

# Optimizing investments in the scientific enterprise by implementing effective data-driven decision making

George Santangelo, Ph.D.  
Director, Office of Portfolio Analysis  
DPCPSI/OD  
National Institutes of Health

# Today's topics

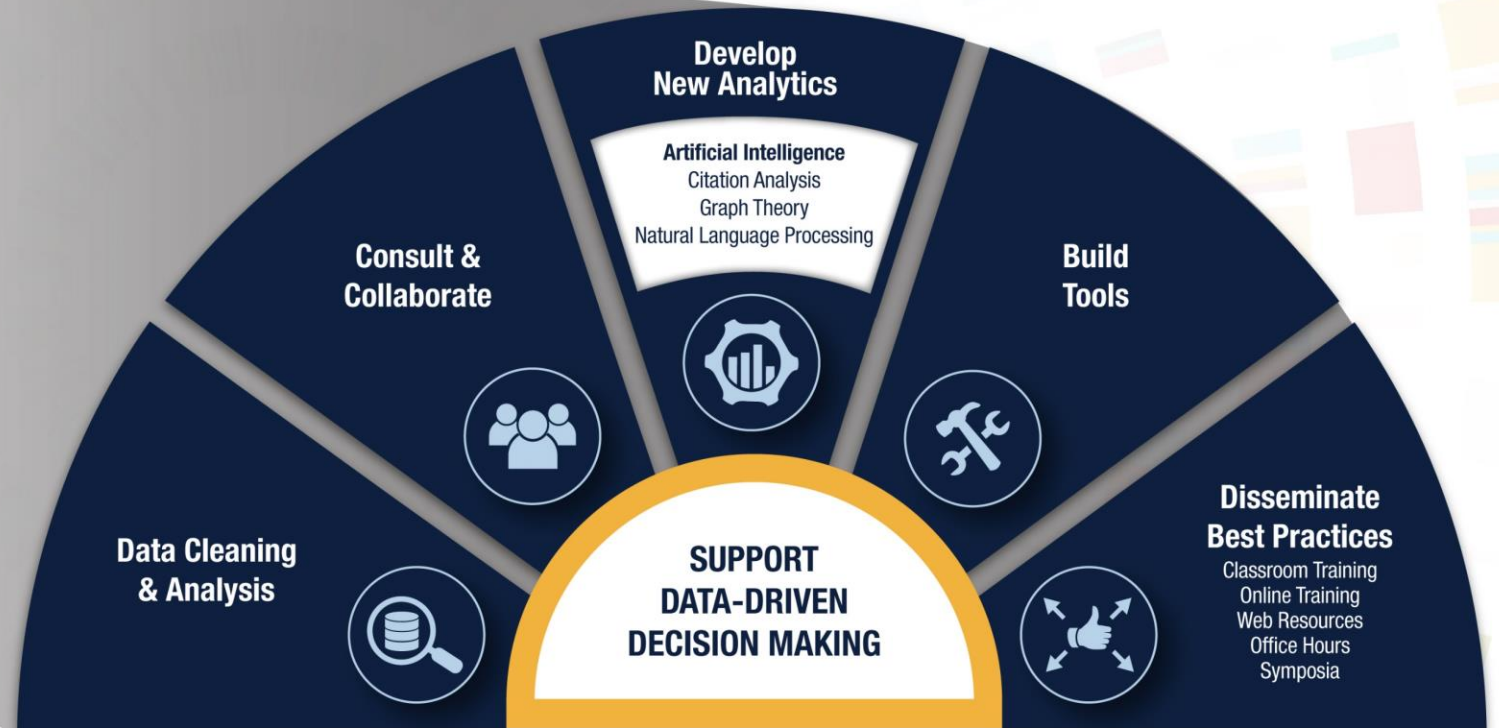
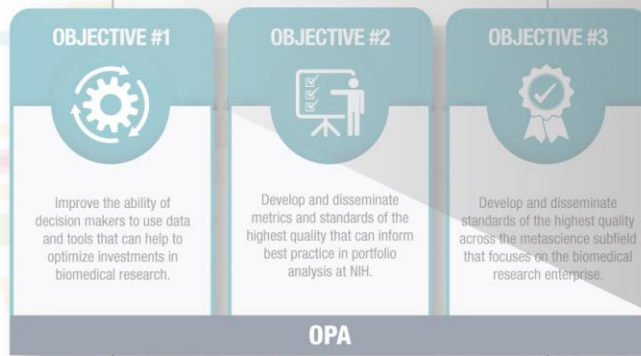
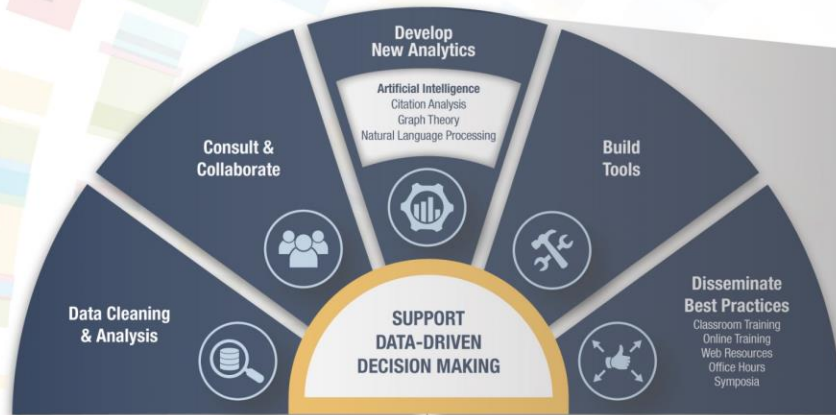
- Implement effective data-driven decision making that improves management of existing and planned portfolios
  - Engage across the enterprise
    - Requirements gathering
    - Build tools (don't buy them)
    - Train, consult, collaborate
- Develop and validate analytics
  - Leverage artificial intelligence/machine learning (AI/ML) tech wherever possible
    - Supervised ML with gold standards (requires high inter-rater reliability!)
    - Language model (LM)-based analysis of scientific content
    - AI/ML that measures productivity, identifies emerging topics, and predicts which topics will produce transformative breakthroughs
- Share analytics, the underlying code, and all input data with the public
  - A public-facing next-gen tool: *iSearch* Analytics

# OFFICE OF PORTFOLIO ANALYSIS

STRATEGIC PLAN, FISCAL YEARS 2021–2025

## OVERARCHING GOAL

To accelerate biomedical research by providing access to improved methods of data-driven decision making



The complete OPA strategic plan can be found on our website:

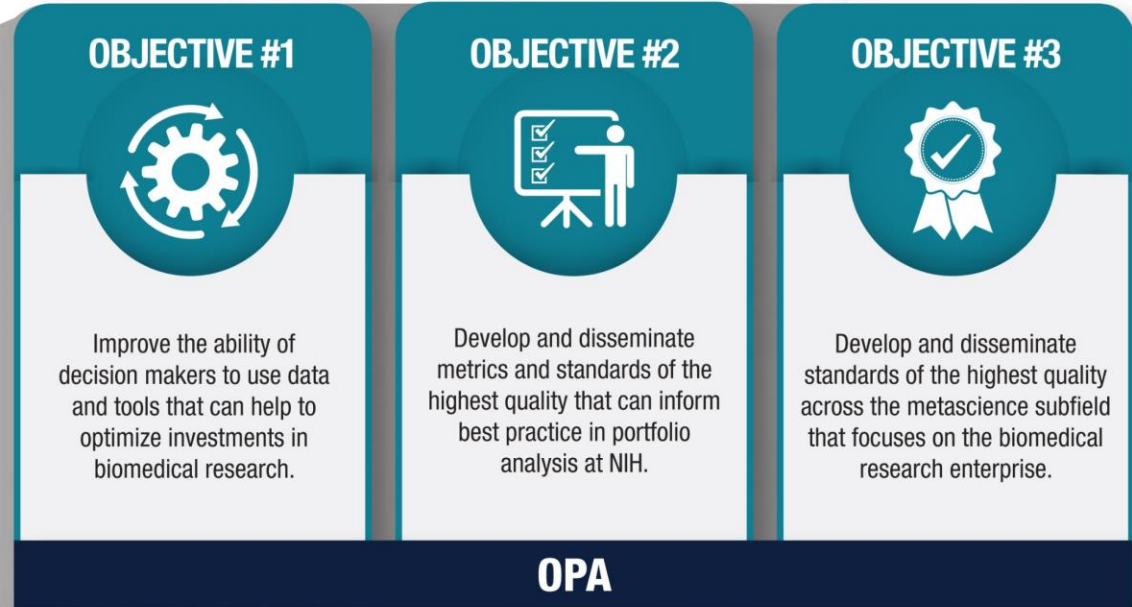
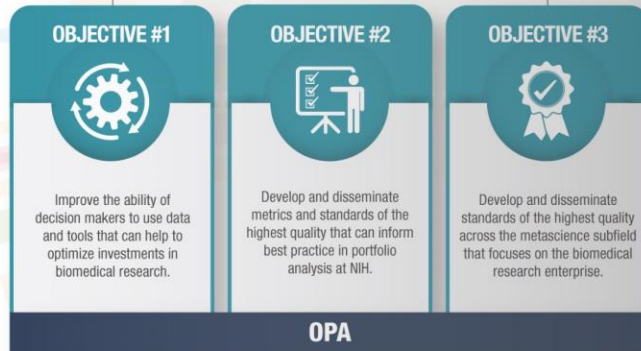
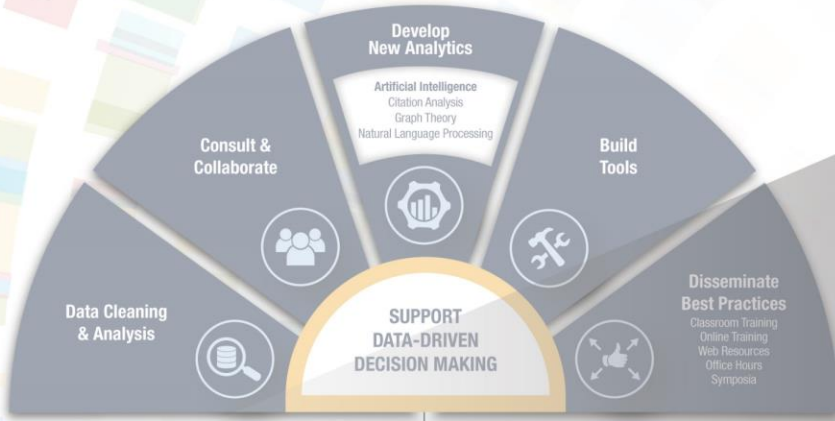
<https://dpcpsi.nih.gov/opa/strategicplan>

# OFFICE OF PORTFOLIO ANALYSIS

STRATEGIC PLAN, FISCAL YEARS 2021–2025

## OVERARCHING GOAL

To accelerate biomedical research by providing access to improved methods of data-driven decision making



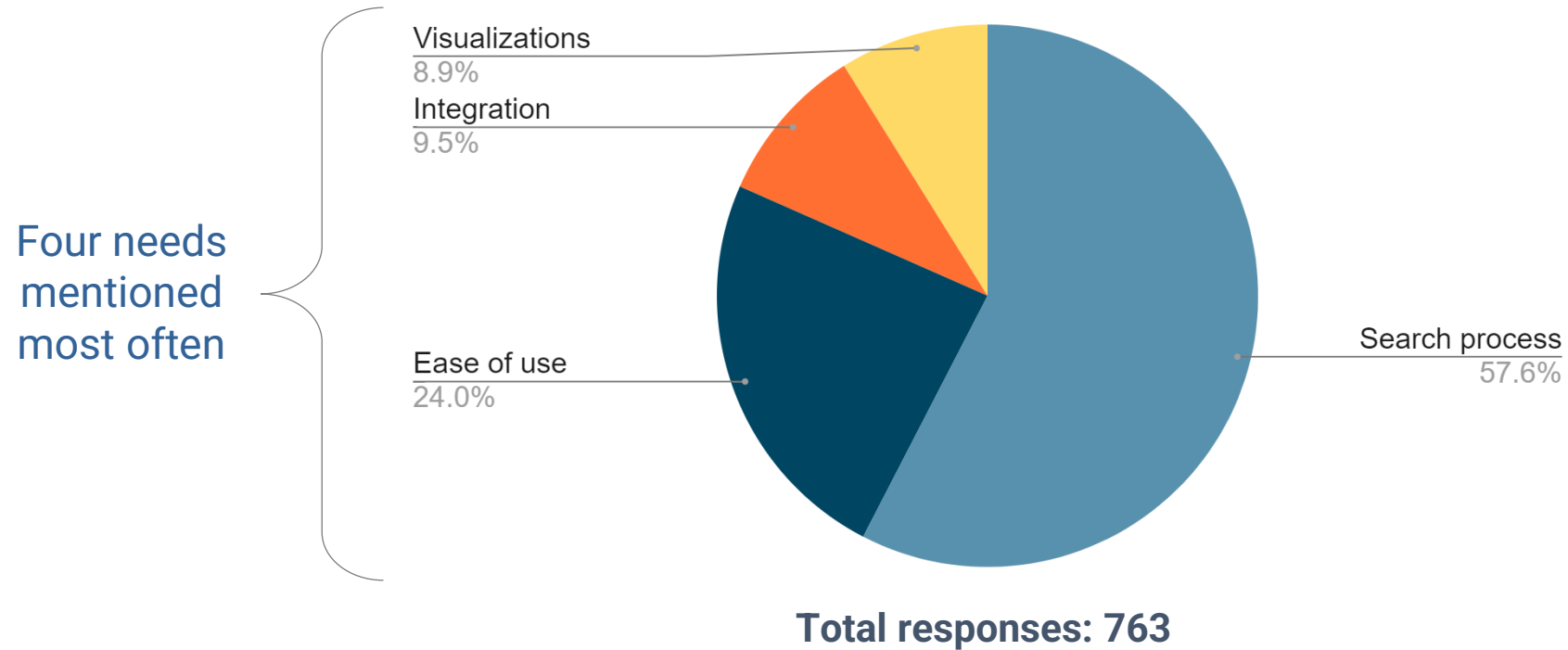
# Today's topics

- Implement effective data-driven decision making that improves management of existing and planned portfolios
  - Engage across the enterprise
    - Requirements gathering
    - Build tools (don't buy them)
    - Train, consult, collaborate
- Develop and validate analytics
  - Leverage artificial intelligence/machine learning (AI/ML) tech wherever possible
    - Supervised ML with gold standards (requires high inter-rater reliability!)
    - Language model (LM)-based analysis of scientific content
    - AI/ML that measures productivity, identifies emerging topics, and predicts which topics will produce transformative breakthroughs
- Share analytics, the underlying code, and all input data with the public
  - A public-facing next-gen tool: *iSearch* Analytics

# Research summary

What do NIH users care about most?

