Measuring Scientific Impact at the NIH

Council of Councils

June 19, 2015

George Santangelo
Director, Office of Portfolio Analysis (OPA)
Division of Program Coordination, Planning,
and Strategic Initiatives (DPCPSI)
National Institutes of Health (NIH)

NIH National Institutes of Health
Office of Portfolio Analysis

Acknowledgments

OPA Analysts

Ian Hutchins
Jean Yuan
Paula Fearon
Carole Christian
Rob Harriman
Patricia Forcinito
Adam Apostoli
Aviva Litovitz
Matt Perkins
Ling Bai

OPA Software Developers

Fai Chan
Kirk Baker
Ehsan Haque
Jason Palmer
Kevin Small

OPA IT Specialist

Chuck Lynch

NIH Center for Information Technology

Calvin Johnson
Krishna Collie

NIH National Library of Medicine

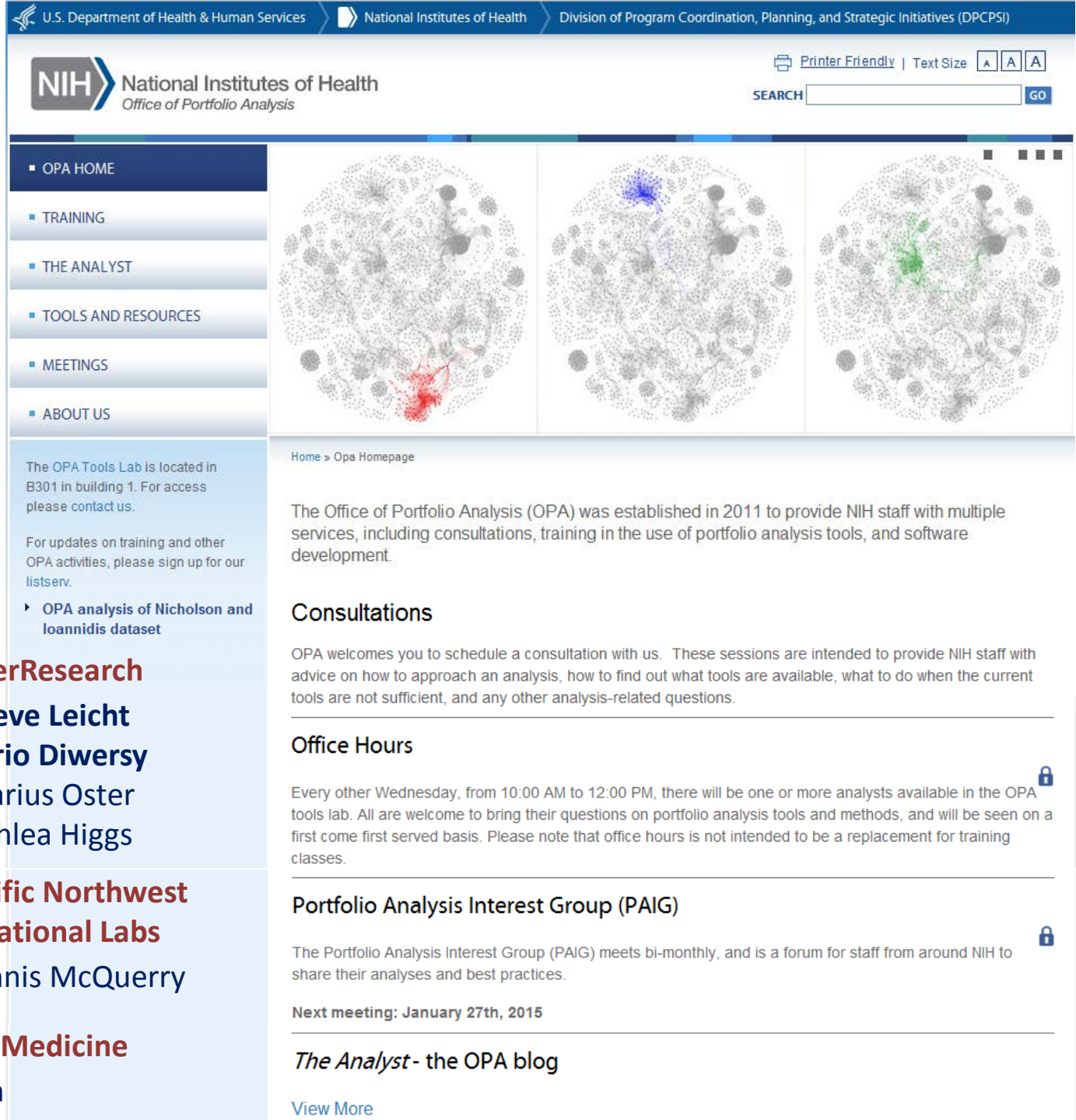
Tom Rindflesch

ÜberResearch

Steve Leicht
Mario Diwersy
Marius Oster
Ashlea Higgs

Pacific Northwest National Labs

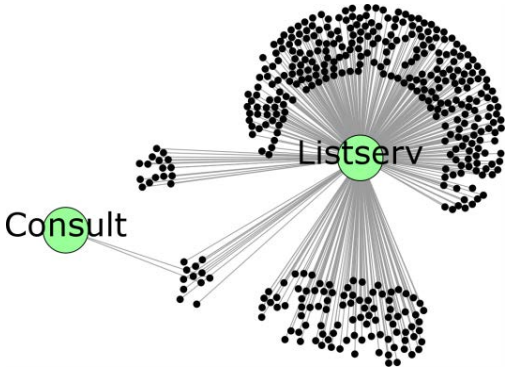
Dennis McQuerry



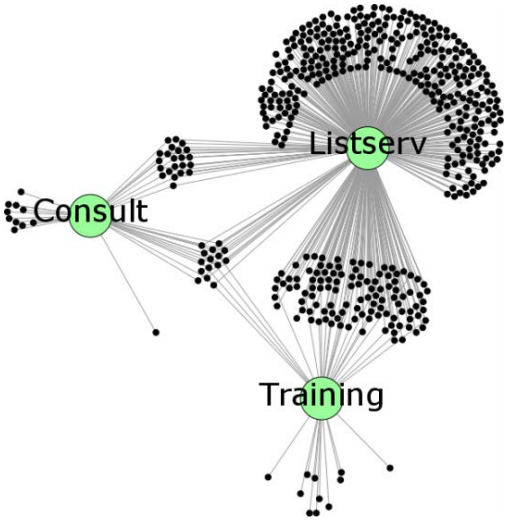
Mission of the Office of Portfolio Analysis: Improving data-driven decision-making at NIH

- **Coordinate portfolio analysis activities at NIH**
  - Conduct analyses for NIH senior leadership
  - Database management and data cleaning
  - Plan and host Seminars, Workshops, and Symposia
  - Create opportunities for crosstalk within the NIH community
    - ❖ Portfolio Analysis Interest Group (PAIG) and blog (*The Analyst*)
- **Consult**
  - Assist NIH staff in the 27 Institutes and Centers (ICs) with analyses
    - ❖ Has resulted in collaborative development of tools, case studies, etc.
- **Train**
  - Both formal classes and ad hoc sessions
  - OPA web site: user manuals, FAQs, instructional videos (under construction)
- **Develop a science of portfolio analysis**
  - Build new tools / approaches and augment pre-existing ones
    - ❖ Primary focus is biomedical research
  - Build a community of experts: government, academia, private sector

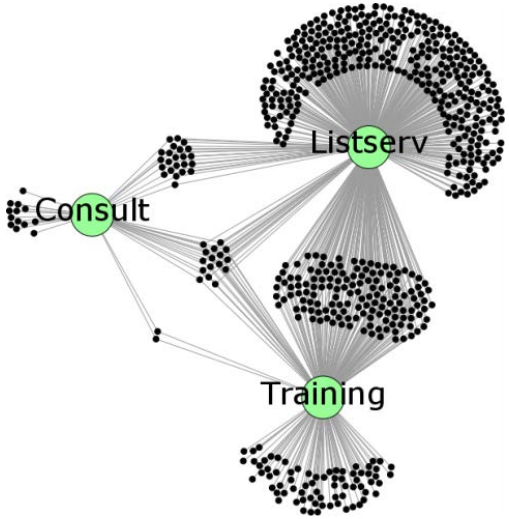
Coordinate, Consult, Train: Building the OPA network



