GWAS resources

Dr. Tinashe Chikowore

University of Witwatersrand

twitter:tchikowore; tinashedoc@gmail.com

Learning outcomes

- Get acquainted with databases that store GWAS summary statistics
  - Disease traits
  - Quantitative traits
  - Ancestry specific
  - Metabolomics

GWAS Catalogue

https://www.ebi.ac.uk/gwas/

Meta-analysis of sub-Saharan African studies provides insights into genetic architecture of lipid traits

Ananyo Choudhury [#] [1], Jean-Tristan Brandenburg [#] [2], Tinashe Chikowore [2] [3], Dhriti Sengupta [2], Palwende Romuald Boua [2] [4], Nigel J Crowther [5], Godfred Agongo [6] [7], Gershim Asiki [8], F Xavier Gómez-Olivé [9], Isaac Kisiangani [8], Eric Maimela [10], Matshane Masemola-Maphutha [11], Lisa K Micklesfield [3], Engelbert A Nonterah [6], Shane A Norris [3], Hermann Sorgho [4], Halidou Tinto [4], Stephen Tollman [9], Sarah E Graham [12], Cristen J Willer [12] [13] [14]; AWI-Gen study; H3Africa Consortium; Scott Hazelhurst [2] [15], Michèle Ramsay [16] [17]

Data availability   Go to: ▸

The full dataset generated in this study is in the EGA [https://ega-archive.org/] database under the study accession code EGAS00001002482. This includes the phenotype dataset EGAD00001006425 and the genotype dataset EGAD00010001996. These datasets are available subject to controlled access through the Data and Biospecimen Access Committee of the H3Africa Consortium. The processed data generated in this study are provided in Supplementary Information and Supplementary Data. Summary statistics reported in the paper are accessible on GWAS Catalog (https://www.ebi.ac.uk/gwas/) at the accession numbers: GCST90101741, GCST90101742, GCST90101743, GCST9010174, GCST90101745, GCST90101746, GCST90101747, GCST90101748. All data that support the findings of this study are available from the corresponding authors on request. Publicly available datasets included in the study are the following: 1000 Genomes Project Phase 3 (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp), UGR meta-analysis summary statistics, GLGC summary Statistics, PAGE consortium summary statistics available at GWAS Catalog (https://www.ebi.ac.uk/gwas/) and gnomAD (https://gnomad.broadinstitute.org/).

Query output

Search results for *GCST90101741*

S **GCST90101741**

Meta-analysis of sub-Saharan African studies provides insights into genetic architecture of lipid traits.

Associations 7

P **Meta-analysis of sub-Saharan African studies provides insights into genetic architecture of lipid traits.**

Choudhury A et al. 2022 Nat Commun PMID:35546142

Associations 52   Studies 8

Study description

Download link

Index of /pub/databases/gwas/summary_statistics/GCST90101001-GCST90102000/GCST90101741

Name	Last modified	Size	Description
Parent Directory		-	
GCST90101741_buildGRCh37.tsv.gz	2022-05-14 00:04	323M	
README.txt	2022-05-14 00:04	1.4K	
harmonised/	2022-07-11 23:14	-	
md5sum.txt	2022-05-14 00:04	109	

Downloading the data

wget http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90101001-GCST90102000/GCST90101741/GCST90101741_buildGRCh37.tsv.gz

```
[tinashe@cream-ce CHAIR]$ zcat GCST90101741_buildGRCh37.tsv.gz | head
variant_id      p_value chromosome      base_pair_location      beta    standard_error effect_allele   other_allele    effect_allele_frequency
rs144434834     0.844322        1       723918  -0.00561981     0.0221663       G       0.890893
rs111533735     0.928312        1       724103  0.00240934      0.0175528       A       0.809232
rs1222283314    0.989279        1       724169  0.000988917     0.0390022       A       0.9630523
rs12069907      0.315583        1       724300  -0.0170174      0.0154994       C       0.711316
rs28692873      0.219178        1       724324  -0.0607959      0.0672038       G       0.9838896
rs578177050     0.283721        1       725211  0.0638079       0.0546656       C       0.981092
[rs111203397    0.0819076       1       725286  0.108617        0.0702693       G       0.9820085
rs1408251607    0.128085        1       725499  -0.05659        0.0455097       A       0.9571677
[rs28454925     0.375904        1       726794  -0.0450522      0.0635831       C       0.9838896
[tinashe@cream-ce CHAIR]$
```

Metabolomics and GWAS

Human Metabolic Individuality

A blog about metabolomics, genomics, and where the two fields meet.

A Table of all published GWAS with metabolomics

👤 by karsten | 🗄 posted in: Resources |

Here you find references to all published mGWAS in humans, ordered by publication date. If a study is missing from this list, please let me know. Please note that I do not apply a strict inclusion criterion for what I would call a GWAS with metabolomics, or rather a GWAS with biochemical traits (such as a GWAS with cholesterol levels). The latter are well covered by the GWAS catalogue.

