Clustering of modENCODE/Reinke ChIP-seq peaks

DC King - Onish lab

Fall 2020

Script version and versions

This repo is from '`bash git remote -v```. UsingR version 4.0.0 (2020-04-24)'.

```
## origin   git@github.com:meekrob/onish_ChIP_R_Analysis.git (fetch)
## origin   git@github.com:meekrob/onish_ChIP_R_Analysis.git (push)
## commit b9ec30eab2f3abb9d00c9d62b7f2726cafe08853
## Author: David <dcking@colostate.edu>
## Date:   Fri Nov 13 15:20:34 2020 -0700
##
##     Adding GO plots and data files.
##  M cluster_heatmaps.Rmd
## diff --git a/cluster_heatmaps.Rmd b/cluster_heatmaps.Rmd
## index 3ff028b..08a672c 100644
## --- a/cluster_heatmaps.Rmd
## +++ b/cluster_heatmaps.Rmd
## @@ -4,7 +4,8 @@ author: "DC King - Onish lab"
##  date: "Fall 2020"
##  output:
##    pdf_document: default
## -  html_document: default
## +  word_document:
## +    kee_md: true
##  ---
##
##  ```{r setup, include=FALSE}
## @@ -28,7 +29,8 @@ library(knitr, quietly=T, warn.conflicts = F)
##  library(grid, quietly=T, warn.conflicts = F)
##  library(VennDiagram, quietly=T, warn.conflicts = F)
##  library(plyranges, quietly=T, warn.conflicts = F) # tidy the GRange datatypes
## -
## +library(parallel)
## +library(future)
##  ```
##
##  ## Script version and versions
## @@ -498,7 +500,7 @@ dfG_max$k6weights[reps_avd$threshold3d_ix] = clusters$k6weights
##
##  Custom tracks are served locally (but world-readable). The chunk currently also maps peaks to genes
##
## -```{r format prepare BED, echo=FALSE}
## +```{r format prepare BED, echo=FALSE,cache=TRUE}
##  dfG_max$LE_nonNormed = (dfG_max$LE_1 + dfG_max$LE_2)/2
```

```
##  dfG_max$L1_nonNormed = (dfG_max$L1_1 + dfG_max$L1_2)/2
##  dfG_max$L3_nonNormed = (dfG_max$L3_1 + dfG_max$L3_2)/2
## @@ -844,8 +846,6 @@ upset(wbid_clusters_fromlist,
##          )
##          )
##        )
## -
## -
##
##  # breakdown of genes with multiple peaks
##  clustersPerGene_rowSums = rowSums(clustersPerGene)
## @@ -937,7 +937,7 @@ mkGO = function(foreground_genes, background_genes) {
##
##  Run the topGO analyses using the above functions.
##
## -``` {r GO term analysis, include=FALSE}
## +``` {r GO term analysis, include=FALSE, cache=TRUE}
##  attach(annotatedPeaks)
##
##  unique.clust_0_wbid = unique((annotatedPeaks$ap %>% filter(k4cluster == 0))$feature)
```

Process data

The source data is a BED file corresponding to the basewise UNION of LE, L1, and L3 IDR peak files, run through an aggregate function (mod'd javaGenomicToolkit) that calculates summary information for each of the above ranges.

The original IDR peaks are rescanned to determine the composition of each peak.

```
## Warning in valid.GenomicRanges.seqinfo(x, suggest.trim = TRUE): GRanges object contains 3 out-of-bou
##   chrIV, chrV, and chrX. Note that ranges located on a sequence whose
##   length is unknown (NA) or on a circular sequence are not considered
##   out-of-bound (use seqlengths() and isCircular() to get the lengths and
##   circularity flags of the underlying sequences). You can use trim() to
##   trim these ranges. See ?`trim,GenomicRanges-method` for more
##   information.
```

Description of data

The total number of merged peaks is `11015`. Now, map the IDR calls into the aggregate dataframe, showing the architecture of the union.

```
## [1] 11015
```

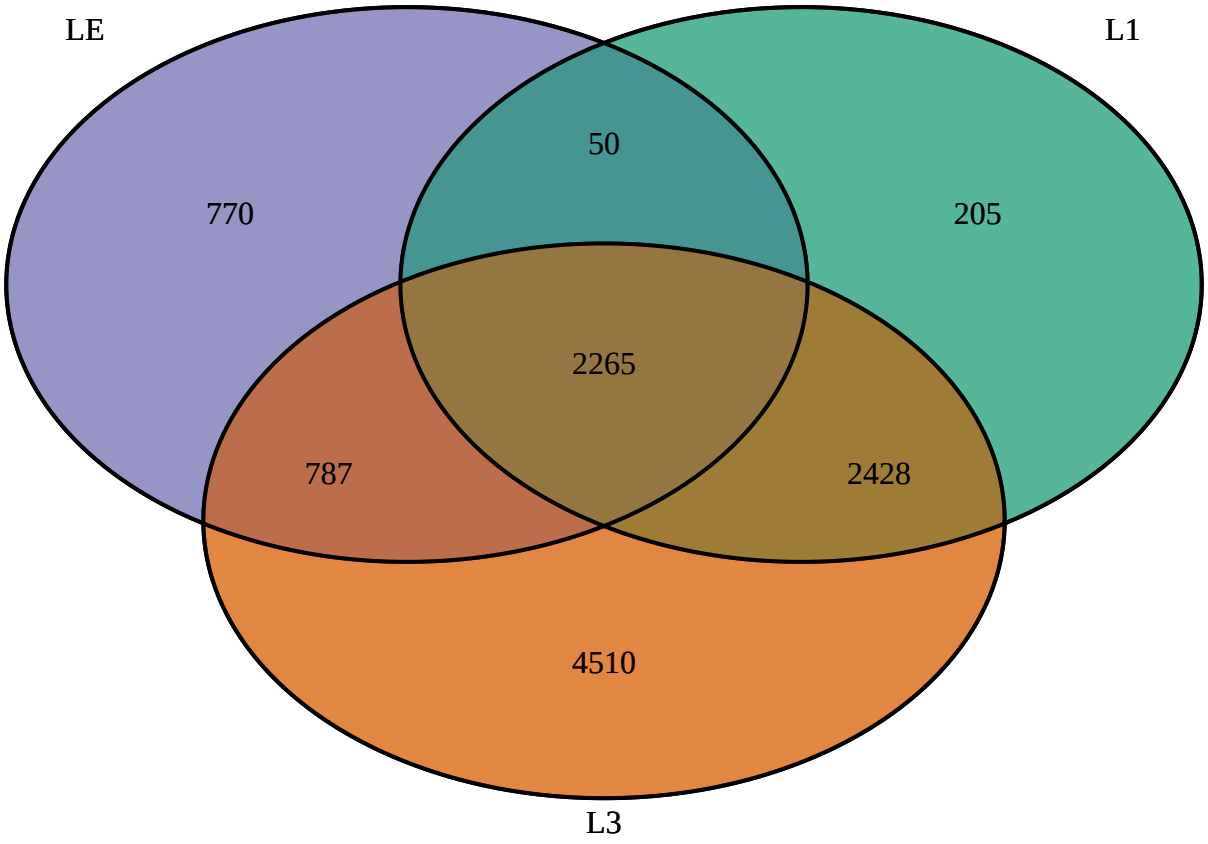
Post-scaling of data

Now the quantitative data represents 3 timepoints, 2 replicates each. Each value is the input and read-depth normalized pileups (signal) computed per basepair. The maximum value within a peak is the value used. There are `11015` peaks, (and therefore rows).

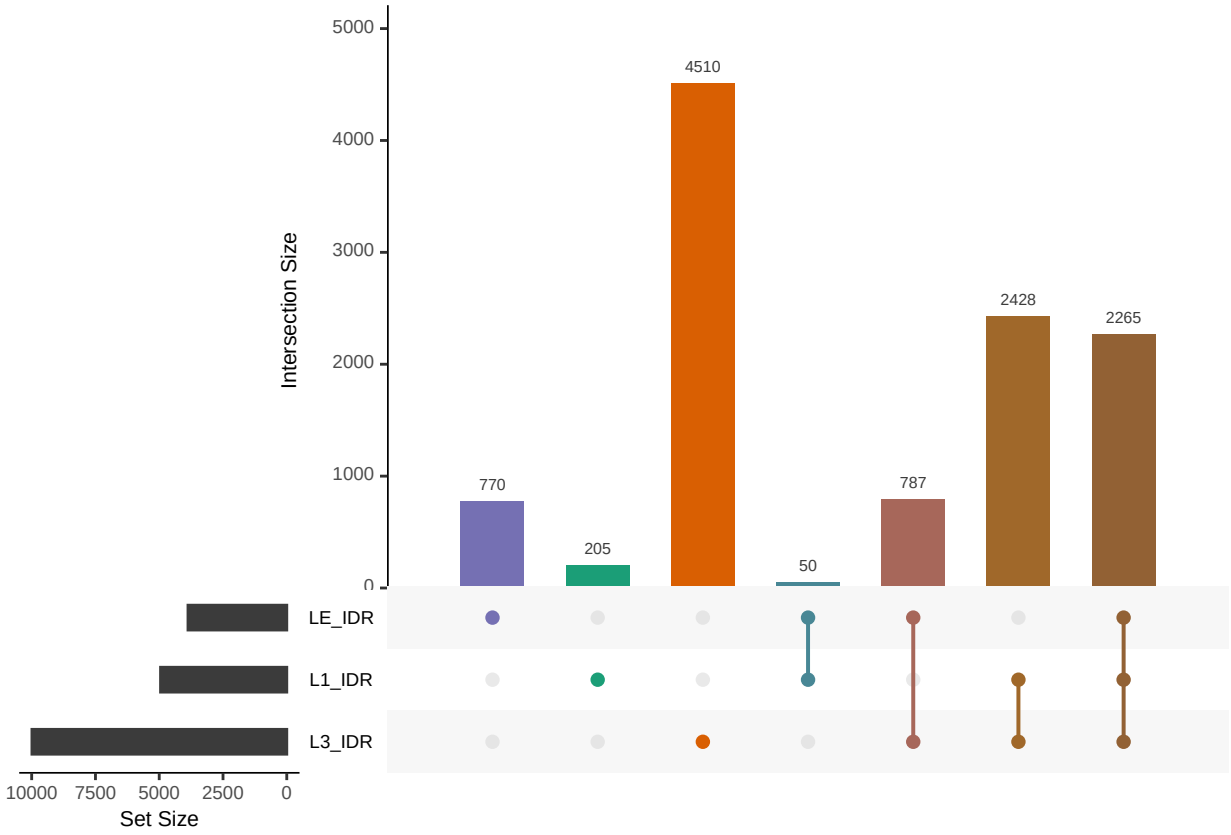
Next we want to filter the lesser 5% invariant rows, and then normalize the values by row.