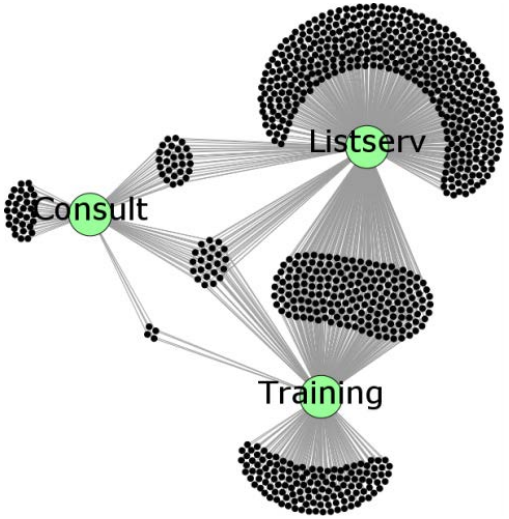
April 2012



April 2013

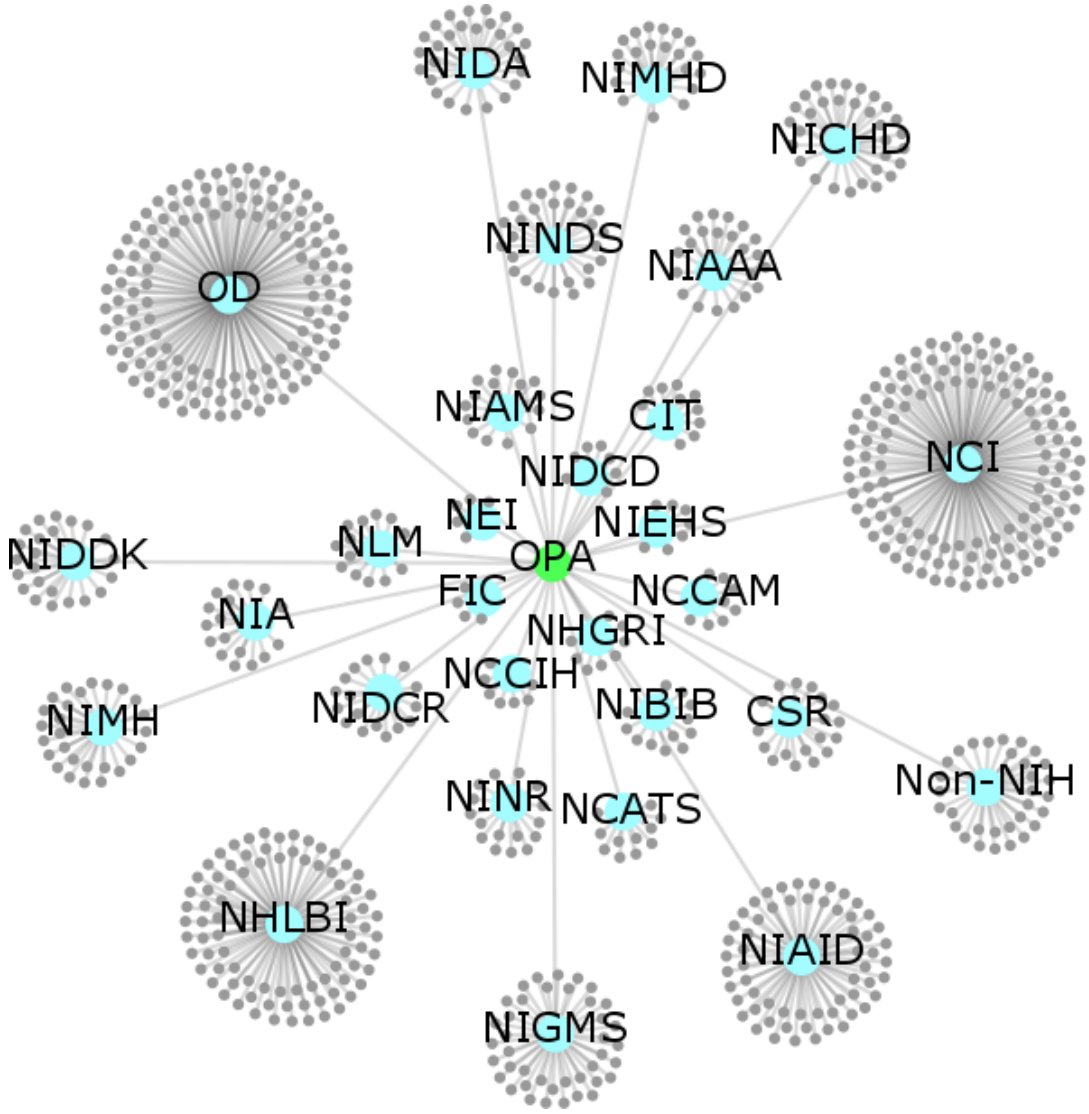


April 2014



April 2015

The OPA network by Institute/Center



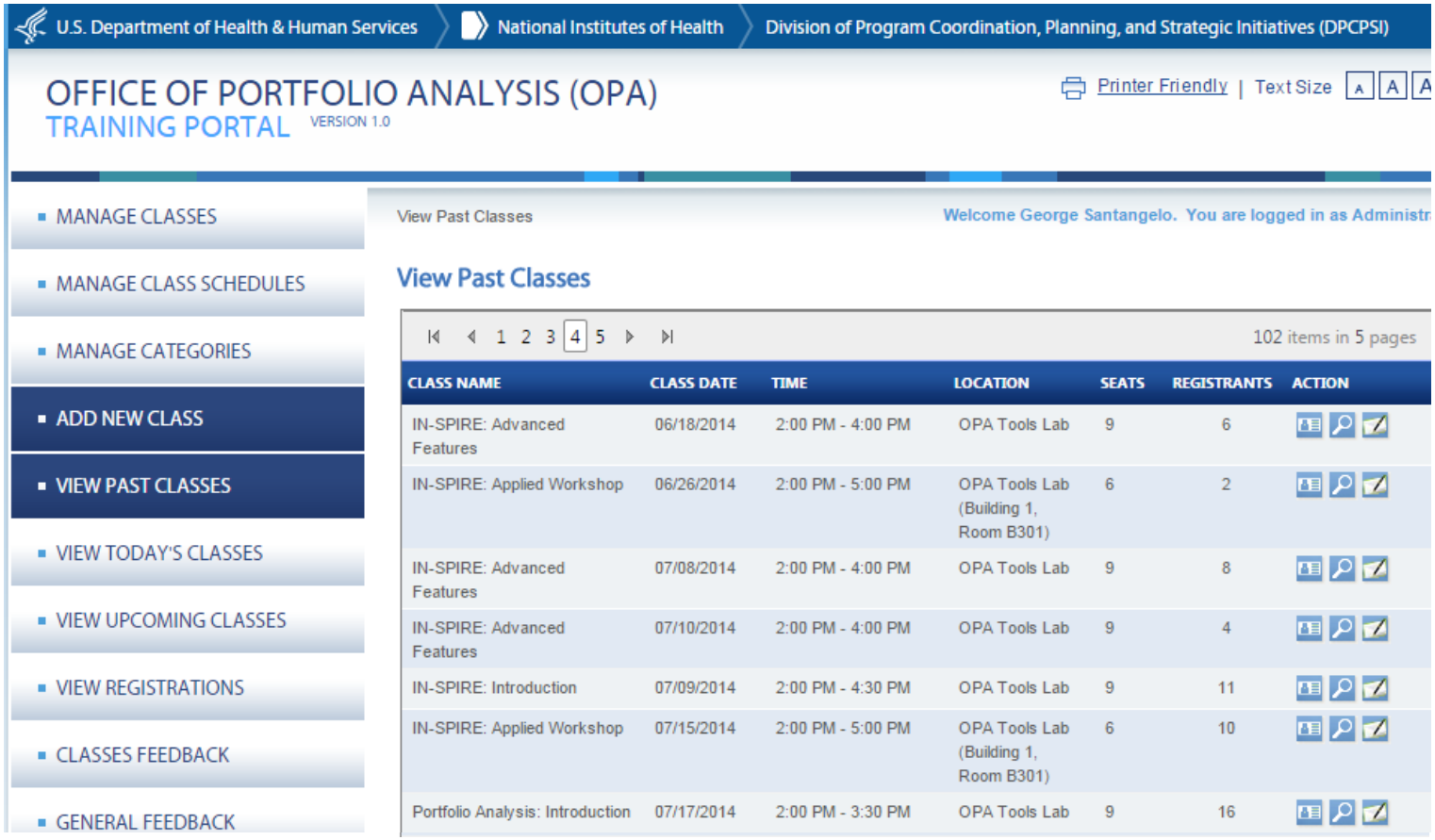
OPA Training

Current classes

Portfolio Analysis: Introduction (PA101)
IN-SPIRE: Introduction
IN-SPIRE: Advanced Features
IN-SPIRE: Applied Workshop

New classes to commence in late Spring 2015

Network Analysis
Bibliometrics



NIH National Institutes of Health
Office of Portfolio Analysis

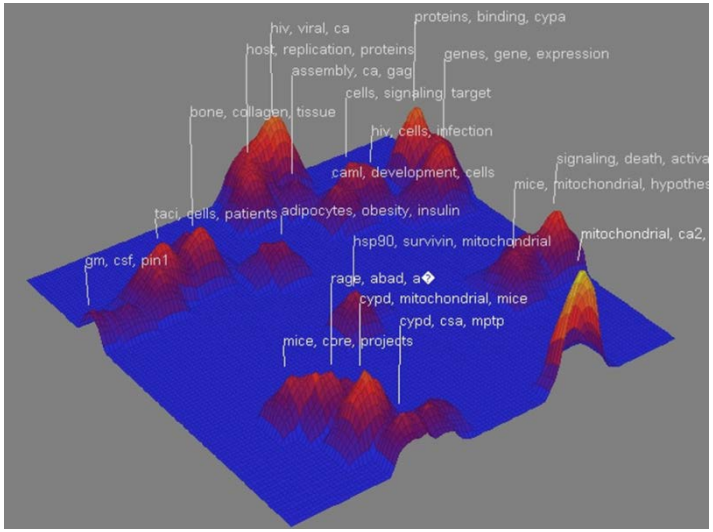
Developing a science of portfolio analysis

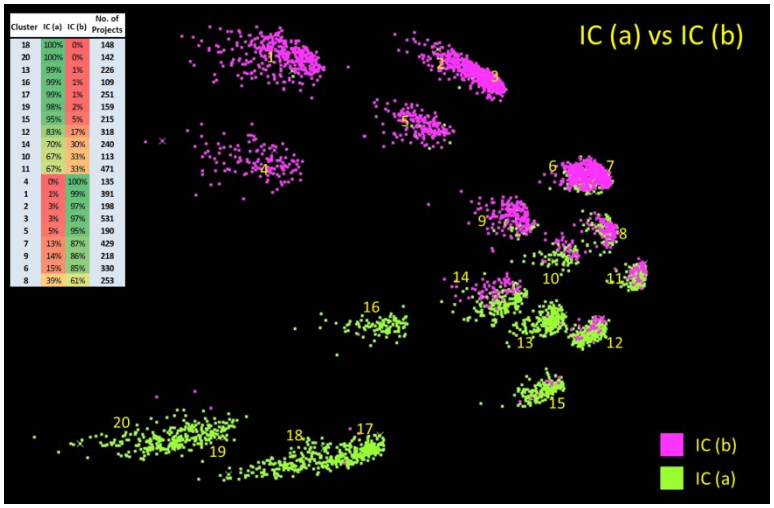
- Use existing data-driven approaches to characterize research investments and the resulting impact
- Develop and deliver effective approaches and methodologies
  - Tools in development:

Functionality	Tool
Content analysis	**IN-SPIRE *et al.***
Efficient disambiguation	***iClean***
Effective bibliometrics	***iCite***
Map translational science	***iTrans***
Track patent, licensing, start-up activity	***iTech***

Content analysis: OPA training and development

1) Visualization/interactive discovery
   - Document clustering with IN-SPIRE
2) Large-scale document clustering and analysis
   - Development of methodology and software





Cluster	IC (a)	IC (b)	No. of Projects
18	100%	0%	148
20	100%	0%	142
13	99%	1%	226
16	99%	1%	109
17	99%	1%	251
19	98%	2%	159
15	95%	5%	215
12	83%	17%	318
14	70%	30%	240
10	67%	33%	113
11	67%	33%	471
4	0%	100%	135
1	1%	99%	391
2	3%	97%	198
3	3%	97%	531
5	5%	95%	190
7	13%	87%	429
9	14%	86%	218
6	15%	85%	330
8	39%	61%	253

Developing a science of portfolio analysis

- Use existing data-driven approaches to characterize research investments and the resulting impact
- Develop and deliver effective approaches and methodologies
  - Tools in development:

Functionality	Tool
Content analysis	IN-SPIRE *et al.*
Efficient disambiguation	*iClean*
Effective bibliometrics	*iCite*
Map translational science	*iTrans*
Track patent, licensing, start-up activity	*iTech*

iCite
Web-based use of the Relative Citation Ratio (RCR) metric

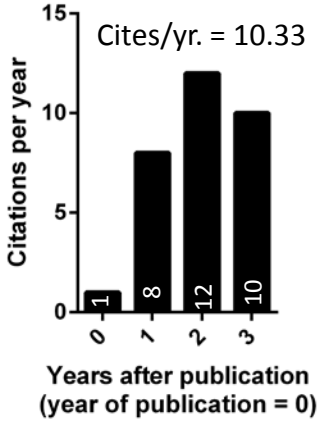
- **Limitations of commonly used bibliometrics in measuring/comparing the value of a publication or group of publications:**
  - ✓ *Publication Counts*: field-dependent, use-independent
  - ✓ *Impact Factor*: journal-level metric
  - ✓ *Citation Rates*: field- and journal-dependent
  - ✓ *h-index*: field-dependent, time-dependent

- **Relative Citation Ratio (RCR)**
  - ✓ Need: An article-level metric that is independent of field, journal, and time
  - ✓ Assumption: Citation of a publication reveals value to or influence on the citer
  - ✓ RCR normalizes citations to each publication's co-citation network

Calculating the Relative Citation Ratio

$$RCR = \frac{Article\ Citation\ Rate}{Expected\ Citation\ Rate}$$

Article Citation Rate
(denominator excludes year of publication)



- Article-level metric
- Number of cites per year
- RCR changes over time with the accrual of new citations
- Scalable to large portfolios containing tens of thousands of articles

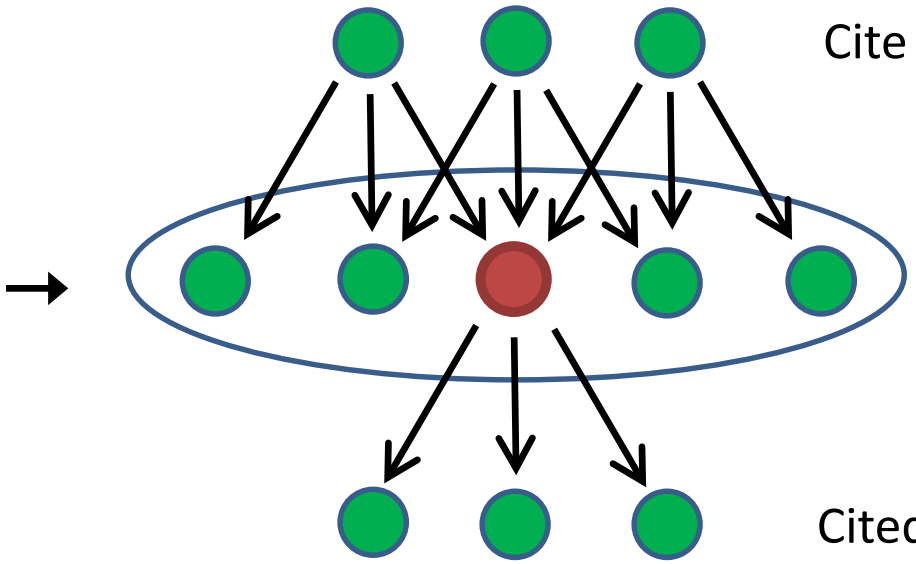
Calculating the Relative Citation Ratio

Expected Citation Rate Part 1: Calculate the Field Citation Rate

= Article of interest

Cite the article of interest

Field Citation Rate
= Average the ***journal citation rates*** for these co-cited articles (includes article of interest)



