



Research that you and others can **reuse & reproduce**

Day 7 | Session 3
July 13, 2021
KCNI Summer School 2021

Abhi Pratap, PhD
Group Head - Digital Health & AI
Independent Scientist, KCNI
Faculty Affiliate, Vector Institute

 @abhishekpratap

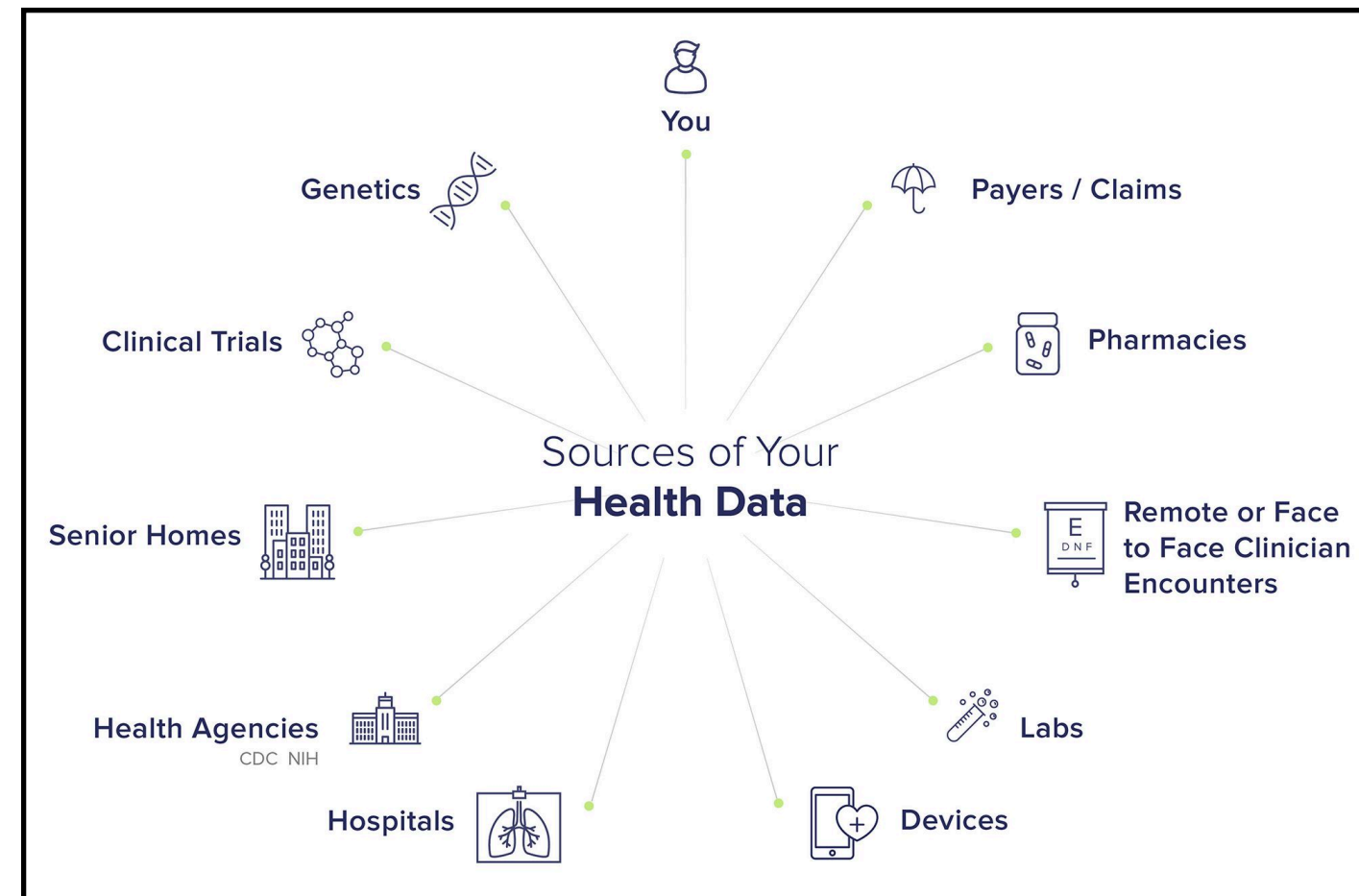
camh | Krembil Center for
Neuroinformatics
Centre for Addiction and Mental Health

Housekeeping - In case you want to follow along



1. Create an account on synapse - www.synapse.org
2. If you want to upload data - Get certified - <https://www.synapse.org/#!Quiz:Certification>

“Context”

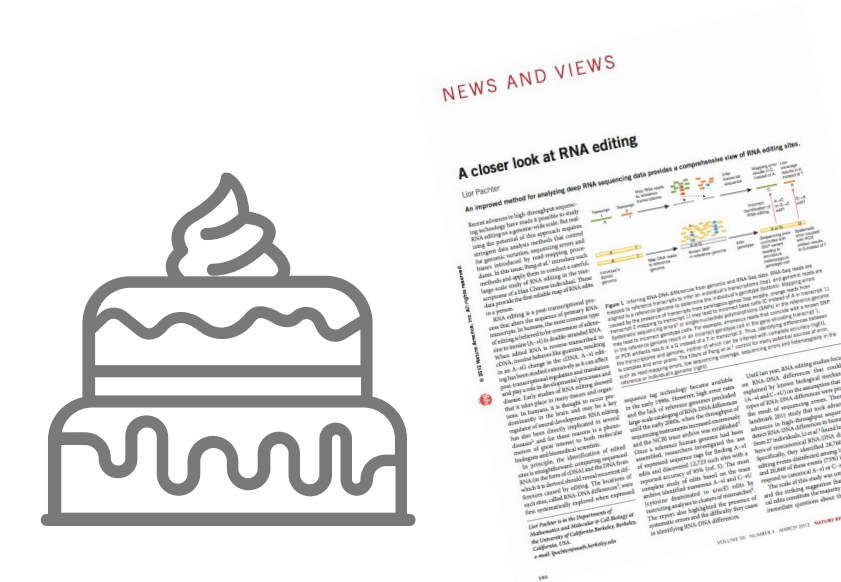


Sources

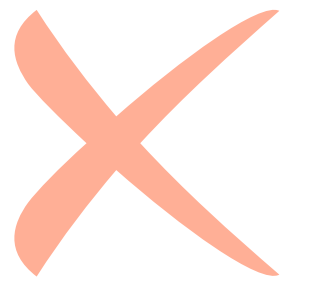
Data

Insights

Recipe | Ingredients

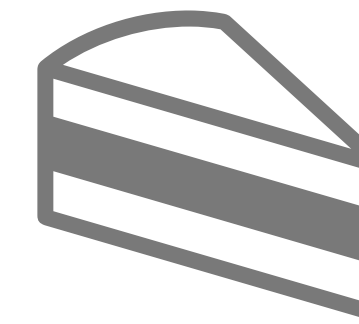


Lab 1 - Original Student



Missing ingredients

Lab 3



Lab 1 - Second Grad Student



Lab 2



Lab 4

School-based smoking prevention in children and adolescents: review of the scientific literature
 L. La Torre, G. Chiarillo, G. Bonaiuto
 November 1, 2012 | Volume 152 | Number 9 | Published online 14 October 2012

Genome Sequence of *Candidatus Nitrososphaera evergladensis* from Group 1.1b Enriched from Everglades Soil Reveals Novel Genomic Features of the Ammonia-Oxidizing Archaea
 Katerina V. Zhurina*, Rezaul Karim, Michael E. Leonard*, Patricia Dor de Queiroz*, Patrick A. O. Connor*, Jennifer C. Green*, William G. Farnham*, Isabella M. Sorensen*, John W. Taylor*
 November 1, 2012 | Volume 152 | Number 9 | Published online 14 October 2012

What is Agricultural Biotechnology Research Capacity in Developing Countries?
 Abstract
 In 1998 INSAR has conducted a biotechnology research indicator survey in their national agricultural research systems in Mexico, India, and China. Findings show that agricultural research systems in Mexico, India, and China have made significant progress in the last decade. However, the survey also identified several key challenges that need to be addressed. The most significant challenges are: (1) limited public sector investment in agricultural research, (2) inadequate infrastructure, (3) limited human resources, and (4) limited access to international markets. The survey also identified several key opportunities for improvement, including: (1) increasing public sector investment, (2) improving infrastructure, (3) increasing human resources, and (4) increasing access to international markets.

What is Agricultural Biotechnology Research Capacity in Developing Countries?
 Abstract
 In 1998 INSAR has conducted a biotechnology research indicator survey in their national agricultural research systems in Mexico, India, and China. Findings show that agricultural research systems in Mexico, India, and China have made significant progress in the last decade. However, the survey also identified several key challenges that need to be addressed. The most significant challenges are: (1) limited public sector investment in agricultural research, (2) inadequate infrastructure, (3) limited human resources, and (4) limited access to international markets. The survey also identified several key opportunities for improvement, including: (1) increasing public sector investment, (2) improving infrastructure, (3) increasing human resources, and (4) increasing access to international markets.

Foot-antigen Protein 1
 Abstract
 The foot-antigen protein 1 (FAP1) is a novel protein that is expressed in the foot of the tick *Ixodes ricinus*. It is a member of the immunoglobulin superfamily and is thought to play a role in the tick's attachment to its host. FAP1 is a type I transmembrane protein with a single extracellular domain that contains a conserved cysteine loop. The extracellular domain of FAP1 is highly similar to that of the foot-antigen protein 2 (FAP2), which is also expressed in the foot of *I. ricinus*. The intracellular domain of FAP1 is highly similar to that of the foot-antigen protein 3 (FAP3), which is also expressed in the foot of *I. ricinus*. The FAP1 protein is thought to play a role in the tick's attachment to its host by interacting with the host's immune system.

Foot-antigen Protein 1
 Abstract
 The foot-antigen protein 1 (FAP1) is a novel protein that is expressed in the foot of the tick *Ixodes ricinus*. It is a member of the immunoglobulin superfamily and is thought to play a role in the tick's attachment to its host. FAP1 is a type I transmembrane protein with a single extracellular domain that contains a conserved cysteine loop. The extracellular domain of FAP1 is highly similar to that of the foot-antigen protein 2 (FAP2), which is also expressed in the foot of *I. ricinus*. The intracellular domain of FAP1 is highly similar to that of the foot-antigen protein 3 (FAP3), which is also expressed in the foot of *I. ricinus*. The FAP1 protein is thought to play a role in the tick's attachment to its host by interacting with the host's immune system.

A closer look at RNA editing
 Abstract
 RNA editing is a post-transcriptional modification of RNA that can alter the amino acid sequence of a protein. It is a common phenomenon in many organisms and is thought to play a role in the regulation of gene expression. RNA editing can occur at several different sites in the RNA molecule, including the 5' and 3' ends, and internal sites. The most common type of RNA editing is cytosine deamination, which converts cytosine to uracil. Other types of RNA editing include adenylation, guanylation, and inosinylation. RNA editing is thought to play a role in the regulation of gene expression by altering the amino acid sequence of a protein, which can affect its function.

A closer look at RNA editing
 Abstract
 RNA editing is a post-transcriptional modification of RNA that can alter the amino acid sequence of a protein. It is a common phenomenon in many organisms and is thought to play a role in the regulation of gene expression. RNA editing can occur at several different sites in the RNA molecule, including the 5' and 3' ends, and internal sites. The most common type of RNA editing is cytosine deamination, which converts cytosine to uracil. Other types of RNA editing include adenylation, guanylation, and inosinylation. RNA editing is thought to play a role in the regulation of gene expression by altering the amino acid sequence of a protein, which can affect its function.

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

School-based smoking prevention in children and adolescents: review of the scientific literature
 L. La Torre, G. Chiarillo, G. Bonaiuto
 November 1, 2012 | Volume 152 | Number 9 | Published online 14 October 2012

Genome Sequence of *Candidatus Nitrososphaera evergladensis* from Group 1.1b Enriched from Everglades Soil Reveals Novel Genomic Features of the Ammonia-Oxidizing Archaea
 Katerina V. Zhurina*, Rezaul Karim, Michael E. Leonard*, Patricia Dor de Queiroz*, Patrick A. O. Connor*, Jennifer C. Green*, William G. Farnham*, Isabella M. Sorensen*, John W. Taylor*
 November 1, 2012 | Volume 152 | Number 9 | Published online 14 October 2012

What is Agricultural Biotechnology Research Capacity in Developing Countries?
 Abstract
 In 1998 INSAR has conducted a biotechnology research indicator survey in their national agricultural research systems in Mexico, India, and China. Findings show that agricultural research systems in Mexico, India, and China have made significant progress in the last decade. However, the survey also identified several key challenges that need to be addressed. The most significant challenges are: (1) limited public sector investment in agricultural research, (2) inadequate infrastructure, (3) limited human resources, and (4) limited access to international markets. The survey also identified several key opportunities for improvement, including: (1) increasing public sector investment, (2) improving infrastructure, (3) increasing human resources, and (4) increasing access to international markets.

What is Agricultural Biotechnology Research Capacity in Developing Countries?
 Abstract
 In 1998 INSAR has conducted a biotechnology research indicator survey in their national agricultural research systems in Mexico, India, and China. Findings show that agricultural research systems in Mexico, India, and China have made significant progress in the last decade. However, the survey also identified several key challenges that need to be addressed. The most significant challenges are: (1) limited public sector investment in agricultural research, (2) inadequate infrastructure, (3) limited human resources, and (4) limited access to international markets. The survey also identified several key opportunities for improvement, including: (1) increasing public sector investment, (2) improving infrastructure, (3) increasing human resources, and (4) increasing access to international markets.

Foot-antigen Protein 1
 Abstract
 The foot-antigen protein 1 (FAP1) is a novel protein that is expressed in the foot of the tick *Ixodes ricinus*. It is a member of the immunoglobulin superfamily and is thought to play a role in the tick's attachment to its host. FAP1 is a type I transmembrane protein with a single extracellular domain that contains a conserved cysteine loop. The extracellular domain of FAP1 is highly similar to that of the foot-antigen protein 2 (FAP2), which is also expressed in the foot of *I. ricinus*. The intracellular domain of FAP1 is highly similar to that of the foot-antigen protein 3 (FAP3), which is also expressed in the foot of *I. ricinus*. The FAP1 protein is thought to play a role in the tick's attachment to its host by interacting with the host's immune system.

