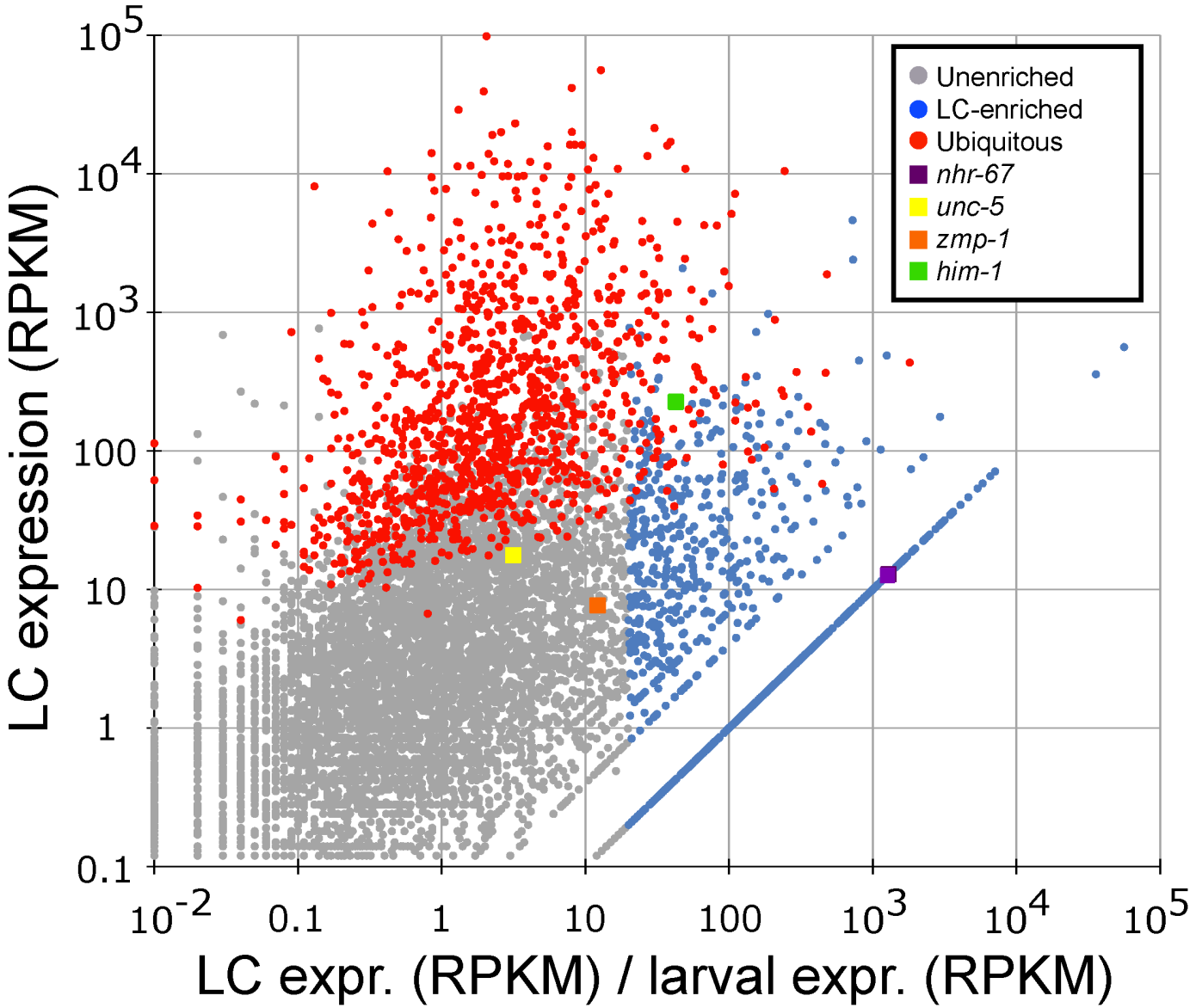
NHR-67-activated, broad expression

Ubiquitous

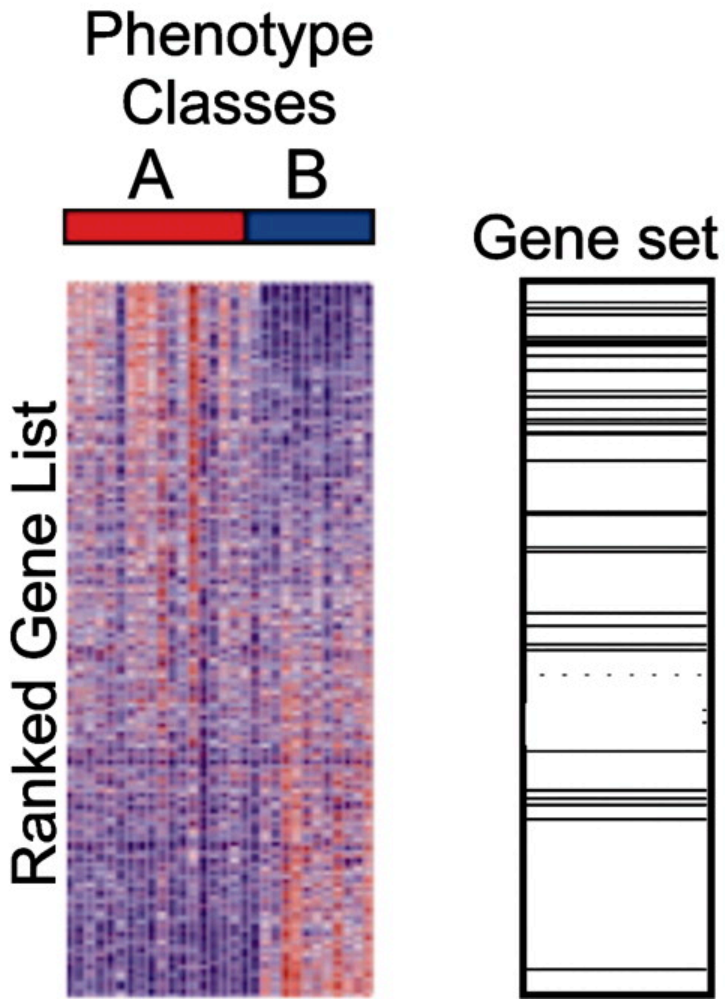
L3-specific

NHR-67-activated, L4-specific

Defining 1,097 LC-enriched genes (14% of 8,011)



Expression ratios vs. gene functions (e.g., GO)



Input	Gene_ID / ontology annotation / associated variable(s)		Annotation (GO format)	
Test	**Hypergeometric** Associated var.: binary	**Wilcoxon rank** Associated var: Floating point	**Binomial** Associated vars: Two integers	**2 x 2 contingency** Associated vars: four integers
Options	Minimum number of genes per category	Top categories (default: biological process, molecular function, cellular component)		
Calculations	1) Sum up gene associated variables per category 2) Perform the chosen test on all selected categories 3) Generate 10,000 sets with permutated gene associated variables 4) Perform the chosen test on all random sets 5) Calculate overall significance by comparing p-value distribution of the data set against the random sets 6) Calculate FDR and FWER per category based on the random sets			
Output	Global p-value, summary statistics Categories with raw p-values			
Refinement	1) Choose raw p-value cut-off 2) Refinement algorithm recursively removes genes contained in significant child categories 3) Generates output table with p-values before and after refinement for each category			

GSEA: Subramanian et al. (2005), Proc. Natl. Acad. Sci. USA. *102*, 15545-15550.
GO terms via FUNC. Ref.: Prüfer et al. (2007), BMC Bioinformatics *8*, 41.

