INTRODUCTION TO EPIGENOME-WIDE ASSOCIATION STUDIES (EWAS)

5. BIOLOGICAL INTERPRETATION OF FINDINGS (THEORY)

EPIGENOME-WIDE ASSOCIATION STUDY (EWAS)

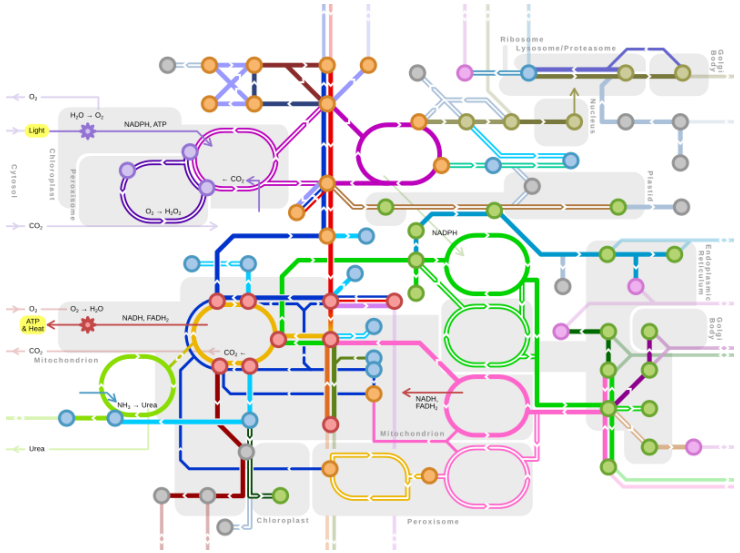
Workflow

1. Scientific question
2. Study population
3. Biological sample
4. DNA methylation data acquisition
5. Quality control of DNA methylation data
6. Epigenome-wide association study (EWAS)
7. Meta-EWAS or replication / validation
8. Biological interpretation

EPIGENOME-WIDE ASSOCIATION STUDY (EWAS)

Workflow

1. Scientific question
2. Study population
3. Biological sample
4. DNA methylation data acquisition
5. Quality control of DNA methylation data
6. Epigenome-wide association study (EWAS)
7. Meta-EWAS or replication / validation
8. **Biological interpretation**



8. BIOLOGICAL INTEPRETATION

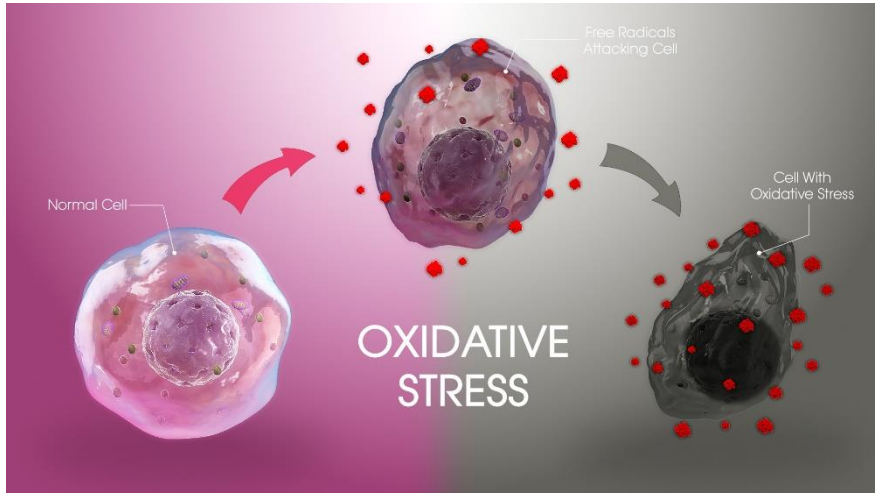
Which is the aim?

To interpret which are the genes or biological functions affected.

