

Understanding the biology of morpholino in zebrafish through integrated gene expression analysis

14/07/2015

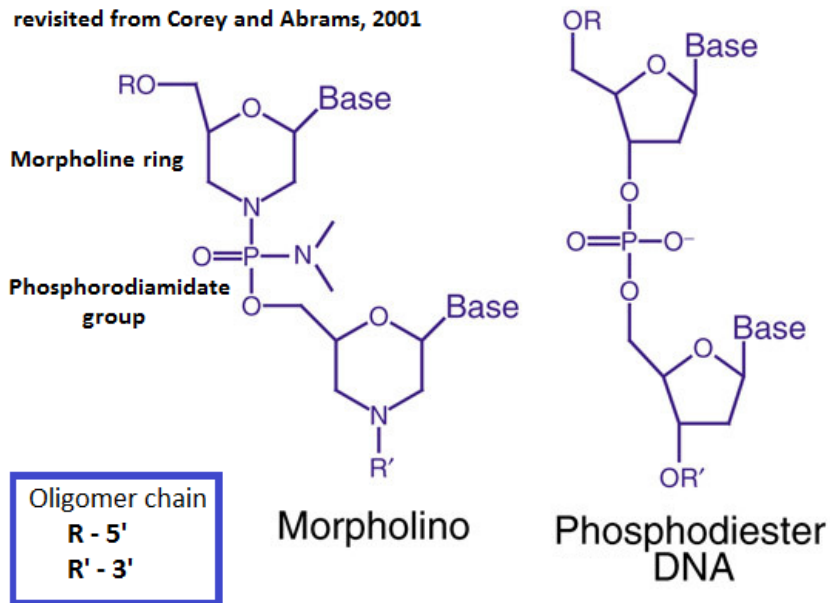
Kristina Gagalova

MSc Bioinformatics and Systems Biology



MORPHOLINO ANTISENSE

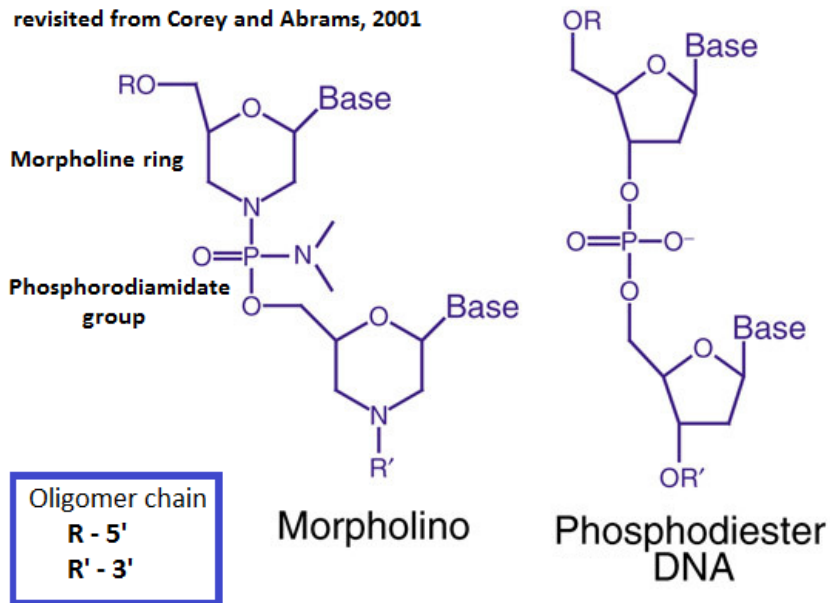
revisited from Corey and Abrams, 2001



The morpholino is an antisense molecule used for gene expression knockdown. It is specifically designed to bind to a target sequence

MORPHOLINO ANTISENSE

revisited from Corey and Abrams, 2001



The morpholino is an antisense molecule used for gene expression knockdown. It is specifically designed to bind to a target sequence

- Non-ionic backbone
- Excellent water solubility
- Resistance to enzymatic degradation

Janson C. and During M. (2006) Morpholinos and Related Antisense Biomolecules, Springer US, Springer-Verlag US
J. Summerton - Morpholinos and PNAs compared

RECENT DEVELOPMENT IN MORPHOLINO ANTISENSE

**Genetic compensation induced by deleterious mutations but
not gene knockdowns.**

**Andrea Rossi^{1,*}, Zacharias Kontarakis^{1,*}, Claudia Gerri¹, Hendrik
Nolte^{1,2}, Soraya Hölper¹, Marcus Krüger^{1,2} and Didier Y.R. Stainier¹**

in press

RECENT DEVELOPMENT IN MORPHOLINO ANTISENSE

Genetic compensation induced by deleterious mutations but not gene knockdowns.

Andrea Rossi^{1,*}, Zacharias Kontarakis^{1,*}, Claudia Gerri¹, Hendrik Nolte^{1,2}, Soraya Hölper¹, Marcus Krüger^{1,2} and Didier Y.R. Stainier¹

in press

Discrepancies between EGFL7 mutant and morphant. The mutant does not show evident phenotype

DISCUSSION

activation of a compensatory network against deleterious mutations

RECENT DEVELOPMENT IN MORPHOLINO ANTISENSE

Genetic compensation induced by deleterious mutations but not gene knockdowns.

Andrea Rossi^{1,*}, Zacharias Kontarakis^{1,*}, Claudia Gerri¹, Hendrik Nolte^{1,2}, Soraya Hölper¹, Marcus Krüger^{1,2} and Didier Y.R. Stainier¹

in press

Discrepancies between EGFL7 mutant and morphant. The mutant does not show evident phenotype

DISCUSSION

activation of a compensatory network against deleterious mutations

Reverse Genetic Screening Reveals Poor Correlation between Morpholino-Induced and Mutant Phenotypes in Zebrafish

Fatma O. Kok,^{1,7} Masahiro Shin,^{1,7} Chih-Wen Ni,^{1,7,8} Ankit Gupta,¹ Ann S. Grosse,¹ Andreas van Impel,² Bettina C. Kirchmaier,^{2,9} Josi Peterson-Maduro,² George Kourkoulis,¹ Ira Male,¹ Dana F. DeSantis,¹ Sarah Sheppard-Tindell,¹ Lwaki Ebarasi,^{3,4} Christer Betsholtz,^{3,4} Stefan Schulte-Merker,^{2,5} Scot A. Wolfe,^{1,6} and Nathan D. Lawson^{1,*}

RECENT DEVELOPMENT IN MORPHOLINO ANTISENSE

Genetic compensation induced by deleterious mutations but not gene knockdowns.

Andrea Rossi^{1,*}, Zacharias Kontarakis^{1,*}, Claudia Gerri¹, Hendrik Nolte^{1,2}, Soraya Hölper¹, Marcus Krüger^{1,2} and Didier Y.R. Stainier¹

in press

Discrepancies between EGFL7 mutant and morphant. The mutant does not show evident phenotype

DISCUSSION

activation of a compensatory network against deleterious mutations

Reverse Genetic Screening Reveals Poor Correlation between Morpholino-Induced and Mutant Phenotypes in Zebrafish

Fatma O. Kok,^{1,7} Masahiro Shin,^{1,7} Chih-Wen Ni,^{1,7,8} Ankit Gupta,¹ Ann S. Grosse,¹ Andreas van Impel,² Bettina C. Kirchmaier,^{2,9} Josi Peterson-Maduro,² George Kourkoulis,¹ Ira Male,¹ Dana F. DeSantis,¹ Sarah Sheppard-Tindell,¹ Lwaki Ebarasi,^{3,4} Christer Betsholtz,^{3,4} Stefan Schulte-Merker,^{2,5} Scot A. Wolfe,^{1,6} and Nathan D. Lawson^{1,*}

Discrepancies between mutant and morphant phenotypes for 19/21 genes

DISCUSSION

morphant phenotypes may be due on off-target effects

MORPHOLINO OFF-TARGET EFFECTS

“Not dependent of Watson/Crick pairing to RNA but instead are due to interactions between the knockdown oligo and extracellular, cell-surface or intracellular structures via a mechanism other than Watson/Crick base-pairing”

James E. Summerton, Morpholino, siRNA, and S-DNA compared: impact of structure and mechanism of action on off-target effects and sequence specificity.

PLoS GENETICS

p53 Activation by Knockdown Technologies

**Mara E. Robu¹, Jon D. Larson¹, Aidas Nasevicius^{1*}, Soraya Beiraghi¹, Charles Brenner², Steven A. Farber³,
Stephen C. Ekker^{1*}**

¹ University of Minnesota, Minneapolis, Minnesota, United States of America, ² Dartmouth Medical School, Lebanon, New Hampshire, United States of America, ³ Carnegie Institute of Washington, Baltimore, Maryland, United States of America

MOTIVATION FOR MORPHOLINO STUDY

- Large amount of gene expression data available from morpholino studies and missing systematic data analysis

MOTIVATION FOR MORPHOLINO STUDY

- Large amount of gene expression data available from morpholino studies and missing systematic data analysis
 - Describe other possible off-target effects

MOTIVATION FOR MORPHOLINO STUDY

- Large amount of gene expression data available from morpholino studies and missing systematic data analysis
 - Describe other possible off-target effects
- New interest in combined knockdown and mutation studies

PROJECT GOALS

Gene expression analysis of embryos treated with morpholino

PROJECT GOALS

Gene expression analysis of embryos treated with morpholino

Unsupervised learning

Find hidden structures or
patterns in unlabeled data

*“Plot” the microarray samples
to find similarities*

PROJECT GOALS

Gene expression analysis of embryos treated with morpholino

Unsupervised learning

Find hidden structures or patterns in unlabeled data

“Plot” the microarray samples to find similarities

Supervised learning

Inferring differences between labeled data

Define differences between the control and morphant samples

Data curation and annotation*

1. Data collection from GEO

The screenshot shows the NCBI GEO Accession Display page for GSE51541. The page header includes the NCBI logo and the GEO logo (Gene Expression Omnibus). Navigation links include HOME, SEARCH, SITE MAP, GEO Publications, FAQ, MIAME, and Email GEO. The breadcrumb trail is NCBI > GEO > Accession Display. A search bar shows the GEO accession GSE51541. The main content area displays the series GSE51541 with a link to Query DataSets for GSE51541. The series details include:

- Status: Public on Mar 01, 2014
- Title: Autophagy is essential for cardiac morphogenesis during vertebrate development
- Organism: [Danio rerio](#)
- Experiment type: Expression profiling by array
- Summary: This study examined the effects of genetic knockdown of autophagy genes on vertebrate cardiac development. We performed microarray studies comparing the hearts of control zebrafish embryos to the hearts of embryos with decreased expression of the autophagy genes atg5, becn1 or atg7. The results provide insight into the role of autophagy in developmental morphogenesis.
- Overall design: Hearts were purified from 3 day-old zebrafish embryos injected with control or autophagy gene-specific morpholino oligonucleotides. RNA was prepared from all samples and hybridized to zebrafish-specific Affymetrix arrays.
- Contributor(s): [Lee E, Ng A, Xavier R, Levine B, Amatruda JF](#)
- Citation(s): Lee E, Koo Y, Ng A, Wei Y et al. Autophagy is essential for cardiac morphogenesis during vertebrate development. *Autophagy* 2014 Apr;10(4):572-87. PMID: [24441423](#)
- Submission date: Oct 22, 2013
- Last update date: Dec 03, 2014
- Contact name: James F Amatruda
- E-mail: james.amatruda@utsouthwestern.edu
- Phone: 214-648-3896
- Organization name: University of Texas Southwestern Medical Center

<http://www.ncbi.nlm.nih.gov/geo/>

* Attach relevant information to the one already available

Data curation and annotation*

1. Data collection from GEO
2. Retrieve the annexed scientific literature, **DISCARD DATASETS WITH MISSING REFERENCE**

NCBI GEO Accession Display

Series GSE51541

Status: Public on Mar 01, 2014

Title: Autophagy is essential for cardiac morphogenesis during vertebrate development

Organism: *Danio rerio*

Experiment type: Expression profiling by array

Summary: This study examined the effects of genetic knockdown of autophagy genes on vertebrate cardiac development. We performed microarray studies comparing the hearts of control zebrafish embryos to the hearts of embryos with decreased expression of the autophagy genes atg5, becn1 or atg7. The results provide insight into the role of autophagy in developmental morphogenesis.

Overall design: Hearts were purified from 3 day-old zebrafish embryos injected with control or autophagy gene-specific morpholino oligonucleotides. RNA was prepared from all samples and hybridized to zebrafish-specific Affymetrix arrays.

Contributor(s): Lee E, Ng A, Xavier R, Levine B, Amatruda JF

Citation(s): Lee E, Koo Y, Ng A, Wei Y et al. Autophagy is essential for cardiac morphogenesis during vertebrate development. *Autophagy* 2014 Apr;10(4):572-87. PMID: 24441423

Submission date: Oct 22, 2013

Last update date: Dec 03, 2014

Contact name: James F Amatruda

E-mail: james.amatruda@utsouthwestern.edu

Phone: 214-648-3896

Organization name: University of Texas Southwestern Medical Center

<http://www.ncbi.nlm.nih.gov/geo/>

* Attach relevant information to the one already available

Data curation and annotation*

1. Data collection from GEO
2. Retrieve the annexed scientific literature, **DISCARD DATASETS WITH MISSING REFERENCE**
3. Carefully go through the article, read and extract useful information
4. Process information - transform variables, retrieve missing information

The screenshot shows the NCBI GEO Accession Display page for GSE51541. The page includes the NCBI logo, GEO logo, and navigation links. The main content area displays the accession number GSE51541 and a search bar. Below the search bar, the series title 'Autophagy is essential for cardiac morphogenesis during vertebrate development' is shown. The page also includes a summary, overall design, contributor(s), citation(s), submission date, last update date, contact name, E-mail, phone, and organization name.

Status	Public on Mar 01, 2014
Title	Autophagy is essential for cardiac morphogenesis during vertebrate development
Organism	Danio rerio
Experiment type	Expression profiling by array
Summary	This study examined the effects of genetic knockdown of autophagy genes on vertebrate cardiac development. We performed microarray studies comparing the hearts of control zebrafish embryos to the hearts of embryos with decreased expression of the autophagy genes atg5, becn1 or atg7. The results provide insight into the role of autophagy in developmental morphogenesis.
Overall design	Hearts were purified from 3 day-old zebrafish embryos injected with control or autophagy gene-specific morpholino oligonucleotides. RNA was prepared from all samples and hybridized to zebrafish-specific Affymetrix arrays.
Contributor(s)	Lee E, Ng A, Xavier R, Levine B, Amatruda JF
Citation(s)	Lee E, Koo Y, Ng A, Wei Y et al. Autophagy is essential for cardiac morphogenesis during vertebrate development. <i>Autophagy</i> 2014 Apr;10(4):572-87. PMID: 24441423
Submission date	Oct 22, 2013
Last update date	Dec 03, 2014
Contact name	James F Amatruda
E-mail	james.amatruda@utsouthwestern.edu
Phone	214-648-3896
Organization name	University of Texas Southwestern Medical Center

<http://www.ncbi.nlm.nih.gov/geo/>

* Attach relevant information to the one already available

Data curation and annotation*

1. Data collection from GEO
2. Retrieve the annexed scientific literature, **DISCARD DATASETS WITH MISSING REFERENCE**
3. Carefully go through the article, read and extract useful information
4. Process information - transform variables, retrieve missing information
5. Check compatibility between different microarray platforms

The screenshot shows the NCBI GEO Accession Display page for GSE51541. The page includes the NCBI logo, GEO logo, and navigation links. The main content area displays the accession number GSE51541 and a search bar. Below the search bar, the series title 'Autophagy is essential for cardiac morphogenesis during vertebrate development' is shown. The page also includes a summary, overall design, contributor(s), citation(s), submission date, last update date, contact name, E-mail, phone, and organization name.

Status	Public on Mar 01, 2014
Title	Autophagy is essential for cardiac morphogenesis during vertebrate development
Organism	Danio rerio
Experiment type	Expression profiling by array
Summary	This study examined the effects of genetic knockdown of autophagy genes on vertebrate cardiac development. We performed microarray studies comparing the hearts of control zebrafish embryos to the hearts of embryos with decreased expression of the autophagy genes atg5, becn1 or atg7. The results provide insight into the role of autophagy in developmental morphogenesis.
Overall design	Hearts were purified from 3 day-old zebrafish embryos injected with control or autophagy gene-specific morpholino oligonucleotides. RNA was prepared from all samples and hybridized to zebrafish-specific Affymetrix arrays.
Contributor(s)	Lee E, Ng A, Xavier R, Levine B, Amatruda JF
Citation(s)	Lee E, Koo Y, Ng A, Wei Y et al. Autophagy is essential for cardiac morphogenesis during vertebrate development. <i>Autophagy</i> 2014 Apr;10(4):572-87. PMID: 24441423
Submission date	Oct 22, 2013
Last update date	Dec 03, 2014
Contact name	James F Amatruda
E-mail	james.amatruda@utsouthwestern.edu
Phone	214-648-3896
Organization name	University of Texas Southwestern Medical Center

<http://www.ncbi.nlm.nih.gov/geo/>

DATASET FEATURES - unsupervised

Study	MO phenotype	MO target
GSE12012	disrupted vessels integrity	miR-126
GSE13157	reduced neuromasts	esr2a
GSE13158	reduced neuromasts	esr2a
GSE16740	missing heart contraction	tnnt2
GSE18830	severe developmental abnormalities	sox2/3/19a/19b
GSE21539	craniofacial defects, reduced pectoral fin growth	esco2
GSE27569	Neural crest (NC) defects	Ovo1
GSE32914	developmental arrest	Sox31
GSE46844	disrupted kidney morphogenesis	mir-34B
GSE51541	cardiac defects	atg5/becl1/atg7
GSE8800	disrupted primitive hematopoiesis, brain and craniofacial defects	C1q-like

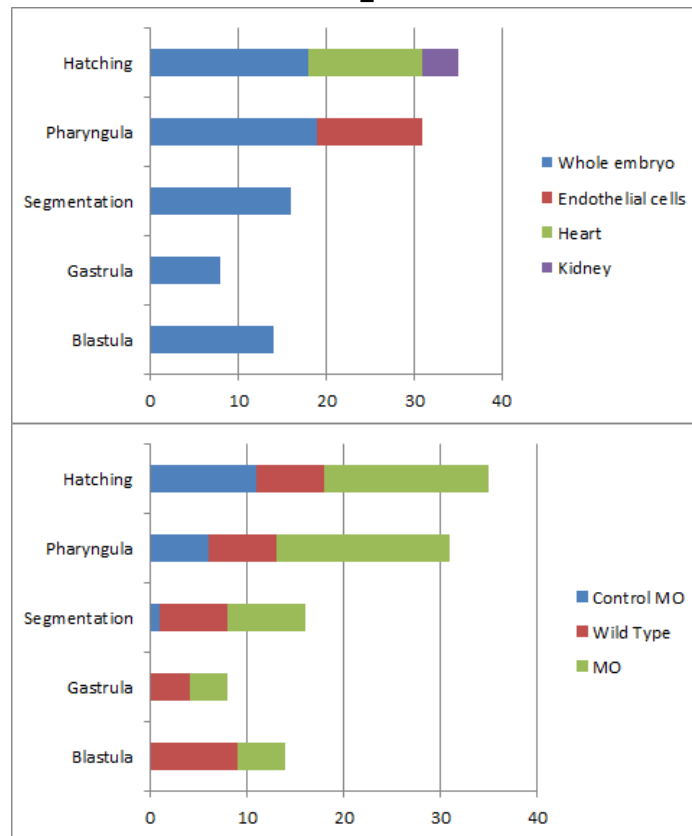
DATASET FEATURES - unsupervised

Study	MO phenotype	MO target
GSE12012	disrupted vessels integrity	miR-126
GSE13157	reduced neuromasts	esr2a
GSE13158	reduced neuromasts	esr2a
GSE16740	missing heart contraction	tnnt2
GSE18830	severe developmental abnormalities	sox2/3/19a/19b
GSE21539	craniofacial defects, reduced pectoral fin growth	esco2
GSE27569	Neural crest (NC) defects	Ovo1
GSE32914	developmental arrest	Sox31
GSE46844	disrupted kidney morphogenesis	mir-34B
GSE51541	cardiac defects	atg5/becl1/atg7
GSE8800	disrupted primitive hematopoiesis, brain and craniofacial defects	C1q-like

