Using causal random forest to determine exposure environmenents with high sexual dimorphism

Exposome Data Challenge/ISGlobal 2021

Alejandro Cáceres

04/28

1 / 26

Can we identify exposure environments with high sexual dimorphism in human development?

Boys and girls develop differently

- Their **immune response** to infections differ from an early age.
- Their **brains** grow at different rates.
- The prevalence of numerous **common diseases**, like *obesity*, is different, with boys having higher risk.

Why studying sex differences?

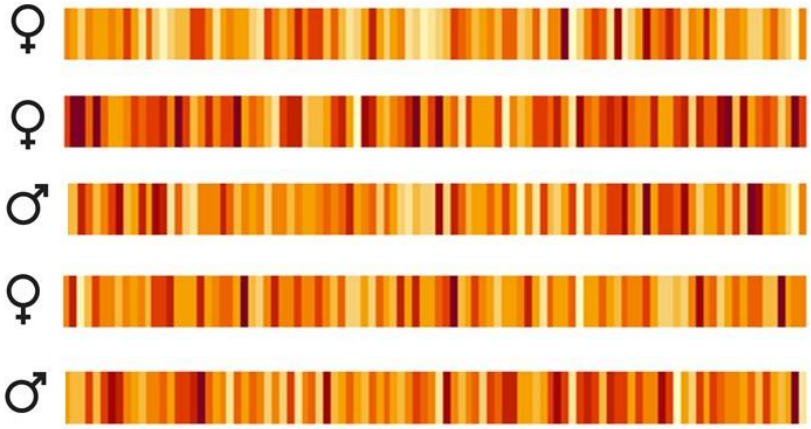
- The **main difference** between people is sex, underlying differences in disease risk and outcome.
- **Precision public health** and medicine aim to apply the best intervention at the right time to the right population.
- Sex differences in disease can inform on **etiology**.
- Age-related diseases like cancer, Alzheimer's and autoimmune diseases are strongly influenced by **sex hormones**.
- Studying **preteens** can inform on non-hormonal mechanisms of sexual dimorphism in disease.

Hypothesis

A co-ocurrence of environmental conditions can be identified from exposomic data such that the differences in BMI between boys and girls is highest.

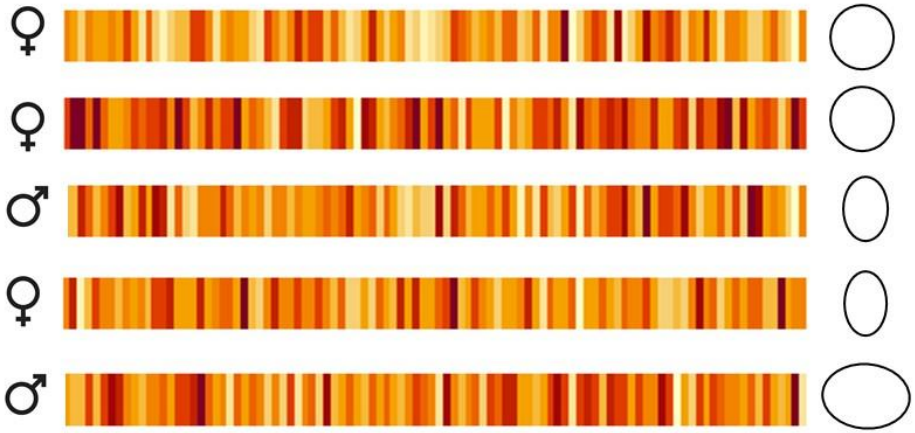
The exposome

High-dimensional exposure data that **overspecifies** each individual.



