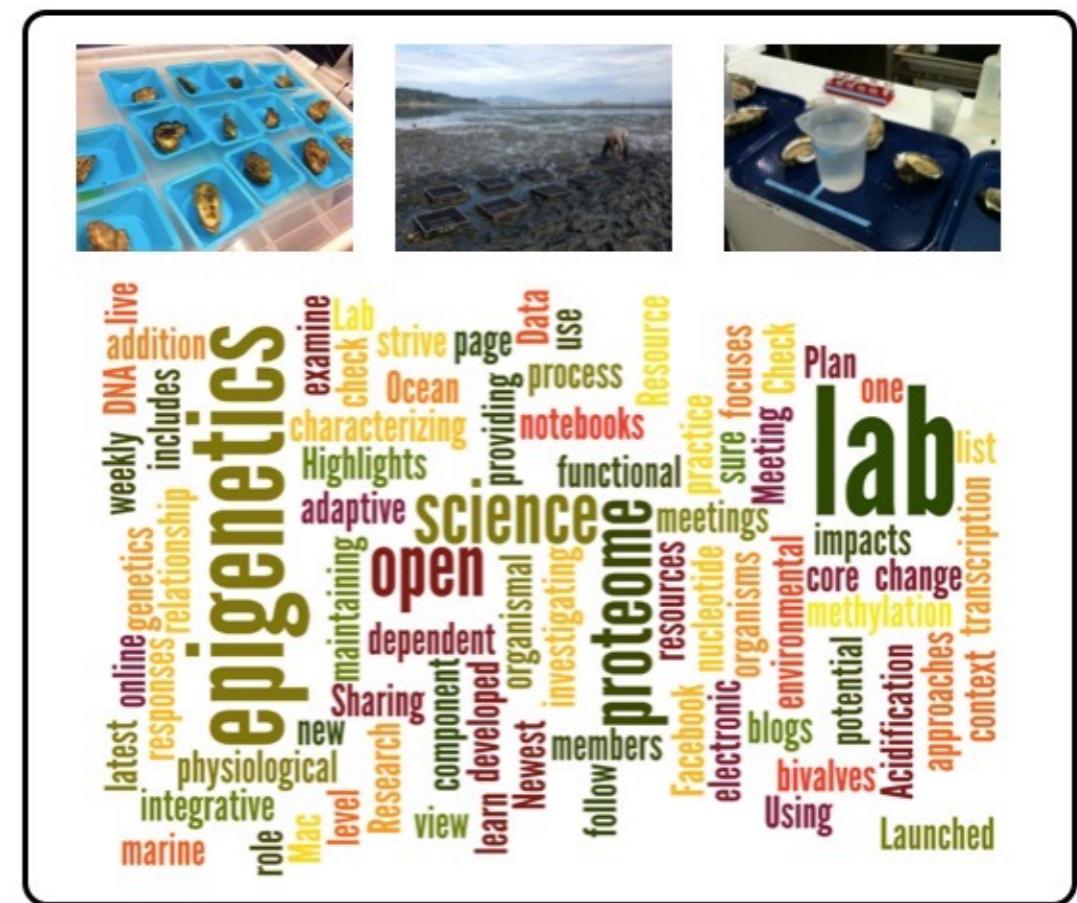


# Does DNA Methylation Facilitate Phenotypic Plasticity in Marine Invertebrates?

Steven Roberts  
Associate Professor  
University of Washington  
School of Aquatic and Fishery Sciences  
**robertslab.info**  
**@sr320**



- 19 October 2015



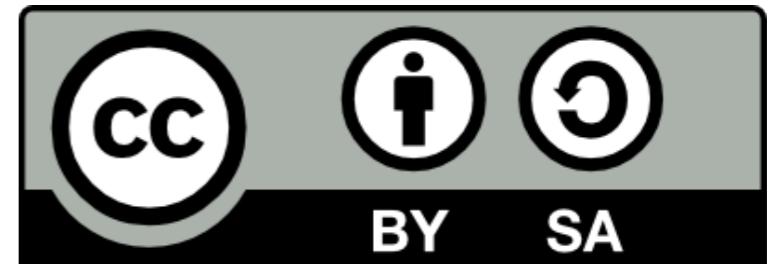
# Open Science

---

- You are free to Share!
- Our lab practices open notebook science



IP[y]: IPython  
Interactive Computing



**OpenNotebookScience**

@ONScience FOLLOWS YOU

Updates from #openscience notebooks,  
brought to you by @openscience. Start  
yours at [onsnetwork.org/joinons/](http://onsnetwork.org/joinons/)

[onsnetwork.org](http://onsnetwork.org)

- Data, Preprints, Proposals, Lab Meetings, Web Cams, Slidedecks

These slides plus links @  
[github.com/sr320/talk-CICESE-2015](https://github.com/sr320/talk-CICESE-2015)

# Background

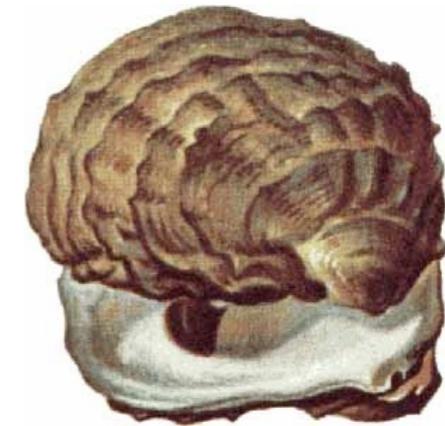
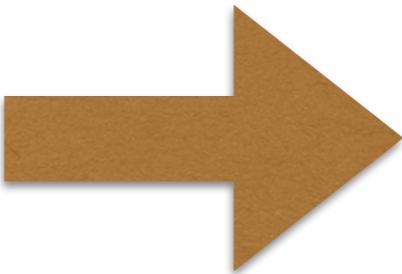
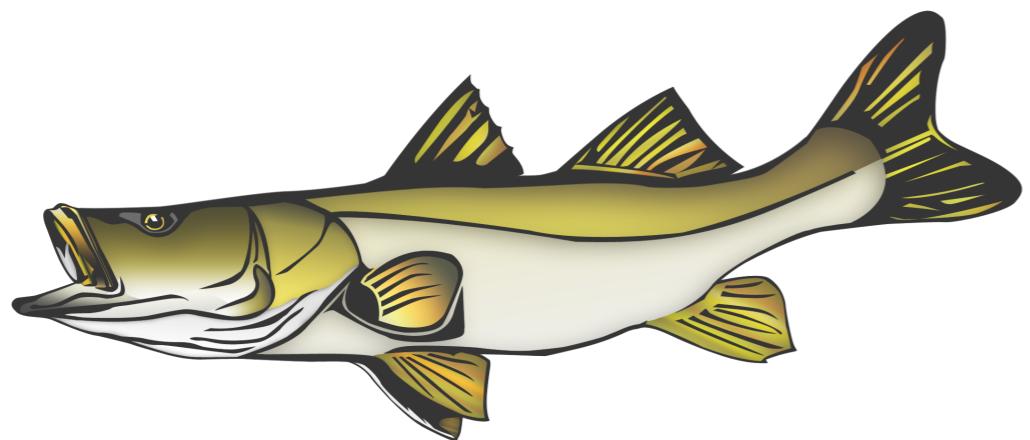
## Physiology

*How fundamental processes work in aquatic species*

# Background

## Physiology

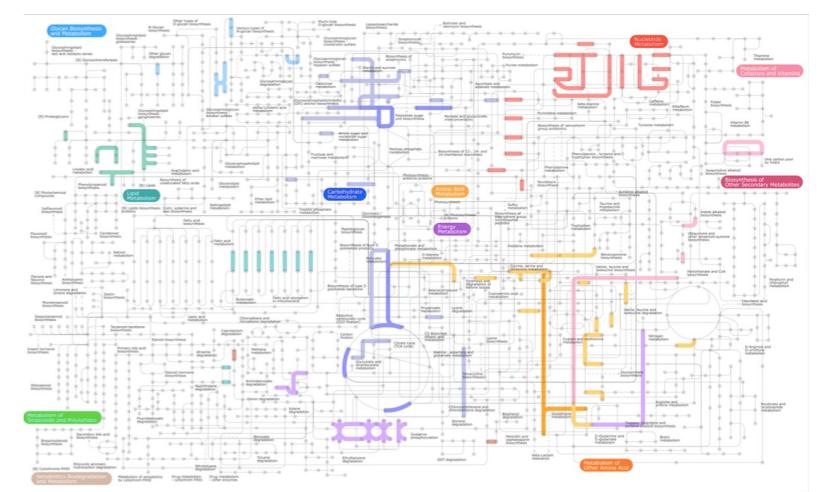
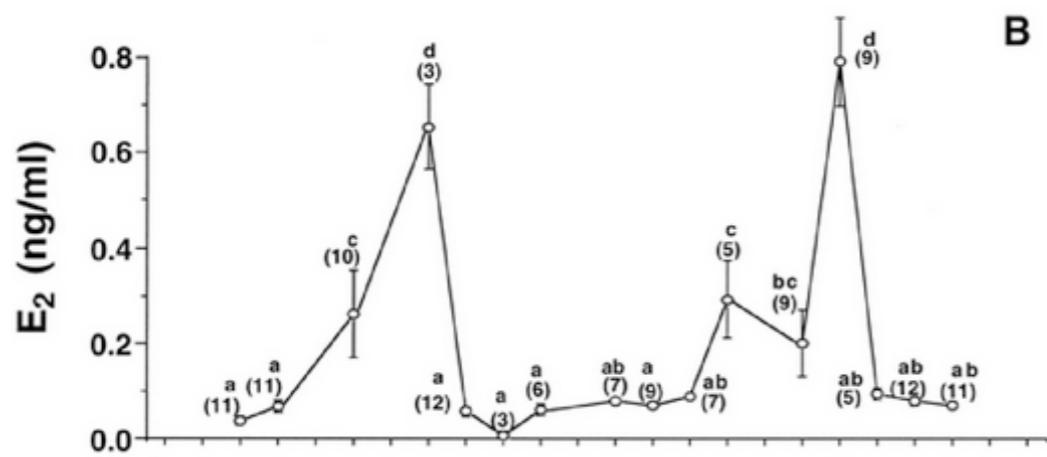
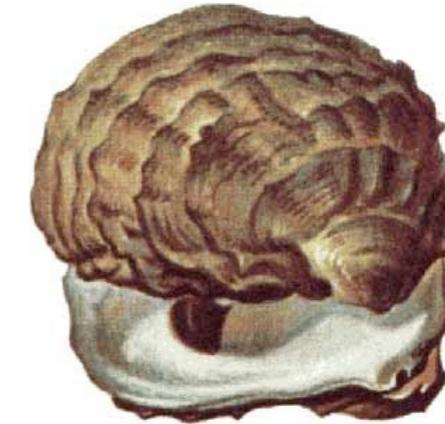
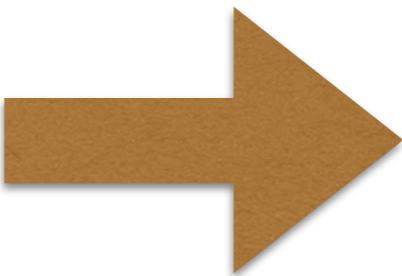
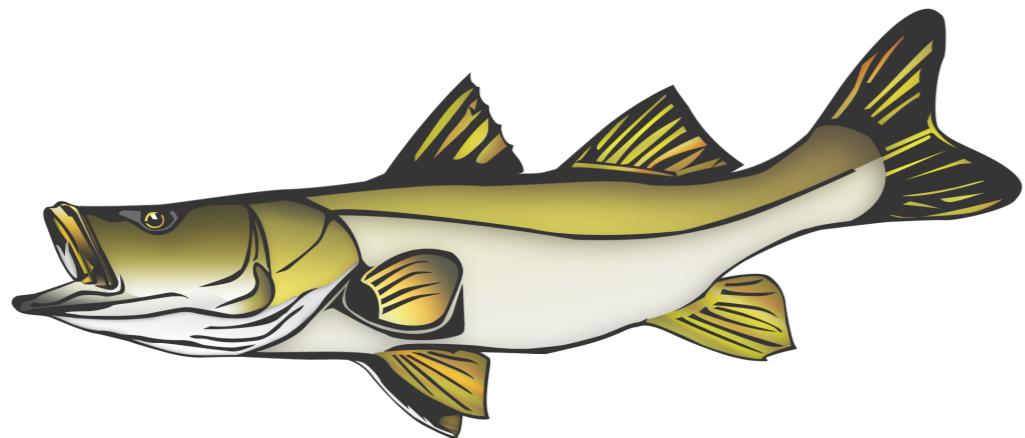
*How fundamental processes work in aquatic species*



# Background

## Physiology

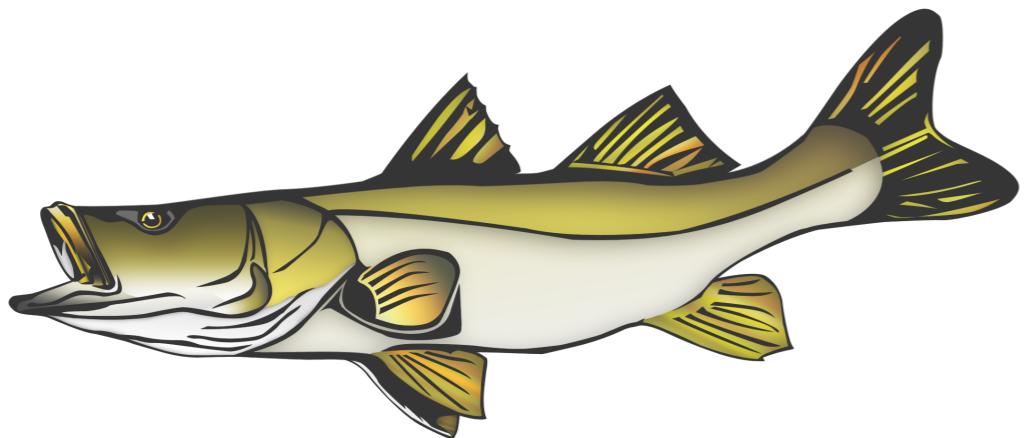
*How fundamental processes work in aquatic species*



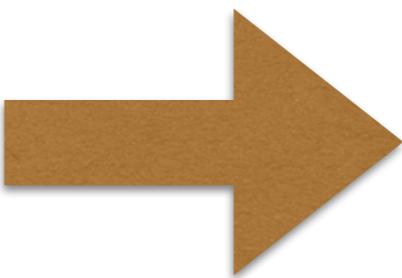
# Background

## Physiology

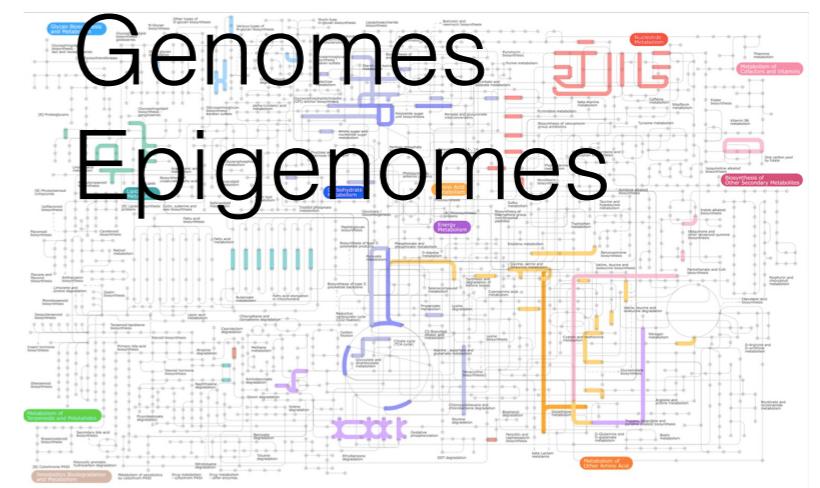
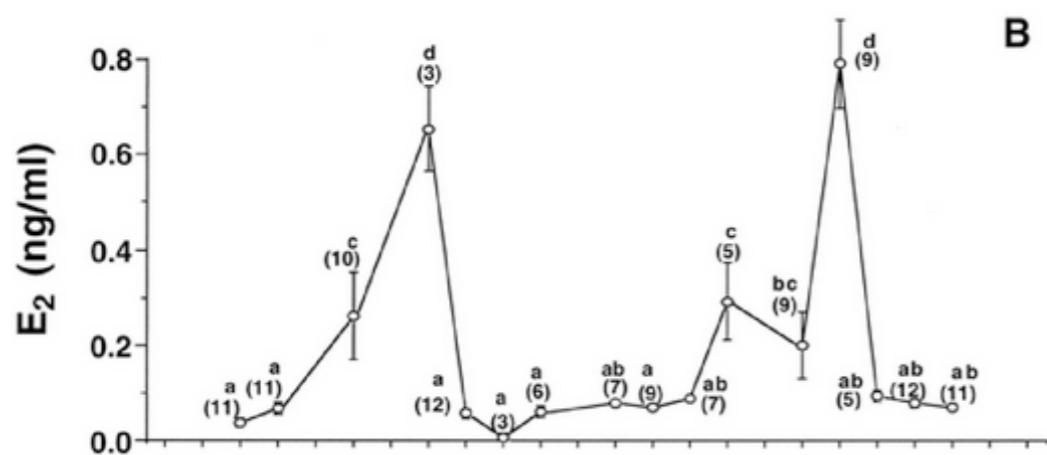
*How fundamental processes work in aquatic species*



Hormones  
Proteins



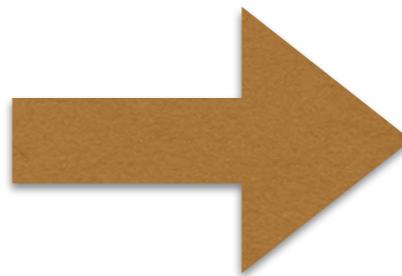
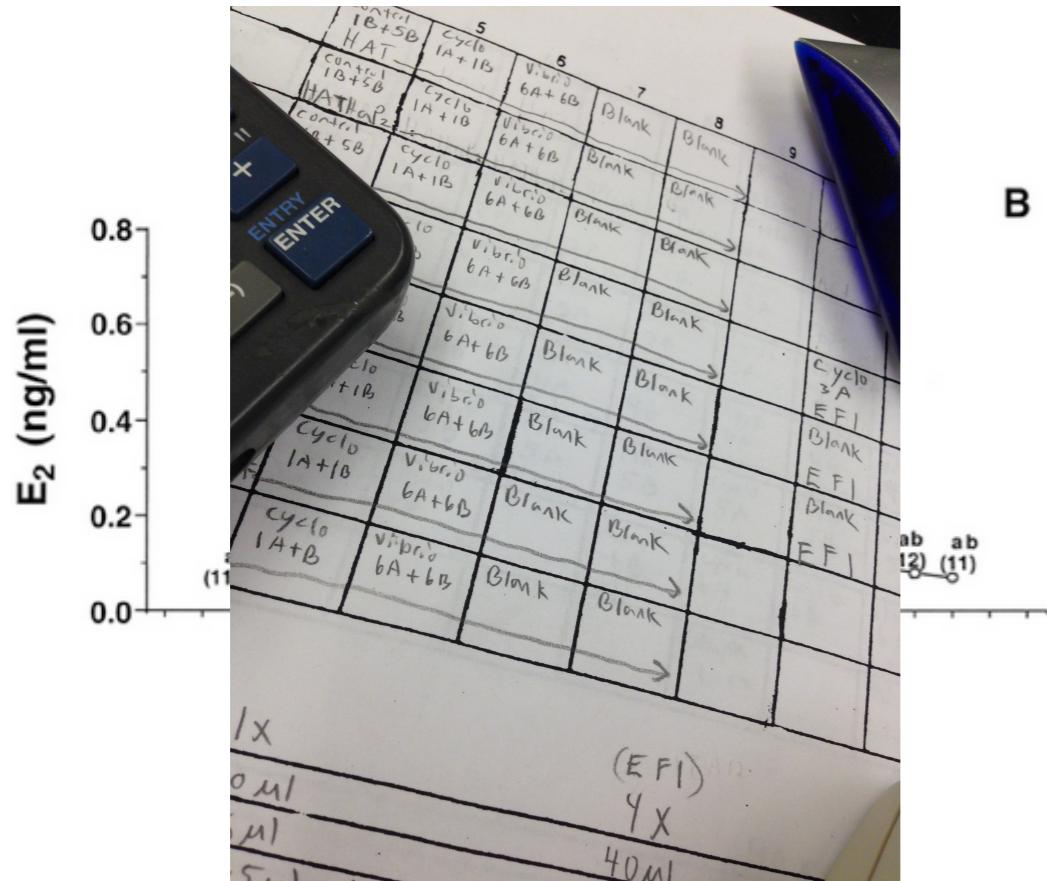
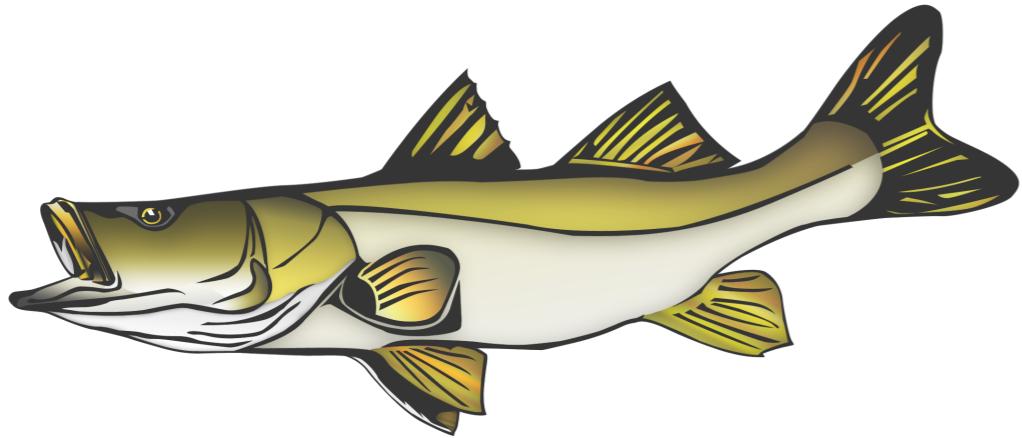
Transcriptomes  
Proteomes  
Genomes  
Epigenomes



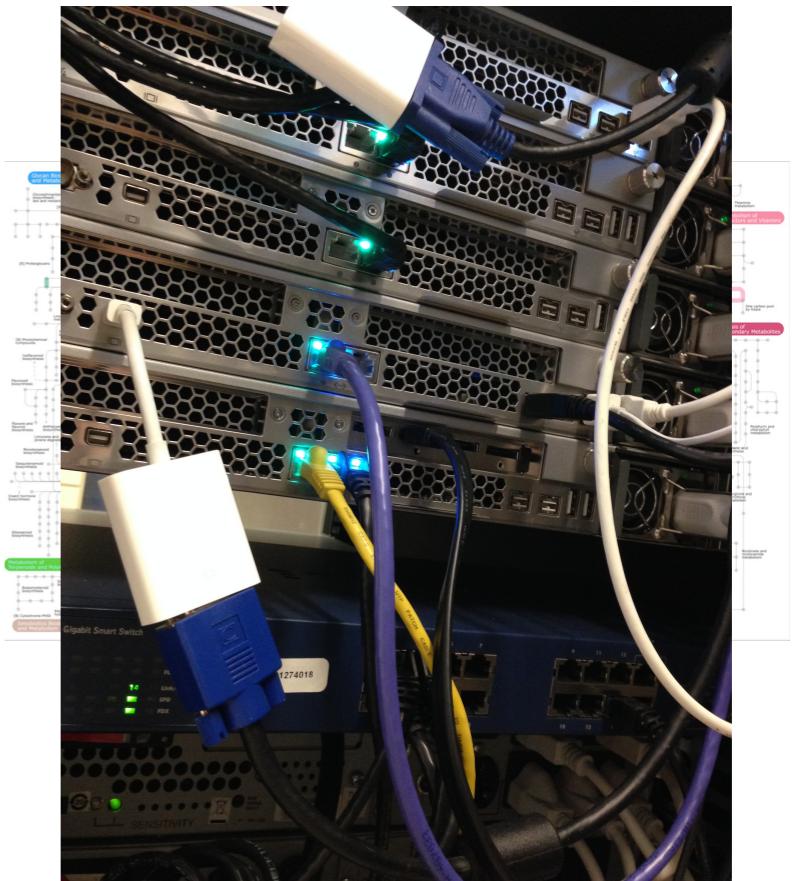
# Background

# Physiology

# *How fundamental processes work in aquatic species*



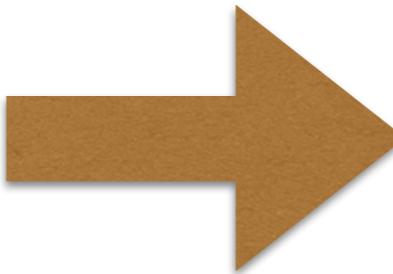
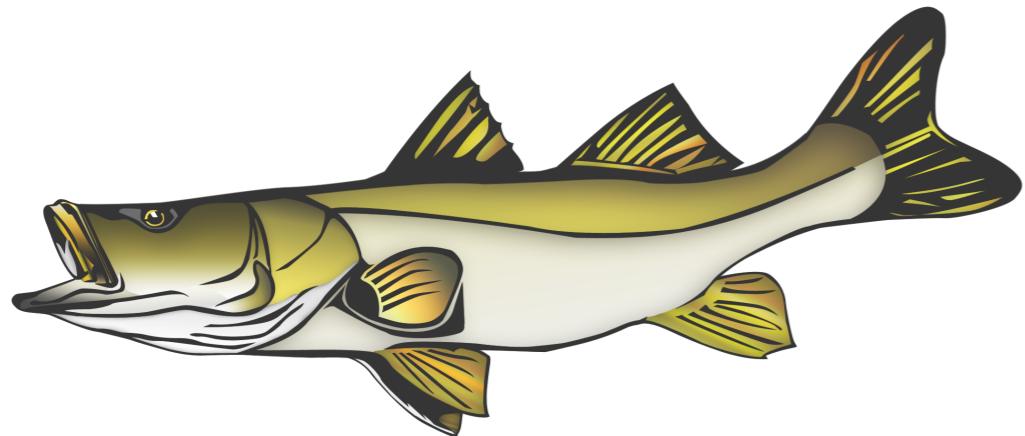
B



# Background

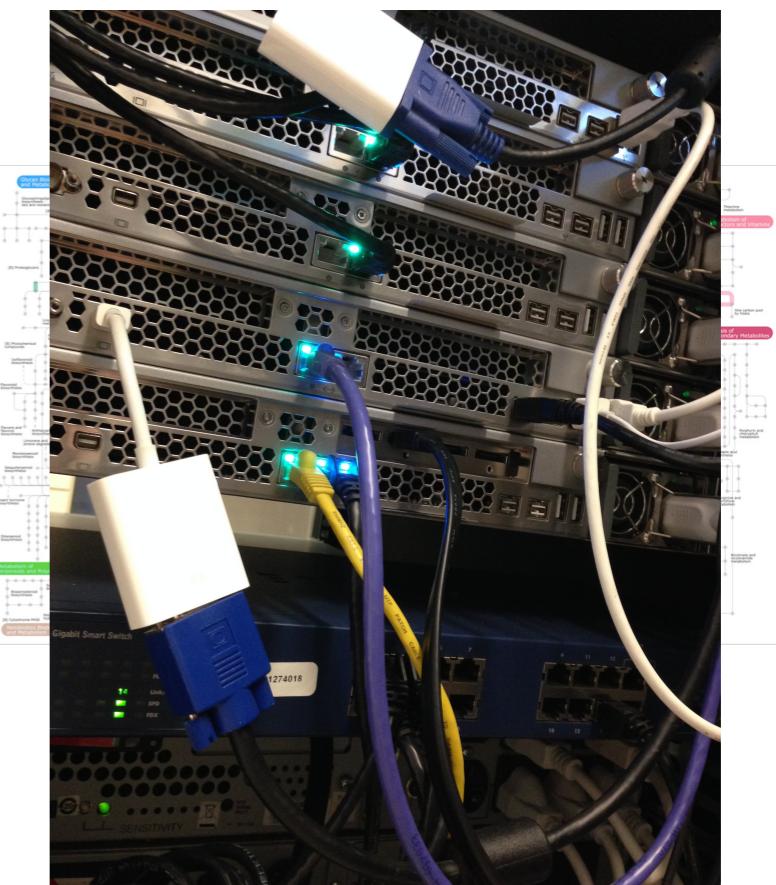
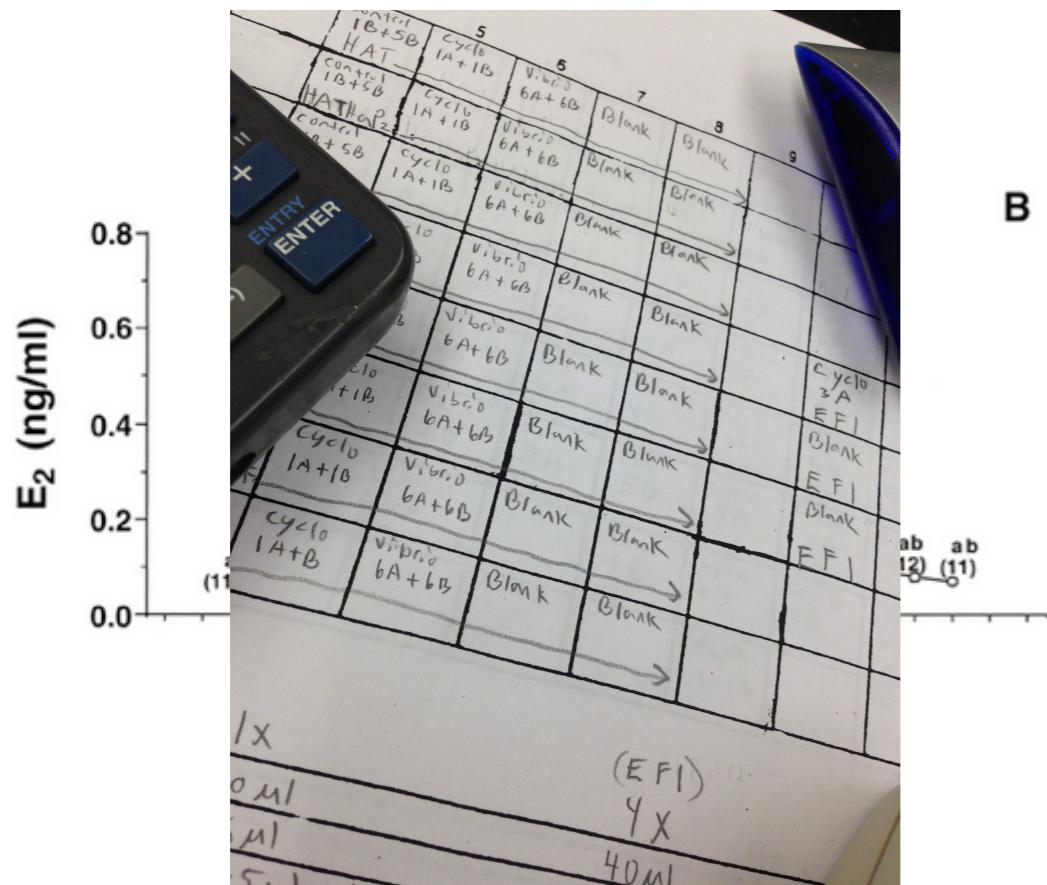
## Physiology

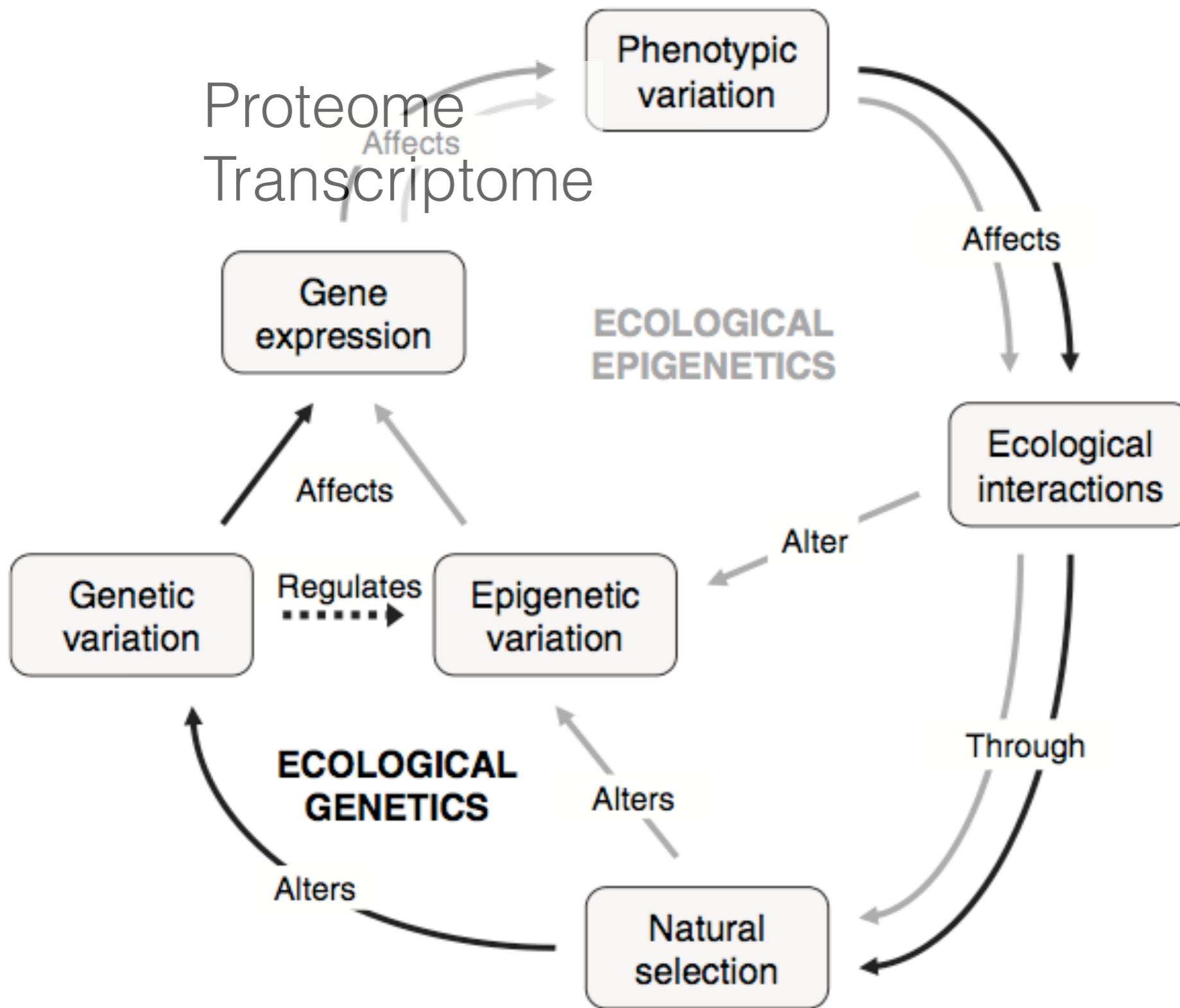
*How fundamental processes work in aquatic species*