Foot-antigen Protein 1
 Abstract
 The foot-antigen protein 1 (FAP1) is a novel protein that is expressed in the foot of the tick *Ixodes ricinus*. It is a member of the immunoglobulin superfamily and is thought to play a role in the tick's attachment to its host. FAP1 is a type I transmembrane protein with a single extracellular domain that contains a conserved cysteine loop. The extracellular domain of FAP1 is highly similar to that of the foot-antigen protein 2 (FAP2), which is also expressed in the foot of *I. ricinus*. The intracellular domain of FAP1 is highly similar to that of the foot-antigen protein 3 (FAP3), which is also expressed in the foot of *I. ricinus*. The FAP1 protein is thought to play a role in the tick's attachment to its host by interacting with the host's immune system.

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

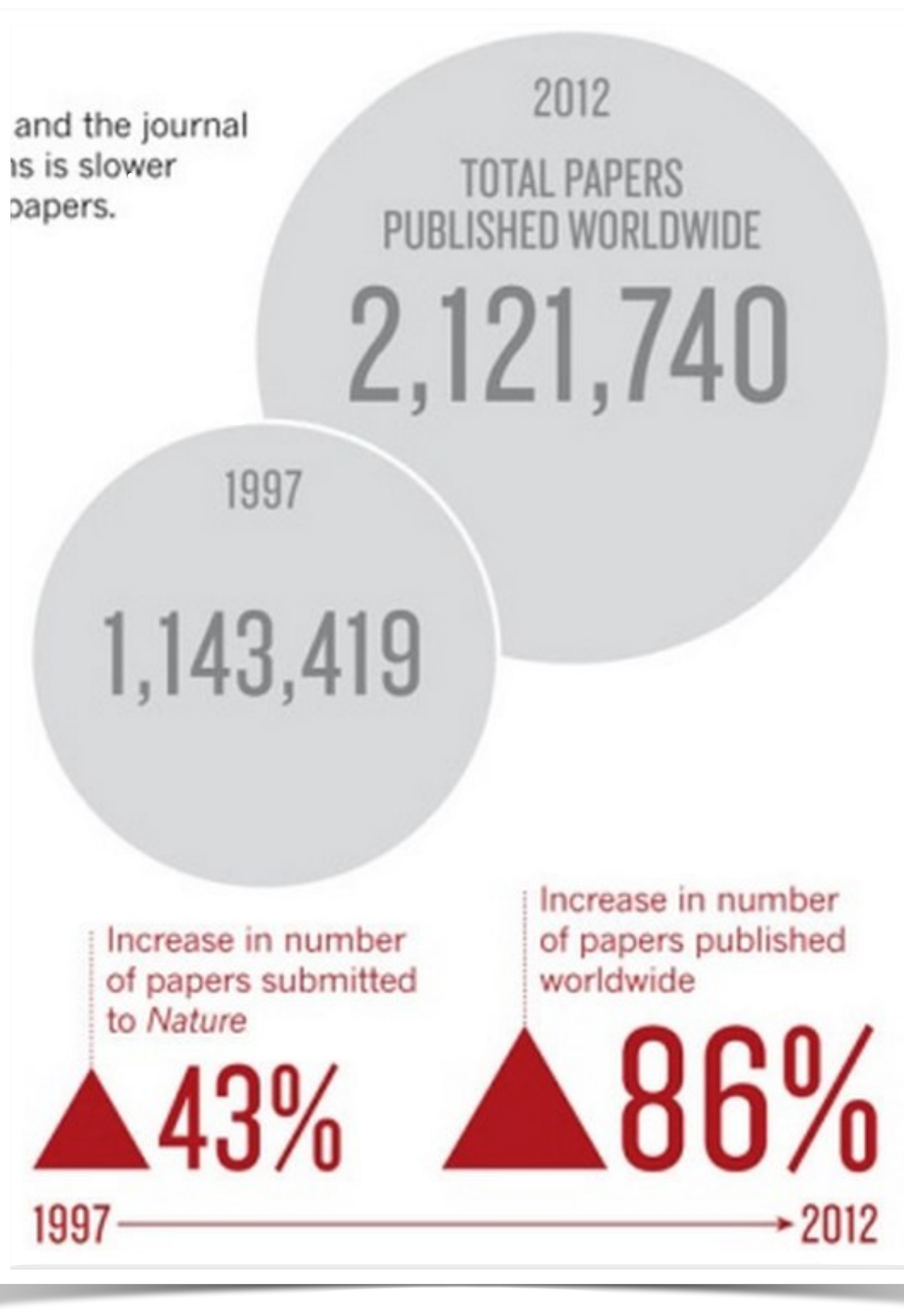
NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing

NEWS AND VIEWS
 A closer look at RNA editing



Science publishing: The golden club

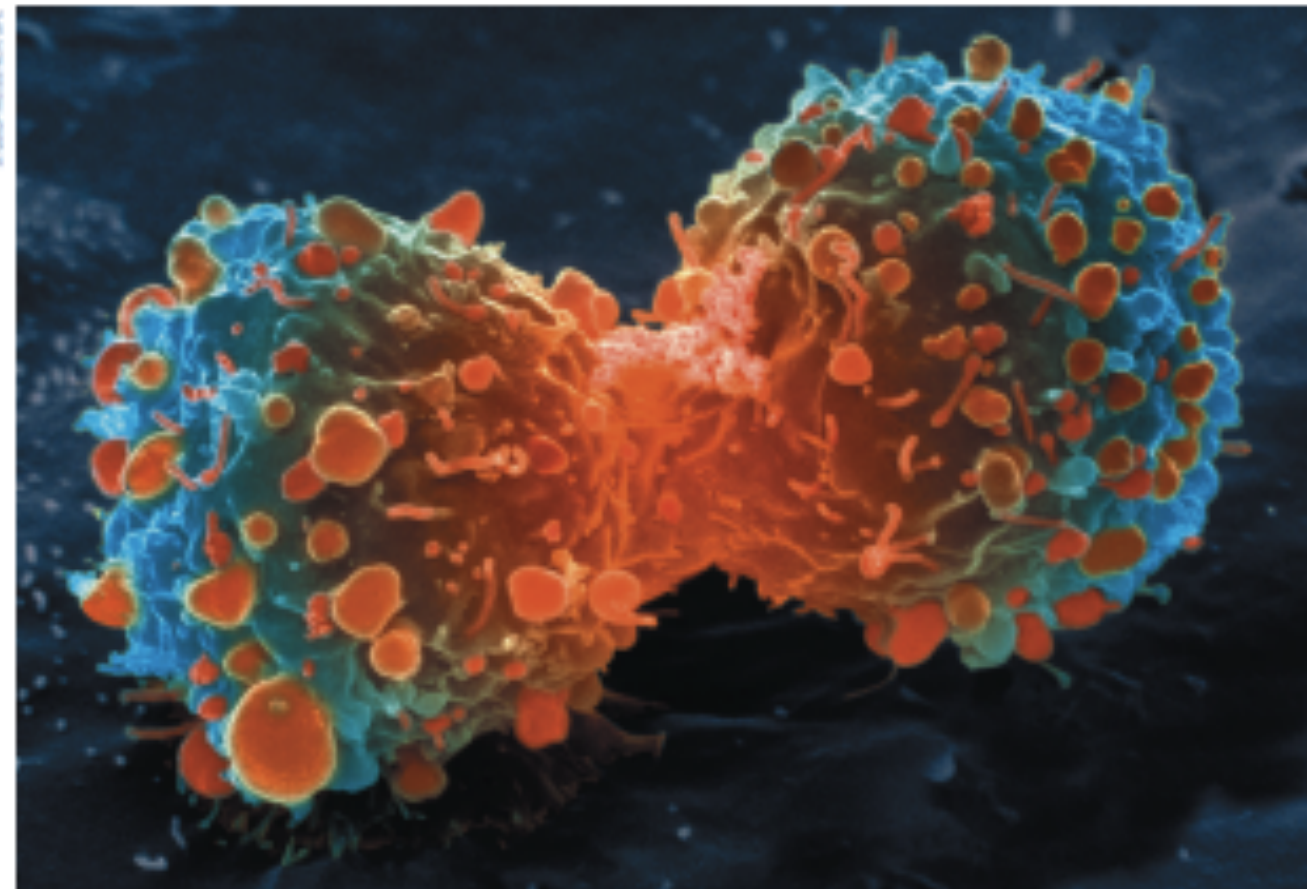
COMMENT

AVIAN INFLUENZA Shift expertise to track mutations where they emerge **p.534**

EARTH SYSTEMS Past climates give valuable clues to future warming **p.537**

HISTORY OF SCIENCE Descartes' lost letter tracked using Google **p.540**

OBITUARY Wylie Vale and an elusive stress hormone **p.542**



Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

Raise standards for preclinical cancer research

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

Efforts over the past decade to characterize the genetic alterations in human cancers have led to a better understanding of molecular drivers of this complex set of diseases. Although we in the cancer field hoped that this would lead to more effective drugs, historically, our ability to translate cancer research to clinical success has been remarkably low. Sadly, clinical

trials in oncology have the highest failure rate compared with other therapeutic areas. Given the high unmet need in oncology, it is understandable that barriers to clinical development may be lower than for other disease areas, and a larger number of drugs with suboptimal preclinical validation will enter oncology trials. However, this low success rate is not sustainable or acceptable, and

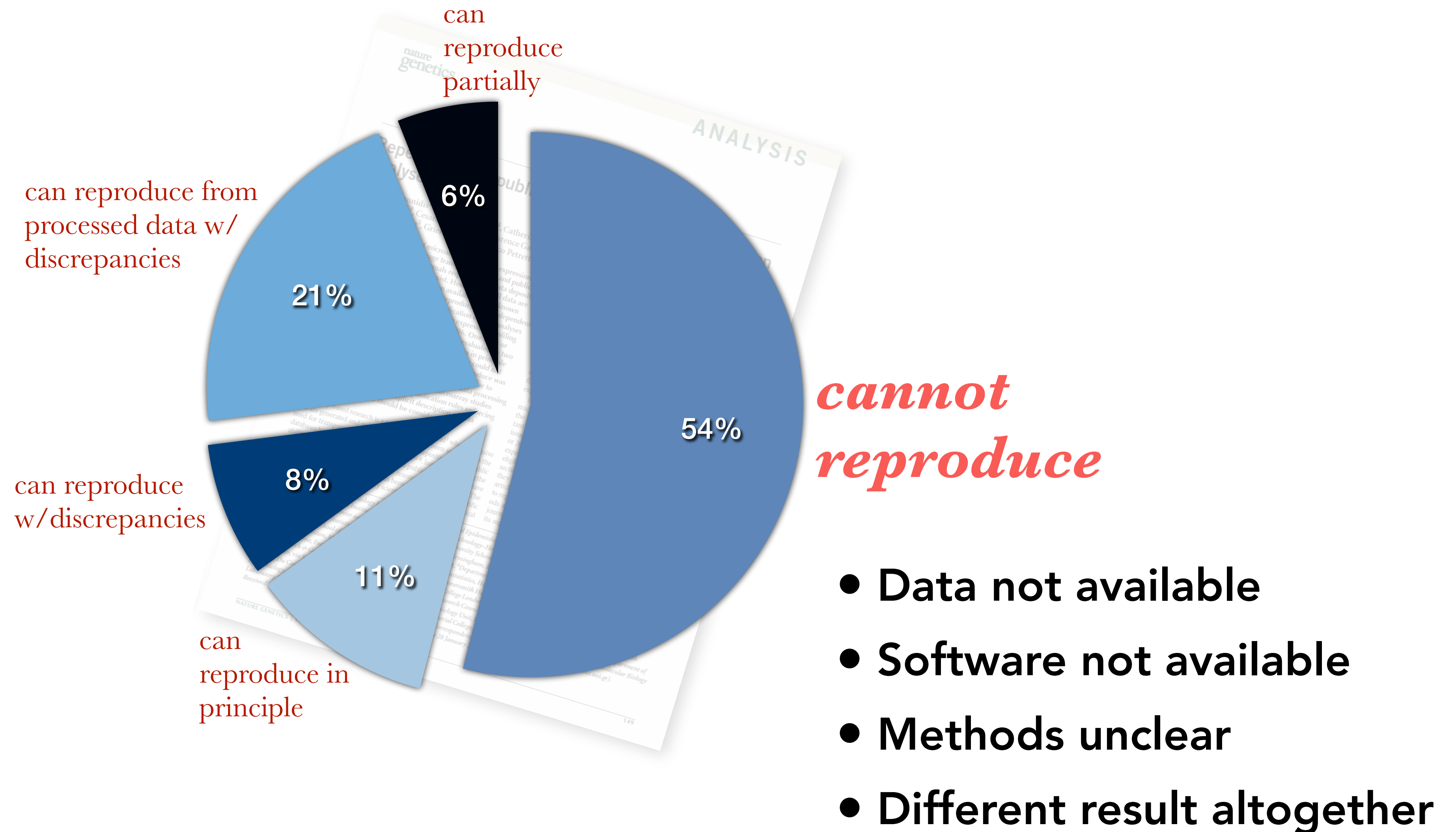
investigators must reassess their approach to translating discovery research into greater clinical success and impact.

Many factors are responsible for the high failure rate, notwithstanding the inherently difficult nature of this disease. Certainly, the limitations of preclinical tools such as inadequate cancer-cell-line and mouse models¹ make it difficult for even ▶

“The biotechnology firm **Amgen**...tried to confirm published findings of **53** 'landmark papers' ... scientific findings were confirmed in only **6 (11%) cases.**”

“**Bayer HealthCare** in Germany last year reported that only about **25%** of published preclinical studies could be **validated** to the point at which projects could continue.”