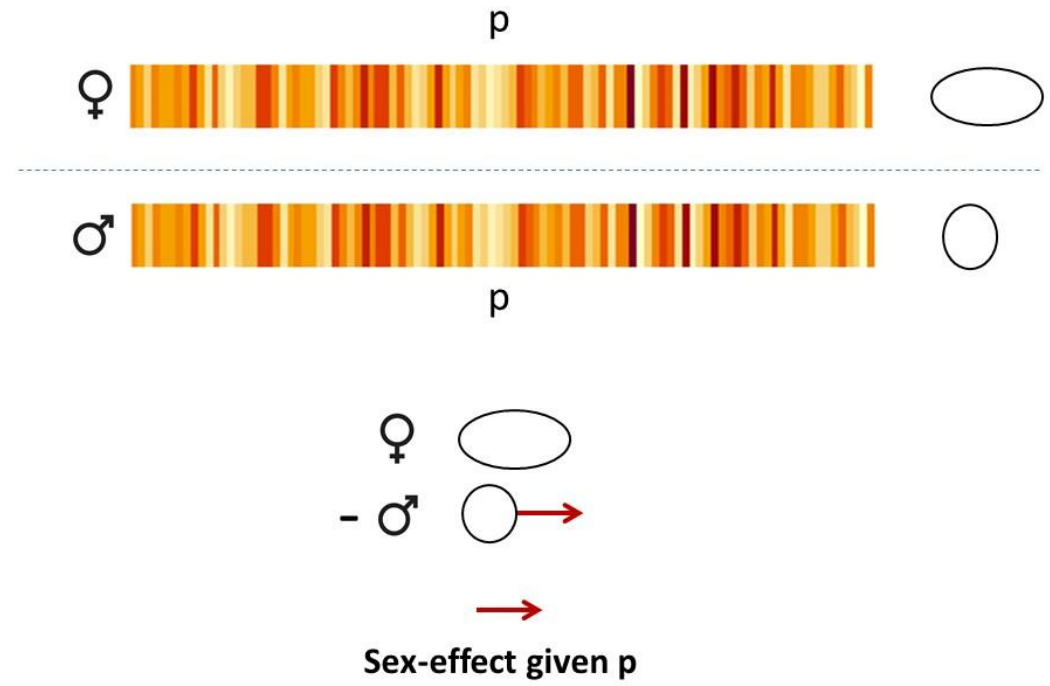
The exposome, sex and BMI

We want to link it with BMI differences between boys and girls.