Control MO* (18) + Wild Type (34) + MO (52)
Total # of Microarrays: 104

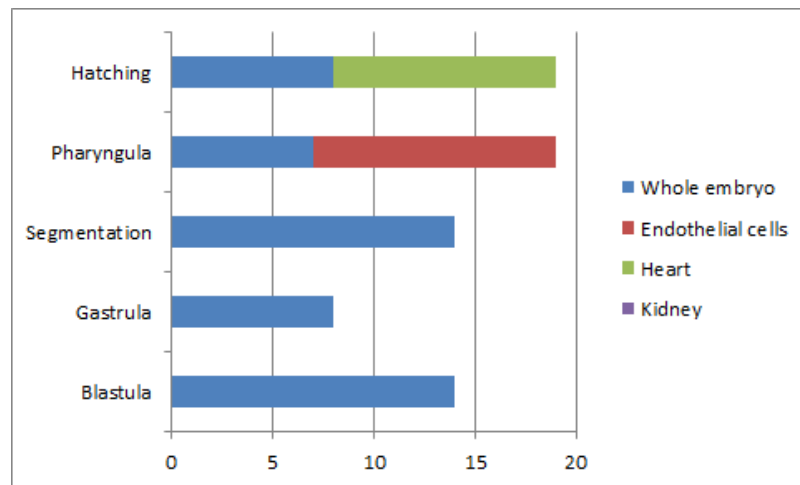


*Standard control



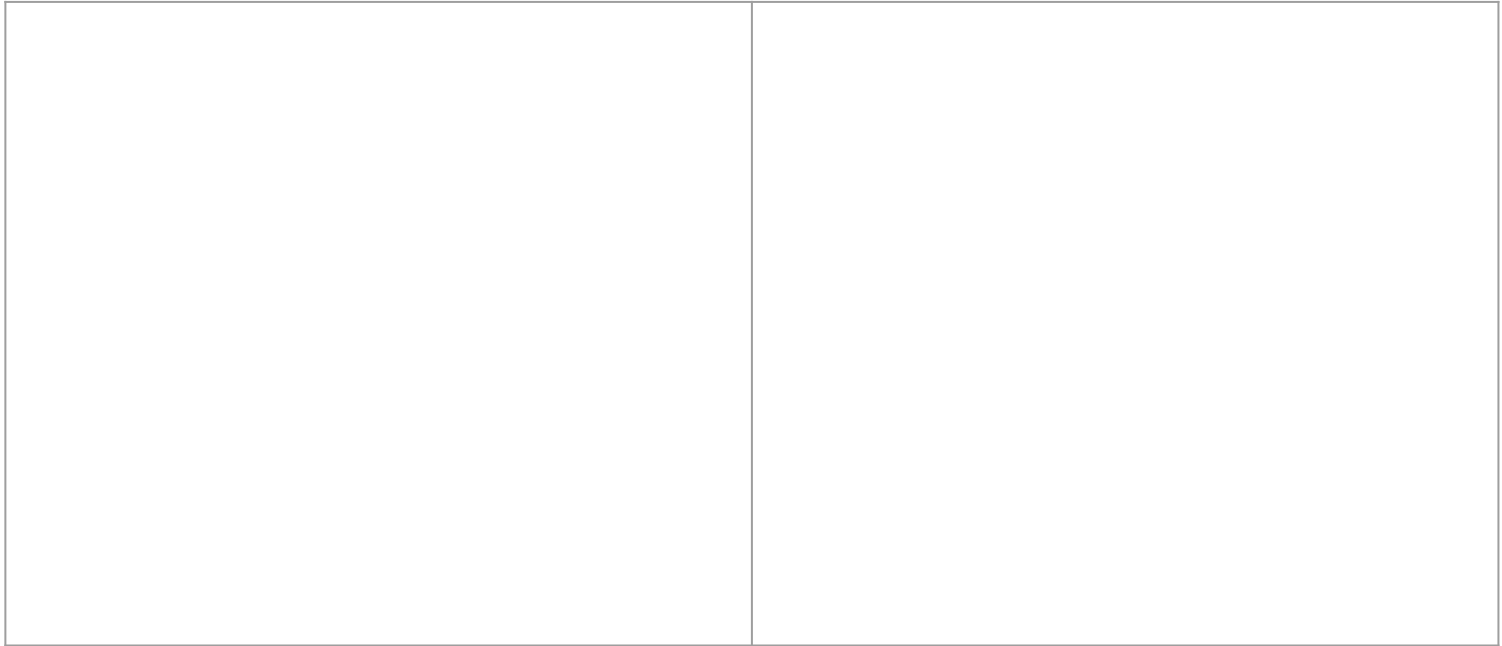
DATASET FEATURES - supervised

Study	MO phenotype	MO target
GSE12012	disrupted vessels integrity	miR-126
GSE13157	reduced neuromasts	esr2a
GSE13158	reduced neuromasts	esr2a
GSE16740	missing heart contraction	tnnt2
GSE18830	severe developmental abnormalities	sox2/3/19a/19b
GSE21539	craniofacial defects, reduced pectoral fin growth	esco2
GSE27569	Neural crest (NC) defects	Ovo1
GSE32914	developmental arrest	Sox31
GSE46844	disrupted kidney morphogenesis	mir-34b
GSE51541	cardiac defects	atg5/becl1/atg7
GSE8800	disrupted primitive hematopoiesis, brain and craniofacial defects	C1q-like



Wild Type (34) + MO (40)
Total # of Microarrays: 74

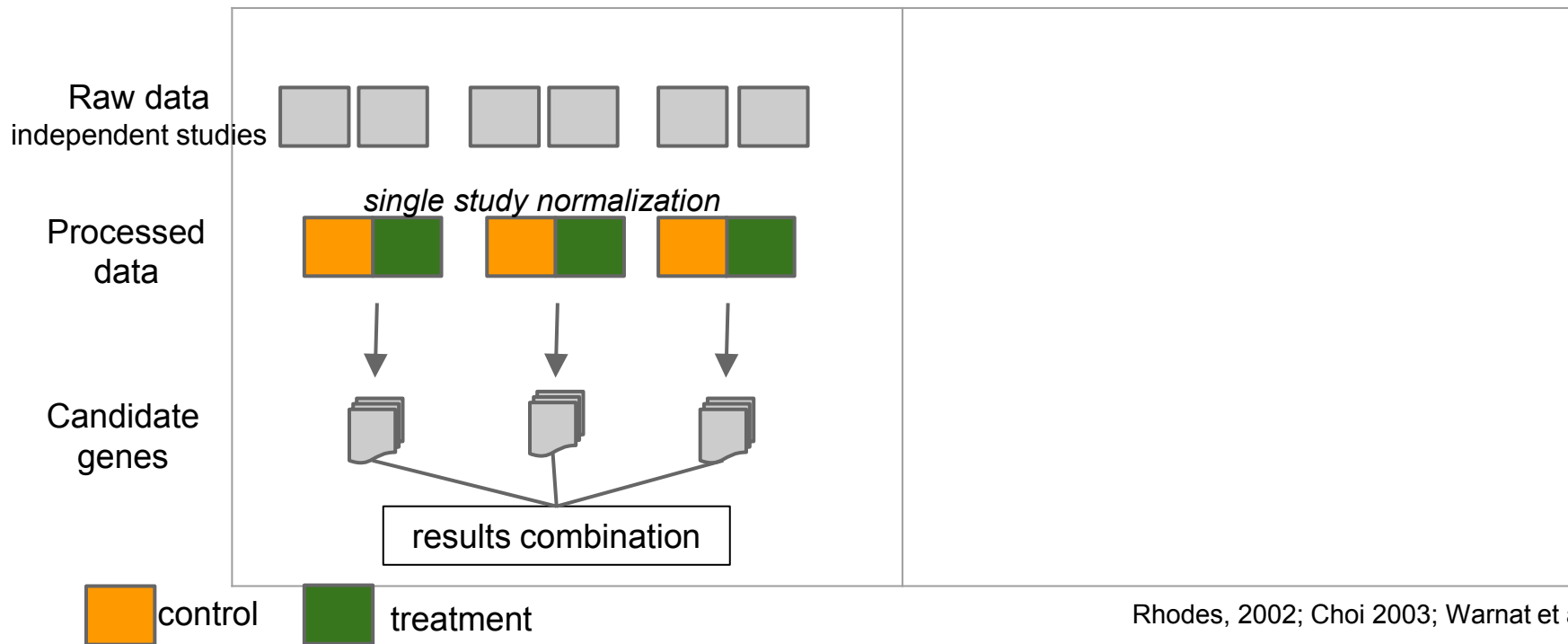
MICROARRAY DATA INTEGRATION



Rhodes, 2002; Choi 2003; Warnat et al., 2005

MICROARRAY DATA INTEGRATION

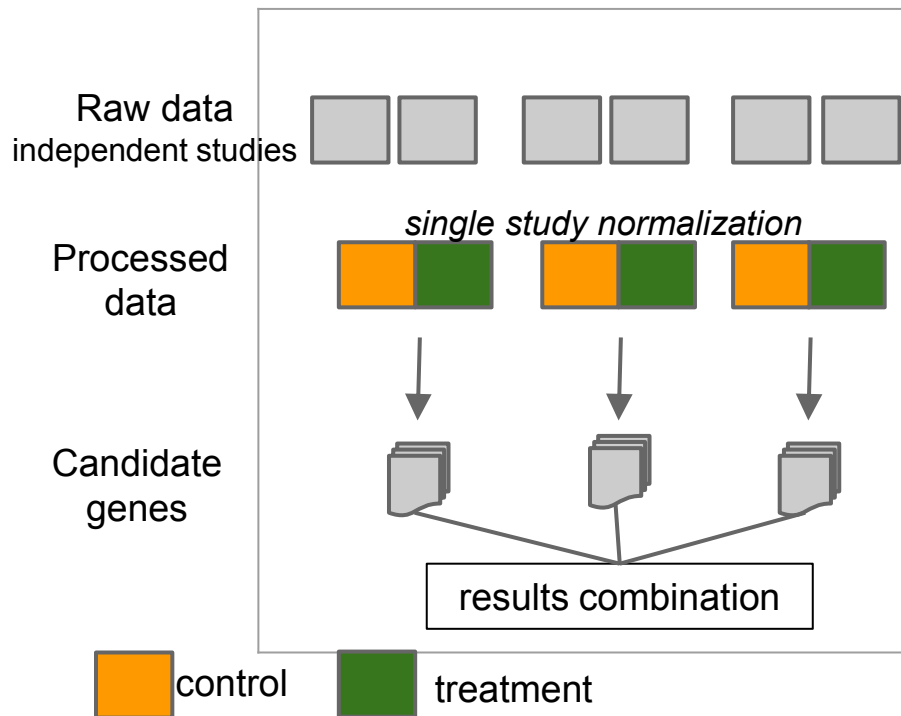
META-ANALYSIS



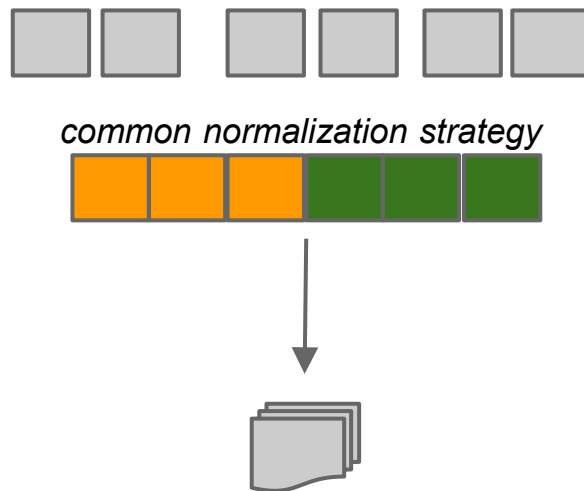
Rhodes, 2002; Choi 2003; Warnat et al., 2005

MICROARRAY DATA INTEGRATION

META-ANALYSIS



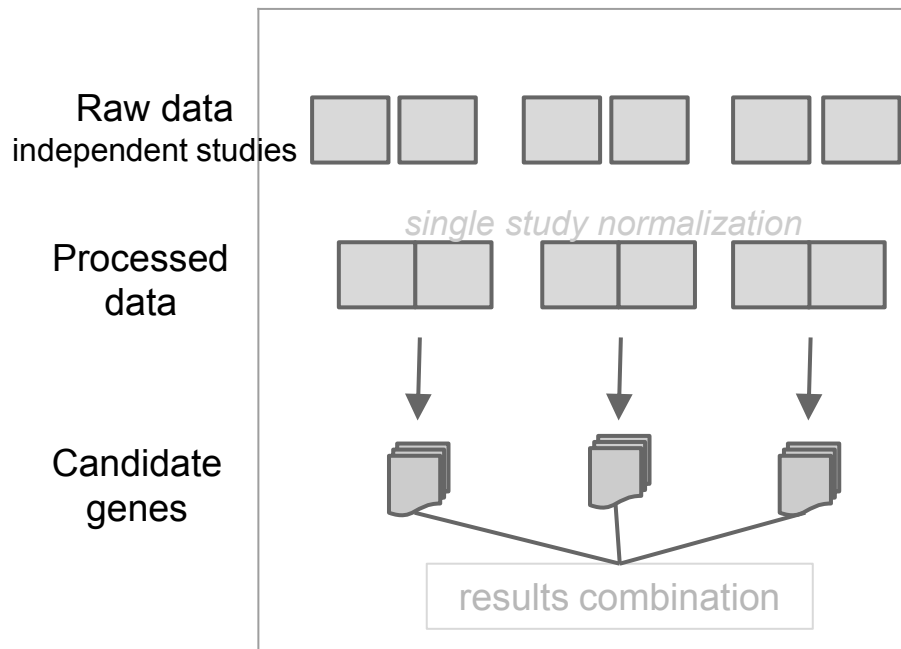
DIRECT INTEGRATION



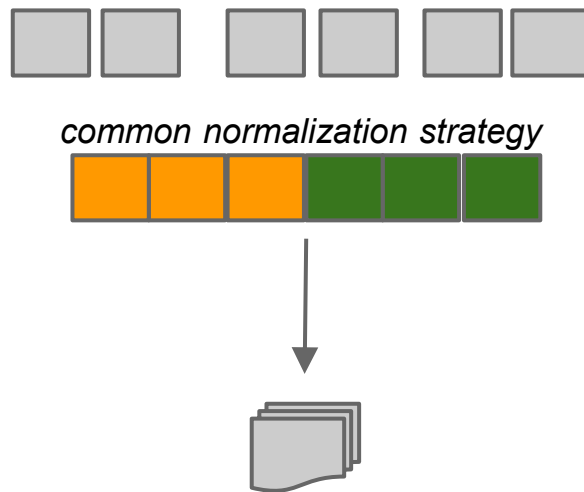
Rhodes, 2002; Choi 2003; Warnat et al., 2005

MICROARRAY DATA INTEGRATION

META-ANALYSIS



DIRECT INTEGRATION



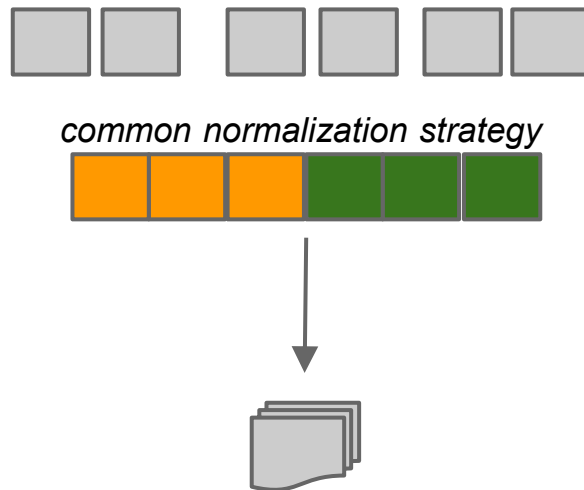
 control  treatment

Rhodes, 2002; Choi 2003; Warnat et al., 2005

MICROARRAY DATA INTEGRATION

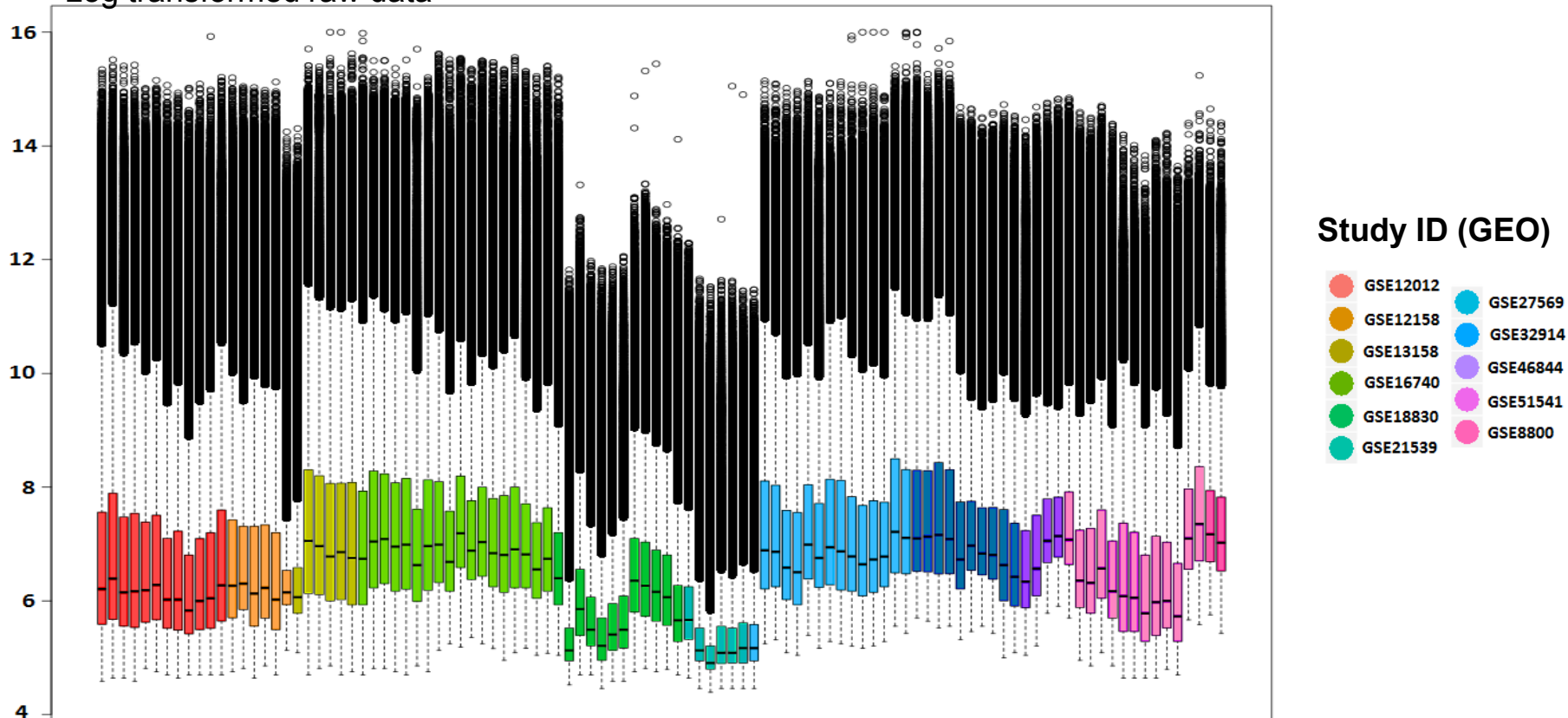
DIRECT INTEGRATION

- Larger sample size: improved generalization
- Correction for differences between protocols
- Application of machine learning approaches