The status quo tolerates poor communication of findings



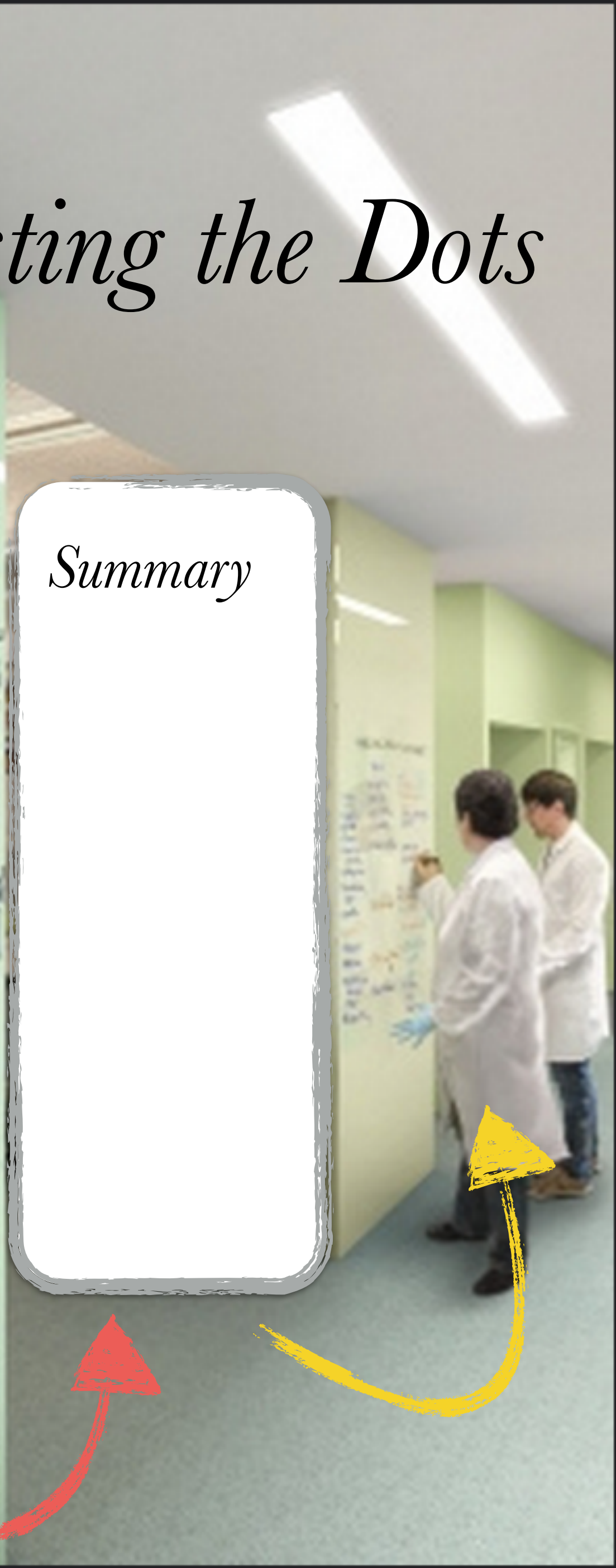
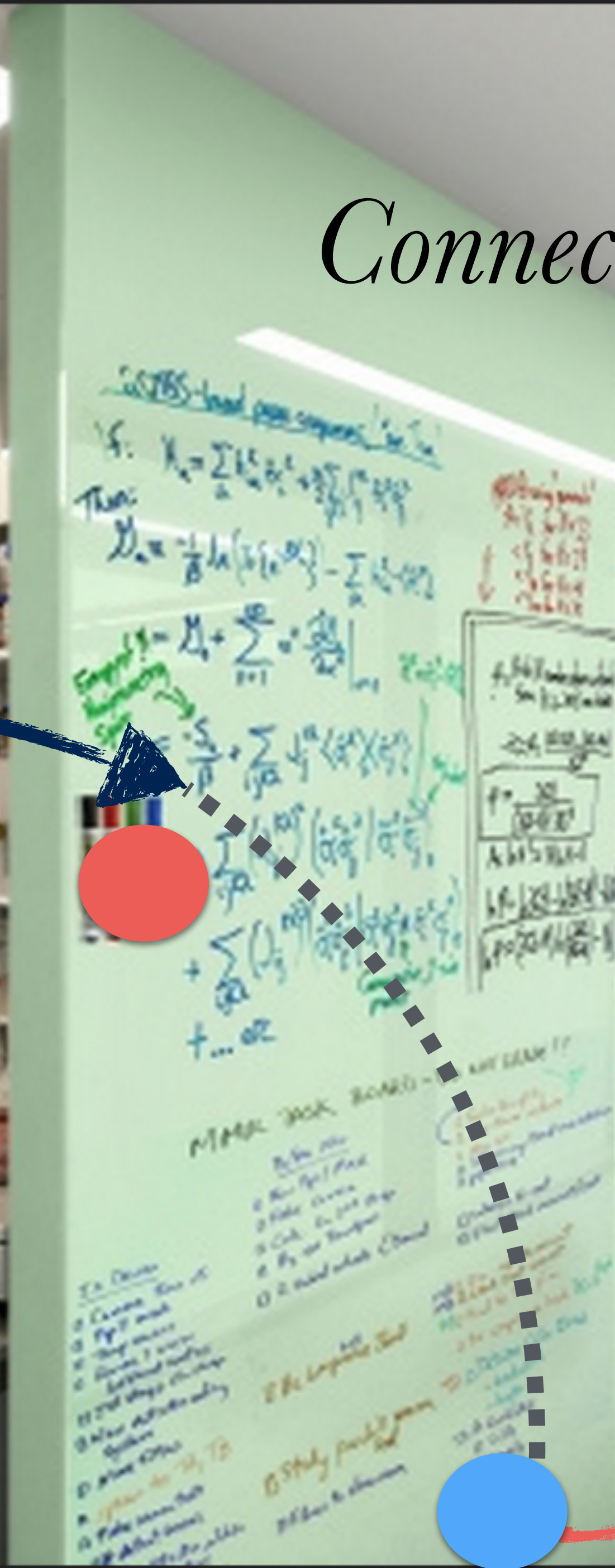
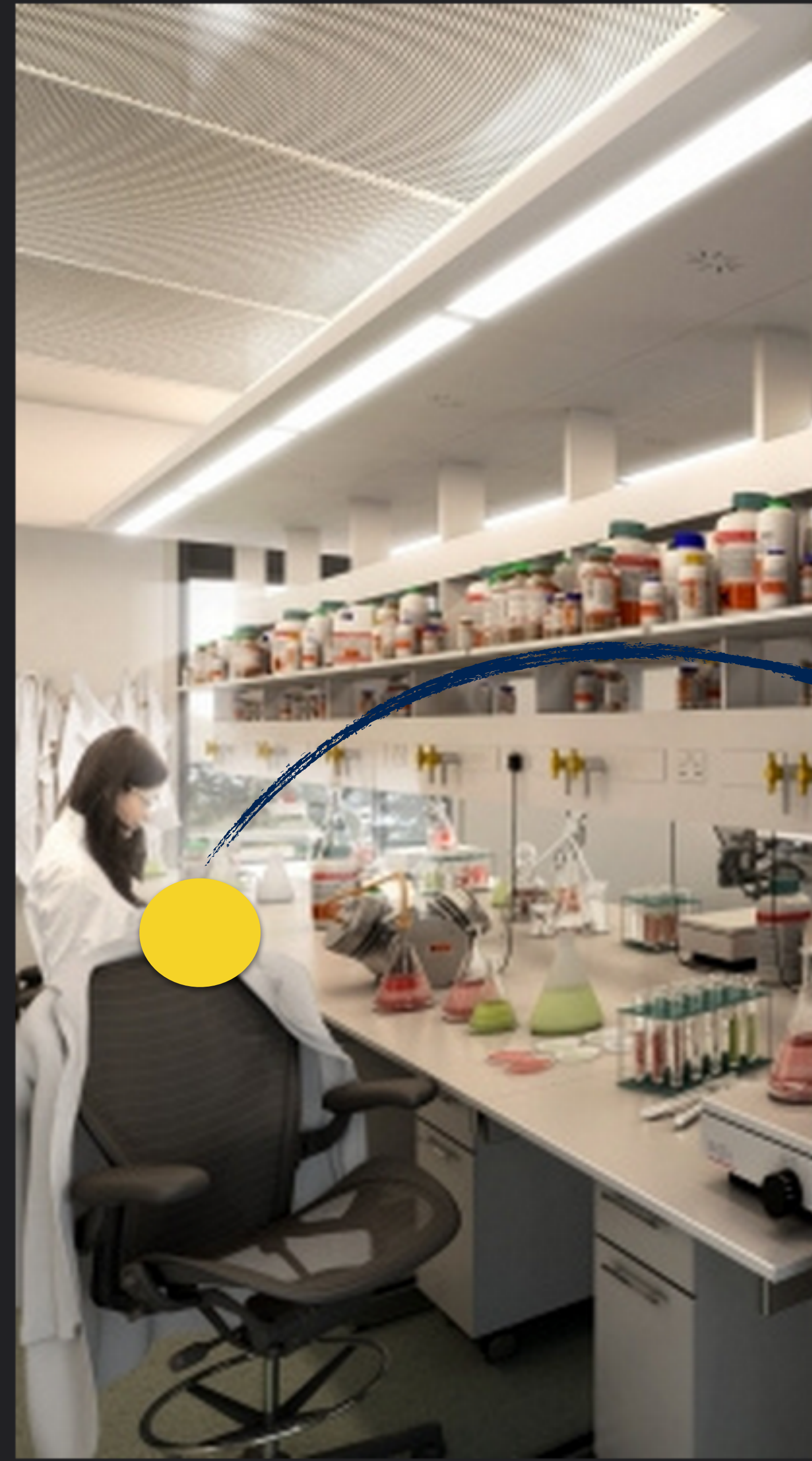
Source

Mixing ratio

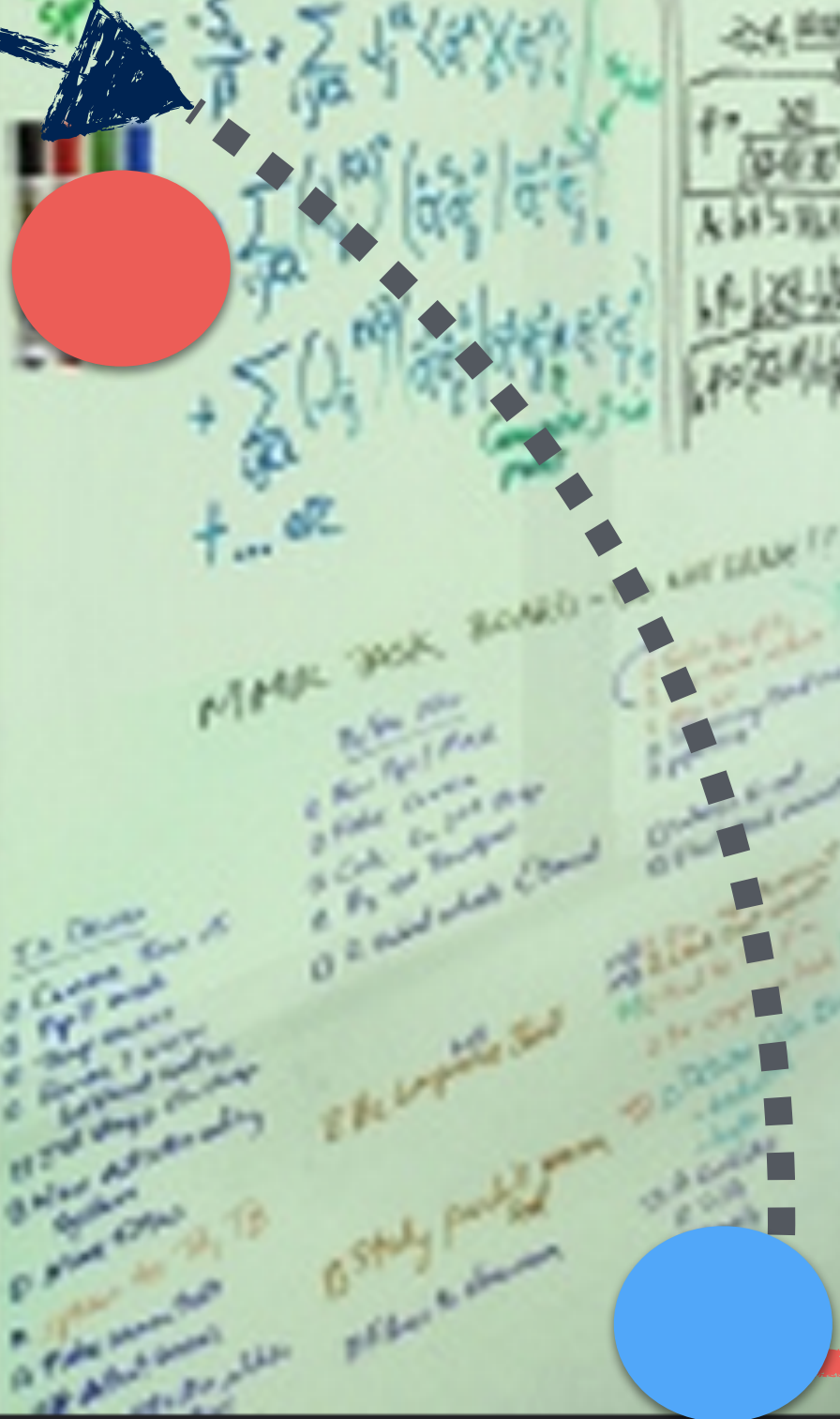
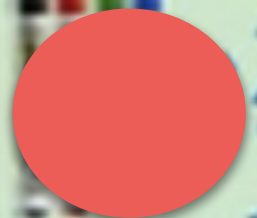
Availability



