The role of young researchers in the knowledge creation process in science

Masatsura IGAMI*, Sadao NAGAOKA**
and John P. WALSH***
EU-SPRI, Milano, 28 June 2012

Understanding Organization of Science

- Increasing interest in organization and economics of science (Stephan, 2010, Cummings and Kiesler).
- Especially with increase in team science (Wuchty, et al., 2007).
- In addition, concerns about scientific labor markets and need to encourage young STEM researchers (NAS, 1998, BRWWG, 2012)

Role of Young Scientists

- Research Question: What is the role of young scientists in scientific research?
  - Are post-docs a conduit for new knowledge and serendipitous findings, or are they largely extensions of the PI? (Vogel, 1999, BRWWG, 2012).
- Vary by field? By country?
- Grad students versus Post-docs?
- What is the impact of papers with young first authors?
  - Types of knowledge used?
  - Impact on field (highly cited)?

Japan-US Scientists Survey

- Survey of Japanese/US authors on scientific papers
- Ask about details of research project that produced the paper
- Including project team composition
- And, knowledge inputs
- Also bibliometric data on citations (top 1% v. others)

The population of the survey

- Articles and letters in Science Citation Indexes-Expanded (Thomson Reuters)
- Time window: 2001 – 2006 (database year)
- Forward citation counts were retrieved on December 31, 2006
- 22 fields in the ESI were adopted
- Sampled about 9000 papers, 1/3 from top 1% in citations (High), and 2/3 from rest of populaiton (Normal)

7

Topics covered

- Motivation and the other basic characteristics of the research project: Stokes quadrants, serendipity, and scientific competition
- Knowledge production process: respondent's roles in the project and uses of external knowledge and geographic location of those knowledge sources, and organization of research project
- Research inputs: project duration, funding and sources of funds
- Composition of the research team: rank, organization type, field, country of origin, research skill/specialty, gender, also number of students trained and personnel hired by project as well as authorship rules
- Outputs: number of other papers, patents, licenses, startup firms;
- Scientists' demographics, family status (marital status, children), education and training, mobility, awards, and publication counts

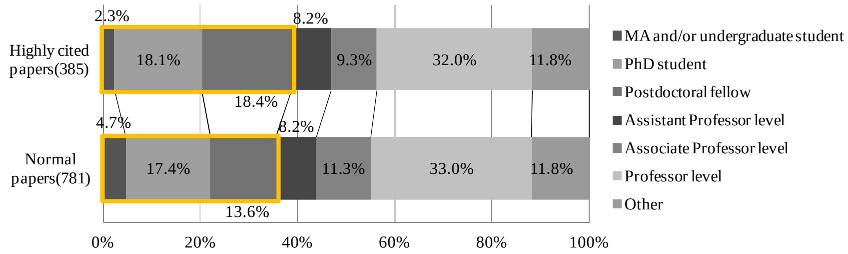
Survey timeline and response rates

- Initial mail outs: September-November, 2010
- Reminder emails: November-December, 2010
- Second (final) reminders: January, 2011
- 2329 responses (as of Jan 28, 2011), out of 8,864 survey targets
- The total response rate is approximately 26% using total mail out as the denominator
  - we may need to adjust this as we finish auditing cases for eligibility
- Fields with over 30% response rate are environment/ecology & geosciences (37%), social science (32.4%) and agriculture, plant, and animal science (31%).
- Response rate in clinical medicine & psychiatry/psychology is the lowest (21%)
- The response rate of H papers seems to be higher than or at least equal to the response rate in N papers in almost all fields.
  - Exception is the computer science & mathematics (3% higher for the N papers)

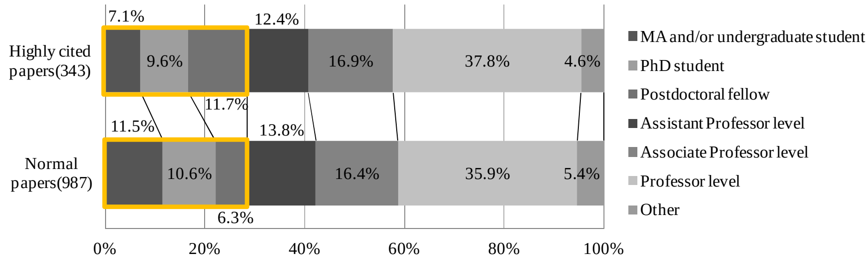
Young Authors

- Any author (Japan/US)
  - Post-doc
  - PhD student
  - MA/undergrat
- First author (Japan/US)
- By field, top v. other (Japan/US)