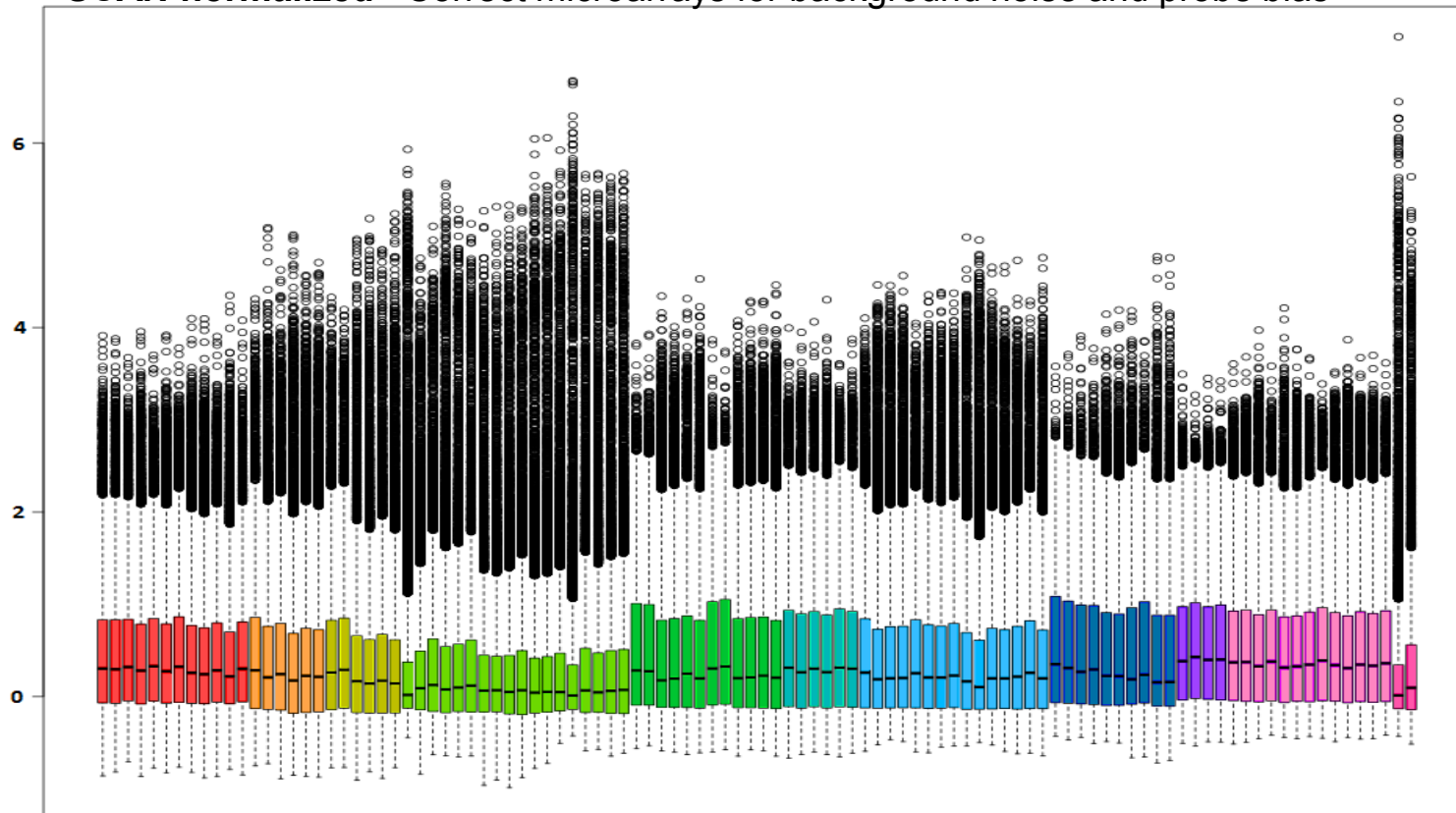


Rhodes, 2002; Choi 2003; Warnat et al., 2005

Log transformed raw data



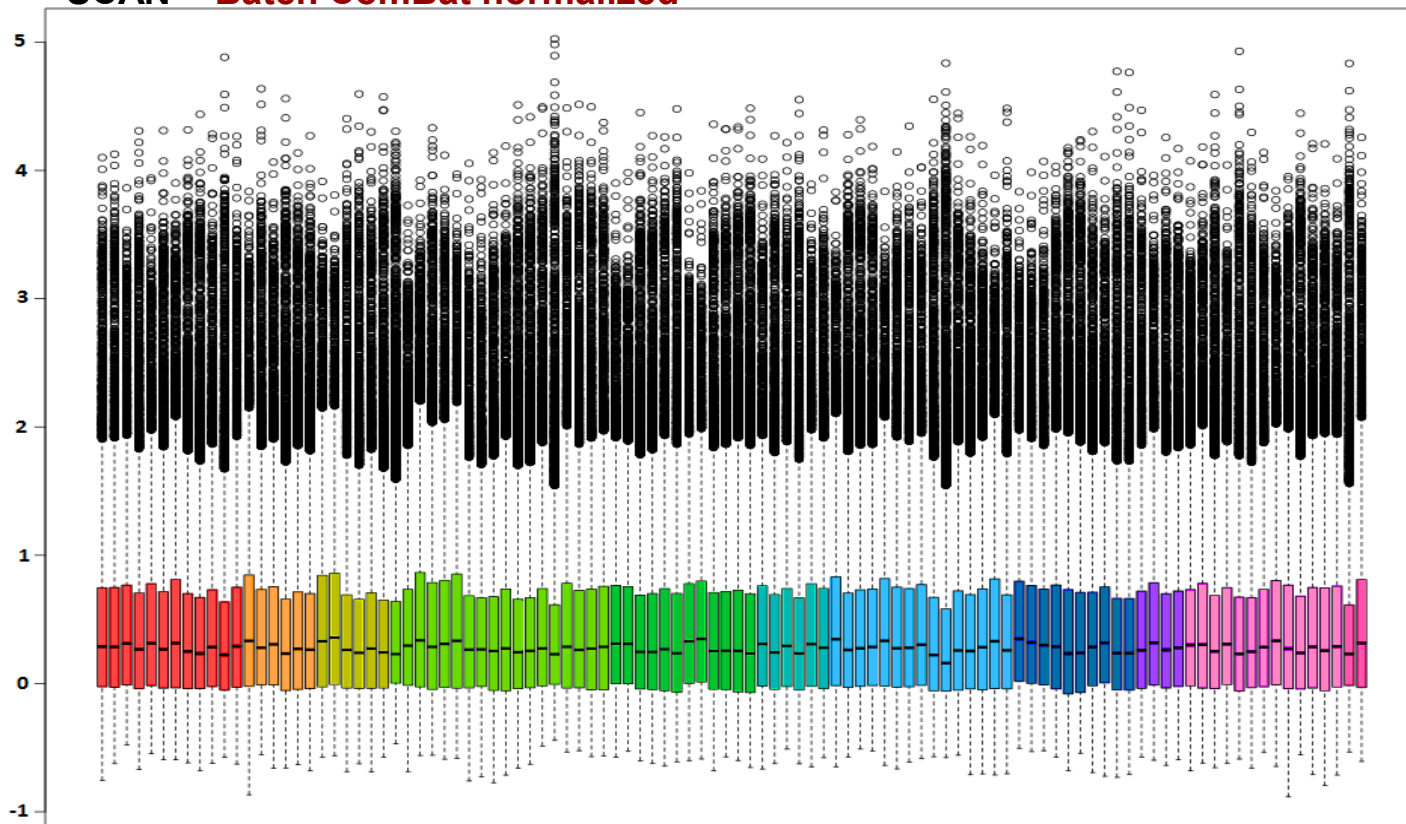
SCAN normalized - Correct microarrays for background noise and probe bias



Study ID (GEO)

- GSE12012
- GSE12158
- GSE13158
- GSE16740
- GSE18830
- GSE21539
- GSE27569
- GSE32914
- GSE46844
- GSE51541
- GSE8800

SCAN + Batch ComBat normalized*



Study ID (GEO)

- GSE12012
- GSE12158
- GSE13158
- GSE16740
- GSE18830
- GSE21539
- GSE27569
- GSE32914
- GSE46844
- GSE51541
- GSE8800

*Adjust for non-biological variability given by the differences in sample preparation and different protocols of hybridization (BATCH EFFECT)

GENES SELECTION IN SUPERVISED LEARNING

	Gene A	Gene B	Gene C	Gene D	Gene E
Sample 1	1	0	1	0	1
Sample 2	1	1	0	0	1
Sample 3	0	0	0	1	0
Sample 4	0	0	1	1	0

⋮

⋮

⋮

⋮

⋮

⋮

Dataset of microarray intensities



control

treatment

1 - High gene expression

0 - Low gene expression

Noisy/irrelevant genes: B, C

Redundant genes: A-E

High variability among the genes

APPROACH: MACHINE LEARNING

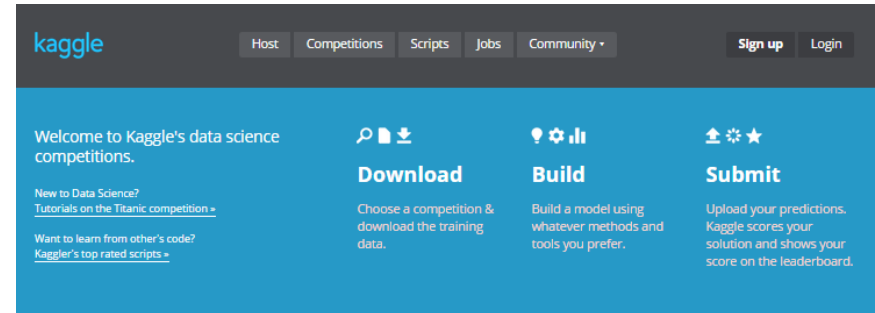
Machine learning: the study of algorithms that can **learn** and make **predictions** on data






Machine learning seeks to learn a model from the dataset and to identify a data structure

APPROACH: MACHINE LEARNING

*Machine learning: the study of algorithms that can **learn** and make **predictions** on data*

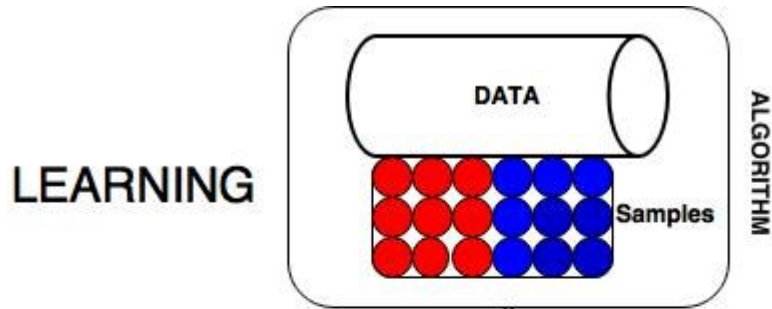
Machine learning seeks to learn a model from the dataset and to identify a data structure



Active Competitions		Competition Name	Reward	Teams	Deadline
All Competitions		 Diabetic Retinopathy Detection Identify signs of diabetic retinopathy in eye images	\$100,000	488	25 days
12 found, 12 active		 Caterpillar Tube Pricing Model quoted prices for industrial tube assemblies	\$30,000	283	60 days
Search competitions		 Avito Context Ad Clicks Predict if context ads will earn a user's click	\$20,000	147	26 days
All competitions Enterable		 Search Results Relevance Predict the relevance of search results from eCommerce sites	\$20,000	1334	4.2 days
Status Active Completed		 ICDM 2015: Drawbridge Cross-Device Connections Identify individual users across their digital devices	\$10,000	131	53 days
Sponsor InClass (student)					

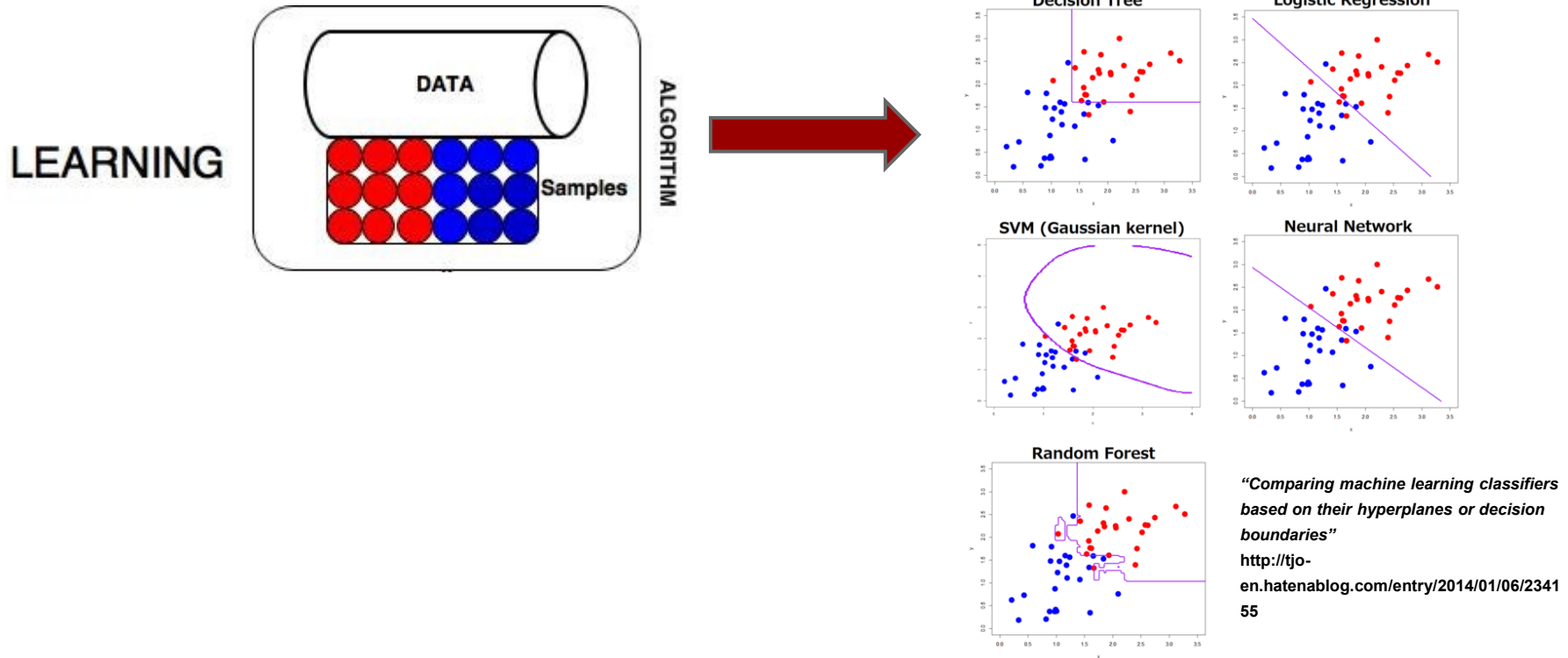
<https://www.kaggle.com/>

APPROACH: MACHINE LEARNING

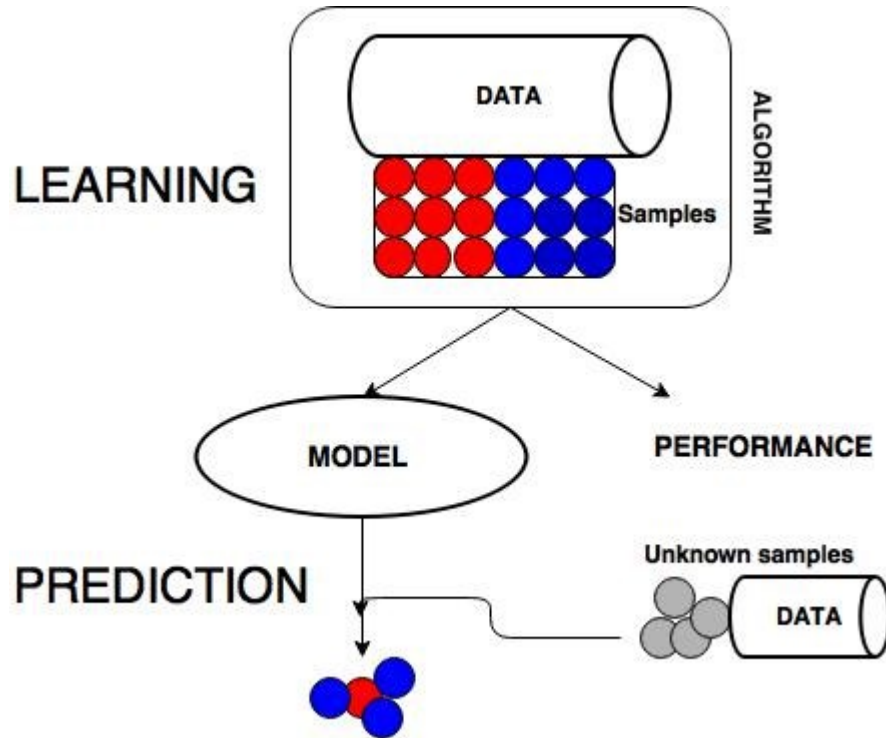


		GENES				
		Gene A	Gene B	Gene C	Gene D
SAMPLES	Sample 1				
	Sample 2				
....						
....						

