Post-EWAS Platforms for Interrogating Results

Corina Lesseur

Assistant Professor

Dept. Environmental Medicine & Public Health

Icahn School of Medicine at Mount Sinai

Outline

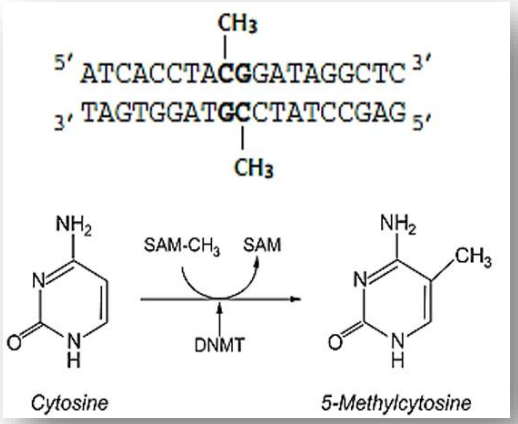
1. DNA methylation & EWAS

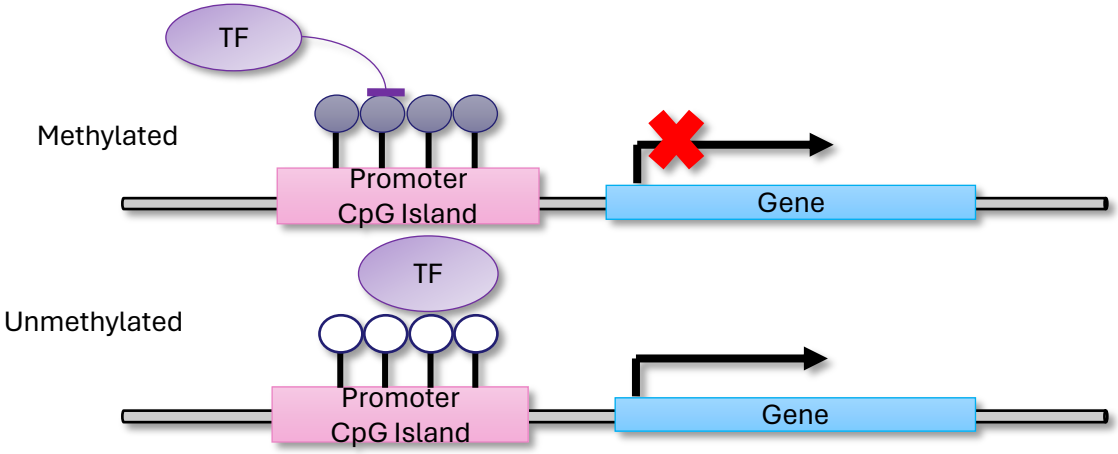
2. Post-EWAS analyses

- Overrepresentation & Gene Set Enrichment Analyses
- Other enrichment analyses
- Molecular quantitative trait loci
- Cell type proportions deconvolution & analyses
- Result annotation with genomics databases

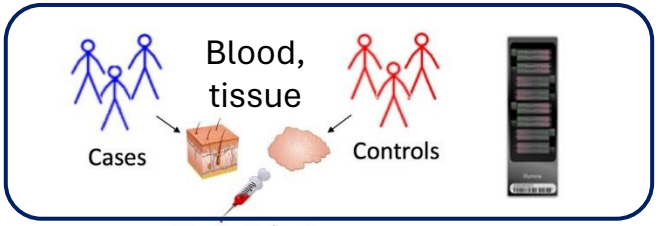
DNA methylation

- Epigenetic marks that do not involve a change sequence and regulate gene expression.
- Addition of a CH_3 to cytosines (5mC) within CpG.
- Essential for cell type differentiation & development.
- Highly *cell-type-specific*.
- Influenced by genotype & environmental exposures (i.e., smoking).
- DNAm changes with age.





Epigenome-wide Association Studies (EWAS)



Infinium
MethylationEPIC
array

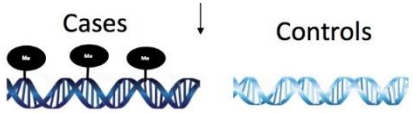
Preprocessing, QC, Normalization

Minfi
ChAMP
wateRmelon

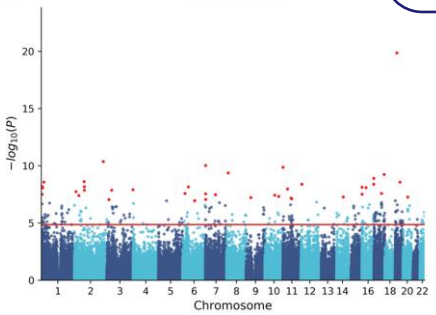
Bioconductor

Minfi
Methylumi
limma

Differential methylation analyses

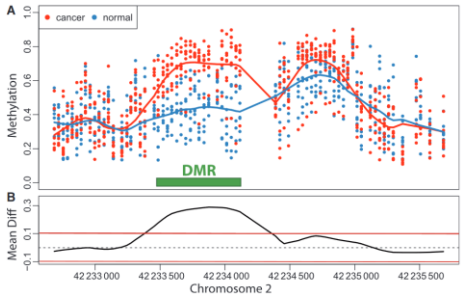


Bumphunter
DMRcate
ipDMR



CpGs Regions

?

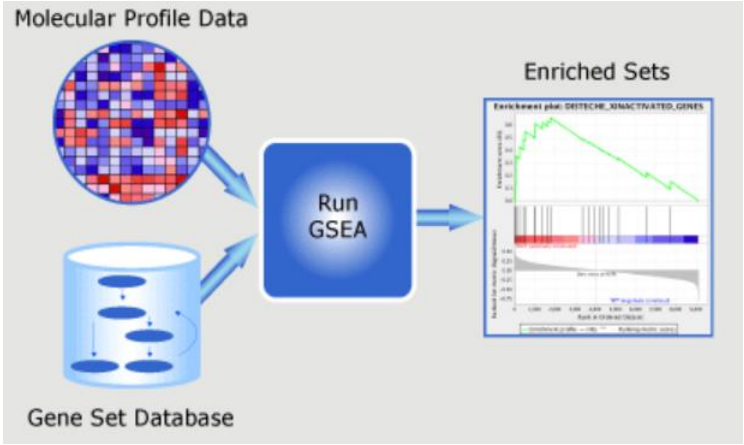


https://en.wikipedia.org/wiki/Epigenome-wide_association_study

Jaffe et al. 2012 Int J Epidemiol. PMID:22422453
Campagna et al 2021 Clin Epigenetics. PMID: 34863305.

Overrepresentation & Gene Set Enrichment Analyses

- Gain a systems-level understanding of the changes in methylation/expression by examining the pathways or gene sets involved.
- Gene sets and pathways are pre-defined in exiting databases (GO, Kegg, Molecular Signatures Database (MSigDB).
- General approaches: Over-Representation Analysis & Functional Class Scoring.



https://www.gsea-msigdb.org/gsea/index.jsp/

Overrepresentation & Gene Set Enrichment Analyses

https://geneontology.org

https://bioconductor.org/packages/release/data/annotation/html/GO.db.html

https://www.genome.jp/kegg/

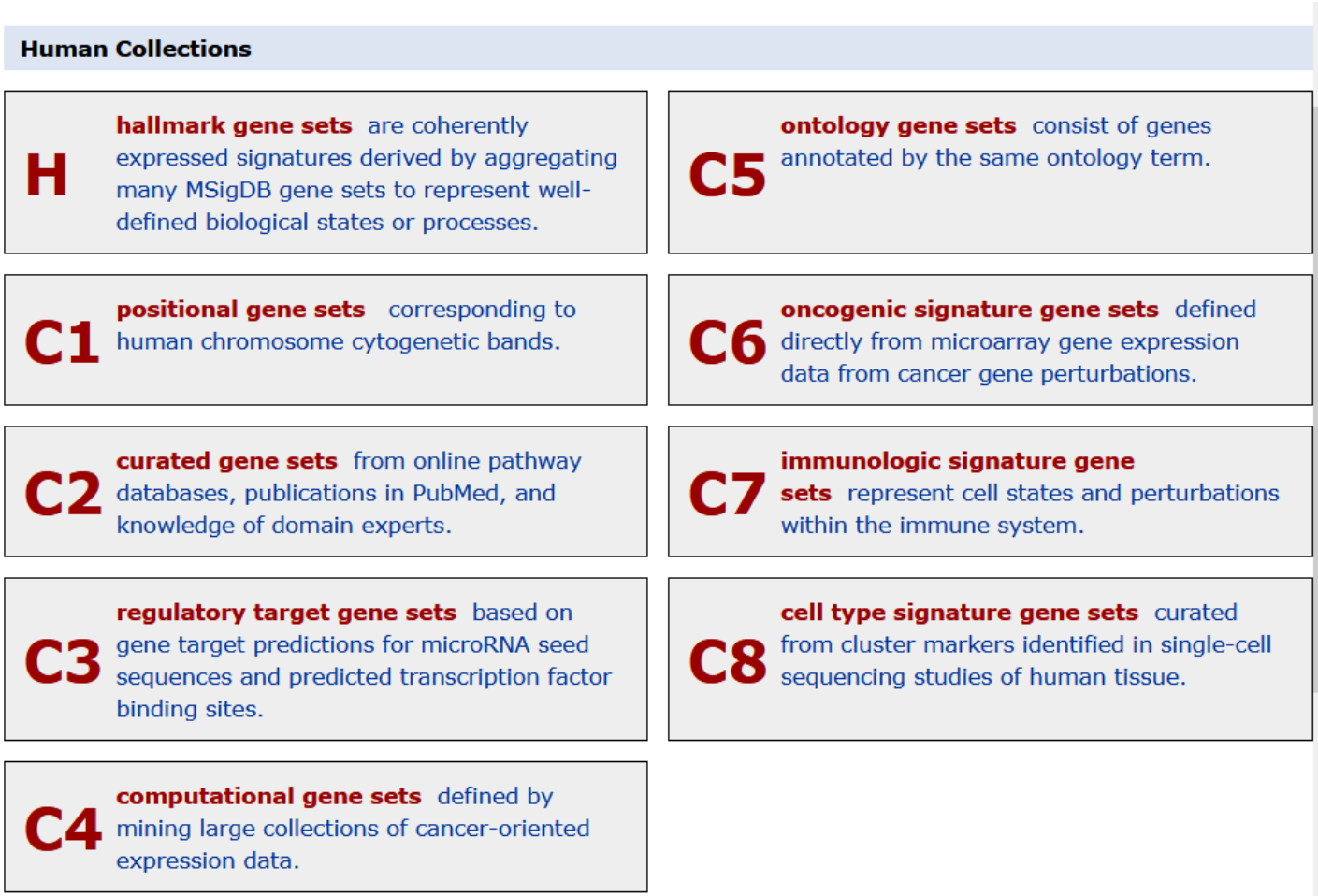
https://www.bioconductor.org/packages//2.12/data/annotation/html/KEGG.db.html

The Molecular Signatures Database (MSigDB)

34,550 gene sets

9 collections

Human & mouse



https://www.gsea-msigdb.org/gsea/msigdb

https://bioconductor.org/packages/release/data/experiment/vignettes/msigdb/inst/doc/msigdb.html

Gene set enrichment analysis for DNA methylation data

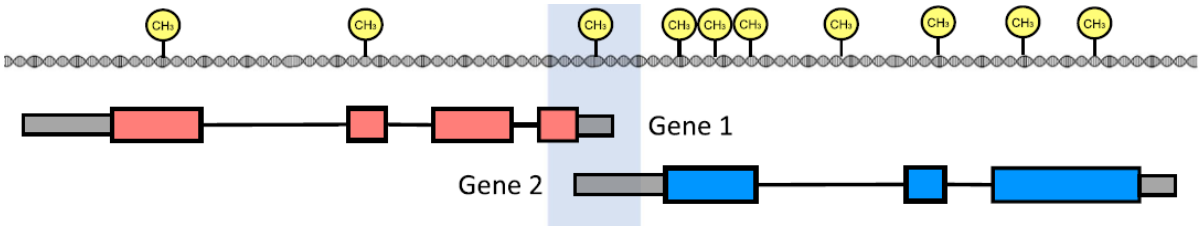
- Methylation occurs anywhere on the genome and is not as directly related to expression.
- How to assign CpGs tp genes?
- Annotate CpGs to genes →differentially methylated →gene list for enrichment.