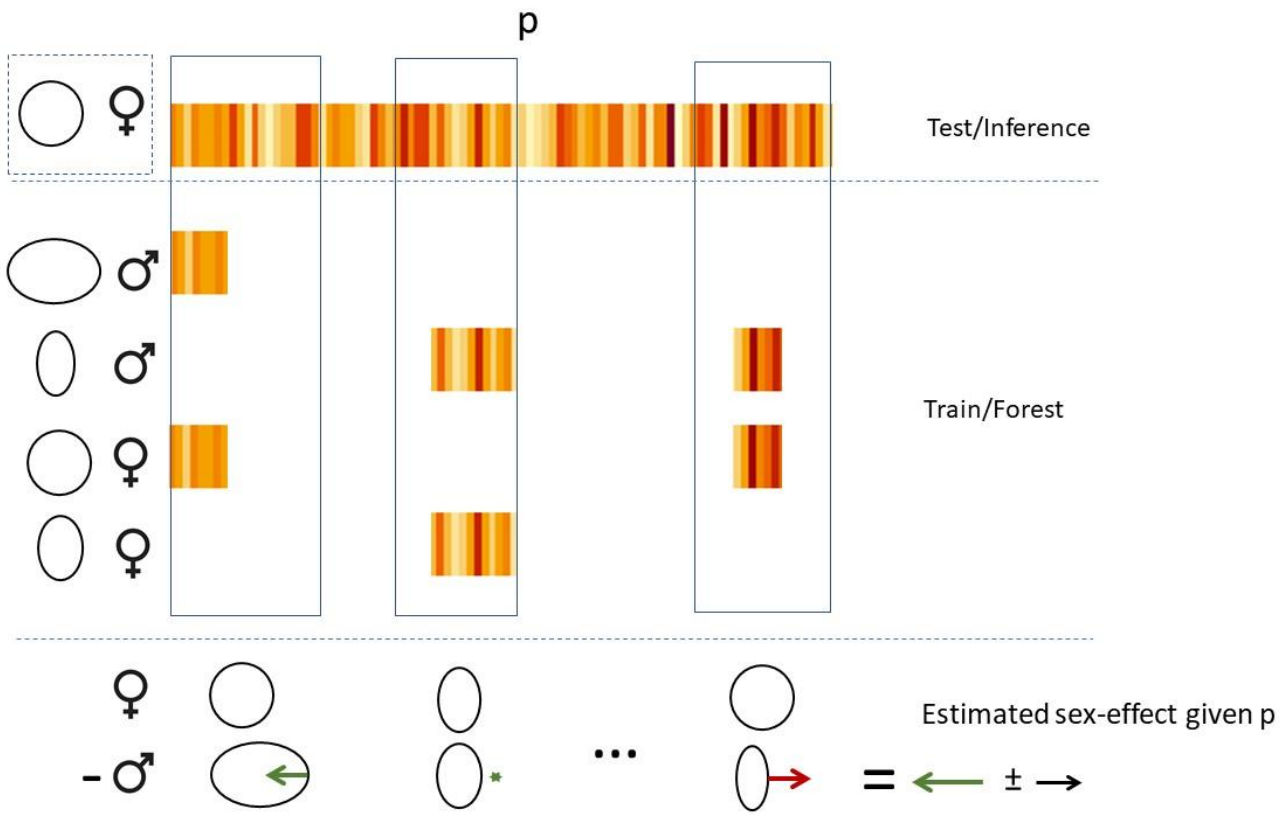
Causal inference

We **propose** to estimate sexual dimorphism as a causal inference problem.



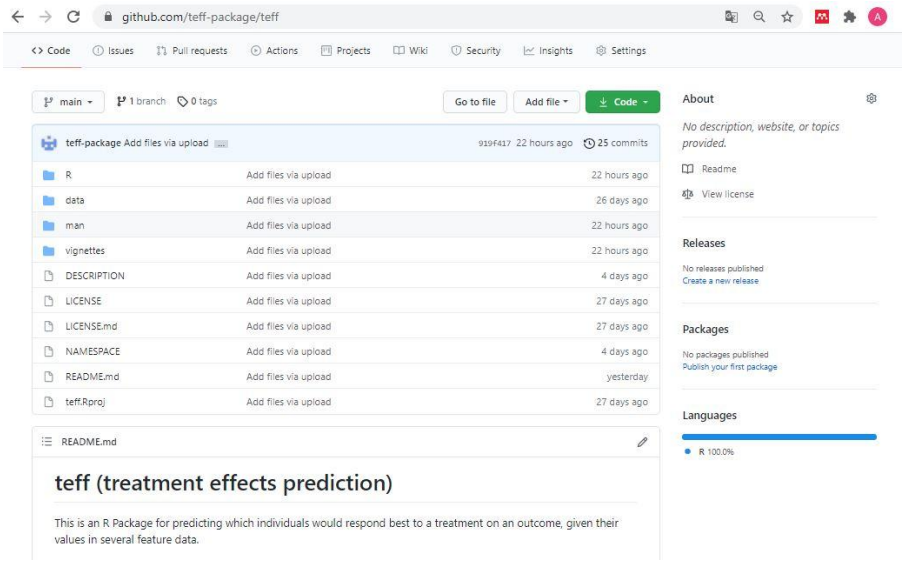
$$\tau(\mathbf{p}) = E[BMI_j(1) - BMI_j(0)|\mathbf{P}_j = \mathbf{p}]$$

Random Causal Forest



Athey, et al. Ann. Statist. 2019, 47(2): 1148-1178

teff (treatment effects prediction)



