

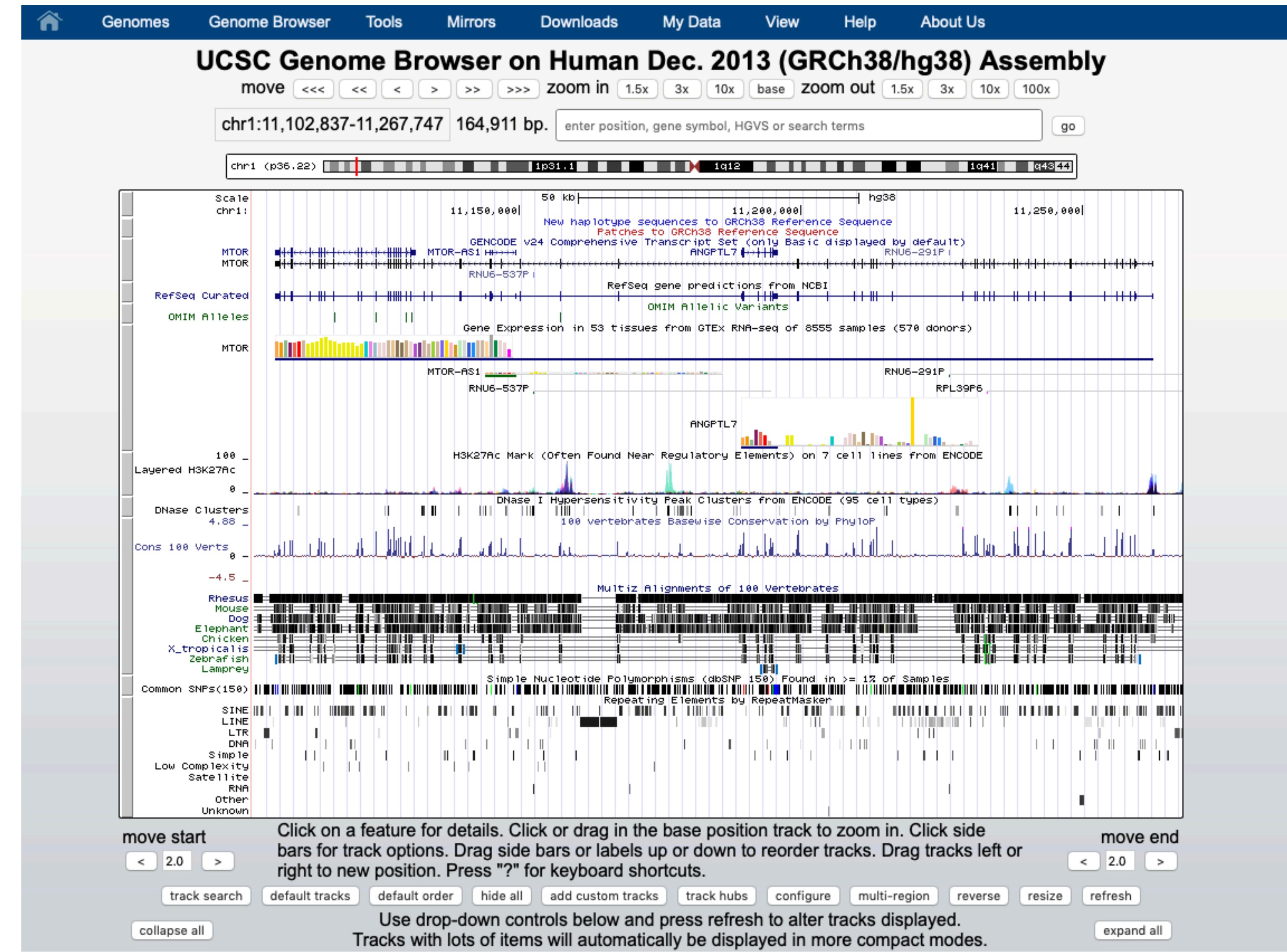
Genome Resources

Sequences | Variants | Interpretations

RESOURCES FOR GENOMICS: UCSC GENOME BROWSER

- ▶ Originated from the Human Genome Project
- ▶ Most widely used general genome browser
- ▶ many default tracks
- ▶ many species
- ▶ customization with "BED" files

genome.ucsc.edu



RESOURCES FOR GENOMICS: HUMAN GENOME RESOURCES AT NCBI

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Log in

Human Genome Resources at NCBI

Download Browse View Learn

Search for Human Genes

Select a chromosome to access the [Genome Data Viewer](#)