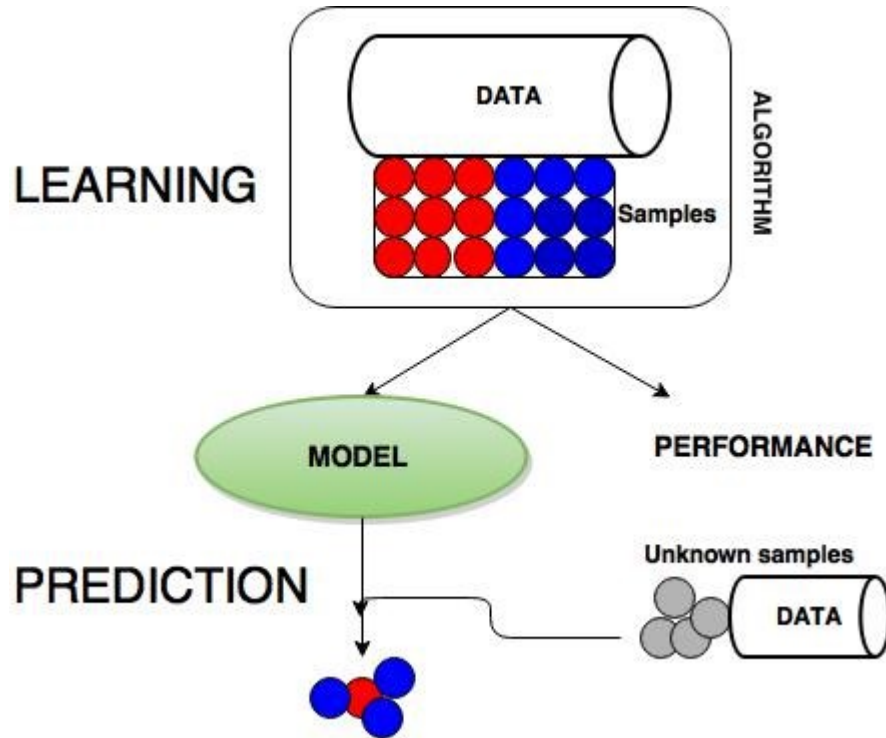
APPROACH: MACHINE LEARNING



APPROACH: MACHINE LEARNING

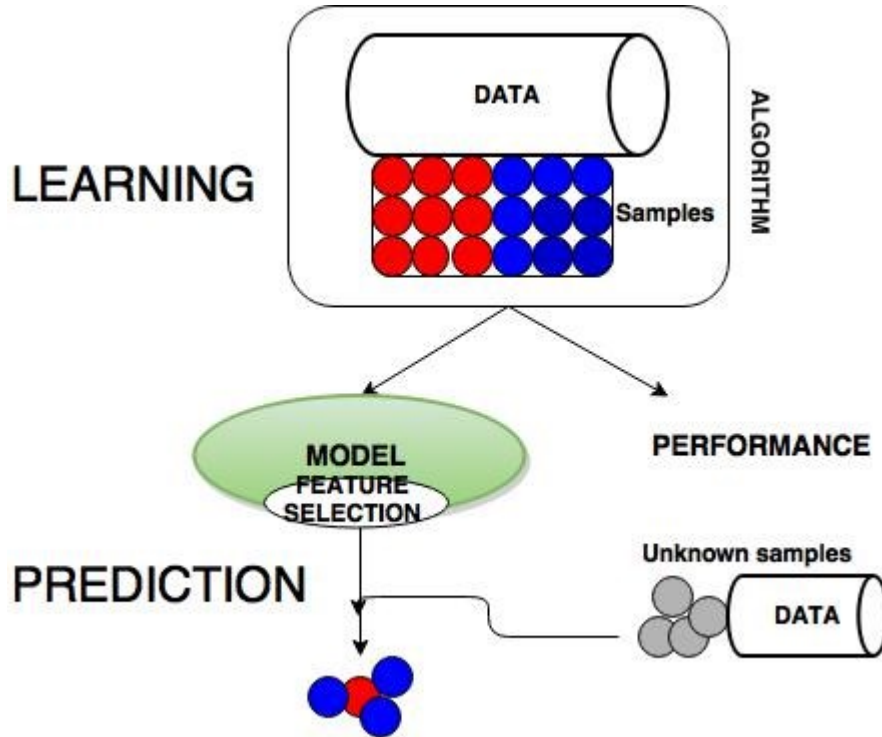


APPROACH: MACHINE LEARNING

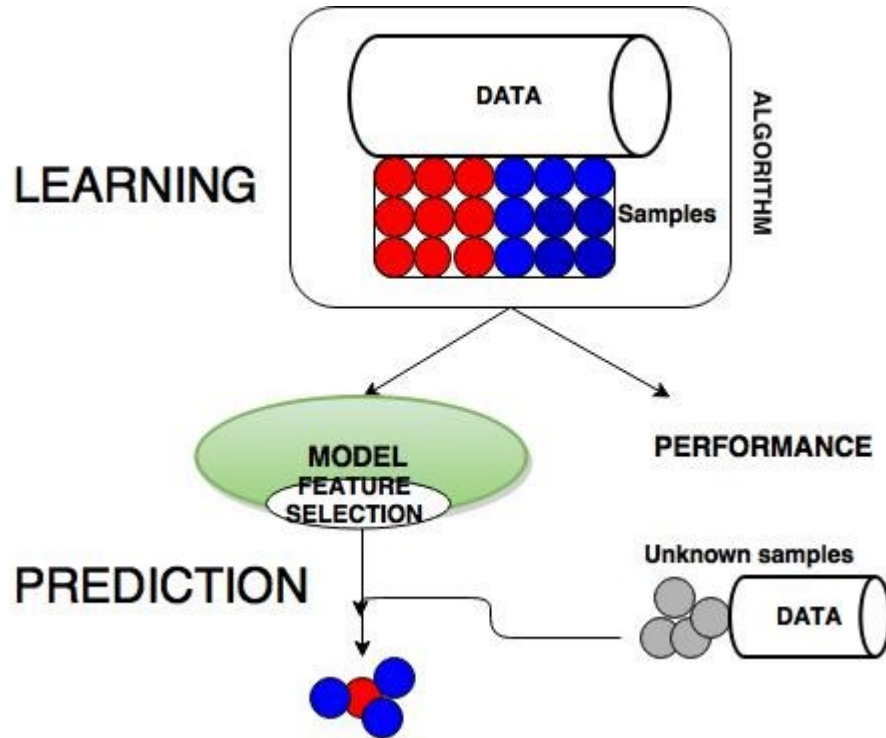


APPROACH: MACHINE LEARNING

The feature selection in the model identifies a small subset of genes which has a high discriminant power between the two groups



APPROACH: MACHINE LEARNING



The feature selection in the model identifies a small subset of genes which has a high discriminant power between the two groups

BioChip J. (2013) 7(1): 57-62
DOI 10.1007/s13206-013-7109-8

Original Research

Identification of cancer-specific biomarkers by using microarray gene expression profiling

Jin Choul Chai¹, Soojun Park², Hyemyung Seo¹, Soo Young Cho³ & Young Seek Lee¹

Received: 25 December 2012 / Accepted: 9 January 2013 / Published online: 20 March 2013
© The Korean BioChip Society and Springer 2013

Biomarkers selection through feature selection for cancer diagnosis and prevention

- Screening tests
- Identification of cancer subtypes

RESULTS

Principal Component Analysis and
treatment effect

(Unsupervised learning)

GSEA from the genes ranked by
Machine Learning

(Supervised learning)

Stage specific gene expression in
morphants

RESULTS

Principal Component Analysis and
treatment effect

(Unsupervised learning)

GSEA from the genes ranked by
Machine Learning

(Supervised learning)

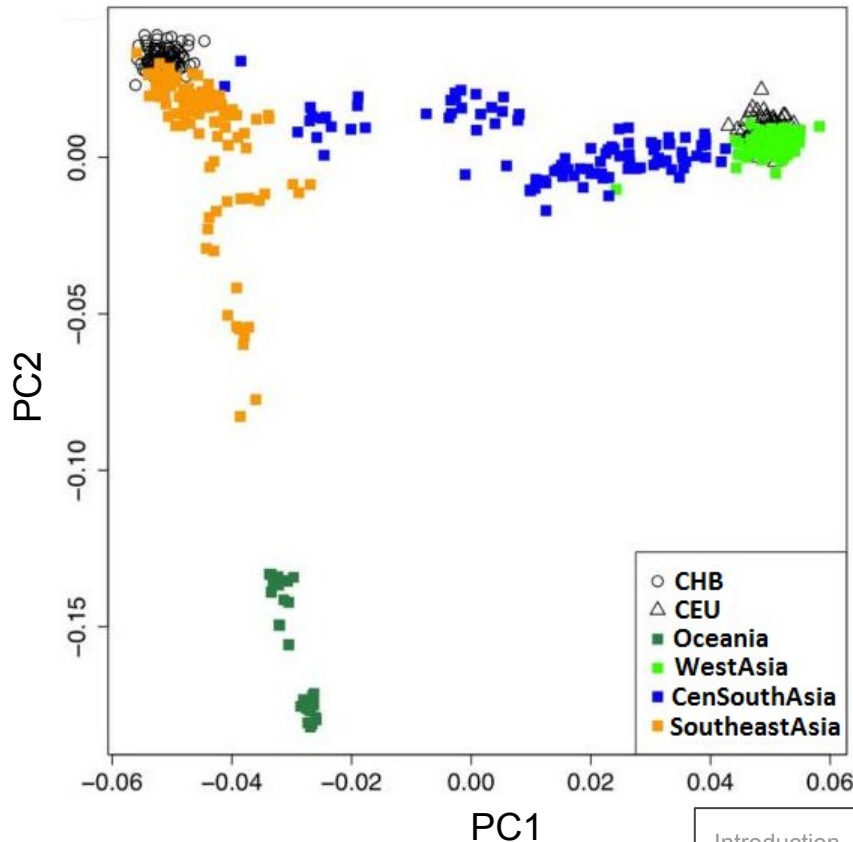
Stage specific gene expression in
morphants

PRINCIPAL COMPONENT ANALYSIS

PCA describes the main sources of variance from the data as principal components (PC)

Every sample is plotted in the space of PC for finding samples grouping or subpopulations

PRINCIPAL COMPONENT ANALYSIS



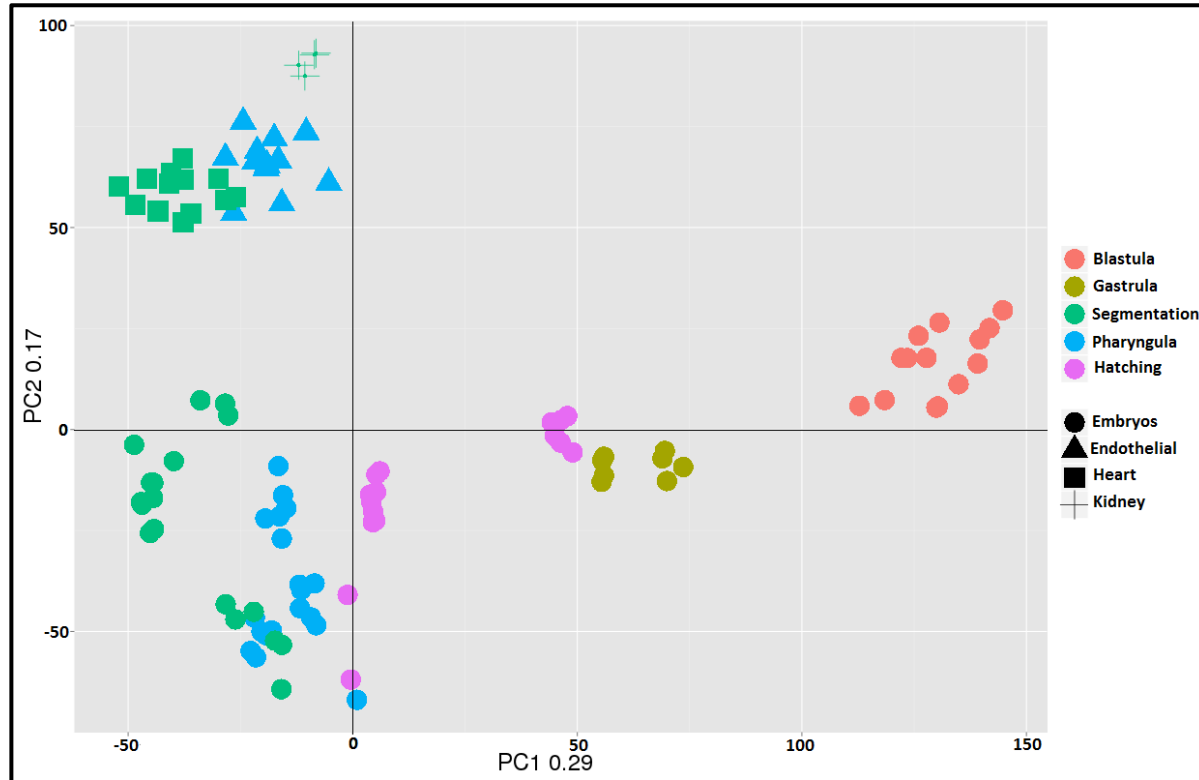
PCA describes the main sources of variance from the data as principal components (PC)

Every sample is plotted in the space of PC for finding samples grouping or subpopulations

Figure 2C

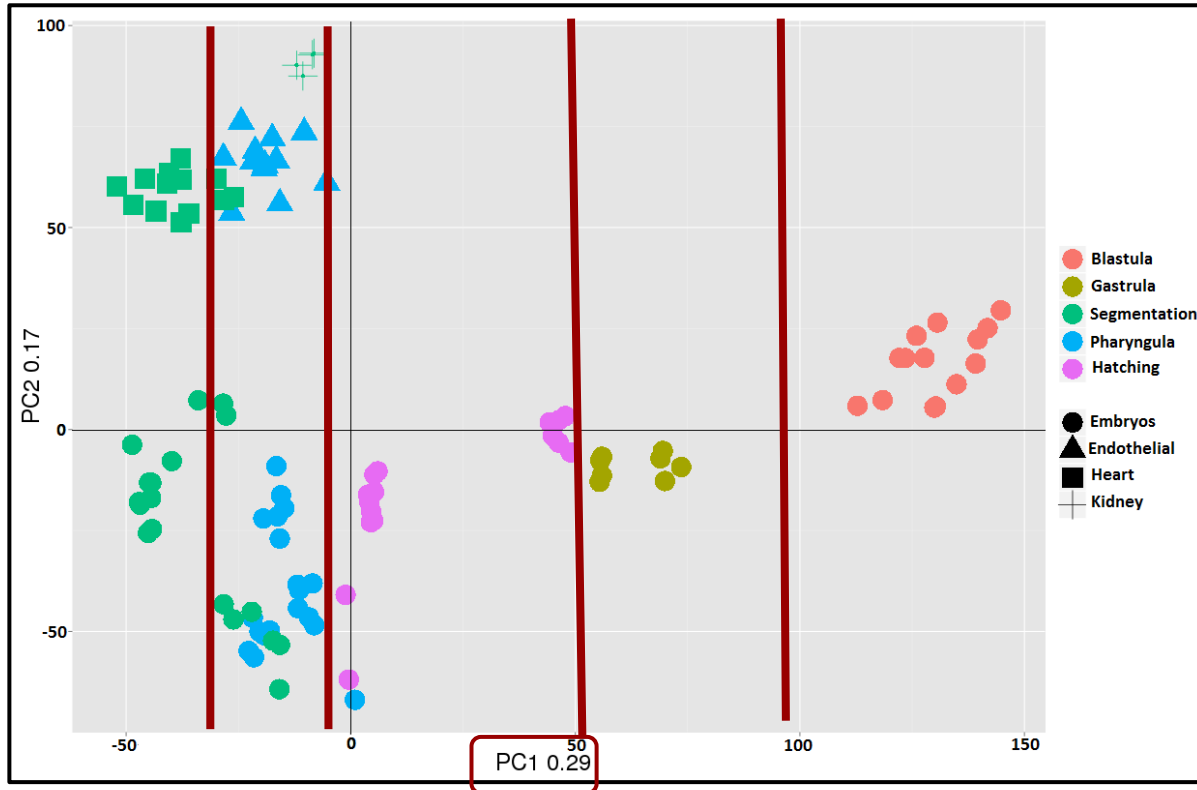
Principal component analysis reveals the 1000 Genomes Project does not sufficiently cover the human genetic diversity in Asia
Dongsheng Lu and Shuhua Xu

PRINCIPAL COMPONENT ANALYSIS



Samples labeled by
tissue source and
dev.stage

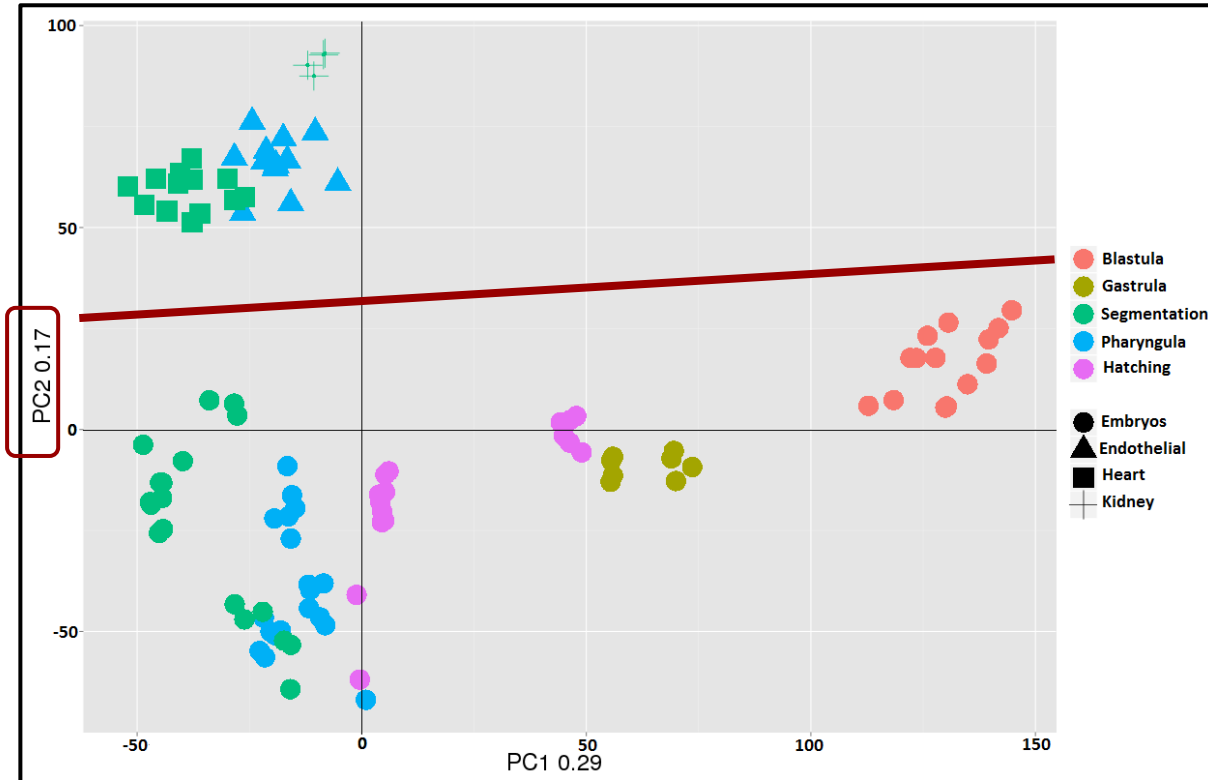
PRINCIPAL COMPONENT ANALYSIS



Samples labeled by
tissue source and
dev.stage

PC1 - separation by
developmental stage

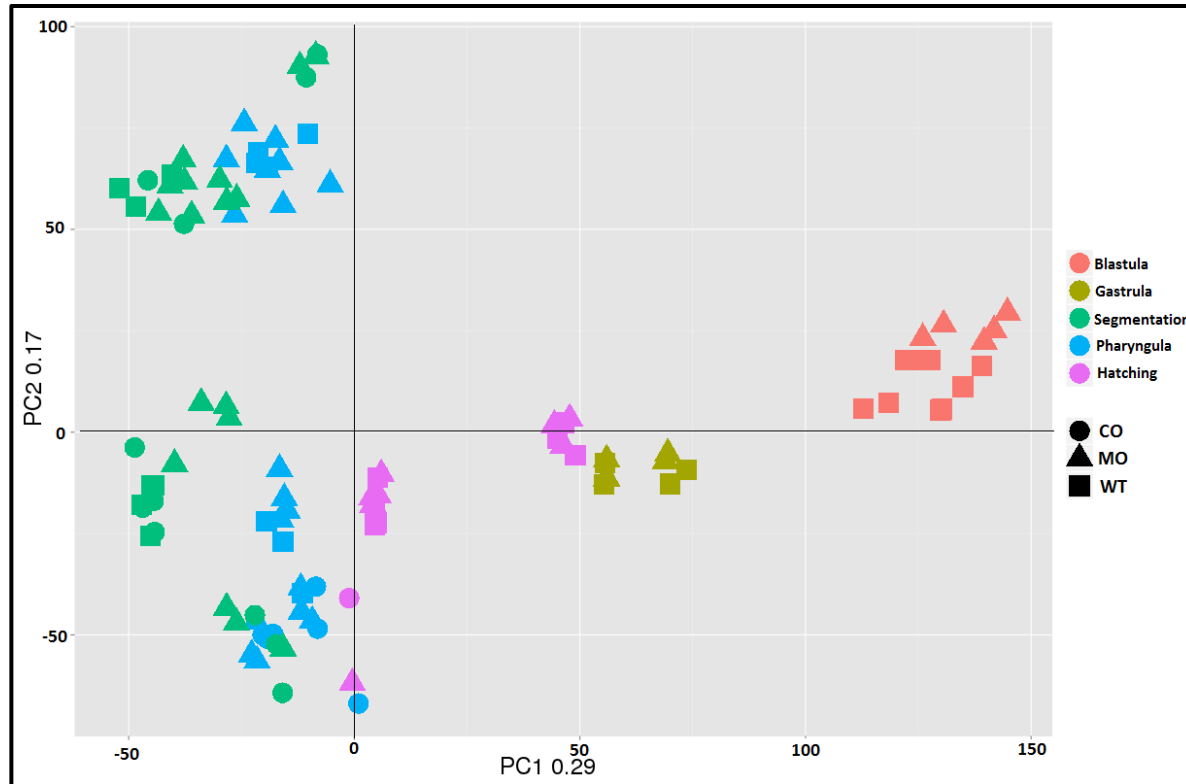
PRINCIPAL COMPONENT ANALYSIS



Samples labeled by
tissue source and
dev.stage

PC2 - separation by
tissue type

PRINCIPAL COMPONENT ANALYSIS



Samples labeled by
treatment and
dev.stage

RESULTS

Principal Component Analysis and
treatment effect

(Unsupervised learning)

GSEA from the genes ranked by
Machine Learning

(Supervised learning)

Stage specific gene expression in
morphants

Gene Set Enrichment Analysis

- Machine learning and feature selection for MO vs WT
- Model and accuracy optimization - selection of the best performing model
- Ranked list of features as output - features with high discriminant power on top positions

Gene Set Enrichment Analysis

- Machine learning and feature selection for MO vs WT
- Model and accuracy optimization - selection of the best performing model
- Ranked list of features as output - features with high discriminant power on top positions

GORilla is a tool for identifying and visualizing enriched GO terms in ranked lists of genes.