Probe number bias:

genes ≠ numbers of CpGs.

Multiple annotation bias:

CpG annotated to >1 gene.

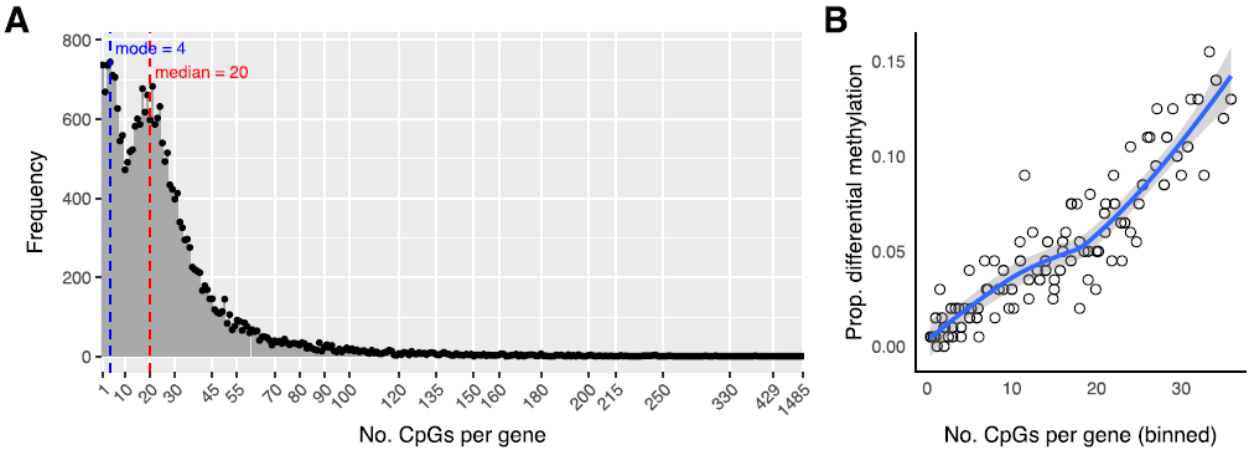


R Bioconductor

MissMethyl::gseameth

ChAMP::ebGSEA

MehylGSA



missMethyl

- CpG list enrichment based on Wallenius' noncentral hypergeometric test.
- *Accounts for probe number and multi-gene biases using weights*.
- Can subset the CpG list to specific genomic regions (i.e. promoters).
- **gometh**: List CpGs with GO.db and Kegg .db annotation packages
- **gsameth**: List CpGs with *other gene sets*.
- **goregion**: List DMRs with GO.db and Kegg .db annotation packages
- **gsaregion**: List DMRs with *other gene sets*.

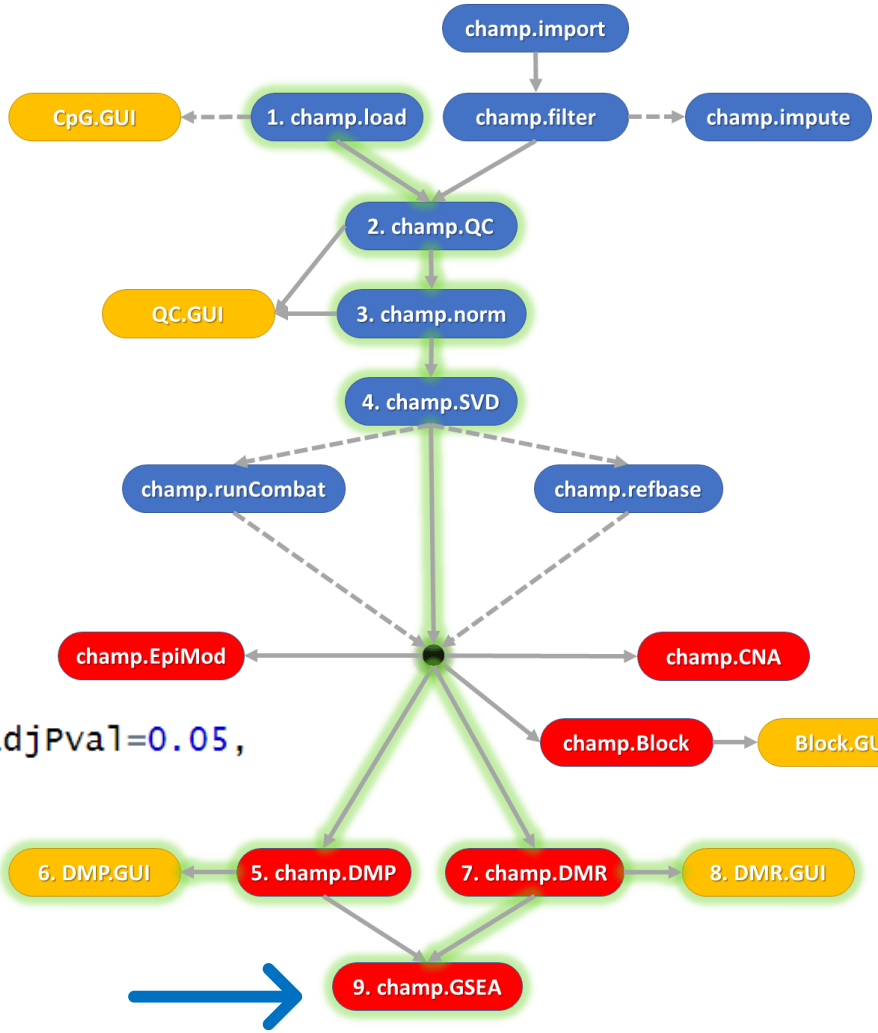
```
> gometh(sigcpgs,all.cpg=NULL,collection = c("GO", "KEGG"),
  array.type="450K",plot.bias = FALSE,prior.prob = TRUE,
  anno = NULL,equiv.cpg = TRUE,fract.counts = TRUE,
  genomic.features = c("ALL", "TSS200", "TSS1500", "Body",
  "1stExon", "3'UTR", "5'UTR","ExonBnd"),sig.genes = FALSE)
```

https://bioconductor.org/packages/release/bioc/html/missMethyl.html
http://oshlacklab.com/methyl-geneset-testing/index.html

Maksimovic et al. 2021 Genome Biology (PMID: 34103055)

ebGSEA

- Adapts a global test to directly rank genes by overall differential methylation level (using all CpGs for that gene).
- **champ.GSEA**



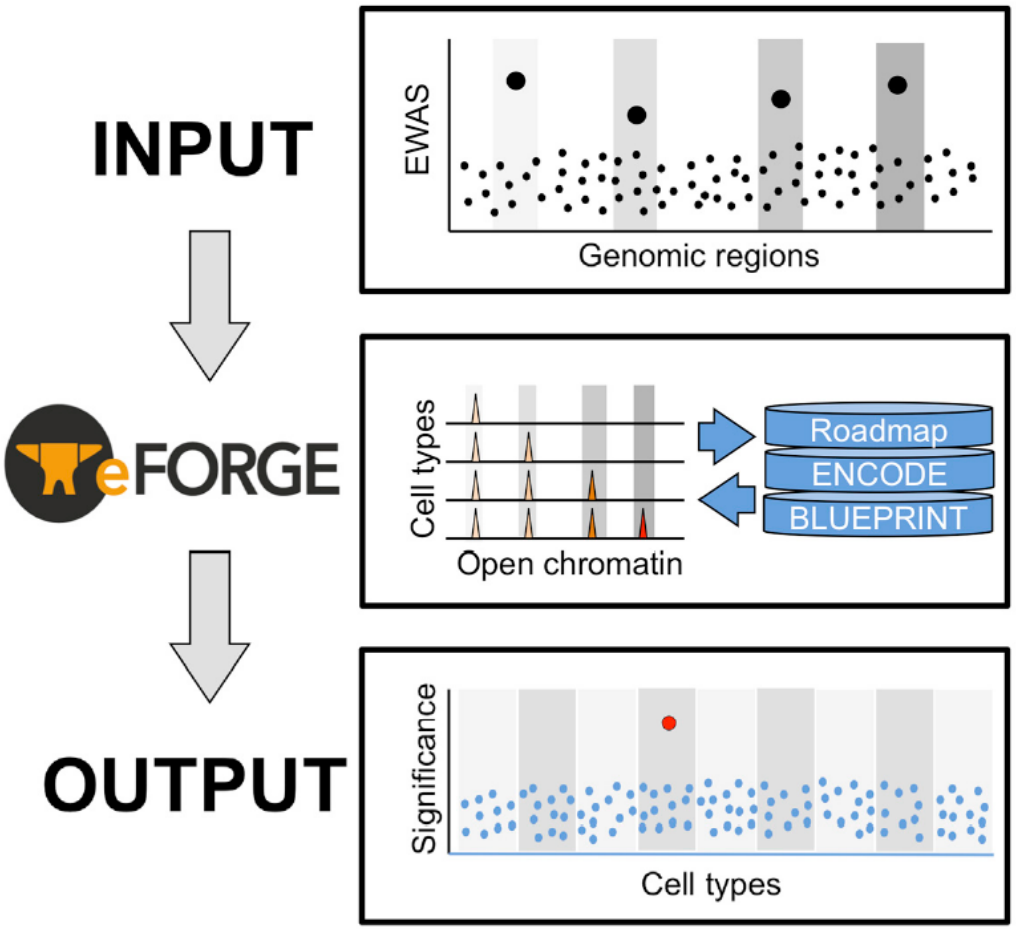

```
> champ.GSEA(beta=myNorm, arraytype="450K",adjPval=0.05,
  method="ebayes",cores=5)
```

https://github.com/aet21/ebGSEA
https://www.bioconductor.org/packages/release/bioc/html/ChAMP.html

