

Mediation of effects of persistent chemicals on the human sperm epigenome

Student: Julia Kornienko

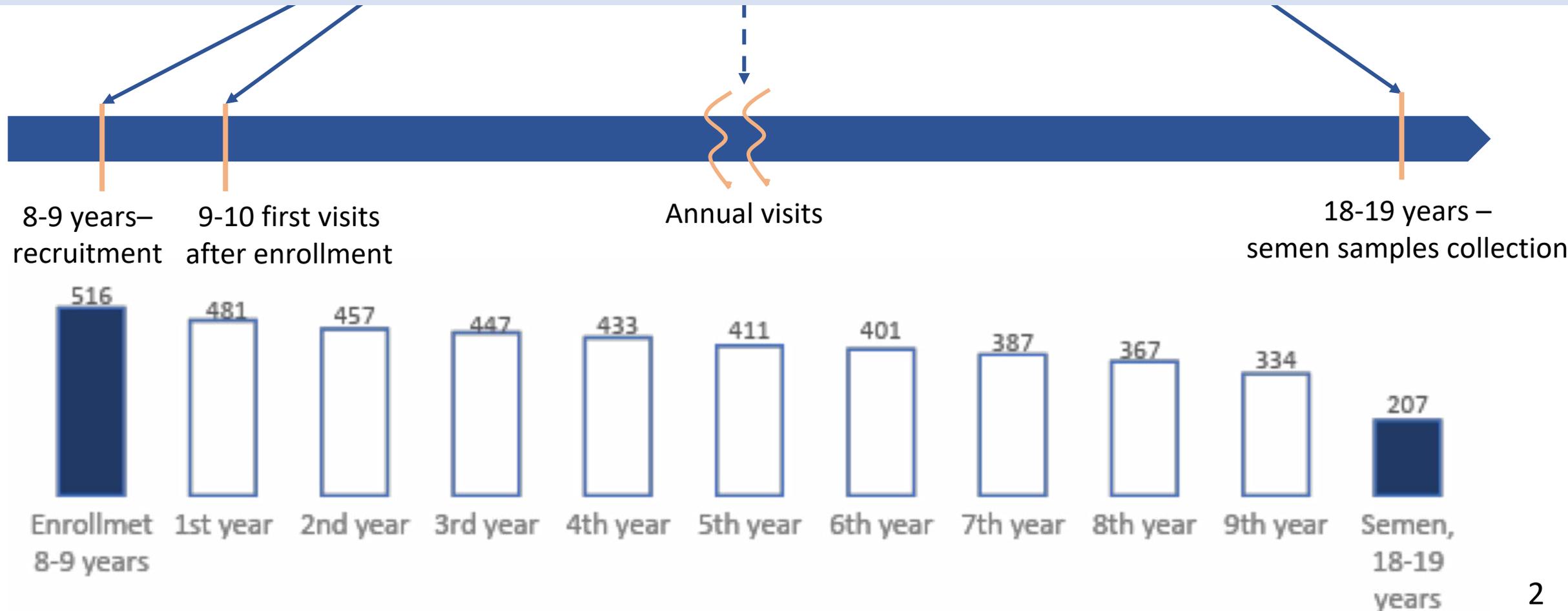
Supervisors: Oleg Sergeev, Yulia Medvedeva;

A.N. Belozersky Research Institute of Physico-Chemical Biology, Moscow State University,

Institute of Bioengineering, Research Center of Biotechnology RAS

The Russian Children's Study (RCS) Male Reproductive Health

Biennially - blood sampling; annually - urine sampling, follow up of growth and puberty, interviewing
=> 20 000+ sample aliquots and 1000+ analyzing parameters in total



Endocrine disrupting chemicals (EDCs)

- are substances that can interfere with any aspect of hormone action at low doses. These disruptions can cause cancer, birth defects, and other developmental disorders as any system in the body controlled by hormones can be affected by exposure to EDCs

- Where do we get them from?