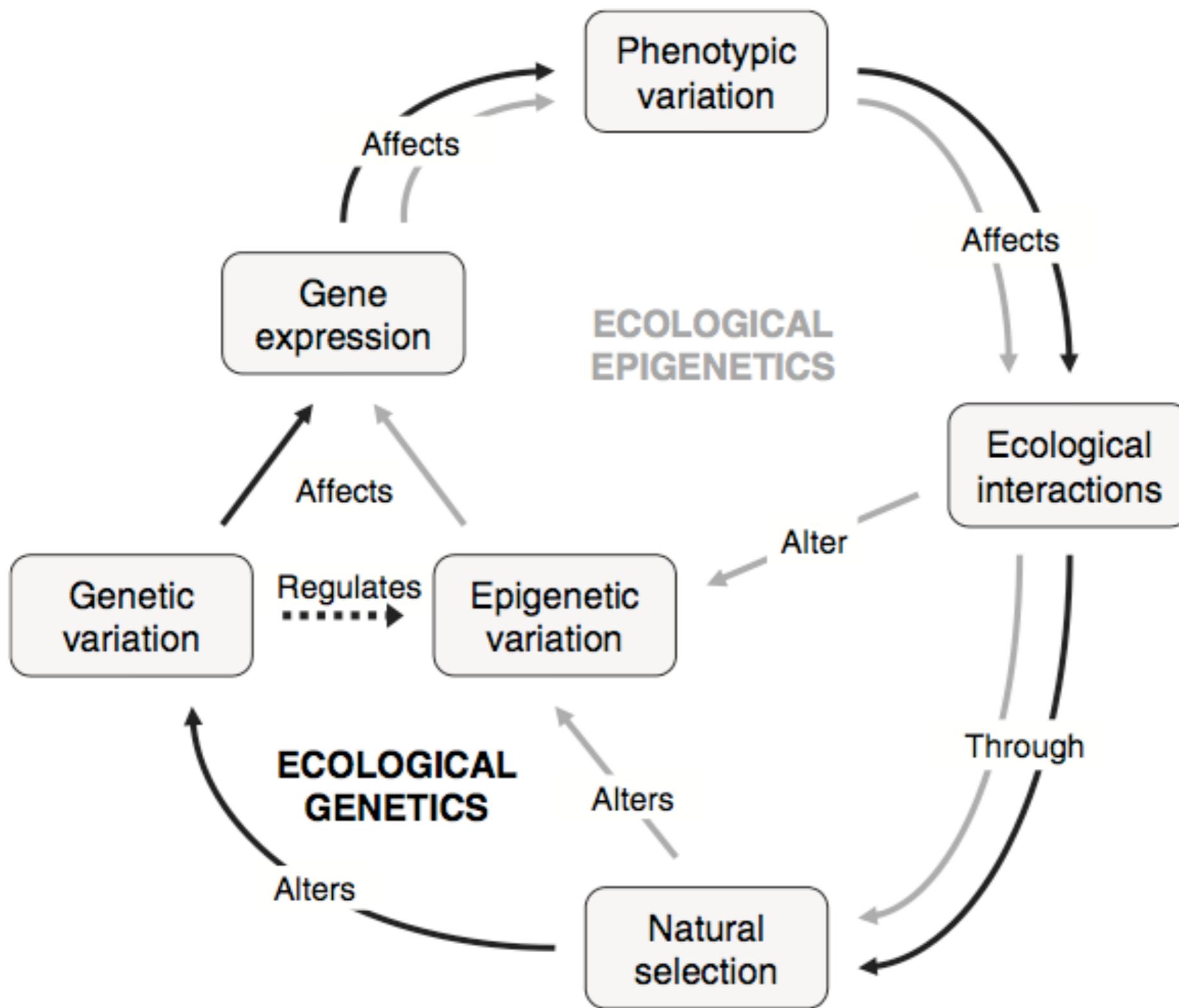


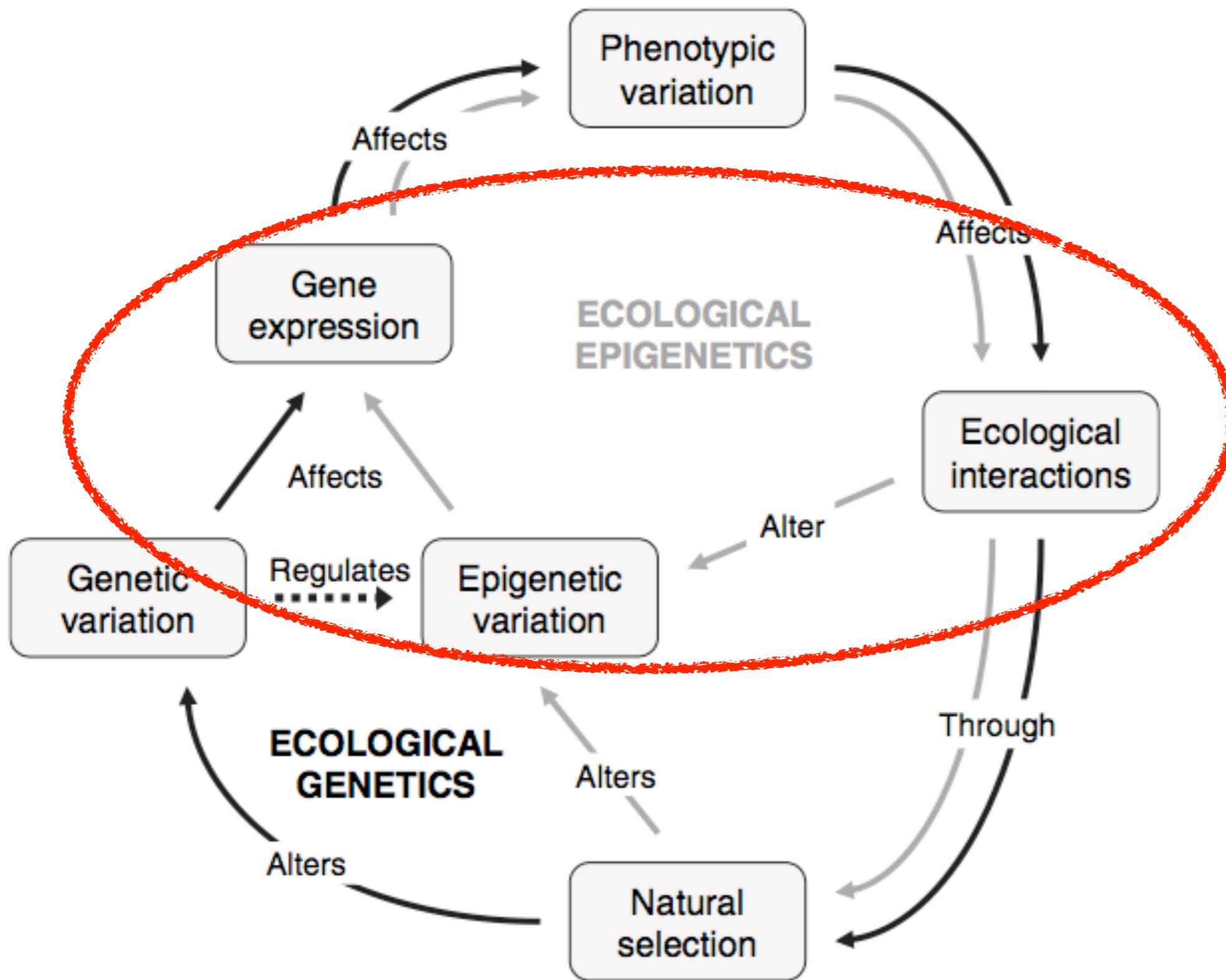
B

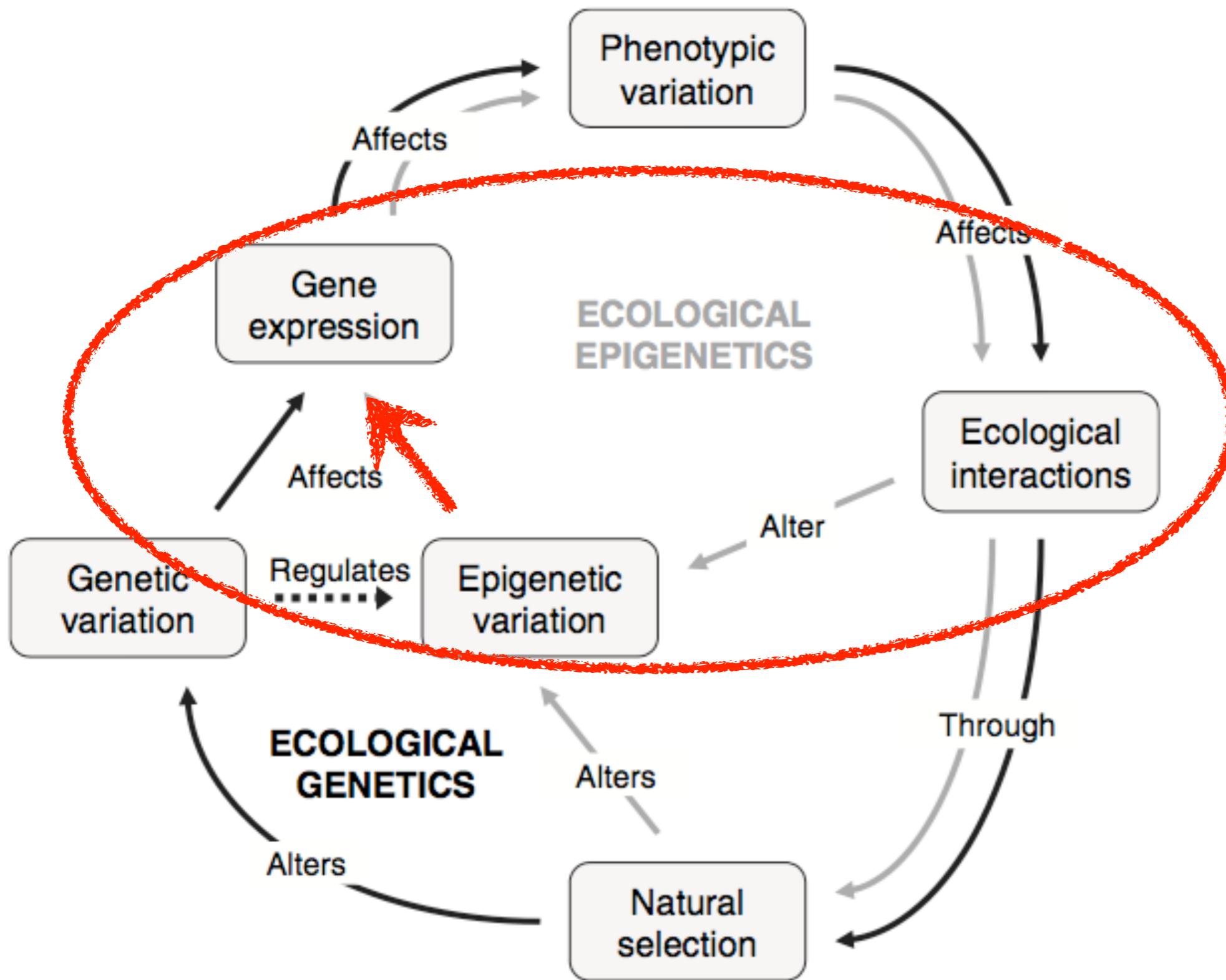
Changing  
Science



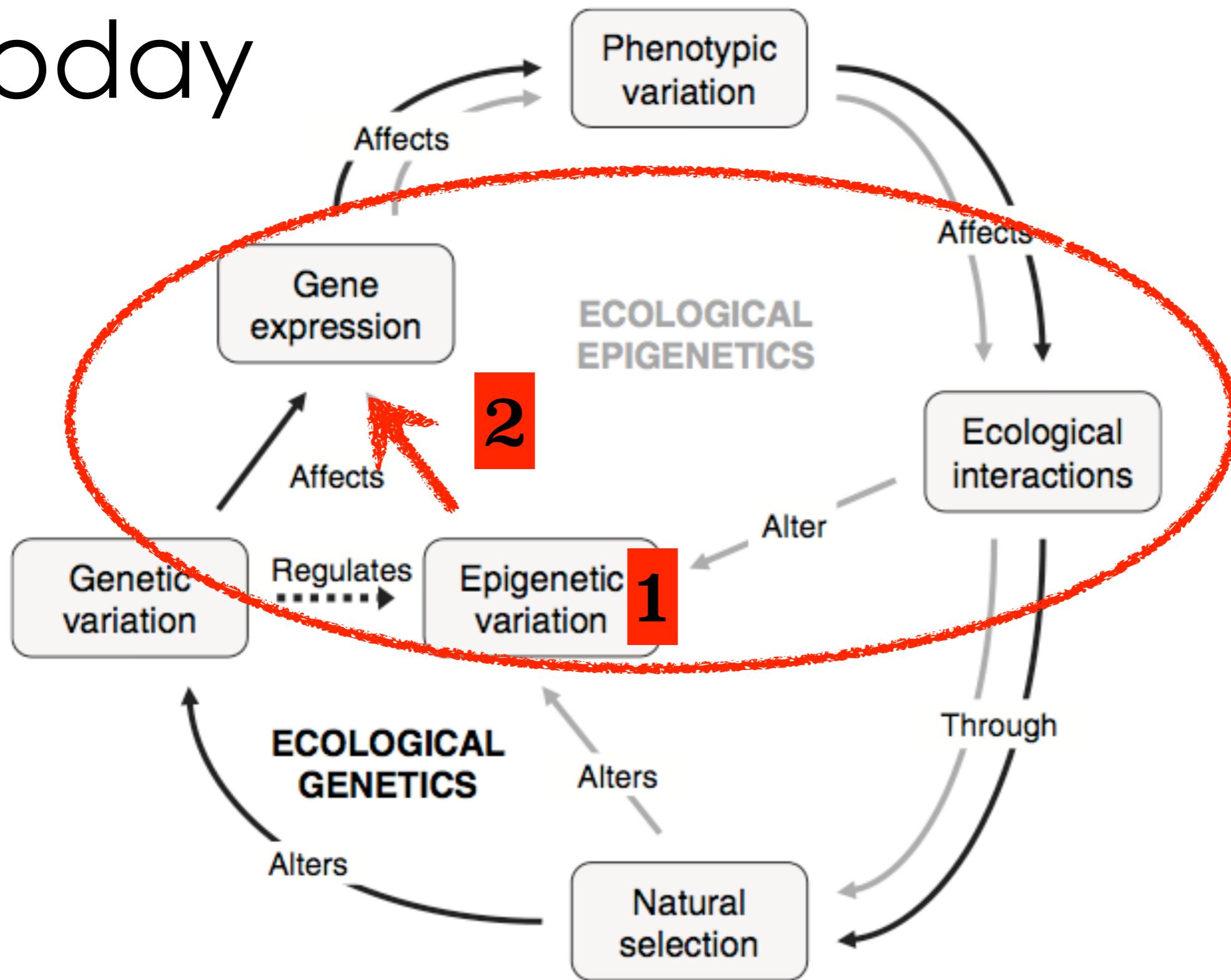








# Today



*Ecology Letters*, (2008) 11: 106–115

doi: 10.1111/j.1461-0248.2007.01130.x

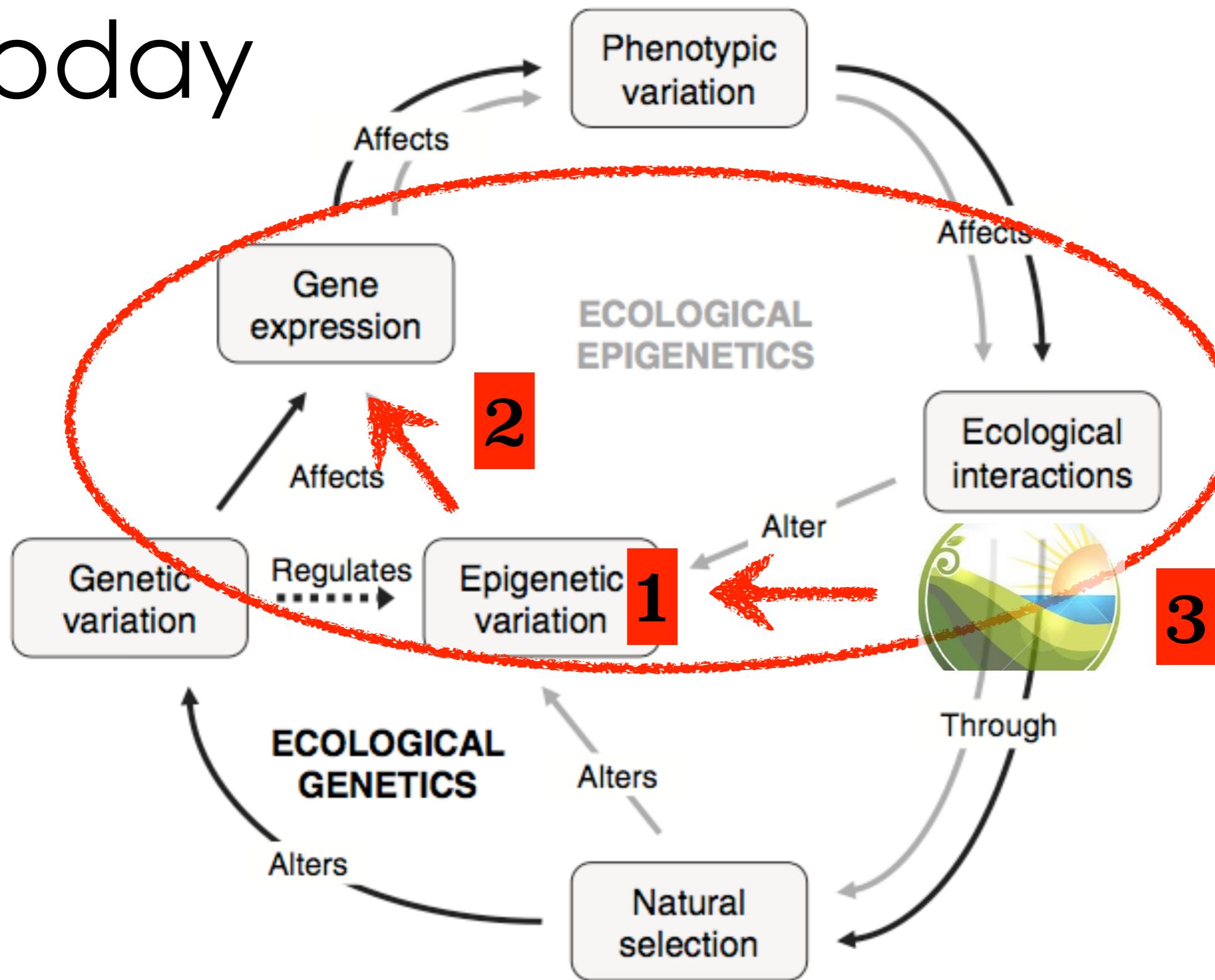
modified from

IDEA AND  
PERSPECTIVE

Epigenetics for ecologists

Oliver Bossdorf,<sup>1,\*</sup> Christina L.  
Richards<sup>2</sup> and Massimo Pigliucci<sup>3</sup>

# Today



*Ecology Letters*, (2008) 11: 106–115

doi: 10.1111/j.1461-0248.2007.01130.x

modified from

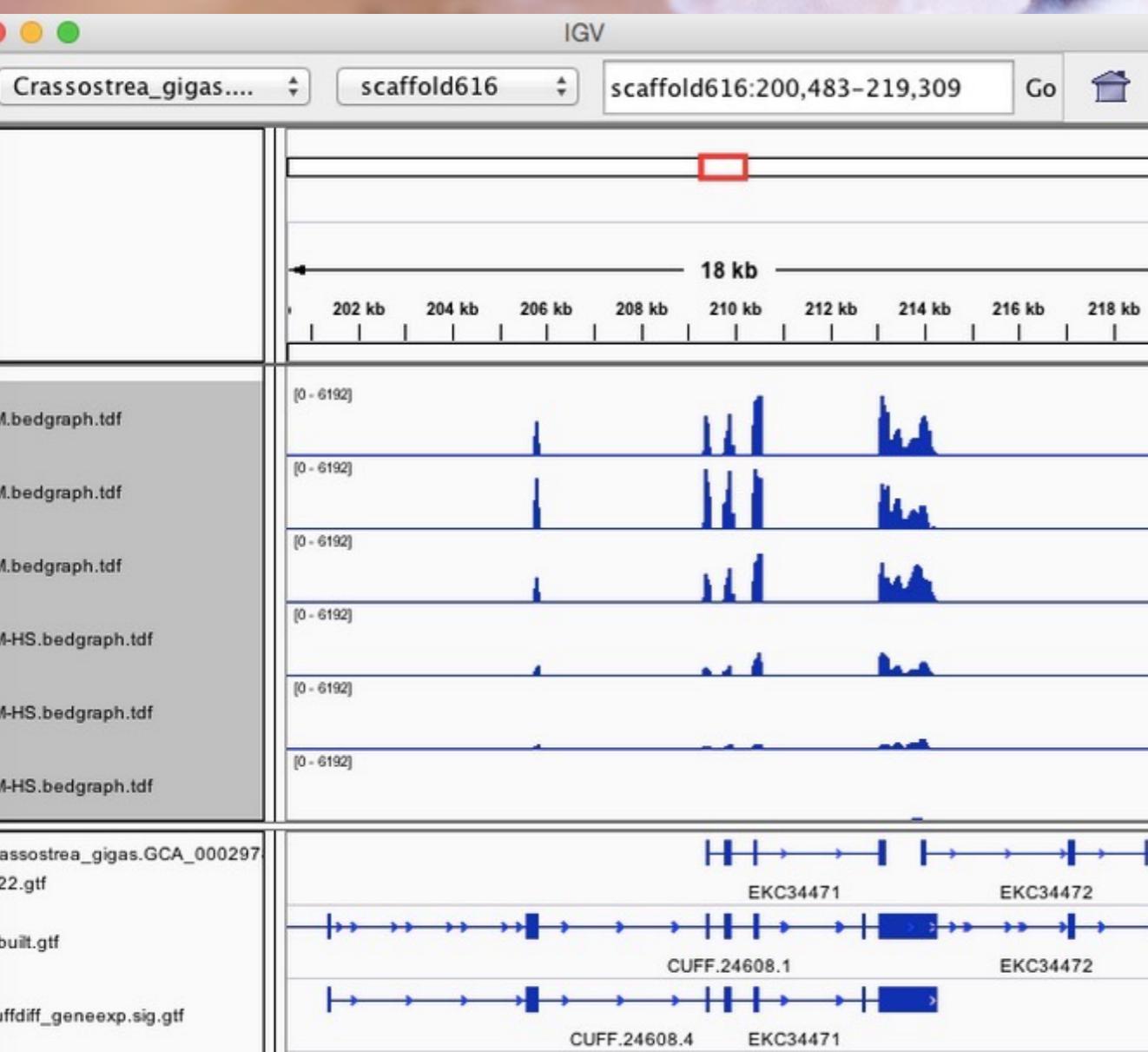
IDEA AND  
PERSPECTIVE

Epigenetics for ecologists

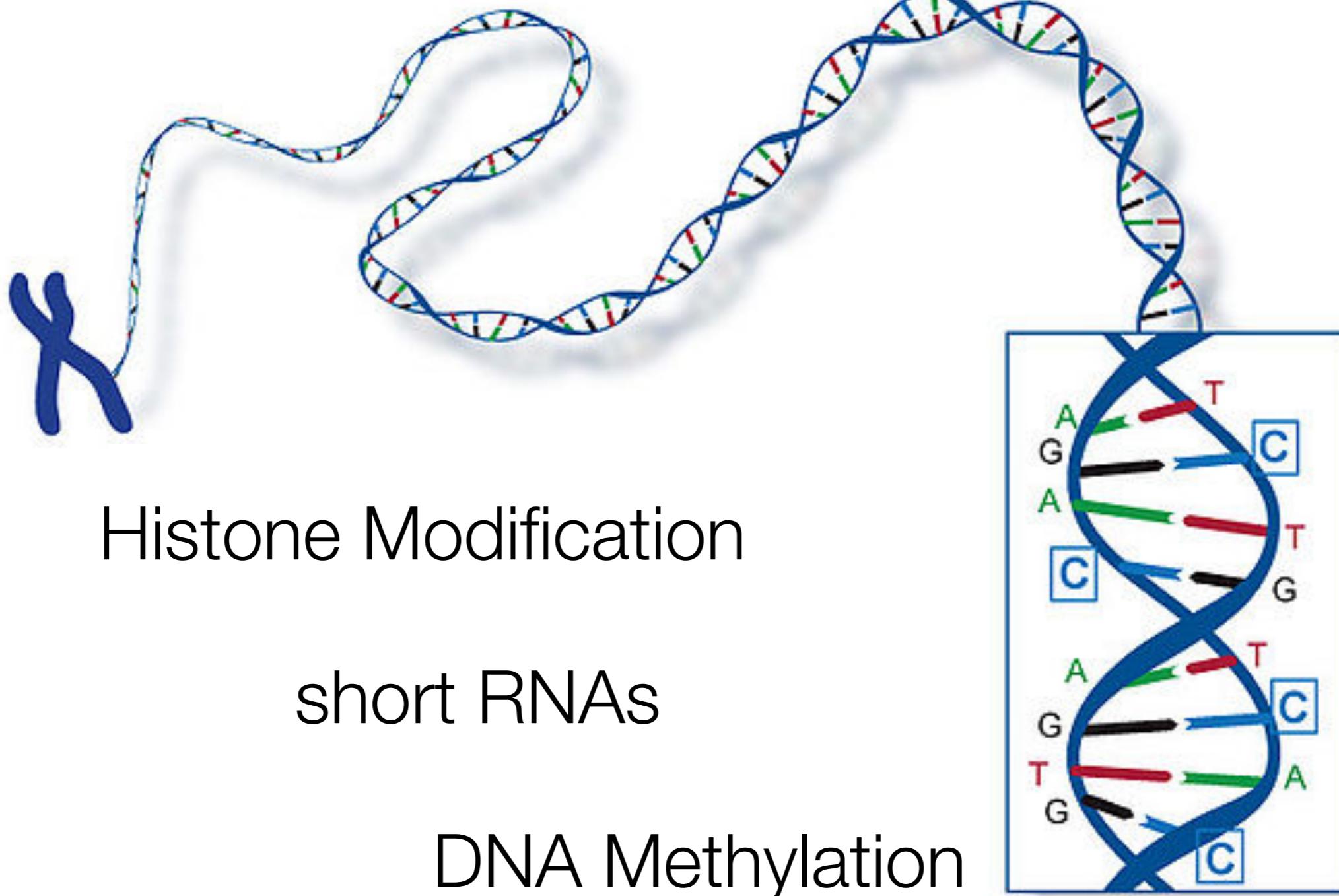
Oliver Bossdorf,<sup>1,\*</sup> Christina L.  
Richards<sup>2</sup> and Massimo Pigliucci<sup>3</sup>



# Genome Resources

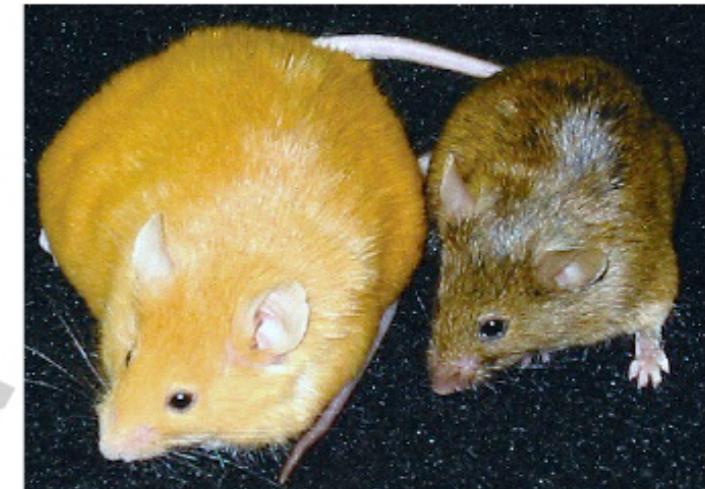


# Epigenetics





# Epigenetics



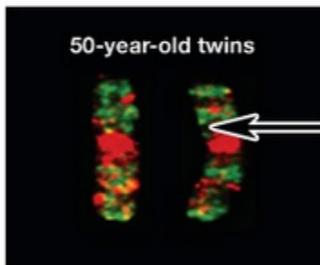
## Nature AND Nurture

### Chromosome 3 Pairs

3-year old twins vs. 50-year-old twins



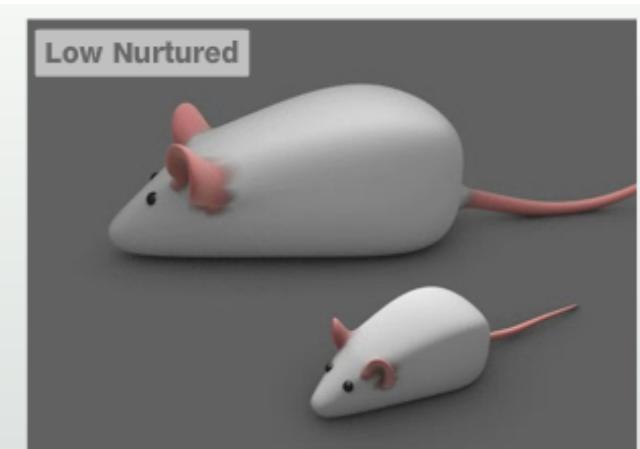
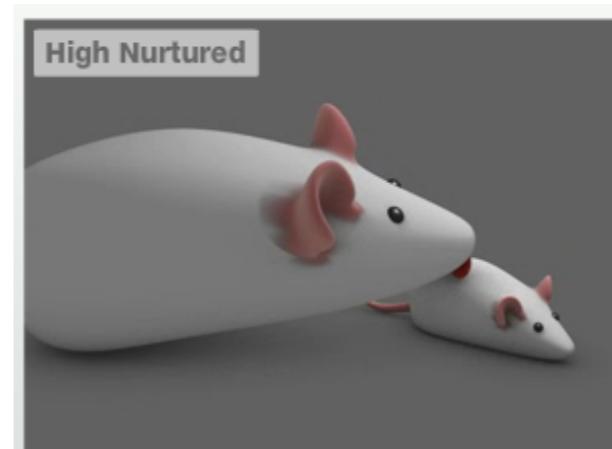
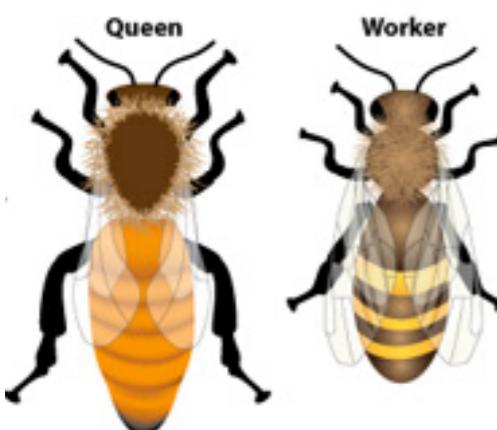
Yellow shows where the twins have epigenetic tags in the same place.



Red and green show where the twins have epigenetic tags in different places.