LC-expressed genes have specific traits

Genes expressed more strongly in LC than in whole larvae:

transcriptional regulation
cytoskeletal protein binding
cell adhesion
intracellular protein transport

GO terms via FUNC. Ref.: Prüfer et al. (2007), BMC Bioinformatics *8*, 41.

LC-expressed genes have specific traits

Genes expressed more strongly in LC than in whole larvae:

transcriptional regulation
cytoskeletal protein binding
cell adhesion
intracellular protein transport
cholinergic synaptic transmission
axon components

GO terms via FUNC. Ref.: Prüfer et al. (2007), BMC Bioinformatics *8*, 41.

LC-expressed genes have specific traits

Genes expressed more strongly in LC than in whole larvae:

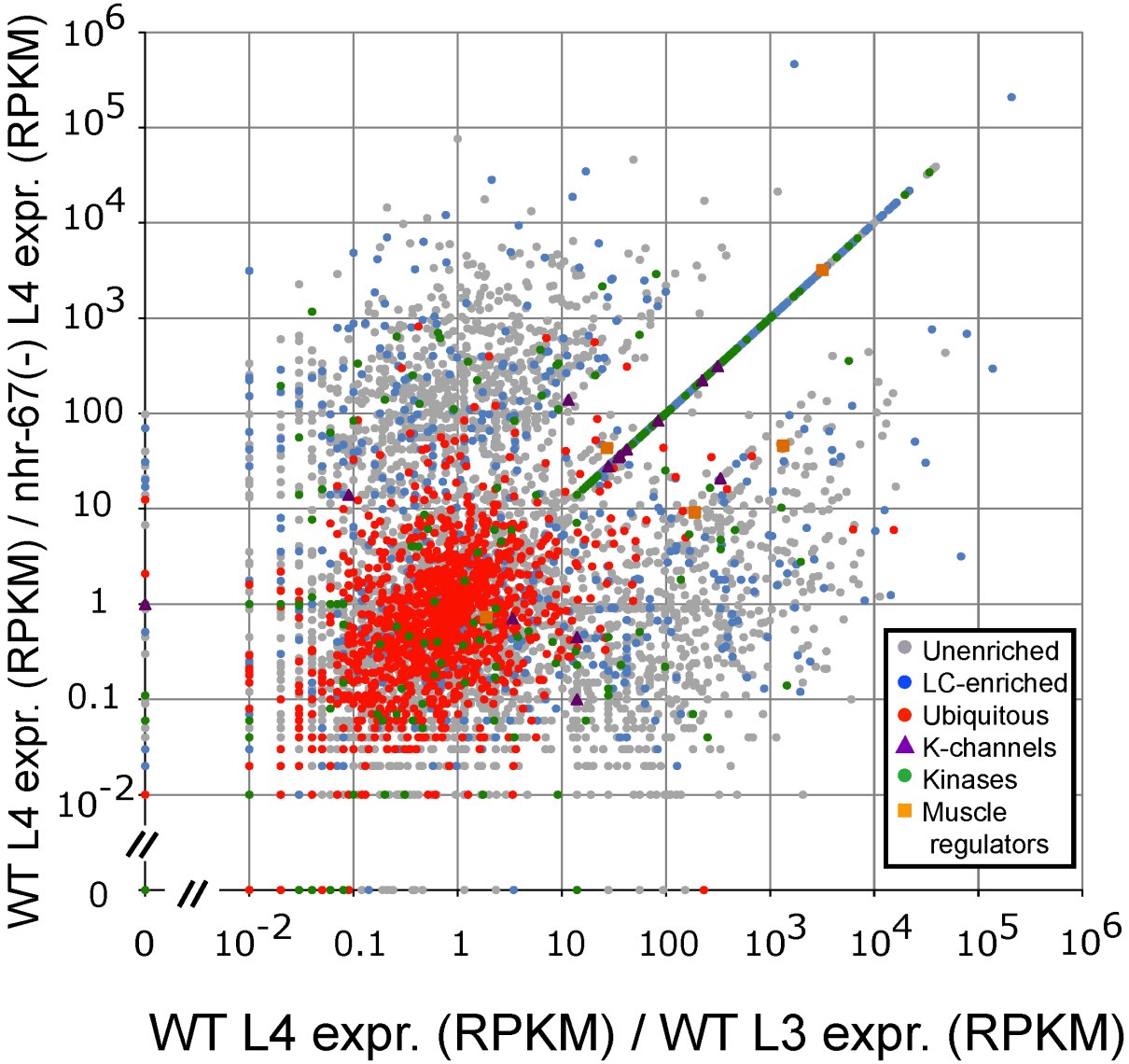
transcriptional regulation
cytoskeletal protein binding
cell adhesion
intracellular protein transport
cholinergic synaptic transmission
axon components

Genes upregulated from L3 to L4 by NHR-67:

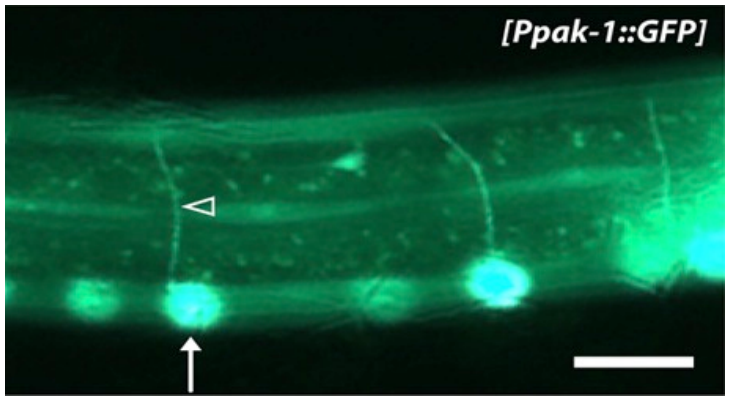
potassium channels
regulators of muscle contraction

GO terms via FUNC. Ref.: Prüfer et al. (2007), BMC Bioinformatics *8*, 41.