It can be run in one of two modes:

1. Searching for enriched GO terms that appear densely at the top of a ranked list of genes or
2. Searching for enriched GO terms in a target list of genes compared to a background list of genes.

For further details see [References](#).

[Running example](#)

[Usage instructions](#)

[GORilla News](#) (Updated March 8th 2018)

Step 1: [Choose organism](#)

Danio rerio (Zebrafish)

Step 2: [Choose running mode](#)

☒ Single ranked list of genes

☐ Two unranked lists of genes (target and background lists)



GORILLA



Gene Set Enrichment Analysis

- Machine learning and feature selection for MO vs WT
- Model and accuracy optimization - selection of the best performing model
- Ranked list of features as output - features with high discriminant power on top positions

Output in Gene Ontology domains:

- **Biological process**
- **Molecular function**
- Cellular component

GORilla is a tool for identifying and visualizing enriched GO terms in ranked lists of genes.

It can be run in one of two modes:

1. Searching for enriched GO terms that appear densely at the top of a ranked list of genes or
2. Searching for enriched GO terms in a target list of genes compared to a background list of genes.

For further details see [References](#).

[Running example](#)

[Usage instructions](#)

[GORilla News](#) (Updated March 8th 2018)

Step 1: [Choose organism](#)

Danio rerio (Zebrafish)

Step 2: [Choose running mode](#)

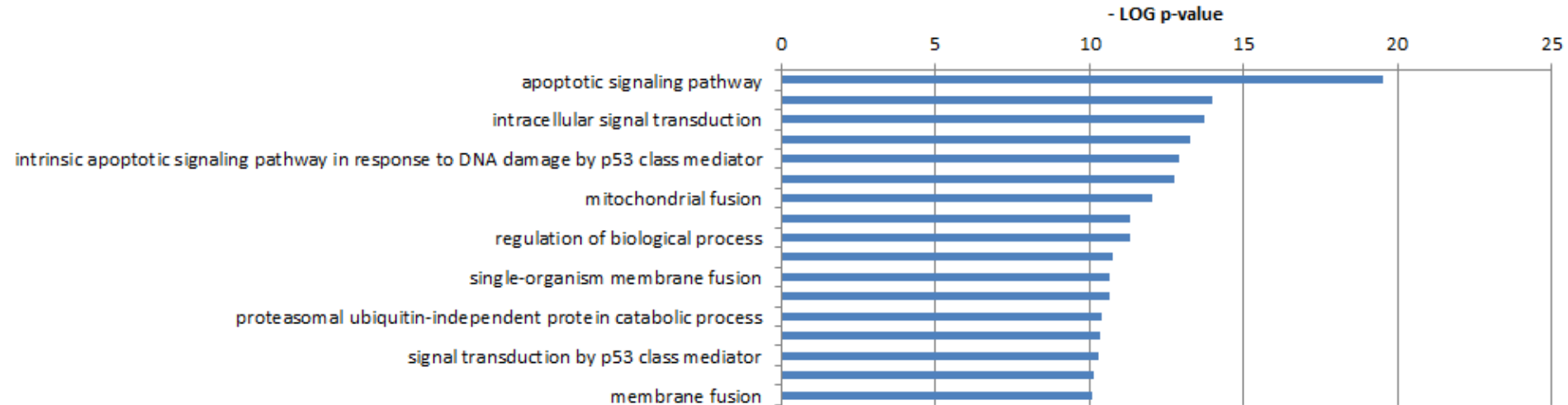
- ☒ Single ranked list of genes ☐ Two unranked lists of genes (target and background lists)



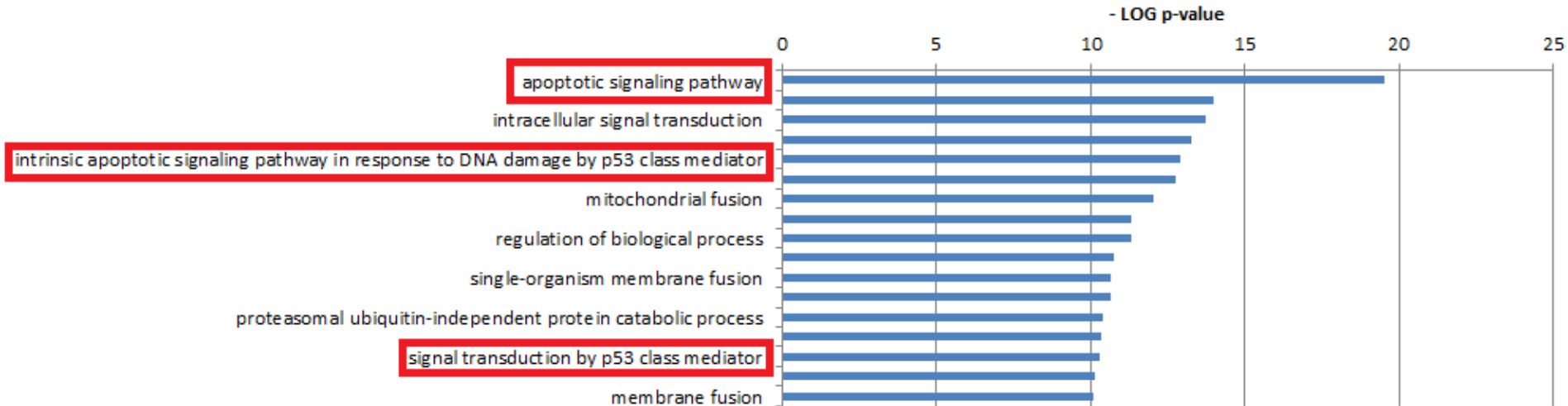
GORILLA



BIOLOGICAL PROCESS



BIOLOGICAL PROCESS



phlda3 - pleckstrin homology-like domain, family a, member 3

dap1b - death associated protein 1b

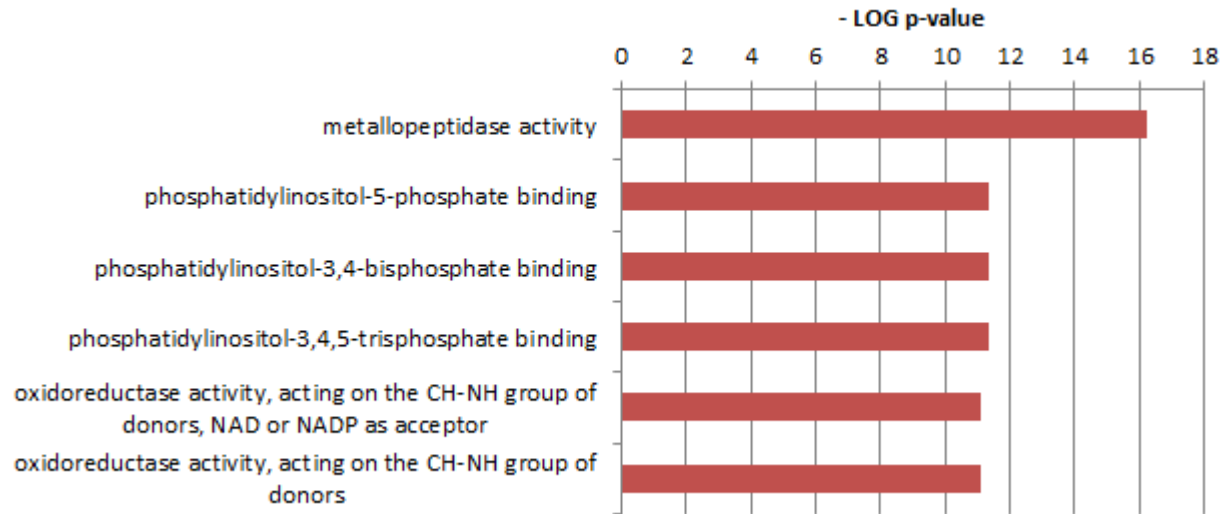
tp53 - tumor protein p53

baxa - bcl2-associated x protein, a

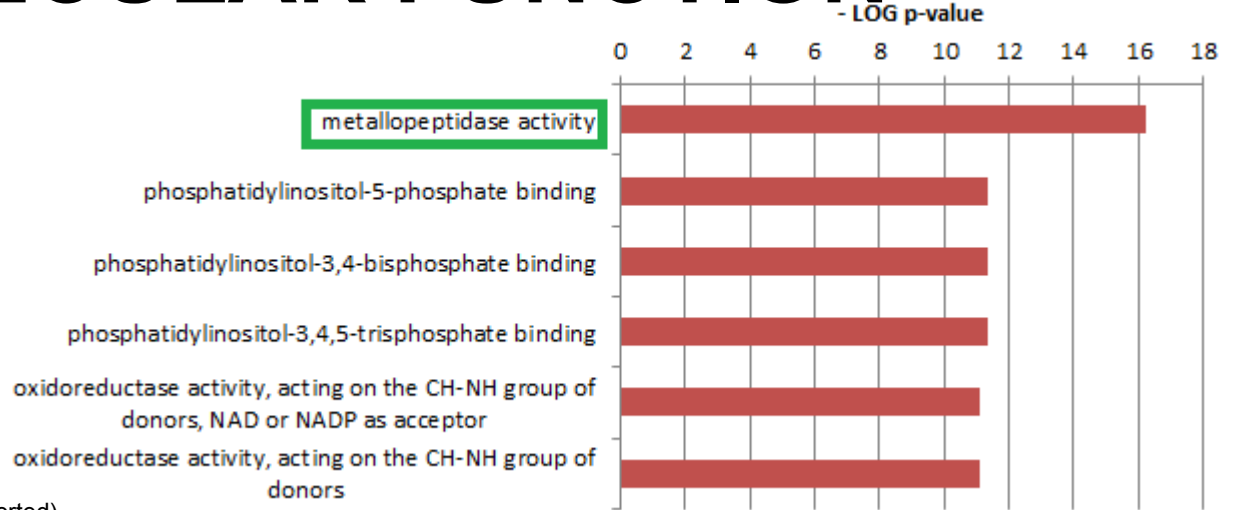
PHLDA3 The molecule interferes with Akt and promotes apoptosis. The protein is induced by DNA-damaging agents in cell lines (Kawase et al., 2009)

baxa pro-apoptotic molecule

MOLECULAR FUNCTION



MOLECULAR FUNCTION



mmp14b - matrix metalloproteinase 14b (membrane-inserted)
mmp11b - matrix metalloproteinase 11b (stromelysin 3)
zgc:152830 - zgc:152830
zmpste24 - zinc metalloproteinase, ste24 homolog
taf2 - taf2 rna polymerase ii, tata box binding protein (tbp)-associated factor
cndp2 - cndp dipeptidase 2 (metalloproteinase m20 family)
mmp9 - matrix metalloproteinase 9
phex - phosphate regulating gene with homologues to endopeptidases on the x chromosome
mmp2 - matrix metalloproteinase 2
ace - angiotensin i converting enzyme (peptidyl-dipeptidase a) 1
cpda - carboxypeptidase d, a
metap1 - methionyl aminopeptidase 1
ide - insulin-degrading enzyme
cpn1 - carboxypeptidase n, polypeptide 1

***Mmp** matrix metalloproteinase (MMP) family involved in the breakdown of extracellular matrix in normal physiological processes*

RESULTS

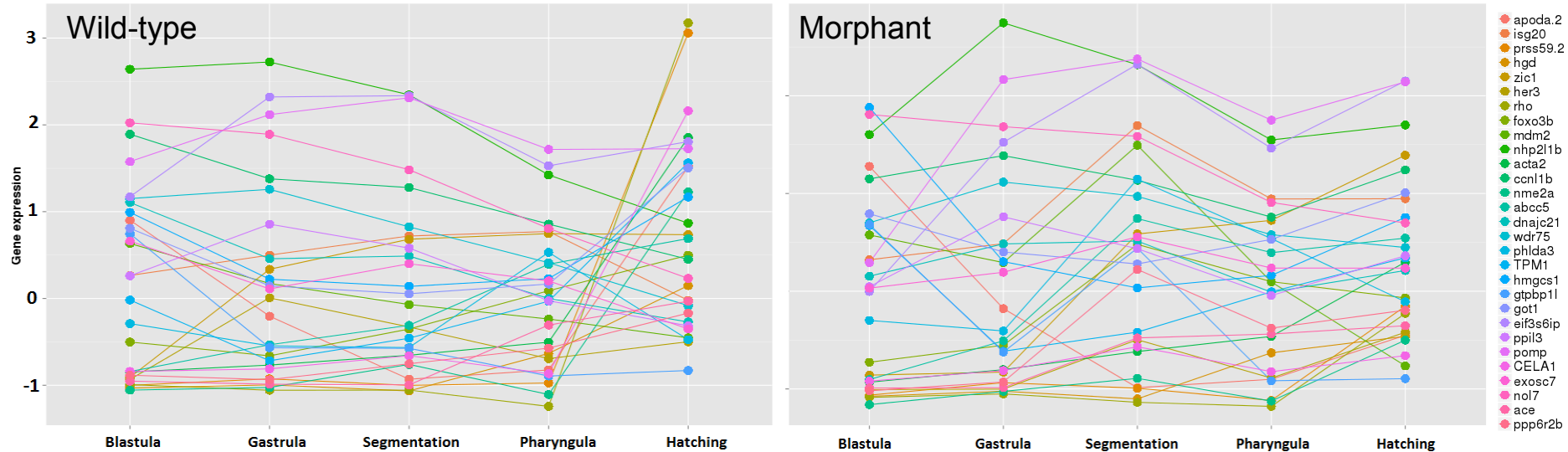
PCA and treatment effect
(Unsupervised learning)

GSEA from the genes ranked by
Machine Learning
(Supervised learning)

Stage specific gene expression in
morphants

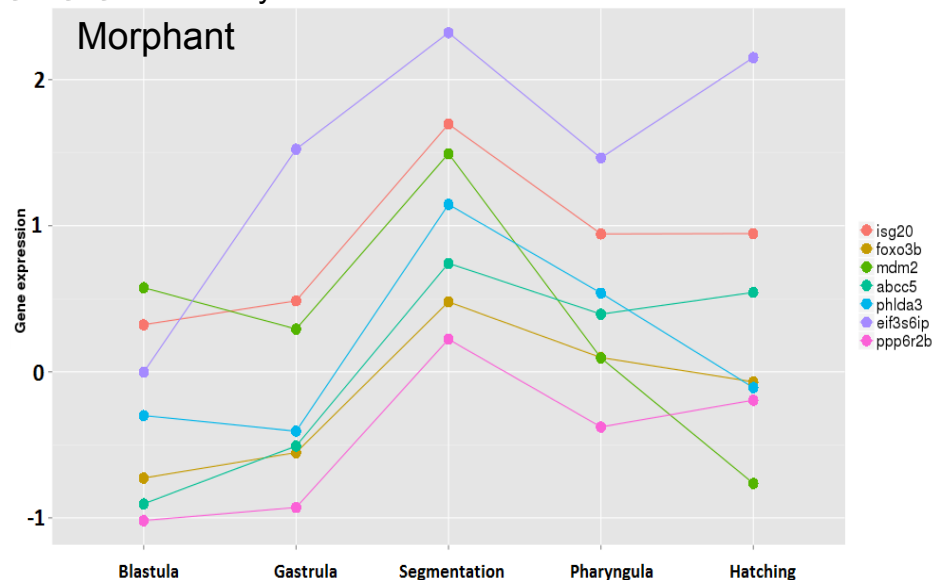
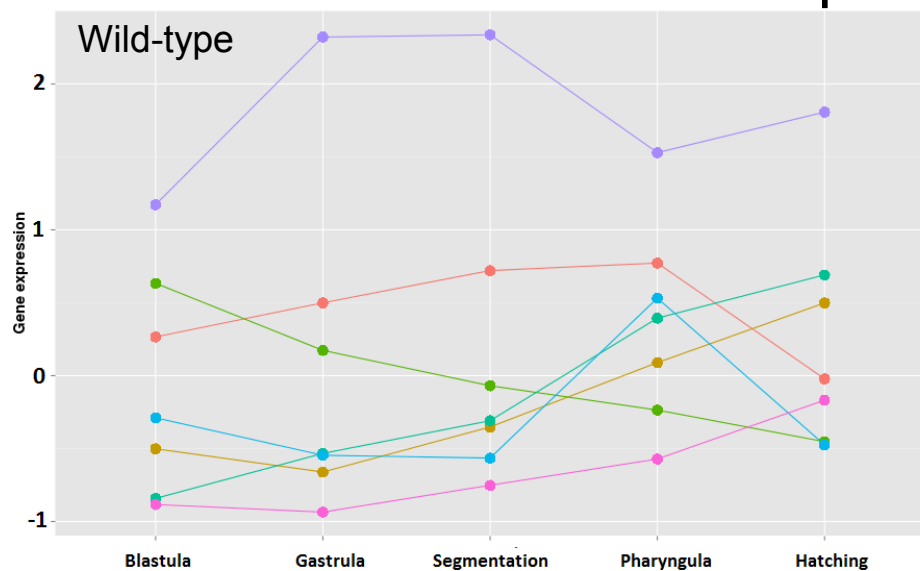
STAGE DEPENDENT EFFECT