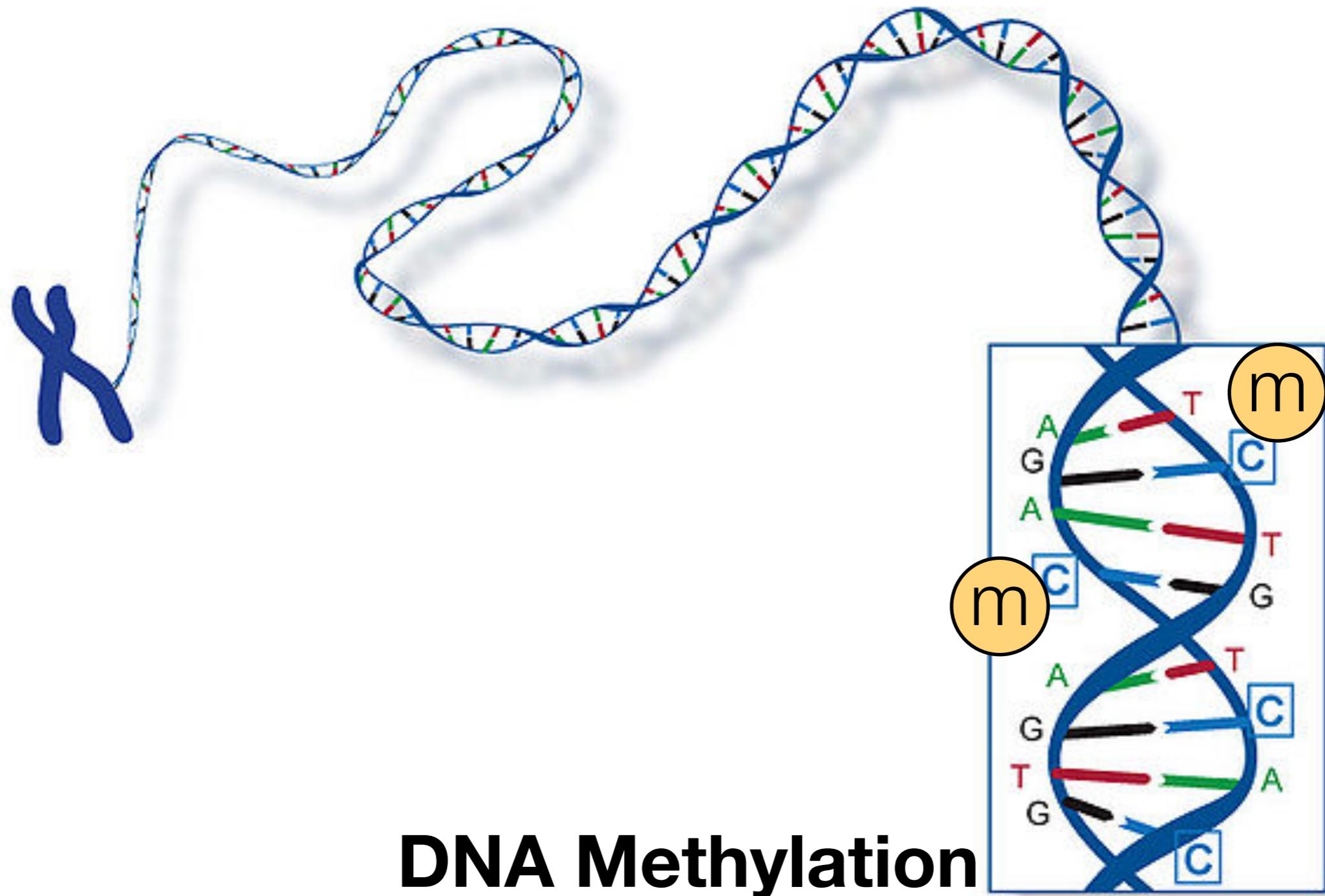


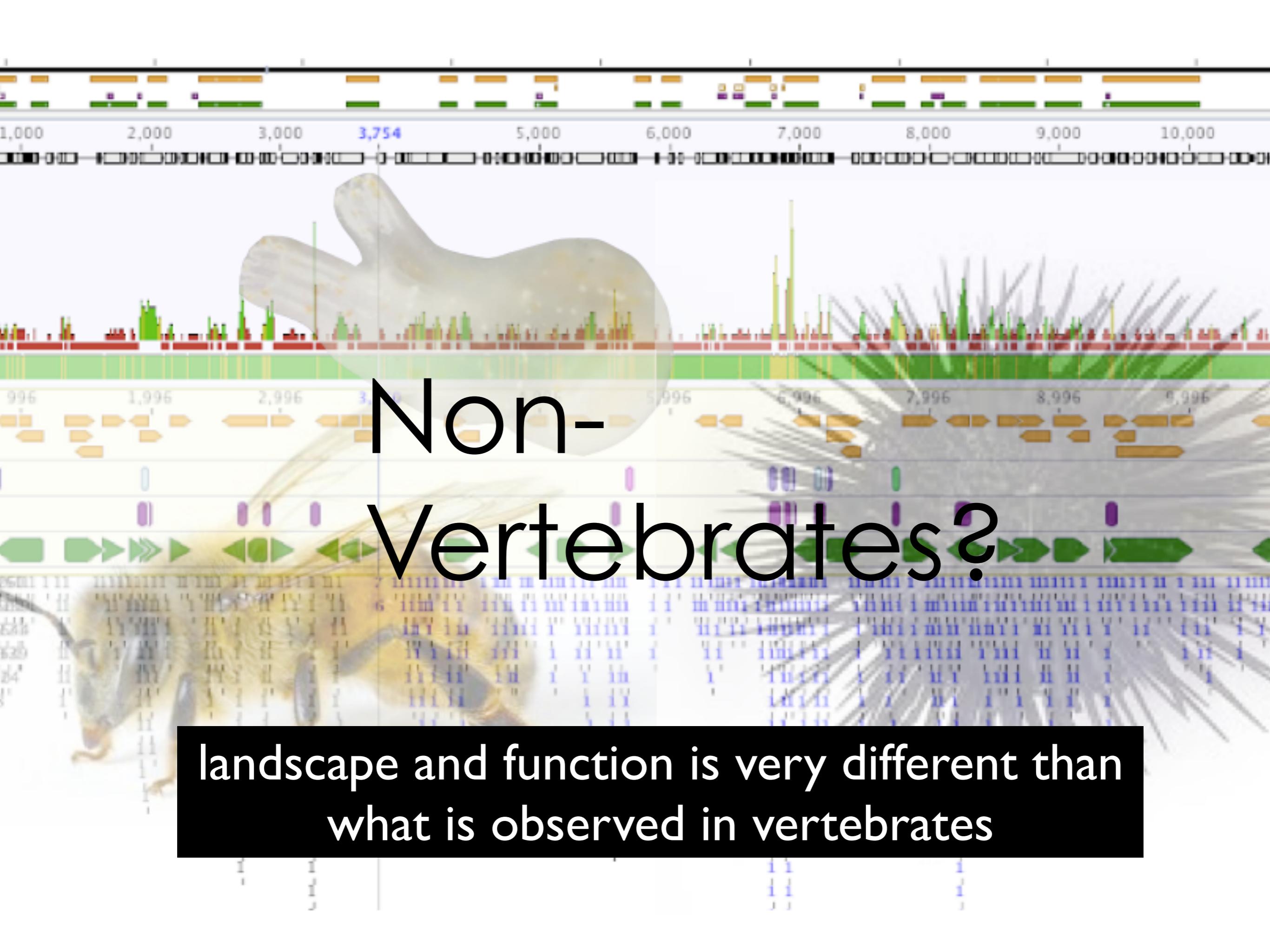
Queen Bee Larvae: Queens are raised in specially constructed cells called "queen cups," which are filled with royal jelly.



These mothers come from a long line of inbred rats, so their genomes are highly similar. But they care for their pups very differently.

AUDIO





# Non- Vertebrates?

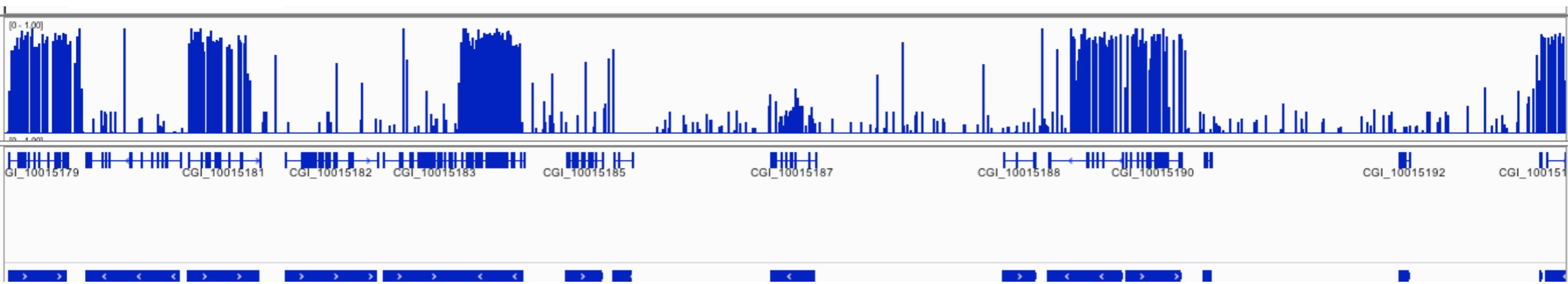
landscape and function is very different than  
what is observed in vertebrates

*Absent in  
several  
model  
organisms*



Oysters?

Epigenetic variation **1**

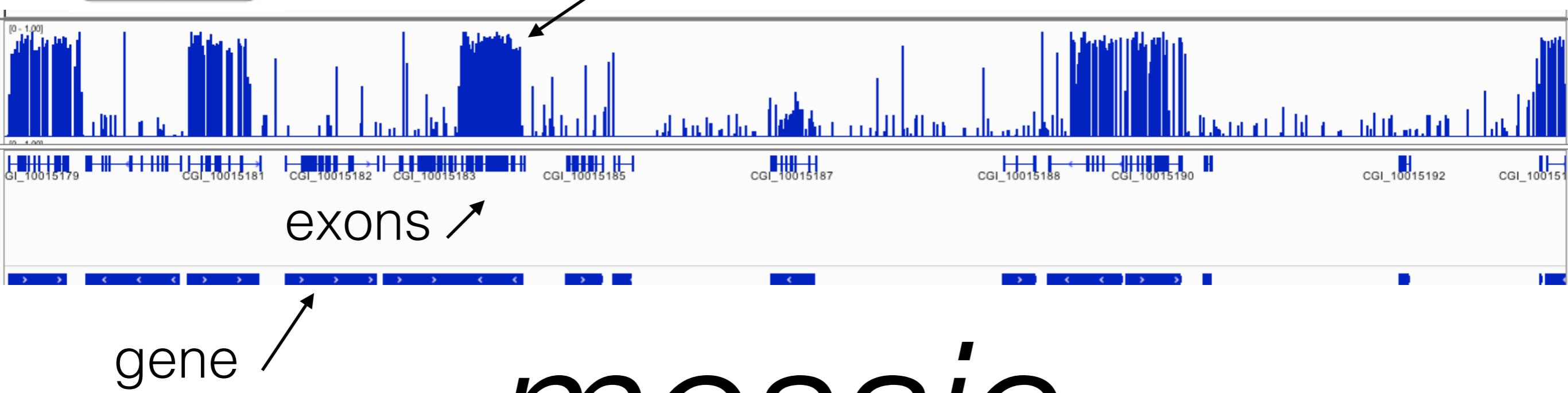


*mosaic*

associated with gene bodies

Epigenetic variation **1**

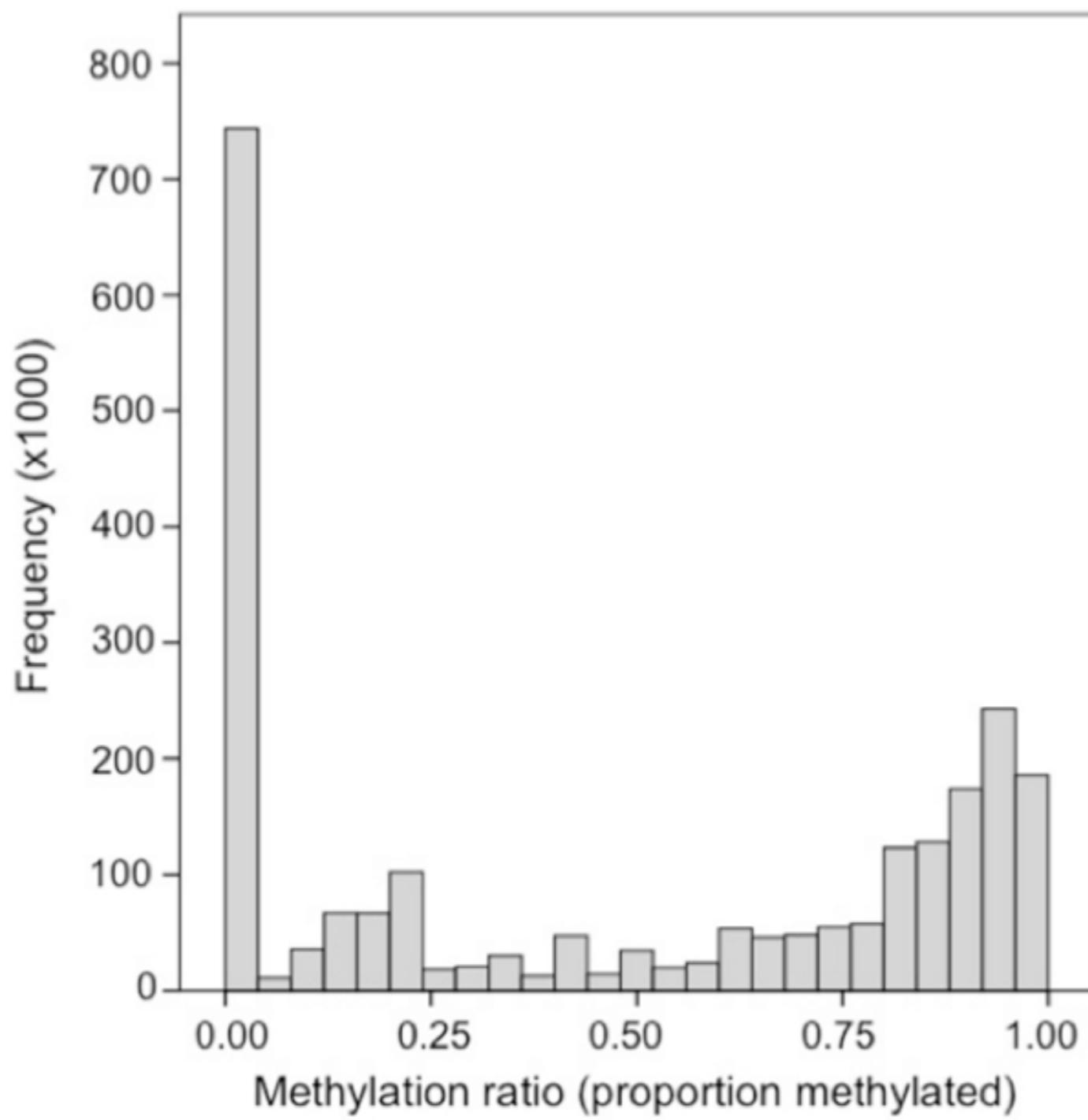
DNA methylation level (0-100%) @ cytosines



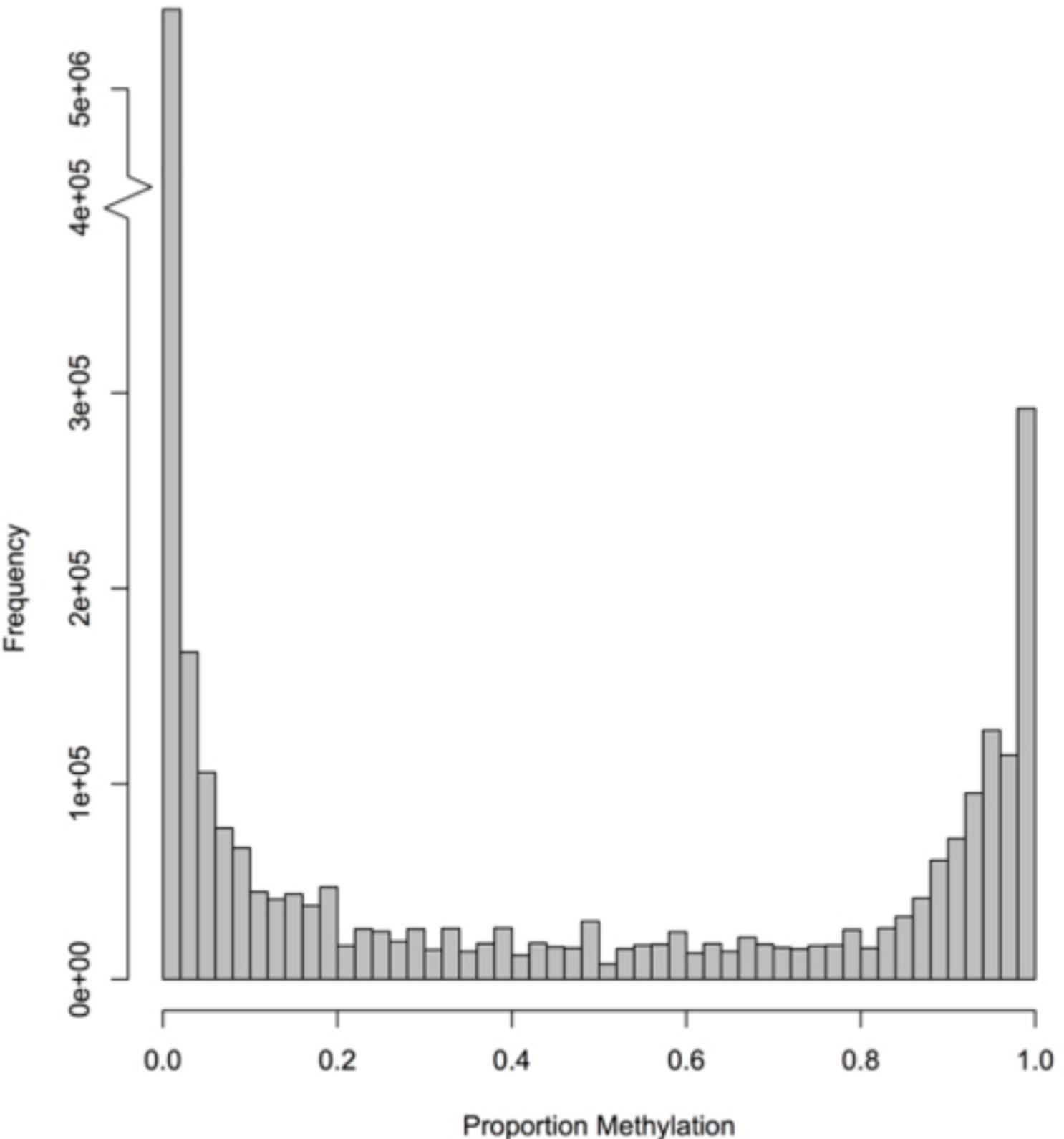
*mosaic*

associated with gene bodies

Epigenetic variation **1**

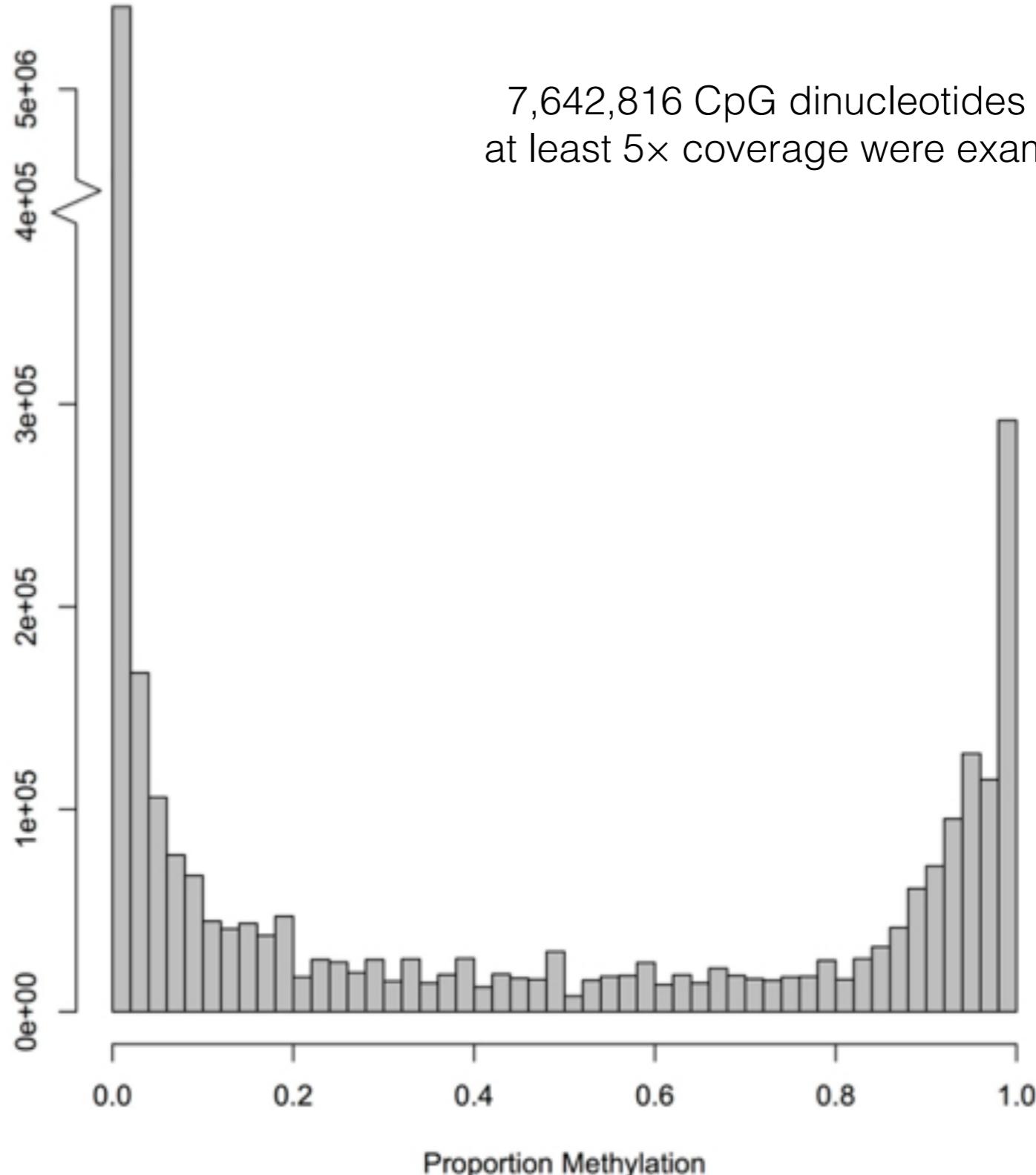
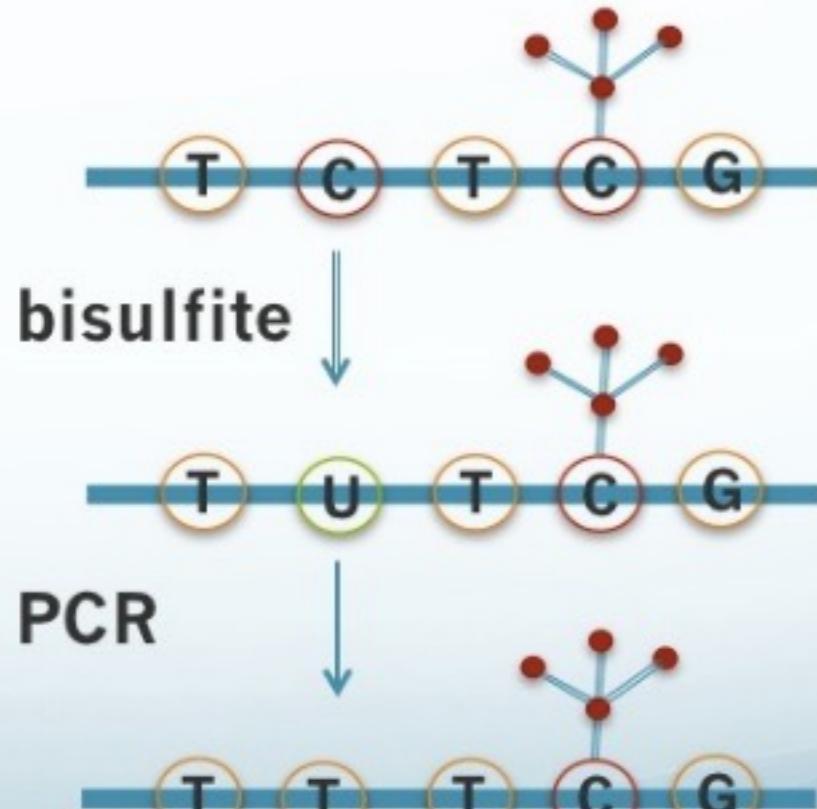


**Figure 1** Frequency distribution of methylation ratios for CpG dinucleotides in oyster gill tissue. A total of 2,625,745 CpG dinucleotides with  $\geq 5 \times$  coverage are represented.



Epigenetic variation **1**

– Bisulfite conversion

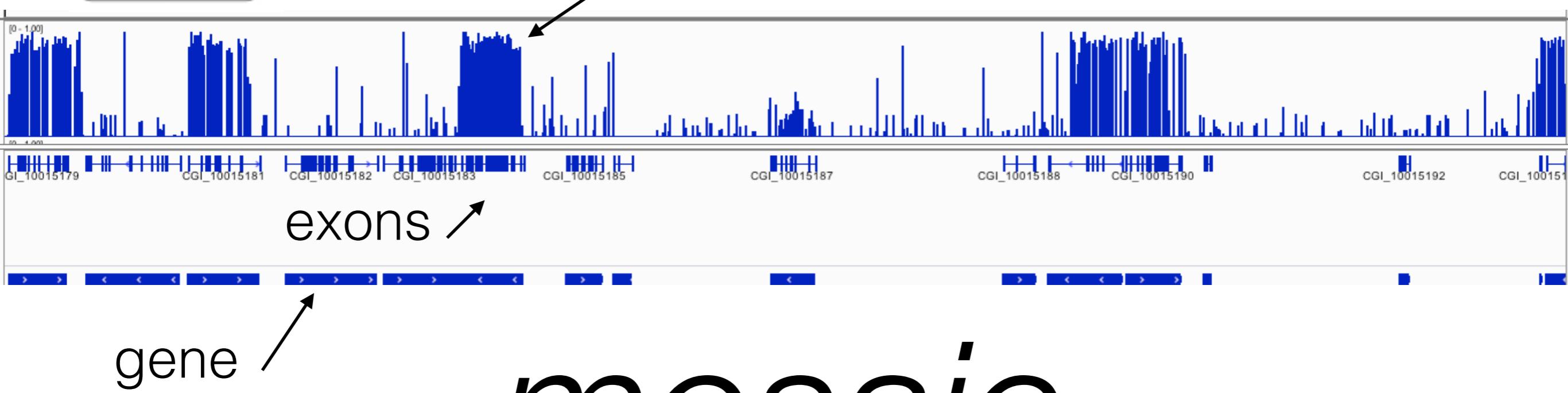


7,642,816 CpG dinucleotides with at least 5x coverage were examined



Epigenetic variation **1**

DNA methylation level (0-100%) @ cytosines

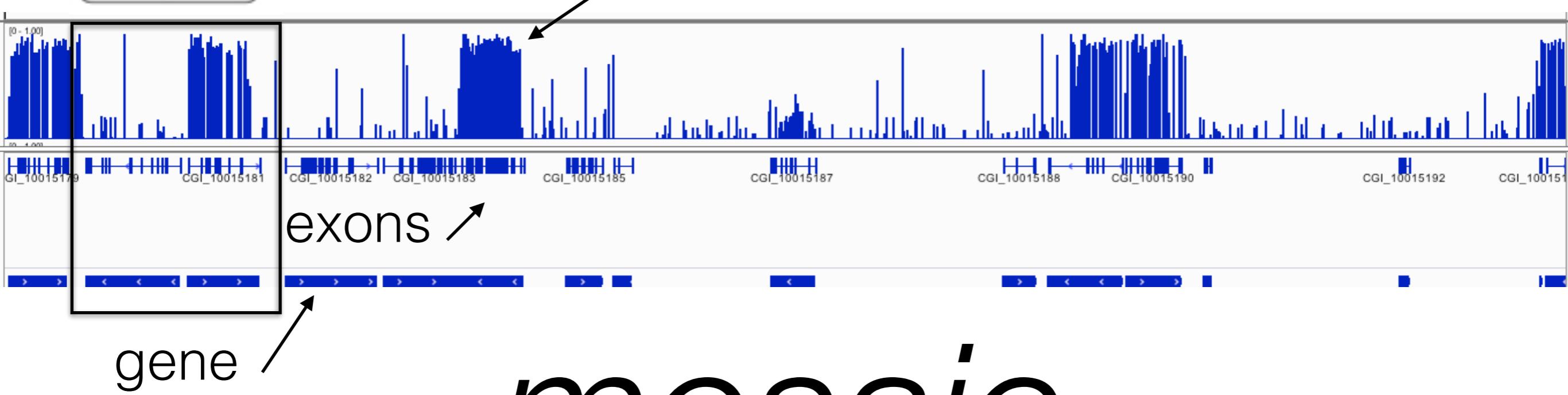


*mosaic*

associated with gene bodies

Epigenetic variation **1**

DNA methylation level (0-100%) @ cytosines

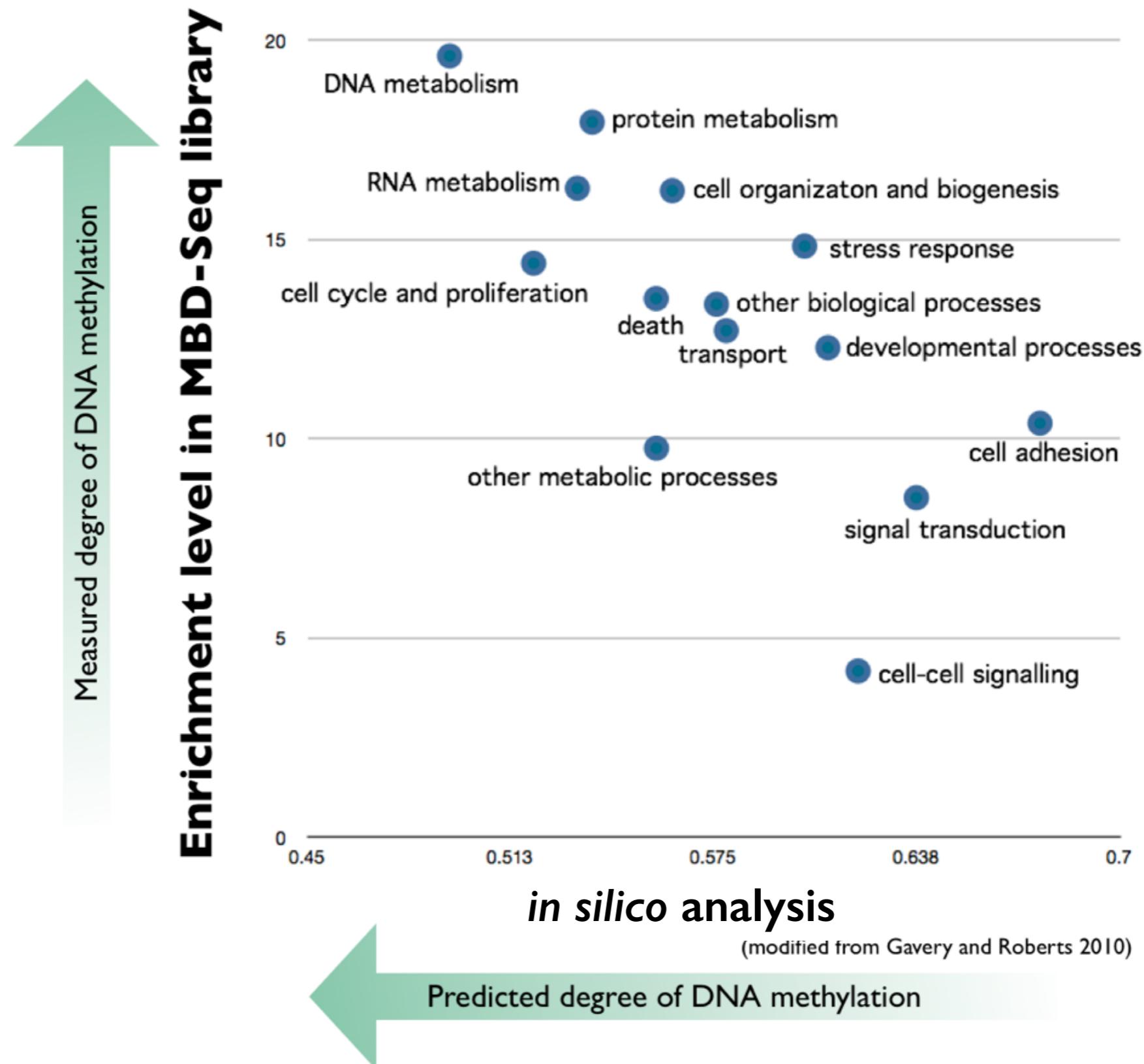


*mosaic*

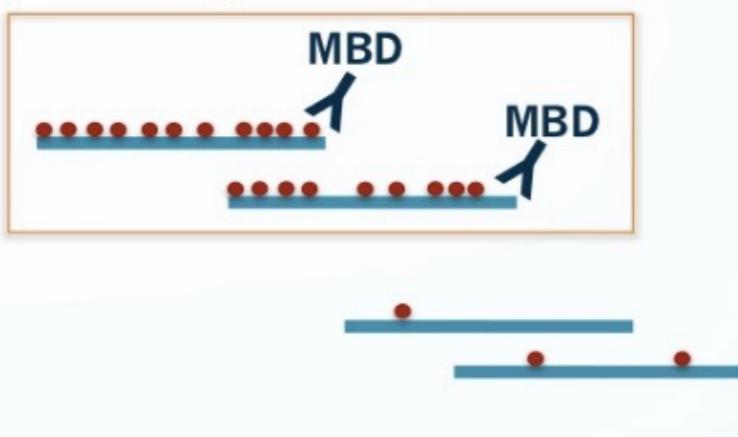
**Why are only a subset of genes methylated?**

associated with gene bodies

Epigenetic variation **1**

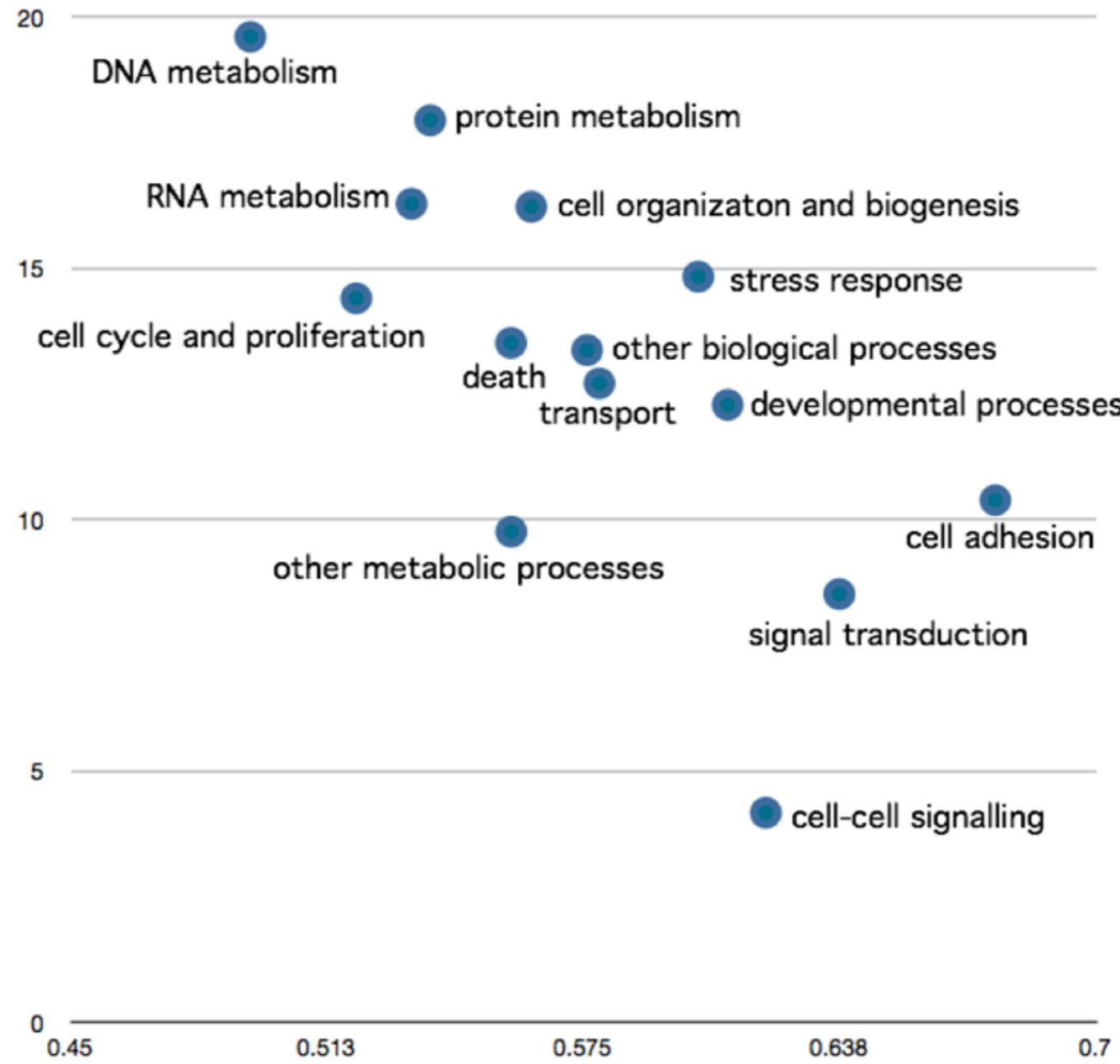


## Epigenetic variation 1



Measured degree of DNA methylation

## Enrichment level in MBD-Seq library



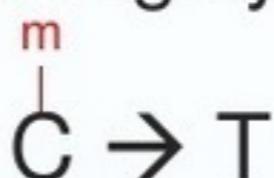
## *in silico* analysis

(modified from Gavery and Roberts 2010)

