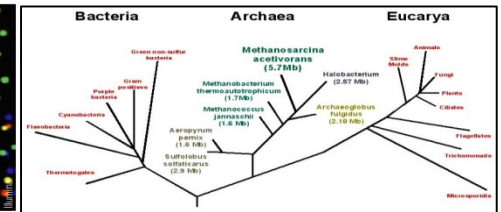
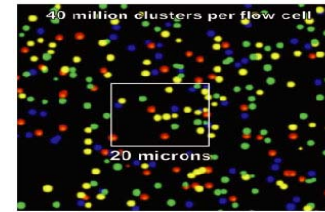




TAACCCTAACCCTAACCCTAACCCTAACCCTA  
 CCTAACCCTAACCCTAACCCTAACCCTAACC  
 CCTAACCCTAACCCTAACCCTAACCCTAAC  
 AACCCTAACCCTAACCCTAACCCTAACCCTA  
 ACCCTAACCCTAACCCTAACCCTAACCCTAAC  
 CTACCCTAACCCTAACCCTAACCCTAACCCTA  
 ACCCTAACCCTAACCCTAACCCTAACCCTAA

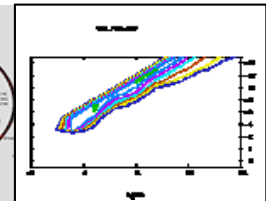
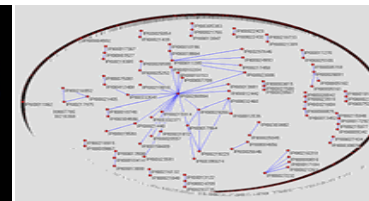
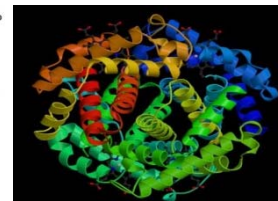
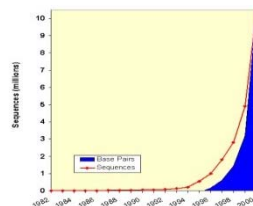
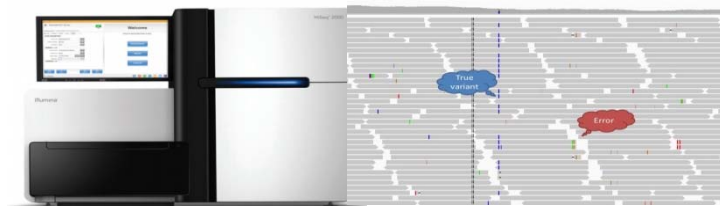


# Case Study 2: From Dry to Wet, an Evolutionary Story

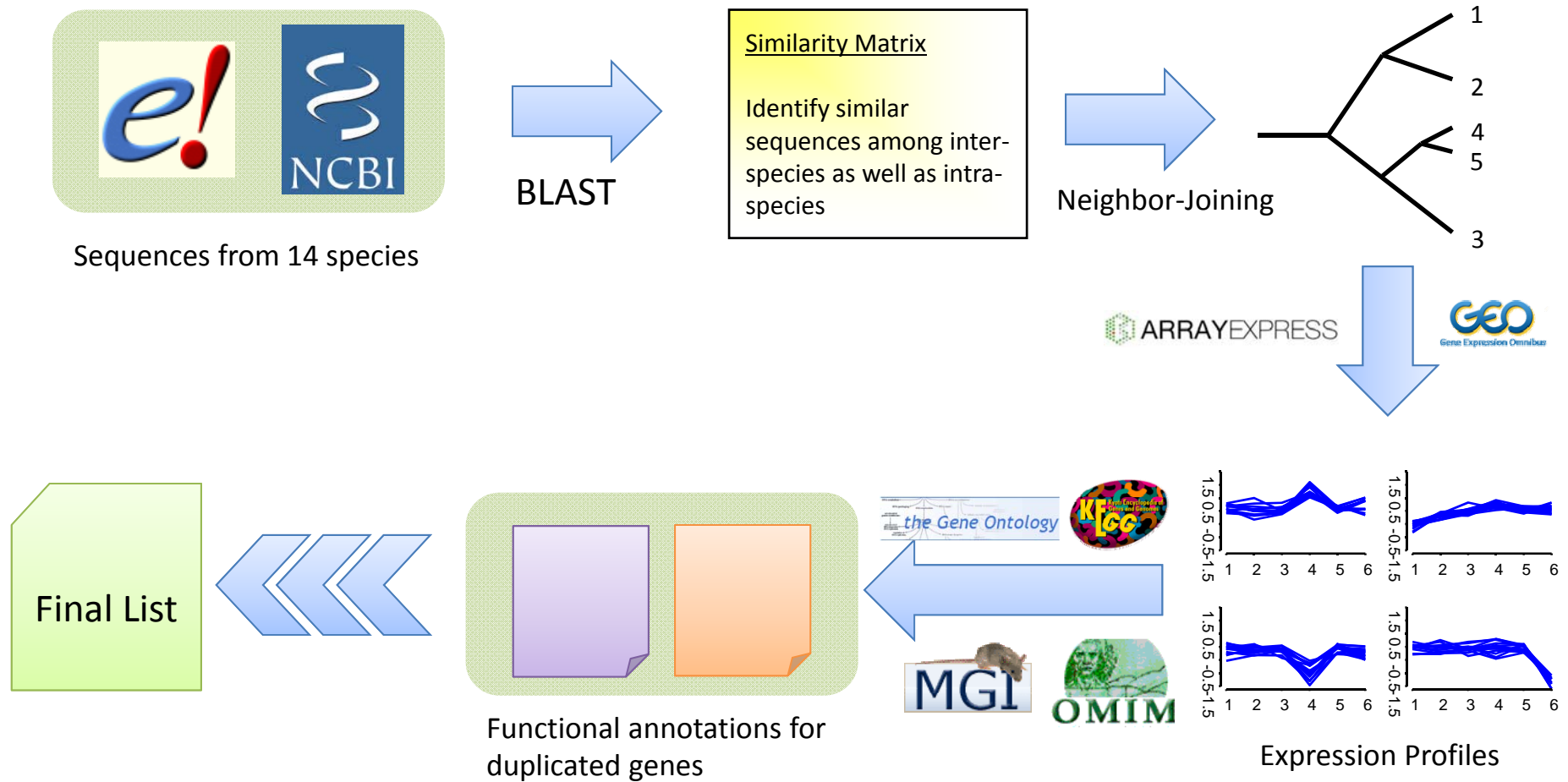
北京大学生物信息学中心 高歌

Ge Gao, Ph.D.