Dong et al. 2019 Bioinformatics (PMID: 30715212)

eFORGE

- Experimentally derived Functional element Overlap analysis of ReGions from EWAS.
- Standalone (Perl /python) & web-based tool.
- Detects enrichment of DNase I hypersensitive sites & chromatin states reference samples (tissues, primary cell types, & cell lines)
- Input CpG probes or bed file.



https://eforge.altiusinstitute.org/
https://eforge-tf.altiusinstitute.org/

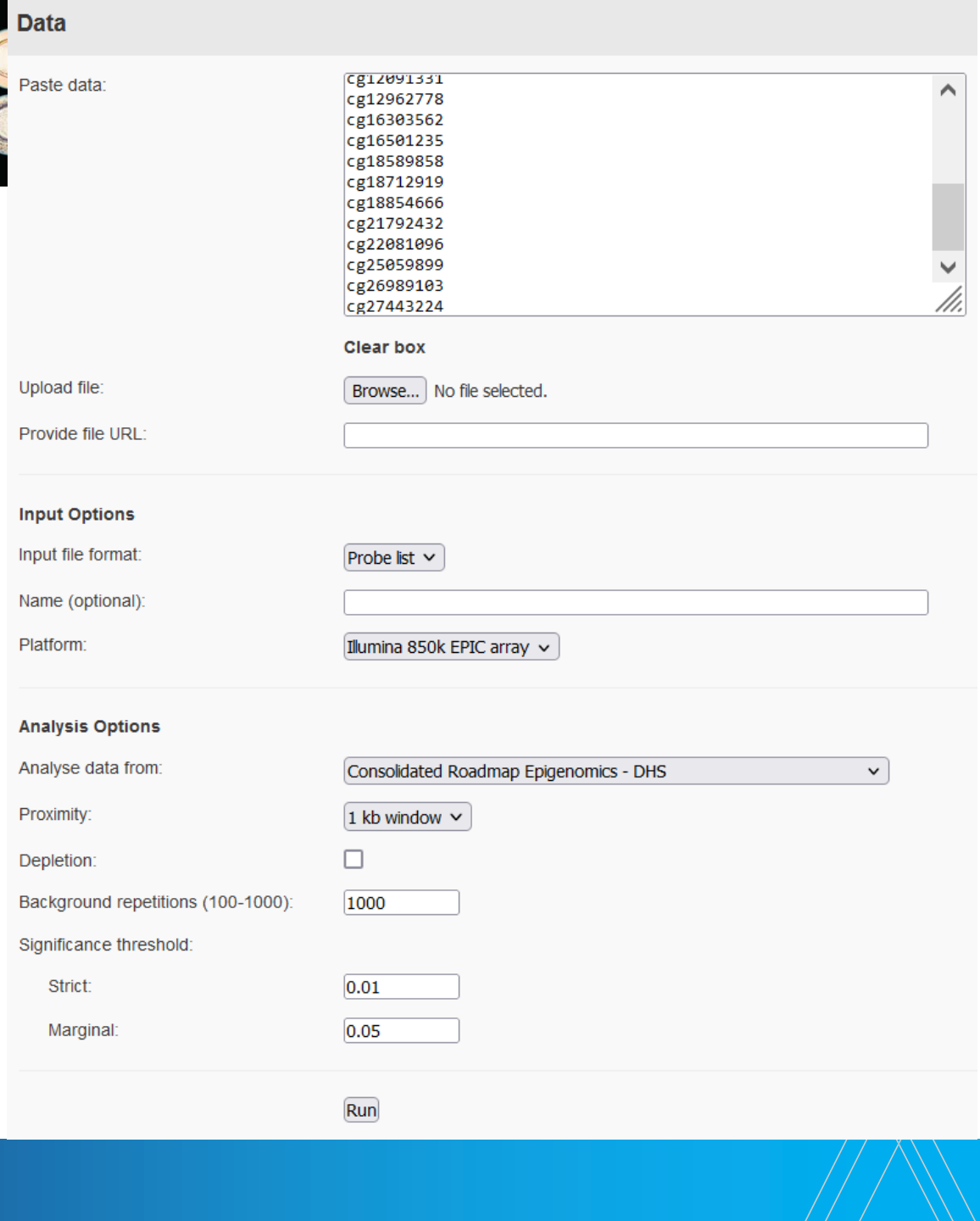
Breeze et al 2016 Cell Rep. PMID: 27851974;
Breeze et al 2019 Bioinformatics PMID: 31161210.

eFORGE

A Tool for Identifying Cell Type-Specific Signal in Epigenomic Data

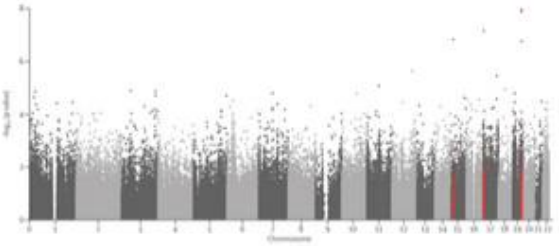
Web version max. 1K probes

https://eforge.altiusinstitute.org/



Breeze et al 2016 Cell Rep. PMID: 27851974;
Breeze et al 2019 Bioinformatics PMID: 31161210.

eFORGE



Published EWAS

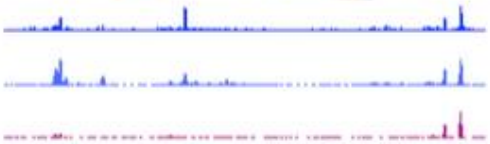
Match 1000 times

485,512 probes

Gene relationship
CpG island relationship

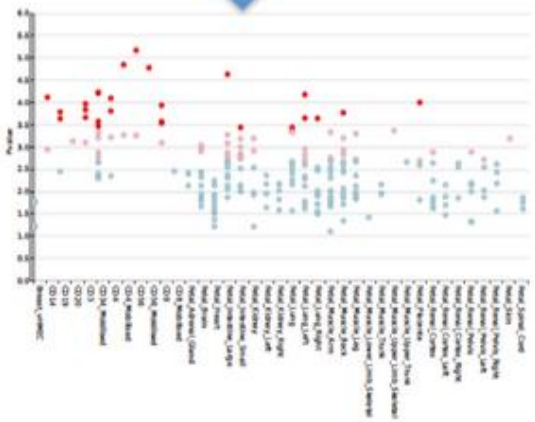
Overlap

Overlap



DNase hotspots from ENCODE or Roadmap

Express enrichment as –log10(p value) of the EWAS MVPs compared to the 450k array probes

