

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](https://twitter.com/sr320)

FIU Seminar - April 23, 2015



epigenetics

latest online genetics responses relationship
physiological integrative marine

new Research level view

maintaining Sharing component developed

adaptive characterizing Highlights

examine check Lab strive page Data use

process notebooks

functional

organismal

investigating members

dependent

Newest

follow proteome

resources

Facebook electronic blogs

nucleotide organisms environmental

sure focuses

meetings

practice

Resource

Plan one

list

impacts core change

methylation potential

bivalves Using

Acidification approaches

context transcription Launched

Check Ocean

Meeting Check

Plan one

list

impacts core change

methylation potential

bivalves Using

Acidification approaches

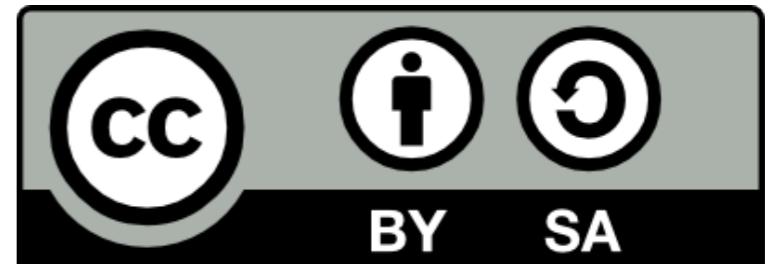
context transcription Launched

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/

onsnetwork.org

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

These slides plus links @
github.com/sr320/talk-fiu-2015

Background

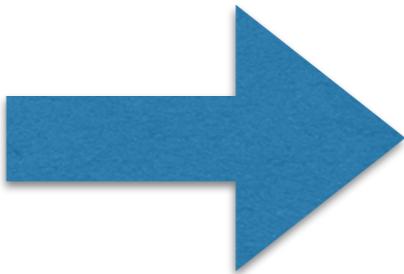
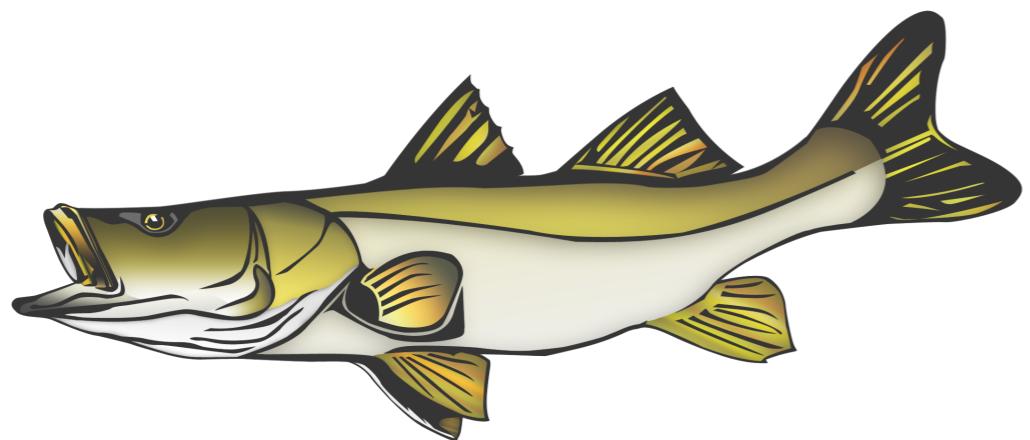
Physiology

How fundamental processes work in aquatic species

Background

Physiology

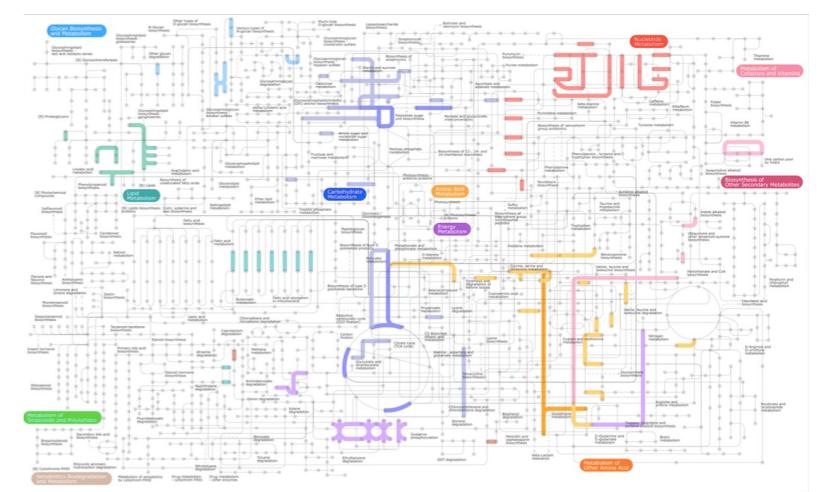
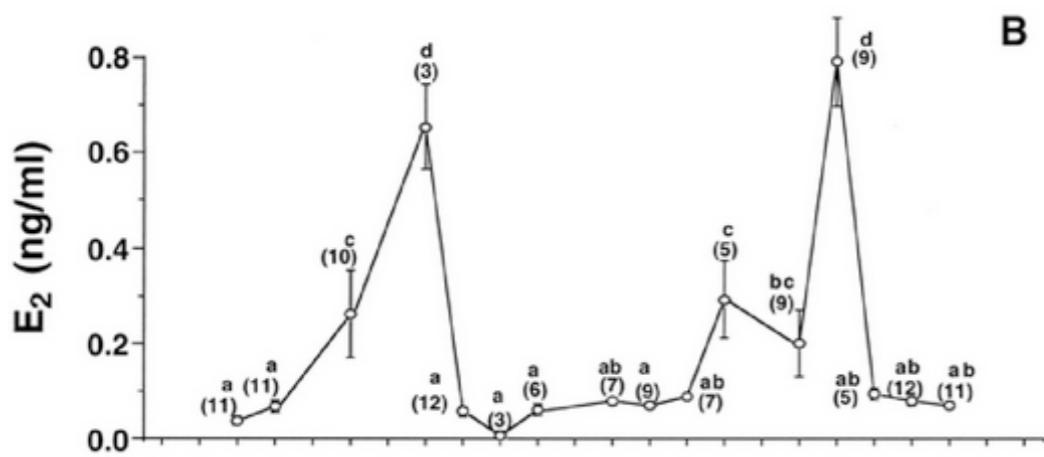
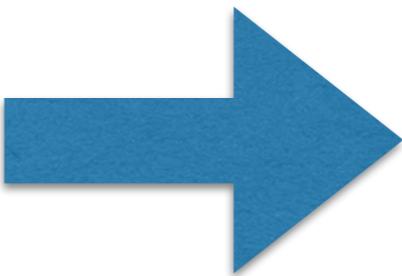
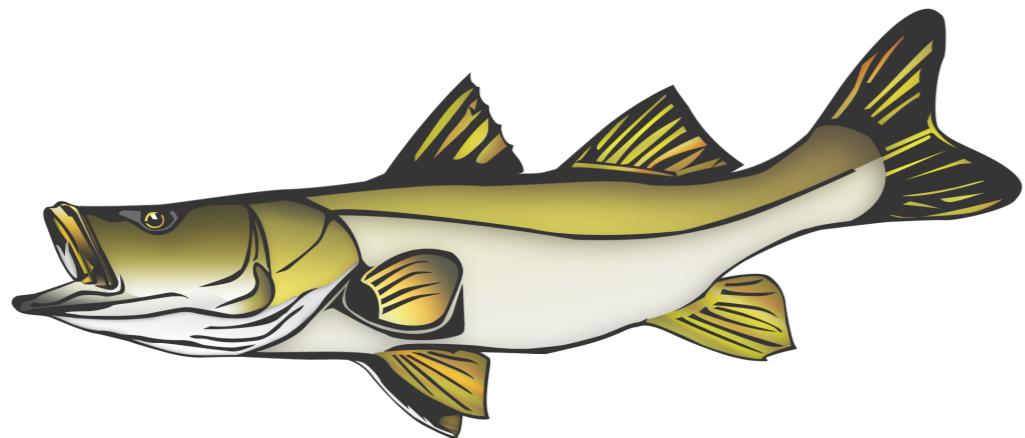
How fundamental processes work in aquatic species



Background

Physiology

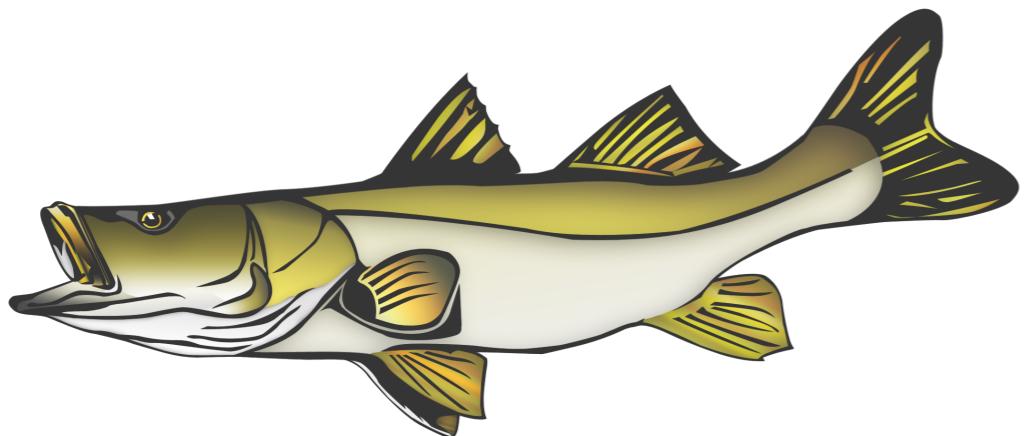
How fundamental processes work in aquatic species



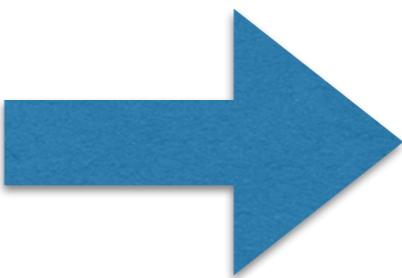
Background

Physiology

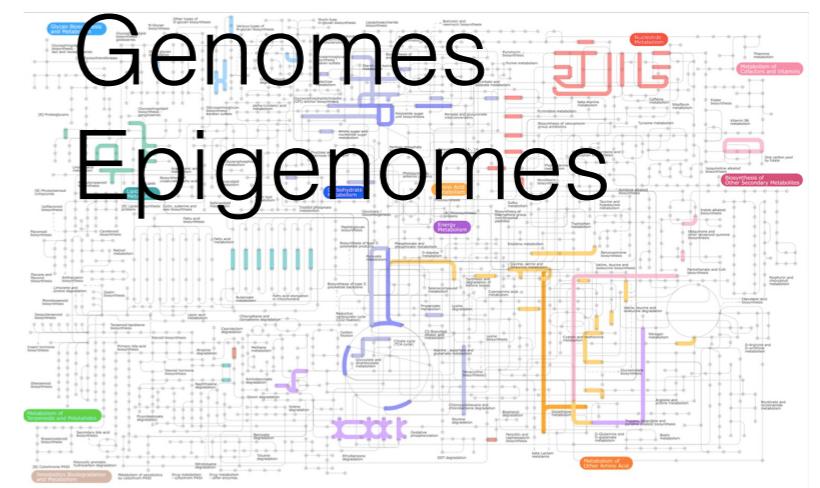
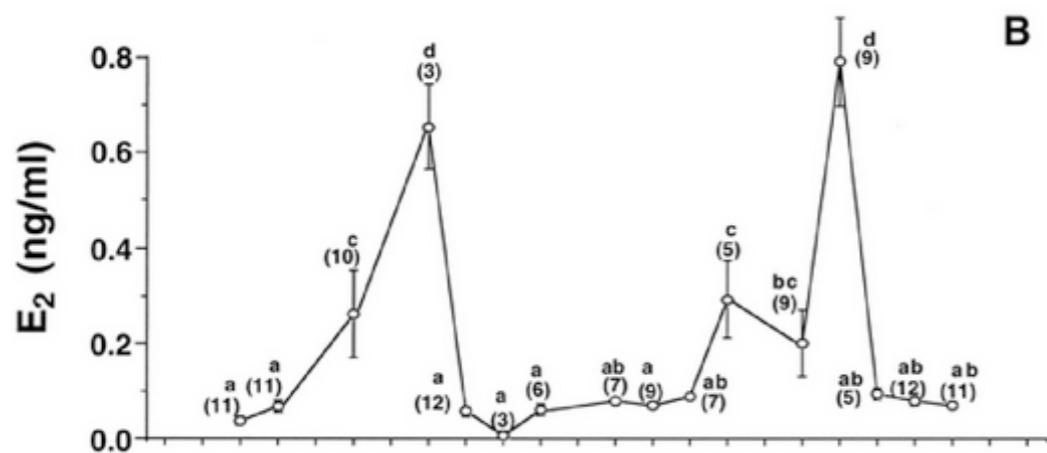
How fundamental processes work in aquatic species



Hormones
Proteins



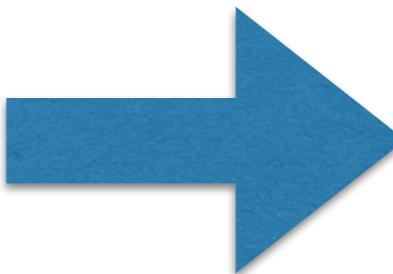
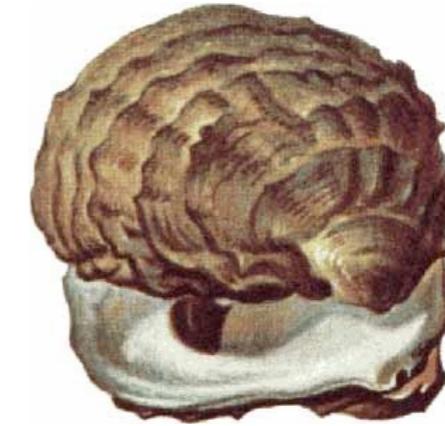
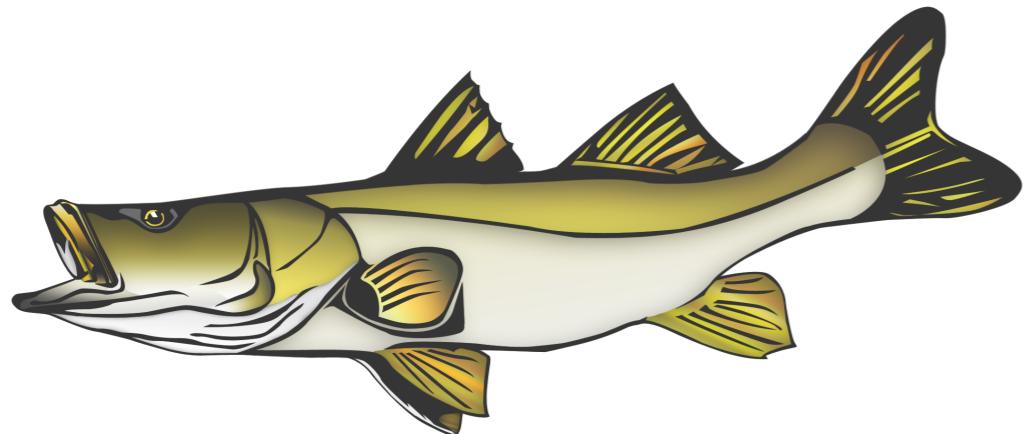
Transcriptomes
Proteomes
Genomes
Epigenomes



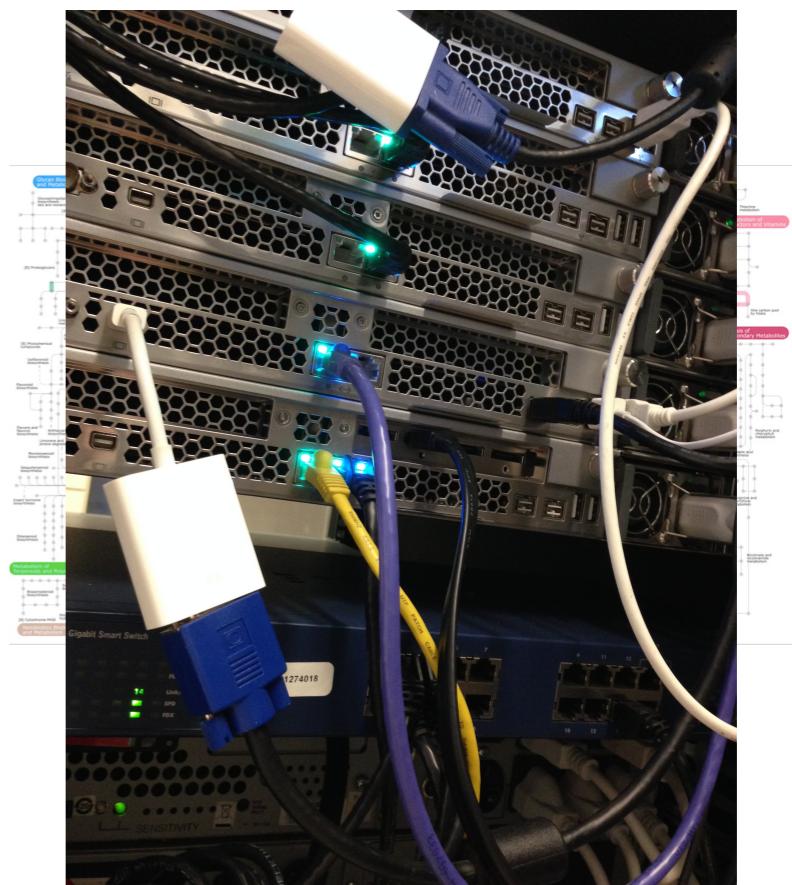
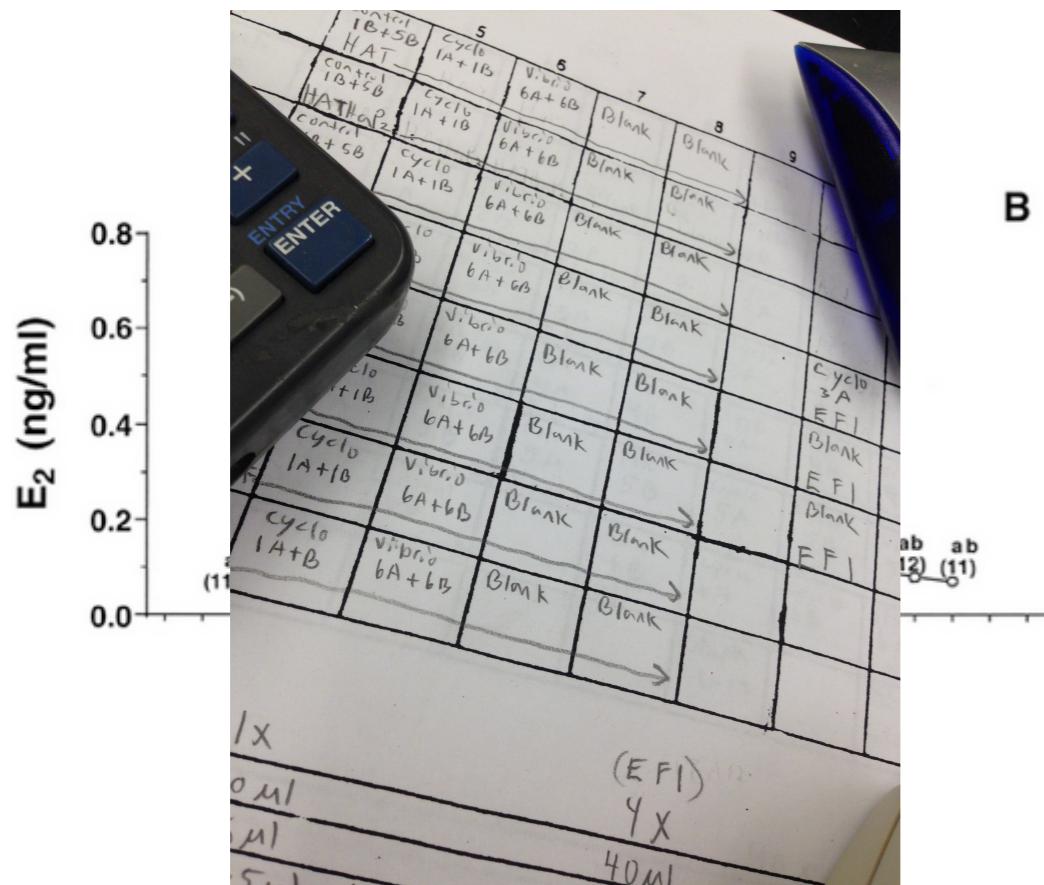
Background

Physiology

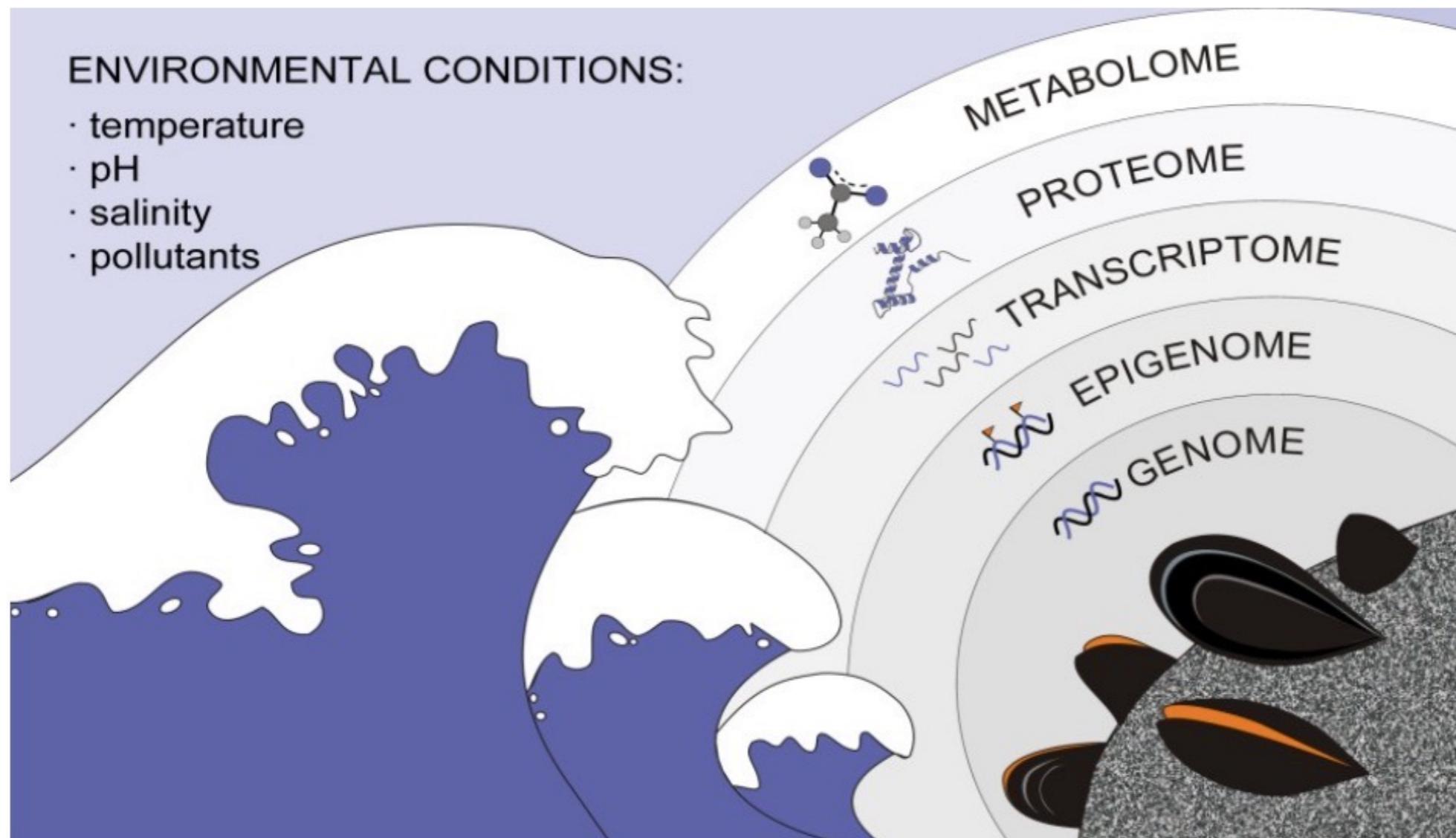
How fundamental processes work in aquatic species



B



Background



Review

Bivalve Omics: State of the Art and Potential Applications for the Biomonitoring of Harmful Marine Compounds

Victoria Suárez-Ulloa ¹, Juan Fernández-Tajes ², Chiara Manfrin ³, Marco Gerdol ³,
Paola Venier ⁴ and José M. Eirín-López ^{1,*}

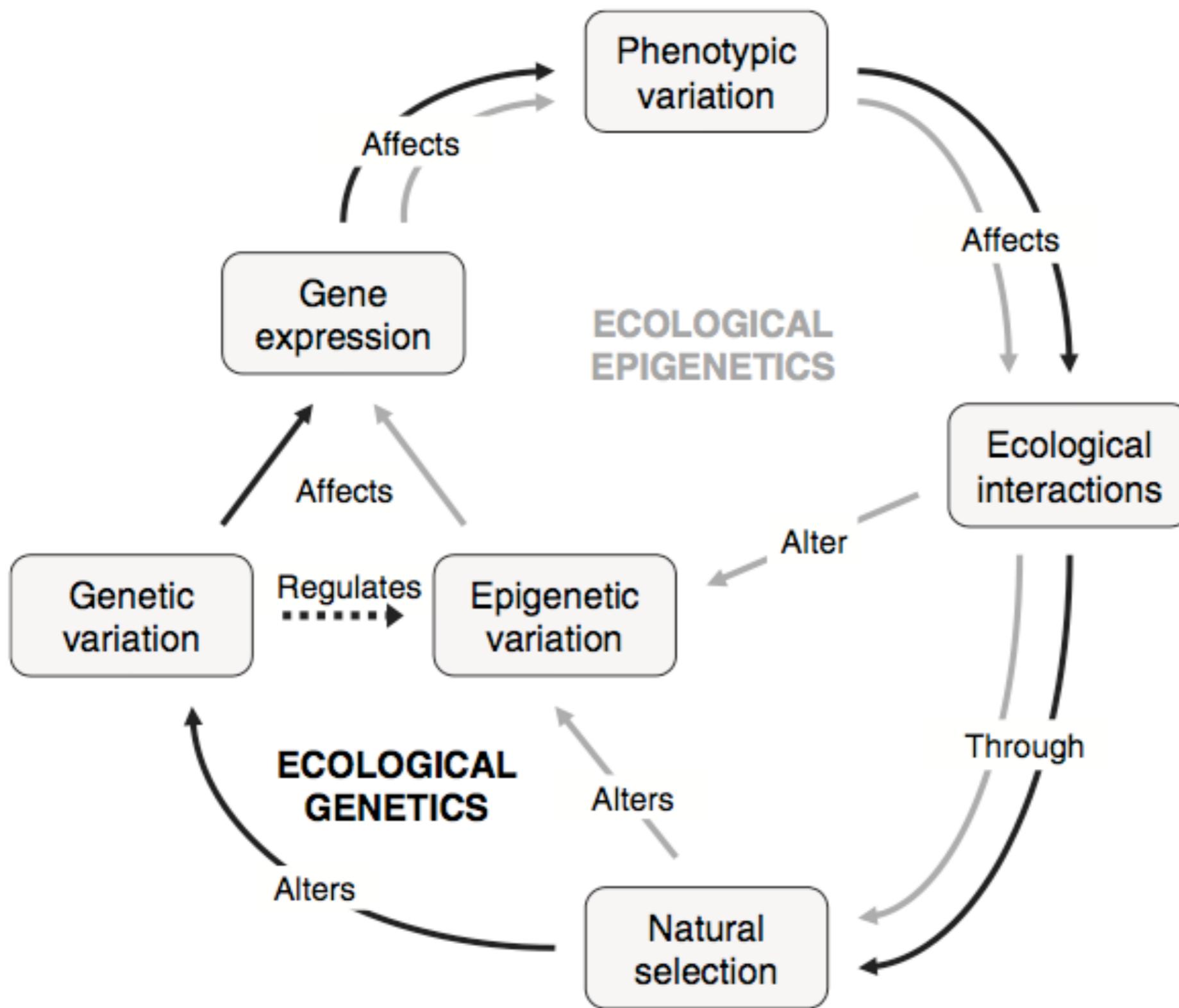
¹ Chromatin Structure and Evolution (CHROMEVOL) Group, Department of Biological Sciences,
Florida International University, North Miami, FL 33181, USA; E-Mail: msuarezu@fiu.edu