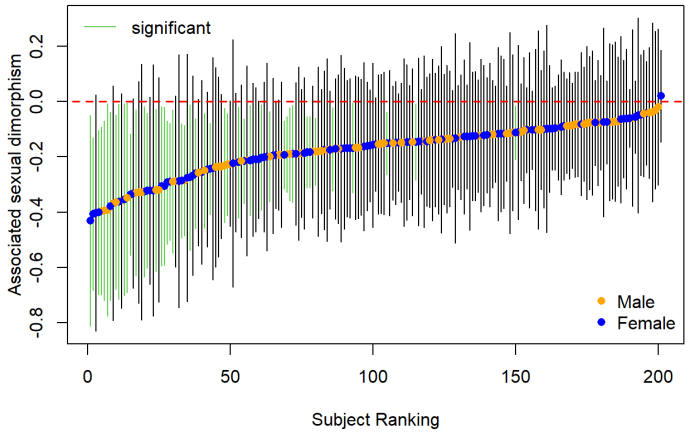
The package uses RCF to estimate the **transcriptomic** groups where the effect of a **treatment** on an **outcome** is maximum.

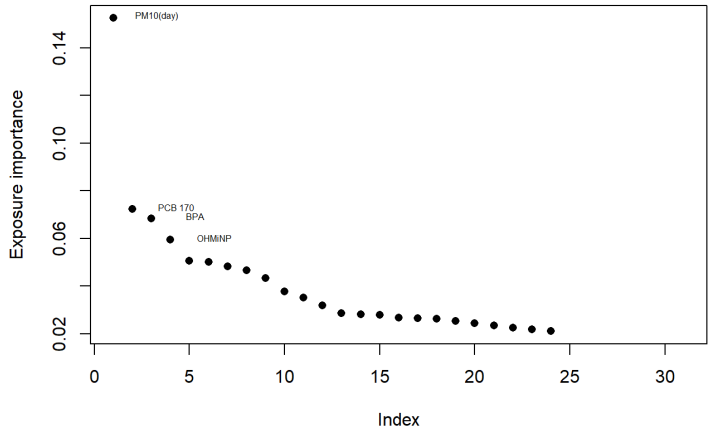
For the data challenge:

- Transcriptomic data → Exposomic data
- Treatment → Sex
- Outcome → BMI

Individuals with significant sex-effects on BMI

We found 46 individuals from 155 (test set) with significantly negative sexual dimorphism in BMI (M>F).



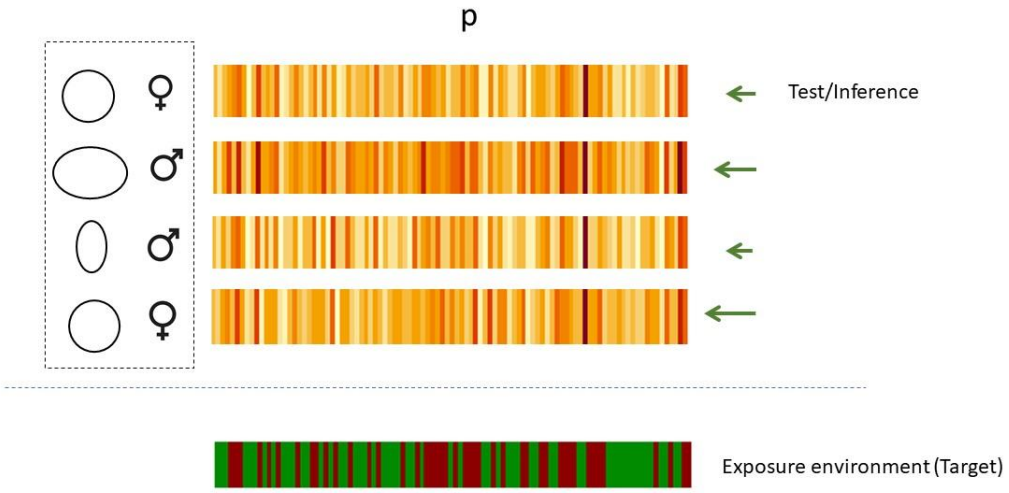


- We **selected** 31 informative exposures: Those whose interaction with sex significantly associated (nominal level) with zBMI. we adjusted by all covariates available in the exposome dataset.

