Skill Obsolescence and Adaptation: Evidence from DNA Sequencing

Danielle Li and Pian Shu

Motivation

Technological change creates opportunities.

- Improved productivity/health/welfare
- Economic growth

...and challenges

- New technologies demand new skills
- Workers often find it difficult to adapt
  - Automation/computerization has contributed to rising inequality (SBTC)
  - Manufacturing workers have been particularly hard hit
  - Retraining programs are limited in their effectiveness

This project: What happens to high skill workers?

We care specifically about high skill workers

- They are the ones generating economic growth
  - Our context: biomedical R&D ($100bn/year spent)
  - $\sim$ 75% of FDA approved drugs cite scientific publications
- We spend a lot of $ on their training ("educate to innovate!")
  - Grants train scientists and engineers (many go to work in industry)
  - Many academic-industry partnerships
  - R&D tax credits subsidize private sector investments in STEM salaries and training

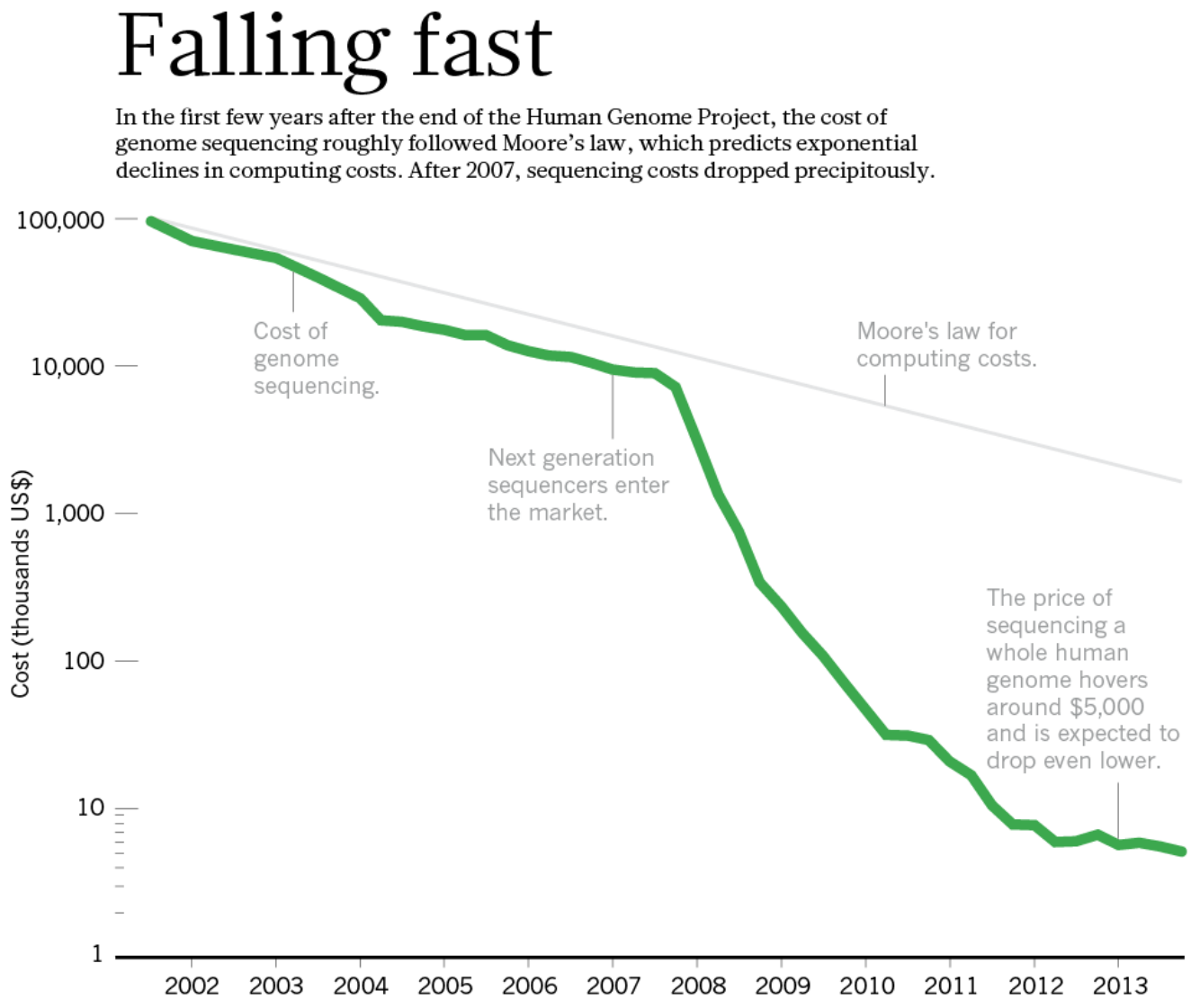
How durable is their human capital?

- More if they have higher general skills
- Less if their skills are more specific or require more investment
- Very little empirical evidence

Research Overview

Context: Biomedical research and DNA sequencing technologies

- Basic and applied research area focused on understanding the genetic foundation of disease
- DNA sequencing is an important tool in this area:
  - Allows researchers to identify differences in genes and associate those differences with diseases and phenotypes
- Lots of technological change
  - Incremental technological change: improvements in traditional technology (Sanger sequencing)
  - Big technological change: introduction of next generation sequencers (NGS)



- DNA sequencing costs fell from \$100M per human genome in 2001 to \$5,000 in 2013 – e.g. by 99.995%

Why is this important?