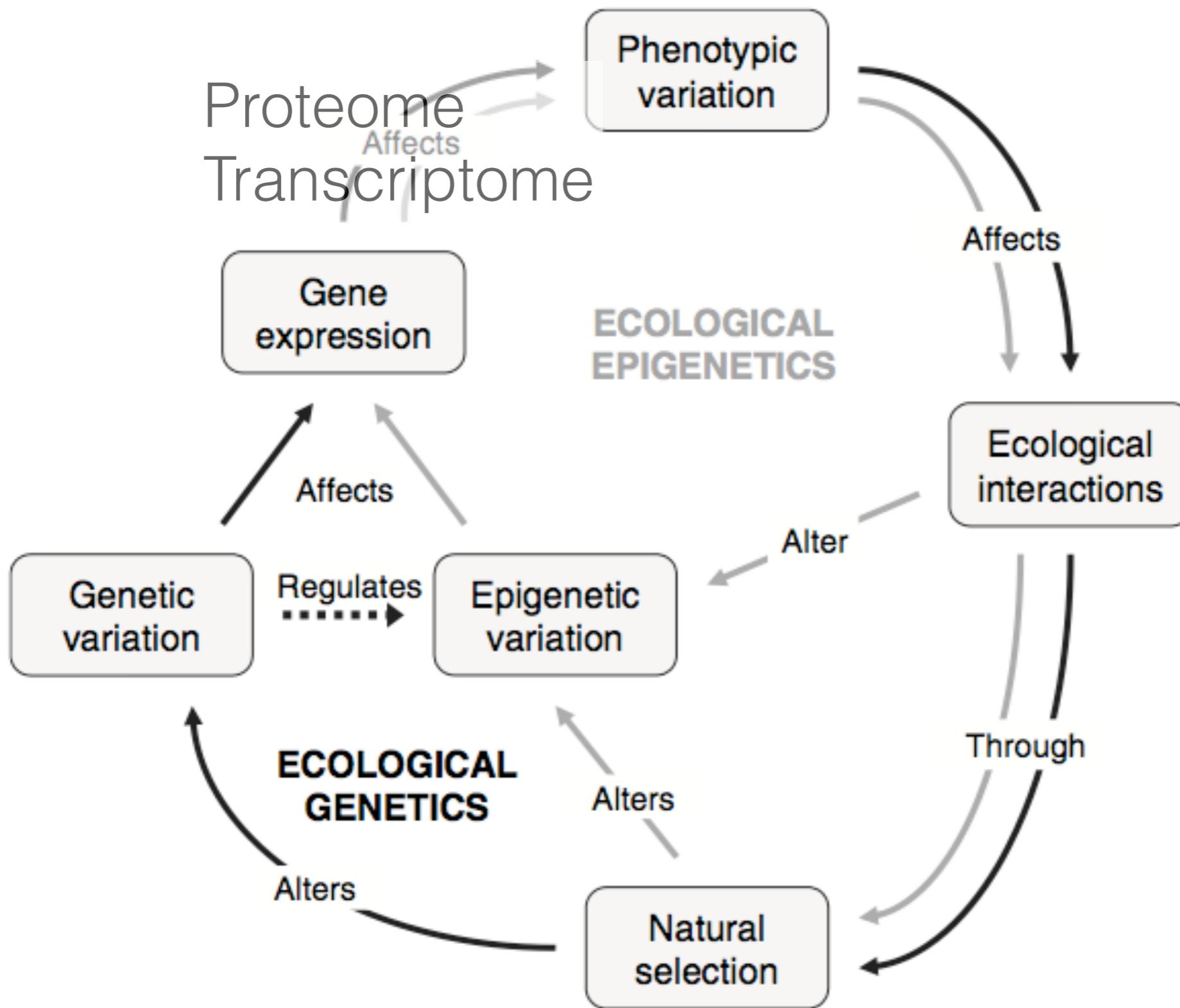
OPEN ACCESS

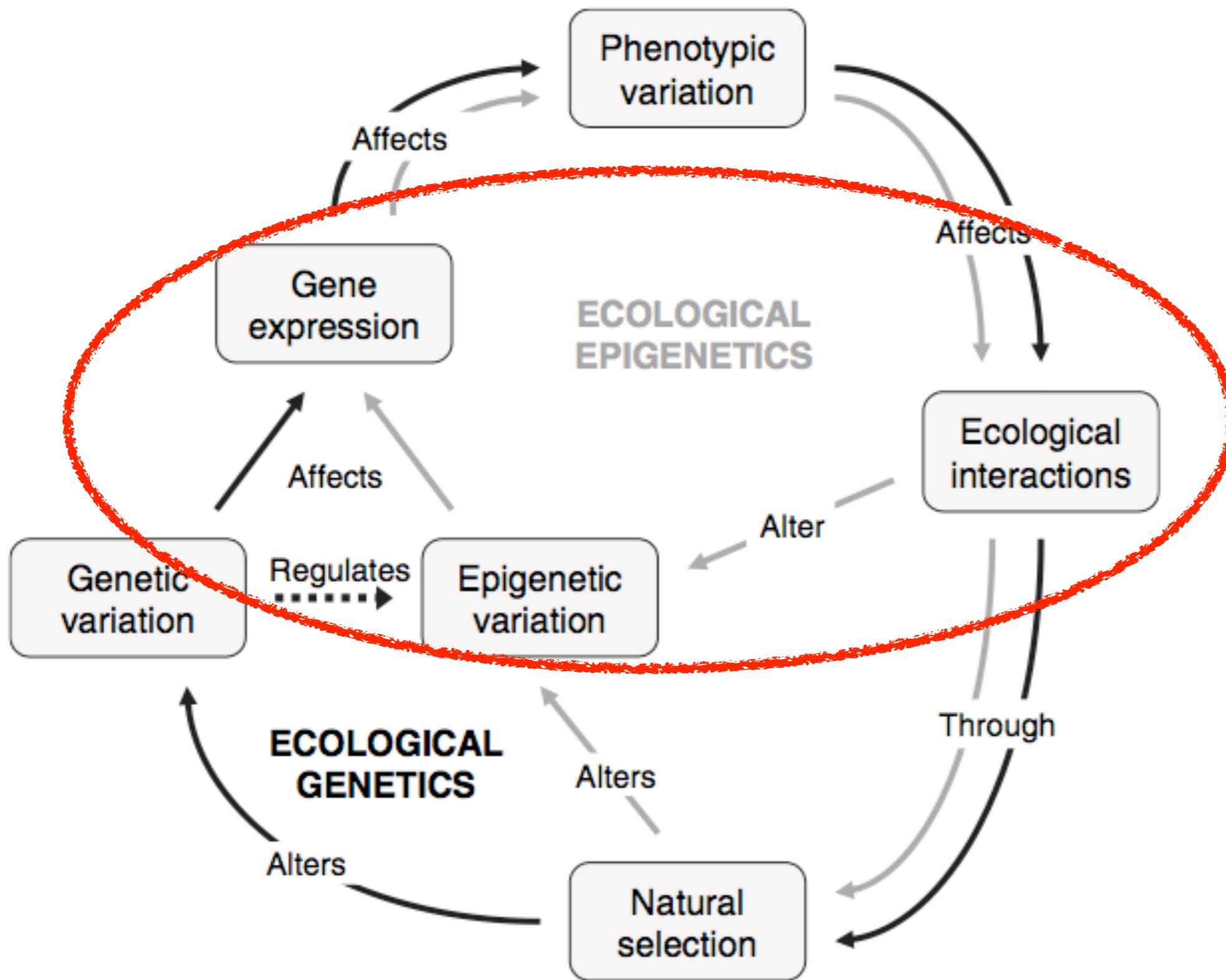
marine drugs

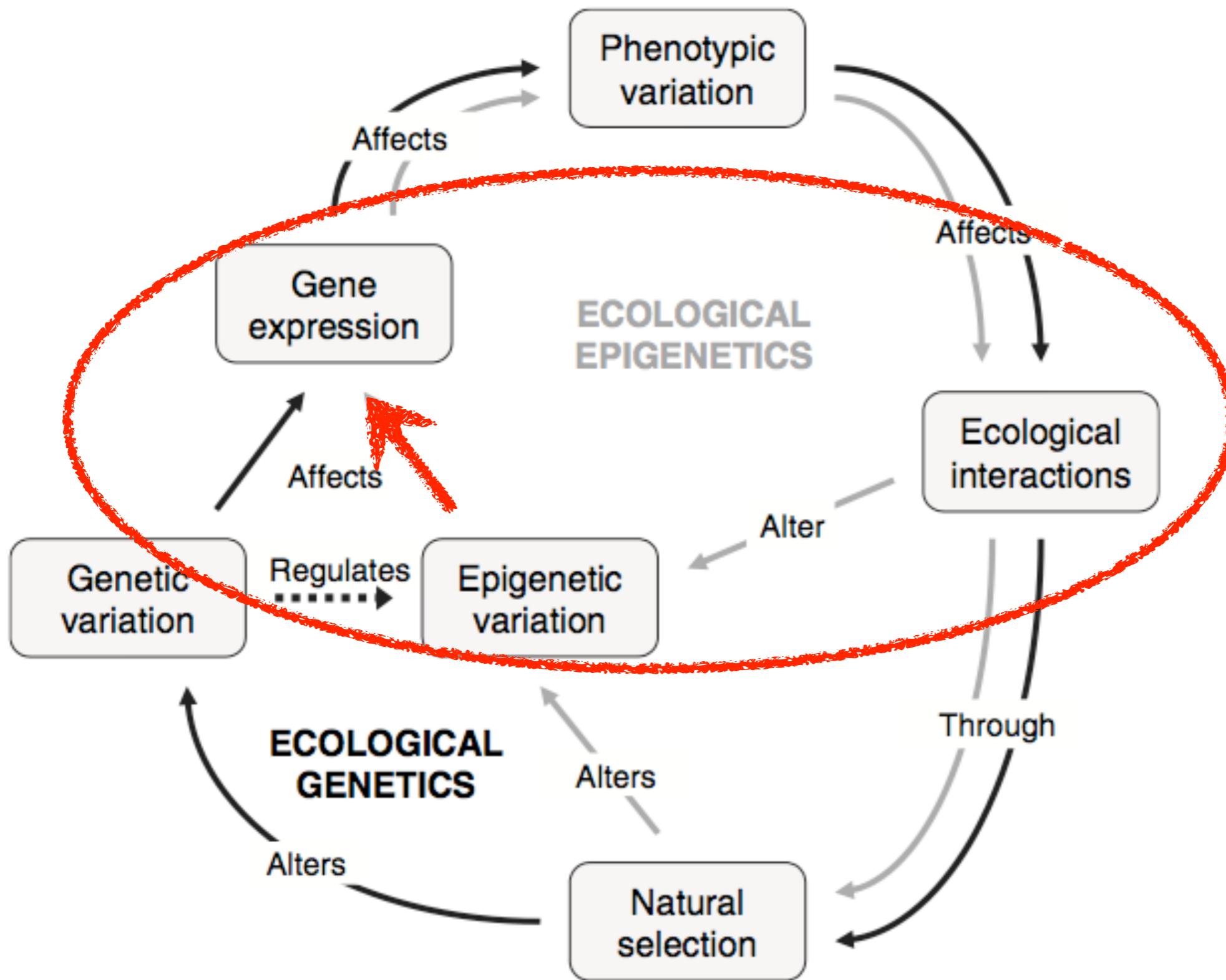
ISSN 1660-3397

www.mdpi.com/journal/marinedrugs

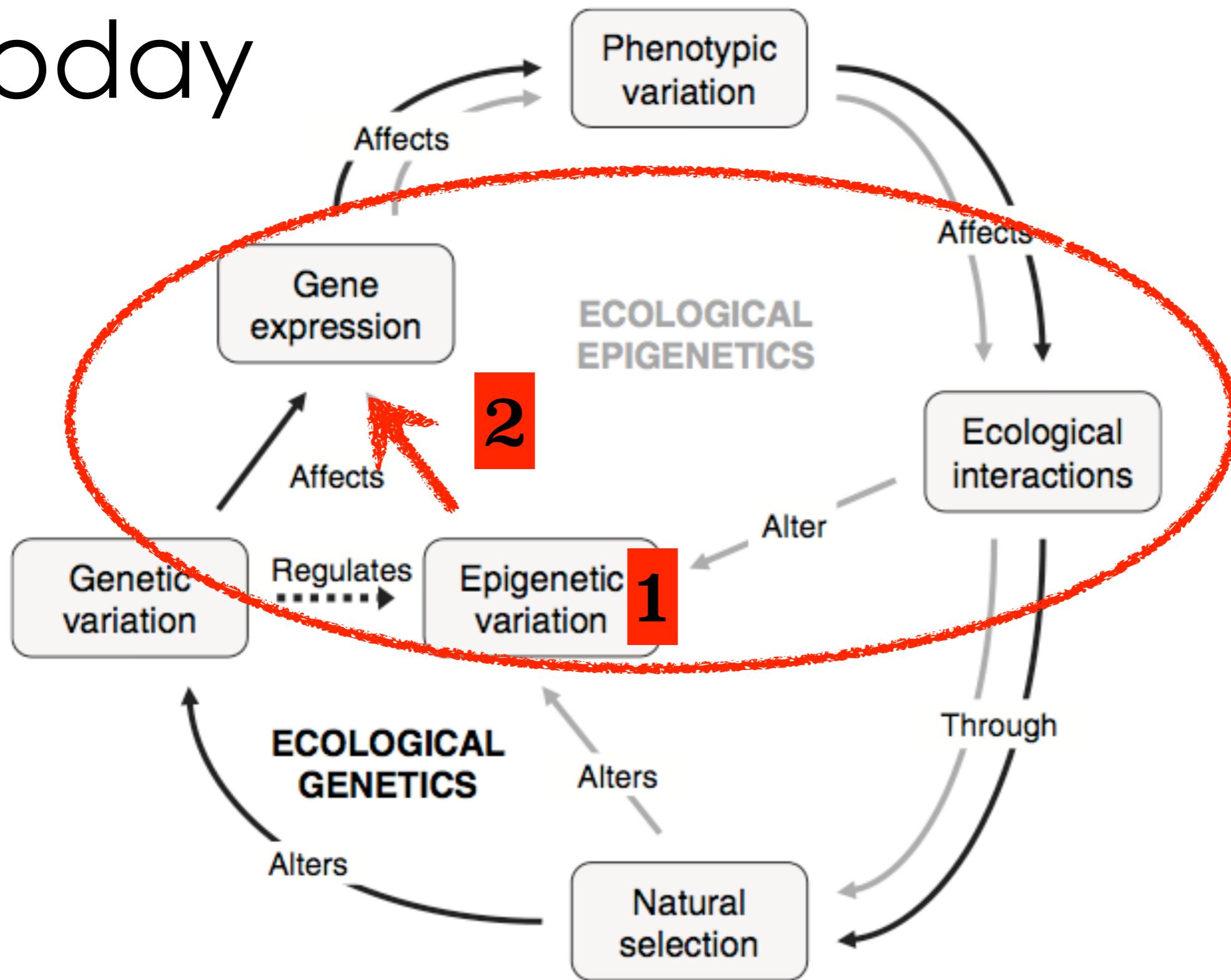








Today



Ecology Letters, (2008) 11: 106–115

doi: 10.1111/j.1461-0248.2007.01130.x

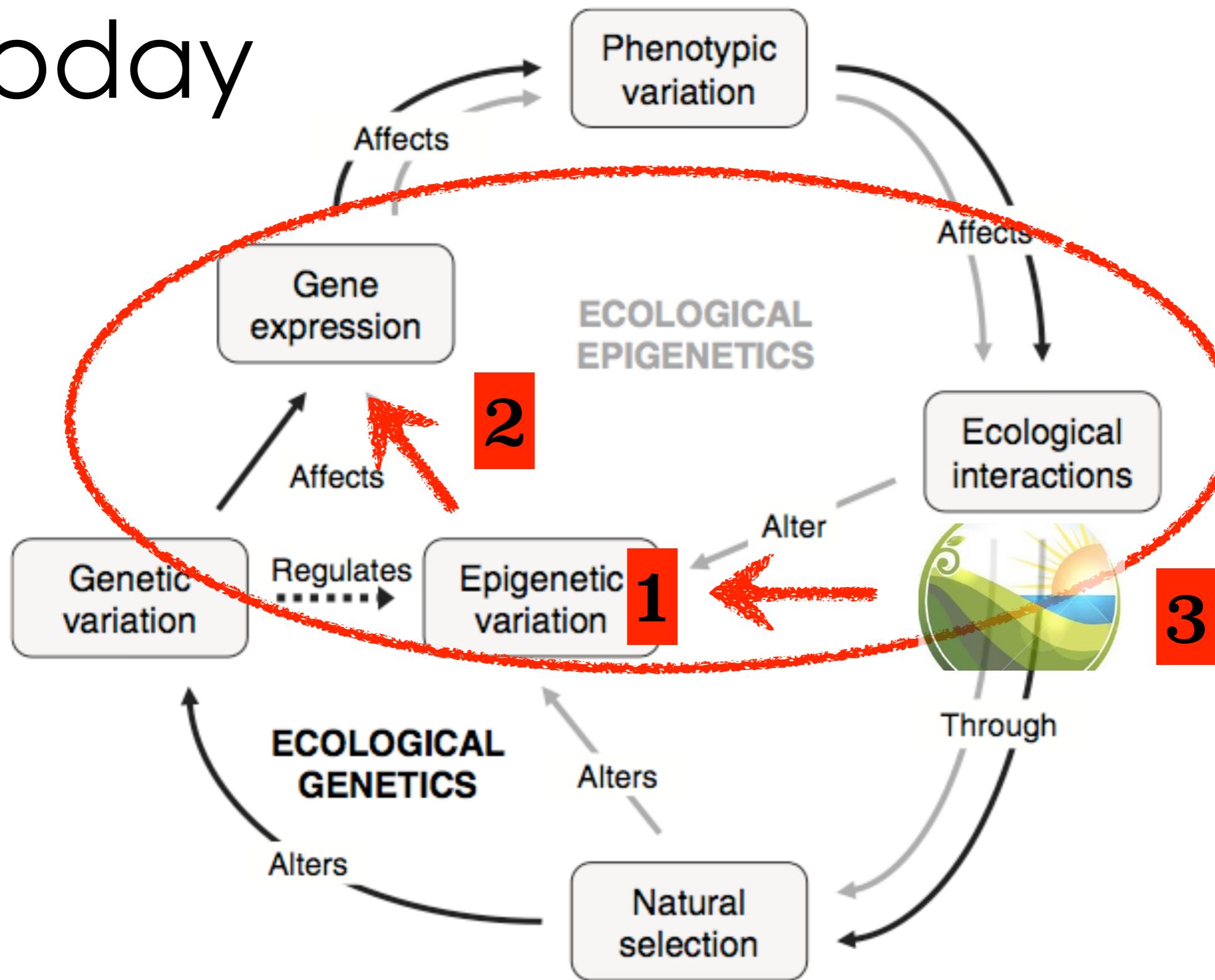
modified from

IDEA AND
PERSPECTIVE

Epigenetics for ecologists

Oliver Bossdorf,^{1,*} Christina L.
Richards² and Massimo Pigliucci³

Today



Ecology Letters, (2008) 11: 106–115

doi: 10.1111/j.1461-0248.2007.01130.x

modified from

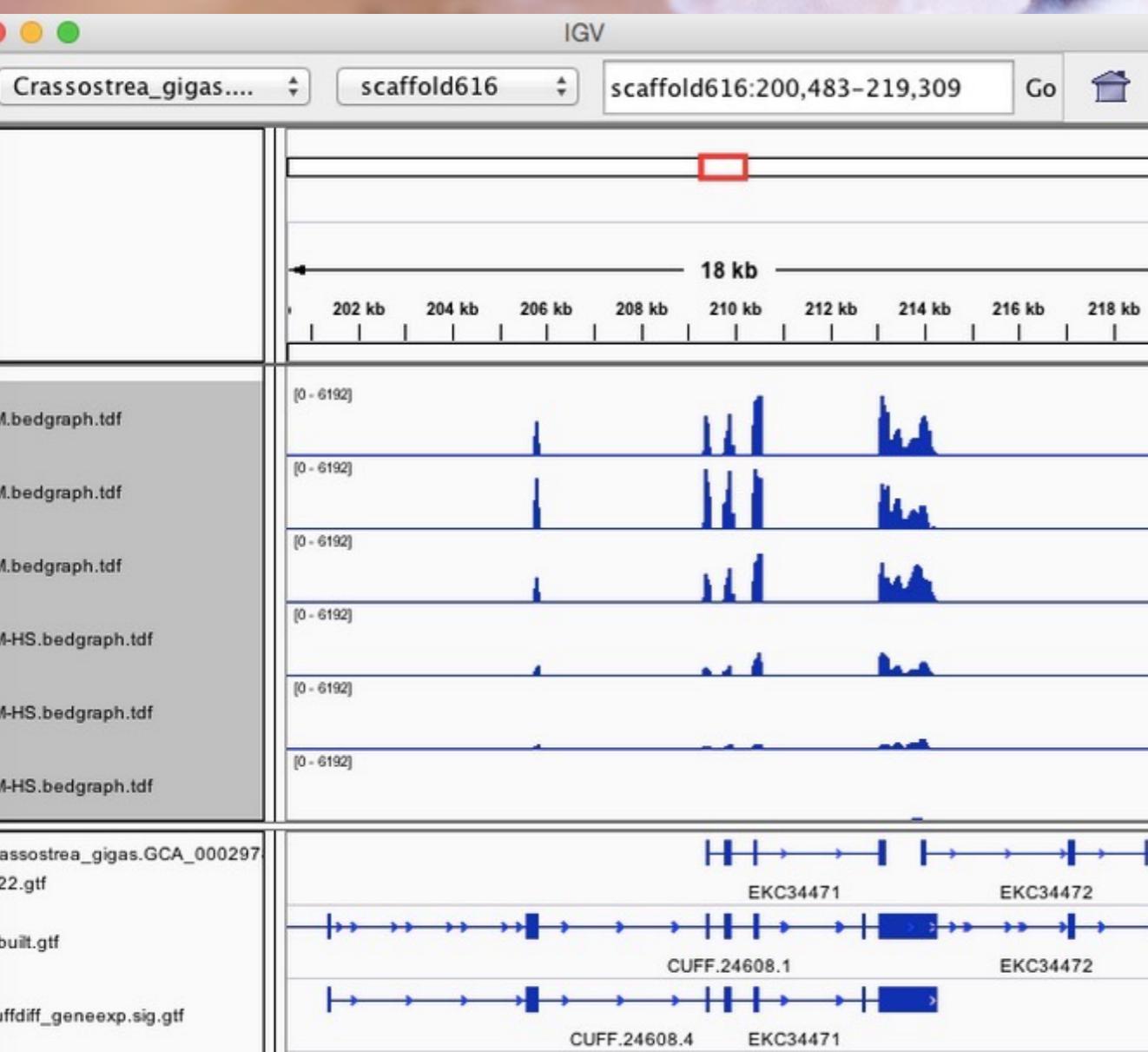
IDEA AND
PERSPECTIVE

Epigenetics for ecologists

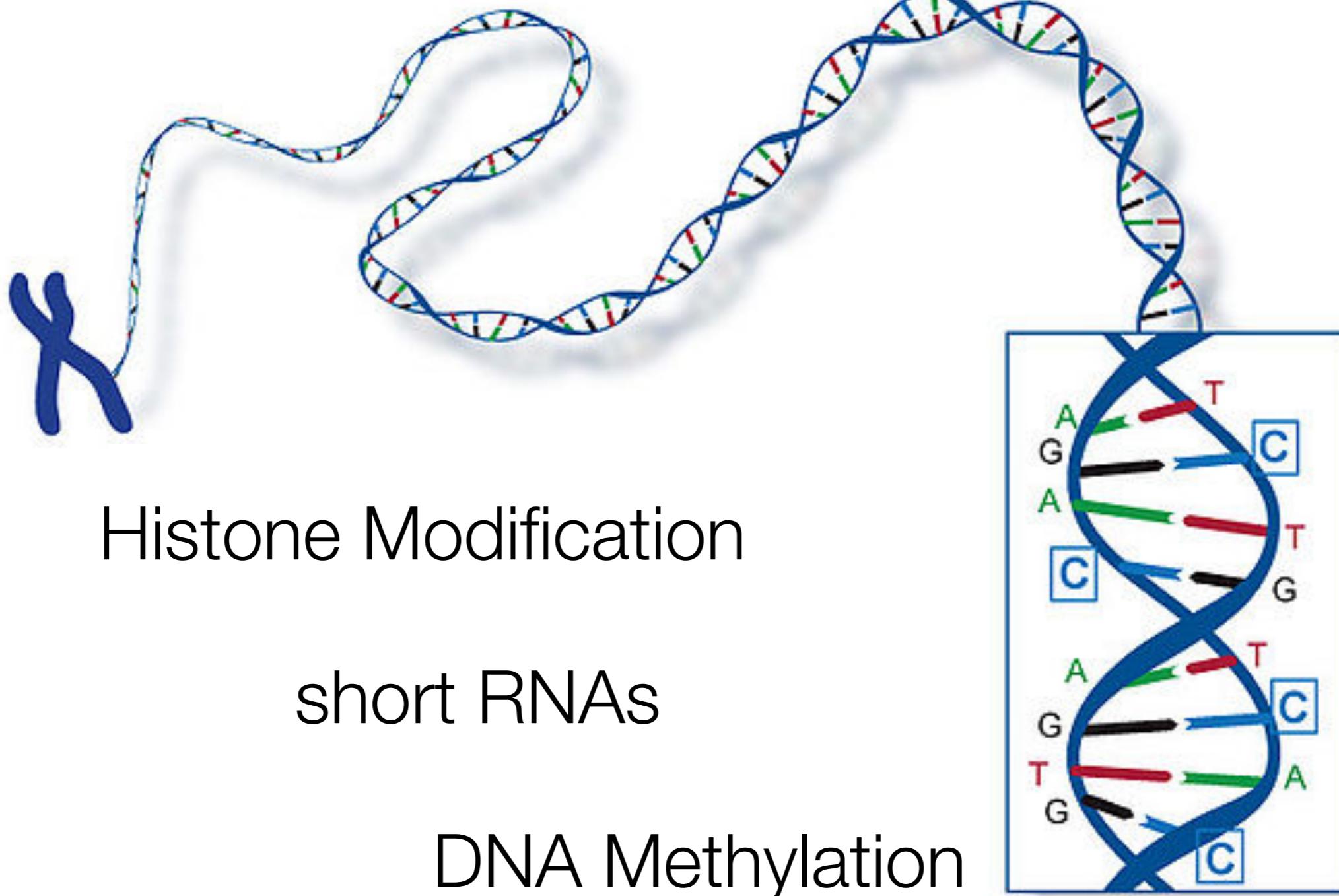
Oliver Bossdorf,^{1,*} Christina L.
Richards² and Massimo Pigliucci³



Genome Resources



Epigenetics



Epigenetics



Photo credit: Flickr, Creative Commons, he-boden



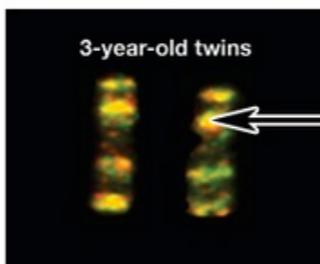
Epigenetics



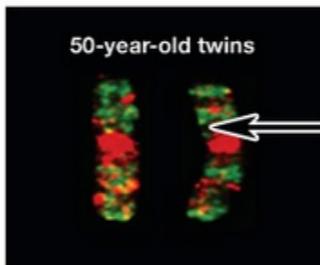
Nature AND Nurture

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.

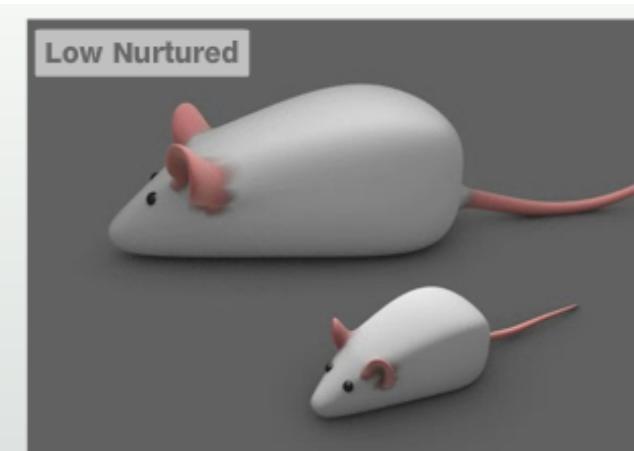
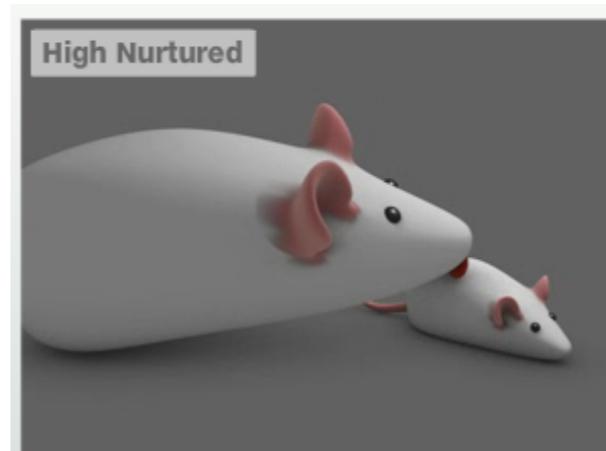
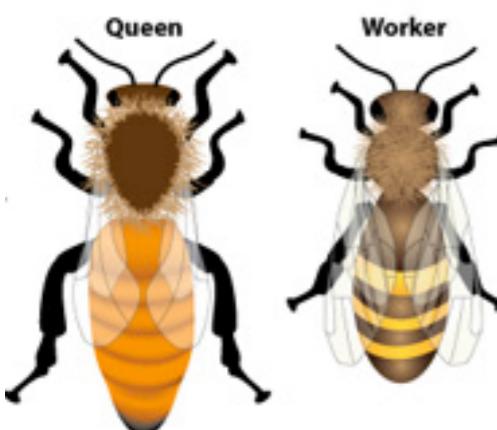
While pregnant, both of their mothers were fed Bisphenol A (BPA) but DIFFERENT DIETS:

The mother of this mouse received a **normal mouse diet**

The mother of this mouse received a diet **supplemented** with choline, folic acid, betaine and vitamin B12

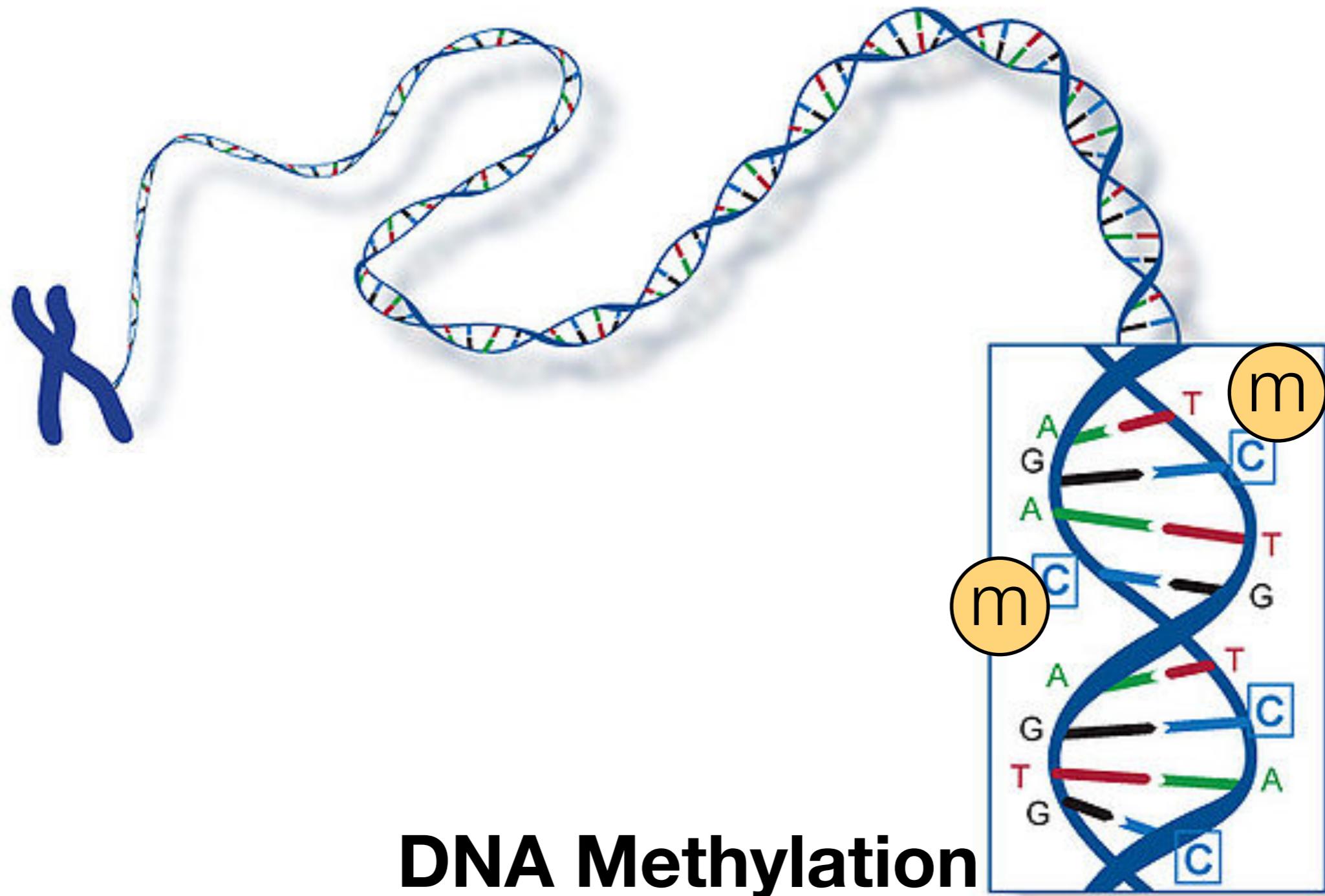


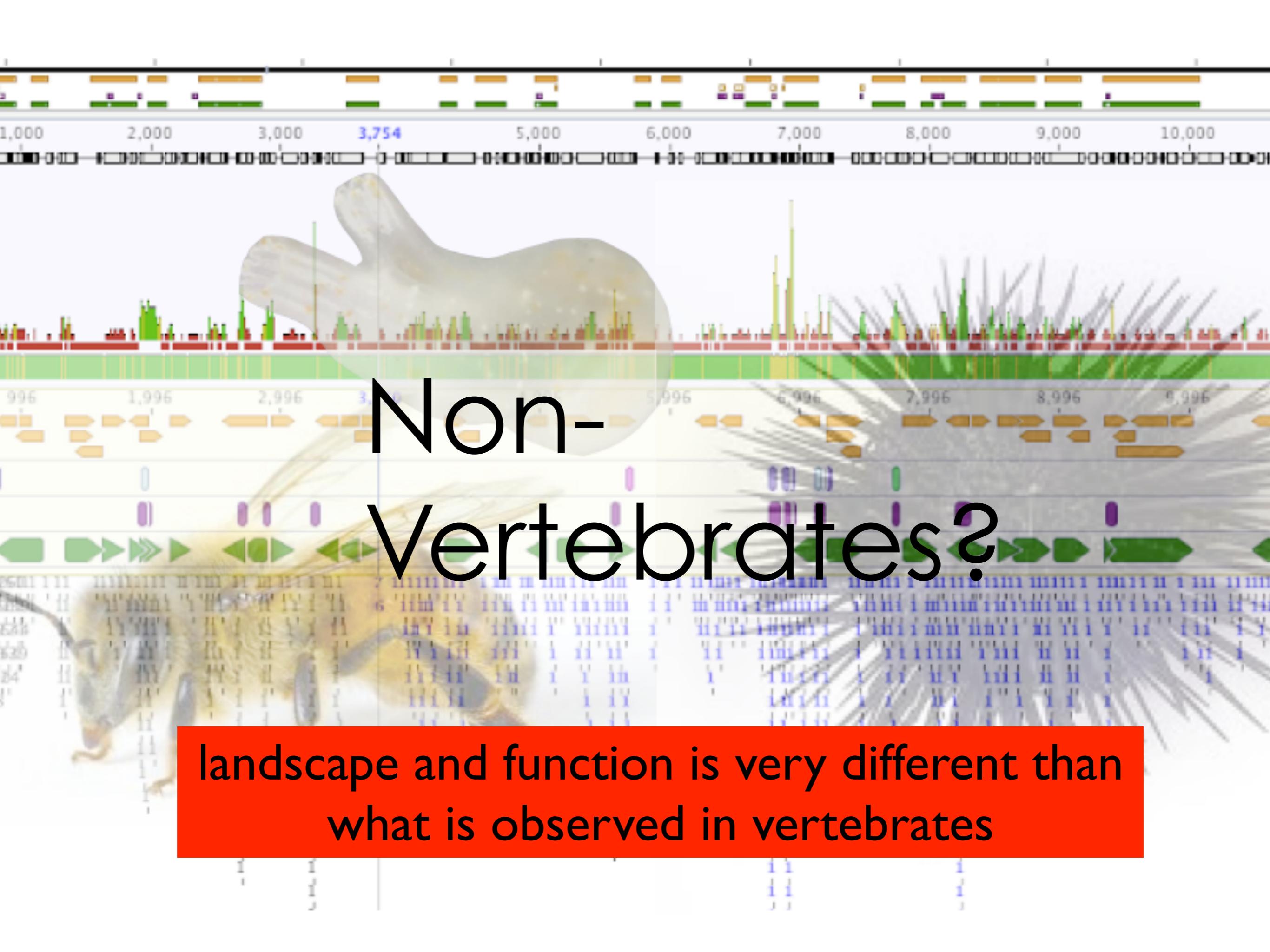
Queen Bee Larvae: Queens are raised in specially constructed cells called "queen cups," which are filled with royal jelly.



