

Step for obtaining gene expression files

- 1) Access the specific dataset page by entering the Accession ID into the GEO database search bar.

The screenshot shows the NCBI GEO homepage. The URL in the address bar is ncbi.nlm.nih.gov/geo/. A search bar at the top contains the accession number "gse159676". The main header reads "Gene Expression Omnibus". Below the header, there is a brief description of what GEO is: "GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles." To the right of the search bar is the GEO logo. Below the search bar, there are three navigation tabs: "Getting Started", "Tools", and "Browse Content". Under "Getting Started", there are links for "Overview" and "Search for Studies at GEO DataSets". Under "Browse Content", there is a link for "Repository Browser".

Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.



Getting Started

Overview

Tools

Search for Studies at GEO DataSets

Browse Content

Repository Browser

- 2) For microarray data, click on the **Series Matrix File** entry to download the integrated file that combines gene expression data with sample metadata.

The screenshot shows the specific dataset page for GSE159676. At the top, there is a blue button labeled "Analyze with GEO2R". Below it, there is a table with various details about the study, such as submission date (Oct 20, 2020), last update date (Mar 16, 2022), contact name (Trine Folseraas), and organization (Oslo University Hospital Rikshospitalet). There is also a section for platforms (GPL6244) and samples (33). In the "Relations" section, there is a link to a BioProject (PRJNA670177). Below this, there is a "Download family" section with options for SOFT, MINiML, and Series Matrix File(s). The "Series Matrix File(s)" option is highlighted with a yellow background. At the bottom, there is a table showing a single supplementary file: GSE159676_RAW.tar, which is 121.4 Mb and has a custom download link (http://) and is in TAR (of CEL) format.

- 3) For RNA-seq data, click the **Series Matrix File** entry to retrieve sample metadata and download the gene expression file provided in the **Supplementary Files** section.

City	Durham
ZIP/Postal code	27710
Country	USA

Platforms (1) [GPL16791 Illumina HiSeq 2500 \(Homo sapiens\)](#)

Samples (368) [GSM6589940 human liver sample A1025223](#)
[More...](#) [GSM6589941 human liver sample 200707904](#)
[GSM6589942 human liver sample 200707919](#)

This SubSeries is part of SuperSeries:
[GSE213623 Hepatocyte Smoothened Activity Controls Susceptibility to Insulin Resistance and Nonalcoholic Fatty Liver Disease](#)

Relations

BioProject	PRJNA881948
SRA	SRP398015

Download family	Format
SOFT formatted family file(s)	SOFT ?
MINIML formatted family file(s)	MINIML ?
Series Matrix File(s)	TXT ?

Supplementary file	Size	Download	File type/resource
GSE213621_FPKMs_allsamples.txt.gz	31.7 Mb	(ftp)(http)	TXT

