https://www.nature.com/articles/s41467-024-50404-y

Data availability

The publicly available genotype, DNA methylation, and gene expression data of GTEx participants used in this study are available in the dbGaP and GEO under accession code phs000424.v8.p2[10,15] and GSE213478[10,15], respectively. The publicly available DNA methylation data of TCGA participants used in this study are available in the NCI Genomic Data Commons Data Portal [https://portal.gdc.cancer.gov/][27]. The publicly available GTEx v8-based gene expression and splicing prediction models used in this study are available in PredictDB [https://p

used in this study ar

cancer GWAS used i

cancer[26], GWAS cata

(accession code phs001736.v2.p1) for renal cell cancer[8], GWAS catalog (accession code GCST004746[34, 48]), UK Biobank data from the Neale lab [https://www.nealelab.is/uk-biobank][35], the FinnGen website [https://r9.finngen.fi/][36], and the Biobank Japan website [https://pheweb.jp/][37] for lung cancer, OCAC website [https://ocac.ccge.medschl.cam.ac.uk/data-projects/] for ovarian cancer[6,38], GWAS catalog

ession code

diction models

10810820)[33]. The

tion or Source Data files. Source data are provided with this paper.

35. Bycroft, C. et al. The UK biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

Article ADS CAS PubMed PubMed Central Google Scholar

37. Sato, G. et al. Pan-cancer and cross-population genome-wide association studies dissect shared genetic backgrounds underlying carcinogenesis. *Nat. Commun.* **14**, 3671 (2023).

Article ADS CAS PubMed PubMed Central Google Scholar

The Biobank Japan website

https://pheweb.jp/

Search for a variant, gene, or phenotype

Phenotypes Top Hits Random Downloads About

BioBank Japan PheWeb (PheWeb.jp)

Search for a variant, gene, or phenotype

Example queries: Type 2 diabetes (T2D) PCSK9 rs671

This website releases genome-wide association study (GWAS) summary statistics of the BioBank Japan Project (BBJ). We provide GWAS results in the Japanese population (mainly from BBJ) using the PheWeb platform, with public access to the full summary statistics.

The BBJ is a prospective genome biobank managed by the Institute of Medical Science, the University of Tokyo. It has collaboratively collected DNA and serum samples from 12 medical institutions across Japan. The project has recruited approximately 260,000 participants, mainly of Japanese ancestry. All study participants were diagnosed with one or more of 51 target diseases. RIKEN Center for Integrative Medical Sciences, the University of Tokyo, and Osaka University contributed to the genotyping of the BBJ samples.

News

March 25, 2025: 3 new GWAS results from Yamamoto, K., et al. (2024), Sasa, N., Kojima, S. & Koide, R., et al. (2025), and Yata, T., Sato, G. & Ogawa, K., et al. (2025).
Phenotypes include Jomon proportion, eHHV-6B positivity, and neuromyelitis optica spectrum disorder (NMOSD).

July 27, 2024: 7 new GWAS results from Sonehara, K. & Yano, Y., et al. (2024), and Suzuki, K., Hatzikotoulas, K., Southam, L., Taylor, HJ., Yin, X., Lorenz, KM. & Mandla, R., et al. (2024).
Phenotypes include recurrent pregnancy loss, and the latest multi-ancestry meta-analysis of T2D.

July 26, 2023: 7 new GWAS results from Akiyama, Y. & Sonehara, K., et al. (2023), and Sato, G., et al. (2023).
Phenotypes include Hunner-type interstitial cystitis, and pan-cancer meta-analyses (e.g., breast cancer).

ICD10 C	Colorectal cancer	BBJ	167,691	8,305	159,386	Download
ICD10 C	Cervical cancer	BBJ	61,581	967	60,614	Download
ICD10 C	Endometrial cancer	BBJ	61,814	1,200	60,614	Download
ICD10 C	Esophageal cancer	BBJ	160,589	1,388	159,201	Download
ICD10 C	Gastric cancer	BBJ	167,122	7,921	159,201	Download
ICD10 C	Hepatic cancer	BBJ	161,323	2,122	159,201	Download
ICD10 C	Lung cancer	BBJ	178,726	4,444	174,282	Download
ICD10 C	Malignant lymphoma	BBJ	177,677	335	177,342	Download
ICD10 C	Ovarian cancer	BBJ	61,457	843	60,614	Download
ICD10 C	Pancreatic cancer	BBJ	159,700	499	159,201	Download
ICD10 C	Pharyngeal and laryngeal cancer	BBJ	178,726	300	178,426	Download
ICD10 C	Prostate cancer	BBJ	90,332	5,672	84,660	Download
ICD10 C	Skin cancer	BBJ	178,726	154	178,572	Download
ICD10 C	Thyroid cancer	BBJ	178,723	361	178,362	Download
ICD10 D	Aplastic anemia	BBJ	174,110	53	174,057	Download