- Products of personal hygiene
- Pesticides
- Medicaments
- Antiseptics
- Plastic
- Air pollution caused by factories
- And other sources

Each person on the planet is being under continuous exposure, which influences many aspects of our physiology including reproductive system

- Why Chapaevsk?

Due to historical production of **organochlorine pesticides**

Dioxins as byproducts

Pesticides, dioxins and PCBs are **EDCs**.

Dioxins – highly toxic by-products of various industrial processes that accumulate in the environment and tissues of animals and people under exposure. They disturb the work of cellular receptors and therefore have many toxic effects.



2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD)

the most toxic dioxin

has known endocrine disrupting properties

What is already known?

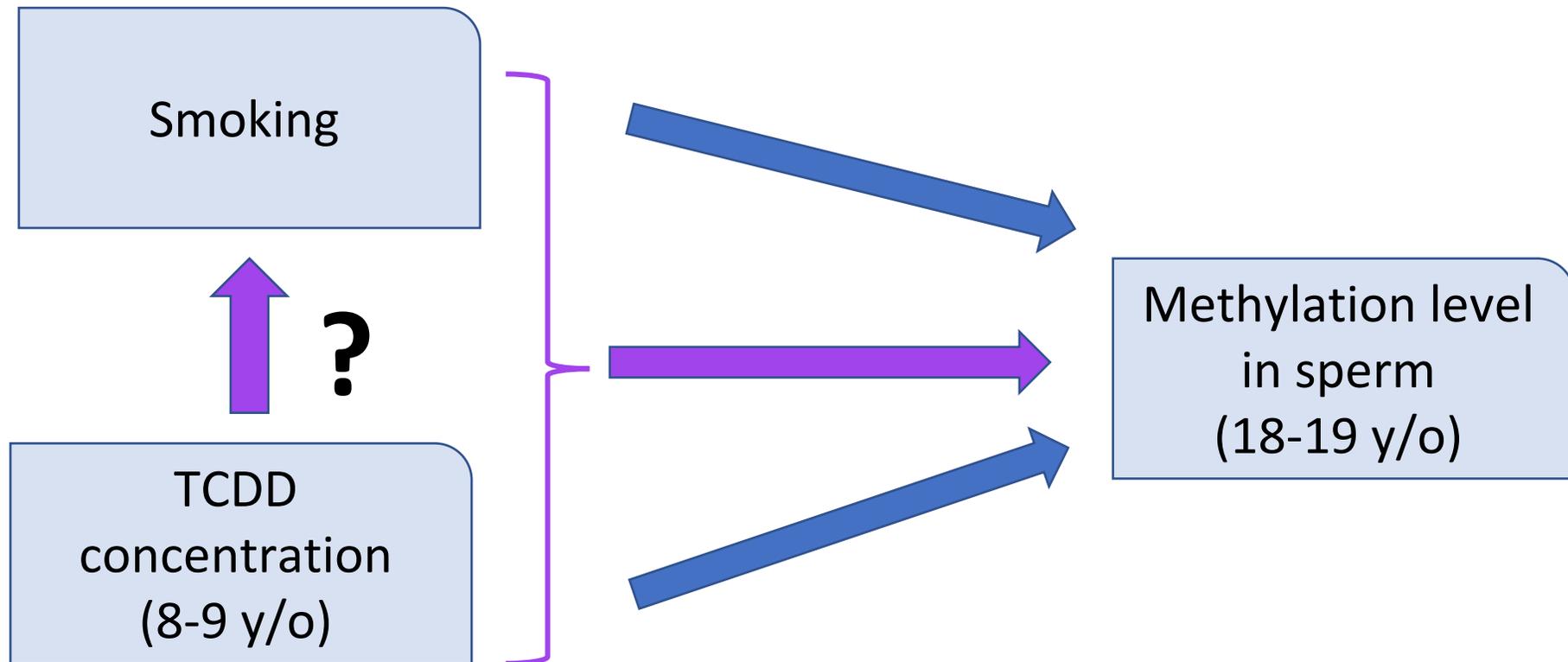
- Peripubertal exposure to TCDD is associated with poorer semen quality (RCS, Minguez-Alarcon et al., 2017)
- 52 differentially methylated regions (DMRs) were identified that distinguished lowest and highest peripubertal serum TCDD concentrations (RCS, Pilsner et al., 2018)

What is needed to be known

- How do other factors influence the methylation level of the human sperm?
- Which factors can mediate the effect of the peripubertal exposure to TCDD?

