Resource Summary Report

Generated by NIF on May 20, 2025

1000 Genomes Project and AWS

RRID:SCR_008801
Type: Tool

Proper Citation

1000 Genomes Project and AWS (RRID:SCR_008801)

Resource Information

URL: http://aws.amazon.com/1000genomes/

Proper Citation: 1000 Genomes Project and AWS (RRID:SCR_008801)

Description: A dataset containing the full genomic sequence of 1,700 individuals, freely available for research use. The 1000 Genomes Project is an international research effort coordinated by a consortium of 75 companies and organizations to establish the most detailed catalogue of human genetic variation. The project has grown to 200 terabytes of genomic data including DNA sequenced from more than 1,700 individuals that researchers can now access on AWS for use in disease research free of charge. The dataset containing the full genomic sequence of 1,700 individuals is now available to all via Amazon S3. The data can be found at: http://s3.amazonaws.com/1000genomes The 1000 Genomes Project aims to include the genomes of more than 2,662 individuals from 26 populations around the world, and the NIH will continue to add the remaining genome samples to the data collection this year. Public Data Sets on AWS provide a centralized repository of public data hosted on Amazon Simple Storage Service (Amazon S3). The data can be seamlessly accessed from AWS services such Amazon Elastic Compute Cloud (Amazon EC2) and Amazon Elastic MapReduce (Amazon EMR), which provide organizations with the highly scalable compute resources needed to take advantage of these large data collections. AWS is storing the public data sets at no charge to the community. Researchers pay only for the additional AWS resources they need for further processing or analysis of the data. All 200 TB of the latest 1000 Genomes Project data is available in a publicly available Amazon S3 bucket. You can access the data via simple HTTP requests, or take advantage of the AWS SDKs in languages such as Ruby, Java, Python, .NET and PHP. Researchers can use the Amazon EC2 utility computing service to dive into this data without the usual capital investment required to work with data at this scale. AWS also provides a number of orchestration and automation services to help teams make their research available to others to remix and reuse. Making the data available via a bucket in Amazon S3 also means that customers can crunch the information using Hadoop via Amazon Elastic MapReduce, and take advantage

of the growing collection of tools for running bioinformatics job flows, such as CloudBurst and Crossbow.

Abbreviations: 1000 Genomes Project and AWS

Synonyms: 1000 Genomes Project and Amazon Web Services, 000 Genomes Project Amazon Web Services, 1000 Genomes Project AWS

Resource Type: data set, data or information resource

Keywords: genomic data, genome, cloud computing, cloud, human, gene, genetic variation, research, dna

Funding:

Resource Name: 1000 Genomes Project and AWS

Resource ID: SCR_008801

Alternate IDs: nlx_144340

Record Creation Time: 20220129T080249+0000

Record Last Update: 20250519T205106+0000

Ratings and Alerts

No rating or validation information has been found for 1000 Genomes Project and AWS.

No alerts have been found for 1000 Genomes Project and AWS.

Data and Source Information

Source: SciCrunch Registry

Usage and Citation Metrics

We found 6982 mentions in open access literature.

Listed below are recent publications. The full list is available at NIF.

Cao C, et al. (2025) GWAShug: a comprehensive platform for decoding the shared genetic basis between complex traits based on summary statistics. Nucleic acids research, 53(D1), D1006.

Rajesh AE, et al. (2025) Machine learning derived retinal pigment score from ophthalmic imaging shows ethnicity is not biology. Nature communications, 16(1), 60.

Xu W, et al. (2025) Deep learning and genome-wide association meta-analyses of bone marrow adiposity in the UK Biobank. Nature communications, 16(1), 99.

Wang G, et al. (2025) Comparative genomic analysis unveiling the mutational landscape associated with premalignant lesions and early-stage gastric cardia cancer. Medicine, 104(2), e40332.

Yu Q, et al. (2025) Causal genes identification of giant cell arteritis in CD4+?Memory t cells: an integration of multi-omics and expression quantitative trait locus analysis. Inflammation research : official journal of the European Histamine Research Society ... [et al.], 74(1), 3.

Bellou E, et al. (2025) Benchmarking Alzheimer's disease prediction: personalised risk assessment using polygenic risk scores across various methodologies and genome-wide studies. Alzheimer's research & therapy, 17(1), 6.

Li R, et al. (2025) Whole genome sequence-based association analysis of African American individuals with bipolar disorder and schizophrenia. medRxiv : the preprint server for health sciences.

Yalcouyé A, et al. (2025) Whole-exome sequencing reveals known and candidate genes for hearing impairment in Mali. HGG advances, 6(1), 100391.

Zhang X, et al. (2025) Prevalence of Transcription Factor 4 Gene Triplet Repeat Expansion Associated with Fuchs' Endothelial Corneal Dystrophy in the United States and Global Populations. Ophthalmology science, 5(1), 100611.

Xi Z, et al. (2025) Clonal hematopoiesis of indeterminate potential is a risk factor of gastric cancer: A Prospective Cohort in UK Biobank study. Translational oncology, 52, 102242.

Lake AM, et al. (2025) Sexual Trauma, Polygenic Scores, and Mental Health Diagnoses and Outcomes. JAMA psychiatry, 82(1), 75.

Sumanaweera D, et al. (2025) Gene-level alignment of single-cell trajectories. Nature methods, 22(1), 68.

Xu J, et al. (2025) Identification of genetic variants of the IL18R1 gene in association with COPD susceptibility. Annals of medicine, 57(1), 2446690.

Wills C, et al. (2025) Relationship between inherited genetic variation and survival from colorectal cancer stratified by tumour location. Scientific reports, 15(1), 2423.

Du W, et al. (2025) Causal associations between iron levels in subcortical brain regions and psychiatric disorders: a Mendelian randomization study. Translational psychiatry, 15(1), 19.

Zhao W, et al. (2025) GoFCards: an integrated database and analytic platform for gain of

function variants in humans. Nucleic acids research, 53(D1), D976.

Huang Y, et al. (2025) RMVar 2.0: an updated database of functional variants in RNA modifications. Nucleic acids research, 53(D1), D275.

Huang Z, et al. (2025) TTC7A Variants Results in Gastrointestinal Defects and Immunodeficiency Syndrome: Case Series and Literature Review. Clinical reviews in allergy & immunology, 68(1), 7.

Cheng Z, et al. (2025) Exploring the Causal Relationship Between Frailty and Chronic Obstructive Pulmonary Disease: Insights From Bidirectional Mendelian Randomization and Mediation Analysis. International journal of chronic obstructive pulmonary disease, 20, 193.

Choi YS, et al. (2025) Assessing the Efficacy of Bortezomib and Dexamethasone for Induction and Maintenance Therapy in Relapsed/Refractory Cutaneous T-Cell Lymphoma: A Phase II CISL1701/BIC Study. Cancer research and treatment, 57(1), 267.