1

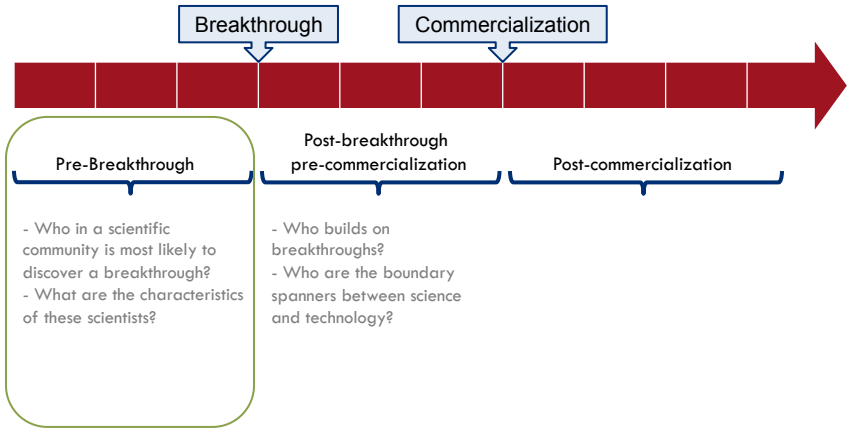
Emergence of Creative Breakthroughs using Author-ity database

Sen Chai
Harvard Business School

UIUC – Workshop on Disambiguation
June 2012

Research Agenda

2



Research Questions

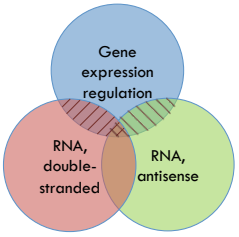
3

- What are the sources of a scientific breakthrough?
- At any given point in time, how accurately can we predict the sources of future breakthroughs?
- Breakthroughs: Notion of impact that encompasses creative novelty and success (Simonton, 1999)
- Motivation
  - Scientific breakthroughs are sources of
    - Economic growth (Jorgenson et al. , 2008)
    - Social benefits (Trajtenberg, 1990)
  - Managerial and policy implications
    - Targeted governmental subsidies and private investment (Lane, 2009)
    - Help firms in identifying and producing more at risk breakthrough research

Analysis

4

- Setting & Data
  - RNA interference (RNAi) - natural mechanism that silences genes by turning them genes on/off
  - Completely new drug class to cure genetic diseases
  - Mechanism understood in 1998, Nobel prize awarded in 2006
  - Author-ity PubMed database for scientific papers
  - Community of scientists working on understanding RNAi mechanism
    - MeSH keywords: Gene expression regulation AND (RNA antisense OR RNA double-stranded)
    - Limited to papers published until 1999
    - 1,551 papers with 3,959 unique authors



Methods

5

- Hybrid methodology
  - Quantitative analysis (with Lee Fleming)
    - assess predictive power of current theories on future breakthroughs with OLS models
    - quasi-maximum likelihood Poisson count models
    - individual author level of analysis
  - Qualitative fieldwork
    - inductively generate theory on the phenomenon of breakthrough emergence
    - interviews of scientists at risk of breakthrough

Quantitative Analysis

6

- Outcome Variables (data in year 1998)
  - Impact
    - number of forward citations
- Control Variables (data from years prior to 1998)
  - Publication History – all papers and papers as first & last author
    - number of publications prior to 1998
    - number of forward citations for publications prior to 1998
  - Affiliation
    - academia vs. industry
    - number of affiliations
  - Affiliation Prestige