Genetics of rheumatoid arthritis contributes to biology and drug discovery

Okada, Y., et al, *Nature* **506** 376–381 (2014)

When using these results, please cite the relevant studies which appear on each phenotype page.

Sakaue, S. & Kanai, M., et al. (2021) | Kanai, M., et al. (2021) | Sakaue, S., et al. (2021) | Okada, Y., et al. (2014)

GWAS summary statistics of BioBank Japan (BBJ)

Citation: Sakaue, S. & Kanai, M., et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53** 1415-1424 (2021).

Detailed description of the file format and columns can be found here.

Link to wget commands for all the files

Category	Phenotype	Cohort	No. samples	No. cases	No. controls	Download link
ICD10 A	Diphtheria infection	BBJ	170,788	541	170,247	Download
ICD10 A	Dysentery	BBJ	178,482	411	178,071	Download
ICD10 A	Nontuberculous Mycobacterial infection	BBJ	172,384	64	172,320	Download
ICD10 A	Pulmonary tuberculosis	BBJ	178,671	7,800	170,871	Download
ICD10 A	Typhoid fever	BBJ	178,328	257	178,071	Download
ICD10 B	Acute hepatitis by Hepatitis A virus	BBJ	166,864	261	166,603	Download
ICD10 B	Chronic hepatitis B	BBJ	171,822	2,234	169,588	Download
ICD10 B	Chronic hepatitis C	BBJ	176,698	7,110	169,588	Download

Columns

Binary traits (disease endpoints and medications)

We provide the output from SAIGE software.
Please also refer to the SAIGE website.

#	column name	Descriptions
1	v	marker name (CHR:POS:REF:ALT)
2	CHR	chromosome
3	POS	position (hg19)
4	SNPID	rsID (if present) or equivalent ID
5	Allele1	REF allele
6	Allele2	ALT allele (This allele is the effect allele.)
7	AC_Allele2	allele count of Allele2 (ALT)
8	AF_Allele2	allele frequency of Allele2 (ALT)
9	imputationInfo	RSQ value in imputation
10	N	sample size
11	BETA	effect size of Allele2
12	SE	standard error f BETA
13	Tstat	score statistic
14	p.value	P value with SPA (suddle point approximation) applied
15	p.value.NA	P value when SPA is not applied
16	Is.SPA.converge	whether SPA is converged or not
17	varT	estimated variance of score statistic with sample related incorporated
18	varTstar	variance of score statistic without sample related incorporated
19	AF.Cases	allele frequency of Allele2 in cases
20	AF.Controls	allele frequency of Allele2 in controls