Tobacco smoking

Significant CpGs

Inflammation



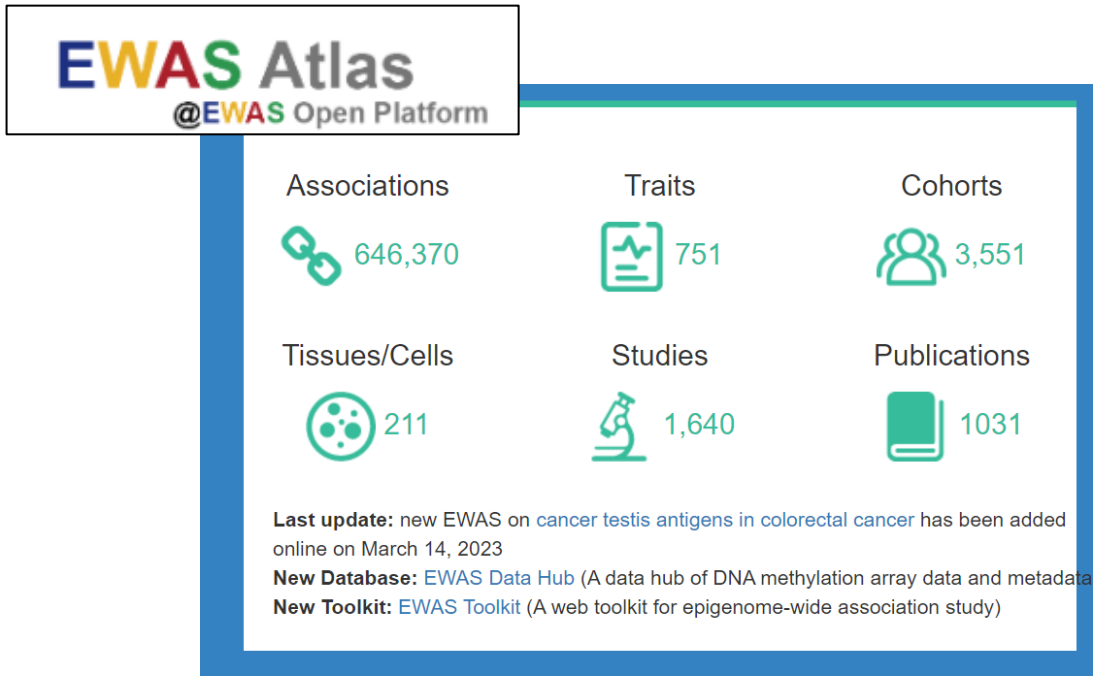
8. BIOLOGICAL INTEPRETATION

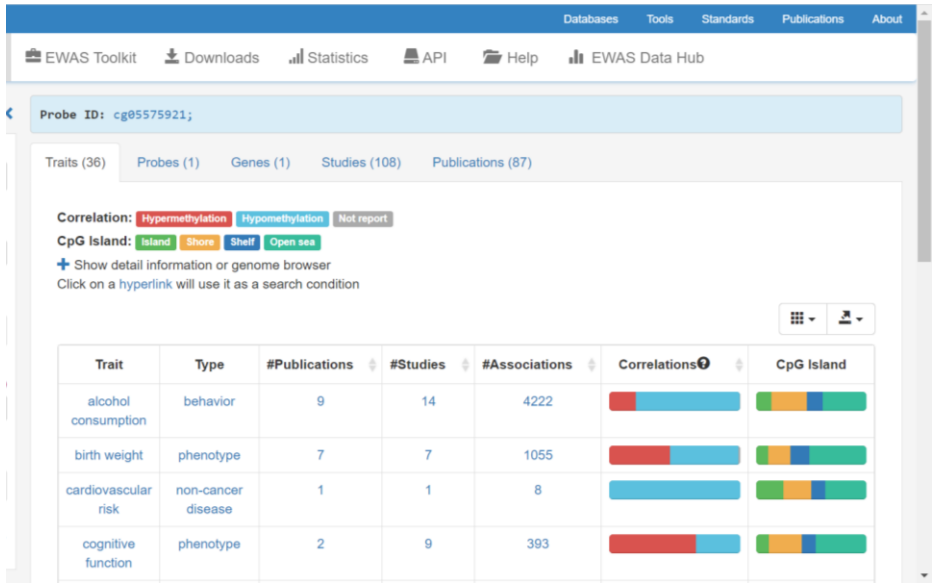
Previous reported associations

- Pubmed
- EWAS catalog (http://www.ewascatalog.org/)
- EWAS atlas (https://ngdc.cncb.ac.cn/ewas/atlas)

