

These Two Mice are Genetically Identical and the Same Age



## Nature AND Nurture

### Chromosome 3 Pairs

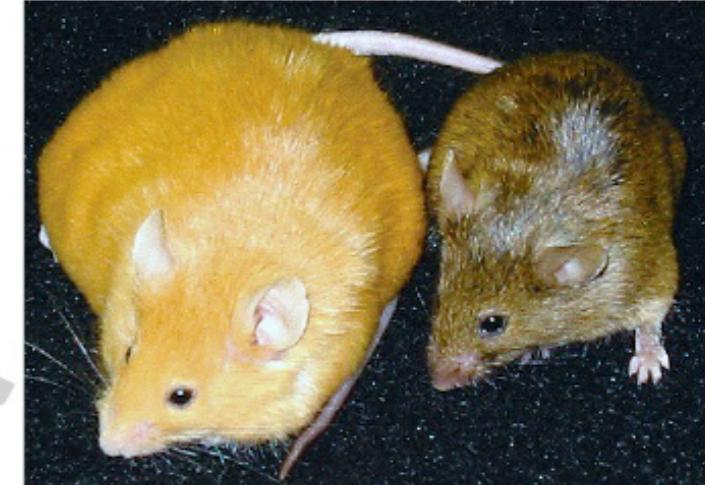
3-year old twins vs. 50-year-old twins



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.



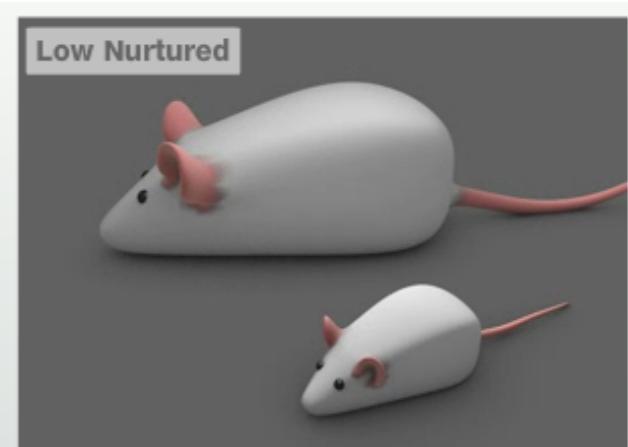
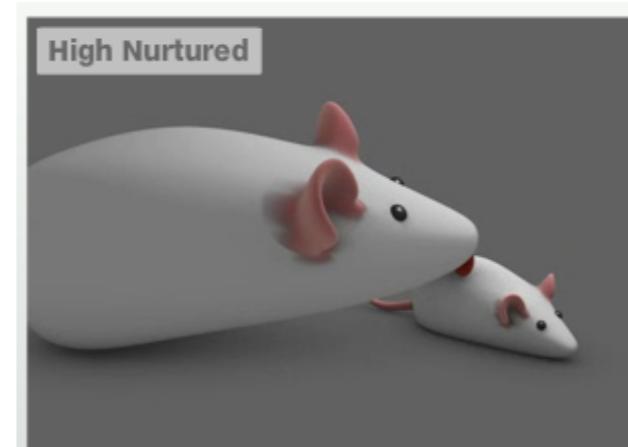
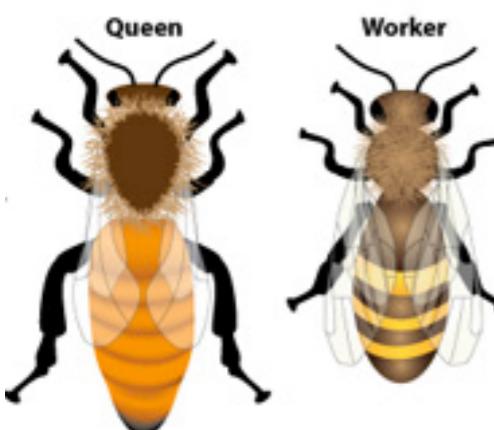
While pregnant, both of their mothers were fed Bisphenol A (BPA) but DIFFERENT DIETS:

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

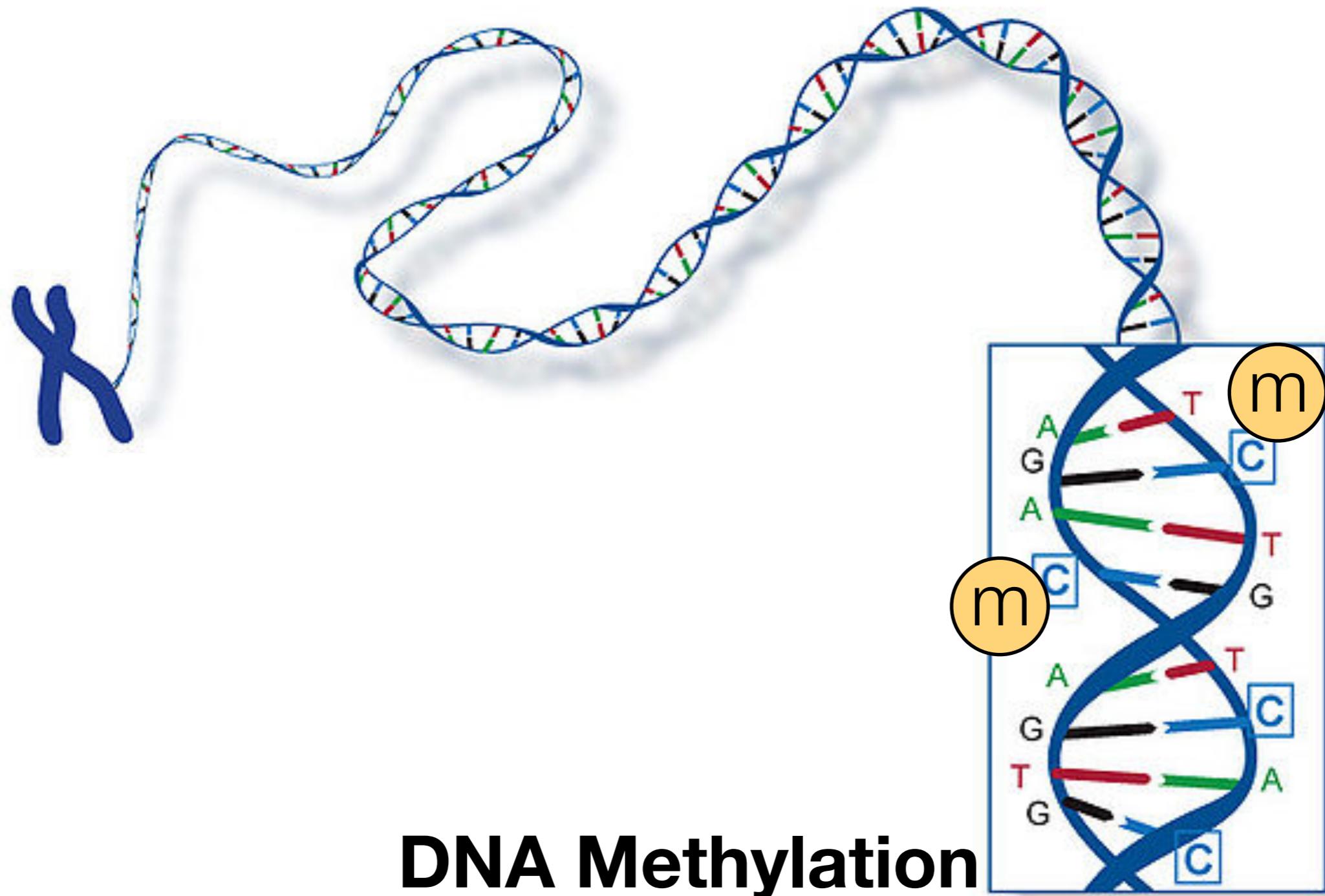


Queen Bee Larvae: Queens are raised in specially constructed cells called "queen cups," which are filled with royal jelly.



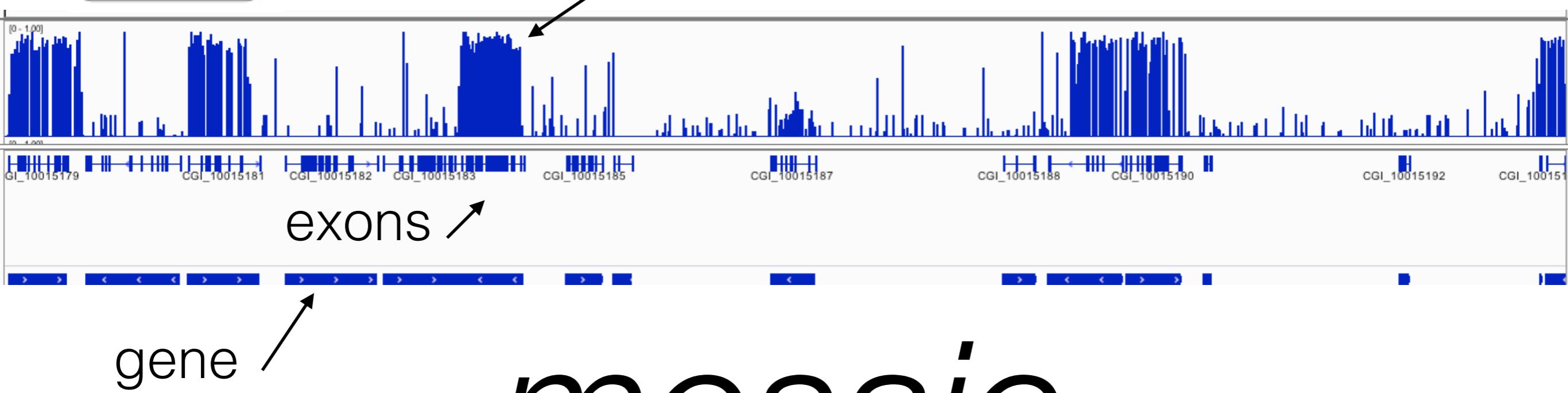
These mothers come from a long line of inbred rats, so their genomes are highly similar. But they care for their pups very differently.

AUDIO



Epigenetic variation **1**

DNA methylation level (0-100%) @ cytosines



*mosaic*

associated with gene bodies

YE mixHYPER.bed

## DMRs in pesticide exposed larvae (hypermethylated)

YE\_mixHYPO.bed

YE mix 22smCG3x

### DMRs in pesticide exposed larvae (hypomethylated)

YE\_mix\_22smCG3x

larvae (mix pesticide exposed) methylation  
larvae (control) methylation



YE\_mixHYPER.bed

DMRs in pesticide exposed larvae  
(hypermethylated)

YE\_mixHYPO.bed

DMRs in pesticide exposed larvae  
(hypomethylated)