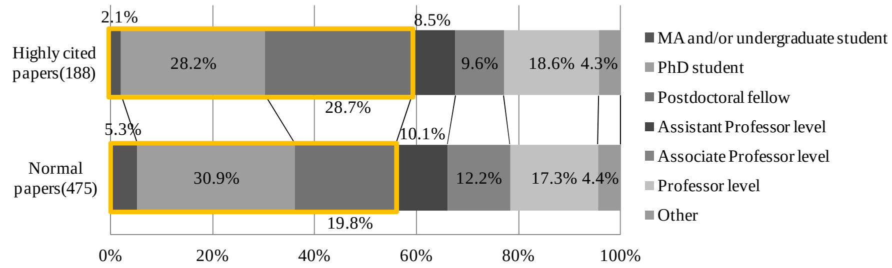
Author composition, all authors, Higher education institutions, US (natural sciences)



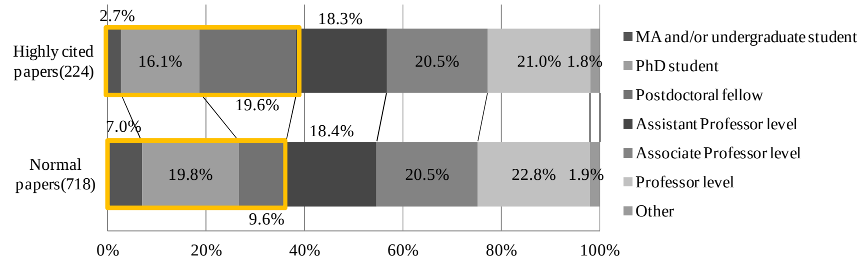
Author composition, all authors, Higher education institutions, Japan (natural sciences)



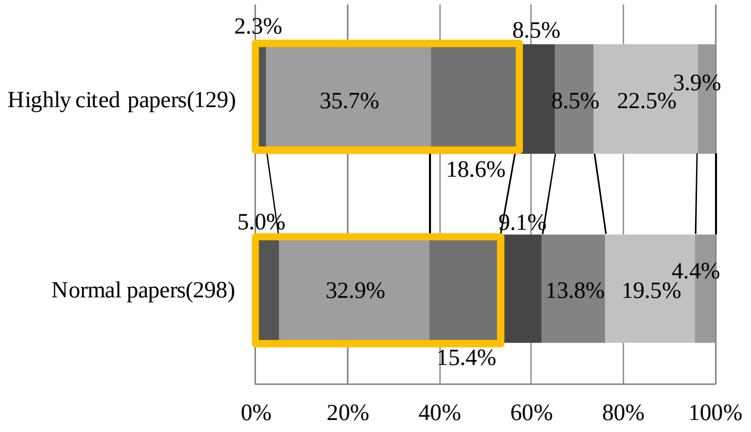
Author composition, first authors, Higher education institutions, US (natural sciences)



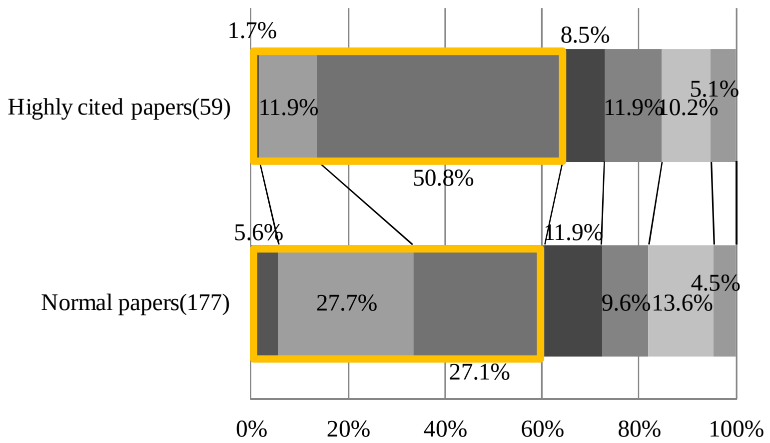
Author composition, first authors, Higher education institutions, Japan (natural sciences)