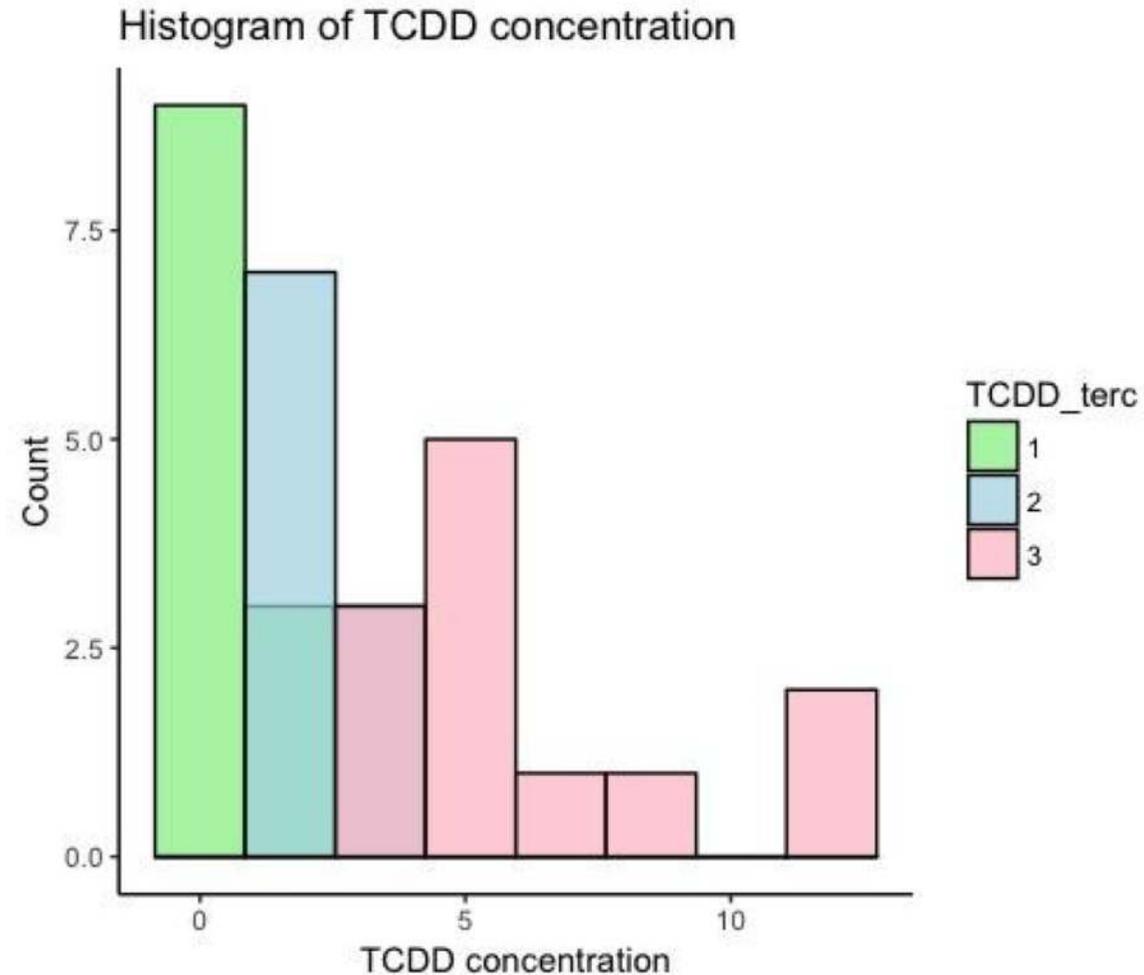
Aim of the project

Analysis of TCDD concentration in prepubertal age and smoking within 6 months before collection of sperm influence on the young adults semen methylation level.