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

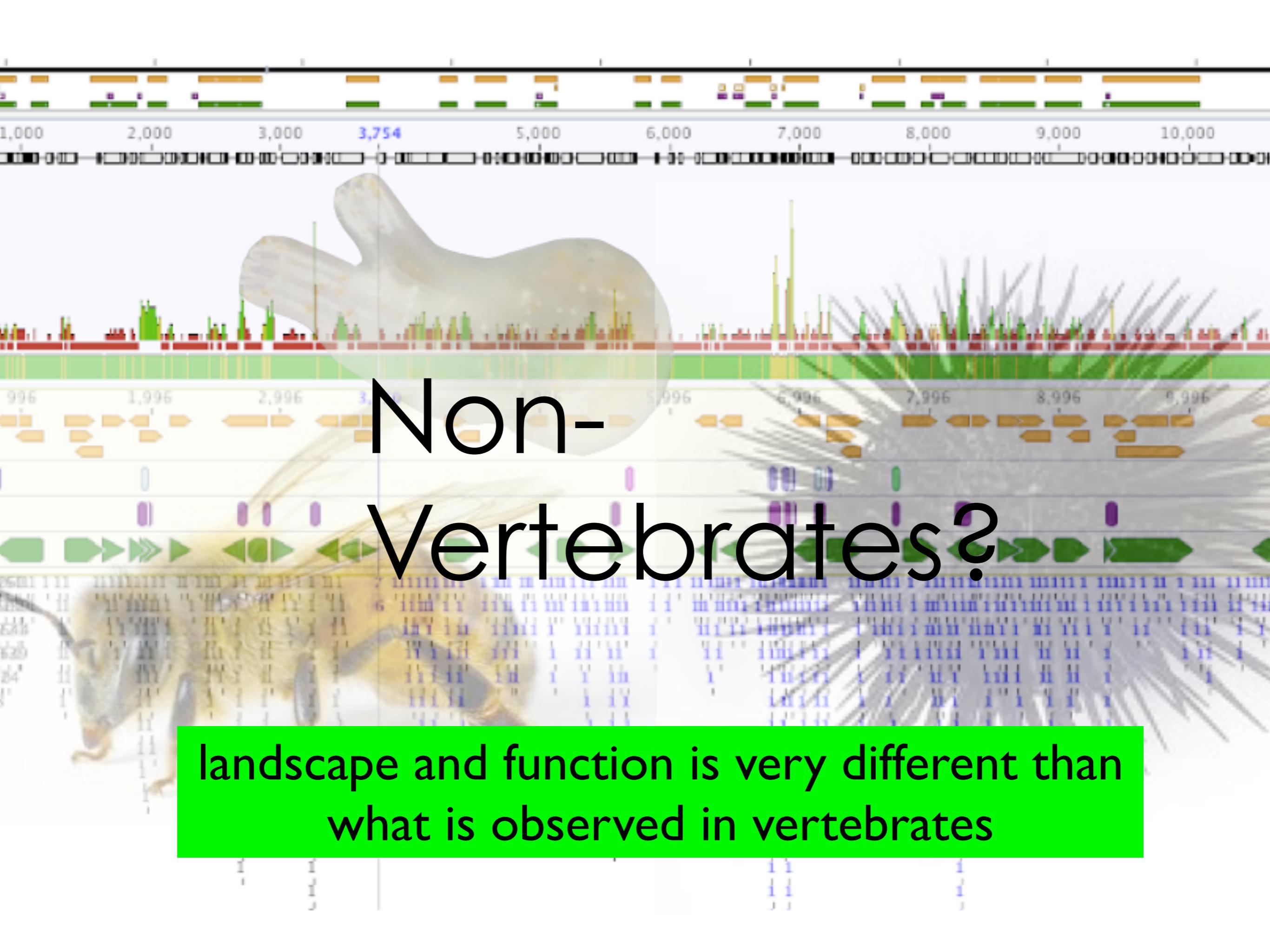
AUDIO





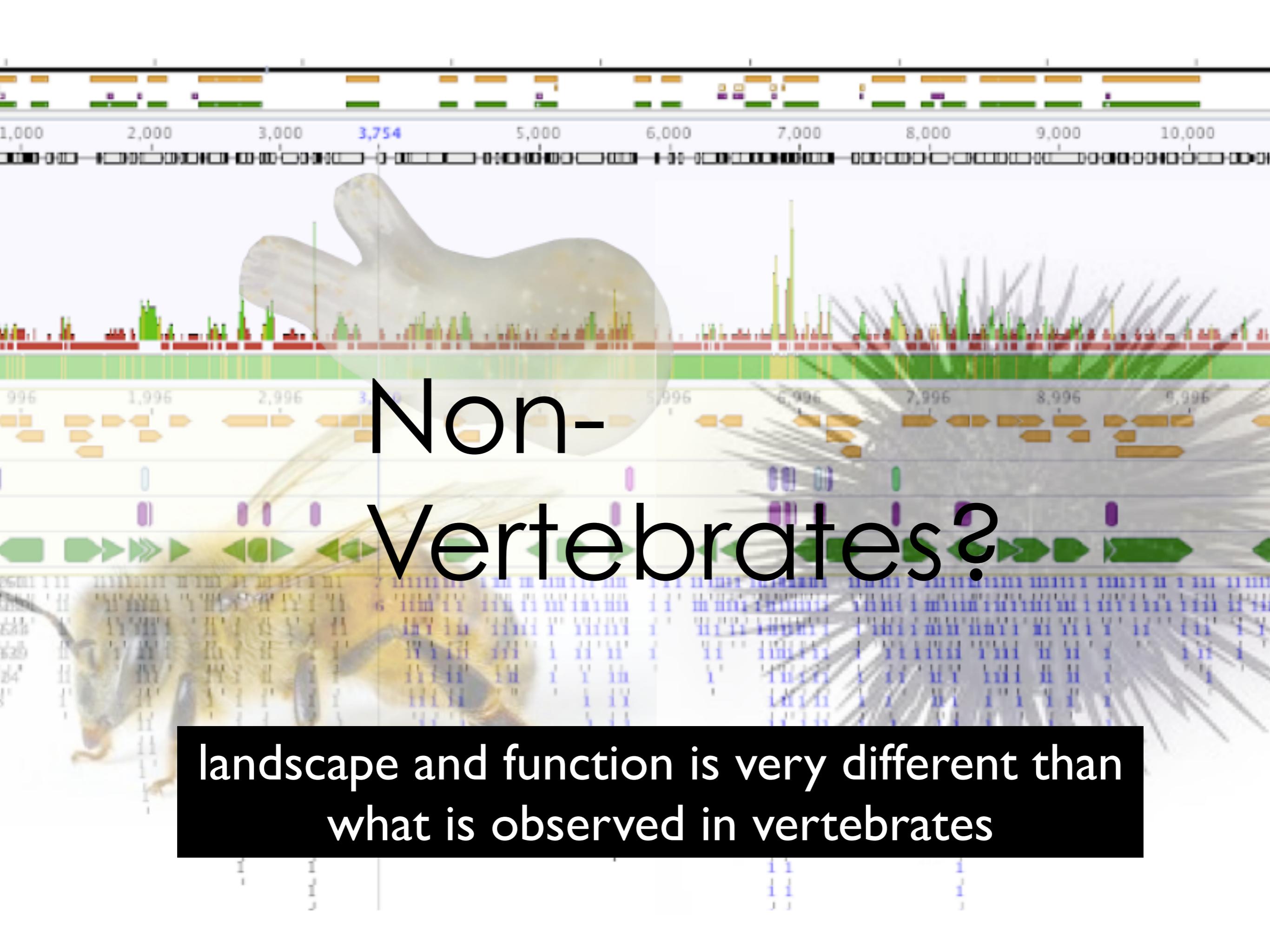
Non- Vertebrates?

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



Non- Vertebrates?

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



Non- Vertebrates?

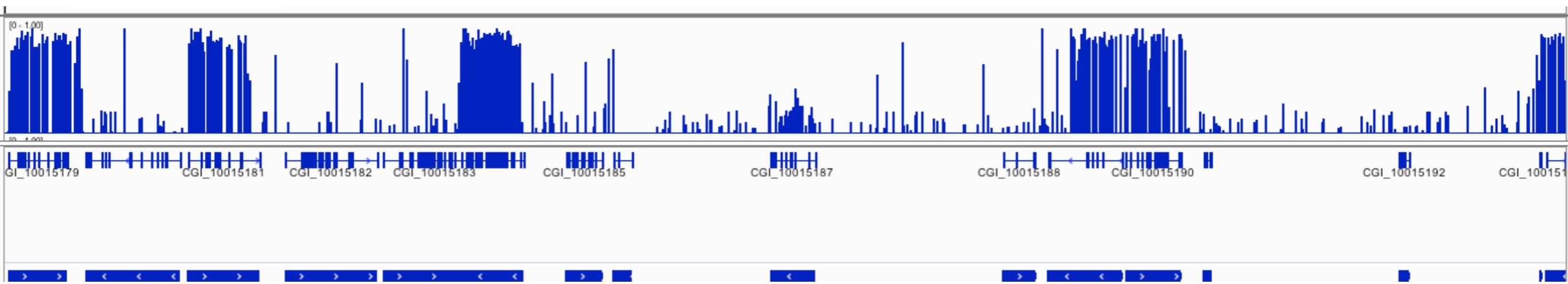
landscape and function is very different than
what is observed in vertebrates

*Absent in
several
model
organisms*



Oysters?

Epigenetic variation **1**

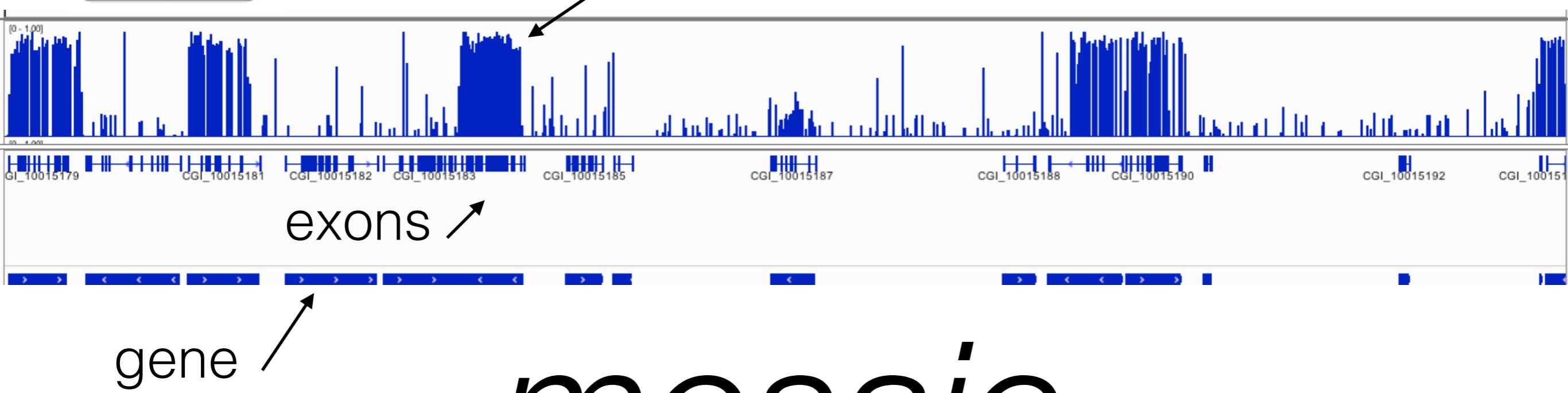


mosaic

associated with gene bodies

Epigenetic variation **1**

DNA methylation level (0-100%) @ cytosines



mosaic

associated with gene bodies

Epigenetic variation **1**

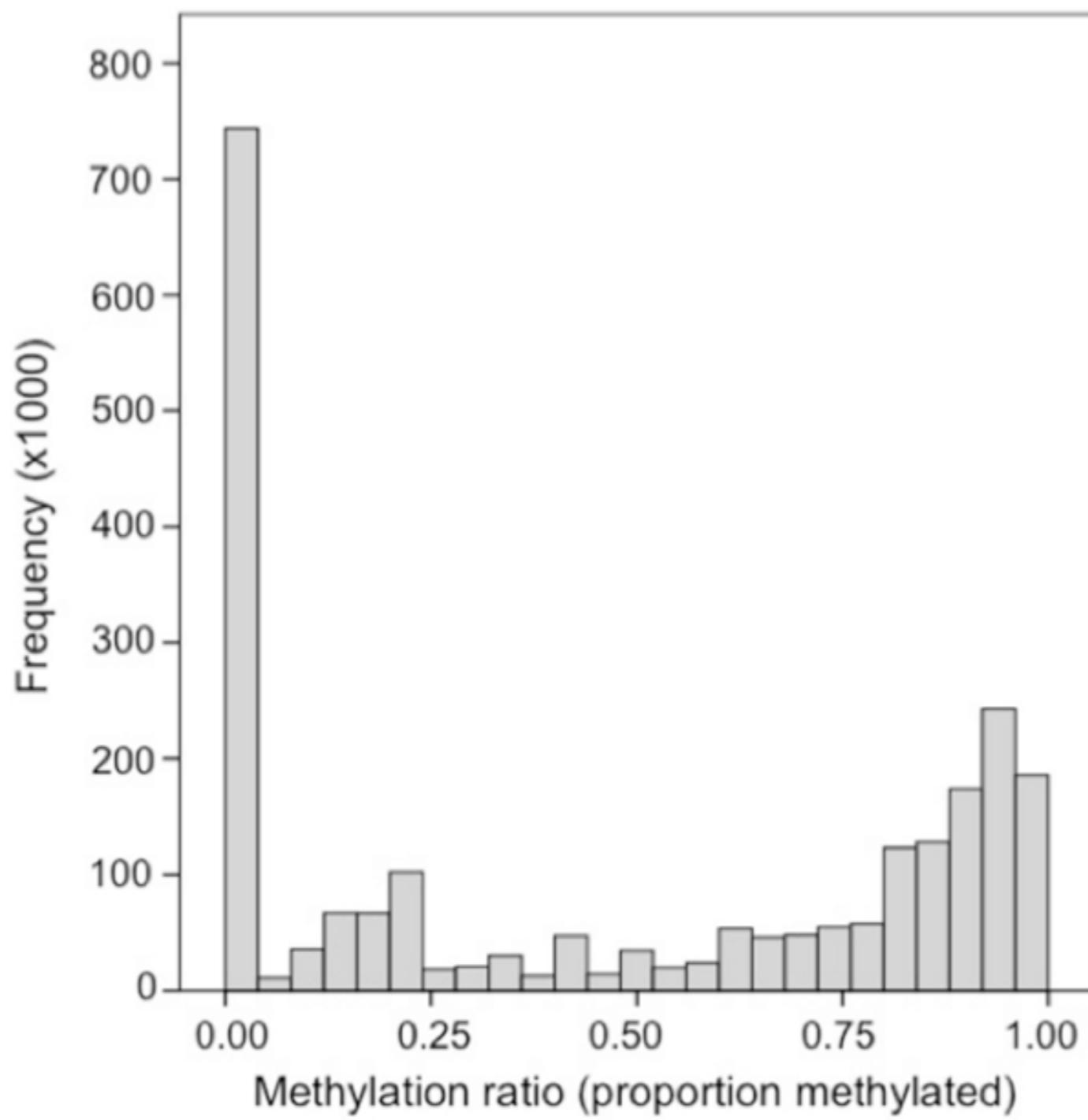
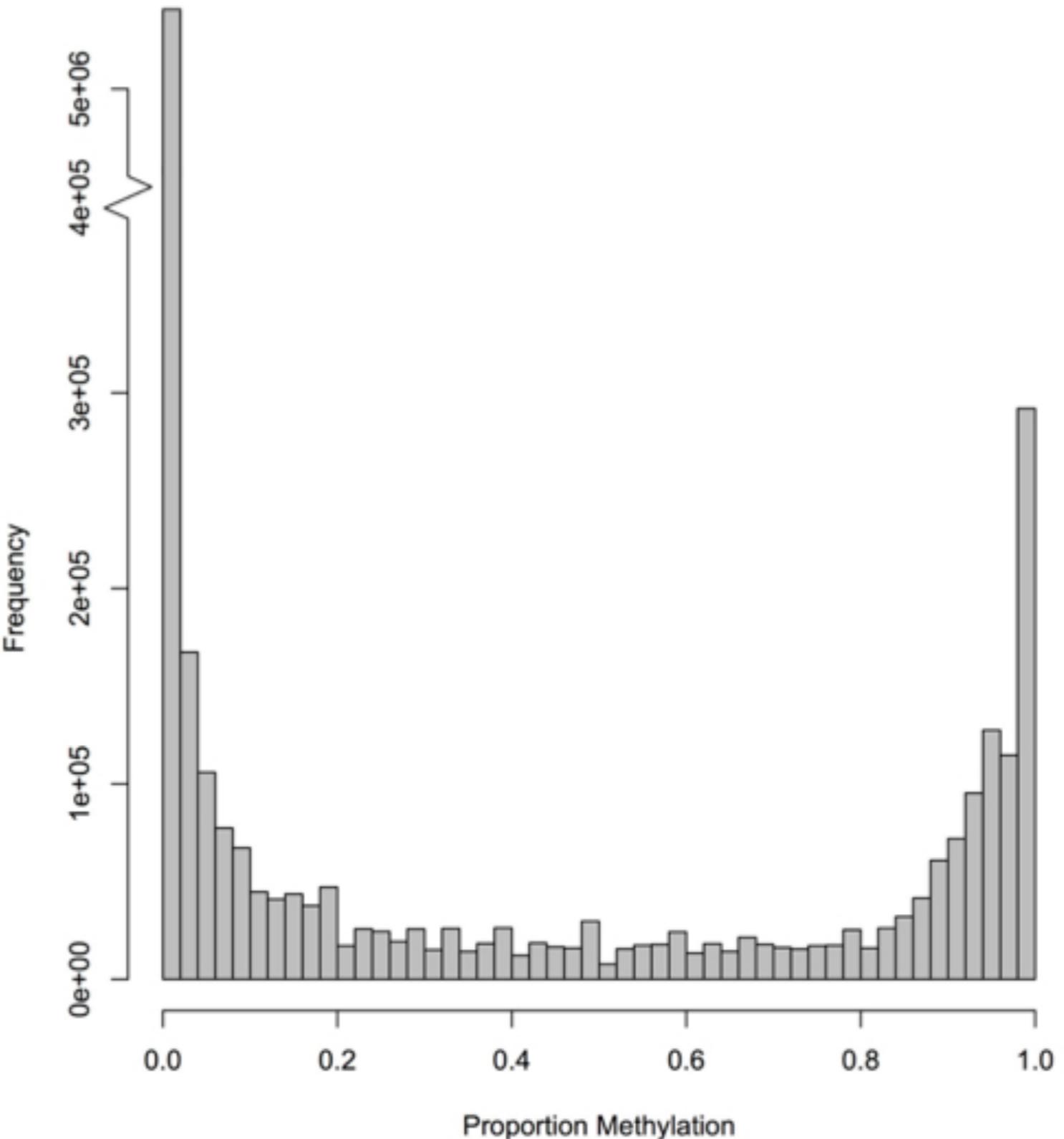
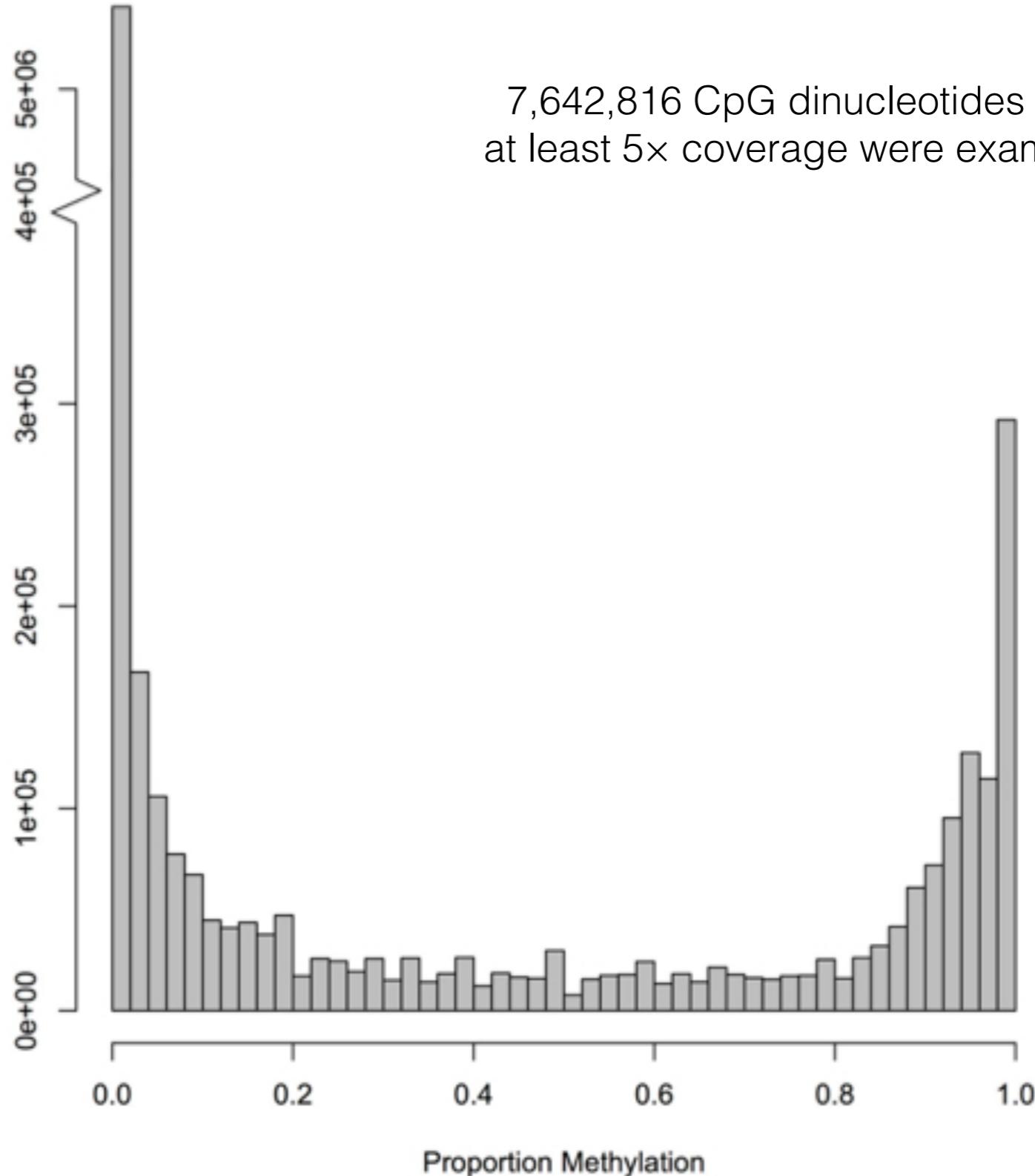
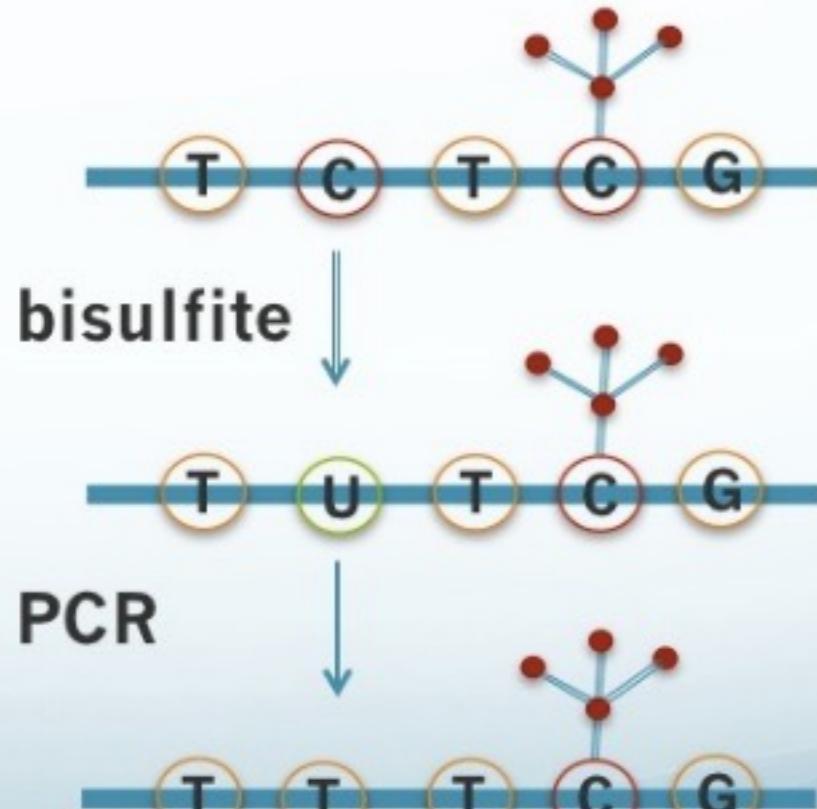


Figure 1 Frequency distribution of methylation ratios for CpG dinucleotides in oyster gill tissue. A total of 2,625,745 CpG dinucleotides with $\geq 5 \times$ coverage are represented.



Epigenetic variation **1**

– Bisulfite conversion

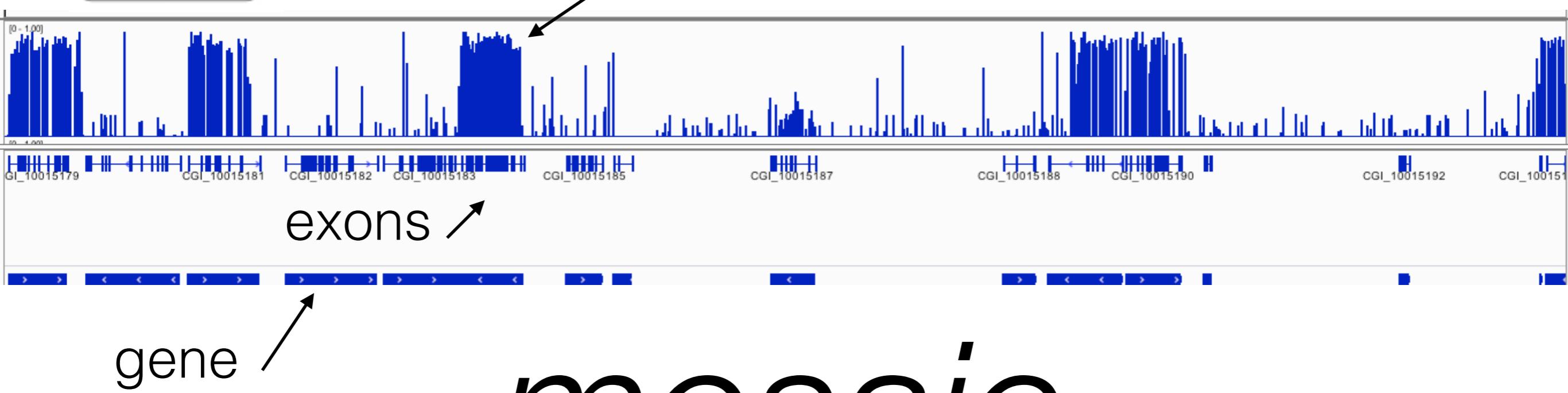


7,642,816 CpG dinucleotides with at least 5x coverage were examined



Epigenetic variation **1**

DNA methylation level (0-100%) @ cytosines

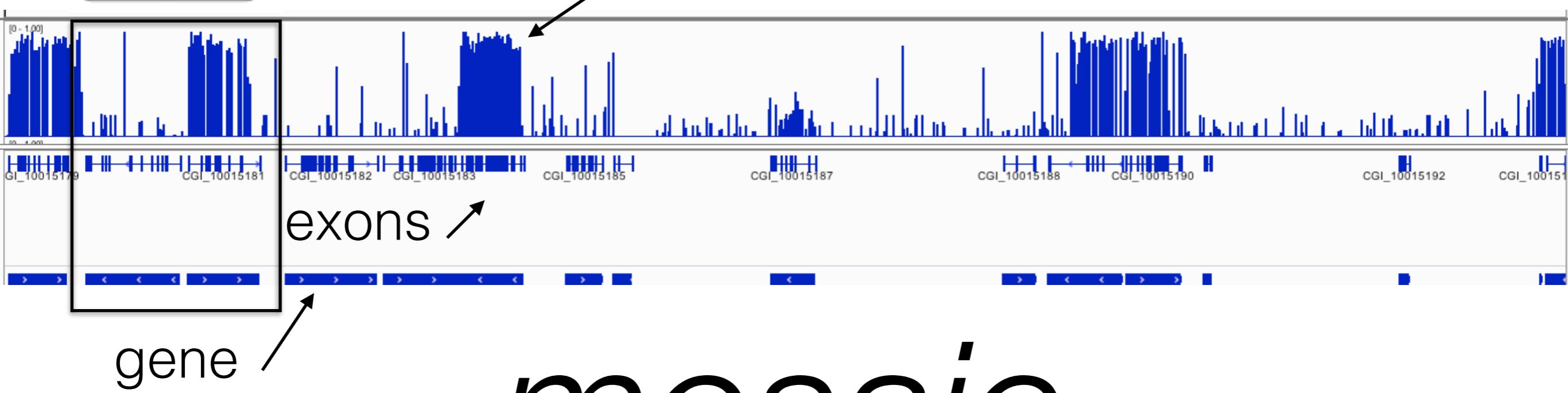


mosaic

associated with gene bodies

Epigenetic variation **1**

DNA methylation level (0-100%) @ cytosines

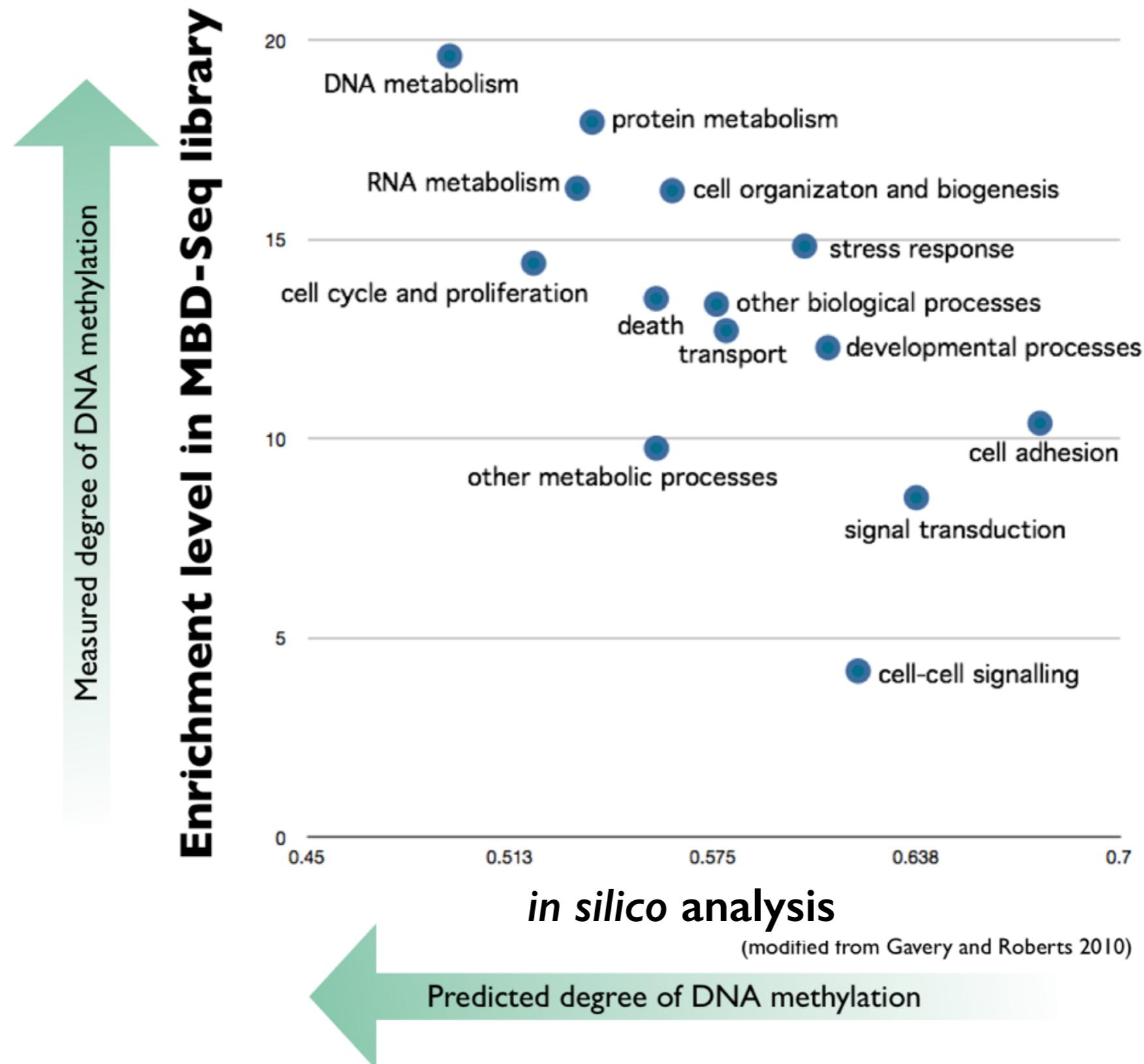


mosaic

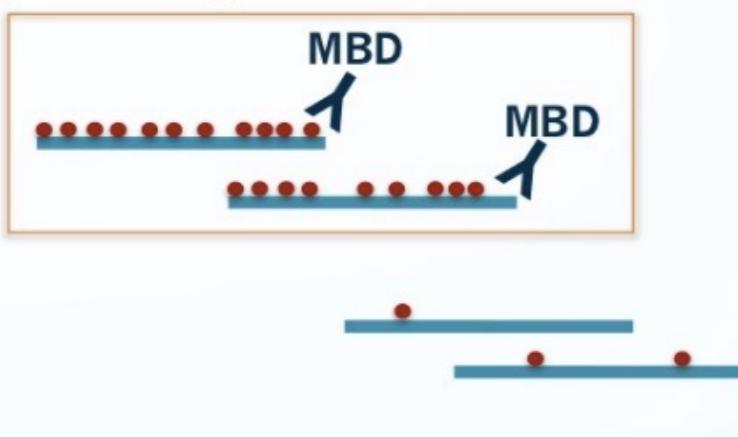
Why are only a subset of genes methylated?

associated with gene bodies

Epigenetic variation **1**

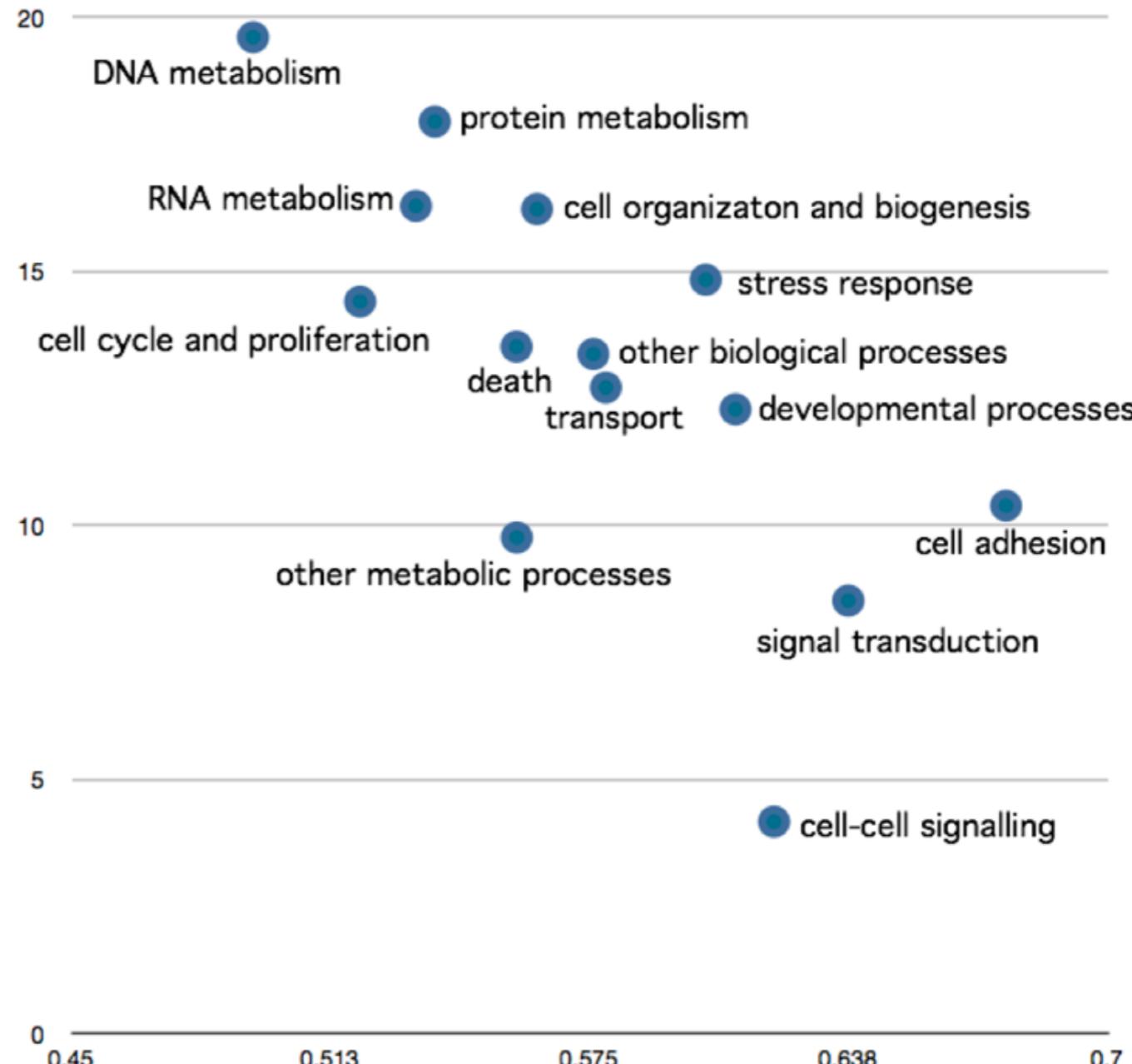


Epigenetic variation 1



Measured degree of DNA methylation

Enrichment level in MBD-Seq library



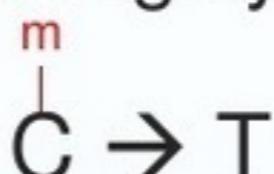
in silico analysis

(modified from Gavery and Roberts 2010)