Co-cited with article of interest:
the co-citation network

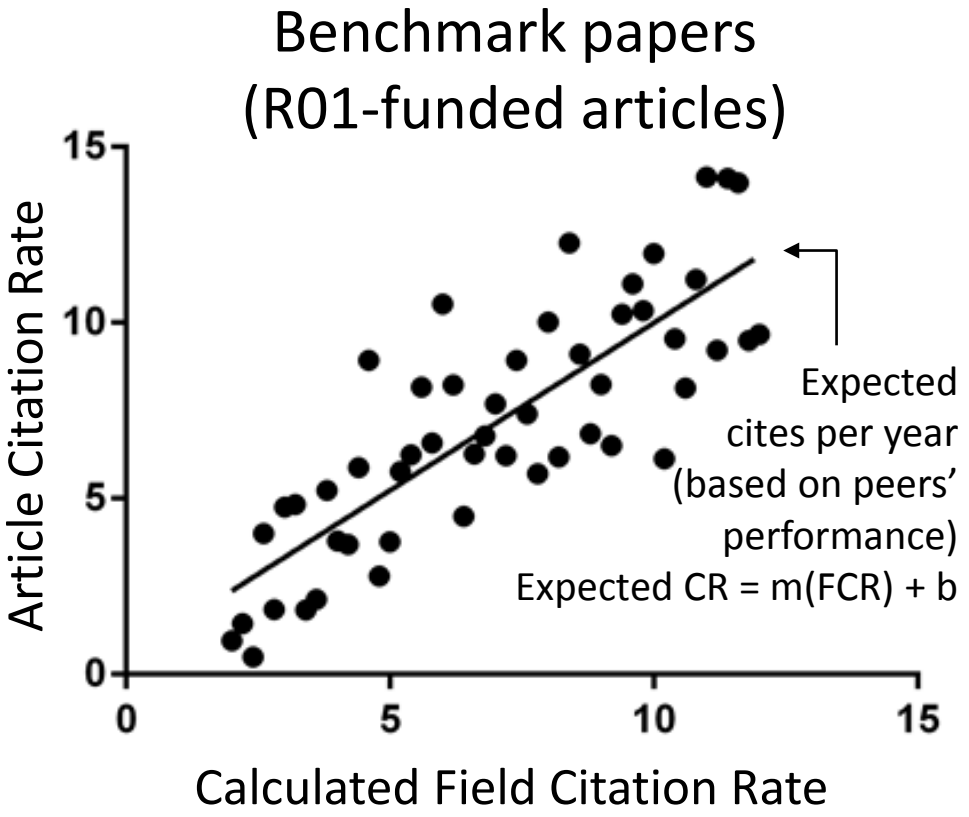
Cited by the article of interest

Calculating the Relative Citation Ratio

Expected Citation Rate Part 2: Benchmark the Field Citation Rate

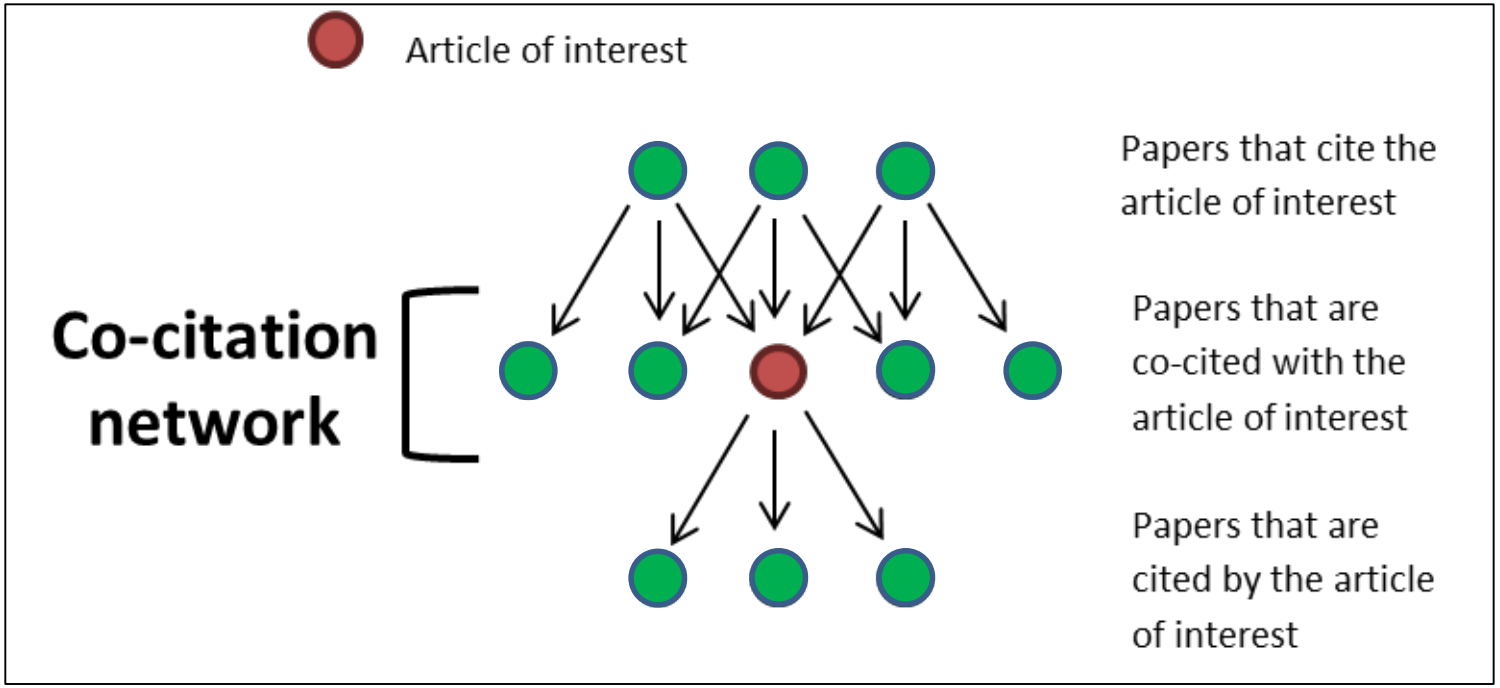
- Use this linear regression equation to transform the Field Citation Rate
- This benchmarks expectations to NIH-funded papers for any Field (avg. = 1.0)

$$RCR = \frac{Article\ Citation\ Rate}{Expected\ Citation\ Rate}$$



RCR

How is the paper of interest cited relative to expectations based on its co-citation network?



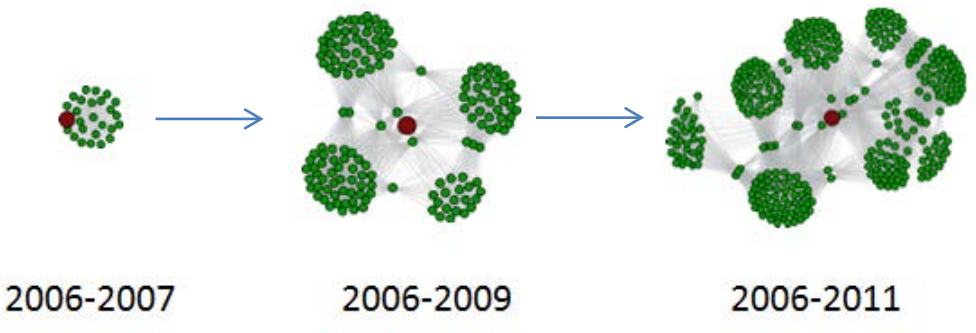
0 = never cited
1 = average
2 = twice the average
>20 = exceptionally highly cited

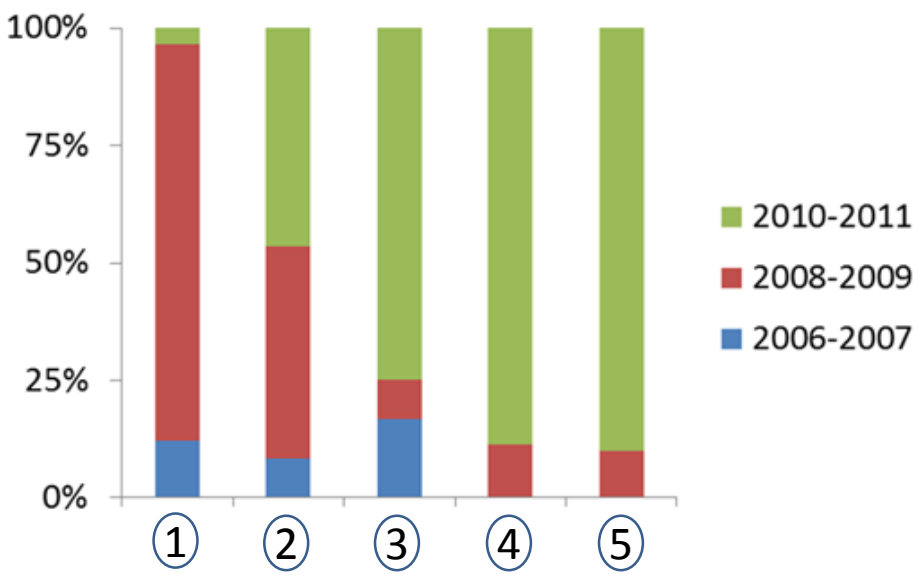
Thomson Reuters Science Citation Index Expanded, 2002-2012

RCR denominator: sample co-citation networks

2006 paper that identified new conotoxin-like peptides of possible clinical utility

Growth of the co-citation network

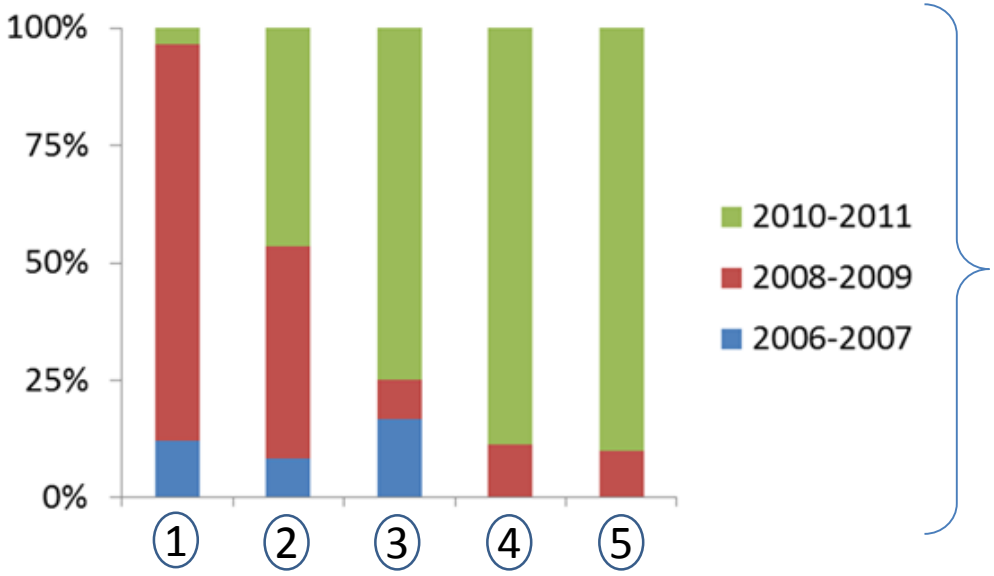




New areas of science appear in the network

1. α-conotoxin mechanisms of action
2. structure and evolution of conotoxins
3. cyclotide biochemistry
4. conotoxin phylogenetics
5. identification and synthesis of lantibiotics

Each article has an RCR denominator that represents its specific field



1. α-conotoxin mechanisms of action
2. structure and evolution of conotoxins
3. cyclotide biochemistry
4. conotoxin phylogenetics
5. identification and synthesis of lantibiotics

- Other denominators–e.g. the Thomson Reuters (TR) Ratio—use the "multidisciplinary" category for all papers in Science, Nature, PLoS ONE, etc.
- Multidisciplinary categories generate the same denominator for very different areas of science
  - structural biology, ecology, neuroscience, climate change, medicine, education, etc.

RCR denominator: sample co-citation networks

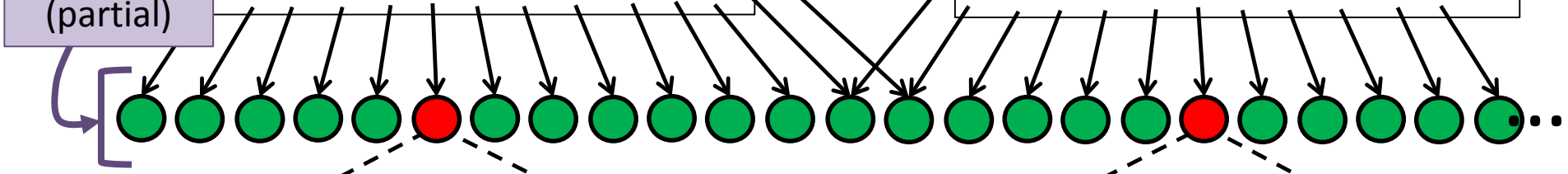
2012 publications citing a Nature 2011 paper on Hepatitis C pathology:

#1 in CELL HOST & MICROBE
#2 in CELL REPORTS
#3 in COLD SPRING HARBOR PERSPECTIVES IN BIOLOGY
#4 in CURRENT OPINION IN STRUCTURAL BIOLOGY
#5 in CURRENT OPINION IN VIROLOGY
#6 in CURRENT OPINION IN VIROLOGY
#7 in CURRENT OPINION IN VIROLOGY
#8 in EMBO JOURNAL
#9 in IMMUNITY
#10 in JOURNAL OF APPLIED CRYSTALLOGRAPHY
#11 in JOURNAL OF BIOLOGICAL CHEMISTRY
#12 in JOURNAL OF VIROLOGY
#13 in MOLECULAR THERAPY
14 in NATURE
#15 in PLOS ONE
#16 in PLOS PATHOGENS
#17 in PLOS PATHOGENS
#18 in PROCEEDINGS NATL ACAD OF SCI USA
#19 in PROCEEDINGS NATL ACAD OF SCI USA
#20 in STRUCTURE

2012 publications citing a PLoS ONE 2011 paper on the immune response to Hepatitis C:

#1 in AMERICAN JOURNAL OF PATHOLOGY
#2 in CNS NEUROSCIENCE & THERAPEUTICS
#3 in EXPERIMENTAL AND MOLECULAR PATHOLOGY
#4 in HEPATOLOGY INTERNATIONAL
#5 in JOURNAL OF TRANSLATIONAL MEDICINE
#6 in JOURNAL OF TRAUMA AND ACUTE CARE
#7 in SURGERY
#8 in JOURNAL OF VIROLOGY
#9 in LEUKEMIA
#10 in PLOS ONE
#11 in PLOS ONE
#12 in PLOS ONE
#13 in PLOS ONE
#14 in PROCEEDINGS NATL ACAD OF SCI USA
#15 in SHOCK
#16 in SHOCK
#17 in TOXICOLOGY AND APPLIED

Co-citation Network (partial)



Nature 2011 Hepatitis C paper

RCR = 3.0

PLoS ONE 2011 Hepatitis C paper

RCR = 4.2

RCR validation study using subject matter experts

Is there a correlation between RCR and expert assessment of value/quality/impact?

- 684 papers published in 2009 supported by R01s and representing a range of RCRs
- Papers assigned based on content of the reviewer's published work
- 537 IRP Investigators were recruited with approval of their SDs
- 3-5 PIs received the same set of 5 publications
  - ✓ Pubs and responses via a secure intranet site

- 6 criteria, each with a 5 point response scale

- Responses
  - ✓ 44.3% Investigators responded
  - ✓ 1028 responses to 561 papers, 290 with ≥2 responses/paper

Relative Citation Ratio – Review Criteria

- **IMPORTANCE**
  Rate whether the question being addressed is important to answer.
  (1 = Not Important, 2 = Slightly Important, 3 = Important, 4 = Highly Important, 5 = Extremely Important)

- **METHODS**
  Rate whether you agree that the methods are appropriate and the scope of the experiments adequate.
  (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree)

- **ROBUSTNESS**
  Rate how robust the study is based on the strength of the evidence presented.
  (1 = Not Robust, 2 = Slightly Robust, 3 = Moderately Robust, 4 = Highly Robust, 5 = Extremely Robust)

- **HUMAN HEALTH RELEVANCE**
  Rate the likelihood that the results could ultimately have a substantial positive impact on human health outcomes.
  (1 = Very unlikely, 2 = Unlikely, 3 = Foreseeable but uncertain, 4 = Probable, 5 = Almost Certainly)