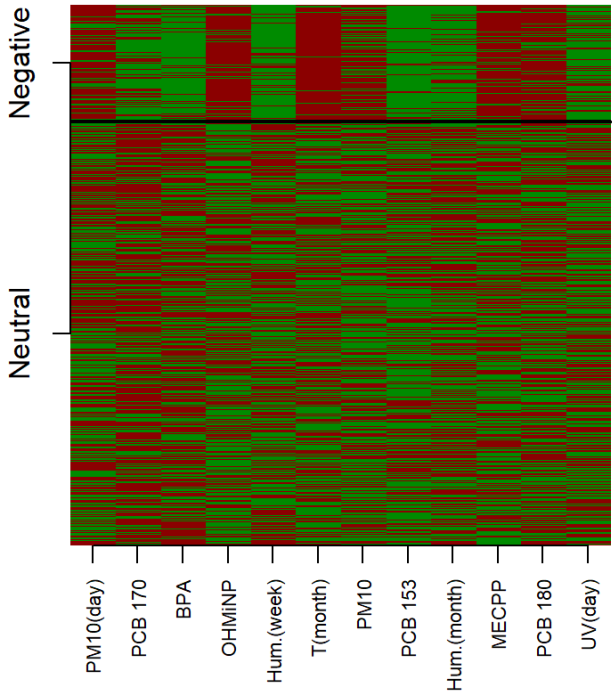
teff Code

```
#...FORMATIG DATA

#data for teff
data4teff <- list(features=data.frame(modexp),
                  teffdata = data.frame(mod))

#appluy RCF
pred <- predicteff(data4teff,
                   featuresinf = rownames(texpsig))

ctrl <- list(lb=c("Male", "Female"),
             wht="bottomright", whs = "topleft")

plotPredict(pred,
            lb="Associated sexual dimorphism",
            ctrl.plot=ctrl)
```

