Title option 1:
Title option 2:
Title option 3:

Abstract

Introduction

Methods

Experiment

Transcriptome Assembly (Trinity)

Sequencing reads from all six libraries were assembled using Trinity (version 2.0.6). As part of the Trinity package reads were quality trimmed (Trimmomatic) and normalized prior to assembly with minimum kmer coverage of 2 and minimum contig length of 200bp. Assessment of transcriptome quality and completeness was performed using Transrate. To further assess completeness of the transcriptome contigs were compared to ???(Cgigas and Moriera Mytilus)???? The Transdecoder package that is part of Trinity was used to predict putative corresponding proteins.

Transcriptome Annotation (Trinity)

Resulting contigs were first compared to the NCBI nucleotide (nt) database to identify any non target taxa sequences (ie bacteria) and these were removed from further analysis. Contigs were annotated by comparing contiguous sequences to the UniProtKB/Swiss-Prot database. Comparisons were made using the BLASTx algorithm with a 1.0E-5 e-value threshold. Genes were classified according to Swiss-Prot Gene Ontology (GO) associations, as well as respective parent categories (GO Slim). Annotation analyses and data are published

Differential Expression

Transcript abundance was determined with Kallisto (ref) as part of a perl script (align_and_estimate_abundance.pl) as part of the Trinity package. Transcript abundances from the six libraries were used to construct the transcript and gene expression matrices (abundance_estimates_to_matrix.pl) used to identify differentially expressed transcripts with EdgeR (dispersion value 0.4).

Long non-coding RNA identification (CLC + online tools)

Results

Experiment

Transcriptome Assembly (Trinity)

Following quality trimming, 792,714,472 (99%) of reads were assembled into 184834 transcripts corresponding to 110408 genes. After removing sequences (668) with significant matches to not Eukaryota taxa 184,166 contigs remained {Consensus Fasta file}.

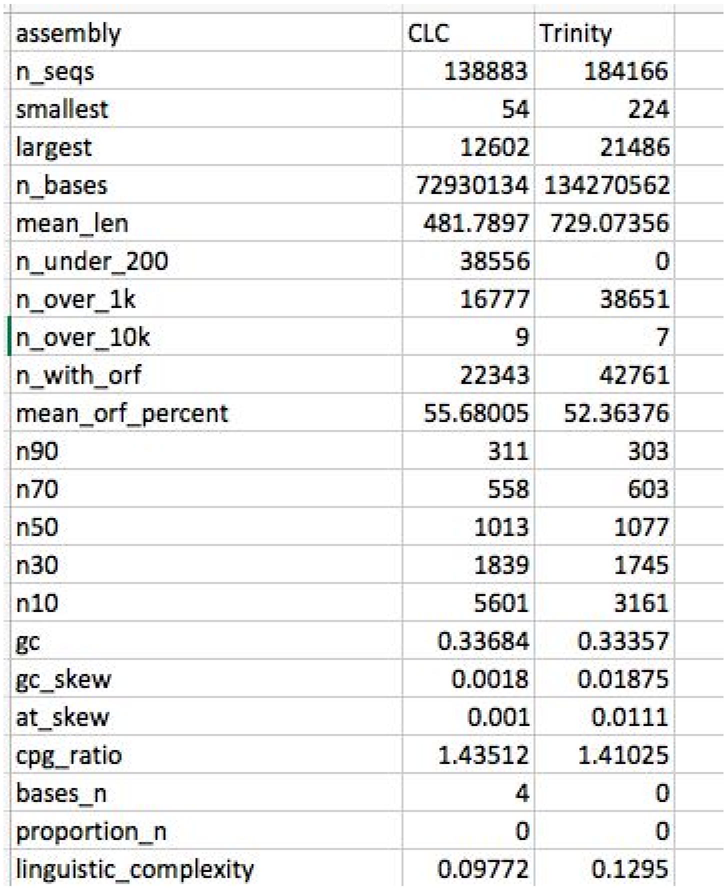
name	explanation	optimum
CRBB hits	the number of reciprocal best hits against the reference using CRB-BLAST. A high score indicates that a large number of real transcripts have been assembled.	As high as possible. The theoretical maximum is the number of contigs (**n seqs**). In practise, the maximum depends on the evolutionary divergence between the assembled species and the reference.
p contigs with CRBB	the proportion of contigs with a CRB-BLAST hit	1

n contigs with CRBB	the number of contigs with a CRB-BLAST hit	n seqs
p references with CRBB	the proportion of references with a CRB-BLAST hit	1
n references with CRBB	the number of references with a CRB-BLAST hit	n seqs
reference coverage	the proportion of reference bases/amino acids covered by a CRB-BLAST hit	As high as possible (see above)
collapse factor	the mean number of reference proteins mapping to each contig. A high score on this metric indicates the assembly contains chimeras or has collapsed gene families.	Dependent on the phylogenomic relationship between the organisms, e.g. whether a genome duplication has taken place.
covX	number of reference proteins with at least X% of their bases covered by a CRB-BLAST hit	All of them

p covX	proportion of reference proteins with at least X% of their bases covered by a CRB-BLAST hit	1

Parameter	Value
CRBB_hits	27302
n_contigs_with_CRBB	27302
p_contigs_with_CRBB	0.14825
rbh_per_reference	1.04649
n_refs_with_CRBB	10912
p_refs_with_CRBB	0.41826
cov25	9526
p_cov25	0.36513
cov50	7595
p_cov50	0.29112
cov75	5512
p_cov75	0.21128
cov85	4478
p_cov85	0.17164
cov95	2897
p_cov95	0.11104
reference_coverage	0.31688