Predicted degree of DNA methylation

# *in silico* approach

- Principle:
  - Methylated cytosines are highly mutable



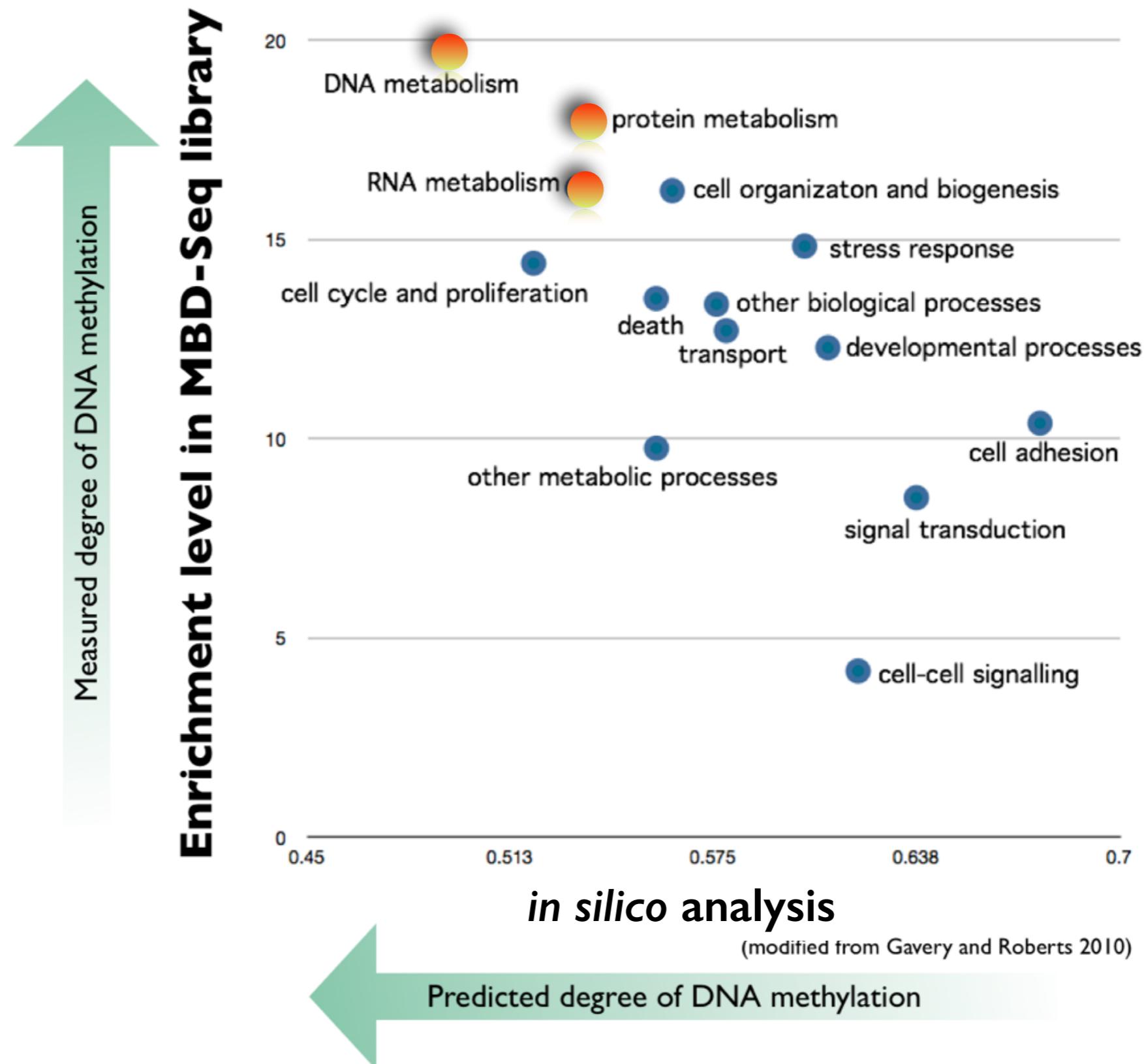
- Methylated regions of DNA are depleted of CpG dinucleotides over evolutionary time (CpG to TpG)

$$\text{CpG O/E} = \frac{\text{CpG observed}}{\text{CpG expected}}$$

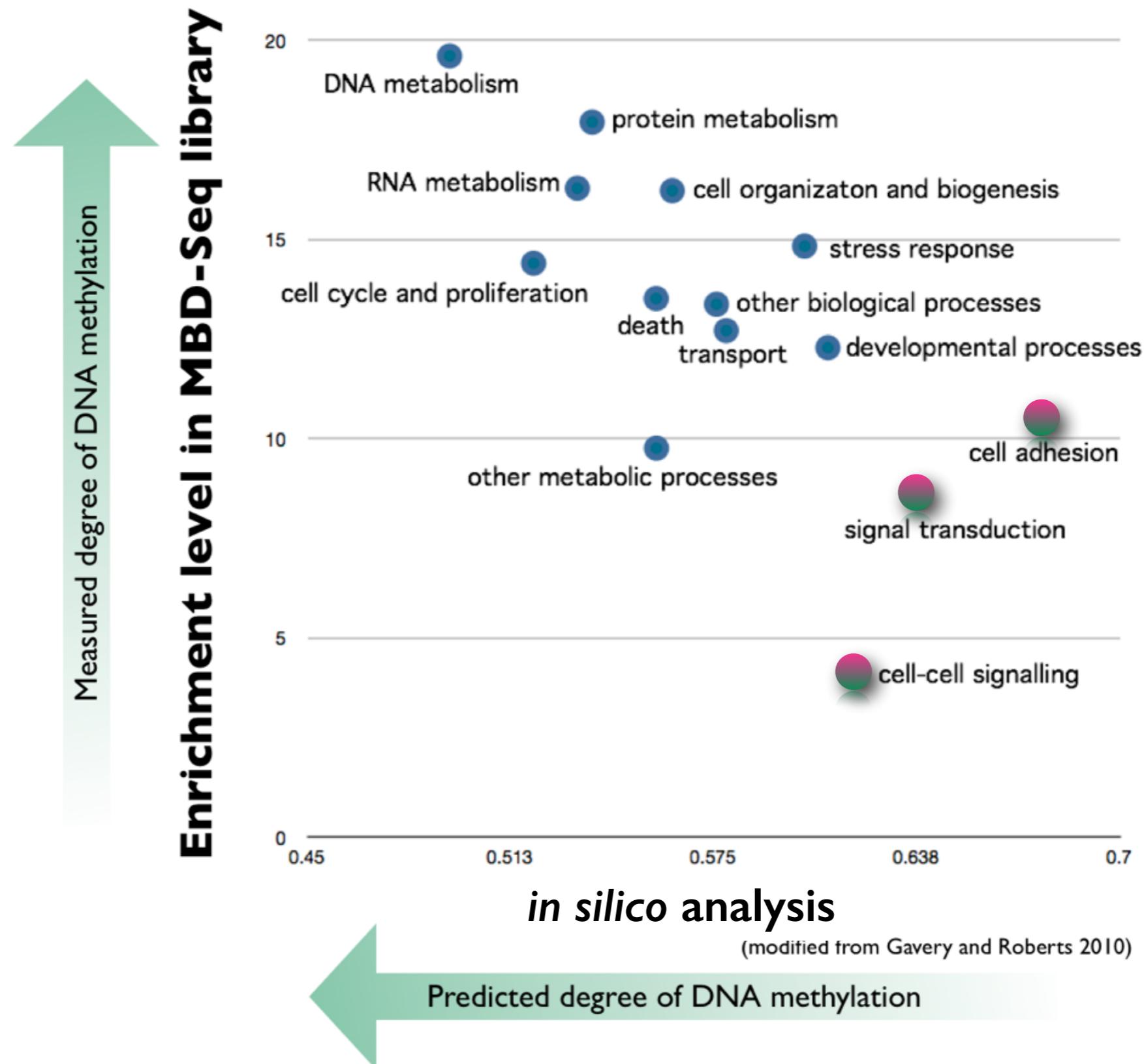


high = unmethylated  
*approaching 1*

Epigenetic variation **1**



Epigenetic variation **1**



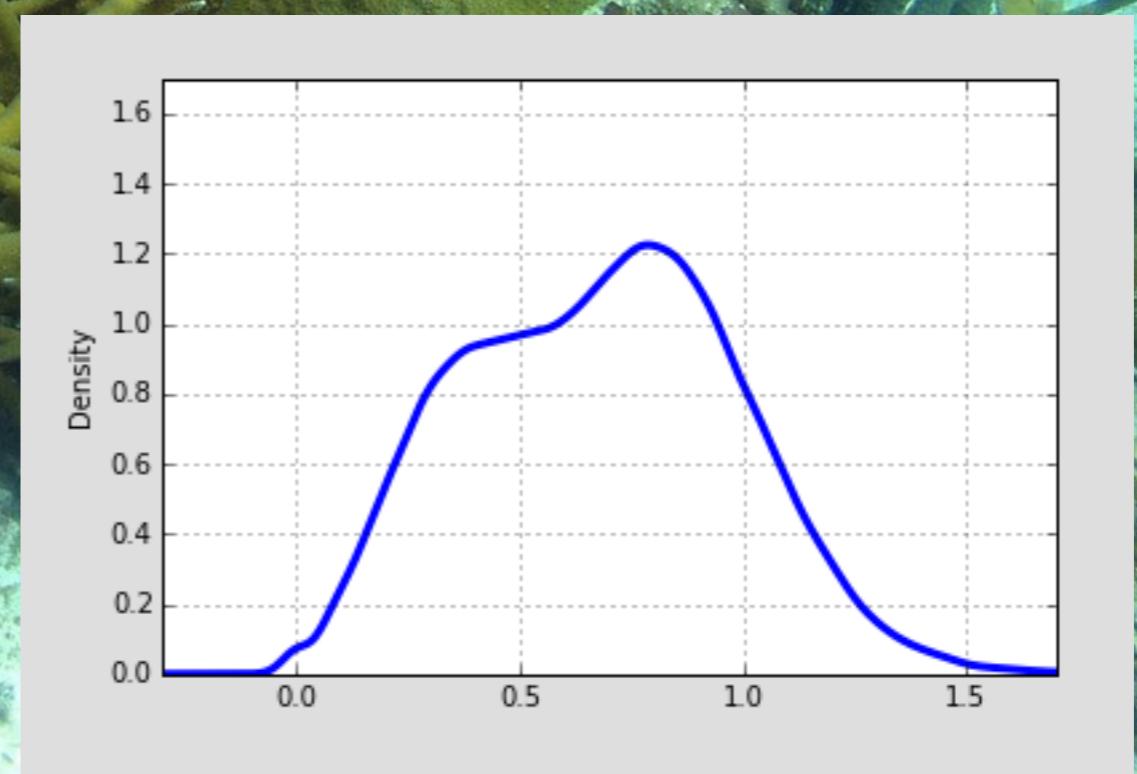
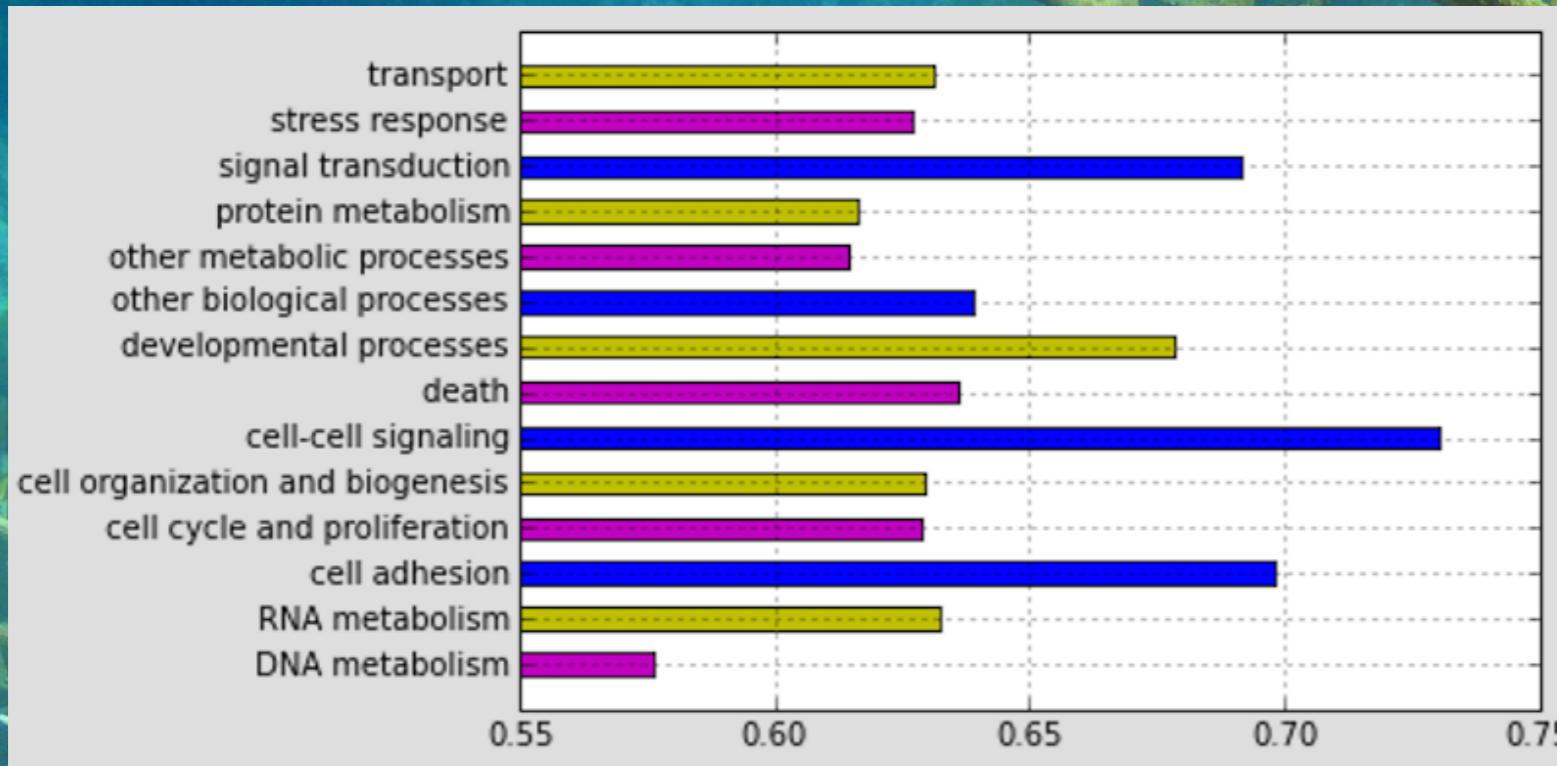
Jay Dimond

*Acropora palmata*

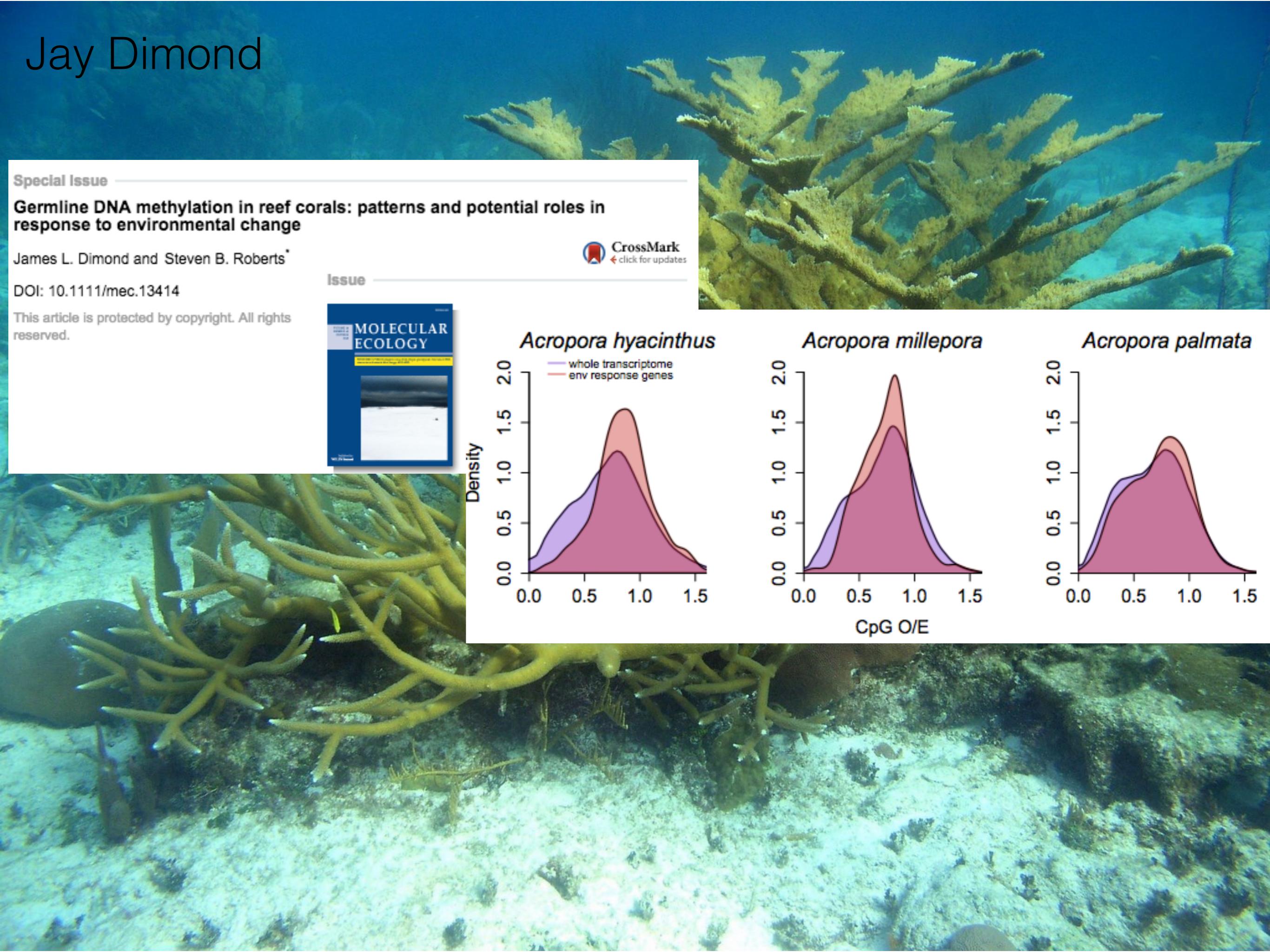
*Acropora cervicornis*

# Jay Dimond

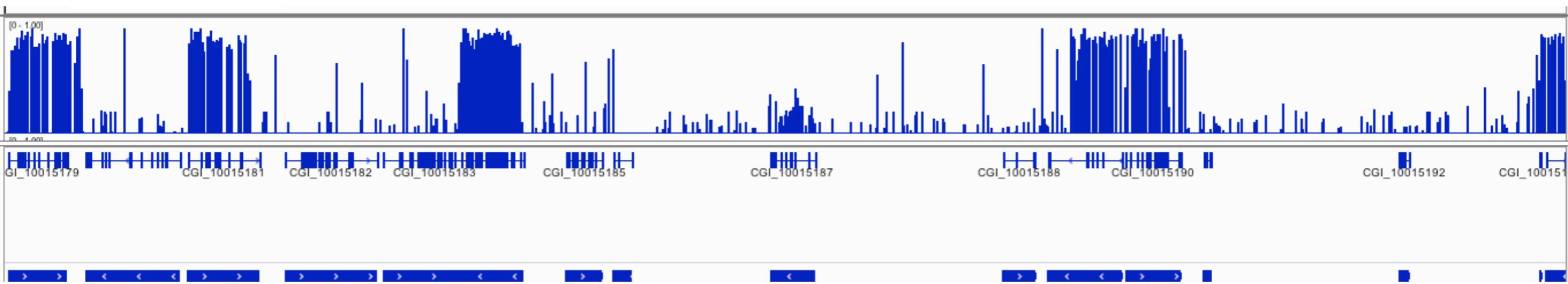
*Acropora palmata*



# Jay Dimond



Epigenetic variation **1**



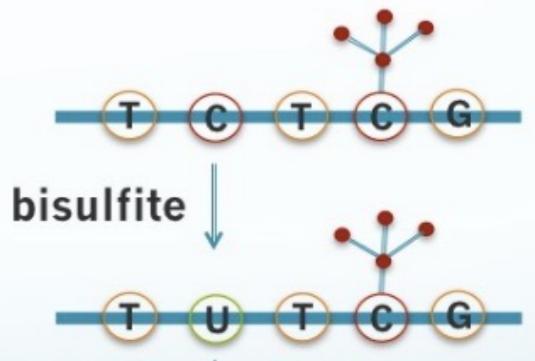
# *mosaic*

associated with gene bodies  
based on gene function

*explanation?*

Epigenetic variation **1**

# Family and Developmental Variation



Sperm &  
Larvae  
(72h & 120h)



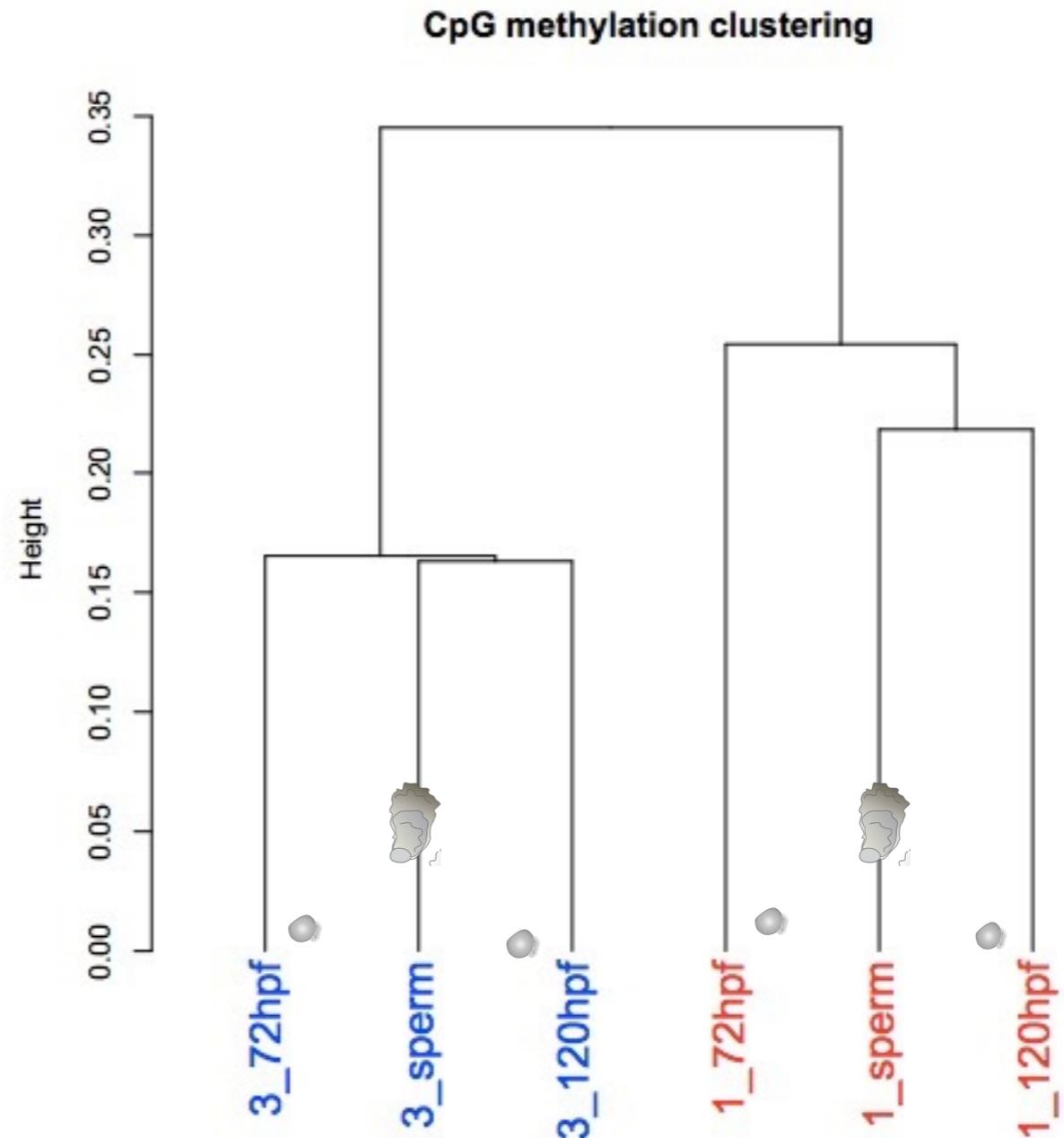
**bioRxiv**  
beta  
THE PREPRINT SERVER FOR BIOLOGY

New Results

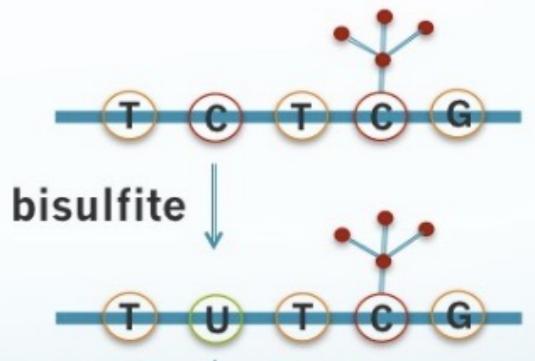
**Indication of family-specific DNA methylation patterns in developing oysters**

Claire E. Olson , Steven B. Roberts

doi: <http://dx.doi.org/10.1101/012831>



# Family and Developmental Variation



## Inheritance



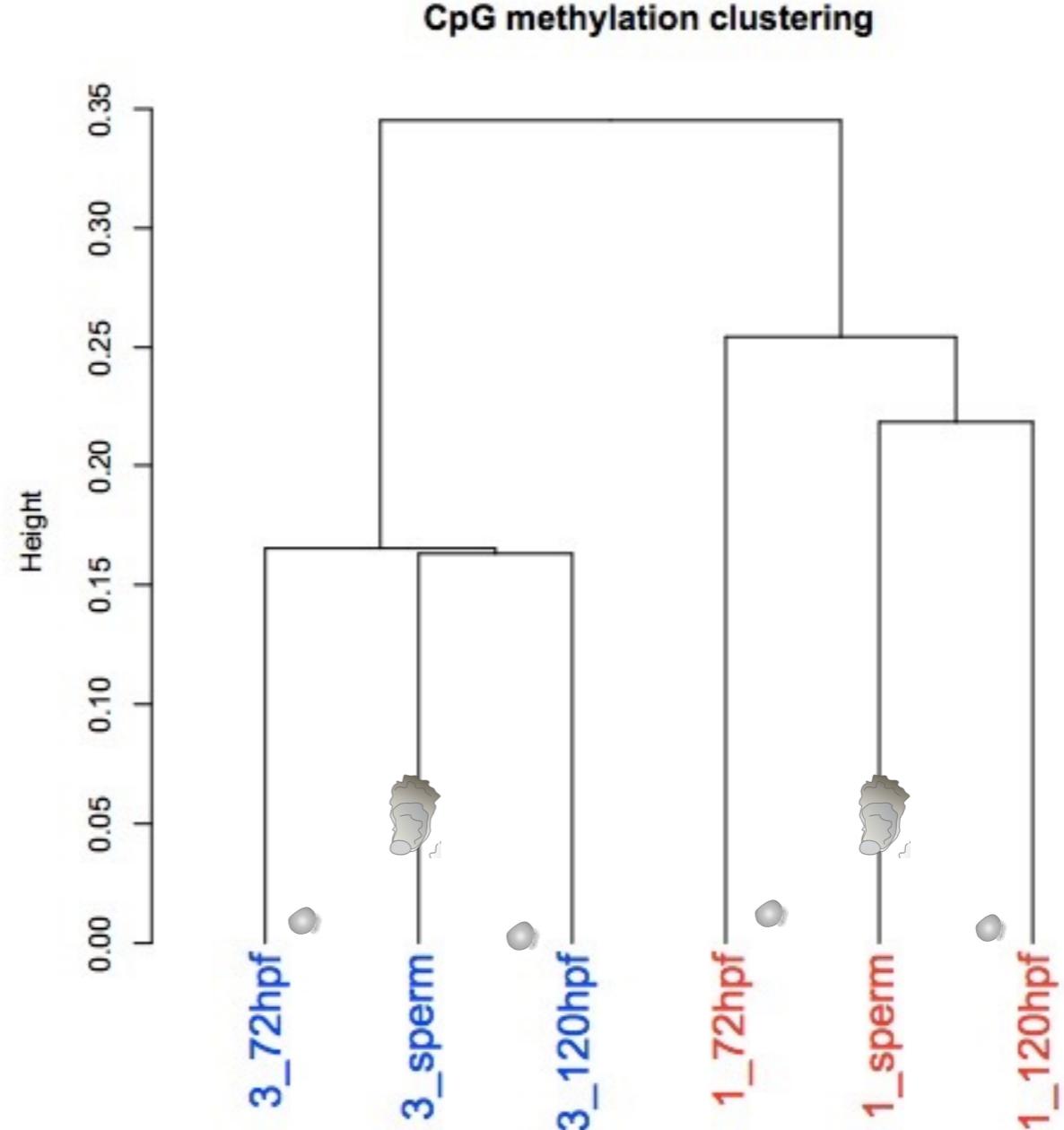
**bioRxiv**  
beta  
THE PREPRINT SERVER FOR BIOLOGY

New Results

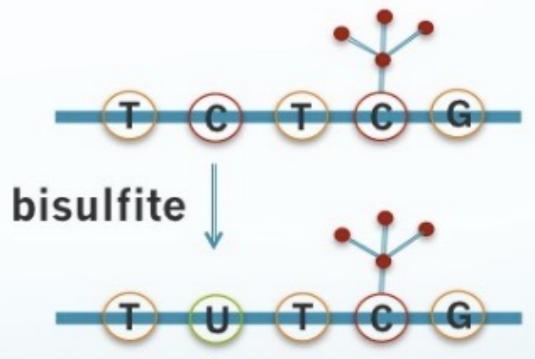
**Indication of family-specific DNA methylation patterns in developing oysters**

Claire E. Olson , Steven B. Roberts

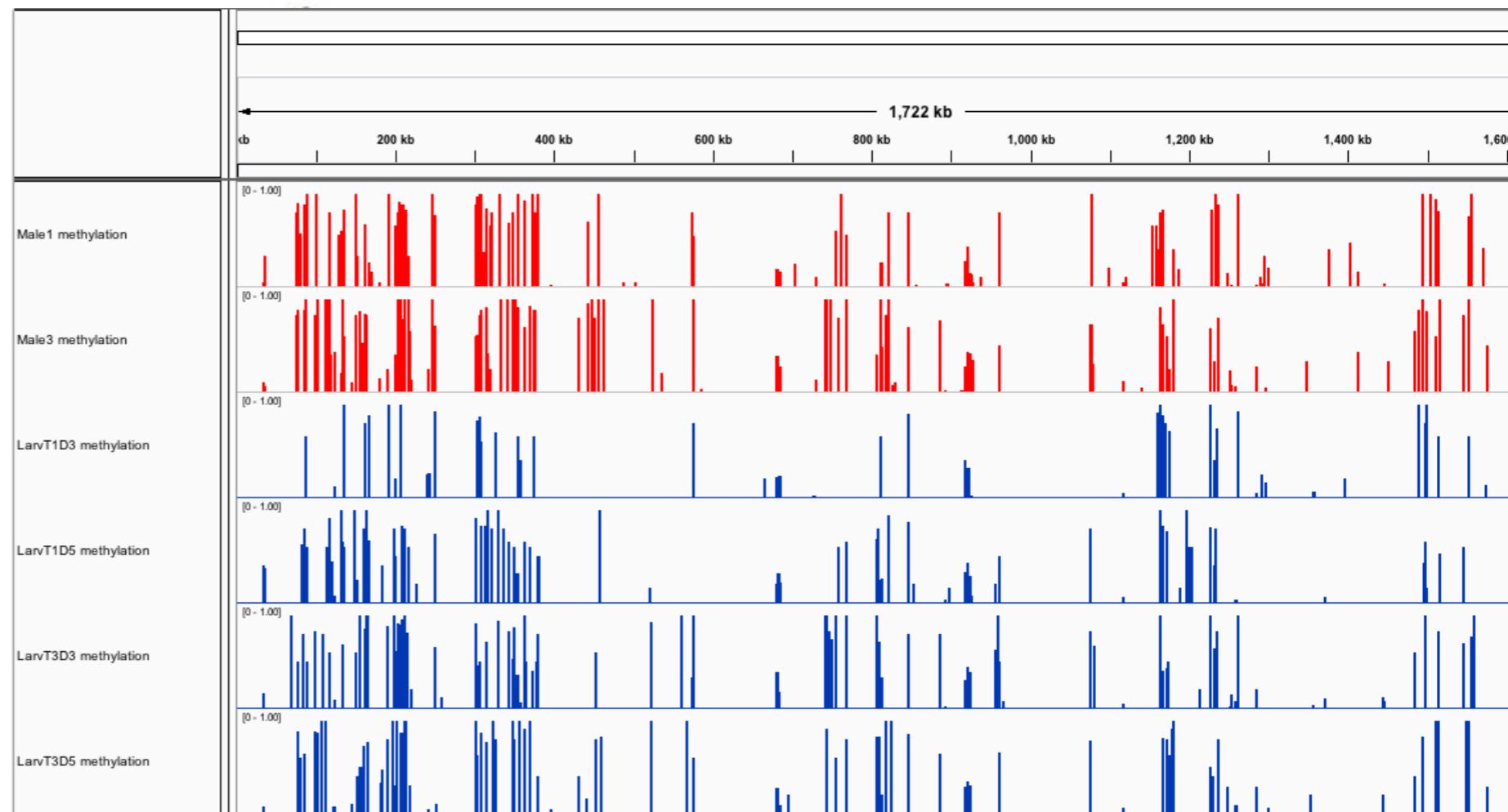
doi: <http://dx.doi.org/10.1101/012831>



# Family and Developmental Variation



Sperm &  
Larvae  
(72h & 120h)