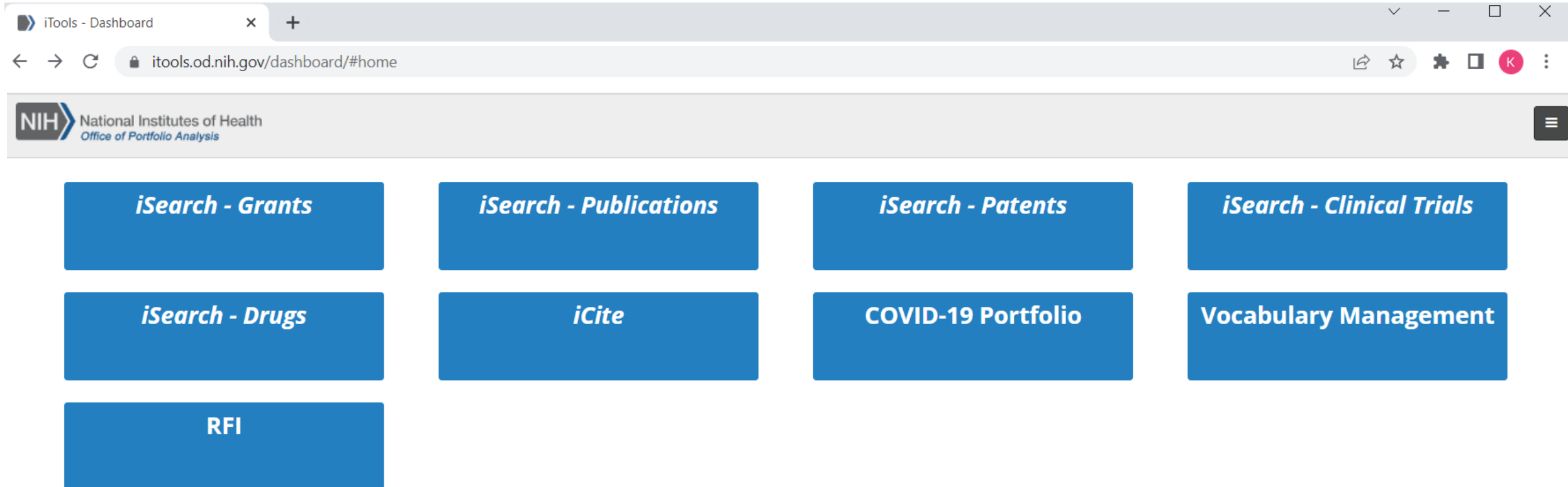
NIH surveys (March 2021 through June 2021)



# NextGen enterprise-wide portfolio analysis at NIH

## The OPA *iSearch* 2.6 dashboard

<https://itools.od.nih.gov/dashboard/#home>



The screenshot displays the NIH iSearch 2.6 dashboard interface. At the top, the browser tab is labeled "iTools - Dashboard" and the address bar shows the URL "itools.od.nih.gov/dashboard/#home". The page header includes the NIH logo and the text "National Institutes of Health Office of Portfolio Analysis". The main content area features a grid of blue buttons for various tools:

- iSearch - Grants*
- iSearch - Publications*
- iSearch - Patents*
- iSearch - Clinical Trials*
- iSearch - Drugs*
- iCite*
- COVID-19 Portfolio
- Vocabulary Management
- RFI

# The OPA training curriculum

- A free service provided to all NIH administrators
- Online registration

iTools are developed and managed by the Office of Portfolio Analysis

The OPA training team provides analytical resources to support the NIH community in performing analyses to the highest standards

## iTools Training Opportunities

### *iSearch*

In this class, you will learn about *iSearch*, NIH's next-generation portfolio analysis platform. You will learn how to access extensively-linked global grants, patents, publications, clinical trials and approved drugs.



- *iSearch: At a Glance* pdf
- View class information and resources
- Register for class

### *iSearch - Portfolio Tool*

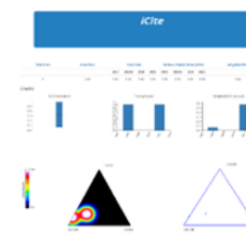
This class will cover how to use the portfolio tool feature available in all of the *iSearch* modules. Existing *iSearch* knowledge is required.



- *iSearch: Portfolio Tool At a Glance* pdf
- View class information and resources
- Register for class

### Measuring Research Outputs (*iCite* - RCR and Translation)

This class will introduce you to the *iCite* module within the OPA iTools dashboard. It will provide an overview of bibliometrics and tools used for investigating the impact of research outputs. It combines the previous 'Bibliometrics (*iCite* & RCR)' and 'Translational Science (iTrans)' classes.



- *iCite 2.0: At a Glance* pdf
- View class information and resources
- Register for class

### RFI Tool

This class will cover how to use the OPA RFI Tool. The RFI Tool streamlines the manual coding of responses to Requests For Information (RFI) or other text data sources e.g. blog comments.



- RFI Tool At a Glance pdf
- View class information and resources
- Register for class

Click Here to Register for a Class



# Driving adoption of data-driven decision making at NIH

## OPA open door consultations

22 NIH Institutes and Centers (ICs) requested OPA consultations between January 2021 and April 2022

### Sample projects:

IC	Project description
NCI	Analysis of NIH funding for surgeon scientists. Results of the analysis were published in the February 2021 issue of the <i>Journal of the American College of Surgeons</i> : <a href="https://www.journalacs.org/article/S1072-7515(20)32577-1/fulltext">https://www.journalacs.org/article/S1072-7515(20)32577-1/fulltext</a>
NCI	Collaboration to characterize NCI metabolomics research
NCI	Collaboration to characterize and evaluate the NCI training portfolio (Fs, Ks, and Ts)
NCI	Collaboration on portfolio analysis of grants that have supported clinical trials conducted in low- and middle-income countries
NCI	Collaboration on bladder cancer portfolio analysis
NIAID	Collaboration on portfolio analysis to characterize the DAIDS portfolio
NIAID	Compare patient-oriented goals to NIH-funded research on eczema
NIAID	Portfolio analysis of collaborative research of NIAID intramural investigators (with each other and with extramural PIs)
NIDDK	Characterization of program officer portfolios for three program divisions
NIDDK	Collaboration to characterize NIDDK DEM portfolio and assignments to POs
NIDDK	Collaboration on the NIDDK obesity portfolio
NHLBI	Collaborating to automate extraction of NHLBI grants and publications
NHLBI	Analysis of the geographical distribution of the NHLBI obesity portfolio
NHGRI	Collaboration on RNome RNA portfolio analysis
NHGRI	Collaboration on NHGRI AI/ML Genomics portfolio analysis
NICHD	Collaboration to analyze NICHD Rehab Specialties
NICHD	Collaboration on Adolescent HIV portfolio analysis
NIDA	Collaboration to characterize the NIDA health disparities portfolio
NIDA	Characterization of the mobile health portfolio
NIMH	Collaboration to characterize the NIMH health disparities portfolio
NIMH	Collaboration on NIMH/RDoC portfolio analysis
NIDCR	Collaboration on the analysis of Science of Behavior Change policy impact
NEI	Analysis of NEI research portfolio from FY2011-FY2020. Results of the analysis were presented to Mike Chiang
CC	Clinical Center portfolio analysis
NCATS	Collaboration to link NCATS PIs to CMS Open Payments database as an outcome measure
NINR	Collaboration to find duplicate or near duplicate applications submitted by the same investigator teams to multiple ICs so the same project is not unknowingly doubly funded
NINR	Collaboration on portfolio analysis of NIH-funded SDoH research
NIEHS	Collaboration on gene-environment interaction portfolio analysis
NINDS	Collaboration on portfolio analysis of BRAIN PI specialties
NIAMS	Analysis of the back pain portfolio