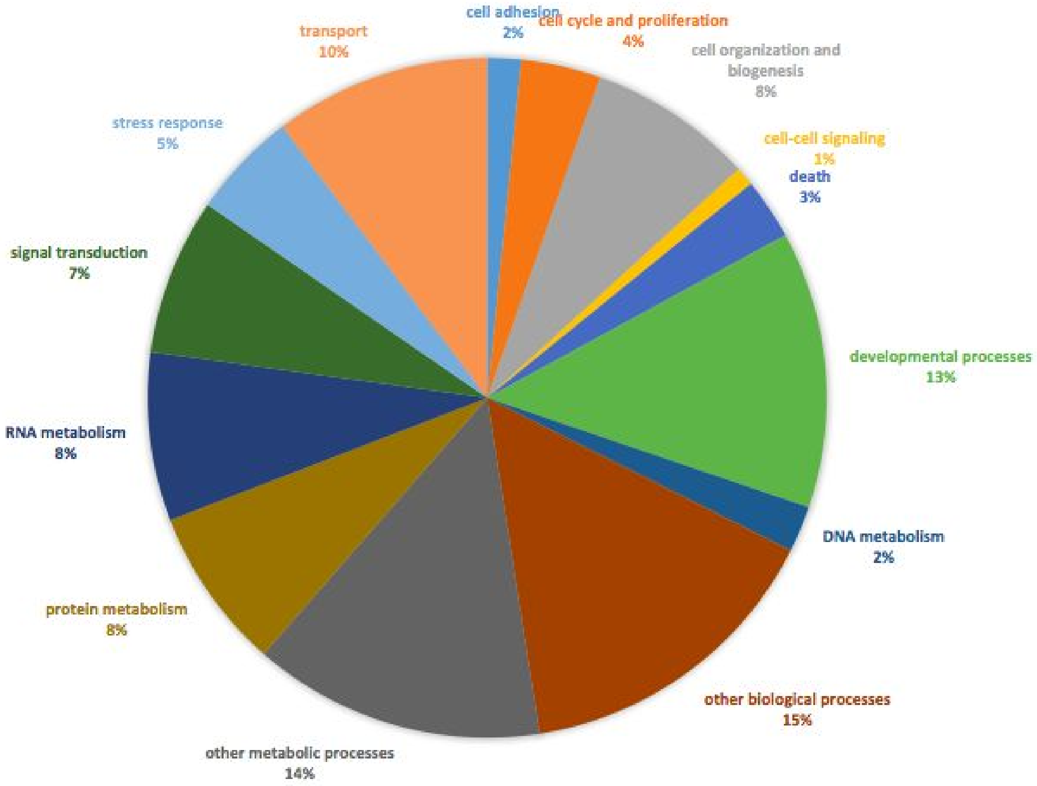
Transcriptome Comparison

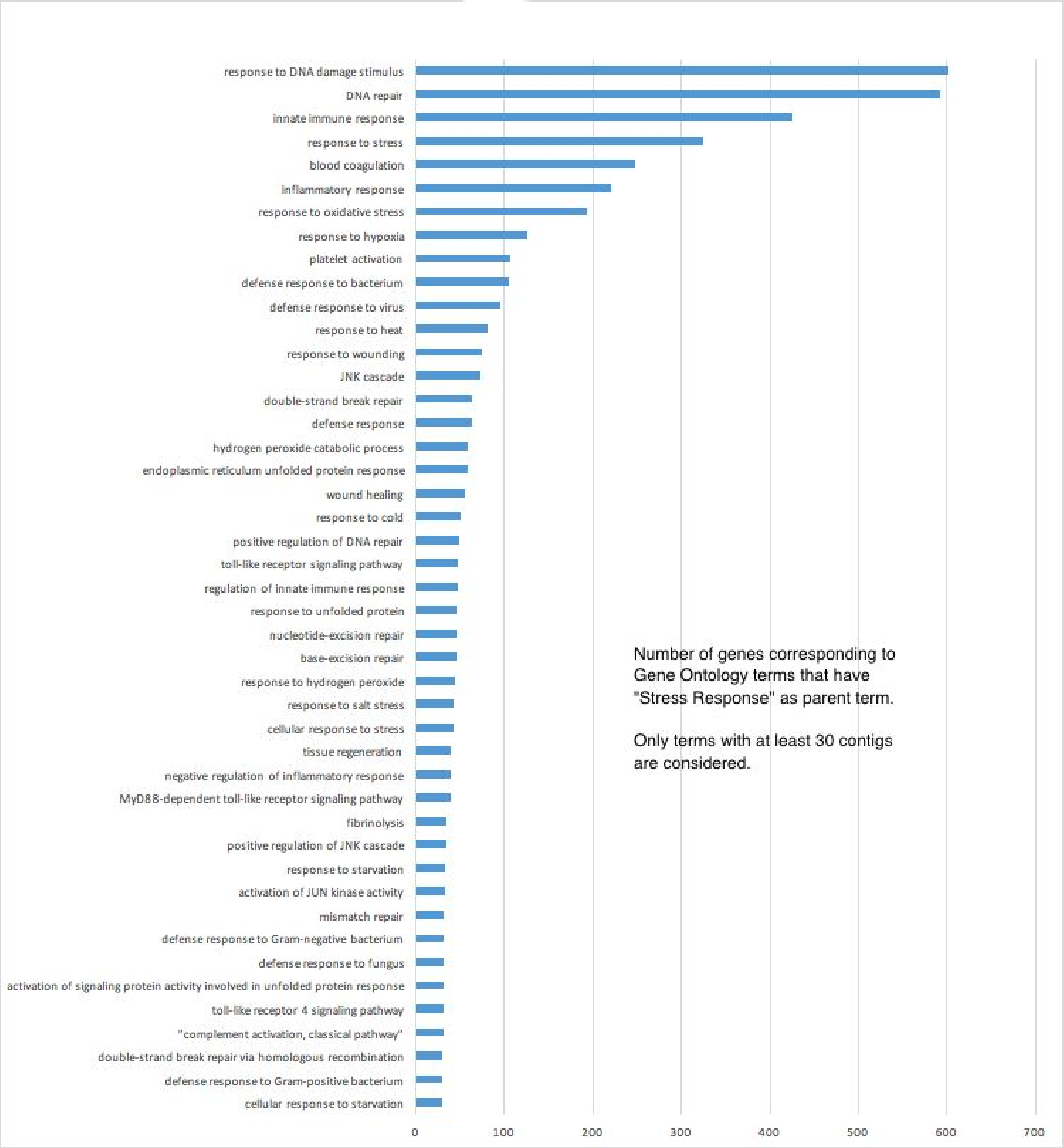


assembly	CLC	Trinity
n_seqs	138883	184166
smallest	54	224
largest	12602	21486
n_bases	72930134	134270562
mean_len	481.7897	729.07356
n_under_200	38556	0
n_over_1k	16777	38651
n_over_10k	9	7
n_with_orf	22343	42761
mean_orf_percent	55.68005	52.36376
n90	311	303
n70	558	603
n50	1013	1077
n30	1839	1745
n10	5601	3161
gc	0.33684	0.33357
gc_skew	0.0018	0.01875
at_skew	0.001	0.0111
cpg_ratio	1.43512	1.41025
bases_n	4	0
proportion_n	0	0
linguistic_complexity	0.09772	0.1295

Transcriptome Annotation (Trinity)

Comparison with Uniprot Swiss-Prot database resulted in 29,445 (non euks removed) contigs with annotations.