**bioRxiv**  
beta

THE PREPRINT SERVER FOR BIOLOGY

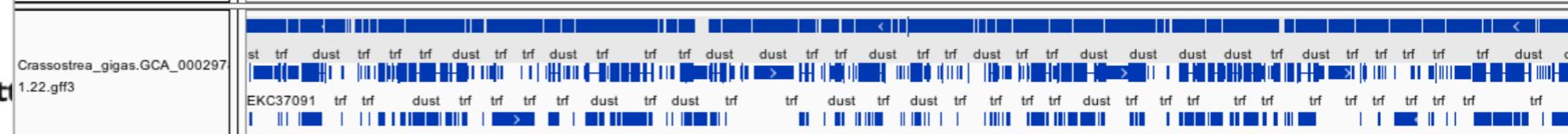
DNA methylation level (0-100%) @ cytosines

New Results

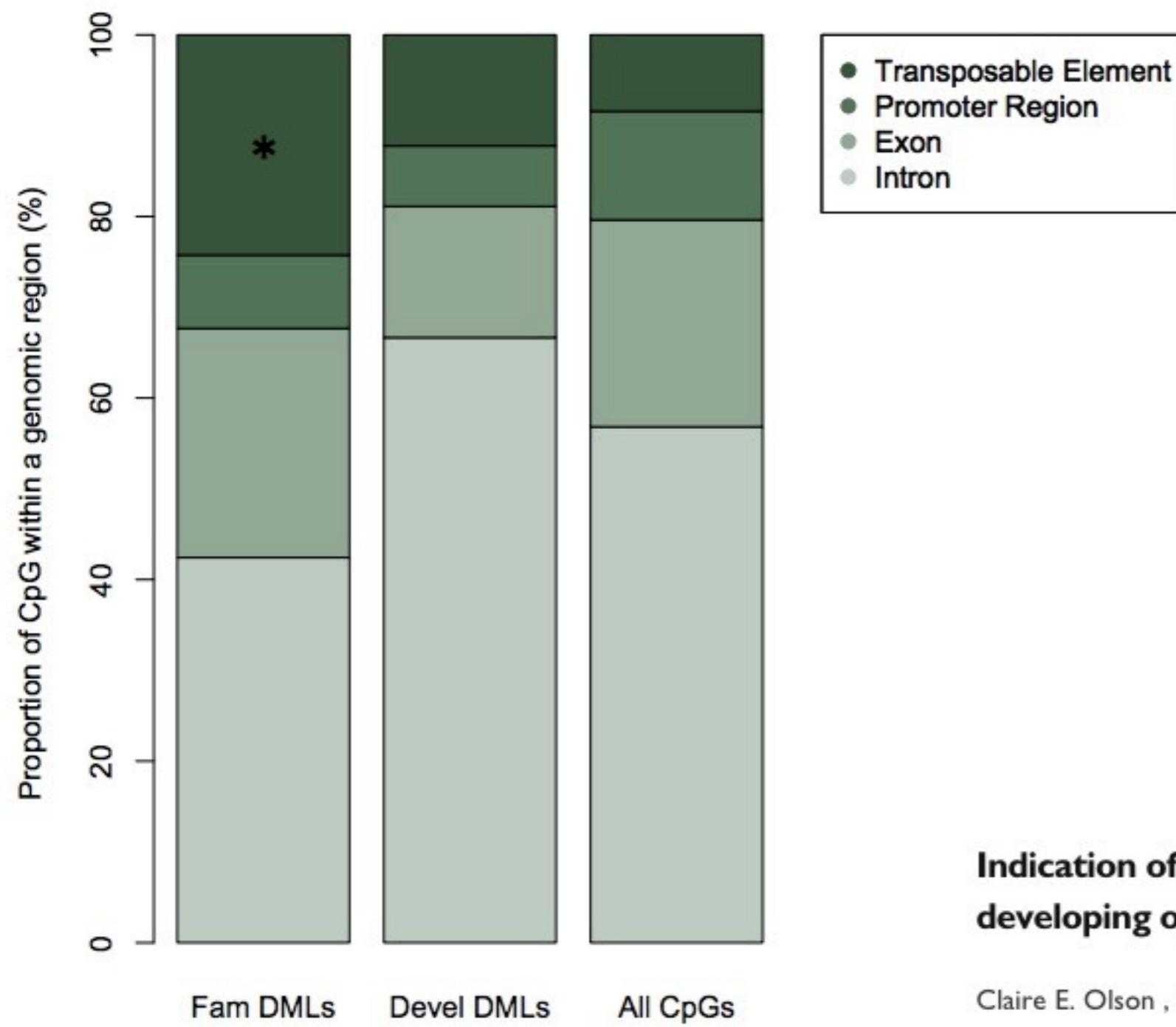
**Indication of family-specific DNA methylation patterns in developing oysters**

Claire E. Olson , Steven B. Roberts

doi: <http://dx.doi.org/10.1101/012831>



# Where are the differences?

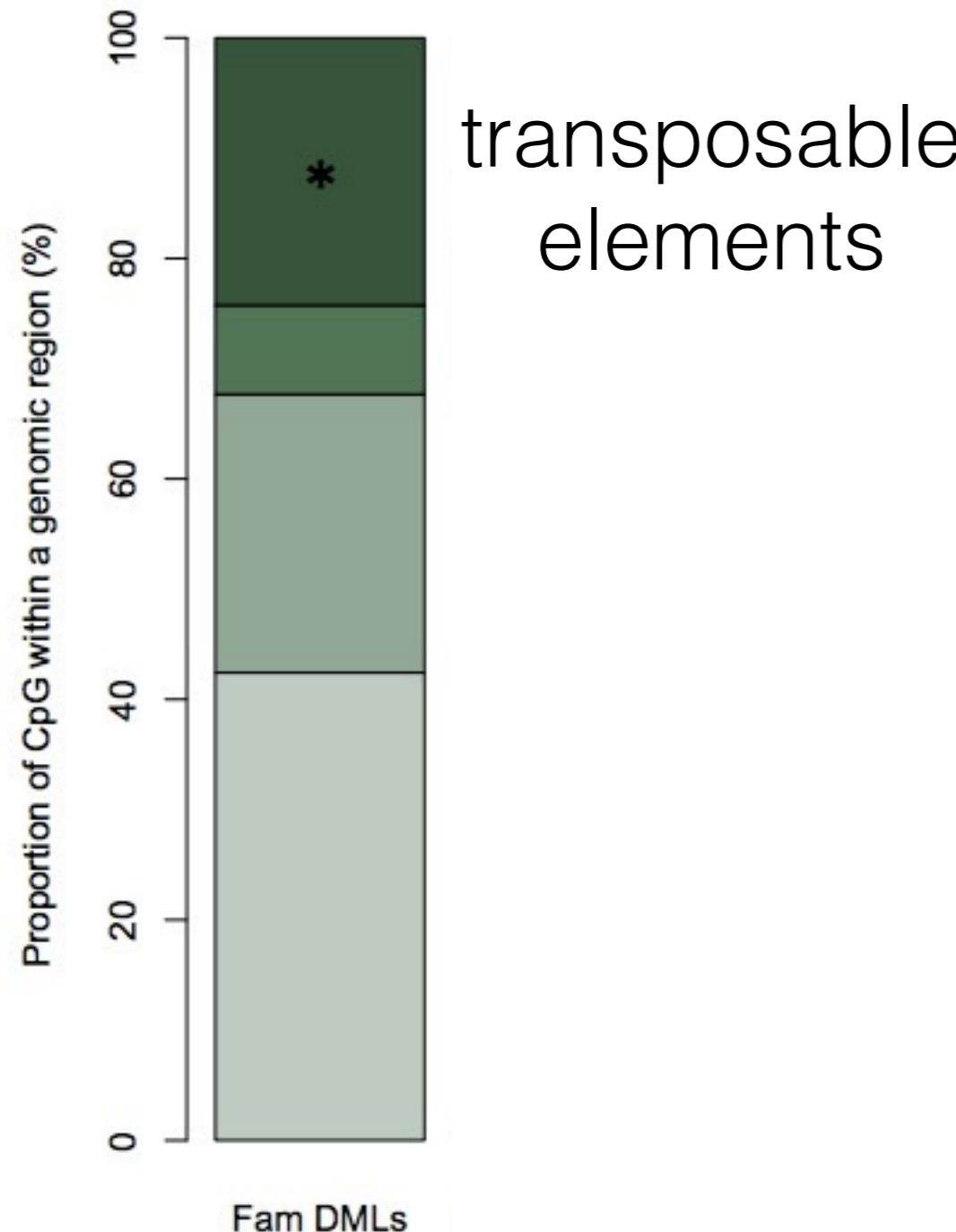


**Indication of family-specific DNA methylation patterns in developing oysters**

Claire E. Olson , Steven B. Roberts

doi: <http://dx.doi.org/10.1101/012831>

Epigenetic variation **1**



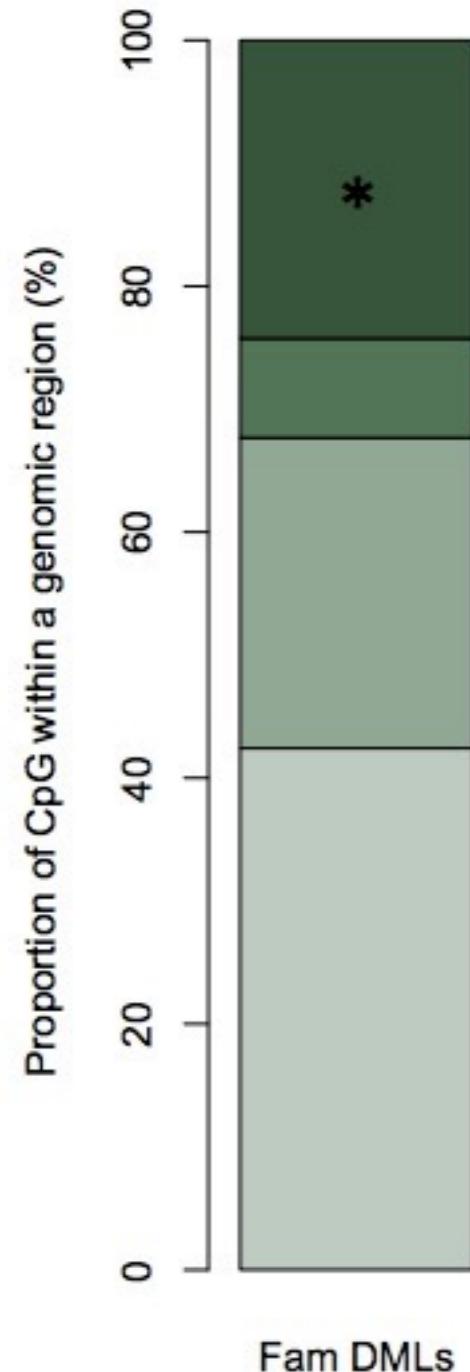
Differentially methylated loci in gene bodies could be lethal or deleterious, as they would alter gene expression.

Another possibility is that differentially methylated loci may provide advantageous phenotypic variation by increasing transposable element mobility.

**Indication of family-specific DNA methylation patterns in developing oysters**

Claire E. Olson , Steven B. Roberts

doi: <http://dx.doi.org/10.1101/012831>



***Assumes some form  
of random process***  
occurring during  
gametogenesis?

**Indication of family-specific DNA methylation patterns in developing oysters**

Claire E. Olson , Steven B. Roberts  
doi: <http://dx.doi.org/10.1101/012831>

- Sparsely (~16 %), mosaic methylated genome
- Gene body methylation correlated with function
- DNA methylation patterns are inherited
- DMRs are predominant in transposable elements

Gene  
expression



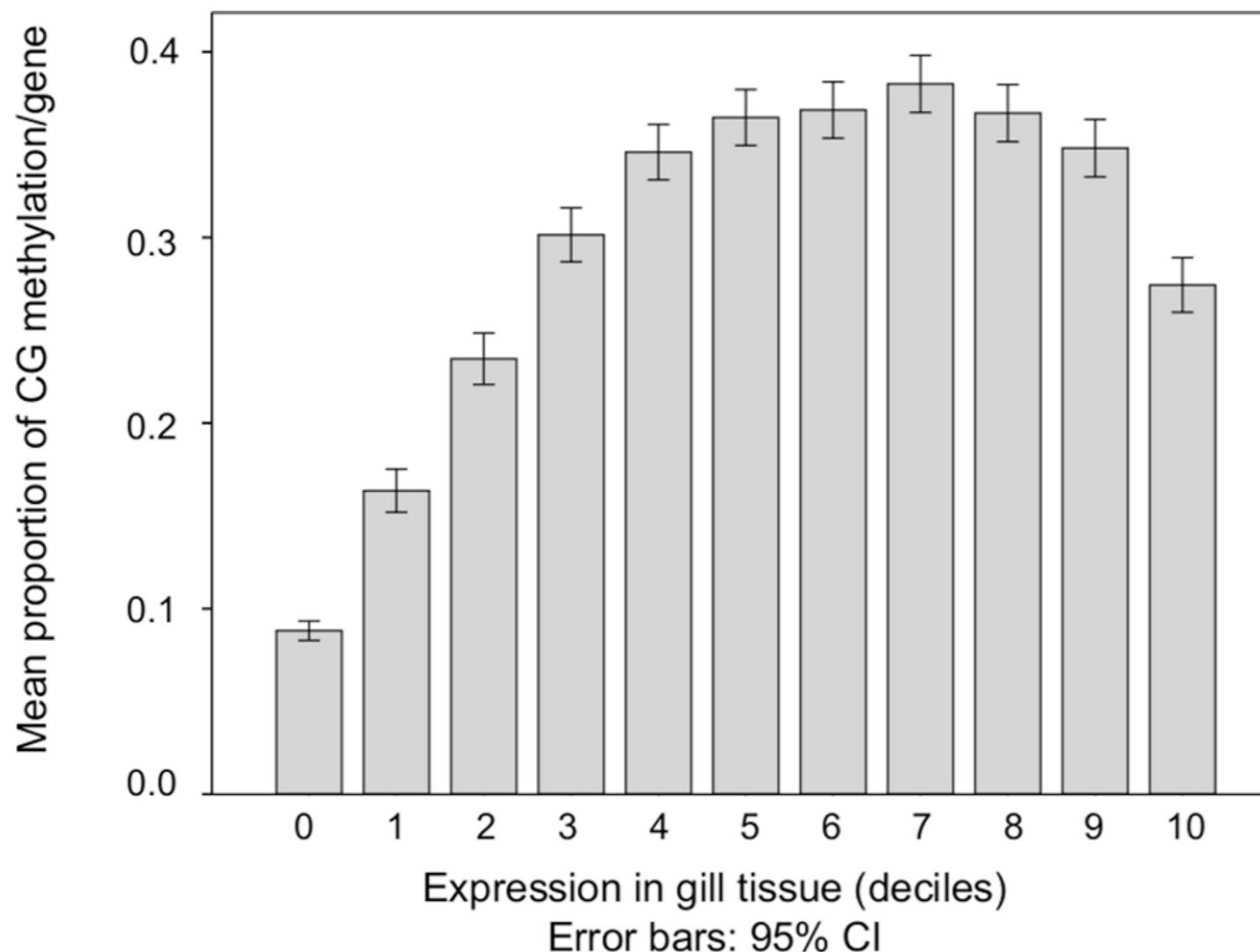
Epigenetic  
variation

# Function?

Gene expression

2

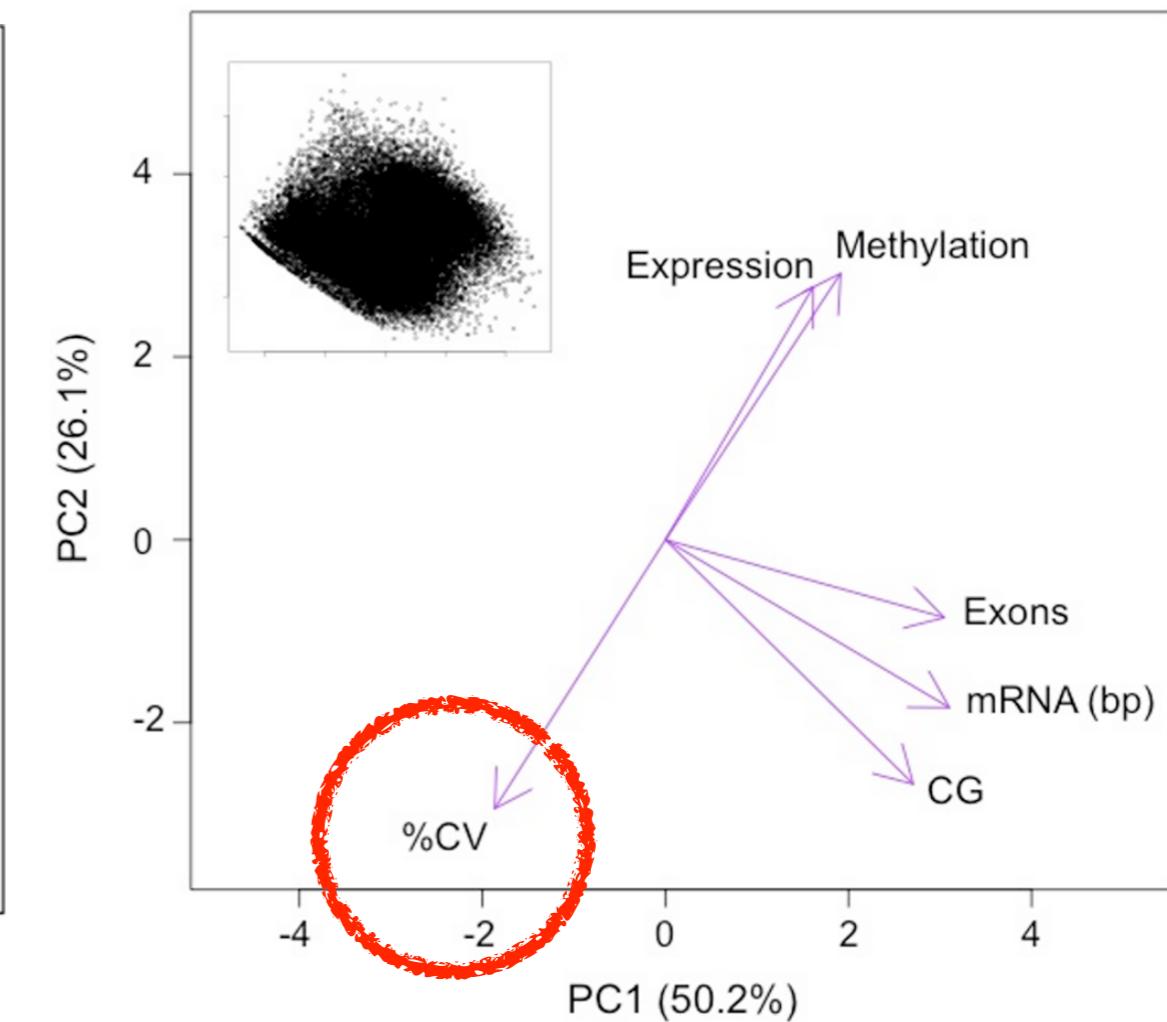
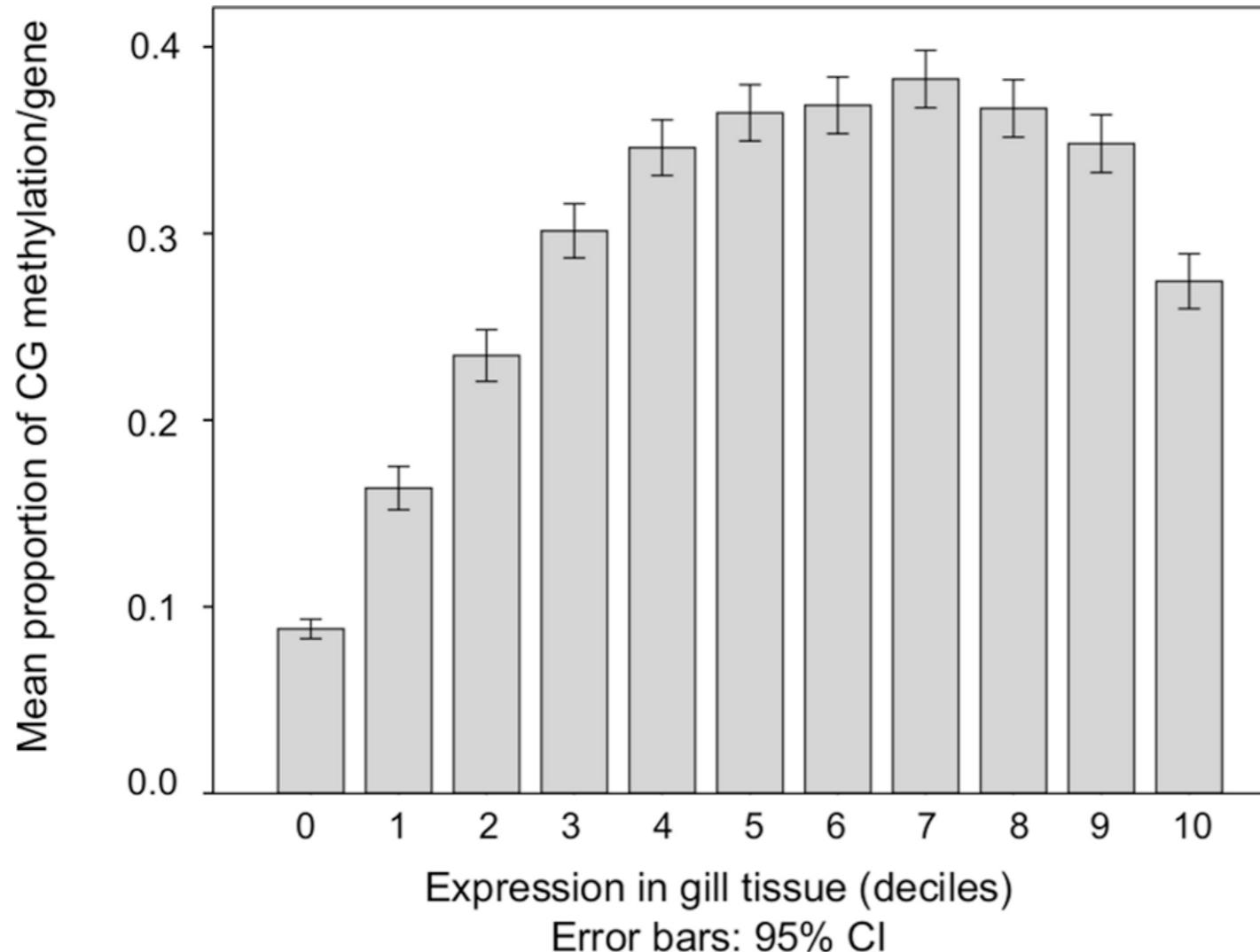
Epigenetic variation



Gene expression

2

Epigenetic variation



## Predominant intragenic methylation is associated with gene expression characteristics in a bivalve mollusc

Mackenzie R. Gavery and Steven B. Roberts

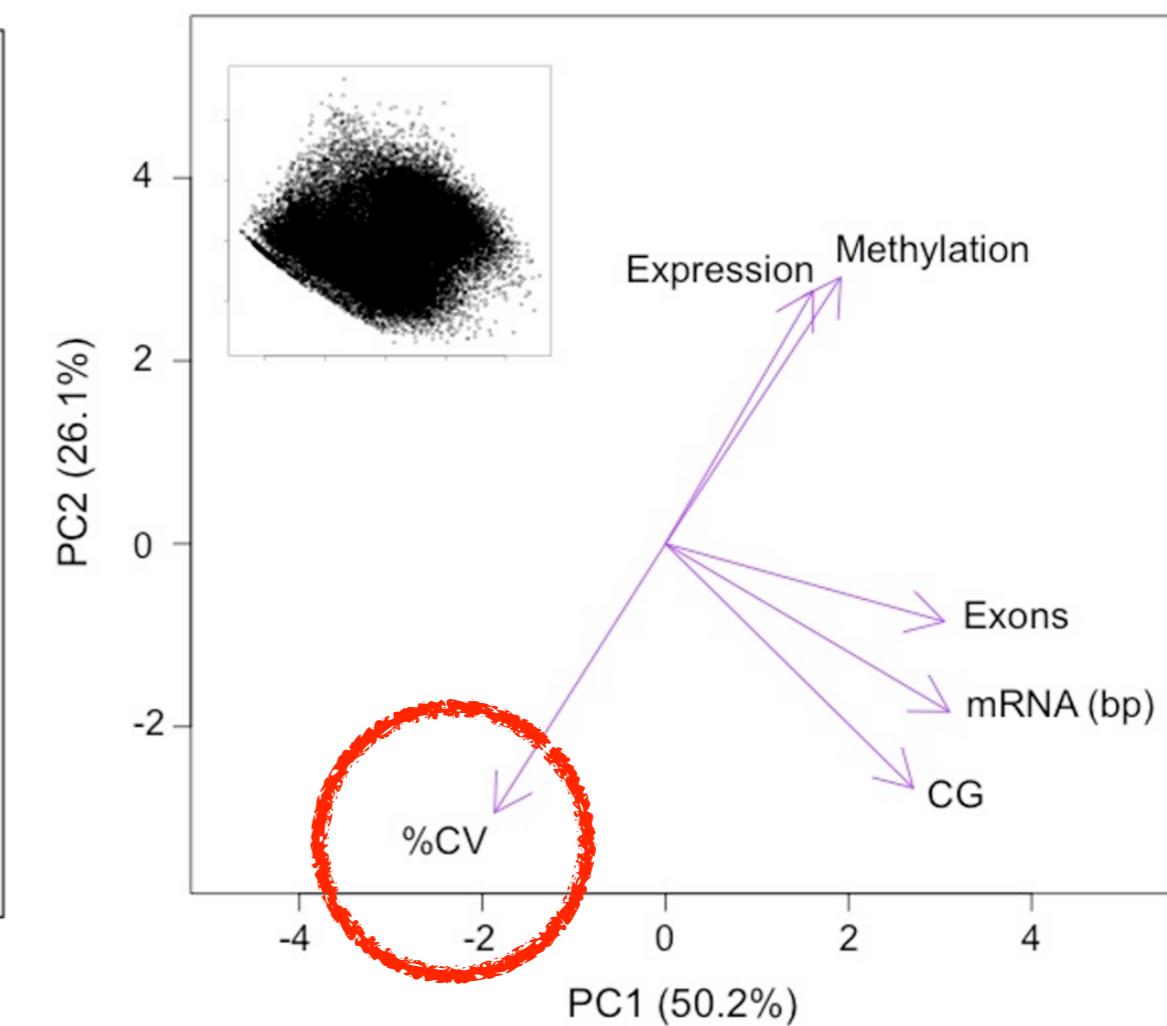
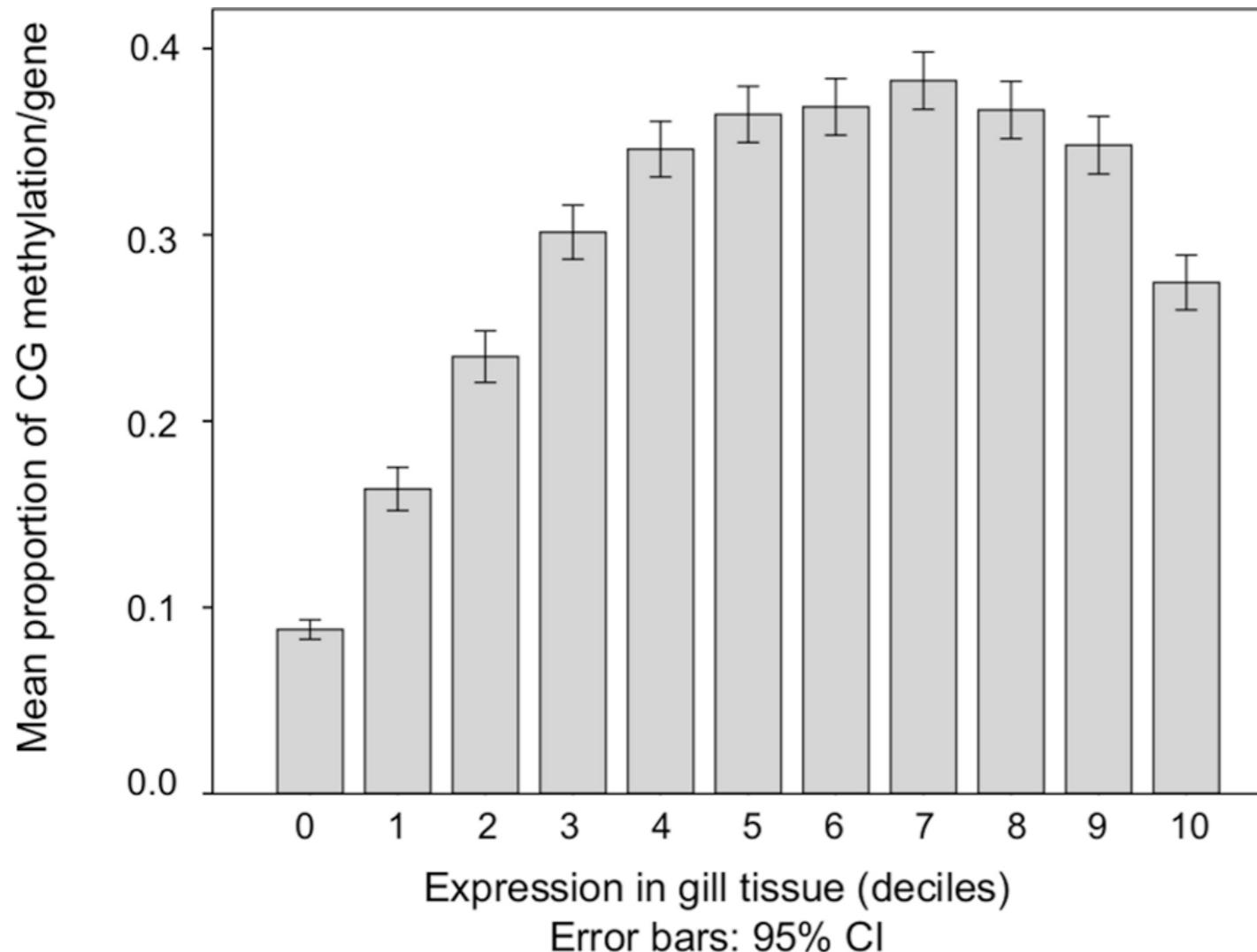
School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA

Gene expression

2

Epigenetic variation

Theory: **Does not** influence expression level  
but rather alternative splicing.



PeerJ

Predominant intragenic methylation is  
associated with gene expression  
characteristics in a bivalve mollusc

Mackenzie R. Gavery and Steven B. Roberts

School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA

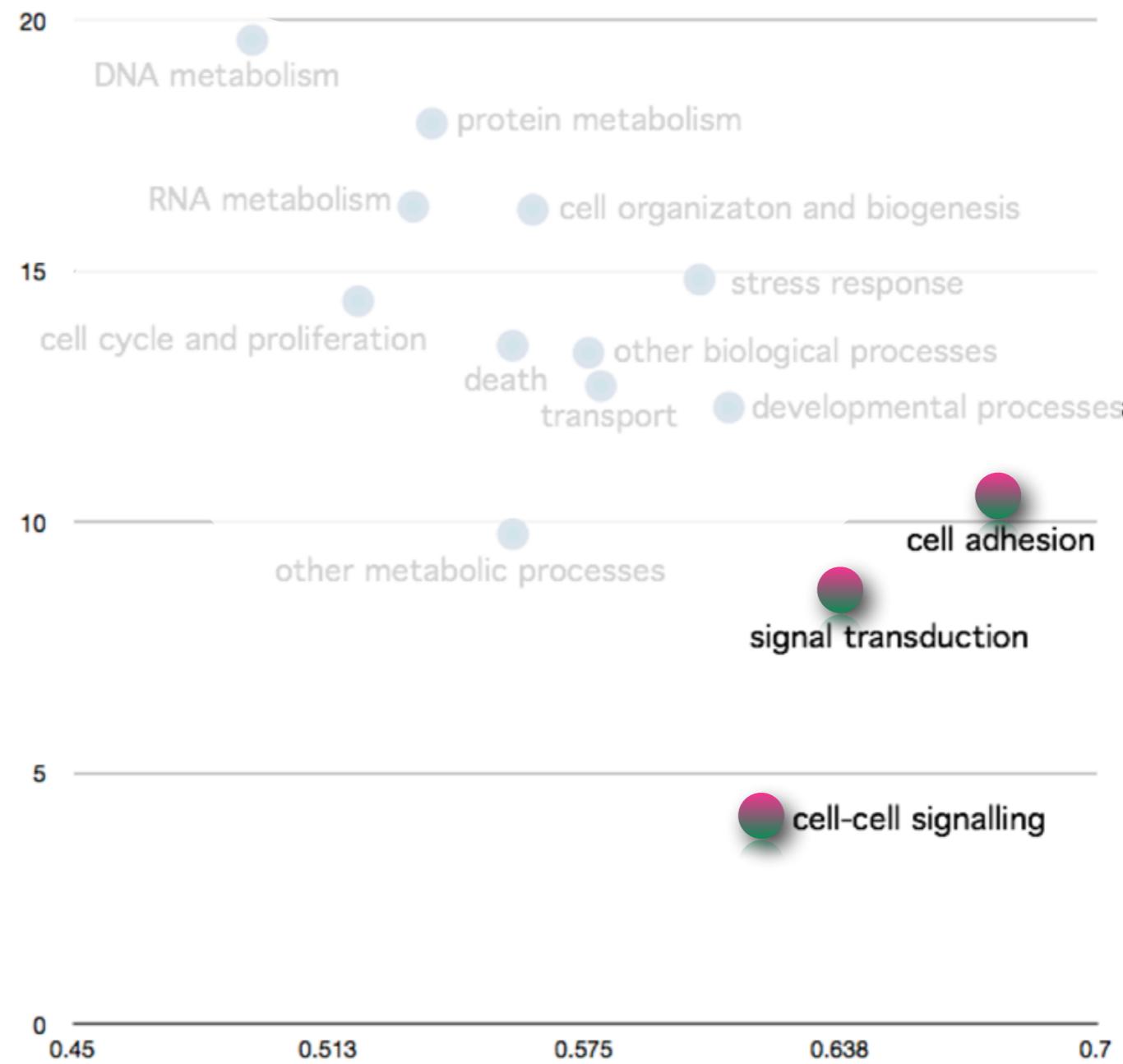


In species that experience a diverse range of environmental conditions, processes have evolved to increase the number of potential phenotypes in a population in order to improve the chances for an individual's survival.