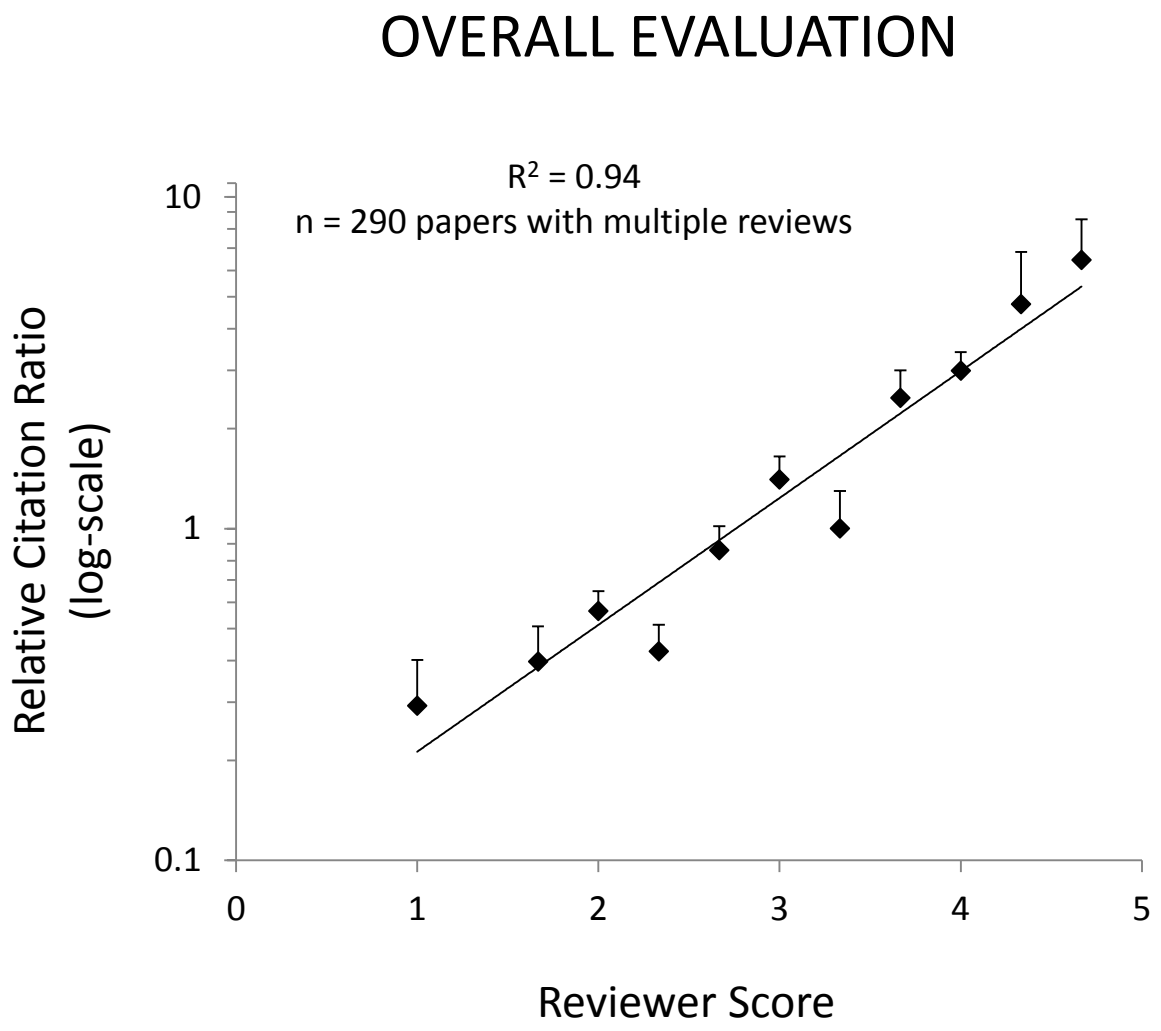
- **LIKELY IMPACT**
  Rate the impact that the research is likely to have or has already had.
  (1 = Minimal Impact, 2 = Some Impact, 3 = Moderate Impact, 4 = High Impact, 5 = Extremely High Impact)

- **OVERALL EVALUATION**
  Provide your overall evaluation of the value and impact of this publication.
  (1 = minimal or no value, 2 = Moderate value, 3 = Average value, 4 = High value, 5= Extremely high value)

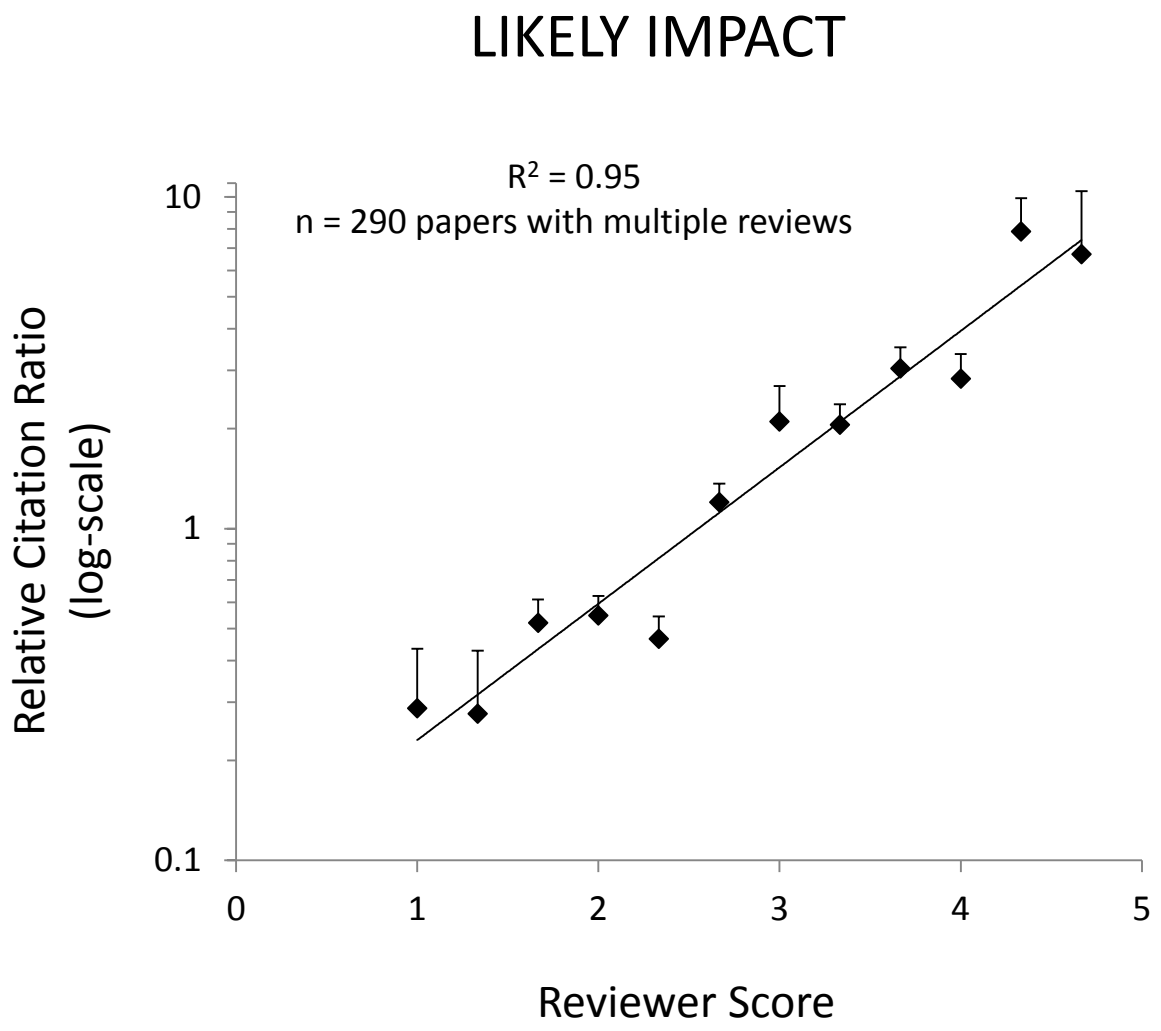
Fields of science represented in the RCR validation study

Cell Biology
Genetics and Genomics
Neuroscience
Chromosome Biology
Developmental Biology
Immunology
Molecular Biology and Biochemistry
Cancer Biology
Stem Cell Research
Molecular Pharmacology
Systems Biology
Epidemiology
Clinical Research
Virology
Computational Biology
Biomedical Engineering & Biophysics
Chemical Biology
Microbiology and Infectious Diseases
Structural Biology
Health Disparities
Social and Behavioral Science

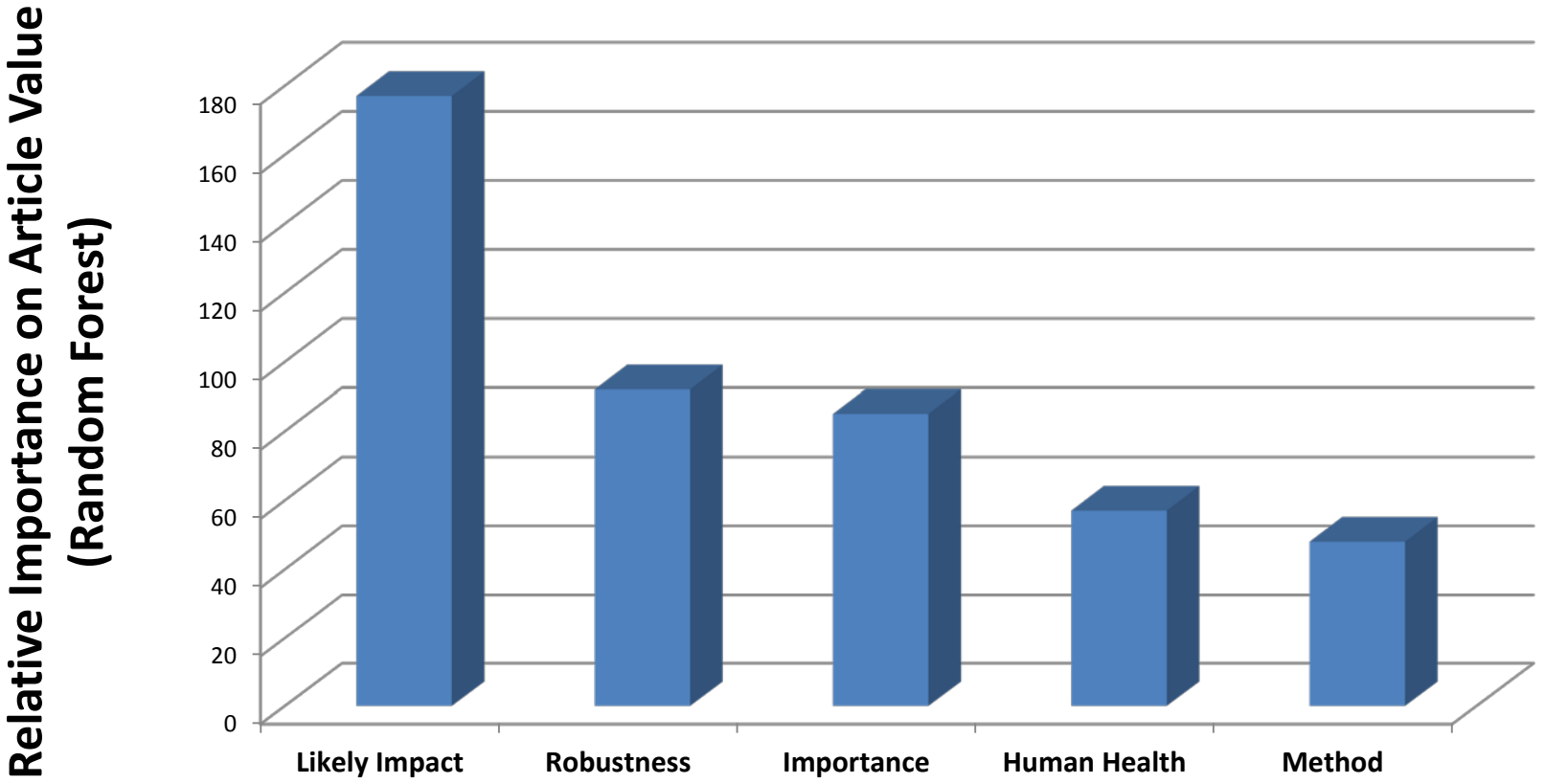
Relative Citation Ratio (RCR) validation study



Relative Citation Ratio (RCR) validation study

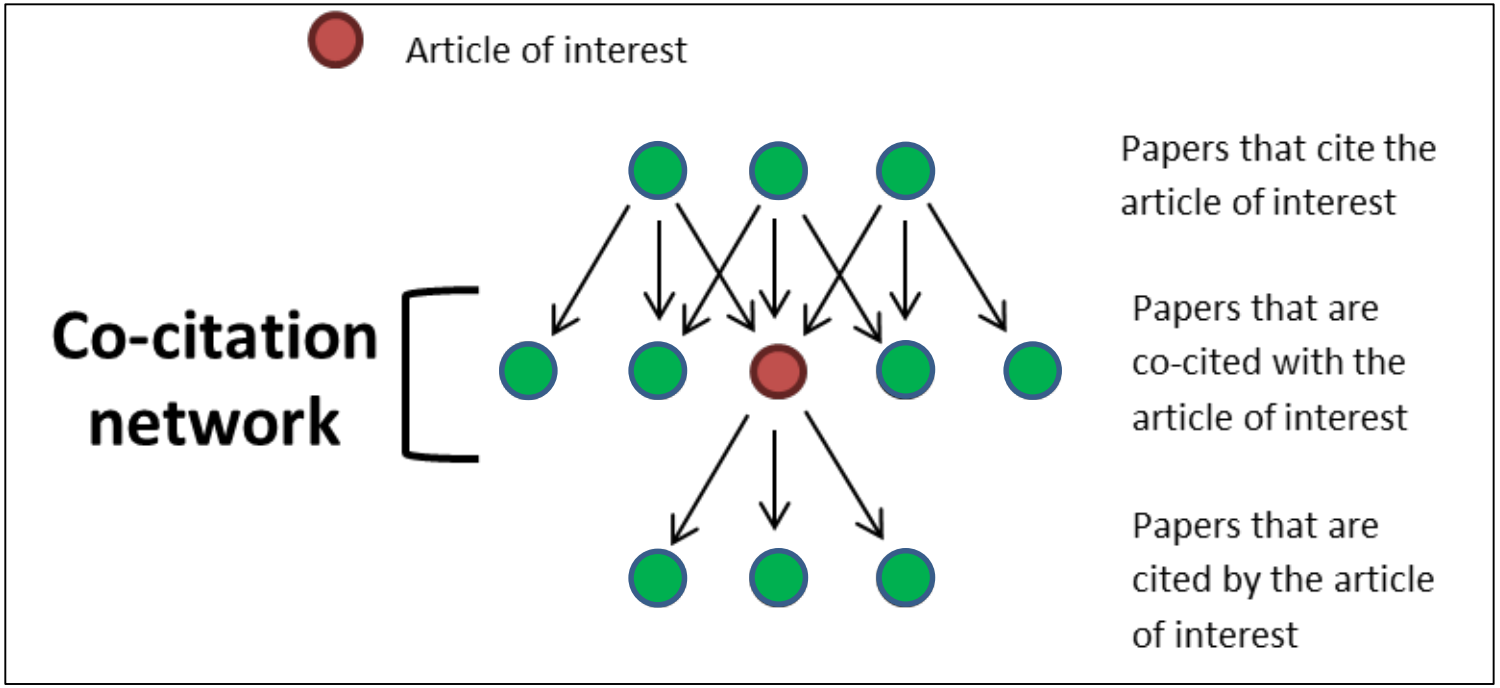


How do reviewers weigh these factors when deciding on a paper's overall value?



RCR

How is the paper of interest cited relative to expectations based on its co-citation network?