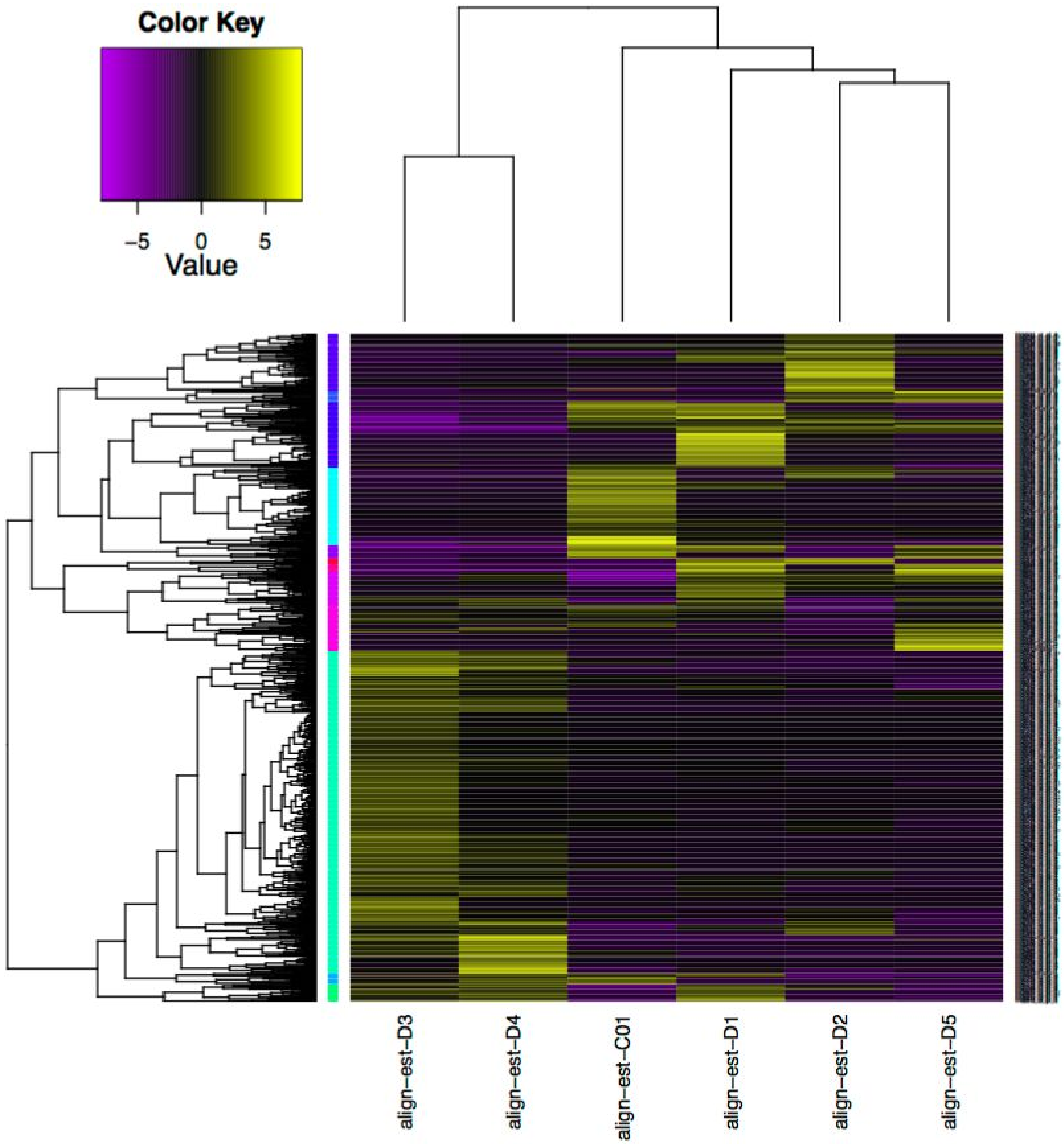

cell adhesion
2%
cell cycle and proliferation
4%
cell organization and biogenesis
8%
cell-cell signaling
1%
death
3%
developmental processes
13%
DNA metabolism
2%
other biological processes
15%
other metabolic processes
14%
protein metabolism
8%
RNA metabolism
8%
signal transduction
7%
stress response
5%
transport
10%

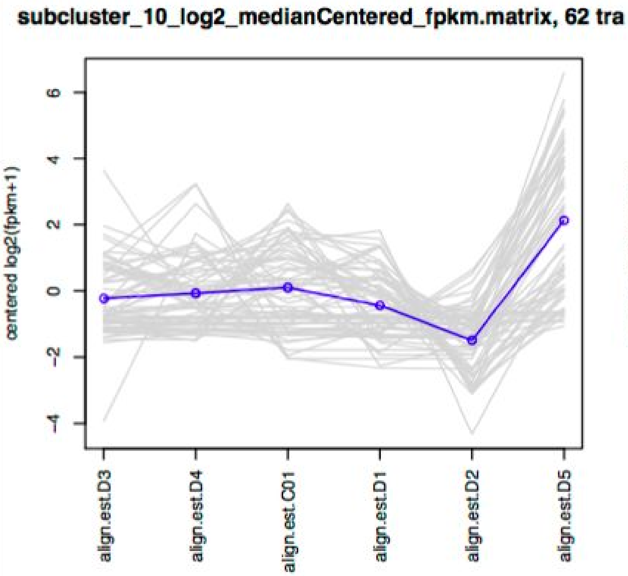


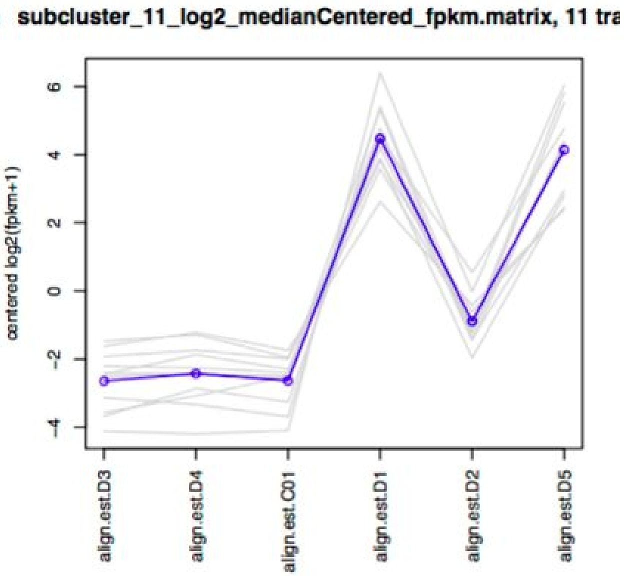
Number of genes corresponding to Gene Ontology terms that have "Stress Response" as parent term.

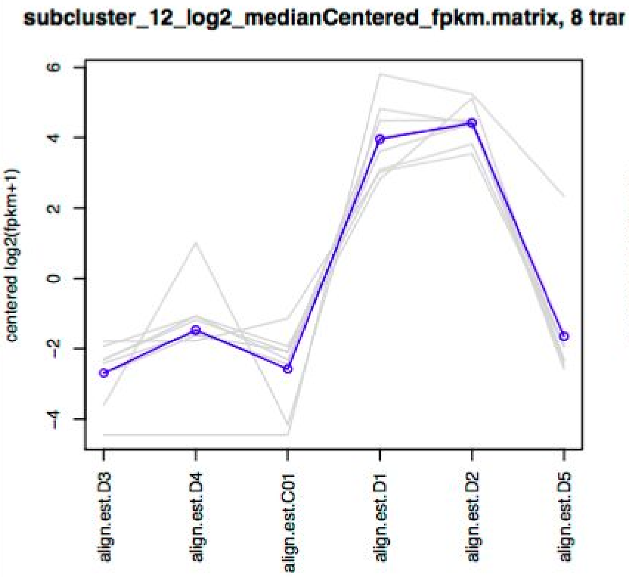
Only terms with at least 30 contigs are considered.