Evaluate overlap between different stages in terms of the stages called in an overlapping location

Using the original IDR files, compare them to the all UNION'd file.



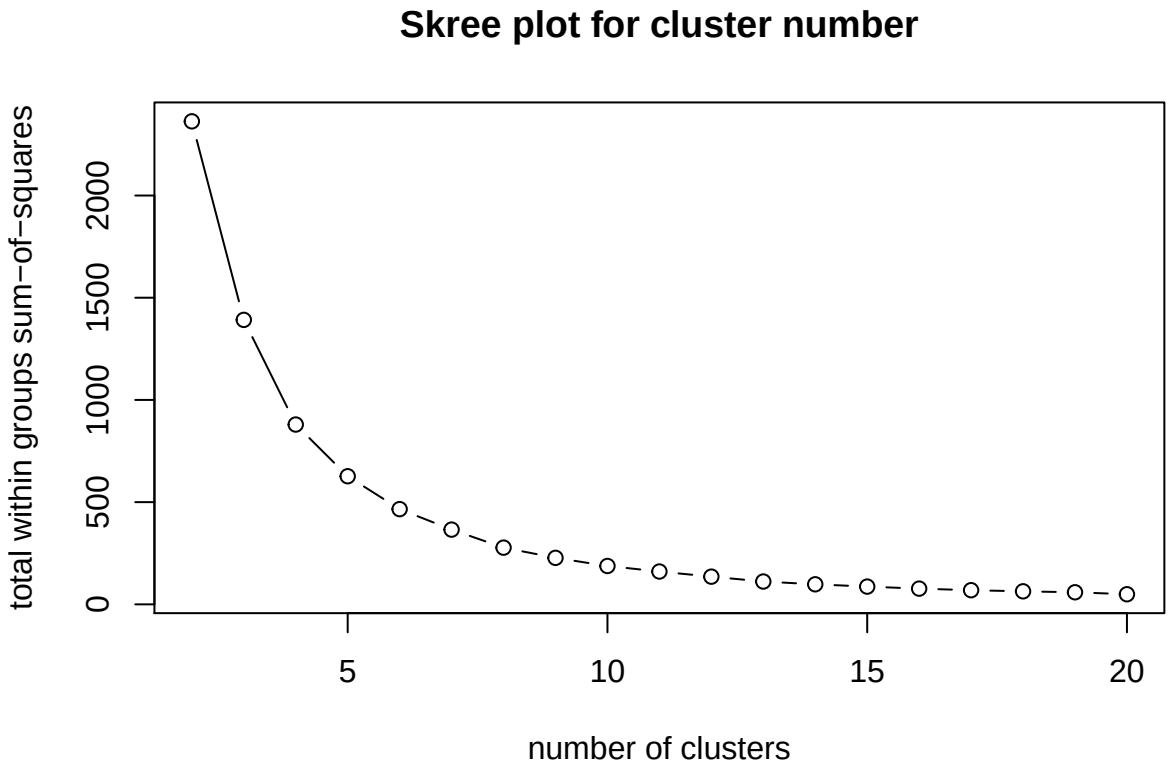

```
## pdf
##   2
```



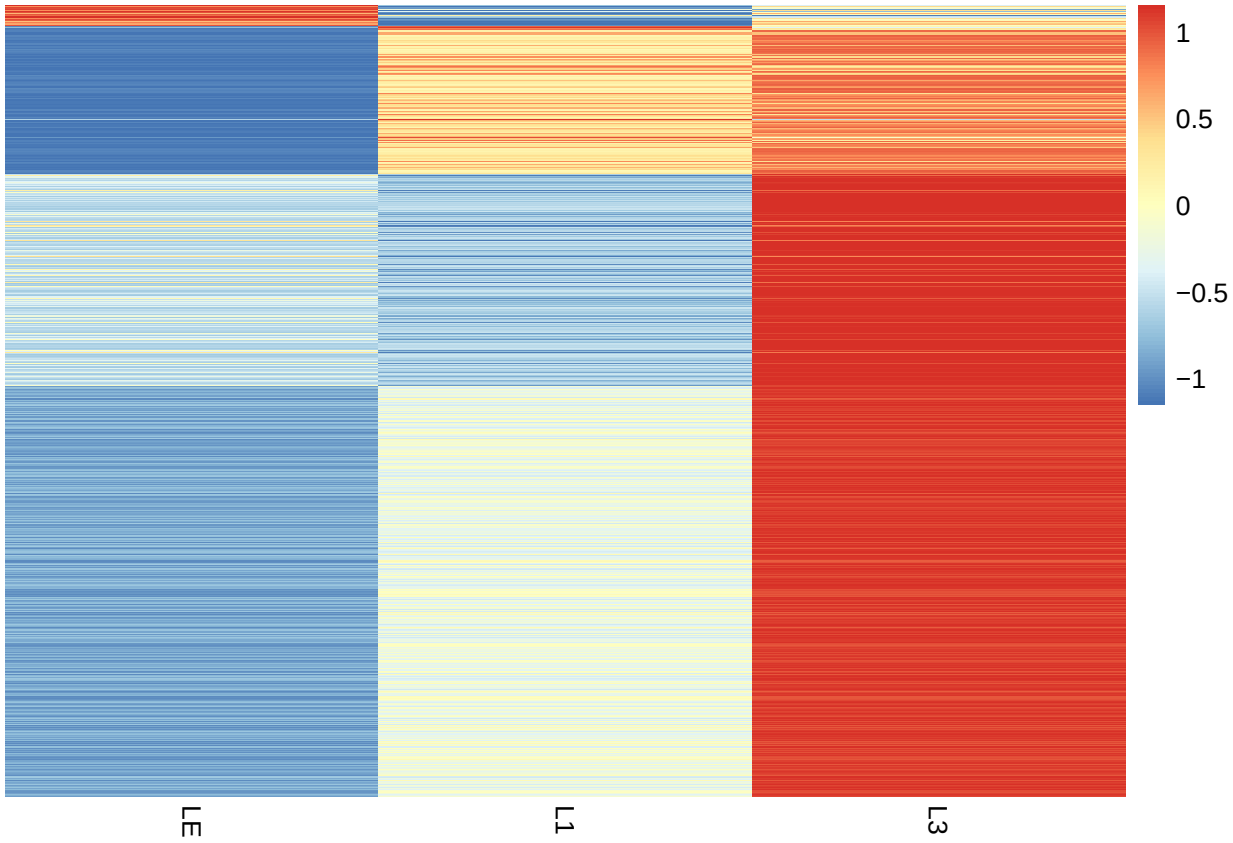
K-means clustering

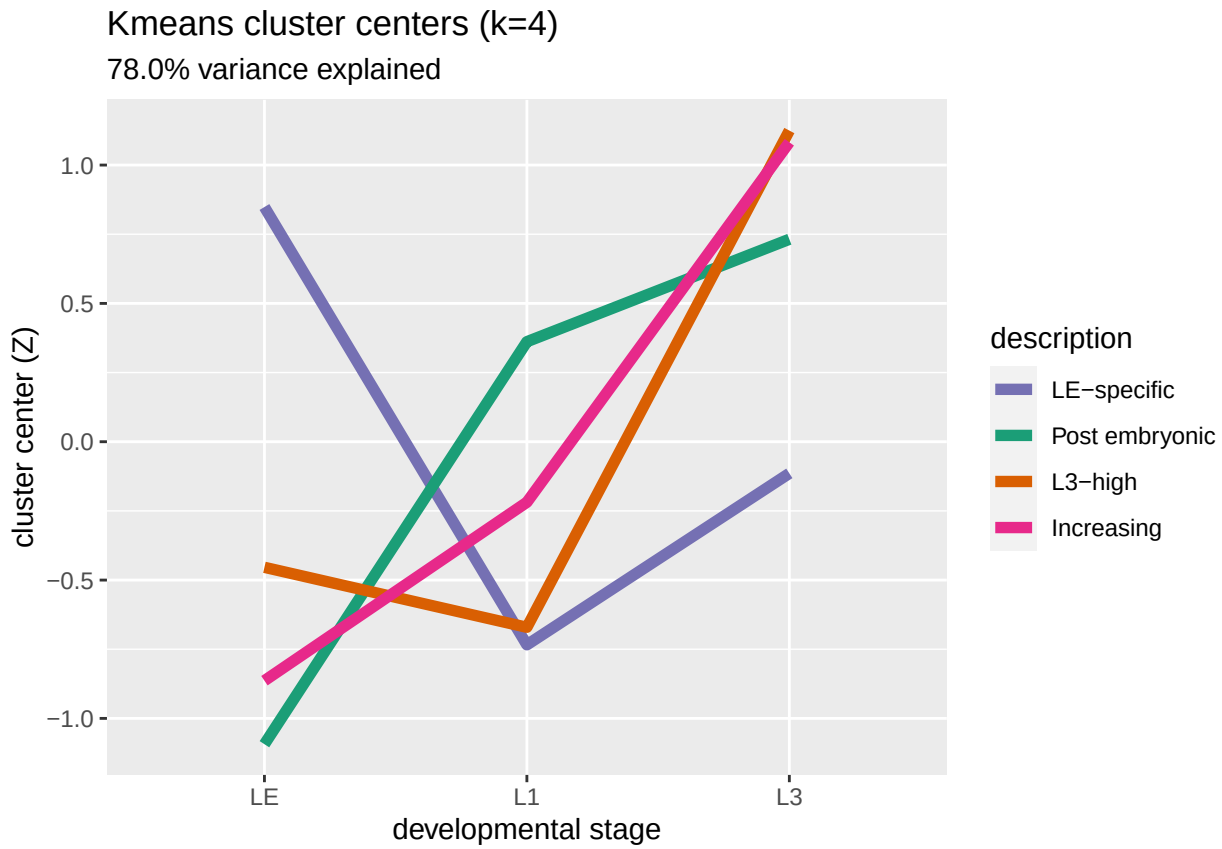
Perform K-means on the peaks that vary throughout the timecourse.