Connecting the Dots

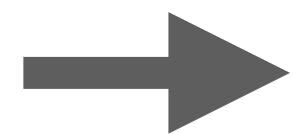


Summary

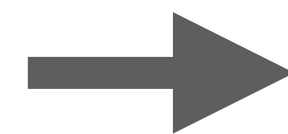


Data

Creation



Curation

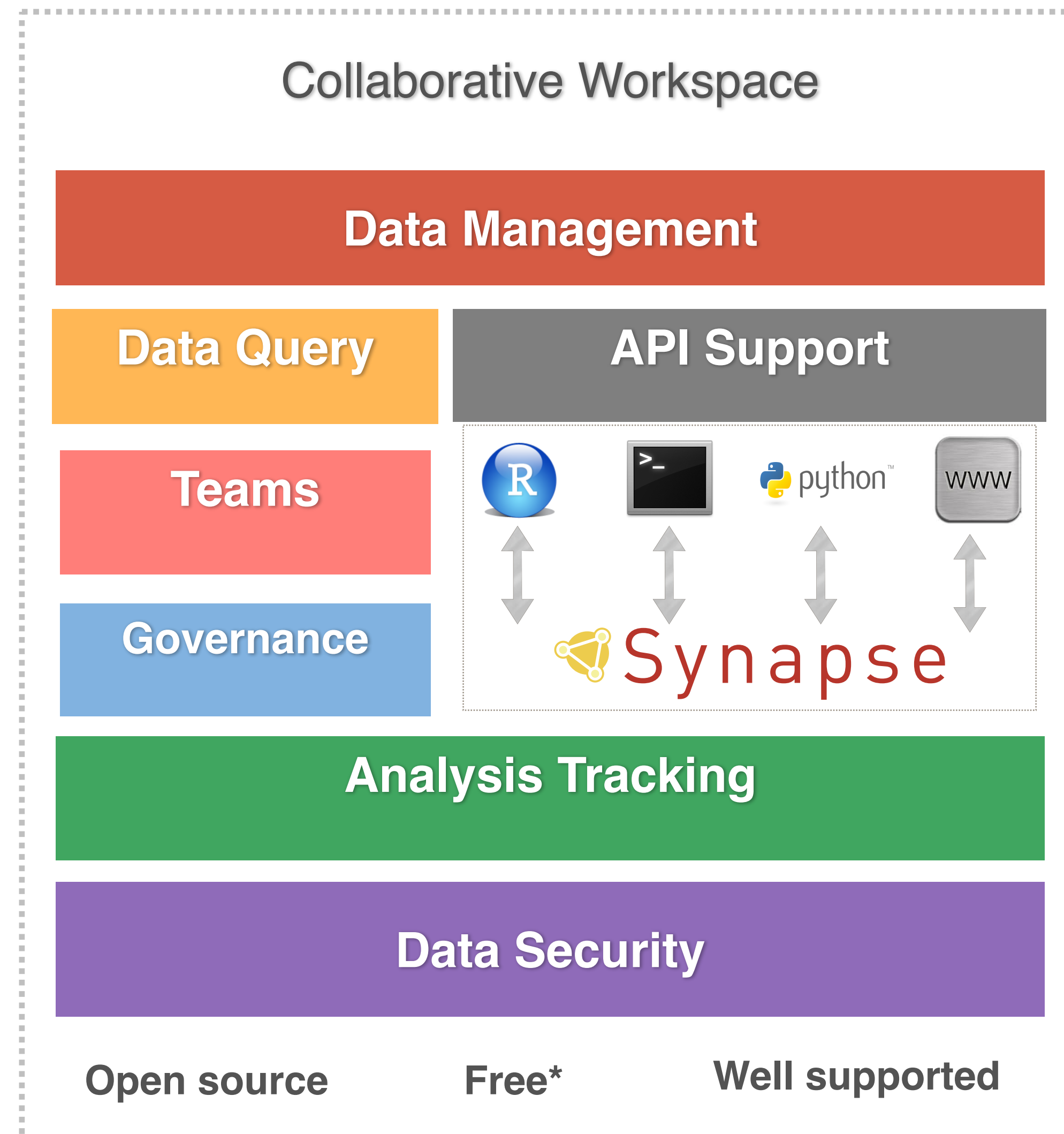


Consumption

www.synapse.org

a tool to improve **transparency** and **reproducibility** of data intensive science by recording analyses in **real-time**

a collection of living research projects **enabling** researchers to contribute to large-scale **collaborative science**



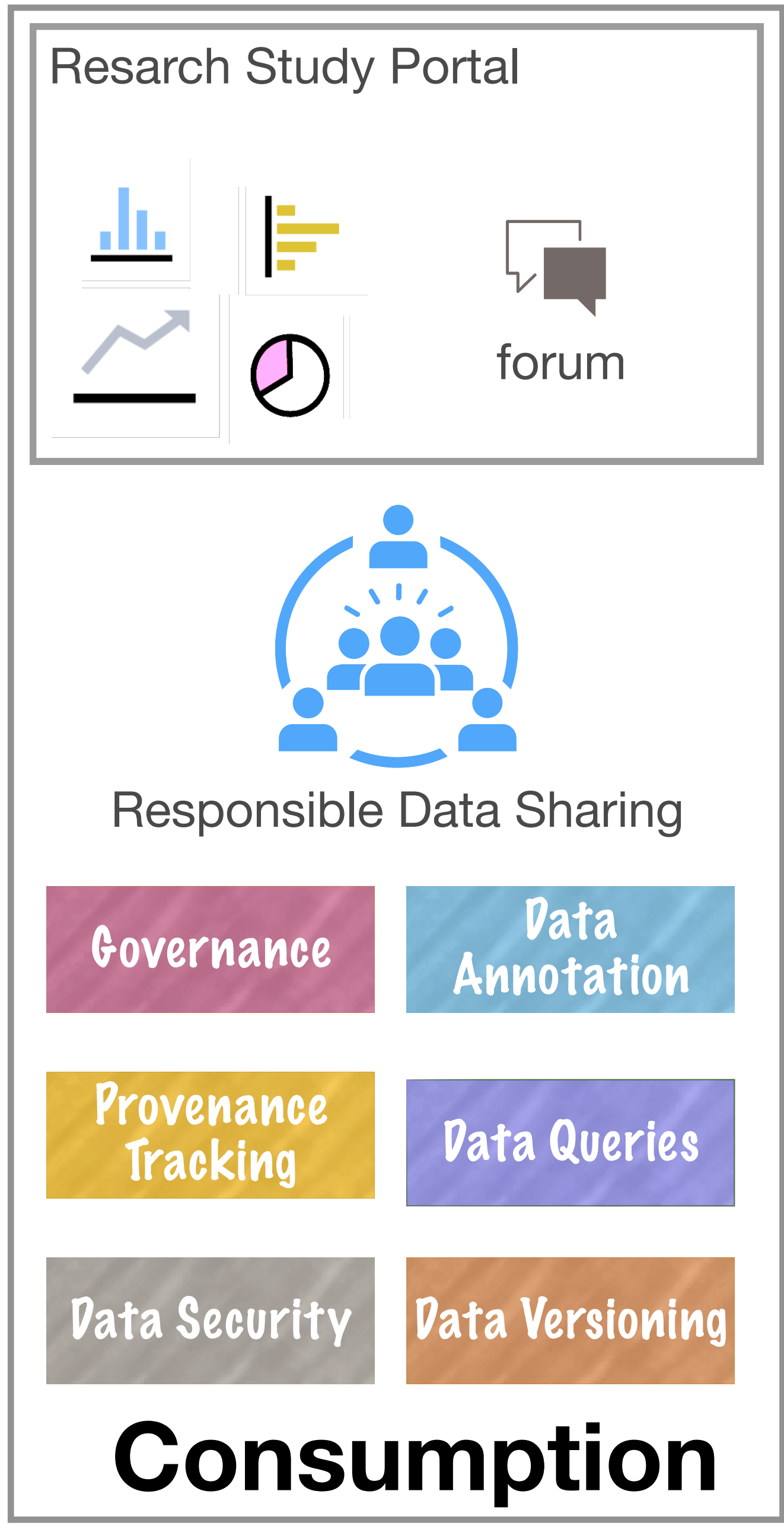
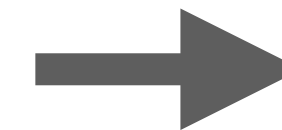
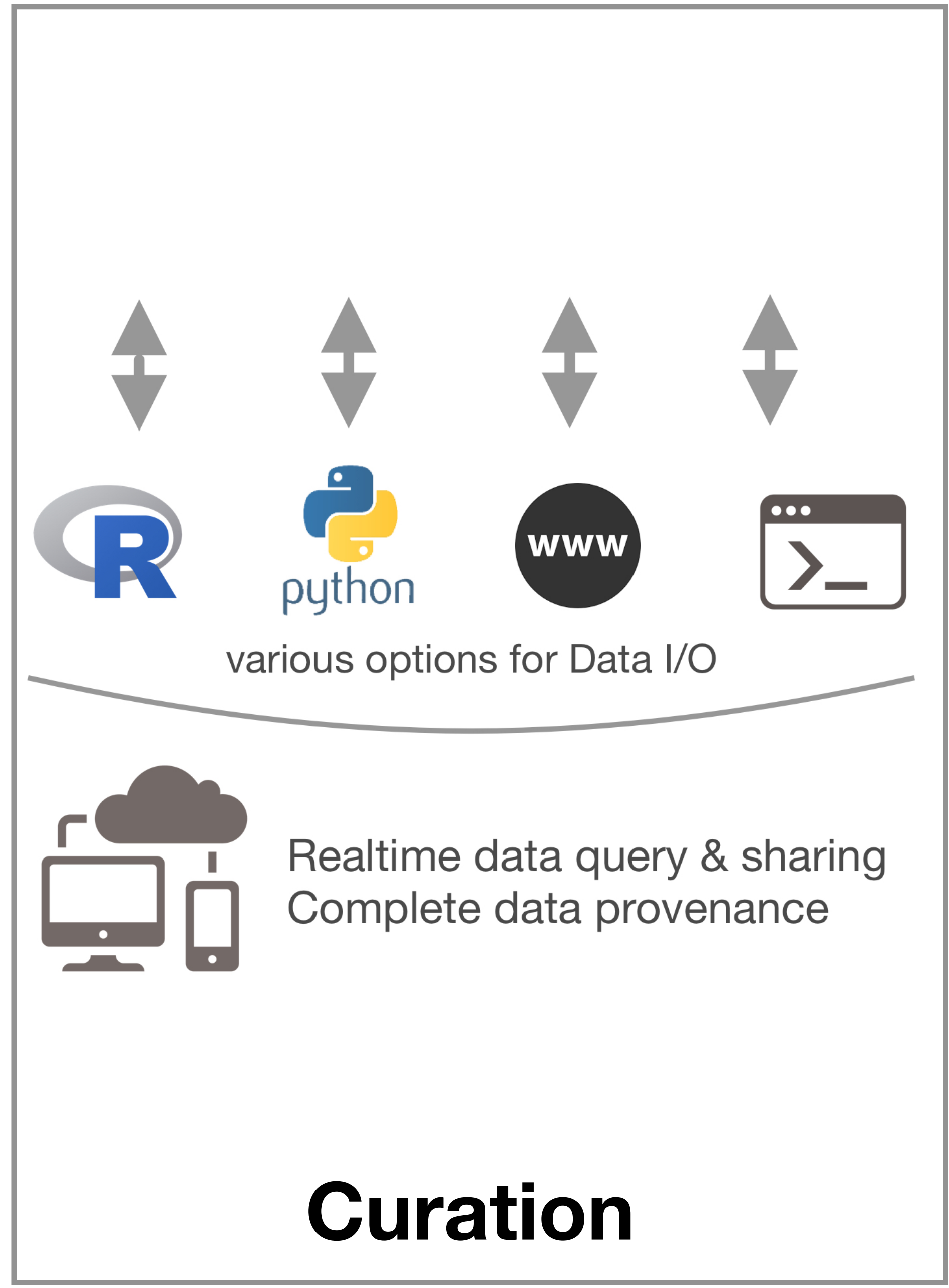
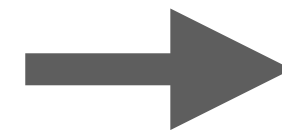
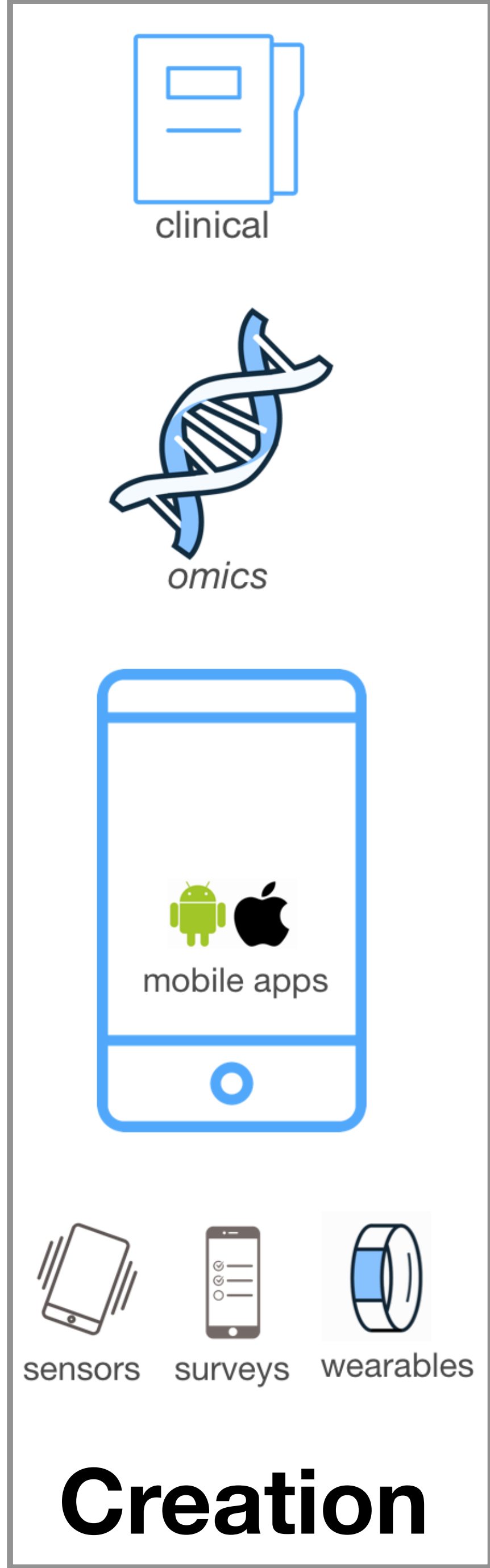
PARTNERS



Many alternatives out there



.....and more





clinical



omics



mobile apps



sensors

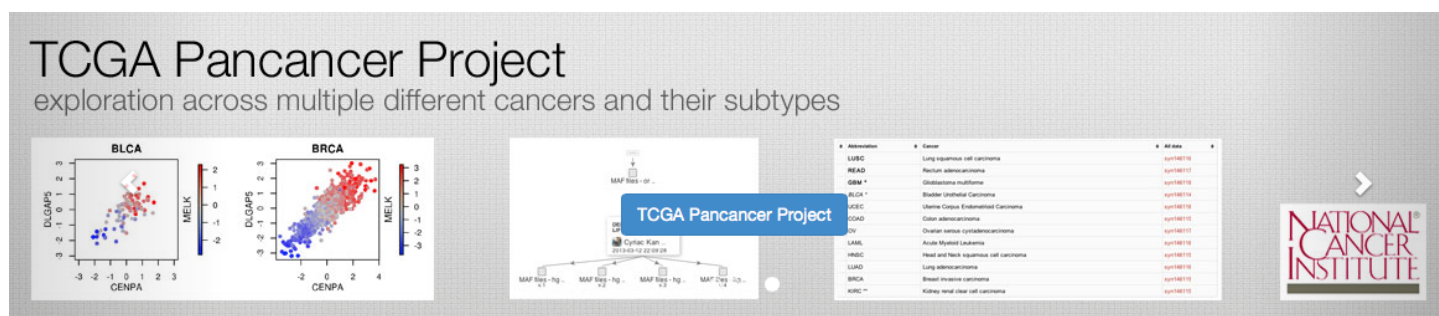


surveys



wearables

Creation



International
Cancer Genome
Consortium



THE NATIONAL CANCER INSTITUTE / NATIONAL INSTITUTES OF HEALTH PRESENTS
UP FOR A CHALLENGE? (U4C)
**STIMULATING INNOVATION IN
BREAST CANCER GENETIC EPIDEMIOLOGY**