NHR-67-dependent gene upregulation in L4



"Axon components" include migratory kinases



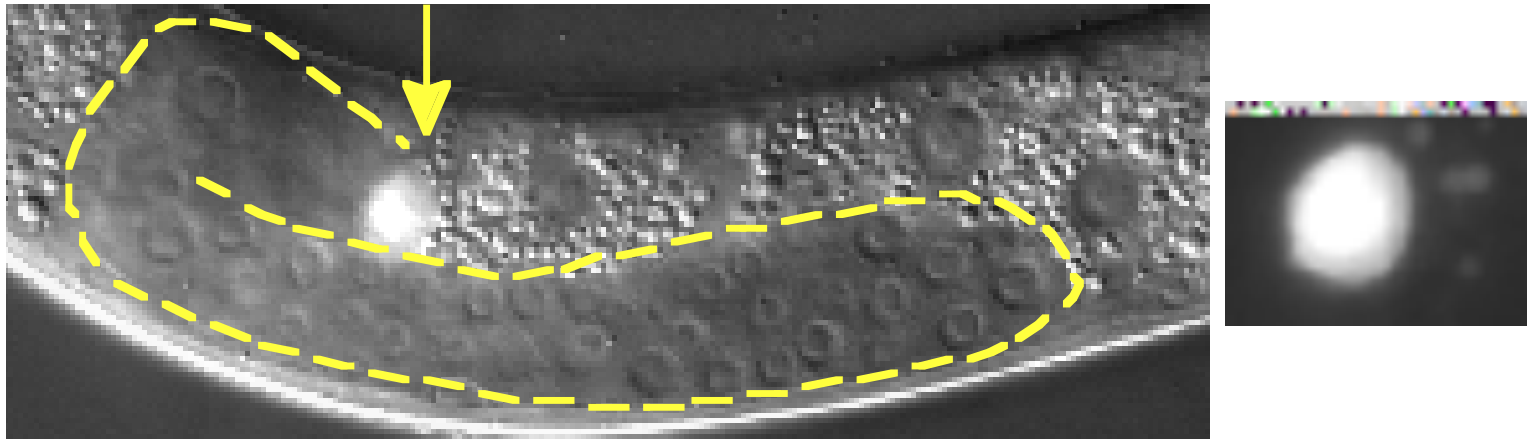
Among the 45 LC genes associated with axon components were *cam-1*, *pak-1*, *sax-1*, *unc-51*, and *vab-1*, which encode protein kinases required for **normal migration or morphology of axons and neurons.**

Another similarity to neurons is that acetycholine and glutamate (AMPA) receptors are expressed in migrating LCs.

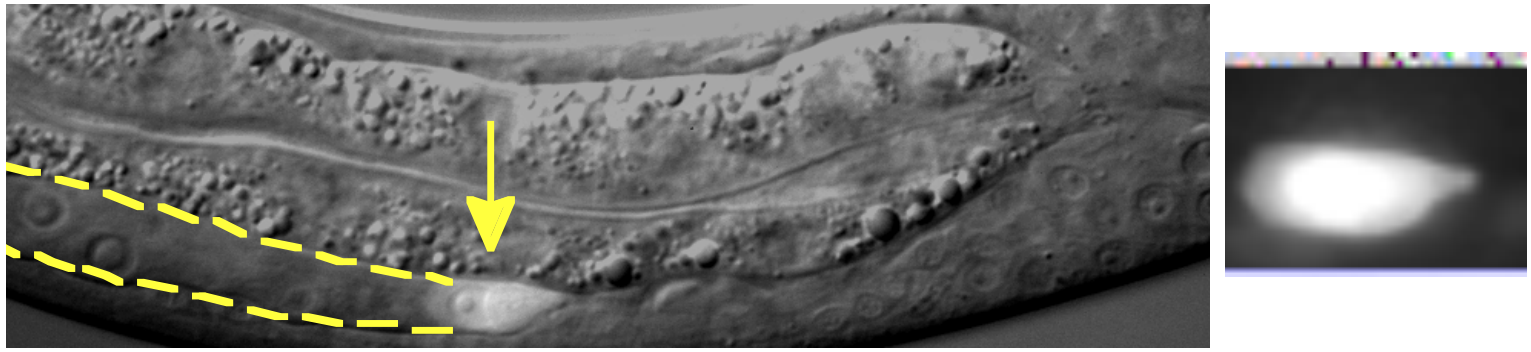
Refs.: Gallegos and Bargmann (2004), Neuron *44*, 239-249; Hayashi et al. (2009), Nat. Neurosci. *12*, 981-987; Lucanic et al. (2006), Development *133*, 4549-4559; Mohamed and Chin-Sang (2006), Dev. Biol. *290*, 164-176; Ogura et al. (2010), Development *137*, 1657-1667.

LCs change shape in the L4 stage

L3 stage



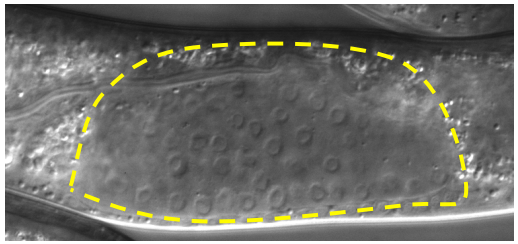
L4 stage



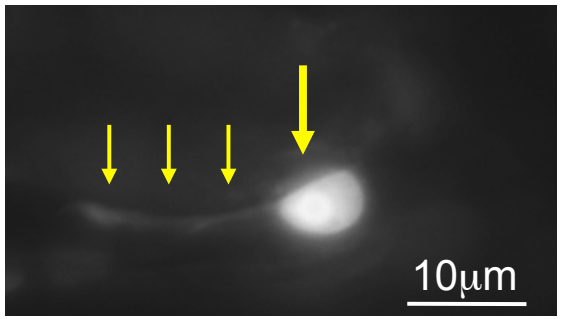
The myosin regulatory light chain genes *mlc-1* and *mlc-2* and the troponin *mup-2* are upregulated by NHR-67 in L4 LCs.

RNAi phenotypes in 45 of 204 genes (22%)

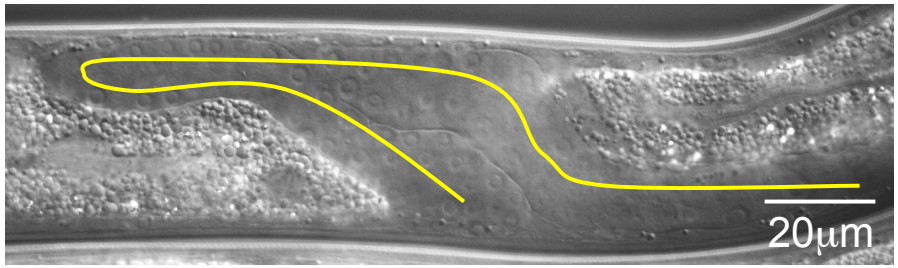
1. **No LC specified**:
   hlh-2 (E/Da HLH)



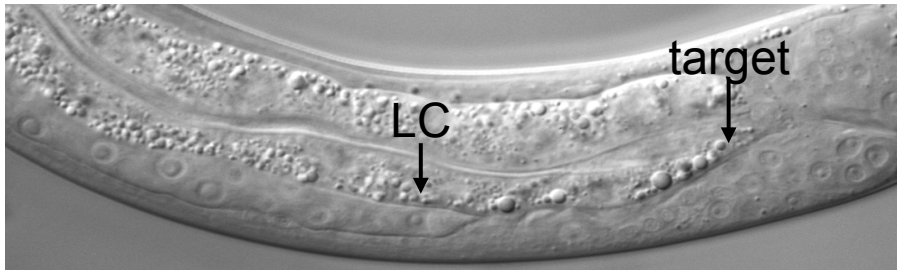
2. **Abnormal LC shape**:
   inft-1 (INF2 formin)



3. **Abnormal path**:
   lim-9 (*limpet*/FHL2)



4. **Slow migration**:
   sphk-1 (sphingosine kinase)



14/48 TFs tested are required by the LC

	RNAi phenotypes:				
Genes:	**Identity**	**Shape**	**Path**	**Slow**	**Gonad**
hlh-2 [E/Da]	**---**				
daf-12 [LXR]		**---**		**---**	**---**
lin-29 [ZNF367]	**---**	**---**	**--**	**---**	