Targetting individuals on an exposure environment

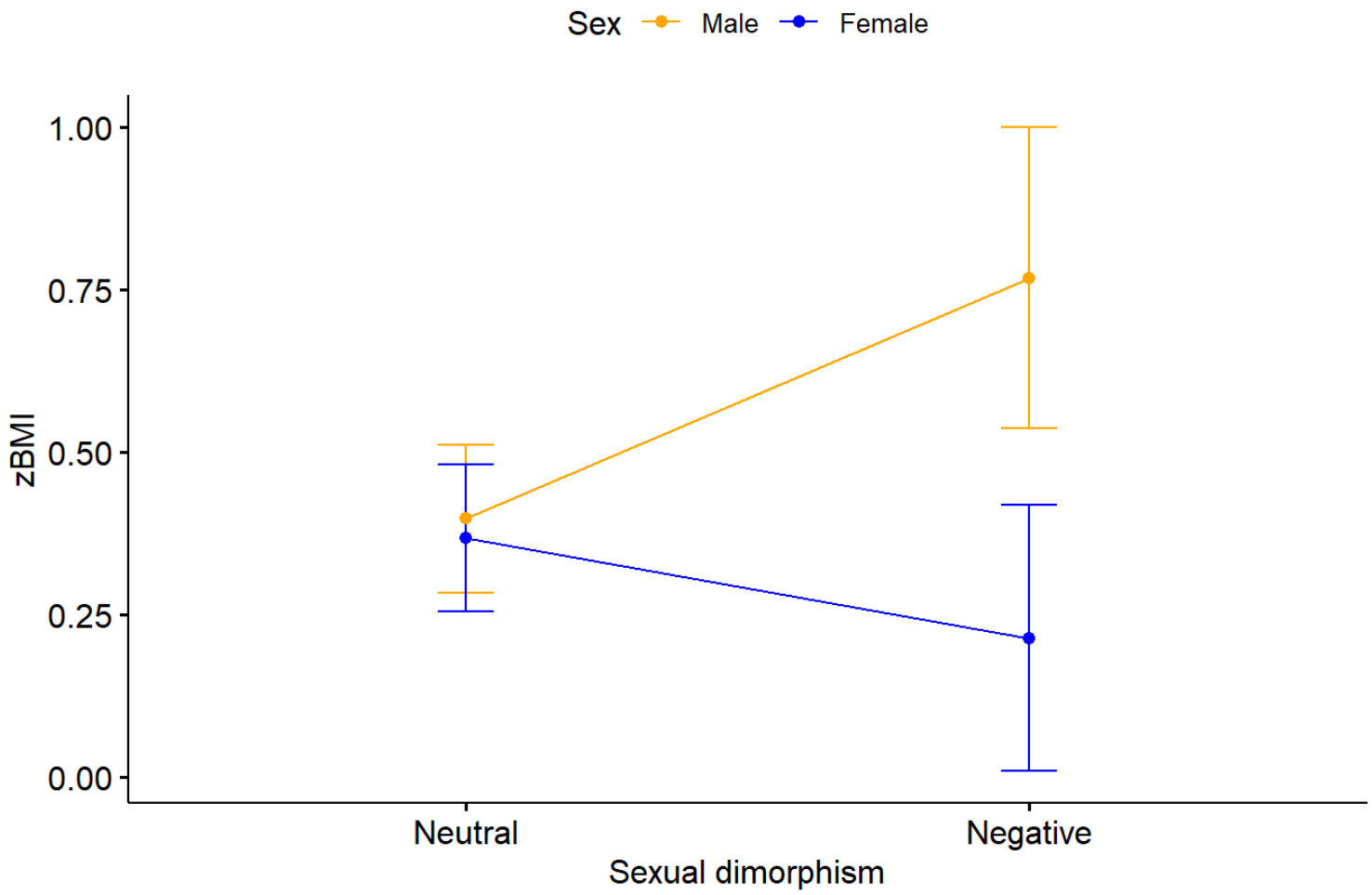


Exposure environment of high dimorphism



- Red: adjusted exposures that were **higher** than the overall mean
- Green: adjusted exposures that were **lower** than the overall mean

Exposure environment of high dimorphism



teff code

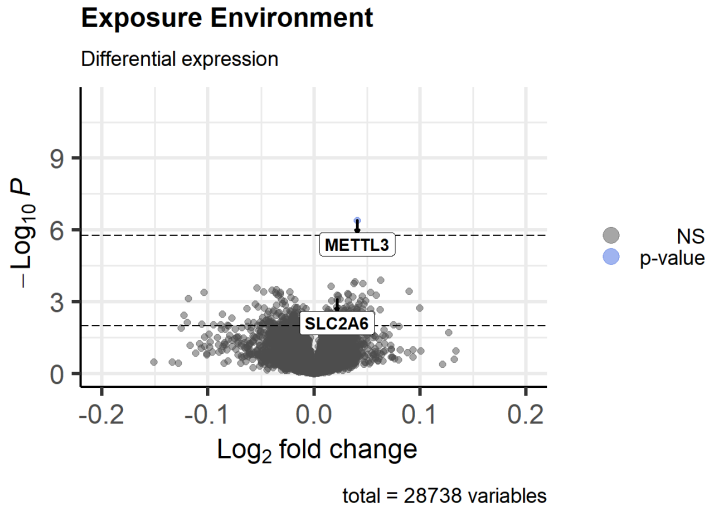
```
pred <- predicteff(data4teff,
                   featuresinf = rownames(texpsig),
                   profile = TRUE, quant = 0.7)

tar <- target(data4teff, pred,
              effect = "negative",
              mat = 0.7,
              lb = code[pred$featurenames$featurenames,"labels"])
```