Predicted degree of DNA methylation

in silico approach

- Principle:
 - Methylated cytosines are highly mutable

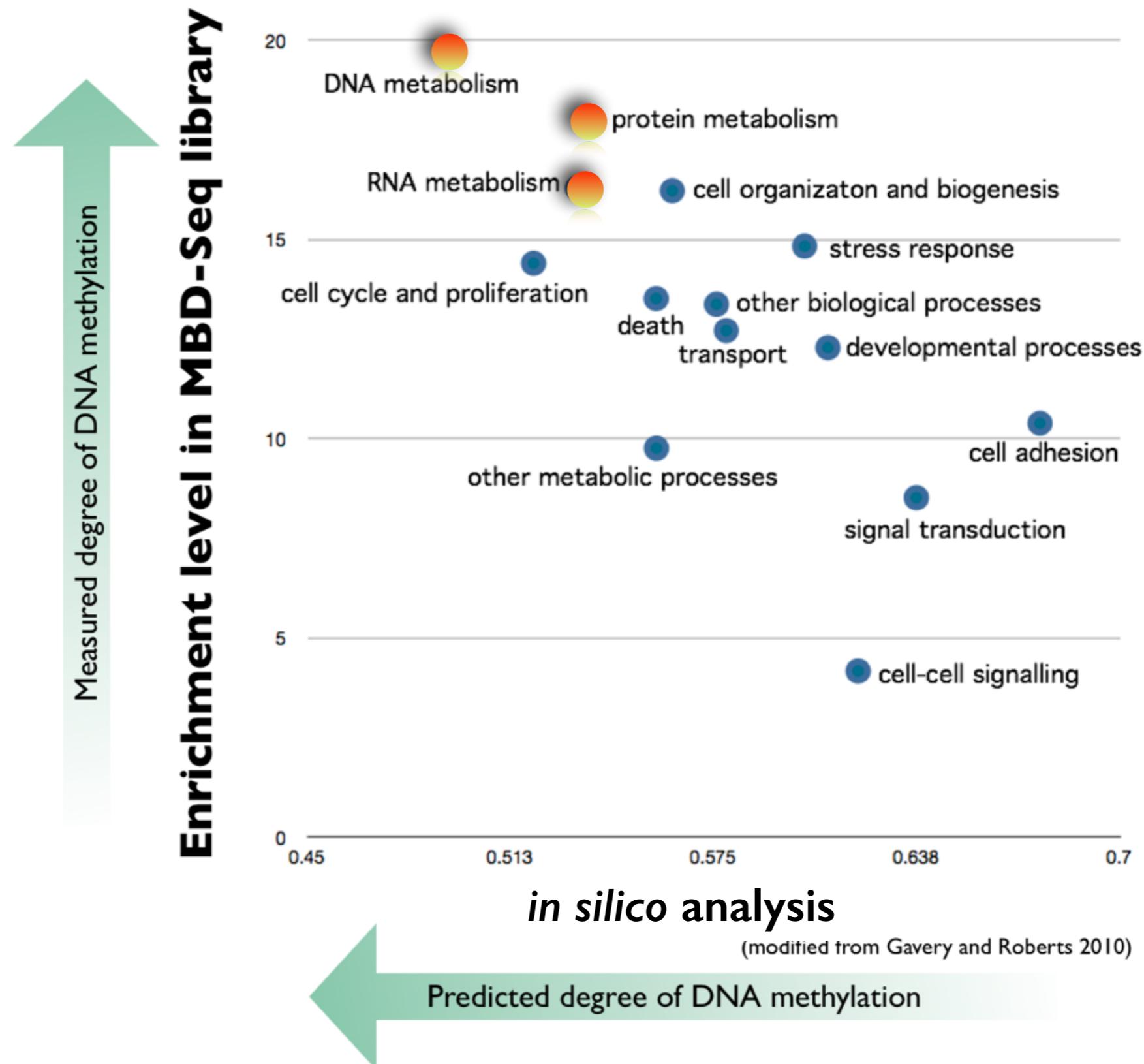


- Methylated regions of DNA are depleted of CpG dinucleotides over evolutionary time (CpG to TpG)

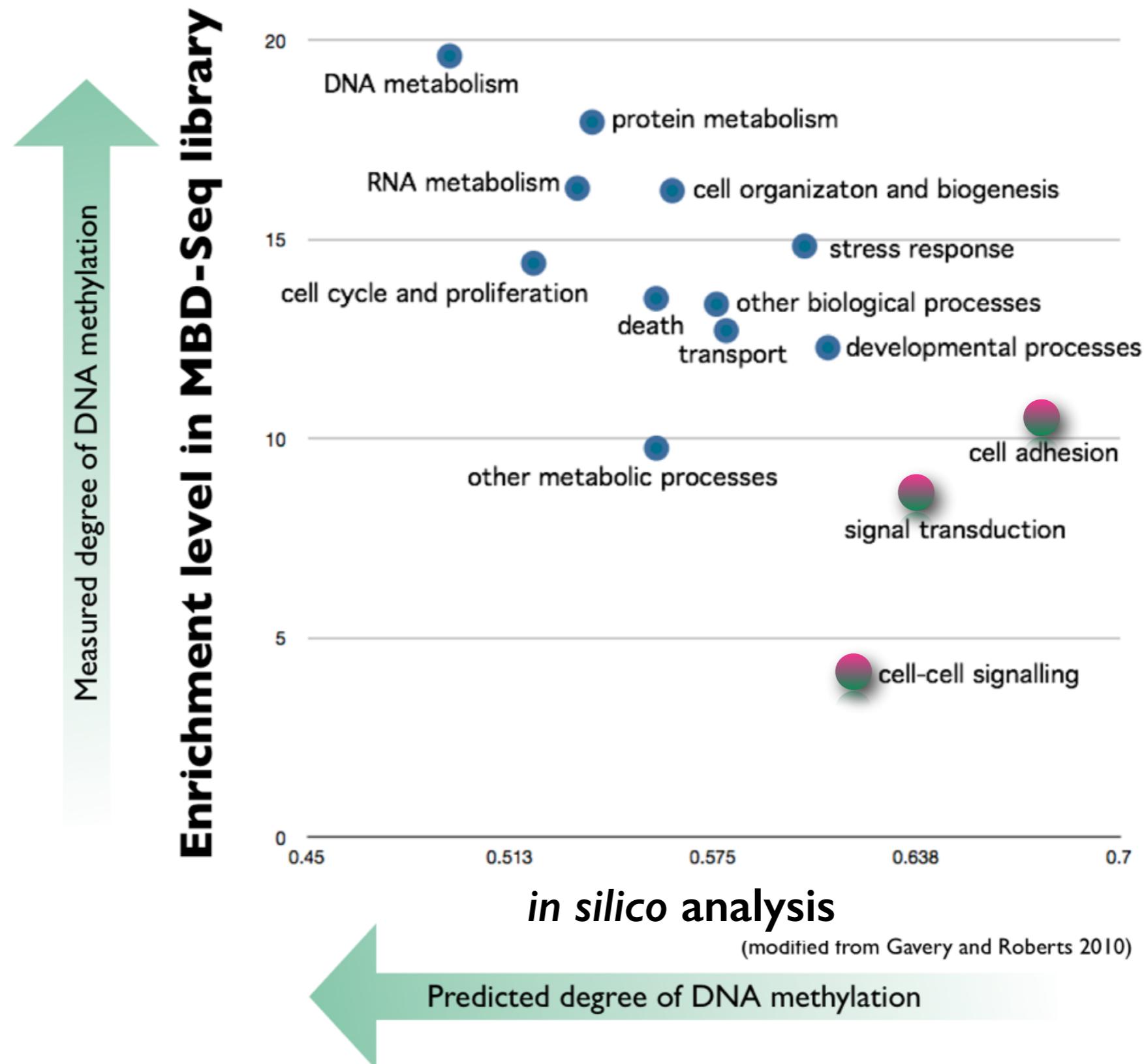
$$\text{CpG O/E} = \frac{\text{CpG observed}}{\text{CpG expected}}$$


high = unmethylated

Epigenetic variation **1**



Epigenetic variation **1**



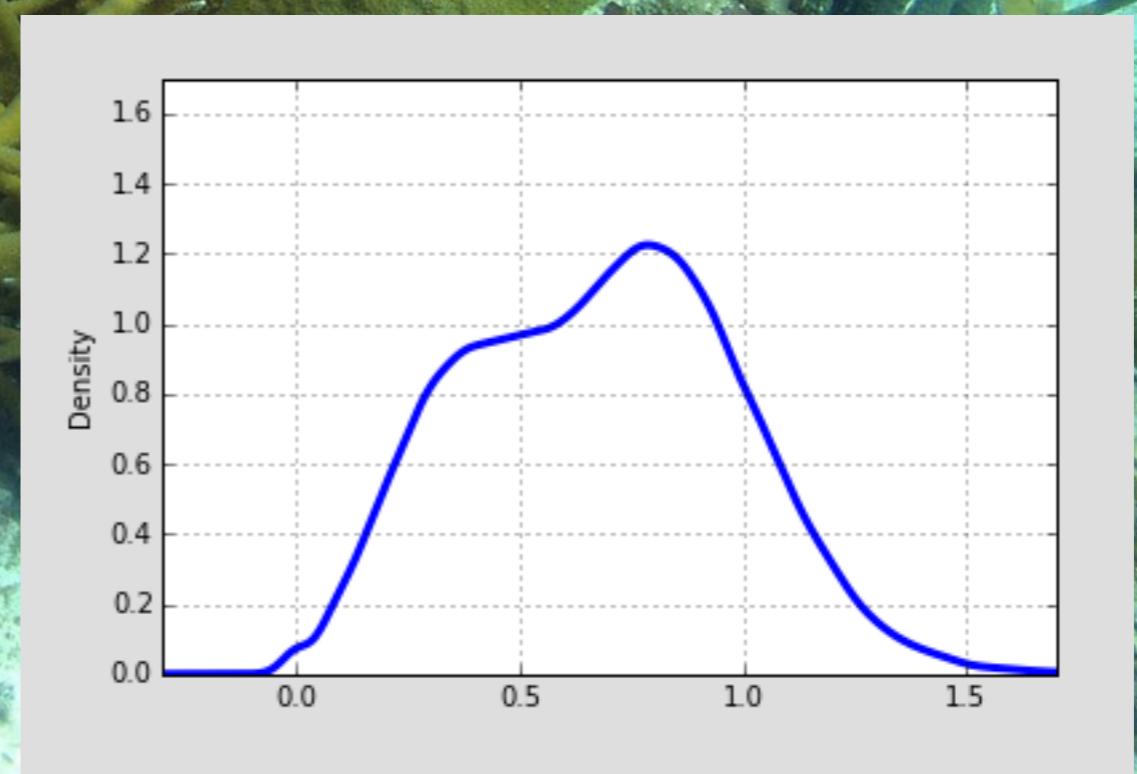
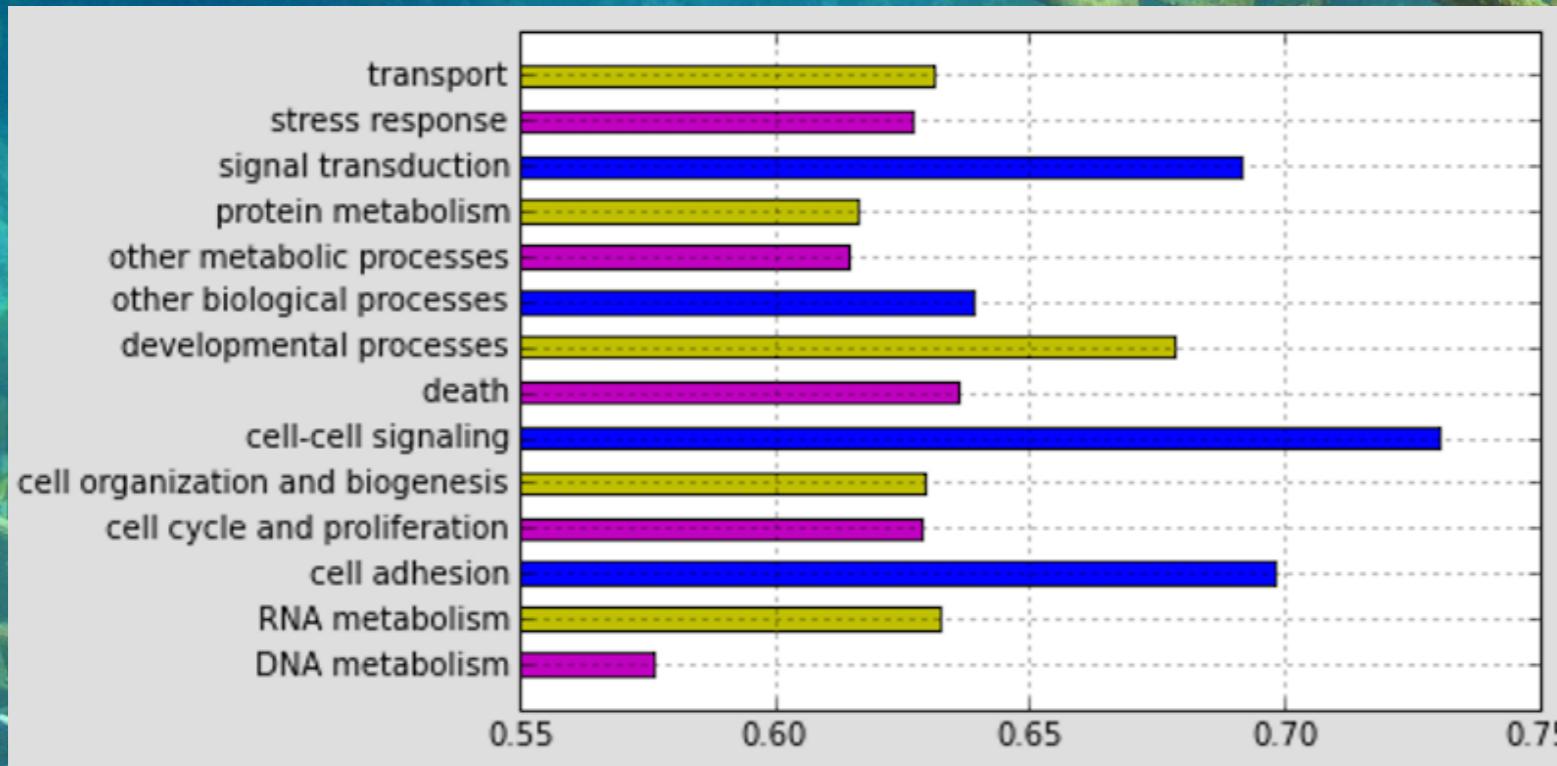
Jay Dimond

Acropora palmata

Acropora cervicornis

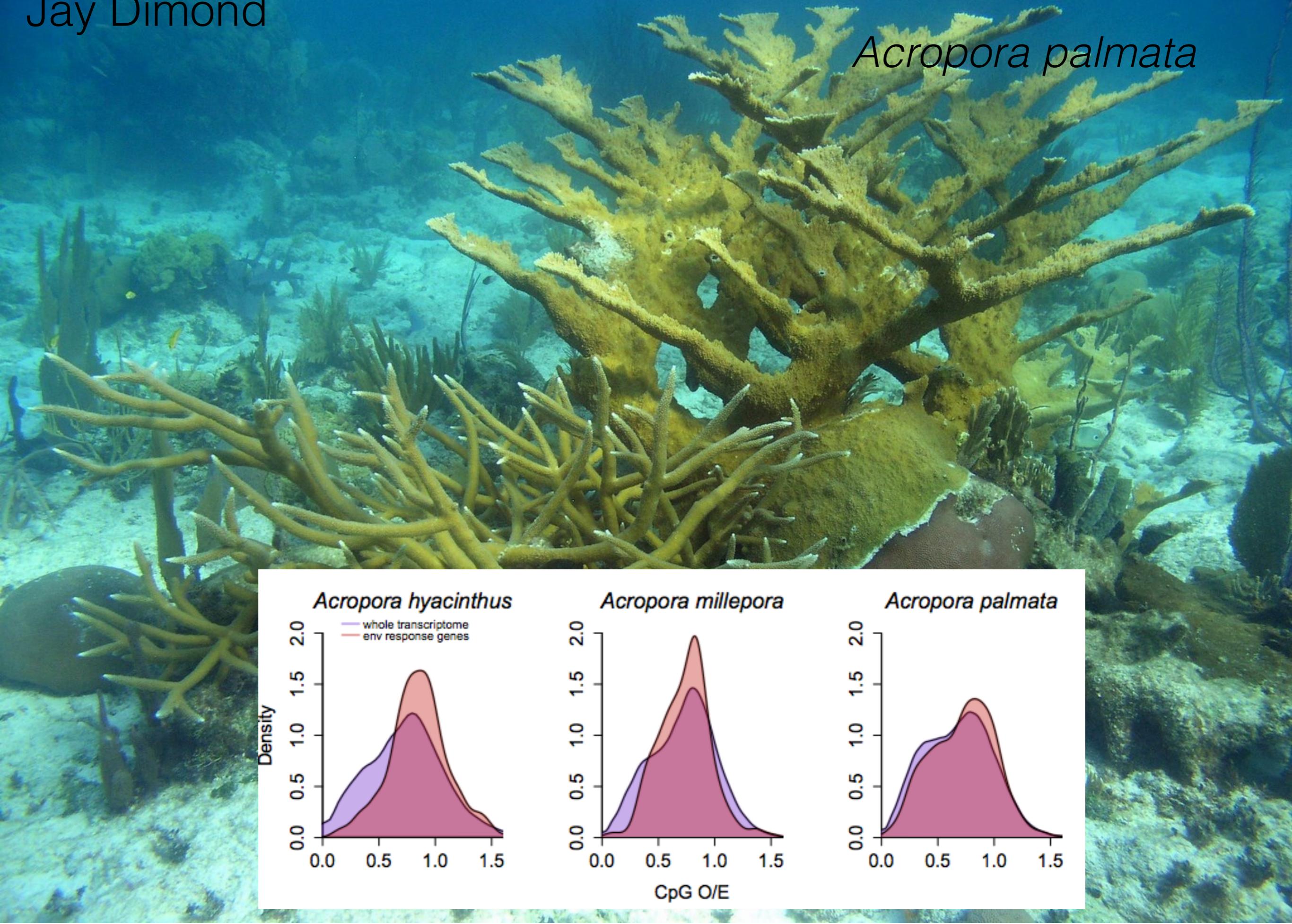
Jay Dimond

Acropora palmata



Jay Dimond

Acropora palmata

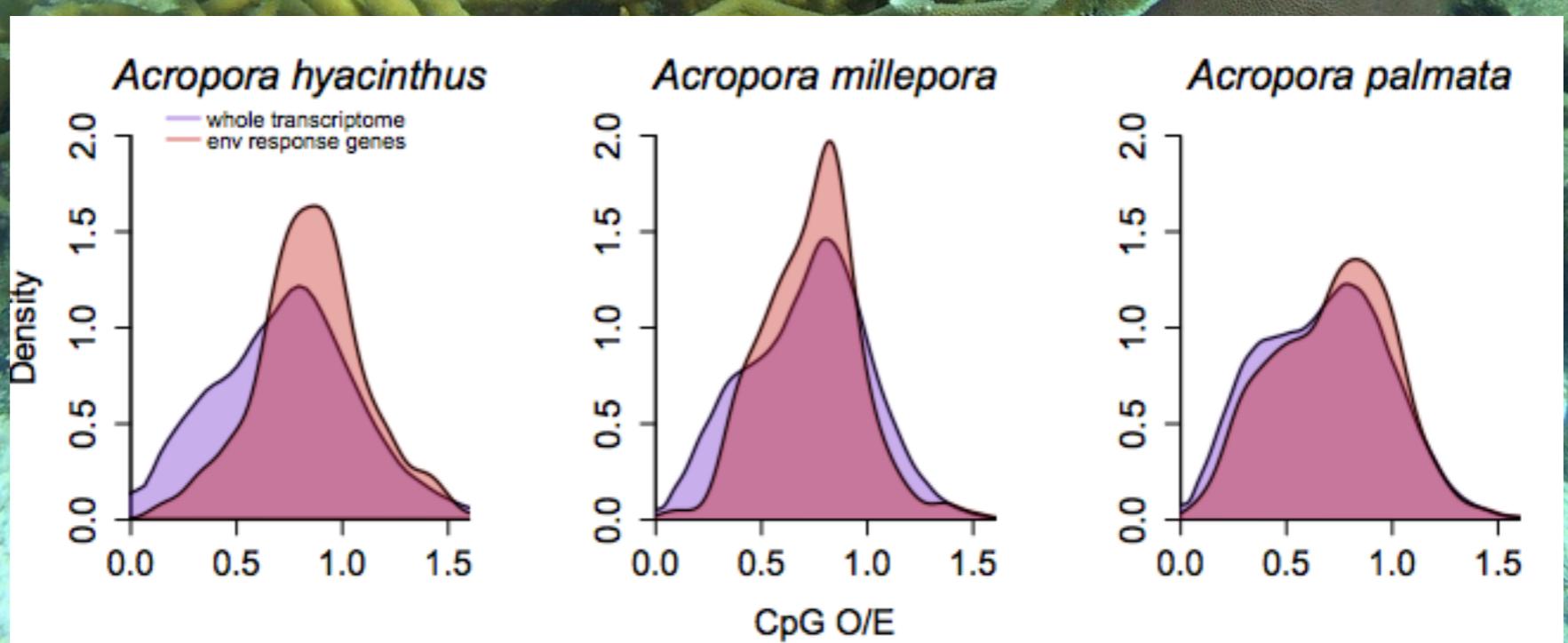


Jay Dimond

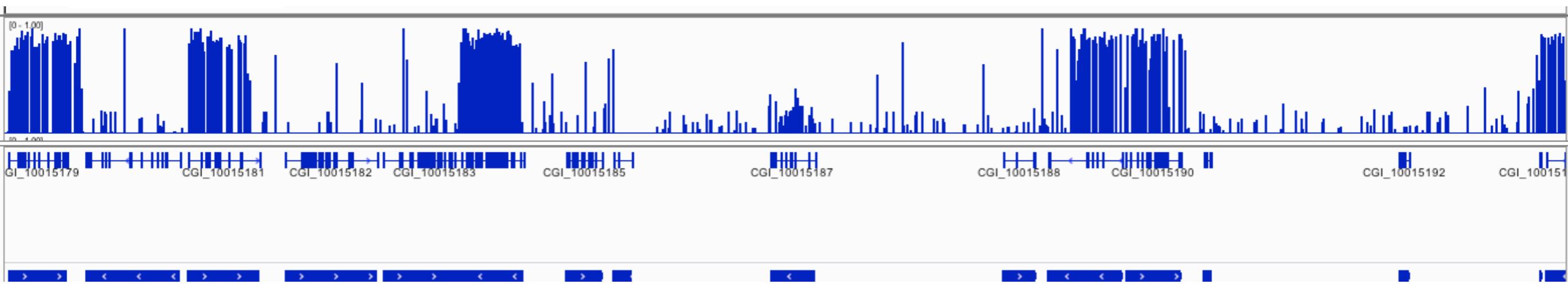


Acropora palmata

Germline DNA methylation in five coral transcriptomes



Epigenetic variation **1**



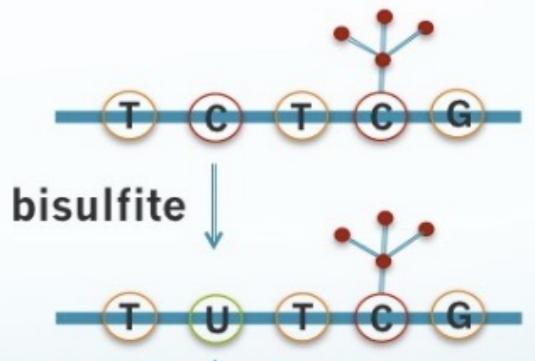
mosaic

associated with gene bodies
based on gene function

explanation?

Epigenetic variation **1**

Family and Developmental Variation



Sperm &
Larvae
(72h & 120h)



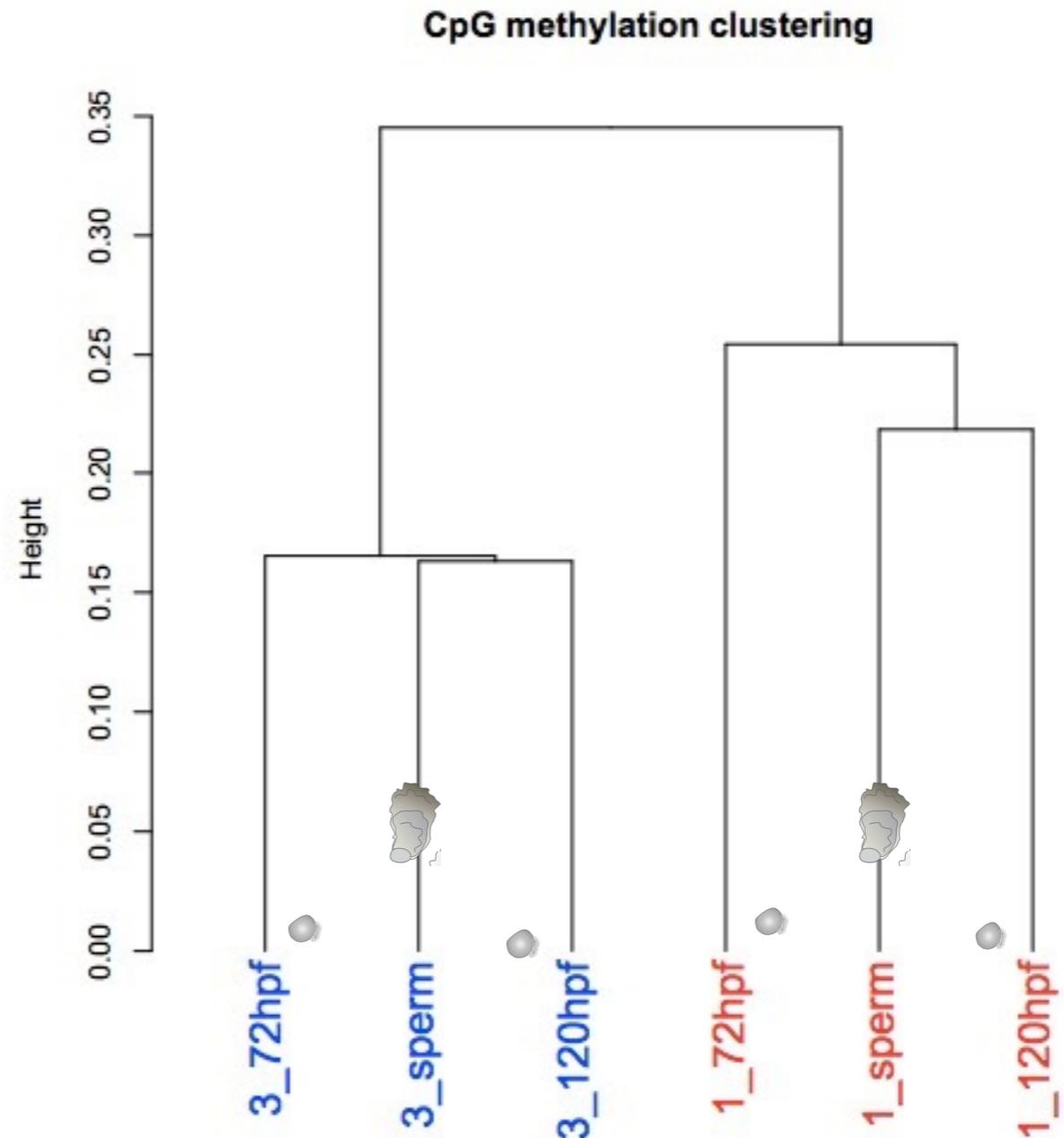
bioRxiv
beta
THE PREPRINT SERVER FOR BIOLOGY

New Results

Indication of family-specific DNA methylation patterns in developing oysters

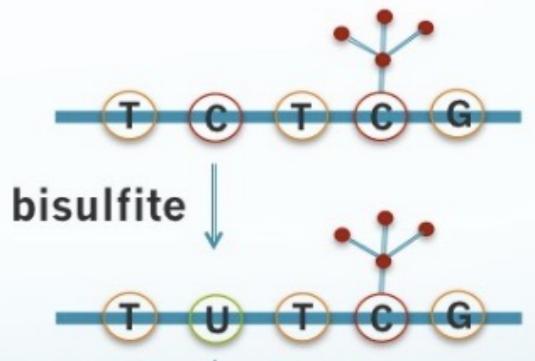
Claire E. Olson , Steven B. Roberts

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

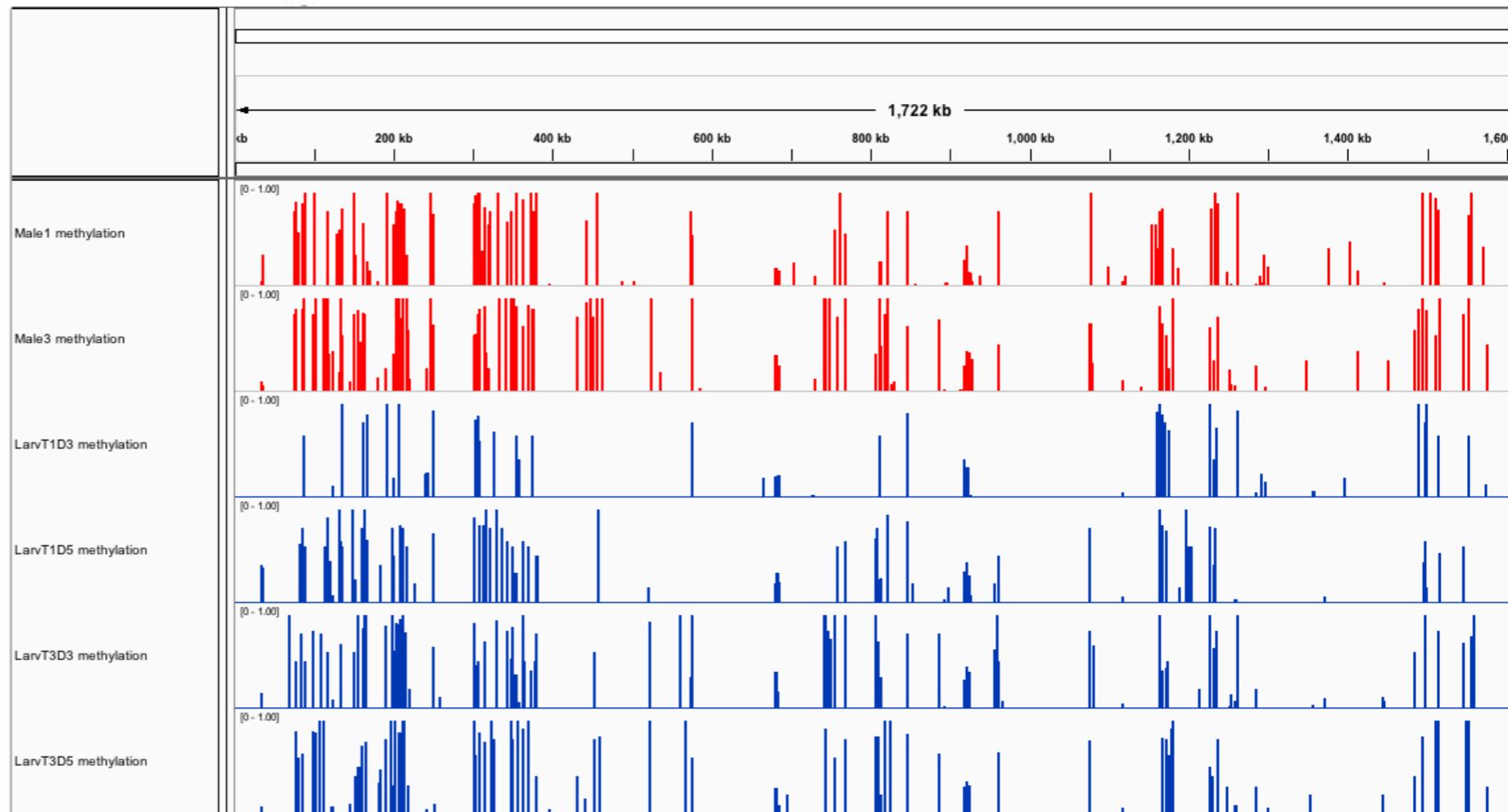


Epigenetic variation 1

Family and Developmental Variation



Sperm &
Larvae
(72h & 120h)



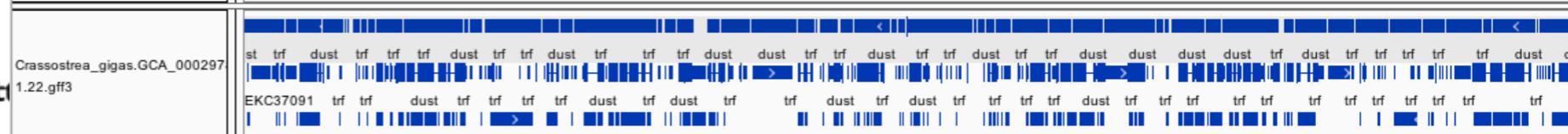
bioRxiv
beta

THE PREPRINT SERVER FOR BIOLOGY

DNA methylation level (0-100%) @ cytosines

New Results

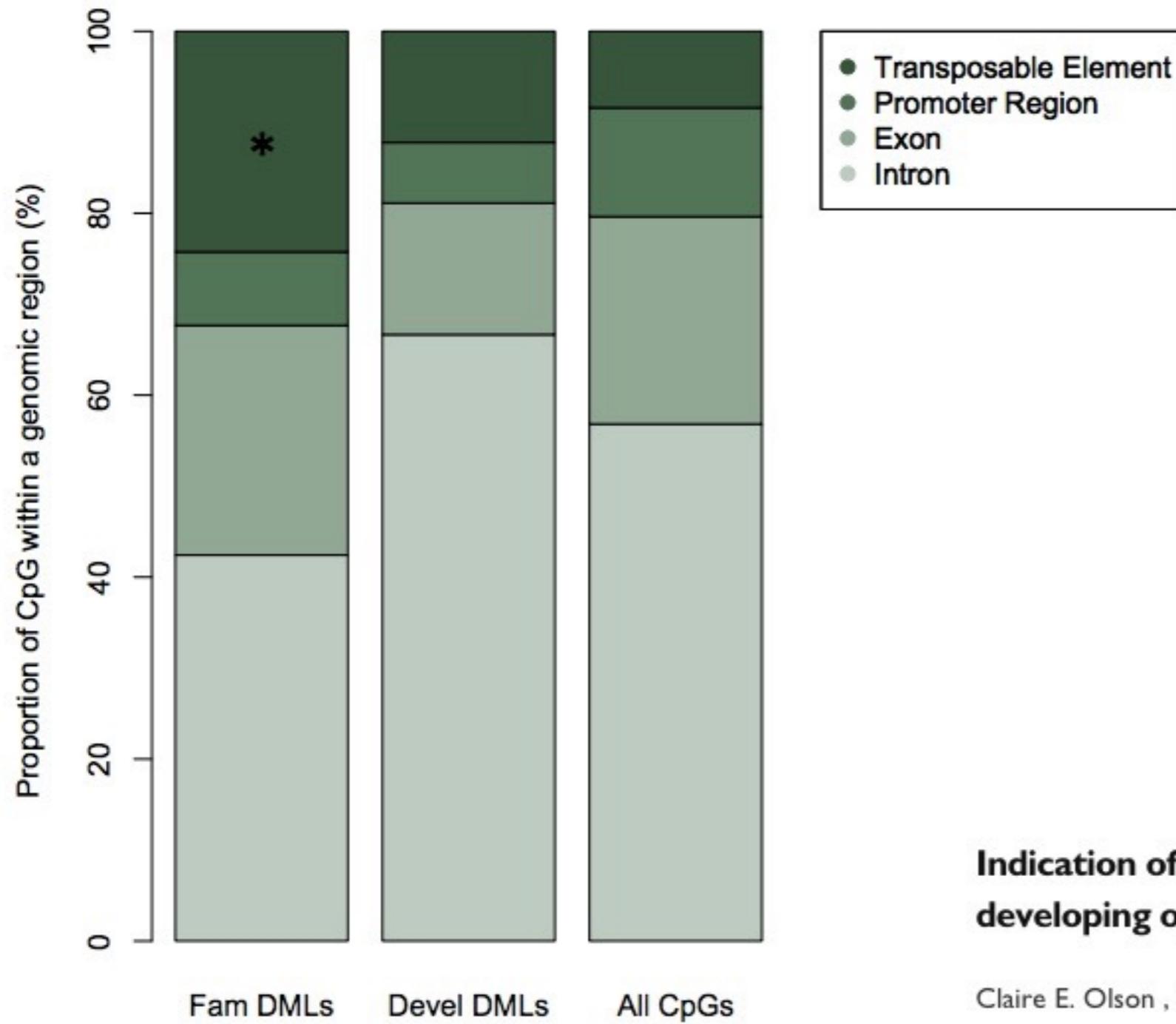
Indication of family-specific DNA methylation patterns in developing oysters