14/48 TFs tested are required by the LC

	RNAi phenotypes:				
Genes:	**Identity**	**Shape**	**Path**	**Slow**	**Gonad**
hlh-2 [E/Da]	---				
daf-12 [LXR]		---		---	---
lin-29 [ZNF367]	---	---	--	---	
hlh-8 [Twist]	-	--		-	
hlh-19	--	-	--	---	-
nhr-97		--	-	---	
egl-5 [AbdB]		--		---	---
nob-1 [AbdB]		-	---	---	
crh-2 [CREB]				---	---
zfp-1 [AF10]		--		---	---
ZC123.3 [ZFHX]		-		---	--
pha-1		---	---	--	
gei-13		---	---	---	
T20F7.1 [Zf]		---	---	---	

14/48 TFs tested are required by the LC

	RNAi phenotypes:				
Genes:	**Identity**	**Shape**	**Path**	**Slow**	**Gonad**
hlh-2 [E/Da]	---				
daf-12 [LXR]		---		---	---
lin-29 [ZNF367]	---	---	--	---	
***hlh-8* [Twist]**	**-**	**--**		**-**	
hlh-10					
crh-2 [CREB]				---	---
zfp-1 [AF10]		--		---	---
ZC123.3 [ZFHX]		-		---	--
pha-1		---	---	--	
gei-13		---	---	---	

Homologs of TWIST are actively required for cell migration during (EMT-like) cancer cell metastasis in mammals, and during *Drosophila* gastrulation.

Refs.: Yang et al. (2004), Cell *117*, 927-939; Kölsch et al. (2007), Science *315*, 384-386.

Some other known migration genes rediscovered

crh-2 =~ *Drosophila* SLBO, C/EBPγ TF
required for border cell migration

sphk-1 == sphingosine kinase;
sphingosine-1-phosphate is generally implicated in migration

lim-9 == mammalian FHL2 inhibits sphingosine kinase, and
is involved in colon cancer invasion and dendritic cell migration

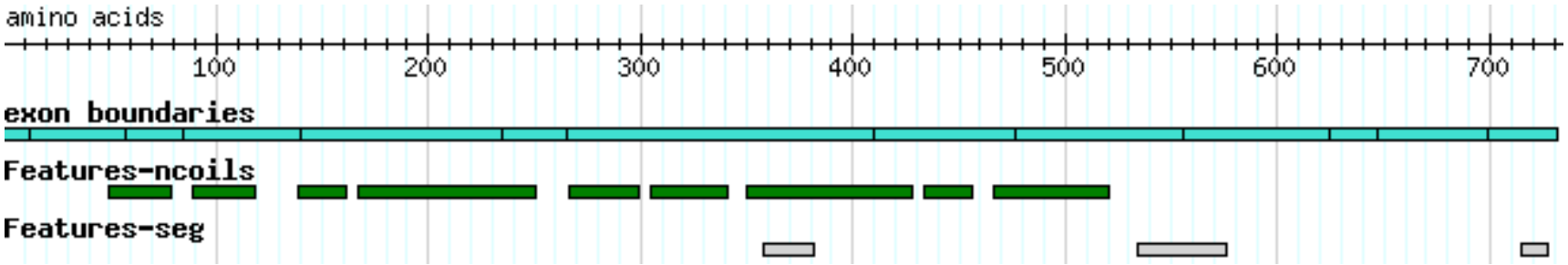
Refs.: Borghese et al. (2006), Dev. Cell *10*, 497-508; Wang et al. (2006), Dev. Cell *10*, 483-495; Hayashi et al. (2009), Arterioscler. Thromb. Vasc. Biol. *29*, 909-914; Fyrst et al. (2010), Nat. Chem. Biol. *6*, 489-497; König et al. (2010), J. Immunol. *185*, 1466-1475; Zhang et al. (2010), Carcinogenesis *31*, 1220-1229.

However, 'conserved unknowns' also found

tpra-1 == TPRA40 (7TM) inhibits embryonic cell division

maea-1 == MAEA/EMP receptor is required for developing erythroblasts to bind macrophages and mature into erythrocytes; *may* also have migratory phenotype in cell culture

cccp-1 ('conserved coiled-coil protein') ==
C10orf118 has no known function whatsoever



Refs.: Aki et al. (2008), J. Cell. Physiol. *217*, 194-206;
Chasis et al. (2008), Blood *112*, 470-478; Manjit Hanspal, pers. comm., 25 Feb 2013.

Potassium channels are, indeed, required for normal LC migration

	RNAi phenotypes:			
Genes:	**Identity**	**Shape**	**Path**	**Slow**
twk-13	**-**	**--**	**---**	**---**
twk-18			**---**	**-**

There are 46 *twk* genes in *C. elegans*, of which 13 are expressed in LCs.

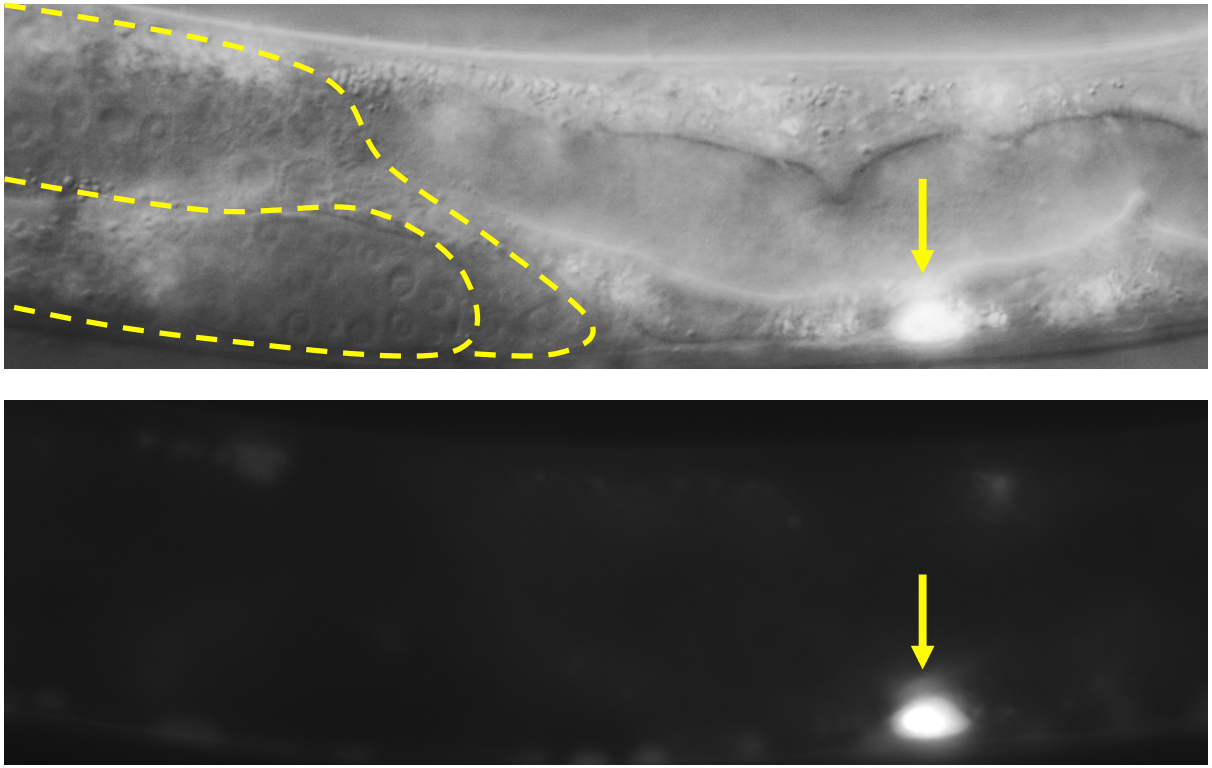
Both *twk-13*(RNAi) and *twk-18*(RNAi) LCs have delayed return from the dorsal to the ventral body wall, a subset of the phenotypes in *nhr-67*(RNAi) LCs.