$p < 0.001$ two-way ANOVA interaction effect



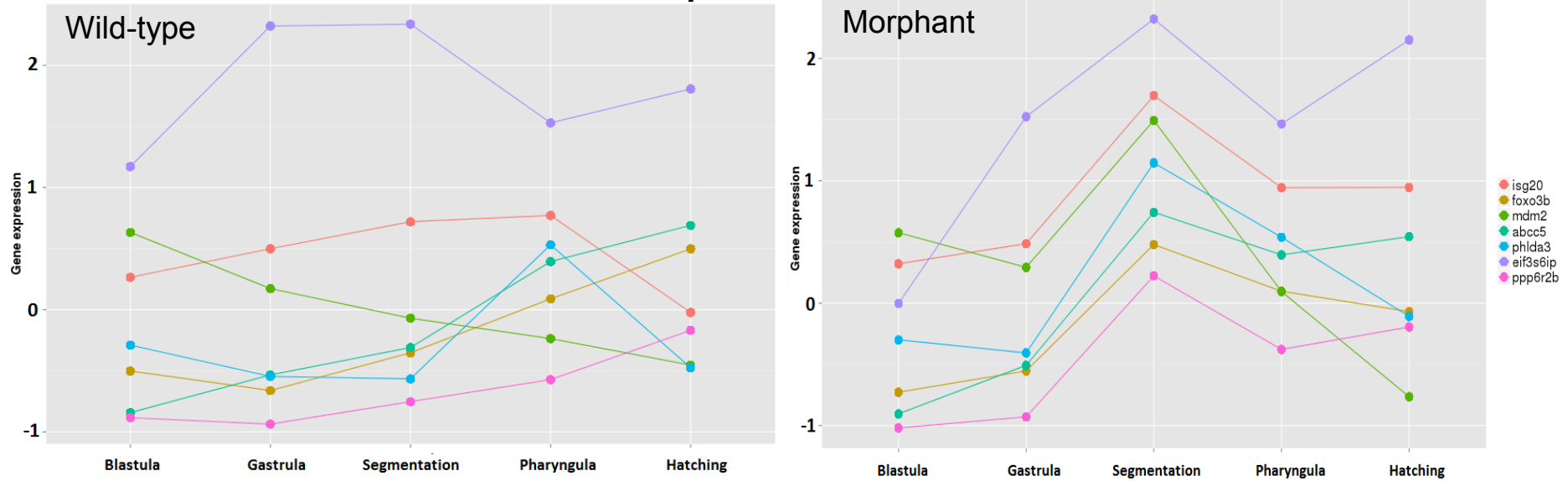
STAGE DEPENDENT EFFECT

$p < 0.001$ two-way ANOVA interaction effect



STAGE DEPENDENT EFFECT

$p < 0.001$ two-way ANOVA interaction effect



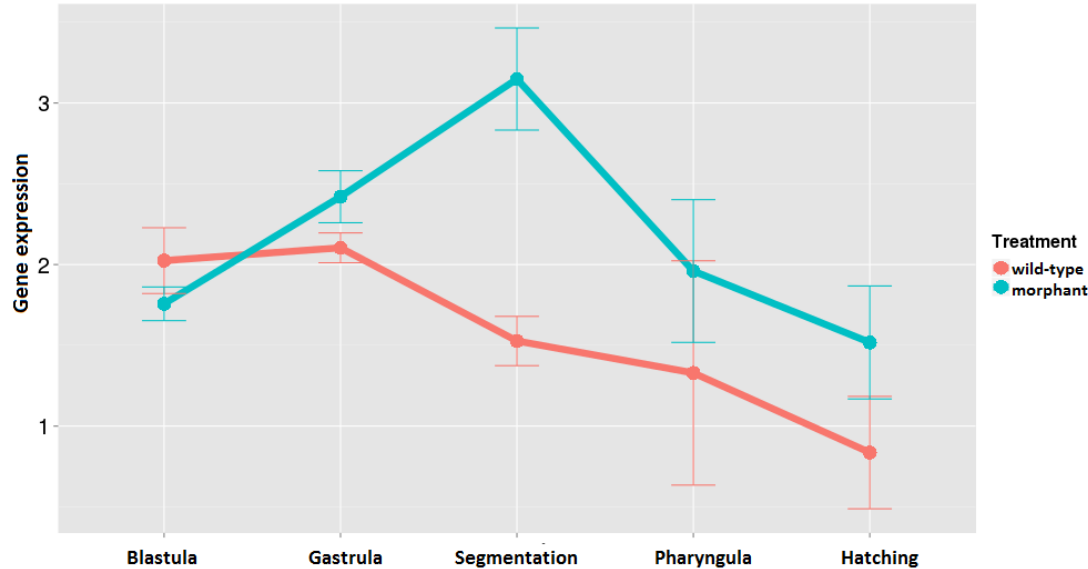
isg20 Interferon Stimulated Exonuclease Gene 20kDa, Interferon-induced antiviral exoribonuclease that acts on single-stranded RNA

abcc5 Member of the superfamily of ATP-binding cassette (ABC) transporters, acts as a multispecific organic anion pump which can transport nucleotide analogs

eif3s6ip Eukaryotic Translation Initiation Factor 3 Subunit 6-Interacting Protein, eukaryotic translation initiation factor 3 (eIF-3) complex, which is required for several steps in the initiation of protein synthesis

STAGE DEPENDENT EFFECT: p53

p-value < 0.05 two-way ANOVA interaction effect



p53 upregulation during the segmentation stage in morphants

p53 in wild type embryos has the highest gene expression at 1 hpf and by 48 hpf the levels of p53 transcript and protein are barely detectable

Cheng et al. 1997; Thisse et al. 2000; Lee et al. 2008

DISCUSSION AND CONCLUSIONS (1)

DATA ANALYSIS APPROACH

Successful microarray processing and integration. Observation of main biological features from unsupervised learning such as developmental stage and tissue type

DISCUSSION AND CONCLUSIONS (1)

DATA ANALYSIS APPROACH

Successful microarray processing and integration. Observation of main biological features from unsupervised learning such as developmental stage and tissue type

Efficient strategy for gene selection in microarrays

- algorithms and feature selection optimization
- identification of the optimal number of features

DISCUSSION AND CONCLUSIONS (2)

BIOLOGICAL RESULTS

- Overall gene expression is not strongly altered by the MO treatment

DISCUSSION AND CONCLUSIONS (2)

BIOLOGICAL RESULTS

- Overall gene expression is not strongly altered by the MO treatment
 - Possible off-target effects observed from integrated dataset
- Upregulation of gene expression in Morphants (data not shown)

DISCUSSION AND CONCLUSIONS (2)

BIOLOGICAL RESULTS

- Overall gene expression is not strongly altered by the MO treatment
 - Possible off-target effects observed from integrated dataset
- Upregulation of gene expression in Morphants (data not shown)
 - Activation of p53 pathway and matrix metalloproteinases

DISCUSSION AND CONCLUSIONS (2)

BIOLOGICAL RESULTS

- Overall gene expression is not strongly altered by the MO treatment
 - Possible off-target effects observed from integrated dataset
- Upregulation of gene expression in Morphants (data not shown)
 - Activation of p53 pathway and matrix metalloproteinases
- Stage specific effect observed during the segmentation stage

FUTURE DIRECTIONS

- Morpholino data analysis for other microarray platforms and technologies

FUTURE DIRECTIONS

- Morpholino data analysis for other microarray platforms and technologies
 - Study involving other control morpholinos
 - mismatch
 - random sequence
 - buffer injection
- MO sequence specific studies in correlation to off-target effects

FUTURE DIRECTIONS

- Morpholino data analysis for other microarray platforms and technologies
 - Study involving other control morpholinos
 - mismatch
 - random sequence
 - buffer injection
- MO sequence specific studies in correlation to off-target effects
 - Validation of isg20, phlda3, mmp upregulation in MO

Thanks To

Didier
Jason
Soraya and Sriram

Andrea R., Javad, Chris, Zach,
Carol, Ziba, Aly, Claudia, Michele M.

...

and all the members of the Lab!



Special thanks to



WEKA
The University
of Waikato



supplementary information

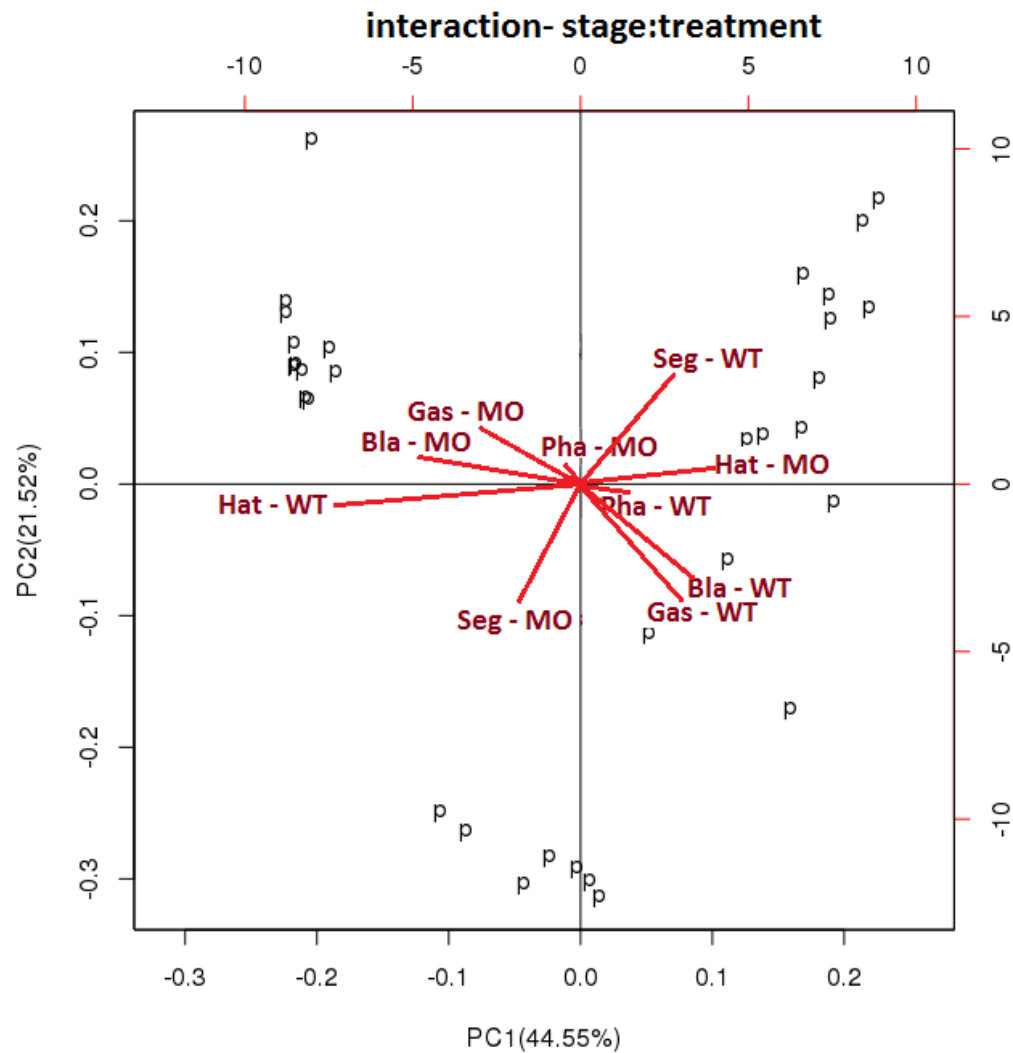
MO validation in studies

Study	RT-qPCR	Luciferase assay	GFP assay	Rescue MO	Western blot	Co-injection with mutated target	None
GSE12012	x						
GSE13157						x	
GSE13158						x	
GSE16740							x
GSE18830		x					
GSE21539			x				
GSE27569							x
GSE32914				x			
GSE46844			x				
GSE51541					x		
GSE8800	x						
	2	1	2	1	1	2	2

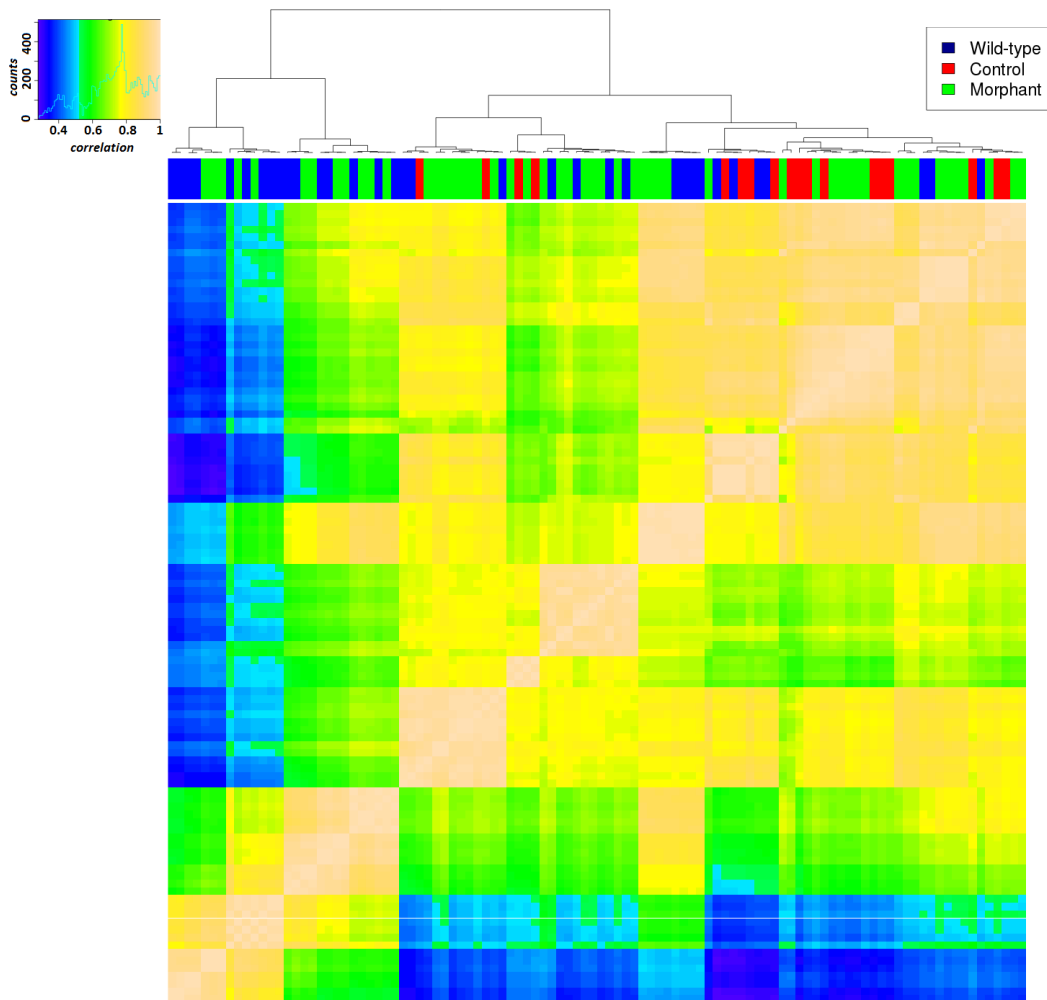
Study details

Study	ng	picoM	Type MO*	Gene
GSE12012	6	-	miRNA	miR- 126
GSE13157	-	0.015/0.05	T	Esr2a
GSE13158	-	0.015/0.05	T	Esr2a
GSE16740	4	-	T	Tnnt2
GSE18830	3.6	-	T	Sox2/3/19a/19b
GSE21539	3	-	T	Ovo1
GSE27569	2	-	T	Esco2
GSE32914	-	1.2	S	Sox31
GSE46844	-	1	miRNA	miR-34B
GSE51541	-	0.65	T	Atg5/Bcn1/Atg7
GSE8800	5	-	T	C1q-like

*Target inactivation strategy: S – splicing, T – translation, miRNA

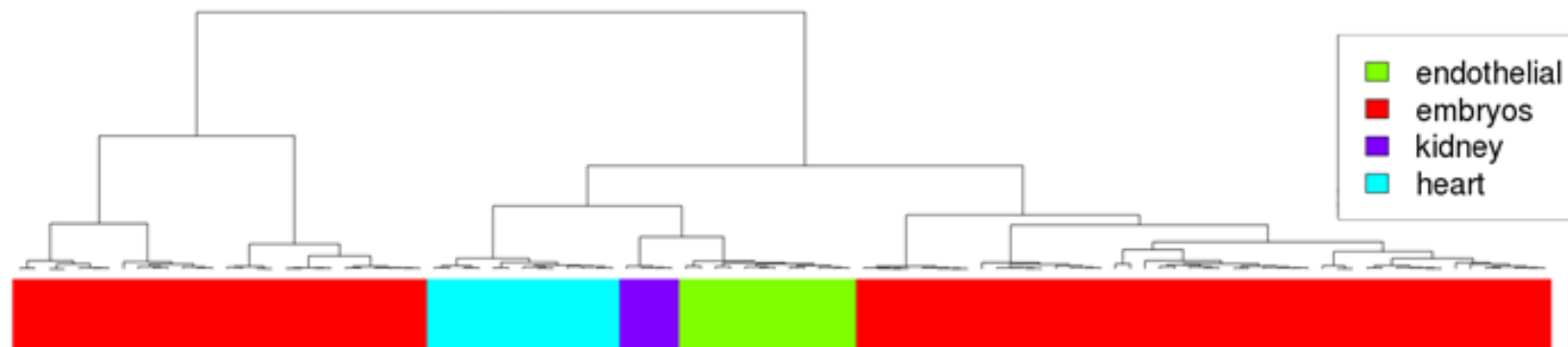
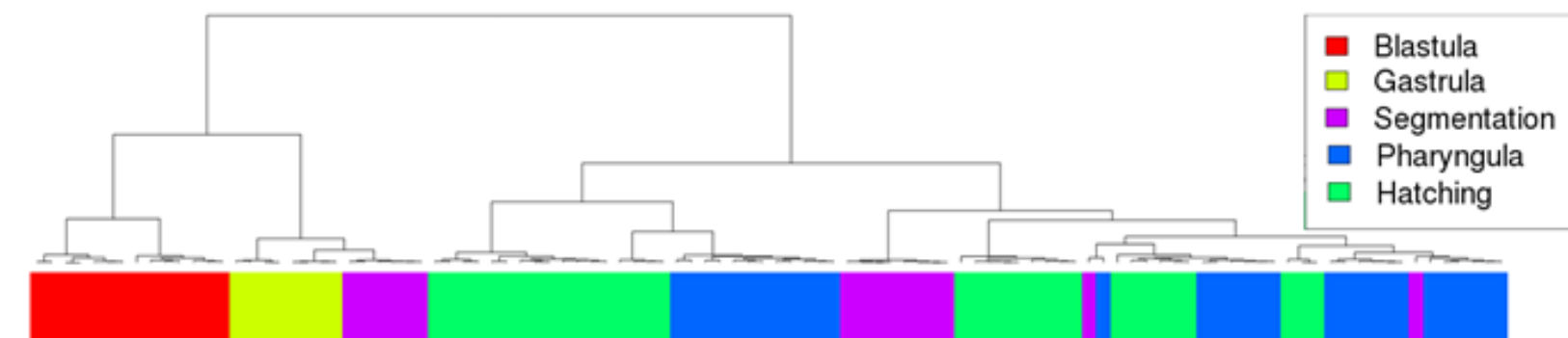