Claire E. Olson, Steven B. Roberts

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

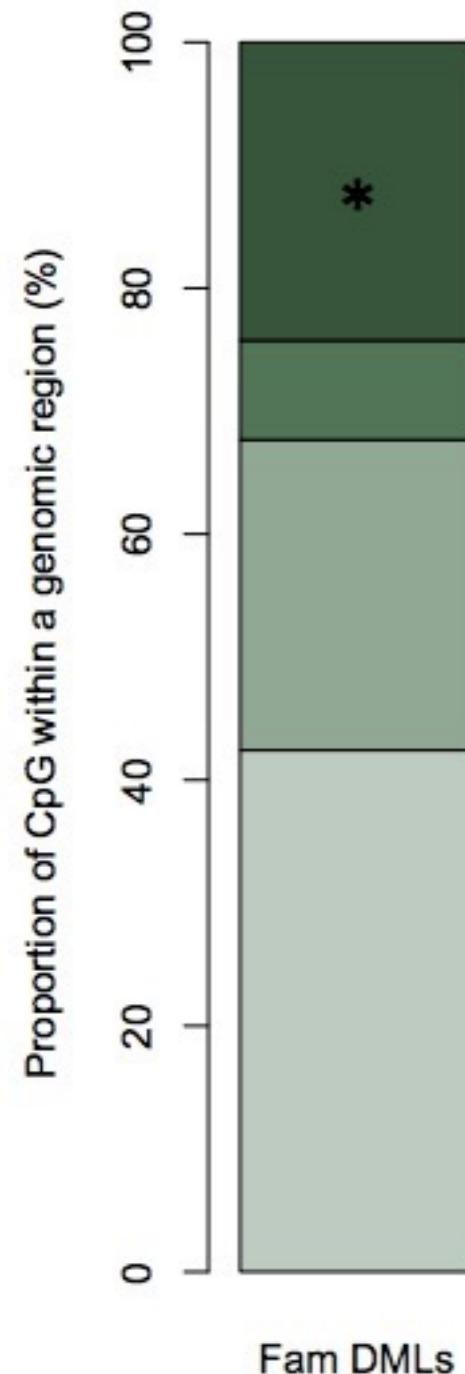
Epigenetic variation **1**



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**



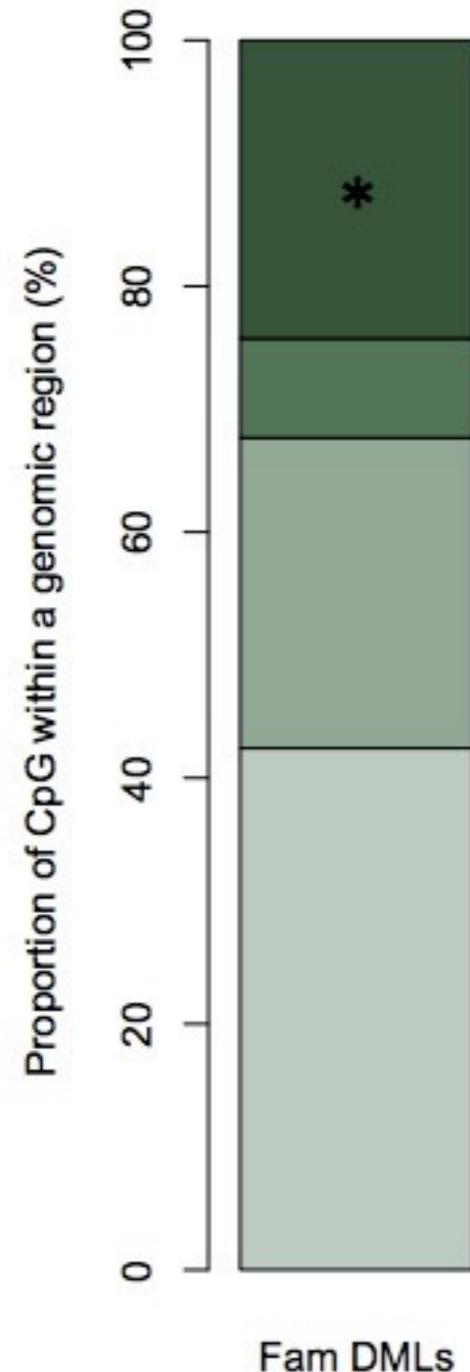
Differentially methylated loci in gene bodies could be lethal or deleterious, as they would alter gene expression.

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>

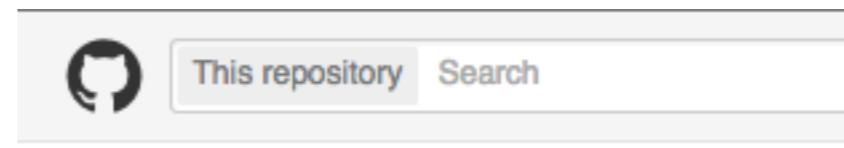
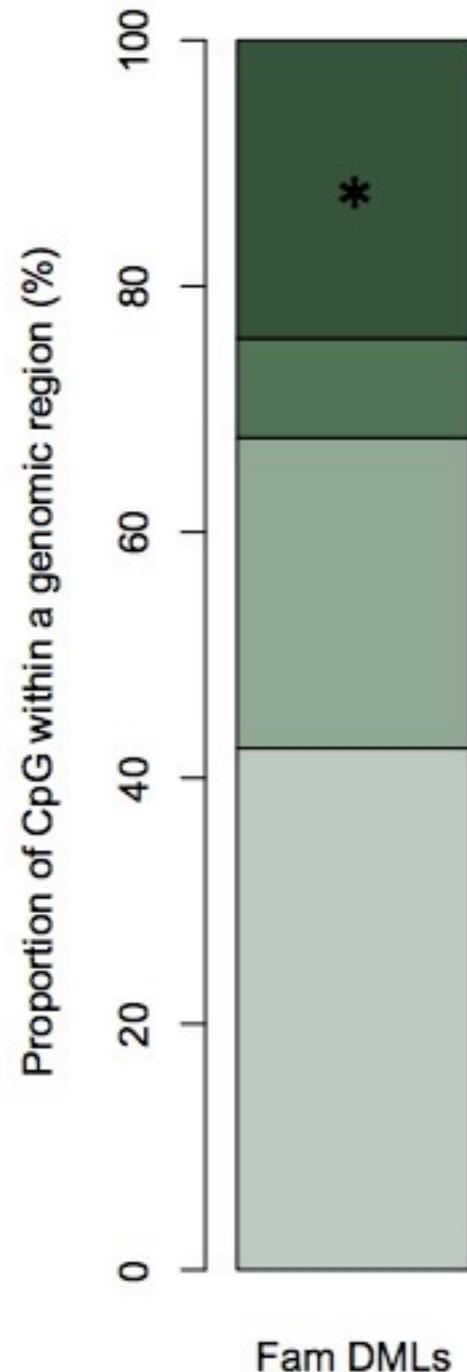


***Assumes some form
of random process***
occurring during
gametogenesis?

Indication of family-specific DNA methylation patterns in developing oysters

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

Epigenetic variation **1**



[che625 / olson-ms-nb](#)



**IPython Notebook for downloading and analyzing data
the manuscript: "Indication of family-specific DNA
methylation patterns in developing oysters"**

bioRxiv preprint - <http://dx.doi.org/10.1101/012831>

To execute the IPython Notebook in its entirety you will need:

- IPython - [install instructions](#)
- BSMAP - [install instructions](#)
- bedtools - [install instructions](#)
- R - [install instructions](#)
- rpy2 (interface to R from Python) - [install instructions](#)

1 Summary

- Sparsely (~16 %) methylated genome
- Limited variation between cell type
- Gene body methylation correlated with function
- Evidence indicates DMRs are predominant in transposable elements

Gene
expression



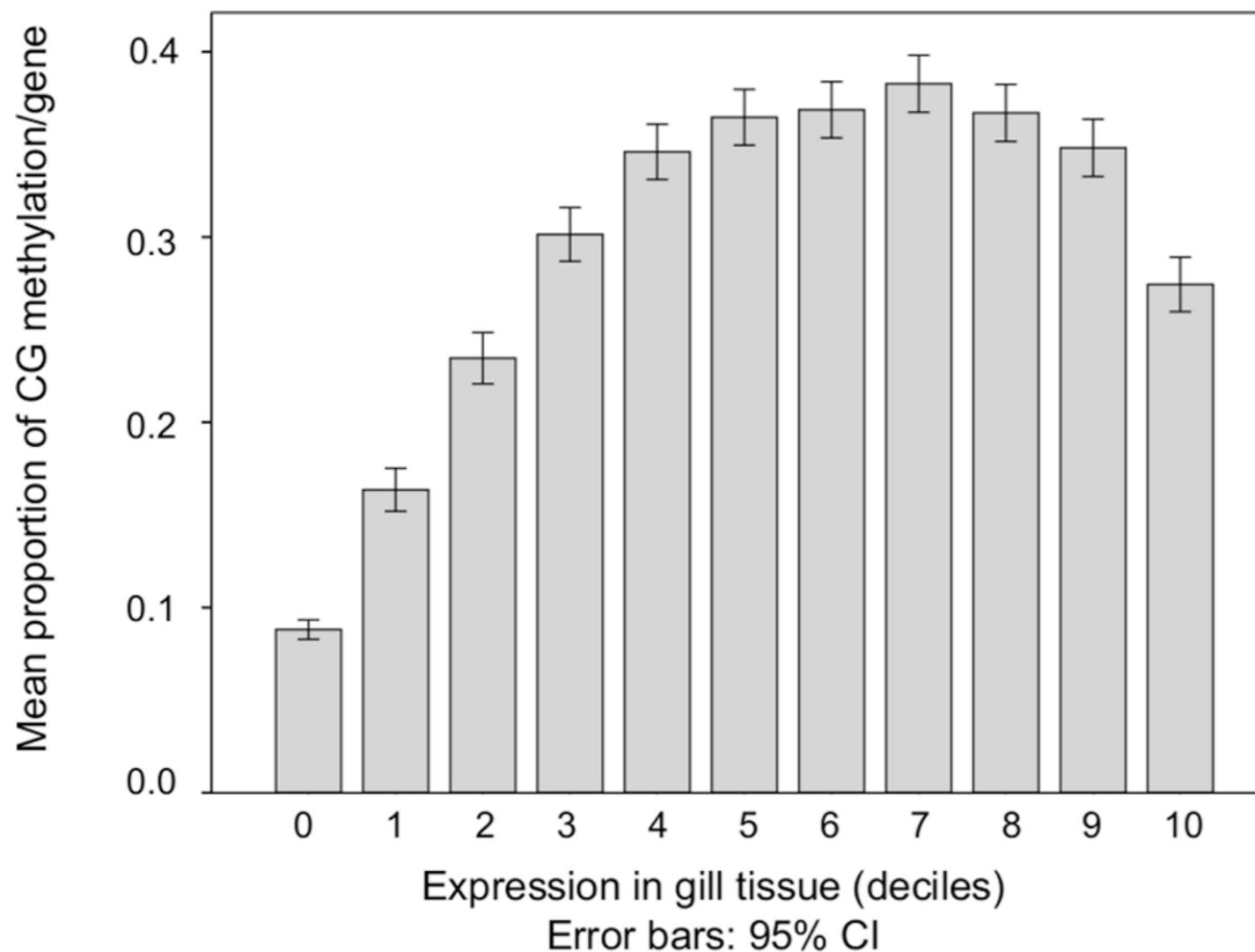
Epigenetic
variation

Function?

Gene expression

2

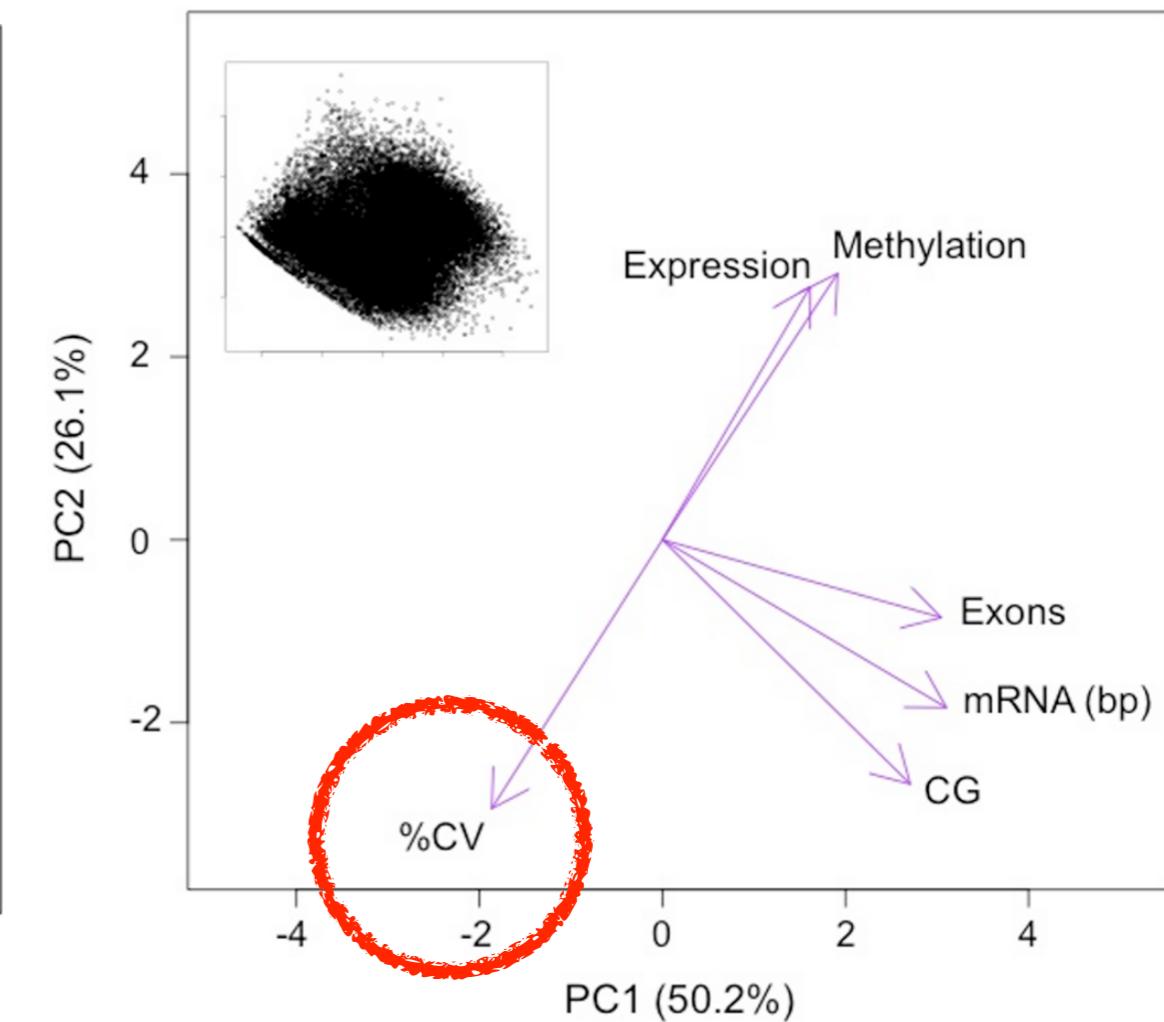
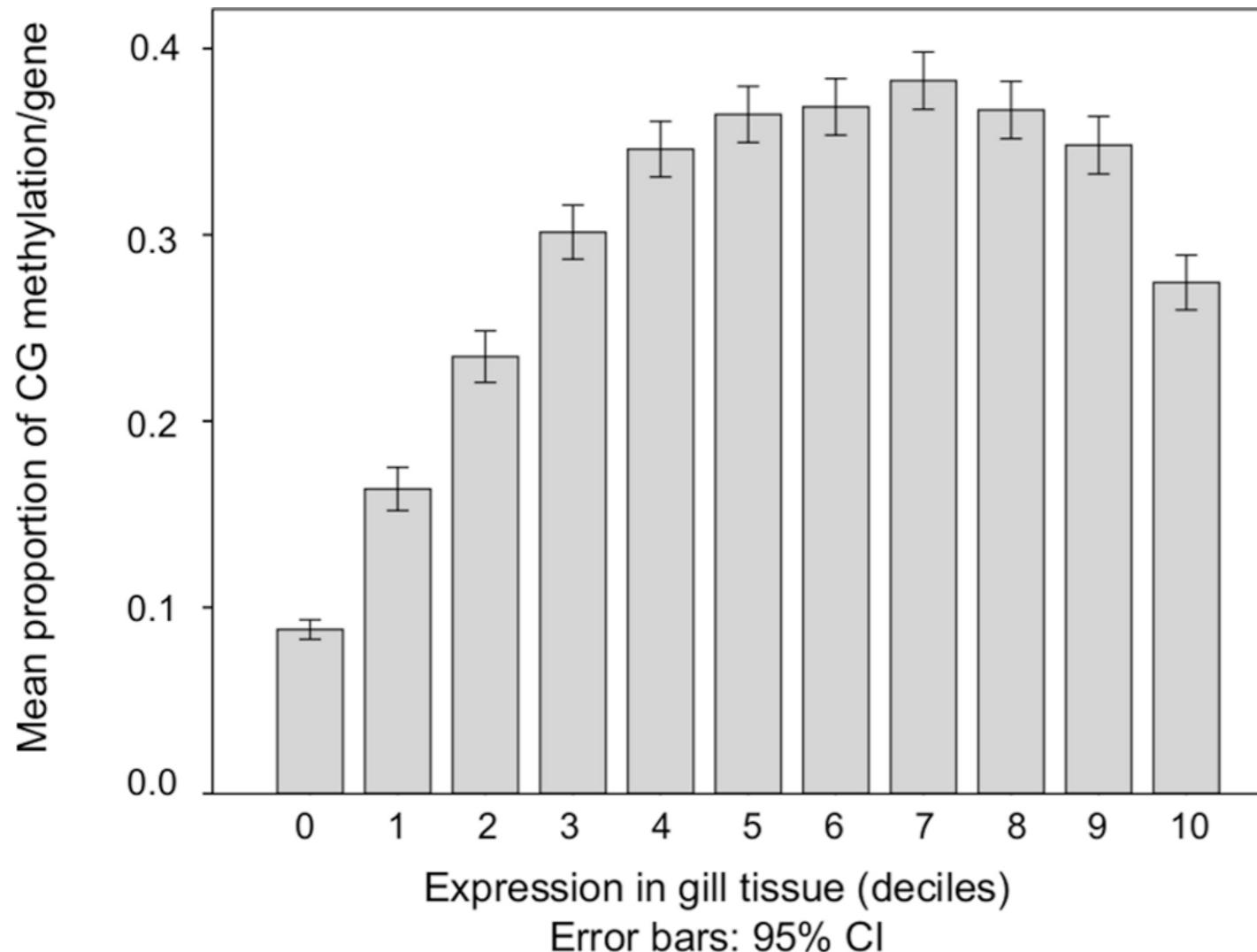
Epigenetic variation



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

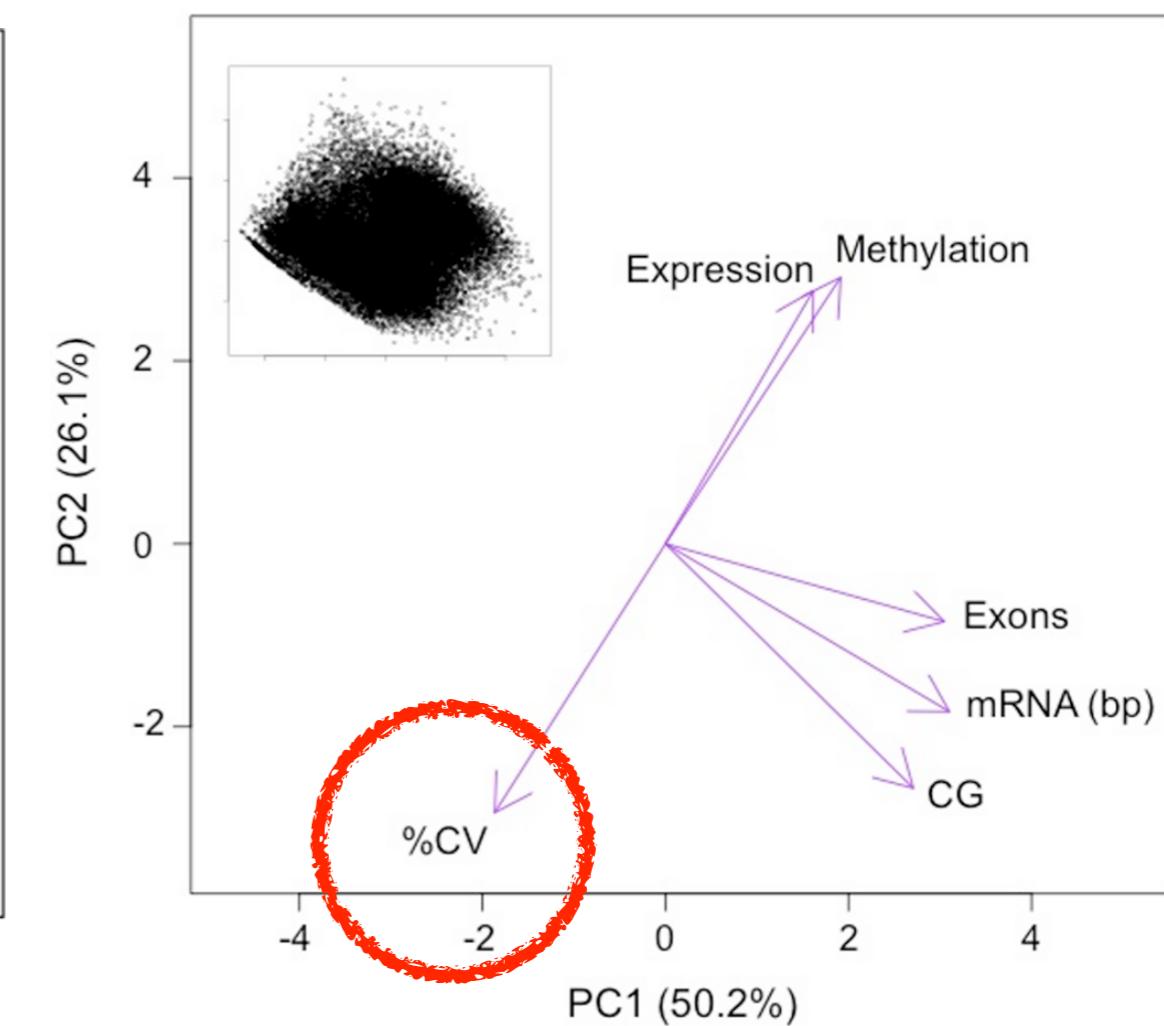
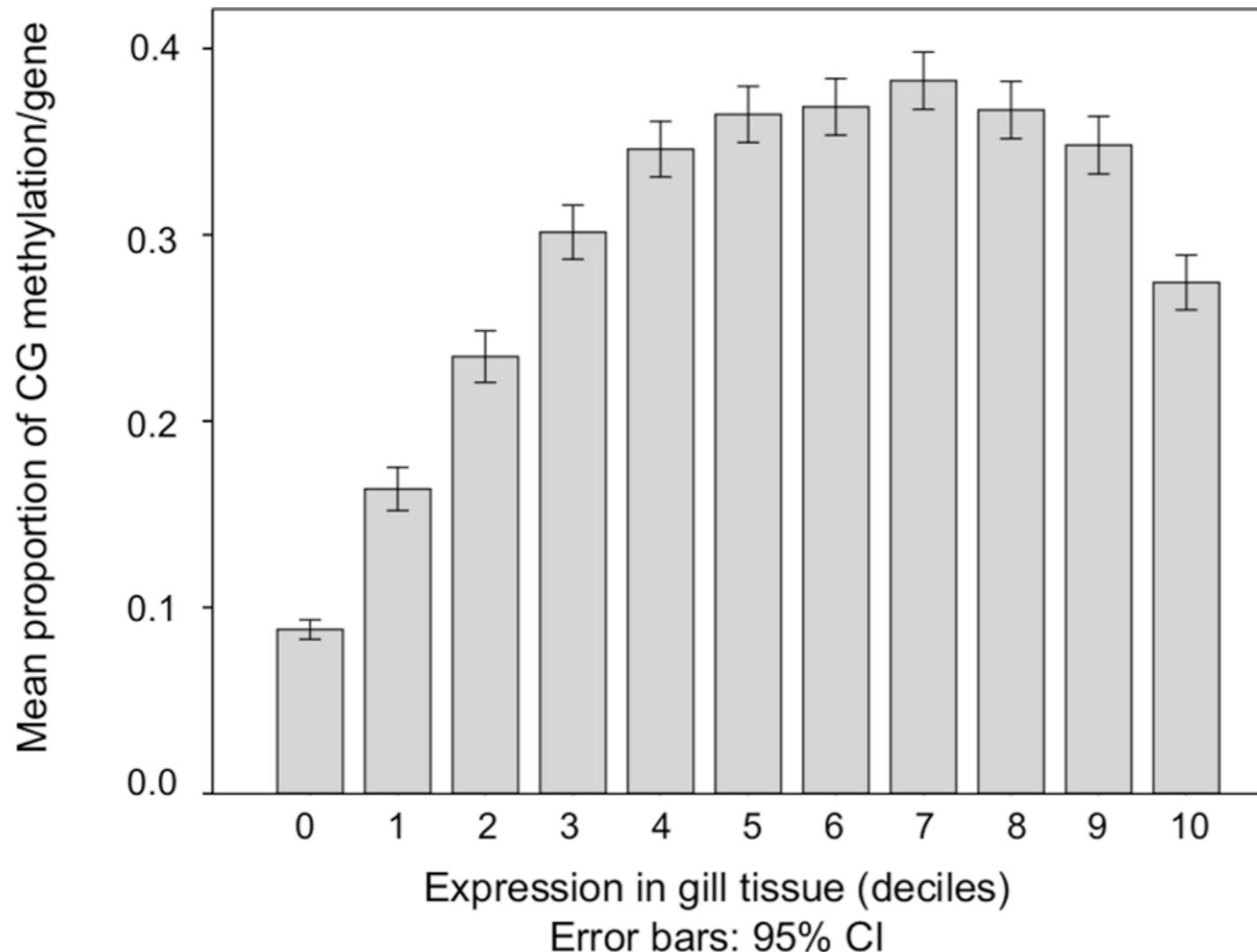
School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA

Gene expression

2

Epigenetic variation

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



What?

A context dependent role for DNA methylation in bivalves

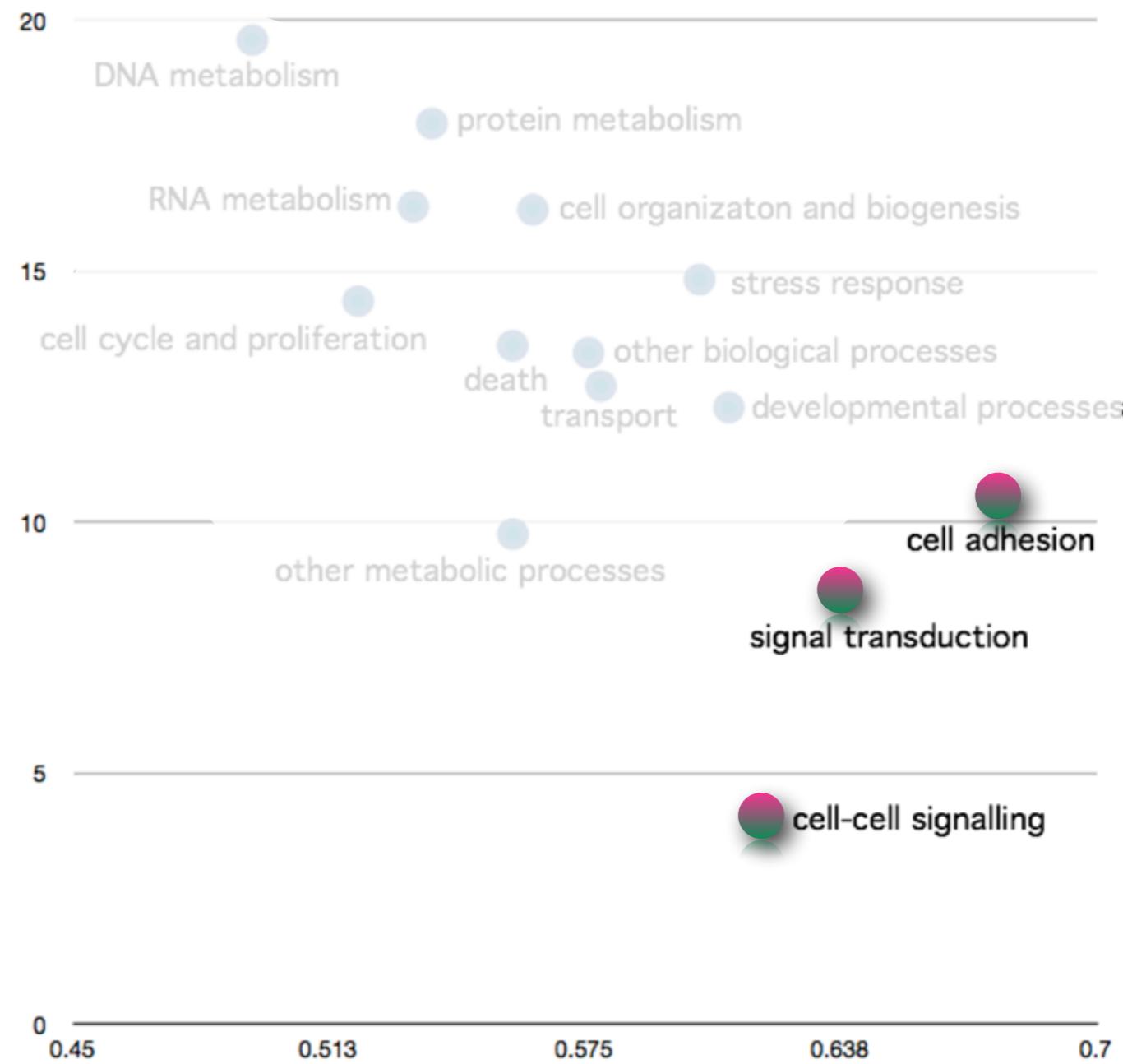
Mackenzie R. Gavery and Steven B. Roberts