Center for Bioinformatics, Peking University



## Computationally screening for function divergent genes involved in early development regulation

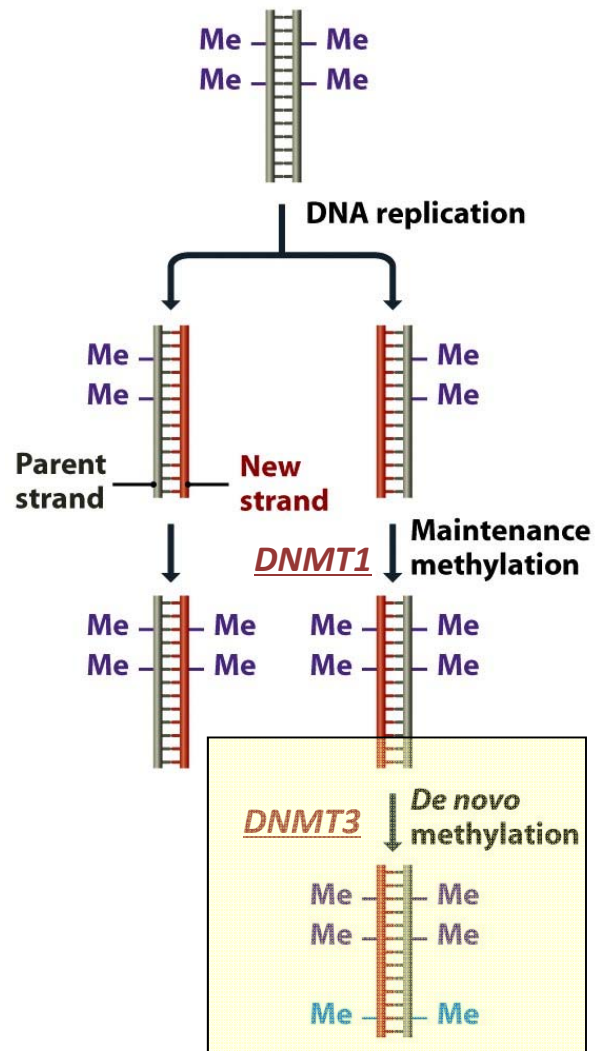


## Computationally screening for function divergenced genes involved in early development regulation

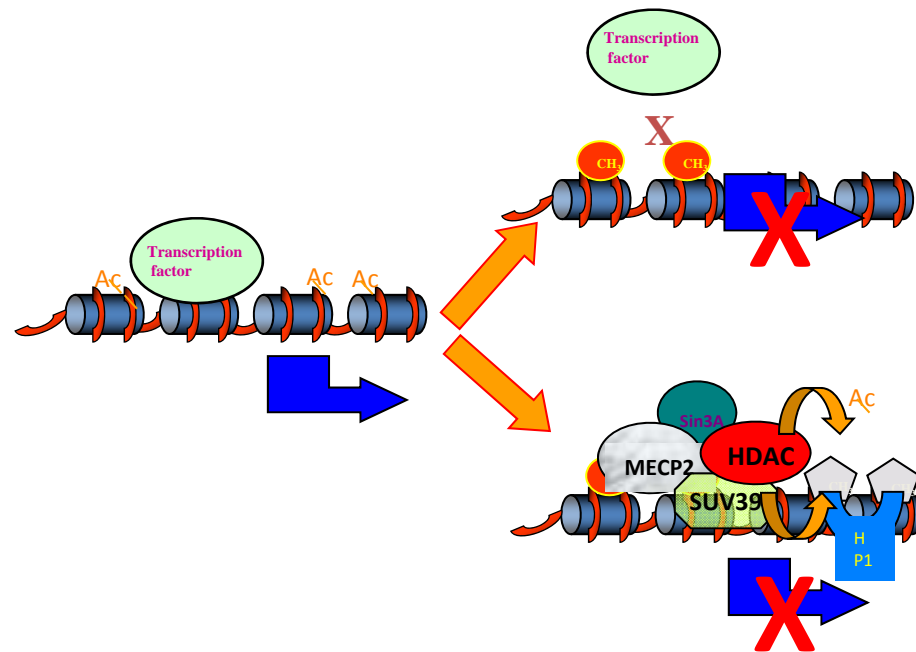
### Computational Genomic Analysis and Bioinformatics through MAPK:

1. Sequence databases constructed directly from the Ensembl website;
2. Each peptide sequence in the database used to search the database using BLAST package.
3. Phylogenetic trees were constructed and paralogous pairs are identified from the resulting alignments based on a minimal amino acid identity (e.g. 50% and 70%) and an overlap of  $\geq 35$  amino acids in the region of local alignment.
4. Coding regions of pairs that meet these criteria will be aligned with the corresponding region and inspected for putative function divergence hallmarks.
5. Local warehouse were searched for further indicators derived from high-throughput data (esp. genetic, genomic, transcriptomic, proteomic and pathway data).

**7 out of 50000+ new paralogous pairs showed clear functional divergence features involved in early development regulation.**



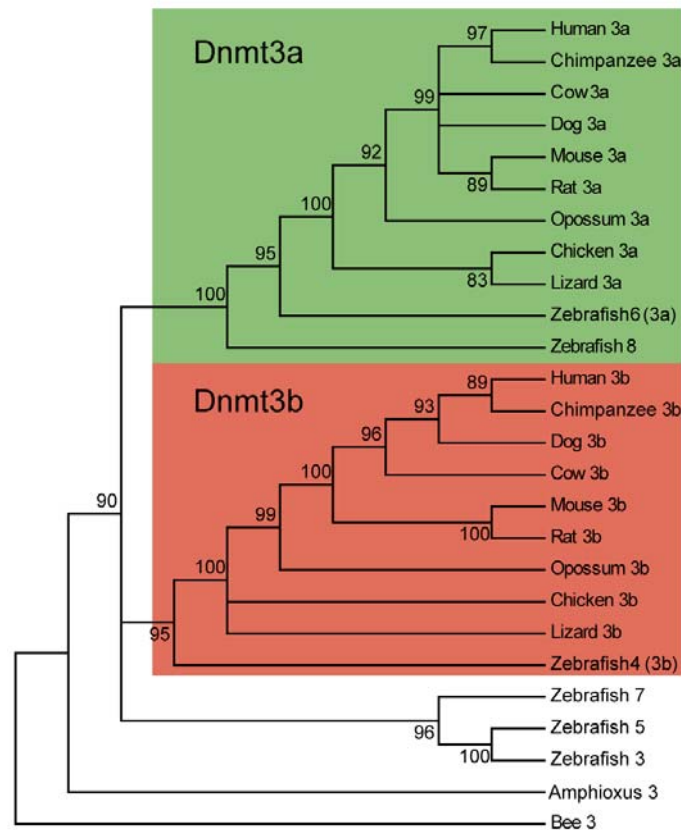
DNA methylation silences gene expression by two mechanisms