EWAS Catalog β
The MRC-IEU catalog of epigenome-wide association studies

EWAS Atlas
@EWAS Open Platform





8. BIOLOGICAL INTEPRETATION

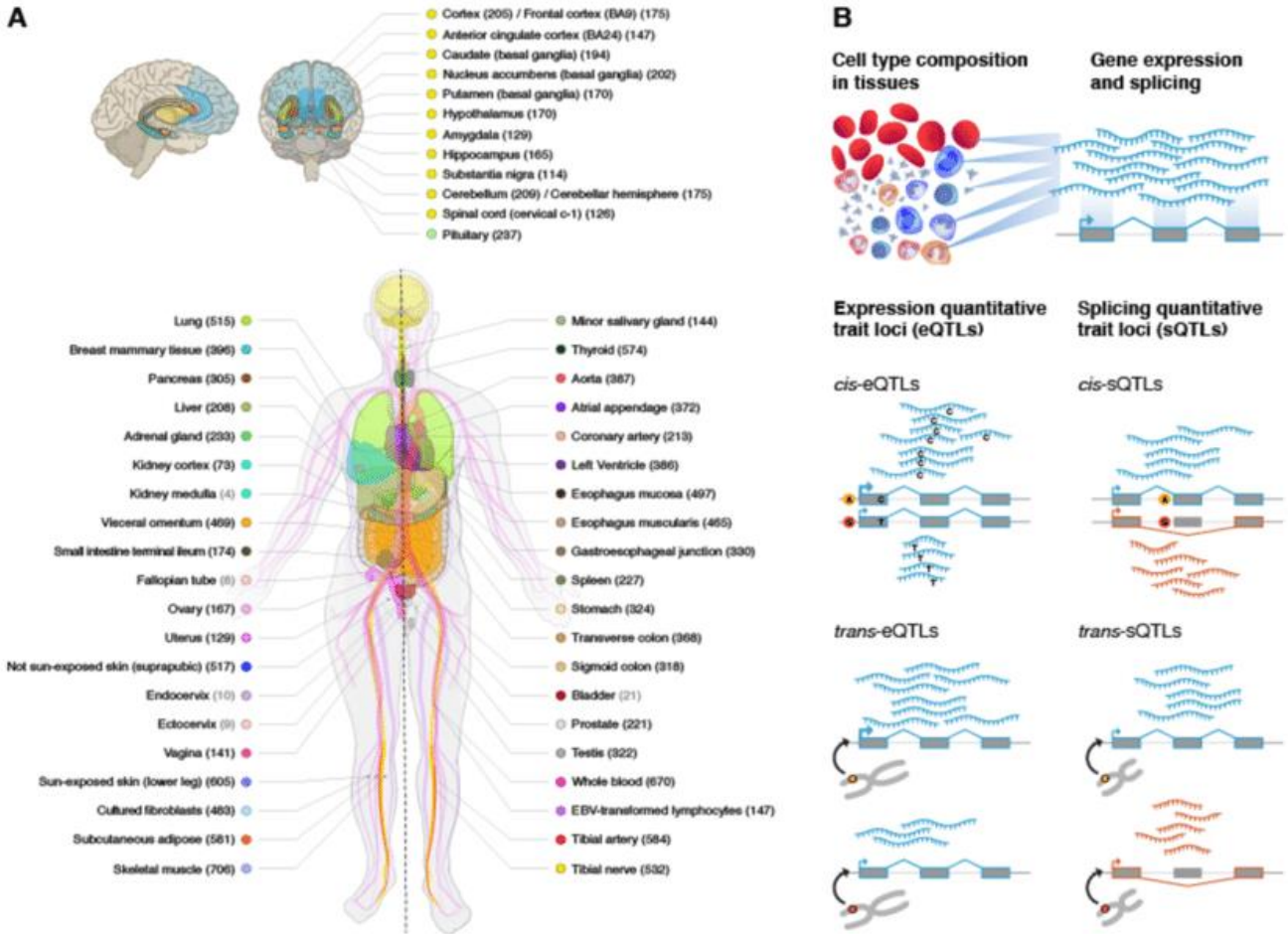
Methylation profile across tissues

The Genotype-Tissue Expression (GTEx)

(https://gtexportal.org/home/)

- Aim: comprehensive public resource to study tissue-specific gene expression and genetic regulation

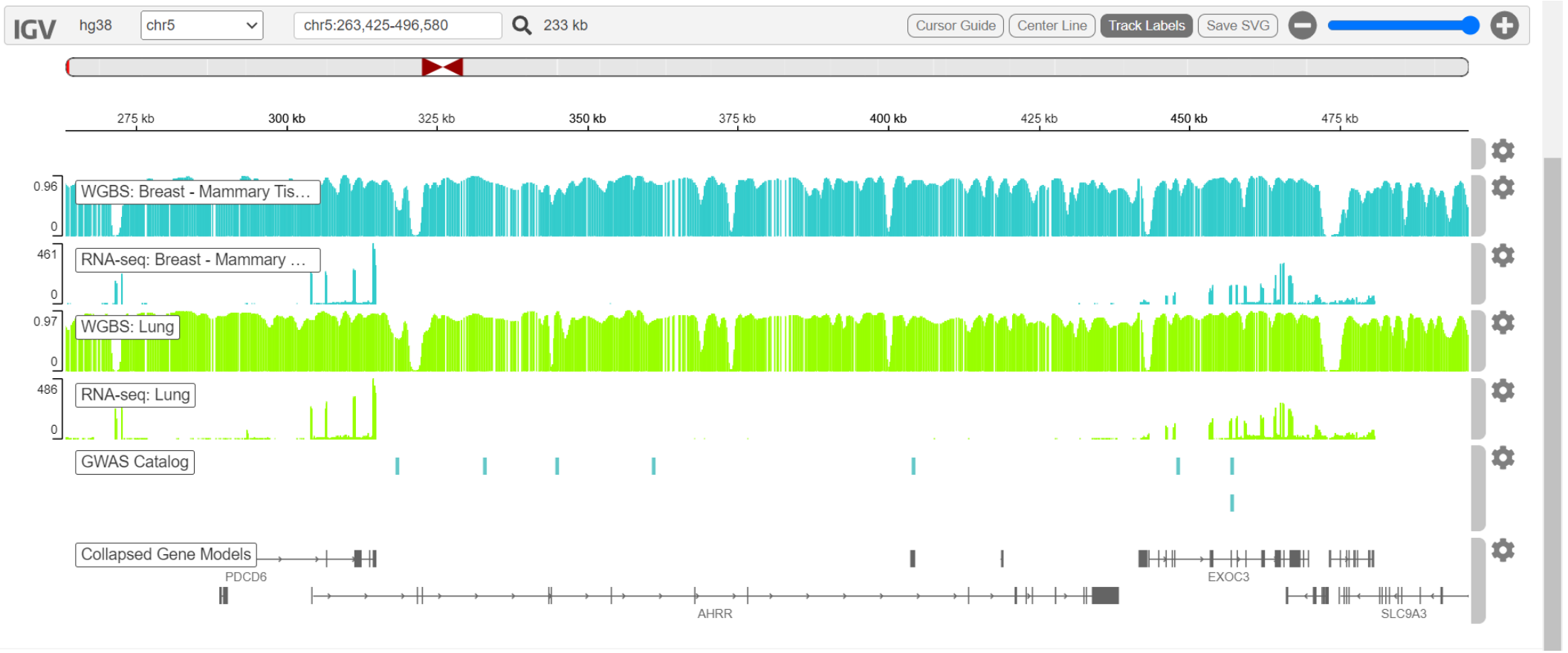
- Now it has incorporated methylation data (WGBS and EPIC array)!



