Software tools for the sharing & use of GWAS summary statistics and polygenic scores

Workshop Introduction

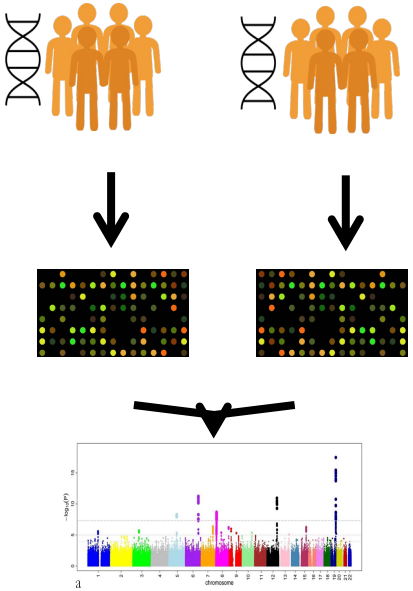
Laura Harris

GWAS Catalog Project Coordinator

ESHG 2024

EMBL-EBI

The Genome Wide Association Studies Catalog

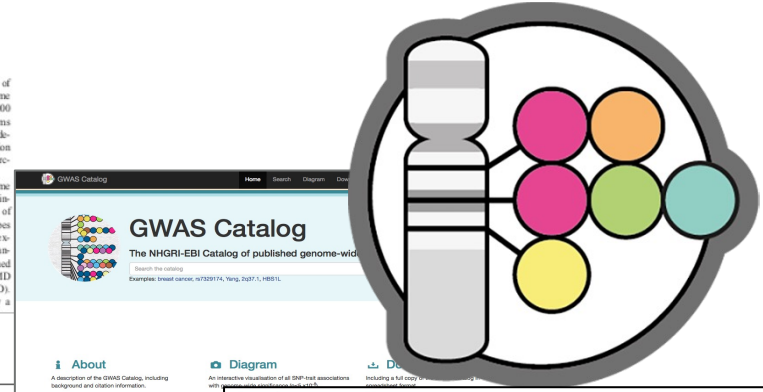


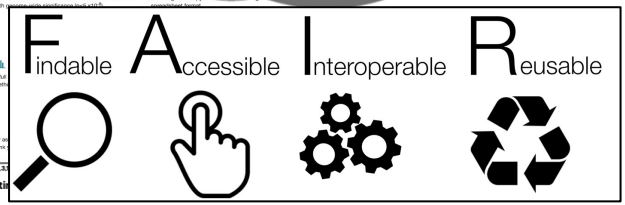
www.sciencemag.org SCIENCE VOL 308 15 APRIL 2005

Complement Factor H Polymorphism in Age-Related Macular Degeneration

Robert J. Klein,[1] Caroline Zeiss,[2]* Emily Y. Chew,[3]* Jen-Yue Tsai,[4]* Richard S. Sackler,[1] Chad Haynes,[1] Alice K. Henning,[5] John Paul SanGiovanni,[3] Shrikant M. Mane,[6] Susan T. Mayne,[7] Michael B. Bracken,[7] Frederick L. Ferris,[3] Jurg Ott,[1] Colin Barnstable,[2] Josephine Hoh[7]†

Age-related macular degeneration (AMD) is a major cause of blindness in the elderly. We report a genome-wide screen of 96 cases and 50 controls for polymorphisms associated with AMD. Among 116,204 single-nucleotide polymorphisms genotyped, an intronic and common variant in the comple-





nature
Article

GWAS and meta-analysis ider
variants underlying critical C

https://doi.org/10.1038/s41586-023-06034-3
Received: 22 November 2022
Accepted: 27 March 2023
Published online: 17 May 2023
Open access
Check for updates

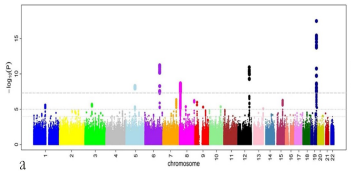
Erola Pairo-Castineira[1,2,3,564], Konrad Rawli
Isar Nassiri[8], Glenn A. McConkey[9], Marie 2
Wilna Oosthuyzen[1,3], Athanasios Kousathanas[10], Anne Richmond[2], Jonathan Millar[1,3,7]
Clark D. Russell[1], Tomas Malinauskas[9], Ryan Thwaites[12], Kirstie Morrice[13], Sean Keatin
David Maslove[14], Alistair Nichol[15], Malcolm G. Semple[16,17], Julian Knight[8],
Manu Shankar-Hari[11,18], Charlotte Summers[19], Charles Hinds[20], Peter Horby[21], Lowell Ling[22],
Danny McAuley[23,24], Hugh Montgomery[25], Peter J. M. Openshaw[12,26], Colin Begg[27],
Timothy Walsh[11], Albert Tenesa[2,3,28], Carlos Flores[29,30,31,32], José A. Riancho[33,34,35],
Augusto Rojas-Martinez[36], Pablo Lapunzina[37,38,39], GenOMICC Investigators*, SCOURGE
Consortium*, ISARICC Investigators*, The 23andMe COVID-19 Team*, Jian Yang[5,6],

