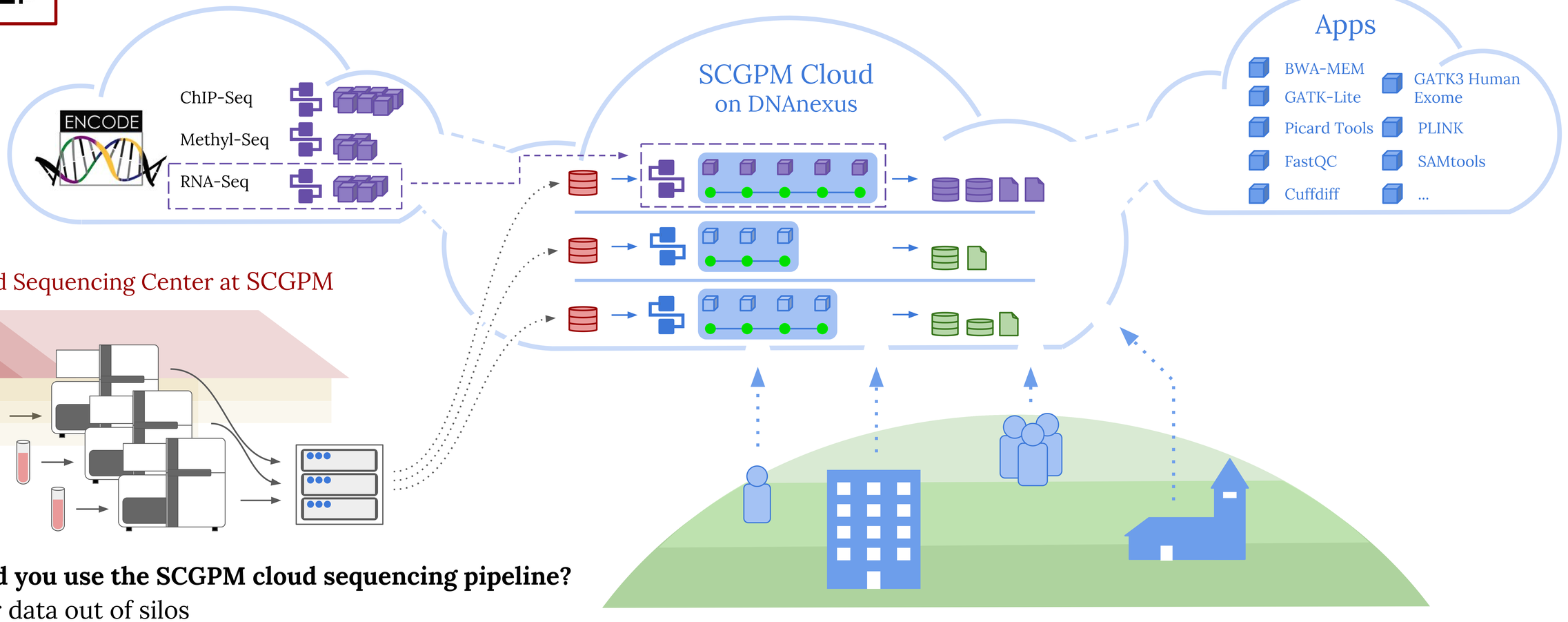
# A cloud-based sequencing pipeline: bringing people and analyses to data

Paul Billing-Ross, Nathaniel Watson, Keith Bettinger, Somalee Datta
Stanford Center for Genomics and Personalized Medicine

Stanford MEDICINE




## Why should you use the SCGPM cloud sequencing pipeline?

### Take your data out of silos

Generating and storing data on a cloud genomics platform provides the ability to share projects with collaborators around the world, with the click of a button. Seamlessly bring together data generators and analyzers.

### Make bioinformatics easy

Plug your sequencing data directly into sophisticated workflows from ENCODE and others to analyze RNA-Seq, Methyl-Seq, and ChIP-Seq data; all using a drag-and-drop web interface or command-line console.

### Scale to Big Data

Using a cloud platform, built on Amazon Web Services, provides limitless storage capacity and the ability to immediately apply vast and parallel computing resources to analyze data.

### Data security

DNAnexus is specifically designed to handle protected health information and provides clinical grade security. All data are encrypted and the platform supports HIPAA compliance.

## Who uses SCGPM sequencing?

SCGPM provides sequencing for several dozen labs at Stanford in addition to researchers at USC, Yale, the University of Chicago and a growing list of research facilities supported by CIRM Center for Excellence in Stem Cell Research (UCSF, UCLA, UCSD, UCSC, Scripps Research Institute, J. Craig Venter Institute, Ludwig Cancer Research, Salk Institute).

## What is SCGPM sequencing used for?

**Stem cell research**: The Center for Excellence in Stem Cell Research is using sequencing data to power studies in multiple fields of research. These projects already span cardiovascular disease, cell differentiation, the study of molecular networks, and many more.

**integrated Personal Omics Profiling (iPoP)**: iPoP combines multiple high-throughput methods to track the genomic, transcriptomic, proteomic, metabolic, and autoantibody profiles of an individual through health and disease. Analyzing this data allows for interpretation of how personal omics changes affect health, and is an important step in personalized medicine.

**Human Microbiome Project (HMP 2):** Researchers at Stanford are applying mixed omics approaches, including genomics, transcriptomics, and proteomics, to understand how changes in the gut microbiome can trigger disease.

**Encyclopedia of DNA elements (ENCODE 3):** The third iteration of ENCODE is coupling the most comprehensive set of antibody assays to-date, with the latest genomic technologies, to expand our encyclopedia of DNA elements.

## How much data can SCGPM sequencing generate?

With 7 HiSeq 2000s, 2 MiSeqs, 2 HiSeq 4000s, and a PacBio RS II; SCGPM is capable of sequencing over 3000 genomes or 22,000 exomes every year.