8. BIOLOGICAL INTEPRETATION

Methylation profile across tissues

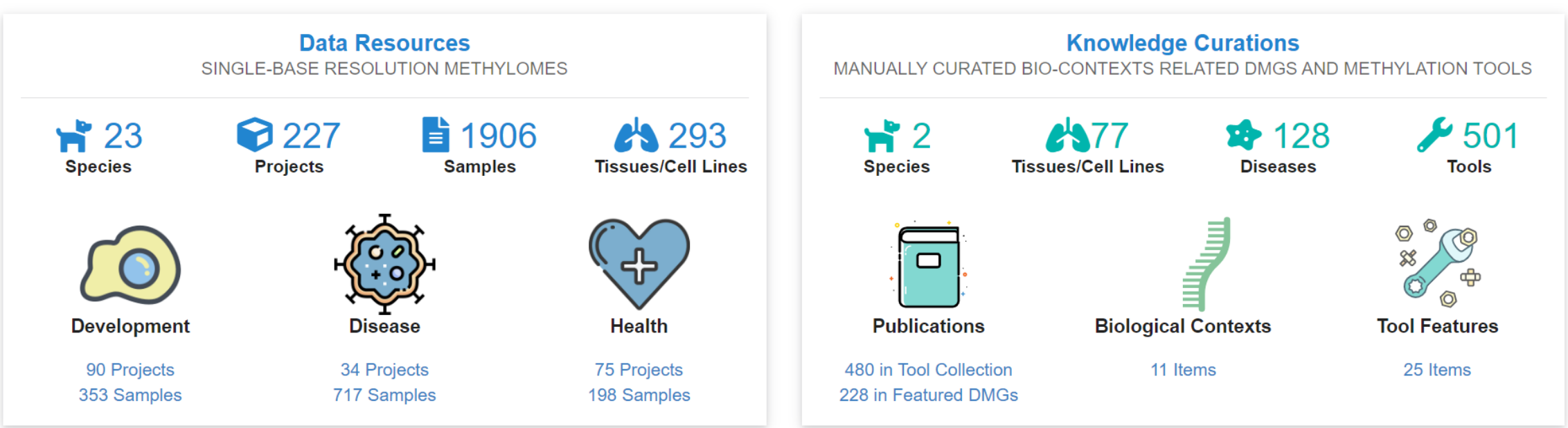
The Genotype-Tissue Expression (GTEx) project (https://gtexportal.org/home/)



8. BIOLOGICAL INTEPRETATION

Methylation profile across species

MetBank (https://ngdc.cncb.ac.cn/methbank/)

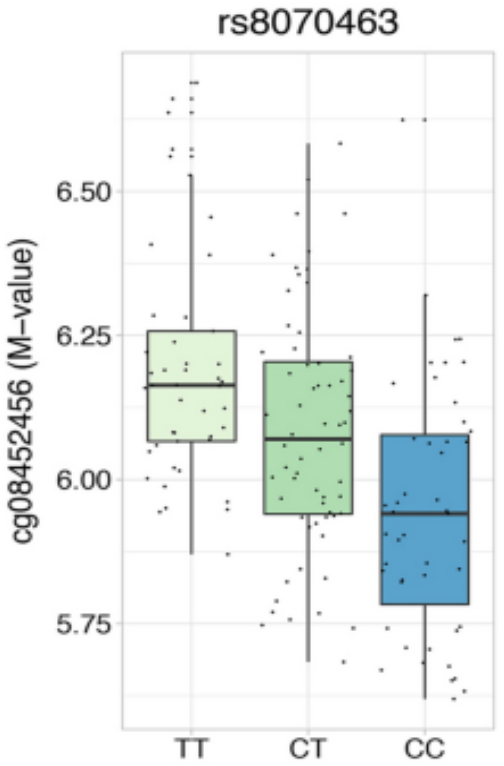


8. BIOLOGICAL INTEPRETATION

Methylation quantitative trait locus (mQTL)

- Genetic variation (SNP) associated with DNA methylation status.
- Tissue specific

- mQTL catalogues for blood
  - GoDMC (http://mqtldb.godmc.org.uk/)
  - ARIES (http://www.mqtldb.org/)

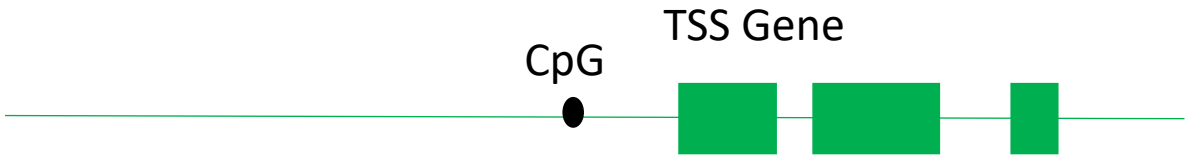


8. BIOLOGICAL INTEPRETATION

Annotation of CpGs to genes

Positional annotation

- “IlluminaHumanMethylationEPICanno.ilm10b2.hg19“ R package
- 1500 upstream and 3’UTR downstream the gene



Functional annotation

- Based on methylation-expression correlations (eQTM)