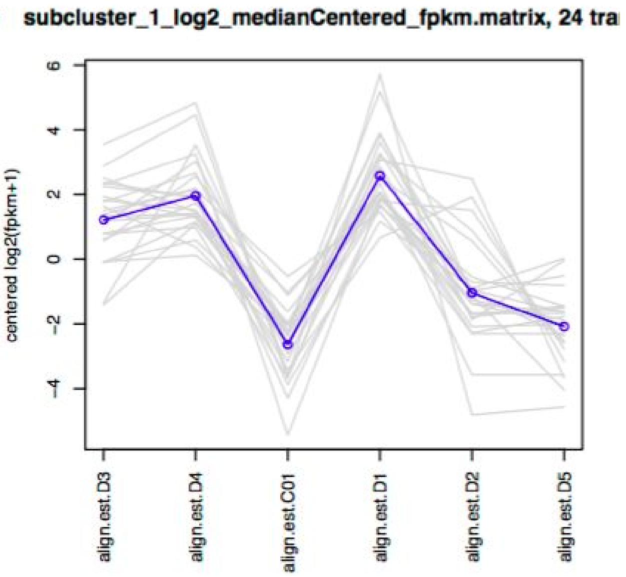
Differential Expression

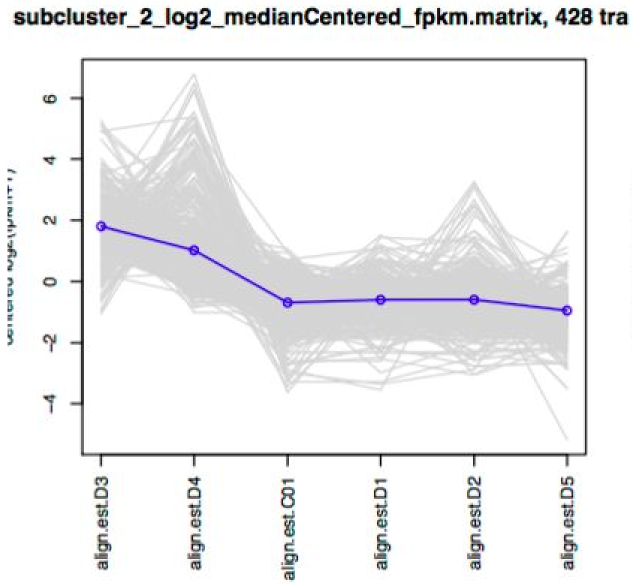

Color Key
−5
0
5
Value
align−est−D3
align−est−D4
align−est−C01
align−est−D1
align−est−D2
align−est−D5

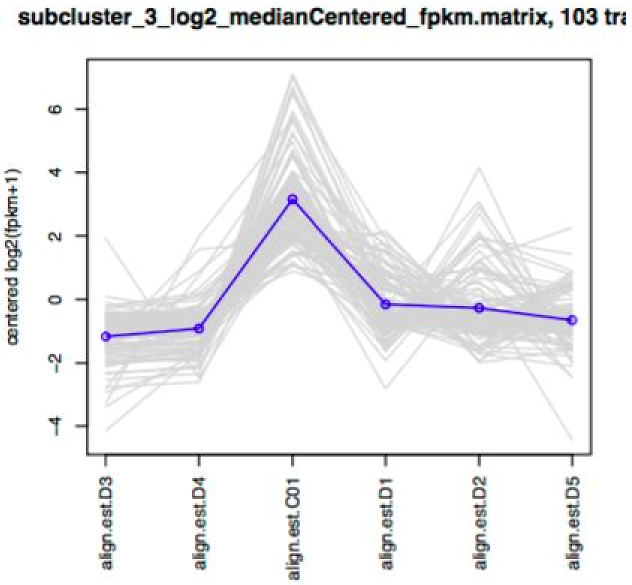

subcluster_10_log2_medianCentered_fpkm.matrix, 62 tra
centered log2(fpkm+1)
6
4
2
0
-2
-4
align.est.D3
align.est.D4
align.est.C01
align.est.D1
align.est.D2
align.est.D5

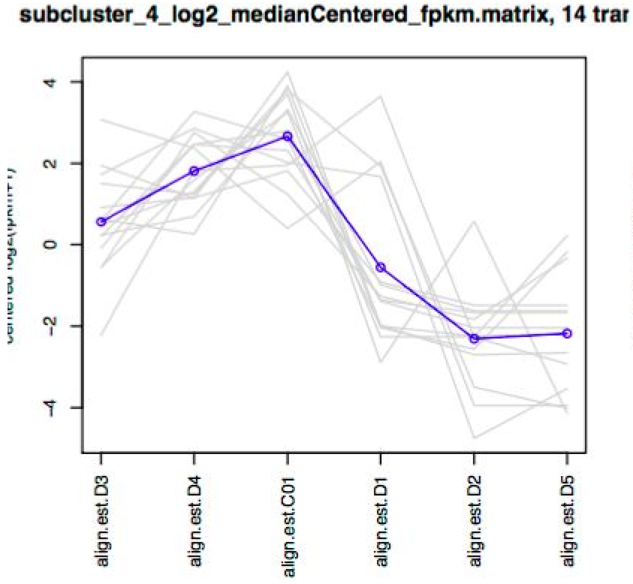

subcluster_11_log2_medianCentered_fpkm.matrix, 11 tra
centered log2(fpkm+1)
6
4
2
0
-2
-4
align.est.D3
align.est.D4
align.est.C01
align.est.D1
align.est.D2
align.est.D5

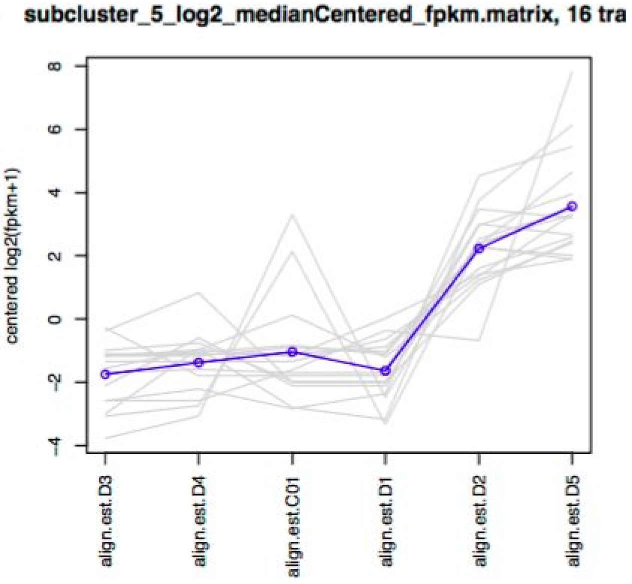

subcluster_12_log2_medianCentered_fpkm.matrix, 8 trar
centered log2(fpkm+1)
6
4
2
0
-2
-4
align.est.D3
align.est.D4
align.est.C01
align.est.D1
align.est.D2
align.est.D5


subcluster_1_log2_medianCentered_fpkm.matrix, 24 tra
centered log2(fpkm+1)
6
4
2
0
-2
-4
align.est.D3
align.est.D4
align.est.C01
align.est.D1
align.est.D2
align.est.D5