Biplot



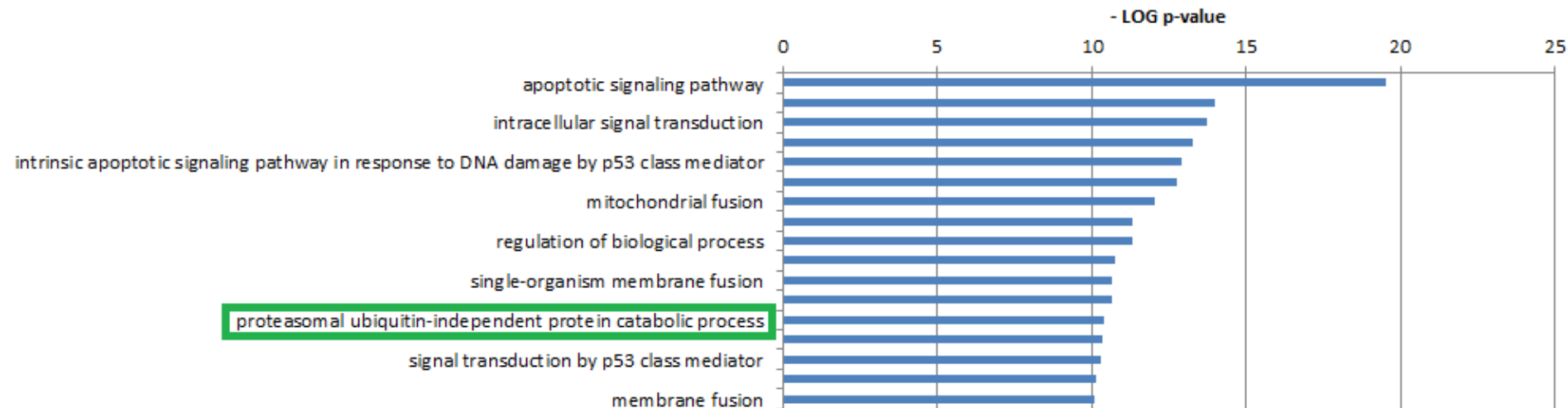
Hierarchical clustering

Samples labeled by
treatment



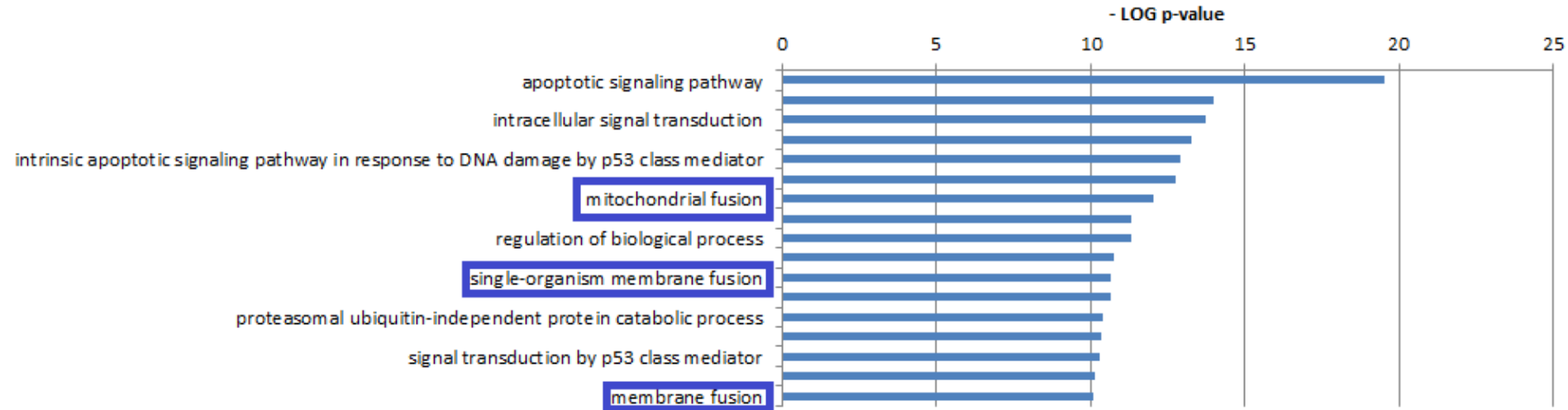
GSEA supplementary

BIOLOGICAL PROCESS



psme4b - proteasome (prosome, macropain) activator subunit 4b
psmb1 - proteasome (prosome, macropain) subunit, beta type, 1
psma6a - proteasome (prosome, macropain) subunit, alpha type, 6a
psmb2 - proteasome (prosome, macropain) subunit, beta type, 2
psma2 - proteasome (prosome, macropain) subunit, alpha type, 2
psma8 - proteasome (prosome, macropain) subunit, alpha type, 8
psmb7 - proteasome (prosome, macropain) subunit, beta type, 7

BIOLOGICAL PROCESS

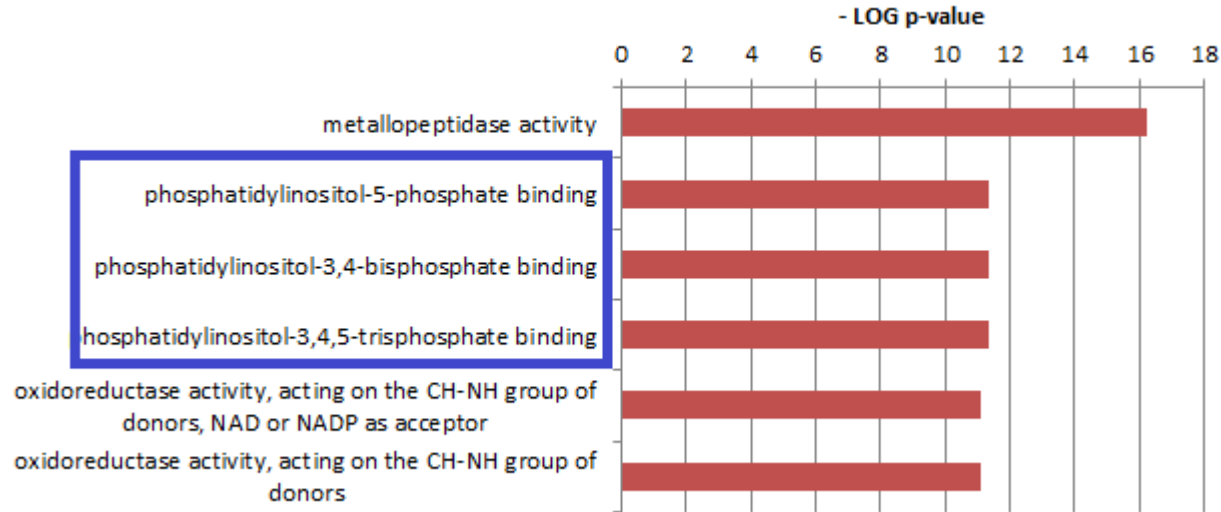


mfn2 - mitofusin 2

bmb - brambleberry

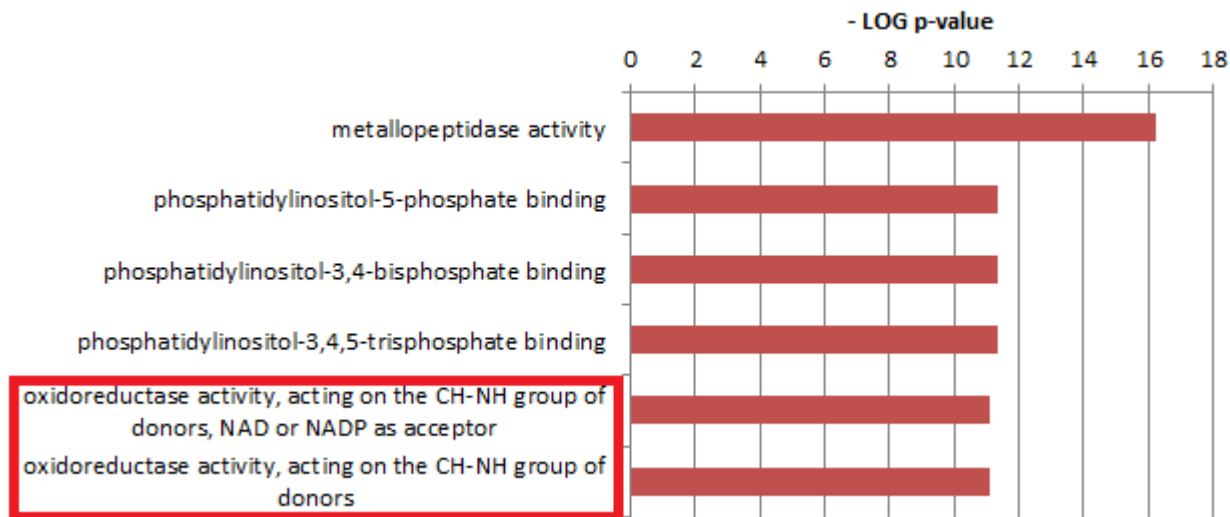
baxa - bcl2-associated x protein, a

MOLECULAR FUNCTION



phlda3 - pleckstrin homology-like domain, family a, member 3

MOLECULAR FUNCTION



mthfd2 - methylenetetrahydrofolate dehydrogenase (nadp+ dependent) 2, methenyltetrahydrofolate cyclohydrolase

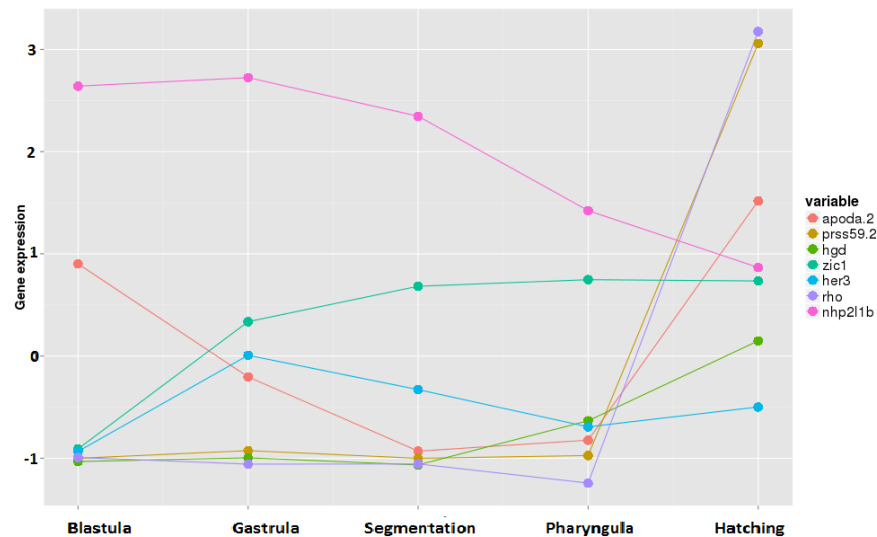
dhfr - dihydrofolate reductase

mthfd1b - methylenetetrahydrofolate dehydrogenase (nadp+ dependent) 1b

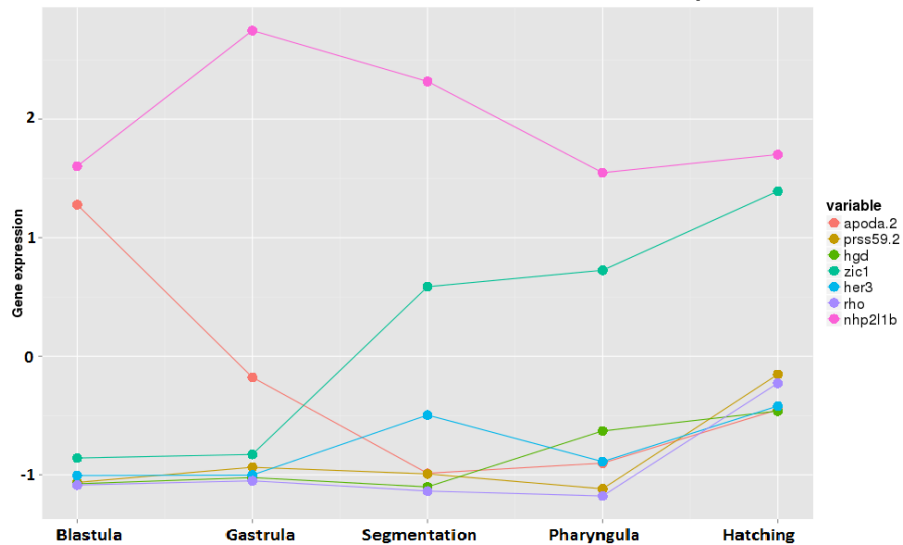
aldh1l2 - aldehyde dehydrogenase 1 family, member 12

STAGE EFFECT (1)

Wild type

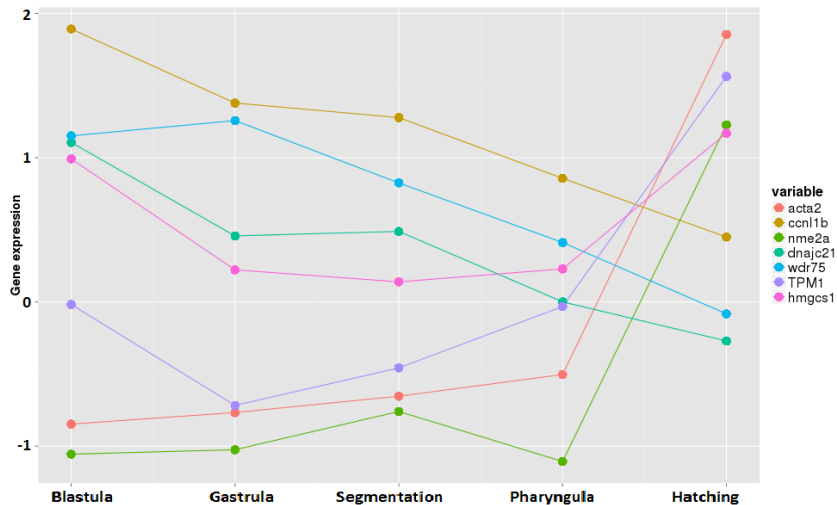


Morphant

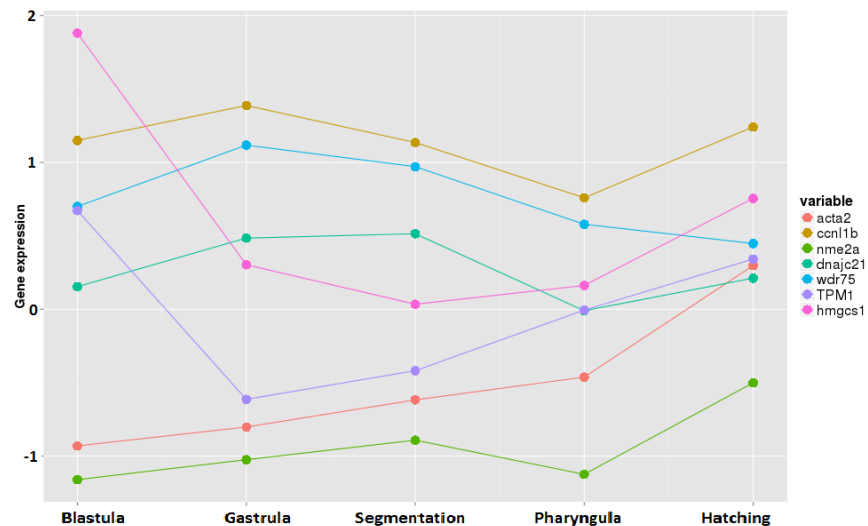


STAGE EFFECT (2)

Wild type

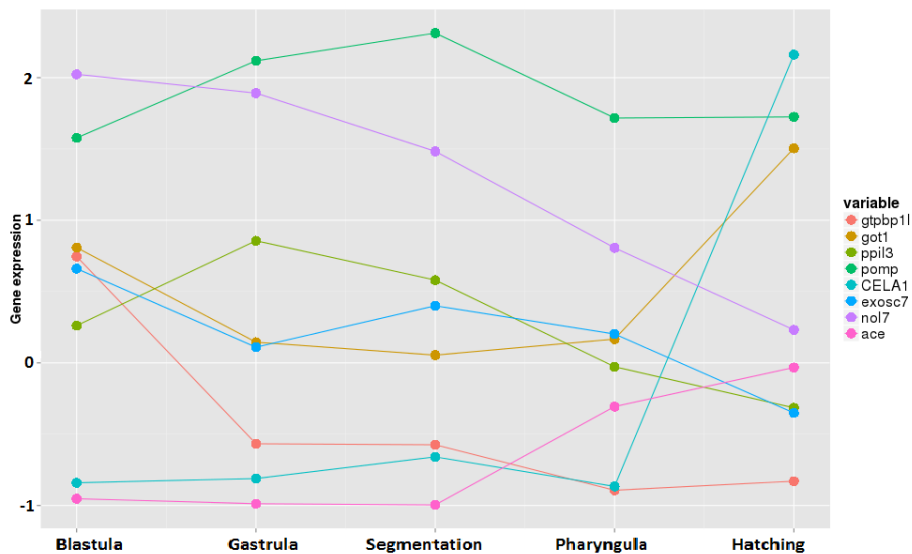


Morphant

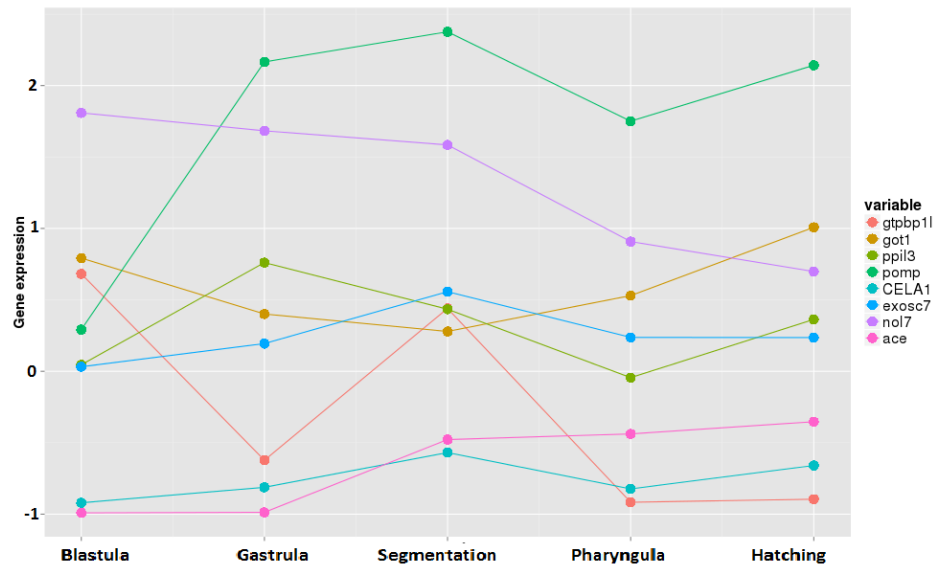


STAGE EFFECT (3)

Wild type



Morphant

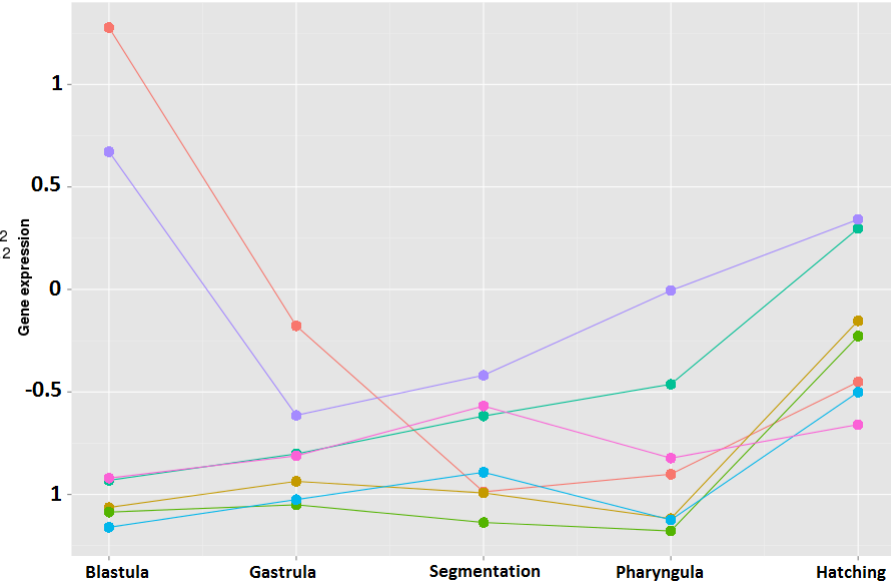
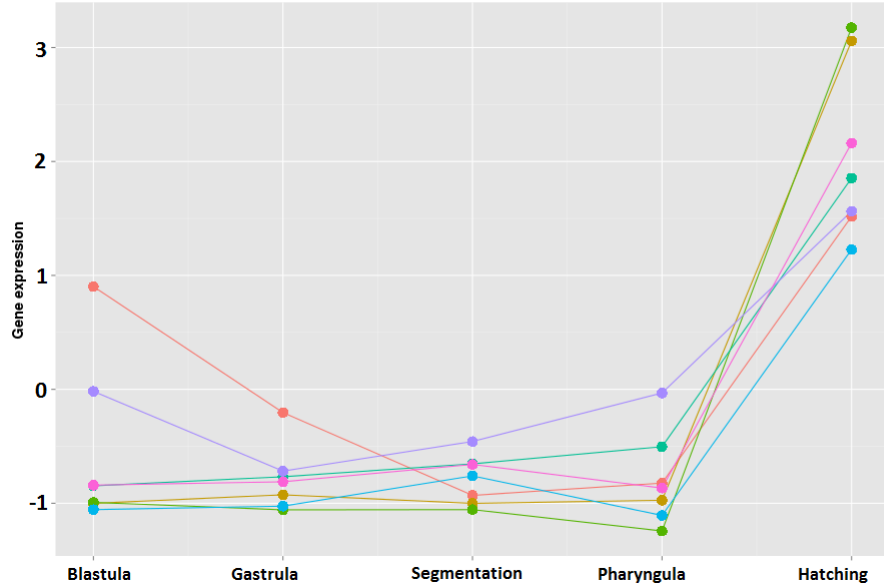