TWK potassium channels have had no known role in migration, but there is evidence, for a few TWKs, that they can be mechanosensory.

Ref.: Mathie et al. (2010), J. Physiol. *588*, 3149-3156.

Cohesin and condensin subunits enable LC-gonad attachment

him-1/SMC1
mix-1/SMC2
smc-3/SMC3
dpy-27/SMC4
smc-4/SMC4



Cohesins can regulate genes as well as join chromosomes; condensins may as well, though this is less studied.

Refs.: Seitan et al. (2006), PLoS Biol. *4*, e242; Wood et al. (2010), Nat. Rev. Genet. *11*, 391-404; Pauli et al. (2010), Curr Biol., 10.1016/j.cub.2010.09.057.

Major sperm proteins expressed in L4-stage LCs

MSP-domain protein family 1

	L3	L4	*nhr-67* (-)	Larvae
D2007.2	5.25	0	0	0.51
F42A9.7	0	1.3	0	11.14
C35E7.9	0	1.59	0	13.77
C35D10.1	0	1.92	0	2.62
F21H7.5	0	2.27	0	17.1
T13F2.12	0	3.41	0	0
ssp-32	0	7.97	0	0.24
F55C5.1	0	8.2	13.06	0.84
W03B1.5	0	10.94	0	1.27
ssp-10	0	14.46	4.15	5.54
ssp-31	0	34.1	0	0.22
C34D4.3	0	38.94	0.54	5.15
ssp-16	0	39.52	3.23	2.64
ssp-19	0	88.92	0	3.31
ssp-11	0	481.57	1.11	26.44
ssp-9	0	774.75	1.13	38.25

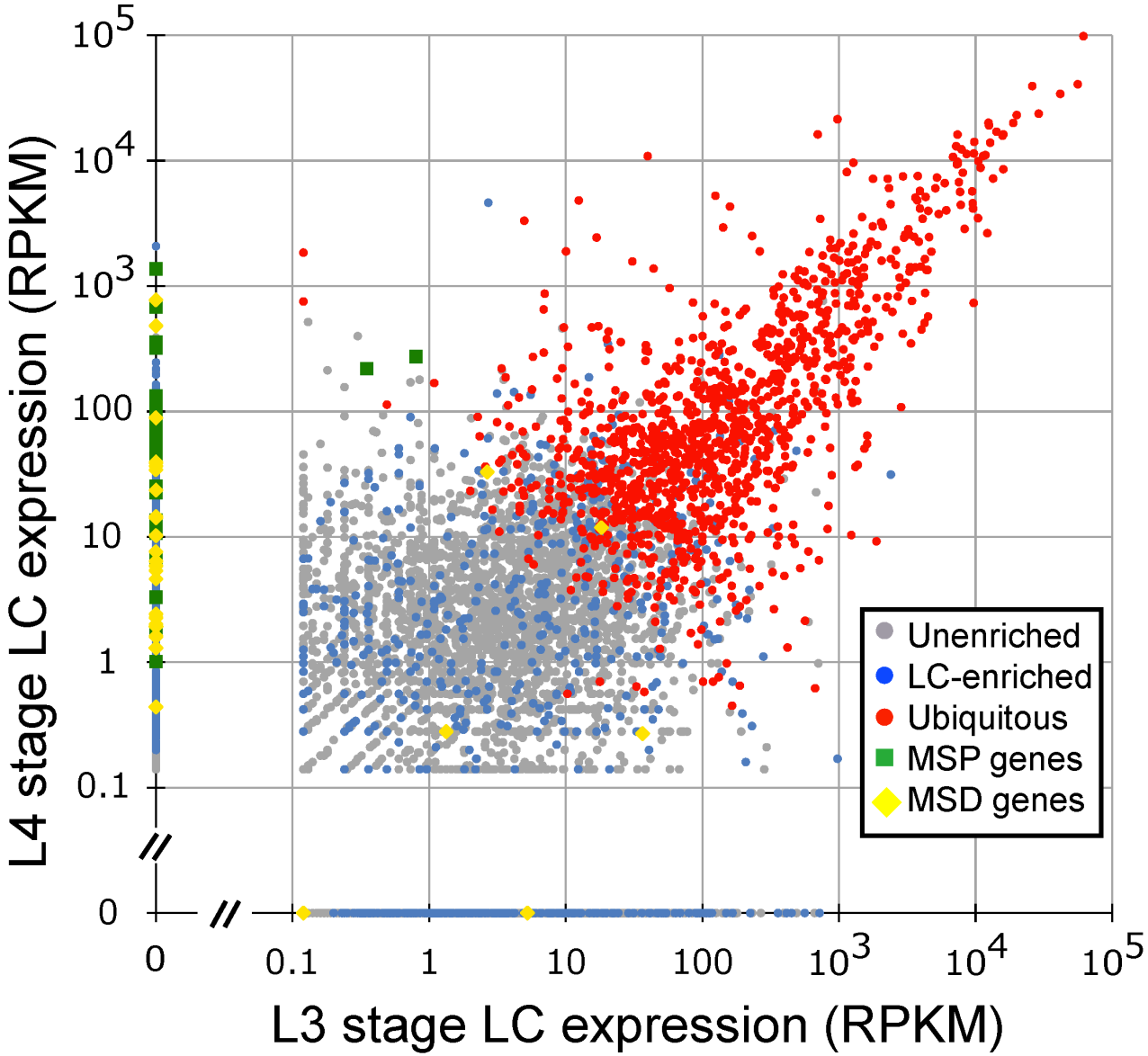
MSP-domain protein family 2

msp-63	0	1.01	1.13	0.84
msp-79	0	6.59	0	4.14
msp-56	0	10.73	0	4.03
msp-77	0	12.01	0	7.19
msp-152	0	12.86	0.51	8.84
msp-10	0	13.57	0	6.79
Y59H11AI	0	14.25	1.83	5.21

