Globus Genomics: A Medical Center's Bioinformatics Core Perspective

Rama Raghavan, Daniel Voss, Brooke L. Fridley

Department of Biostatistics, University of Kansas Medical Center

Outline

- Background
  - BISR of University of Kansas
  - Problems of Bioinformatics Cores
- Approach
  - Problems with current infrastructure
  - New approach tested with Globus
- Pilot project experience
- Future plans

Outline

- Background
  - BISR of University of Kansas
  - Problems of Bioinformatics Cores
- Approach
  - Problems with current infrastructure
  - New approach tested with Globus
- Pilot project experience
- Future plans

BISR of University of Kansas

- The **Biostatistics and Informatics Shared Resource (BISR)** is a critical and highly utilized resource which supports the research of KUCC members.
  - **Bioinformatics / Statistical Genomics**
  - **Study Design**
  - **Statistical Oversight and Analyses**
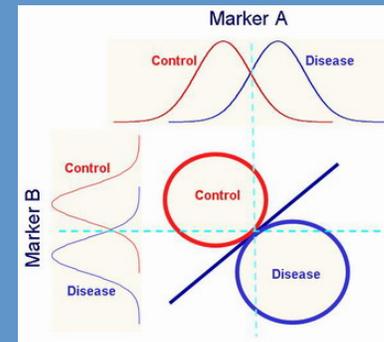 - **Clinical Research Informatics and Data Management**

Bioinformatics-Statistics "continuum"
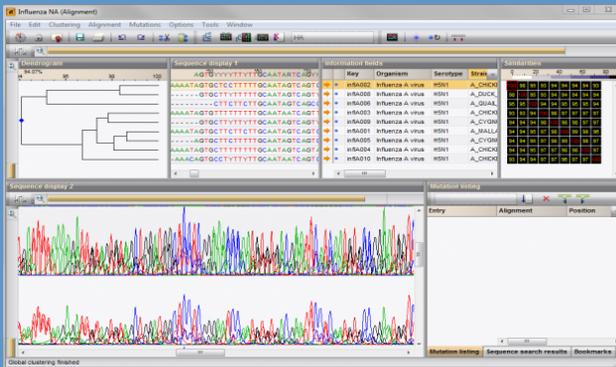
Algorithms for processing next-generation sequence data

Experimental Design

Differential Analysis

Modeling & Prediction

New statistical methods

Bioinformatics

Biostatistics

Problems of Bioinformatics Cores

- Complex hardware and sophisticated software play an important role in bioinformatics

- Technology and infrastructure – one size does not fit all

- Research labs often lack the storage, computing power and technical know-how to cope with the current deluge of genomic data.
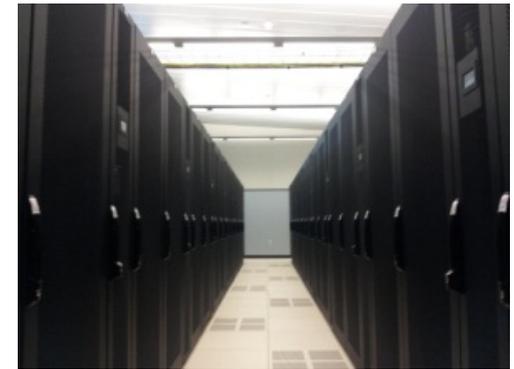
Outline