Gene expression

Epigenetic variation  
2



Roberts and Gavery 2012

Gene expression

2

Epigenetic variation

cell adhesion

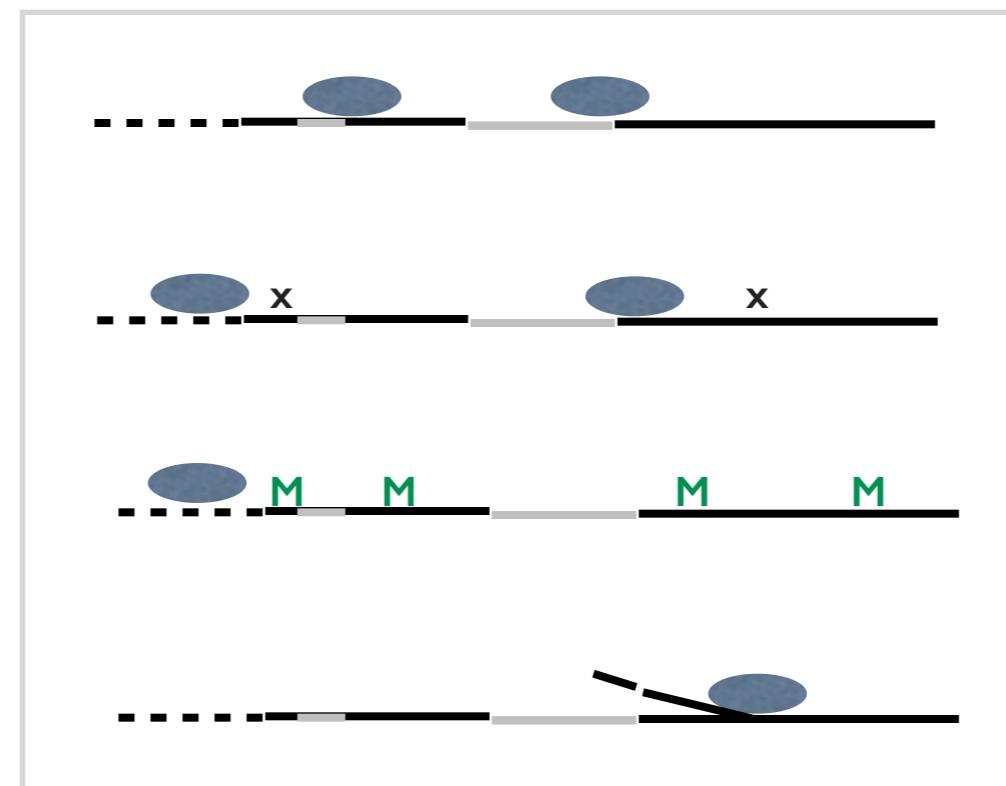
signal transduction

cell-cell signalling

**sparse methylation**

*tissue / temporal specific and inducible genes*

## Transcriptional opportunities



alternative start sites

sequence mutation  
change AA, premature stop codon

conventional transcription  
*transient methylation*

alternate transcript  
exon skipping

Gene  
expression



Epigenetic  
variation

promoter exon intron exon



Gene  
expression



Epigenetic  
variation

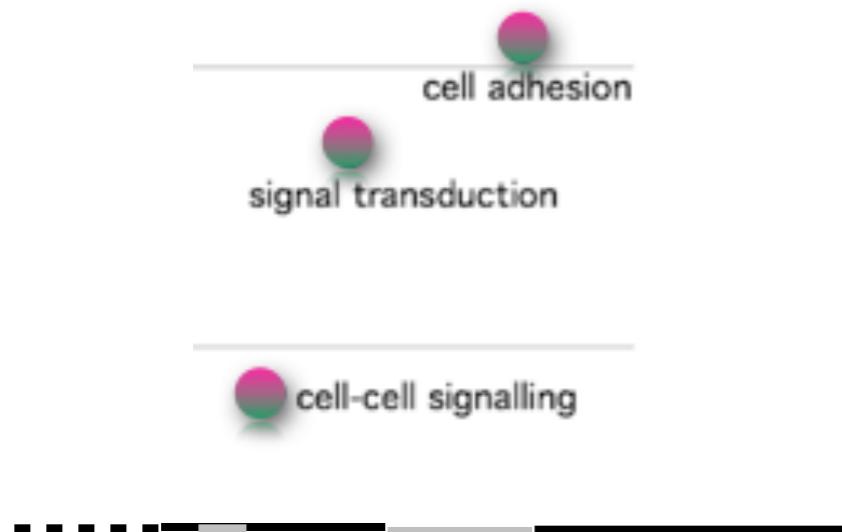
transcript



Gene expression

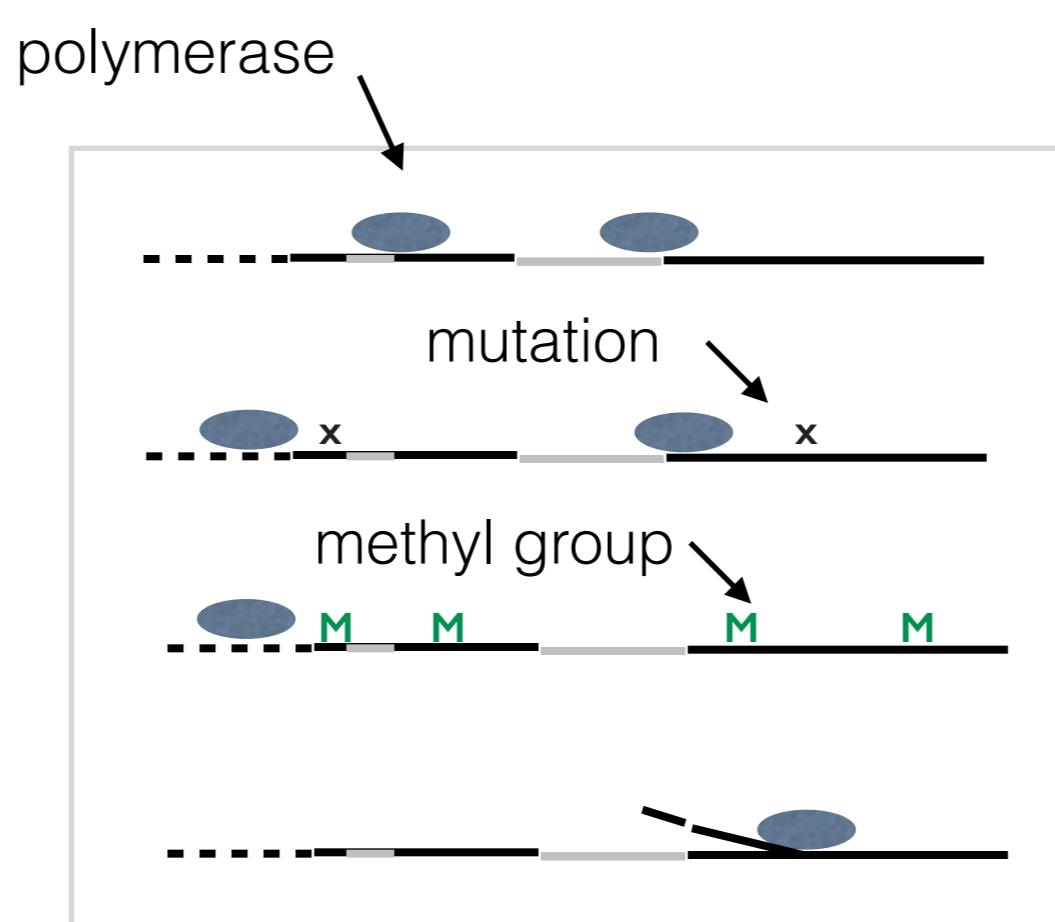
2

Epigenetic variation



## sparse methylation

*tissue / temporal specific and inducible genes*



Gene  
expression



Epigenetic  
variation

## Transcriptional opportunities



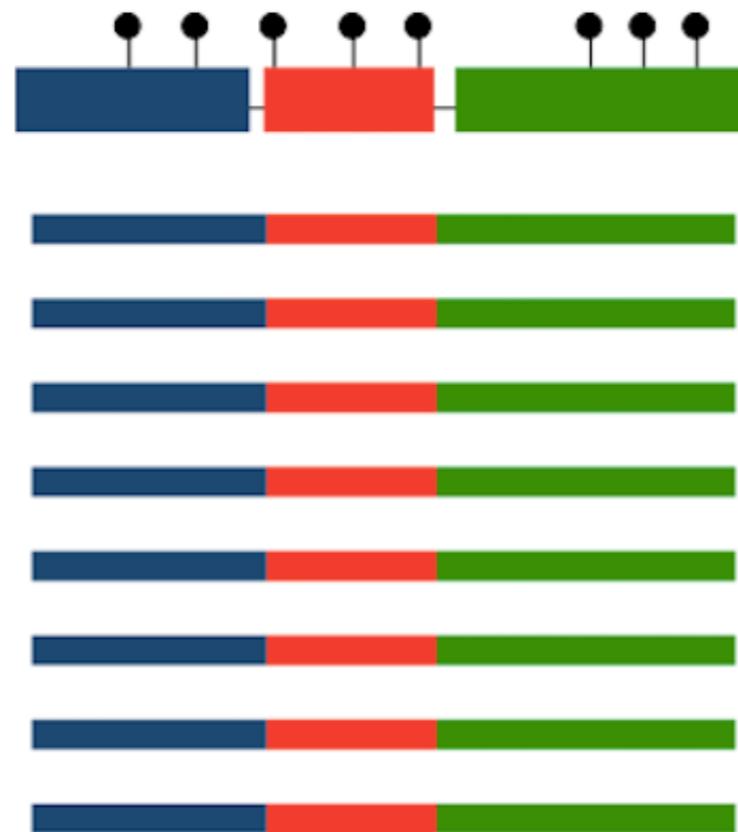
**germline methylation**

*ubiquitously expressed,  
critical genes*

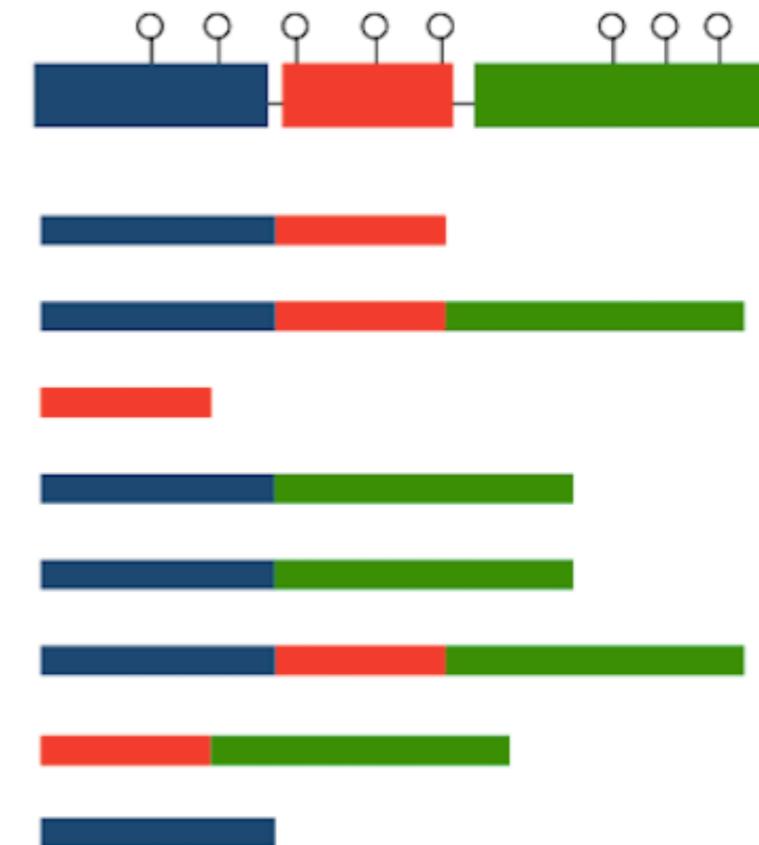
methyl group



conventional transcription  
5' UTR promotor



housekeeping

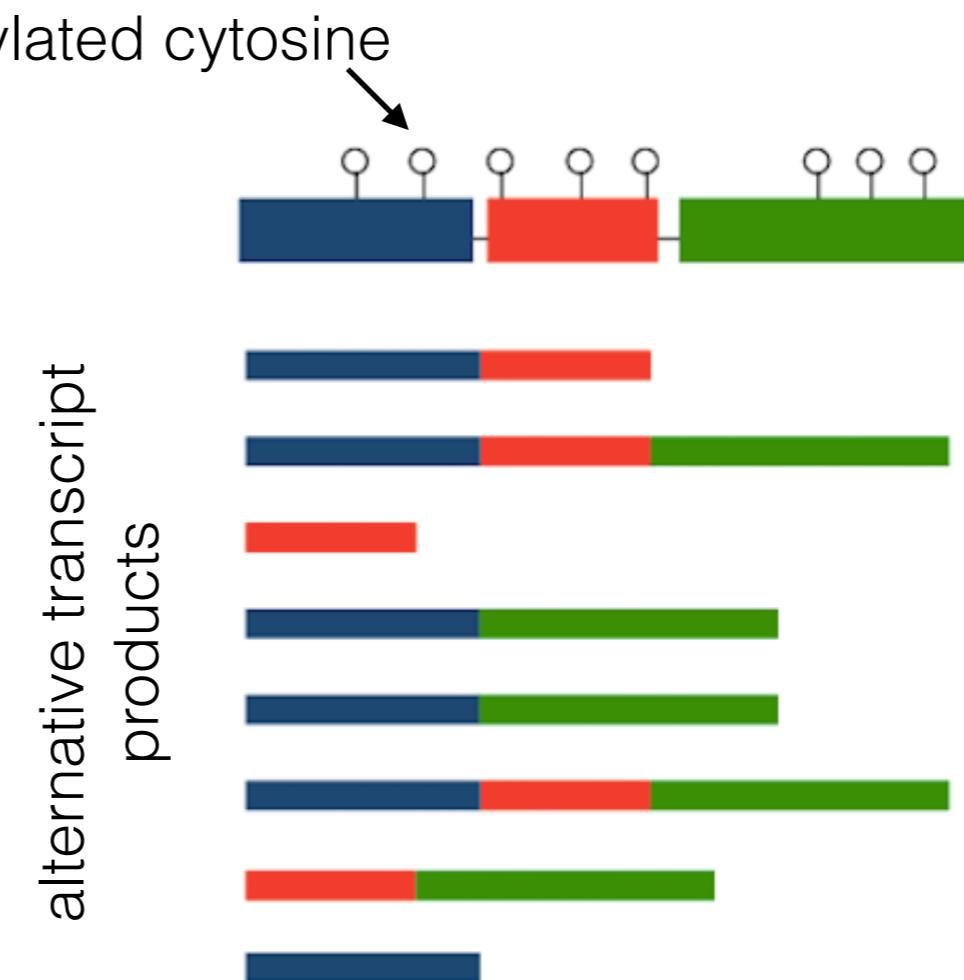
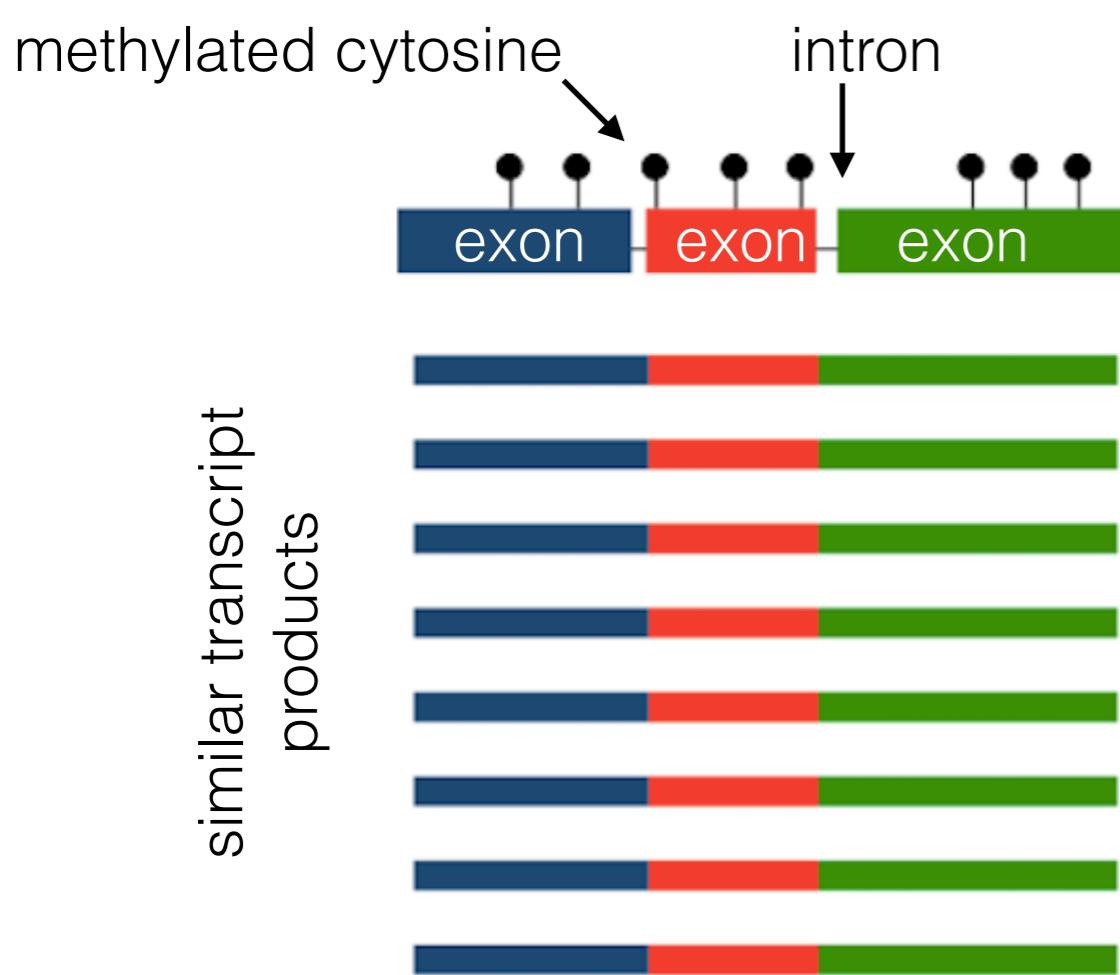


response to  
change

# A context dependent role for DNA methylation in bivalves

Mackenzie R. Gavery and Steven B. Roberts

Advance Access publication date 7 January 2014



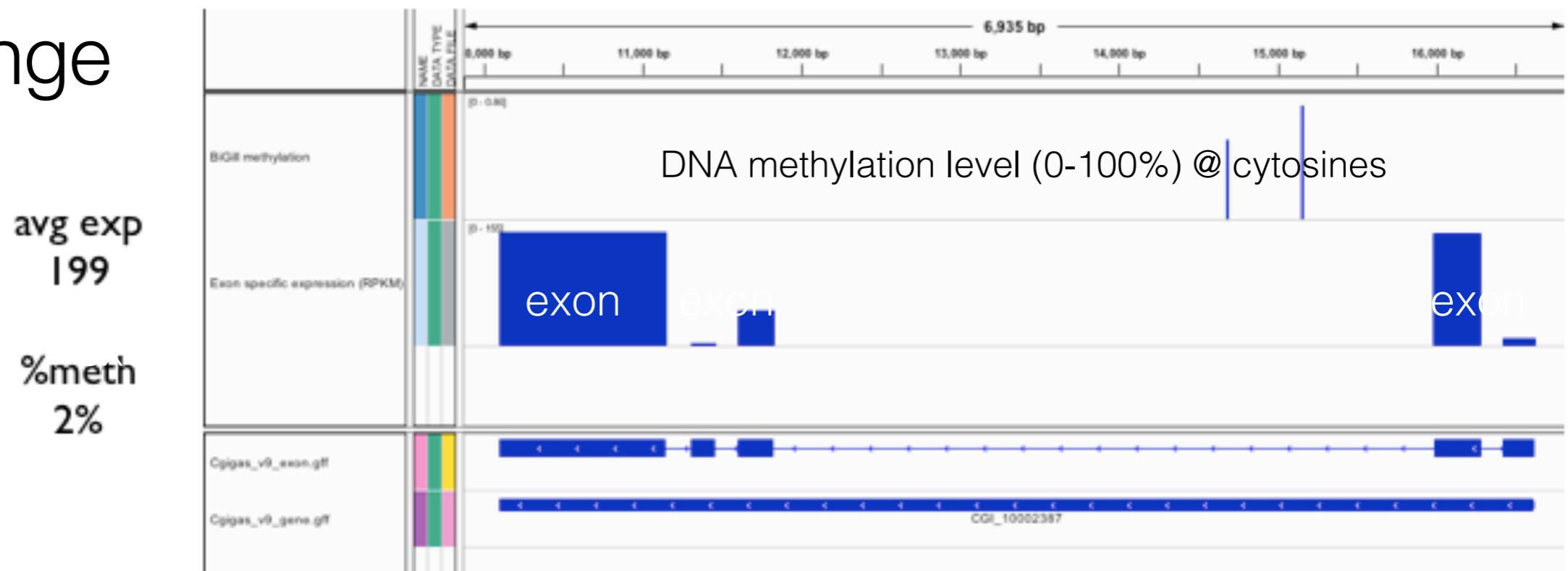
# A context dependent role for DNA methylation in bivalves

Mackenzie R. Gavery and Steven B. Roberts

Advance Access publication date 7 January 2014

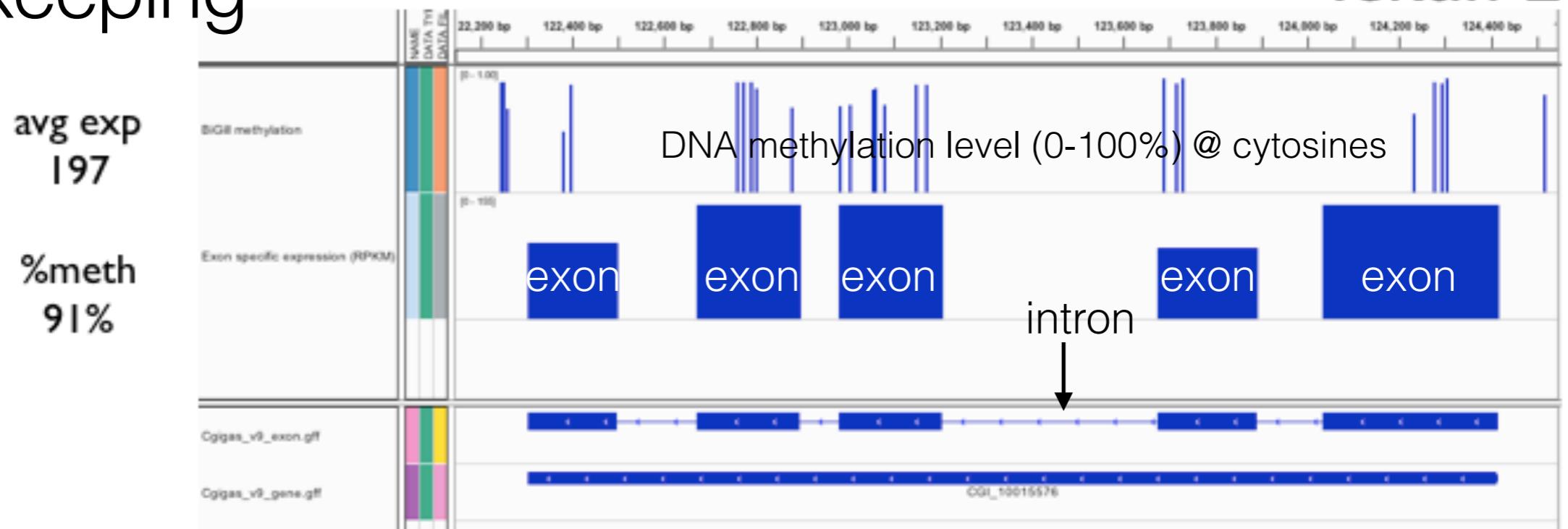
response to  
change

# Heat shock 70 kDa protein 12A



housekeeping

# Tektin-2



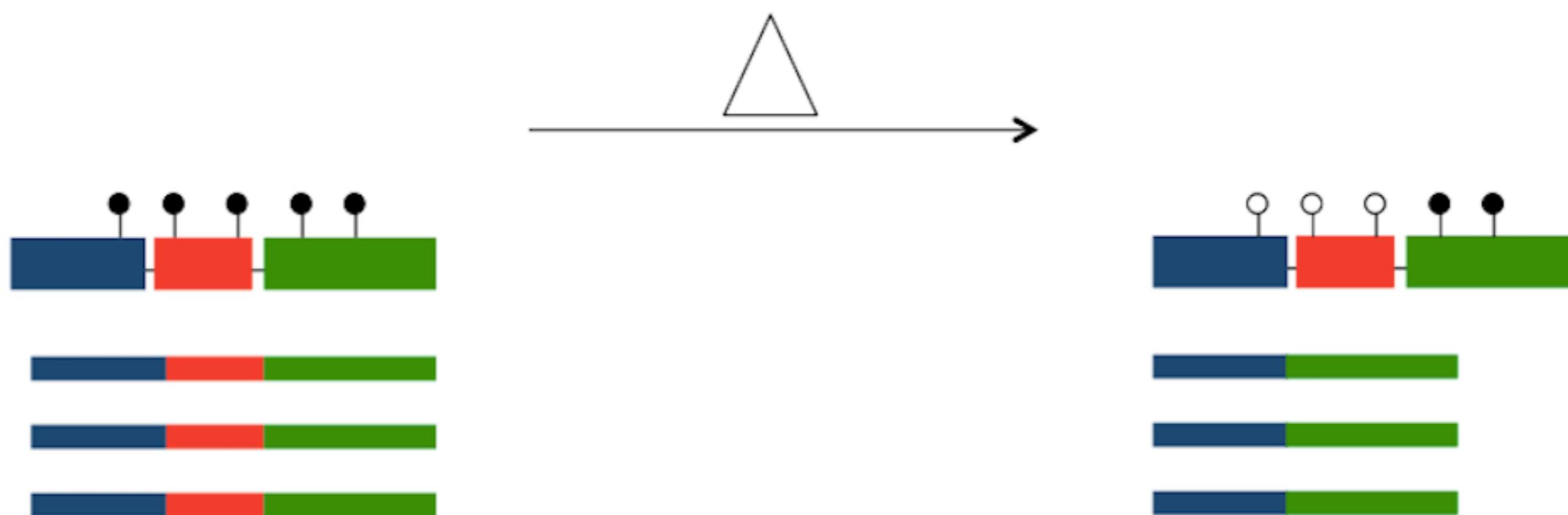


# Targeted Regulation

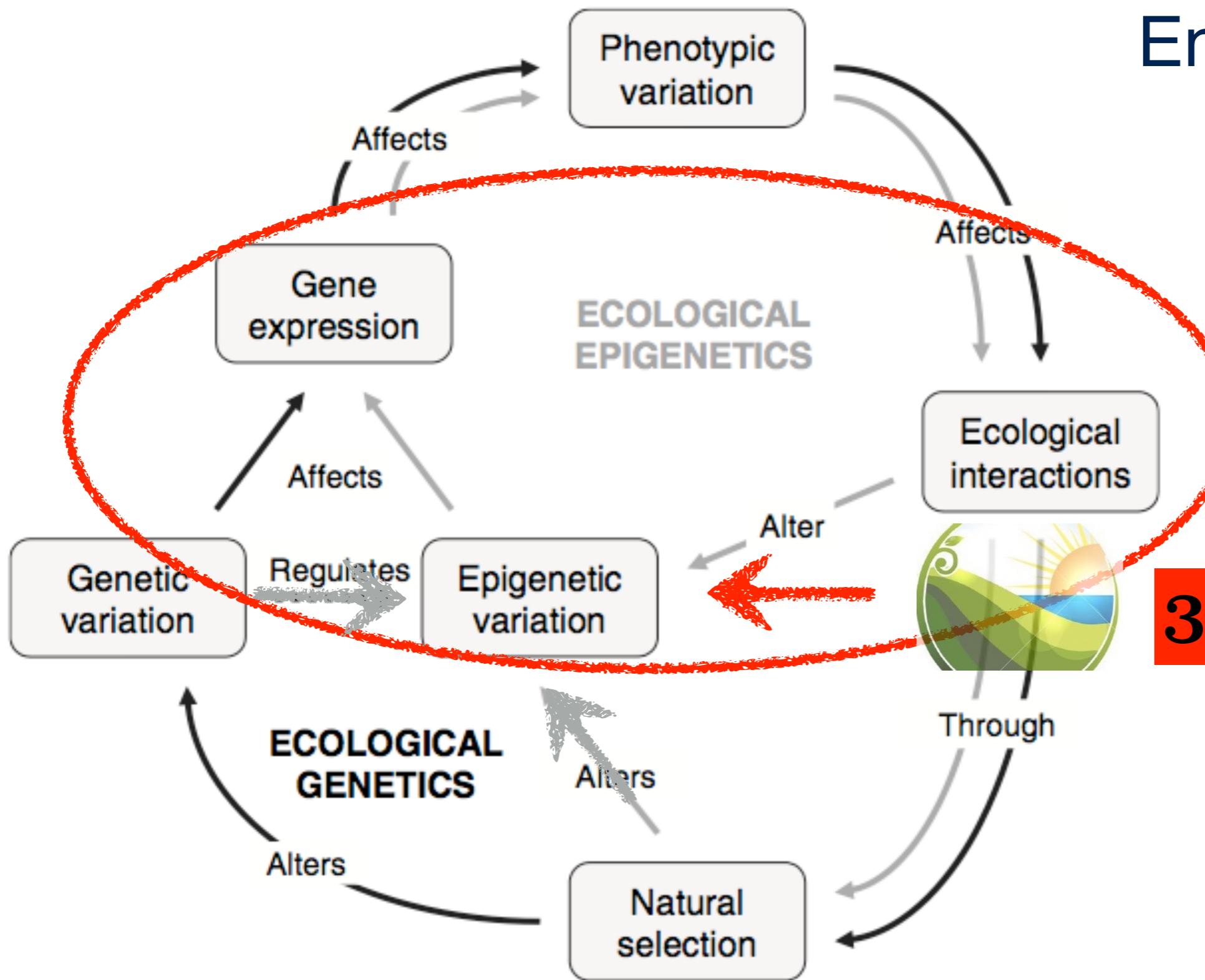
## A context dependent role for DNA methylation in bivalves