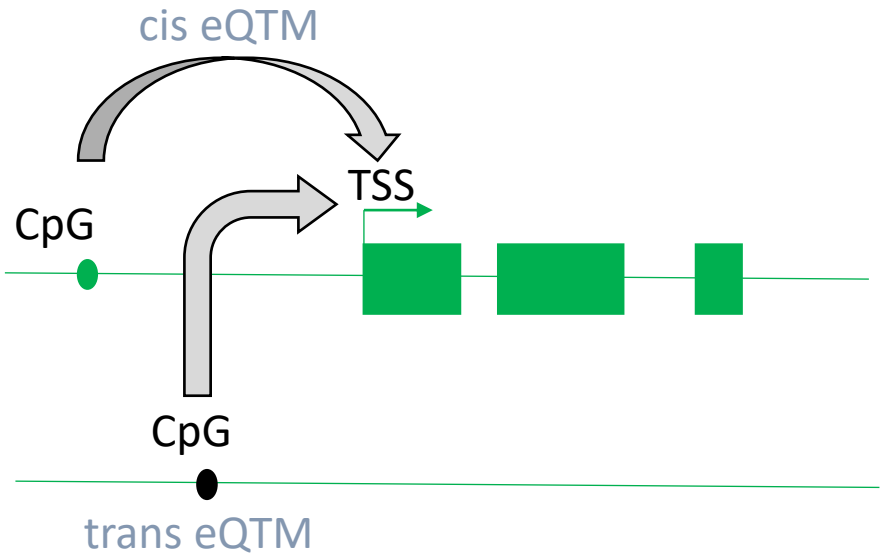
8. BIOLOGICAL INTEPRETATION

Expression quantitative trait methylation (eQTM)

- Specific position of the genome whose DNA methylation status is associated with the expression of a given gene.
- Cis and trans

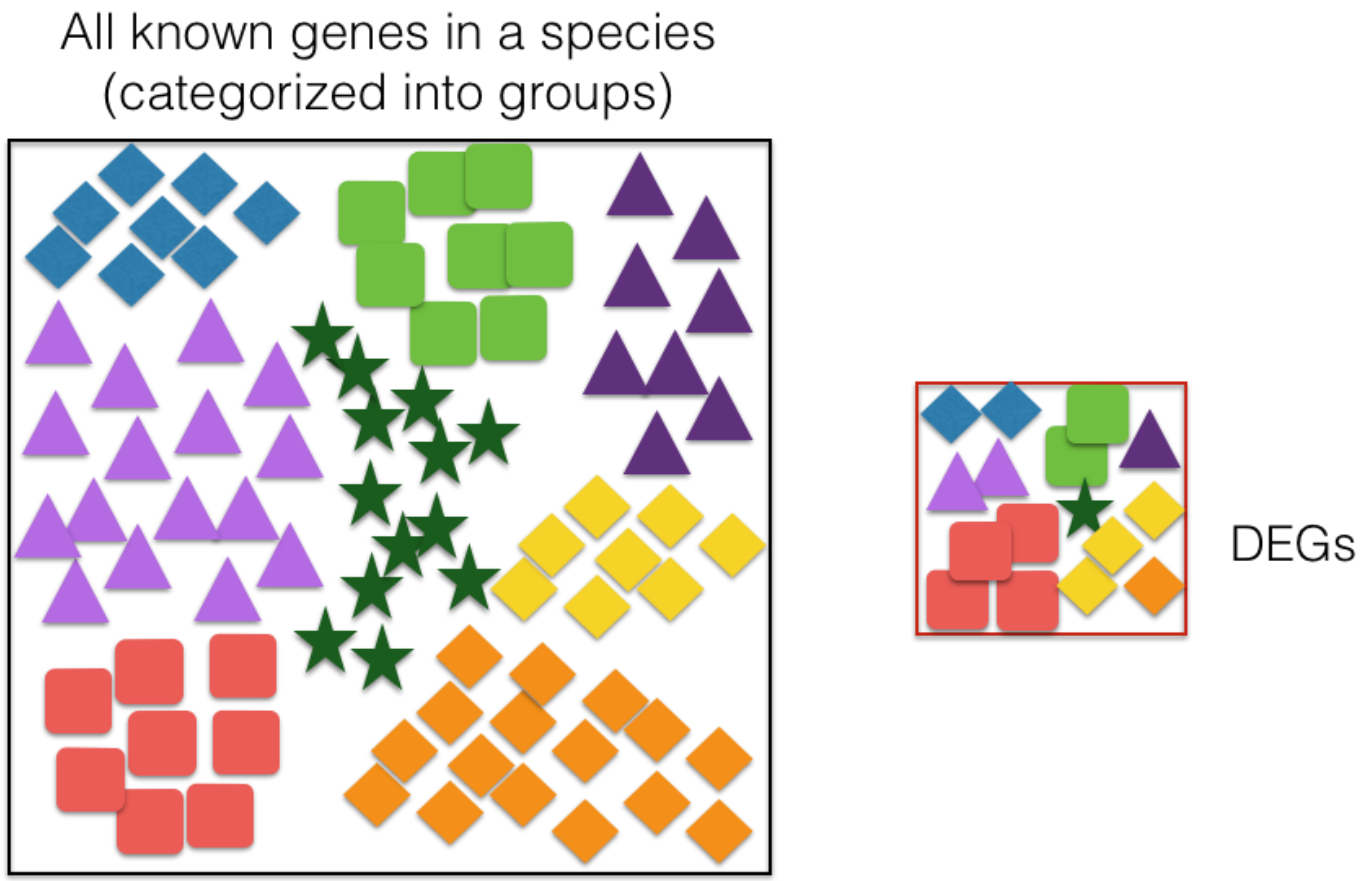
- Catalogues
  - Tissue specific
  - Mostly in blood
  - HELIX cis eQTM catalog in child blood