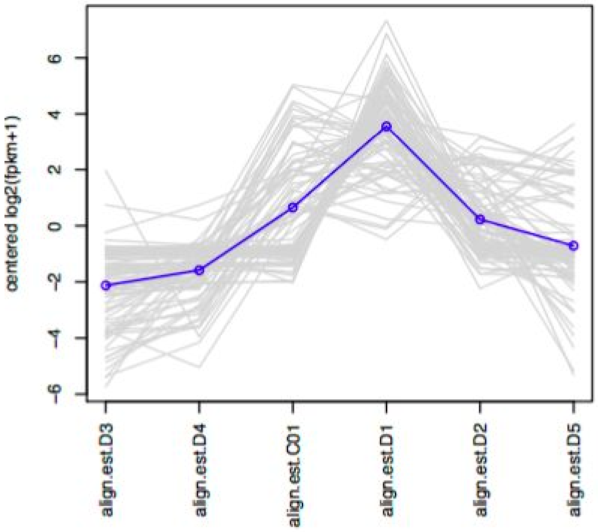

subcluster_2_log2_medianCentered_fpkm.matrix, 428 tra
centered log2(fpkm+1)
6
4
2
0
-2
-4
align.est.D3
align.est.D4
align.est.C01
align.est.D1
align.est.D2
align.est.D5


subcluster_3_log2_medianCentered_fpkm.matrix, 103 tra
centered log2(fpkm+1)
6
4
2
0
-2
-4
align.est.D3
align.est.D4
align.est.C01
align.est.D1
align.est.D2
align.est.D5

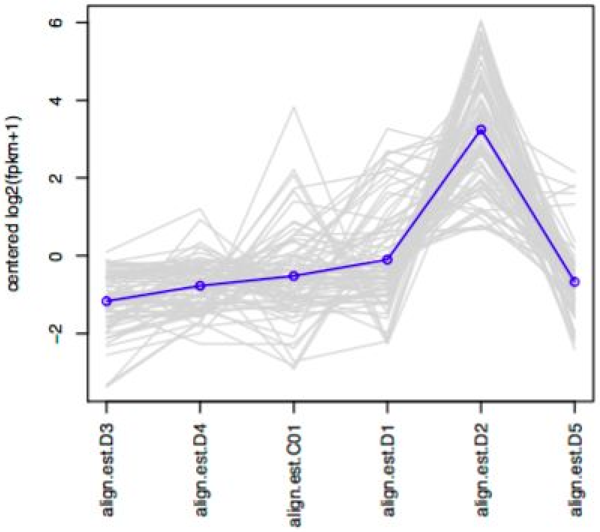

subcluster_4_log2_medianCentered_fpkm.matrix, 14 trar
centered log2(fpkm+1)
4
2
0
-2
-4
align.est.D3
align.est.D4
align.est.C01
align.est.D1
align.est.D2
align.est.D5


subcluster_5_log2_medianCentered_fpkm.matrix, 16 tra
centered log2(fpkm+1)
8
6
4
2
0
-2
-4
align.est.D3
align.est.D4
align.est.C01
align.est.D1
align.est.D2
align.est.D5

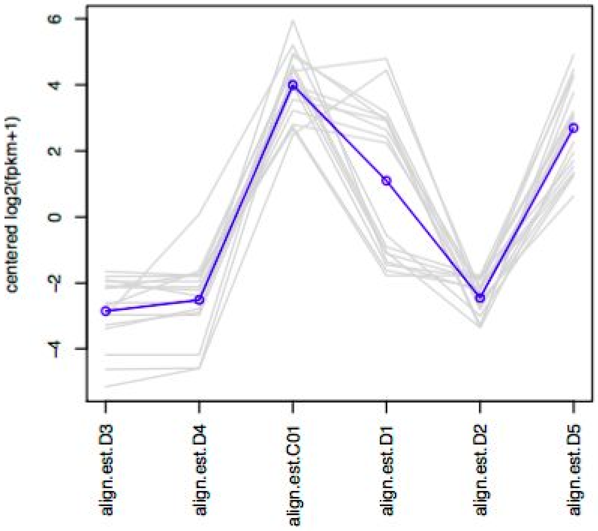
subcluster_6_log2_medianCentered_fpkm.matrix, 86 trar



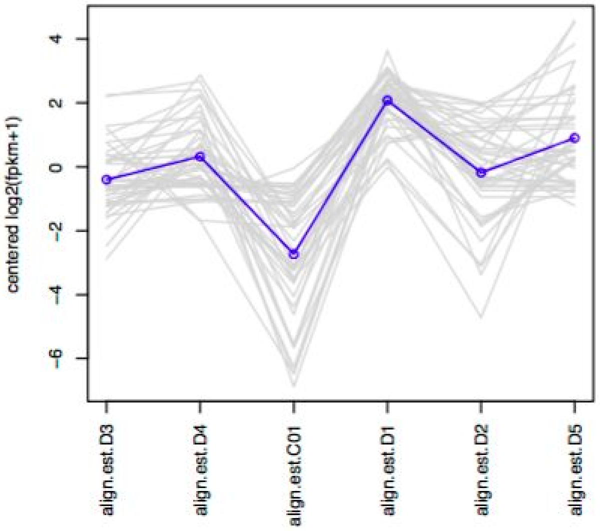
subcluster_7_log2_medianCentered_fpkm.matrix, 76 tr