Advance Access publication date 7 January 2014

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

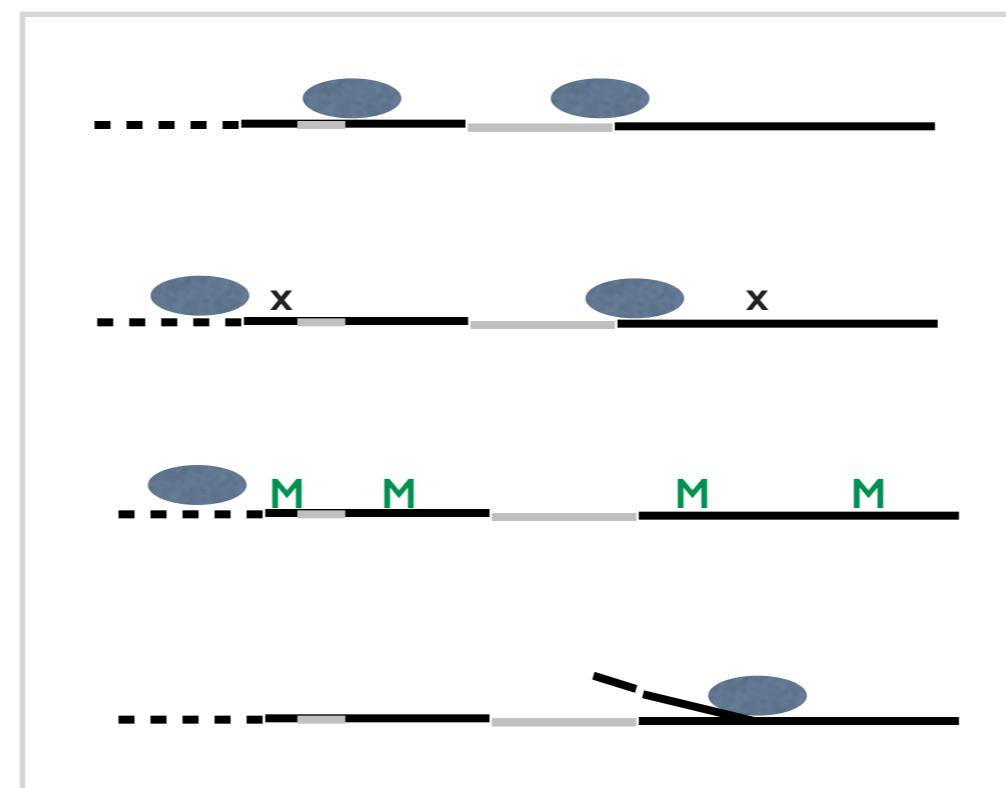
signal transduction

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

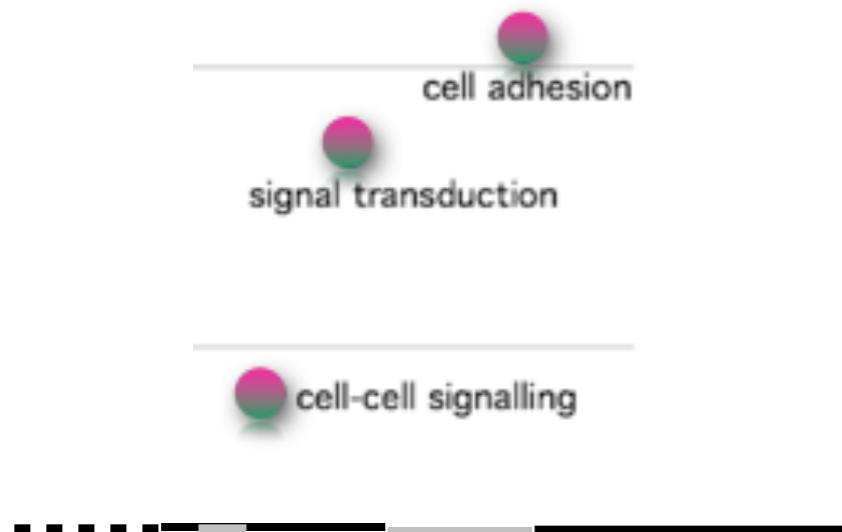
transcript



Gene expression

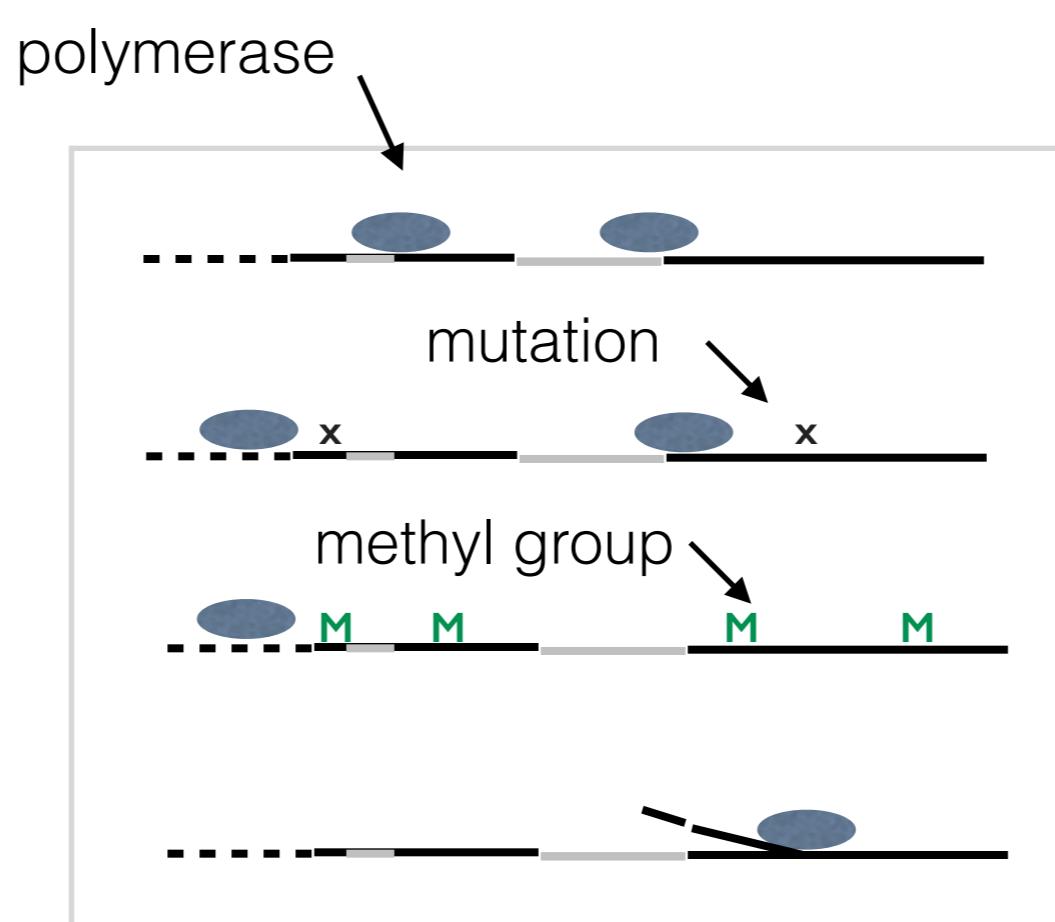
2

Epigenetic variation



sparse methylation

tissue / temporal specific and inducible genes



Gene
expression



Epigenetic
variation

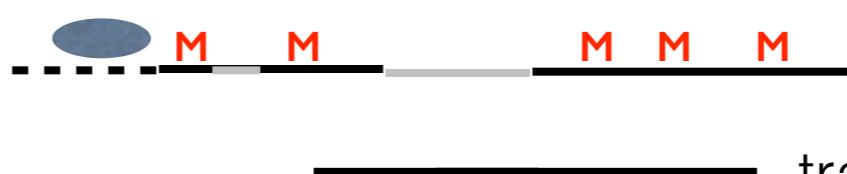
Transcriptional opportunities



germline methylation

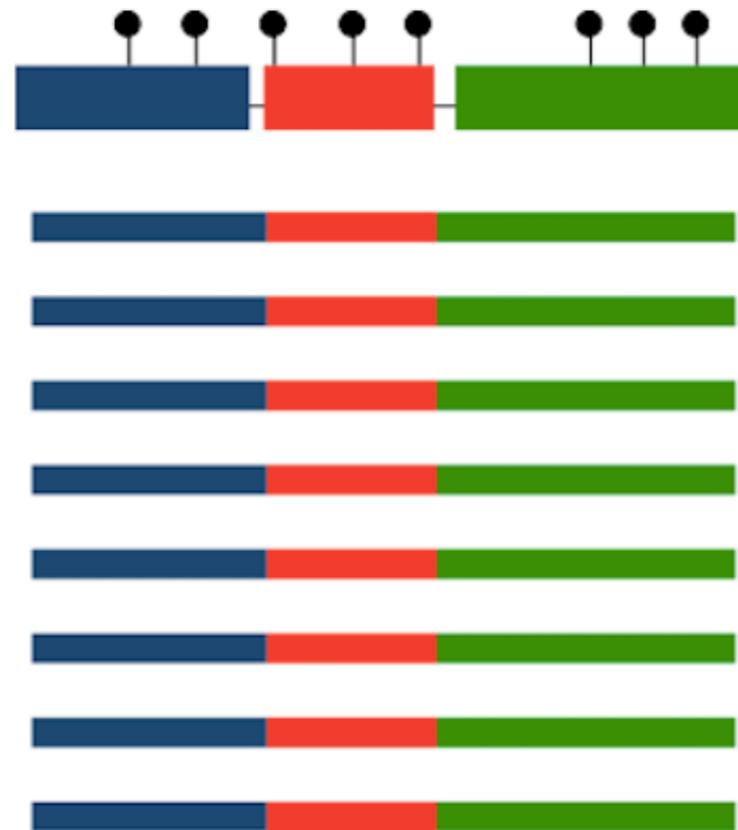
*ubiquitously expressed,
critical genes*

methyl group

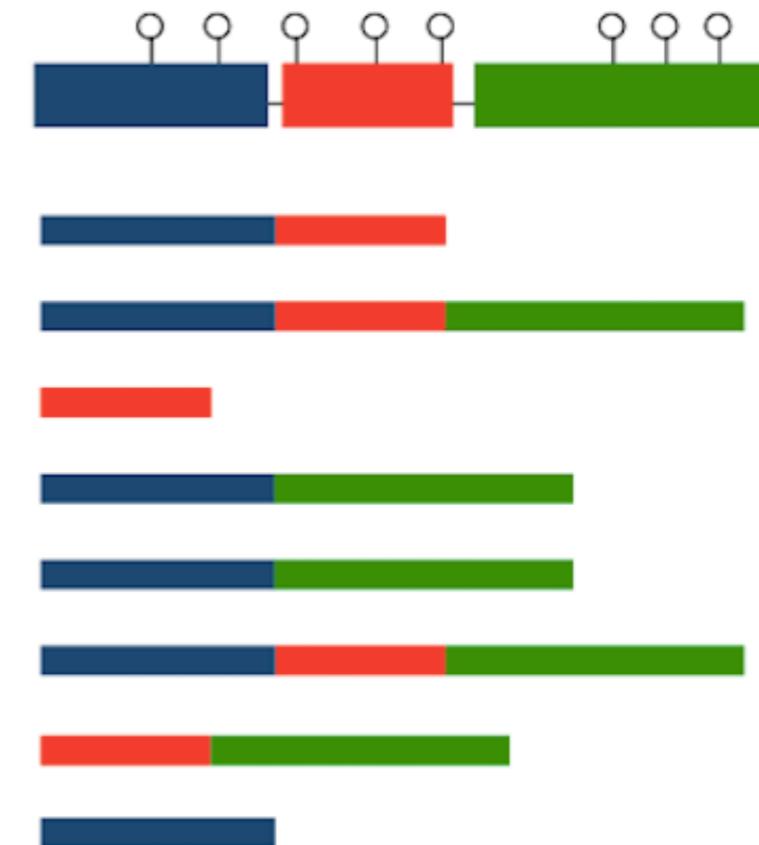


conventional transcription
5' UTR promotor

transcript



housekeeping



response to
change

A context dependent role for DNA methylation in bivalves

Mackenzie R. Gavery and Steven B. Roberts

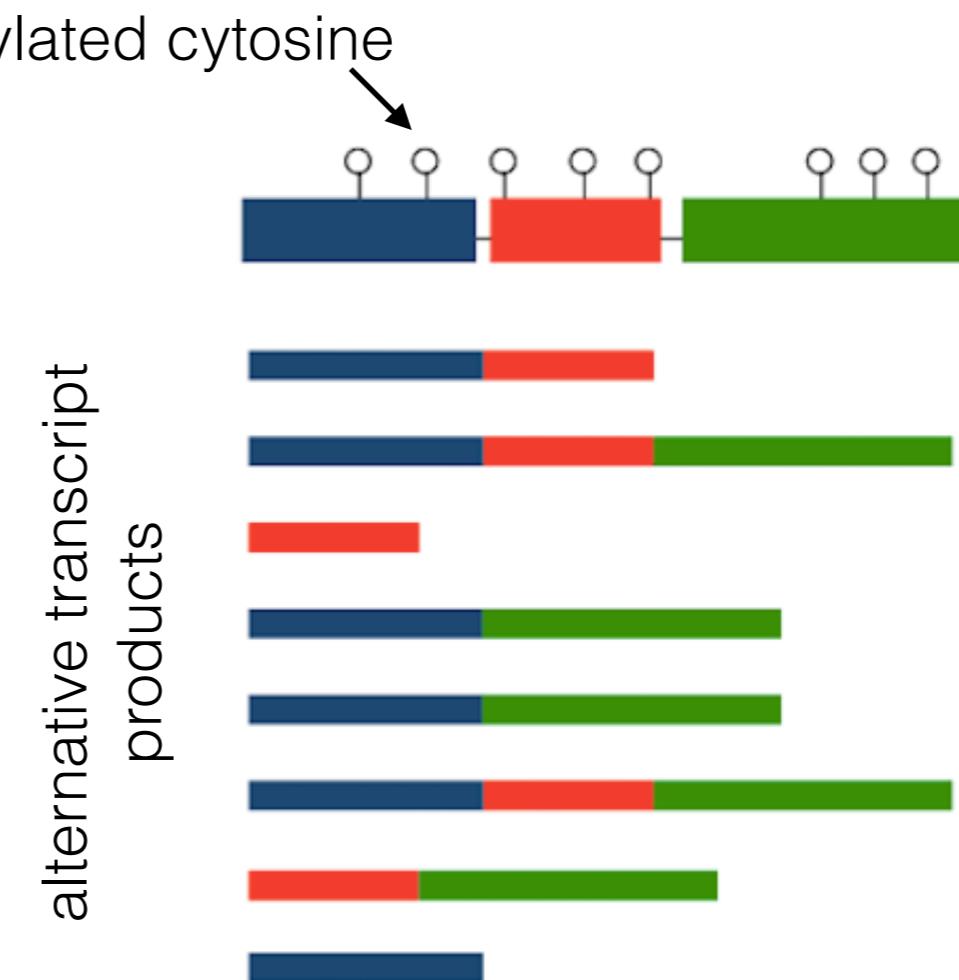
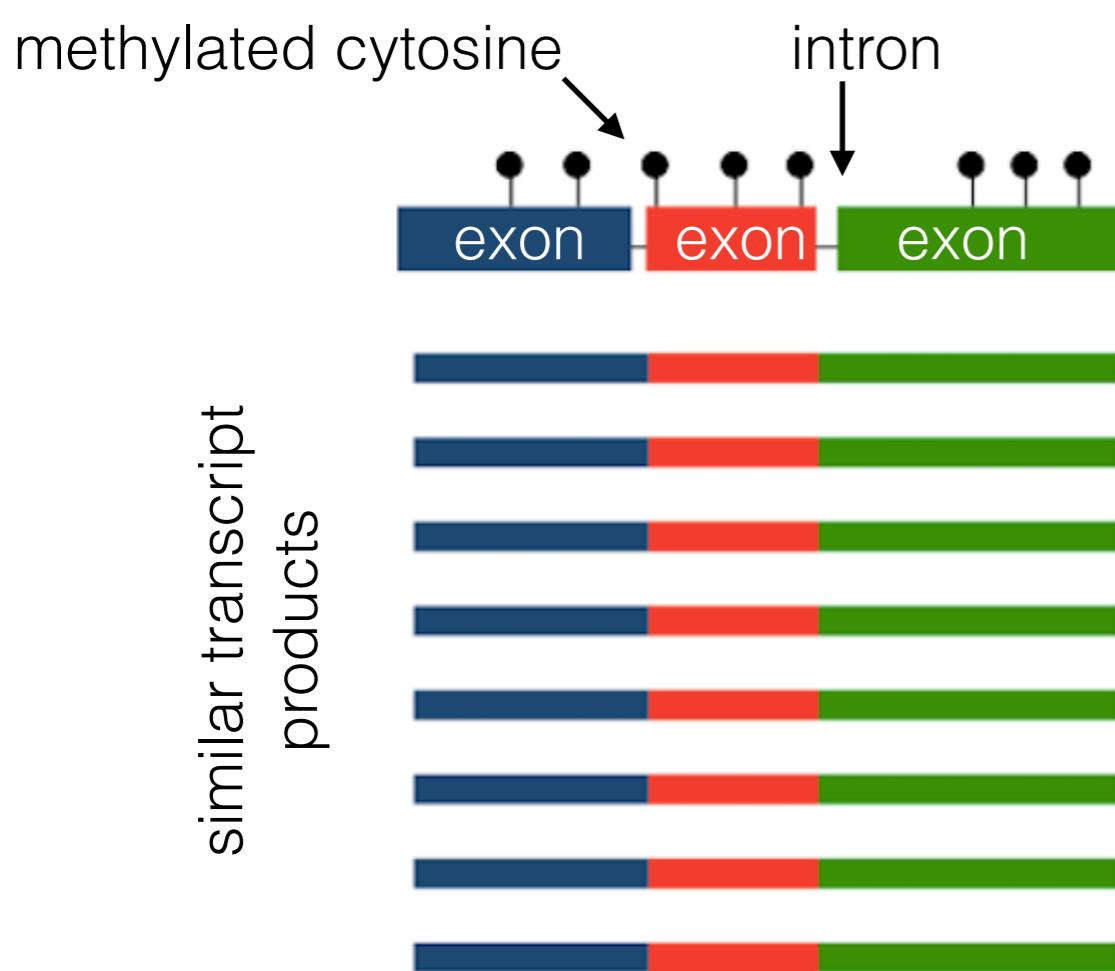
Advance Access publication date 7 January 2014

Stochastic Variation

A context dependent role for DNA methylation in bivalves

Mackenzie R. Gavery and Steven B. Roberts

Advance Access publication date 7 January 2014



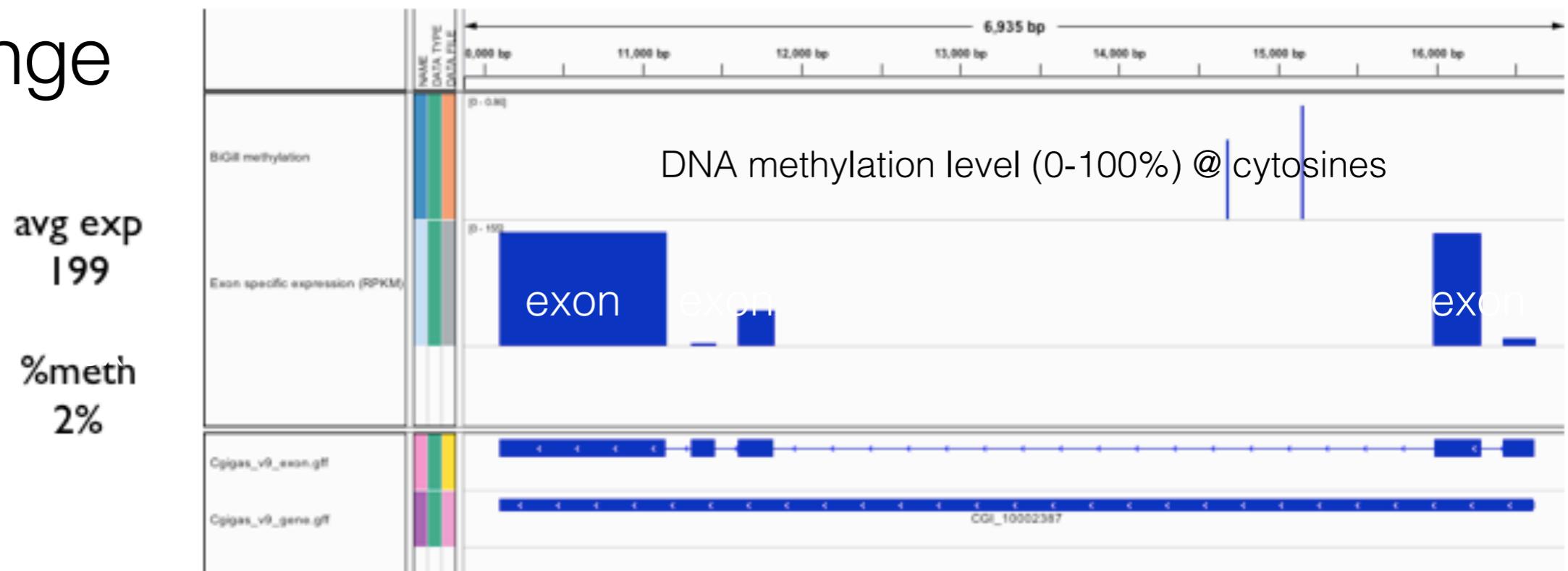
housekeeping

similar transcript
products

response to
change

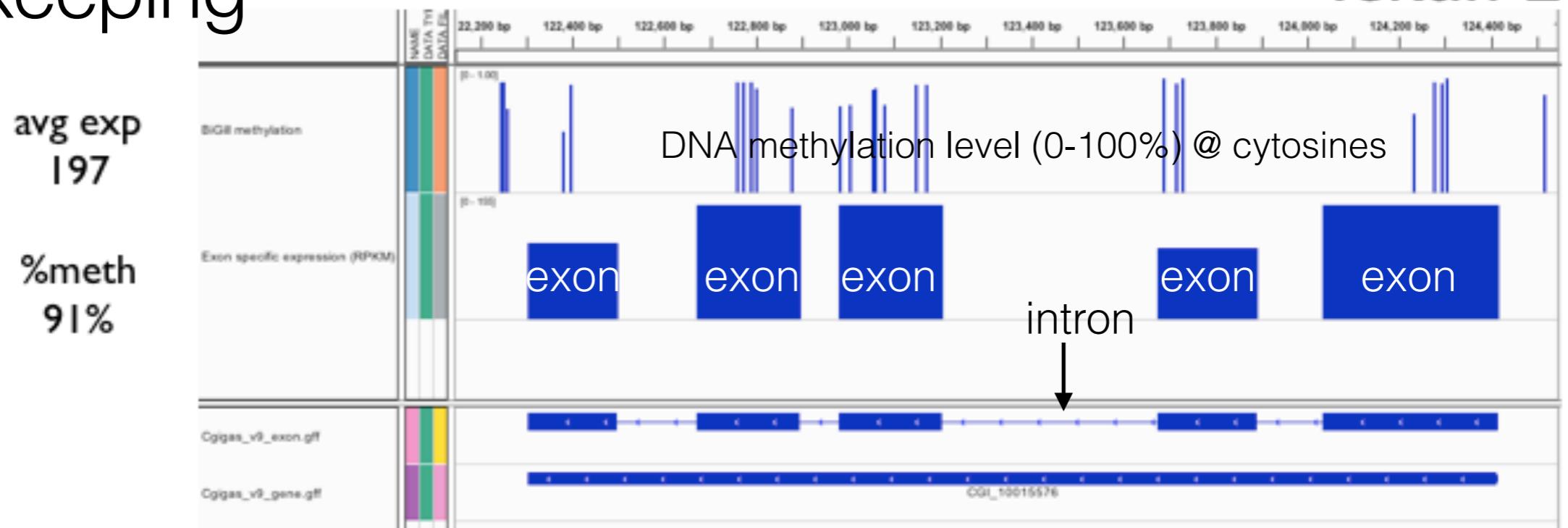
response to
change

Heat shock 70 kDa protein 12A



housekeeping

Tektin-2



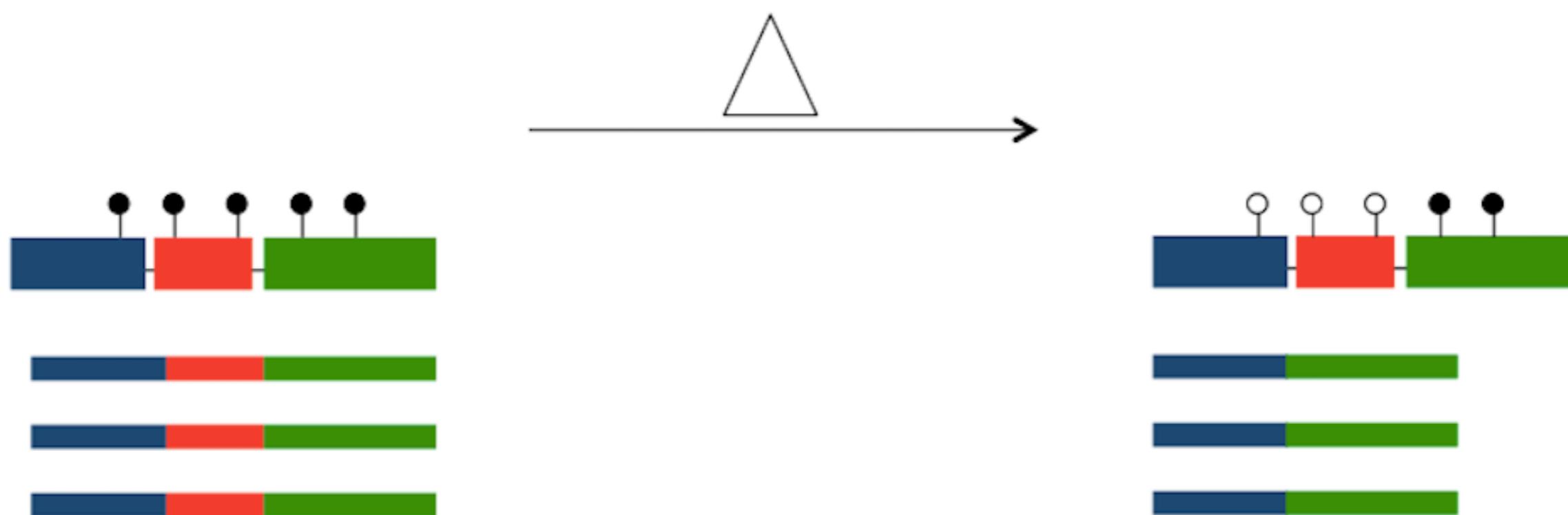


Targeted Regulation

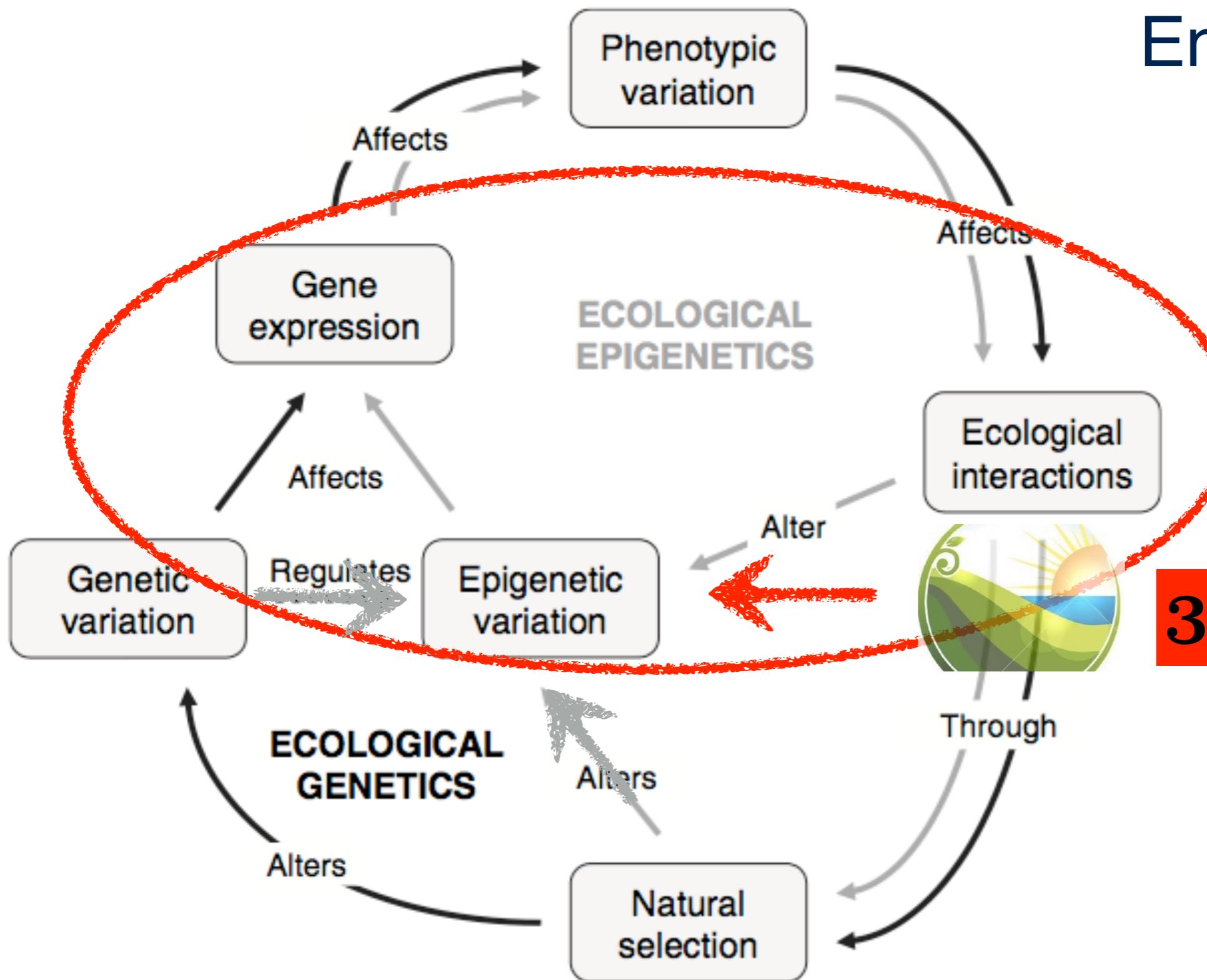
A context dependent role for DNA methylation in bivalves

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

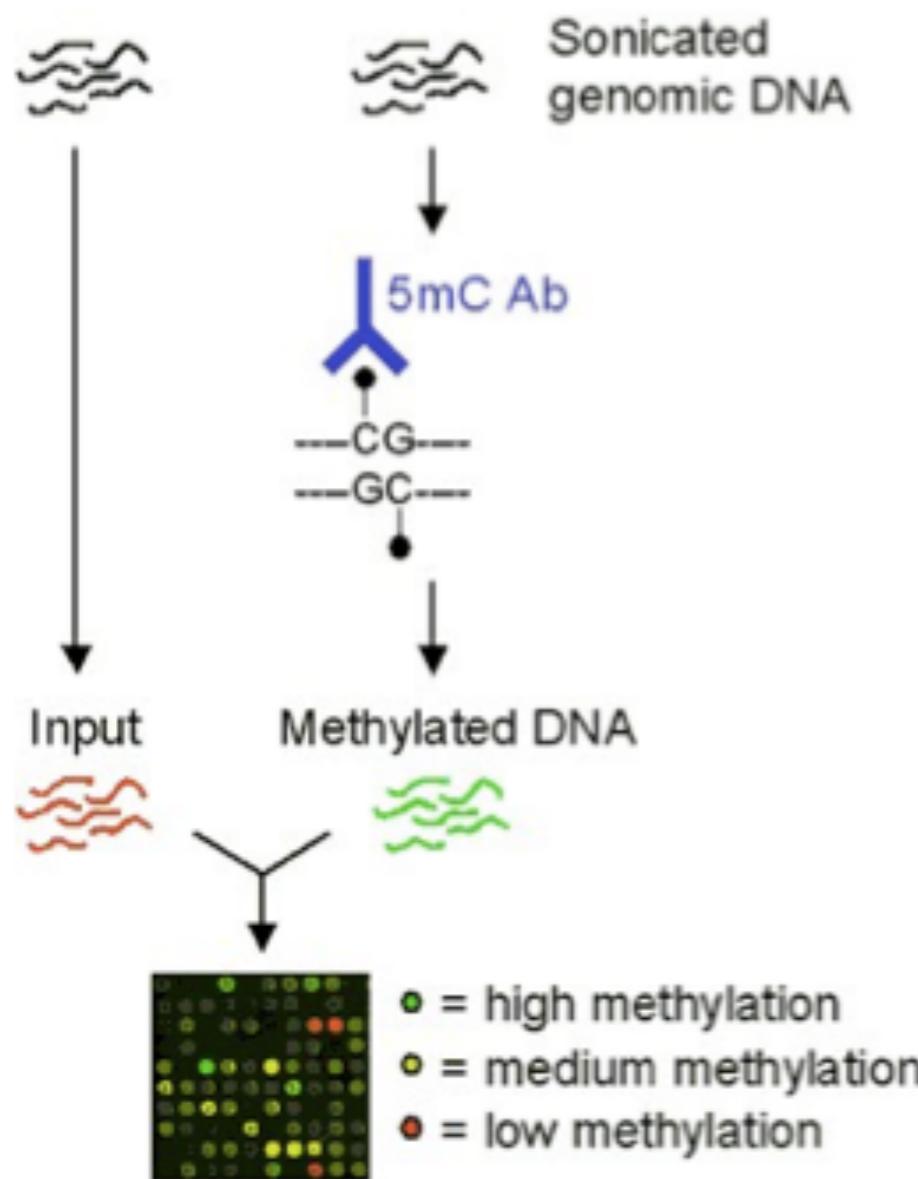
IDEA AND
PERSPECTIVE

Epigenetics for ecologists

Oliver Bossdorf,^{1,*} Christina L.
Richards² and Massimo Pigliucci³

Very new data

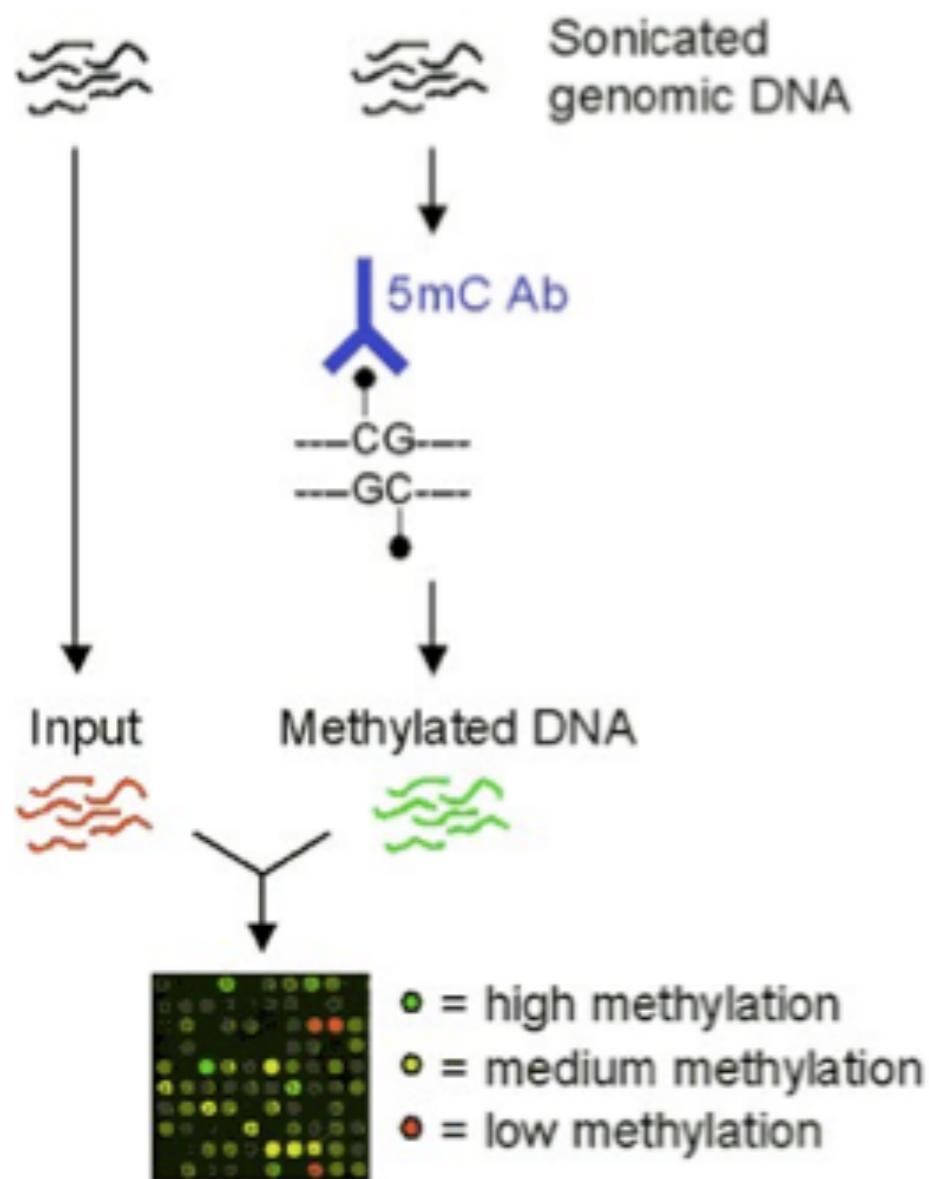
Environmental impact (Estrogens)



- 96 regions that were hypermethylated in EE2
 - 90 are in genes
 - 52 of these cross exon/intron boundary another
 - 32 are in introns
 - 6 are just in exons
- 287 regions that were hypomethylated in EE2
 - 256 are in genes
 - 138 cross exon/intron boundary
 - 114 are in introns only and
 - 4 are just in exons