(Modified from Moshe Szyf's slide)



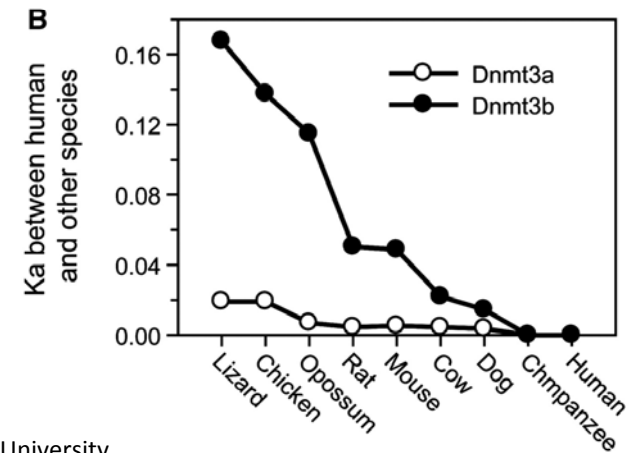
# DNMT3 raised around the separation of vertebrates



(Shen *et al. Nucle Acids Res.* 2010)

**A**

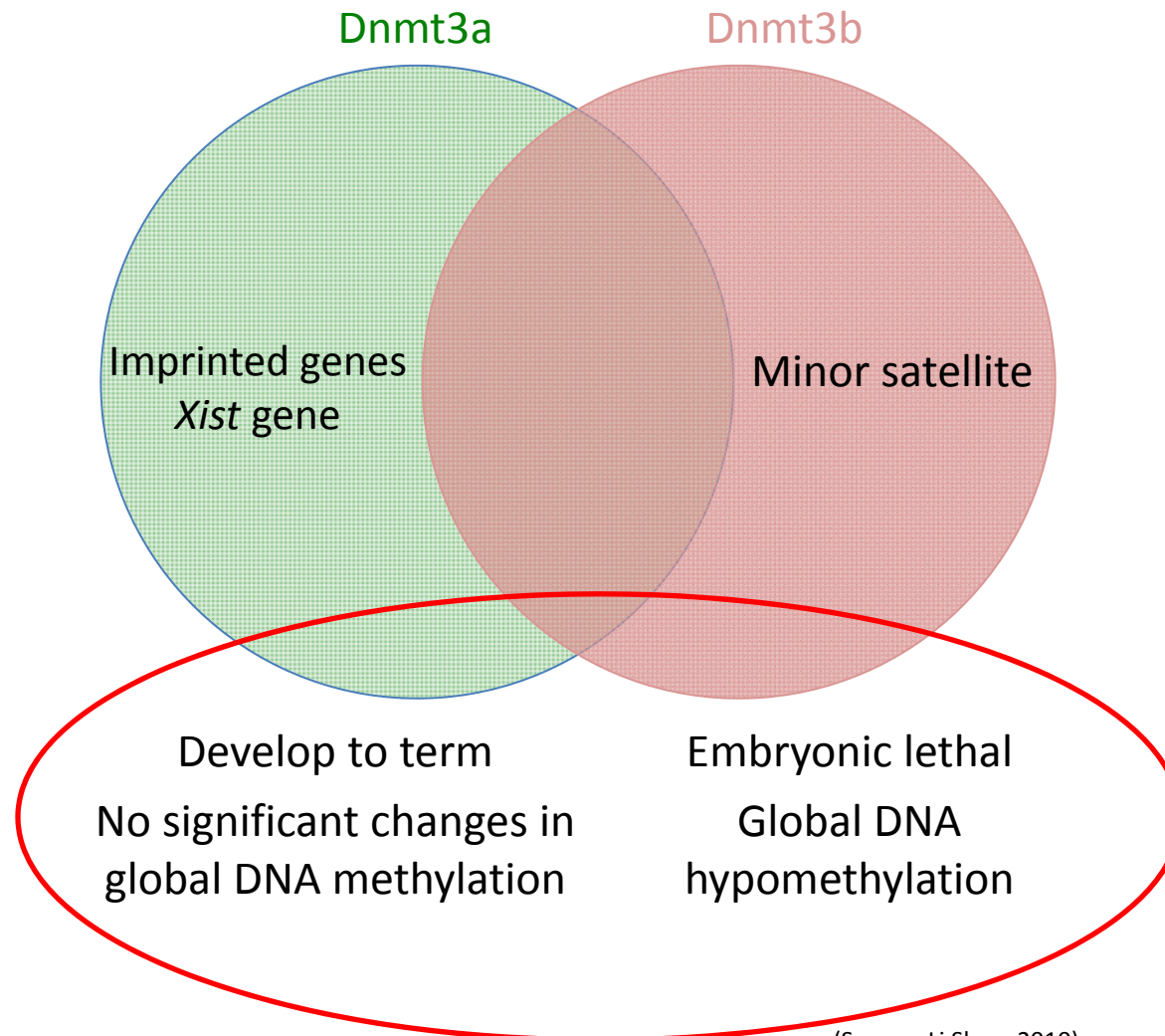
	Synonymous			Non-Synonymous		
	S <sub>3a</sub>	S <sub>3b</sub>	P-value	N <sub>3a</sub>	N <sub>3b</sub>	P-value
Human	221	228	0.377	294	343	0.025*
Chimpanzee	222	227	0.413	294	343	0.025*
Dog	227	236	0.340	298	342	0.041*
Cow	230	222	0.366	299	347	0.029*
Mouse	220	228	0.342	300	345	0.040*
Rat	227	229	0.460	298	340	0.049*
Opossum	247	257	0.319	298	357	0.011*
Chicken	226	208	0.196	307	350	0.047*
Lizard	231	229	0.463	305	353	0.031*