名稱

hum0197.v3.BBJ.LuC.v1.zip

名稱

GWASsummary_LuC_Japanese_SakaueKanai2020.auto.txt.gz

GWASsummary_LuC_Japanese_SakaueKanai2020.chrX.txt.gz

名稱

GWASsummary_LuC_Japanese_SakaueKanai2020.auto.txt

v	CHR	POS	SNPID	Allele1	Allele2	AC_Allele2	AF_Allele2	imputationInfo	N	BETA	SE	Tstat	p.value	p.value.NA	Is.SPA.converge	varT	varTstar	AF.Cases
1:751343:T:A	1	751343	rs28544273	T	A	52797.9333332697	0.147706358709057	0.902226705653162	178726	0.0380859409477793	0.0321097919454924	36.9393889						
1:751756:T:C	1	751756	rs28527770	T	C	52632.1843136993	0.147242662829413	0.907137991857088	178726	0.0381053732436682	0.0320342400035706	37.1327719						
1:752566:G:A	1	752566	rs3094315	G	A	301245.603921563	0.842758199482904	0.998880155492734	178726	-0.0358871260393195	0.029660987403796	-40.791292						
1:752894:T:C	1	752894	rs3131971	T	C	303703.800000067	0.849635195774726	0.93587401089489	178726	-0.0375661558692187	0.0312299545534736	-38.517141						
1:753405:C:A	1	753405	rs3115860	C	A	304361.062745183	0.851473939844184	0.913916501556503	178726	-0.0379453034220357	0.0318101314022378	-37.499640						
1:753425:T:C	1	753425	rs3131970	T	C	304126.050980472	0.850816476003694	0.918229880659406	178726	-0.0379477189885624	0.0316764611289227	-37.819202						
1:756285:ATCCACCCTGTCTACACTACCTGCTTGTCCAGCAGG:A	1	756285	1:756285:ATCCACCCTGTCTACACTACCTGCTTGTCCAGCAGG:A	ATCCACCCTGTCTACACTACCTGCTTGTCCAGCAGG	A	50416.22352923												
1:759293:T:A	1	759293	rs10157329	T	A	43938.5254900521	0.12292147054724	0.771496296636453	178726	0.0325772644136663	0.0377233887933103	22.8924995						
1:759700:T:C	1	759700	rs3115852	T	C	302343.541176511	0.845829765049602	0.928756966180498	178726	-0.0372128891488308	0.031135131554094	-38.387690						
1:759837:T:A	1	759837	rs3115851	T	A	304130.537254949	0.850829026708339	0.895896293886641	178726	-0.0384109732799781	0.032144691598347	-37.173785						
1:760912:C:T	1	760912	rs1048488	C	T	301848.098039239	0.844443724022355	0.921597399175042	178726	-0.0370258424318253	0.0312000477309129	-38.035964						
1:761147:T:C	1	761147	rs3115850	T	C	301743.003921583	0.84414971498714	0.920199606330509	178726	-0.0369741121768211	0.0312110254219712	-37.956108						
1:761752:C:T	1	761752	rs1057213	C	T	303562.109803952	0.849238806340298	0.886881818526478	178726	-0.0382236500760109	0.0322423104083215	-36.768833						
1:762273:G:A	1	762273	rs3115849	G	A	303142.494117567	0.848064898553	0.842635749605786	178726	-0.0473150603895612	0.0332913484155606	-42.691029384						
1:762589:G:C	1	762589	rs3115848	G	C	305481.372548958	0.854608094370594	0.846074025770228	178726	-0.0478040431087389	0.0336556266144027	-42.203577						
1:762592:C:G	1	762592	rs3131950	C	G	305479.015686211	0.854601500862246	0.845840034295337	178726	-0.0477851312037589	0.0336608661386646	-42.173749						
1:762601:T:C	1	762601	1:762601:T:C	T	C	305509.90980386	0.854687929578963	0.846023581239321	178726	-0.0478011927877414	0.0336628732859178	-42.182893						
1:762632:T:A	1	762632	rs3131948	T	A	305587.274509749	0.854904363410328	0.846596609304415	178726	-0.0476855842876394	0.0336638505497148	-42.078430						
1:764191:T:G	1	764191	rs7515915	T	G	49648.8078428354	0.138896433207355	0.82274564684342	178726	0.0486385783569644	0.0347404059043576	40.3005613						

UK Biobank data (from the Neale lab)

https://www.nealelab.is/uk-biobank

Improving the health of future generations

[1ST AUGUST 2018] WE'RE THRILLED TO ANNOUNCE AN UPDATED GWAS ANALYSIS OF THE UK BIOBANK.

LEARN MORE

GWAS RESULTS

Multi-ancestry analysis:

GWAS analysis of 7,221 phenotypes across 6 continental ancestry groups in the UK Biobank. This effort was led by Alicia Martin, Hilary Finucane, Mark Daly and Ben Neale, lead analysts Konrad Karczewski and Elizabeth Atkinson, with contributions from team members at ATGU. The summary statistics have been made available on the **Pan UKBB website.**

GWAS round 2 :

Results shared 1st August 2018.

Imputed genotypes from HRC plus UK10K & 1000 Genomes reference panels as released by UK Biobank in March 2018.

GWAS round 2 **results can be found here**

GWAS round 2 **Github code repository**

FAQ

Got questions about our GWAS of the UK Biobank?