Download

	GRCh38	GRCh37
Reference Genome Sequence	Fasta	Fasta
RefSeq Reference Genome Annotation	gff3	gff3
RefSeq Transcripts	Fasta	Fasta
RefSeq Proteins	Fasta	Fasta
ClinVar	vcf	vcf
dbSNP	vcf	vcf
dbVar	vcf	vcf

www.ncbi.nlm.nih.gov/projects/genome/guide/human/

- ▶ Entry point for genome reference data
- ▶ Human genome assemblies
- ▶ Human variant collections (dbVar, ClinVar, dbSNP) for download

RESOURCES FOR GENOMICS: ENSEMBL

- ▶ Entry point for many genome data services and collections
- ▶ Downloads ("BioMart"), REST API

The screenshot shows the Ensembl Human (GRCh38.p14) homepage. At the top, there is a navigation bar with links to BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is also present. Below the navigation bar, the text "Human (GRCh38.p14) ▾" is displayed.

The main content area is divided into several sections:

- Search Human (Homo sapiens)**: Includes a search bar for "Search all categories" and "Search..." with a "Go" button. Below it, examples like PPP2R2A or 8:26291508-26372680 or rs699 or osteoarthritis are shown.
- Genome assembly: GRCh38.p14 (GCA_000001405.29)**: Includes links for More information and statistics, Download DNA sequence (FASTA), Convert your data to GRCh38 coordinates, and Display your data in Ensembl. It also shows a "View karyotype" icon and an "Example region" icon.
- Gene annotation**: Includes links for More about this genebuild, Download FASTA files for genes, cDNAs, ncRNA, proteins, Download GTF or GFF3 files for genes, cDNAs, ncRNA, proteins, and Update your old Ensembl IDs. It also shows icons for Pax6, INS, FOXP2, BRCA2, DMD, and ssh.
- Comparative genomics**: Includes links for More about comparative analysis, Download alignments (EMF). It shows an "Example gene tree" icon.
- Variation**: Includes links for More about variation in Ensembl, Download all variants (GVF), Variant Effect Predictor, and Ve!P. It shows an "Example variant" icon with the sequence ATCGAGCT, ATCCAGCT, ATCGAGAT.
- Regulation**: Includes links for More about the Ensembl regulatory build and microarray annotation, Experimental data sources, and Download all regulatory features (GFF). It shows an "Example regulatory feature" icon.
- Example transcript**: Shows an icon of a gene transcript with a start site and a stop site.
- Example structural variant**: Shows an icon of a DNA double helix with a structural variation indicated by a red arrow.

www.ensembl.org/Homo_sapiens/Info/Index

Where to find genome *variant* data ...

Reference Resources for Human Genome Variants

NCBI:dbSNP



- single nucleotide polymorphisms (SNPs) and multiple small-scale variations
- including insertions/deletions, microsatellites, non-polymorphic variants

NCBI:dbVAR



- genomic structural variation
- insertions, deletions, duplications, inversions, multinucleotide substitutions, mobile element insertions, translocations, complex chromosomal rearrangements

NCBI:ClinVar



- aggregates information about genomic variation and its relationship to human health

EMBL-EBI:EVA



- open-access database of all types of genetic variation data from all species

Ensembl



- portal for many things genomic...

RESOURCES FOR CANCER GENOMICS

COSMIC
Catalogue of somatic mutations in cancer

Home ▾ Resources ▾ Curation ▾ Tools ▾ Data ▾ News ▾ Help ▾ About ▾ Search COSMIC... Login ▾

COSMIC v79, released 14-NOV-16

COSMIC, the Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.

Start using COSMIC by searching for a gene, cancer type, mutation, etc. below, or by browsing a region of the human genome using the map to the right.

eg: *Braf, COLO-829, Carcinoma, V600E, BRCA-UK, Campbell* **SEARCH**

R Resources

Key COSMIC resources

- Cell Lines Project
- COSMIC
- Whole Genomes
- Cancer Gene Census
- Drug Sensitivity
- Mutational Signatures
- GRCh37 Cancer Archive

T Tools

Additional tools to explore COSMIC

- Cancer Browser
- Genome Browser
- GA4GH Beacon
- CONAN

C Expert Curation

High quality curation by expert postdoctoral scientists

- Drug Resistance
- Cancer Gene Census
- Curated Genes
- Gene Fusions
- Genome-Wide Screens

D Data

Further details on using COSMIC's content

- Downloads
- License
- Submission
- Genome Annotation
- Datasheets
- Help
- FAQ

Browse the [genomic landscape](#) of cancer

Cancer Gene Census Update

7 genes have been added to the [Cancer Gene Census](#) -

- EPAS1 - Endothelial PAS domain protein 1.
- PTPRT - Protein tyrosine phosphatase, receptor type T.
- PPM1D - Protein phosphatase, Mg²⁺/Mn²⁺ dependent 1D.
- BTK - Bruton tyrosine kinase.
- PREX2 - Phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 2.
- TP63 - Tumour protein p63.
- QKI - QKI, KH domain containing RNA binding.