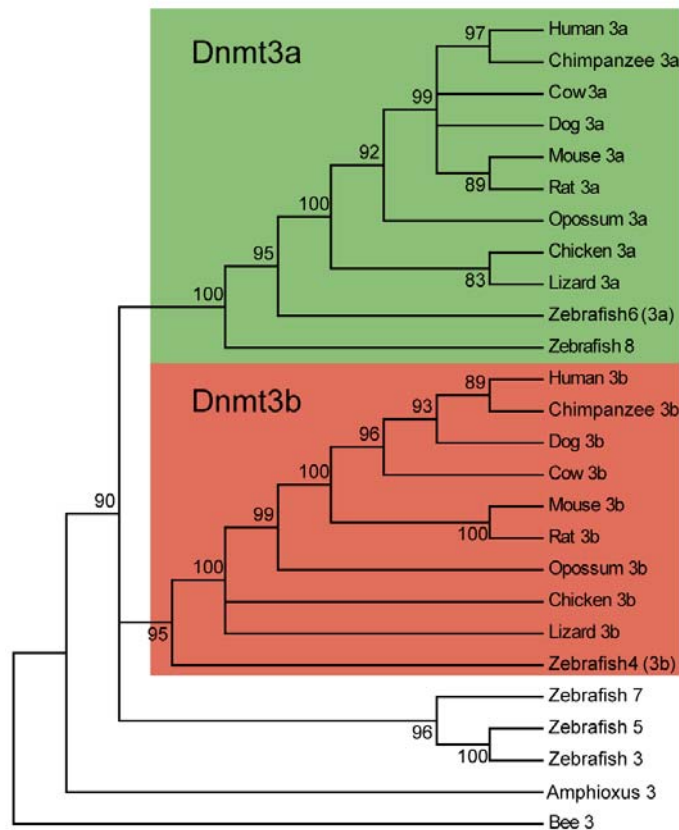
Copyright © Peking University

Specific  
targets

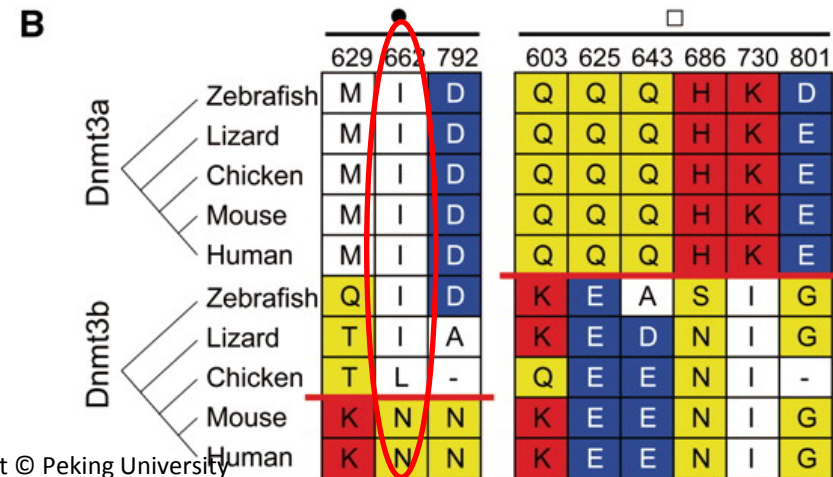
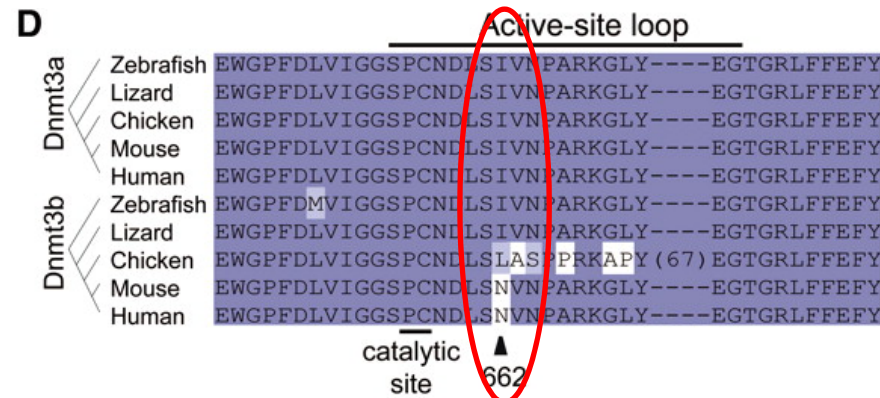
Phenotype  
of KO mice



# A mammalian DNMT3b-specific amino acid change appeared near catalytic site

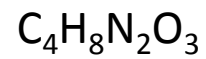
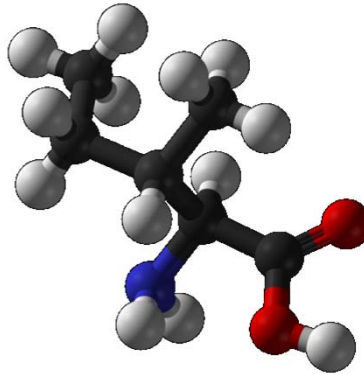


(Shen *et al. Nucleic Acids Res.* 2010)



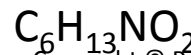
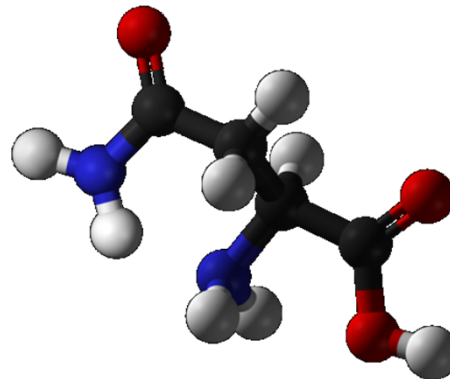
Copyright © Peking University