What are the data to analyze within the project:

- 34 samples with different concentrations of TCDD in the blood at the time of enrollment into the study
- Data regarding lifestyle habits of each of 34 chosen participants
- Methylation levels of 2 611 773 CpGs present in at least one of 34 samples



What are the data to analyze within the project:

- 34 samples with different concentrations of TCDD in the blood at the time of enrollment into the study
- Data regarding lifestyle habits of each of 34 chosen participants
- Methylation levels of 2 611 773 CpGs present in at least one of 34 samples

In the LAST year, have you smoked a cigarette, even a few puffs?

Have you smoked at least 100 cigarettes (5 packs) in your life?

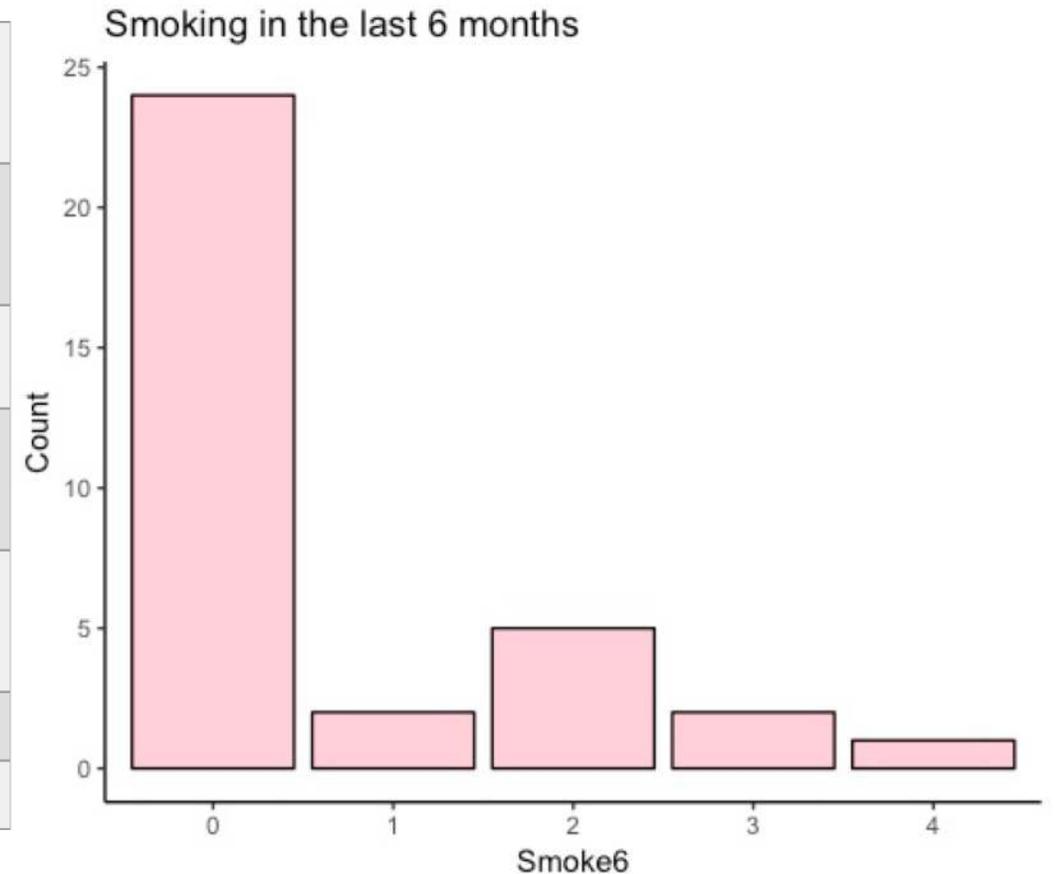
How long ago did you smoke your last cigarette?