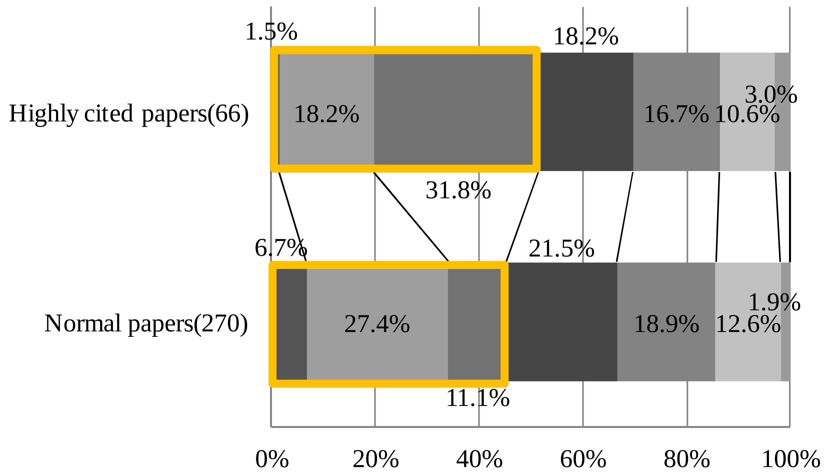
Author composition, first authors, Higher education institutions, US (physical sciences)



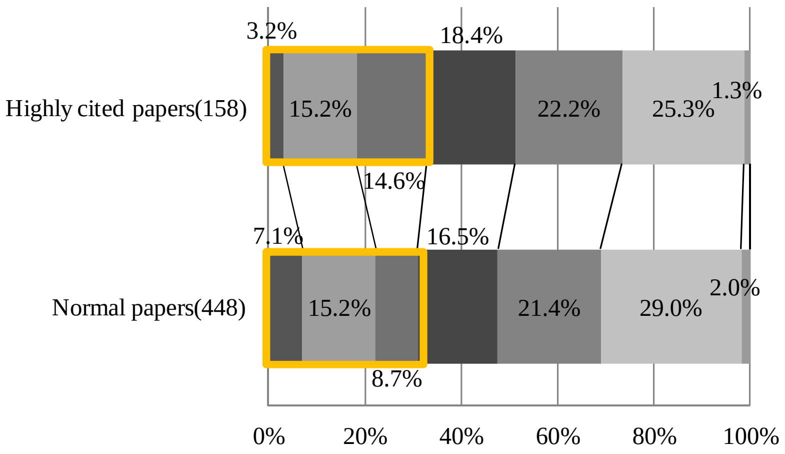
Author composition, first authors, Higher education institutions, US (life sciences)



Author composition, first authors, Higher education institutions, Japan (life sciences)



Author composition, first authors, Higher education institutions, Japan (physical sciences)