(Omics, Clinical data + Model to data approaches)



Parkinsons



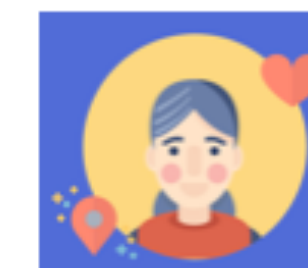
Cardiovascular



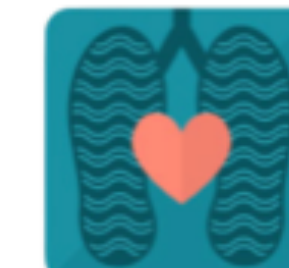
MoleMapper



Asthma



Multiple
Sclerosis



AllofUS
CRF



Samsung
my BP lab

www.synapse.org/digitalhealth

(PROs, sensor-based functional tests, wearables, passive data etc)



clinical



omics



mobile apps



sensors

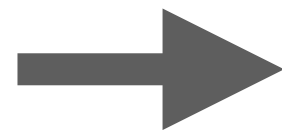


surveys



wearables

Creation



www.synapse.org



python



various options for Data I/O



Realtime data query & sharing
Complete data provenance

Curation



www.synapse.org



various options for Data I/O



Realtime data query & sharing
Complete data provenance

Curation

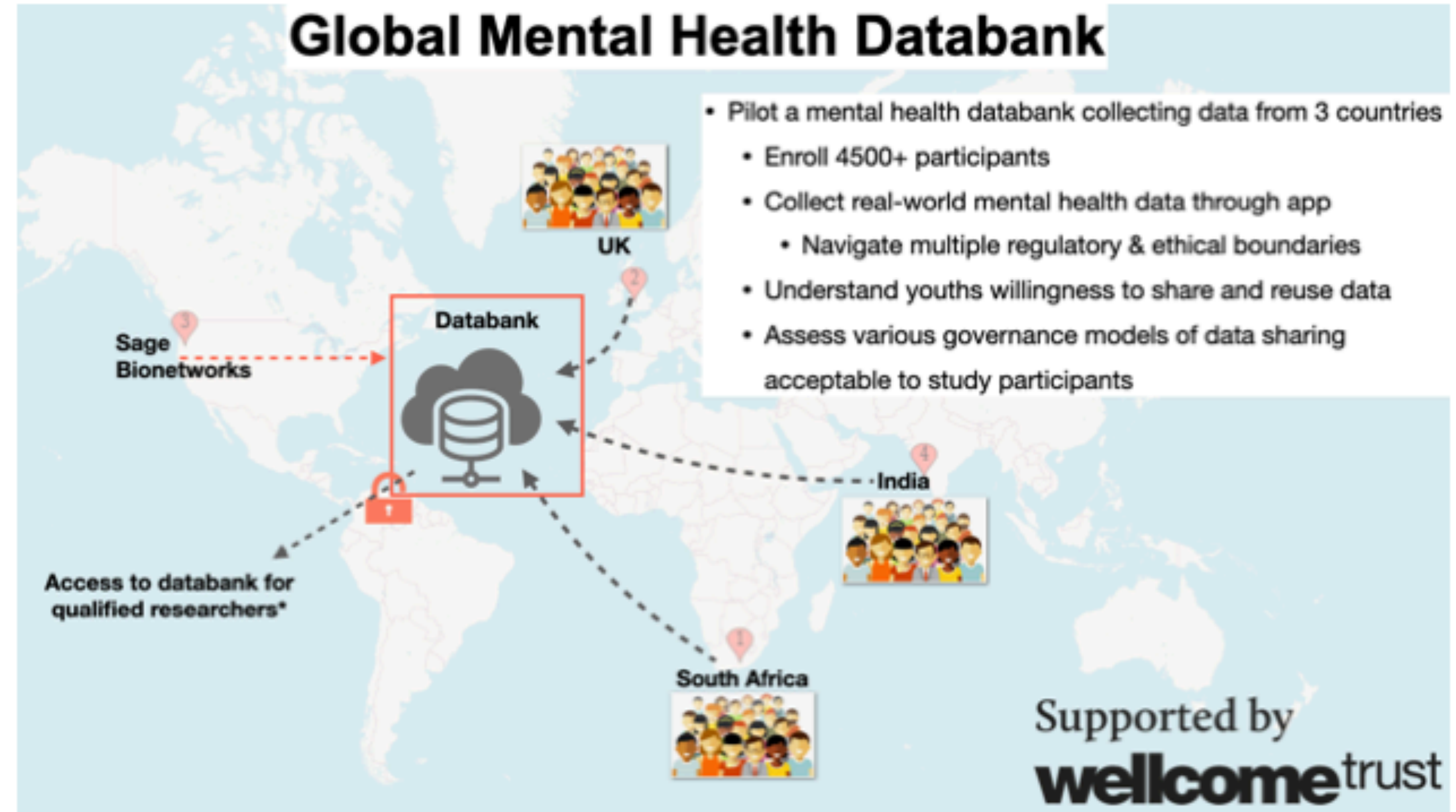
Wellcome GMHD

UniqueID - syn123456

Project

Summary

Global Mental Health Databank



- Pilot a mental health databank collecting data from 3 countries
- Enroll 4500+ participants
- Collect real-world mental health data through app
 - Navigate multiple regulatory & ethical boundaries
- Understand youths willingness to share and reuse data
- Assess various governance models of data sharing acceptable to study participants

Data

Folders

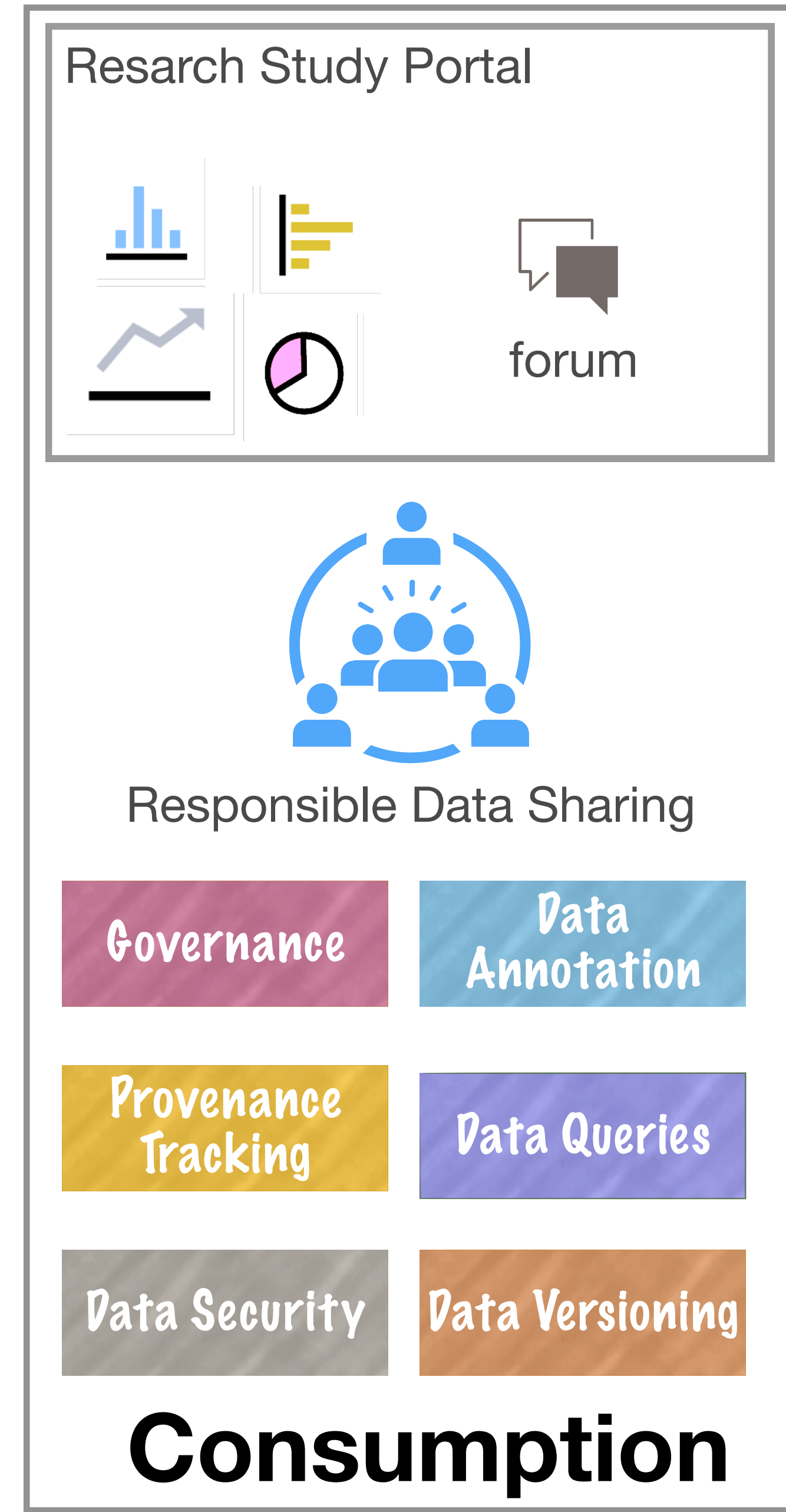
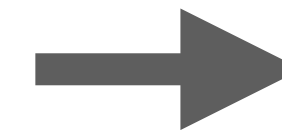
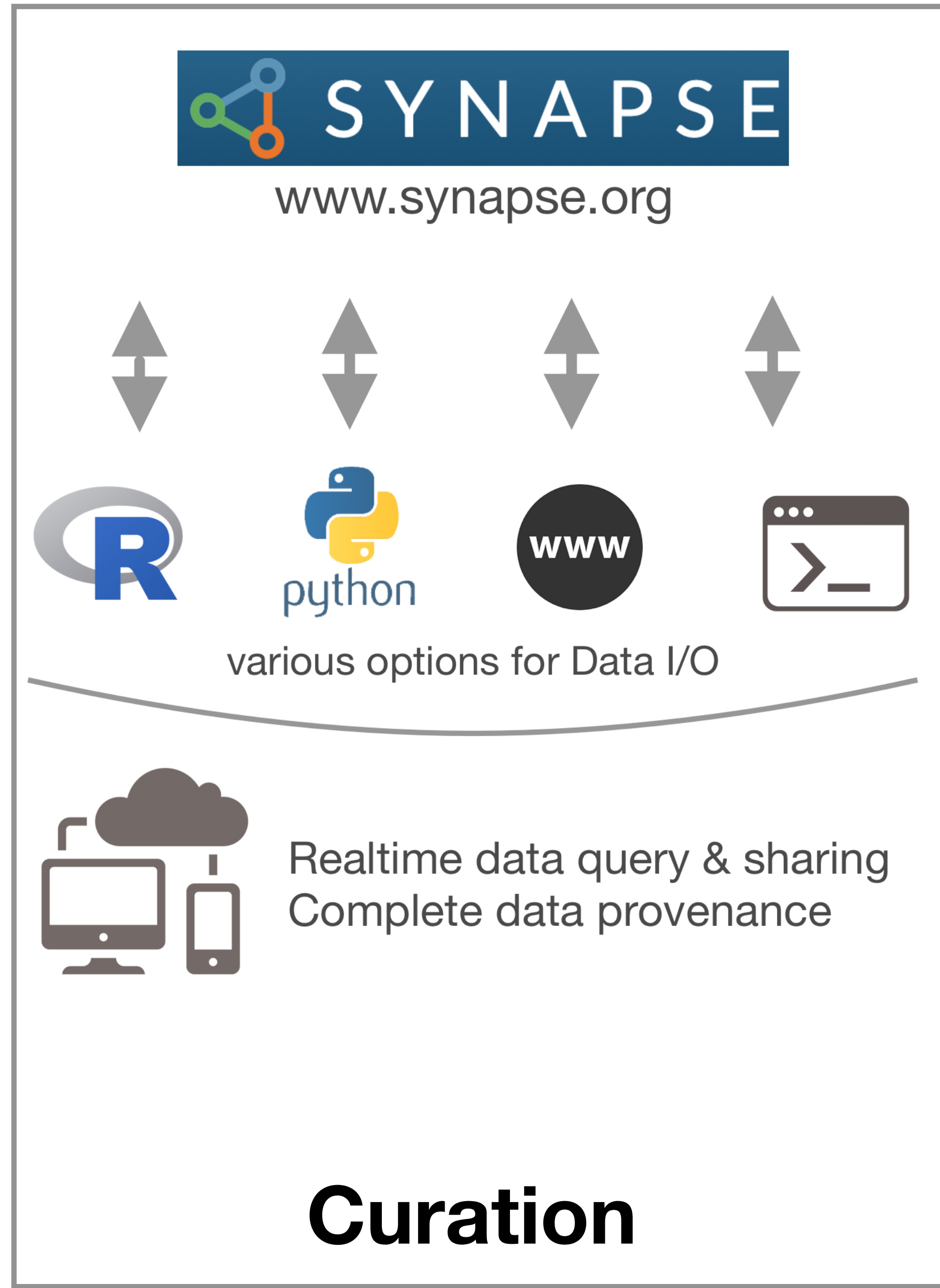
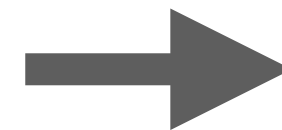
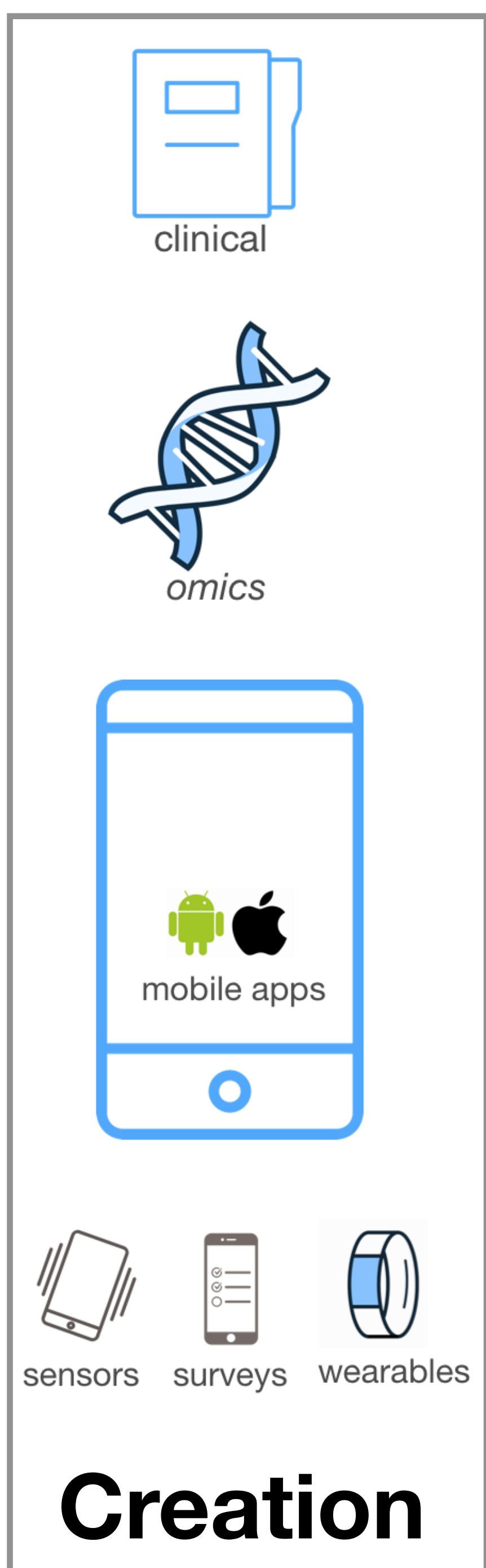
raw data	syn48788
results	syn4785