YE\_mix\_22smCG3x

larvae (mix pesticide exposed) methylation

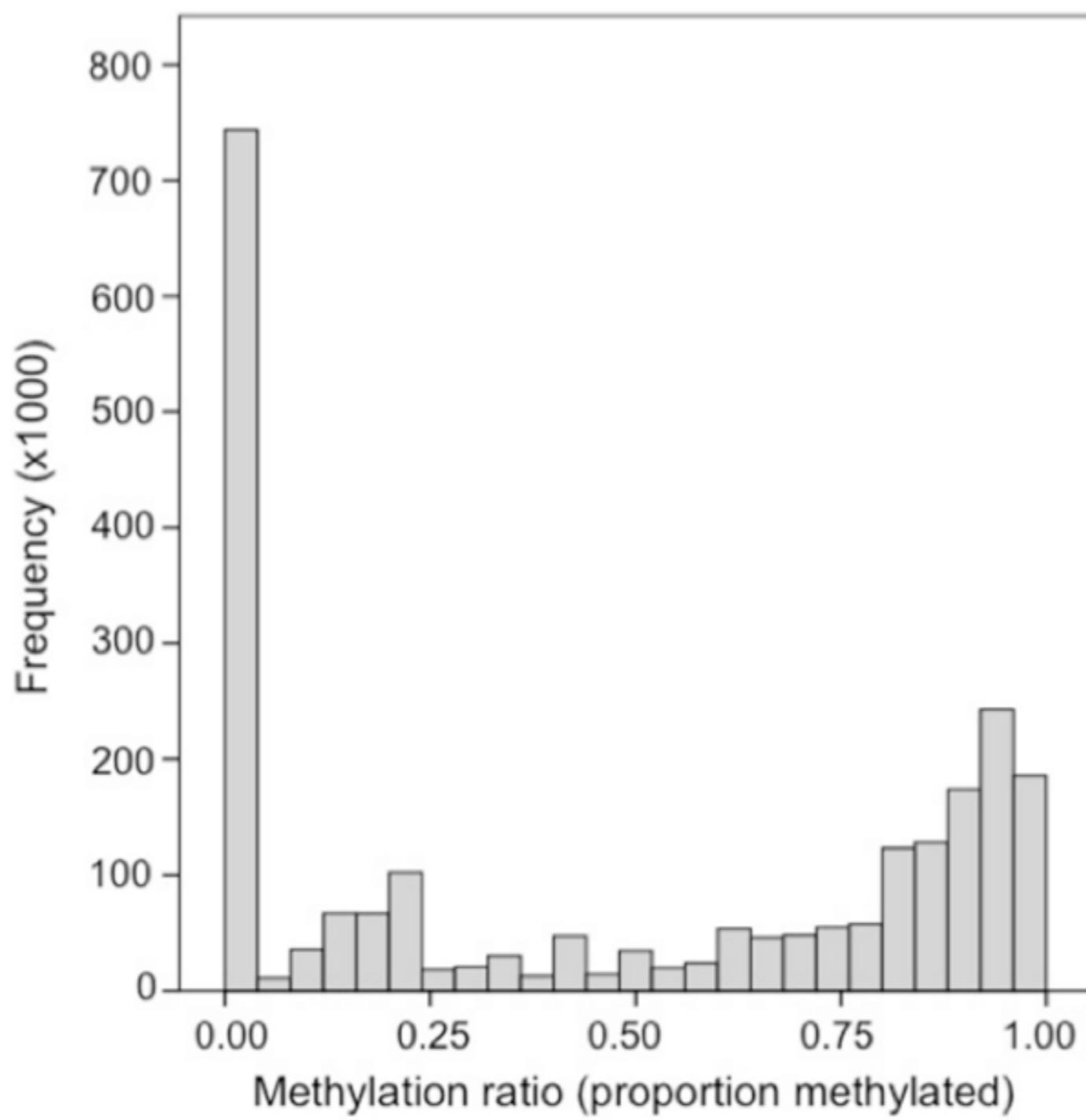
YE\_control\_22smCG3x

larvae (control) methylation



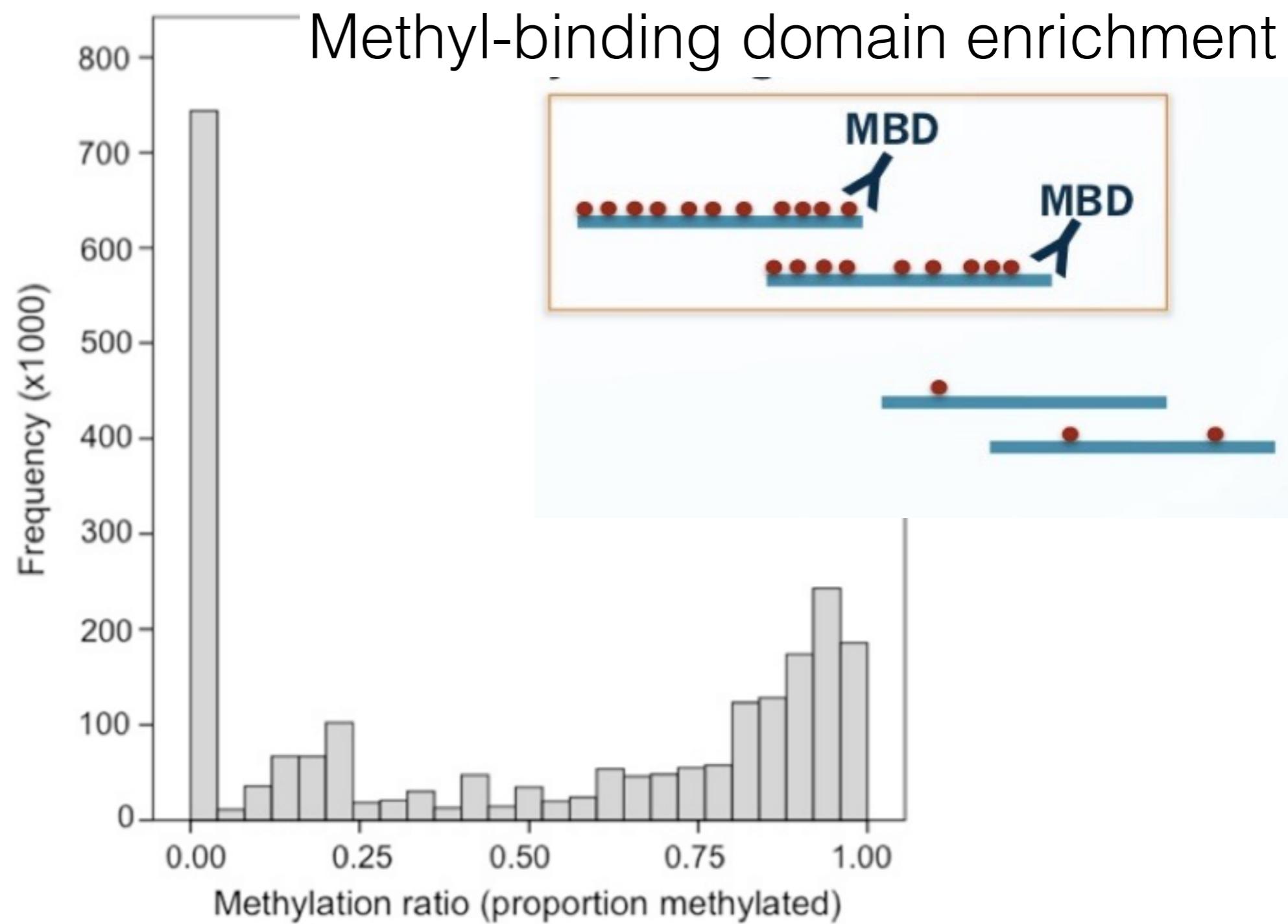
D-128-95-149-192:course-fish546-2016 sr320\$ head -50  
@HWI-D00743:54:C8M0MACXX:1:1101:1208:2248 1:N:0:GATCA  
GTTGGTGTGAGGGAGTGTATGATAAGTGGTTAATATTAAATTTAGTAT  
+  
????DDDDAD?DDD@E2C2<<<ABEDE31@EEDDD?<B9DD>DDEIAD9??  
@HWI-D00743:54:C8M0MACXX:1:1101:1484:2156 1:N:0:GATCA  
TNTGTTAAAATAAATAGAGAATTAAAAATTGAAGTTGTTGAAGTAGATAA  
+  
C#1ADDDDFFFHHHJIIIIJJJJJJJJJJIGGJFGIIJJIIIJGHHHIIH  
@HWI-D00743:54:C8M0MACXX:1:1101:1456:2206 1:N:0:GATCA  
AATGGATGTATAATGTTATTAAATAATTTTTATGAAAATATATGTTTT  
+  
?B@DFDFFFHFHFFGGFGIFIGHIIIIIGGHIIJIIJIIJIIJIGHGGA  
@HWI-D00743:54:C8M0MACXX:1:1101:1943:2168 1:N:0:GATCA  
AAATAGTTAGTGAAATATTAAATTGTTATTGTGATGTAGTGTGTTAT  
+  
@@@FFFDD?DHDHGGIIJIIIEFEGIIJHHGJJIEFEIJIHGIBGHHIIJJ  
@HWI-D00743:54:C8M0MACXX:1:1101:1932:2206 1:N:0:GATCA  
TACGAGTCATGTTAGTGTAGTTATTAAATGGTTATTTATGTGATGG  
+  
CCCFABDFGHHBEGGAEIJHAFIJJIIIIHHGIIIIJEGGDGHJG<  
@HWI-D00743:54:C8M0MACXX:1:1101:2097:2223 1:N:0:GATCA  
TAGTAGTTAAGGATTGGTTGGAAATAGTATGGTGTAAATTTATGTA  
+  
@C@DFEEDHHGHJIIJCEFEGGHGHAEHIG@EGHIIJJIIJJIGJJF  
@HWI-D00743:54:C8M0MACXX:1:1101:2579:2189 1:N:0:GATCA  
GCGGTTATGATTGTTGTGTTATTAGTAAATGAAATTGTAAGTTATAGA  
+  
@CCFDDDFGGGGHGJIGJII9FHI@GCHEHIIIEIDGIHFFFCD?DFIIGG  
@HWI-D00743:54:C8M0MACXX:1:1101:2768:2171 1:N:0:GATCA  
GCCGGGGTTGGTTGTAAATTAAATTATATTGTAAAATTACGATATA  
+  
CCCCFFFFFFHHF?FHIEGIIJJGIIJJJJJJJJEGIJJGGHHFFFDE  
@HWI-D00743:54:C8M0MACXX:1:1101:2791:2190 1:N:0:GATCA  
TAATACAGTAAATTGTTTTATTGTATTAGAATTGTTTTATGTTA  
+  
CCCCFFFFHDFHHJJFGHJJIEECBEIIIIIIJJIIHG>DFHIGHGH  
@HWI-D00743:54:C8M0MACXX:1:1101:3033:2168 1:N:0:GATCA  
CTCAAGTGATCGGATTAGTTATATTGTTAAAATTTAATTAA  
+  
CCCCFFFBDFGGHGIIJJGJFHIJJJJJJJJJJJJJJJJJJJJJJJJJJ  
@HWI-D00743:54:C8M0MACXX:1:1101:3282:2145 1:N:0:GATCA  
GNGAGTTGGTTGTAAAGTTATTGATGATGTTATGTTAATAGAAA  
+  
C#4ADDDFGFFFHHHGIJJJJHJJJJJJJJJJFIIIIHIEHGHGIJJJJJJ  
@HWI-D00743:54:C8M0MACXX:1:1101:3394:2168 1:N:0:GATCA  
TAGTGGTGGTTAATTAAATGATTGGGGATTTTTATGTTATATATAT  
+  
BCCDFEBDHHFHHIJJGEHIEIJGIHGIEHIIJJJJICIJJJJJFJJJ

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.

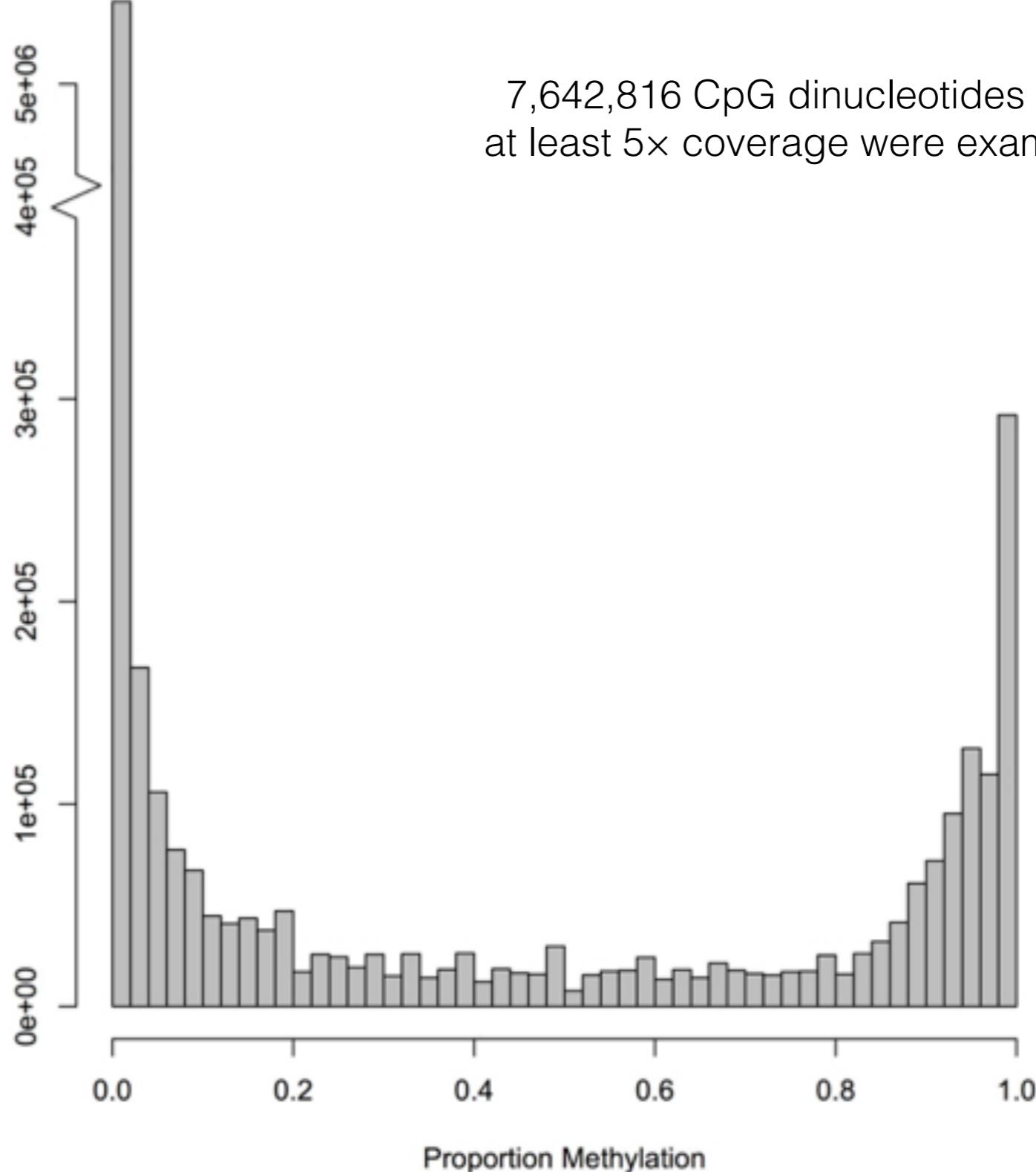
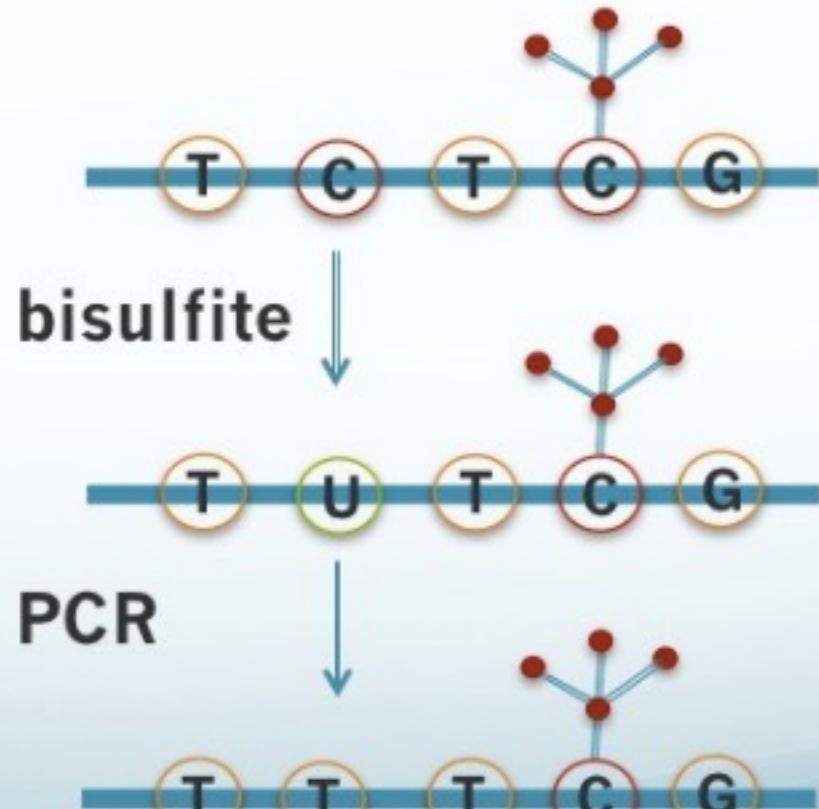
Epigenetic variation **1**