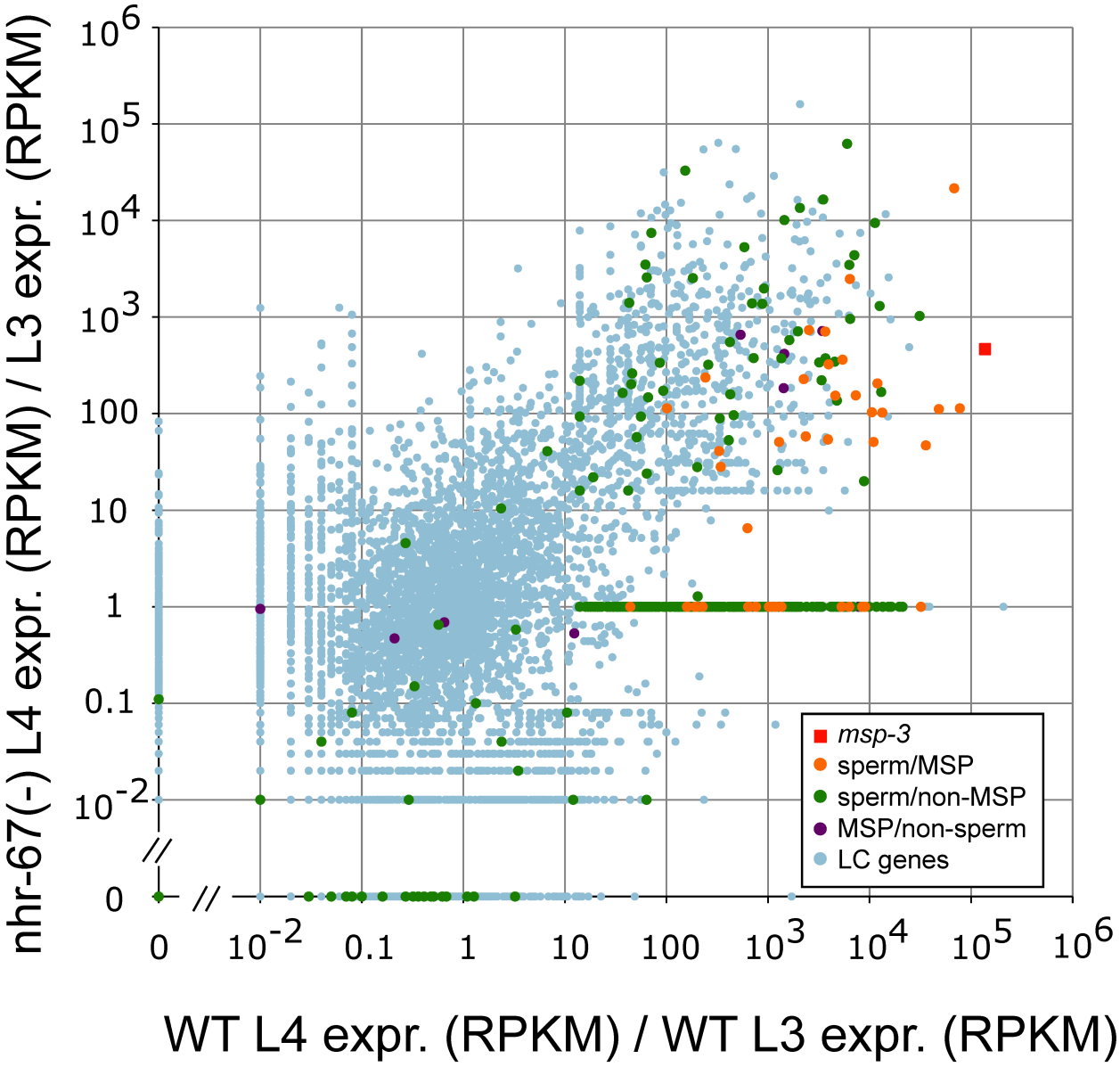
MSP-domain protein family 2

	L3	L4	*nhr-67* (-)	Larvae
msp-113	0	22.47	2.27	7.38
ZK546.3	0	23.53	0.58	4.23
msp-142	0	25.31	7.31	6.59
Y59E9AR.	0	34.27	7.15	16.43
Y59E9AR.	0	36.62	7.06	16.02
msp-31	0	46.04	1.53	12.69
msp-81	0	47.1	0	22.41
msp-78	0	53.04	0	9.62
msp-51	0	54.22	3.61	21.6
msp-36	0	63.18	0	13.48
msp-76	0	63.98	24.73	29.09
msp-55	0	73.16	1.54	20.13
msp-38	0	77.84	0	20.46
msp-50	0	84.66	0	38
msp-64	0	90.54	0	5.48
msp-57	0	105.35	1.03	63.28
msp-33	0	109.43	0.51	18.84
msp-59	0	119.63	2.05	64.57
msp-65	0	133.43	1.02	49.23
msp-53	0.35	219.67	2.28	39.48
msp-19	0.8	273.78	22.52	29.01
msp-49	0	320.51	0	20.03
msp-45	0	358.54	0.47	17.23
msp-40	0	681.52	215.36	27.99
msp-3	0	1372.22	4.64	17.97

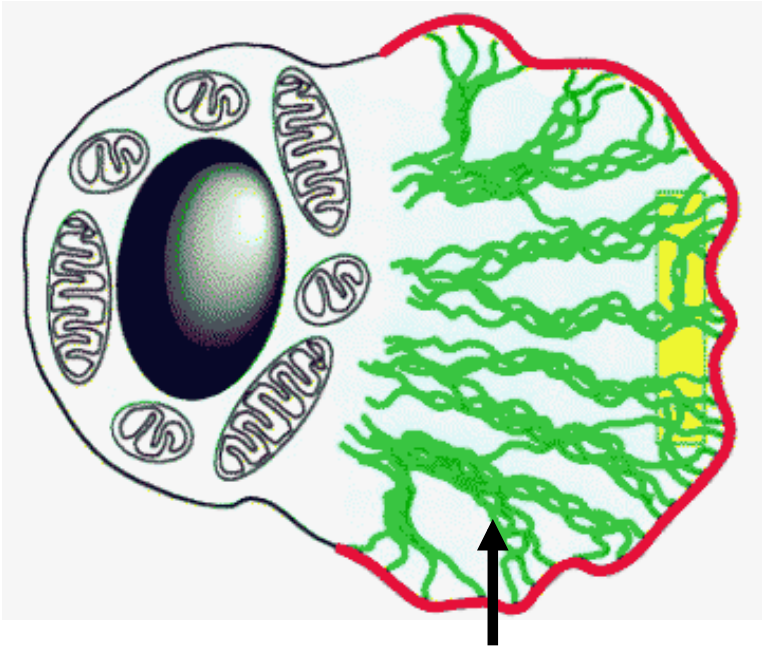
MSP and MSD genes are heavily L4-specific