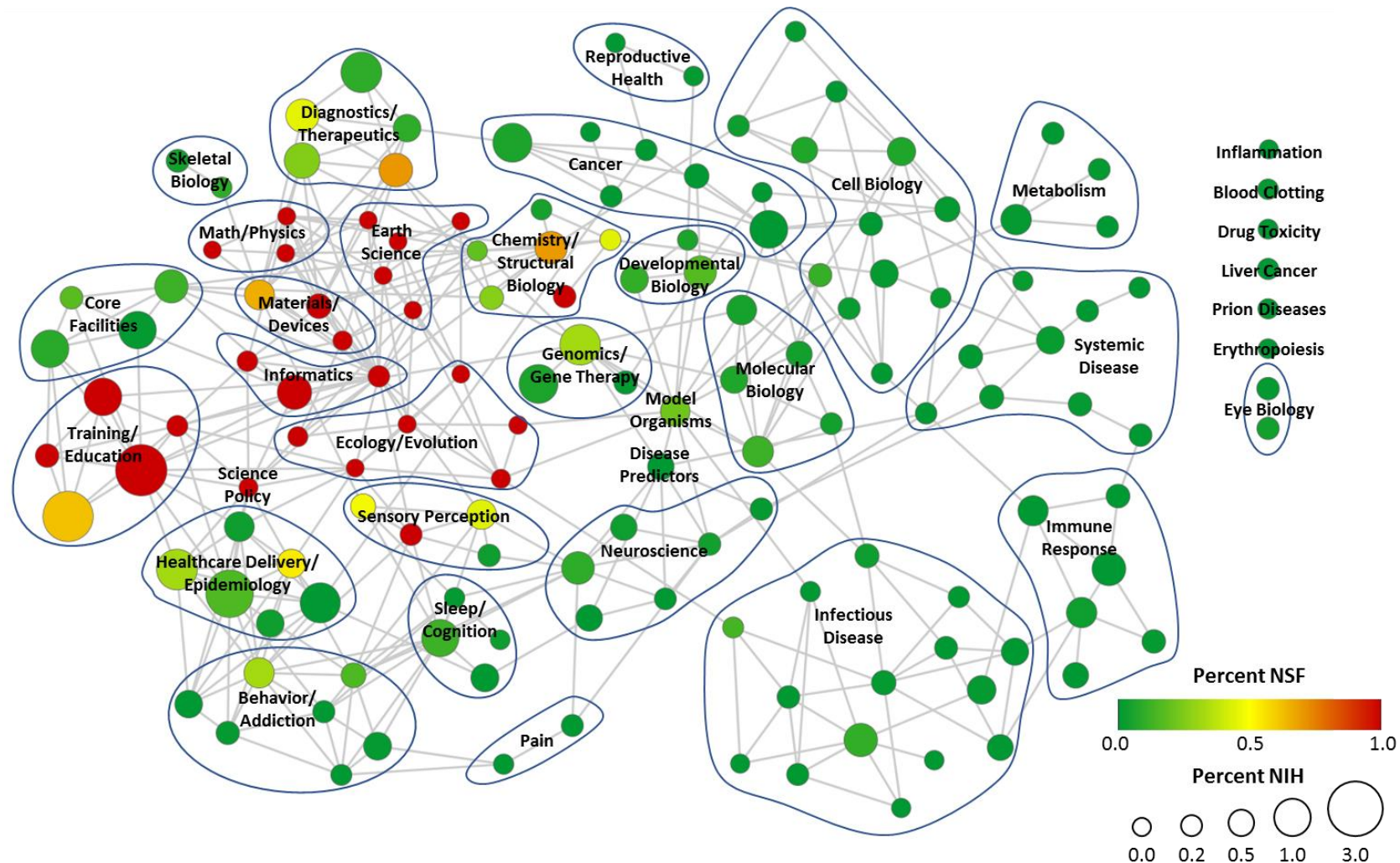
# Today's topics

- Implement effective data-driven decision making that improves management of existing and planned portfolios
  - Engage across the enterprise
    - Requirements gathering
    - Build tools (don't buy them)
    - Train, consult, collaborate
- **Develop and validate analytics**
  - Leverage artificial intelligence/machine learning (AI/ML) tech wherever possible
    - Supervised ML with gold standards (requires high inter-rater reliability!)
    - Language model (LM)-based analysis of scientific content
    - AI/ML that measures productivity, identifies emerging topics, and predicts which topics will produce transformative breakthroughs
- Share analytics, the underlying code, and all input data with the public
  - A public-facing next-gen tool: *iSearch Analytics*

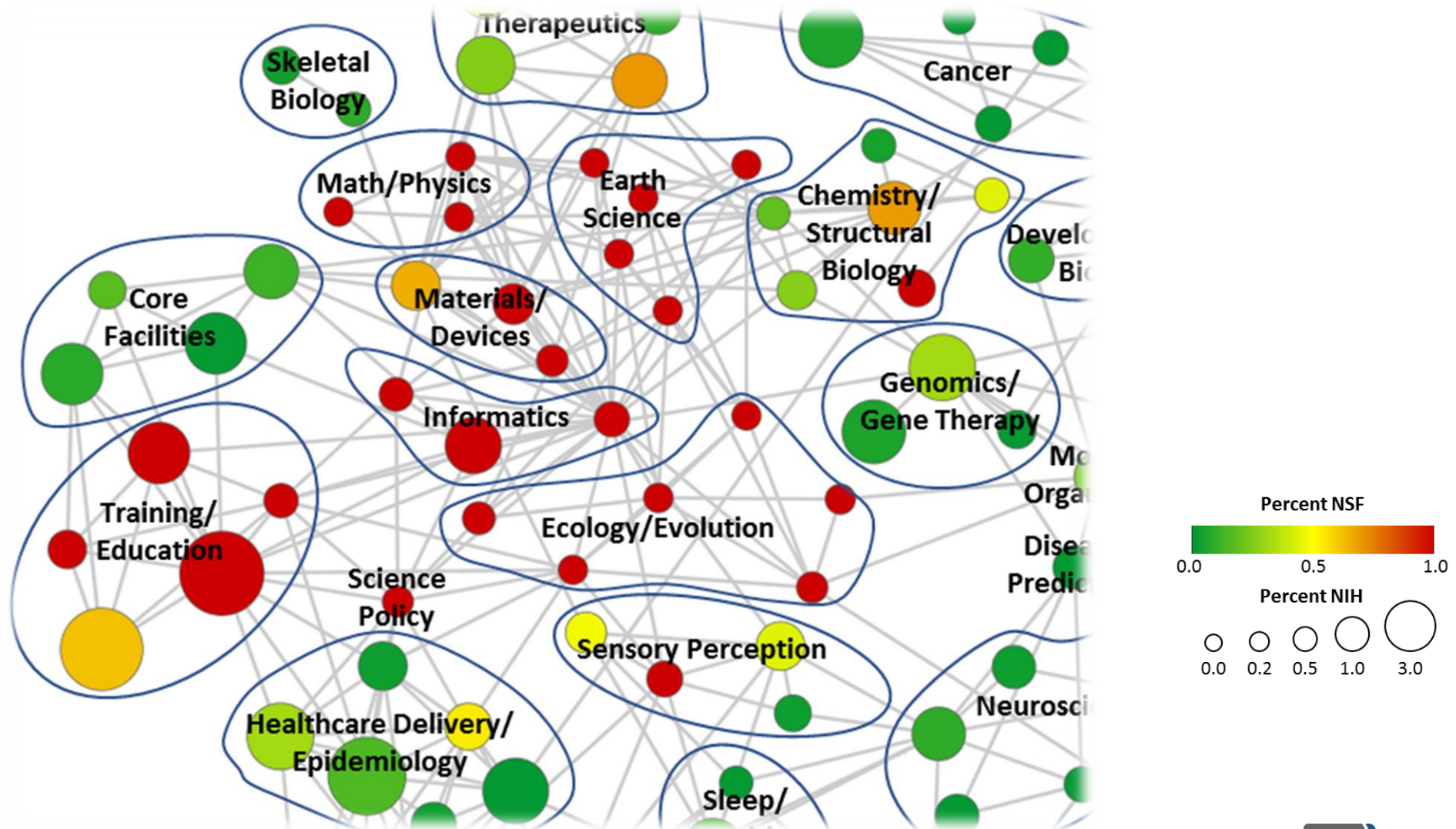
# Today's topics

- Implement effective data-driven decision making that improves management of existing and planned portfolios
  - Engage across the enterprise
    - Requirements gathering
    - Build tools (don't buy them)
    - Train, consult, collaborate
- **Develop and validate analytics**
  - Leverage artificial intelligence/machine learning (AI/ML) tech wherever possible
    - Supervised ML with gold standards (requires high inter-rater reliability!)
    - Language model (LM)-based analysis of scientific content
    - AI/ML that measures productivity, identifies emerging topics, and predicts which topics will produce transformative breakthroughs
- Share analytics, the underlying code, and all input data with the public
  - A public-facing next-gen tool: *iSearch Analytics*