**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

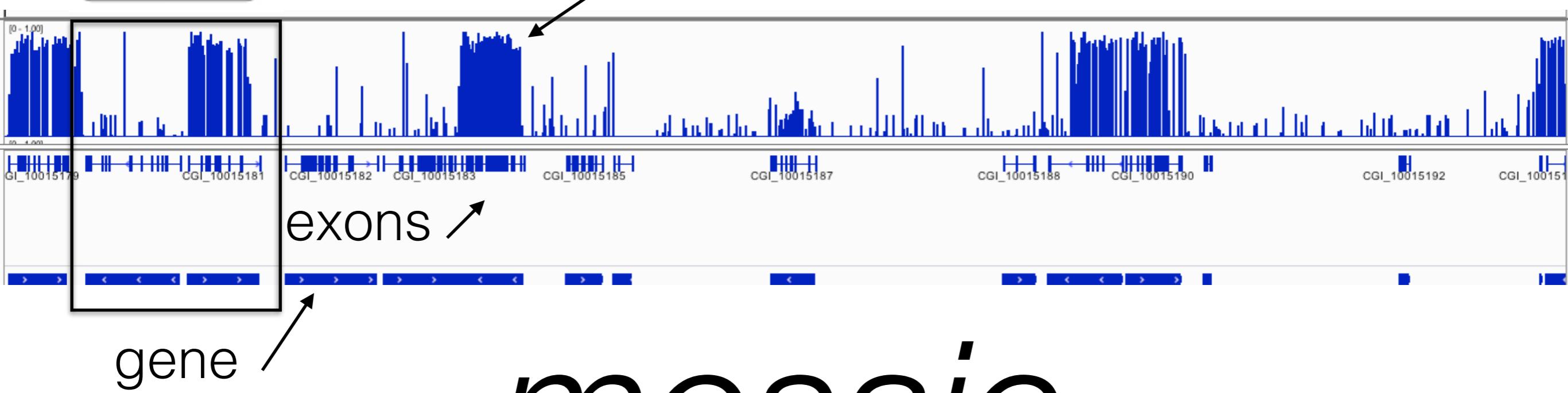


7,642,816 CpG dinucleotides with at least 5x coverage were examined



Epigenetic variation **1**

DNA methylation level (0-100%) @ cytosines



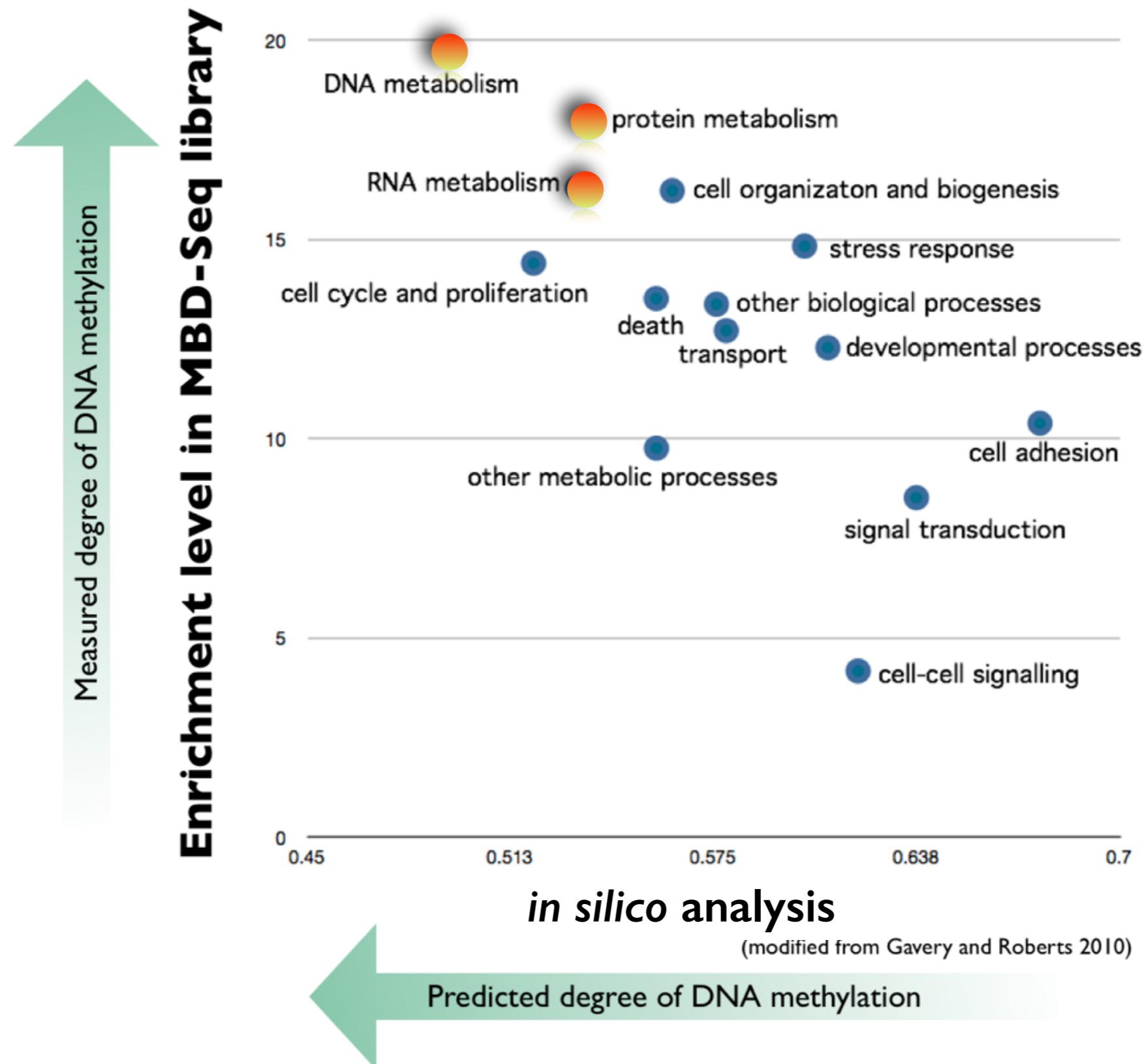
*mosaic*

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

associated with gene bodies

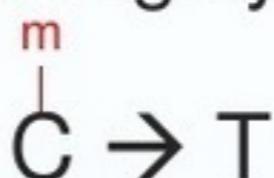


Epigenetic variation **1**



# *in silico* approach

- Principle:
  - Methylated cytosines are highly mutable



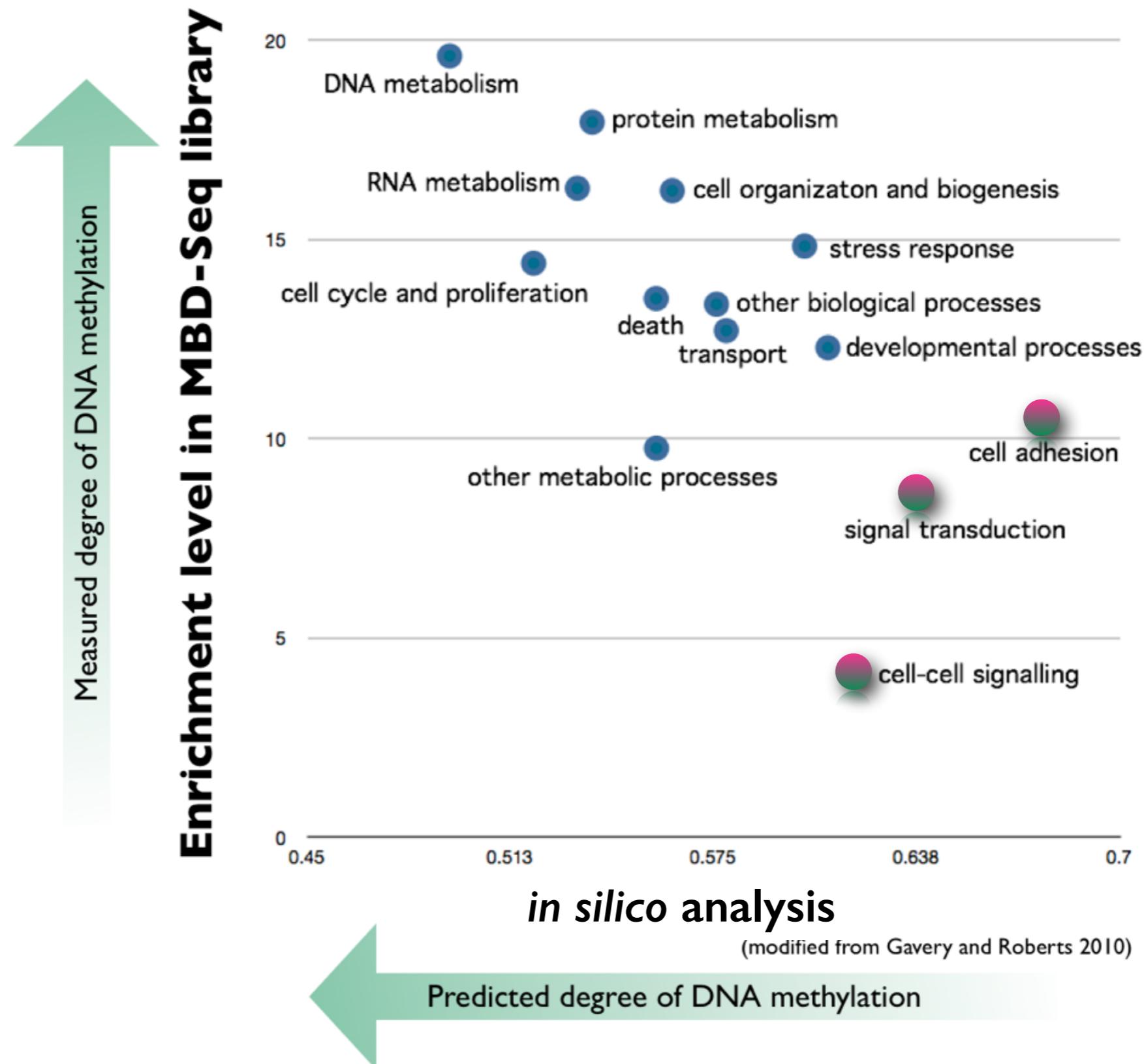
- Methylated regions of DNA are depleted of CpG dinucleotides over evolutionary time (CpG to TpG)

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



high = unmethylated  
*approaching 1*

Epigenetic variation **1**



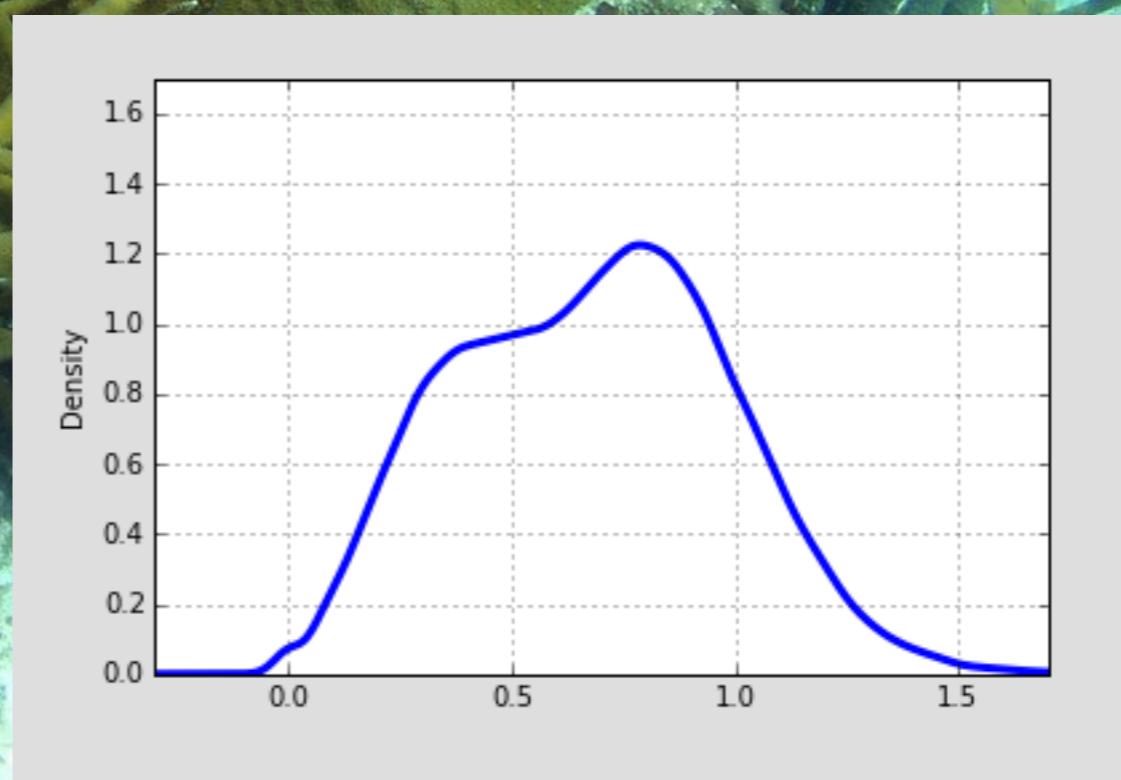
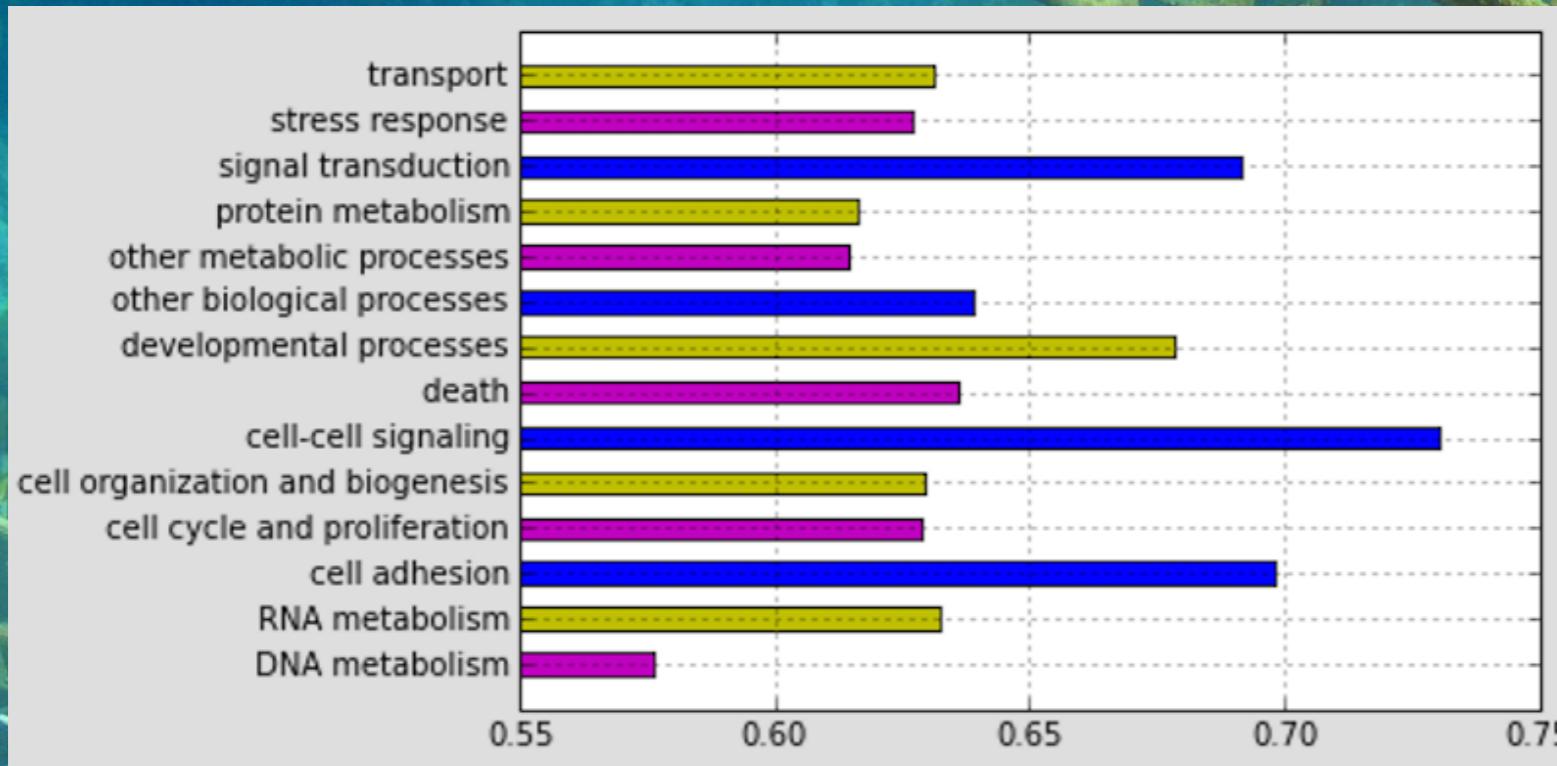
Jay Dimond

*Acropora palmata*

*Acropora cervicornis*

# Jay Dimond

*Acropora palmata*



# Jay Dimond

Special Issue

## Germline DNA methylation in reef corals: patterns and potential roles in response to environmental change

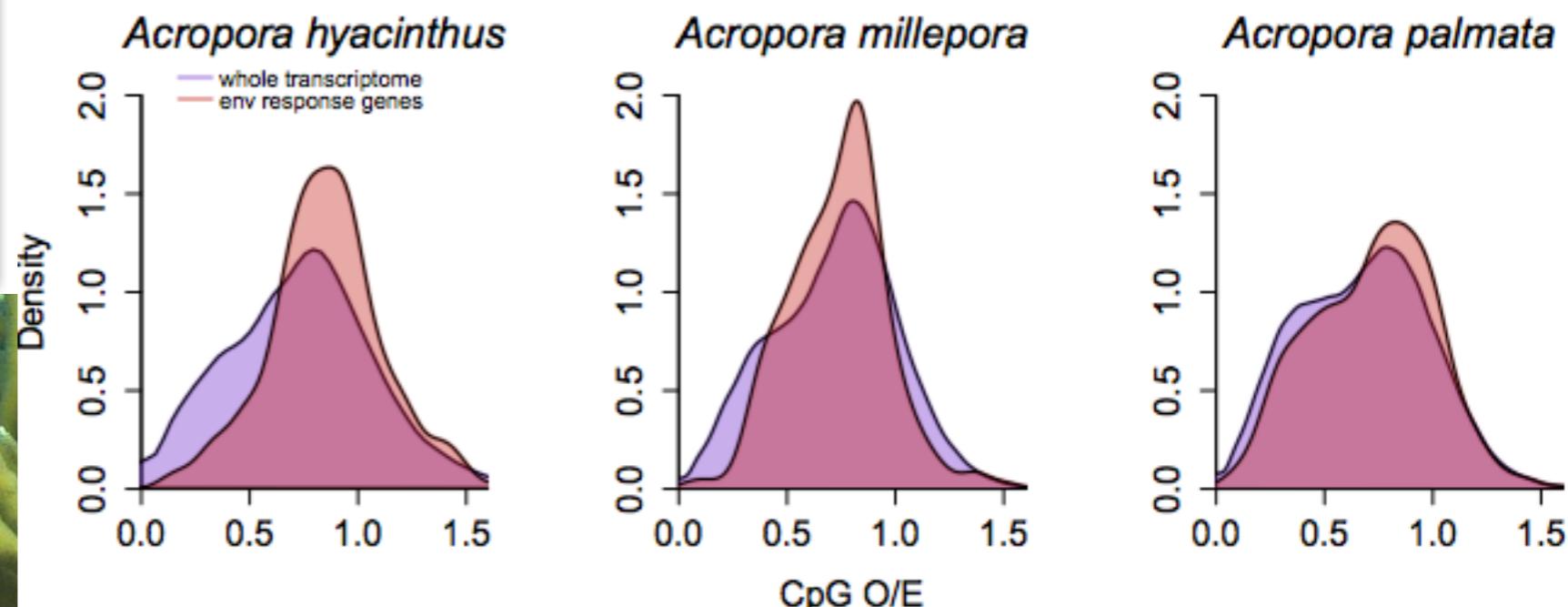
James L. Dimond and Steven B. Roberts\*

DOI: 10.1111/mec.13414

This article is protected by copyright. All rights reserved.