What samples, markers and phenotypes we used? The format of our GWAS results files? Use permissions of the results? How to cite our results?

Read our FAQs.

THE TEAM

AI 摘要

GRCh37 和hg19 是同一套人类基因组参考序列的两种不同命名方式。 GRCh37 是基因组参考联合会(Genome Reference Consortium) 发布的版本，而hg19 是 University of California at Santa Cruz (UCSC) 基于GRCh37 整理的版本。 它们在核心染色体上的序列基本相同，但其他contig 的命名方式可能有所不同。

72	**<phenotype_code>.gwas.imputed_v3.{both_sexes,female,male}.tsv.bgz**		
73	These are the GWAS results files (e.g., "K50.gwas.imputed_v3.both_sexes.tsv.bgz").		
74			
75	**Contents:**		
76	variant	string	Variant identifier in the form "chr:pos:ref:alt", where "ref" is aligned to the forward strand of GRCh37 and "alt" is the effect allele (use this to join with variant annotation file).
77	minor_allele	string	The minor allele (alt allele is not always minor).
78	minor_AF	float	Frequency of the minor allele in the n_complete_samples defined for this phenotype.
79	expected_case_minor_AC	float	(Optional) For case/control phenotypes, calculated as (2 * minor_AF * n_cases).
80	expected_min_category_minor_AC	float	(Optional) For categorical phenotypes with less than 5 categories, calculated as (2 * minor_AF * number of samples in smallest category).
81	low_confidence_variant	boolean	Flag indicating low confidence results based on the following heuristics: - Case/control phenotypes: expected_case_minor_AC < 25 or minor_AF < 0.001. - Categorical phenotypes with less than 5 categories: expected_min_category_minor_AC < 25 or minor_AF < 0.001. - Quantitative phenotypes: minor_AF < 0.001.
82	n_complete_samples	int	Number of samples defined for this phenotype.
83	AC	float	Allele count of alt allele calculated on dosages within n_complete_samples.
84	ytx	float	Dot product of phenotype vector y and genotype vector x (alt allele count in cases for case/control phenotypes).
85	beta	float	Estimated effect size of alt allele.
86	se	float	Estimated standard error of beta.
87	tstat	float	t-statistic of beta estimate (= beta/se).
88	pval	float	p-value of beta significance test.
89			
90			
91			
92			
93			
94			
95			
96			

README | Manifest 201807 | DEPRECATED - DROPBOX Manifest 2 | Description Lookup | md5s for files

```
> library(data.table) #fread
> #install.packages("R.utils")
> library(R.utils) #gunzip
>
> data.raw <- gunzip("22140.gwas.imputed_v3.both_sexes.tsv.bgz", "22140.gwas.imputed_v3.both_sexe
s.tsv")
> data <- fread("22140.gwas.imputed_v3.both_sexes.tsv")
> head(data)
         variant minor_allele    minor_AF expected_case_minor_AC low_confidence_variant
1:   1:15791:C:T            T 0.00000e+00              0.00000000                    TRUE
2:   1:69487:G:A            A 1.15143e-05              0.00255617                    TRUE
3:   1:69569:T:C            C 2.12235e-04              0.04711610                    TRUE
4:  1:139853:C:T            T 1.13861e-05              0.00252772                    TRUE
5: 1:692794:CA:C            C 1.11409e-01             24.73280000                    TRUE
6:  1:693731:A:G            G 1.16398e-01             25.84050000                   FALSE
   n_complete_samples          AC     ytx          beta          se      tstat     pval
1:              91787     0.00000  0.0000           NaN         NaN        NaN      NaN
2:              91787     2.11373  0.0000 -1.54509e-03 0.024525300 -0.0629998 0.949767
3:              91787    38.96080  0.0000 -1.26638e-03 0.005790820 -0.2186870 0.826895
4:              91787     2.09020  0.0000 -1.53306e-03 0.024525800 -0.0625079 0.950159
5:              91787 20451.80000 25.1922  2.51863e-05 0.000284583  0.0885024 0.929478
6:              91787 21367.70000 21.2941 -2.79860e-04 0.000269534 -1.0383100 0.299128
> |
```

Variant = chr:pos:ref:alt
Beta=effect size of alt allele

Harmonization of GWAS data

- Ensure the reference and alternate alleles of each variant are defined consistently across datasets
- Ambiguous: A/T or G/C