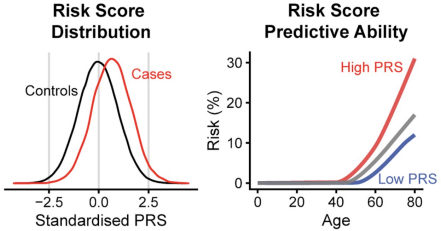
www.ebi.ac.uk/gwas

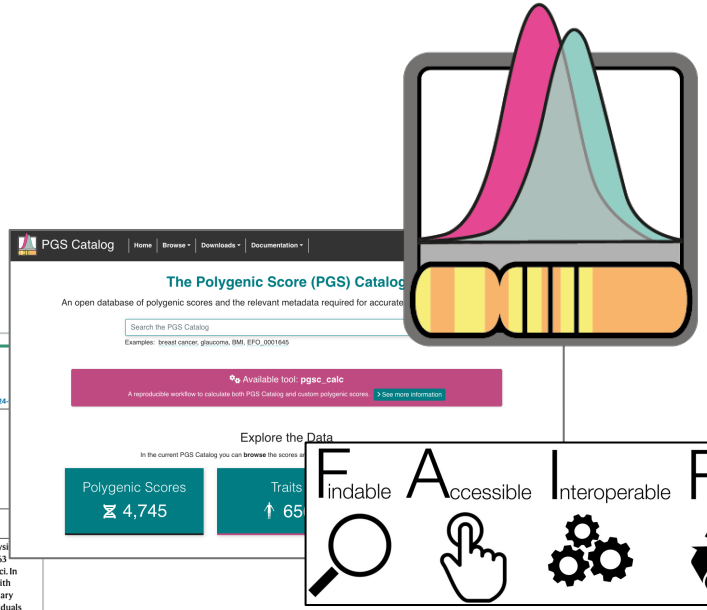
General training on data access via the GWAS Catalog website & API:
https://www.ebi.ac.uk/training/online/courses/gwas-catalogue-exploring-snp-trait-associations/

EMBL-EBI

The Polygenic Score Catalog







www.pgscatalog.org

Lambert et al 2024, *The Polygenic Score Catalog: new functionality and tools to enable FAIR research*
www.medrxiv.org/content/10.1101/2024.05.29.24307783v1

EMBL-EBI

Knowledgebase & Deposition resources

pre-publication

Journal publication

genomewide summary stats

metadata

lead SNPs

GWAS Catalog

The NHGRI-EBI Catalog of published genome-wide association studies

Search the catalog

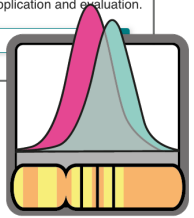
Examples: breast cancer, rs7329174, Yang, Zx37.1, H6R1L

The Polygenic Score (PGS) Catalog

An open database of polygenic scores and the relevant metadata required for accurate application and evaluation.

