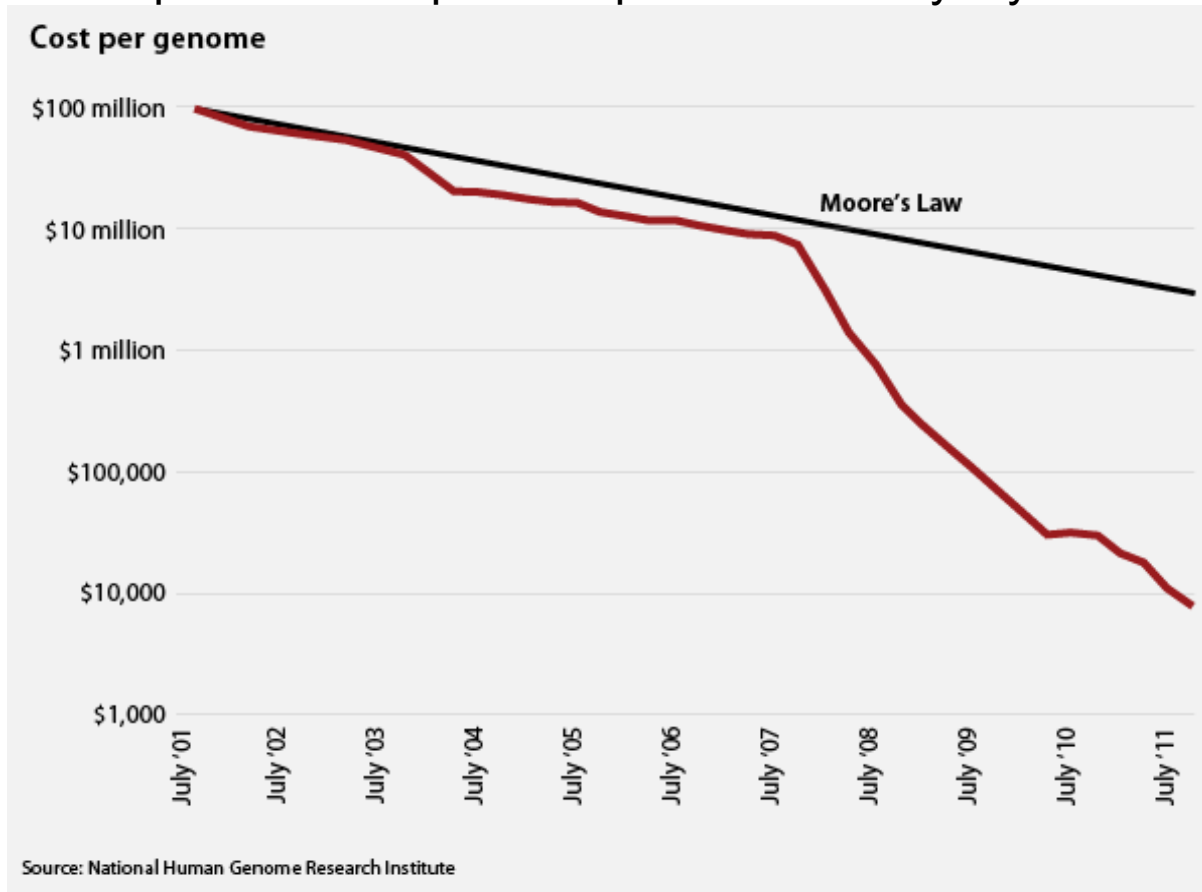# Medical Informatics: Introduction

Gil Alterovitz
Center for Biomedical Informatics
Harvard Medical School/MIT
gil@mit.edu

- **Moore's law**- The number of transistors (basic computational units) that can be put on a computer chip doubles every 2 years.