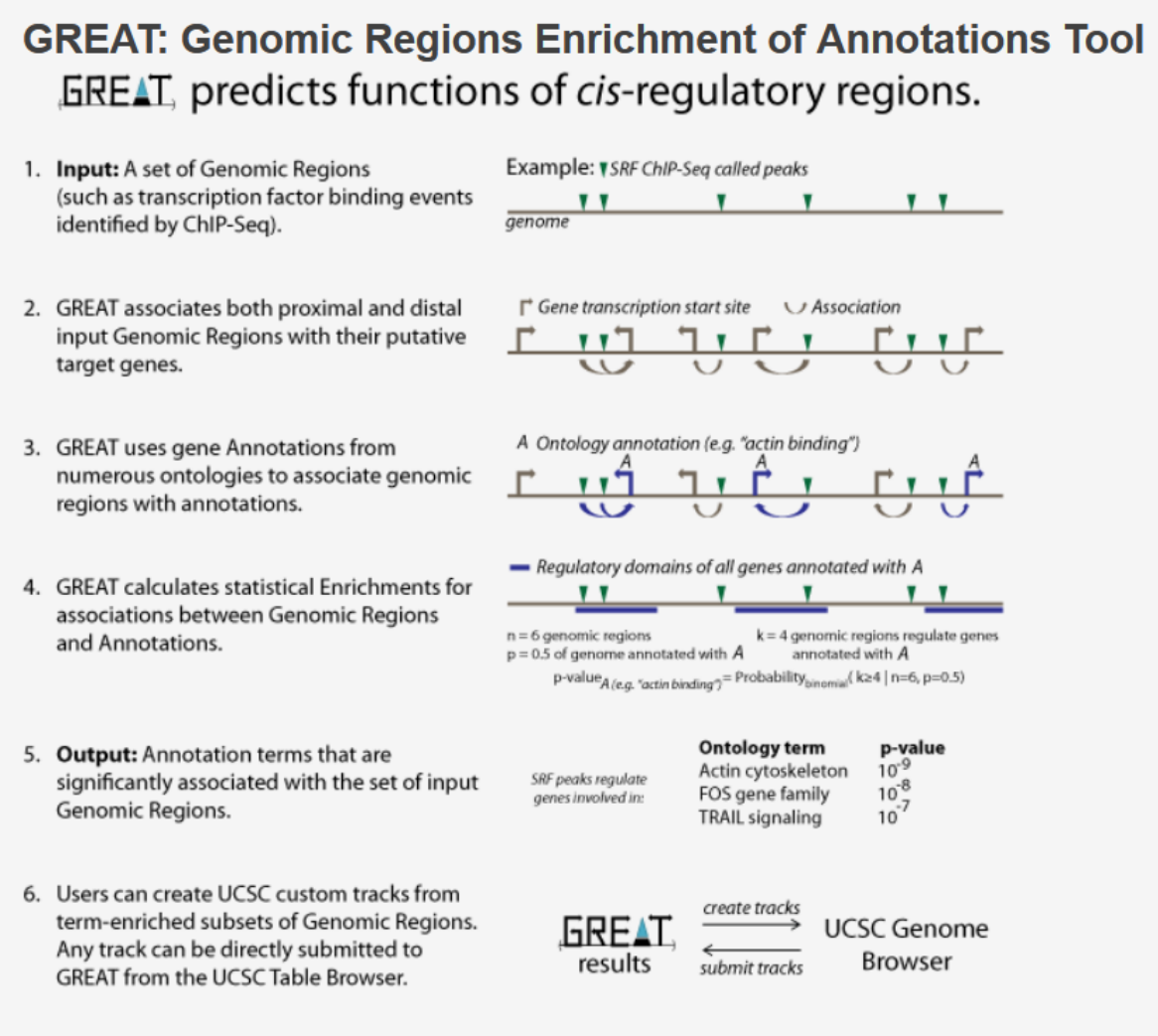


*Manhattan plot from BMI Dick et al. EWAS, Lancet, 2014

https://eforge.altiusinstitute.org/?documentation

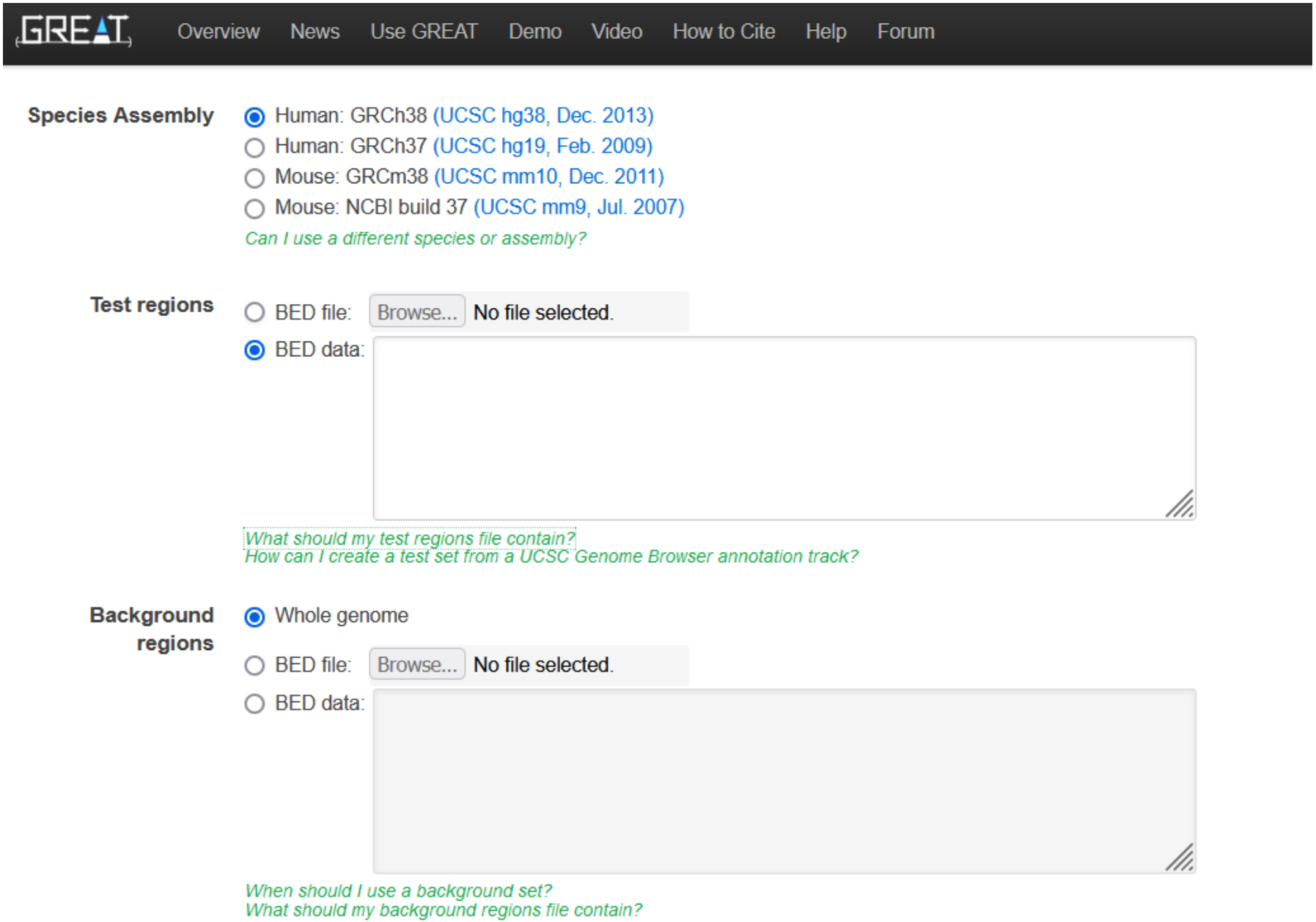
Breeze et al 2016 Cell Rep. PMID: 27851974

GREAT: Genomic Regions Enrichment of Annotations Tool



http://great.stanford.edu/public/html/splash.php

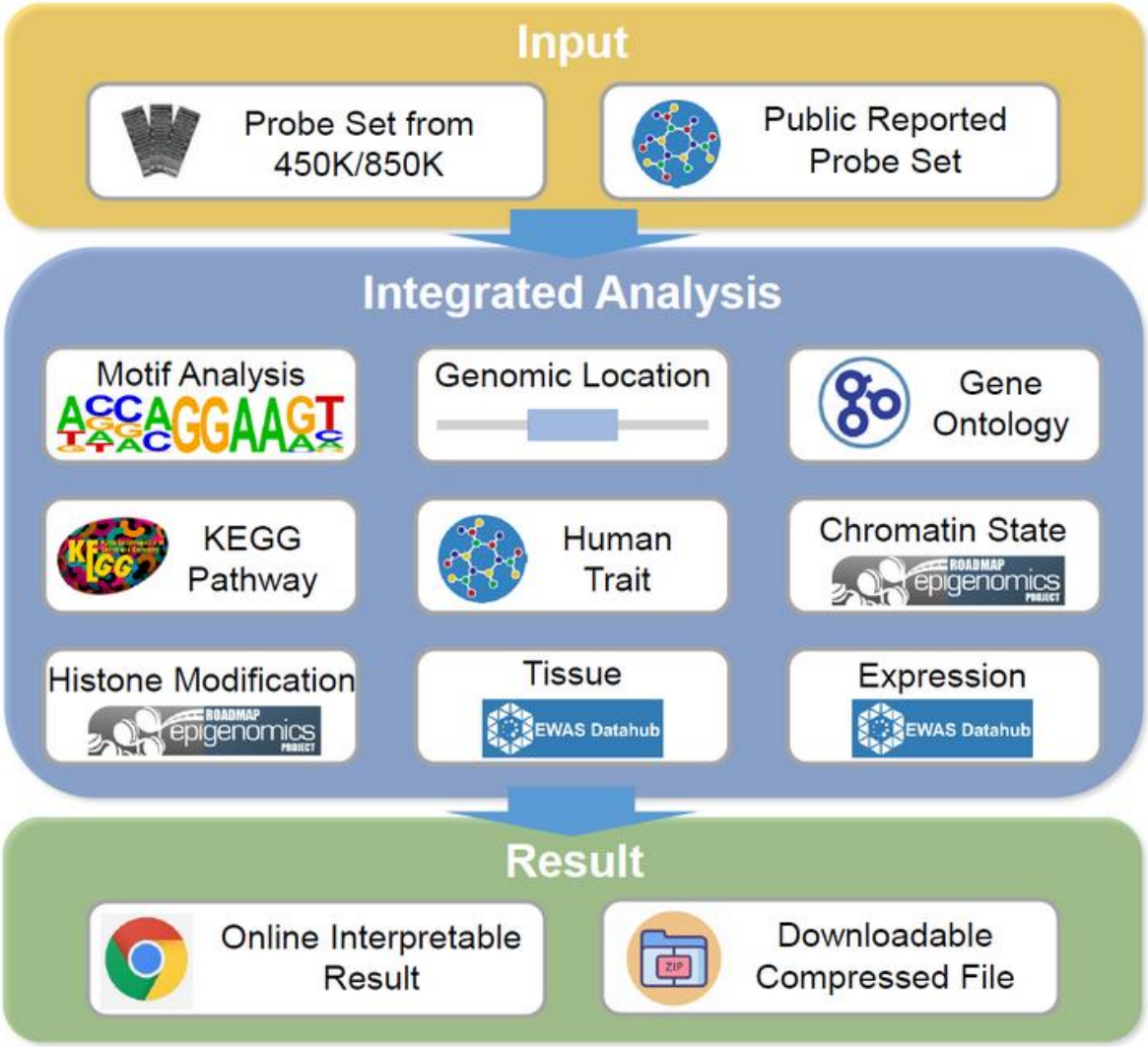
McLean et al 2010 Nat Biotechnol. PMID: 20436461