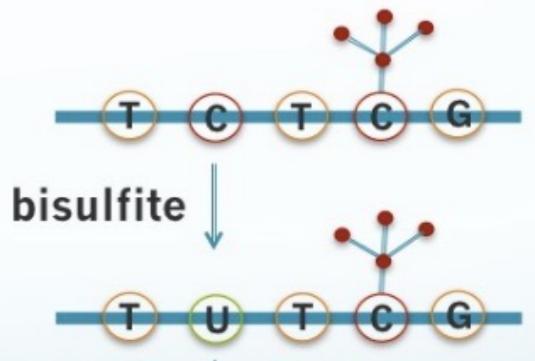
Issue





Epigenetic variation 1

# Family and Developmental Variation



Sperm &  
Larvae  
(72h & 120h)



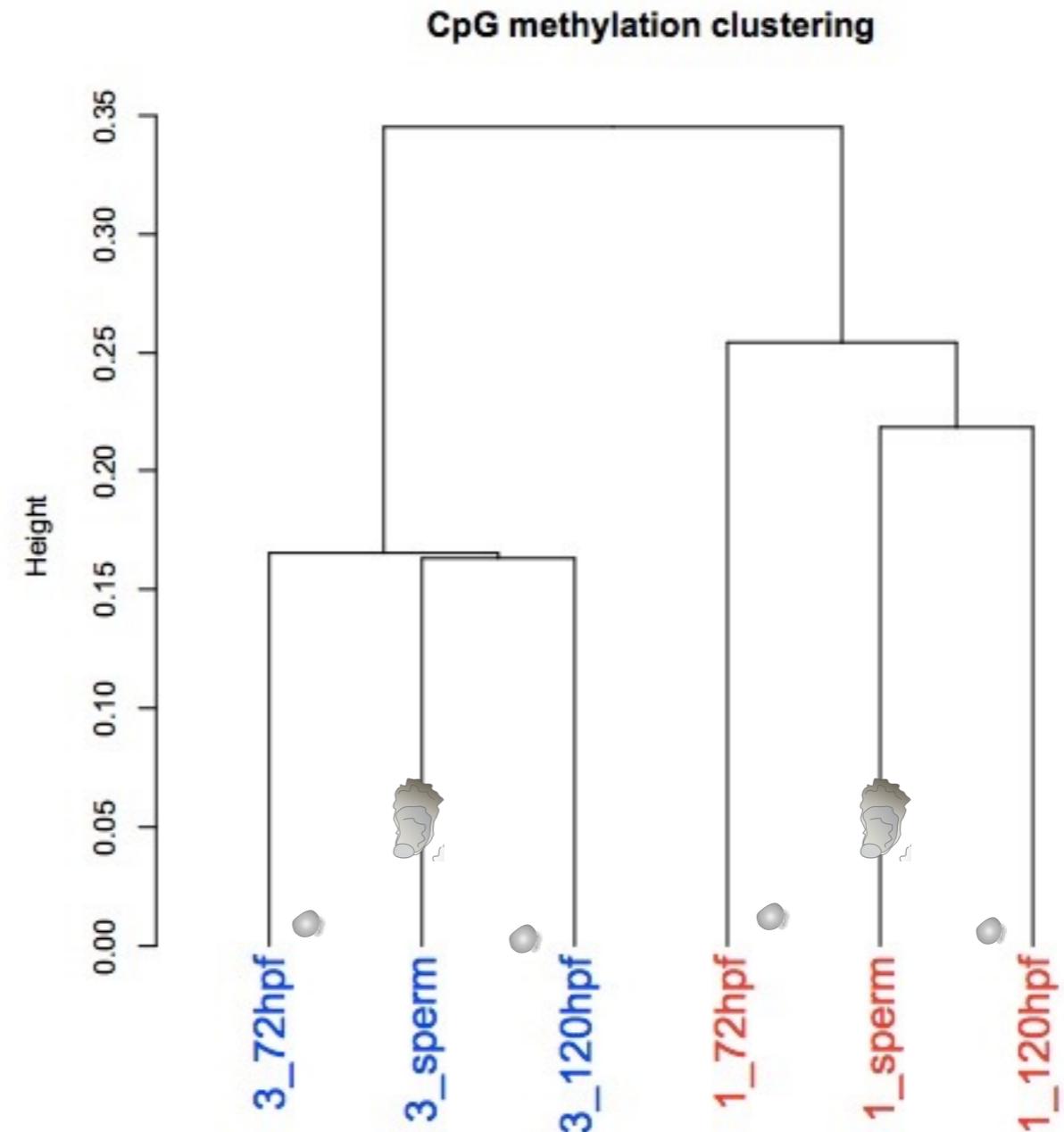
**bioRxiv**  
beta  
THE PREPRINT SERVER FOR BIOLOGY

New Results

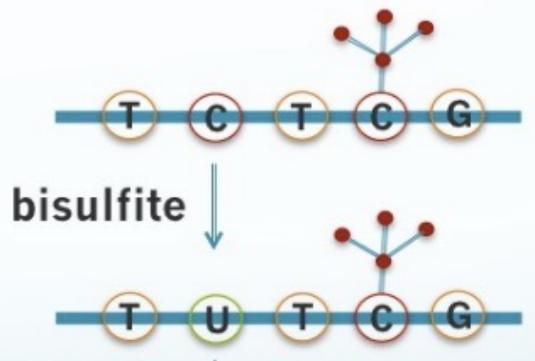
**Indication of family-specific DNA methylation patterns in developing oysters**

Claire E. Olson , Steven B. Roberts

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



# Family and Developmental Variation



## Inheritance



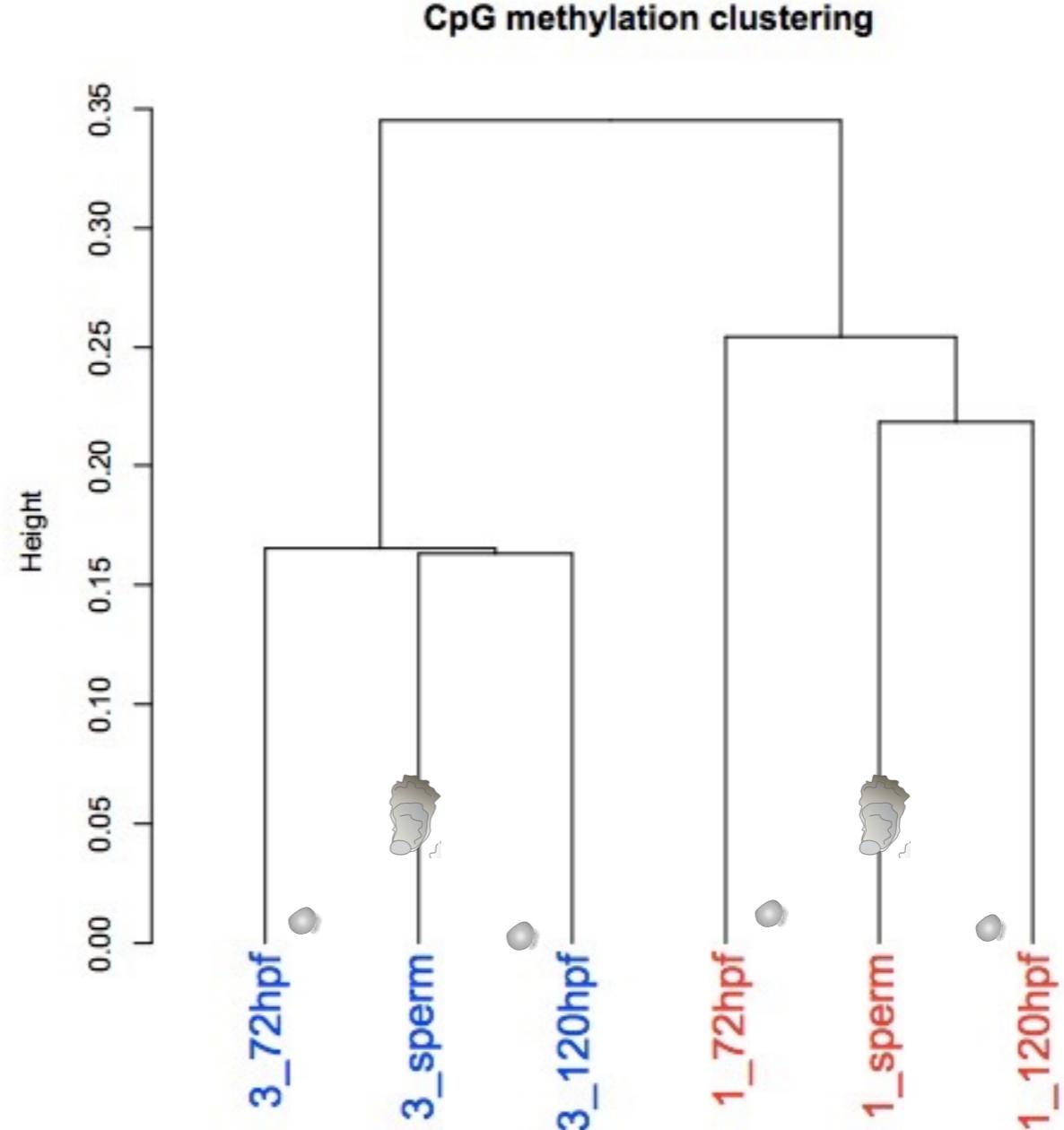
**bioRxiv**  
beta  
THE PREPRINT SERVER FOR BIOLOGY