STAGE DEPENDENT EFFECT (pattern)

Wild-type

$p < 0.001$

Morphant



apoda.2 - apolipoprotein, duplicate 2

prss59.2 - protease, serine, 59, tandem duplicate 2

rho - rhodopsin

acta - actin alpha 2

nme2a - nucleoside diphosphate kinase

TPM1 - tropomyosin 1 (alpha)

CELA1 - Chymotrypsin-Like Elastase Family, Member 1

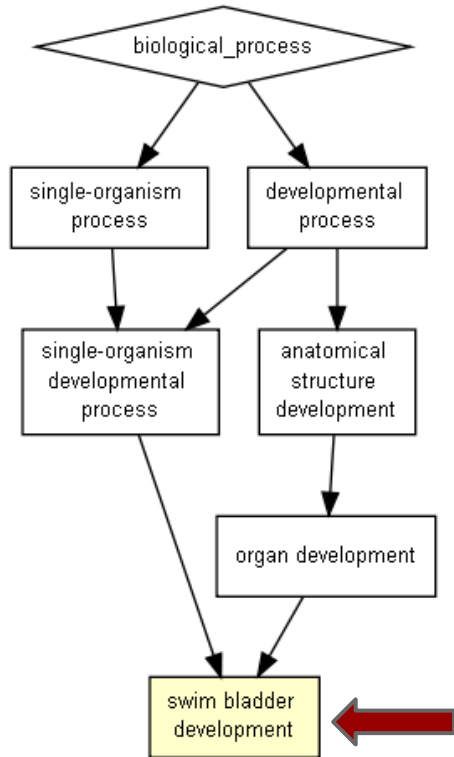
Introduction

Material and Methods

Results

Discussion and conclusions

Control vs WT



Feature selection performed on 31 samples (14 WT and 17 CO)

Control - GeneTools Standard Control Oligo

Samples from Pharyngula and Hatching stages

Biological process

Mass Spectrometry

genes: vegfaa (24 hpf), dll4 (24 hpf), tnnt2 (36 hpf)

3 biological replicates for morphants

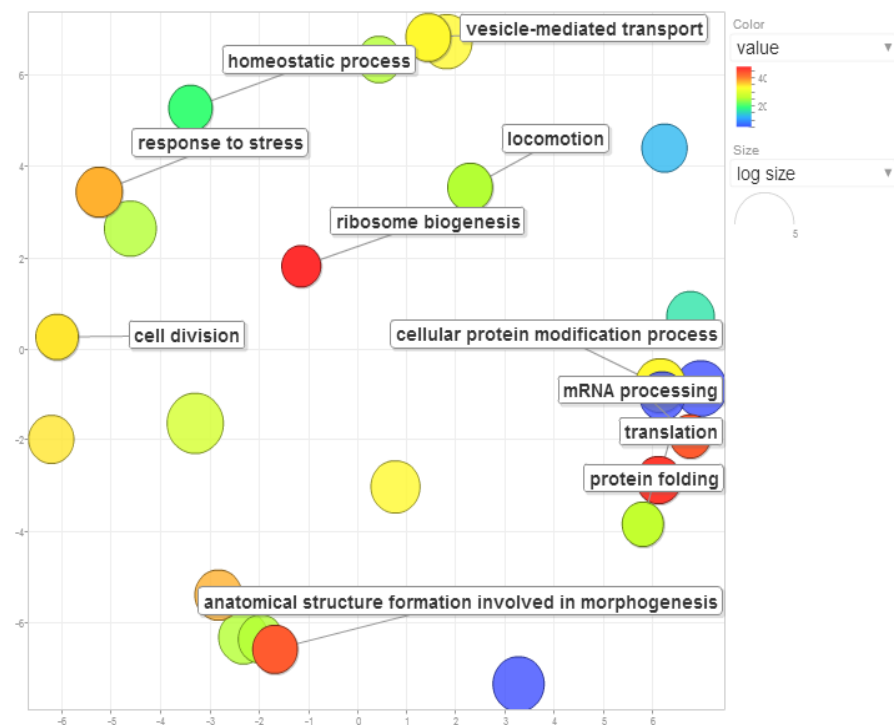
wild type as controls - 3 replicates for 24 hpf, 3 replicates for 36 hpf

24 hpf

Upregulated in WT



Upregulated in MO

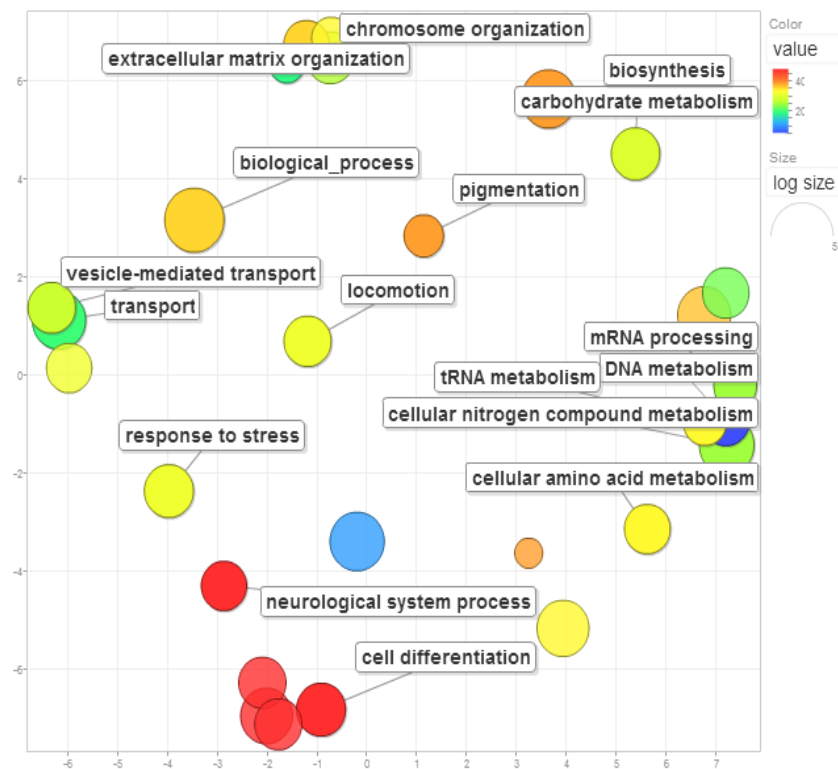


36 hpf

Upregulated in WT



Upregulated in MO



extra slides

PROBLEM: how to identify differentially expressed genes on a big dataset?

Dalman et al. *BMC Bioinformatics* 2012, **13**(Suppl 2):S11
<http://www.biomedcentral.com/1471-2105/13/S2/S11>



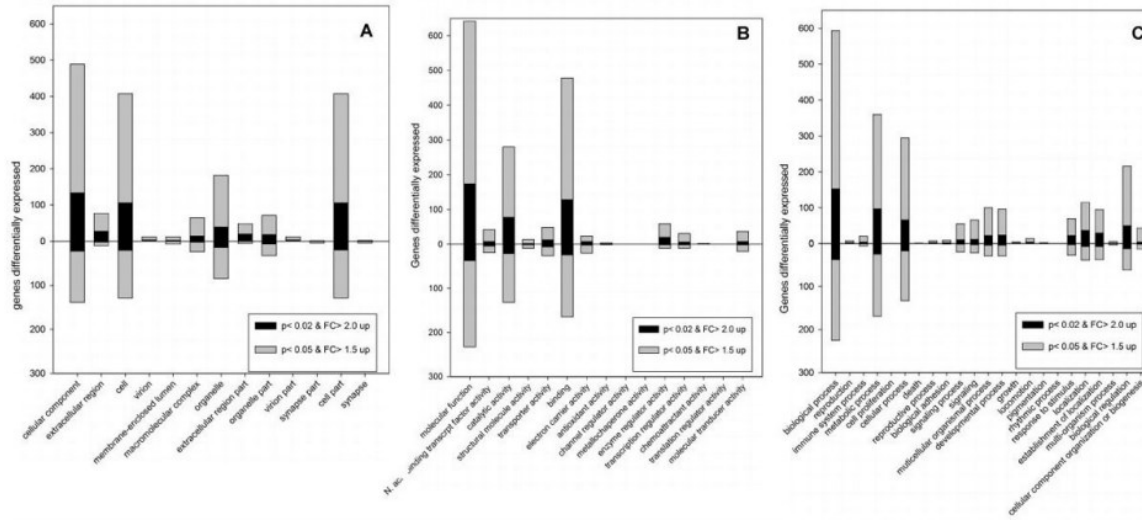
PROCEEDINGS

Open Access

Fold change and p-value cutoffs significantly alter microarray interpretations

Mark R Dalman^{1*}, Anthony Deeter², Gayathri Nimishakavi², Zhong-Hui Duan²

From Great Lakes Bioinformatics Conference 2011
Athens, OH, USA. 2-4 May 2011



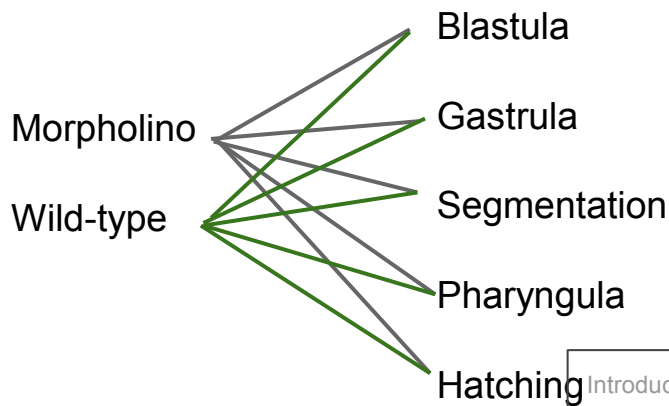
Which are the optimal fold change and p-value thresholds?

How to deal with high variability in the integrated gene expression dataset?

STAGE DEPENDENT EFFECT

INGREDIENTS:

- One measurement variable (microarray gene intensity)
- Two nominal variables (Treatment and Developmental stage)
- Each value of one nominal variable in combination with each value of the other nominal variable



Two-way ANOVA

adapted for microarrays datasets

Interaction effect

Data collection

NCBI Resources How To kristina1987 My NCBI Sign Out

GEO DataSets GEO DataSets (zebrafish morpholino OR zebrafish knockdown) AND "Danio rerio"[porgn: __txid7955] Search

Show additional filters Display Settings: Summary, 20 per page, Sorted by Default order Send to: Filters: Manage Filters

Clear all

Entry type
DataSets (2)
Series (72)

Organism
Select ...

Study type
clear
✓ Expression profiling by array
More ...

Author
Select ...

Attribute name
tissue
strain
More ...

Publication dates
1 year
Custom range...

Clear all

Show additional filters

Results: 1 to 20 of 74
Filters activated: Expression profiling by array Clear all

1. [Transcription factor Ovo1 depletion effect on 12 hour post-fertilization embryos](#)
Analysis of wild-type embryos injected with Ovo1 morpholino antisense oligos. At 12 hpf, the Ovo1 morphants begin showing neural crest (NC) defects wherein a subset of NC cells aggregates in the dorsal midline above the neural tube. Results provide insight into the role of Ovo1 in NC migration.
Organism: **Danio rerio**
Type: **Expression profiling by array, count, 2 genotype/variation sets**
Platform: GPL1319 Series: GSE21539 6 Samples
Download data: GEO (CEL)
DataSet Accession: GDS3719 ID: 3719
[PubMed](#) [Full text in PMC](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)

2. [Homeobox protein Hoxb1a deficiency effect on the embryonic hindbrain](#)
Analysis of hindbrain rhombomere 4 tissues of embryos deficient for the homeobox protein Hoxb1a. Hoxb1a confers regional identity to rhombomere 4 in hindbrain development. Results provide insight into the mechanisms controlling the development of regional identity in the hindbrain.
Organism: **Danio rerio**
Type: **Expression profiling by array, count, 2 strain, 2 tissue sets**
Platform: GPL3721 Series: GSE5199 9 Samples
Download data: GEO
DataSet Accession: GDS2575 ID: 2575
[PubMed](#) [Similar studies](#) [GEO Profiles](#) [Analyze DataSet](#)

3. [The functional study of zebrafish rig1 gene in NNV infection](#)
(Submitter supplied) Transcriptional profiling of ZF-4 cells comparing controle morpholino-infection cells with

Top Organisms [Tree](#)
Danio rerio (74)
Homo sapiens (1)

Find related data
Database: Select
Find items

Search details
(((("zebrafish"[MeSH Terms] OR "Danio rerio"[Organism] OR zebrafish[All Fields]) AND morpholino[All Fields]) OR ((("zebrafish"[MeSH Terms] OR "Danio rerio"[Organism] OR

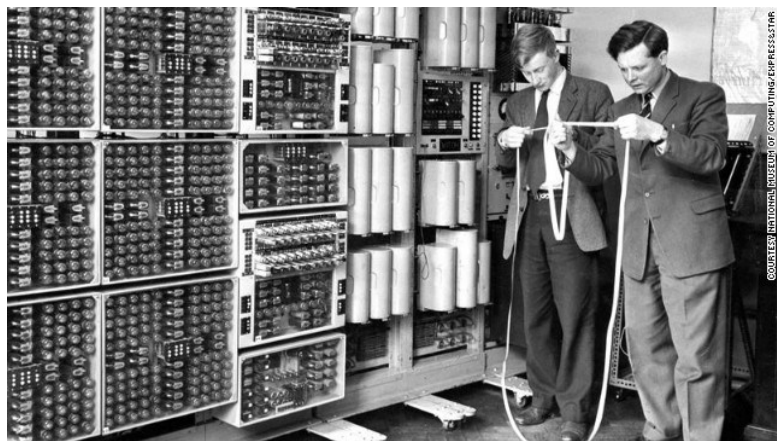
Search See more...

Recent activity
[Turn Off](#) [Clear](#)
Q (zebrafish morpholino OR zebrafish knockdown) AND "Danio rerio"[p GEO DataSets
Q (zebrafish morpholino OR zebrafish knockdown) AND "Danio rerio"[p GEO DataSets
Q (zebrafish morpholino OR zebrafish

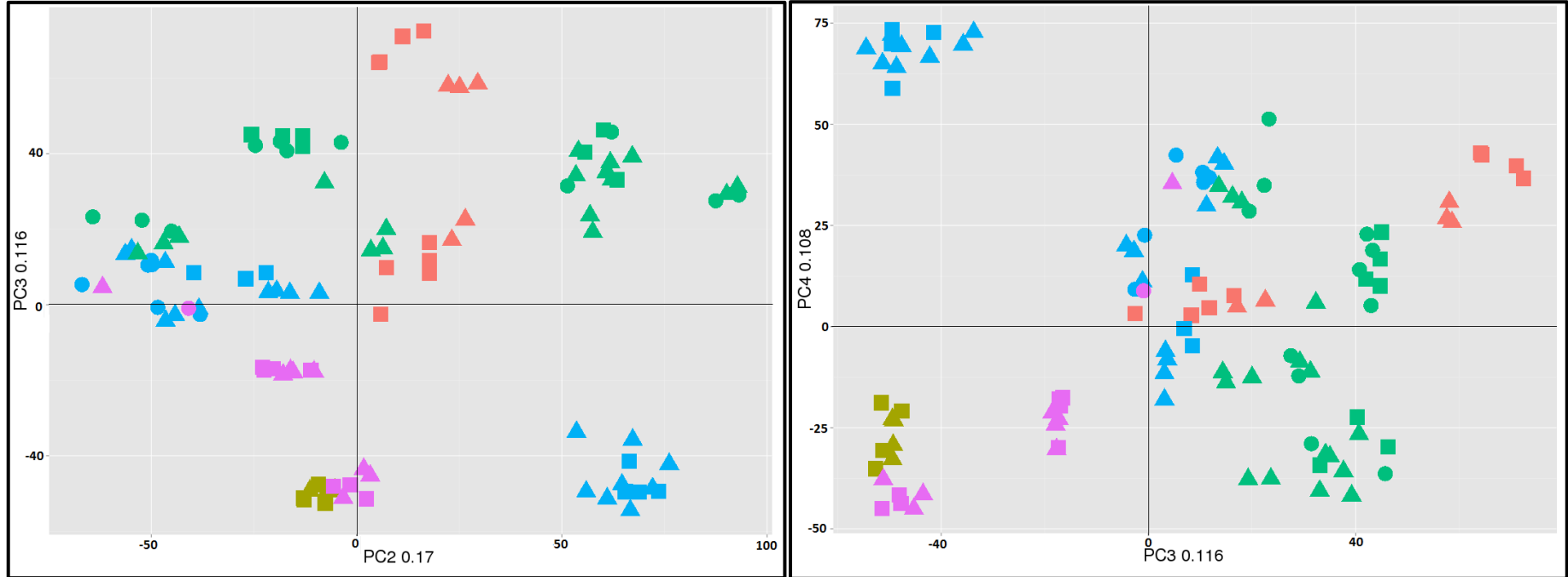


Gene Expression Omnibus

<http://www.ncbi.nlm.nih.gov/geo/>

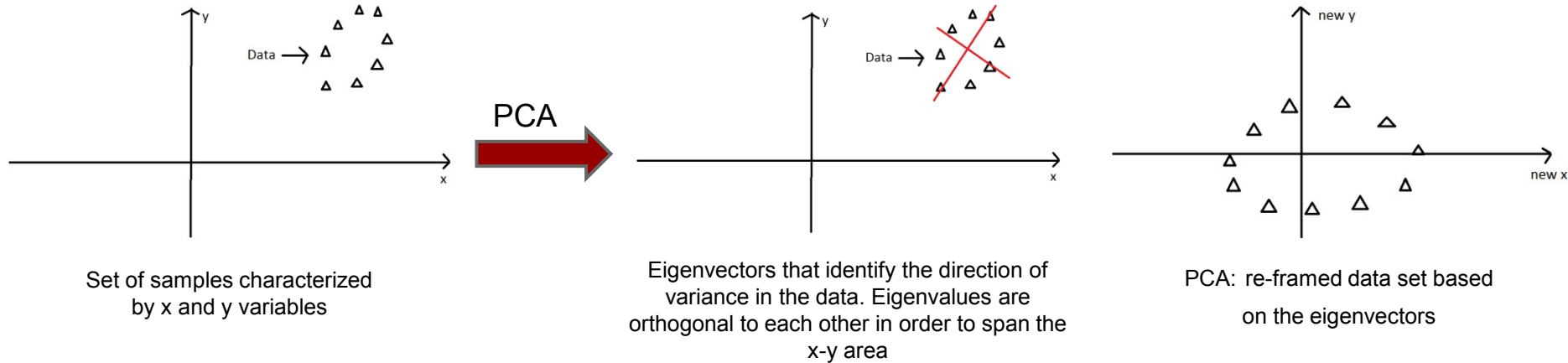


PRINCIPAL COMPONENT ANALYSIS



PCA does not highlight grouping for treatment: absent data pattern for the first five PCA components
The results are confirmed also with hierarchical clustering

PRINCIPAL COMPONENT ANALYSIS



Principal Component Analysis 4 Dummies: Eigenvectors, Eigenvalues and Dimension Reduction

Note that nothing has been done to the data itself. We're just looking at it from a different angle