

---

## Search for correlation between methylation profiles and social status based on data on orphans adopted in the USA

Scientific advisors: Dobrynin P., Komissarov A.

Student: Grigorev K.

# Project premise

---

## Correlation between socioeconomic status and DNA methylation.

- Associations with early-life socio-economic position in adult DNA methylation  
(*Szyf M, McGill University*)
- Profiling of Childhood Adversity-Associated DNA Methylation Changes  
(*Zhang H, **Yale University***)
- Childhood adversity and DNA methylation of genes  
(*Bick J, Naumova O, **Yale University***)

# Project goal and experiment design

---

Detect systematic change in methylation patterns (if any) in orphans moving from Russia to the U.S.

**Core** group:

Russian orphans **adopted** by U.S. families and brought to the U.S.

**Control** group:

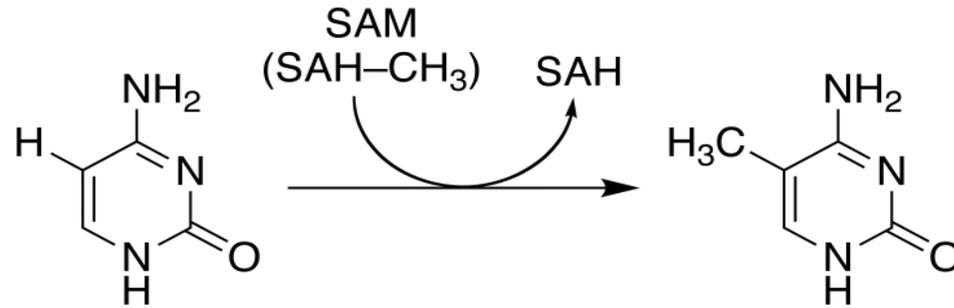
Russian children who moved to the U.S. with their **biological** parents.

Subject age: **toddlers**

Time interval between sequencing: **6 months**

# Cytosine methylation

---

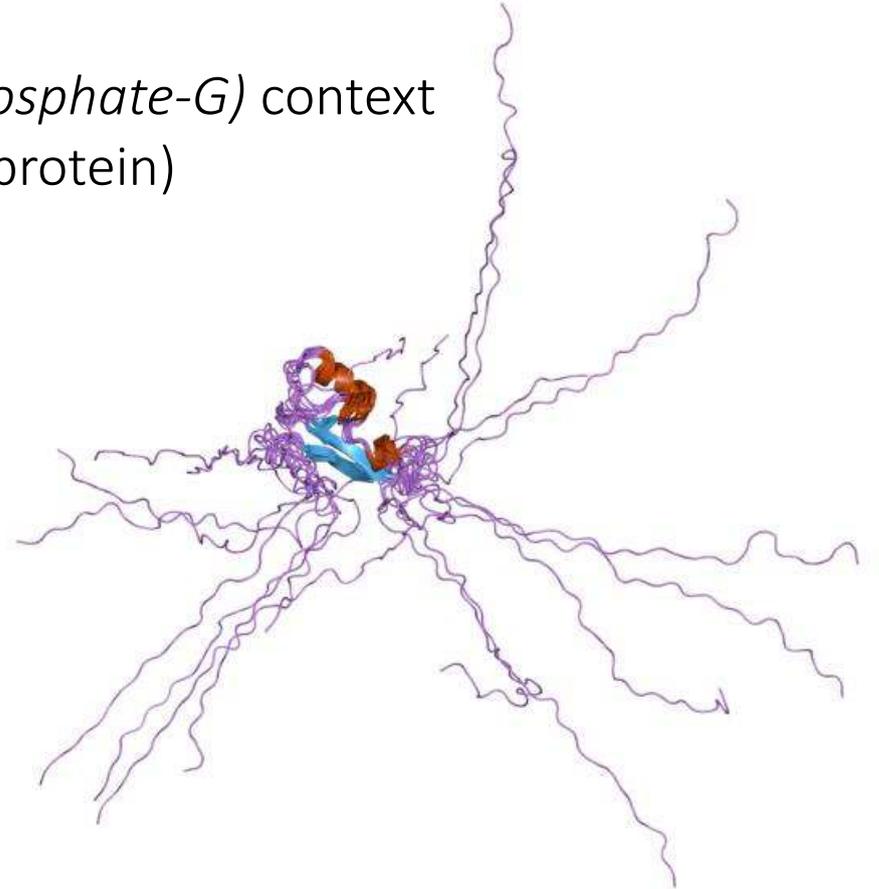


- Typically occurs in CpG (*C-phosphate-G*) context
- Is involved in regulation of gene transcription