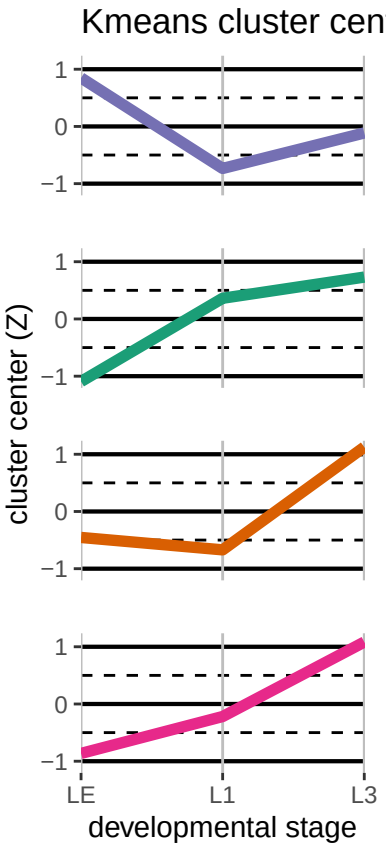
```
## Warning: Quick-TRANSfer stage steps exceeded maximum (= 523200)
```


Skree plot for cluster number
total within groups sum−of−squares
0
500
1000
1500
2000
5
10
15
20
number of clusters

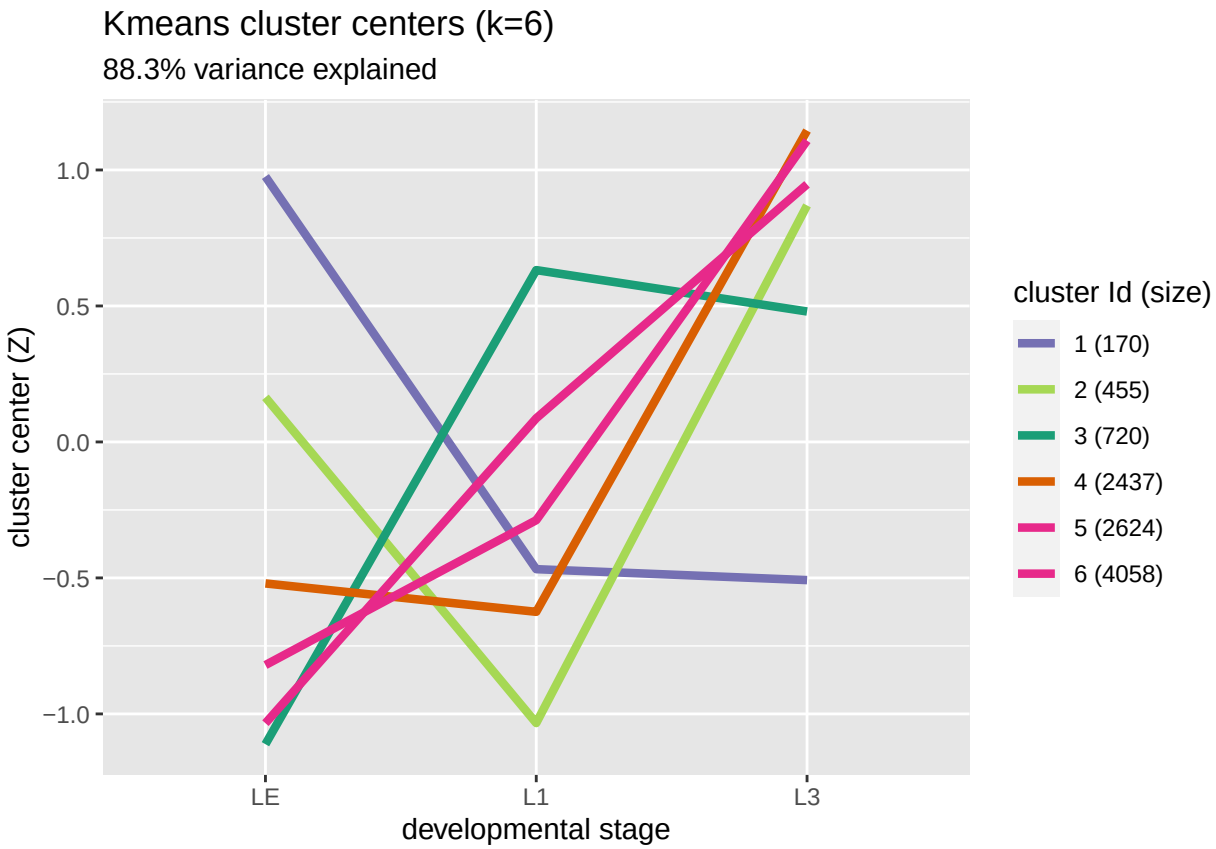

eans clusters at std. dev. > 0.383, (excludes lower 5.0%)
Cluster: 1 Size: 276
Cluster: 2 Size: 5428
Cluster: 3 Size: 2797
Cluster: 4 Size: 1963
1
0.5
0
−0.5
−1
LE
L1
L3