  (https://helixomics.isglobal.org/)



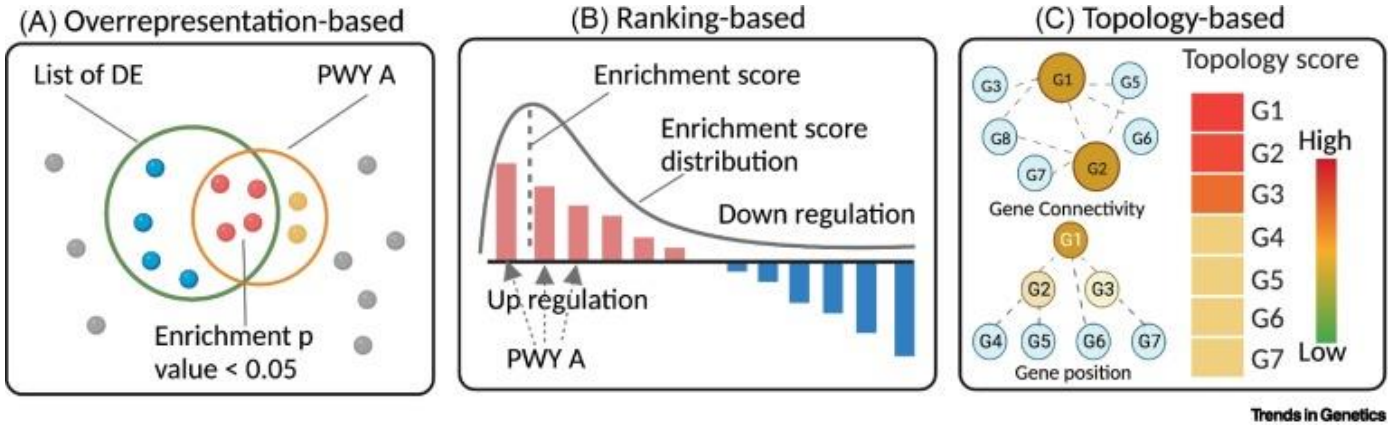
8. BIOLOGICAL INTEPRETATION

Functional enrichment analyses



8. BIOLOGICAL INTEPRETATION

Functional enrichment analyses approaches



Overrepresentation method

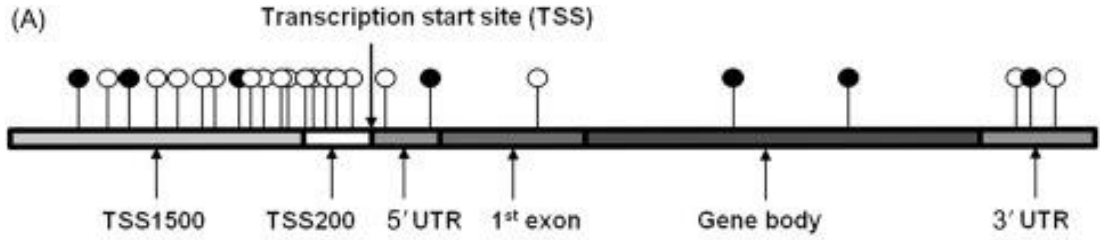
- Contingency table
- Test: Chi2, Fisher, …
- Correction for multiple-testing
- Correction for array bias (missmethyl R package)

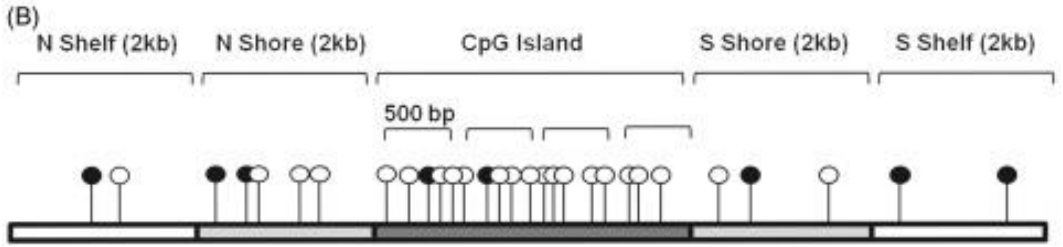
	User Genes	Genome	
In Pathway	3	37	40
Not in Pathway	297	29663	29960
	300	29700	30000