**Isoleucine (I)**  
异亮氨酸



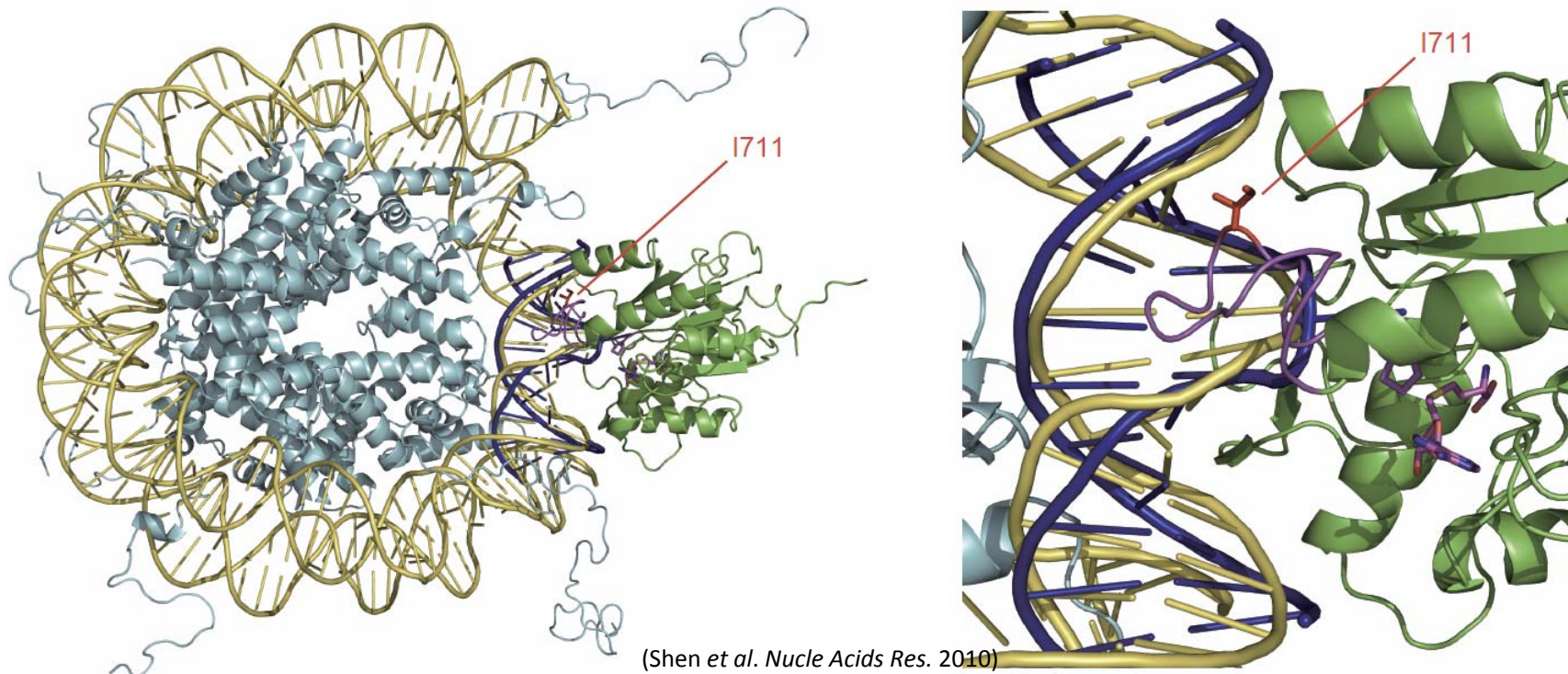
- non-Polar,
- Hydropathy index = 4.5
- pI at 25° C = 6.04

**Asparagine (N)**  
天冬酰胺



- Polar,
- Hydropathy index = -3.5
- pI at 25° C = 10.76





Structural analysis suggested that the I→N results in a tighter enzyme-DNA interaction

*Observation*  
I→N in mammalian Dnmt3b



Tighter interaction with phosphate group  
of DNA backbond



*Hypothesis*

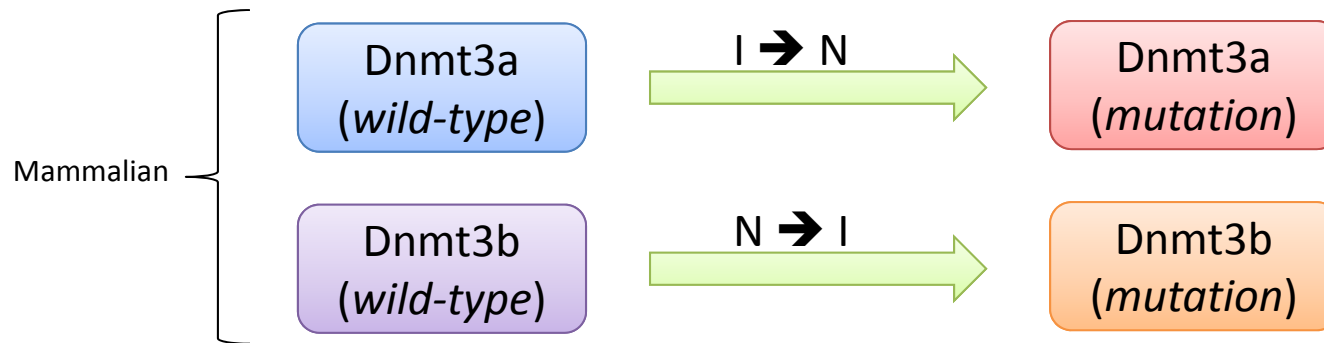
- 1) Higher methylation activity in **mammalian Dnmt3b** than of **Dnmt3a** and **non-mammalian Dnmt3b**
- 2) which is the direct result of changed physicochemical property.