Exposure environment of high dimorphism

```
tar
```

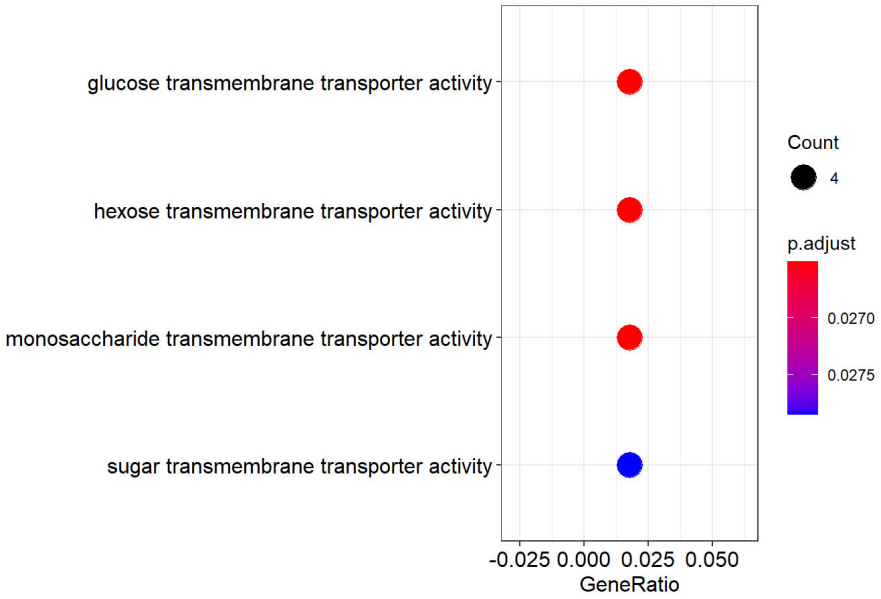
```
## object of class: tarteff
##
## classification into
##   negative treatment effect: 1
##   neutral: 0
##
##   0   1
## 790 217
##
## interaction fitted model: gaussian
##     Estimate   Std. Error      t value     Pr(>|t|)
## -0.524140672  0.177261958 -2.956870600  0.003180596
```

Molecular correlates of the exposure environment



- *METTL3* a subunit of the most abundant **methyltransferase** of mRNA (m6A).
- Upregulation of *METTL3* is involved in **adipogenesis** and in M1 **magrophague** polarization (PMC6066751, PMID: 31365297).
- *FTO* (fat mass and obesity-associated protein) **reverses** the action of m6A (PMID: 28002401).
- *METTL3* also regulates **chromosome-X inactivation** in females.

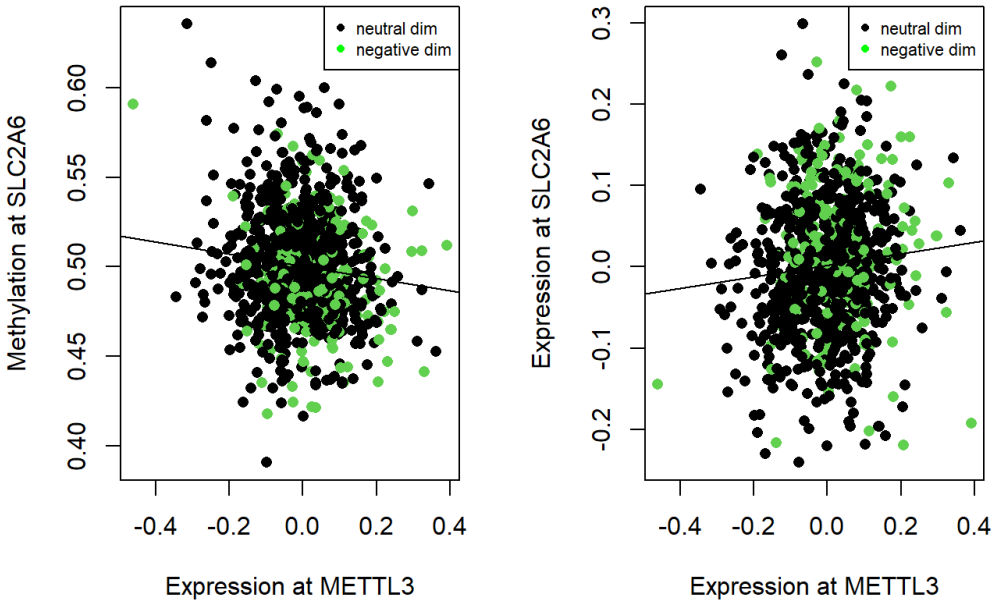
Molecular correlates of the exposure environment



These erichment associations were led by 4 genes:

- *SLC5A9*, *SLC2A8*, *SLC2A6* and *SLC5A1*.
- They encode **glucose transporters**: GLUT9, GLUT8, GLUT6 and SGLT1.