For full details, see the [Datasheet](#).

RESOURCES FOR GENOMICS: CLINGEN

- ▶ "The Genomic Variant WG brings together representatives from the Sequence and Structural Variant communities for focused discussions on resolving discrepancies in variant interpretation and creating consistent curation guidelines."
- ▶ Interpreted genome variants with disease association

The screenshot shows the ClinGen Clinical Genome Resource website. At the top right is a search bar with the placeholder "Search our Knowledge Base for genes and diseases..." and a magnifying glass icon. Below the search bar are navigation links: About ClinGen, Working Groups, Resources, GenomeConnect, Share Your Data (highlighted in blue), and Curation Activities. The main banner features a blue background with a blurred image of laboratory glassware and a computer screen displaying genetic data. The text "Defining the clinical relevance of genes & variants for precision medicine and research..." is centered above three large numbers: 1496 (ClinGen Curated Genes), 31 (Expert Groups), and 10446 (Expert Reviewed Variants in ClinVar). To the right of these numbers is a magnifying glass icon labeled "Knowledge Base Search". Below the banner, the tagline "Sharing Data. Building Knowledge. Improving Care." is displayed, followed by a description of ClinGen's mission. Six call-to-action boxes are arranged in a grid below:

- ClinGen-ClinVar Partnership (Icon: DNA helix inside a circle)
- How to share genomic & health data (Icon: DNA helix inside a circular arrow)
- Learn about ClinGen curation activities (Icon: Computer monitor with DNA helix)
- GenomeConnect Patient Registry (Icon: Three DNA helices)
- View ClinGen's Resources & Tools (Icon: Computer monitor with multiple windows)
- Get Involved (Icon: Computer keyboard, mouse, and notepad)

clinicalgenome.org

The ClinGen and ClinVar Partnership

Both provide resources to support genomic interpretation

- ▶ ClinVar (an NCBI database/resource) is used as basis for curated variant <-> disease associations in ClinGen
- ▶ ClinGen - a funded project (application/funding limited)
- ▶ ClinVar - an internal NIH resource (dependent on political "goodwill")

ClinGen - A Program

An NIH funded project

Building a central resource that defines the clinical relevance of genes and variants

ClinGen is addressing the following critical questions:

- Is the gene associated with disease?
- Is the variant pathogenic?
- Is the variant/gene information actionable?

Encouraging data sharing

- Promote lab submissions to ClinVar
- Facilitate patient data sharing through GenomeConnect



Assessing the clinical **validity** and **actionability** of genes and their relationship to diseases

ClinVar- A Database

Funded by intramural NIH funding

Freely accessible and downloadable public archive of reports of the relationship between variants and conditions

Maintained by the National Center for Biotechnology Information (NCBI)



Expertly **curating** and **interpreting** variants

- Provide curated knowledge to ClinVar and on clinicalgenome.org

Expert Curation

Supporting **sharing** of variants interpretations

- Interpretations of the clinical significance of variants
- Submitter information
- Supporting evidence and individual level data, when available

clinicalgenome.org

ClinGen

Find out more online...

ClinVar

RESOURCES FOR CANCER GENOMICS

National Cancer Institute U.S. National Institutes of Health | www.cancer.gov

CANCER GENOME ANATOMY PROJECT

CGAP How To

Tools

CGAP Info

- Educational Resources
- Slide Tour
- Team Members
- References

CGAP Data

Quick Links:

- ICG
- NCI Home
- NCICB Home
- NCBI Home
- OCG

Genes **Chromosomes** **Tissues** **SAGE Genie** **RNAi** **Pathways**

Cancer Genome Anatomy Project (CGAP)

The NCI's Cancer Genome Anatomy Project sought to determine the gene expression profiles of normal, precancer, and cancer cells, leading eventually to improved detection, diagnosis, and treatment for the patient. Resources generated by the CGAP initiative are available to the broad cancer community. Interconnected modules provide access to all CGAP data, bioinformatic analysis tools, and biological resources allowing the user to find "in silico" answers to biological questions in a fraction of the time it once took in the laboratory.