New Results

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

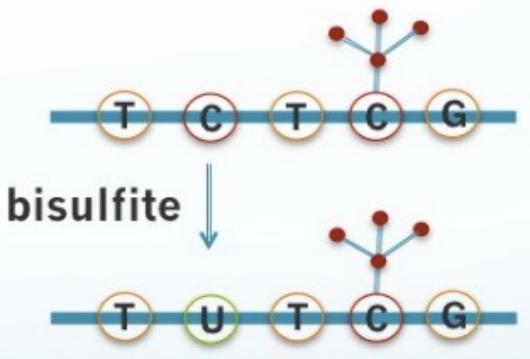
Claire E. Olson , Steven B. Roberts

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

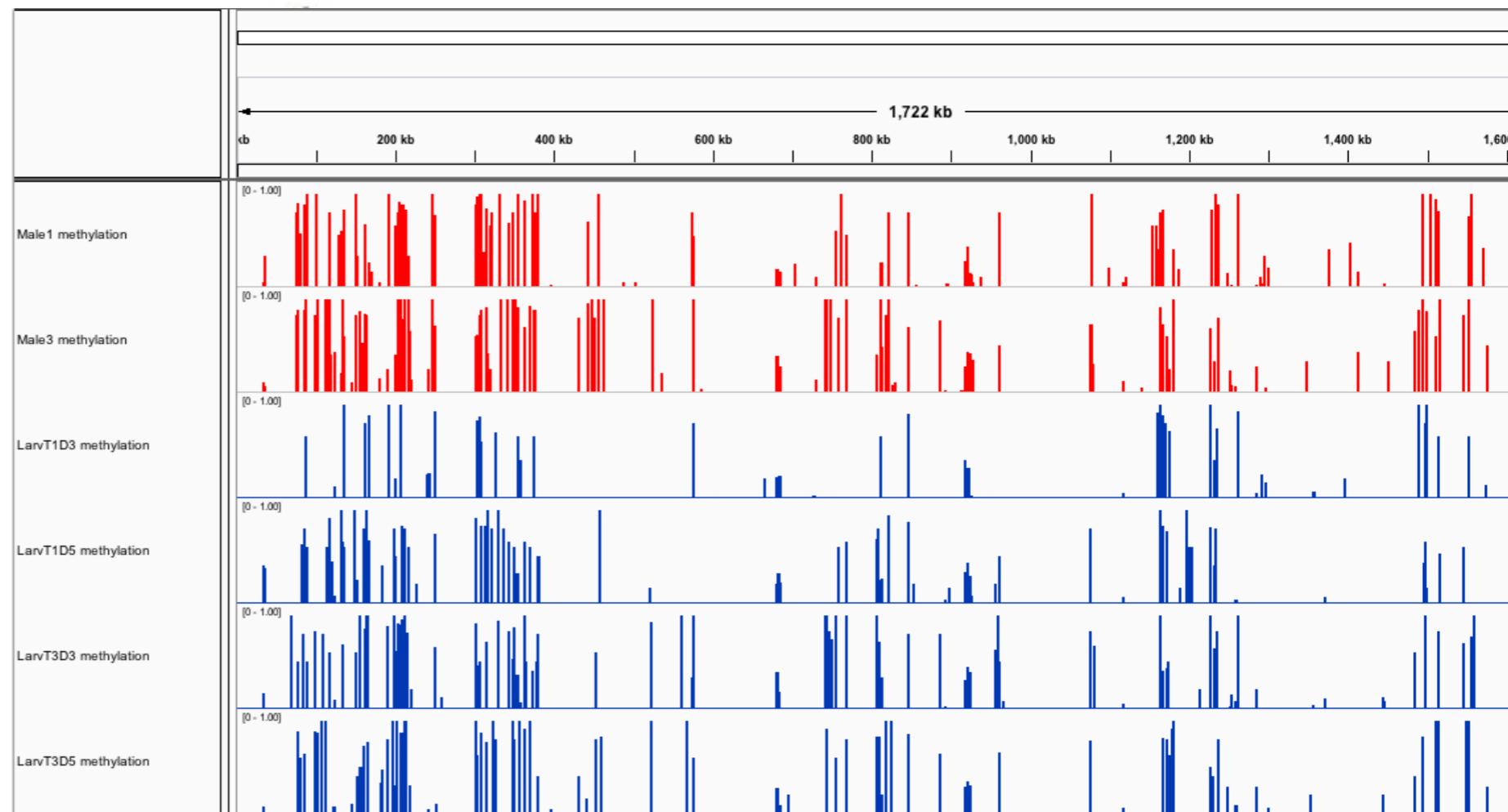


## Epigenetic variation 1

# *Family and Developmental Variation*



# Sperm & Larvae (72h & 120h)



DNA methylation level (0-100%) @ cytosines



bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

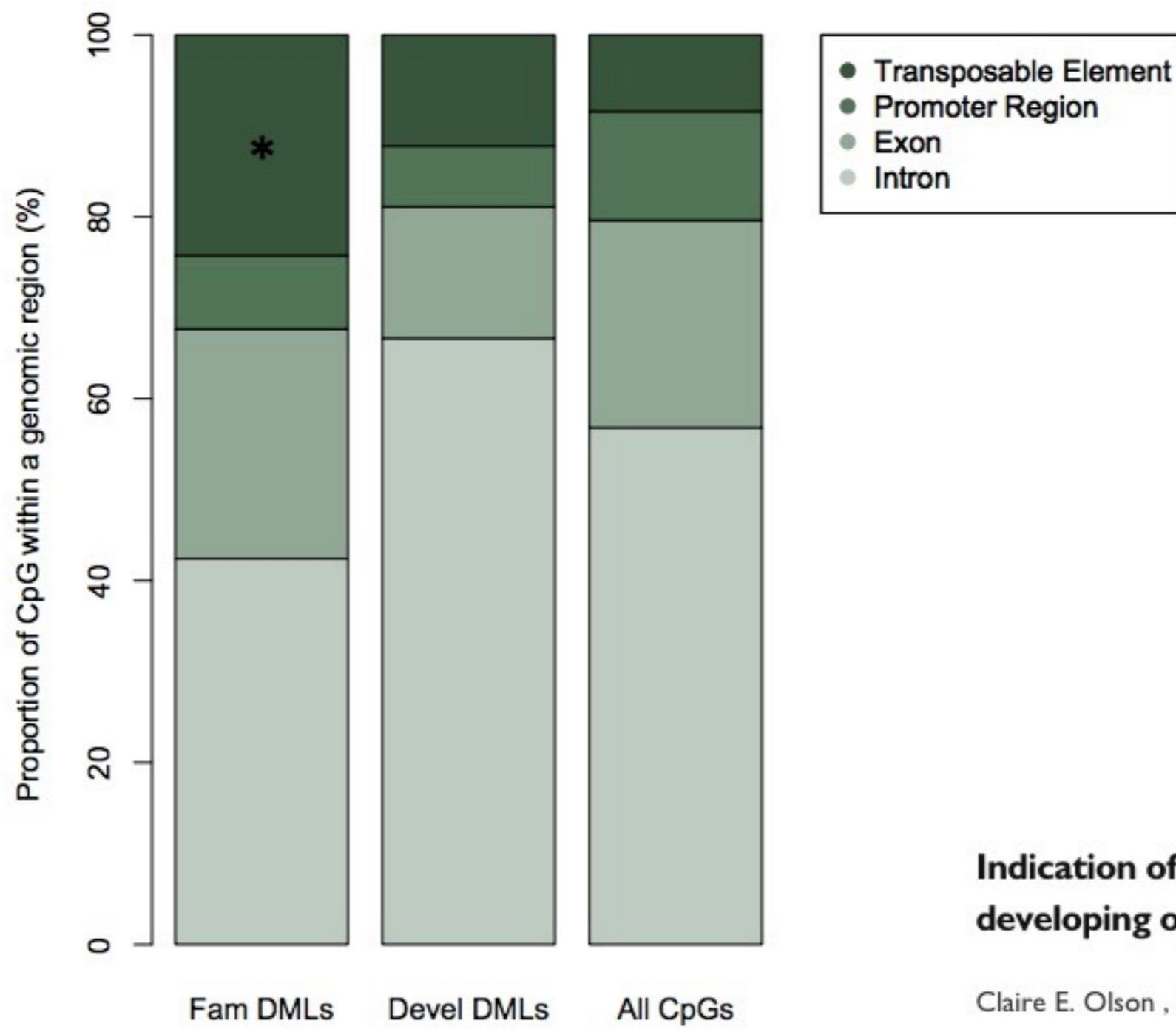
New Results

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

Claire E. Olson , Steven B. Roberts

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

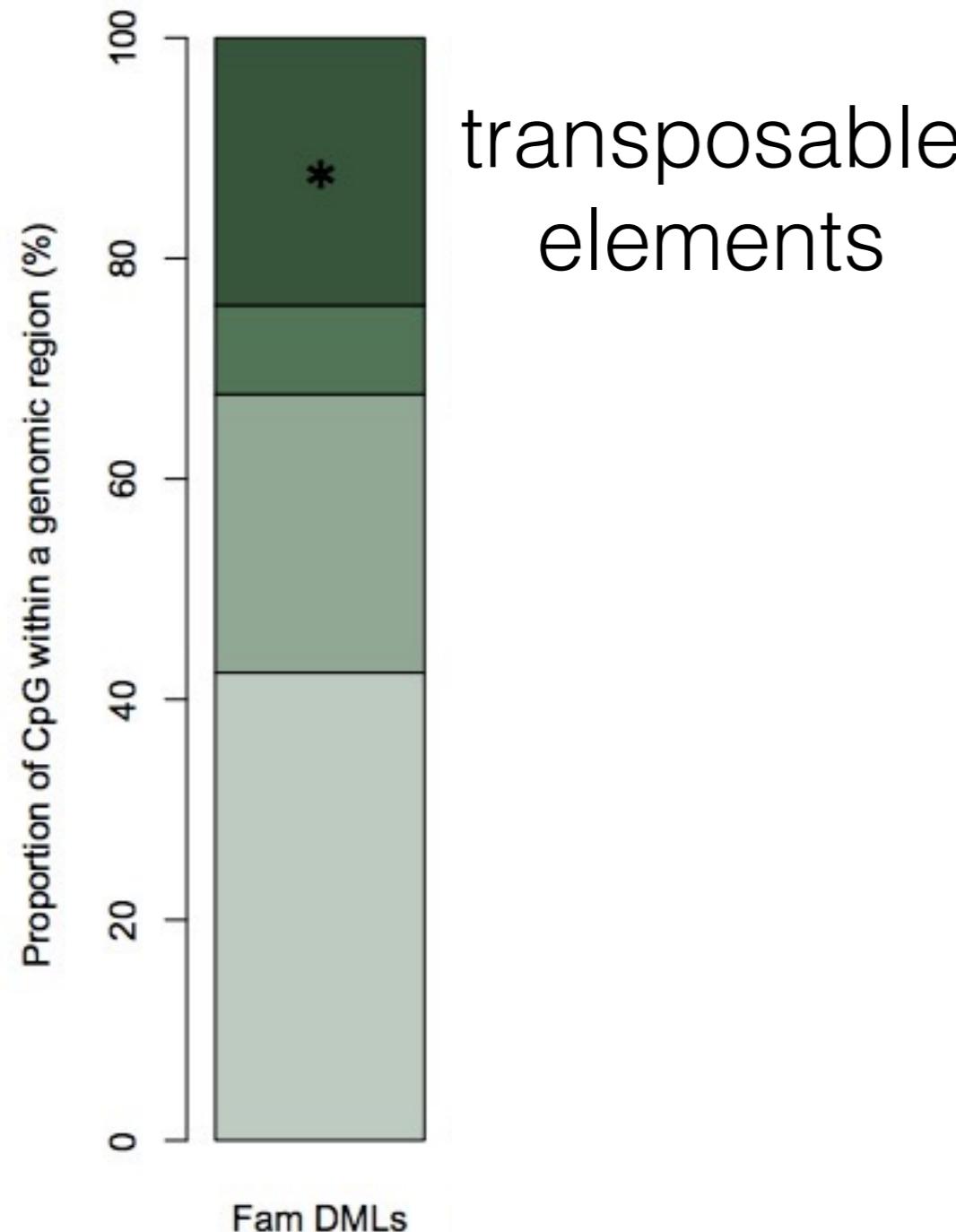
# Where are the differences?



**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**



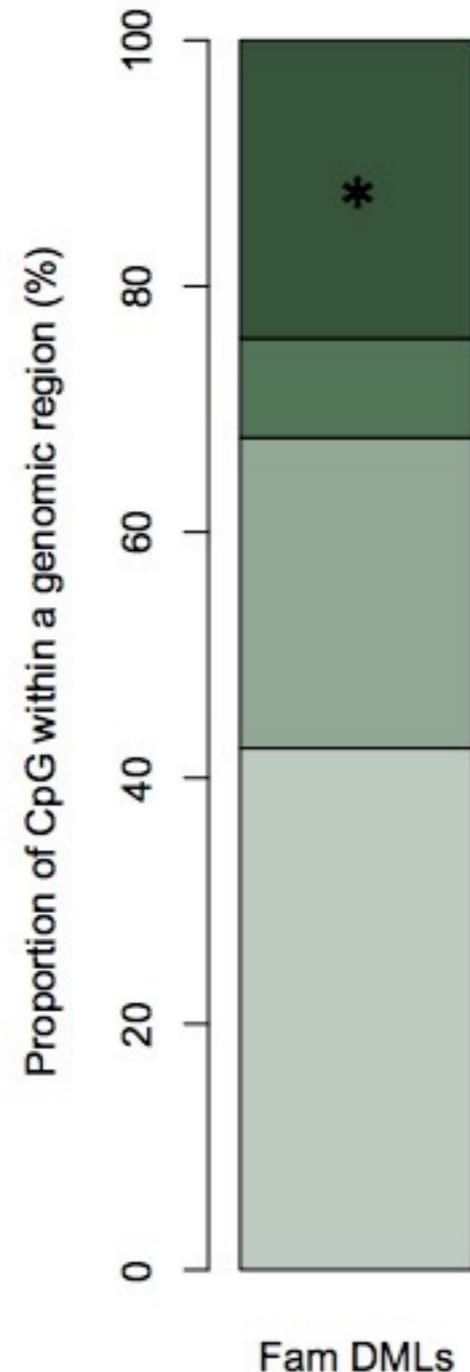
Differentially methylated loci in gene bodies could be lethal or deleterious, as they would alter gene expression.

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>

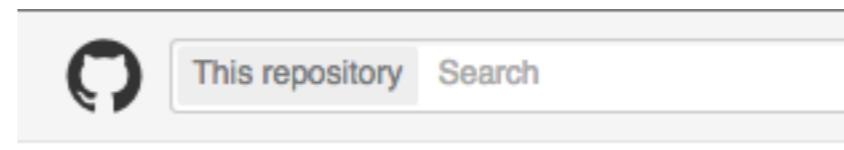
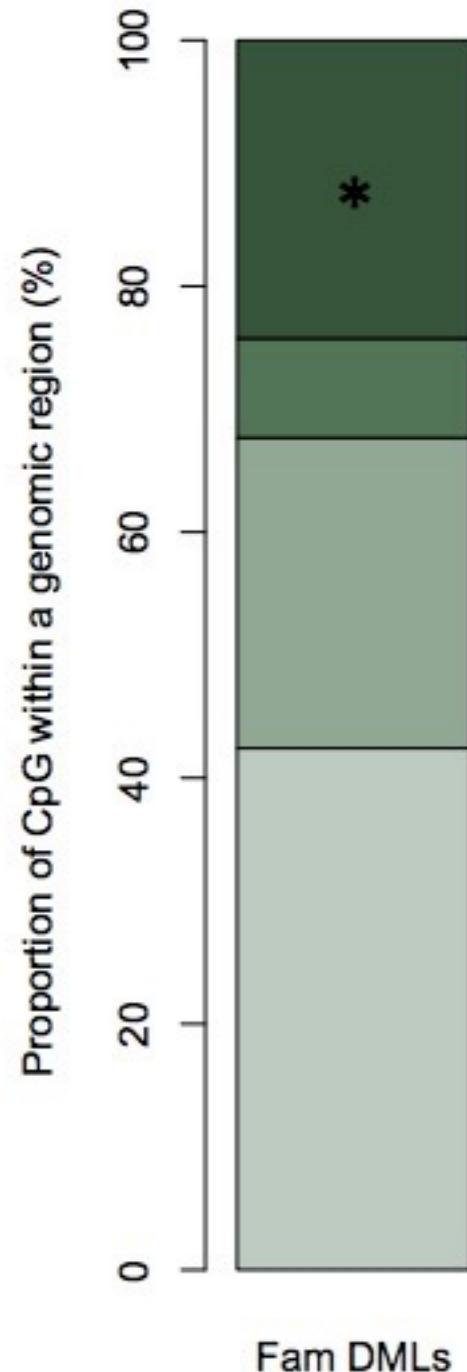


***Assumes some form  
of random process***  
occurring during  
gametogenesis?

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

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

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](#)

- 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?

**Methylation  
Enhanced  
Random  
Variation  
*theory***



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



*in silico analysis*

(modified from Gavery and Roberts 2010)