Cost per genome

Moore's Law

Source: National Human Genome Research Institute

- **Beyond Moore's law-** We are now beating Moore's law in genomics. Now, significantly more than double the number of bases (basic elements of genetics) can be sequenced every 2 years at a fixed cost.
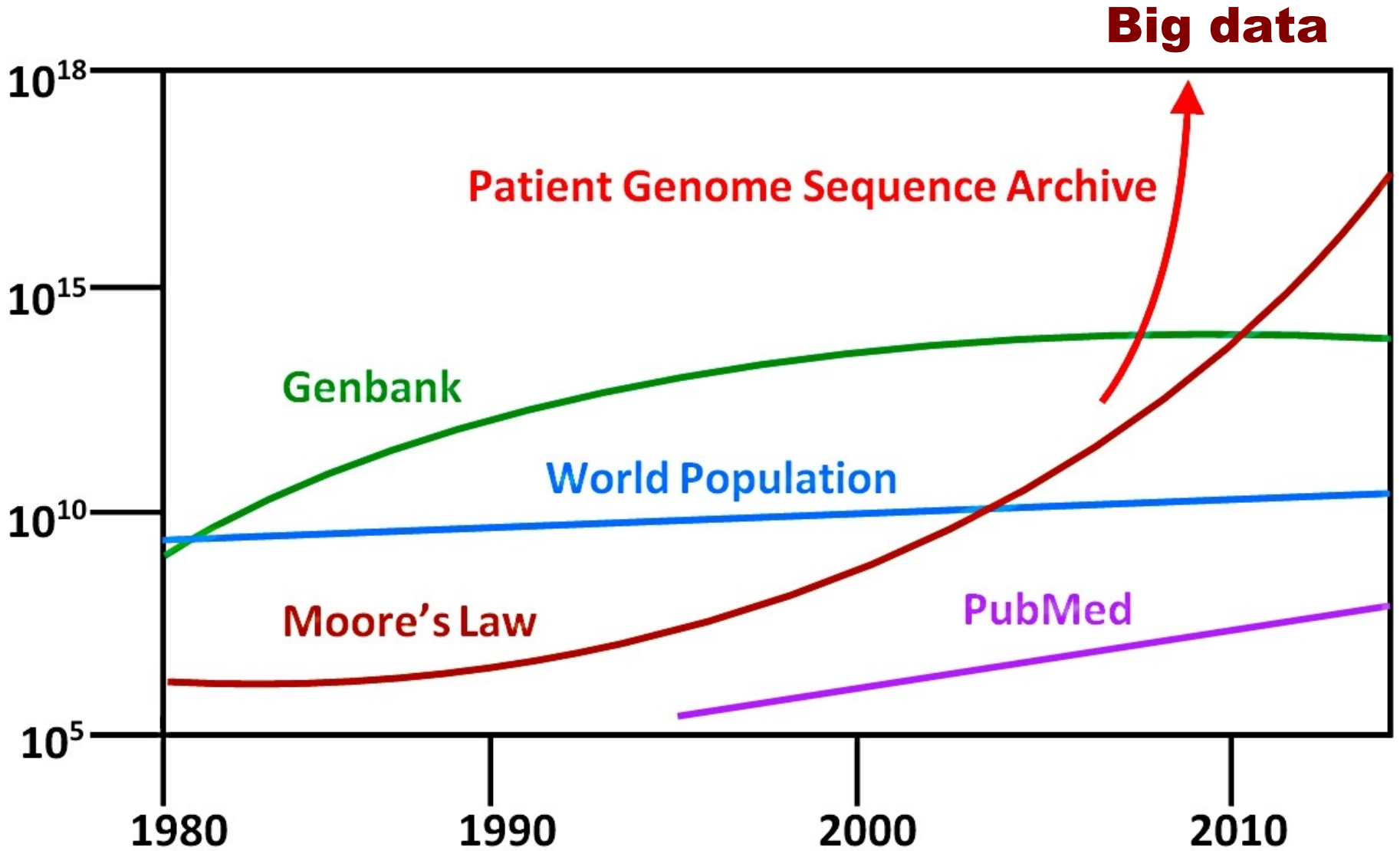
Figure 1: Approximate Growth of Different Data Populations

CHARACTERISTICS OF BIG DATA

Volume:
The sheer amount of data generated or data intensity that must be ingested, analyzed, and managed to make decisions based on complete data analysis.

Velocity:
How fast data is being produced and changed and the speed with which data must be received, understood and processed.

Variety:
Both structured and unstructured data generated by a wide range of sources.

Veracity:
The quality and provenance of received data.