The CGAP Website

Interconnected modules provide access to all CGAP data, bioinformatic analysis tools, and biological resources allowing the user to find "in silico" answers to biological questions in a fraction of the time it once took in the laboratory.

Genes Gene information, clone resources, SNP500Cancer, GAI, and transcriptome analysis.

Chromosomes FISH-mapped BAC clones, SNP500Cancer, and the Mitelman database of chromosome aberrations.

Tissues cDNA library information, methods, and EST-based gene expression analysis.

Pathways Diagrams of biological pathways and protein complexes, with links to genetic resources for each known protein.

RNAi RNA-interference constructs, targeted specifically against cancer relevant genes. New addition: Validated set of shRNAs.

International Cancer Genome Consortium

Home Cancer Genome Projects Committees and Working Groups Policies and Guidelines Media

ICGC Cancer Genome Projects

Committed projects to date: 89

Sort by: Project

Biliary Tract Cancer Japan	Biliary Tract Cancer Singapore	Bladder Cancer China
Bladder Cancer United States	Blood Cancer China	Blood Cancer Singapore
Blood Cancer South Korea	Blood Cancer United States	Blood Cancer United States
Blood Cancer United States	Blood Cancer United States	Bone Cancer France
Bone Cancer United Kingdom	Bone Cancer United States	Brain Cancer Canada
Brain Cancer China	Brain Cancer United States	Brain Cancer United States
Breast Cancer China	Breast Cancer European Union / United Kingdom	Breast Cancer France
Breast Cancer Mexico	Breast Cancer South Korea	Breast Cancer South Korea

ICGC Goal: To obtain a comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types and/or subtypes which are of clinical and societal importance across the globe.

[Read more »](#)

Launch Data Portal »

Apply for Access to Controlled Data »

Announcements

23/August/2016 - The ICGC Data Coordination Center (DCC) is pleased to announce ICGC data portal data release 22 (<http://dcc.icgc.org>).

ICGC data release 22 in total comprises data from more than 16,000 cancer donors spanning 70 projects and 21 tumour sites.

17/April/2016 - ICGCmed is pleased to announce the release of its white paper (<http://icgcmed.org>).

The International Cancer Genome Consortium for Medicine (ICGCmed) will link genomics data to clinical information, health and responses to therapies.

18/November/2015 - The International Cancer Genome Consortium (ICGC) PanCancer dataset generated by the PanCancer Analysis of Whole Genomes (PCAWG) study is now available on Amazon Web Services (AWS), giving cancer researchers access to over 2,400 consistently analyzed genomes corresponding to over 1,100 unique ICGC donors (<https://icgc.org/icgc-in-the-cloud>).

RESOURCES FOR GENOMICS - THEY MAY BREAK SOMETIMES ...

NCBI Resources How To Sign in to NCBI

We are sorry, but the page you requested is no longer available.

NCBI's SKY-CGH site has been retired.

The public data from this resource can be downloaded from our [FTP server](#) and will soon be available in the [dbVar database \(SKY-CGH\)](#).

You are here: NCBI > National Center for Biotechnology Information Write to the Help Desk

Skip Navigation

GETTING STARTED RESOURCES POPULAR FEATURED NCBI INFORMATION

NCBI Education	Chemicals & Bioassays	PubMed	Genetic Testing Registry	About NCBI
NCBI Help Manual	Data & Software	Bookshelf	PubMed Health	Research at NCBI
NCBI Handbook	DNA & RNA	PubMed Central	GenBank	NCBI News
Training & Tutorials	Domains & Structures	PubMed Health	Reference Sequences	NCBI FTP Site
Submit Data	Genes & Expression	BLAST	Gene Expression Omnibus	NCBI on Facebook
	Genetics & Medicine	Nucleotide	Map Viewer	NCBI on Twitter
	Genomes & Maps	Genome	Human Genome	NCBI on YouTube
	Homology	SNP	Mouse Genome	
	Literature	Gene	Influenza Virus	
	Proteins	Protein	Primer-BLAST	
	Sequence Analysis	PubChem	Sequence Read Archive	
	Taxonomy			
	Variation			

Cancer Genome Anatomy Project (CGAP)

The NCI's [Cancer Genome Anatomy Project](#) sought to determine the gene expression profiles of normal, precancer, and cancer cells for diagnosis, and treatment for the patient. Resources generated by the CGAP initiative are available to the broad cancer community. Data, bioinformatic analysis tools, and biological resources allowing the user to find "in silico" answers to biological questions in a timely manner.