Quantitative Analysis

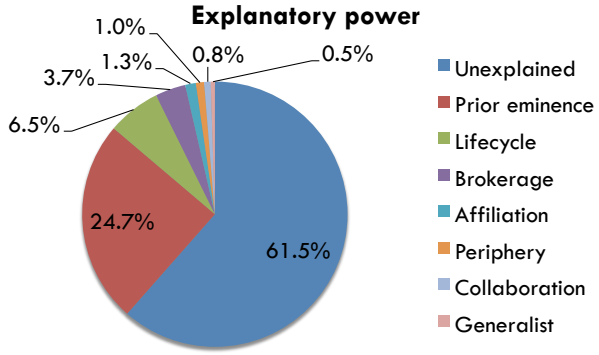
7

- Explanatory Variables (data from years prior to 1998)
  - Brokerage vs. Cohesion – constraint
    - Brokerage: nexus position that facilitates information control
    - Cohesion: trust that allows richer lateral diffusion
  - Collaborative vs. Individual researcher - # co-authors
    - Collaboration: idea selection and feedback
    - Individual: coordination costs
  - Periphery vs. Core – technical core & collaborative core
    - Periphery: ignorance of prevailing assumptions and theories
    - Core: more influx and faster flow of information from social ties
  - Specialist vs. Generalist – publication depth
    - Specialist: deep knowledge optimally recombine components
    - Generalist: not bound to the current thinking in the focal field
  - Lifecycle – experience
    - Early career: less weighted down with established beliefs
    - Late career: burden of knowledge to be at frontier of science

Results

8

- *Brokers*, scientists at the *technical periphery* and relatively *mature* researchers are consistently associated with higher risk of breakthrough discovery.
- Theoretical themes account for 13% of variance



Qualitative Inductive Analysis

9

- Motivation
  - More than 60% of variance is still unexplained
- Goal
  - Induct explanations of individual influences on breakthrough discovery
  - Uncover other influencing factors missed in the current literature

Qualitative Analysis

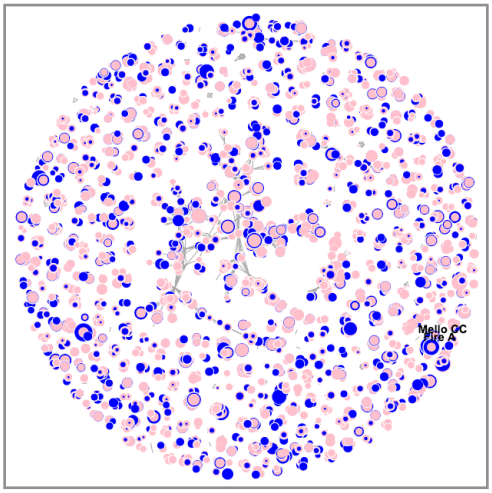
10

- Qualitative interviews of 16 scientists around and at center of breakthrough
- Sampling of interviewees
  - Residual analysis of impact model
    - Scientists who overperformed, and should have done well but underperformed
  - Conference attendance at the inception of the field
- Interviews are semi-structured lasting from 60 to 100 minutes
  - Breakthrough & circumstances around the breakthrough
  - Scientist's productivity and research focus at the time
  - Community definition and characteristics

Interviewee sampling – Residual analysis

11

- Scientific collaborative network maps
  - Node: author
  - Edge: co-author relationship
- The predicted versus actual impact (number of forward citations) of RNAi community scientists in 1998
  - Blue: actual impact
  - Pink: predicted impact
- Interviewees
  - Overperformers: blue ring
  - Underperformers: pink ring



Nature of Scientific Research

12

- Element of truth seeking in science
  - Multiple observations before understanding the mechanism of the phenomenon
- Unexpected observations are dismissed as artifact and ignored
  - Antidogmatic
  - Fear of being wrong
  - Main hypotheses and results are unaffected
- Lack of transparency in reported results

Antidogmatic & Fear of being wrong

13

"Cause if you think about it if you were sitting in a lab in the middle of nowhere injecting dsRNA into c.elegans, and seeing it having an effect, a really good effect, a really strong effect on gene expression and it doesn't work with single-stranded RNA, and ***no one has ever seen this before, you can't write this up****. You must have put out a few fingers to see whether anyone have heard of anything before. "*

Main hypothesis and results are unaffected

14

"[…]When you have a well defined system and it's telling you something you don't understand, it isn't consistent with the way you've designed the system then something is new in the system. It's paying attention to that [bizarre phenomenon] and ***not pushing it out of the way as you went towards your more conventional hypothesis driven science****. That meant the difference between the genius and good science."*

Lack of transparency in reported results

15

"It's almost like each of us has a little piece of the puzzle but by the time we are ready to show the puzzle piece to the audience we've filed off some of the pieces we don't like about it and now of course it doesn't fit. The other guy has got the other piece of the puzzle but of course ***it doesn't fit cause we have changed the shape of it.***"

Increasing breakthroughs – theme 1

16

- Increase number of effective attempts
  - Designing well-controlled experiments
    - Confidence in experimental results
  - Seeking for evolutionary conservatism – making parallels between model organisms as double-checking mechanism
    - Literature
    - Conferences
    - Proximity to other labs
    - Multiple organisms in one single lab