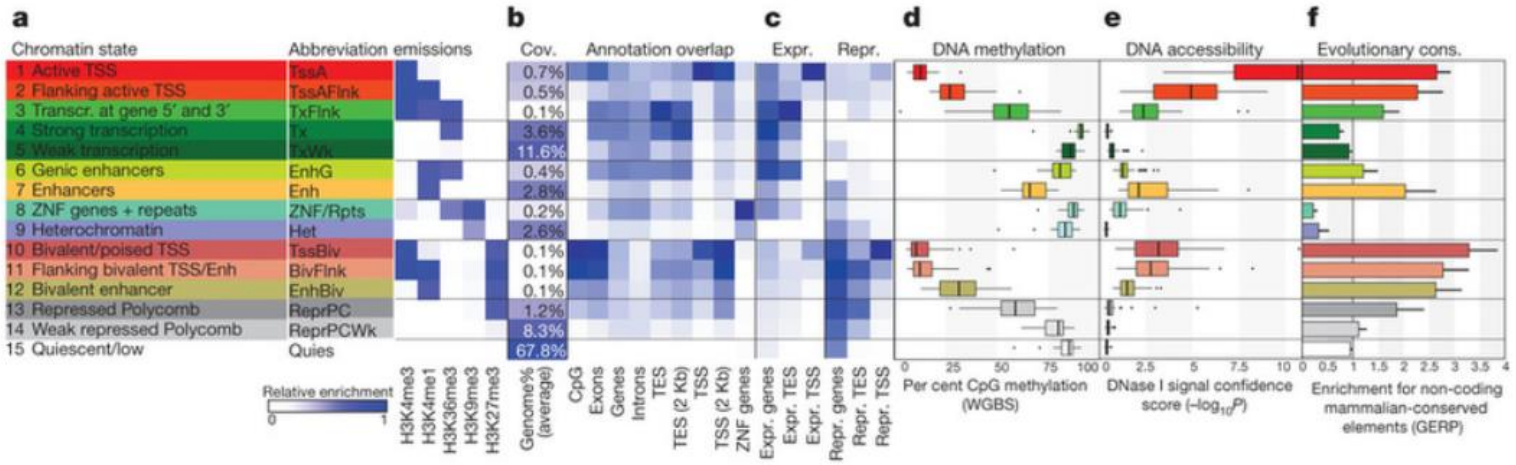
8. BIOLOGICAL INTEPRETATION

DB molecular enrichment (CpG list)

- Relative gene positions (Illumina)
- CpG island relative positions (Illumina)
- Chromatin states (ROADMAP)
- Imprinted regions (literature)
- Tissues and cell types
- …



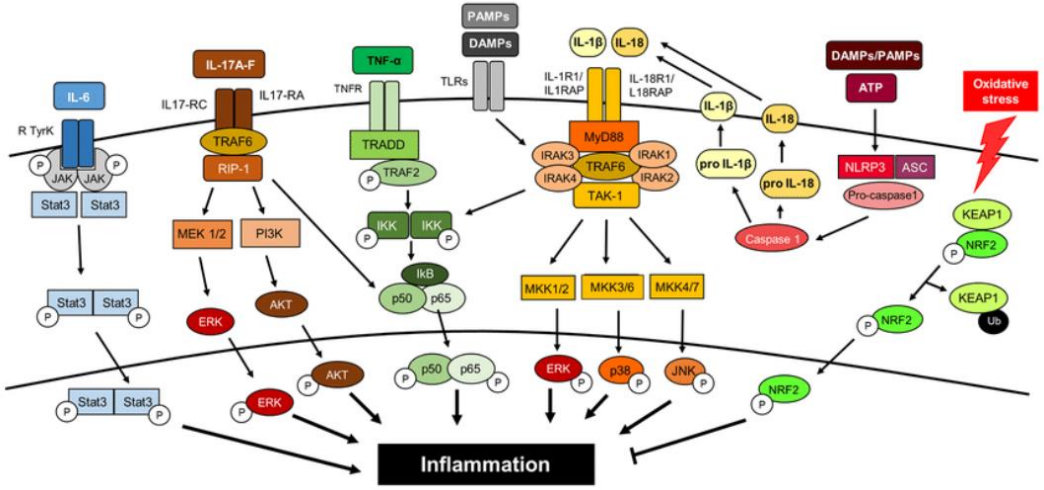




8. BIOLOGICAL INTEPRETATION

Databases for gene-set enrichment (gene list)

- Gene ontology (GO) terms (http://geneontology.org/)
  - Molecular function (ie. ubiquitinization)
  - Cellular component (ie. cell membrane)
  - Biological process (ie. DNA repair)
- Pathways or gene-sets
  - Kyoto Encyclopedia of Genes and Genomes (https://www.genome.jp/kegg/)
  - Reactome (https://reactome.org/)
- Diseases
  - DisGeNET (https://www.disgenet.org/)
- Tissues and cell types
  - Human Protein Atlas (https://www.proteinatlas.org/)
- Transcription factor regulation
  - ENCODE and ChEA Consensus TFs from ChIP-X
- …



8. BIOLOGICAL INTEPRETATION

Tools

CpG list

- **eFORGE (https://eforge.altiusinstitute.org/): CpG list -> tissue**
- eFORGE TF (https://eforge-tf.altiusinstitute.org/): CpG list -> transcription factors
- missmethyl R package: CpG list -> gene list (internally) -> gene-set
  -> controls for array bias

Gene list

- **ClusterProfiler R package: gene list -> gene-set, GO-terms**
- enrichR (https://maayanlab.cloud/Enrichr/): gene list -> gene-set, disease, transcription factors …
- ConsensusPath (http://cpdb.molgen.mpg.de/): gene list -> gene-set and others
- DAVID (https://david.ncifcrf.gov/): gene list -> gene-set

INTRODUCTION TO EPIGENOME-WIDE ASSOCIATION STUDIES (EWAS)

