As DNA sequence data is generated at breakneck pace, scientists and clinicians are finding that cloud computing provides a flexible, low-cost alternative to internal computational resources.

**AN OVERVIEW**

**Data Producers Are Cranking …**

These broad categories of organizations generate the majority of DNA sequence data:

- **SEQUENCING FACILITIES**: ranging from facilities with industrial-scale genome sequencing pipelines to small cost centers, with one in two sequencing using a small academic department.

- **CLINICAL DIAGNOSTICS LABS**: These labs, either commercial or those based within academic medical centers, produce genomic data from patients to facilitate patient diagnosis and treatment.

- **RESEARCH LABS**: University, nonprofit, and biopharmaceutical research centers, and small companies are all producing increasing amounts of information as a result of genome-scale research activities.

… While Costs Drop and Capacity Rockets

Sequencing costs are falling faster than Moore’s Law, while production ability is higher than ever.

**COST PER GENOME IS FALLING**

<table>
<thead>
<tr>
<th>Year</th>
<th>Cost per Genome</th>
<th>Cost/Million Bases</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>$310</td>
<td>$310</td>
</tr>
<tr>
<td>2015</td>
<td>$131</td>
<td>$131</td>
</tr>
</tbody>
</table>

**CAPACITY IS INCREASING**

<table>
<thead>
<tr>
<th>Year</th>
<th>Seq per Genome</th>
<th>Seq Capacity (billion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>3</td>
<td>2.1</td>
</tr>
<tr>
<td>2013</td>
<td>4</td>
<td>4.3</td>
</tr>
</tbody>
</table>

**Cloud Providers Are the Obvious Choice**

Groups generating sequence data can store it locally in a cluster or private cloud, or can upload it to a cloud computing service. Groups that choose local storage must purchase hardware and storage to handle the highest peaks of usage. If these peaks are only偶尔, running costs vary tremendously for local and cloud computing.

**CASE STUDY:** HGSC at Baylor College of Medicine

What Is CHARGE?

The Cohort for Heart and Aging Research in Genomic Epidemiology (CHARGE) is a collaborative study of Asian human genetic effects, heart disease, and aging.

Cloud Computing Enables Big Science

Faced with a data analysis project beyond the scope of internal compute power, the Baylor Human Genome Sequencing Center ported its Mercury analysis pipeline—HGSC-MERCURY—cloud as a modular set of apps and algorithms developed in infancy for the NIH Gen2E project and designed to harness the cloud’s scale.

Cloud to the DNA

DMAnexus, the DNA cloud platform for human data research, is able to cloud solution that made data accessible to the 300 researchers in the CHARGE Consortium.

**DMAnexus**

- **300 researchers involved in CHARGE**
- **3,751 whole genomes analyzed**
- **10,771 exomes analyzed**
- **2.4 million computational core hours used**
- **20,800 core hours at peak computational demand**
- **400TB results generated for further analysis**

Brought to you by DNAexus

DMAnexus is a pioneer in cloud-based DNA data management and analysis solutions, offering cloud storage from Amazon Web Services and offering a suite of analysis and pipeline-building tools to its platform-as-a-service.

Funded by Google, Perkin Elmer, TPG Growth, and First Round Capital

The DMAnexus team consists of experts in bioinformatics, cloud computing, genomics and bioengineering. Clients include major academic medical centers, genome centers, and diagnostic providers.

Operates on the premise, "someone can do it your data and managed it,"