NGS changed how sequencing worked and what it could do

1. Sanger Sequencing (1980s to mid 2000s)
   - Based on PCR reactions going down a single strand of DNA
   - Accurate but slow, reactions are labor intensive to set up
2. Next generation sequencing (mid 2000s to present)
   - Fast parallel reads of DNA
   - Much cheaper and easier to automate or outsource
   - "Like Stata for biologists"

NGS changed what questions could be asked and what skills were important

- E.g., get medical histories for 100 people, sequence all of them, data mine for differences in DNA and correlate with differences in health outcomes
- Increases return to data analytics
- Decreases emphasis on "wet bench" lab skills

Research Questions

1. What is the impact of technological change on individual scientist productivity?
   - Who wins? Who loses?
   - What happens to research direction and team composition?
   - Adoption of new technologies / questions

2. Do new technologies challenge or amplify existing hierarchies?
   - Does having more experience and investments in old technologies help or hinder transitions?
   - Who are the stars before and after?

3. What kinds of resources aid adaptation?
   - Access to collaborators and new human capital

What is the impact of technological change on individual scientist productivity?

NGS is more useful to biomedical researchers focused on DNA than those who are not

- Treatment group: those who are "at risk" of being directly affected by NGS (more on this later)
- Control group: Biomedical researchers that were not.

Simple DD design

$$Y_{it} = a_0 + a_1 \text{Post NGS}_t \times \text{At risk}_i + \delta_t + \delta_i + e_{it}$$

- Y_{it} include: publication/grant outcomes, exit from publishing, use of NGS technology (inferred from abstracts)

Data Construction I: Sample and Treatment

Want to identify all biomedical researchers publishing prior to the introduction of NGS:

1. Begin with sample of all NIH applicants 1993-2006 (not necessarily awardees) ($\sim$ 150,000 people)
2. Collect their coauthors (will cover graduate students and other faculty collaborators)
3. (1.5 mill unique Scopus IDs and counting....)

Identify scientists "at risk" of being exposed to changes in sequencing technology:

1. Categorize publications as related to sequencing using MeSH keywords
   - "Sequence Analysis"
   - "Oligonucleotide Array Sequence Analysis" etc.
2. By author, construct proportion of pre-NGS sequencing-related publications
3. Can also define at risk by use of DNA instead, etc.

Data Construction II: Outcome Variables

Want to measure individual productivity:

- Citations, publications, etc.
- Grant funding

Adaptive behaviors:

- Change in collaborators
- Adoption of new technologies, based on abstracts and keywords
  - "High-Throughput Nucleotide Sequencing"
  - "Genome-Wide Association Study" etc.

And career outcomes:

- Entry/exit from publishing

Do new technologies challenge existing hierarchies?

When do technological "disruptions" also "disrupt" economic hierarchies?

- Established scientists/institutions have more resources: should be able to transition faster
- Established scientists/institutions have made greater technology specific investments: should make transition slower
- Can new technologies reset the playing field?

Compare differential treatment effects among exposed scientists (either as DDD or split DD):

$$Y_{it} = a_0 + a_1 \text{Post NGS}_t \times \text{Established}_i + \delta_t + \delta_i + e_{it}$$

- Established_i is defined by funding and publication success in pre-period

Do new technologies challenge existing hierarchies?

Want to distinguish:

1. Differences in adaptation may reflect differences in general skills
   - E.g. If established groups do better under both regimes, it may be due to higher unobserved general quality

2. Or differences in technology specific skills/investment
   - E.g. if established groups are more reluctant to adopt the new technology and other groups catch up as a result
   - To test: hold general human capital fixed, but shift degree of specific investment
     - Randomly force some universities to make investments in the old sequencing technology
     - Examine how scientists at these institutions adapt to NGS, relative to scientists at treatment institutions

Do new technologies challenge existing hierarchies?

We compare groups with comparable general skills, but differential access to established technology

- In 1996, the NIH established 6 Human Genome Project sequencing centers
  - Washington U, Whitehead, Baylor, TIGR, University of Washington, Stanford
- But there were many comparable institutions that were not awarded
  - Duke, Columbia, U Chicago, Texas SW Medical Center, Berkeley, Scripps, Johns Hopkins, etc.

Instrument being established with being at a treated institution

$$Y_{it} = a_0 + a_1 \text{Post NGS}_t \times \text{Established}_i + \delta_t + \delta_i + e_{it}$$

- Asks whether being established aids adaptation among scientists at comparable institutions
- Caveat: being assigned a genome center may have increased general quality of these institutions

What kinds of resources aid adaptation? Access to collaborators

NGS increased returns to data analysis

- Scientists need to 1) actually learn these skills or 2) find someone who knows them
- Returns to specialization and teamwork

Universities with similar biology programs differ in their access to computer science, statistics, and bioinformatics collaborators

- Quality of departments
- Physical proximity of offices

Are molecular biologists with access to quantitative collaborators better able to adapt to technological change?

$$Y_{it} = a_0 + a_1 \text{Sequencing Cost}_t \times \text{Collaborators}_i + \delta_t + \delta_i + e_{it}$$

Next Steps

- Collecting publications for some scientists
- Coding "at risk" set (suggestions for other ways to do this?)
- Other ways to measure change in research direction?
- Physical proximity of offices