Source: TechAmerica Foundation

Design: An interface for communicating genomic information

* **Presenter:** Yishen (Tom) Chen
* **Title:** The SMART Genomic API
  - ✓ Getting genomic and clinical components to talk to each other.
  - ✓ Paved way for clinical genomic apps.
* **Result:** Work being integrated into International standards by international standard organization (Fast Healthcare Interoperability Resources HL7).
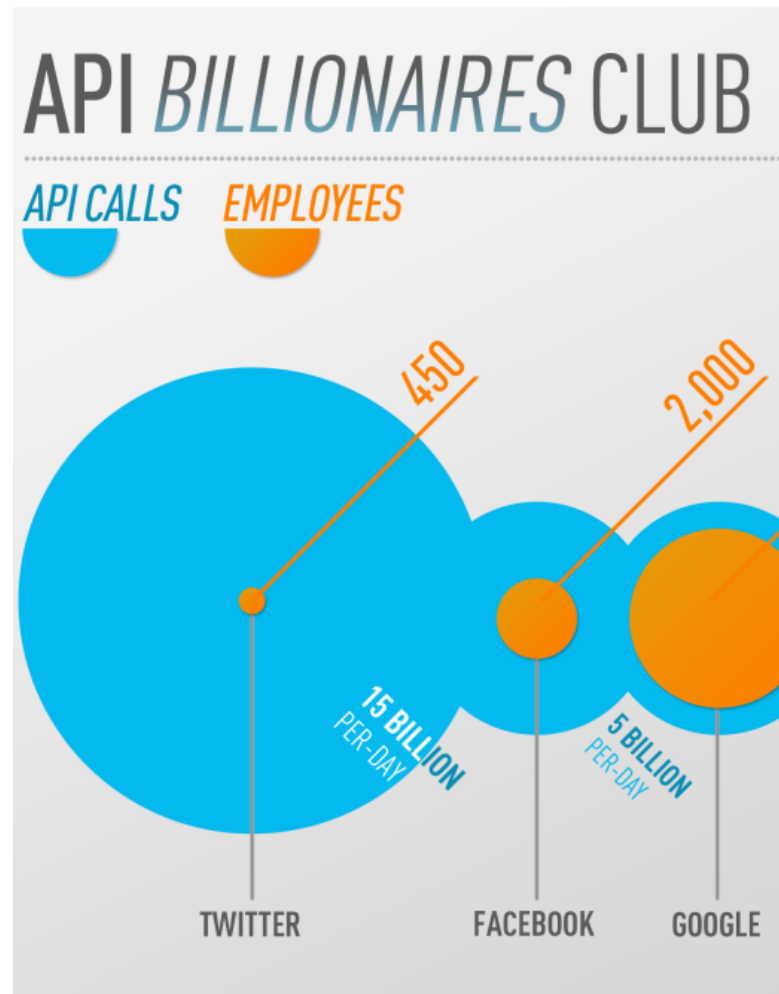
From Design to Implementation

✳ **Presenter:** Xi (Steve) Chen

✳ **Title:** Genomics development library

✳ **Result:** Created API software development kit for Android. Created sample app released on Android Market.

Translational Medicine

✳ **Presenter:** *John Zhang*

✳ **Title:** Integrating genomic, clinical, and patient questionnaire information for breast cancer diagnosis and treatment

✳ **Result:** First app to integrate clinical-grade genomics, clinical records, and family history for for cancer risk prediction

would you say you had a dominant or recessive character

gene therapy

Predictive Medicine

✳ **Presenter:** Andrew Li and Arul Prasad

✳ **Title:** Exploration of disordered proteins related to drug resistance in Hepatitis B virus and lung cancer

✳ **Result:** Discovered gene-based drug resistance significantly associated with two classes of protein disorder.

caution Super Bug!

Harvard-MIT Division of
Health Sciences & Technology

MIT COMPUTER SCIENCE AND
ARTIFICIAL INTELLIGENCE LABORATORY
CSAIL

Children's Hospital
Informatics Program

Harvard
Medical School

R&D Cycle

**Yishen (Tom)
Chen** *SMART
Genomics API*

Design

Xi (Steve) Chen
Genomics development library

Implementation

**Predictive
Medicine**

**Translational
Medicine**

Andrew Li & Arul Prasad
*Exploration of disordered proteins
related to drug resistance*

John Zhang
*Integrating genomic, clinical, and patient
questionnaire information for breast cancer
diagnosis and treatment*