Dr. Li Shen  
Research Fellow, Harvard

**Test!**

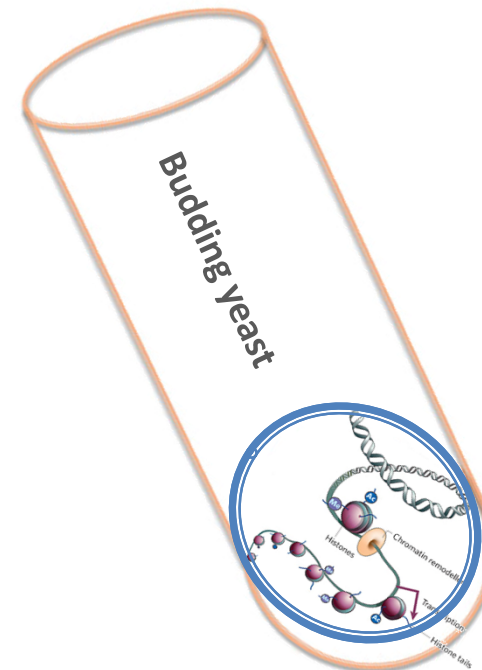
# Test the hypothesis with wet experiences

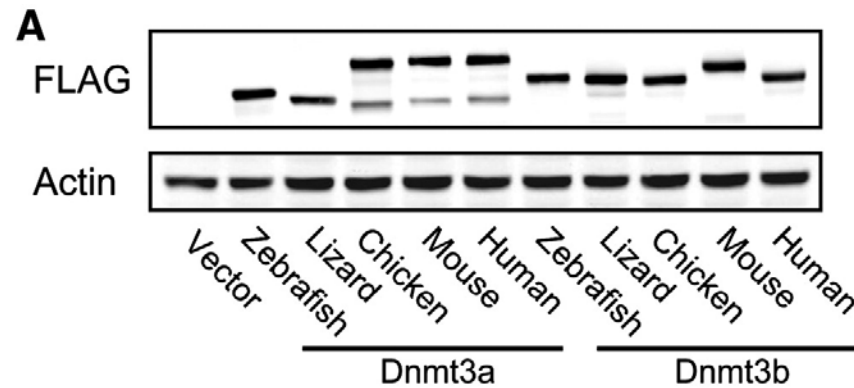


*"Site-directed mutagenesis is a molecular biology method that is used to make specific and intentional changes to the DNA sequence of a gene and any gene products. Also called site-specific mutagenesis or oligonucleotide-directed mutagenesis, it is used for investigating the structure and biological activity of DNA, RNA, and protein molecules, and for protein engineering. With decreasing costs of oligonucleotide synthesis, artificial gene synthesis is now occasionally used as an alternative to site-directed mutagenesis." (Source: wikipedia.org)*

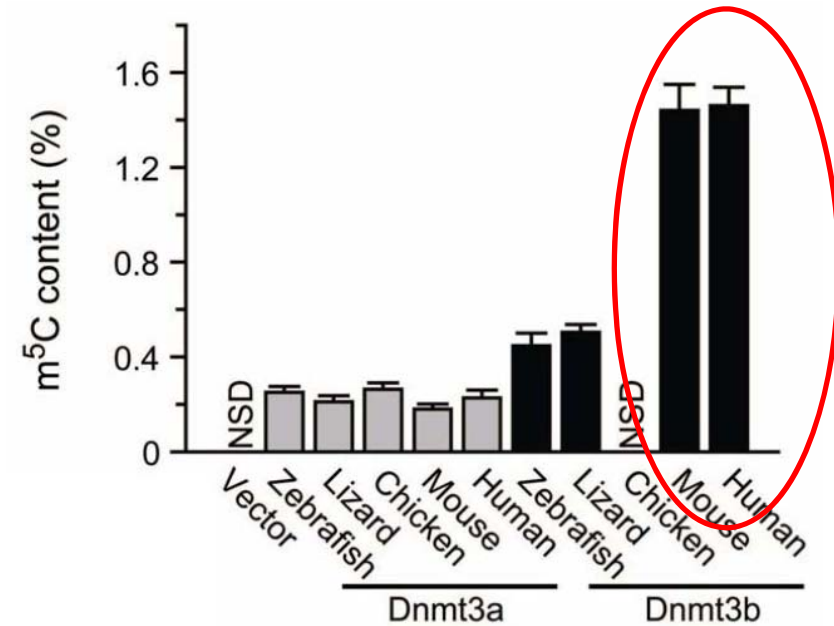
# Using budding yeast as the "*in vivo* test tube"

Epigenetic features	mammals	budding yeast
Chromatin	Yes	Yes
Histone acetylation	Yes	Yes
H3K4	Yes	Yes
H3K36	Yes	Yes
H3K79	Yes	Yes
SWI/SNF complexes	Yes	Yes
CHD1 ATPase	Yes	Yes
SWR1 ATPase	Yes	Yes
ISWI ATPase	Yes	Yes
Endogenous methylation	Yes	No





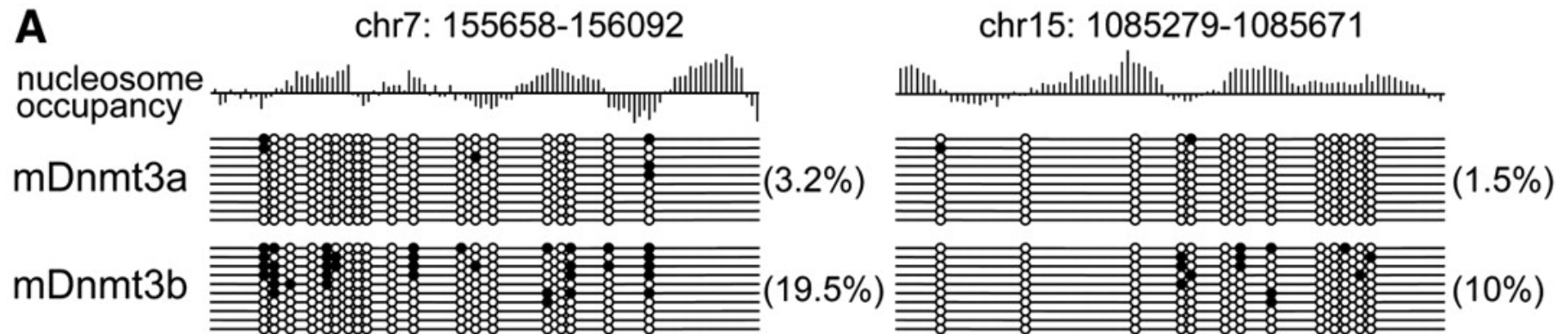
(Shen *et al.* *Nucle Acids Res.* 2010)



Mammalian Dnmt3b possesses higher chromatin DNA methylation activity than Dnmt3a and non-mammalian Dnmt3b