# Cytosine methylation: MBD-Seq

---

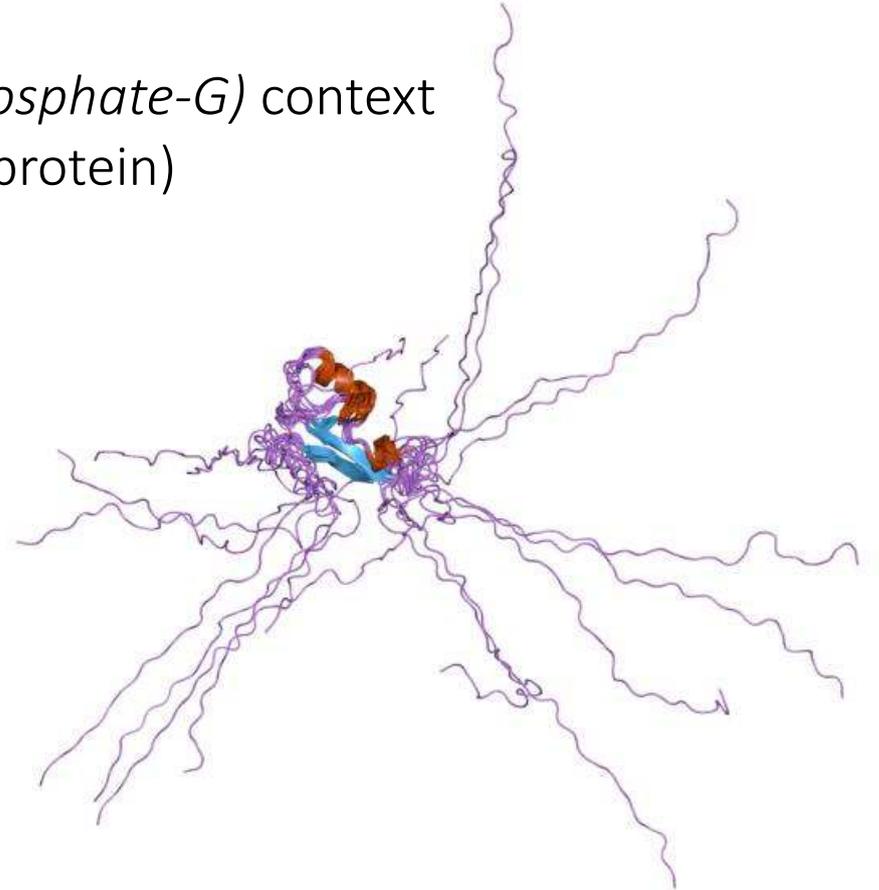
1. Methylation typically occurs in CpG (*C-phosphate-G*) context
2. **MBD** = Methyl-CpG Binding Domain (in a protein)
3. Can use MBD in ChIP-seq
4. ???
5. Methylated region detection



# Cytosine methylation: MBD-Seq

---

1. Methylation typically occurs in CpG (*C-phosphate-G*) context
2. **MBD** = Methyl-CpG Binding Domain (in a protein)
3. Can use MBD in ChIP-seq
4. ~~???~~ Short read alignment, **MACS2**
5. Methylated region detection



# MBD-Seq: what's in the data

---

MACS2 output (*various formats inc. bed*):