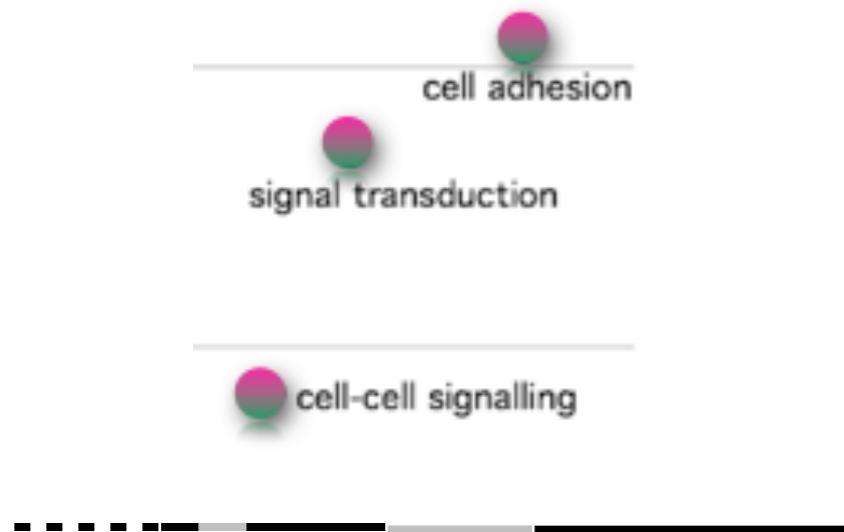
Predicted degree of DNA methylation

Roberts and Gavery 2012

Gene expression

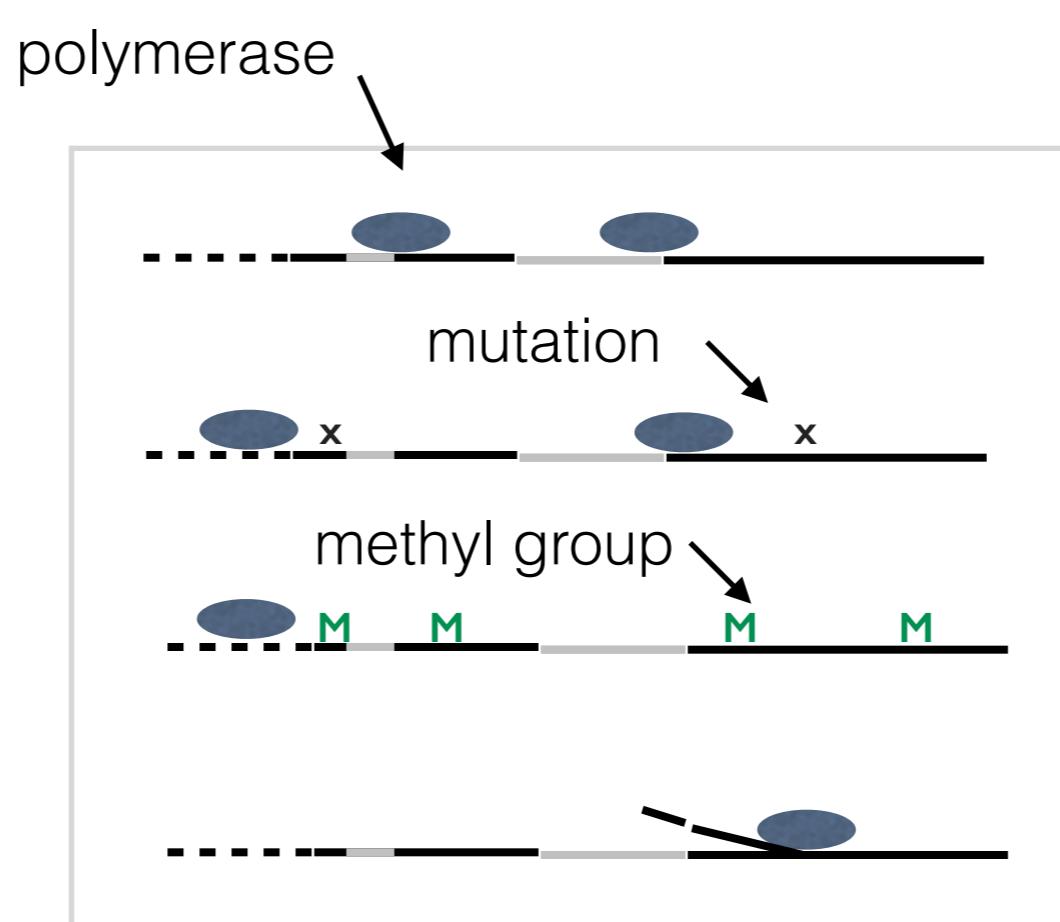
2

Epigenetic variation

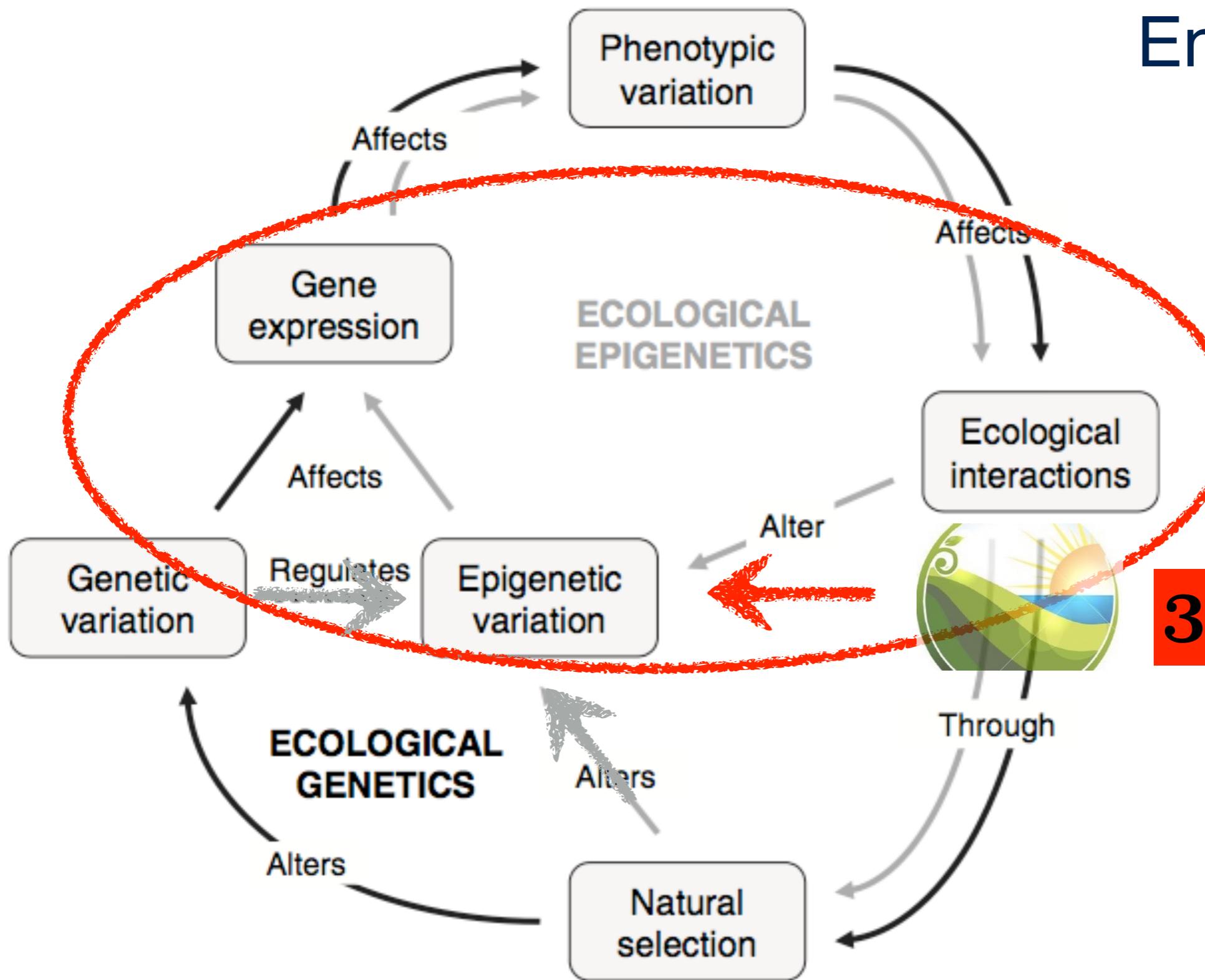


## sparse methylation

tissue / temporal specific and inducible genes



# Environmental Influence



*Ecology Letters*, (2008) 11: 106–115

doi: 10.1111/j.1461-0248.2007.01130.x

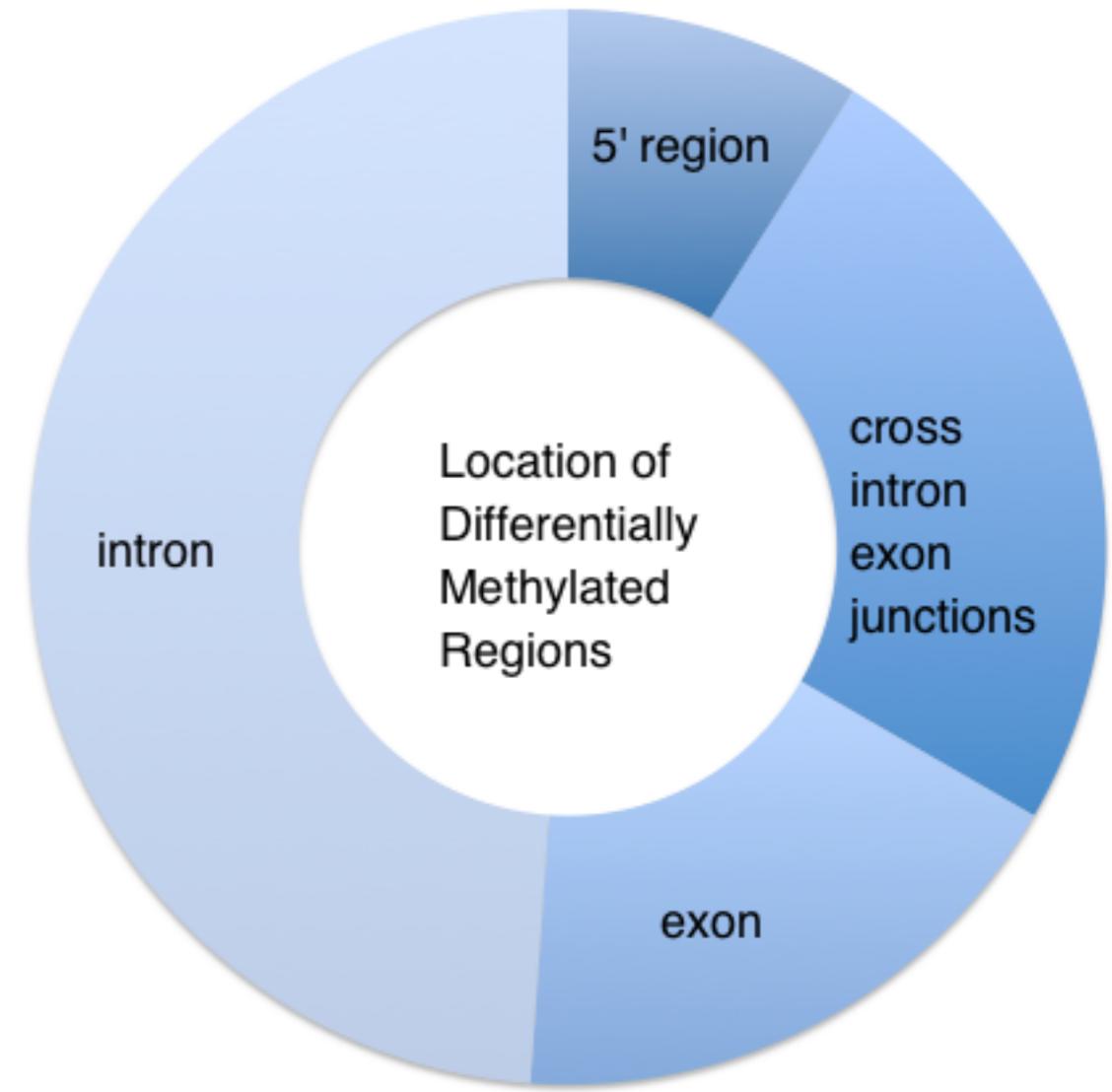
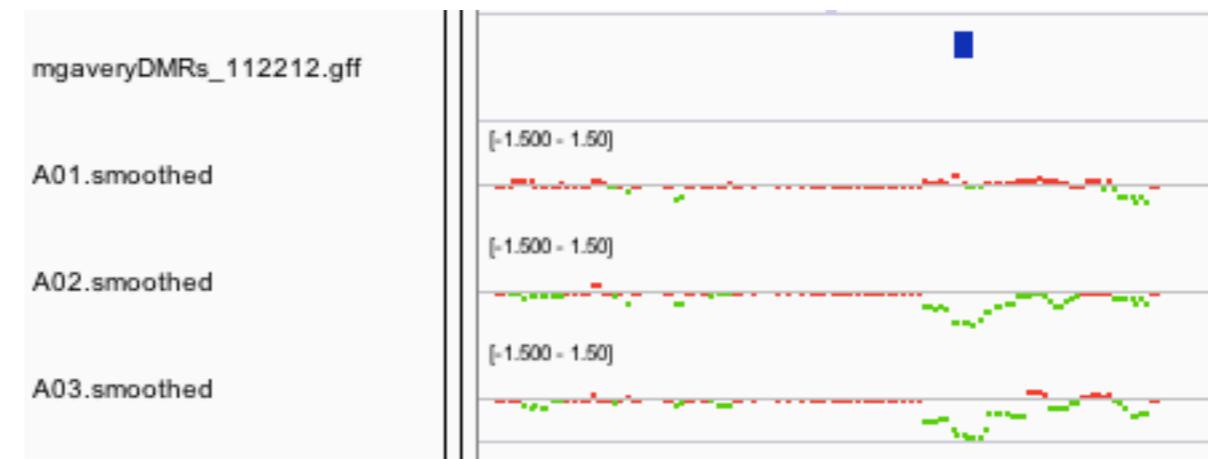
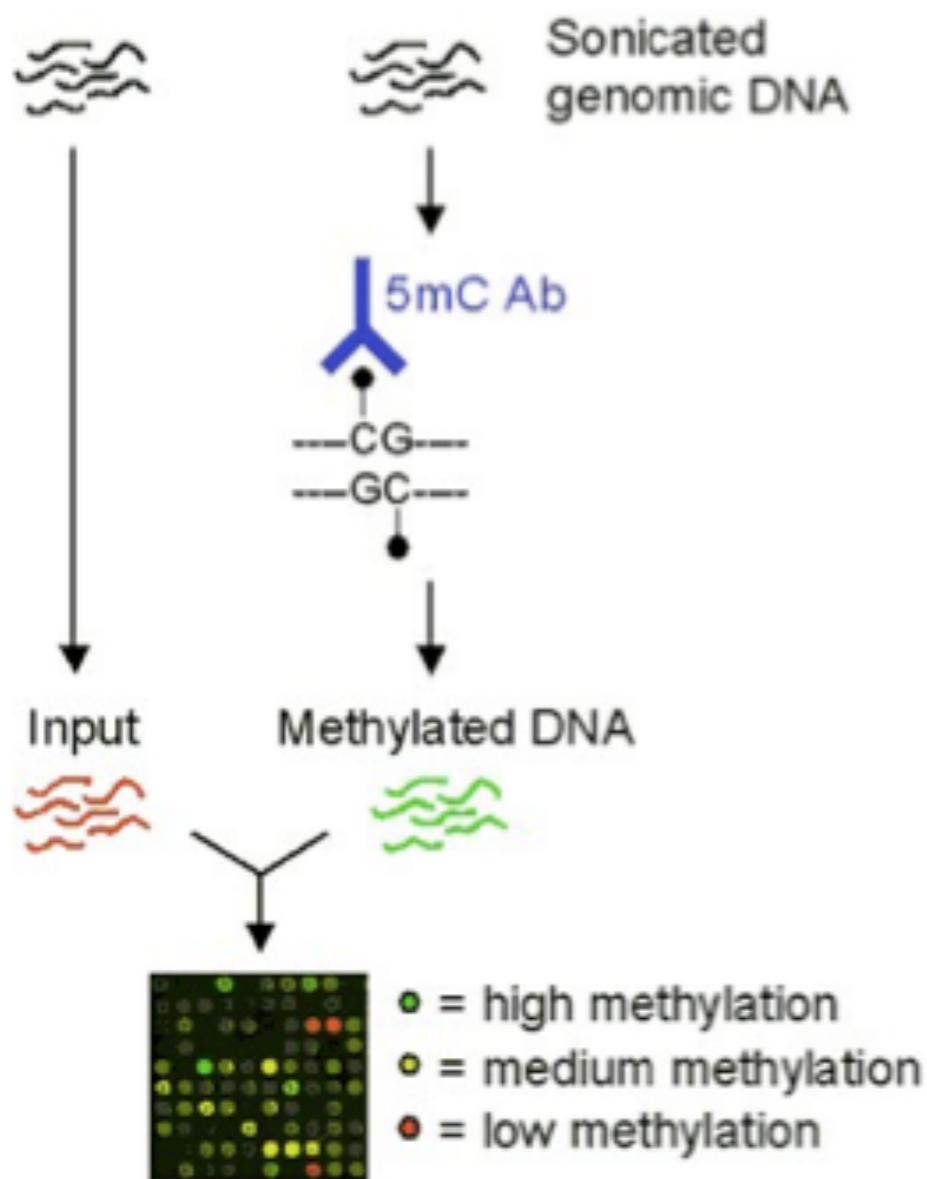
modified from