Files/Tables

PHQ-9	syn85656
Demog	syn47858

Synapse python API demo

Tutorial / Demo : bit.ly/synapse-1



Data Consumption - Example

https://www.synapse.org/brighten

Brighten Study Public Researcher Portal

Wiki Files **Tables** Discussion Docker

Tables » PHQ-9

PHQ-9

Synapse ID: syn18405275 Access: Add Conditions for Use Report Issue

SELECT * FROM syn18405275



participant_id	week	sum_phq9	phq9Date	phq9_1	phq9_2	phq9_3	phq9_4	phq9_5	phq9_6	phq9_7	phq9_8	phq9_9
BLUE-00048	1	7	2014-08-08	0	1	1	2	1	0	0	0	0
BLUE-00050	2	7	2014-08-23	1	1	1	1	1	1	1	0	0
BLUE-00050	3	4	2014-08-30	1	0	1	1	1	0	0	0	0
BLUE-00050	4	5	2014-09-08	1	1	1	1	1	0	0	0	0
BLUE-00050	6	5	2014-09-22	1	1	1	1	1	0	0	0	0
BLUE-00050	8	8	2014-10-05	1	1	1	1	2	1	1	0	0
BLUE-00050	10	5	2014-10-19	0	1	1	1	1	1	1	0	0




Access Options






- Web UI
- R
- Python
- ...more...



Data Consumption - Example 1


https://www.synapse.org/mpower


 mPower Public Researcher Portal 


Synapse ID: syn4993293  DOI: 10.7303/syn4993293 Storage Location: Synapse Storage  Project Settings 


Wiki  Files  Tables  Discussion  Docker 

 Wiki Tools 

mPower Public Researcher Portal 

- 1 - Accessing the mPower data
- 2 - Data description 
- 3 - mPower research community
- 4 - Publications
- 5 - FAQs
- 6 - Contact us

Edit Order 



mPower: Mobile Parkinson Disease Study



[Data description](#) [Accessing the data](#) [Research community](#) [FAQs](#) [Publications](#)




OVERVIEW

Welcome to the mPower Public Researcher Portal, the data access point for the mPower mobile Parkinson Disease study. mPower is an mobile application-based study piloting new approaches to monitoring key indicators of Parkinson Disease progression and diagnosis by supplementing traditional behavioral symptom measurements with novel metrics gleaned from sensor-rich mobile devices. As a scalable, inexpensive, and non-invasive method for frequent measurement and tracking of symptoms, the [Parkinson mPower app](#) has been able to survey a large, longitudinal cohort of volunteers with PD and controls. The goal of this study is to understand the frequency and degree of variation of patient symptoms, the sources of those variations, and the potential modulators of those variations. The insights gained through these data may help develop personalized interventions to mitigate the dynamic variability in the disease.



Data Consumption - Example 1


https://www.synapse.org/mpower


 mPower Public Researcher Portal 



Synapse ID: syn4993293  DOI: 10.7303/syn4993293 Storage Location: Synapse Storage  Project Settings 


[Wiki !\[\]\(eedf39744dbed0fff689e599c7c73f71_img.jpg\)](#) [Files !\[\]\(d299ba3f0ed3478fa8cf5fa846bbd4e6_img.jpg\)](#) [Tables !\[\]\(e4e88b6a7edbea288a42dec204de976d_img.jpg\)](#) [Discussion !\[\]\(b9af56f51a895a6022f70a41d80f0562_img.jpg\)](#) [Docker !\[\]\(6b34e0dcbb526bd4d791aa9cca595801_img.jpg\)](#)

 Wiki Tools 

mPower Public Researcher Portal 

- 1 - Accessing the mPower data
- 2 - Data description 
- 3 - mPower research community
- 4 - Publications
- 5 - FAQs
- 6 - Contact us



mPower: Mobile Parkinson Disease Study

[Data description](#) [Accessing the data](#) [Research community](#) [FAQs](#) [Publications](#)

OVERVIEW

Welcome to the mPower Public Researcher Portal, the data access point for the mPower mobile Parkinson Disease study. mPower is an mobile application-based study piloting new approaches to monitoring key indicators of Parkinson Disease progression and diagnosis by supplementing traditional behavioral symptom measurements with novel metrics gleaned from sensor-rich mobile devices. As a scalable, inexpensive, and non-invasive method for frequent measurement and tracking of symptoms, the [Parkinson mPower app](#) has been able to survey a large, longitudinal cohort of volunteers with PD and controls. The goal of this study is to understand the frequency and degree of variation of patient symptoms, the sources of those variations, and the potential modulators of those variations. The insights gained through these data may help develop personalized interventions to mitigate the dynamic variability in the disease.



 Project Settings 

Wiki 

Files 

Tables 

Discussion 

Docker 

Tables » Sample Walking Activity



Download Options 

 Table Tools 

Sample Walking Activity

Synapse ID: syn5713119 





Access: 

 Report Issue 

SELECT * FROM syn5713119

Query



recordId	healthCode	createdOn	appVersion	phoneInfo	accel_walking_outbound.json.items
0dcb75f3-4b4c-4ede-884c-e04a1a8...	9a3b504e-624f-46f7-abf5-fea091d4...	03/10/2015 10:06 AM	version 1.0, build 7	iPhone 6	 accel_walking_outbound.json.items-97685ee9-3eb0-43e5-9879-ce5b31e0371b8296129926155346845.tmp
bf5ad14c-7c32-421d-8b10-53906bf...	8746d33b-b03f-4879-a0ad-13cc2f9...	03/12/2015 1:20 PM	version 1.0, build 7	iPhone 6 Plus	 accel_walking_outbound.json.items-96b7d33b-6baa-495a-83ca-caf86b4023987371500467451125723.tmp
9a042f44-c47b-4bcc-ba8d-041c5cb...	5bee7cd6-e984-4867-b618-ad9646...	03/13/2015 9:34 AM	version 1.0, build 7	iPhone 6	 accel_walking_outbound.json.items-2f6c5f88-e048-4290-b9aa-d29773ea487e4269585026506265516.tmp
4ab5ea3e-fbfb-4fdb-ad9c-42aacca1...	e26facdc-f8bc-4be2-969a-876a063...	03/14/2015 6:59 AM	version 1.0, build 7	iPhone 6	 accel_walking_outbound.json.items-bcaab5b5-882d-4972-8123-34e5fa447c3a864091809024668349.tmp

Data Consumption - Example 2

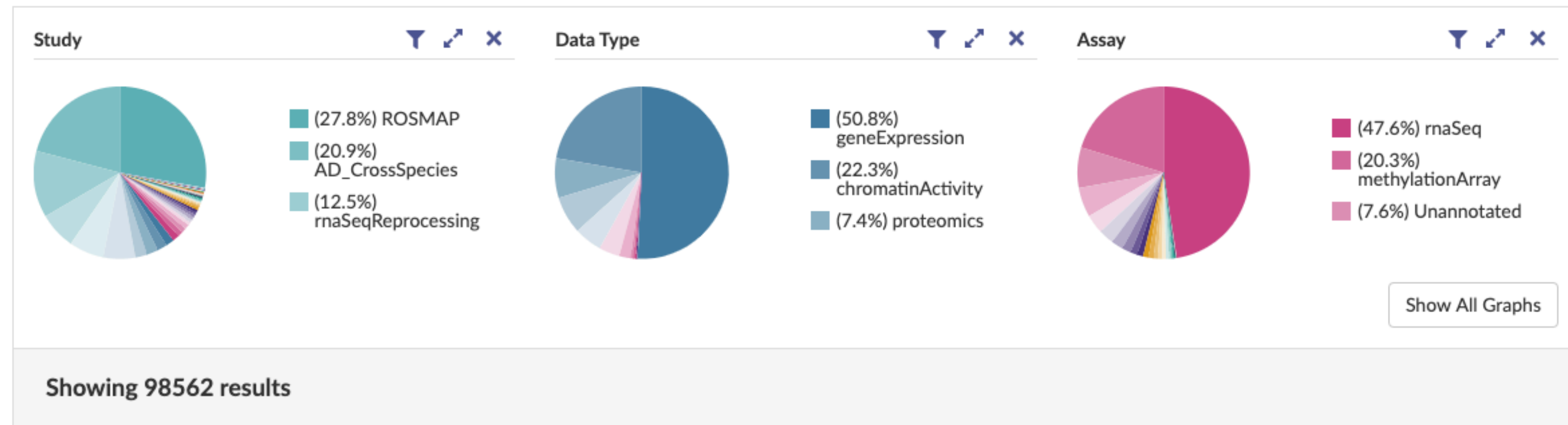
https://adknowledgeportal.synapse.org/Explore/Data



EXPLORE

- PROGRAMS
- PROJECTS
- STUDIES
- DATA**
- PUBLICATIONS
- PEOPLE
- EXPERIMENTAL TOOLS
- COMPUTATIONAL TOOLS
- RESULTS

Data (98,562)



Showing 98562 results

Study	Access	Id	Study	Data Type	Assay	Organ	Tis
<input checked="" type="checkbox"/> All	Request Access	chr1.chop.dosage.gz	ROSMAP	genomicVariants	snpArray		
<input type="checkbox"/> ROSMAP 27413	Request Access	chr2.chop.dosage.gz	ROSMAP	genomicVariants	snpArray		
<input type="checkbox"/> AD_CrossSpecies 20630	Request Access	chr4.chop.dosage.gz	ROSMAP	genomicVariants	snpArray		
<input type="checkbox"/> rnaSeqReprocessing 12326	Request Access	chr8.chop.dosage.gz	ROSMAP	genomicVariants	snpArray		
<input type="checkbox"/>	Request Access	chr9.chop.dosage.gz	ROSMAP	genomicVariants	snpArray		
<input type="checkbox"/>	Request Access	chr6.chop.dosage.gz	ROSMAP	genomicVariants	snpArray		



Welcome to dHealth

A Digital Health Knowledge Portal to enable the discovery and download of digital and mobile health data, tools, and benchmarked outcomes and digital biomarkers.

Supported by SAGE BIONETWORKS



Collections

10

Studies leveraging digital health technologies, as well as analytical efforts and benchmarking challenges.

EXPLORE

Data

124

Raw and processed data from the studies catalogued in this portal.

EXPLORE

Tools

1

Computational resources for the processing and analysis of digital health data.

EXPLORE

Publications

27

Lessons learned and models built from digital health data, documented in peer-reviewed journal articles.

EXPLORE

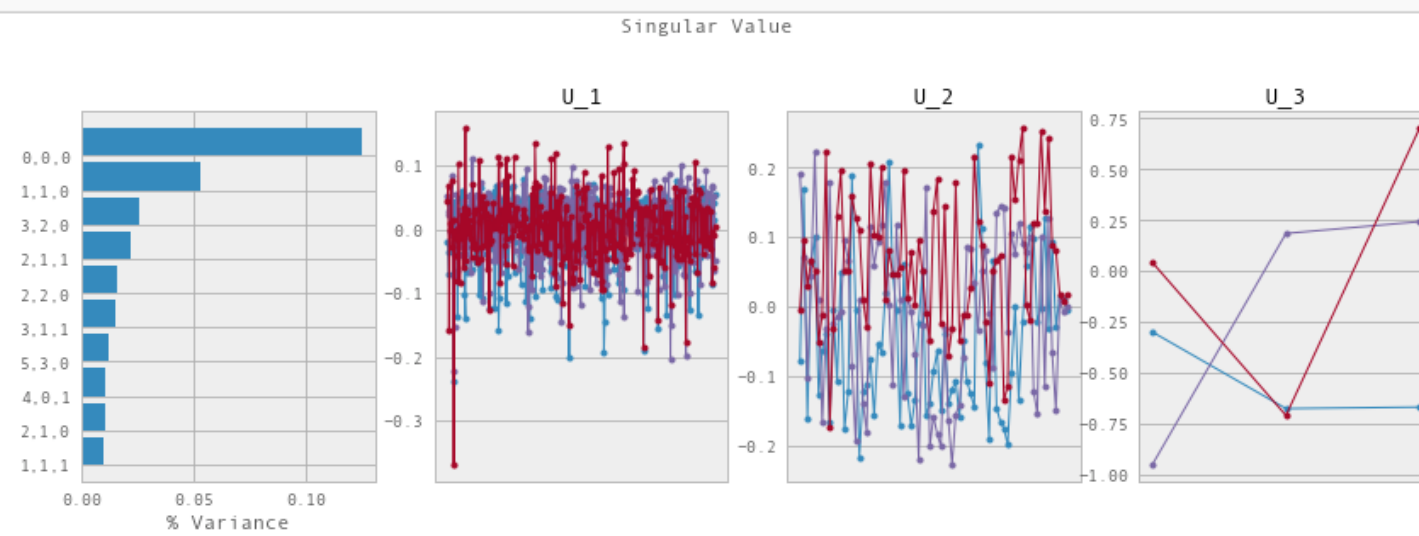
Tractable Data Analysis

tremor_subtyping.ipynb [Download](#) [other download options](#)

Download file as: tremor_HOSVD.ipynb | 2.102 MB | Synapse Storage [md5](#)