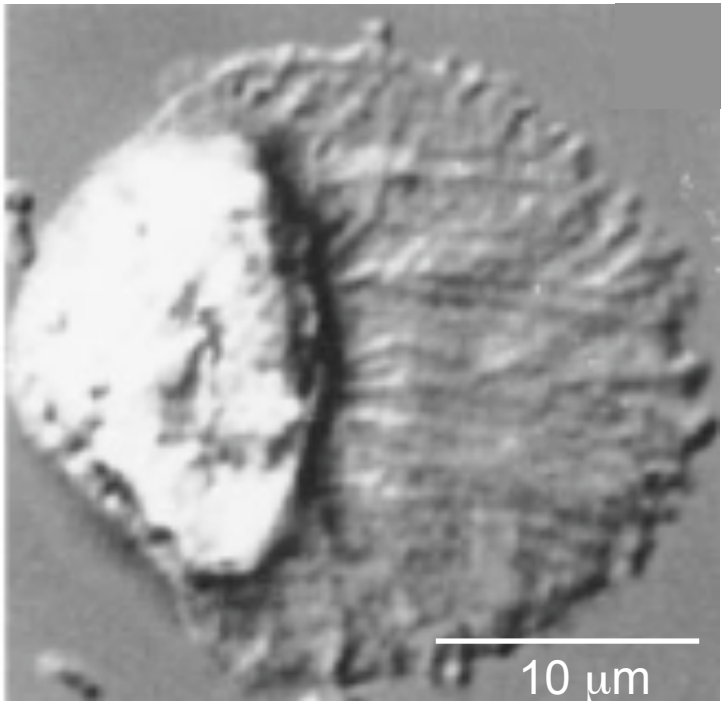


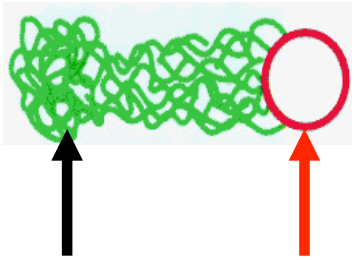
MSP and MSD genes are NHR-67-dependent

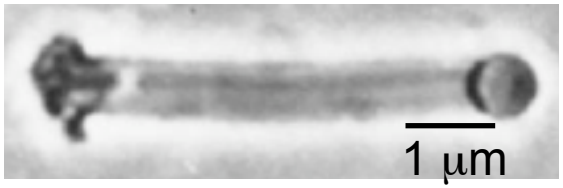


MSP, not actin, drives nematode sperm motility





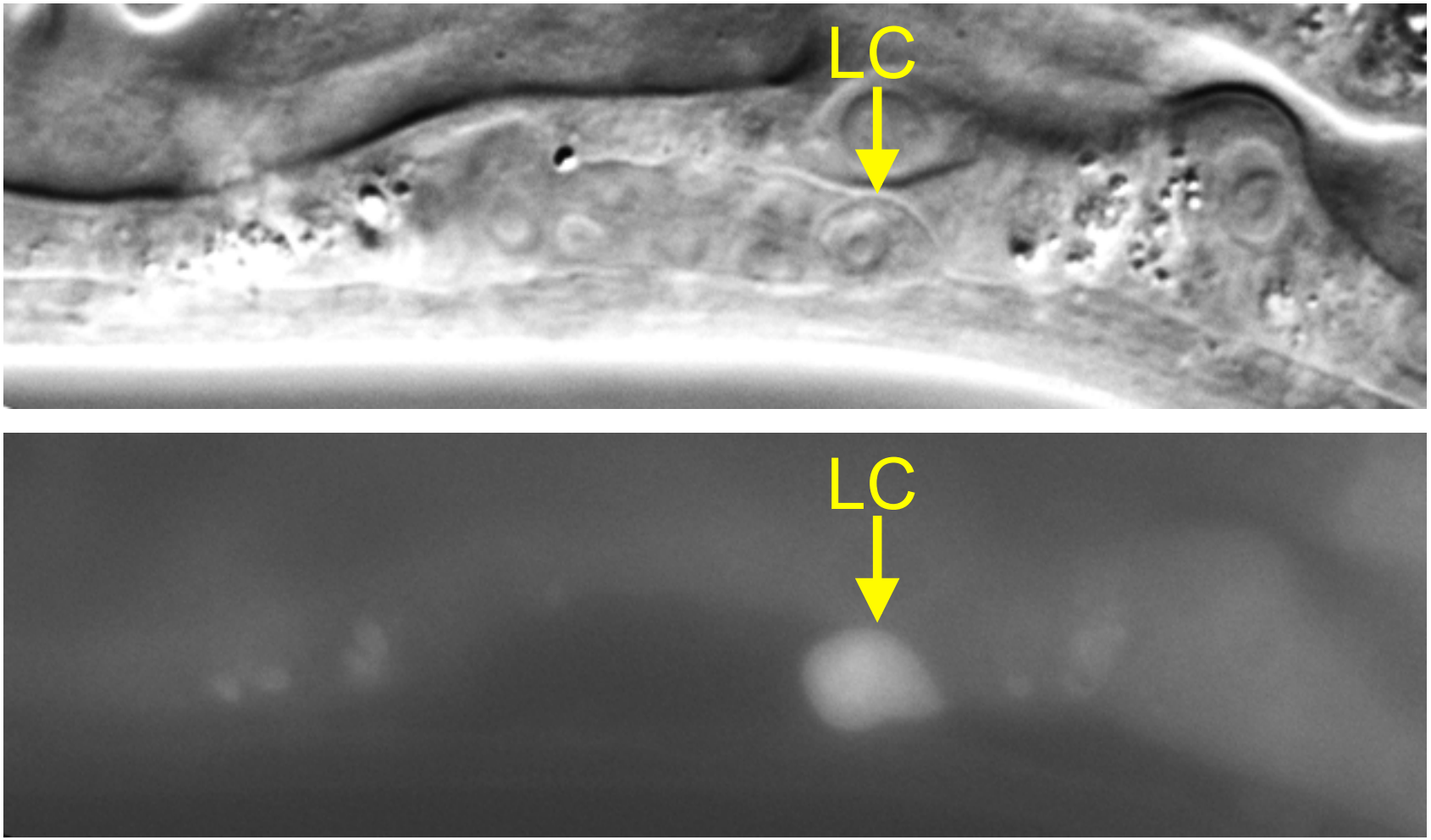




MSP fibers and membrane vesicle

Refs.: Italiano et al. (1996), Cell, *84*, 105-114; Theriot (1996), Cell, *86*, 1-4.