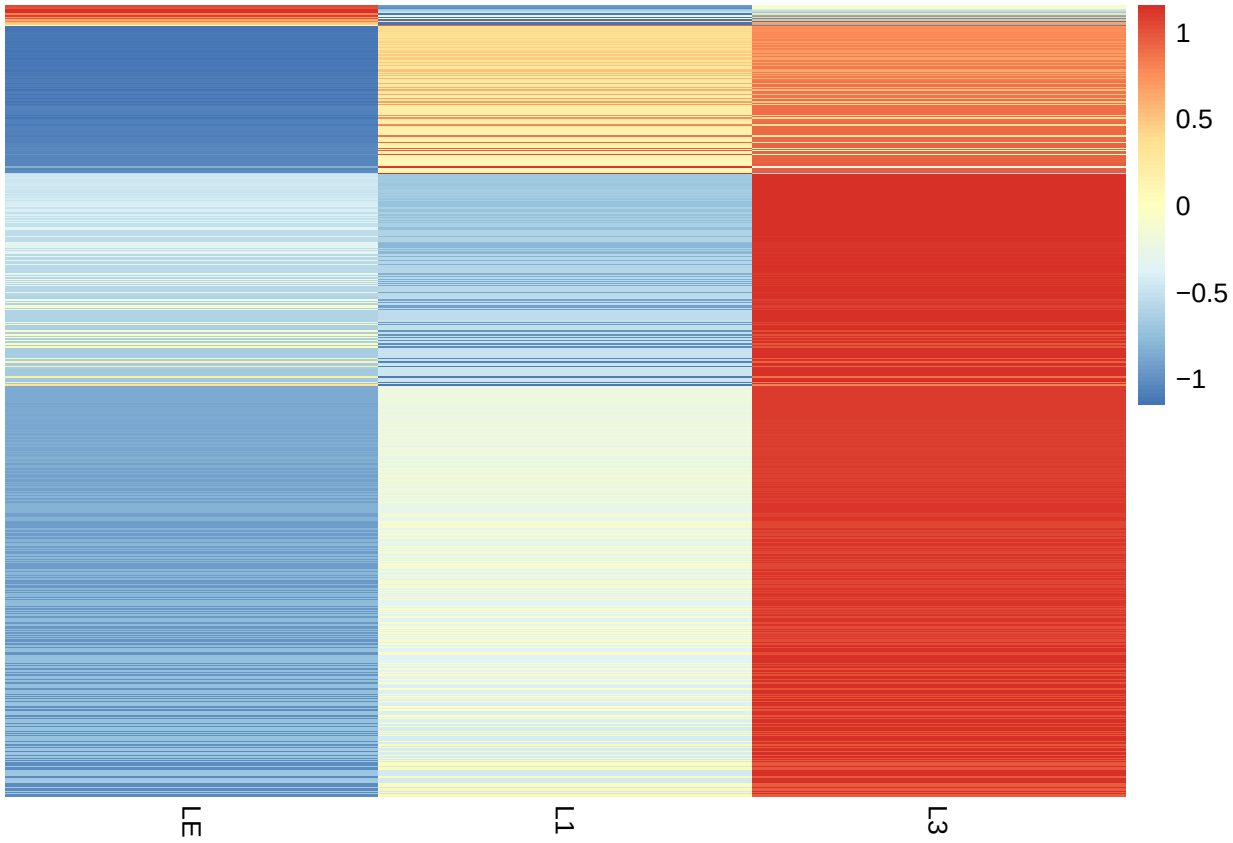

1
0.5
0
−0.5
−1
LE
L1
L3

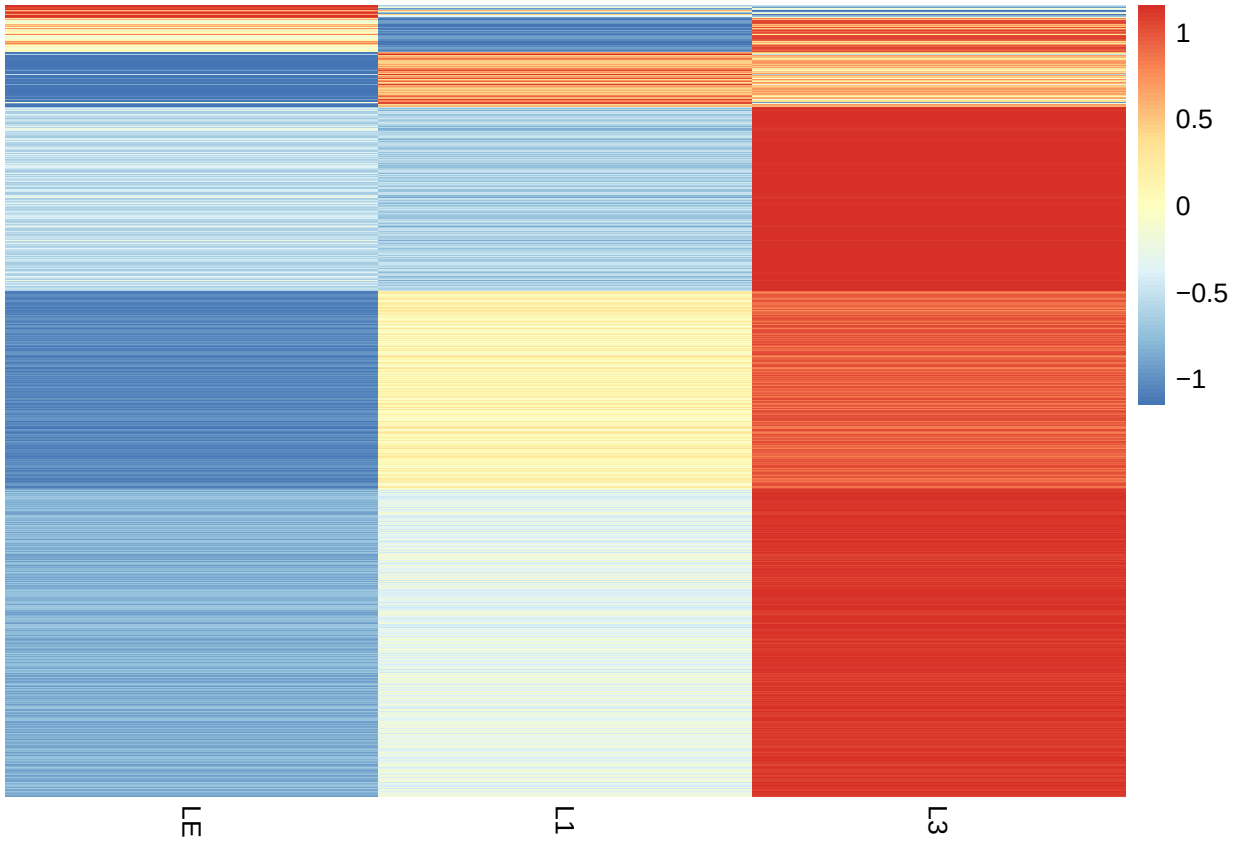

Kmeans cluster centers (k=4)
78.0% variance explained
cluster center (Z)
1.0
0.5
0.0
−0.5
−1.0
LE
L1
L3
developmental stage
description
LE−specific
Post embryonic
L3−high
Increasing




```
## Using label as id variables
```


Kmeans cluster centers (k=6)
88.3% variance explained
cluster center (Z)
1.0
0.5
0.0
−0.5
−1.0
LE
L1
L3
developmental stage
cluster Id (size)
1 (170)
2 (455)
3 (720)
4 (2437)
5 (2624)
6 (4058)


1
0.5
0
−0.5
−1
LE
L1
L3

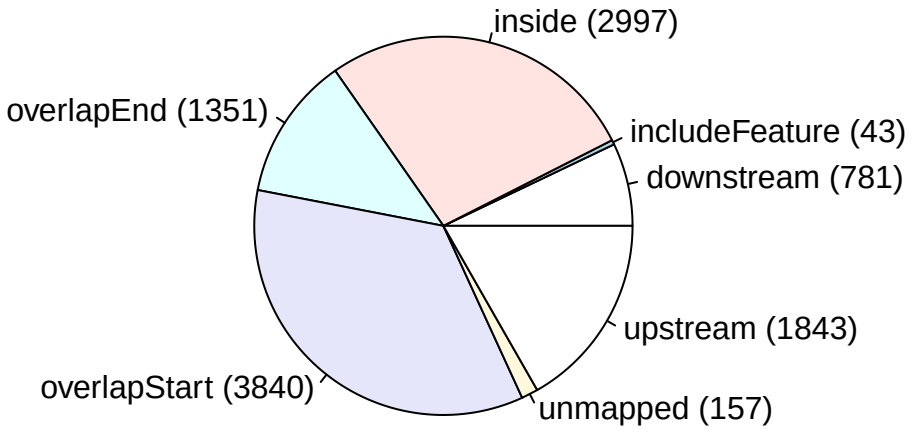


Prepare data structures to output bigBed files.

Custom tracks are served locally (but world-readable). The chunk currently also maps peaks to genes through the "scripts/getCodingGenes.R".

```
## Warning in annotatePeakInBatch(peaks, AnnotationData = all_CDS_genes, bindingRegions = c(-within_gene
##                       in the AnnotationData.

## Warning in FUN(extractROWS(unlisted_X, IRanges(X_elt_start[i], X_elt_end[i])), : not all the seqnames
##                       in the AnnotationData.

## Warning in FUN(extractROWS(unlisted_X, IRanges(X_elt_start[i], X_elt_end[i])), : not all the seqnames
##                       in the AnnotationData.

## Warning in FUN(extractROWS(unlisted_X, IRanges(X_elt_start[i], X_elt_end[i])), : not all the seqnames
##                       in the AnnotationData.

## Warning in annotatePeakInBatch(no_overlap_peaks, AnnotationData = all_CDS_genes, : not all the seqnar
##                       in the AnnotationData.

## Warning in annotatePeakInBatch(no_overlap_peaks, AnnotationData = all_CDS_genes, : not all the seqnar
##                       in the AnnotationData.

## Warning in annotatePeakInBatch(no_overlap_peaks, AnnotationData = all_CDS_genes, : not all the seqnar
##                       in the AnnotationData.

## Warning in annotatePeakInBatch(no_overlap_peaks, AnnotationData = all_CDS_genes, : not all the seqnar
##                       in the AnnotationData.
```

```
##    user  system elapsed
##   4.377   0.067   4.451
```



The closest feature to a peak is mapped unless farther than 5KB.
Feature" means the peak completely encompasses the gene (might be a consequence o

A BASH chunk, bedToBigBed

Run UCSC-user apps tools.

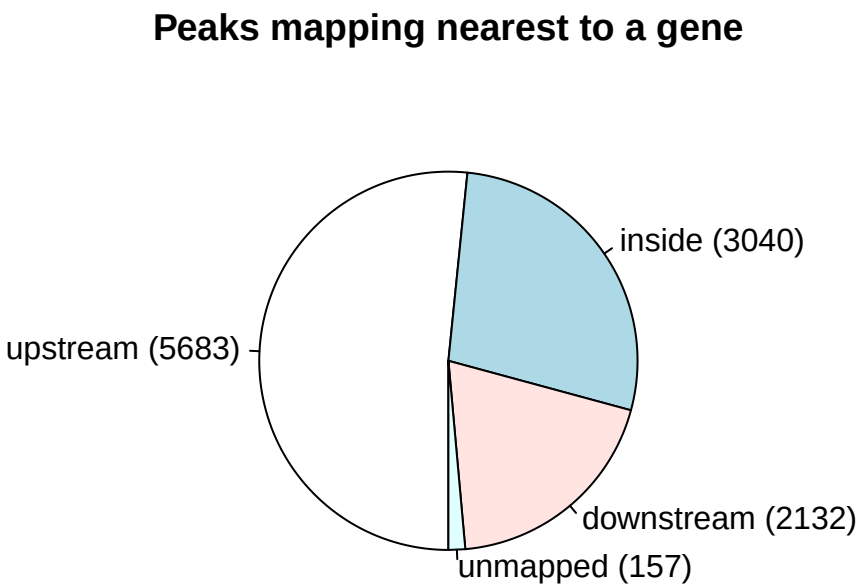
```
## pass1 - making usageList (6 chroms): 2 millis
## pass2 - checking and writing primary data (11012 records, 12 fields): 28 millis
## Sorting and writing extra index 0: 2 millis
```