http://great.stanford.edu/public/html/splash.php

McLean et al 2010 Cell Rep. PMID: 27851974

EWAS Toolkit
EWAS Open Platform

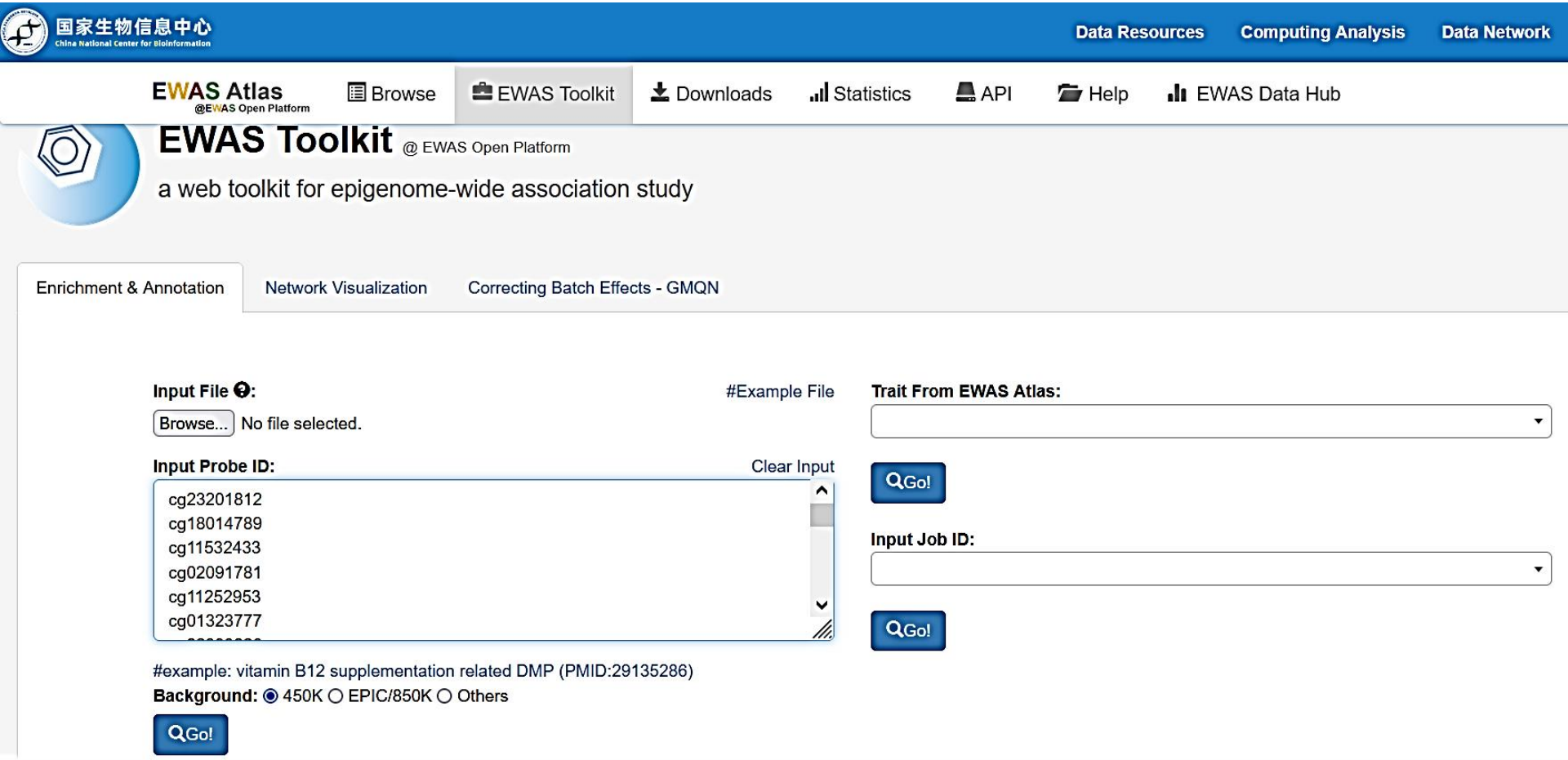


https://ngdc.cncb.ac.cn/ewas/

Xiong et al 2022. Nucleic Acids Res. PMID: 34718752

EWAS Toolkit
EWAS Open Platform

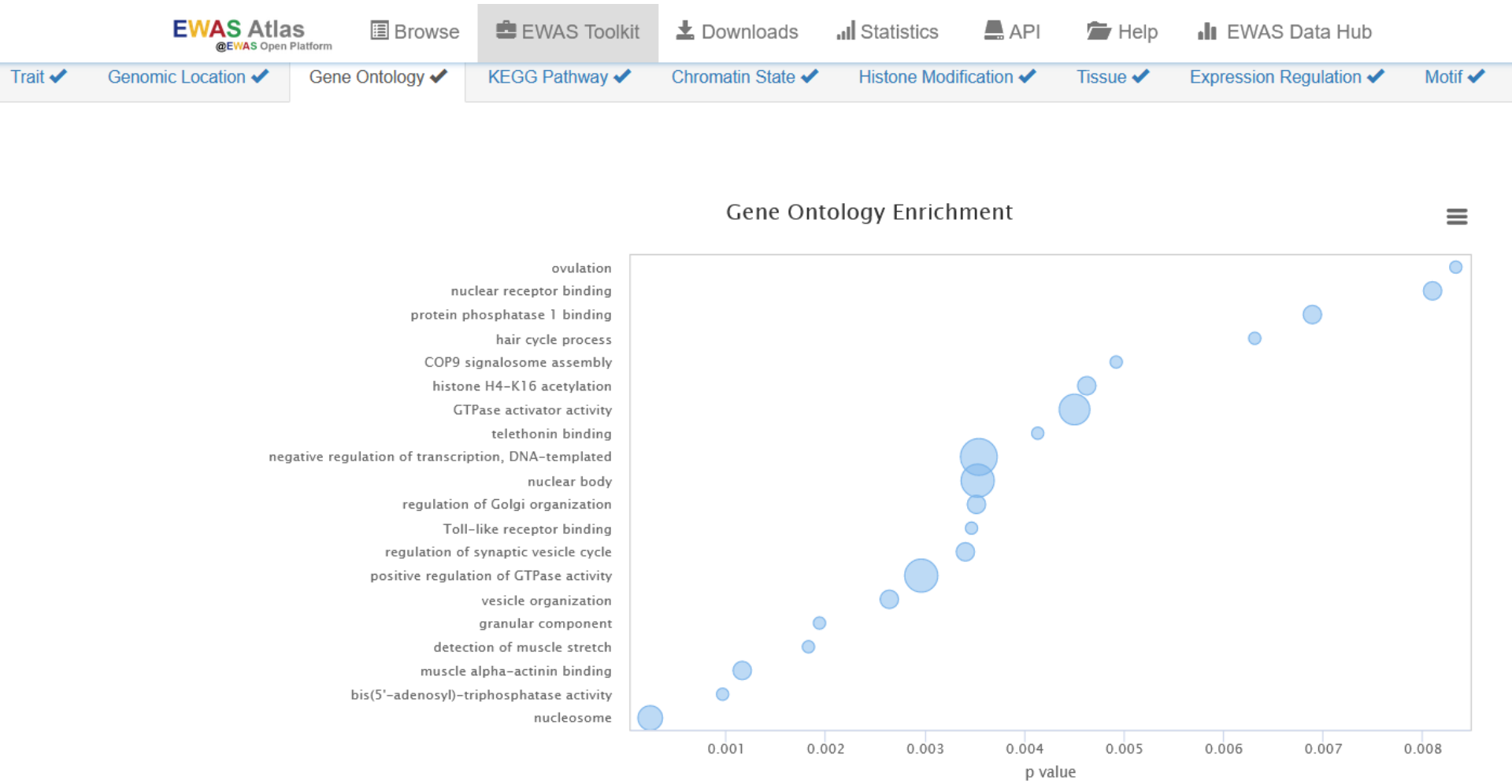
Web toolkit for epigenome-wide association studies



https://ngdc.cncb.ac.cn/ewas/

Xiong et al 2022. Nucleic Acids Res. PMID: 34718752

EWAS Toolkit
EWAS Open Platform



Xiong et al 2022. Nucleic Acids Res. PMID: 34718752

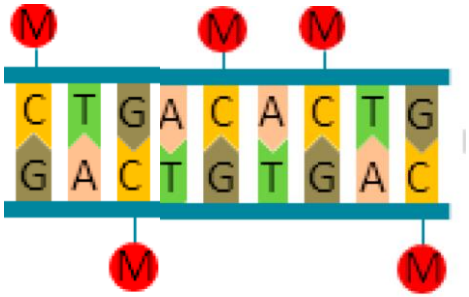
https://ngdc.cncb.ac.cn/ewas/documentation#/toolkit

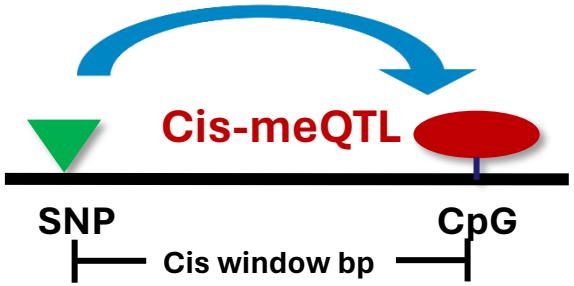
Outline

1. DNA methylation & EWAS
2. Post-EWAS analyses
   - Overrepresentation & Gene Set Enrichment Analyses
   - Other enrichment analyses
   - Molecular quantitative trait loci
   - Cell type proportions deconvolution & analyses
   - Result annotation with other genomics databases

Molecular Quantitative Trait Loci (molQTL)

- Methylation quantitative trait loci (meQTL)
- ≈ 45% of CpGs in 450K array



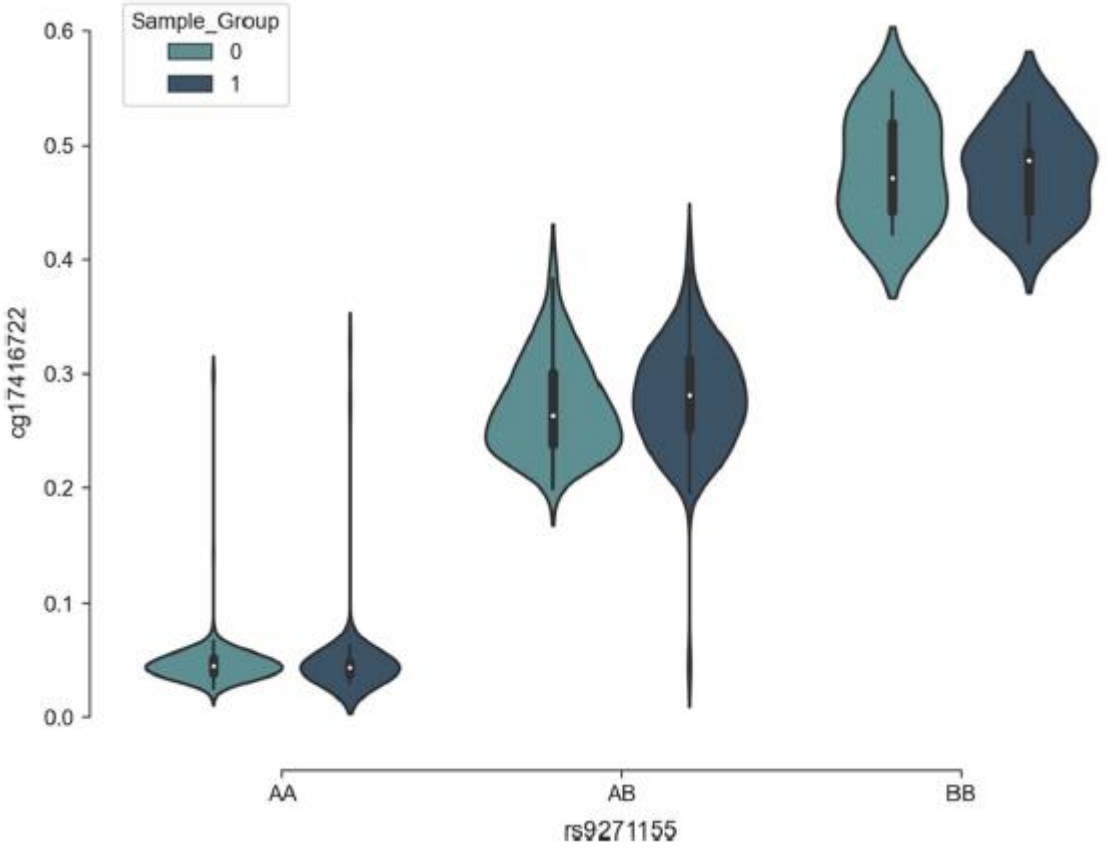


SNP	%Meth.
AA	30
AG	35
GG	45

Molecular Quantitative Traits	
SNP → methylation	**mQTL or meQTLs**
SNP → gene expression	**eQTLs**
5mC → expression	**eQTM**