5. BIOLOGICAL INTERPRETATION OF FINDINGS (THEORY)

FUNCTIONAL ENRICHEMENT ANALYSIS OF CURRENT AND FORMER SMOKING

Data: results from meta-analysis

- Array: 450K
- Tissue: blood
- Ancestry: White European
- Sex: males and females
- Smoking: never, former, current
- Age: yes
- Array batch: yes
- Cells: yes

Input (current and former): results meta-analysis

Output (current and former): list FDR CpG, list FDR genes, gene-set, GO terms, and tissue/cell type enrichment plot and table

Tool: annotation / ClusterProfiler R package, eFORGE (online), EWAS catalogue/atlas (online)

Questions:

1. Which are the enriched GO terms in current smokers? and in former smokers?
2. Which are the enriched Reactome pathways in current smokers? and in former smokers?
3. Which are the enriched diseases in current smokers? and in former smokers?
4. Which are the enriched tissues in current smokers? and in former smokers?
5. Has the top CpG previously been reported in relation to smoking or other traits?

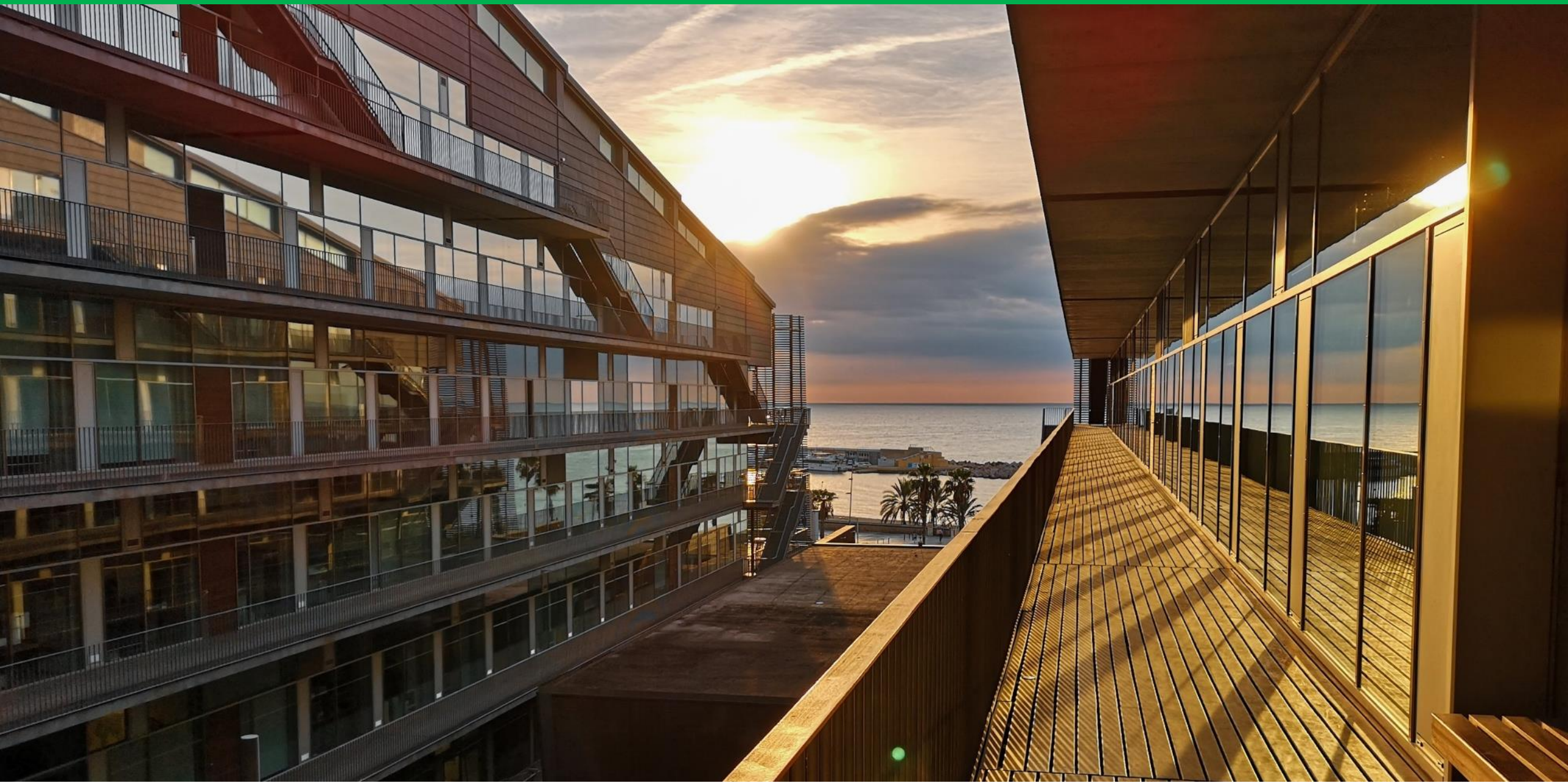
INTRODUCTION TO EPIGENOME-WIDE ASSOCIATION STUDIES (EWAS)

5. CLOUSURE OF THE COURSE AND COMMENTS

COMMENTS OR QUESTIONS

Any comments or questions?

Teachers:

- Juan Ramon González (juanr.gonzalez@isglobal.org)
- Mariona Bustamante (mariona.bustamante@isglobal.org)
- Laura Balagué (laura.balague@isglobal.org)
- Sofía Aguilar (sofia.aguilar@isglobal.org)

THANKS FOR TAKING PART IN THE COURSE!