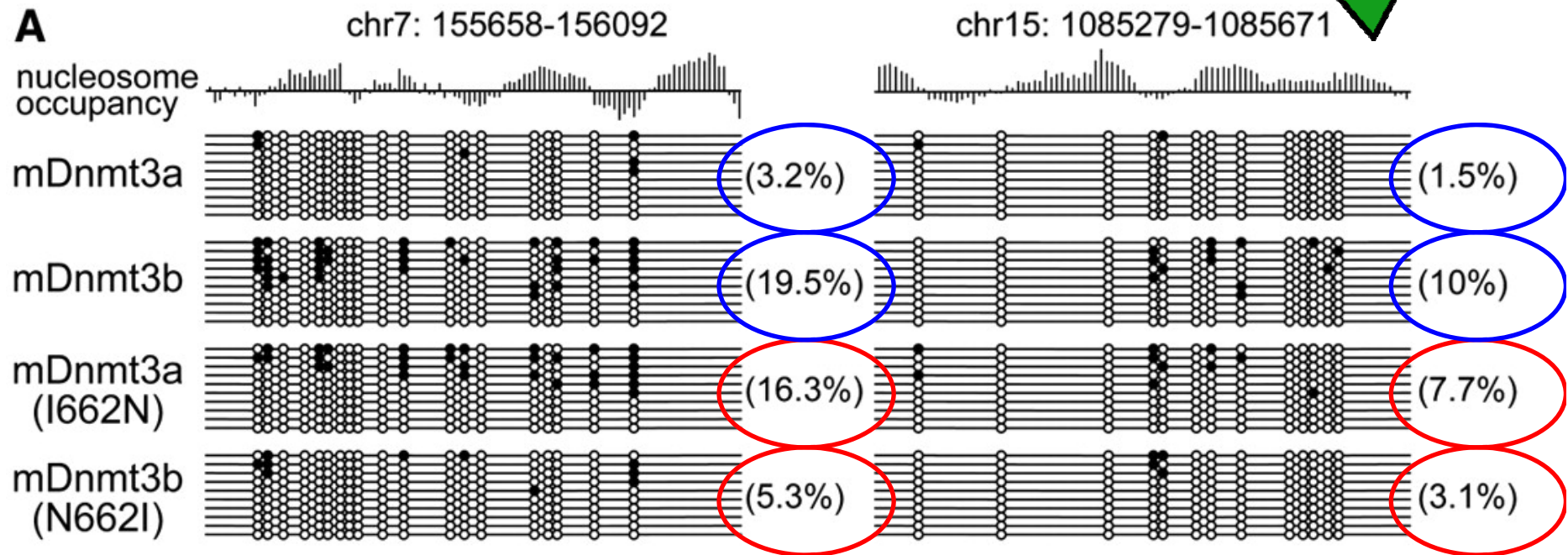
Would I662N substitution accounts for the increased nucleosome DNA methylation activity in mammalian DNMT3b?



(Shen *et al. Nucle Acids Res.* 2010)

Copyright © Peking University

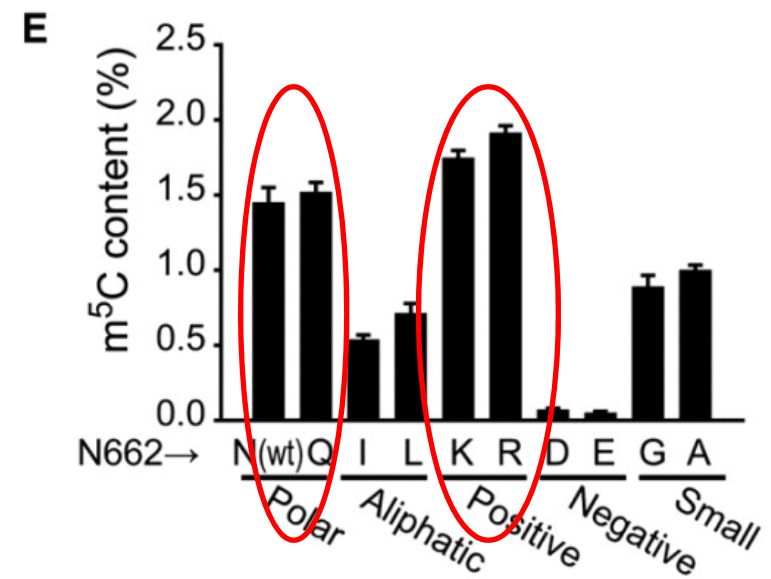
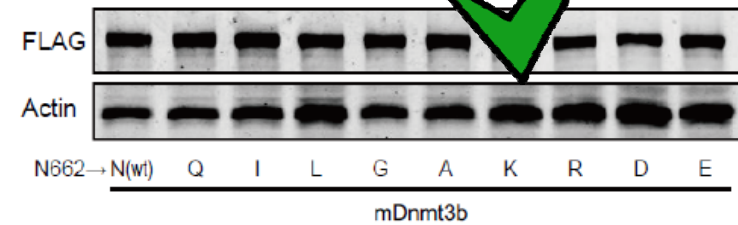
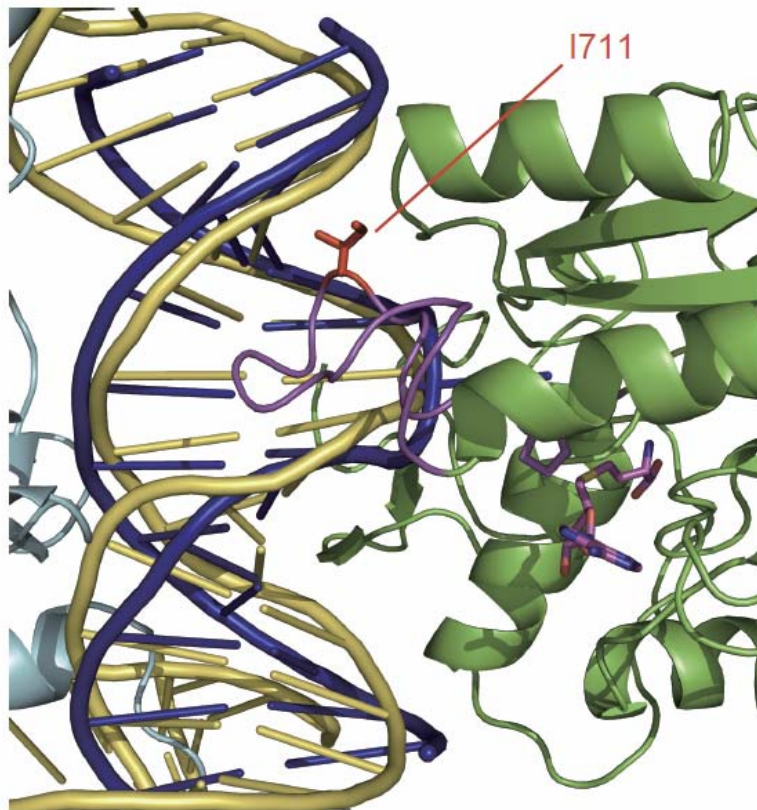
Would I662N substitution accounts for the increased nucleosome DNA methylation activity in mammalian DNMT3b?