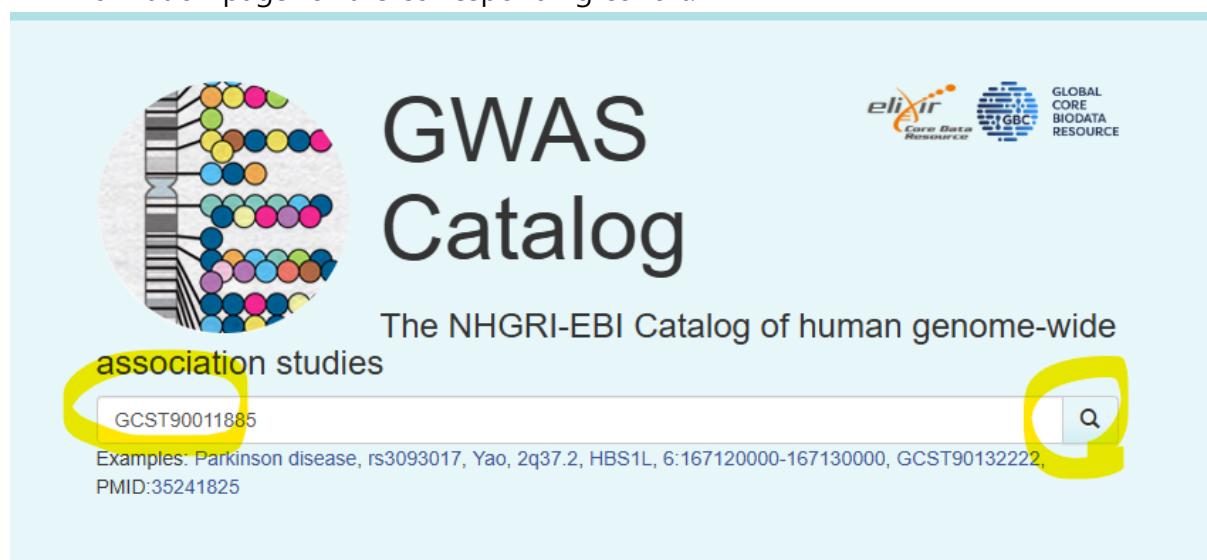
[SRA Run Selector \[?\]\(#\)](#)

Raw data are available in SRA

Processed data are available on Series record

Step for obtaining GWAS summary statistics

- 1) Search for the Accession ID on the GWAS Catalog homepage to access the detailed information page for the corresponding cohort.



The GWAS Catalog is a resource for human genome-wide association studies. It is a collaboration between the NHGRI and EBI. The homepage features a circular graphic of chromosomes and the text "The NHGRI-EBI Catalog of human genome-wide association studies". A search bar at the top right contains the accession ID "GCST90011885". Below the search bar, there is a note: "Examples: Parkinson disease, rs3093017, Yao, 2q37.2, HBS1L, 6:167120000-167130000, GCST90132222, PMID:35241825". Logos for elixir and GBC are also present.

- 2) Click the FTP Download link located in the bottom-right section of the metadata to access the directory where the cohort dataset is stored.

Publication information

PubMed ID	32298765	Journal	J Hepatol	View in PubMed
First author	Anstee QM	Publication date	2020-04-13	View in Europe PMC
Title	Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically-characterised cohort.			
Authors	Anstee QM, Darlay R, Cockell S , Meroni M, Govaere O, Tiniakos D, Burt AD, Bedossa P, Palmer J, Liu YL, Althal GP, Allison M, Yki-Järvinen H, Vacca M, Dufour JF, Invernizzi P, Prati D, Ekstedt M, Kechagias S, Francque S, Petta S, Bugianesi E, Clement K, Ratziu V, Schattenberg JM, Valenti L... Show more >			
Full Summary Statistics				

Available data: [Associations](#) 18 [Studies](#) 7 [Full summary statistics](#) 1 [Download Associations](#)

Studies with summary statistics 1

First author	Study accession	Pub. date	Journal	Title	Reported trait	Trait(s)	Background trait(s)	Discovery sample number	Replication sample number	Association count	Summary statistics
Anstee QM	GCST90011885	2020-04-13	J Hepatol	Genome-wide association study of non-alcoholic... fatty liver disease	Nonalcoholic fatty liver disease	non-alcoholic fatty liver disease	-	• 19264 European	• 1504 European	2	FTP Download

- 3) Select and download the file with comparably large size, as datasets containing the complete SNP set typically exceed 300 MB.

Index of /pub/databases/gwas/summary_stat

	Name	Last modified	Size	Description
	Parent Directory			
	GCST90011885_buildGRCh37.tsv.gz	2025-02-10 01:35	337M	
	GCST90011885_buildGRCh37.tsv.gz-meta.yaml	2025-02-10 01:35	1.0K	
	README.txt	2025-02-10 01:35	288	
	harmonised/	2025-06-26 15:44	-	
	md5sum.txt	2025-02-10 01:35	140	