0 = never cited
1 = average
2 = twice the average
>20 = exceptionally highly cited

Thomson Reuters Science Citation Index Expanded, 2002-2012

RCR vs. Thomson Reuters (TR) Ratio
Identifying potentially problematic outliers due to low Expected CPY

Analysis of 35,837 articles: all NIH (R01) pubs in 2009
Denominator = Expected CPY

TR Ratio

Journal categories used to calculate denominator

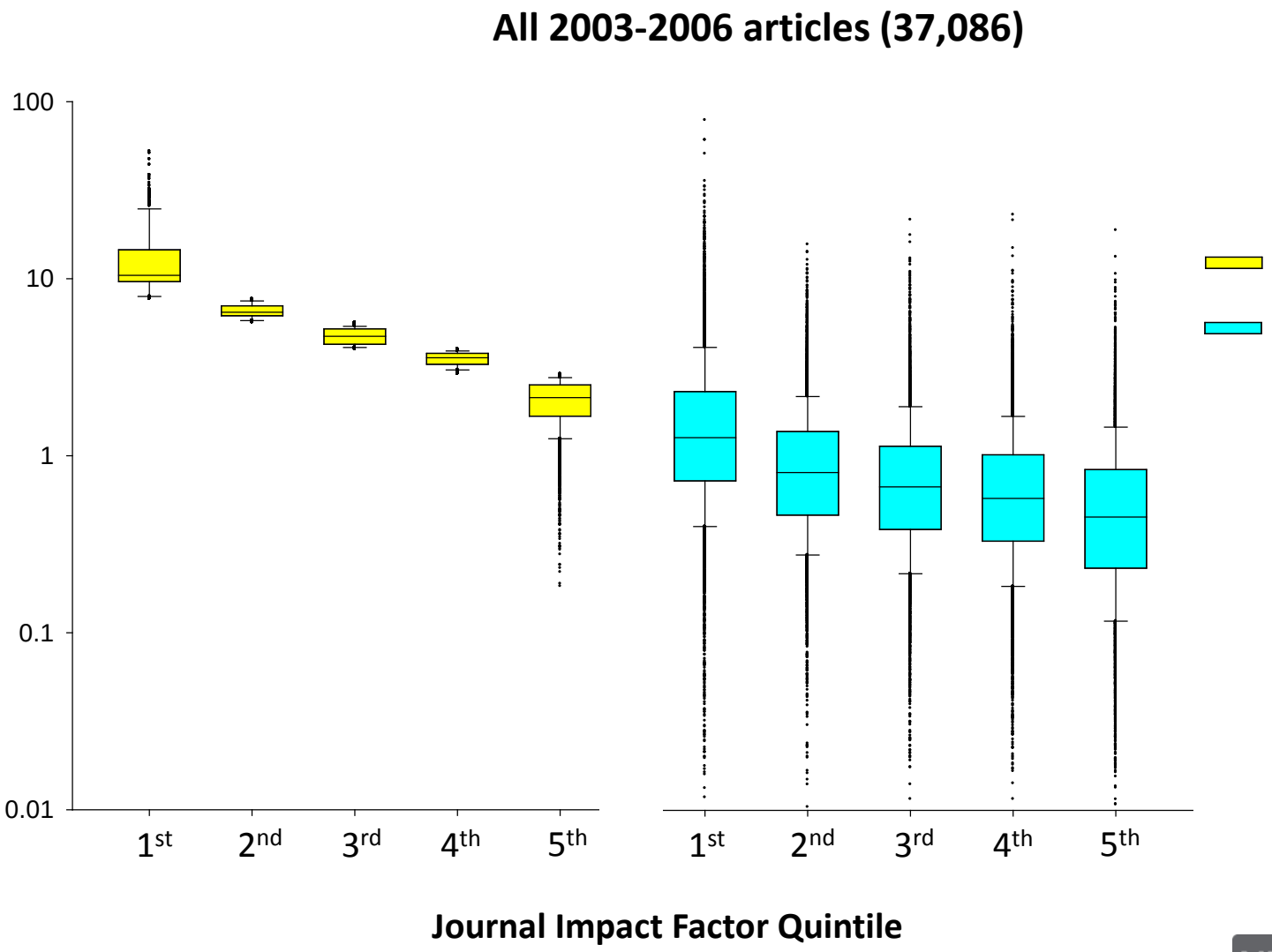
- **565** pubs with Expected CPY < 2.0
- TR Ratio ranged from 0 to **23.4**
- **Average TR Ratio = 1.6**

RCR

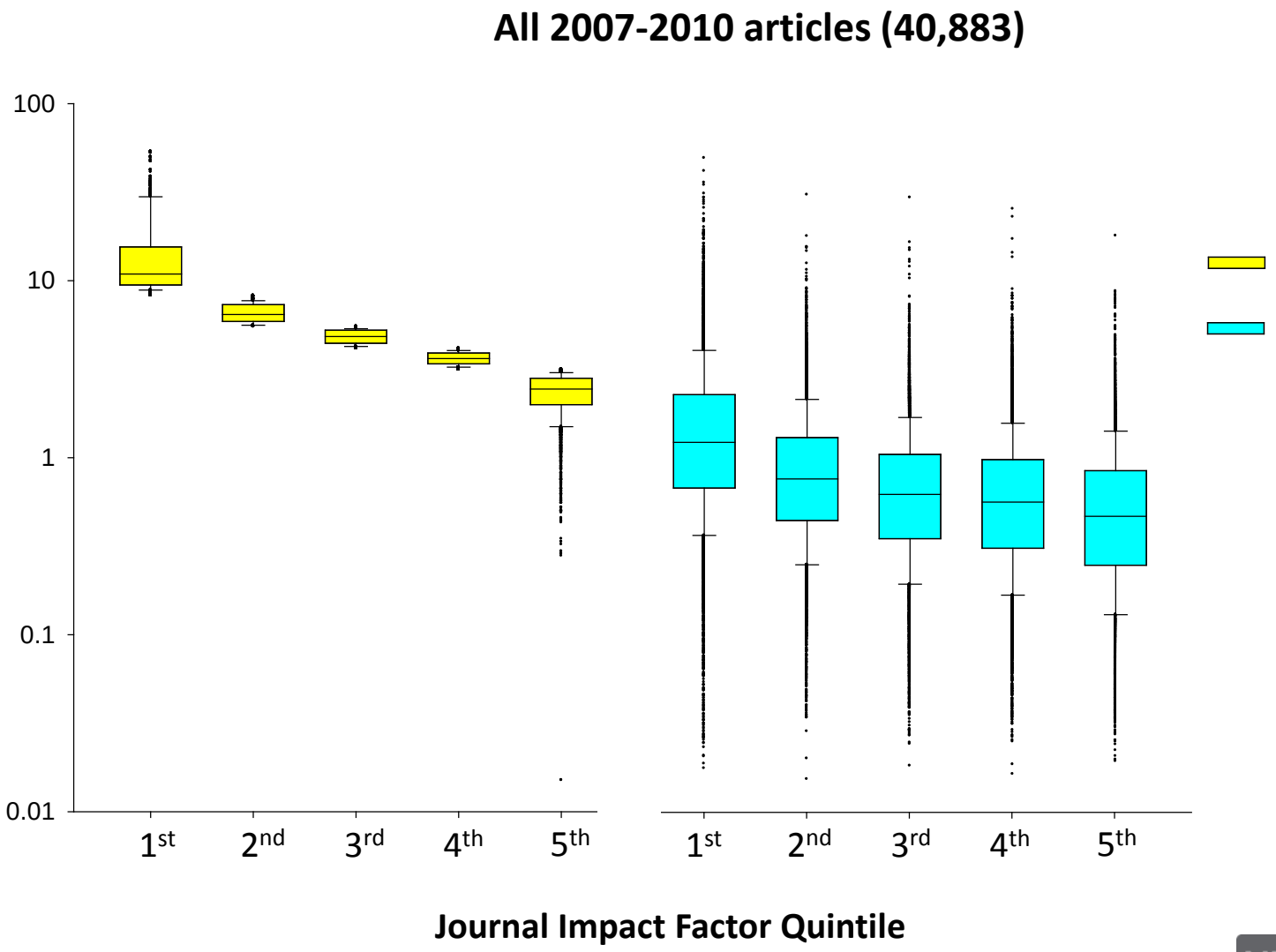
Co-citation networks used to calculate denominator

- **147** pubs with Expected CPY < 2.0
- RCR ranged from 0 to **3.7**
- **Average RCR = 0.24**

Publications of NIH investigators with continuous funding from FY2003 to FY2010



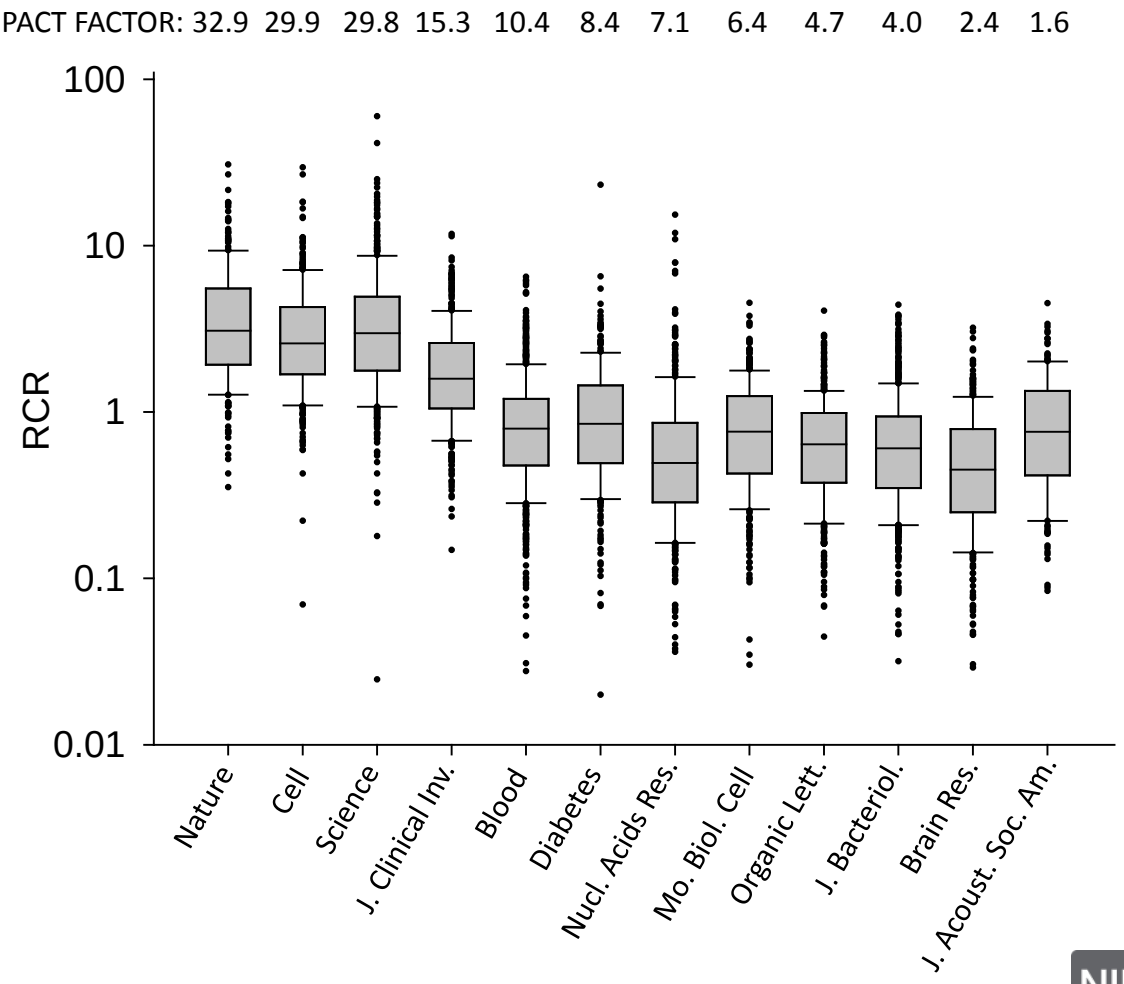
Publications of NIH investigators with continuous funding from FY2003 to FY2010



NIH National Institutes of Health
Office of Portfolio Analysis

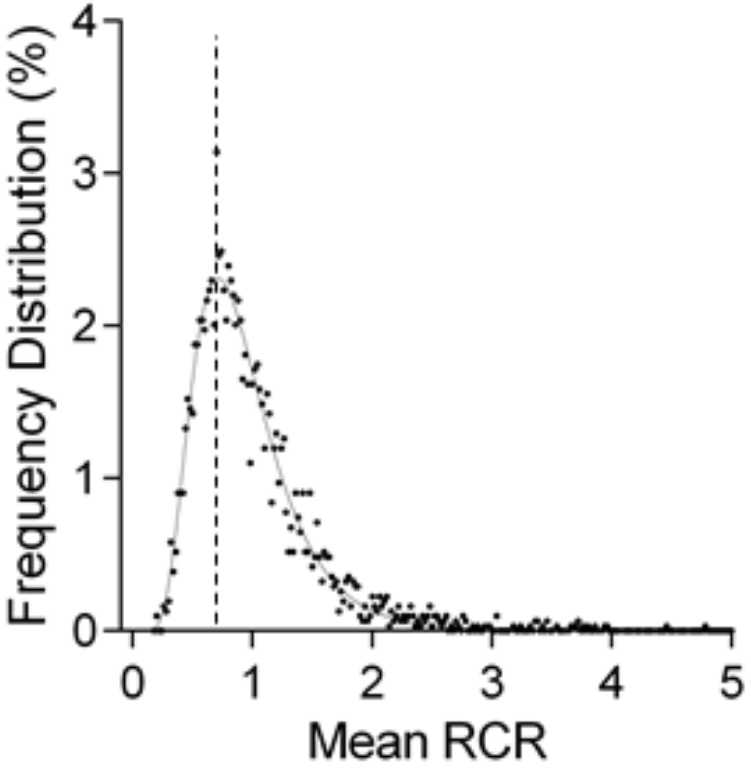
Publications of NIH investigators with continuous funding from FY2003 to FY2010

2003-2010 articles: Selected journal IFs from 32.9 to 1.6



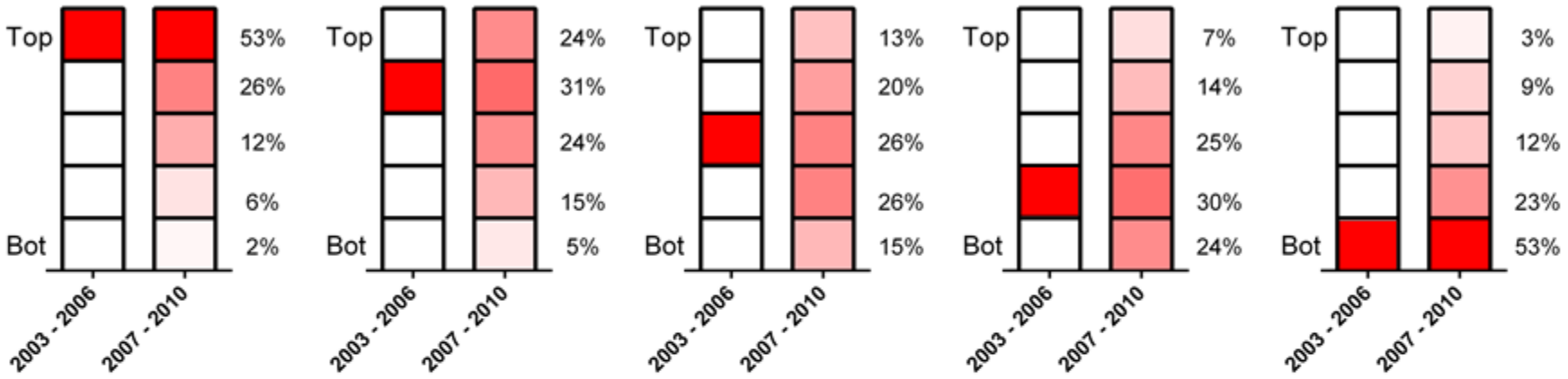
Does the RCR of NIH PIs fluctuate over time?