Search the PGS Catalog

Examples: breast cancer, glaucoma, BMI, EFO_0001645



scoring file

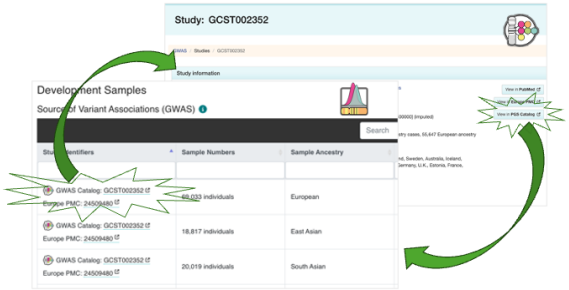
metadata

development & evaluation samples

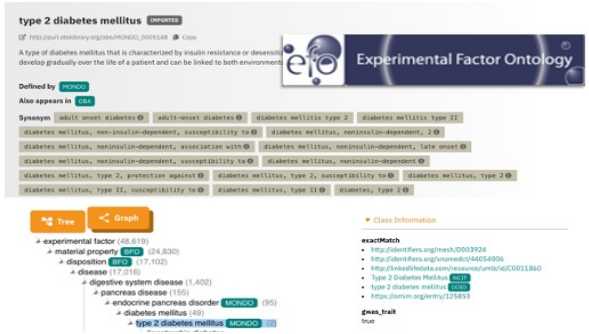
EMBL-EBI

Sibling resources

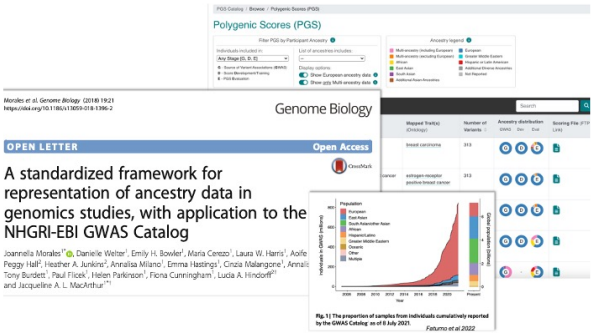
PGS linked to source GWAS & vice versa



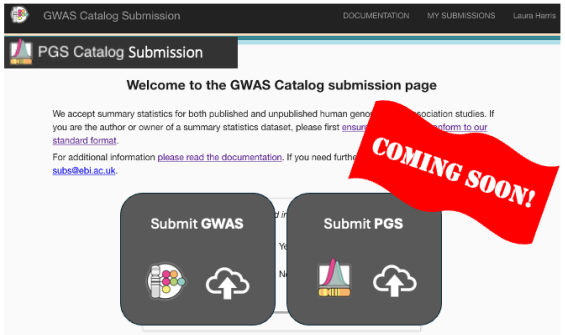
Shared trait ontology



Standard framework for sample description

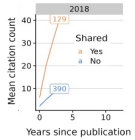


Shared data submission portal

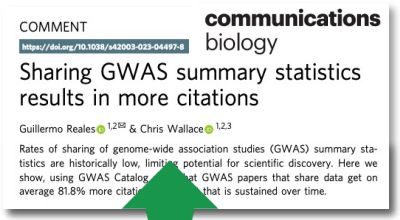


EMBL-EBI

Why share your data in the Catalogs?



More citations!


COMMENT
communications biology
https://doi.org/10.1038/s42003-023-04497-8
Sharing GWAS summary statistics results in more citations
Guillermo Reales[1,2✉] & Chris Wallace[1,2,3]
Rates of sharing of genome-wide association studies (GWAS) summary statistics are historically low, limiting potential for scientific discovery. Here we show, using GWAS Catalog [illegible] GWAS papers that share data get on average 81.8% more citati[illegible] that is sustained over time.

Annotated

Linked

Added value

Piped

Standardised

Reusability

Harmonised

37 → 38

Integrated with tools

pgsc_calc

EMBL-EBI

Workshop outline

- 1. Introduction
- 2. Formatting, validating and submitting GWAS summary statistics
- 3. Using GWAS Catalog summary statistics
- 4. Submitting data to the PGS Catalog
- 5. Using the PGS Catalog Calculator

Prep for workshop

- https://github.com/EBISPOT/eshg-2024-workshop/

If you are interested in **formatting & validating GWAS summary stats for submission**

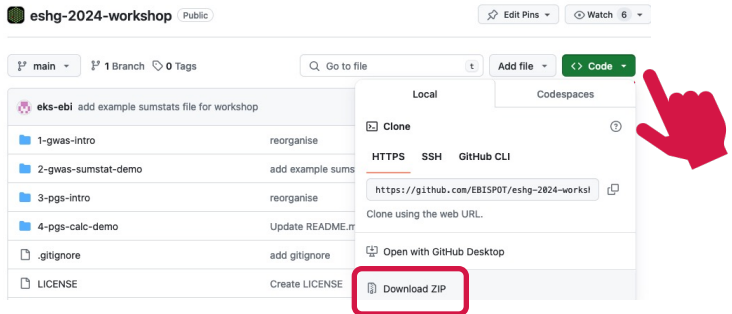
1. If you have a Chromium-based browser installed on your computer (e.g. Google Chrome, Microsoft Edge)…..
2. Go to folder 2
3. Download the example summary statistics (eshg2024_example_sumstats.tsv) file and save it to a new folder on your computer.

If you are interested in **calculating polygenic scores using pgsc_calc**

1. If you have a laptop with Linux or MacOS and >16GB RAM…
2. Go to folder 4
3. Follow the installation tutorial – requires installing Docker or Nextflow (dependent on wifi speed!)

To **access the entire workshop materials**

- Download entire repo from github



EMBL-EBI

Meet the team!

Laura Harris
GWAS Catalog Project Lead
EMBL-EBI

Sam Lambert
PGS Catalog Project Lead
University of Cambridge/EMBL-EBI

Ben Wingfield
Senior Bioinformatician
EMBL-EBI

Elliot Sollis
Senior Curator
EMBL-EBI

National Human Genome Research Institute

National Institute of Diabetes and Digestive and Kidney Diseases

10 YEAR ANNIVERSARY Open Targets 2014-2024

INTERVENE

GLOBAL CORE BIODATA RESOURCE

elixir Core Data Resource

UNIVERSITY OF CAMBRIDGE

EMBL-EBI

Questions?

www.ebi.ac.uk/gwas
gwas-info@ebi.ac.uk
www.pgscatalog.org
pgs-info@ebi.ac.uk

EMBL-EBI