# Using AI/ML to identify overlap between the NIH and NSF portfolios



# Using AI/ML to identify overlap between the NIH and NSF portfolios

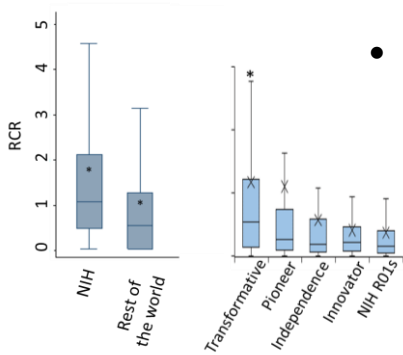


# Today's topics

- Implement effective data-driven decision making that improves management of existing and planned portfolios
  - Engage across the enterprise
    - Requirements gathering
    - Build tools (don't buy them)
    - Train, consult, collaborate
- **Develop and validate analytics**
  - Leverage artificial intelligence/machine learning (AI/ML) tech wherever possible
    - Supervised ML with gold standards (requires high inter-rater reliability!)
    - Language model (LM)-based analysis of scientific content
    - AI/ML that measures productivity, identifies emerging topics, and predicts which topics will produce transformative breakthroughs
- Share analytics, the underlying code, and all input data with the public
  - A public-facing next-gen tool: *iSearch Analytics*

# OPA AI to track and predict the impact of NIH decision-making

Track and parameterize:



- Influence using bibliometric data

The Relative Citation Ratio (RCR)

- Hutchins BI et al. *PLoS Biology* 2016 14:e1002541
- Hutchins BI et al. *PLoS Biology* 2017 15:e2003552

The publicly available OPA tool

iCite 2.0

Influence module

Translation module

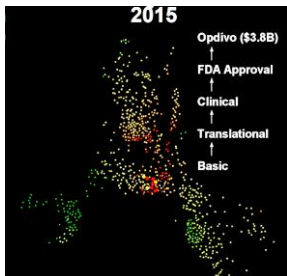
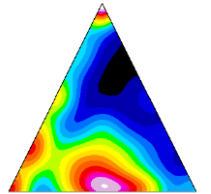
Open Citation Collection

Hutchins et al. *PLOS Biology* 2019 17:e3000385

- Translational progress / clinical trials (CTs) and tech transfer / patents

The triangle of biomedicine, APT scores

- Hutchins BI et al. *PLoS Biology* 2019 17(10):e3000416

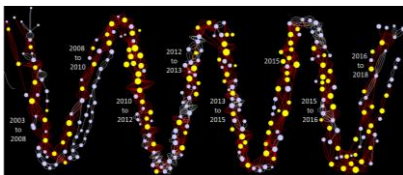
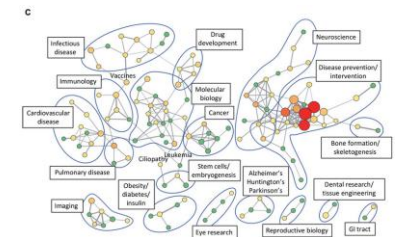


- Development of drugs and devices

Disambiguated drug and lead compound name, FDA data

- Language model (LM) analysis of NIH investments and publications

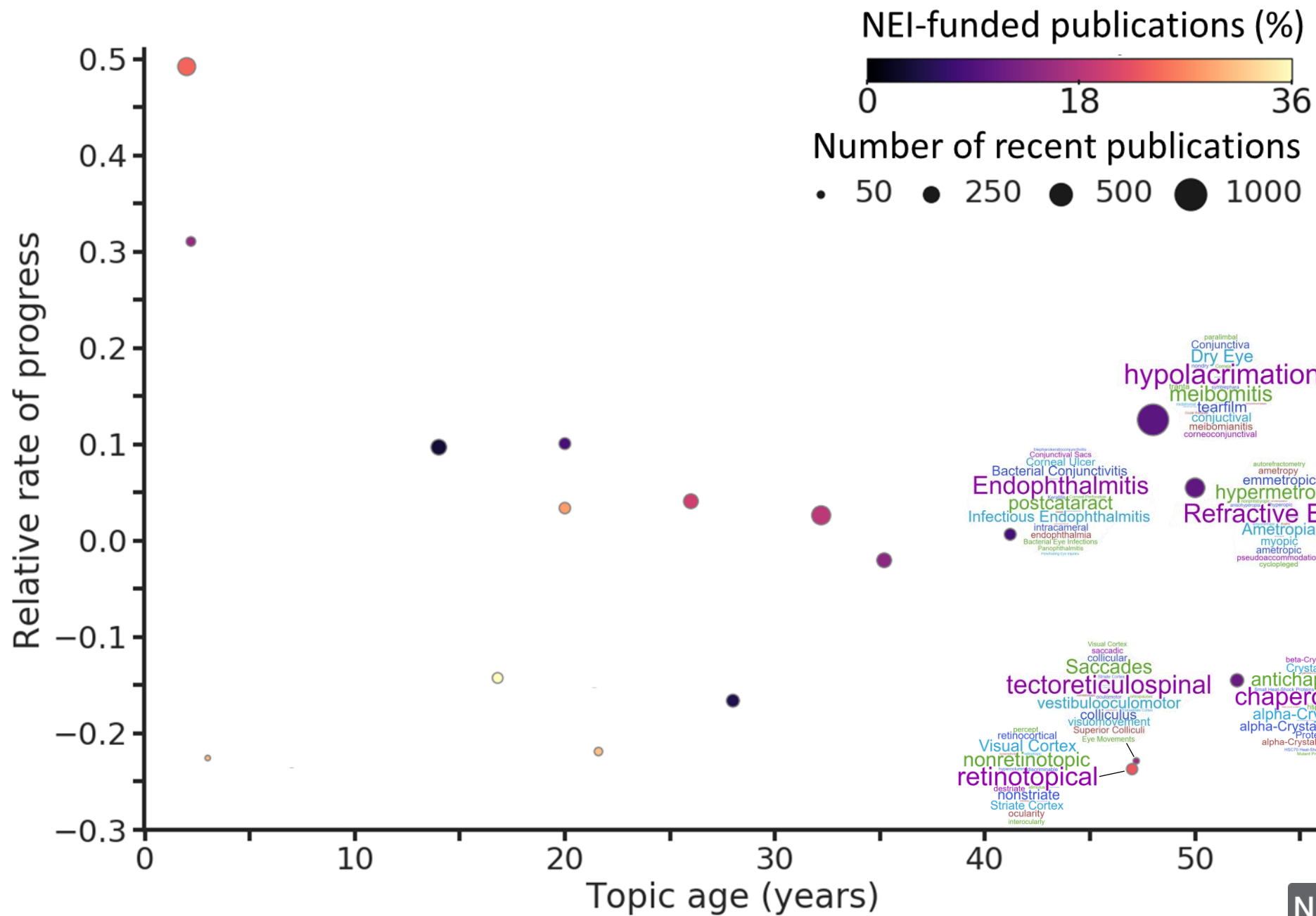
- Hoppe et al. *Science Advances* 2019 5:eaaw7238