When you are smoking how many cigarettes per month do you smoke?

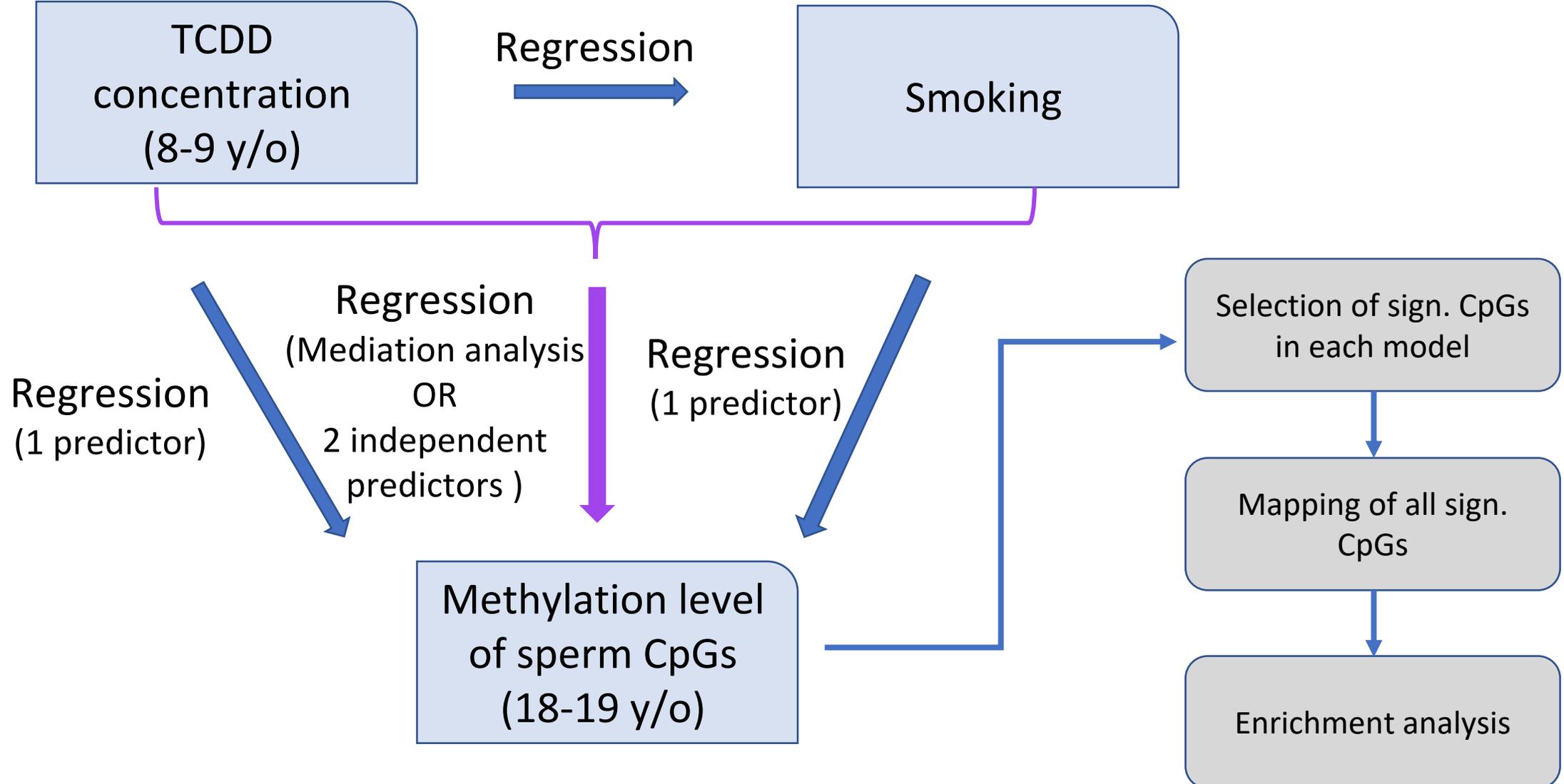
When you are smoking how many cigarettes do you smoke per one day?