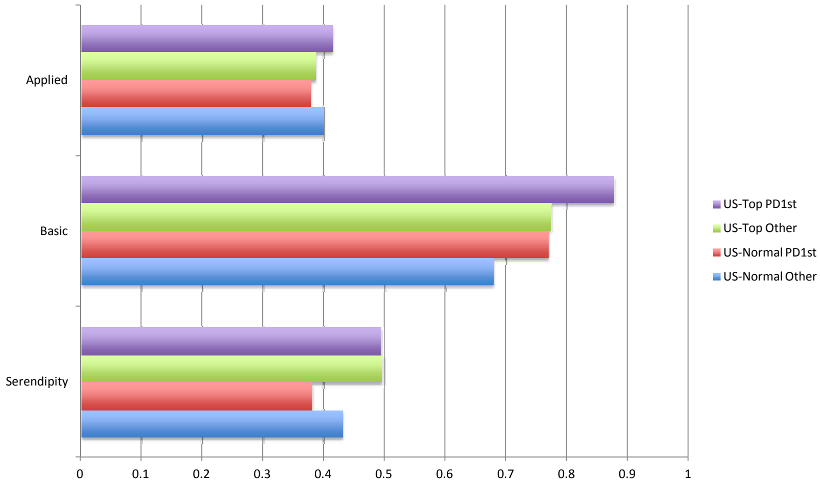
Findings

- Young researchers make up a significant share of authors, and disproportionate share of first authors
- In life science top papers in the US, post-doctoral researchers dominate (over half)
  - Over 30% in Japan as well
- Post-doctoral researchers are key to life science laboratories
  - However, may reflect labor surplus as these may be frustrated aspirants to assistant professor roles

Project Types

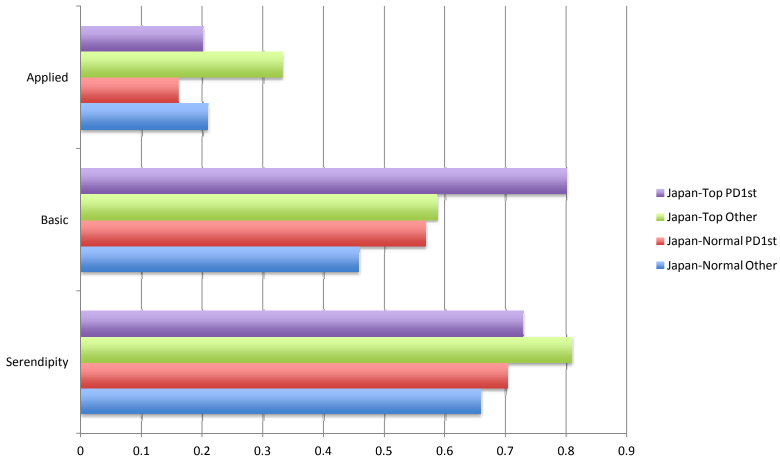
- Do PD projects specialize in basic, applied, Pasteur’s quadrant?
- And, are they more (or less) likely to generate serendipitous findings

Project Types, by post-doc papers, top, US



Note: difference in **basic**, top and normal, significant (.05)

Project Types, by post-doc papers, top, Japan

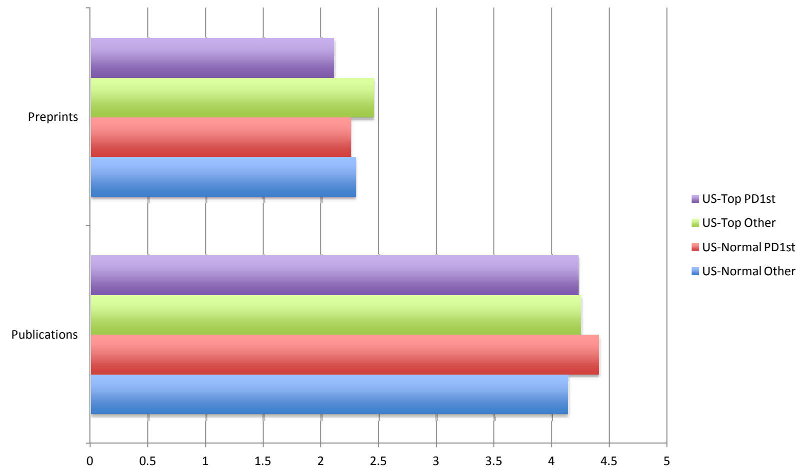


Note: difference in **basic**, top and normal, significant (.05), and in **applied**, top, sig, (.05).

