Searching EWAS catalog
https://www.ewascatalog.org/

Standard operating procedure 4

EWAS catalog publication

REVISED **The EWAS Catalog: a database of epigenome-wide association studies [version 2; peer review: 2 approved]**

Thomas Battram ✉, Paul Yousefi iD, Gemma Crawford, Claire Prince, Mahsa Sheikhali Babaei,

Gemma Sharp, Charlie Hatcher, María Jesús Vega-Salas, Sahar Khodabakhsh,

Oliver Whitehurst iD, Ryan Langdon, Luke Mahoney, Hannah R. Elliott iD, Giulia Mancano,

Matthew A. Lee iD, Sarah H. Watkins, Abigail C. Lay iD, Gibran Hemani iD, Tom R. Gaunt,

Caroline L. Relton, James R. Staley [*], Matthew Suderman [*]

https://wellcomeopenresearch.org/articles/7-41

Description of the data stored in EWAS catalog

Study inclusion

Studies are eligible for inclusion in the EWAS Catalog if:

• Include at least 100,000 CpG sites in the analysis.

• Have a sample size of at least 100 individuals.

Study exclusion

Studies were not included in the EWAS Catalog if:

• The DNA methylation data studied was not genome-wide.

• The study does not include any new EWAS data.

CpG inclusion

Association results with CpGs are eligible for inclusion in the EWAS Catalog if:

• The association has $p < 1 \times 10^{-4}$.

• The analysis was performed genome-wide.

Description of the data stored in EWAS catalog (continued)

Data extraction
The data for the EWAS Catalog are manually extracted from the published literature. The information extracted falls into four categories: study information, analysis information, participants information, and CpG results.

Study information
The following pieces of study and publication information are extracted:
• Author - the first author of the publication (surname then initials).
• Consortium - the name of the consortium.
• PMID - the PubMedID of the publication.
• Date - the date the paper was published (YY-MM-DD).
• Trait - the name of the trait.
• EFO - the corresponding ontology term(s) for the trait.
• Analysis - description of the analysis performed.
• Source - the table where the result can be found in the paper.

Analysis information

The following pieces of information on the analysis are extracted:

- Outcome - the outcome of the analysis.
- Exposure - the exposure of the analysis.
- Covariates - the covariates adjusted for in the analysis.
- Outcome_Units - the units of the outcome.
- Exposure_Units - the units of the exposure.
- Methylation_Array - the array used to measure the methylation.
- Tissue - the tissue in which the methylation was measured.
- Further_Details - any other relevant details of the analysis.

Description of the data stored in EWAS catalog (continued)

Participants information

The following pieces of information on the participants are extracted:

- N - the total number of participants used in the analysis.
- N_Cohorts - the total number of cohorts used in the analysis.
- Categories - the total number of individuals in each category (e.g. 200 smokers, 150 never smokers).
- Age - the age group participants belonged to.
- Sex - sex of individuals used in the analysis.
- Ancestry - ancestry of the individuals used in the analysis.

CpG results

The following information on the CpG associations are extracted:

- CpG - the CpG site.
- Beta - the effect estimate.
- SE - the standard error of beta.
- P - *p*-value.
- Details - any additional details on the analysis (e.g. sub-trait).

Searching EWAS catalog

Use cardiovascular disease-related keywords stored in this file
https://github.com/slavailn/CVD_genomics/blob/main/CVDs_and_keywords.md to search EWAS catalog and download the data as shown below:

Search for a term

Searching EWAS catalog

Download relevant results

Searching EWAS catalog

The results will be displayed as a table of CpG methylation levels associated with the trait of interest. Scroll to the bottom of the page and press "Download" to retrieve the results

Search results for *coronary heart disease*

Author	PMID	Outcome	Exposure	Tissue	Analysis	N	CpG	Location	Gene	Beta	P
Agha G	31424985	Coronary heart disease	DNA methylation	Whole blood	Meta analysis	11461	cg22617878	chr3:10417183	ATP2B2	-0.37	2E-08
Agha G	31424985	Coronary heart disease	DNA methylation	Whole blood	Meta analysis	11461	cg13827209	chr9:101912842	TGFBR1	0.27	3.8E-08
Agha G	31424985	Coronary heart disease	DNA methylation	Whole blood	Meta analysis	11461	cg14185717	chr9:16864746	BNC2	-0.29	1.4E-07
Agha G	31424985	Coronary heart disease	DNA methylation	Whole blood	Meta analysis	11461	cg10307345	chr11:18771567	PTPN5	-0.15	1.9E-07
Agha G	31424985	Coronary heart disease	DNA methylation	Whole blood	Meta analysis	11461	cg13822123	chr2:54197256	PSME4	0.41	2E-07
Agha G	31424985	Coronary heart disease	DNA methylation	Whole blood	Meta analysis	11461	cg23245316	chr2:3260005	TSSC1	-0.47	2.2E-07
Agha G	31424985	Coronary heart disease	DNA methylation	Whole blood	Meta analysis	11461	cg24977276	chr7:74105270	GTF2I	-0.33	2.5E-07
Agha G	31424985	Coronary heart disease	DNA methylation	Whole blood	Meta analysis	11461	cg24447788	chr19:795310	-	-0.27	4.3E-07

⬇ Download   ⟵ **Download data**