PIs with eight years of continuous R01 funding (FY2003 to FY2010)

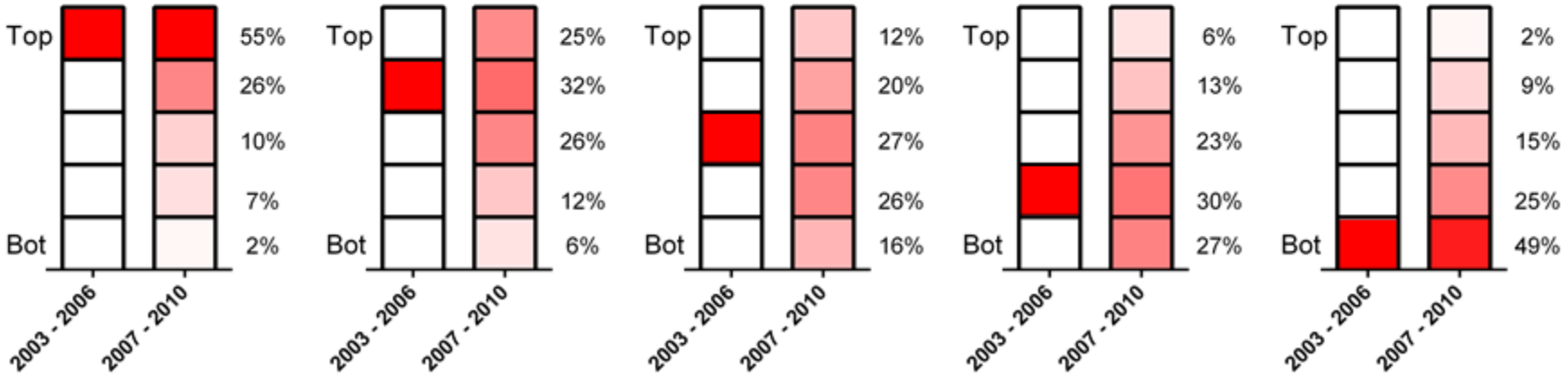


Scientific mobility of PIs (3089 total) with continuous R01 funding through two consecutive 4-year periods

Ranked by RCR (in quintiles)

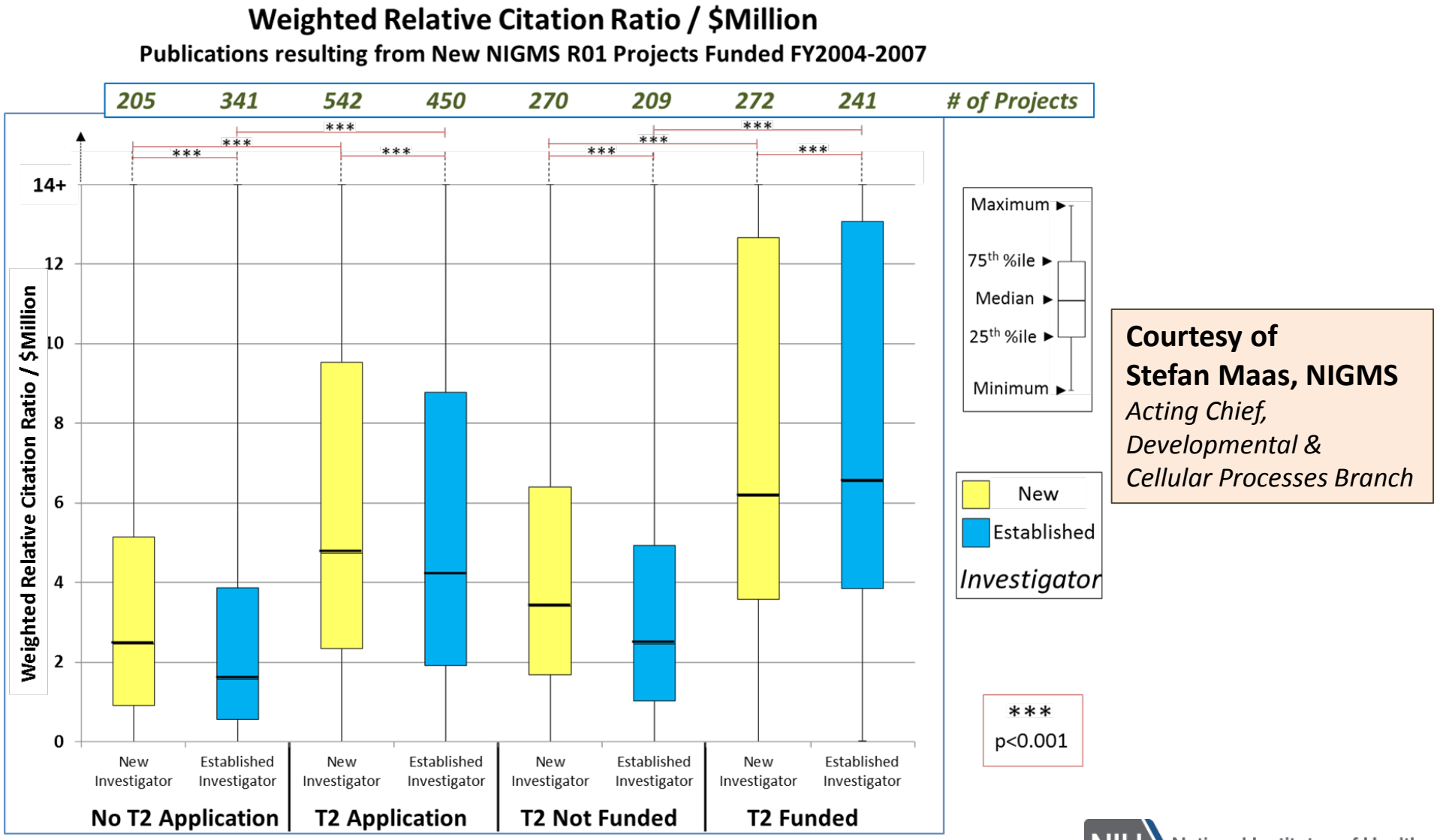


Ranked by weighted RCR (RCR times total # of pubs; in quintiles)



NIH National Institutes of Health
Office of Portfolio Analysis

Use of RCR at NIGMS: R01 projects submitted for competitive renewal (T2) have higher weighted RCRs



Courtesy of
Stefan Maas, NIGMS
Acting Chief,
Developmental &
Cellular Processes Branch

NIH National Institutes of Health
Office of Portfolio Analysis

Uses of RCR to compare impact

- Validated use:
  - ✓ compare individual publications within a network
- Additional uses:
  - ✓ compare output of cohorts (portfolios, programs, mechanisms…)
  - ✓ flagging low impact for inspection
  - ✓ to follow trends
- Possible misuses of RCR or any bibliometric assessment:
  - x determining importance of the endeavor
  - x predicting long-term impact of the work
  - x evaluating effectiveness of “downstream” (e.g. patentable) applications

Developing a science of portfolio analysis

- Use existing data-driven approaches to characterize research investments and the resulting impact
- Develop and deliver effective approaches and methodologies
  - Tools in development:

Functionality	Tool
Content analysis	**IN-SPIRE *et al.***
Efficient disambiguation	***iClean***
Effective bibliometrics	***iCite***
Map translational science	***iTrans***
Track patent, licensing, start-up activity	***iTech***

iTrans: Using Medical Subject Heading (MeSH) terms to track bench to bedside trends in scientific knowledge

Griffin Weber
Identifying translation science within the triangle of biomedicine
J Trans Med 11:126 (2013)

Human

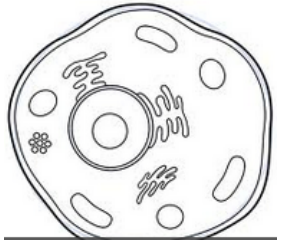
Clinical trials

Mutations in genes of interest discovered in human patients

Candidate pharmacological modulators developed

Pharmaceuticals tested in animal research

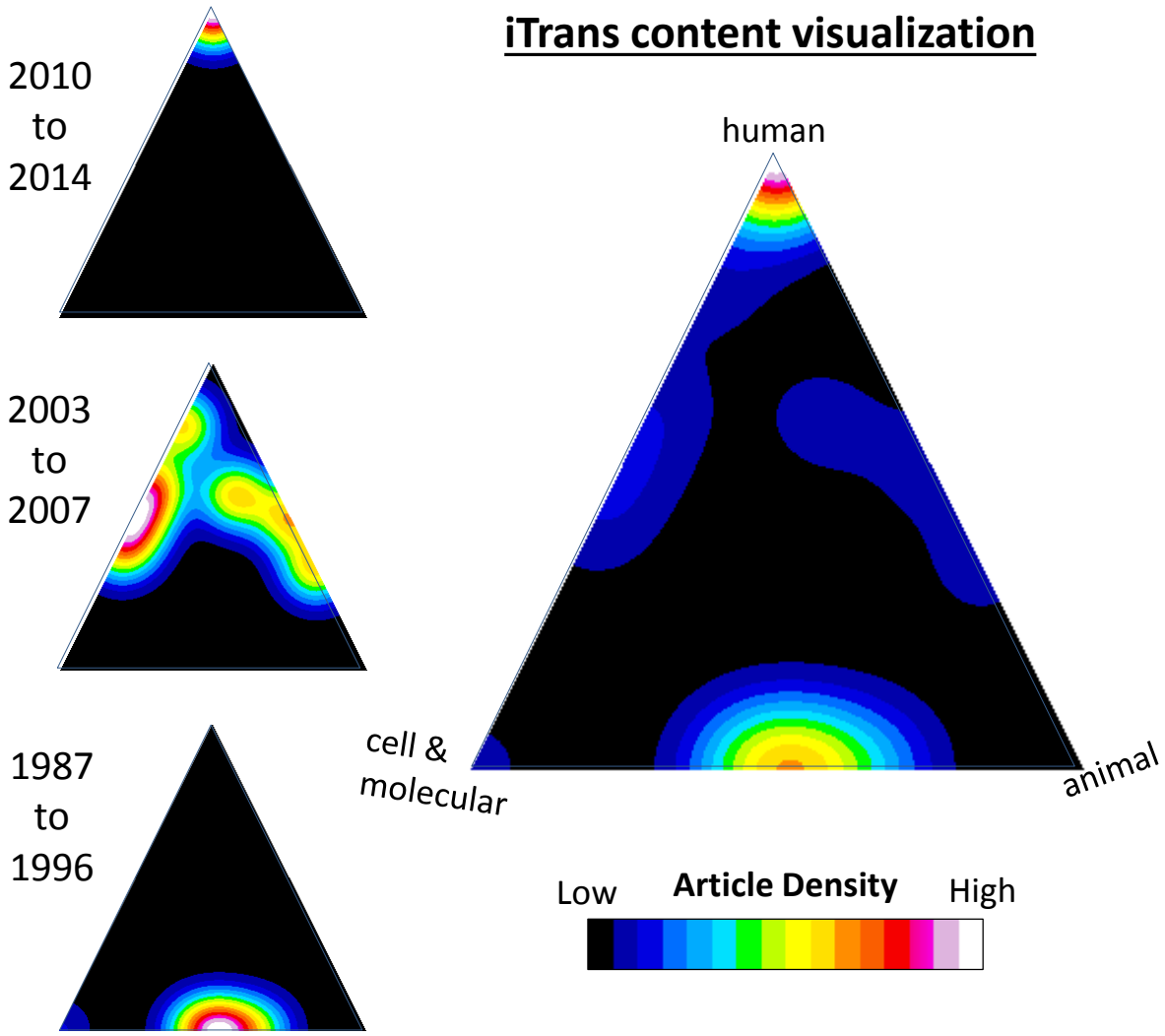
Biological mechanism of disease elucidated in model systems