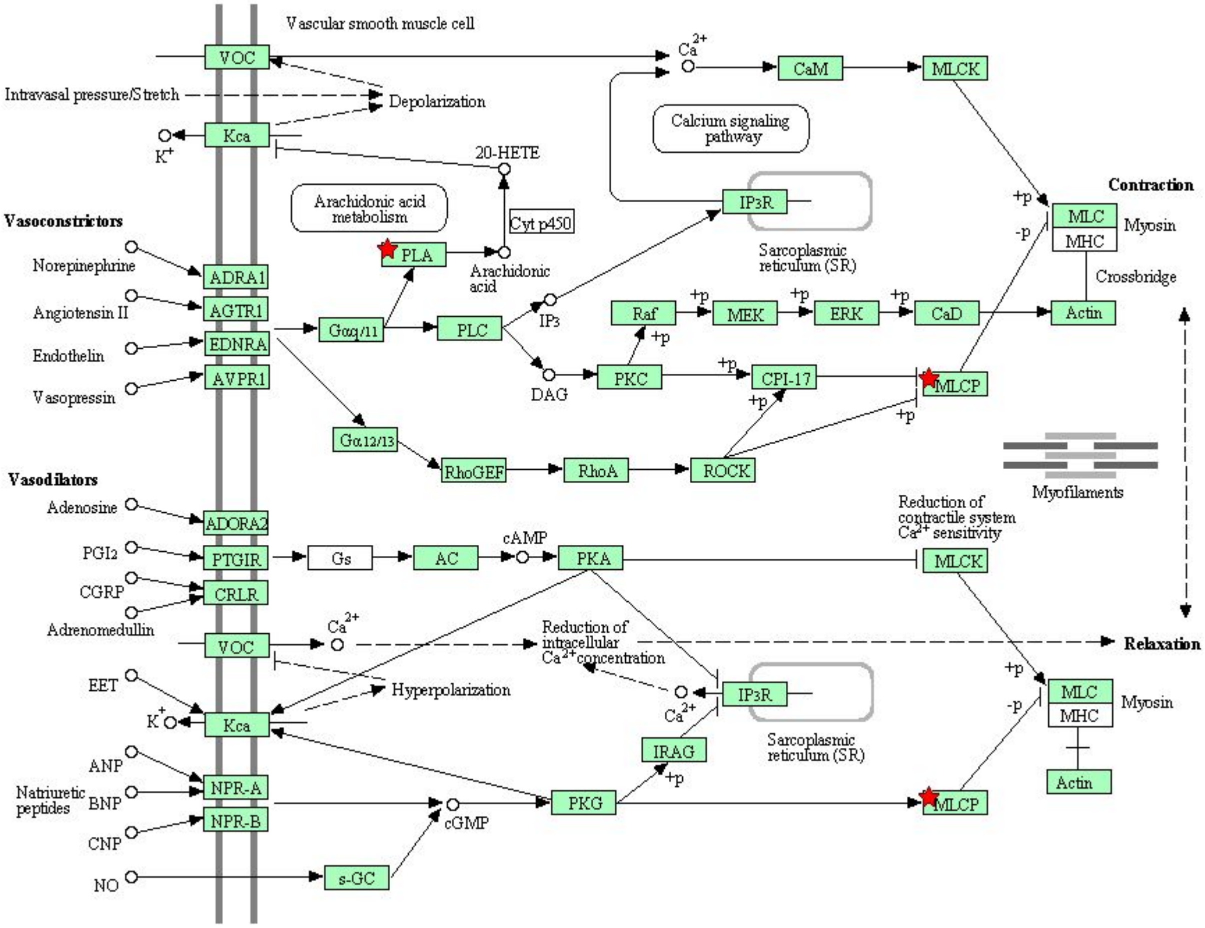
subcluster_8_log2_medianCentered_fpkm.matrix, 17 trar