- Background
  - BISR of University of Kansas
  - Problems of Bioinformatics Cores
- Approach
  - Problems with current infrastructure
  - New approach tested with Globus
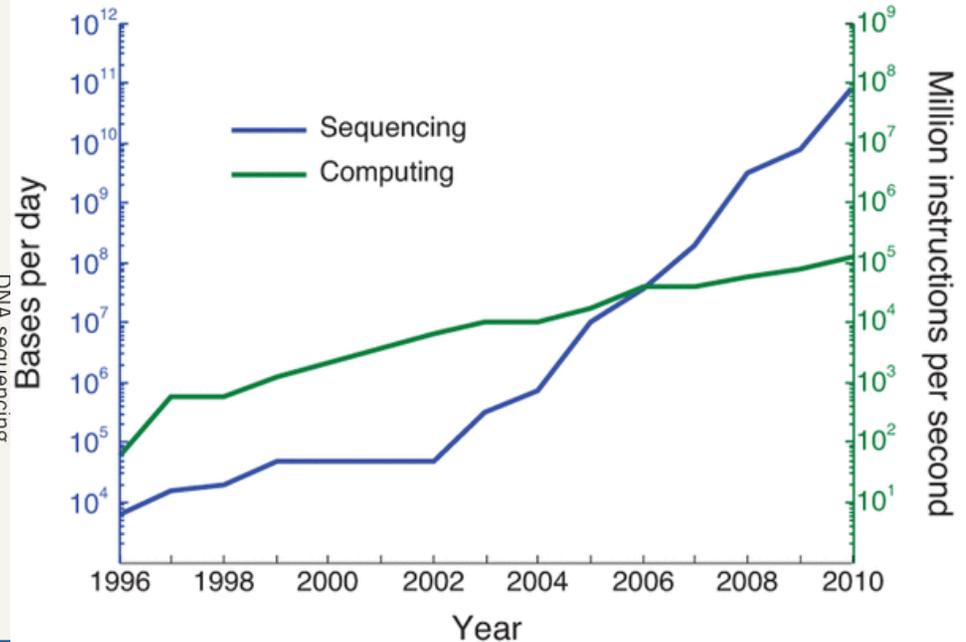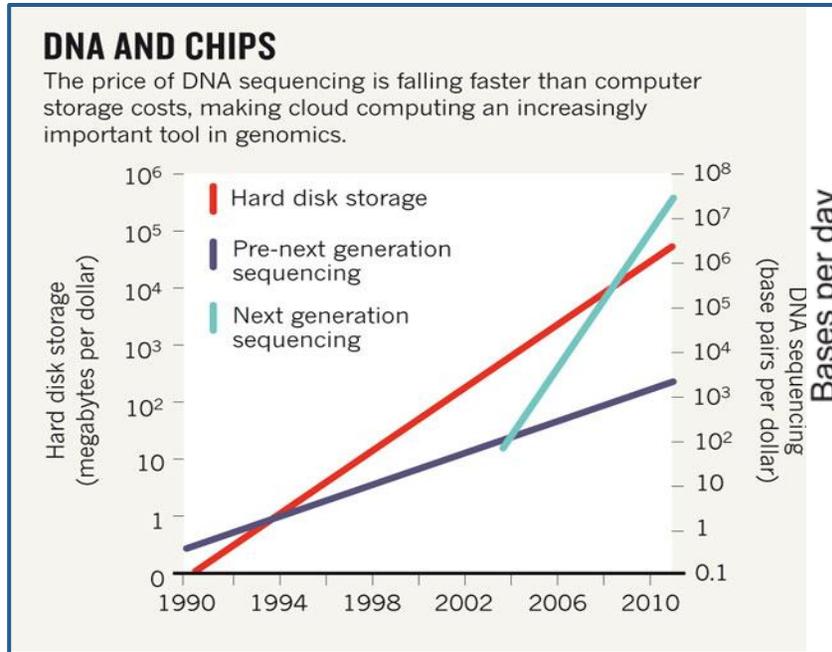- Pilot project experience
- Future plans

Current infrastructure

- Invested in computing and storage at the Advance Computing Facility (ACF)

- Investigated and tested the use of Cloud computing for analysis of de-identified data

- Yet, unable to fulfill the current computation needs of bioinformatics projects

The big problem

DNA AND CHIPS
The price of DNA sequencing is falling faster than computer storage costs, making cloud computing an increasingly important tool in genomics.

Loh, Po-Ru, Michael Baym, and Bonnie Berger. "Compressive genomics."*Nature biotechnology* 30.7 (2012): 627-630.
Next-generation sequencing: adjusting to data overload, Monya Bayer

New Approach

Samples

Investigator

FASTQ Files /
Pre-processed data

amazon
web services™

globus
genomics

Sequencing Center
(Internal or External)

BISR/KINBRE
CORES (Dr. Fridley)

New Discovery

Processed data
(.VCF, .BAM, .BE
D, count table)

Advance Compute
Facility (ACF) at KU-L

Outline

- Background
  - BISR of University of Kansas
  - Problems of Bioinformatics Cores
- Approach
  - Problems with current infrastructure
  - New approach tested with Globus
- Pilot project experience
- Future plans

Pilot

- Tested different NGS data processing tools

- Data upload/download

- Multiple data type analysis (RNASeq, Exome)

- Testing automated pipelines

- Constructing and testing new workflows

- Batch-mode job submission

- Globus transfer from pre-integrated endpoints

- Globus technical support

RNASeq analysis workflow

In process

- Consolidated billing coupled with Amazon
- Setting up endpoints
- Legal paper work
- If everything goes well, a planned approach

Outline

- Background
  - BISR of University of Kansas
  - Problems of Bioinformatics Cores
- Approach
  - Problems with current infrastructure
  - New approach tested with Globus
- Pilot project experience
- Future plans

Future plans

- Bioinformatics Analysis at K-INBRE and KUCC
  - Extend to use to other cores at KU and KUMC
- Teaching students how to process NGS data
- Provide Bioinformatics Analysis Services to the researchers in the region using Globus to transfer data and Globus Genomics for the workflows and tracking of projects.

Future

- Larger translational research
- Faster innovations

The stage is set for a great revolution in genomic science

Thank you!