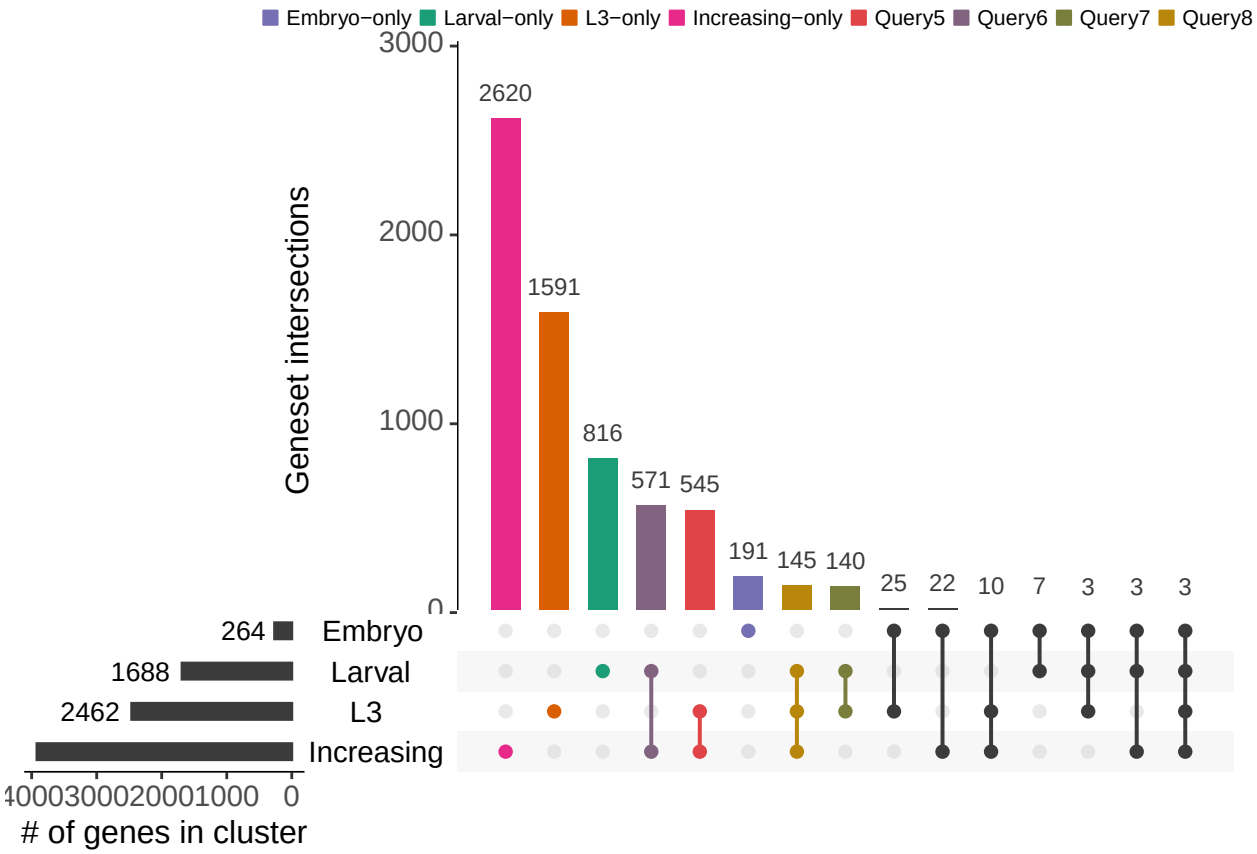
Plot gene mapping stats.

Pie chart for the relative locations and counts of how the peaks mapped to genes.

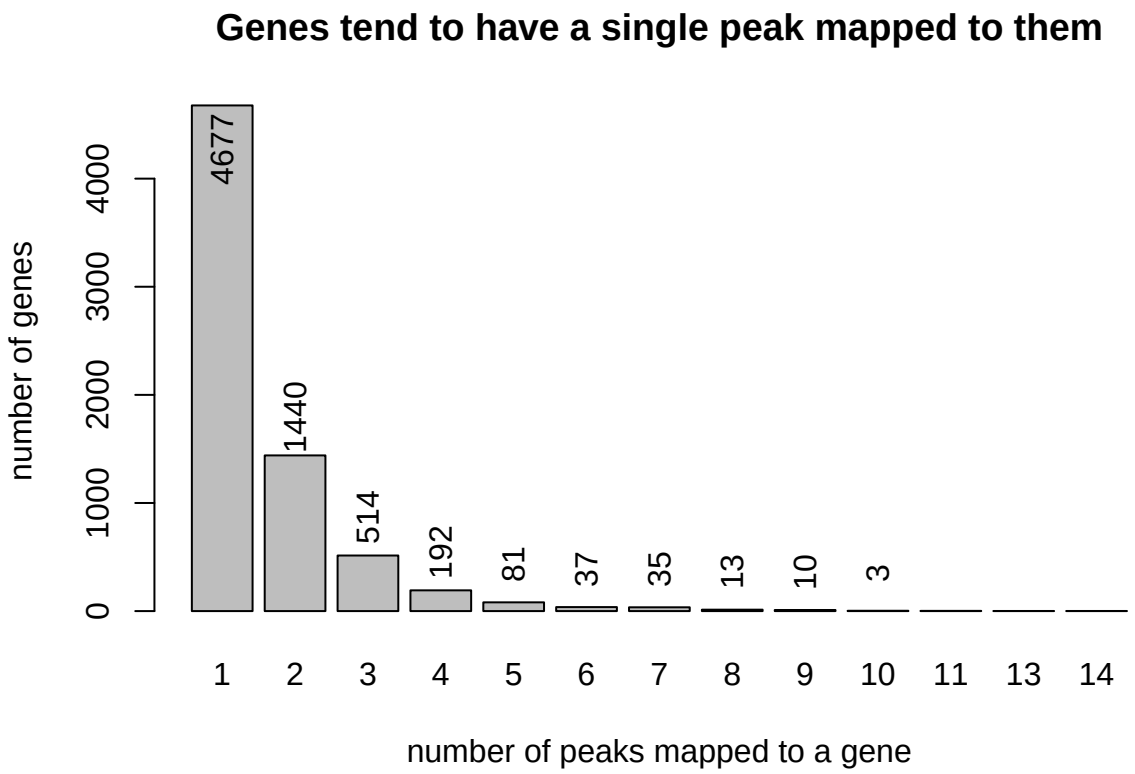
```
## pdf
##   2
```

```
## [1] 525
## [1] 264
## [1] 1688
## [1] 2462
## [1] 3919
```


Peaks mapping nearest to a gene
inside (3040)
upstream (5683)
downstream (2132)
unmapped (157)


Embryo−only
Larval−only
L3−only
Increasing−only
Query5
Query6
Query7
Query8
Geneset intersections
3000
2000
1000
0
2620
1591
816
571
545
191
145
140
25
22
10
7
3
3
3
264
Embryo
1688
Larval
2462
L3
Increasing
4000
3000
2000
1000
0
# of genes in cluster

```
## [1] 2329
```



GO data and processing functions

Data is from paramart, the functions are necessary to make the topGO analysis easier to repeat across the clusters.

```
##    user  system elapsed
##   0.609   0.008   3.943
```

```
##    user  system elapsed
##   0.458   0.006   0.464
```

Perform GO-term enrichment analyses

Run the topGO analyses using the above functions.

Output GO term enrichment results. Also, write files to paste into Revigo analysis, which collapses terms.

Paste output two-column files (GO-term p-like-value) into the form at. http://revigo.irb.hr/index.jsp

Table 1: Dataset versus genome

GO.ID	Term	Annotated	Significant	Expected	pval	DB
GO: 0005737	cytoplasm	3730	1959	1415.03	0.0e+00	CC
GO: 0005634	nucleus	2541	1209	963.97	0.0e+00	CC

GO.ID	Term	Annotated	Significant	Expected	pval	DB
GO: 0005829	cytosol	556	324	210.93	0.0e+00	CC
GO: 0045087	innate immune response	230	169	102.05	0.0e+00	BP
GO: 0008340	determination of adult lifespan	253	165	112.26	0.0e+00	BP
GO: 0005739	mitochondrion	624	328	236.72	0.0e+00	CC
GO: 0045121	membrane raft	48	40	18.21	0.0e+00	CC
GO: 0043231	intracellular membrane-bounded organelle	4042	2044	1533.40	0.0e+00	CC
GO: 0002119	nematode larval development	456	276	202.33	0.0e+00	BP
GO: 0005783	endoplasmic reticulum	388	218	147.19	0.0e+00	CC
GO: 0009792	embryo development ending in birth or eg...	305	184	135.33	0.0e+00	BP
GO: 0005777	peroxisome	66	52	25.04	0.0e+00	CC
GO: 0046872	metal ion binding	1689	872	732.41	0.0e+00	MF
GO: 0016323	basolateral plasma membrane	46	34	17.45	2.6e-06	CC
GO: 0051082	unfolded protein binding	71	50	30.79	3.4e-06	MF
GO: 0031581	hemidesmosome assembly	15	15	6.66	5.0e-06	BP
GO: 0005759	mitochondrial matrix	134	71	50.84	5.5e-06	CC
GO: 0040018	positive regulation of multicellular org...	49	37	21.74	8.9e-06	BP
GO: 0005764	lysosome	108	68	40.97	1.1e-05	CC
GO: 0022625	cytosolic large ribosomal subunit	50	34	18.97	1.5e-05	CC
GO: 0016324	apical plasma membrane	72	47	27.31	1.9e-05	CC
GO: 0005789	endoplasmic reticulum membrane	247	135	93.70	2.3e-05	CC
GO: 0005524	ATP binding	972	482	421.49	2.4e-05	MF
GO: 0003707	steroid hormone receptor activity	43	32	18.65	3.5e-05	MF
GO: 0012501	programmed cell death	165	92	73.21	4.5e-05	BP
GO: 0005788	endoplasmic reticulum lumen	24	19	9.10	4.5e-05	CC
GO: 0055114	oxidation-reduction process	645	351	286.19	4.8e-05	BP
GO: 0043401	steroid hormone mediated signaling pathw...	43	32	19.08	6.0e-05	BP

GO.ID	Term	Annotated	Significant	Expected	pval	DB
GO: 0050829	defense response to Gram-negative bacter...	21	18	9.32	1.1e-04	BP
GO: 0005623	cell	103	58	39.07	1.1e-04	CC
GO: 0000287	magnesium ion binding	48	34	20.81	1.1e-04	MF
GO: 0003824	catalytic activity	4165	2005	1806.09	1.9e-04	MF
GO: 0048471	perinuclear region of cytoplasm	85	49	32.25	2.0e-04	CC
GO: 0000045	autophagosome assembly	38	30	16.86	2.2e-04	BP
GO: 0003723	RNA binding	576	300	249.77	2.2e-04	MF
GO: 0030421	defecation	48	35	21.30	2.3e-04	BP
GO: 0030170	pyridoxal phosphate binding	51	35	22.12	2.3e-04	MF
GO: 0005667	transcription regulator complex	143	72	54.25	2.6e-04	CC
GO: 0005096	GTPase activator activity	77	49	33.39	2.6e-04	MF
GO: 0016311	dephosphorylation	272	116	120.69	3.1e-04	BP
GO: 0042802	identical protein binding	115	73	49.87	3.3e-04	MF
GO: 0008270	zinc ion binding	555	280	240.67	3.3e-04	MF
GO: 0016887	ATPase activity	275	140	119.25	4.2e-04	MF
GO: 0030173	integral component of Golgi membrane	36	24	13.66	4.4e-04	CC
GO: 0015020	glucuronosyltransferase activity	39	26	16.91	4.9e-04	MF
GO: 0040010	positive regulation of growth rate	24	19	10.65	5.4e-04	BP
GO: 0035973	aggrephagy	9	9	3.99	6.6e-04	BP
GO: 0006099	tricarboxylic acid cycle	26	20	11.54	7.5e-04	BP
GO: 0000793	condensed chromosome	133	61	50.46	7.6e-04	CC
GO: 0006631	fatty acid metabolic process	121	89	53.69	8.6e-04	BP