Very new data

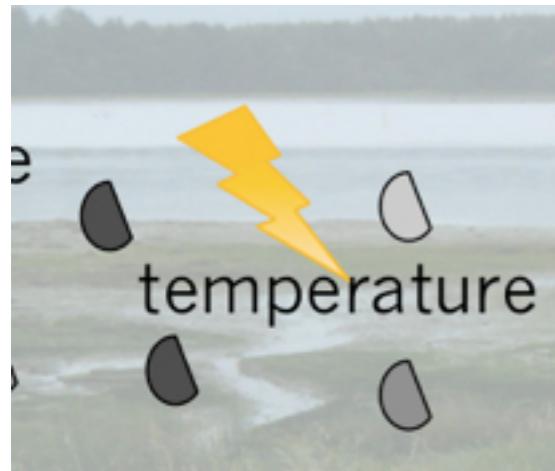
Environmental impact (Estrogens)



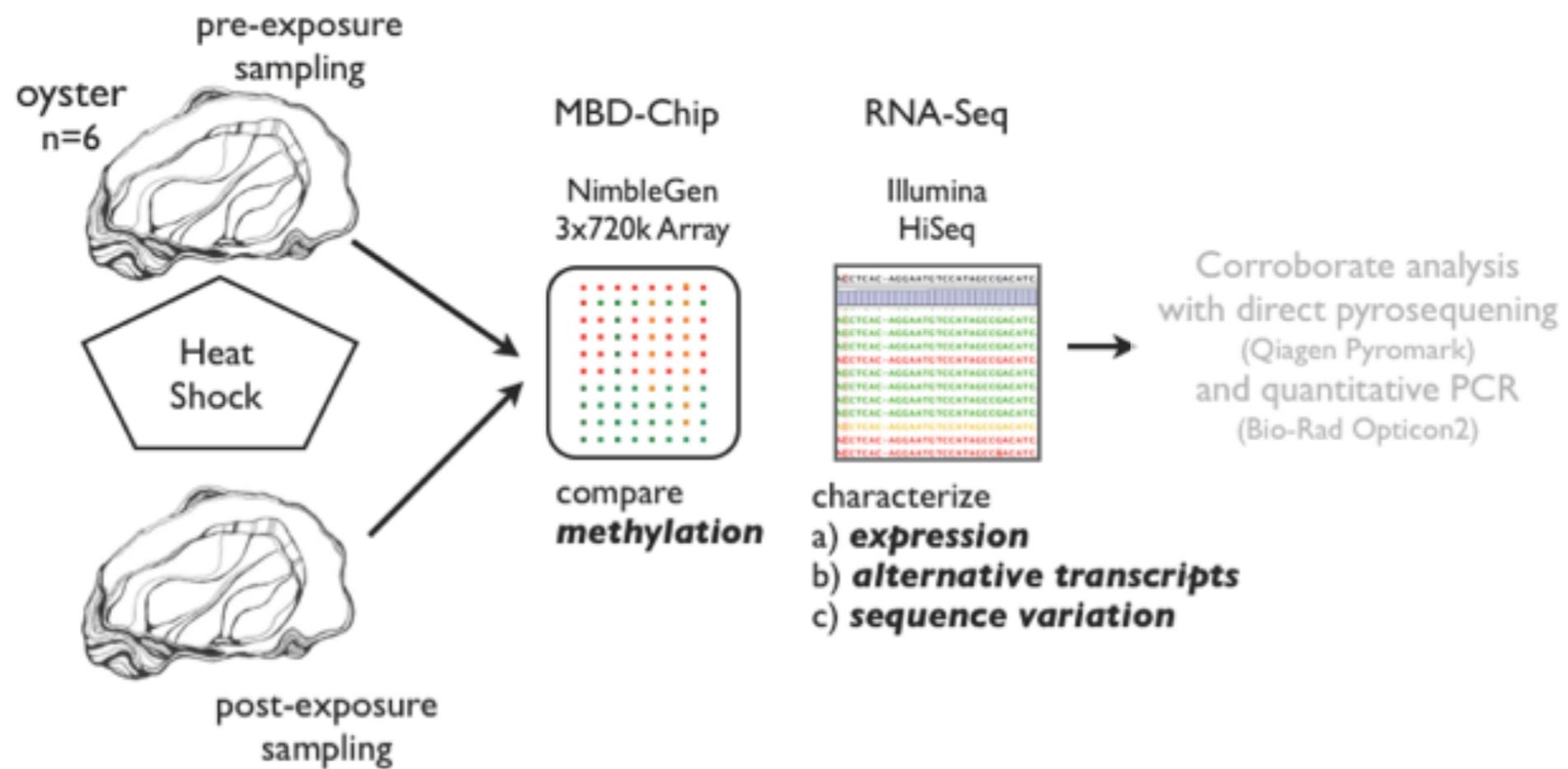
Exposure results in significant hypomethylation

Very new data

Environment and gene expression

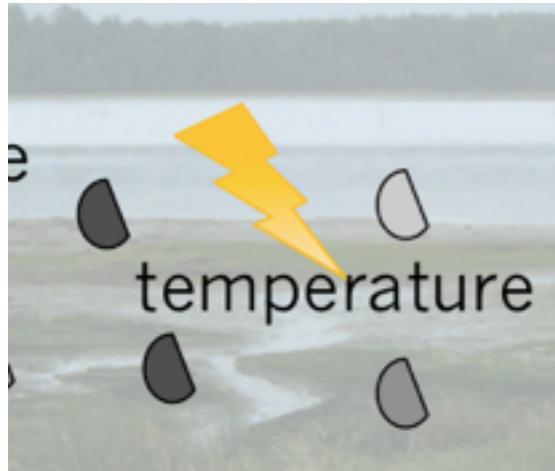


stochastic or targeted?



Very new data

Environment and gene expression

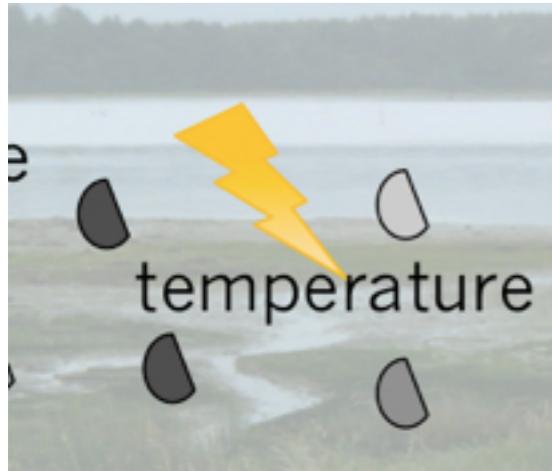


stochastic or targeted?

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

Very new data

Environment and gene expression



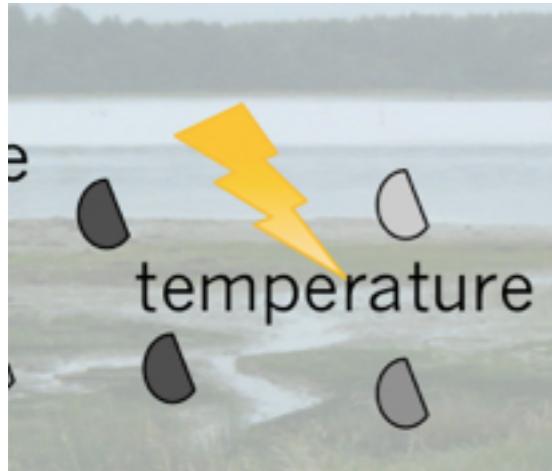
stochastic or targeted?

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

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

Very new data

Environment and gene expression

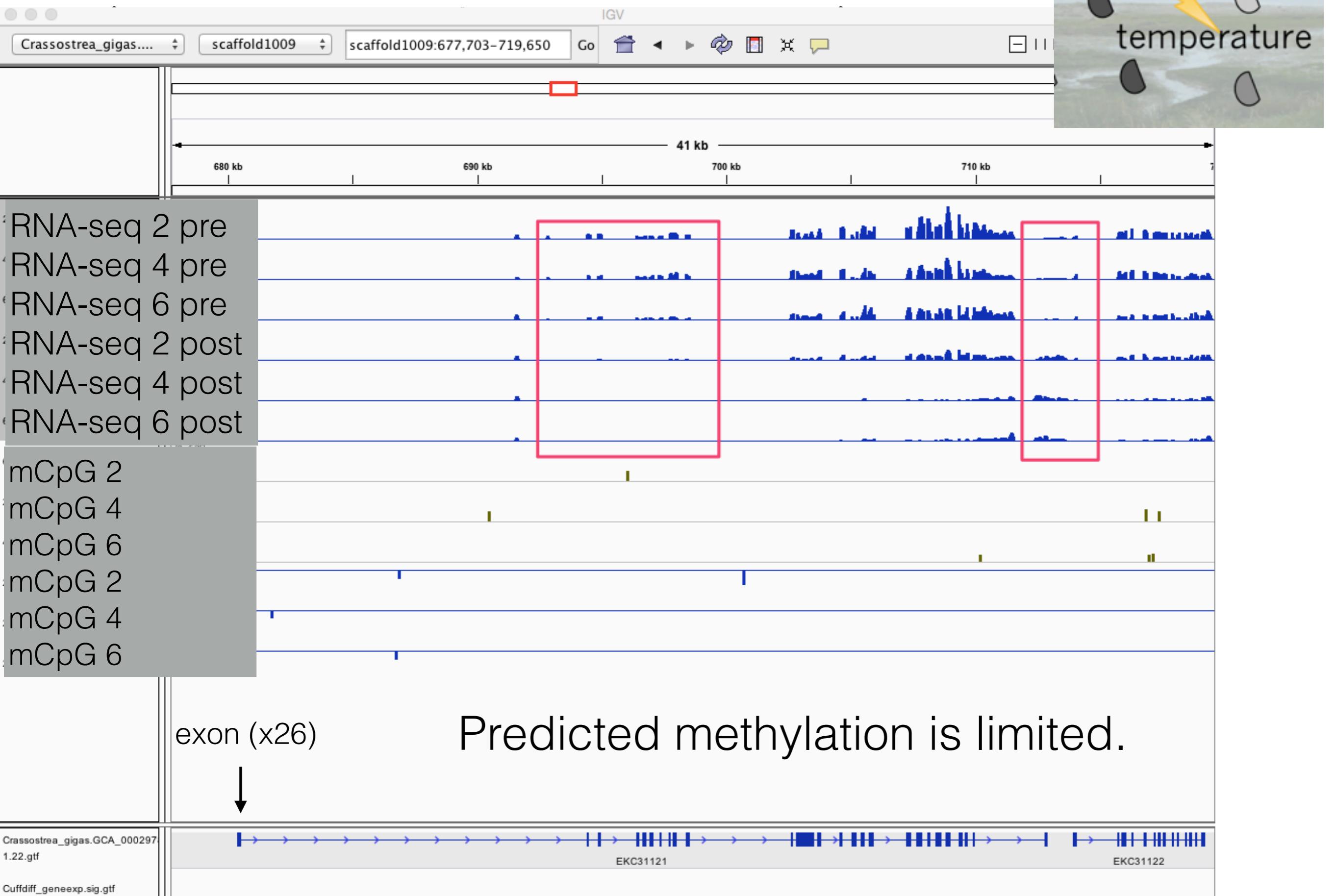


*stochastic or targeted
or ..?*

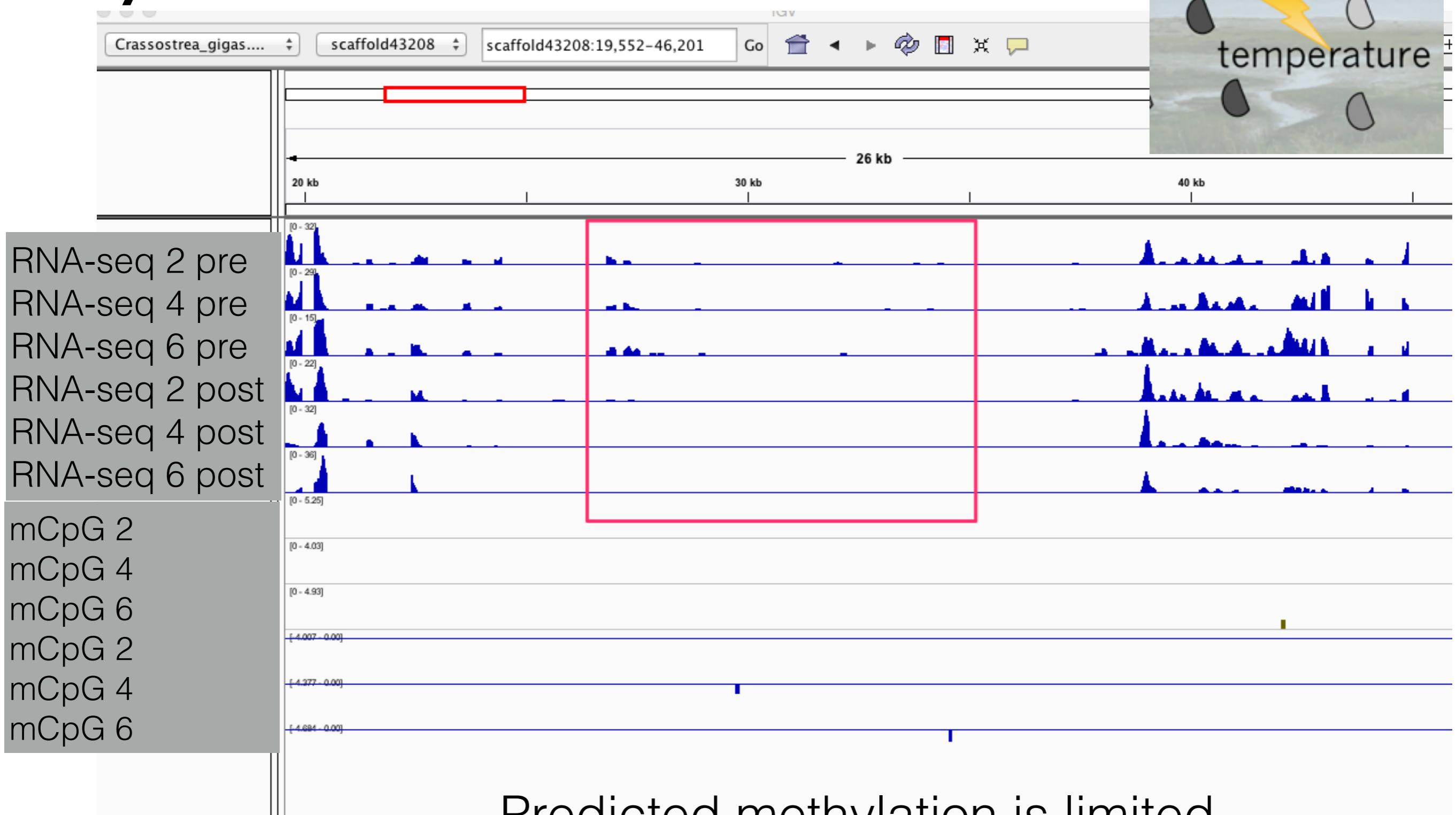
Oyster	Hypo-methylated	Hyper-methylated
2	7224	2803
4	6560	3587
6	7645	4044

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

Very new data



Very new data

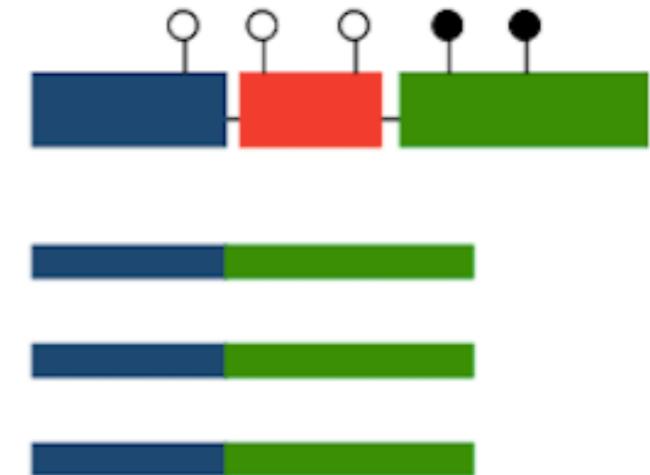
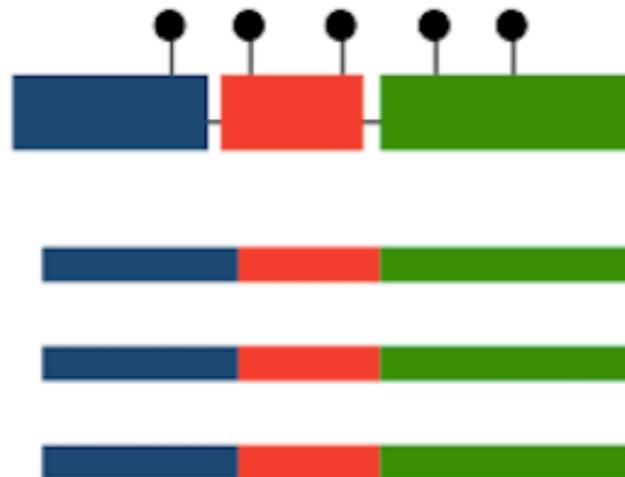


Predicted methylation is limited.

Gene
expression

2

Epigenetic
variation



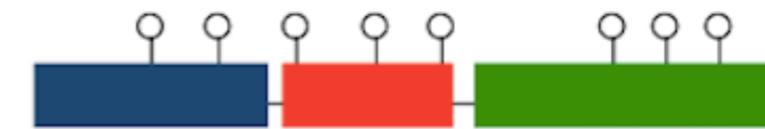
not in this experiment

not even consistent methylation changes at loci level

Gene
expression

2

Epigenetic
variation



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

Take Home

Next Steps

Dig into the
heat-shock data

Transposable
Elements

Consider other
epigenetic process

Next Steps

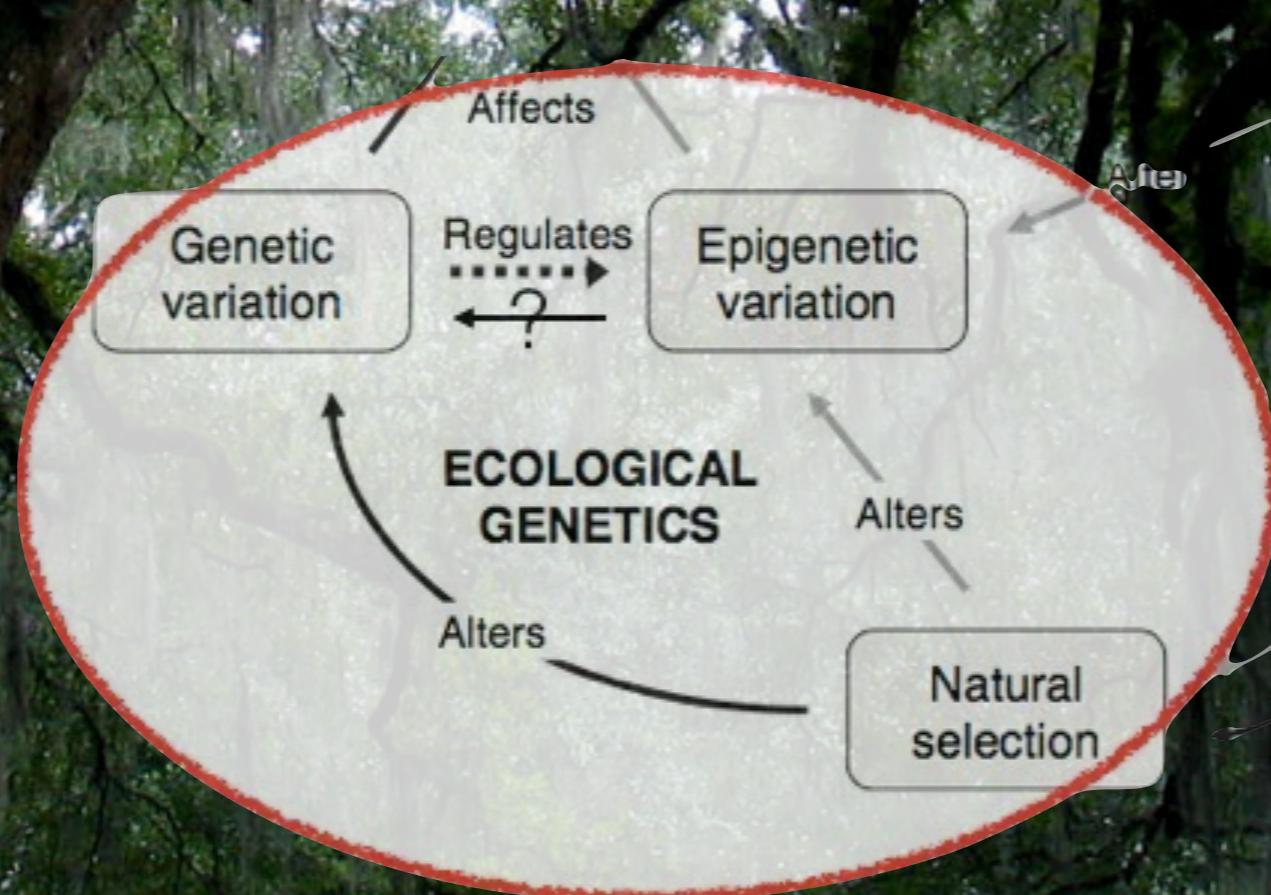


Photo credit: Flickr, Creative Commons, csessums

~~Very new data~~
Heritability
Plasticity
Local Adaptation

*Genetics
versus
Epigenetics*

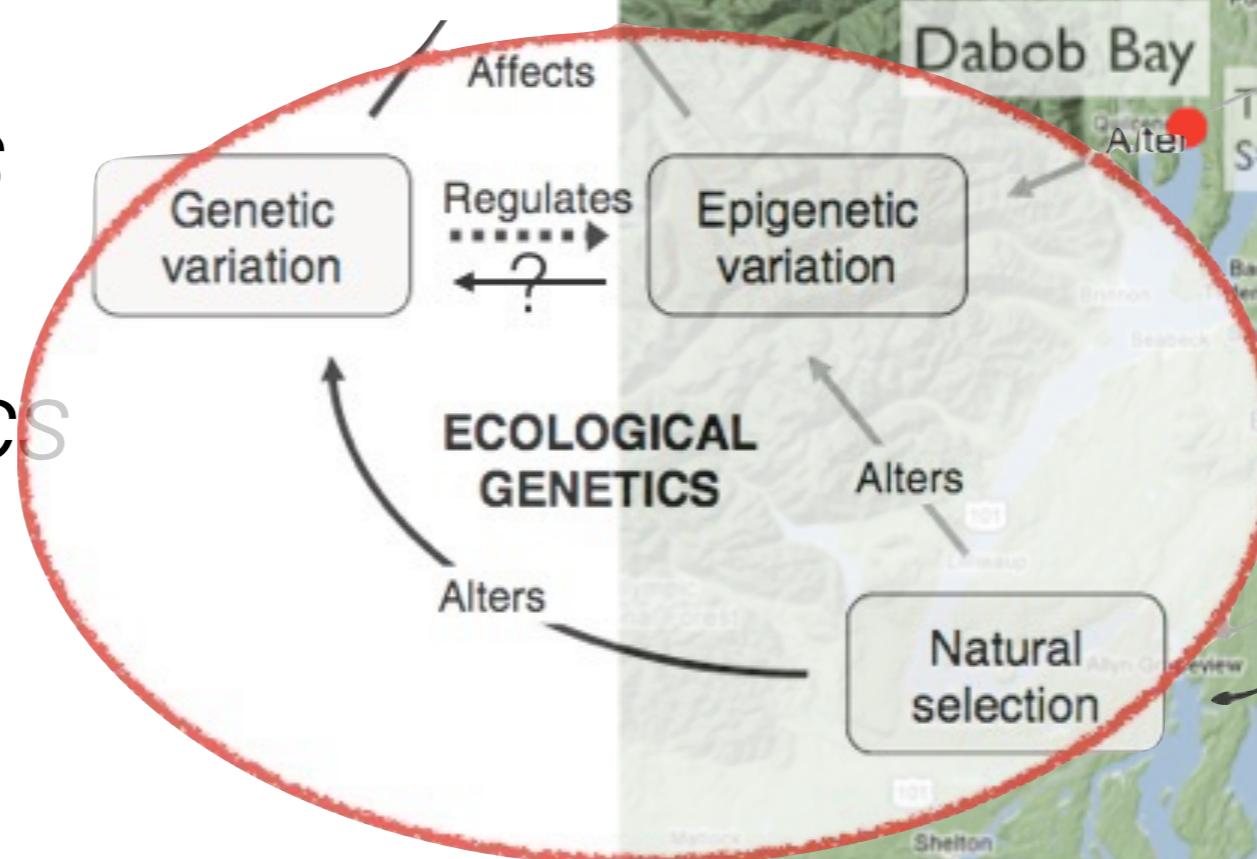


Common Garden Experiment

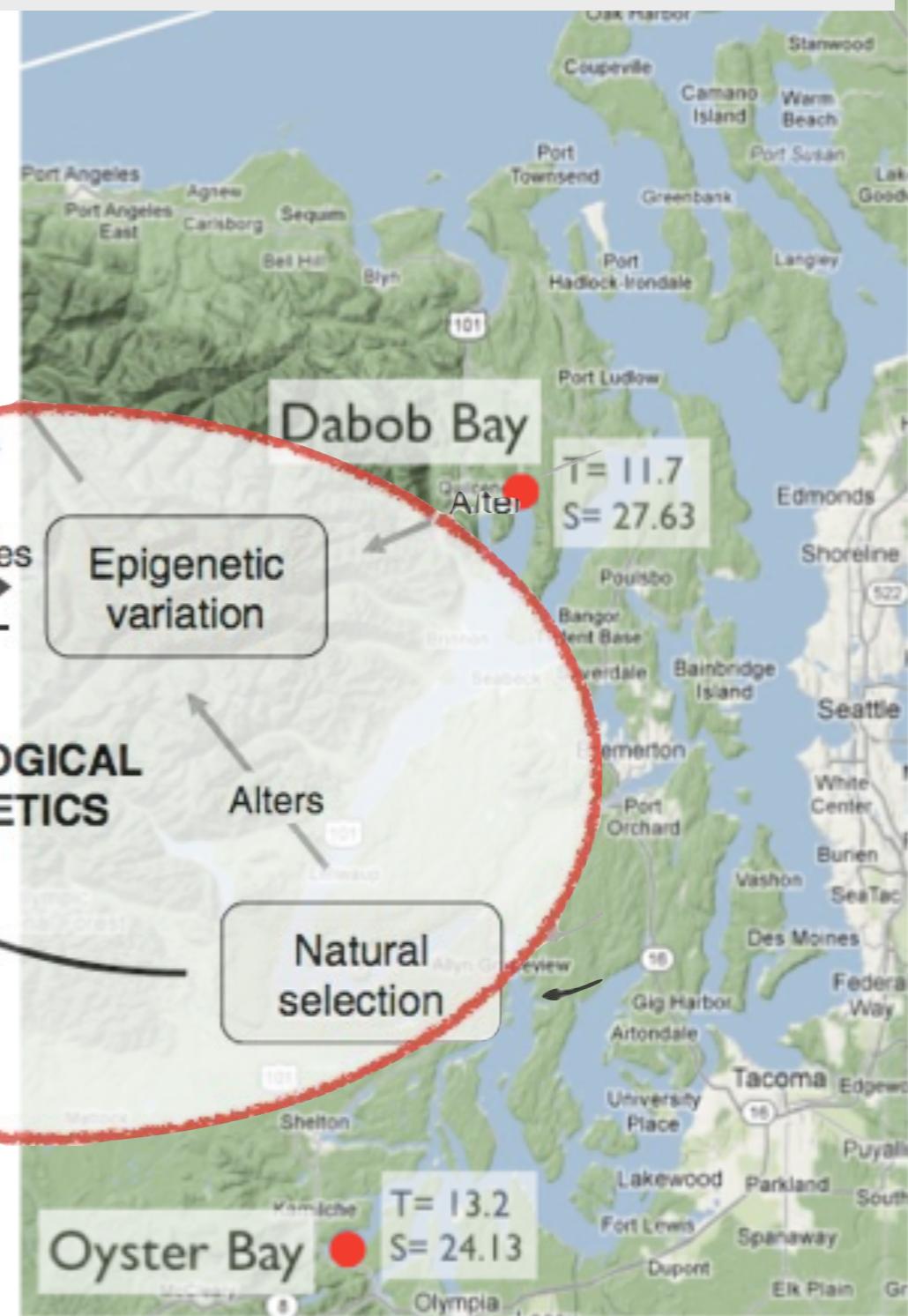


~~Very new data~~
Heritability
Plasticity
Local Adaptation

Genetics
versus
Epigenetics



Common Garden Experiment



Open Science

about science

web-native scholarship

Sharing



Photo credit: Flickr, Creative Commons, speechless

Example



Public Sigenae Contig Browser Oyster

Search *e!* Oyster: Anything

e.g. [BQ426804.p.cg.8](#), [snp_EW777925_683](#)

Based on Ensembl release 40 - Aug 2006

Use Ensembl to...

- Search Contig Browser
- Data mining [BioMart]
- Export data
- Download data
- Digital Differential Display
- Venn Diagrams

Docs and downloads

- About Contig Browser
- Contig Browser data

Other links

- Home
- Sigenae
- Agenae



Browse the *Crassostrea gigas* contigs database

About the *Crassostrea gigas* contigs database



Assembly

The Oyster EST contig browser aims to produce and maintain an automatic annotation of Oyster EST libraries. This database [GigasDatabase](#) was initiated within the frame of the [AquaFirst](#) European project, it now gathers EST sequences

produced by a [Marine Genomics Europe](#) project (GOCE-CT-2004-505403) and a [Genoscope project](#). [GigasDatabase](#) is regularly updated in the context of the ANR project "Gametogenes" (ANR-08-GENM-041).

Annotation

Contigs were annotated searching sequence homologies against following databases:

- ▶ UniProtKB/Swiss-Prot Release 2011_03 of 08-Mar-2011
- ▶ RefSeq Protein Index Blast of 09-Jan-2011
- ▶ Pfam Release 24.0 of Jul-2009
- ▶ Sigenae Oyster Contigs V8
- ▶ RefSeq RNA Index Blast of 09-Jan-2011
- ▶ TIGR Fugu FGI 3.0

Statistics

Assembly: pcg8, Mar 2011

Number of base pairs (Contigs): 67,875,621

Number of sequences (EST/mRNA): 1,013,570

Total number of contigs: 82,312

Number of singlets: 20,148

Polymorphism: Mar 2011

Number of putative SNPs: 29,493

Number of contigs including SNP: 10,392

Example Data Points

Jump directly to the example entry points:

Contig: [BQ426804.p.cg.8](#)

SNP: [snp_EW777925_683](#)

Example

S!  **Public Sigenae Contig Browser Oyster** Search e| Oyster: Anything Go e.g. [BQ426804.p.cg.8](#), [snp_EW777925_683](#)

Based on Ensembl release 40 - Aug 2006

Use Ensembl to... Browse the *Crassostrea gigas* contigs database

Search Contig Browser

Table 1. Summary of assembly and RNA-Seq statistics for de novo assembly and reference mapping (GigasDatabase v8).

	De novo assembly	Reference mapping	
Assembly	Mapped reads	8,407,963	29,107,760
	Unmapped reads	36,944,698	16,244,901
	Contigs	18,510	77,433
	Average contig length	276	554
	Average contig coverage	62	16
	Contigs annotated to GO Slim	3931	7296
RNA-Seq	Differentially expressed genes	2991	427
	Enriched GO biological process	15	3

About the *Crassostrea gigas* contigs

Statistics

Assembly: pcg8, Mar 2011

Number of base pairs (Contigs): 67,875,621

Number of sequences (EST/mRNA): 1,013,570

Total number of contigs: 82,312

Number of singlets: 20,148

Polymorphism: Mar 2011

Number of putative SNPs: 29,493

Number of contigs including SNP: 10,392

Contigs were annotated searching sequence homologies against following databases:

Characterizing short read sequencing for gene discovery and RNA-Seq analysis in *Crassostrea gigas*

Mackenzie R. Gavery, Steven B. Roberts *

INRA

▶ Sigenae Oyster Contigs V8
▶ RefSeq RNA Index Blast of 09-Jan-2011
▶ TIGR Fugu FGI 3.0

Example Data Points

Jump directly to the example entry points:

Contig: [BQ426804.p.cg.8](#)

SNP: [snp_EW777925_683](#)

Example



Based on Ensembl release 40 - Aug 2006

Use Ensembl to...

Search Contig Browser

Browse the Crassostrea gigas genome

About the Crassostrea gigas genome

Table 1. Summary of assembly and RNA-Seq statistics for de novo assembly (v8).