[Read more about CGAP](#) and access the many valuable resources.

Cancer Genome Characterization Initiative (CGCI)

The [Cancer Genome Characterization \(CGC\) Initiative](#): Assessing the use of new genomics technologies to strategically characterize tumors. Groups involved with the CGCI Initiative make all of their data available through a publicly accessible database. Cancer CGCI incorporates genomic characterization methods including exome and transcriptome analysis using second generation sequencing to identify genetic changes leading to cancer.

[Read more about the CGC Initiative](#) and how the project is enabling the next generation of discovery through rapid data release and analysis.

Download Plugin: [Windows](#) [Mac OS X](#) [Linux](#)

National Center for Biotechnology Information, U.S. National Library of Medicine
8600 Rockville Pike, Bethesda MD, 20894 USA
[Policies and Guidelines](#) | [Contact](#)

NATIONAL LIBRARY OF MEDICINE NATIONAL INSTITUTES OF HEALTH USA.gov

A Service of the National Cancer Institute

as of 2018-09-19

VARIANT RESOURCES FOR CANCER GENOMICS

Resource name	Primary institute	Constituent Knowledge base	Cancer focused	Therapeutic evidence	Predisp. evidence	Diagnostic evidence	Prognostic evidence	Variant emphasis	URL
Cancer Genome Interpreter (CGI)	Institute for Research in Biomedicine, Barcelona, Spain	x	x	x				Somatic	https://www.cancergenomeinterpreter.org/home
Clinical Interpretation of Variants in Cancer (CIViC)	Washington University School of Medicine (WashU)	x	x	x	x	x	x	All variants	www.civicdb.org
JAX Clinical Knowledgebase (CKB)	The Jackson Laboratory	x	x	x	x	x	x	Somatic	https://ckb.jax.org/
Molecular Match	Molecular Match	x	x	x			x	Somatic	https://app.molecularmatch.com/
OncoKB	Memorial Sloan Kettering Cancer Center	x	x	x				Somatic	http://oncokb.org/#/
Precision Medicine Knowledgebase (PMKB)	Weill Cornell Medical College	x	x	x	x	x	x	Somatic	https://pmkb.weill.cornell.edu/
BRCA exchange	GA4GH	x	x		x			Germline	http://brcaexchange.org/
Cancer Driver Log (CanDL)	Ohio State University (OSU) / James Cancer Hospital		x	x				Somatic	https://cndl.osu.edu/
Gene Drug Knowledge Database	Synapse		x	x		x	x	Somatic	https://www.synapse.org/#!Synapse:syn2370773/wiki/62707
MatchMiner	Dana-Farber Cancer Institute		x					Somatic	http://bcb.dfci.harvard.edu/knowledge-systems/
COSMIC Drug Resistance Curation	Wellcome Trust Sanger Institute		x	x				Somatic	http://cancer.sanger.ac.uk/cosmic/drug_resistance
My Cancer Genome	Vanderbilt University		x	x		x	x	Somatic	https://www.mycancergenome.org/
NCI Clinical Trials	National Cancer Institute of the National Institutes of Health		x					Somatic	www.cancer.gov/about-cancer/treatment/clinical-trials
Personalized Cancer Therapy Database	MD Anderson Cancer Center		x	x	x	x	x	Somatic	https://pct.mdanderson.org/#/home
ClinGen Knowledge Base	ClinGen				x			Germline	https://www.clinicalgenome.org/resources-tools/
ClinVar	National Center for Biotechnology Information (NCBI)			x	x			All variants	http://www.ncbi.nlm.nih.gov/clinvar/
Pharmacogenomics Knowledgebase (PharmGKB)	Stanford University			x				Germline	https://www.pharmgkb.org/
The Human Gene Mutation Database (HGMD)	Institute of Medical Genetics in Cardiff				x			Germline	http://www.hgmd.cf.ac.uk

Beyond a Single Resource: Federation

Cell Genomics

CellPress
OPEN ACCESS

Commentary

International federation of genomic medicine databases using GA4GH standards

Adrian Thorogood,^{1,2,*} Heidi L. Rehm,^{3,4} Peter Goodhand,^{5,6} Angela J.H. Page,^{4,5} Yann Joly,² Michael Baudis,⁷ Jordi Rambla,^{8,9} Arcadi Navarro,^{8,10,11,12} Tommi H. Nyronen,^{13,14} Mikael Linden,^{13,14} Edward S. Dove,¹⁵ Marc Fiume,¹⁶ Michael Brudno,¹⁷ Melissa S. Cline,¹⁸ and Ewan Birney¹⁹