- Rate of scientific progress and emergence

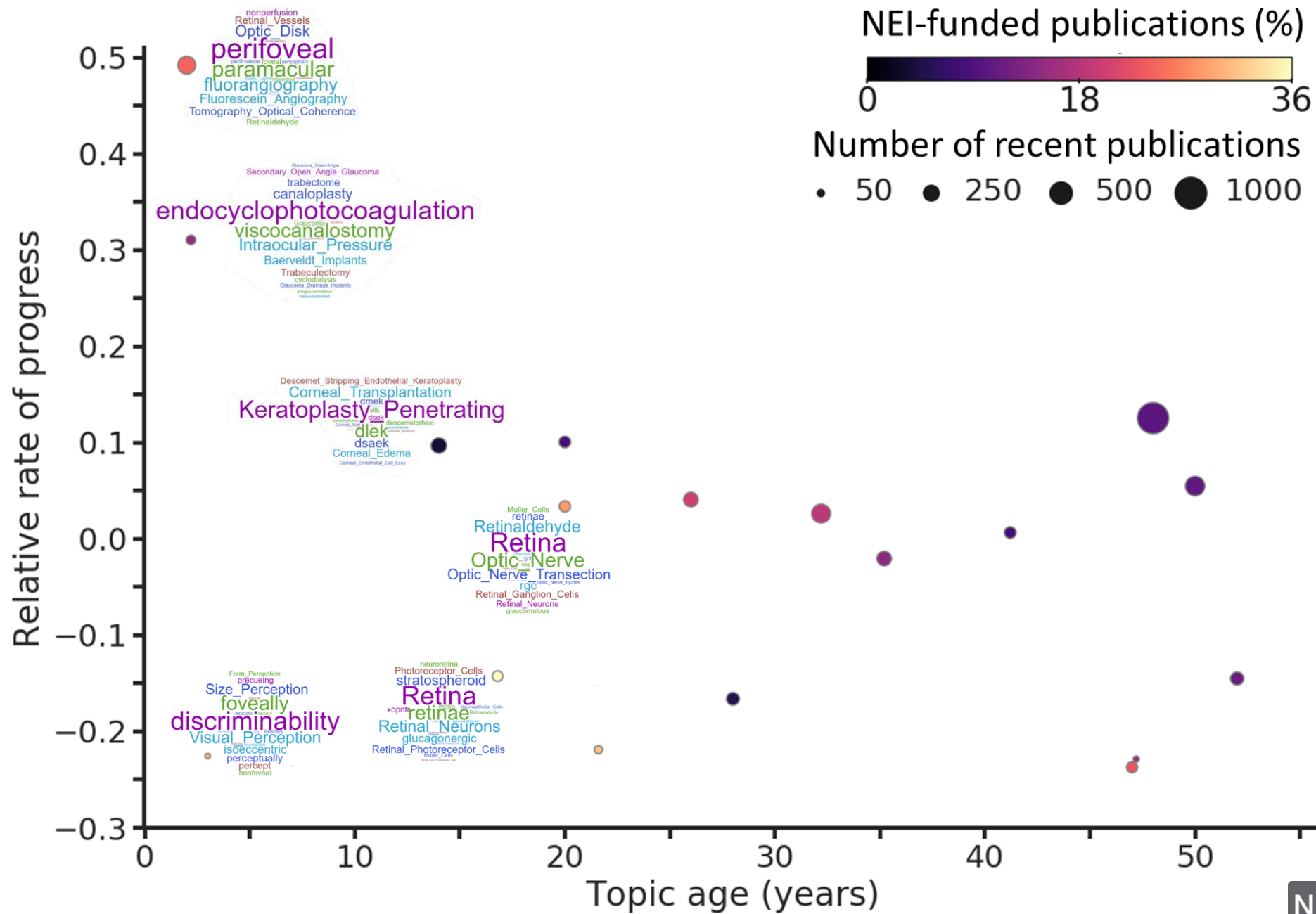
Prediction of transformative breakthroughs in biomedical research

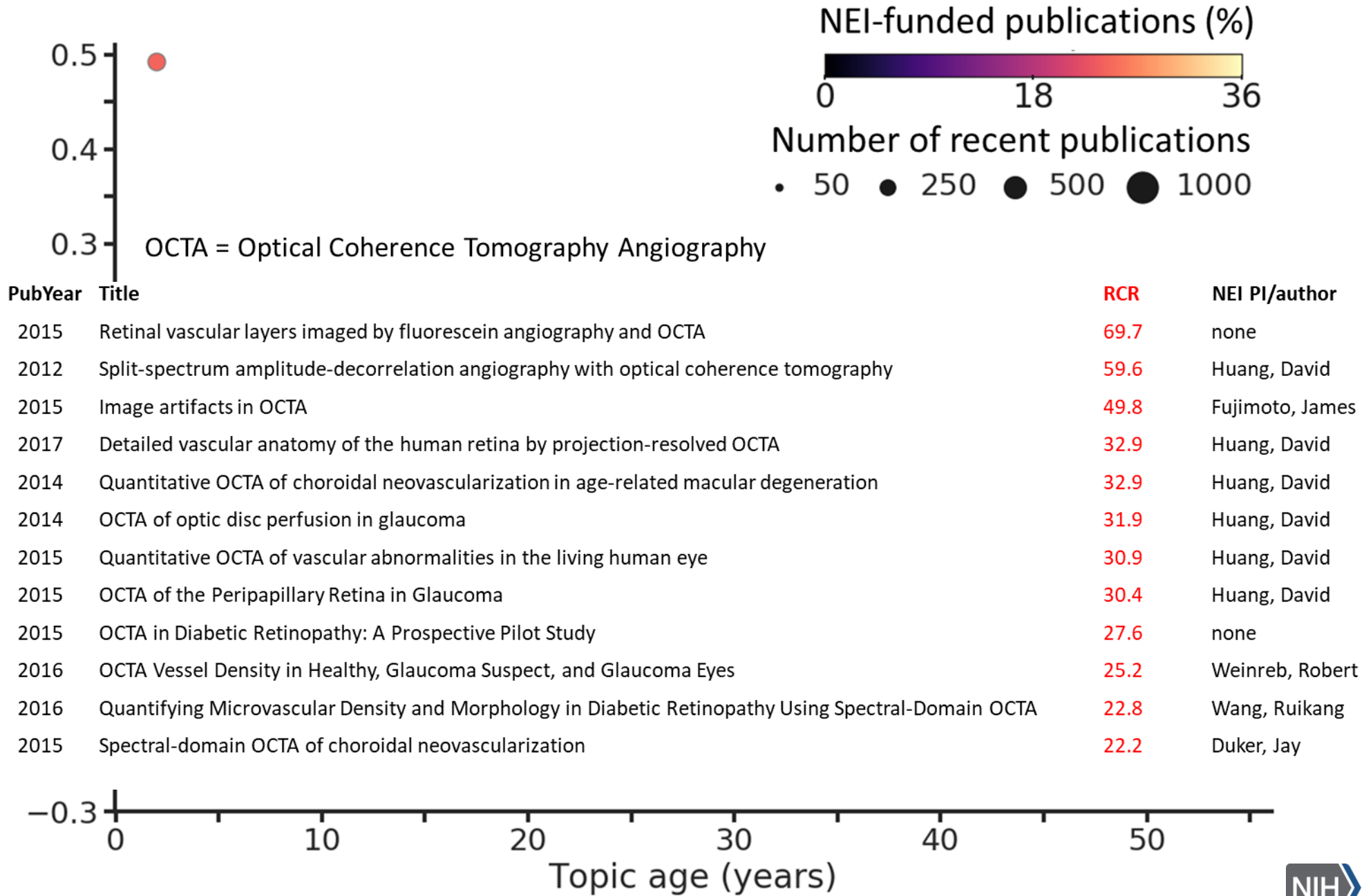
- Santangelo et al. U.S. Patent Application No. 63/257,818