		De novo assembly	Referer
Assembly	Mapped reads	8,407,963	29,107,400
	Unmapped reads	36,944,698	16,244,000
	Contigs	18,510	77,433
	Average contig length	276	554
	Average contig coverage	62	16
	Contigs annotated to GO Slim	3931	7296
RNA-Seq	Differentially expressed genes	2991	427
	Enriched GO biological process	15	3

HOME

Sigenae

Agenae

Contigs were annotated search against following databases:

- UniProtKB/Swiss-Prot Reference Proteins 2011
- RefSeq Protein Index Blast
- Pfam Release 24.0 of June 2011
- Sigenae Oyster Contigs 1.0
- RefSeq RNA Index Blast
- TIGR Fugu FGI 3.0

Research article

Highly accessed

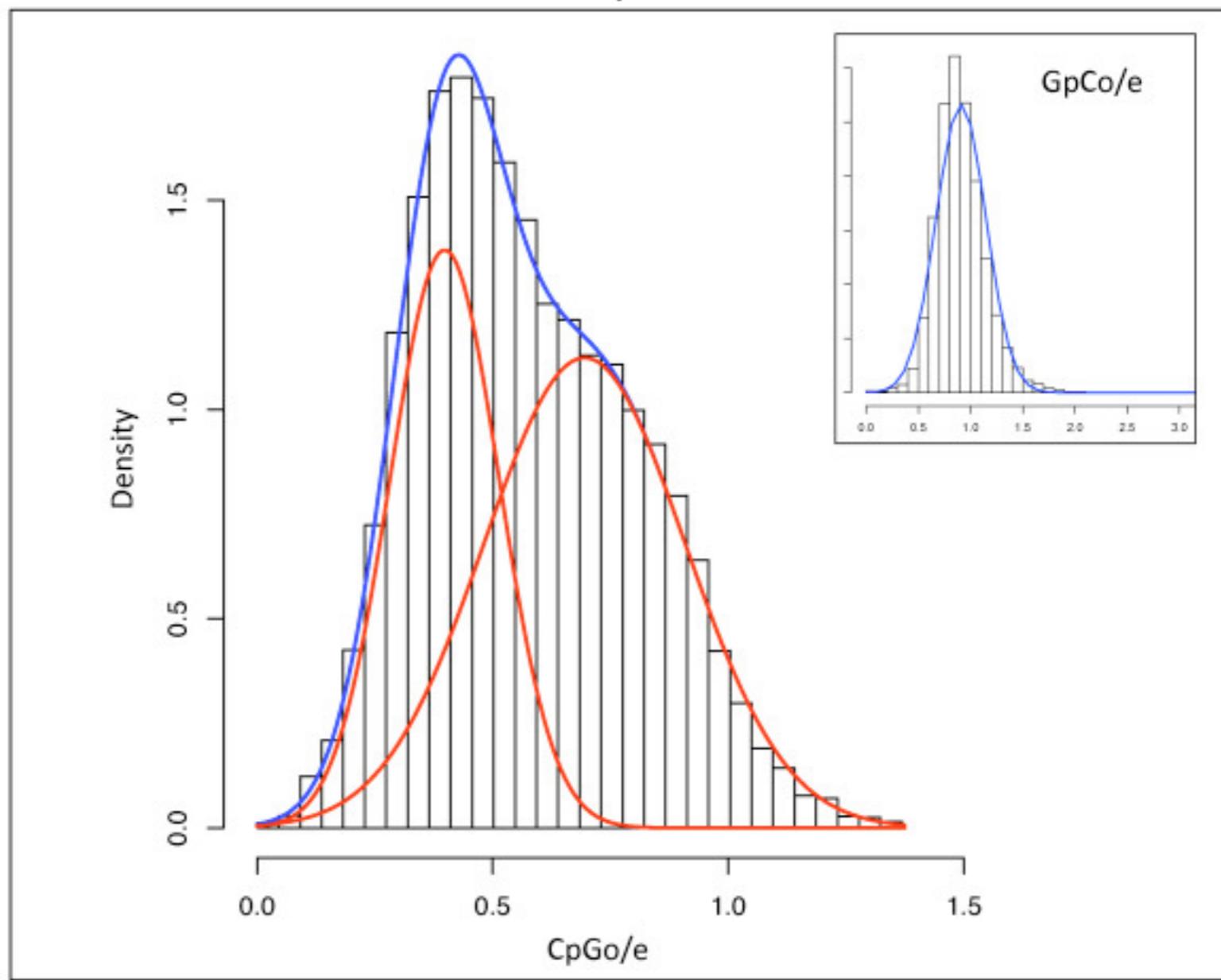
Open Access

DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*)

Mackenzie R Gavery and Steven B Roberts*

DOI:

<http://dx.doi.org/10.1186/1471-2105-12-500>



Idea Generation

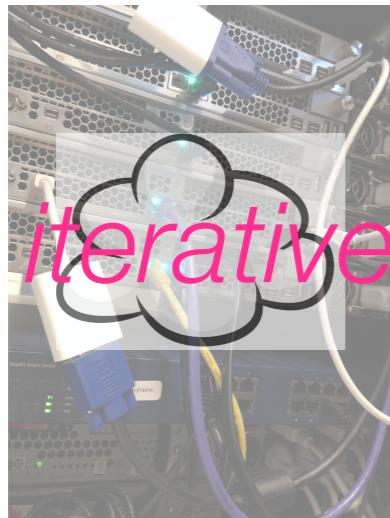
Publication

Data Acquisition and Analysis

Publication

Publication

Publication

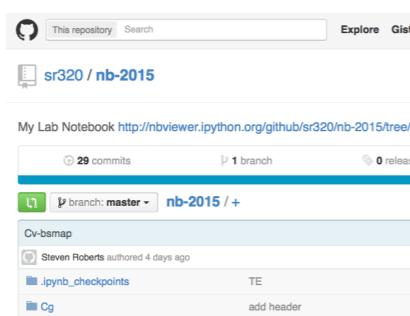


lab notebooks



wikis for everyone

IP[y]: IPython
Interactive Computing



Open Notebook Science

Open Notebook Science

... there is a URL to a laboratory notebook that is freely available and indexed on common search engines. It does not necessarily have to look like a paper notebook but it is essential that all of the information available to the researchers to make their conclusions is equally available to the rest of the world.

— Jean-Claude Bradley

Idea Generation Data Acquisition and Analysis

Publication
Publication
Publication
Publication



lab notebooks



wikis for everyone

IP[y]: IPython
Interactive Computing

"When you're working in the open, only 9 in the ENTIRE world needs to solve a given problem"

TOREV

Sean-Claude



OpenNotebookScience

This repository Search Explore Gist

sr320 / nb-2015

My Lab Notebook <http://nbviewer.ipython.org/github/sr320/nb-2015/tree/r>

29 commits 1 branch 0 releases

branch: master nb-2015 / +

Cv-bsmap

Steven Roberts authored 4 days ago

.ipynb_checkpoints TE

.Cg add header

Automating a Workflow: Beyond Blast - to GO Slim

The concept is that you can take a fasta file in a working directory and end up with GO slim information all within a single notebook that is automated. Current writing (and overwriting) as scratch file to SQLShare. Assumptions are that you are working in a directory with fasta file named query.fa. And blast algorithms are available.

```
In [13]: #allows plots to be shown inline
%pylab inline

Populating the interactive namespace from numpy and matplotlib

In [4]: #Setting Working Directory
wd="/Volumes/web/whale/fish546/qpx_go_val"
#Setting directory of Blast Databases
dbd="/Volumes/Bay3/Software/ncbi-blast-2.2.29+/db/"
#Database name
dbn="uniprot_sprot_r2013_12"
#Blast algorithm
ba="blastx"
#Location of SQLShare python tools: you can empty ("") if tools are in PATH
spd="/Users/sr320/sqlshare-pythonclient/tools/"

In [5]: cd {wd}
/Volumes/web/whale/fish546/qpx_go_val

In [5]: !{ba} -query query.fa -db {dbd}{dbn} -out {dbn}_{ba}_out.tab -evalue 1E-50 -num_threads 4 -max_hsps_per_
BLAST Database error: No alias or index file found for protein database [/Volumes/Bay3/Software/ncbi-bla
/db/uniprot_sprot_r2013_12] in search path [/Volumes/web/whale/fish546/pipeline_test_dir4::]

In [6]: !head -1 {dbn}_{ba}_out.tab
QPX_transcriptome_v1_Contig_2 sp|P52712|CBPX_ORYSJ 43.75 416 213 12 2095 869
7 3e-98 326

In [7]: !tail -1 {dbn}_{ba}_out.tab
QPX_transcriptome_v1_Contig_2 sp|P52712|CBPX_ORYSJ 43.75 416 213 12 2095 869
7 3e-98 326

In [8]: #Translate pipes to tab so SPID is in separate column for Joining
!tr ' ' '\t' <{dbn}_{ba}_out.tab> {dbn}_{ba}_out2.tab

In [9]: !head -1 {dbn}_{ba}_out2.tab
QPX_transcriptome_v1_Contig_2 sp|P52712|CBPX_ORYSJ 43.75 416 213 12 2095 869
7 3e-98 326

In [10]: !tail -1 {dbn}_{ba}_out2.tab
QPX_transcriptome_v1_Contig_2 sp|P52712|CBPX_ORYSJ 43.75 416 213 12 2095 869
7 3e-98 326

In [11]: !python {spd}singleupload.py -d scratchblast_out {dbn}_{ba}_out2.tab
QPX_transcriptome_v1_Contig_2 sp|P52712|CBPX_ORYSJ 43.75 416 213 12 2095 869
7 3e-98 326

In [12]: #Sets GO aspect
!python {spd}fetchdata.py -s "SELECT * FROM [sr320@washington.edu].[scratchblast_out]blast Left Join [sr320@washington.edu].[scratchblast_out]go Slim Right Join [sr320@washington.edu].[scratchblast_out]go Slim AS go ON go.go_id = blast.go_id WHERE blast.spid = 'P52712' AND go.goslim_id = 'GO:0003674'"

In [13]: !head justslim.txt
QPX_transcriptome_v1_Contig_2 sp|P52712|CBPX_ORYSJ 43.75 416 213 12 2095 869
7 3e-98 326

In [14]: !tail -1 justslim.txt
QPX_transcriptome_v1_Contig_2 sp|P52712|CBPX_ORYSJ 43.75 416 213 12 2095 869
7 3e-98 326

In [15]: from pandas import *
```

Idea Generation

Publication

Data Acquisition and Analysis

Publication

Publication

Publication



lab notebooks



IP[y]: IPython
Interactive Computing



This repository Search Explore Gist

sr320 / nb-2015

My Lab Notebook <http://nbviewer.ipython.org/github/sr320/nb-2015/tree/r>

29 commits 1 branch 0 releases

branch: master nb-2015 / +

Cv-bsmap

Steven Roberts authored 4 days ago

.ipynb_checkpoints TE

Cg add header

Search or type a command Explore Gist Blog Help



sr320 Contributions Repositories Public activity

Popular repositories

ipython_nb	1 ★
LabDocs	1 ★
qdod	1 ★
dgo	0 ★
fish310	0 ★

Joined on May 15, 2013

GitHub

[nbviewer](#) [FAQ](#) [IPython](#)

[ipython_nb](#) /

Name

- [◀ sr320's repositories](#)
- [📁 .ipynb_checkpoints](#)
- [📁 examples](#)
- [📁 fish546](#)
- [📁 img](#)
- [📁 tools](#)
- [📄 BSMAP2MK_workflow.ipynb](#)
- [📄 BSMAP2view_larvae.ipynb](#)
- [📄 BSMAP2view_larvae_c.ipynb](#)
- [📄 BiGill_CpG_Eensembl.ipynb](#)
- [📄 BiGill_Gene_Methylation.ipynb](#)

Open Notebook Science

Hello,
I was doing a search for what concentrations of MgSO₄ to use as a bivalve relaxant and came across Katie's Notebook. What a cool idea!
Please say thanks to Katie :). My boss said MgSO₄ but most of the literature was MgCl₂.... So g'day from the University of Queensland in Brisbane, Australia.

Cheers

KA

email June 22, 2013

Sharing

Reproducibility

usable data formats

Platforms

FigShare



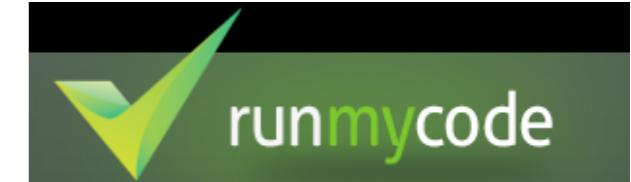
Dryad



GitHub



runmycode



Sharing

Reproducibility

usable data formats

looking for patterns

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.	http://goo.gl/oqCxl	https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Cgigas_gene_fasta
qDOD_Zhang_Gil_gene_RNA-seq	Gill RNA-seq data (gene based)	http://goo.gl/8oISR	https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Zhang_Gil_gene_RNA-seq
qDOD_Zhang_Mgo_gene_RNA-seq	Male Gonad RNA-seq data (gene based)	http://goo.gl/6buVz	https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Zhang_Mgo_gene_RNA-seq





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.



eScience Institute

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.

DATA

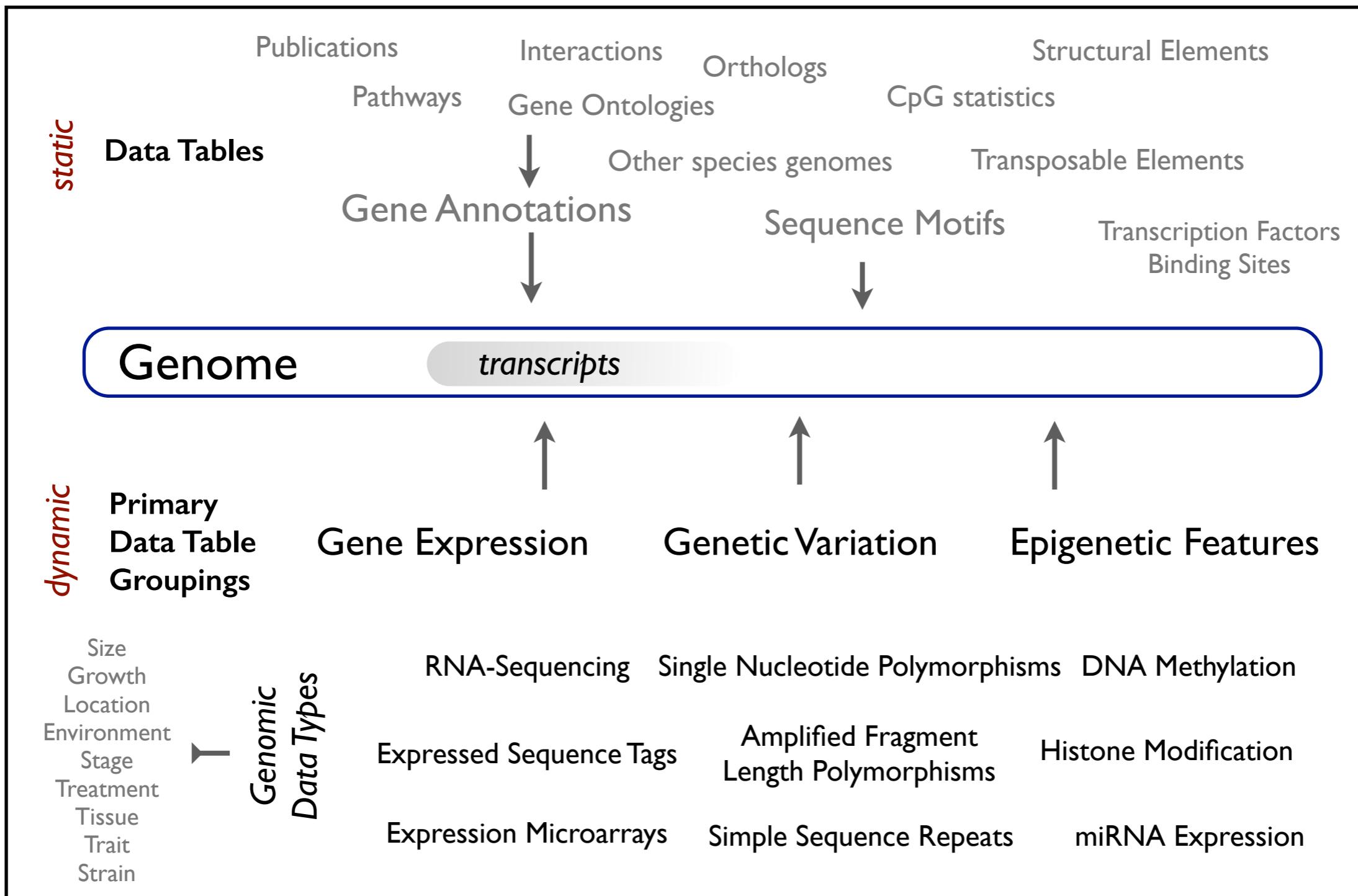
qDOD online

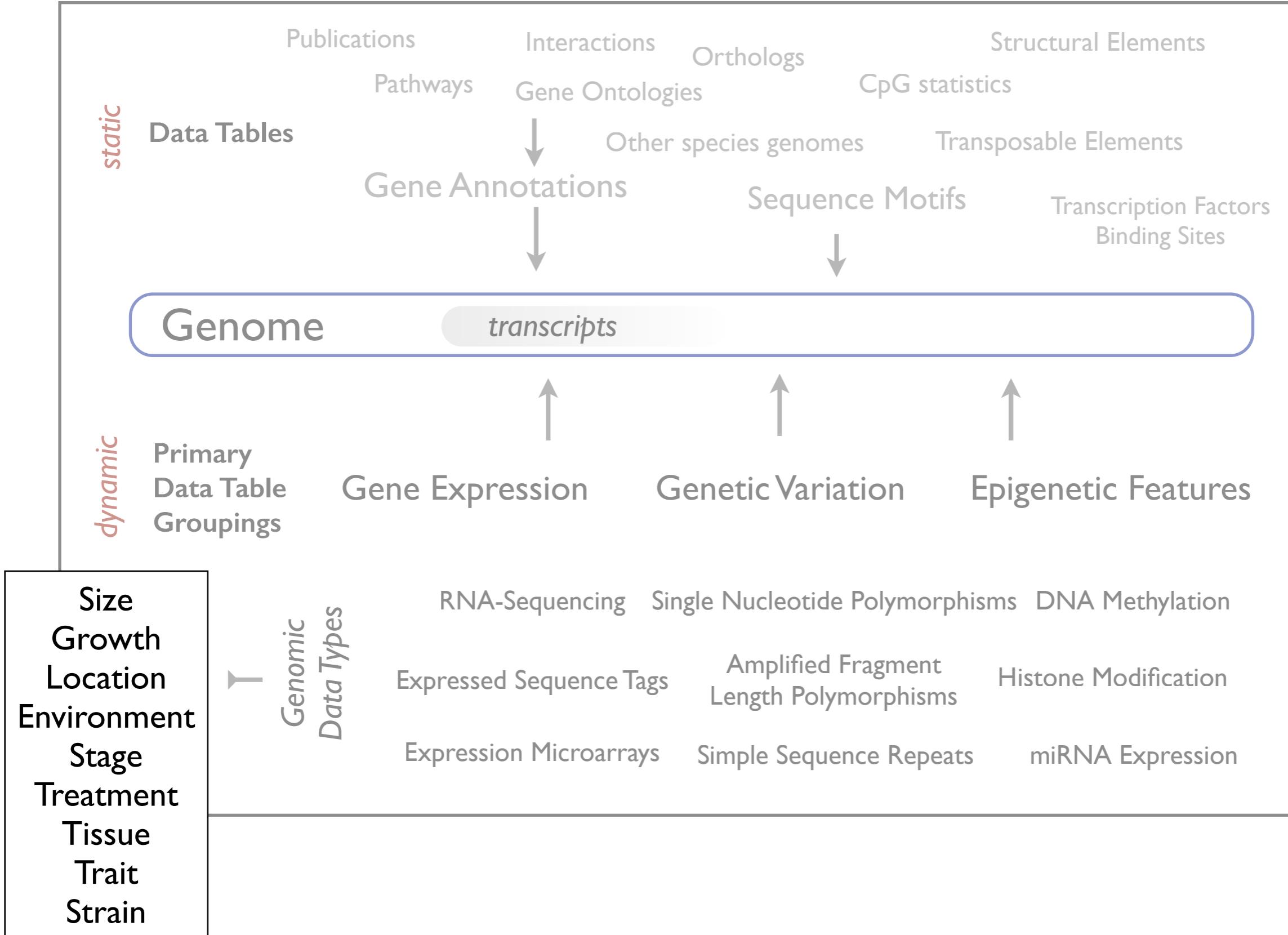
qDOD_Cgigas_gene_fasta	sequence fasta file. Exon only.	http://goo.gl/oqCxl	https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Cgigas_gene_fasta
qDOD_Zhang_Gil_gene_RNA-seq	Gill RNA-seq data (gene based)	http://goo.gl/8oISR	https://sqlshare.escapevelocity.net/datasets/qdod/qDOD_Zhang_Gil_gene_RNA-seq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

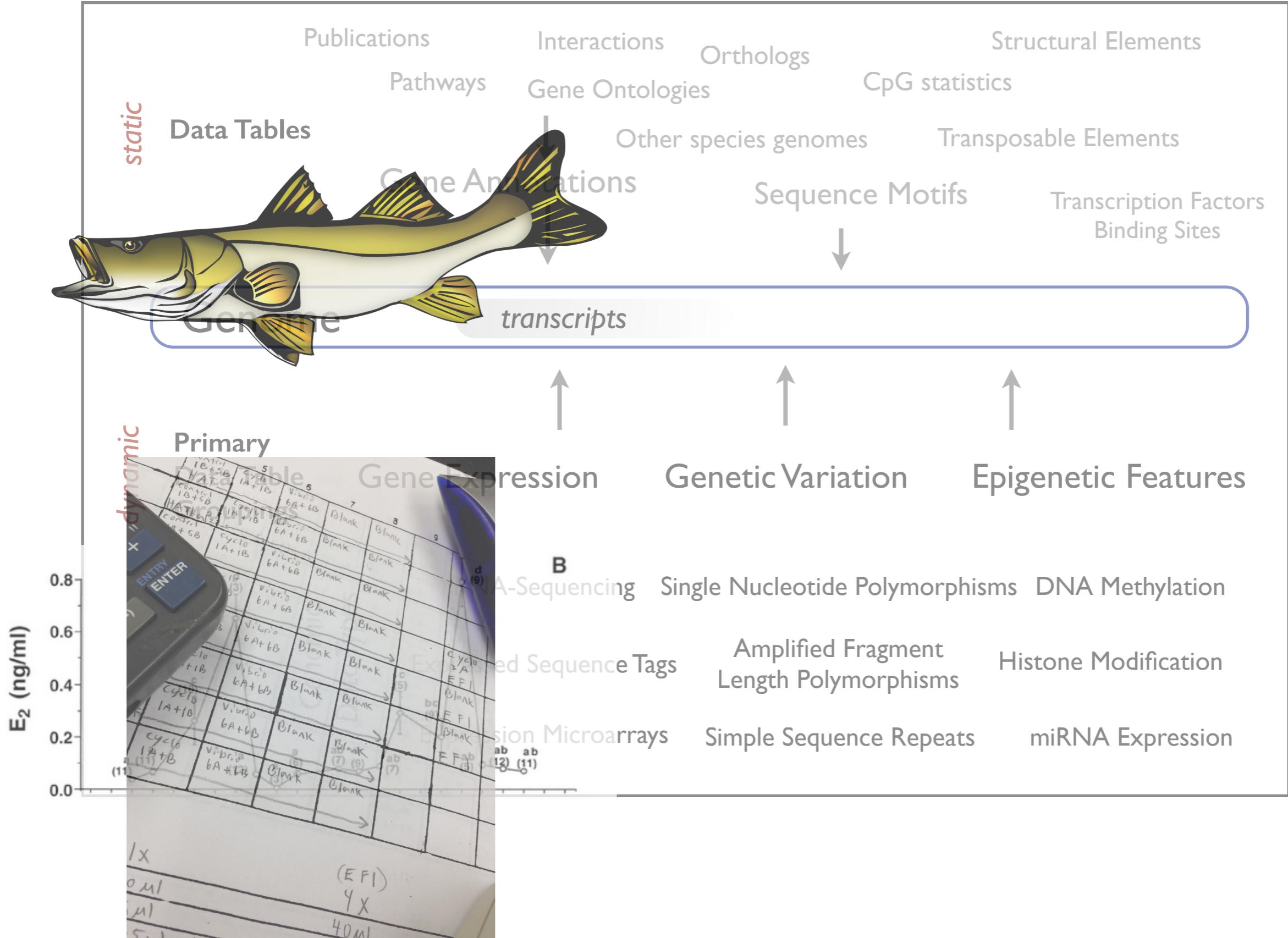
Your datasets
All datasets
Shared datasets
Recent activity... 293
Recently viewed »
Upload dataset
New query
YOUR TOP VIEWED
qDOD_scaffol... 35
oyster_v9_mR... 28
oyster_v9_CD... 23
BiGO_betty_pl... 16
qDOD_Cgigas... 16
POPULAR TAGS
protein 47
oyster 45
proteomics 21
bioinformatics 21
oa 20

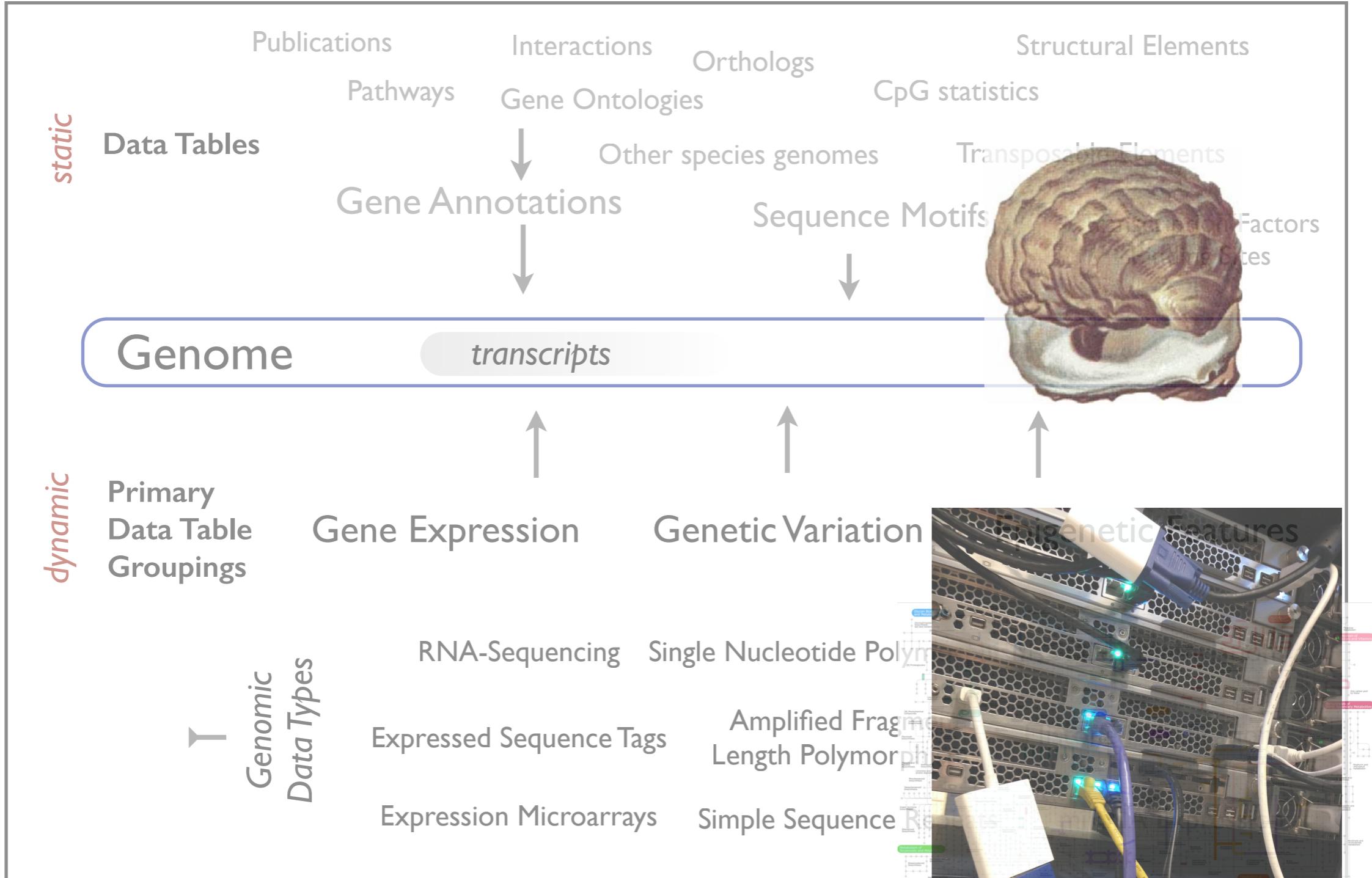
Tagged Datasets | qdod

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









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

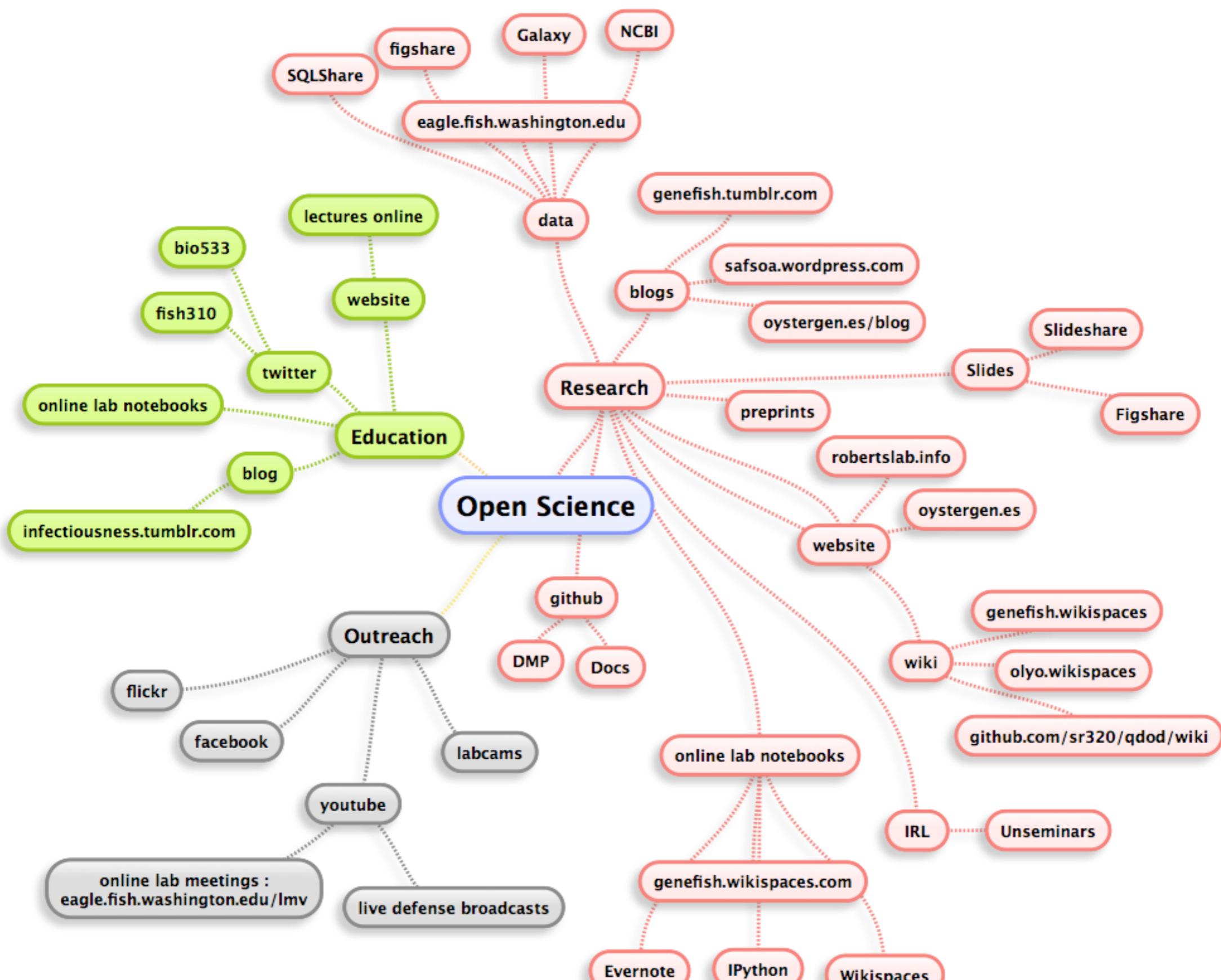
iPlant Galaxy

Notebooks

Rationale

Platforms

Open Science



everything else...

BioCode's Notes

Biology

Computational Proteomics & Bioinformatics

Home Software & Tools Programming Tips Core Manuscript Polls E

Environment

Wednesday, 19 February 2014

In the ERA of science communication, Why you need Twitter, Professional Blog and ImpactStory?

Molecular

Yasset Perez-Riverol en Wednesday, February 19, 2014

Data Analysis

eScience

iPlant Galaxy

Notebooks

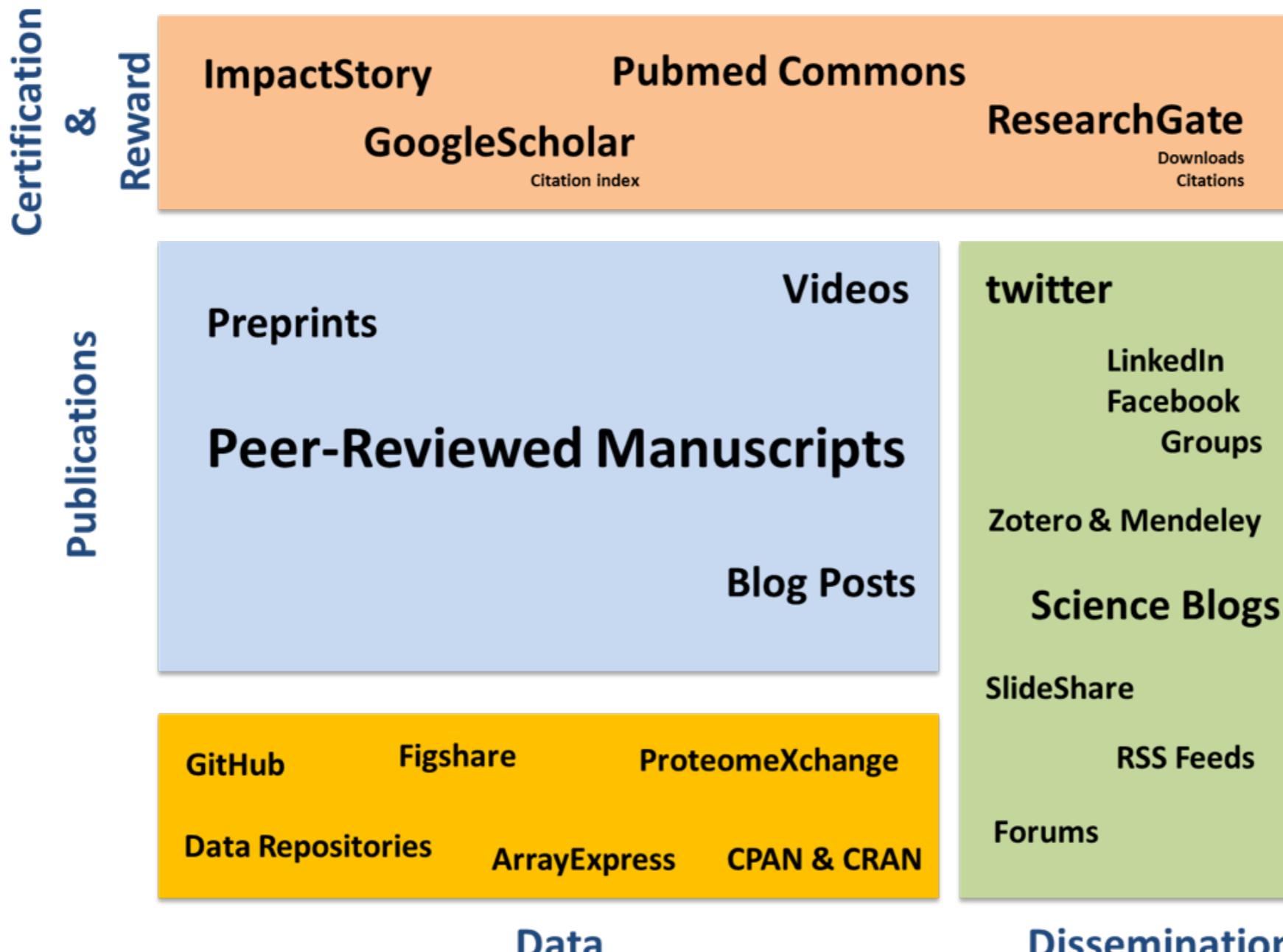
Rationale

Platforms

Open Science

Data

everything else...



Data

start them young



Acknowledgements

Mackenzie Gavery

Claire Olson

Sam White

Brent Vadopalas

Jake Heare

Jay Dimond

Bill Howe

Dan Halperin



slides, data & more @
github.com/sr320/talk-fiu-2015