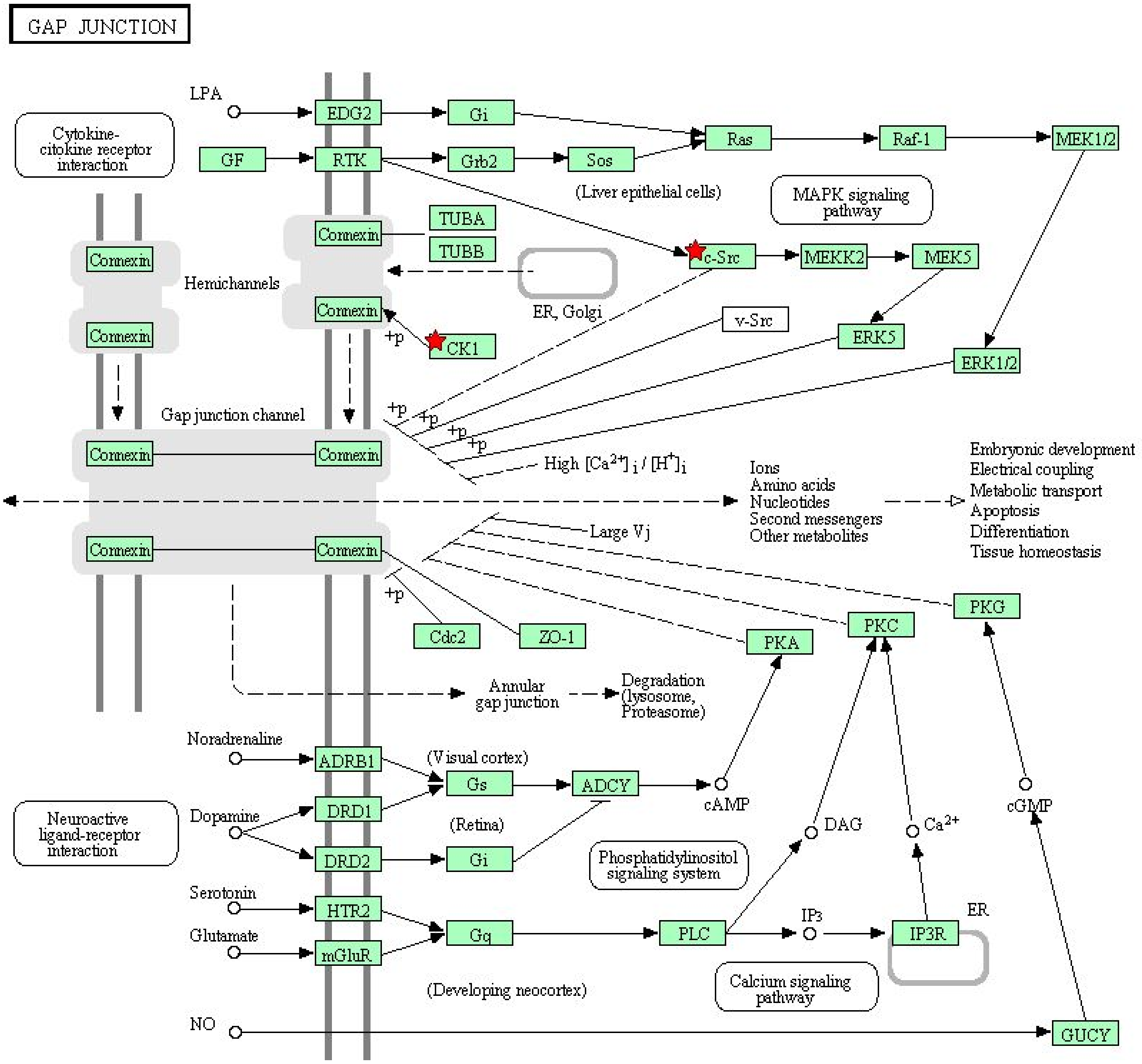


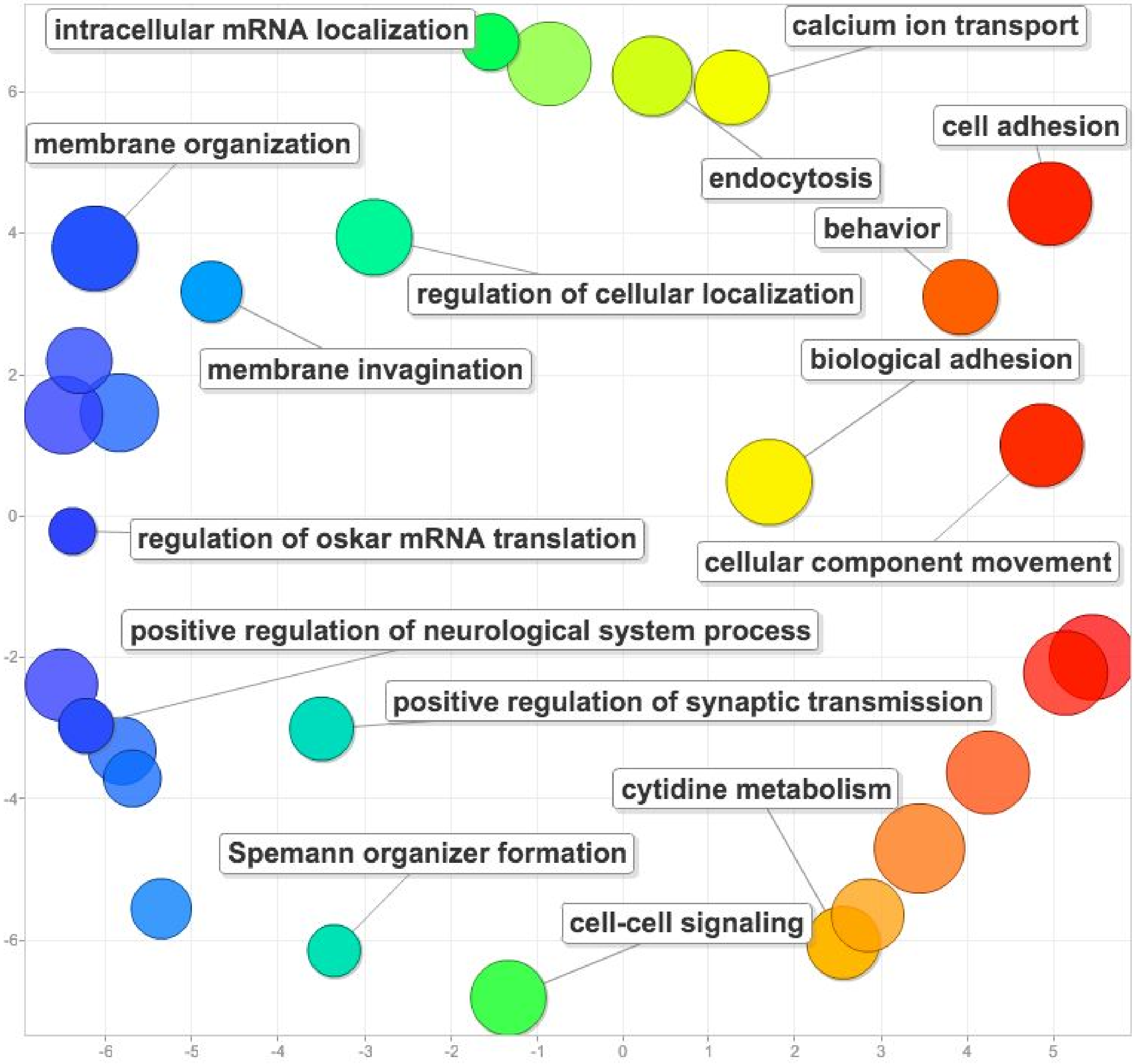
subcluster_9_log2_medianCentered_fpkm.matrix, 43 tr