IDEA AND  
PERSPECTIVE

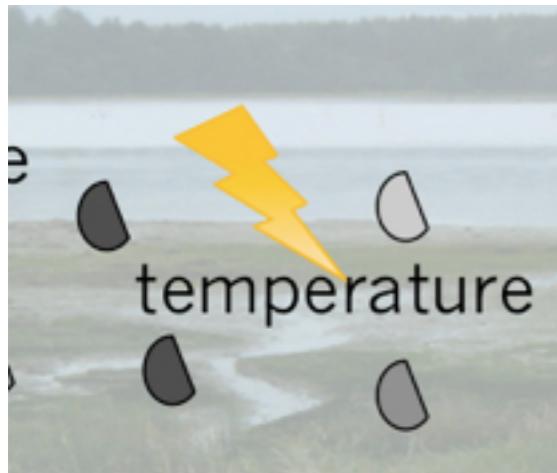
Epigenetics for ecologists

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

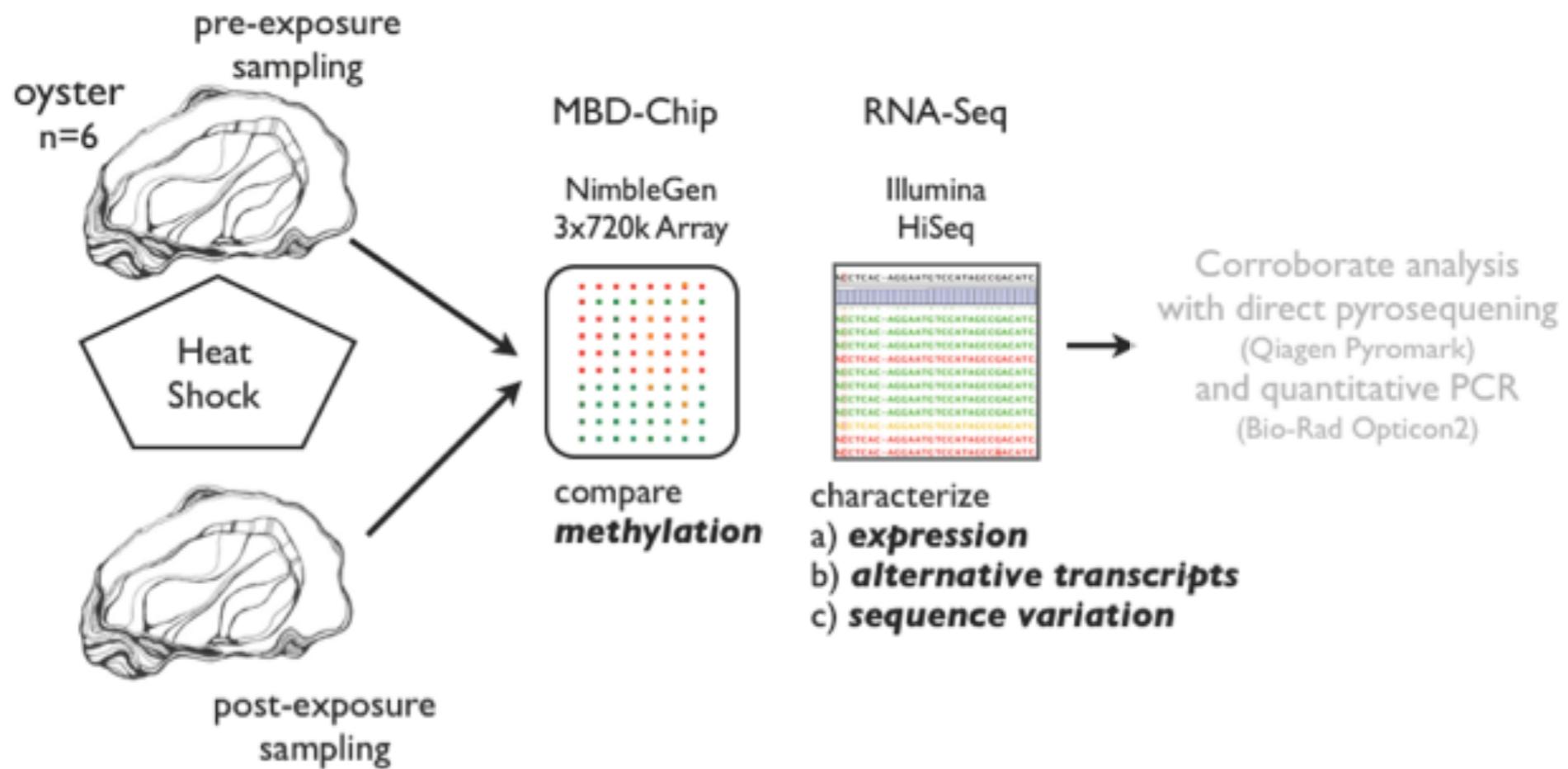
# Environmental impact (Estrogens)



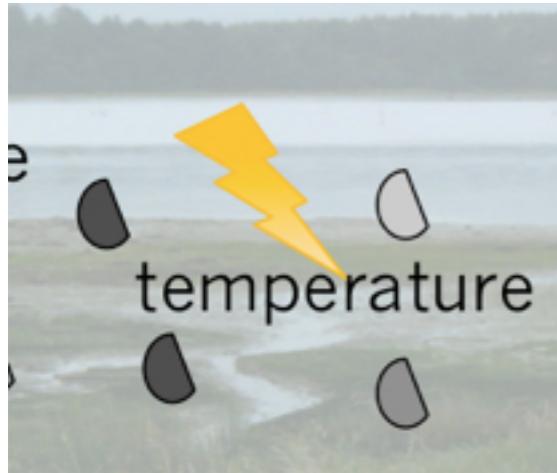
# Environment and gene expression



*stochastic or targeted?*



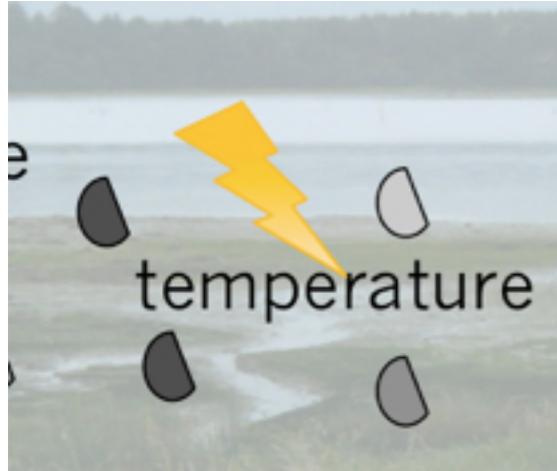
# Environment and gene expression



*stochastic or targeted?*

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

# Environment and gene expression

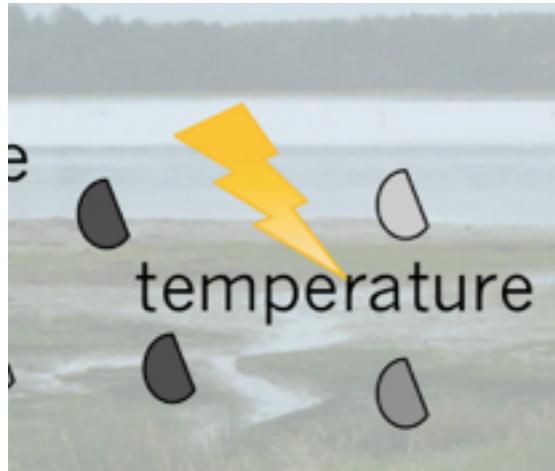


*stochastic or targeted?*

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

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

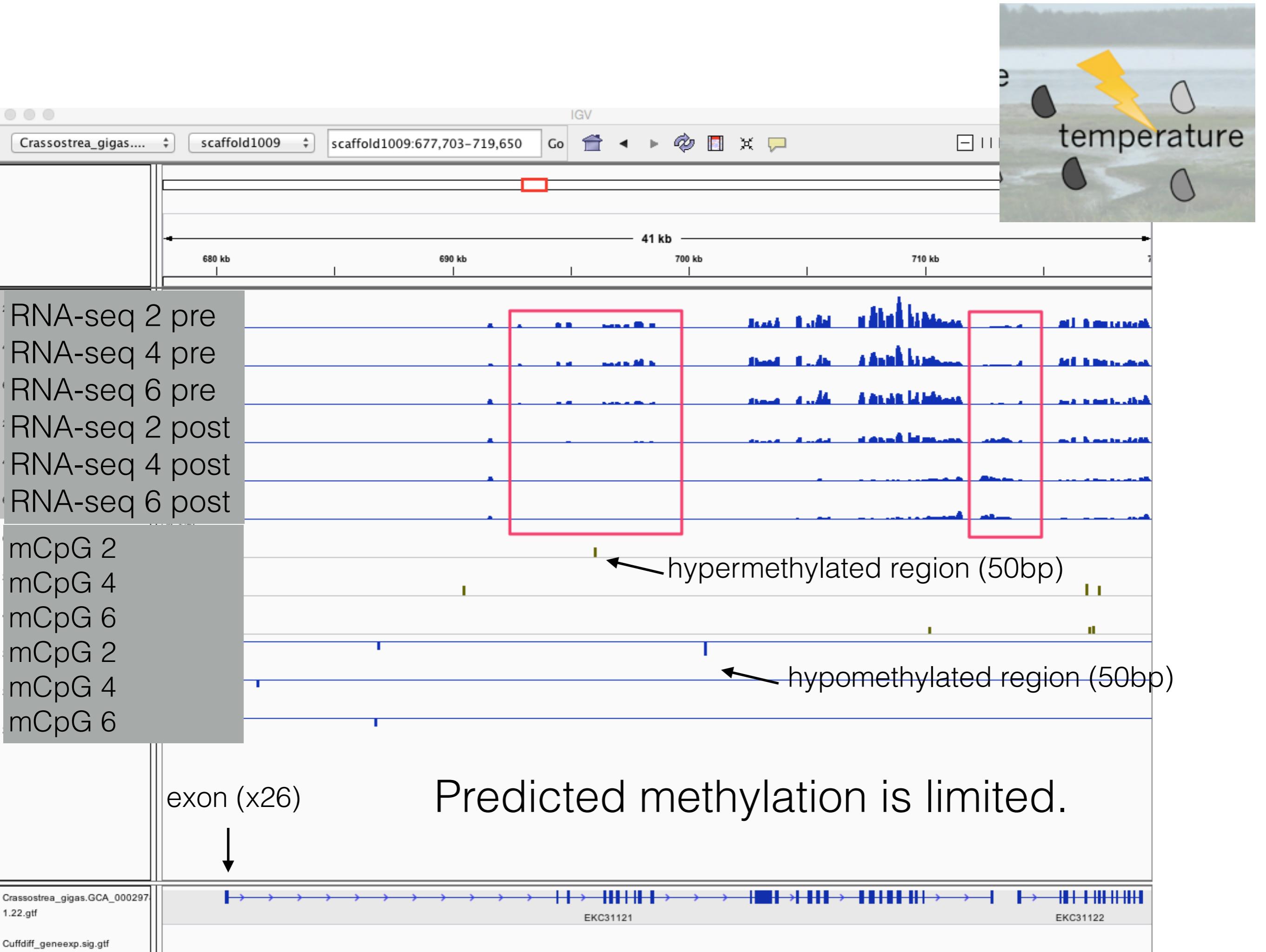
# Environment and gene expression



*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.







# 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.

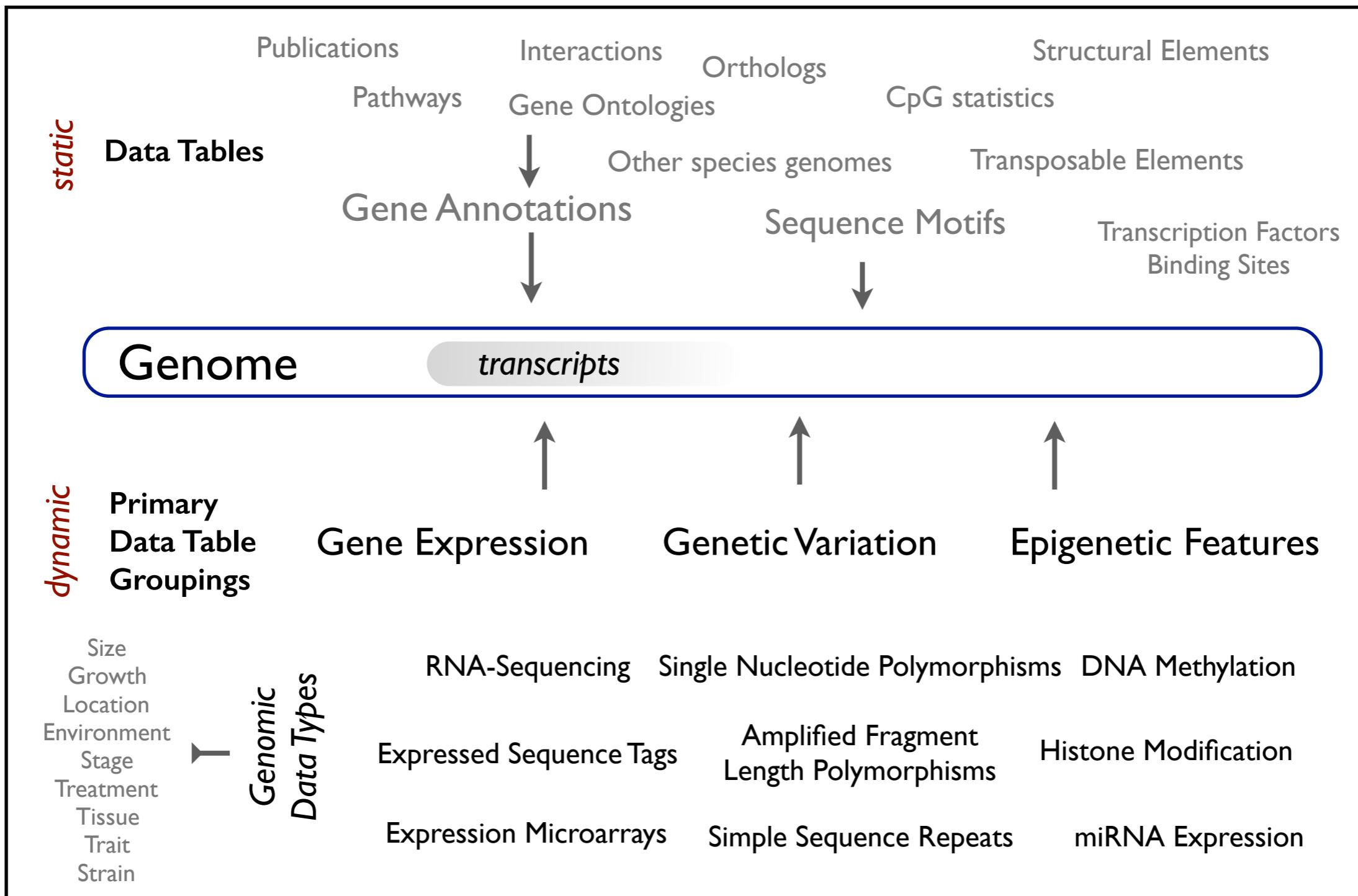
### qDOD online

<a href="#">qDOD_Cgigas_gene_fasta</a>	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>
<a href="#">qDOD_Zhang_Gil_gene_RNA-seq</a>	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>
<a href="#">qDOD_Zhang_Mgo_gene_RNA-seq</a>	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



Your datasets  
All datasets  
Shared datasets  
Recent activity... 0  
Recently viewed »

Upload dataset  
New query

## Phel\_uniprot\_sprot\_sql.tab

Last modified: Jan 5, 2015 2:29 PM

sr320@washington.edu

[Click here to add a description](#)

[Click here to add a tag](#)

```
SELECT * FROM [sr320@washington.edu].[table_Phel_uniprot_sprot_sql.tab7D708]
```

[Edit dataset](#)[Derive dataset](#)[Create snapshot](#)[More actions ▾](#)

DATASET PREVIEW Rows 1 - 100 of 10513 | Columns 14 of 14

<< first < prev 1 [2](#) [3](#) [4](#) [5](#) [next >](#) [last >>](#)