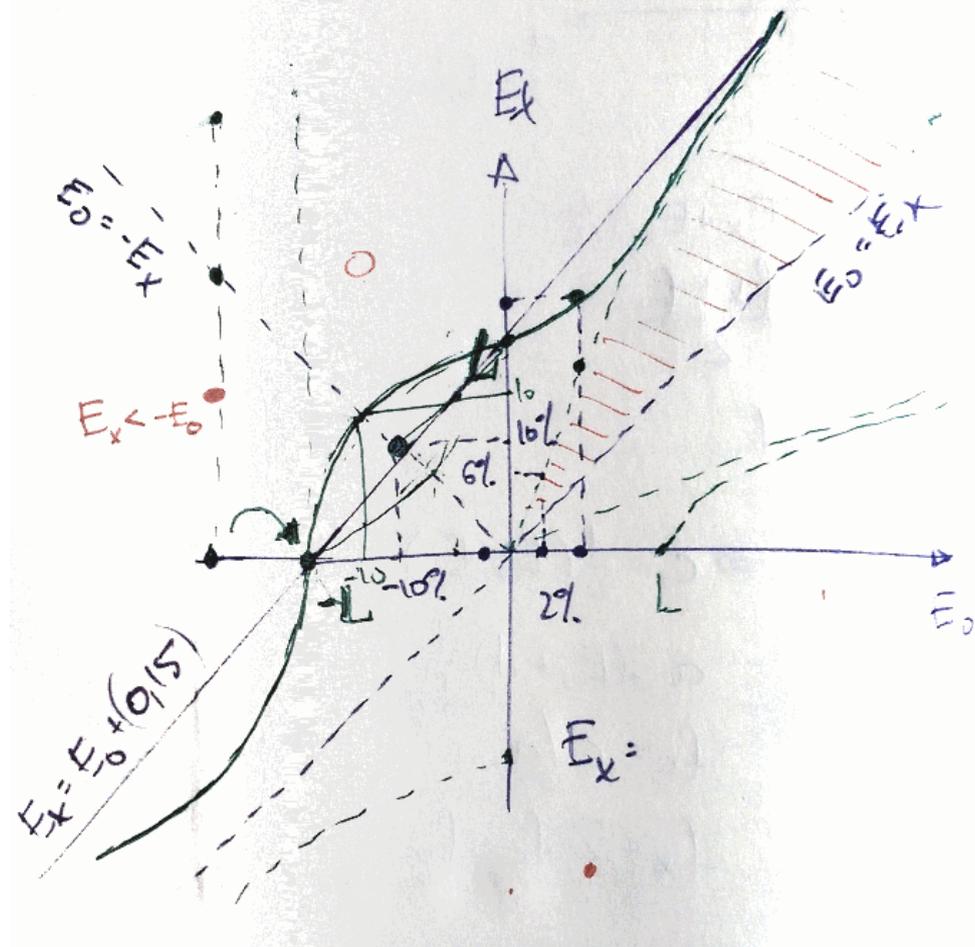
```
chr  start  stop  ...  fold_enrichment  ...
```

General idea:

Find regions where `fold_enrichment` changes:

- in the *same* direction for all the *core* subjects
- in the *different* direction for all the *control* subjects

$\Delta_{\text{enrichment}}$ : how different is "different"?



# Finding affected genes

---

- Intersect found regions with:
  - regions of all known genes
  - upstreams of all known genes (shot at promoters)
- Collect statistics:
  - genes with biggest amount of changed regions (*in uniform direction!*)
  - genes with biggest cumulative *significance* of changes

## Examples:

|                 |       |                  |                 |
|-----------------|-------|------------------|-----------------|
| ENSG00000155093 | 15/-9 | inside,          | 1.35778573742   |
| ENSG00000153707 | 8 /-8 | inside,          | 0.513472427213  |
| ENSG00000179242 | 7 /-7 | inside,          | 0.713335805813  |
| ENSG00000196338 | 2 /-2 | upstream,inside, | 1.7097332516    |
| ENSG00000267369 | 2 /-2 | upstream,inside, | 0.950590042702  |
| ENSG00000259009 | 2 /-2 | inside,          | 2.08073993732   |
| ENSG00000204394 | 1 /-1 | inside,          | 0.0405047542049 |