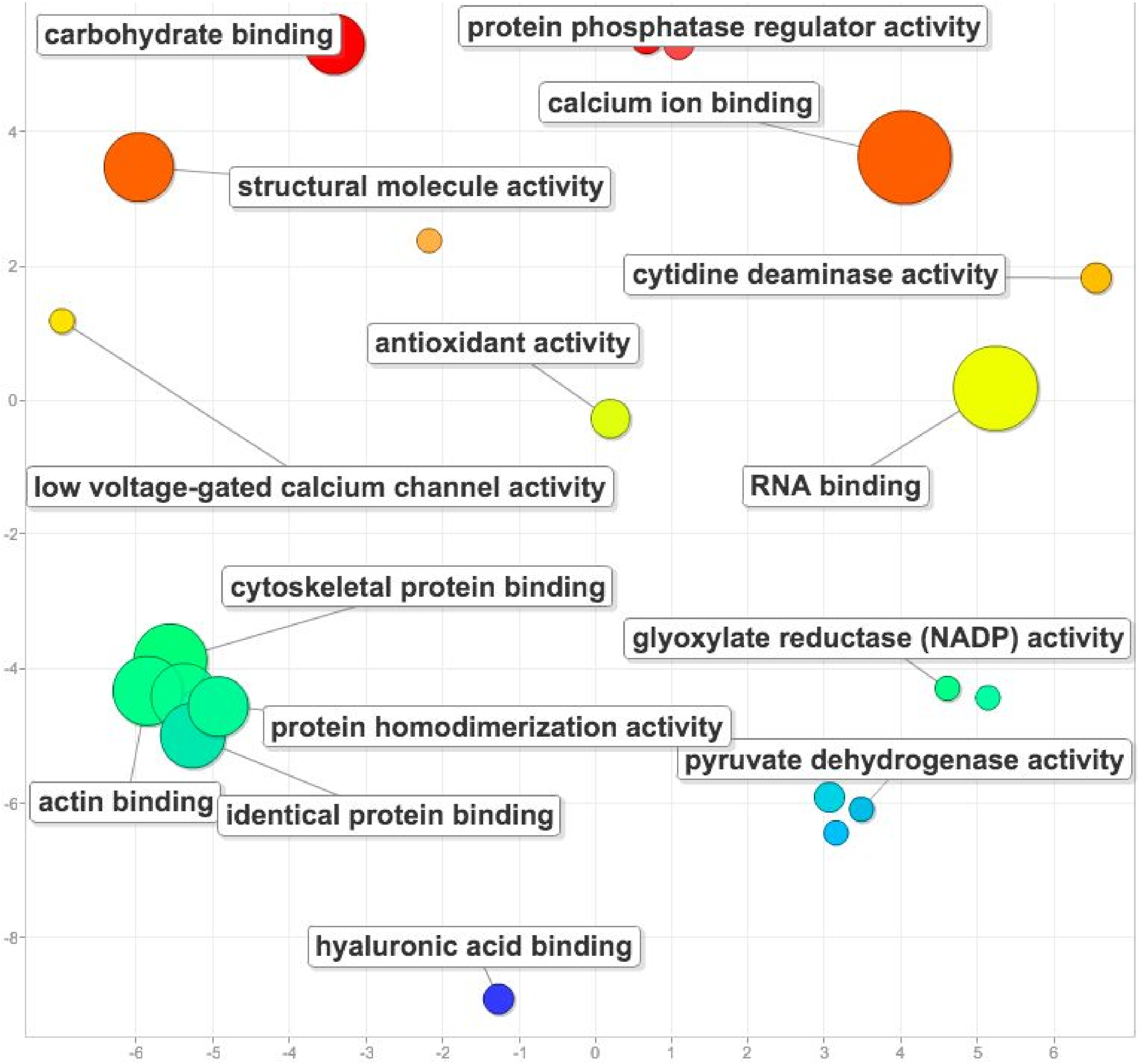


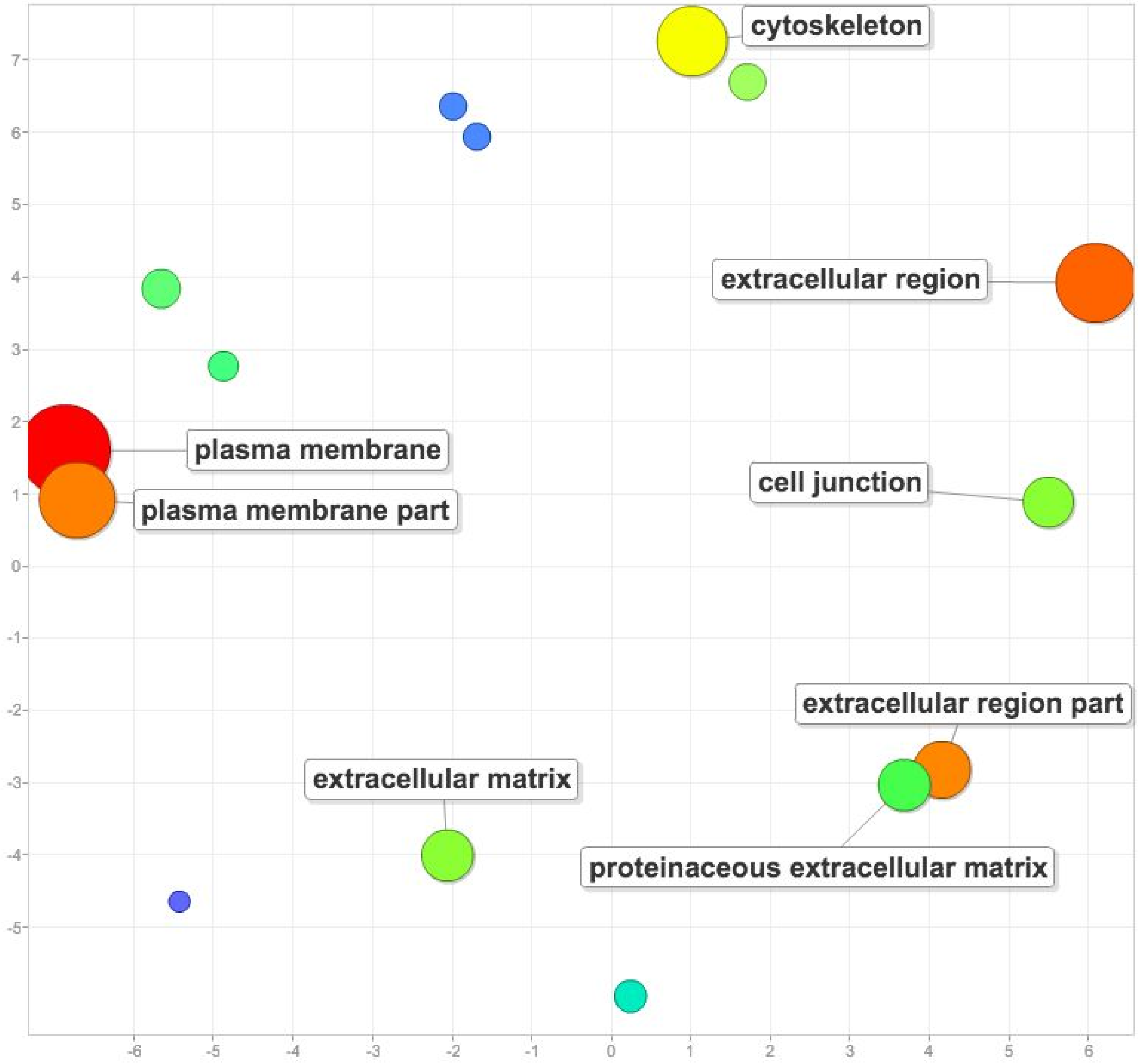
VASCULAR SMOOTH MUSCLE CONTRACTION




GAP JUNCTION
LPA
EDG2
Gi
Ras
Raf-1
MEK1/2
Cytokine-citokine receptor interaction
GF
RTK
Grb2
Sos
(Liver epithelial cells)
MAPK signaling pathway
TUBA
TUBB
Connexin
Hemichannels
c-Src
MEKK2
MEK5
ER, Golgi
v-Src
ERK5
+p
CK1
ERK1/2
Gap junction channel
+p
+p
+p
+p
High [Ca2+]i / [H+]i
Ions
Amino acids
Nucleotides
Second messengers
Other metabolites
Embryonic development
Electrical coupling
Metabolic transport
Apoptosis
Differentiation
Tissue homeostasis
Large Vj
+p
Cdc2
ZO-1
PKA
PKC
PKG
Annular gap junction
Degradation (lysosome, Proteasome)
Noradrenaline
ADRB1
(Visual cortex)
Gs
ADCY
cAMP
cGMP
Neuroactive ligand-receptor interaction
Dopamine
DRD1
DRD2
(Retina)
Gi
DAG
Ca2+
Phosphatidylinositol signaling system
Serotonin
HTR2
Glutamate
mGluR
Gq
PLC
IP3
IP3R
ER
(Developing neocortex)
Calcium signaling pathway
NO
GUCY


intracellular mRNA localization
calcium ion transport
membrane organization
cell adhesion
endocytosis
behavior
regulation of cellular localization
membrane invagination
biological adhesion
regulation of oskar mRNA translation
cellular component movement
positive regulation of neurological system process
positive regulation of synaptic transmission
cytidine metabolism
Spemann organizer formation
cell-cell signaling
6
4
2
0
-2
-4
-6
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5


carbohydrate binding
protein phosphatase regulator activity
calcium ion binding
structural molecule activity
cytidine deaminase activity
antioxidant activity
RNA binding
low voltage-gated calcium channel activity
cytoskeletal protein binding
glyoxylate reductase (NADP) activity
protein homodimerization activity
pyruvate dehydrogenase activity
actin binding
identical protein binding
hyaluronic acid binding
4
2
0
-2
-4
-6
-8
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6



Long non-coding identification (CLC + online tools)

Table I. Long non-coding sequences after data filtering steps.

Item	Number of obtained sequences	Number of discarded sequences
De novo assembly	138833	0
Coverage (average coverage of contigs > 50)	38609	100224

ORF identification (sequence with ORF > 200 were discarded)	23492	15117
Coding potential (CPAT)	22308	1148
Contig length (> 250 bp)	16012	6296
blastX against mollusca proteins	12714	3298
Conserved Domains Search	12346	368
blastn against nr genbank database	8505	3841

Discussion