Molecular correlates of the exposure environment



- Increased expression of *METTL3* associated with more methylation (cg16190350) and less expression of *SLC2A6*.
- *SLC2A6* is involved in the activation of **inflammatory macrophages** M1 (PMID: 30700586)

Molecular correlates of the exposure environment

Association of **immune cell abundance** with BMI and the interaction between the exposure environment and sex.

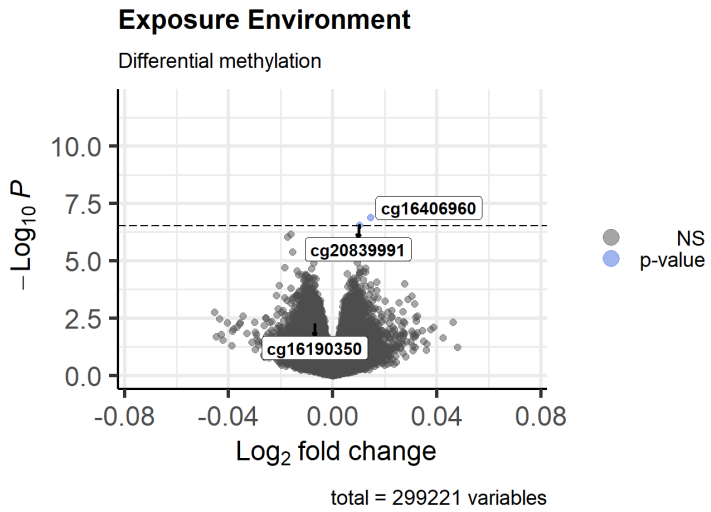
```
##             BMI      P envExp*sex      P
## NK_6     0.0002 0.8255     0.0010 0.9126
## Bcell_6  0.0001 0.9415     0.0070 0.3883
## CD4T_6  -0.0063 0.0002     0.0010 0.9415
## CD8T_6  -0.0029 0.0255     0.0209 0.0469
## Gran_6   0.0079 0.0070    -0.0248 0.2883
## Mono_6   0.0017 0.0267    -0.0046 0.4452
```

Molecular correlates of the exposure environment

Association of immune cell abundance with *SLC2A6* and the interaction between *METTL3* and sex.

```
##           SLC2A6      P METTL3*sex      P
## NK_6    -0.0115 0.4441    -0.0652 0.0059
## Bcell_6 -0.0005 0.9719     0.0356 0.1082
## CD4T_6  -0.0077 0.7464     0.0237 0.5271
## CD8T_6  -0.0167 0.3611     0.0148 0.6077
## Gran_6   0.0110 0.7864     0.0007 0.9911
## Mono_6   0.0232 0.0253    -0.0062 0.7074
```

Molecular correlates of the exposure environment



Significant methlylome-wide CpGs in genes with immune function:

- cg16406960 is at *GNLY* that enocodes granulysin an **antimicrobial** peptide relased by cytolytic T cells
- cg20839991 is at *HLA-DPB2*, major **Histocompatibility** Complex, Class II, DP Beta 2.

Final remaks

- We identified an **environment of negative dimorphism in BMI** (M>F) that could be linked to meaningful molecular mechanisms.
- These included genes involved in **immune response** and **adipogenenis**.
- The 12-exposure environment included increments in PM10, low meterological temperature, high humidity and the reduction and increment of specific organochlorines, phthalates and phenols.
- The exposure environment **should be validated** with external data and simplified (standardized) for public health interventions.
- The method is sensitive on **variable selection**.
- We found exposure environments of high dimorphism for **IQ** and **asthma** but they could not be linked to molecular function.
- We **did not find** an exposure environments of high dimorphism for **obesity**, but the BMI exposure environment significantly modulated the obeseity dimorphism.

Further question

To which extent immflamation from environmental exposures can trigger obesity?

Thank you for your attention

Find the slides and the code online.

Alejandro Cáceres@BRGE

ISGlobal Bioinformatics Research Group in Epidemiology (BRGE)
alejandro.caceres@isglobal.org