Conferences

17

- Actively seek input on unexpected results

  "[You can] talk to people about ***some surprising thing that you're finding and get input and be able to test ideas with****."*

- Passively find others' results that corroborate your own

  "And it only requires you going along to one seminar. We've been clearly influenced. ***We had a theory, we didn't have any confidence in it, and this guy from Harvard shows up and talked about something utterly different, and you think that's worth doing a few experiments****."*

Research ecosystem

18

- Proximity to other labs

  "If you're at a place like MIT where there are ***labs that have the expertise in each of these systems usually in the same building or across the street****, it's very easy for students and post-docs to start a project in these systems and get help from their friends in the labs."*

- Multiple organisms in one lab

  "And for example, what I really liked [in one of the labs I was working at] is that even in a single lab we were working on ten different organisms. […] ***We could not only go and find the details of silencing we were studying in each organism****, we could also make the parallel and trying to find what was common between these different mechanisms, how did it start, how did it evolve."*

Increasing breakthroughs – theme 2

19

- Increase search variance and widen search space
  - Exploiting and exploring at the fringes - Ambidexterity

    *"You know, if you're running [a lab of] 30 people you can do **some things in the fringe** if you've got really bright people."*

  - Teaching

    *"Because when you teach you need **to read about things which you are not directly involved in** […] For example, I have one paper which has been cited more than six hundred times, and this paper actually came from the fact that I was teaching."*

Summary of findings

20

- Increase number of effective attempts given same search space
  - Designing well-controlled experiments
  - Attending conferences
  - Building conducive research ecosystem
- Increase variance of search and widen search space
  - Ambidexterity
  - Teaching

Operationalizing new measures

21

- Conferences
  - Conference attendance
- Proximity of other labs
  - Size of scientists' department
- Multiple organisms
  - Mix of organisms in MeSH keywords
- Ambidexterity – exploration & exploitation
  - Survey data
- Teaching
  - Cross-disciplinary courses taught

Questions & Feedback

22

Supplemental materials

23

Interview questions

24

- Breakthrough
  - In the period of 1997-1998 were you and your peers aware that a breakthrough was about to be discovered? Was there excitement due to a potential impactful discovery?
  - Were scientists trying to solve a specific puzzling mechanism or did they just happen to stumble on the RNAi mechanism by chance while looking for something else?
  - Were there many teams working towards solving the same problem? Was there racing?
  - Do you feel like the breakthrough could have been made earlier? Why? What was the missing link that prevented it?
  - Was the discovery and its results a surprise? In terms of simplicity or complexity of the solution, in terms of who made the discovery?
  - Before you chose your research direction, how do you evaluate the potential impact of your research? How?
  - What papers or findings spurred your interest in RNAi research? What works had a decisive influence on your research interests?
  - What experiments, field or prior breakthroughs do you believe paved the road to the discovery? What inventions (tools), environment fostered the discovery?
  - Were you aware of the similar co-suppression and quelling results obtained in plants and fungi? / As a plant scientist did you think that co-suppression and quelling would be present in animals?

Interview questions (ctd.)

25

- Community
  - Was there a defined community of RNAi scientists prior to breakthrough?
  - How would you define the community of RNAi scientists prior to breakthrough? Which subfields of biology came together to form such a community?
  - How would you characterize this community? Social, open or collective?
  - How open was the community of scientists working towards solving this discovery? Was there an informal group established that frequently communicated and shared their ideas? Or were results withheld?
  - What kind of conference/research seminars did you attend at the time, was it phenomenom-based, organism-based or something else?
  - How do you think about conferences? What role do conferences play in your research?
  - In your opinion, did the breakthrough come from within the community or from outside?
  - In your opinion, who were the big contenders in the community to discover the mechanism to RNAi? Why?
- Diffusion and commercialization
  - What makes a breakthrough widely recognized and used?
  - Why do you think the RNAi breakthrough was commercialized fairly quickly compared to other breakthroughs in biology?
  - How do you decide whether or not to patent a piece of research?
- Theory Building
  - Explain the theoretical perspectives of breakthrough emergence in the social science literature, and ask whether and why they agree or disagree?
  - According to your experience on the patterns of discovery, do you think the theories I presented are comprehensive or whether alternative explanations exist?