Table 2: Not changing versus whole dataset

GO.ID	Term	Annotated	Significant	Expected	pval	DB
GO:0008083	growth factor activity	3	3	0.23	0.00047	MF
GO:0030628	pre-mRNA 3'-splice site binding	3	3	0.23	0.00047	MF
GO:0007399	nervous system development	180	31	14.42	0.00200	BP

GO.ID	Term	Annotated	Significant	Expected	pval	DB
GO:0008045	motor neuron axon guidance	18	6	1.44	0.00200	BP
GO:0008584	male gonad development	5	3	0.40	0.00450	BP
GO:0030150	protein import into mitochondrial matrix	5	3	0.40	0.00450	BP
GO:0043025	neuronal cell body	87	13	6.81	0.00530	CC
GO:0004831	tyrosine-tRNA ligase activity	2	2	0.16	0.00606	MF
GO:0005125	cytokine activity	2	2	0.16	0.00606	MF
GO:0030121	AP-1 adaptor complex	2	2	0.16	0.00610	CC
GO:0071565	nBAF complex	2	2	0.16	0.00610	CC
GO:0089701	U2AF	2	2	0.16	0.00610	CC
GO:0009190	cyclic nucleotide biosynthetic process	14	3	1.12	0.00640	BP
GO:0006437	tyrosyl-tRNA aminoacylation	2	2	0.16	0.00640	BP
GO:0010696	positive regulation of mitotic spindle p...	2	2	0.16	0.00640	BP
GO:0010862	positive regulation of pathway-restricte...	2	2	0.16	0.00640	BP
GO:0060395	SMAD protein signal transduction	2	2	0.16	0.00640	BP
GO:0072321	chaperone-mediated protein transport	2	2	0.16	0.00640	BP
GO:0035335	peptidyl-tyrosine dephosphorylation	29	7	2.32	0.00660	BP
GO:0043022	ribosome binding	11	4	0.86	0.00771	MF
GO:0030687	preribosome, large subunit precursor	11	4	0.86	0.00780	CC
GO:0000723	telomere maintenance	11	4	0.88	0.00840	BP
GO:0001054	RNA polymerase I activity	7	3	0.55	0.01298	MF
GO:0005736	RNA polymerase I complex	7	3	0.55	0.01310	CC
GO:0007419	ventral cord development	7	3	0.56	0.01400	BP
GO:0019784	NEDD8-specific protease activity	3	2	0.23	0.01725	MF
GO:0030993	axonemal heterotrimeric kinesin-II compl...	3	2	0.23	0.01740	CC
GO:1990423	RZZ complex	3	2	0.23	0.01740	CC
GO:0007274	neuromuscular synaptic transmission	6	3	0.48	0.01810	BP
GO:0032968	positive regulation of transcription elo...	3	2	0.24	0.01820	BP
GO:0048730	epidermis morphogenesis	3	2	0.24	0.01820	BP
GO:0071108	protein K48-linked deubiquitination	3	2	0.24	0.01820	BP
GO:1903394	protein localization to kinetochore invo...	3	2	0.24	0.01820	BP
GO:1905342	positive regulation of protein localizat...	3	2	0.24	0.01820	BP
GO:0005524	ATP binding	482	50	37.58	0.01869	MF
GO:0005819	spindle	41	6	3.21	0.01970	CC
GO:1904115	axon cytoplasm	8	3	0.63	0.01980	CC
GO:0008355	olfactory learning	8	3	0.64	0.02110	BP
GO:0000381	regulation of alternative mRNA splicing,...	14	4	1.12	0.02130	BP
GO:0003677	DNA binding	456	35	35.55	0.02388	MF
GO:0005089	Rho guanyl-nucleotide exchange factor ac...	15	4	1.17	0.02488	MF
GO:0005515	protein binding	1622	154	126.46	0.02844	MF
GO:0005160	transforming growth factor beta receptor...	4	2	0.31	0.03273	MF
GO:0008353	RNA polymerase II CTD heptapeptide repea...	4	2	0.31	0.03273	MF
GO:0005744	TIM23 mitochondrial import inner membran...	5	3	0.39	0.03280	CC
GO:0035023	regulation of Rho protein signal transdu...	16	4	1.28	0.03400	BP
GO:0002225	positive regulation of antimicrobial pep...	4	2	0.32	0.03450	BP
GO:0006265	DNA topological change	4	2	0.32	0.03450	BP
GO:0008543	fibroblast growth factor receptor signal...	4	2	0.32	0.03450	BP
GO:0010754	negative regulation of cGMP-mediated sig...	4	2	0.32	0.03450	BP

Table 3: Embryo cluster versus whole dataset

GO.ID	Term	Annotated	Significant	Expected	pval	DB
GO:0005001	transmembrane receptor protein tyrosine ...	2	2	0.07	0.0012	MF
GO:0033563	dorsal/ventral axon guidance	14	4	0.55	0.0016	BP
GO:0010172	embryonic body morphogenesis	27	5	1.05	0.0034	BP
GO:0005604	basement membrane	9	3	0.35	0.0041	CC
GO:0048665	neuron fate specification	3	2	0.12	0.0044	BP
GO:0004115	3',5'-cyclic-AMP phosphodiesterase activ...	4	2	0.14	0.0068	MF
GO:0005576	extracellular region	181	11	7.06	0.0077	CC
GO:0007416	synapse assembly	24	3	0.94	0.0086	BP
GO:0040022	feminization of hermaphroditic germ-line	4	2	0.16	0.0086	BP
GO:0097376	interneuron axon guidance	4	2	0.16	0.0086	BP
GO:0004930	G protein-coupled receptor activity	71	7	2.45	0.0111	MF
GO:0038007	netrin-activated signaling pathway	5	2	0.19	0.0140	BP
GO:0048589	developmental growth	120	6	4.68	0.0140	BP
GO:0009952	anterior/posterior pattern specification	36	4	1.40	0.0202	BP
GO:0001228	DNA-binding transcription activator acti...	7	2	0.24	0.0222	MF
GO:0070006	metalloaminopeptidase activity	7	2	0.24	0.0222	MF
GO:0007218	neuropeptide signaling pathway	17	3	0.66	0.0264	BP
GO:0005747	mitochondrial respiratory chain complex ...	17	3	0.66	0.0265	CC
GO:0008045	motor neuron axon guidance	18	3	0.70	0.0308	BP
GO:0038023	signaling receptor activity	187	16	6.46	0.0329	MF
GO:0019900	kinase binding	72	5	2.49	0.0342	MF
GO:0000035	acyl binding	1	1	0.03	0.0345	MF
GO:0000036	acyl carrier activity	1	1	0.03	0.0345	MF
GO:0003827	alpha-1,3-mannosylglycoprotein 2-beta-N-...	1	1	0.03	0.0345	MF
GO:0004504	peptidylglycine monooxygenase activity	1	1	0.03	0.0345	MF
GO:0004566	beta-glucuronidase activity	1	1	0.03	0.0345	MF
GO:0004616	phosphogluconate dehydrogenase (decarbox...	1	1	0.03	0.0345	MF
GO:0004719	protein-L-isoaspartate (D-aspartate) O-m...	1	1	0.03	0.0345	MF
GO:0004818	glutamate-tRNA ligase activity	1	1	0.03	0.0345	MF
GO:0004935	adrenergic receptor activity	1	1	0.03	0.0345	MF
GO:0004989	octopamine receptor activity	1	1	0.03	0.0345	MF
GO:0005005	transmembrane-ephrin receptor activity	1	1	0.03	0.0345	MF
GO:0005518	collagen binding	1	1	0.03	0.0345	MF
GO:0008226	tyramine receptor activity	1	1	0.03	0.0345	MF
GO:0008467	[heparan sulfate]-glucosamine 3-sulfotra...	1	1	0.03	0.0345	MF
GO:0008474	palmitoyl-(protein) hydrolase activity	1	1	0.03	0.0345	MF
GO:0016155	formyltetrahydrofolate dehydrogenase act...	1	1	0.03	0.0345	MF
GO:0018738	S-formylglutathione hydrolase activity	1	1	0.03	0.0345	MF
GO:0019799	tubulin N-acetyltransferase activity	1	1	0.03	0.0345	MF
GO:0032216	glucosaminyl-phosphatidylinositol O-acyl...	1	1	0.03	0.0345	MF
GO:0038062	protein tyrosine kinase collagen recepto...	1	1	0.03	0.0345	MF
GO:0048101	calcium- and calmodulin-regulated 3',5'-...	1	1	0.03	0.0345	MF
GO:1990890	netrin receptor binding	1	1	0.03	0.0345	MF
GO:0007219	Notch signaling pathway	22	3	0.86	0.0360	BP
GO:0006730	one-carbon metabolic process	8	2	0.31	0.0362	BP
GO:0050769	positive regulation of neurogenesis	36	4	1.40	0.0386	BP
GO:0003388	neuron development involved in amphid se...	2	2	0.08	0.0387	BP
GO:0032878	regulation of establishment or maintenan...	8	2	0.31	0.0388	BP
GO:0000390	spliceosomal complex disassembly	1	1	0.04	0.0390	BP
GO:0000493	box H/ACA snoRNP assembly	1	1	0.04	0.0390	BP