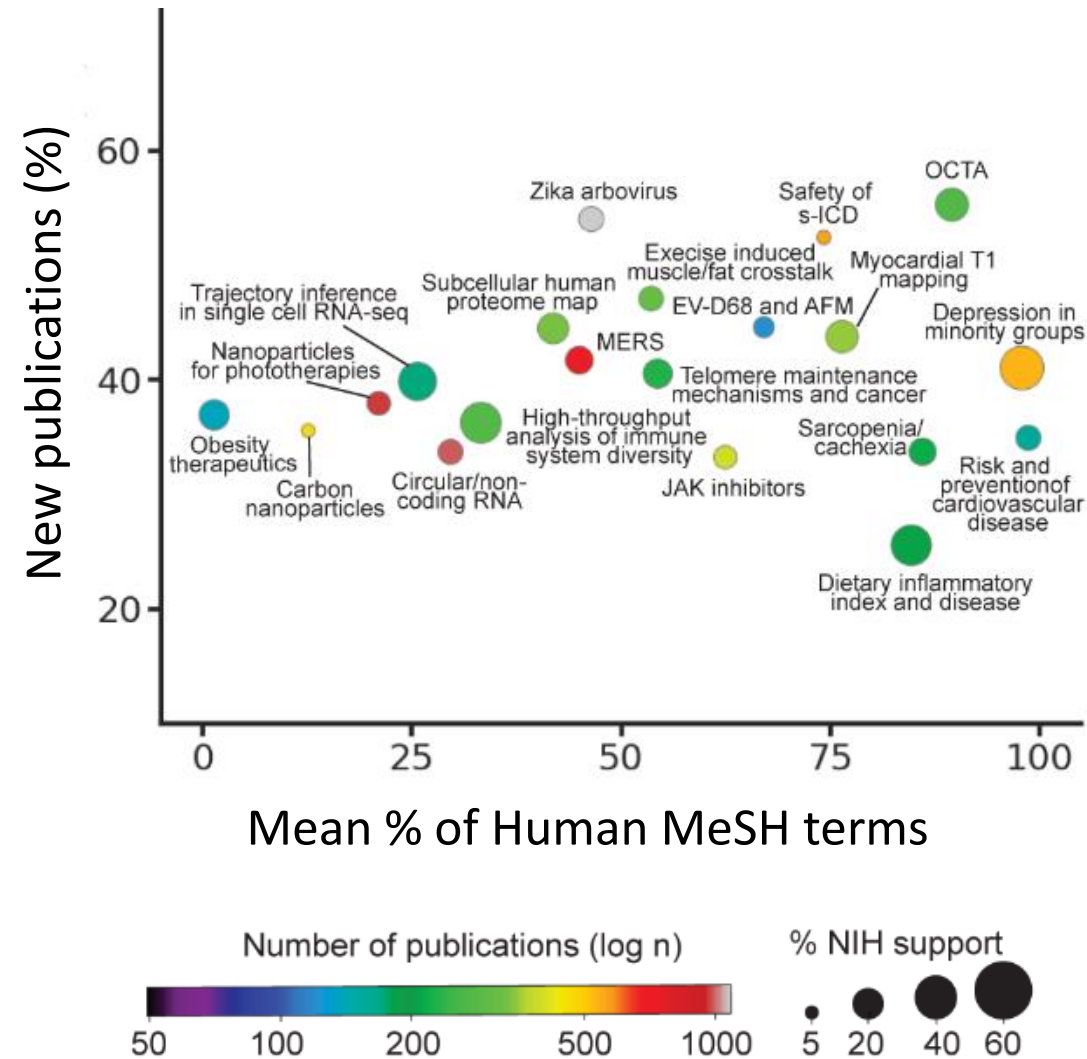








# Topics that signaled transformative breakthroughs between 2014 and 2017




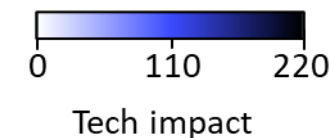
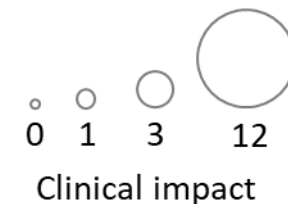
# Emergence of mRNA vaccines: Flow of information

- 1 Cationic liposome-mediated RNA transfection  
RCR 8.19 *Proc Natl Acad Sci* 1989
- 2 Characterization of a messenger RNA polynucleotide vaccine vector  
RCR 2.26 *Cancer Res* 1995
- 3 In vivo application of RNA leads to induction of specific cytotoxic T lymphocytes and antibodies  
RCR 3.12 *Eur J Immunol* 2000
- 4 Polarization of immunity induced by direct injection of naked sequence-stabilized mRNA vaccines  
RCR 1.46 *Cell Mol Life Sci* 2004
- 5 Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA  
RCR 12.70 *Immunity* 2005
- 6 Results of the first phase I/II clinical vaccination trial with direct injection of mRNA  
RCR 2.75 *J Immunother* 2008
- 7 Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability  
RCR 7.86 *Mol Ther* 2008
- 8 Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation  
RCR 3.45 *Nucleic Acids Res* 2010

- 9 Nucleoside modifications in RNA limit activation of 2'-5'-oligoadenylate synthetase and increase resistance to cleavage by RNase L  
RCR 2.43 *Nucleic Acids Res* 2011
- 10 Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA  
RCR 5.75 *Nucleic Acids Res* 2011
- 11 Protective efficacy of in vitro synthesized, specific mRNA vaccines against influenza A virus infection  
RCR 6.60 *Nat Biotechnol* 2012
- 12 Validation of the wild-type influenza A human challenge model H1N1pdMIST: an A(H1N1)pdm09 dose-finding investigational new drug study  
RCR 4.05 *Clin Infect Dis* 2015
- 13 Optimization of Lipid Nanoparticle Formulations for mRNA Delivery in Vivo with Fractional Factorial and Definitive Screening Designs  
RCR 7.44 *Nano Lett* 2015
- 14 Expression kinetics of nucleoside-modified mRNA delivered in lipid nanoparticles to mice by various routes  
RCR 8.56 *J Control Release* 2015

- 15 Preclinical and Clinical Demonstration of Immunogenicity by mRNA Vaccines against H10N8 and H7N9 Influenza Viruses  
RCR 12.92 *Mol Ther* 2017
- 16 Safety and immunogenicity of a mRNA rabies vaccine in healthy adults: an open-label, non-randomised, prospective, first-in-human phase 1 clinical trial  
RCR 7.50 *Lancet* 2017

Clinical trial or guideline = 



1



1989 1995 2000 2004 2008 2012 2016 2018

# Today's topics

- Implement effective data-driven decision making that improves management of existing and planned portfolios
  - Engage across the enterprise
    - Requirements gathering
    - Build tools (don't buy them)
    - Train, consult, collaborate
- Develop and validate analytics
  - Leverage artificial intelligence/machine learning (AI/ML) tech wherever possible
    - Supervised ML with gold standards (requires high inter-rater reliability!)
    - Language model (LM)-based analysis of scientific content
    - AI/ML that measures productivity, identifies emerging topics, and predicts which topics will produce transformative breakthroughs
- Share analytics, the underlying code, and all input data with the public
  - A public-facing next-gen tool: *iSearch Analytics*

# *iSearch* Analytics

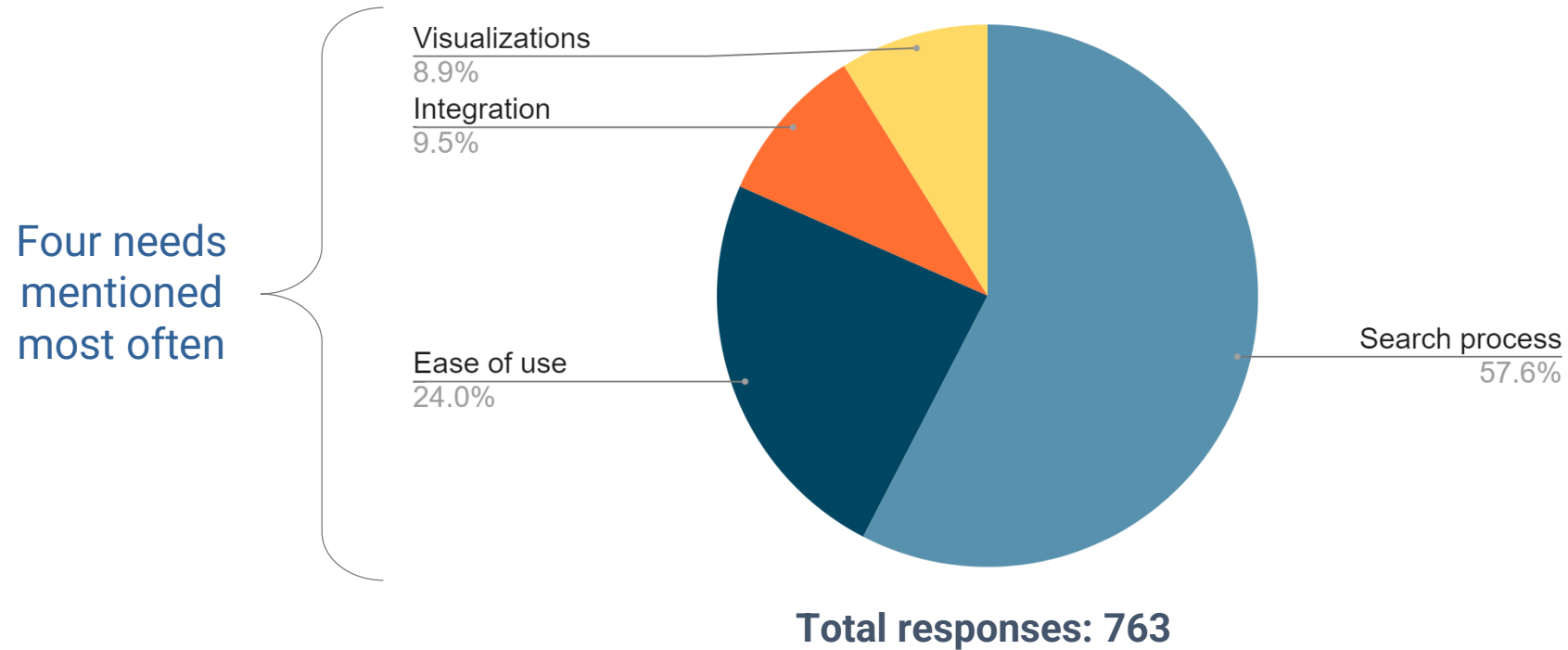
look for our public beta launch in September!