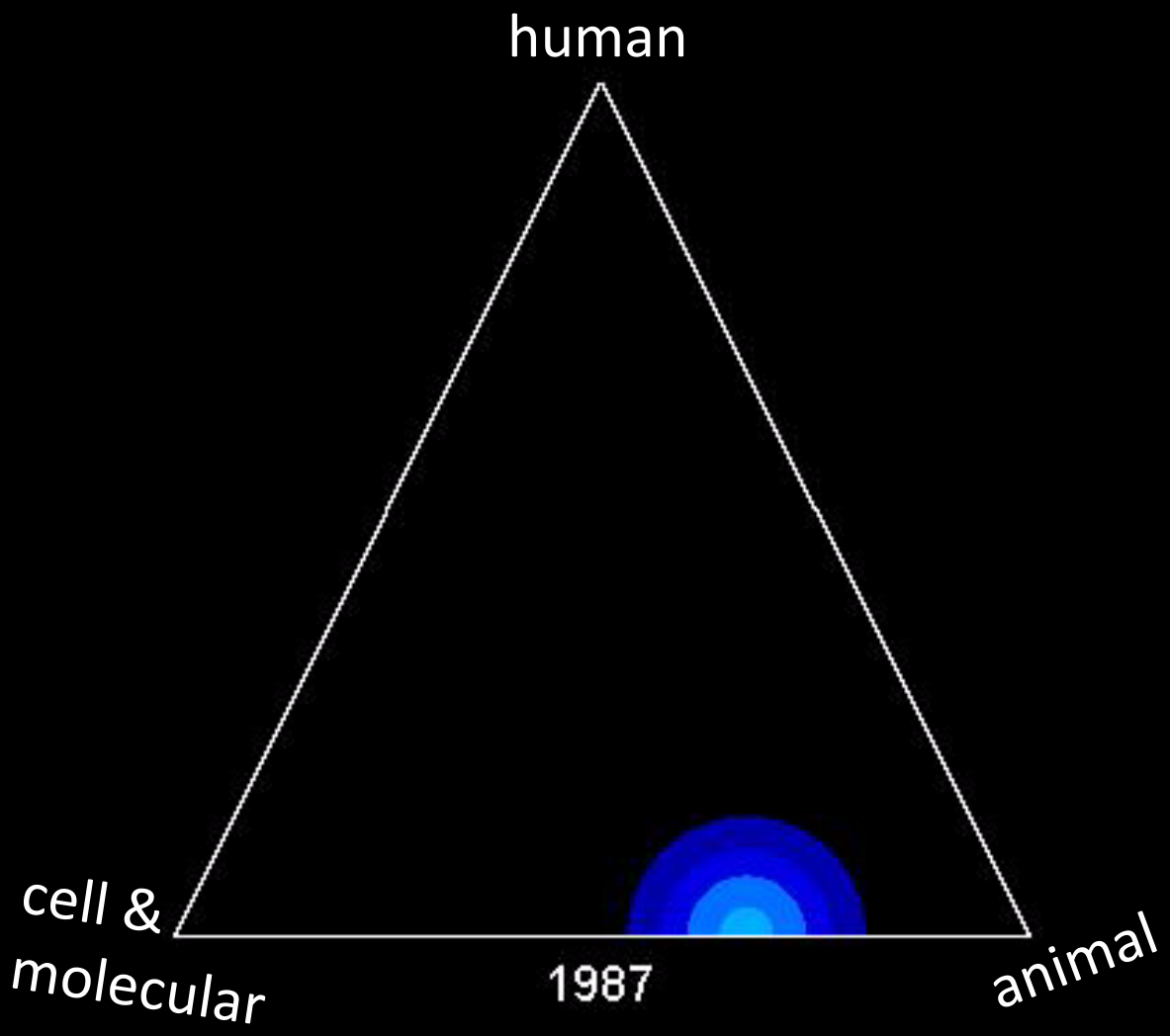
Cellular/ Molecular

Animal

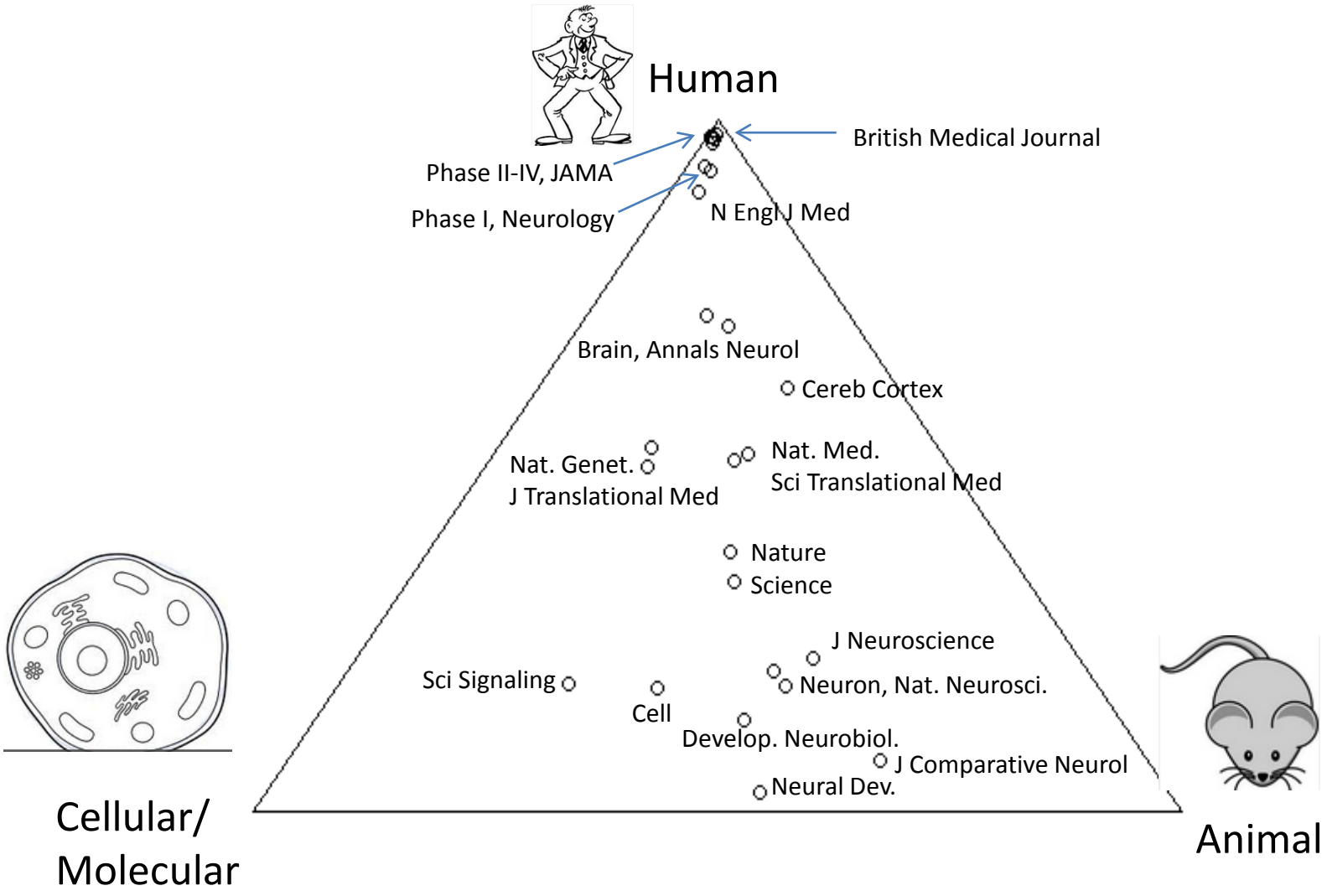
iTrans tracks translational development of cancer immunotherapeutic agents



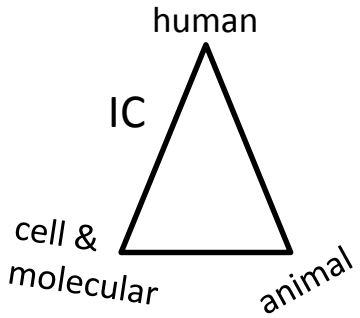
NIH National Institutes of Health
Office of Portfolio Analysis


human
cell &
molecular
1987
animal

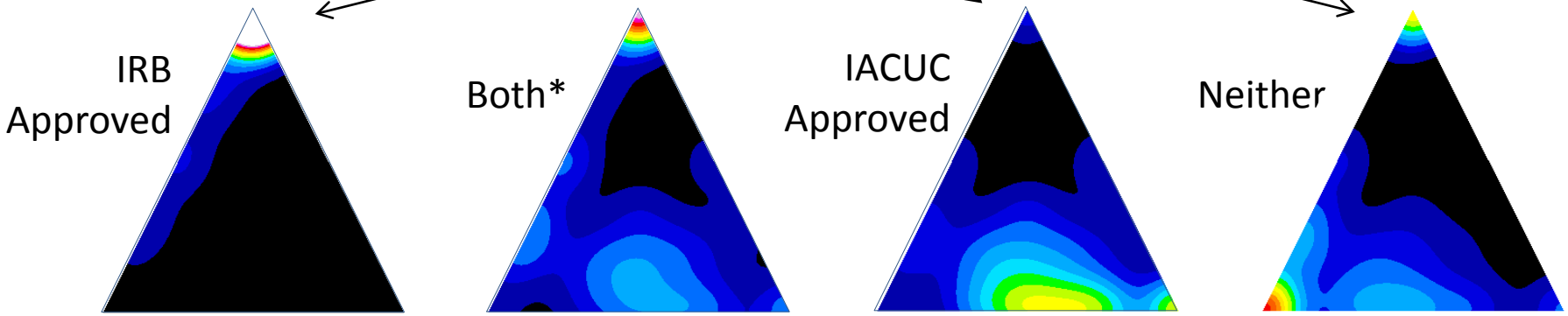
iTrans: Using Medical Subject Heading (MeSH) terms to track bench to bedside trends in scientific knowledge



iTrans: Using MeSH terms to characterize the NIH publication landscape

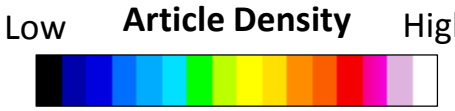


All NIH publications funded by active research grants in 2010



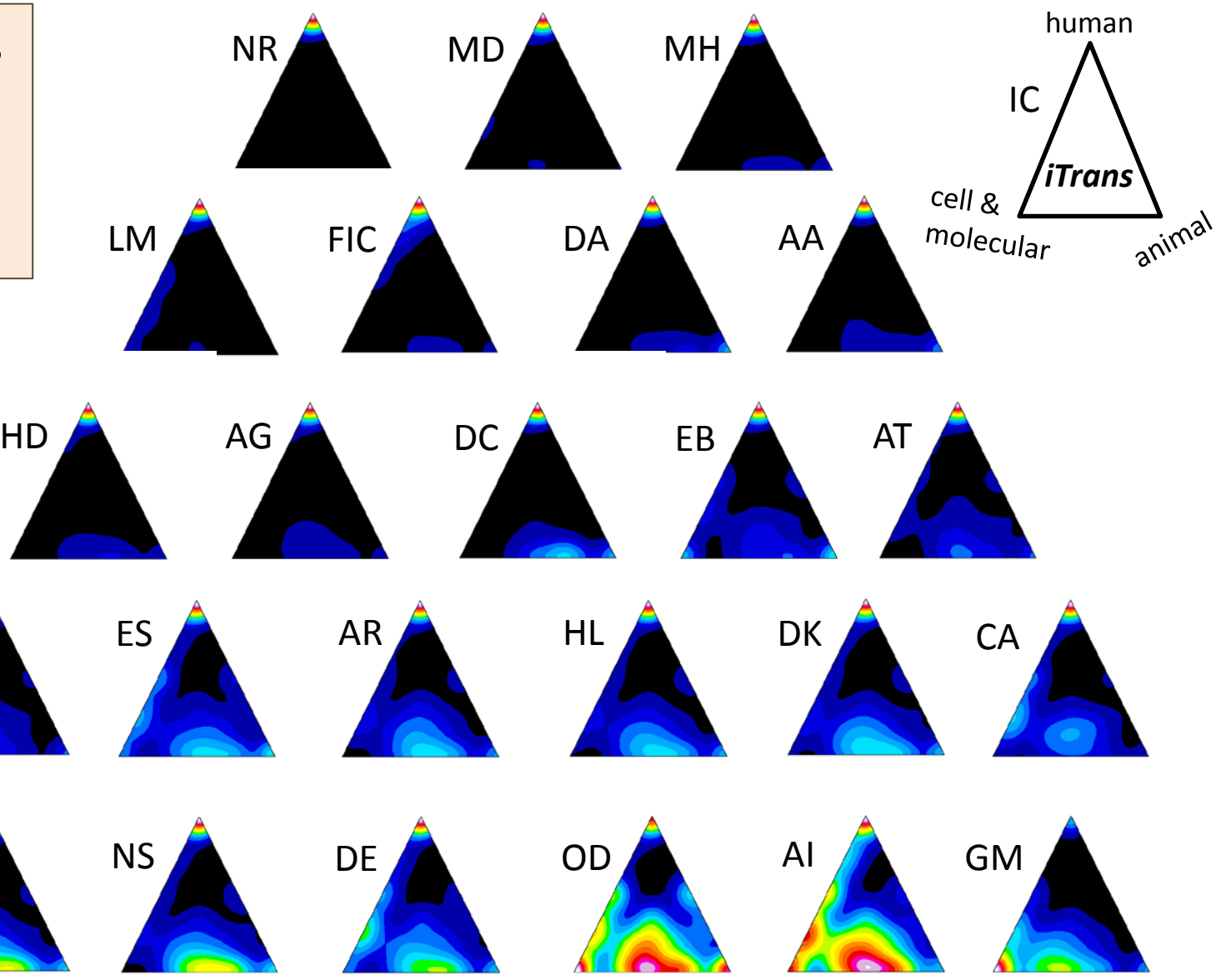
*Both = both IRB- and IACUC-approved

*Neither = not requiring either IRB or IACUC approval



NIH National Institutes of Health
Office of Portfolio Analysis

All publications funded by active research grants in 2012



AA	NIAAA
AG	NIA
AI	NIAID
AR	NIAMS
AT	NCCIH
CA	NCI
DA	NIDA
DC	NIDCD
DE	NIDCR
DK	NIDDK
EB	NIBIB
ES	NIEHS
EY	NEI
GM	NIGMS
HD	NICHD
HG	NHGRI
HL	NHLBI
LM	NLM
MD	NIMHD
MH	NIMH
NR	NINR
NS	NINDS
FIC	Fogarty Int'l Ctr