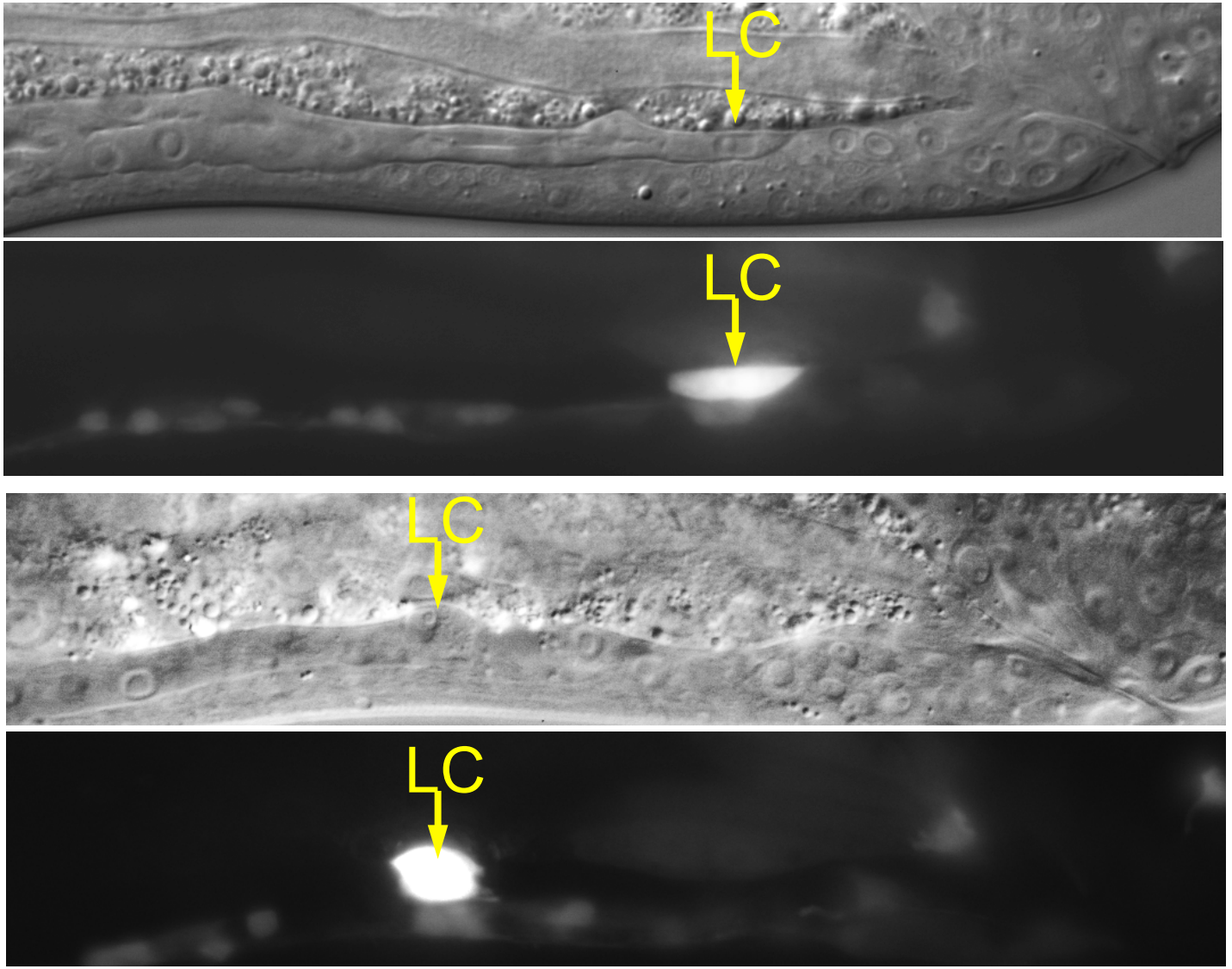
MSP::YFP is expressed in L4-stage linker cells



msp-3(RNAi) LCs are slow and abnormally round

wild type

msp-3 RNAi



Conclusions from linker cell RNA-seq

1. TWK potassium channels might have mechanosensory role in cell migration.

2. MSPs can no longer be considered only 'sperm' proteins, and thus should be considered as possible motor proteins in other cell types or organisms.

3. Other novel conserved proteins may also act during migration of animal cells, e.g., MAEA.

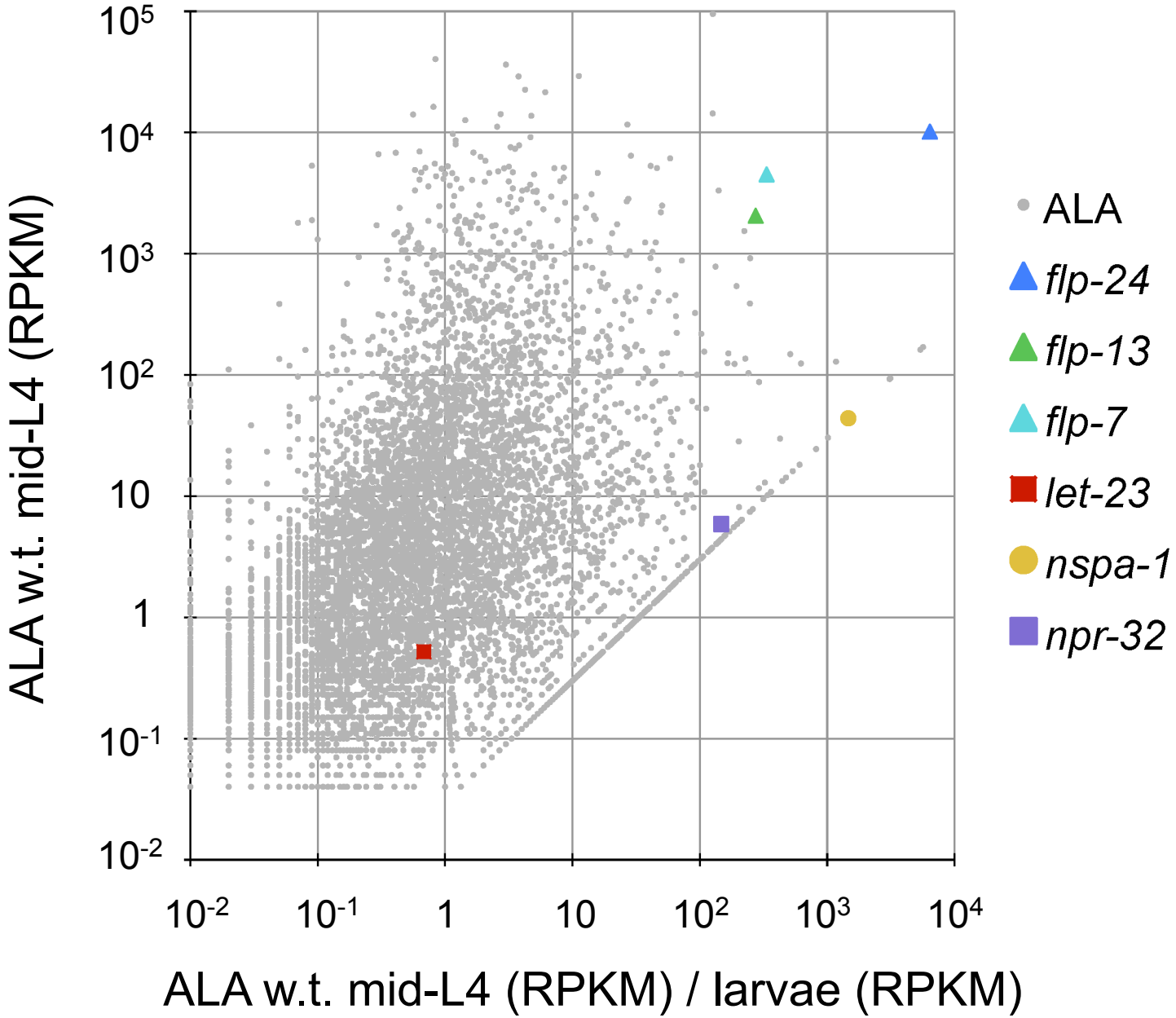
More general conclusions

1. Highly dynamic single cells can be transcriptionally profiled.

2. Gene sets are large but manageable.

3. Profiling is surprise-driven rather than hypothesis-driven.

8,133 genes in neuroendocrine ALA neurons



Thanks:

Mihoko Kato	Linker cell dissections, RNAi, and GFP
Steven Kuntz	Dissection of *ceh-13*/*lin-39* Hox elements
Miriam Goodman	AFD/ASER/PLML dissections
Yen-Ping Hsueh	AWC RNA-seq
Elly Chow	ALA RNA-seq
Ali Mortazavi	Early RNA-seq analyses
Brian Williams	cDNA library construction
Igor Antoshechkin	Optimized sequencing protocols
Jacobs Genome Center	Illumina sequencing

Profile of a chemosensory neuron (AWC)