This table was initially published in Kastenmüller et al., Genetics of human metabolism: an update. Hum. Mol. Genet. 2015 and has been **updated as of 20 January 2023**. If you are interested in associations at specific loci you may also use the SNiPA block annotation tool [read this post], and if you are looking for associations specific to certain metabolites, try the Metabolomics GWAS Server [read this post].

http://www.metabolomix.com/list-of-all-published-gwas-with-metabolomics/

UK Biobank

Pan-ancestry genetic analysis of the UK Biobank

The UK Biobank is a collection of a half million individuals with paired genetic and phenotype information that has been enormously valuable in studies of genetic etiology for common diseases and traits. However, most genome-wide analyses of this dataset use only the European ancestry individuals. Analyzing a more inclusive and diverse dataset increases power and improves the potential for discovery. Here, we present a multi-ancestry analysis of 7,228 phenotypes, across 6 continental ancestry groups, for a total of 16,131 genome-wide association studies. We release these summary statistics freely to the community ahead of publication.

https://pan.ukbb.broadinstitute.org/

GLGC

Global Lipids Genetics Consortium Results

Result Files

Ancestry-specific GWAS summary statistics for HDL-C, LDL-C, nonHDL-C, TC and TG (90 files plus README)

Trans-ancestry GWAS summary statistics for HDL-C, LDL-C, nonHDL-C, TC and C (20 files plus README)

LDL-C polygenic score weights for ALL, Admixed African, East Asian, European, Hispanic and South Asian ancestries (12 files plus README)

Trans-ancestry credible sets for HDL-C, LDL-C, nonHDL-C, TC and TG (5 files plus README)

Noncoding variant prioritization at lipid loci

ChrX GWAS summary statistics for HDL-C, LDL-C, nonHDL-C, TC and TG (180 files)

Sex-specific GWAS summary statistics for HDL-C, LDL-C, nonHDL-C, TC and TG (20 files)

Reference

Graham et al. (2021) The power of genetic diversity in genome-wide association studies of lipids. In press at Nature.

Kanoni et al. medRxiv (2021) Implicating genes, pleiotropy and sexual dimorphism at blood lipid loci through multiancestry meta-analysis.

Ramdas et al. bioRxiv (2021) A multi-layer functional genomic analysis to understand noncoding genetic variation in lipids.

Description

The Global Lipids Genetics Consortium aggregated GWAS results from 1,654,960 individuals from 201 primary studies representing five genetic ancestry groups: Admixed African or African (AdmAFR, N=99.4k, 6.0% of sample), East Asian (EAS, N=146.5k, 8.9%), European (EUR, N=1.32m, 79.8%), Hispanic (HIS, N=48.1k, 2.9%), and South Asian (SAS, N=41.0k, 2.5%). We performed GWAS for five blood lipid traits: low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), total cholesterol (TC), and non-high-density lipoprotein cholesterol (nonHDL-C). We performed meta-analysis within each ancestry using RAREMETAL. For the meta-analysis of all ancestries we performed meta-analysis of single cohort files using MR-MEGA to account for heterogeneity in variant effect sizes on lipids between ancestry groups. For the all ancestry analysis, fixed-effects meta-analysis was carried out with METAL to estimate effect sizes and se. The meta-analysis summary statistics

https://csg.sph.umich.edu/willer/public/glgc-lipids2021/

GIANT

https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files

MAGIC

Home page

Steering committee

Cohorts

Publications

Data downloads

Ongoing projects

MAGIC (the Meta-Analyses of Glucose and Insulin-related traits Consortium)

MAGIC (the Meta-Analyses of Glucose and Insulin-related traits Consortium) represents a collaborative effort to combine data from multiple GWAS to identify additional loci that impact on glycemic and metabolic traits.

MAGIC investigators have initially studied fasting glucose, fasting insulin, 2h glucose and HBA1c, as well as performed meta-analysis of more sophisticated measures of insulin secretion and sensitivity. Through these efforts, dozens of loci influencing these traits have been idenfified, a subset of which also influence risk of type 2 diabetes.

We are releasing results from our meta-analyses and these can be accessed through the data downloads pages.

https://magicinvestigators.org/

Next steps

GWAS summary data ecosystem

Data ingestion

User GWAS results → QC reports / dbSNP rsid alignment / hg19 strand alignment ← Published GWAS summary data

Data storage

HPC environment
- LD reference files
- GWAS VCF file store

Fast remote access
- Elasticsearch database
- Neo4j meta data
- LD reference files
- dbSNP reference

RESTful API

Data retrieval

- pygwasvcf python pkg
- gwasvcf R pkg
- gwasglue R pkg
- ieugwasr R pkg
- ieugwaspy python pkg

Analytical tools

Genetic architecture	Mendelian randomization	Finemapping	Colocalisation
LDSC	TwoSampleMR	FINEMAP	coloc
PRSice	MendelianRandomization	PAINTOR	HEIDI
	RadialMR	CAVIAR	eCAVIAR
	TRYX	GCTA-COJO	S-Predixcan
	MRMix	ABF	
	MR-PRESSO	JAM	
	GSMR	SuSiE	
	mr.raps		