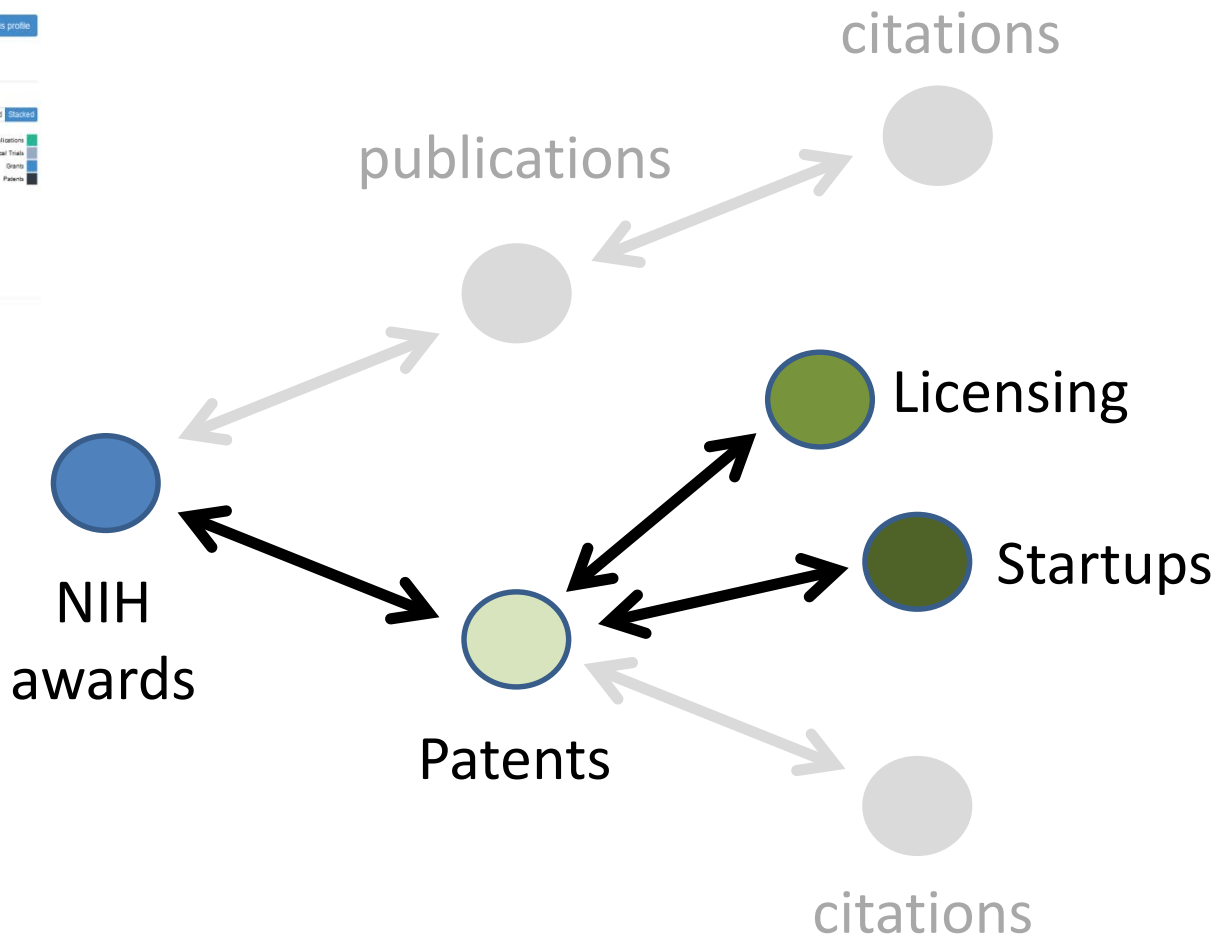
National Institutes of Health
Office of Portfolio Analysis

Developing a science of portfolio analysis

- Use existing data-driven approaches to characterize research investments and the resulting impact
- Develop and deliver effective approaches and methodologies
  - Tools in development:

Functionality	Tool
Content analysis	**IN-SPIRE *et al.***
Efficient disambiguation	***iClean***
Effective bibliometrics	***iCite***
Map translational science	***iTrans***
Track patent, licensing, start-up activity	***iTech***

iTech: Patents, licensing and start-ups
Connecting the dots (bidirectionally)

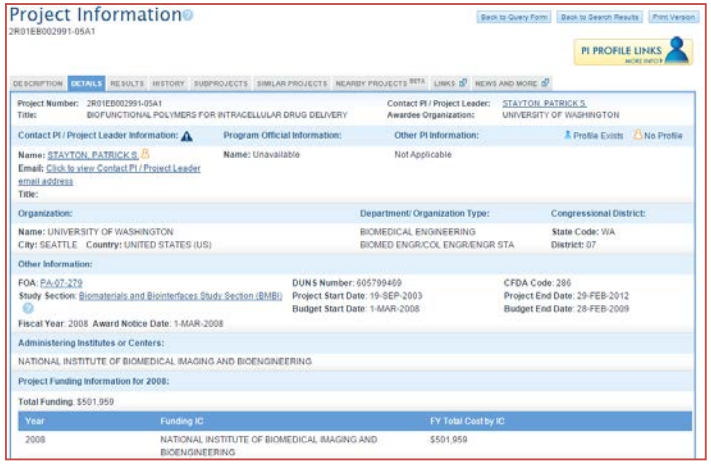


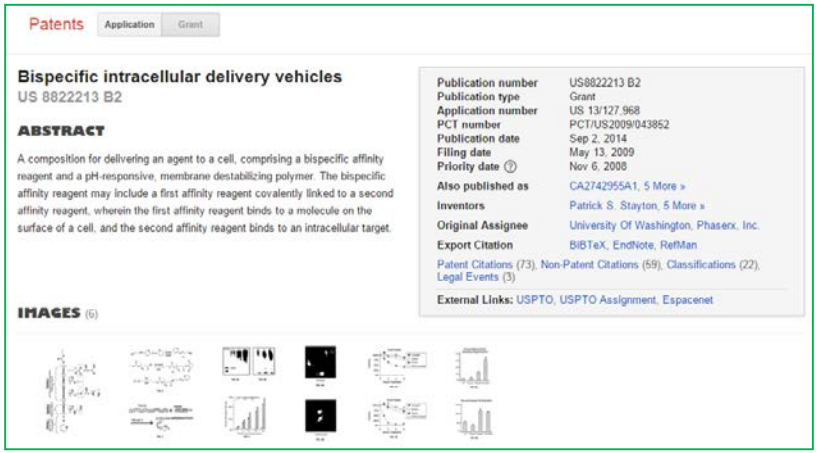
Case 1: Linking Patent with Grant - Dr. Stayton from U of Washington

Invention Profile: Prof. Patrick Stayton

- Highly productive in inventions, with total of 124 patents filed globally since 1995
- 93% funding from NIH

Linking award R01EB002991 *"Biofunctional Polymers for Intracellular Drug Delivery"* with Patent US8822213 (also filed as CA2742955A1, WO2010053596A)





Case 1: Tracking Dr. Stayton's Patent to Startup *PhaseRX*

Dr. Stayton founded PhaseRx in 2008, using his patented technology (US8822213)

PhaseRx
Unlocking the value of RNA

About | mRNA Technology | Pipeline | Partnering | News | Careers | Contact

About
Management Team
Board of Directors
Founders

Pat S. Stayton, PhD
Allan S. Hoffman, PhD
Oliver W. Press, MD, PhD
Robert W. Overell, PhD
Paul H. Johnson, PhD

Pat S. Stayton, PhD
Washington Research Foundation
Professor of Bioengineering, University of Washington

Dr. Stayton currently serves as the Washington Research Foundation Professor in the Department of Bioengineering at the University of Washington. He received his BS in Biology (summa cum laude, Chemistry minor) from Illinois State University in 1984, his PhD in Biochemistry from the University of Illinois in 1989, and was a Postdoctoral Research Associate at the Beckman Institute for Advanced Science and Technology, also at the University of Illinois.

Dr. Stayton has been elected as a Fellow of the American Institute for Medical and Biological Engineering and has been the recipient of the Clemson Award from the Society For Biomaterials and the CRS-Cygnus Recognition Award from the Controlled Release Society. He served as Chair of the Gordon Conference on Drug Carriers in Medicine and Biology in 2010. He has also been awarded the College of Engineering's Faculty Innovator Award, the Distinguished Teacher and Mentor Award from the Department of Bioengineering, and an Honorary Award from the College of Engineering's Minority Science and Engineering Program.

Dr. Stayton is the Director of the Center for Intracellular Delivery of Biologics, and his eclectic research group works at the interface of fundamental molecular science and applied molecular bioengineering. The group has projects aimed at elucidating basic principles underlying biomolecular recognition and connected projects applying those principles to drug delivery and diagnostic technologies through the development of new biohybrid materials that merge the impressive recognition and biofunctional properties of biomolecules with the impressive responsiveness and chemical versatility of "smart" polymers. The group has published over 200 research papers and has been funded by the NIH, the NSF, NASA, The Bill and Melinda Gates Foundation, and by the Coulter Foundation.

00, Seattle WA 98119 206.805.6300 Terms of Use Site Design

The Seattle Times **Business**

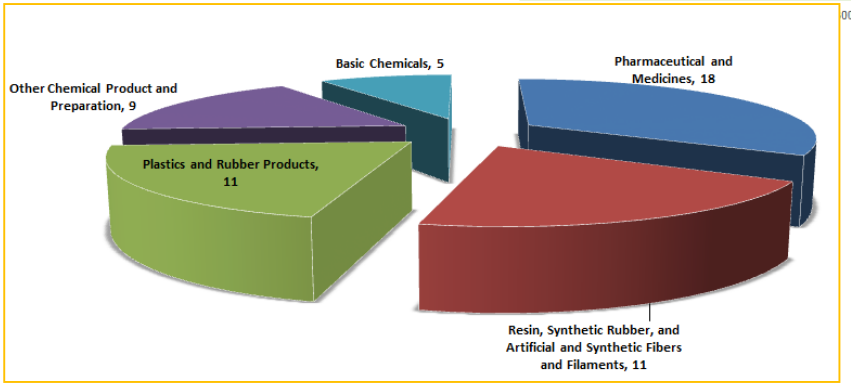
Business

$19 million for biotech startup PhaseRx

Originally published February 28, 2008 at 12:00 am | *Updated February 27, 2008 at 9:21 pm*

Venture capitalists have agreed to nurture PhaseRx, a nascent Seattle biotech, with funding of up to $19 million to pursue a pioneering technique that could help turn off harmful genes.

PhaseRx continued to grow, total 62 Patents/applications since 2008



NIH National Institutes of Health
Office of Portfolio Analysis

Case 4: Licensing - U of Cincinnati's Patent Licensed to CardioDX

Dr. Liggett's P50HL052318 award "Beta Adrenergic Receptor Variants in Heart Failure", funded in 1995-2004

DESCRIPTION | DETAILS | RESULTS | NEARBY PROJECTS BETA | LINKS | NEWS AND MORE

Parent Project Number: 2P50HL052318-06 **Sub-Project ID:** 0002
Title: BETA ADRENERGIC RECEPTOR VARIANTS IN HEART FAILURE
Contact PI / Project Leader: LIGGETT, STEPHEN B
Awardee Organization: UNIVERSITY OF CINCINNATI

Contact PI / Project Leader Information: **Program Official Information:** **Other PI Information:** Profile Exists / No Profile

Name: LIGGETT, STEPHEN B
Email: Unavailable
Title:
Name: Unavailable
Not Applicable

Organization: **Department/ Organization Type:** **Congressional District:**

Name: UNIVERSITY OF CINCINNATI
City: CINCINNATI **Country:** UNITED STATES (US)
Unavailable
Domestic Higher Education
State Code: OH
District: 01

Other Information:

FOA:
Study Section:
Fiscal Year: 2000 **Award Notice Date:**
DUNS Number: 041064767
Project Start Date: 14-FEB-2000
Budget Start Date:
CFDA Code:
Project End Date: 31-JAN-2001
Budget End Date:

Administering Institutes or Centers:

NATIONAL HEART, LUNG, AND BLOOD INSTITUTE

Project Funding Information for 2000:

Total Funding: $244,100

Year	Funding IC	FY Total Cost by IC
2000		$244,100

- Dr. Liggett filed patent US 7449292 on "Methods for predicting relative efficacy of a beta blocker therapy based on a B1-adrenergic receptor polymorphism"
- **This patent was missed by RePORTER / ExPORTER**
- In 2006 this patent was exclusively licensed to CardioDx®, Inc. (Founded in 2004), a cardiovascular genomic diagnostics company located in CA

genomeweb

Business & Policy | Technology | Research | Clinical | Disease Areas

Home » Business, Policy & Funding » Business News » CardioDx Raises $35M

CardioDx Raises $35M

Dec 18, 2014 | a GenomeWeb staff reporter

NEW YORK (GenomeWeb) – CardioDx today announced the close of a $35 million financing round.

The funds will be used to expand the commercial use of Corus CAD and to support the company's development of other genomic tests for coronary artery disease, CardioDx said. Corus CAD is a gene expression-based test for assessing non-diabetic patients who display symptoms suggestive of obstructive CAD.

Along with CardioDx's existing investors, Alberta Investment Management participated in the round.

In July the Redwood City, Calif.-based company said in a regulatory document that it raised $21 million toward a targeted goal of $25 million. The $35 million figure includes the July financing.

Conclusions

- Quantitative data-driven approaches can inform scientific portfolio management
  - Effective data cleaning is critical: garbage in, gospel out
- The Relative Citation Ratio (RCR) is a validated, article-level replacement for widely used but inaccurate and/or imprecise measures of scholarly influence
- Tracking outcomes and measuring impact of investments in biomedical research requires methods to monitor translation of basic and/or patented discoveries into improvements in human health

- OPA HOME
- TRAINING
- THE ANALYST
- TOOLS AND RESOURCES
- MEETINGS
- ABOUT US

Home » OPA Home » Tools and Resources

Resource Center

Are you interested in details of what tools are available to NIH staff and how to get hold of them? Would you like to see examples of how tools can be used in analyses? Do you need a refresher on a tool that you took a class for? ...Welcome to the OPA Resource Center.

OPA is currently developing computational tools that you can access from your desktop. Tools currently in development include:

Tool	Functionality	Beta Release Date
iClean	Efficient disambiguation	Spring, 2015
iCite	Effective bibliometrics	Spring, 2015
iTrans	Map translational science	Fall, 2015
iTech	Track patent, licensing, and start-up activity	Summer, 2016

Let us know if you are interested in beta-testing one or more of these tools in the coming months. In your email, please indicate which tool(s) you are interested in testing.

The OPA Tools Lab is located in B301 in building 1. For access please contact us.

For updates on training and other OPA activities, please sign up for our listserv.

- **OPA analysis of Nicholson and Ioannidis dataset**

Questions?
Comments?