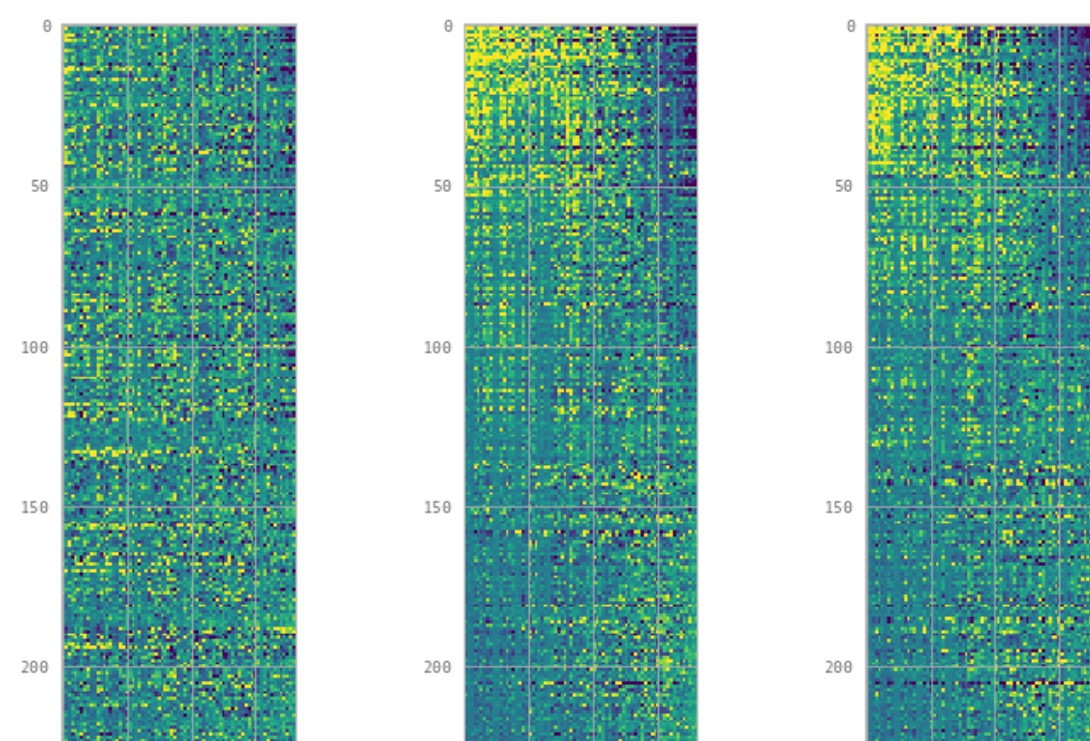
Synapse ID: syn12169237 [Conditions for use: None \(change\)](#) [report issue](#)

Preview



```
In [8]: ax = plt.figure(figsize=(12,12));
idx1 = np.argsort(Un[0][:,0])
idx2 = np.argsort(Un[1][:,0])
plt.subplot(1,3,1);
plt.imshow(T[idx1,:,0][:,idx2],vmin=-2, vmax=2)
plt.subplot(1,3,2);
plt.imshow(T[idx1,:,1][:,idx2],vmin=-2, vmax=2)
plt.subplot(1,3,3);
plt.imshow(T[idx1,:,2][:,idx2],vmin=-2, vmax=2)
```

Out[8]: <matplotlib.image.AxesImage at 0x1121ce9d0>



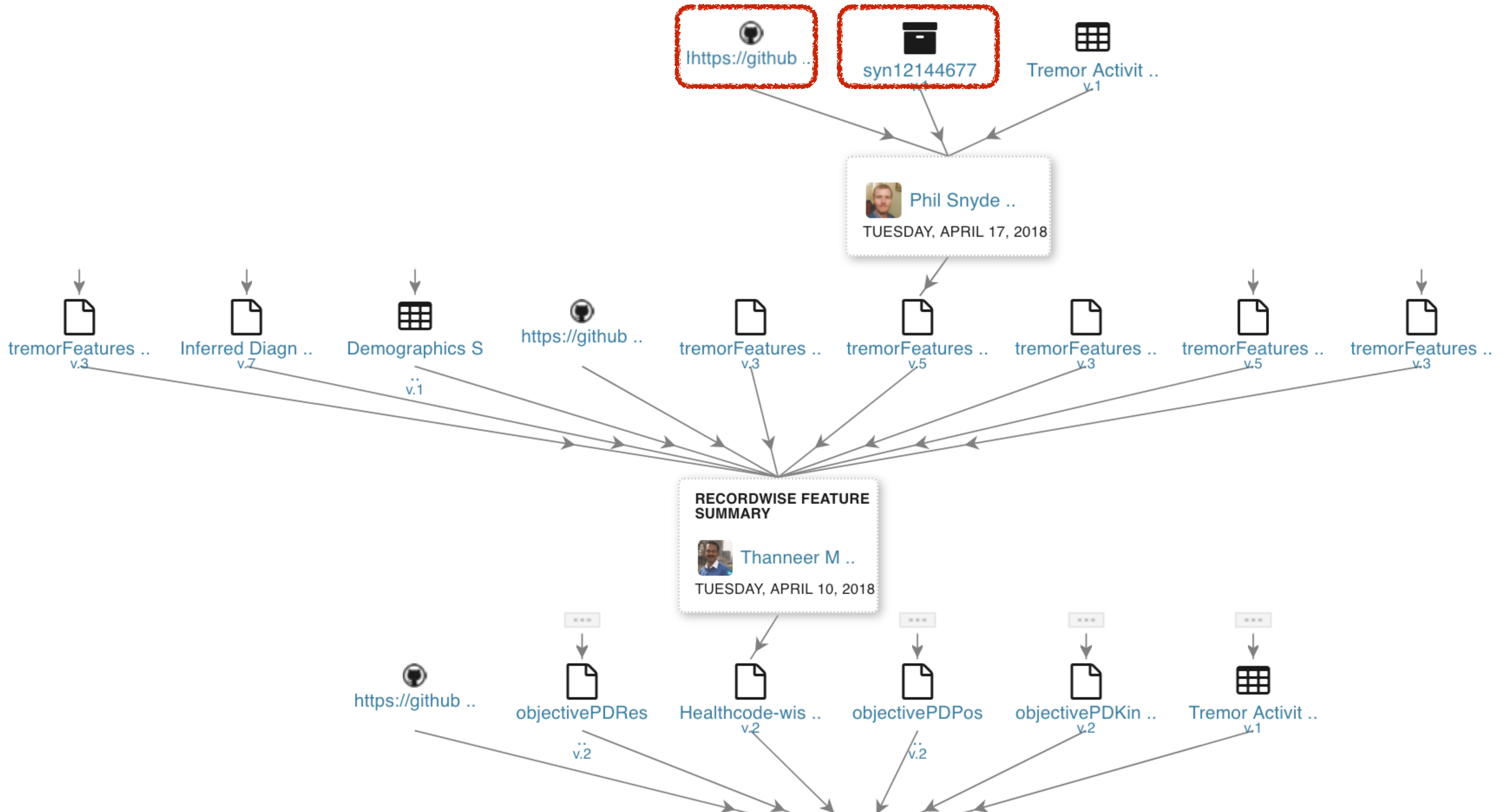
Provenance



open collaborative methodologies



open collaborative methodologies



open collaborative methodologies



reusable



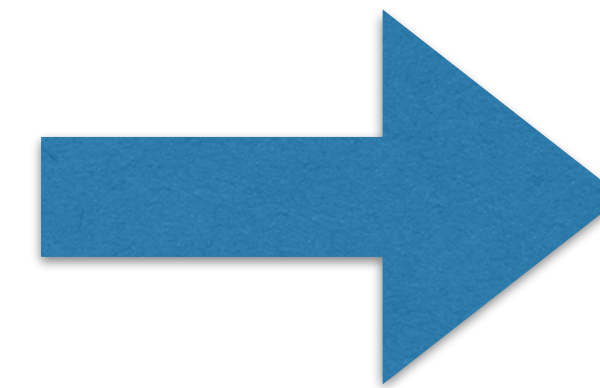
[https://github](https://github.com)



syn12144677

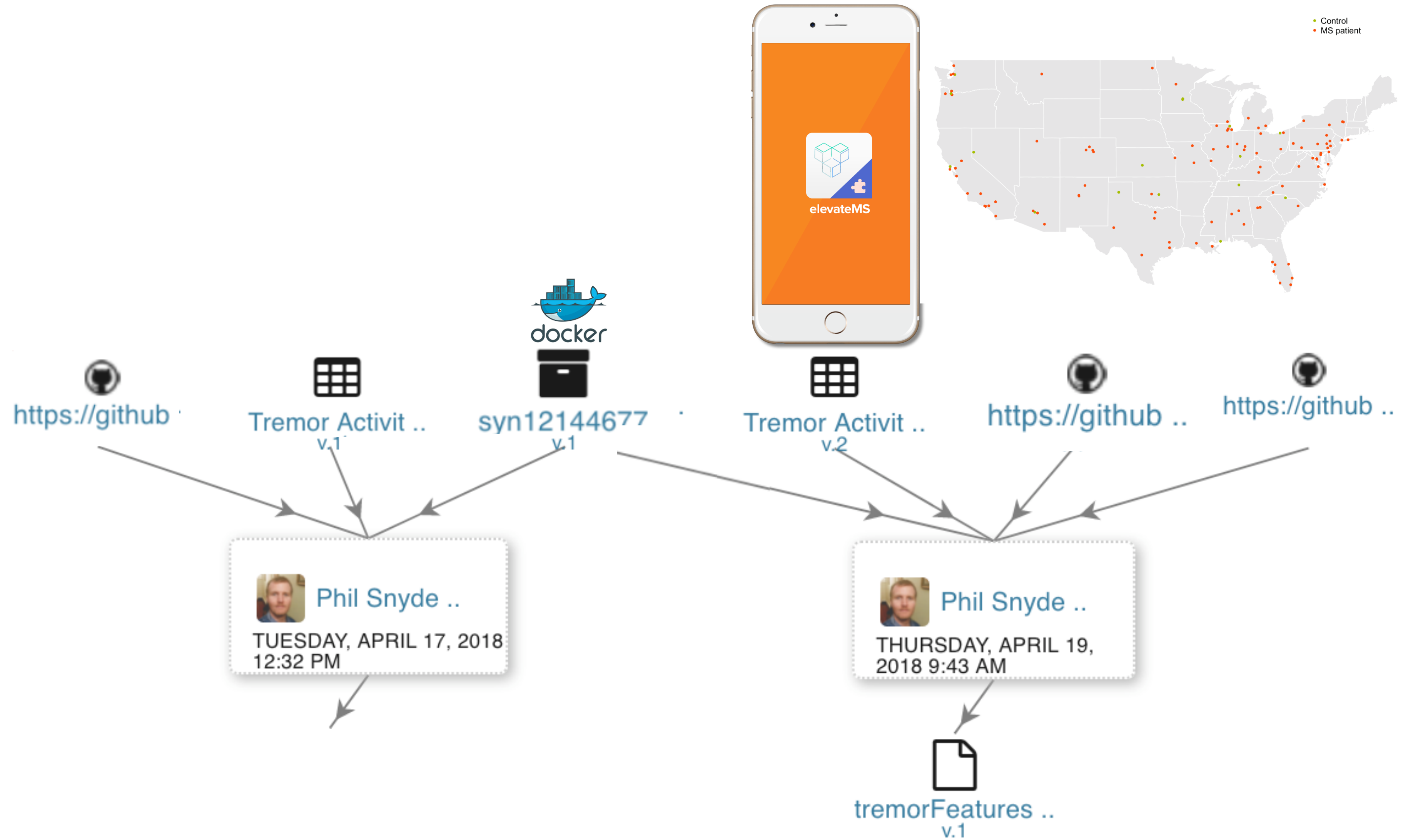
 Sage-Bionetworks / mhealthtools

```
10 process_medicationChoiceAnswers <- function(  
11   tryCatch({  
12     d <- jsonlite::fromJSON(json_file)  
13     data.frame(medication = paste(unique(d$medication),  
14       collapse = "+"), medicationTime = paste(unique(d$medicationTime),  
15       collapse = "+"))  
16   }, error = function(err) {  
17     data.frame(medication = "NA", medicationTime = "NA")  
18   })  
19 }
```



Cloud enabled workflow

reusable



Tractable Analysis

bit.ly/synapse-2



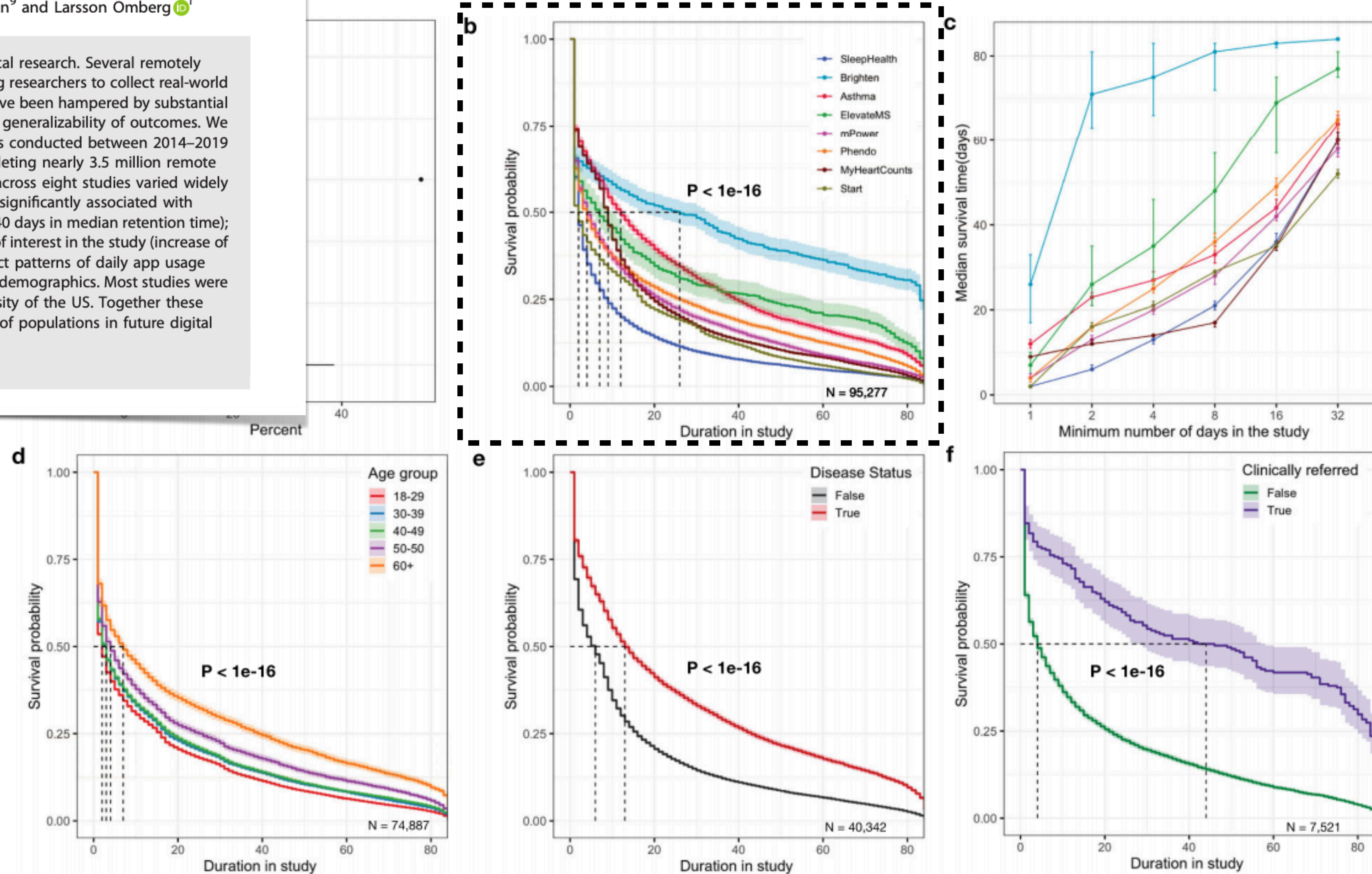
ARTICLE OPEN
Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants

Abhishek Pratap^{1,2}, Elias Chaibub Neto¹, Phil Snyder¹, Carl Stepnowsky^{3,4}, Noémie Elhadad⁵, Daniel Grant⁶, Matthew H. Mohebbi⁷, Sean Mooney², Christine Suver¹, John Wilbanks¹, Lara Mangravite¹, Patrick J. Heagerty⁸, Pat Areán⁹ and Larsson Omberg¹

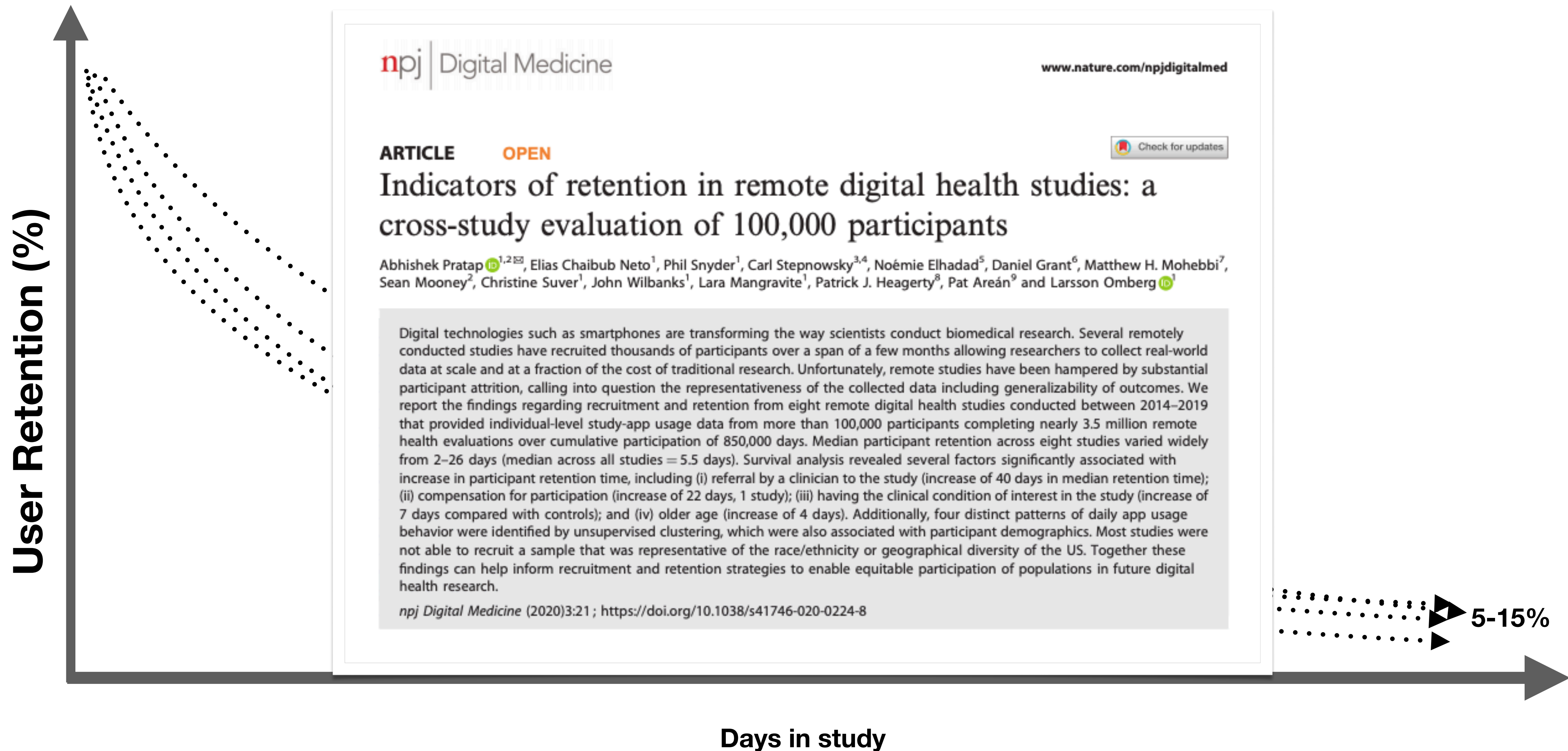
Digital technologies such as smartphones are transforming the way scientists conduct biomedical research. Several remotely conducted studies have recruited thousands of participants over a span of a few months allowing researchers to collect real-world data at scale and at a fraction of the cost of traditional research. Unfortunately, remote studies have been hampered by substantial participant attrition, calling into question the representativeness of the collected data including generalizability of outcomes. We report the findings regarding recruitment and retention from eight remote digital health studies conducted between 2014–2019 that provided individual-level study-app usage data from more than 100,000 participants completing nearly 3.5 million remote health evaluations over cumulative participation of 850,000 days. Median participant retention across eight studies varied widely from 2–26 days (median across all studies = 5.5 days). Survival analysis revealed several factors significantly associated with increase in participant retention time, including (i) referral by a clinician to the study (increase of 40 days in median retention time); (ii) compensation for participation (increase of 22 days, 1 study); (iii) having the clinical condition of interest in the study (increase of 7 days compared with controls); and (iv) older age (increase of 4 days). Additionally, four distinct patterns of daily app usage behavior were identified by unsupervised clustering, which were also associated with participant demographics. Most studies were not able to recruit a sample that was representative of the race/ethnicity or geographical diversity of the US. Together these findings can help inform recruitment and retention strategies to enable equitable participation of populations in future digital health research.

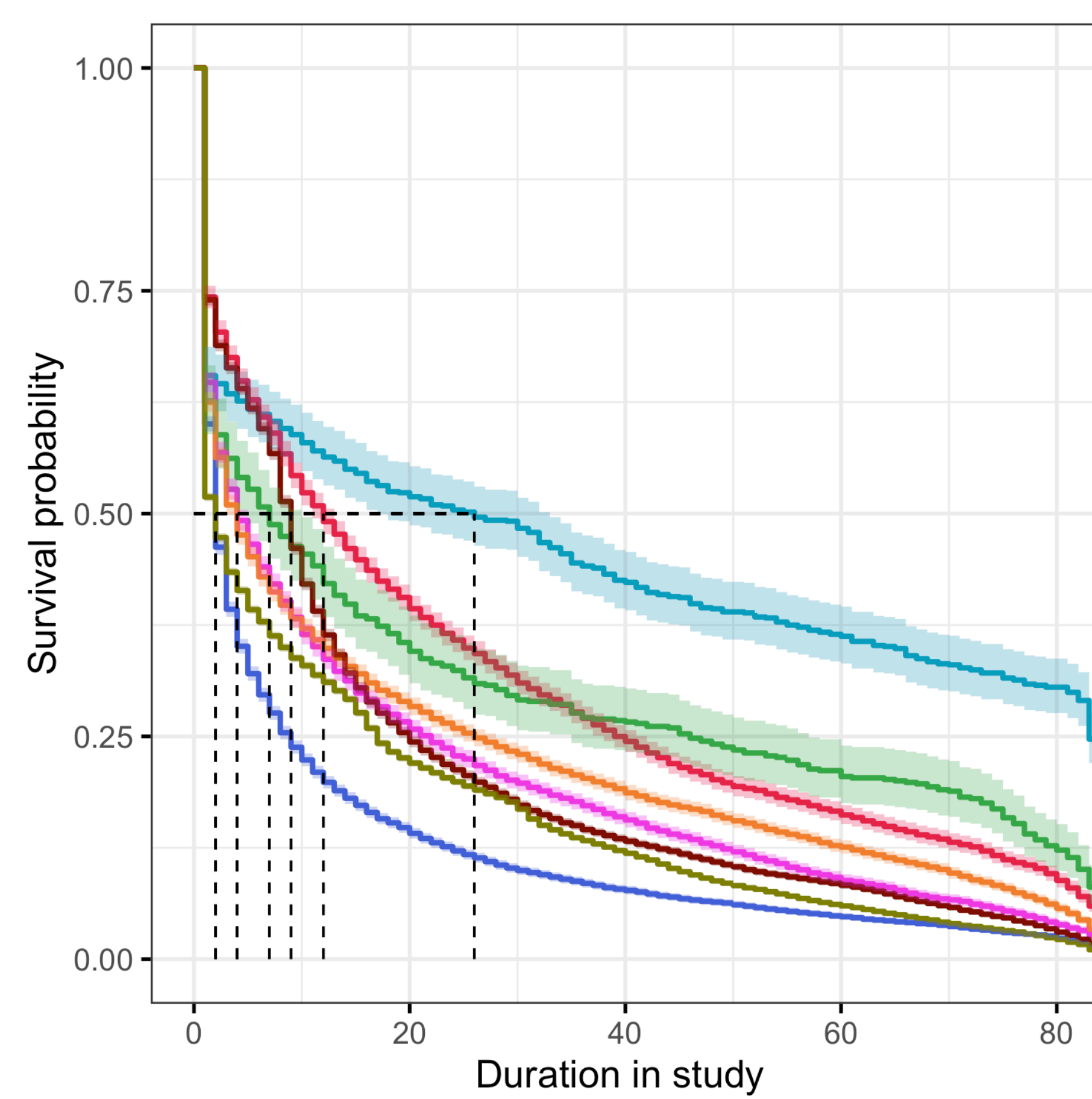
npj Digital Medicine (2020)3:21; <https://doi.org/10.1038/s41746-020-0224-8>

A. Pratap et al.

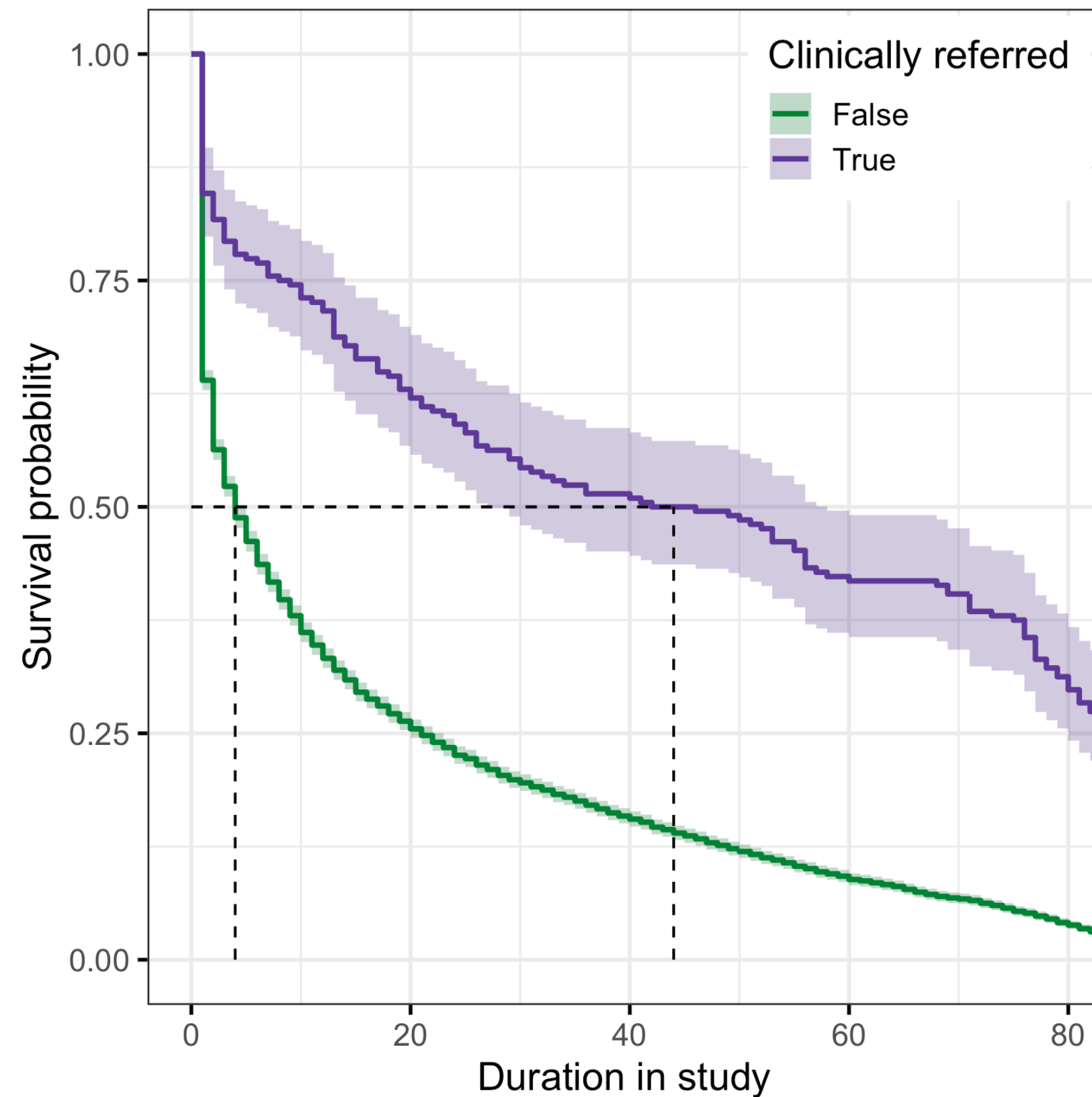


Building digital tech alone is not enough





- ~50% of participants leave the studies within the first 7-10 days
- Targeted Comms during the first week may help



- Participants referred by clinical sites/partners engage for significantly longer time