# Research summary

What do NIH users care about most?

NIH surveys (March 2021 through June 2021)





# Research summary

Who are the prospective public users?

How can we best meet their analytical needs?

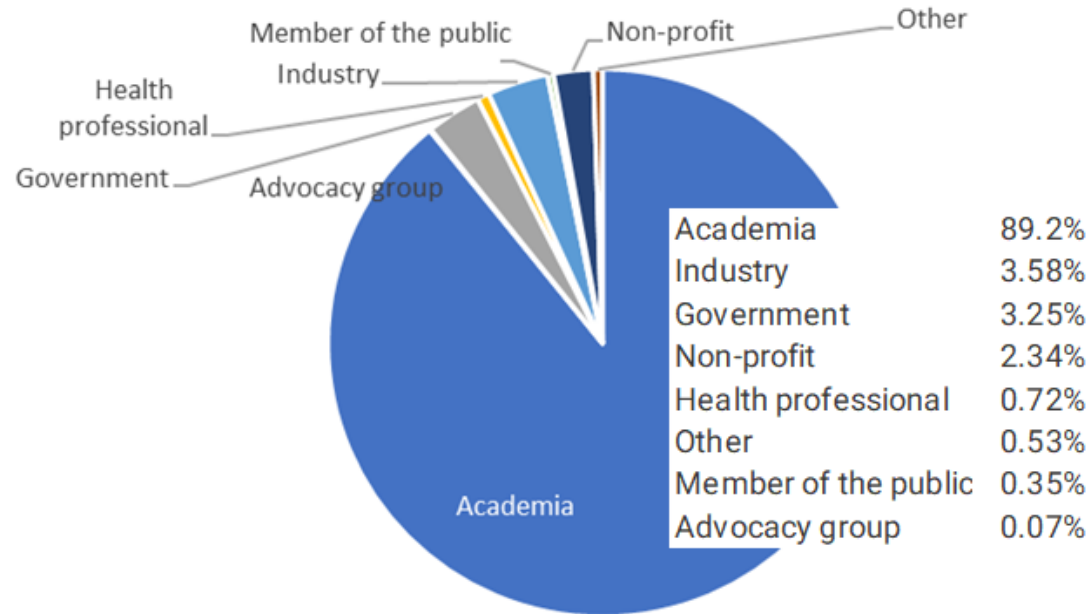
## September 2021 Request for Information (RFI) public responses



A total of **6850** people responded to the RFI

**1360** agreed to being contacted for follow-up

## Respondent categories



## Interviews & User testing

- NIH Staff
- Staff of other government agencies
- Public RFI respondents who agreed to be contacted



# *iSearch Analytics*

## A new public-facing scientific resource

### Transformative functionality

Three major new features



#### **Literature Expansion**

Goes beyond PubMed:  
Adds preprints and  
continues progress towards  
comprehensive coverage of publications



#### **Person Disambiguation\***

Provides highly cleaned  
person-level data and metrics



#### **Visualizations Reimagined**

Language Model (LM) organizes  
the information for users  
Automated, multi-layered,  
interactive displays with  
AI-generated labels to guide  
searches and exploration

\*Yu et al. (2021) The effect of mentee and mentor gender on scientific productivity of applicants for NIH training fellowships. bioRxiv 10.1101/2021.02.02.429450

# *iSearch* Analytics

A new public-facing scientific resource

## Transformative functionality

Examples of the analytical power of this new tool



**Map research topics,  
drill down on sub-topics  
(Grants, pubs, CTs)**

- Just “click around”
- Use keywords
- Optimize searches by using both strategies
- Find overlapping investments
- Find research gaps

# *iSearch* Analytics

## A new public-facing scientific resource

### Transformative functionality

Examples of the analytical power of this new tool



#### **Map research topics, drill down on sub-topics (Grants, pubs, CTs)**

- Just “click around”
- Use keywords
- Optimize searches by using both strategies
- Find overlapping investments
- Find research gaps



#### **Find experts in any area of biomedical research**

- Identify the subset funded by NIH, VA, FDA, CDC etc.
- Find all of their co-authors
- For the awardees above, find all of their co-PIs



#### **Review the careers of individuals or groups**

- Publications
- Preprints
- Clinical trials
- Patents
- NIH support

# *iSearch* Analytics

## A new public-facing scientific resource

### Transformative functionality

Examples of the analytical power of this new tool



#### **Map research topics, drill down on sub-topics (Grants, pubs, CTs)**

- Just “click around”
- Use keywords
- Optimize searches by using both strategies
- Find overlapping investments
- Find research gaps



#### **Find experts in any area of biomedical research**

- Identify the subset funded by NIH, VA, FDA, CDC etc.
- Find all of their co-authors
- For the awardees above, find all of their co-PIs



#### **Review the careers of individuals or groups**

- Publications
- Preprints
- Clinical trials
- Patents
- NIH support



#### **Characterize research outputs and measure impact**

- Highly influential pubs and/or those with
- High translational potential
- Verified clinical impact
- Verified tech impact

# iSearch Analytics

## A new public-facing scientific resource

### Transformative functionality

Examples of the analytical power of this new tool



#### Map research topics, drill down on sub-topics (Grants, pubs, CTs)

- Just “click around”
- Use keywords
- Optimize searches by using both strategies
- Find overlapping investments
- Find research gaps



#### Find experts in any area of biomedical research

- Identify the subset funded by NIH, VA, FDA, CDC etc.
- Find all of their co-authors
- For the awardees above, find all of their co-PIs



#### Review the careers of individuals or groups

- Publications
- Preprints
- Clinical trials
- Patents
- NIH support



#### Characterize research outputs and measure impact

- Highly influential pubs and/or those with
- High translational potential
- Verified clinical impact
- Verified tech impact



#### Explore topics predicted to produce future scientific breakthroughs

- Find the very recent seminal papers on each topic
- Characterize the investments that funded or are funding each topic

# Today's topics

- Implement effective data-driven decision making that improves management of existing and planned portfolios
  - Engage across the enterprise
    - Requirements gathering
    - Build tools (don't buy them)
    - Train, consult, collaborate
- Develop and validate analytics
  - Leverage artificial intelligence/machine learning (AI/ML) tech wherever possible
    - Supervised ML with gold standards (requires high inter-rater reliability!)
    - Language model (LM)-based analysis of scientific content
    - AI/ML that measures productivity, identifies emerging topics, and predicts which topics will produce transformative breakthroughs
- Share analytics, the underlying code, and all input data with the public
  - A public-facing next-gen tool: *iSearch* Analytics