Villicaña and Bell 2021. Genome Biol. PMID: 33931130

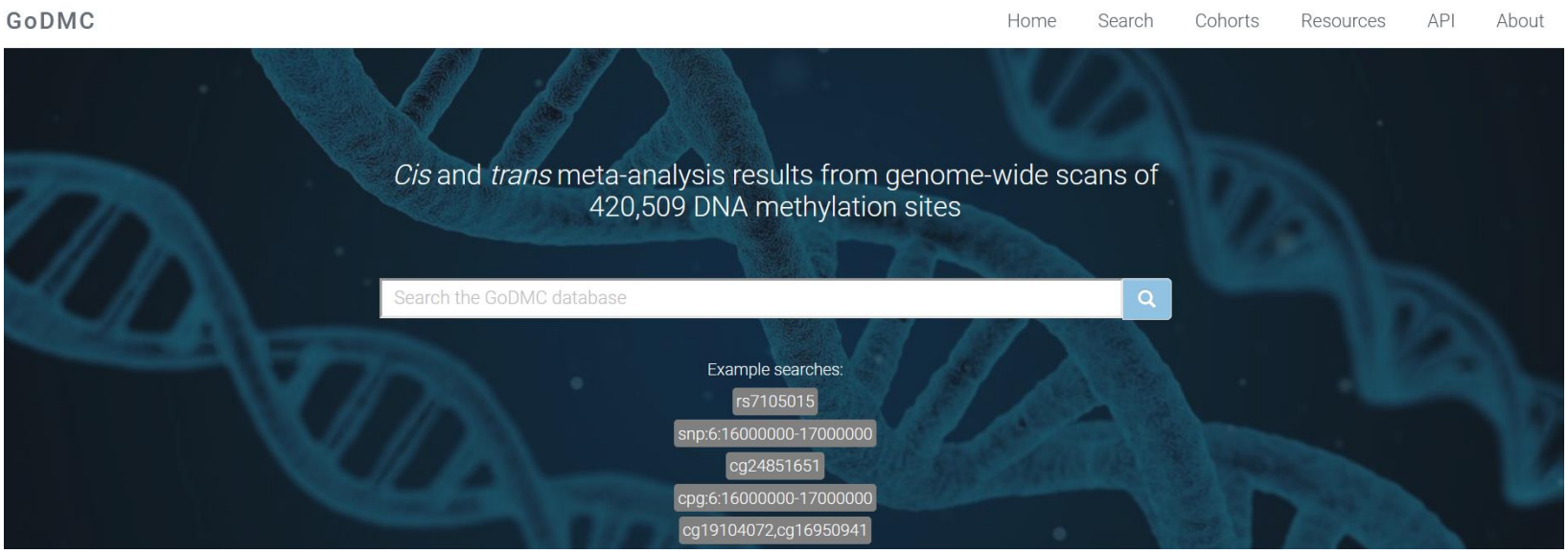
Methylation Quantitative Trait Loci



Campagna et al 2021 Clin Epigenetics. PMID: 34863305.

Genetics of DNA Methylation Consortium (GoDMC)

- A multi-cohort meta-analysis of **blood meQTLs**
- >32,851 participants
- Database http://mqtldb.godmc.org.uk/
- API http://api.godmc.org.uk/v0.1

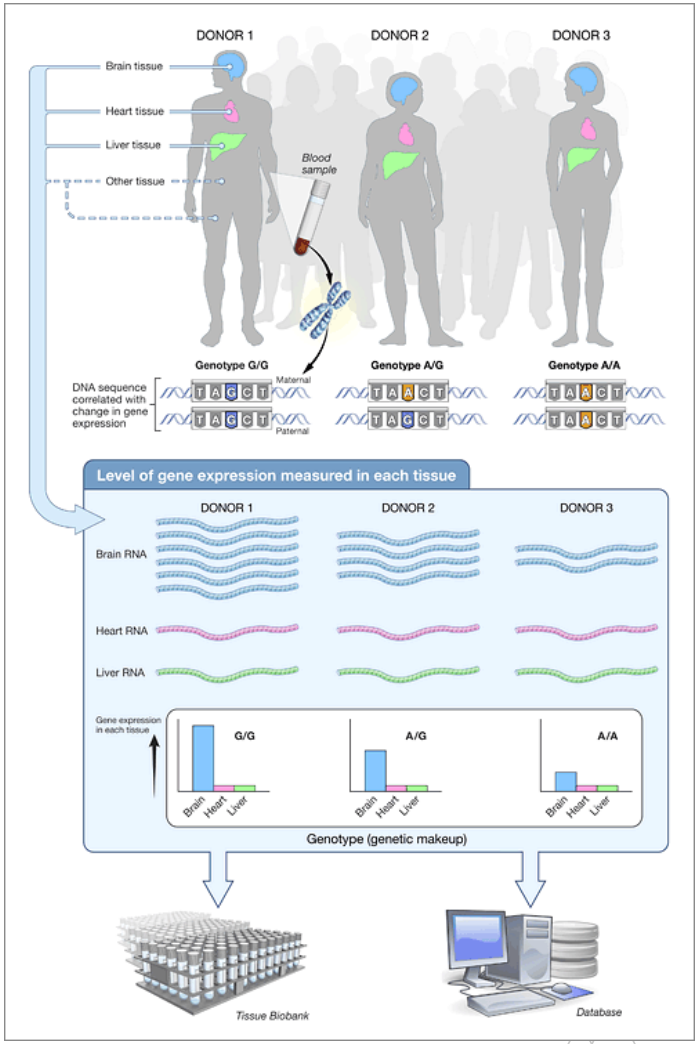


http://www.godmc.org.uk/

Min et al 2021 Nat Genet. PMID: 34493871

The Genotype-Tissue Expression (GTEx) Project

- Ongoing effort to build a comprehensive public resource to study tissue-specific gene expression and regulation.
- Samples were collected from 54 non-diseased tissue sites across nearly 1000 individuals.



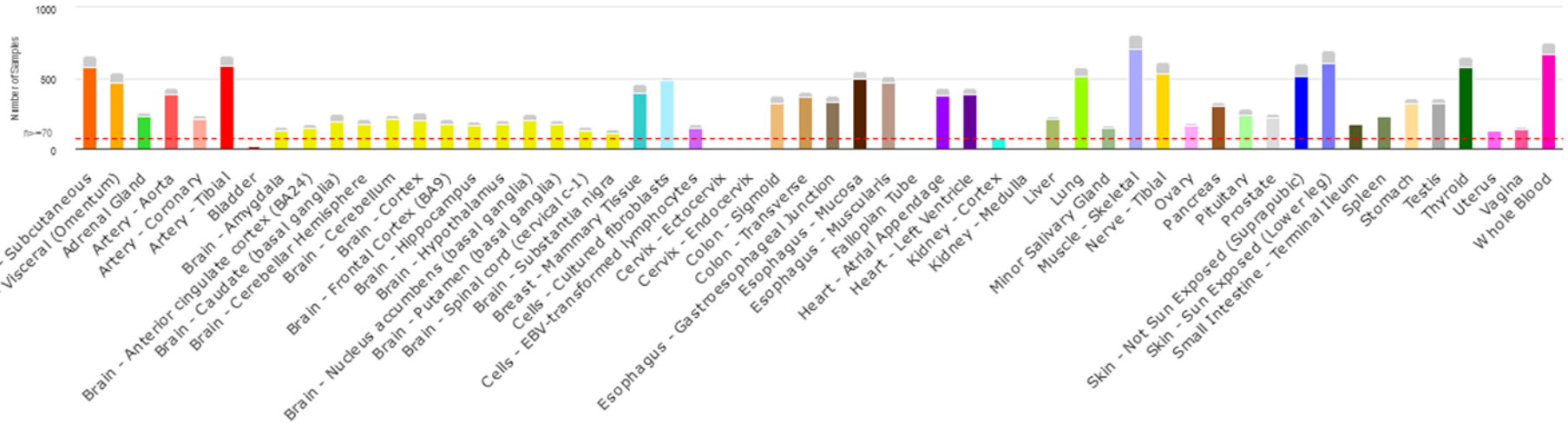
GTEx Portal

Home Downloads Expression Single Cell QTL IGV Browser Tissues & Histology

https://gtexportal.org/home/

The GTEx Consortium. Science. 2015 May 8;348(6235):648-60

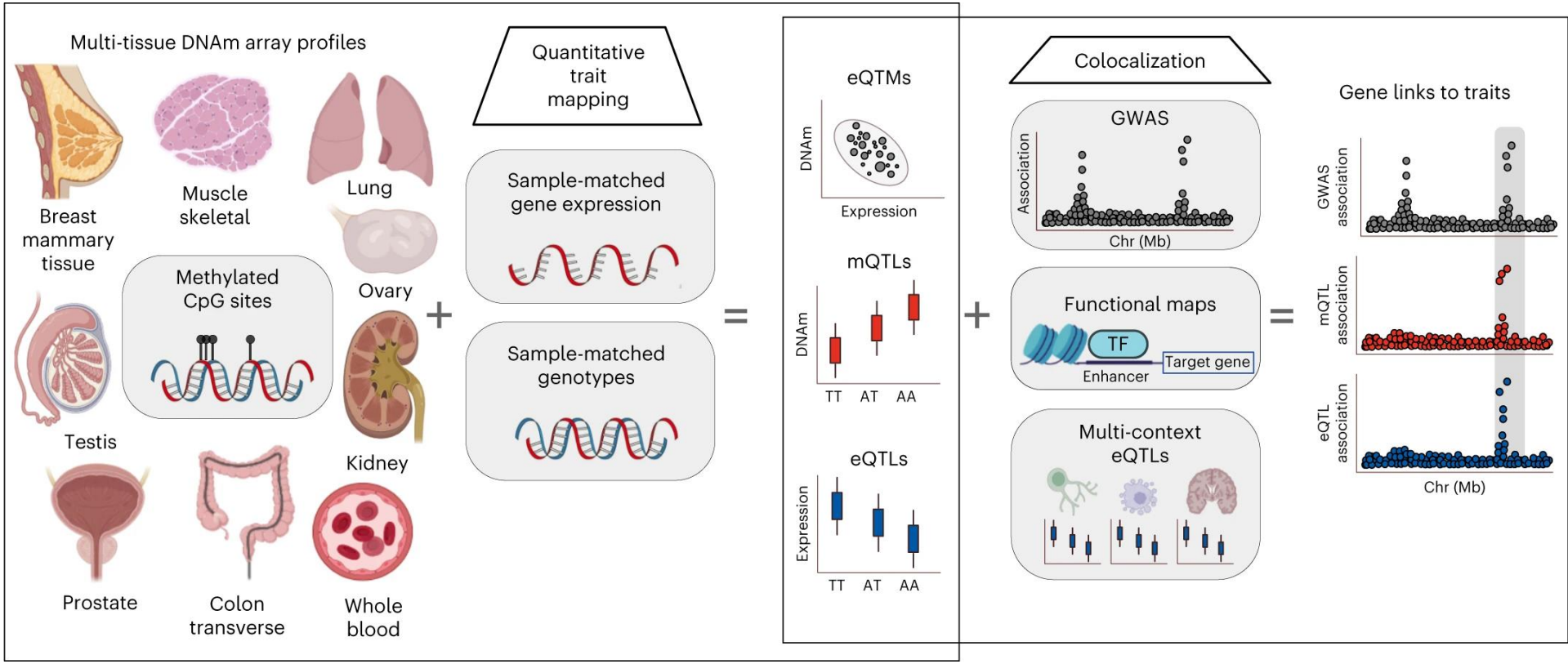
The Genotype-Tissue Expression (GTEx) Project