How often do you smoke?

How deeply do you inhale the smoke?

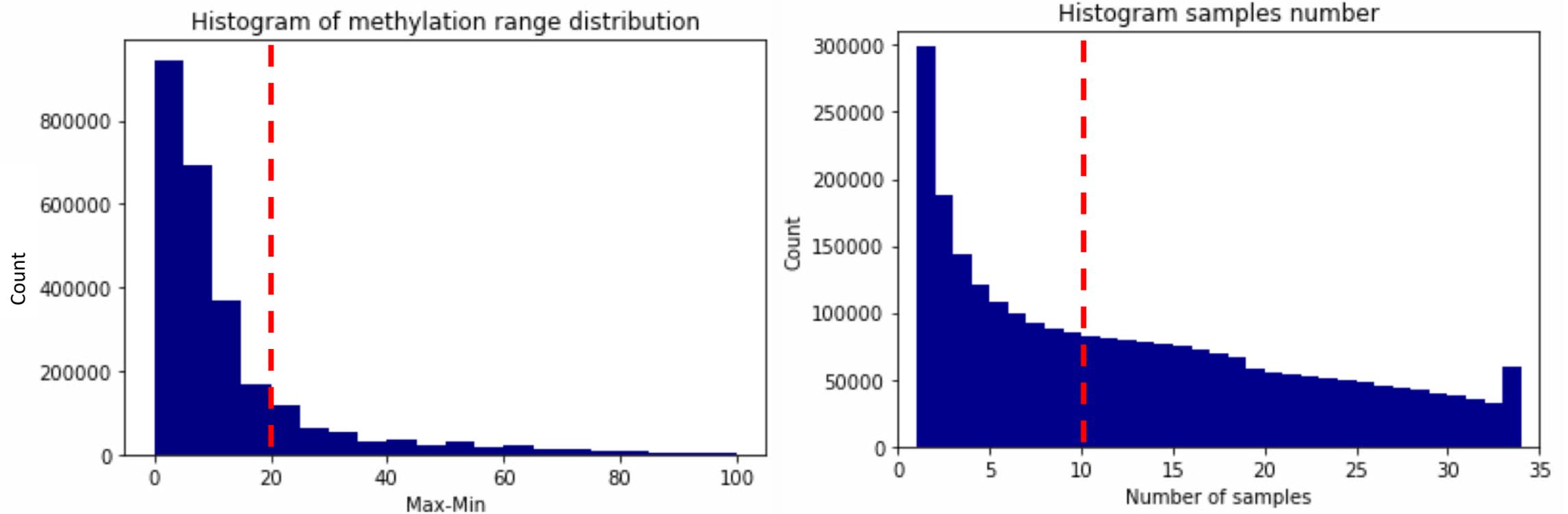


Project outline:



Results of regressions:

1. Making adequate subset of dataframe containing information about 2 611 773 CpGs methylation levels



⇒ **307 538 CpGs were selected for further analysis**

2. Performing the regressions

(either scipy (for one continuous variable as a predictor) or statsmodels (all others))

Results of regressions:

3. Can the smoking habit act as a mediator of TCDD exposure effects on human sperm epigenome?

- When smoking is a binary variable (did the person smoke within 6 months before semen collection) – **yes**
- When smoking is a range variable (did and how much did the person smoke within 6 months before semen collection) – **no**

For further analysis the range variable was chosen as more informative one

4. Selection of CpGs, whose methylation level is significantly dependent on the predictors

- **P value** of the overall model with fdr correction is less than **0.05**
- Absolute value of **R²** is more than **0.7**

Model	CpGs significant total	Common with TCDD+Smoke6 model
TCDD	29	11
Smoke6	1131	842
TCDD+Smoke6	1097	

Results of regressions:

5. Mapping of all significant CpGs found in each model to the human genome:

- Promoter regions: 2000 upstream:1000 downstream from TSS (Ensembl)
- Intraenhancer regions (Fantom5)
- Intragenic regions (hg19 annotation)