Mackenzie R. Gavery and Steven B. Roberts

Advance Access publication date 7 January 2014



# Environmental Influence



*Ecology Letters*, (2008) 11: 106–115

doi: 10.1111/j.1461-0248.2007.01130.x

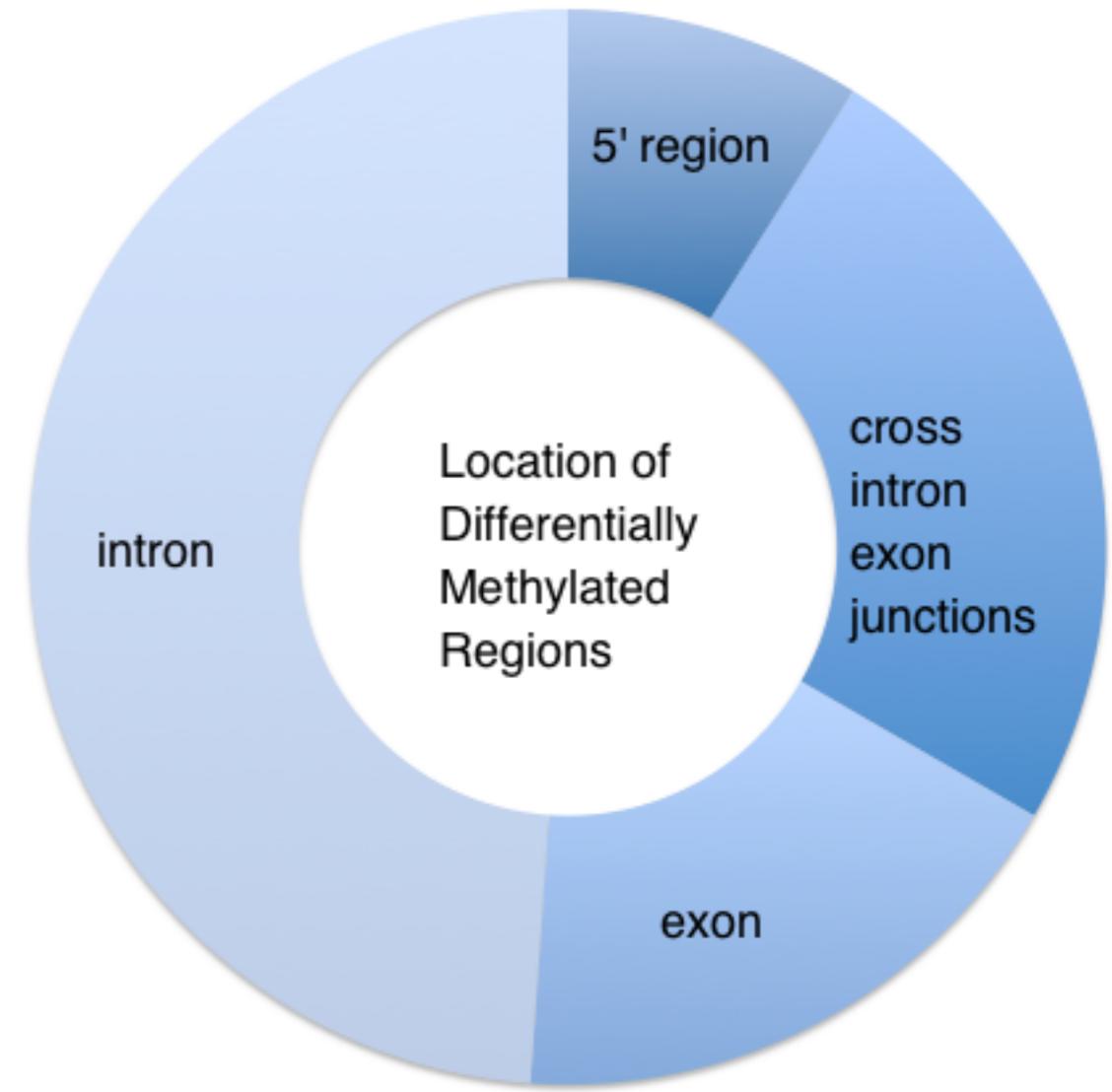
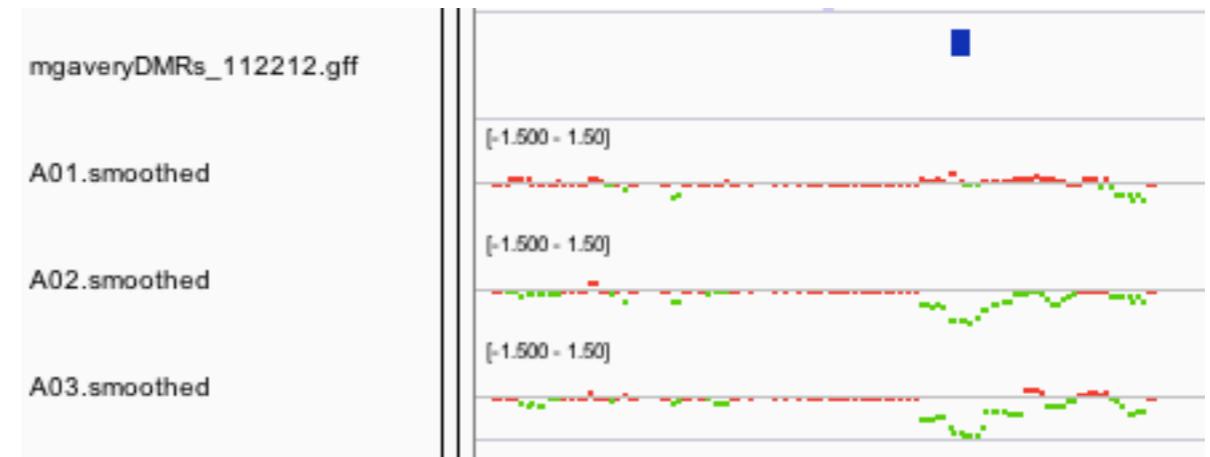
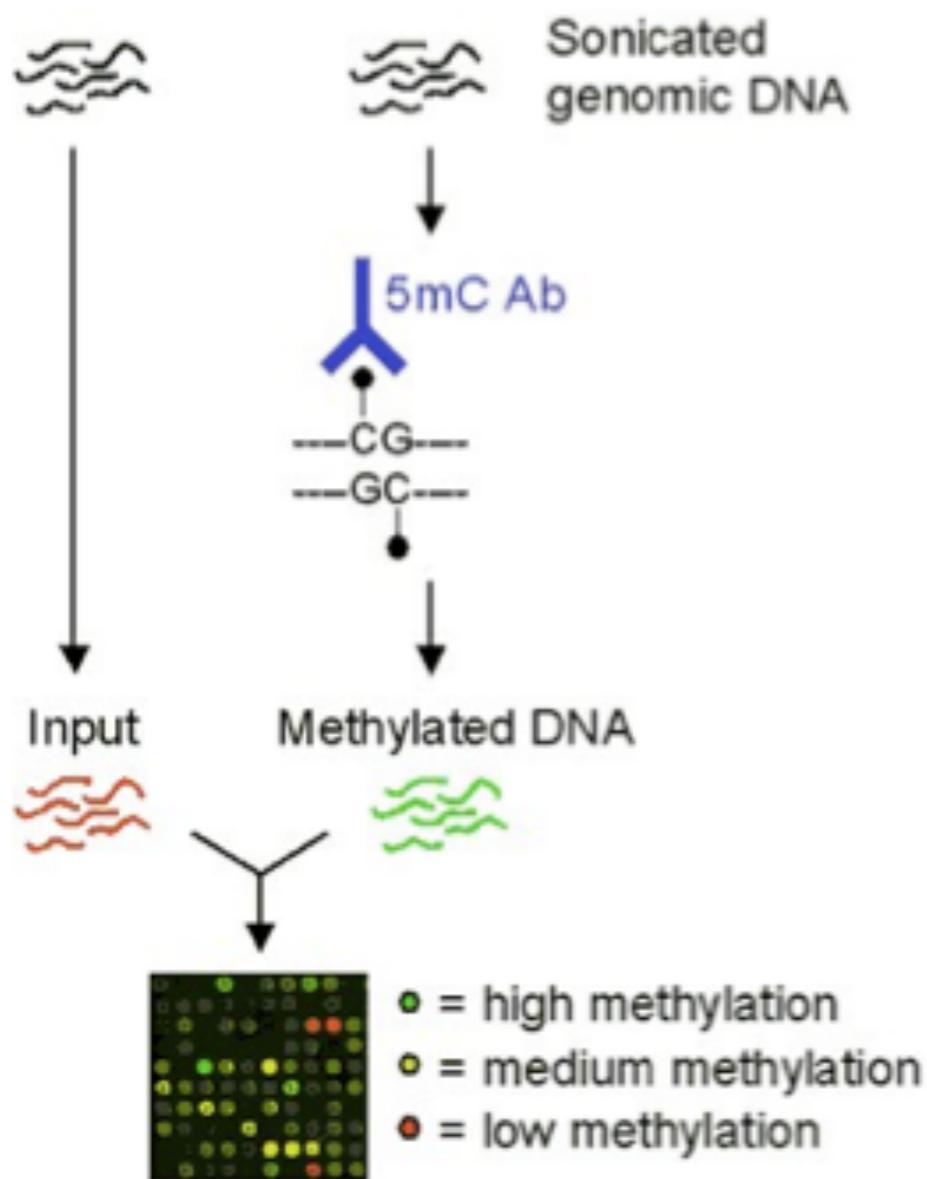
modified from

IDEA AND  
PERSPECTIVE

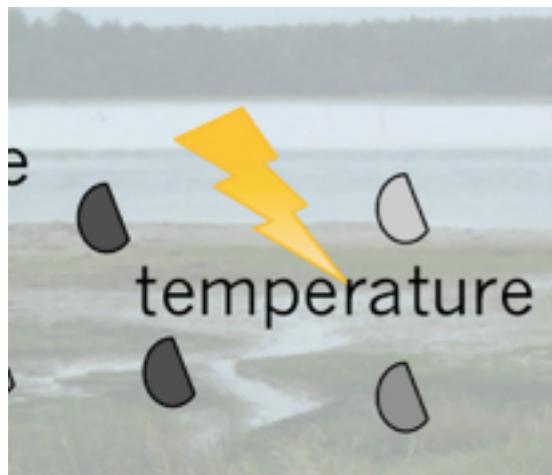
Epigenetics for ecologists

Oliver Bossdorf,<sup>1,\*</sup> Christina L.  
Richards<sup>2</sup> and Massimo Pigliucci<sup>3</sup>

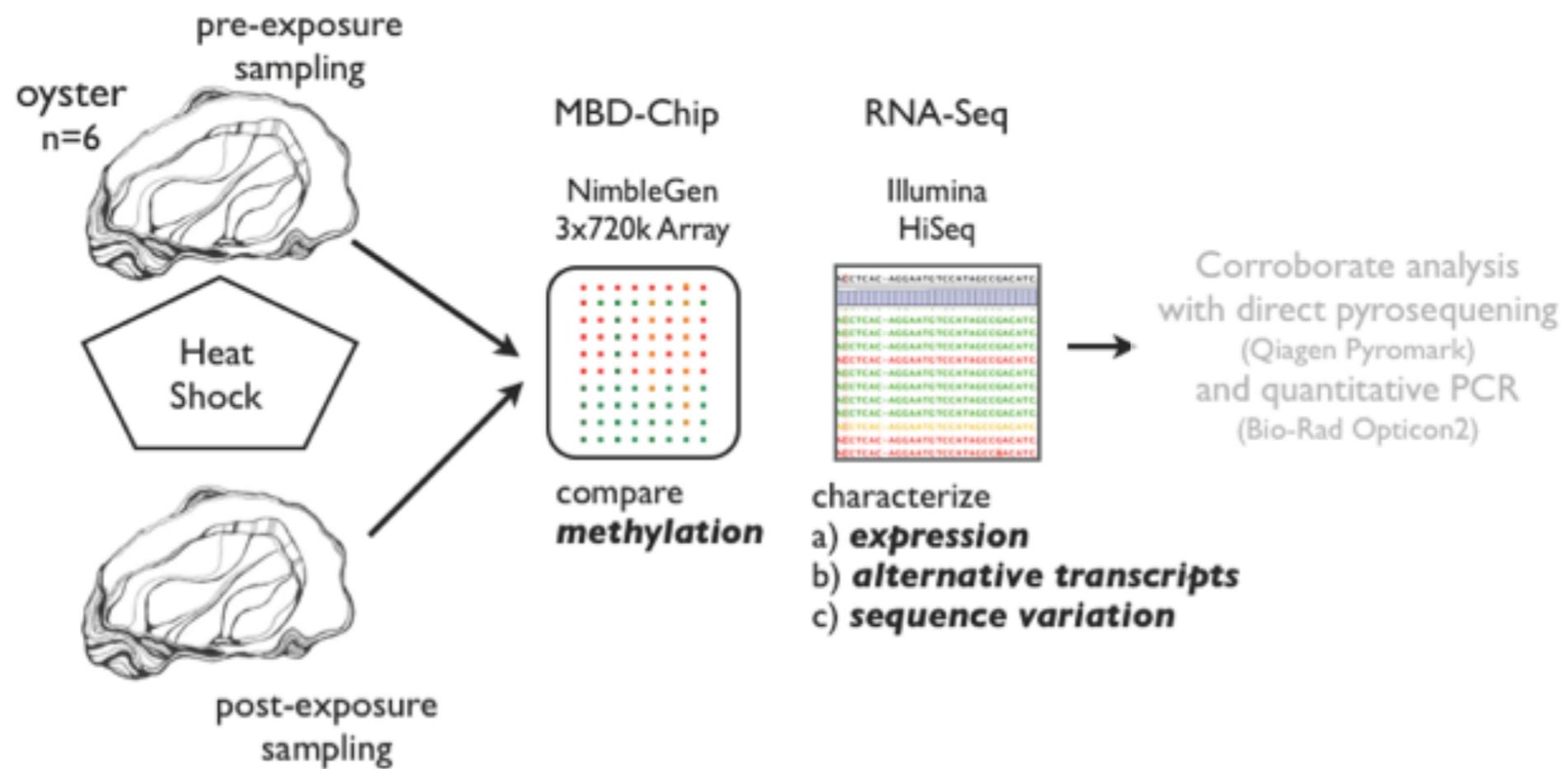
# Environmental impact (Estrogens)



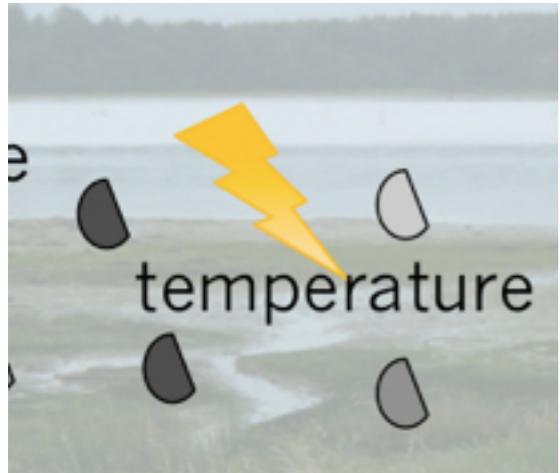
# Environment and gene expression



# *stochastic or targeted?*



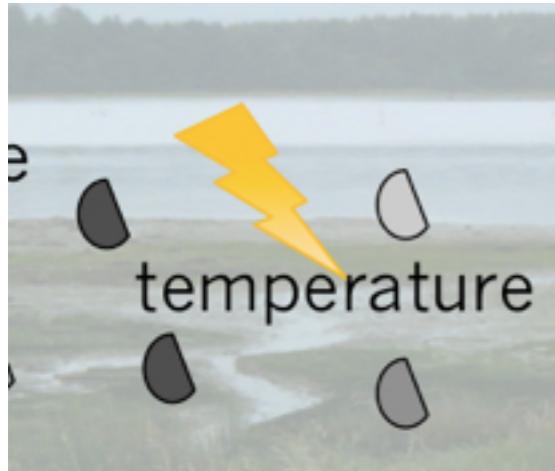
# Environment and gene expression



*stochastic or targeted?*

Oyster	Hypo-methylated	Hyper-methylated
2	7224	2803
4	6560	3587
6	7645	4044

# Environment and gene expression

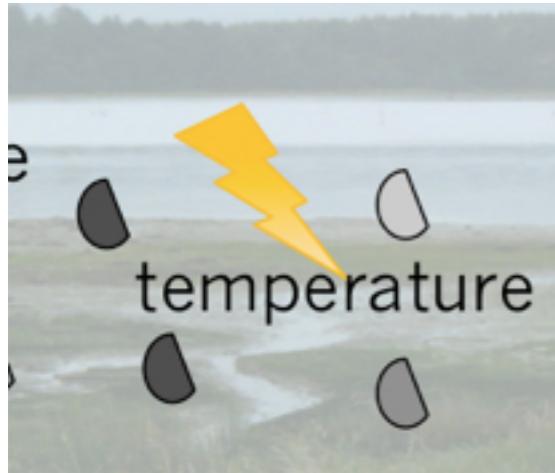


*stochastic or targeted?*

Oyster	Hypo-methylated	Hyper-methylated
2	7224	2803
4	6560	3587
6	7645	4044

No obvious association  
with genome feature  
including *differentially*  
*expressed*  
*genes*

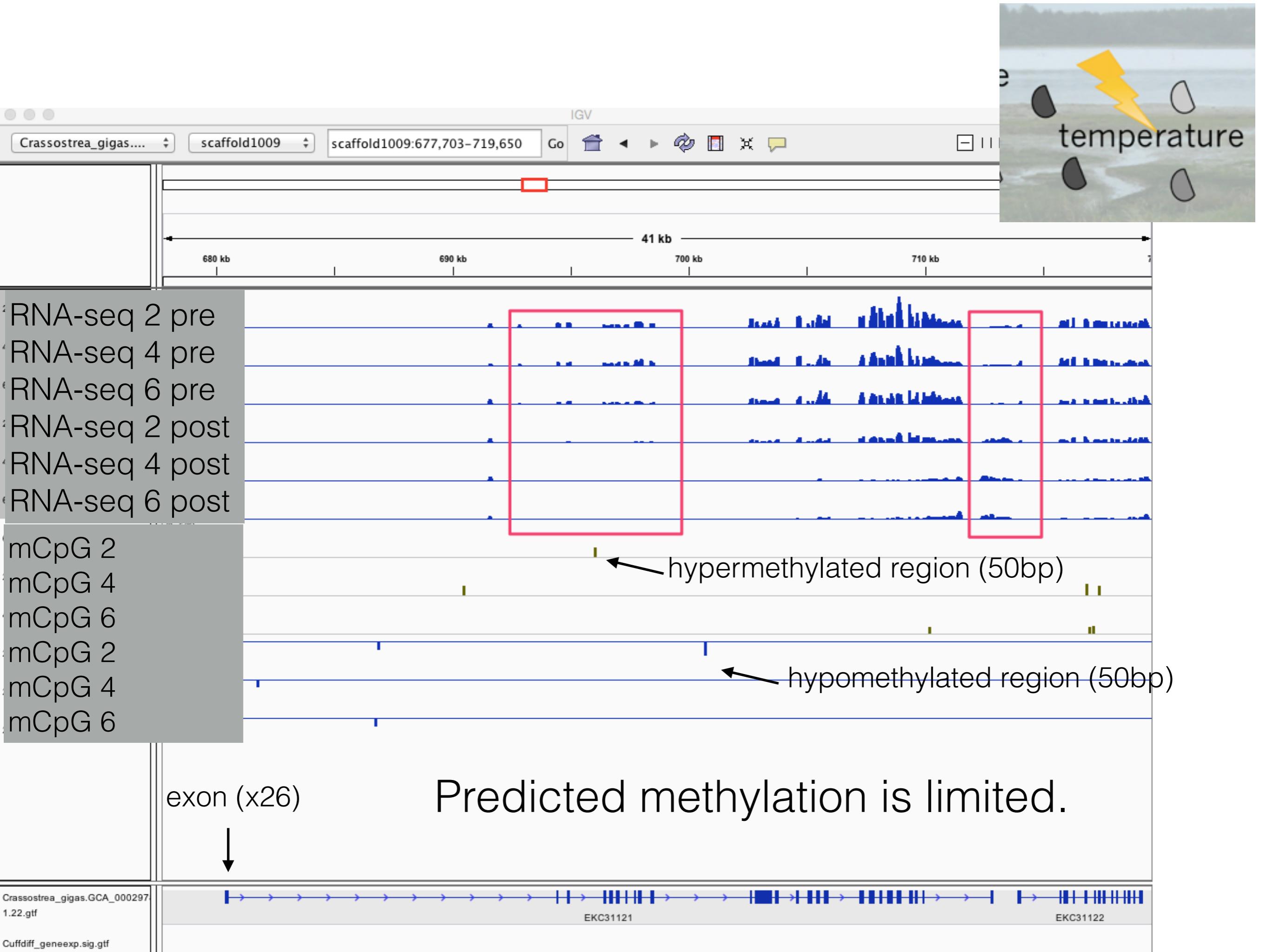
# Environment and gene expression



*stochastic or targeted  
or ..?*

Oyster	Hypo-methylated	Hyper-methylated
2	7224	2803
4	6560	3587
6	7645	4044

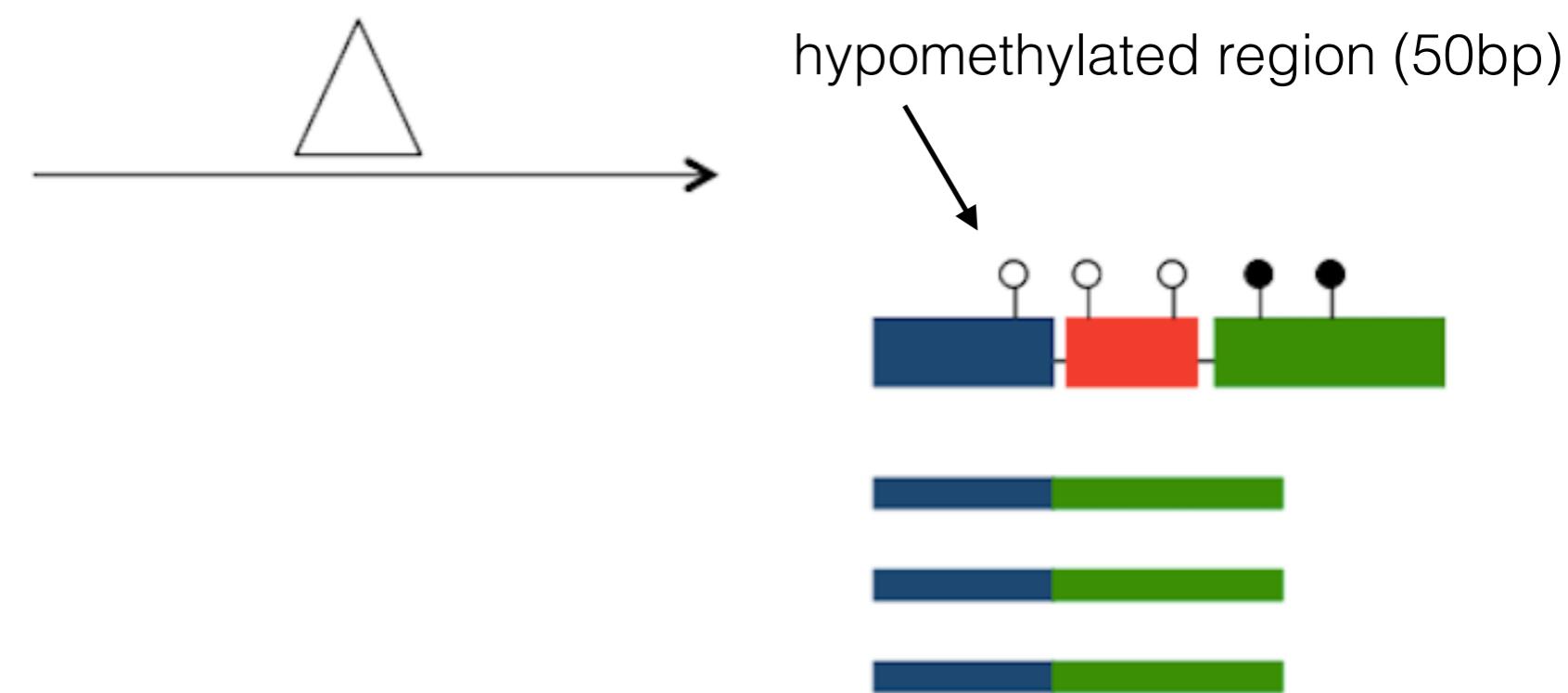
Changes in methylation (either direction) are more prevalent in introns, repeats, and transposable elements.



Gene  
expression

2

Epigenetic  
variation



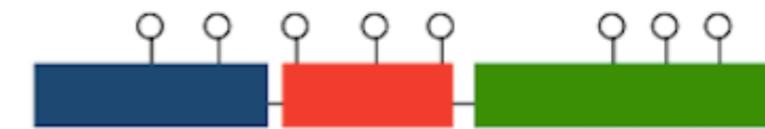
**not** in this experiment

not even consistent methylation changes at loci level

Gene  
expression

2

Epigenetic  
variation



more questions....  
interesting but what is controlling?

# Take Home

*Oyster genome has a fantastic degree of diversity contributing to phenotypic plasticity & adaptation potential.*

# Take Home

*Oyster genome has a fantastic degree of diversity contributing to phenotypic plasticity & adaptation potential.*

- large gene families
- very high mutation rate (snp/50bp)
- numerous exons per gene (potential for alternatives)
- genome full of repeats region
- high number of transposable elements
- lack of methylation of transposable suggest mobility
- family variation of methylation
- limited methylation environmental response genes  
is associated with spurious transcription
- inheritance of epigenetic marks as mechanism of improved adaptation

A photograph of a dense tropical forest. In the foreground, a large tree trunk with dark, textured bark is visible on the left. To its right, several thick branches of another tree are covered in bright green moss. The background is filled with more trees and foliage, creating a lush, dark green environment.

## Next Steps

Transposable  
Elements

Consider other  
epigenetic process

# Next Steps

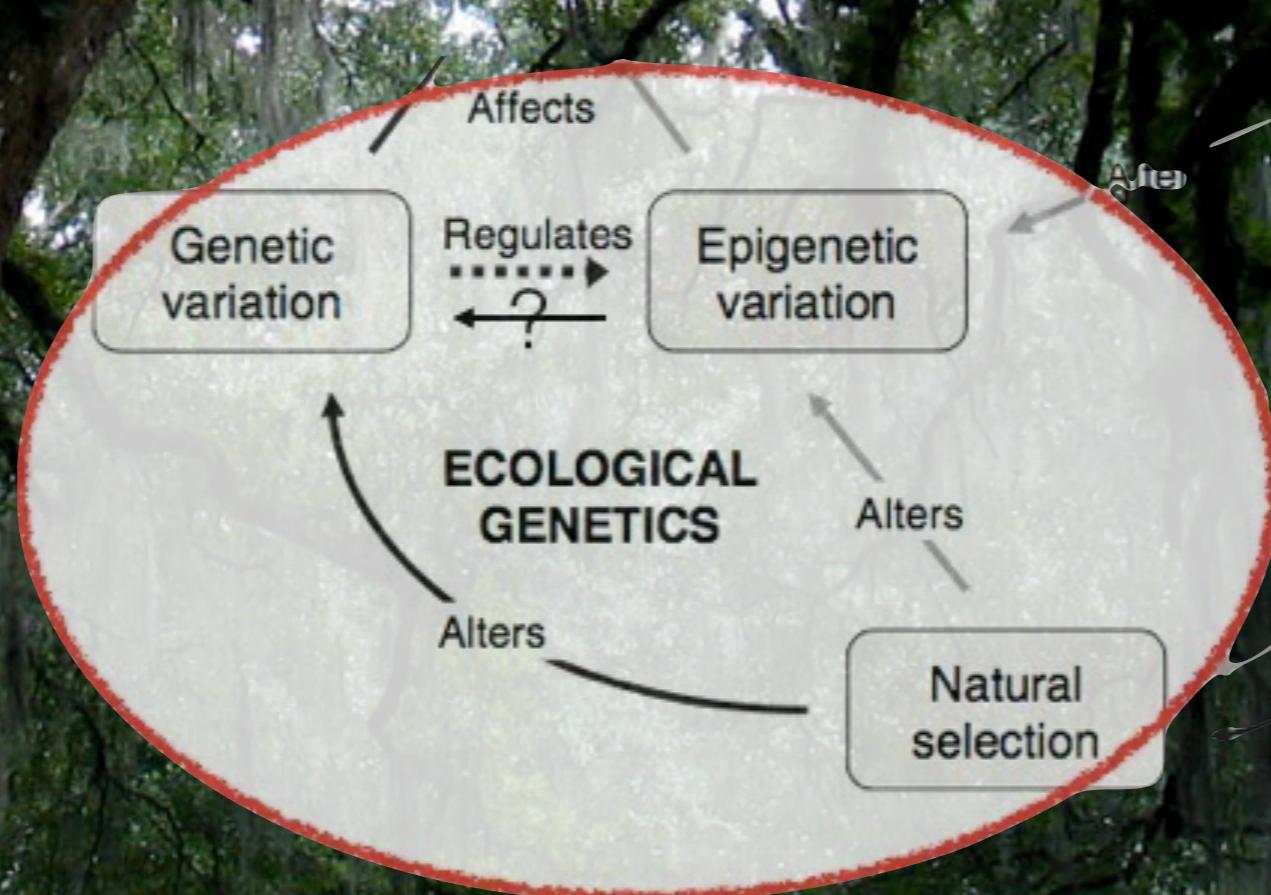


Photo credit: Flickr, Creative Commons, csessums

~~Very new data~~  
Heritability  
Plasticity  
Local Adaptation

*Genetics  
versus  
Epigenetics*

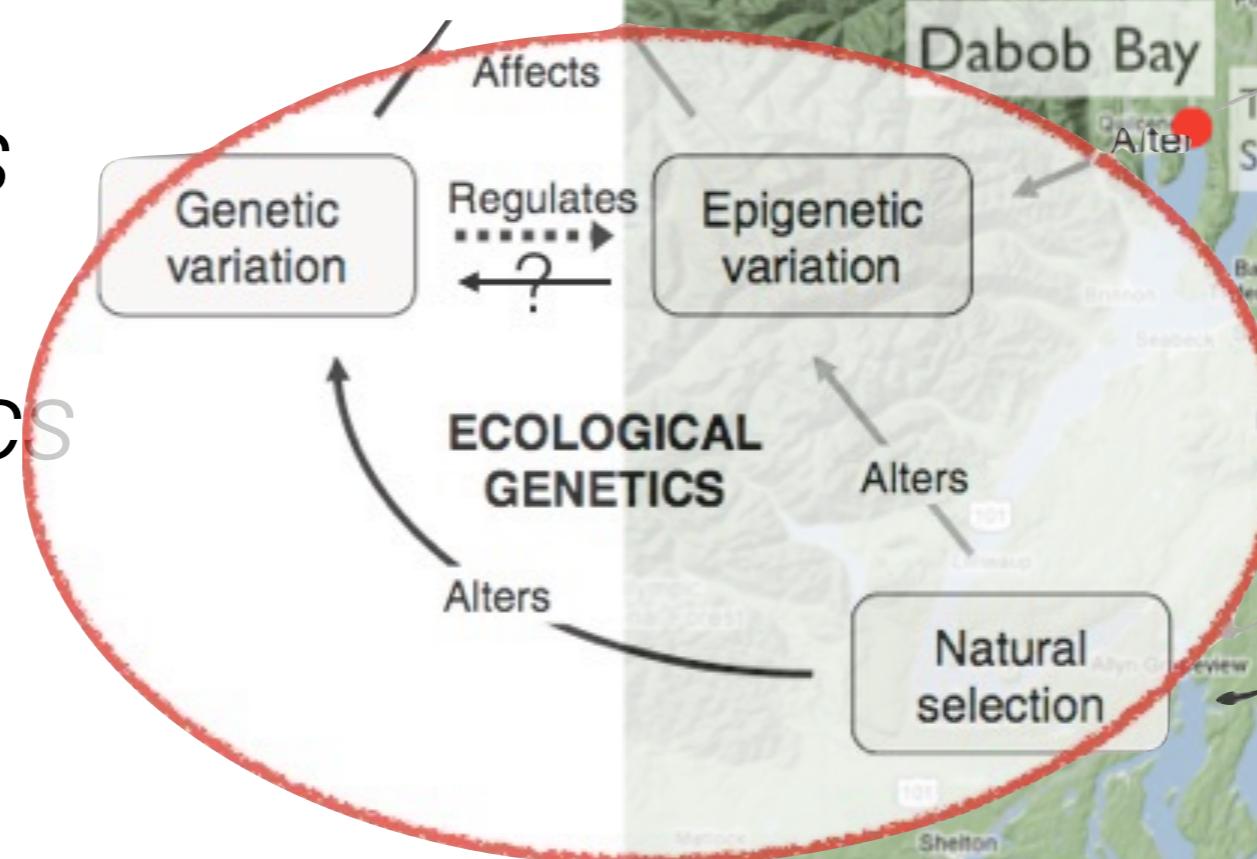


## Reciprocal Transplant Experiment

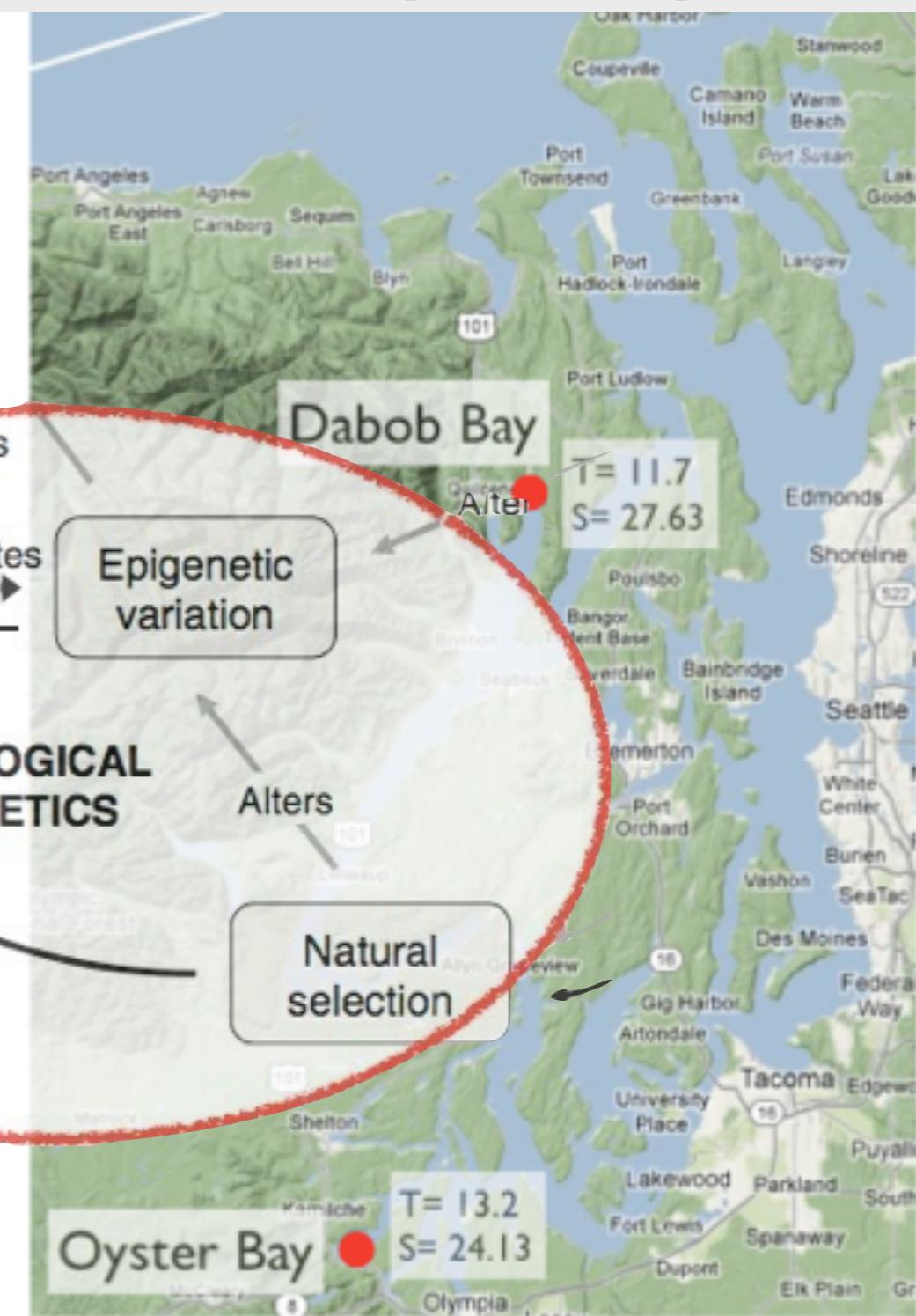


~~Very new data~~  
Heritability  
Plasticity  
Local Adaptation

Genetics  
*versus*  
Epigenetics



## Reciprocal Transplant Experiment



# Open Science

about science

web-native scholarship

# Sharing



Photo credit: Flickr, Creative Commons, speechless

# Open Data

# Open Methods

# Open Data

Sharing raw data

# Approach



Data Lake

# Raw Data

1) As sequencing facility provides data, files are downloaded to our local NAS (owl), in the root `nightingales` directory.  
<http://owl.fish.washington.edu/nightingales/>

2) The Nightingales Google Spreadsheet is updated.

3) Update the Nightingales Google Fusion Table with new information from the Nightingales Google Spreadsheet. This is accomplished by:

- deleting all rows in the Nightingales Google Fusion Table (Edit > Delete all rows)
- Importing data from the Nightingales Google Spreadsheet (File > Import more rows...)

4) Once metadata is included in the Google Fusion Table [Nightingales](#), respective data files are moved to subdirectory labelled with first letter of genus followed by species ie `c_gigas`. Check url in Nightingales Fusion table to ensure it is accurate.