https://gtexportal.org/home/

https://gtexportal.org/home/downloads/adult-gtex/qtl

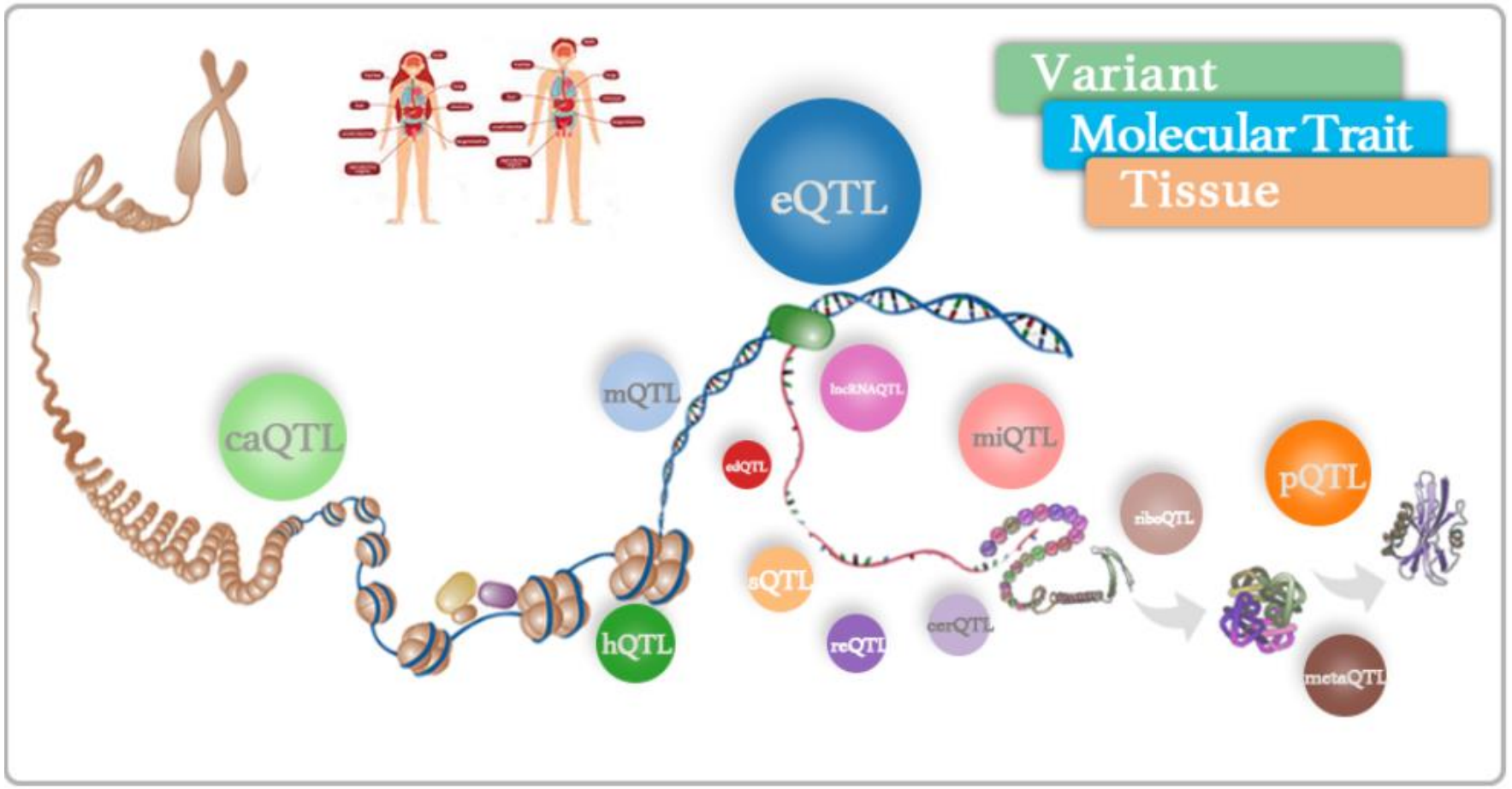
DNA methylation QTL mapping across diverse human tissues



https://gtexportal.org/home/downloads/egtex/methylation

Oliva et al. 2022 Nat Gen. PMID: 36510025

QTLbase



http://www.mulinlab.org/qtlbase

Zheng et al 2020 Nucleic Acids Res PMID: 31598699

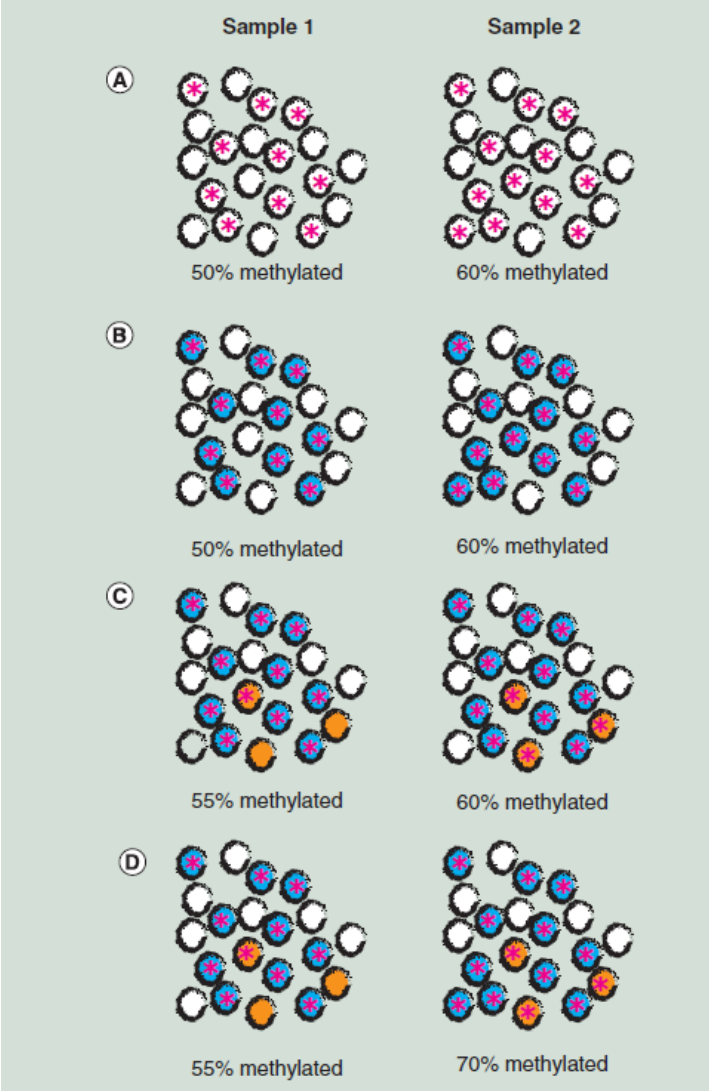
Outline

1. DNA methylation & EWAS
2. Post-EWAS analyses
   - Overrepresentation & Gene Set Enrichment Analyses
   - Other enrichment analyses
   - Molecular quantitative trait loci
   - Cell type proportions deconvolution & analyses
   - Result annotation with other genomics databases

Cellular Heterogeneity in EWAS

- Most biological samples are mixtures of cells with distinct methylation patterns.
- EWAS aim to identify DMPs, cell type proportions may also vary between cases and controls.
- Cell-type deconvolution methods:
  - Reference-free
  - Reference-based

R packages: EpiDISH, RefFreeEWAS

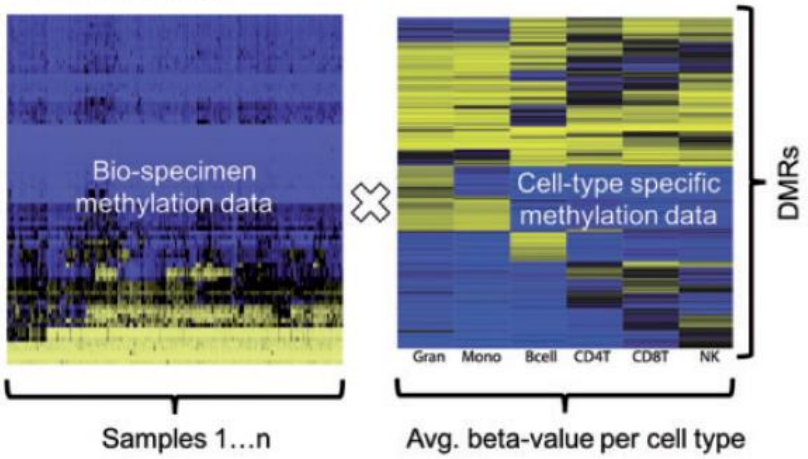


PMID: 28749184

Houseman et al. 2012 Bioinformatics 22568884
Houseman et al. 2019 Bioinformatics PMID: 27358049
Zheng et al 2018. Nat Methods. PMID: 30504870.

Cell proportion estimates in analyses

1. **Reference-based cell type deconvolution using immune cell DMRs (e.g. Houseman method)**



Result: a matrix of samples with estimated immune cell type proportions

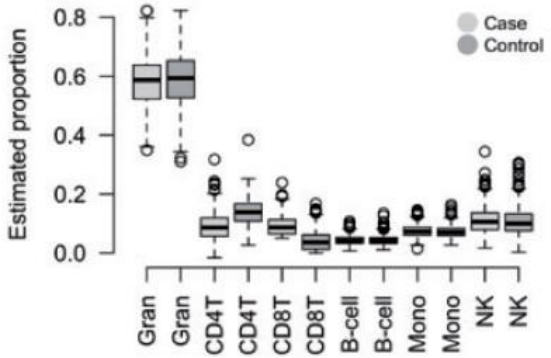
	Gran	Mono	B-cell	CD4T	CD8T	NK
Samples 1…n	$Gran_1$	$Mono_1$	$B\text{-}cell_1$	$CD4T_1$	$CD8T_1$	NK_1
	$Gran_2$	$Mono_2$	$B\text{-}cell_2$	$CD4T_2$	$CD8T_2$	NK_2
	…	…	…	…	…	…
	$Gran_n$	$Mono_n$	$B\text{-}cell_n$	$CD4T_n$	$CD8T_n$	NK_n

Immune proportion estimates for samples 1…n

2. **Use cell type proportion estimates in analyses of outcome or phenotype of interest.**

Examples:

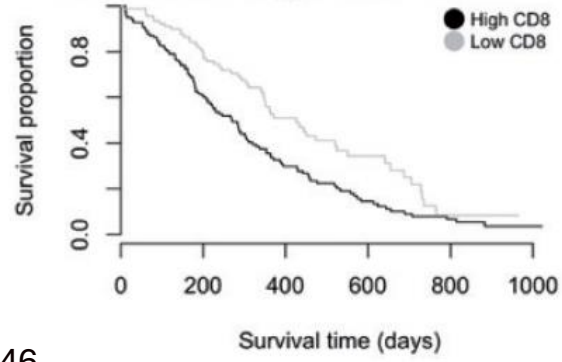
i. Compare cell type proportions in cases with controls



ii. Calculate cell type ratio (e.g. NLR), and test association of NLR with variables of interest (e.g. survival)