Table 4: Larval cluster versus whole dataset

GO.ID	Term	Annotated	Significant	Expected	pval	DB
GO:0030246	carbohydrate binding	115	57	27.95	0.0000000	MF
GO:0045087	innate immune response	169	65	40.17	0.0000032	BP
GO:0045121	membrane raft	40	22	9.25	0.0000054	CC
GO:0000981	DNA-binding transcription factor activit...	58	25	14.10	0.0001900	MF
GO:0051603	proteolysis involved in cellular protein...	168	46	39.93	0.0010000	BP
GO:0004190	aspartic-type endopeptidase activity	17	11	4.13	0.0012700	MF
GO:0030163	protein catabolic process	203	61	48.25	0.0017000	BP
GO:0016936	galactoside binding	4	4	0.97	0.0034700	MF
GO:0004185	serine-type carboxypeptidase activity	8	6	1.94	0.0035900	MF
GO:0005882	intermediate filament	12	8	2.77	0.0047000	CC
GO:0005764	lysosome	68	26	15.72	0.0057000	CC
GO:0042761	very long-chain fatty acid biosynthetic ...	9	6	2.14	0.0076000	BP
GO:0043401	steroid hormone mediated signaling pathw...	32	14	7.61	0.0099000	BP
GO:0008270	zinc ion binding	280	85	68.05	0.0100800	MF
GO:0000977	RNA polymerase II transcription regulato...	67	27	16.28	0.0103600	MF
GO:0003707	steroid hormone receptor activity	32	14	7.78	0.0120900	MF
GO:0050830	defense response to Gram-positive bacter...	15	8	3.57	0.0126000	BP
GO:0030590	first cell cycle pseudocleavage	5	4	1.19	0.0129000	BP
GO:0048268	clathrin coat assembly	5	4	1.19	0.0129000	BP
GO:0042157	lipoprotein metabolic process	36	9	8.56	0.0130000	BP
GO:0043413	macromolecule glycosylation	56	12	13.31	0.0131000	BP
GO:0031054	pre-miRNA processing	3	3	0.71	0.0134000	BP
GO:0035461	vitamin transmembrane transport	3	3	0.71	0.0134000	BP
GO:0043327	chemotaxis to cAMP	3	3	0.71	0.0134000	BP
GO:0070574	cadmium ion transmembrane transport	3	3	0.71	0.0134000	BP
GO:0070781	response to biotin	3	3	0.71	0.0134000	BP
GO:0072488	ammonium transmembrane transport	3	3	0.71	0.0134000	BP
GO:1905803	negative regulation of cellular response...	3	3	0.71	0.0134000	BP
GO:0022857	transmembrane transporter activity	374	105	90.90	0.0135400	MF
GO:0004523	RNA-DNA hybrid ribonuclease activity	5	4	1.22	0.0140000	MF
GO:0005381	iron ion transmembrane transporter activ...	5	4	1.22	0.0140000	MF
GO:0008107	galactoside 2-alpha-L-fucosyltransferase...	5	4	1.22	0.0140000	MF
GO:0036459	thiol-dependent ubiquitinyl hydrolase ac...	20	5	4.86	0.0141300	MF
GO:0004022	alcohol dehydrogenase (NAD+) activity	3	3	0.73	0.0143300	MF
GO:0005542	folic acid binding	3	3	0.73	0.0143300	MF
GO:0015086	cadmium ion transmembrane transporter ac...	3	3	0.73	0.0143300	MF
GO:0032051	clathrin light chain binding	3	3	0.73	0.0143300	MF
GO:0071949	FAD binding	18	9	4.37	0.0159100	MF
GO:0017147	Wnt-protein binding	10	6	2.43	0.0170400	MF
GO:0016887	ATPase activity	140	39	34.03	0.0186300	MF
GO:0030148	sphingolipid biosynthetic process	25	9	5.94	0.0219000	BP
GO:0043235	receptor complex	27	11	6.24	0.0255000	CC
GO:0006508	proteolysis	328	88	77.97	0.0259000	BP
GO:0040034	regulation of development, heterochronic	15	8	3.57	0.0291000	BP
GO:0005576	extracellular region	181	57	41.84	0.0293000	CC
GO:0008340	determination of adult lifespan	165	50	39.22	0.0298000	BP
GO:0007606	sensory perception of chemical stimulus	36	13	8.56	0.0298000	BP
GO:0042594	response to starvation	32	11	7.61	0.0301000	BP
GO:0004672	protein kinase activity	193	60	46.91	0.0308500	MF
GO:0007168	receptor guanylyl cyclase signaling path...	6	4	1.43	0.0314000	BP

Table 5: L3 cluster versus whole dataset

GO.ID	Term	Annotated	Significant	Expected	pval	DB
GO:0003735	structural constituent of ribosome	81	54	30.44	0.0000001	MF
GO:0006412	translation	181	102	68.81	0.0000003	BP
GO:0003697	single-stranded DNA binding	17	15	6.39	0.0000110	MF
GO:0005840	ribosome	92	61	34.37	0.0001100	CC
GO:0000786	nucleosome	37	25	13.82	0.0001800	CC
GO:0005634	nucleus	1209	539	451.69	0.0002100	CC
GO:0008340	determination of adult lifespan	165	85	62.73	0.0002200	BP
GO:0000398	mRNA splicing, via spliceosome	63	39	23.95	0.0002800	BP
GO:0022625	cytosolic large ribosomal subunit	34	23	12.70	0.0003100	CC
GO:0003677	DNA binding	456	198	171.37	0.0003200	MF
GO:0046982	protein heterodimerization activity	48	30	18.04	0.0003800	MF
GO:0003688	DNA replication origin binding	8	8	3.01	0.0003900	MF
GO:0006334	nucleosome assembly	29	20	11.03	0.0006300	BP
GO:0009792	embryo development ending in birth or eg...	184	91	69.95	0.0008000	BP
GO:0033290	eukaryotic 48S preinitiation complex	7	7	2.62	0.0010100	CC
GO:0005737	cytoplasm	1959	831	731.90	0.0021800	CC
GO:0071013	catalytic step 2 spliceosome	18	13	6.72	0.0024600	CC
GO:0022627	cytosolic small ribosomal subunit	16	12	5.98	0.0024700	CC
GO:0019843	rRNA binding	23	16	8.64	0.0025900	MF
GO:0005669	transcription factor TFIID complex	6	6	2.24	0.0027000	CC
GO:0016282	eukaryotic 43S preinitiation complex	6	6	2.24	0.0027000	CC
GO:0051123	RNA polymerase II preinitiation complex ...	9	8	3.42	0.0029900	BP
GO:0000462	maturation of SSU-rRNA from tricistronic...	12	10	4.56	0.0039000	BP
GO:0046872	metal ion binding	872	338	327.70	0.0043700	MF
GO:0005814	centriole	8	7	2.99	0.0054400	CC
GO:0046716	muscle cell cellular homeostasis	8	7	3.04	0.0060900	BP
GO:0005687	U4 snRNP	5	5	1.87	0.0072500	CC
GO:0080008	Cul4-RING E3 ubiquitin ligase complex	5	5	1.87	0.0072500	CC
GO:0046686	response to cadmium ion	13	9	4.94	0.0078900	BP
GO:0000727	double-strand break repair via break-ind...	5	5	1.90	0.0079100	BP
GO:0000028	ribosomal small subunit assembly	10	8	3.80	0.0085700	BP
GO:0045944	positive regulation of transcription by ...	112	57	42.58	0.0089400	BP
GO:0000794	condensed nuclear chromosome	21	13	7.85	0.0101100	CC
GO:0044877	protein-containing complex binding	113	64	42.47	0.0105600	MF
GO:0005686	U2 snRNP	7	6	2.62	0.0128800	CC
GO:0017025	TBP-class protein binding	7	6	2.63	0.0133100	MF
GO:0007212	dopamine receptor signaling pathway	8	7	3.04	0.0141500	BP
GO:0005884	actin filament	15	12	5.60	0.0163800	CC
GO:0005682	U5 snRNP	9	7	3.36	0.0165400	CC
GO:0003743	translation initiation factor activity	27	16	10.15	0.0178500	MF
GO:0001732	formation of cytoplasmic translation ini...	9	7	3.42	0.0183700	BP
GO:0005852	eukaryotic translation initiation factor...	6	6	2.24	0.0193800	CC
GO:0034719	SMN-Sm protein complex	4	4	1.49	0.0194400	CC
GO:0042555	MCM complex	4	4	1.49	0.0194400	CC
GO:0003746	translation elongation factor activity	11	8	4.13	0.0194500	MF
GO:0005829	cytosol	324	155	121.05	0.0196400	CC
GO:0004861	cyclin-dependent protein serine/threonin...	4	4	1.50	0.0199000	MF
GO:0017124	SH3 domain binding	4	4	1.50	0.0199000	MF
GO:0035196	production of miRNAs involved in gene si...	8	6	3.04	0.0208100	BP
GO:0030866	cortical actin cytoskeleton organization	15	11	5.70	0.0208400	BP