# Affected gene clusters by function

Enter **Gene Ontology**

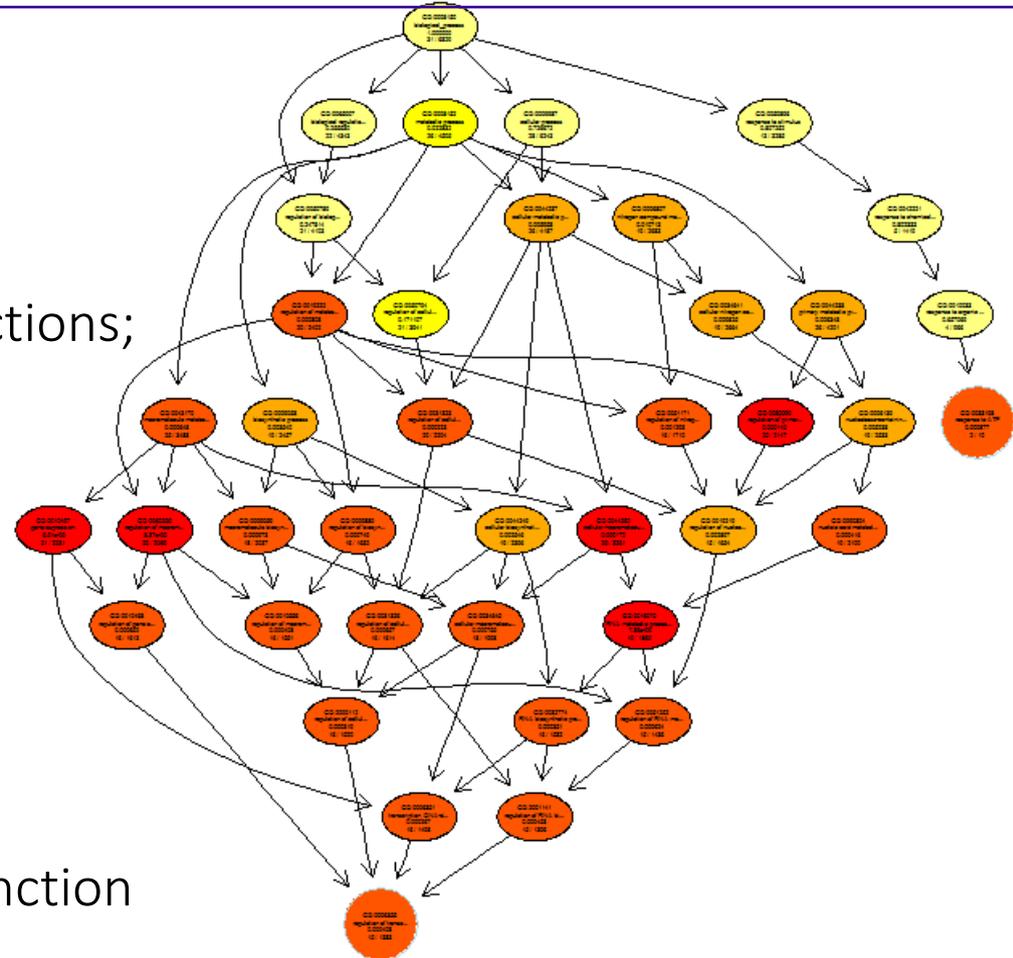
Reports functions for a gene;

Provides an XML file of relations of all functions;

=> we can find full “ancestry” of functions for any gene

=> we can sort functions by frequency in our dataset

=> we can find clusters with interesting function



# Affected gene clusters by function

---

Examples of found clusters:

G0:0032502 (developmental process): 30 genes

G0:0048513 (organ development): 16 genes

G0:0009653 (anatomical structure morphogenesis): 13 genes

G0:0007417 (central nervous system development): 11 genes

Examples of genes:

- G0:0007417 => ENSG00000075213 (**SEMA3A**) [4 regions, 0.21 significance]  
neuronal pattern development
- G0:0032502 => ENSG00000153707 (**PTPRD**) [8 regions, 0.51 significance]  
neurite growth, ADHD / restless legs syndrome
- G0:0048513 => ENSG00000196498 (**NCOR2**) [3 regions, 0.91 significance]  
transcription activity mediator

# Conclusion

---

## Results:

- A pipeline was created for detection of methylation profile changes
- A successful proof-of-concept run was executed on a sample dataset
- Significant change in methylation was found in a number of genes with important function

## Future plans:

- Apply the pipeline to the entire dataset
- Expand the pipeline to suit the needs of further research
- Assess biological significance of obtained results

# Orphans' methylomes

---

Thank you!  
Any questions?