(Shen *et al. Nucle Acids Res.* 2010)

Copyright © Peking University

Does the changed physicochemical property matter?



*Observation*

I→N in mammalian Dnmt3b

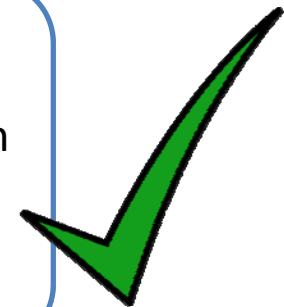


Tighter interaction with phosphate group  
of DNA backbond



*Hypothesis*

- 1) Higher methylation activity in **mammalian Dnmt3b** than of **Dnmt3a** and **non-mammalian Dnmt3b**
- 2) which is the direct result of changed physicochemical property.



*Observation*

I→N in mammalian Dnmt3b



Tighter interaction with phosphate group  
of DNA backbond



*Hypothesis*

- 1) Higher methylation activity in mammalian Dnmt3b than of Dnmt3a and non-mammalian Dnmt3b
- 2) which is the direct result of changed physicochemical property.

**Selective Advantage?**



A (hypothesis) connection of the chromatin DNA methylation activity of Dnmt3b with the density of repetitive sequences in the genome?

Percentage of repeats in the genome	Chromatin DNA methylation activity of Dnmt3b
--	---



~ 40-50%

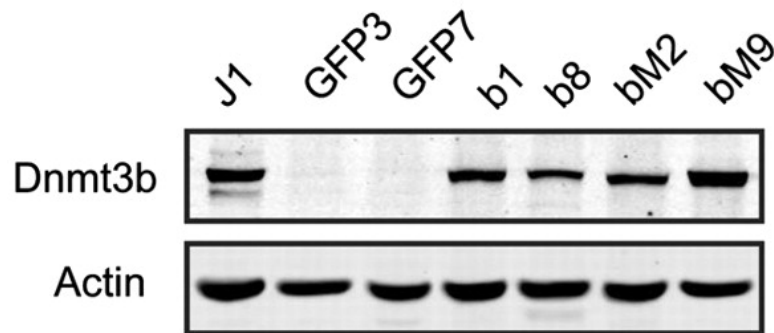
highest



~10%

inactive

(Source: Li Shen, 2010)

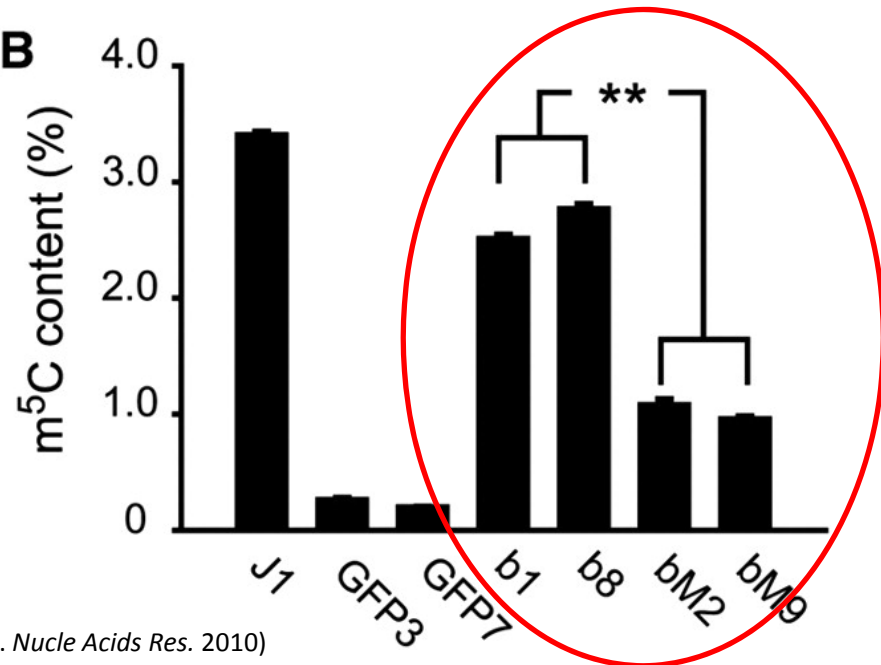
**A**

GFP: GFP3, GFP7

mDnmt3b: b1, b8

mDnmt3b(N662I): bM2, bM9

(Shen *et al.* *Nucle Acids Res.* 2010)

**B**

The substitution (I662N) is crucial for mammalian Dnmt3b to efficiently methylate repetitive sequences in mammalian cells

*Observation*

I→N in mammalian Dnmt3b



Tighter interaction with phosphate group  
of DNA backbond



*Hypothesis*

- 1) Higher methylation activity in **mammalian Dnmt3b** than of **Dnmt3a** and **non-mammalian Dnmt3b**
- 2) which is the direct result of changed physicochemical property.



More  
effectively  
silent  
repetitive  
sequences?

# Summary

- Evolution-guided bioinformatics analysis successfully identified interesting genes involved in early development regulation showed clear functional novelty during evolution, and also provided strong hints for the key substitution, its biochemical effort, and the eventually functional significance.
- Key single substitution could result in significant functional novelty and help novel gene (re-)wired itself into existing circuits.
- An integrated, genome-scale bioinformatic analysis combined with targeted experimental assay is effective in studying complex biological system.

## **Bioinformatics: an interdisciplinary field that develop and apply computer and computational technologies to study biomedical questions**

- As a technology, bioinformatics is a powerful technology to manage, search, and analyze k
- As a methodology, bioinforma driven, genome-wide, and sys new hypotheses, find new pat functional elements.





# 生物信息学：导论与方法

## Bioinformatics: Introduction and Methods



<https://www.coursera.org/course/pkubioinfo>