**Nightingales**

Imported at Wed Jun 25 07:25:38 PDT 2014 from Nightin  
Add Attribution - Edited on December 15, 2014

File Edit Tools Help Rows 1 ▾

Filter ▾ No filters applied. Sorted by SeqDat

◀ ▶ 1-100 of 153 ▶

SeqID
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

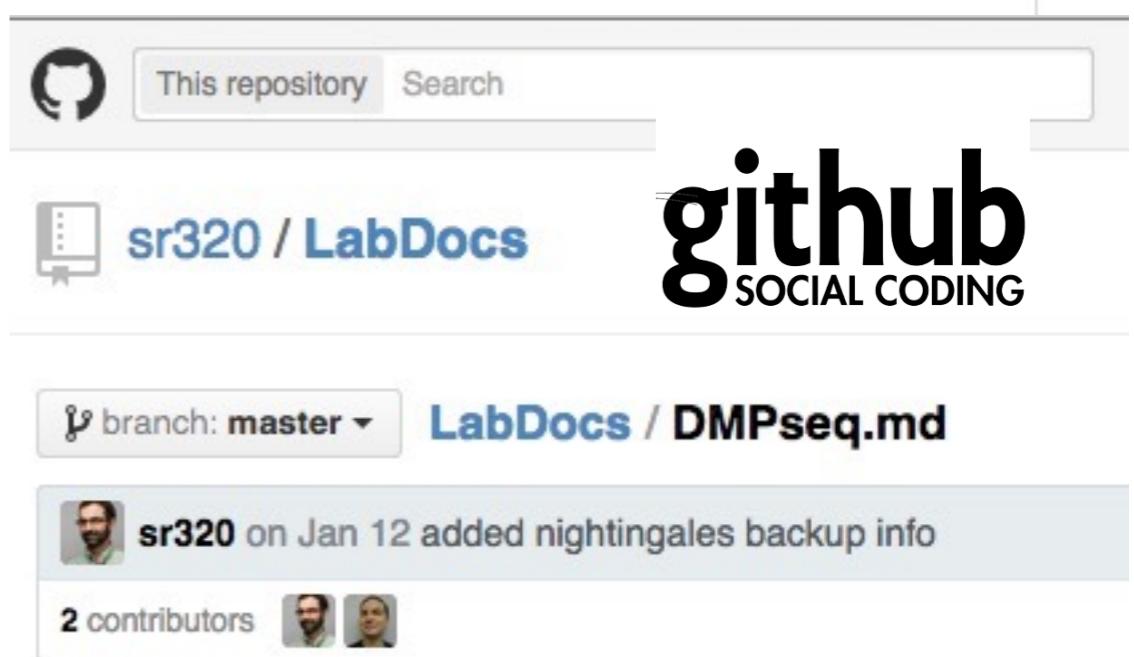
This repository Search

sr320 / LabDocs

branch: master LabDocs / DMPseq.md

sr320 on Jan 12 added nightingales backup info

2 contributors



# Querying Disparate Datasets

## Querying Disparate Oyster Datasets | qDOD

The goal of this project is to produce a web-based interface for querying and visualizing *Crassostrea gigas* genomic datasets. This site serves as a portal for documenting our efforts, providing user access, as well as a means to gather feedback.

DATA

### Preliminary Phase: Aggregating Datasets

Using SQLShare as a platform we have already begun to aggregate and format data. Anyone can view (and contribute) using the tag "qdod". Below is a table describing some of the relevant datasets. "Snapshot" provides you with a screenshot of the data in SQLShare and "Direct Link" brings you directly to the data in SQLShare. You can also open the table in a new webpage.

#### qDOD online

qDOD_Cgigas_gene_fasta	sequence fasta file. Exon only.	<a href="http://goo.gl/oqCxl">http://goo.gl/oqCxl</a>	<a href="https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Cgigas_gene_fasta">https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Cgigas_gene_fasta</a>
qDOD_Zhang_Gil_gene_RNA-seq	Gill RNA-seq data (gene based)	<a href="http://goo.gl/8oISR">http://goo.gl/8oISR</a>	<a href="https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Zhang_Gil_gene_RNA-seq">https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Zhang_Gil_gene_RNA-seq</a>
qDOD_Zhang_Mgo_gene_RNA-seq	Male Gonad RNA-seq data (gene based)	<a href="http://goo.gl/6buVz">http://goo.gl/6buVz</a>	<a href="https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Zhang_Mgo_gene_RNA-seq">https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Zhang_Mgo_gene_RNA-seq</a>



# DATA

## Preliminary Phase: Aggregating Datasets

Using SQLShare as a platform we have already begun to aggregate and format data. Anyone can view (and contribute) using the tag "qdod". Below is a table describing some of the relevant datasets. "Snapshot" provides you with a screenshot of the data in SQLShare and "Direct Link" brings you directly to the data in SQLShare. You can also open the table in a new webpage.

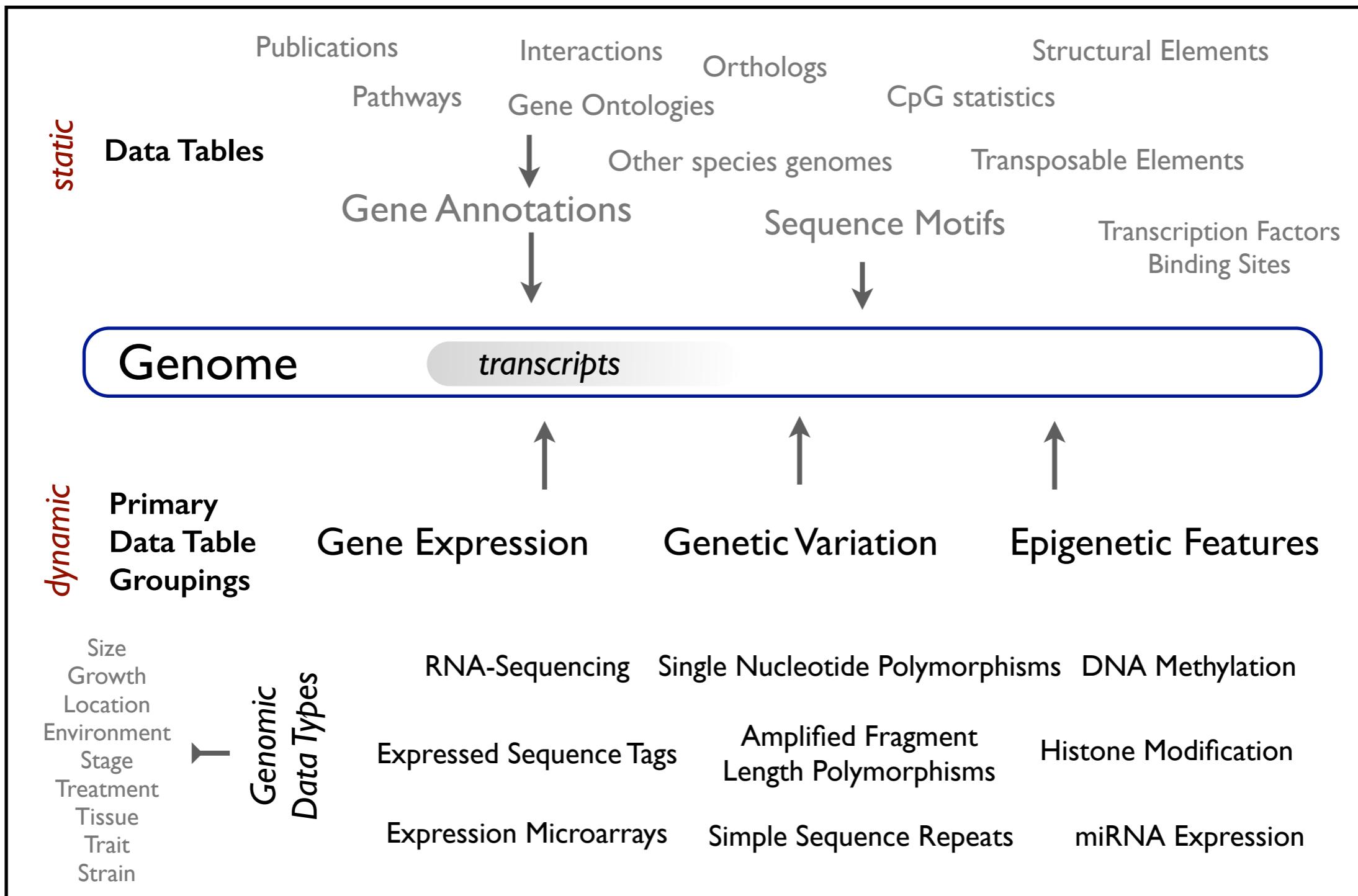
### qDOD online

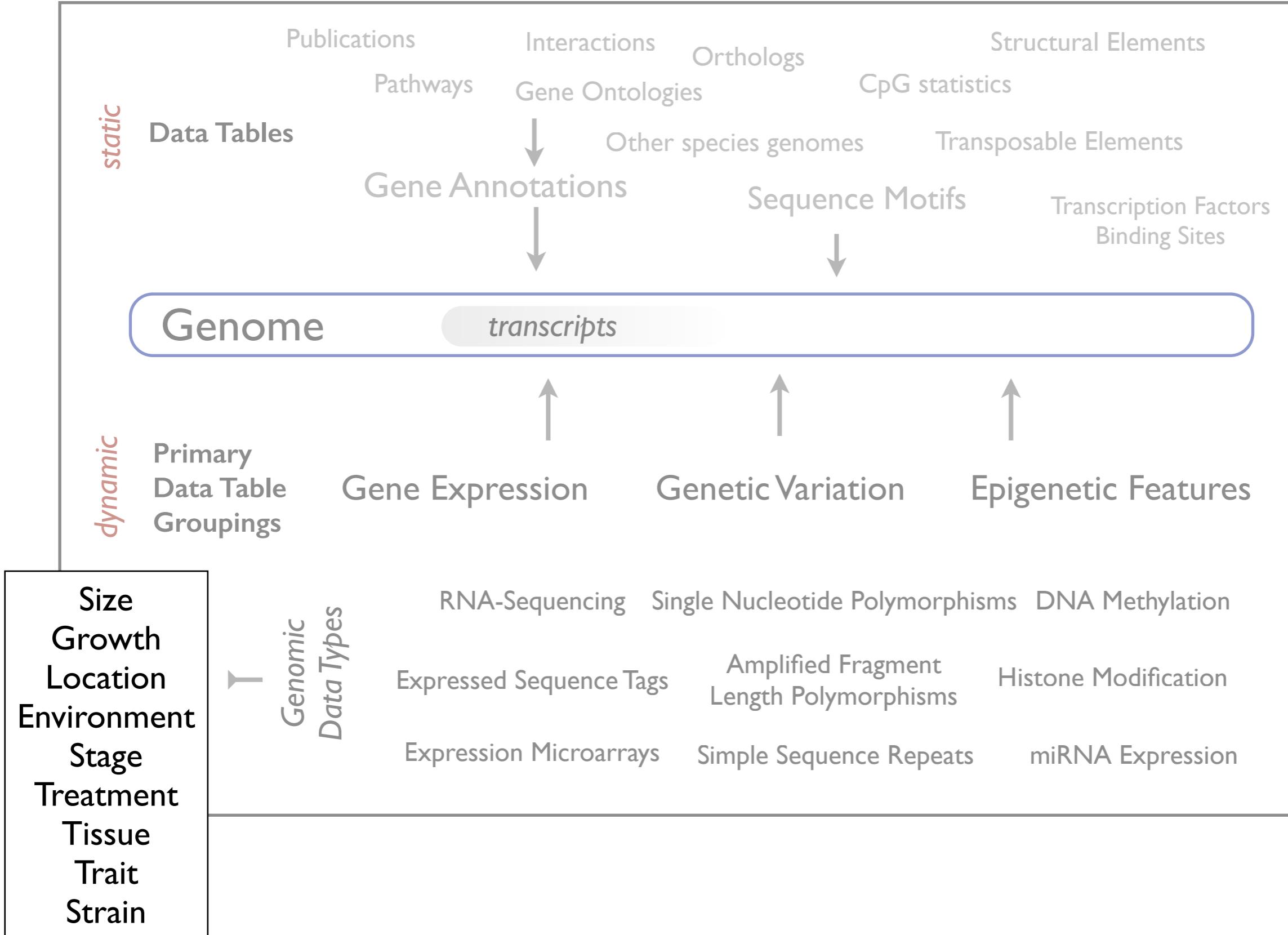
qDOD_Cgigas_gene_fasta	sequence fasta file. Exon only.	<a href="http://goo.gl/oqCxl">http://goo.gl/oqCxl</a>	<a href="https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Cgigas_gene_fasta">https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Cgigas_gene_fasta</a>
qDOD_Zhang_Gil_gene_RNA-seq	Gill RNA-seq data (gene based)	<a href="http://goo.gl/8oISR">http://goo.gl/8oISR</a>	<a href="https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Zhang_Gil_gene_RNA-seq">https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Zhang_Gil_gene_RNA-seq</a>
qDOD_Zhang_Mgo_gene_RNA-seq	Male Gonad RNA-seq data (gene based)	<a href="http://goo.gl/6buVz">http://goo.gl/6buVz</a>	<a href="https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Zhang_Mgo_gene_RNA-seq">https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Zhang_Mgo_gene_RNA-seq</a>

Your datasets
All datasets
Shared datasets
Recent activity... 293
Recently viewed »
Upload dataset
New query
YOUR TOP VIEWED
qDOD_scaffol... 35
oyster_v9_mR... 28
oyster_v9_CD... 23
BiGO_betty_pl... 16
qDOD_Cgigas... 16
POPULAR TAGS
protein 47
oyster 45
proteomics 21
bioinformatics 21
oa 20

## Tagged Datasets | qdod

Name	Sharing /
BiGO_Methylation_oysterv9_GFF GFF format file with percent methylation (score) information for bsmap gff gonad methylation oyster qdod sperm	sr320
qDOD_gene_length Sequence lengths of all genes in the oyster genome. This only includes trans fasta gene length oyster qdod	sr320
BiGill_methratio_v9_A.txt Methylation ratio information (MBD-Seq) from oyster gill tissue. Combin bsmap epigenetic gill mbd methylation oyster qdod	sr320
qDOD_Protein_Sequences Amino acid sequence for all proteins (v9) fasta oyster protein qdod	sr320
qDOD_oyster_gene_exon_number Number of exons for all genes in oyster genome (v9) exon oyster qdod	sr320
Zhang_et al_SuppTable14 S14. Transcriptomic representation of genes (RPKM) at different develo oyster qdod zhang	sr320
qDOD_Cgigas_gene_fasta Tabular format of Cgigas gene sequence fasta file Derived using Data fasta oyster qdod	sr320
qDOD_Zhang_Gil_gene_RNA-seq Paired end Gill RNA-Seq data from Zhang et al 2012. Exported gill oyster qdod rna-seq zhang	sr320
Mgo_RNAseq_transcript Paired end Male Gonad RNA-Seq data from Zhang et al 2012 Exported f oyster qdod rna-seq zhang	sr320
qDOD_Zhang_Mgo_gene_RNA-seq Paired end Male Gonad RNA-Seq data from Zhang et al 2012 gonad oyster qdod rna-seq sperm zhang	sr320





```
SELECT cgslim.CGI_ID, Description, evalue, SPID, GOID, term, GOSlim_bin, sequence
  FROM [sr320@washington.edu].[qDOD_Cgigas_GO_GOslim] cgslim
LEFT JOIN [sr320@washington.edu].[qDOD_Cgigas_gene_fasta] cfg
    ON cgslim.CGI_ID = cfg.CGI_ID
Where term LIKE '%methyl%'
OR
term LIKE '%histone%'
```

What genes are associated with epigenetics?

What genes are associated with immune response?

Which genetic markers are associated with

fast growth?

resilience?

disease tolerance?

# Open Methods

# Open Methods

reproducibility

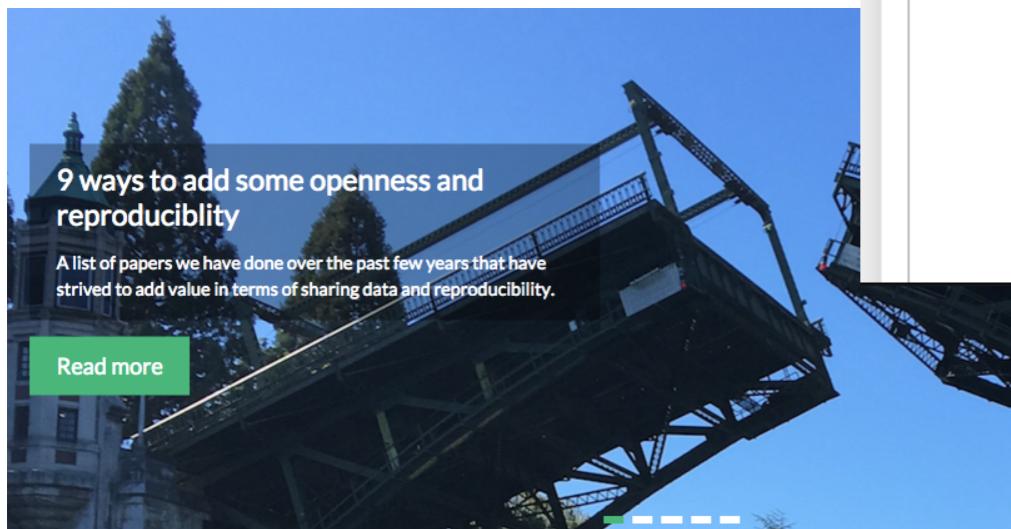
# Data Acquisition and Analysis

---

Open  
Notebook  
Science

# Data Acquisition and Analysis

# Open Notebook Science



jupyter Welcome3 (autosaved)

File Edit View Insert Cell Kernel Help Python 3

Cell Toolbar: None

## Welcome

Welcome to Pineapple, the next generation scientific notebook.

### Run Python code

```
In [1]: 2 ** 64
```

```
Out[1]: 18446744073709551616
```

### You can make plots right in the notebook

```
In [2]: %matplotlib inline
import matplotlib.pyplot as plt
import numpy as np
x = np.linspace(0, 10)
plt.plot(x, np.sin(x));
```

jldimond / Coral-CpG

Unwatch 2

Repository associated with article "Germline DNA methylation in reef corals: patterns and potential roles in response to environmental change"

22 commits 1 branch 0 releases 2 contributors

Branch: master +

Update READMEs

jldimond authored 5 days ago latest commit d94c7b5aec

analyses Update READMEs 5 days ago

data typo 7 days ago

ipynb Update READMEs 5 days ago

README.md typo 7 days ago

README.md

Analyses and data associated with manuscript:  
*Germline DNA methylation in six species of reef corals: patterns and potential roles in response to environmental change*

File Edit View Insert Cell Kernel Help



Run

Run and Select Below

Run and Insert Below

Run All

Run All Above

Run All Below

Cell Toolbar: None

## Fasta2Slim

This IPython notebook is personal preferences. As

- IPython
- NCBI Blast
- SQLShare Python CLI

structured means to annotate sequences using UniProt/SwissProt requires the user has the following software installed ...

Cell Type

Current Output

All Output

### Instructions for use.

In a working directory of your choosing place query fasta file, naming as `query.fa`. Edit the cell below, providing th

Identify the location of the blast database you would like to use and indicate path in the cell below.

Identify the location of your `sqlshare-pythonclient/tools` and indicate path in the cell below.

Change the input to the `usr` variable to reflect your SQLShare user account.

```
In [2]: #Location Variables  
wd="~/Desktop/test/"  
  
db="/Volumes/Bay3/Software/ncbi-blast-2.2.29+/db/uniprot_sprot_r2013_12"  
  
sqls="~/sqlshare-pythonclient/tools/"  
  
usr="sr320@washington.edu"
```

```
In [254]: !head {wd}query.fa
```

```
In [2]: !head {wd}query.fa
```

```
>PiuraChilensis_v1_contig_1
ATTACAATACGAAGTAAAATAGATAACGTAAAAATAATCTTGGTGCTGGATGATCGATC
AAGTTCACCAATATTTATTGTAAAAAATCATTCTAACAGCATGAAATCGTGTACAATG
TATAAACAAAGCAAATATATAACACTAAAGCAAGAGGGCGTAAGTGGGGGGTGGGTGAGA
GTAAAAAAATTCAAACATGTCAAATACCCCGGCCTAGCCTTAAAAGCACCATGGACTTCT
GCCTTCAATAAGCATAAAATTAAAACACCTAATACACAATGAATATACAGATAAAACAGA
TTTATGAATAGTTGGTGTACATCTTTACAGCCATAAGCCTTCATTGCTTCCAAACG
TATAAAATCTGACTTGGAACAAATATACAGCCATGAGATATGACACAGCGAGCACTACAAT
ATATATTATCTTGTACTATACAGCCTGTACAAGAAAATTCTGGAATTGTCTTCACAAGA
GACAGAAAAATAGTTGCAATGTGAATGCTAGTCTACTATTGATCACAATTGGATAGAAA
```

```
In [3]: #number of sequences
!fgrep -c ">" {wd}query.fa
```

282

## Blast

```
In [4]: !blastx \
-query {wd}query.fa \
-db {db} \
-max_target_seqs 1 \
-max_hsps 1 \
-outfmt 6 \
-evalue 1E-05 \
-num_threads 2 \
-out {wd}blast_sprot.tab
```

## Number of matched sequences:

```
In [5]: !wc -l {wd}blast_sprot.tab
```

```
211 /Users/sr320/Desktop/test/blast_sprot.tab
```

```
In [6]: !tr '||' "\t" <{wd}blast_sprot.tab> {wd}blast_sprot_sql.tab  
!head -1 {wd}blast_sprot.tab  
!echo SQLShare ready version has Pipes converted to Tabs ....  
!head -1 {wd}blast_sprot_sql.tab
```

PiuraChilensis_v1_contig_3	sp Q6P9A1 ZN530_HUMAN	33.33	105	61
3 825 1118 414	516 1e-07 57.4			
SQLShare ready version has Pipes converted to Tabs ....				
PiuraChilensis_v1_contig_3	sp Q6P9A1 ZN530_HUMAN	33.33	105	
61 3 825 1118	414 516 1e-07 57.4			

## Joining in SQL Share

```
In [7]: !python {sqls}singleupload.py \  
-d _blast_sprot \  
{wd}blast_sprot_sql.tab
```

```
processing chunk line 0 to 211 (0.000264167785645 s elapsed)  
pushing /Users/sr320/Desktop/test/blast_sprot_sql.tab...  
parsing 983DD315...  
finished _blast_sprot
```

```
In [8]: !python {sqls}fetchdata.py \  
-s "SELECT Column1, term, GOSlim_bin, aspect, ProteinName FROM [{usr}].[_blast_sprot]md left join [samwhite@washington.edu].[UniprotProtNamesReviewed_yes20130610]sp on md.Column3=sp.SPID left join [sr320@washington.edu].[SPID and GO Numbers]go on md.Column3=go.SPID left join [sr320@washington.edu].[GO_to_GOslim]slim on go.GOID=slim.GO_id where aspect like 'P'" \  
-f tsv \  
-o {wd}GODEscriptions.txt
```

## Plot GoSlim terms

```
In [10]: pylab inline
```

Populating the interactive namespace from numpy and matplotlib

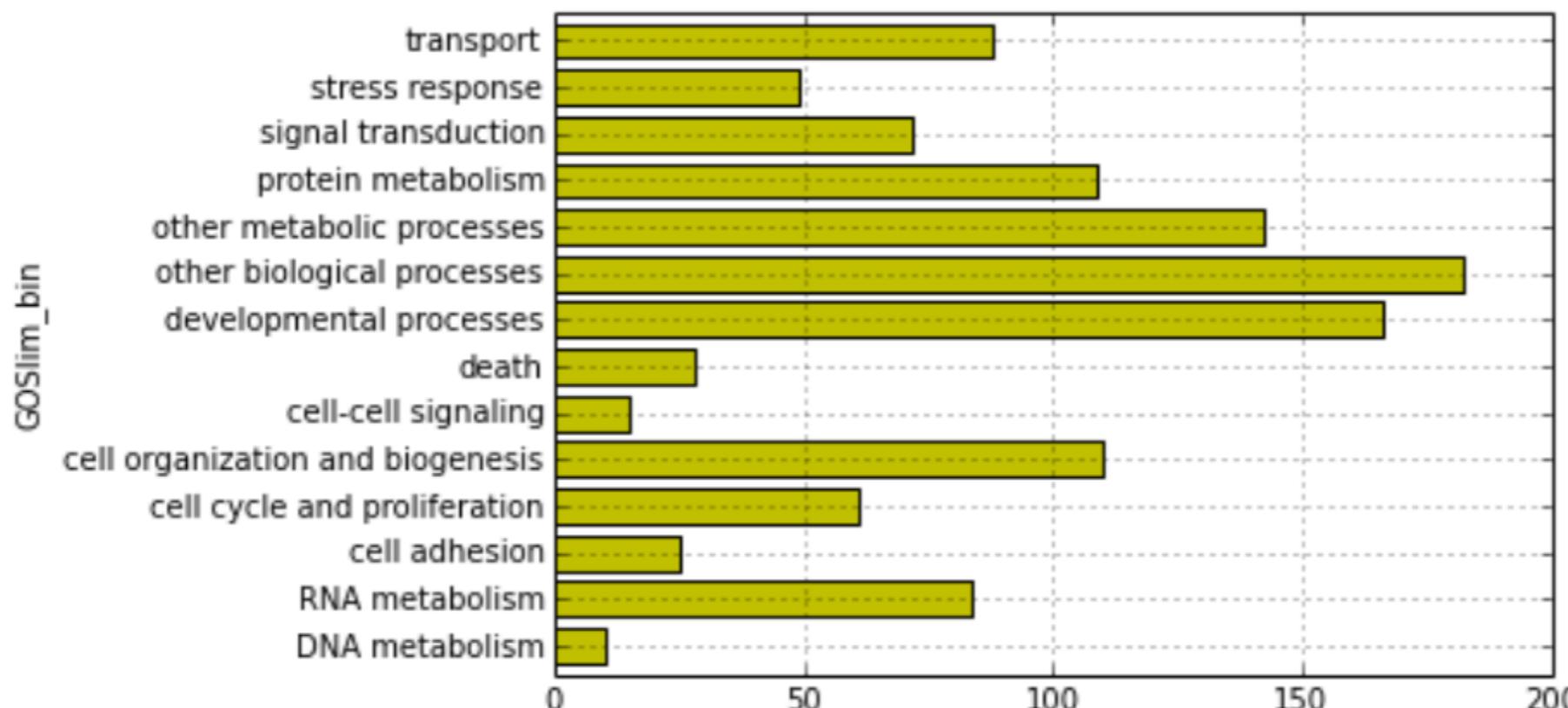
```
In [11]: cd {wd}
```

```
/Users/sr320/Desktop/test
```

```
In [12]: from pandas import *
```

```
gs = read_table('GOdescriptions.txt')
```

```
In [13]: gs.groupby('GOSlim_bin').Column1.count().plot(kind='barh', color=list('y'))
)
savefig('GOSlim.png', bbox_inches='tight')
```



# Open Science Philosophy

Transparency with limited effort

*will try just about anything*

Biology

Environment

Molecular

Data Analysis

eScience

iPlant Galaxy

Notebooks

Rationale

Platforms

**Open Science**

Data

everything else...

Biology

Environment

Molecular

Data Analysis

eScience

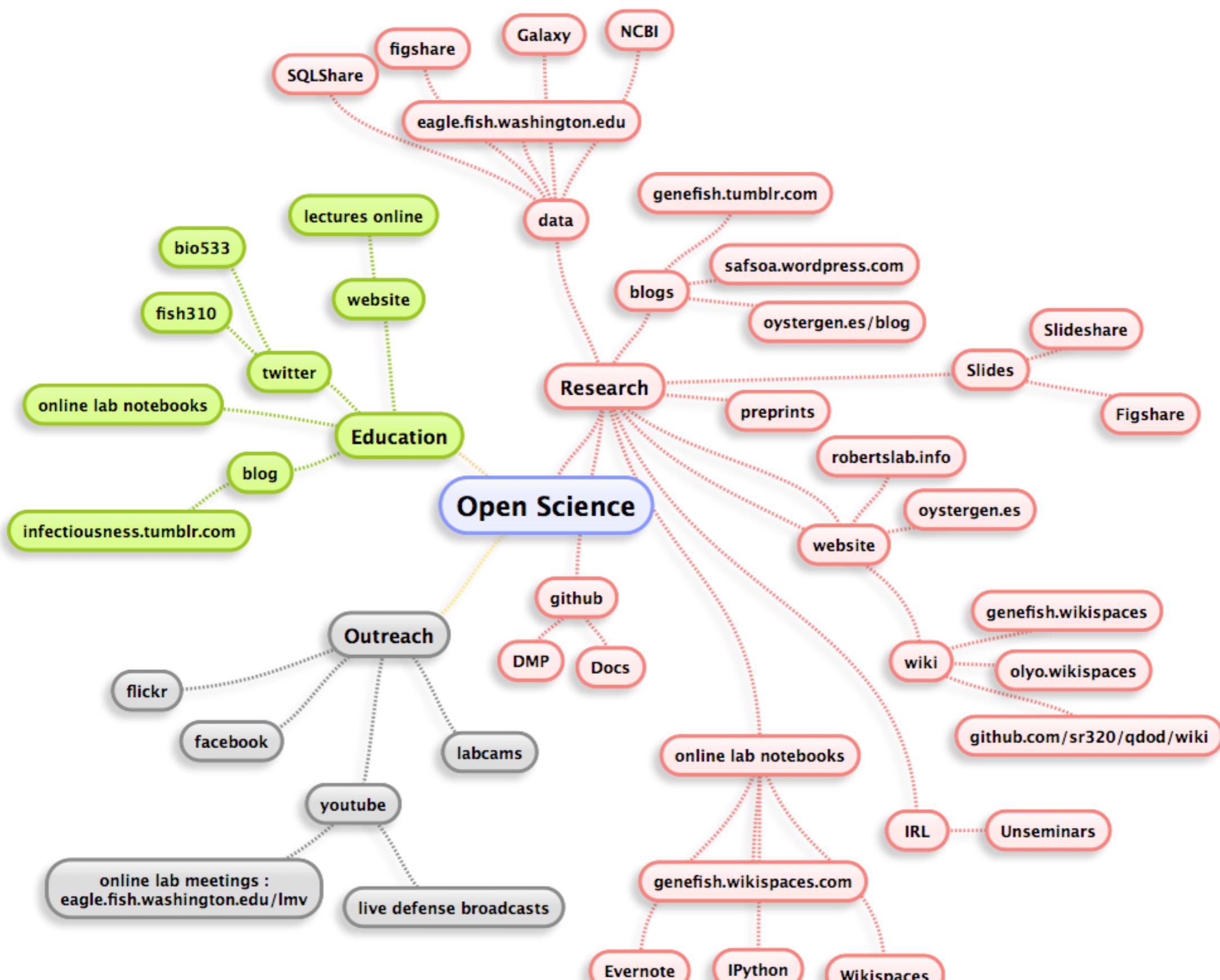
iPlant Galaxy

Notebooks

Rationale

Platforms

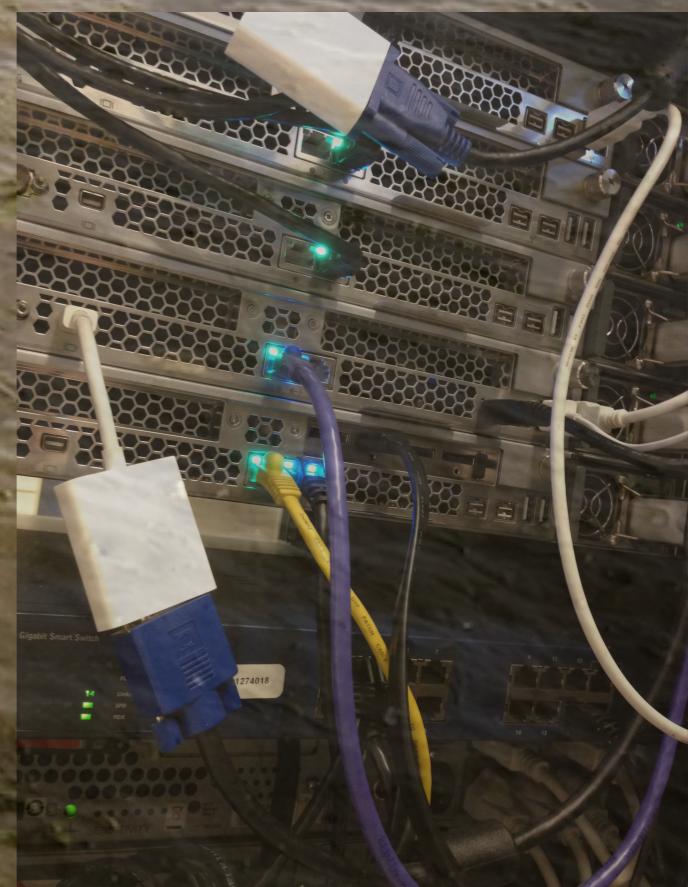
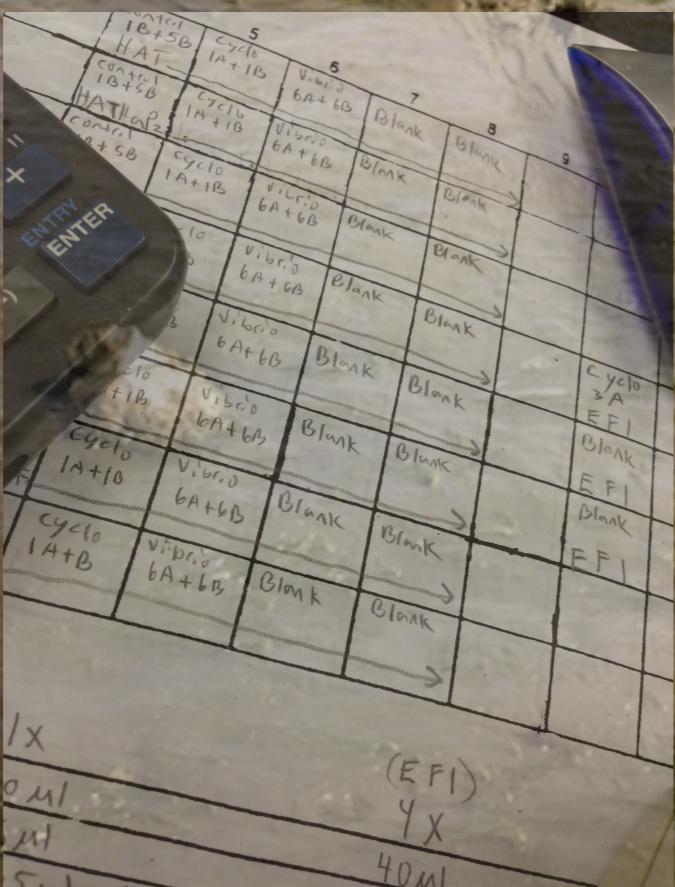
**Open Science**



everything else...

start them young

# Changing Science



# Acknowledgements

Mackenzie Gavery

Claire Olson

Sam White

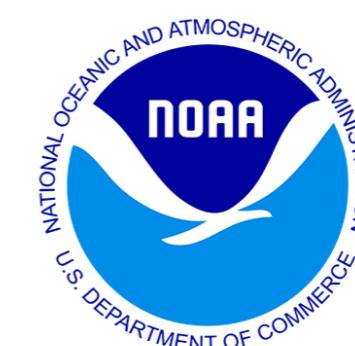
Brent Vadopalas

Jake Heare

Jay Dimond

Bill Howe

Dan Halperin



slides, data & more @  
[github.com/sr320/talk-CICESE-2015](https://github.com/sr320/talk-CICESE-2015)