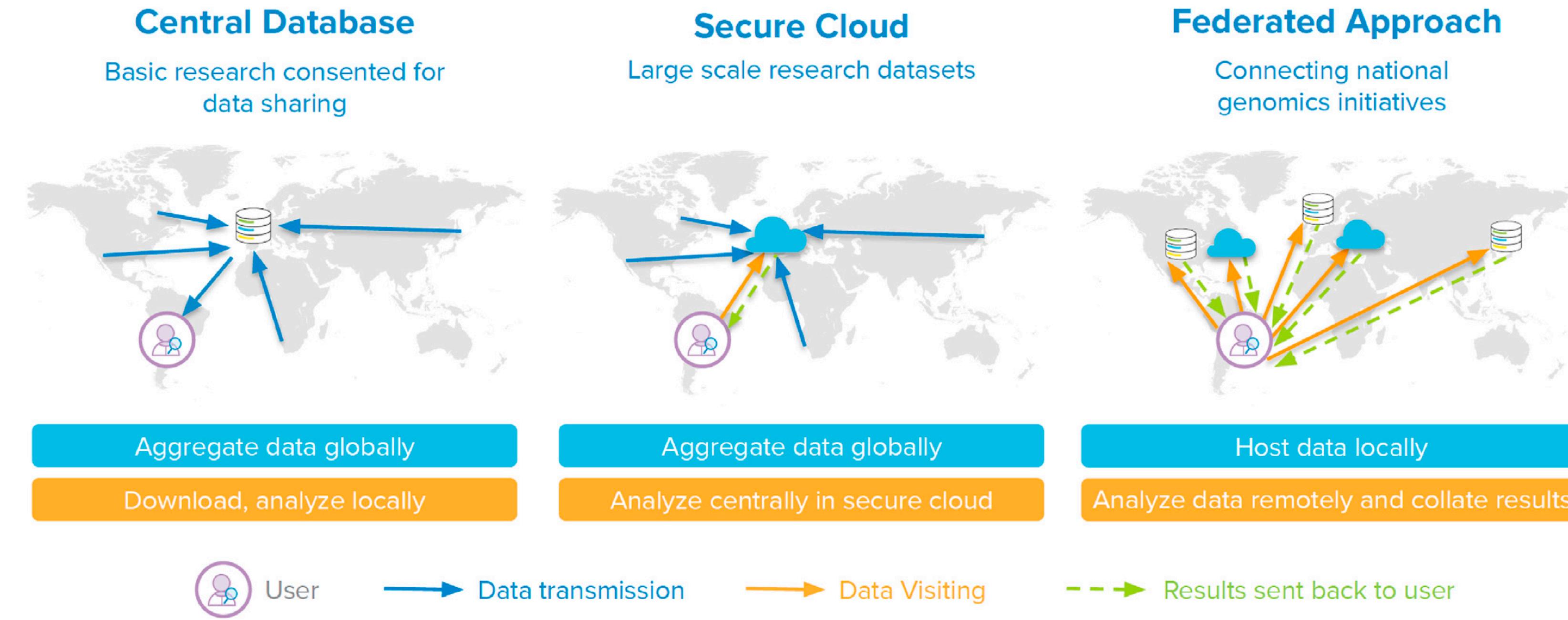


Figure 1. Data sharing approaches: Central database, secure cloud, and federated

Central database: Data from multiple sources are pooled in a central database. Researchers download copies of data and analyze them in their own computing environment.

Secure cloud: Data from multiple sources are pooled in a central cloud environment. Researchers remotely visit data and run their analyses in the cloud and download the result.

Federation: Data remain within locally controlled databases and computing environments, which may be cloud environments. Researchers remotely visit data, run their analyses at each site, and receive a local result, which can then be aggregated.

Task: Exploring Genome Resources

- primary deposition databases
- interpreted databases (e.g. variant annotations...)
- suggestion: VICC paper (Wagner et al.)
 - Wagner et al (2020): A harmonized meta-knowledgebase of clinical interpretations of somatic genomic variants in cancer
- make some notes about different genome resources and their primary use
 - ➡ Don't think only "human" _(___/__