Resource Summary Report

Generated by SPARC SAWG on Sep 17, 2026

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 or information resource, data set

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

Resource Name: 1000 Genomes Project and AWS

Resource ID: SCR_008801

Alternate IDs: nlx_144340

Record Creation Time: 20220129T080249+0000

Record Last Update: 20260912T200318+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 7076 mentions in open access literature.

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

Kore P, et al. (2026) Interplay between Genetic Ancestry, Self-reported Race and Ethnicity, and Clinical Factors in Pediatric Acute Lymphoblastic Leukemia: A REDIAL Consortium Report. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 35(5), 710.

Arthur TD, et al. (2025) Multiomic QTL mapping reveals phenotypic complexity of GWAS loci and prioritizes putative causal variants. Cell genomics, 5(3), 100775.

MacLehose RF, et al. (2025) CYP2D6 Phenotype and Breast Cancer Outcomes: A Bias Analysis and Meta-Analysis. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 34(2), 224.

Zhang JT, et al. (2025) Follow-up Analysis Enhances Understanding of Molecular Residual Disease in Localized Non-Small Cell Lung Cancer. Clinical cancer research : an official journal of the American Association for Cancer Research, 31(7), 1305.

Stanciu A, et al. (2025) Personalized ctDNA detection and genomic profiling in the NeoRHEA Study. NPJ breast cancer, 11(1), 137.

Siu JHY, et al. (2025) Early lymph node T follicular helper cell signalling hub drives influenza vaccine response in an ancestrally diverse cohort. EBioMedicine, 122, 106036.

Johnson SH, et al. (2025) Tumor ploidy determination in low-pass whole genome sequencing and allelic copy number visualization using the Constellation Plot. Genome biology, 26(1), 132.

Goshayeshi L, et al. (2025) Germline variants in patients from the Iranian hereditary colorectal cancer registry. Cancer cell international, 25(1), 140.

Wonkam A, et al. (2025) FLT1 and other candidate fetal haemoglobin modifying loci in sickle cell disease in African ancestries. Nature communications, 16(1), 2092.

Yang J, et al. (2025) Unveiling tumor-infiltrating immune cell-driven immune-mediated drug resistance in clear cell renal cell carcinoma: prognostic insights and therapeutic strategies. Discover oncology, 16(1), 259.

Li H, et al. (2025) Modeling gene interactions in polygenic prediction via geometric deep learning. Genome research, 35(1), 178.

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.

Yalcouy?? 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.

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

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

Ji L, et al. (2025) Identification of Schizophrenia-Risk Regulatory Variant rs1399178 in the Non-coding Region With Its Impact on NRF1 Binding. CNS neuroscience & therapeutics, 31(3), e70275.

Park J, et al. (2025) Rare genetic associations with human lifespan in UK Biobank are enriched for oncogenic genes. Nature communications, 16(1), 2064.

Wang ZZ, et al. (2025) Genetic insights into the shared molecular mechanisms of Crohn's disease and breast cancer: a Mendelian randomization and deep learning approach. Discover oncology, 16(1), 198.

Shi W, et al. (2025) Polyunsaturated fatty acids may not be helpful for people with osteoarthritis: a two-sample Mendelian randomization analysis. Scientific reports, 15(1), 6065.