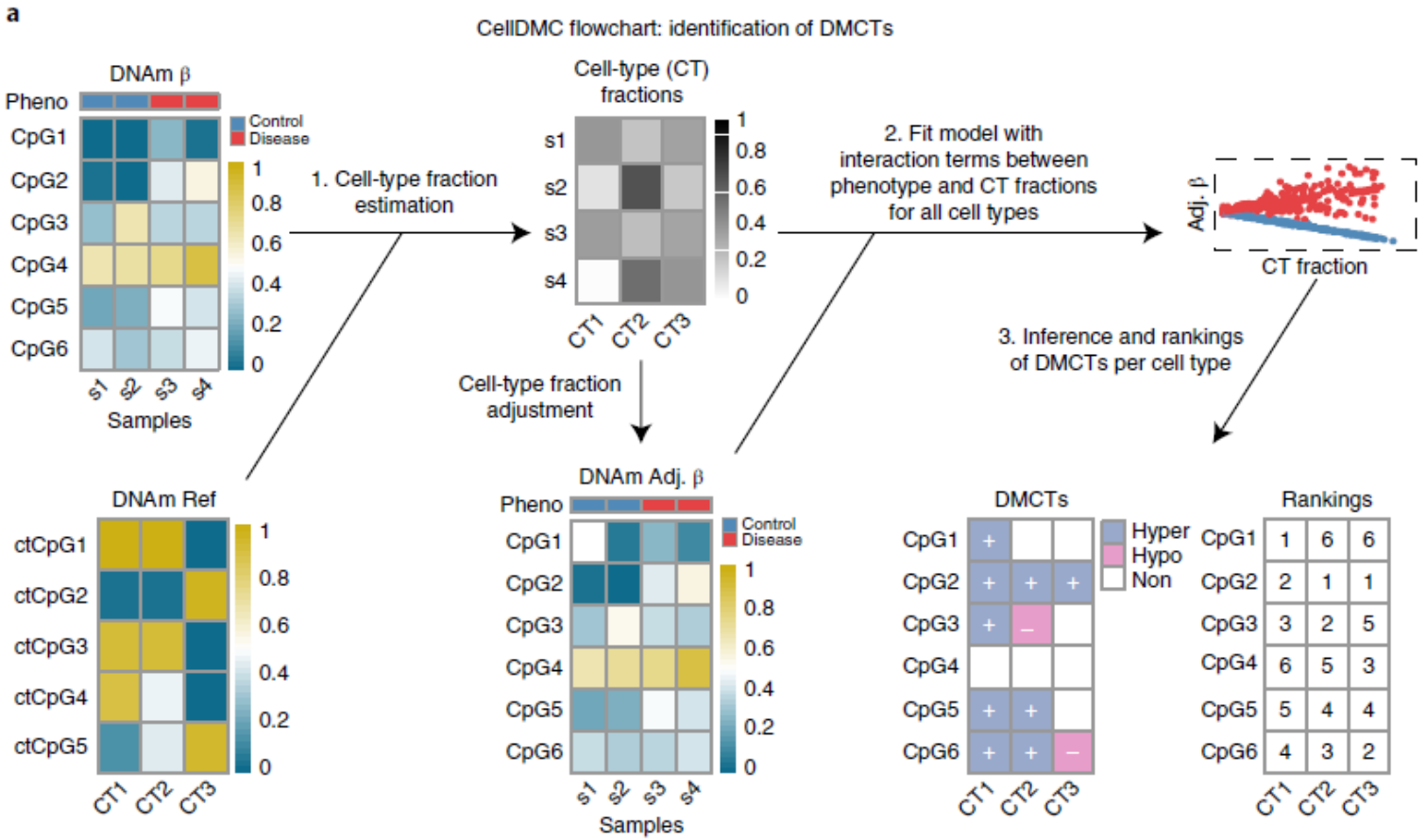
$$\text{NLR} = \frac{\text{Proportion of Granulocytes}}{\text{Proportion of Lymphocytes (B-cells + CD4T +CD8T + NK)}}$$

iii. Test relation of cell type proportion (e.g. CD8) or ratio with survival outcome (e.g. overall survival)



Titus et al 2017 Hum Mol Genet. PMID: 28977446

Identification of differentially methylated cell types in EWAS



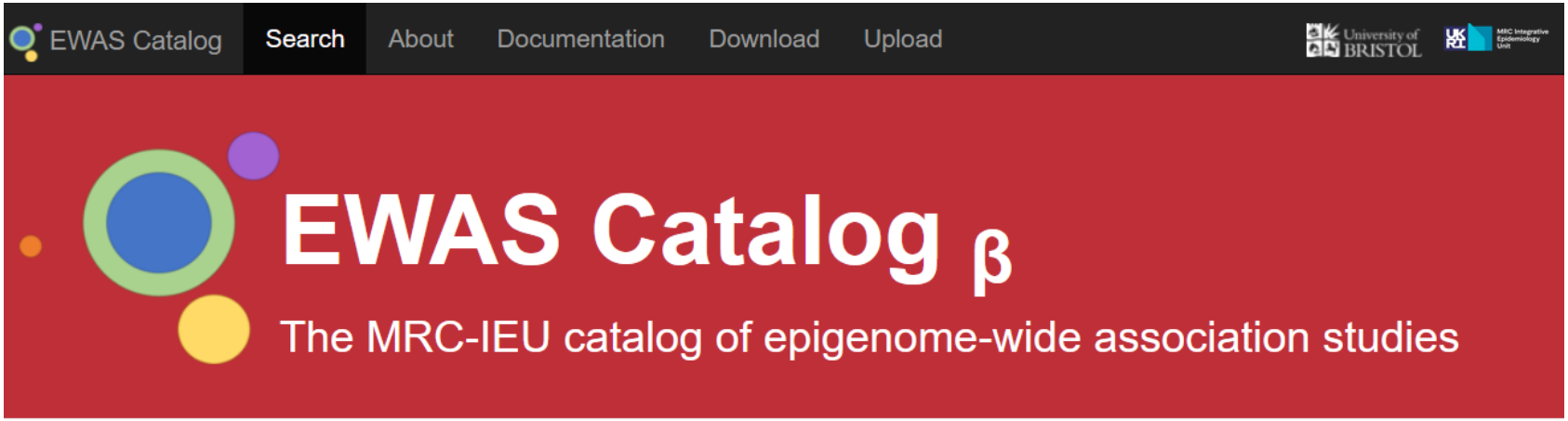
EpiDISH::CellDMC

Zheng et al 2018. Nat Methods. PMID: 30504870.

Outline

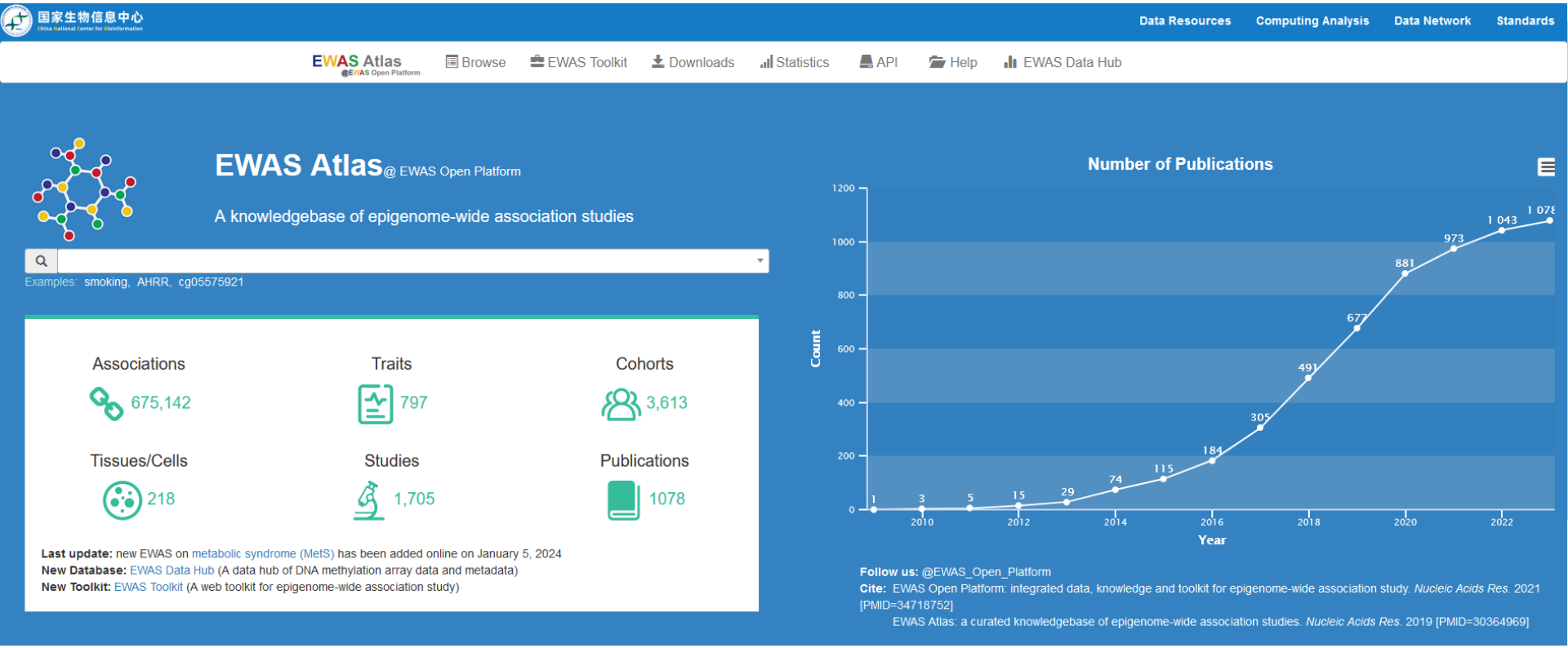
1. DNA methylation & EWAS
2. Post-EWAS analyses
   - Overrepresentation & Gene Set Enrichment Analyses
   - Other enrichment analyses
   - Molecular quantitative trait loci
   - Cell type proportions deconvolution & analyses
   - Result annotation with other genomics databases

EWAS Catalog



https://www.ewascatalog.org/

EWAS Atlas
EWAS Open Platform



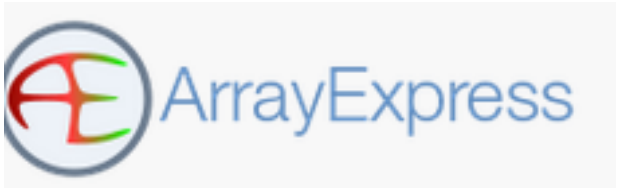
https://ngdc.cncb.ac.cn/ewas/

Xiong et al 2022. Nucleic Acids Res. PMID: 34718752

Public genomics data repositories

https://www.ncbi.nlm.nih.gov/geo/

https://bioconductor.org/packages/release/bioc/html/GEOquery.html



https://www.ebi.ac.uk/biostudies/arrayexpress

https://ngdc.cncb.ac.cn/ewas/datahub/index

Questions