

# DNAnexus Supports Data Management & Genomic Analysis for a Global Research Consortium

*“The DNAnexus global network provides hundreds of researchers at institutions worldwide secure and immediate access and use of ENCODE’s results.”*

**Richard Daly**

CEO of DNAnexus

## OVERVIEW

In June 2015, DNAnexus announced that Stanford University, the Data Coordination Center (DCC) for the National Institutes of Health (NIH)-funded ENCYclopedia of DNA Elements (ENCODE) Project, a flagship functional genomics consortium funded by the National Human Genome Research Institute (NHGRI) at the NIH, has adopted the company’s cloud genomics platform to support data analysis and sharing for its Phase 3 project. The DNAnexus platform supports the DCC bioinformatics analysis of ENCODE data, making the consortium’s bioinformatics methods available to the broader research community.

## A CHALLENGING OBJECTIVE

The goal of the ENCODE Project is to comprehensively catalog candidate functional and regulatory elements of the human genome, and provide a foundation for studying the genomic basis of human biology and disease. The ENCODE Consortium includes data production centers and computational biologists at a number of biomedical institutes across North America and Europe.

The ENCODE project is comprised of three phases. Phase 1 (2003-2007) was a pilot effort for a small portion of the genome. Phase 2 (2007-2012) scaled up to full genome wide analyses. Now in Phase 3, the ENCODE Consortium project is using next-generation technologies and methods to expand the size and scope of catalog content created in earlier phases.

## SUMMARY

### MISSION

The ENCODE Project is producing a comprehensive catalog of candidate functional and regulatory elements of the human genome, and it is providing a foundation for studying the genomic basis of human biology and disease.

### PARTICIPANTS

Stanford University  
ENCODE Consortium Members  
DNAnexus

### WEBSITES

[cherrylab.stanford.edu](http://cherrylab.stanford.edu)  
[www.encodeproject.org](http://www.encodeproject.org)  
[www.dnanexus.com](http://www.dnanexus.com)

### PROJECT

International Research Consortium

### CHALLENGE

The ENCODE Data Coordination Center (DCC) at Stanford University requires a secure and scalable environment to consistently process thousands of datasets and allow collaboration around petabyte-sized genomic analysis results.

### SOLUTION

DNAnexus, a secure and unified platform connecting thousands of scientists globally.

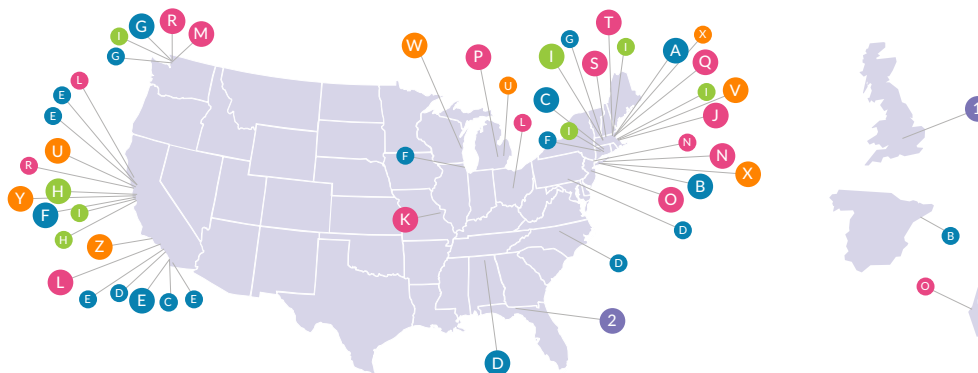
## MORE THAN SCALE

Although a scalable solution capable of processing thousands of datasets was a key requirement for the DCC, this was not the only capability needed. The development of version-controlled ENCODE pipelines is a priority in the current phase of the ENCODE project to ensure that data release to the public are consistently processed. Tasked with centralizing the project's raw and processed sequence-based data using uniform metadata standards and bioinformatics analysis, the DCC will also take advantage of the DNAnexus platform to supply the Consortium with:

- a secure and unified platform already connecting thousands of scientists around the world,
- transparency, reproducibility, and data provenance for consistency amongst ENCODE pipelines and results.

## SUPPORTING THE ENCODE PROJECT NETWORK

Stanford University, the ENCODE DCC serves as a data warehouse and processing hub for the ENCODE Project.



### Production Groups

- A** Broad Institute
- B** Cold Spring Harbor  
Centre for Genomic Regulation (CRG);
- C** University of Connecticut Health Center;  
UCSD
- D** HudsonAlpha; Pennsylvania State; UC Irvine;  
Duke; Caltech
- E** UCSD; Salk Institute; Joint Genome Institute;  
Lawrence Berkeley National Laboratory;  
UCSD
- F** Stanford; University of Chicago; Yale
- G** University of Washington;  
Fred Hutchinson Cancer Research Center;  
University of Massachusetts Medical School

### Data Coordination Center

- H** Stanford; UCSC

### Data Analysis Center

- I** University of Massachusetts Medical School;  
Yale; MIT; Stanford; Harvard; University of  
Washington

### Technology Development Groups

- J** MIT
- K** Washington University, St.Louis
- L** USC; Ohio State University; UC, Davis
- M** University of Washington
- N** Sloan-Kettering; Weill Cornell Medical College
- O** Princeton; Weizmann
- P** University of Michigan
- Q** Broad Institute
- R** University of Washington; UCSF
- S** Advanced RNA Technologies, LLC
- T** Harvard

### Computational Analysis Groups

- U** Berkeley; Wayne State University
- V** MIT
- W** University of Wisconsin
- X** Sloan-Kettering; Broad Institute
- Y** Stanford
- Z** UCLA

### Affiliated Groups

- 1** Wellcome Trust Sanger Institute
- 2** Florida State University

Adapted from: NHGRI

[www.encodeproject.org/about/contributors/](http://www.encodeproject.org/about/contributors/)

## DNAnexus

DNAnexus combines expertise in cloud computing and bioinformatics to create the global network for genomic medicine. DNAnexus provides security, scalability, and collaboration for enterprises and organizations that are pursuing genomic-based approaches to health in order to accelerate medical discovery. DNAnexus is supporting customers around the world that are tackling some of the most challenging and exciting opportunities in human health.