Table 6: Increasing cluster versus whole dataset

GO.ID	Term	Annotated	Significant	Expected	pval	DB
GO: 0006355	regulation of transcription, DNA-templat...	455	299	252.77	0.0000007	BP
GO: 0003700	DNA-binding transcription factor activit...	257	180	145.30	0.0000008	MF
GO: 0043565	sequence-specific DNA binding	262	177	148.13	0.0000046	MF
GO: 0045087	innate immune response	169	116	93.88	0.0008500	BP
GO: 0046777	protein autophosphorylation	22	17	12.22	0.0032200	BP
GO: 0010494	cytoplasmic stress granule	9	9	4.98	0.0048000	CC
GO: 0032436	positive regulation of proteasomal ubiqu...	23	19	12.78	0.0062200	BP
GO: 0009058	biosynthetic process	1095	630	608.30	0.0081600	BP
GO: 0009116	nucleoside metabolic process	28	20	15.55	0.0089900	BP
GO: 0042048	olfactory behavior	20	17	11.11	0.0090200	BP
GO: 0016324	apical plasma membrane	47	36	25.98	0.0101000	CC
GO: 0018105	peptidyl-serine phosphorylation	47	34	26.11	0.0105900	BP
GO: 0006357	regulation of transcription by RNA polym...	210	134	116.66	0.0126500	BP
GO: 0045167	asymmetric protein localization involved...	13	11	7.22	0.0151500	BP
GO: 0055085	transmembrane transport	413	235	229.43	0.0161500	BP
GO: 0006636	unsaturated fatty acid biosynthetic proc...	7	7	3.89	0.0162600	BP
GO: 0055114	oxidation-reduction process	351	214	194.99	0.0166200	BP
GO: 0016310	phosphorylation	350	212	194.43	0.0174800	BP
GO: 0042803	protein homodimerization activity	27	21	15.27	0.0180000	MF
GO: 0040024	dauer larval development	57	44	31.67	0.0187400	BP
GO: 0008270	zinc ion binding	280	175	158.31	0.0210000	MF
GO: 0010468	regulation of gene expression	649	403	360.54	0.0238000	BP
GO: 0045179	apical cortex	6	6	3.32	0.0285000	CC
GO: 0000132	establishment of mitotic spindle orienta...	25	19	13.89	0.0288400	BP
GO: 0061063	positive regulation of nematode larval d...	36	24	20.00	0.0292000	BP

GO.ID	Term	Annotated	Significant	Expected	pval	DB
GO: 0019367	fatty acid elongation, saturated fatty a...	6	6	3.33	0.0293100	BP
GO: 0034625	fatty acid elongation, monounsaturated f...	6	6	3.33	0.0293100	BP
GO: 0034626	fatty acid elongation, polyunsaturated f...	6	6	3.33	0.0293100	BP
GO: 0048557	embryonic digestive tract morphogenesis	16	12	8.89	0.0293200	BP
GO: 0005634	nucleus	1209	653	668.37	0.0296000	CC
GO: 0009922	fatty acid elongase activity	6	6	3.39	0.0330000	MF
GO: 0102336	3-oxo-arachidoyl-CoA synthase activity	6	6	3.39	0.0330000	MF
GO: 0102337	3-oxo-cerotoyl-CoA synthase activity	6	6	3.39	0.0330000	MF
GO: 0102338	3-oxo-lignoceronyl-CoA synthase activity	6	6	3.39	0.0330000	MF
GO: 0102756	very-long-chain 3-ketoacyl-CoA synthase ...	6	6	3.39	0.0330000	MF
GO: 0042626	ATPase-coupled transmembrane transporter...	52	37	29.40	0.0340000	MF
GO: 0016323	basolateral plasma membrane	34	23	18.80	0.0358000	CC
GO: 0005667	transcription regulator complex	72	43	39.80	0.0360000	CC
GO: 0035556	intracellular signal transduction	274	169	152.21	0.0371600	BP
GO: 0000003	reproduction	437	250	242.77	0.0372700	BP
GO: 0009986	cell surface	12	9	6.63	0.0399000	CC
GO: 1990573	potassium ion import across plasma membr...	10	9	5.56	0.0411100	BP
GO: 0007614	short-term memory	9	8	5.00	0.0411600	BP
GO: 0009786	regulation of asymmetric cell division	9	8	5.00	0.0411600	BP
GO: 0005912	adherens junction	18	14	9.95	0.0429000	CC
GO: 0030659	cytoplasmic vesicle membrane	48	28	26.54	0.0440000	CC
GO: 0002119	nematode larval development	276	181	153.33	0.0447000	BP
GO: 0012501	programmed cell death	92	59	51.11	0.0453300	BP
GO: 0006629	lipid metabolic process	302	193	167.77	0.0462900	BP
GO: 0016311	dephosphorylation	116	74	64.44	0.0465000	BP

```
write.table(o$all[,c(1,8)], "gene_lists/all.REVIGO/all.forREVIGO.txt",quote=F,col.names=F,row.names=F)
write.table(o$K0[,c(1,8)], "gene_lists/K0.REVIGO/K0.forREVIGO.txt",quote=F,col.names=F,row.names=F)
write.table(o$K1[,c(1,8)], "gene_lists/K1.REVIGO/K1.forREVIGO.txt",quote=F,col.names=F,row.names=F)
write.table(o$K2[,c(1,8)], "gene_lists/K2.REVIGO/K2.forREVIGO.txt",quote=F,col.names=F,row.names=F)
write.table(o$K3[,c(1,8)], "gene_lists/K3.REVIGO/K3.forREVIGO.txt",quote=F,col.names=F,row.names=F)
write.table(o$K4[,c(1,8)], "gene_lists/K4.REVIGO/K4.forREVIGO.txt",quote=F,col.names=F,row.names=F)
```

```
pdf("GOplots.pdf")

o[['all']] %>% rename(term_ID=GO.ID) -> okAll
bp=(read.csv('gene_lists/all.REVIGO/all.BP.69.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]
cc=(read.csv('gene_lists/all.REVIGO/all.CC.69.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]
mf=(read.csv('gene_lists/all.REVIGO/all.MF.69.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]

okAll%>% filter(term_ID %in% c(bp,cc,mf)) %>% arrange(`Rank in elim`,elim) %>% select(Term, term_ID, eli
xx$Term = factor(xx$Term, levels=rev(xx$Term))
ggplot(xx) +
  geom_col(aes(nlogp,Term,fill=DB),color='black' )+
  theme(legend.position = 'none') +
  theme_bw() + scale_fill_manual("ontology",values=c('#404040','grey','white')) +
  ggtitle("Terms enriched in genes mapped to all peaks (versus genome)")



o[['K0']] %>% rename(term_ID=GO.ID) -> ok0

bp=(read.csv('gene_lists/K0.REVIGO/K0.BP.69.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]
cc=(read.csv('gene_lists/K0.REVIGO/K0.CC.69.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]
mf=(read.csv('gene_lists/K0.REVIGO/K0.MF.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]

ok0 %>% filter(term_ID %in% c(bp,cc,mf)) %>% arrange(`Rank in elim`,elim) %>% select(Term, term_ID, elin
xx$Term = factor(xx$Term, levels=rev(xx$Term))
ggplot(xx) +
  geom_col(aes(nlogp,Term,fill=DB),color='black' )+
  theme(legend.position = 'none') +
  theme_bw() + scale_fill_manual("ontology",values=c('#404040','grey','white')) +
  ggtitle("Terms enriched in genes mapped to 'not changing' peaks")



o[['K1']] %>% rename(term_ID=GO.ID) -> ok1

bp=(read.csv('gene_lists/K1.REVIGO/K1.BP.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]
cc=(read.csv('gene_lists/K1.REVIGO/K1.CC.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]
mf=(read.csv('gene_lists/K1.REVIGO/REVIGO.MF.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]

ok1 %>% filter(term_ID %in% c(bp,cc,mf)) %>% arrange(`Rank in elim`,elim) %>% select(Term, term_ID, elin
xx$Term = factor(xx$Term, levels=rev(xx$Term))
ggplot(xx) +
  geom_col(aes(nlogp,Term,fill=DB),color='black' )+
  theme(legend.position = 'none') +
  theme_bw() + scale_fill_manual("ontology",values=c('#404040','grey','white')) +
  ggtitle("Terms enriched in genes mapped to embryo-high peaks")
```

```
o[['K2']] %>% rename(term_ID=GO.ID) -> ok2

bp=(read.csv('gene_lists/K2.REVIGO/K2.BP.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]
cc=(read.csv('gene_lists/K2.REVIGO/K2.CC.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]
mf=(read.csv('gene_lists/K2.REVIGO/K2.MF.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]

ok2 %>% filter(term_ID %in% c(bp,cc,mf)) %>% arrange(`Rank in elim`,elim) %>% select(Term, term_ID, elim
xx$Term = factor(xx$Term, levels=rev(xx$Term))
ggplot(xx) +
  geom_col(aes(nlogp,Term,fill=DB),color='black' )+
  theme(legend.position = 'none') +
  theme_bw() + scale_fill_manual("ontology",values=c('#404040','grey','white')) +
  ggtitle("Terms enriched in genes mapped to larval-high peaks")

o[['K3']] %>% rename(term_ID=GO.ID) -> ok3

bp=(read.csv('gene_lists/K3.REVIGO/K3.BP.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]
cc=(read.csv('gene_lists/K3.REVIGO/K3.CC.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]
mf=(read.csv('gene_lists/K3.REVIGO/K3.MF.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]


ok3 %>% filter(term_ID %in% c(bp,cc,mf)) %>% arrange(`Rank in elim`,elim) %>% select(Term, term_ID, elim
xx$Term = factor(xx$Term, levels=rev(xx$Term))
ggplot(xx) +
 geom_col(aes(nlogp,Term,fill=DB),color='black' ) +
  theme(legend.position = 'none') +
  theme_bw() + scale_fill_manual("ontology",values=c('#404040','grey','white')) +
  ggtitle("Terms enriched in genes mapped to L3-high peaks")


o[['K4']] %>% rename(term_ID=GO.ID) -> ok4

bp=(read.csv('gene_lists/K4.REVIGO/K4.BP.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]
cc=(read.csv('gene_lists/K4.REVIGO/K4.CC.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]
mf=(read.csv('gene_lists/K4.REVIGO/K4.MF.csv') %>% filter(eliminated == 0) %>% select(term_ID))[[1]]

ok4 %>% filter(term_ID %in% c(bp,cc,mf)) %>% arrange(`Rank in elim`,elim) %>% select(Term, term_ID, elim

xx$Term = factor(xx$Term, levels=rev(xx$Term))
ggplot(xx) +
 geom_col(aes(nlogp,Term,fill=DB),color='black' ) +
  theme(legend.position = 'none') +
  theme_bw() + scale_fill_manual("ontology",values=c('#404040','grey','white')) +
  ggtitle("Terms enriched in genes mapped to Increasing peaks")

dev.off()
```

```
## pdf
##   2
```