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13	Column14
Phel_contig_100	sp	Q16513	PKN2_HUMAN	81.33	332	61	1	7935	6940	653	983	5E-162	537
Phel_contig_1000	sp	Q8R4U2	PDIA1_CRIGR	53.62	442	201	2	199	1512	31	472	5E-146	464
Phel_contig_10006	sp	Q9Y2H9	MAST1_HUMAN	70.93	289	82	1	861	1	434	722	6E-132	415
Phel_contig_10021	sp	Q96MU7	YTDC1_HUMAN	60.85	212	82	1	1115	1750	294	504	1E-73	258
Phel_contig_10025	sp	Q8BFQ8	PDDC1_MOUSE	55.56	216	92	2	60	704	8	220	2E-70	230
Phel_contig_10030	sp	Q5RA95	NDRG3_PONAB	44.7	264	139	1	2409	1618	69	325	3E-69	240
Phel_contig_10031	sp	Q8LEV3	Y2060_ARATH	23.11	251	142	9	254	916	344	573	5E-07	57
Phel_contig_10034	sp	B9EKI3	TMF1_MOUSE	45.89	717	354	7	3458	1344	392	1086	2E-161	521
Phel_contig_10039	sp	Q5FVB0	ATX10_XENTR	34.36	390	246	5	3	1166	93	474	1E-59	211
Phel_contig_10040	sp	Q6EAS5	GLGB_HORSE	67.75	555	177	1	2143	479	140	692	0	751
Phel_contig_10044	sp	Q13075	BIRC1_HUMAN	27.19	217	140	6	738	91	405	604	5E-12	73.9
Phel_contig_10047	sp	Q12767	K0195_HUMAN	46.72	122	61	2	6	368	1238	1356	1E-18	88.6
Phel_contig_10049	sp	Q54PZ6	Y5897_DICDI	25.82	275	177	7	1752	988	17	284	2E-20	99
Phel_contig_10050	sp	Q5R7A2	ALG1_PONAB	52.27	440	196	6	2351	1056	26	459	2E-150	457
Phel_contig_10053	sp	Q13075	BIRC1_HUMAN	24.71	255	165	5	738	1	405	641	2E-13	76.6
Phel_contig_10054	sp	P55175	Y601_SYN3	40.94	254	143	6	839	93	13	264	7E-54	182
Phel_contig_10058	sp	P20749	BCL3_HUMAN	46.15	247	122	5	44	775	121	359	6E-58	204
Phel_contig_10068	sp	Q8BVE8	NSD2_MOUSE	62.76	333	118	2	3	995	999	1327	1E-127	406
Phel_contig_10069	sp	Q5XXB5	MSH2_CHLAE	65.11	599	202	2	3	1790	334	928	0	780
Phel_contig_10071	sp	A1A546	ISX_MOUSE	53.85	52	24	0	156	1	79	130	4E-11	65.5

<< first < prev 1 [2](#) [3](#) [4](#) [5](#) [next >](#) [last >>](#)



Your datasets  
All datasets  
Shared datasets  
Recent activity... 1  
Recently viewed »

Upload dataset  
New query

## YOUR TOP VIEWED

qDOD Cgigas ...	19
uniprot-review...	19
GPL11353_array	14
filt3_M3	14
UniprotProtNa...	12

## POPULAR TAGS

proteomics	357
pnitzsch	189
oa	169
orbitrap	131
oyster	63
published	62
seaflow	52
protein	51
input	47
swissprot	38
techtrip	34
qdod2	30
bioinformatics	26
ssgcid	24
skyline	24
oceanography	23
blast	21
qdod	18
rna-seq	16

**Phel\_uniprot\_sprot\_sql.tab** < Viewable by everyone

Last modified: Jan 5, 2015 2:29 PM sr320@washington.edu

*Click here to add a description* *Click here to add a tag*

```
SELECT * FROM [sr320@washington.edu].[table_Phel_uniprot_sprot_sql.tab7D708]
```

**Actions:** Edit dataset Derive dataset Create snapshot More actions ▾

**DATASET PREVIEW** Rows 1 - 100 of 10513 | Columns 14 of 14

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13	Column14
Phel_contig_100	sp	Q16513	PKN2_HUMAN	81.33	332	61	1	1	1	1	1	1	1
Phel_contig_1000	sp	Q8R4U2	PDIA1_CRIGR	53.62	442	201	2	2	2	2	2	2	2
Phel_contig_10006	sp	Q9Y2H9	MAST1_HUMAN	70.93	289	82	1	1	1	1	1	1	1
Phel_contig_10021	sp	Q96MU7	YTDC1_HUMAN	60.85	212	82	1	1	1	1	1	1	1
Phel_contig_10025	sp	Q8BFQ8	PDDC1_MOUSE	55.56	216	92	2	2	2	2	2	2	2
Phel_contig_10030	sp	Q5RA95	NDRG3_PONAB	44.7	264	139	1	1	1	1	1	1	1
Phel_contig_10031	sp	Q8LEV3	Y2060_ARATH	23.11	251	142	9	9	9	9	9	9	9
Phel_contig_10034	sp	B9EKI3	TMF1_MOUSE	45.89	717	354	7	7	7	7	7	7	7
Phel_contig_10039	sp	Q5FVB0	ATX10_XENTR	34.36	390	246	5	5	5	5	5	5	5
Phel_contig_10040	sp	Q6EAS5	GLGB_HORSE	67.75	555	177	1	1	1	1	1	1	1
Phel_contig_10044	sp	Q13075	BIRC1_HUMAN	27.19	217	140	6	6	6	6	6	6	6
Phel_contig_10047	sp	Q12767	K0195_HUMAN	46.72	122	61	2	2	2	2	2	2	2
Phel_contig_10049	sp	Q54PZ6	Y5897_DICDI	25.82	275	177	7	7	7	7	7	7	7
Phel_contig_10050	sp	Q5R7A2	ALG1_PONAB	52.27	440	196	6	6	6	6	6	6	6
Phel_contig_10053	sp	Q13075	BIRC1_HUMAN	24.71	255	165	5	5	5	5	5	5	5
Phel_contig_10054	sp	P55175	Y601_SYNY3	40.94	254	143	6	6	6	6	6	6	6
Phel_contig_10058	sp	P20749	BCL3_HUMAN	46.15	247	122	5	5	5	5	5	5	5
Phel_contig_10068	sp	Q8BVE8	NSD2_MOUSE	62.76	333	118	2	2	2	2	2	2	2
Phel_contig_10069	sp	Q5XXB5	MSH2_CHLAE	65.11	599	202	2	2	2	2	2	2	2
Phel_contig_10071	sp	A1A546	ISX_MOUSE	53.85	52	24	0	0	0	0	0	0	0

**Actions:** << first < prev 1 2 3 4 5 next > last >>

DERIVE file0 close x

```
SELECT * FROM [sr320@washington.edu].[Phel_uniprot_sprot_sql]
```

**Actions:** Execute query

No table available!

## DERIVE

close x

```
SELECT * FROM [sr320@washington.edu].[Phel_uniprot_sprot_sql].  
left join  
[sr320@washington.edu].[uniprot-reviewed_wGO_010714]des  
on phel.Column3=des.Entry
```

Execute query

DATASET PREVIEW Rows 1 - 100 of 100 or more | Columns 27 of 27

Column1	Column2	Column3	Column4	Column5
Phel_contig_11186	sp	Q5RCU0	TF2AA_PONAB	77.94
Phel_contig_11111	sp	Q9Y2B5	VP9D1_HUMAN	44.91
Phel_contig_1111	sp	P80146	SEPR_THESR	45.51

Column13	C
5E-162	5
5E-146	4
6E-132	4
1E-73	2
2E-70	2
3E-69	2
5E-07	5
2E-161	5
1E-59	2
0	7
5E-12	7
1E-18	8
2E-20	9
2E-150	4
2E-13	7
7E-54	1
6E-58	2
1E-127	4
0	7
4E-11	6

more actions ▾

- Your datasets
- All datasets
- Shared datasets
- Recent activity...
- Recently viewed »

 Upload dataset

[New query](#)

YOUR TOP VIEWED

uniprot-review... 19

gDOD Gaigas ... 18

GPL11353\_array 14

filt3\_M3 14

UniprotProtNa... 12

## Start a New Query

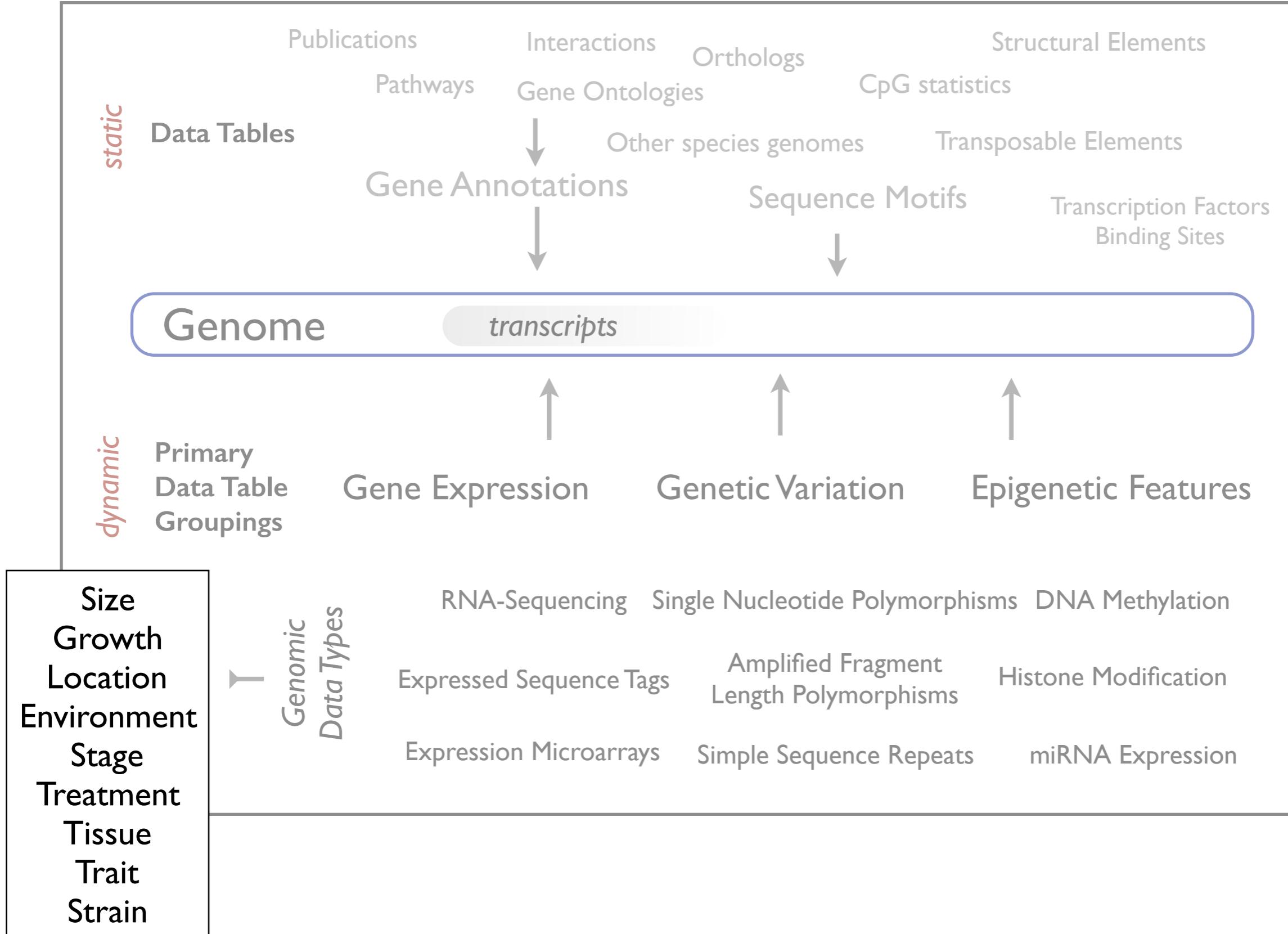
**SQL Query** (*E.g. select \* from table*)

```
SELECT * FROM [sr320@washington.edu].[Phel_uniprot_sprot_sql.tab]phel
left join
[sr320@washington.edu].[uniprot-reviewed_wGO_010714]des
on phel.Column3=des.Entry
```

[Run query](#) [Save query](#) [Download results](#)

DATASET PREVIEW Rows 1 - 100 of 100 or more | Columns 27 of 27

Column12	Column13	Column14	Entry	Entry name	Gene ontology IDs	Interacts with	Cross-reference (GO)	Gene ontology (GO)	Status	InterPro	Pathway	Protein names	Gene names	Organism
5e-162	537	Q16513	PKN2_HUMAN		GO:0005524; GO:0043297; GO:0043296; GO:0006915; GO:0007155; GO:0007049; GO:0051301; GO:0032154; GO:0005737; GO:0005856; GO:0010631; GO:0042826; GO:0030027; GO:0016020; GO:0030496; GO:0005634; GO:0032467; GO:0045931; GO:0004697; GO:0004674; GO:2000145; GO:0006355; GO:0007165; GO:0006351	Q15118		ATP binding; apical junction assembly; apical junction complex; apoptotic process; cell adhesion; cell cycle; cell division; cleavage furrow; cytoplasm; cytoskeleton; epithelial cell migration; histone deacetylase binding; lamellipodium; membrane; midbody; nucleus; positive regulation of cytokinesis; positive regulation of mitotic cell cycle; protein kinase C activity; protein serine/threonine kinase activity; regulation of cell motility; regulation of transcription, DNA-dependent; signal transduction; transcription, DNA-dependent	reviewed	IPR000961; IPR000008; IPR011072; IPR011009; IPR017892; IPR000719; IPR017441; IPR002290; IPR008271;		Serine/threonine-protein kinase N2 (EC 2.7.11.13) (PKN gamma) (Protein kinase C-like 2) (Protein-kinase C-related kinase 2)	PKN2 PRK2 PRKCL2	Homo sapiens (Human)
5e-146	464	Q8R4U2	PDIA1_CRIGR		GO:0045454; GO:0005788; GO:0006662; GO:0042470; GO:0005886; GO:0003756; GO:0015035; GO:0006457			cell redox homeostasis; endoplasmic reticulum lumen; glycerol ether metabolic process; melanosome; plasma membrane; protein disulfide isomerase activity; protein disulfide oxidoreductase activity; protein folding	reviewed	IPR005788; IPR005792; IPR005746; IPR012336; IPR017937; IPR013766;		Protein disulfide-isomerase (PDI) (EC 5.3.4.1) (Prolyl 4-hydroxylase subunit beta) (p58)	P4HB PDIA1	Cricetulus griseus (Chinese hamster) (Cricetulus barabensis griseus)
6e-132	415	Q9Y2H9	MAST1_HUMAN		GO:0005524; GO:0005737; GO:0005856; GO:0007010; GO:0007243; GO:0000287; GO:0005886;	P00533		ATP binding; cytoplasm; cytoskeleton; cytoskeleton organization; intracellular protein kinase cascade; magnesium ion binding; plasma membrane; protein serine/threonine kinase	reviewed	IPR000961; IPR011009; IPR015022; IPR023142; IPR001478; IPR000719; IPR002290;		Microtubule-associated serine/threonine-protein kinase 1 (EC 2.7.11.1) (Syntrphin-associated serine/threonine-protein kinase)	MAST1 KIAA0973 SAST	Homo sapiens (Human)



```
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?



## Fasta2Slim

This IPython notebook is intended to serve as a structured means to annotate sequences using UniProt/SwissProt database. The notebook can be easily modified to personal preferences. As developed, the notebook requires the user has the following software installed ...

- IPython
  - NCBI Blast
  - SQLShare Python Client

## **Instructions for use.**

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

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
```

```
>PiuraChilensis_v1_contig_1  
ATTTACAATACGAAGTAAAATAGATAACGTAAAATAATCTGGTGCTGGATGATCGATC  
AAGTTCAACCAATATTTATTGTAAAAAAATCATTCTAACAGCATGAAATCGTGTACAATG  
TATAAACAAAGCAAATATATAACACTAAAGCAAGAGGGCGTAAGTGGGGGGTGGGTGAGA  
GTAAAAAAATTCAAACATGTCAAATACCCCGGCCTAGCCTTAAAGCACCATGGACTTCT  
CCCTTGAAATACGATTAATTAAGACCTTATGACGATTAATGACGATTAACGAC
```

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
```

```
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'
```

## 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')
```