Model	Promoters	Enhancers	Intragenic	Total
TCDD	1	0	12	13
Smoke6	50	8	648	706
TCDD+Smoke6	47	13	623	683

Results of regressions:

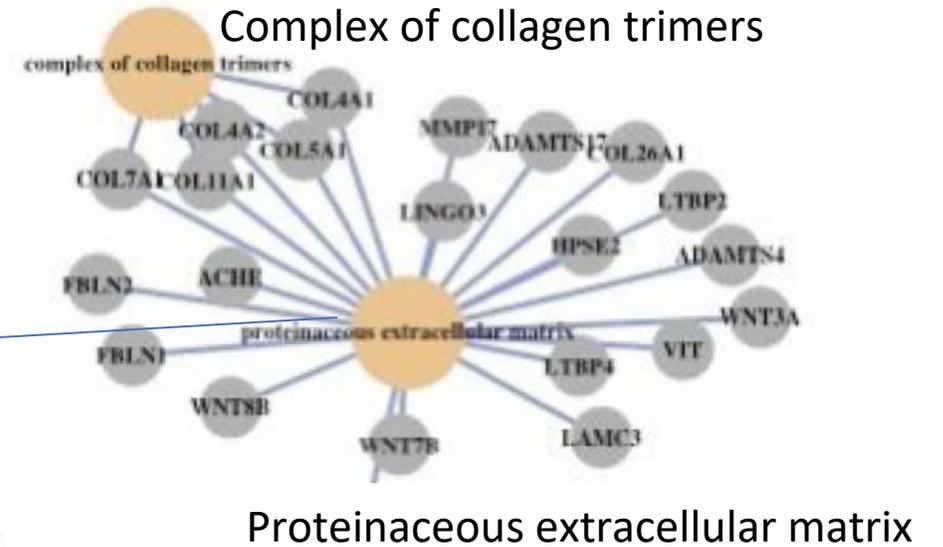
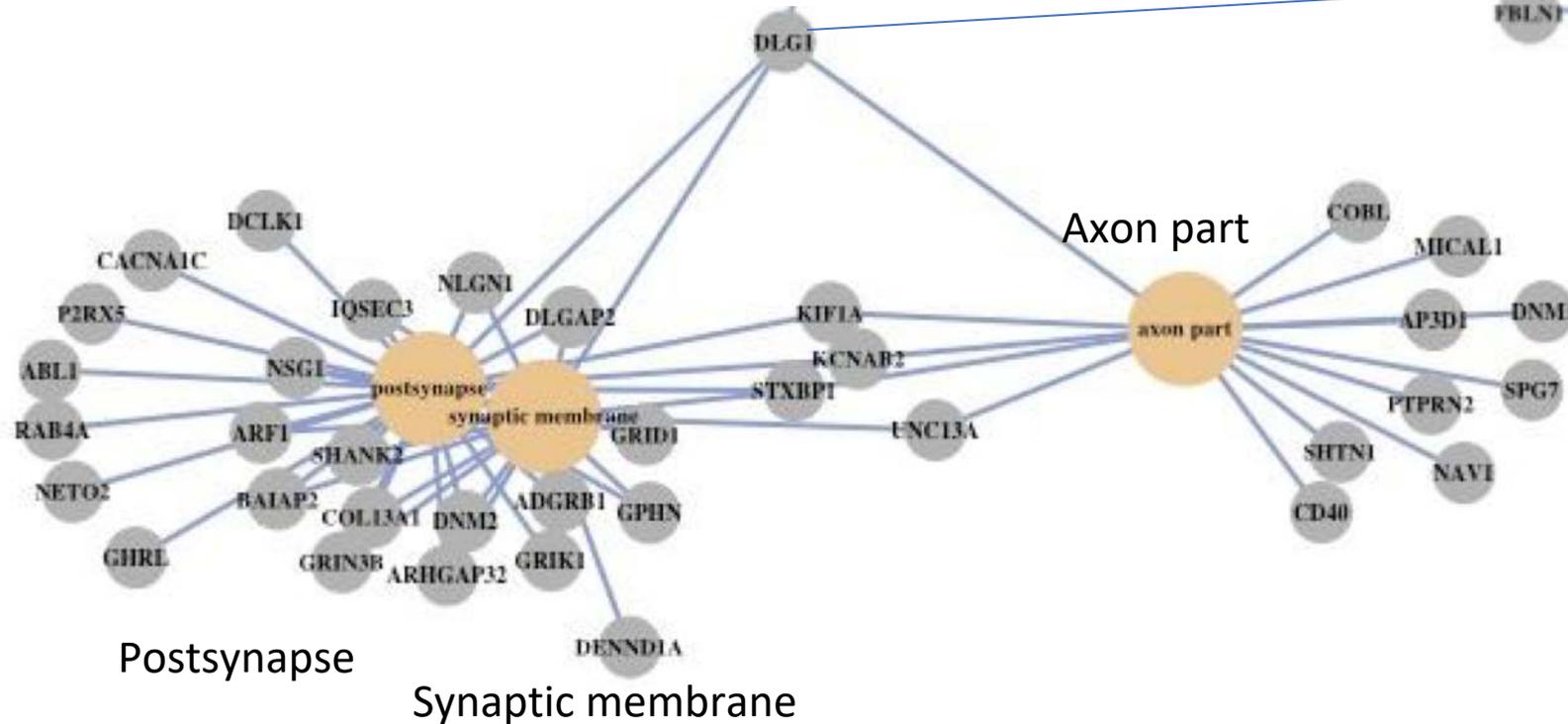
6. Enrichment analysis for all genes, associated with significant CpGs found in each model:

PathwayID	Pathway_category	TCDD	Smoke6	TCDD+Smoke6
R-HSA-8948216	Collagen chain trimerization	n/s	n/s	8/44
GO:0017053	transcriptional repressor complex	n/s	n/s	9/83
hsa04211	Longevity regulating pathway	n/s	9/89	10/89
GO:0098644	complex of collagen trimers	n/s	4/20	5/20
GO:0005578	proteinaceous extracellular matrix	n/s	20/365	22/365
GO:0098794	postsynapse	n/s	29/424	26/424
GO:0097060	synaptic membrane	n/s	20/295	19/295
GO:0033267	axon part	n/s	14/181	14/181
GO:0045211	postsynaptic membrane	n/s	17/225	15/225
GO:0009898	cytoplasmic side of plasma membrane	n/s	14/170	n/s
GO:0014069	postsynaptic density	n/s	15/204	n/s
hsa04072	Phospholipase D signaling pathway	n/s	12/146	n/s
R-HSA-157118	Signaling by NOTCH	2/73	n/s	n/s
GO:0005819	spindle	3/318	n/s	n/s
GO:0016323	basolateral plasma membrane	2/207	n/s	n/s
GO:0016580	Sin3 complex	1/13	n/s	n/s
GO:0071565	nBAF complex	1/14	n/s	n/s
GO:0000790	nuclear chromatin	2/340	n/s	n/s
	Genes in GO:	6	467	469

Results of regressions:

6. Enrichment analysis for all genes, associated with significant CpGs found in each model:

GO enrichment for TCDD+Smoking6 model



Summary:

1. Linear regression models for the influences of TCDD concentration in prepubertal age and smoking within 6 months before collection of sperm on the young adults semen methylation level were built (both for separate predictors and for their combination).
2. Predictors were considered to be independent on each other.
3. Only CpGs whose methylation levels were significantly dependent on the TCDD concentration and/or smoking, were chosen for further analysis.
4. All significant CpGs were mapped to either gene, promoter or enhancer regions.
5. Enrichment analysis for all genes associated with CpGs, whose methylation levels were significantly dependent on one or both of predictors, was performed.
6. Some interesting findings are needed to be analyzed in more details.

Thank you!

https://github.com/yukornienko/mediation_analysis