Knowledge

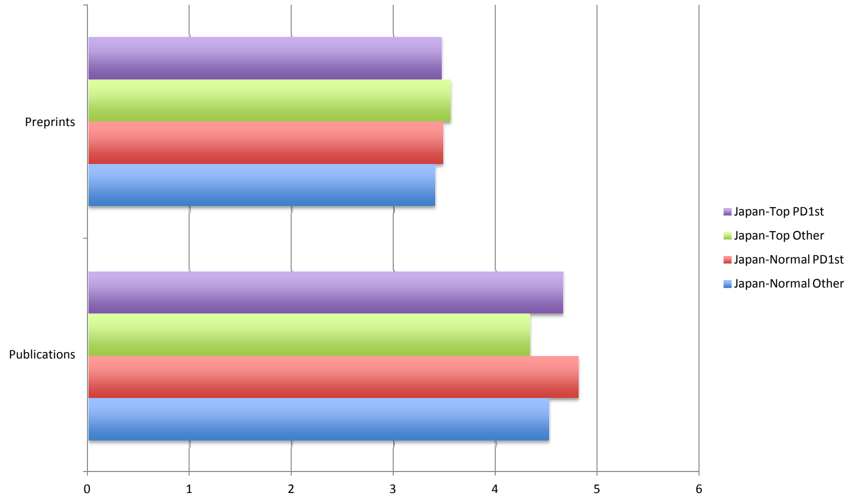
- Are young researchers key to bringing new knowledge into the lab (Vogel, 1999)?
- Measures of uses of outside knowledge
  - Access current published literature
  - Access pre-print literature
- Compare young author papers to others

Knowledge Sources, by post-doc papers, top, US



Note: difference for **publication**, normal significant (.05)
Difference in **preprints**, top significant (.05)

Knowledge Sources, by post-doc papers, top, Japan



Note: difference for **publication**, significant for top (.10), normal (.05)

Findings

- We find that post-doc papers are more likely to be focused on fundamental principles/understanding (basic), somewhat less likely to focus on solving practical problems
- We also find that Top post-doc papers in the US are more likely to result in a patent application (21% v. 11%, p<.05), though not in Japan
- We find some evidence that post-doc papers are more likely to access outside knowledge from publications (post-docs may be more current on literature, or more active in reading recent publications)
- May be somewhat less plugged in to preprint literature (weak evidence, with only US top as significant

Post-docs and Project Performance

- Are post-doc's papers more likely to produce top quality papers?
- And, does this vary by context (field, country)?
- Important for understanding contribution of post-docs to science system
- Regress "top" paper on post-doc first author, controlling for project size, industry ties, basic, field.
  - Field-specific regressions (life v. physical)

Logistic regression of top paper on post-doc first author

		US			JAPAN	
	ALL (field dummies)	LIFE	PHYSICAL	ALL (field dummies)	LIFE	PHYSICAL
Post-doc 1st	.64*** (.20)	.88*** (.27)	12 (.32)	.84*** (.21)	.92*** (.32)	.46 (.31)
Authors	.23*** (.03)	.26*** (.04)	.19 (.04)	.16*** (.04)	.25*** (.03)	.08* (.04)
Ln(months)	-.05 (.09)	-.10 (.12)	-.08 (.12)	-.01) (.07)	.18 (.12)	-.14* (.08)
Funds/100K	.01** (.00)	.01*** (.00)	-.00 (.00)	.02*** (.00)	.01* (.01)	.02*** (.01)
Industry funds	-.73** (.32)	-.55 (.43)	-.89* (.50)	.07 (.17)	.00 (.27)	.08 (.23)
Company co-author	-.26 (.42)	-1.33 (.81)	/39 (.47)	-.02 (.26)	.43 (.41)	-.01 (.34)
Basic	.72*** (.22)	.49* (.30)	.78** (.31)	.59*** (.17)	.68** (.26)	.33 (.20)
N	732	369	363	1089	501	588
Chi-sq	91.2	55.3	27.3	80.6	81.0	22.0

Predicting top papers

- Post-doc papers (first author) are more likely to be top papers
  - In US and Japan
- However, this is only the case in life sciences, not physical sciences, engineering
  - Results largely robust if do detailed field by field analyses
- Related to organization of labs in different fields? Labor market?
  - For example, large biology labs may be organized as loosely coupled decentralized projects (subcontracting?) while in physical sciences, labs may be organized in a more integrated, centralized project (Vogel, 1999; Swatez)

Conclusions

- Young researchers make up a significant share of authors, including first authors
- Post-doc authored papers more likely to be top quality
- But, this effect largely limited to bio-medical fields
- Next questions: drivers of composition?
  - Equipment needs as driver of team composition?
- Explanation of differential impact of PD across fields (and within fields)
  - Centralized/decentralized?
  - Interdependence?
  - Heterogeneity?