Cohen Veterans Bioscience

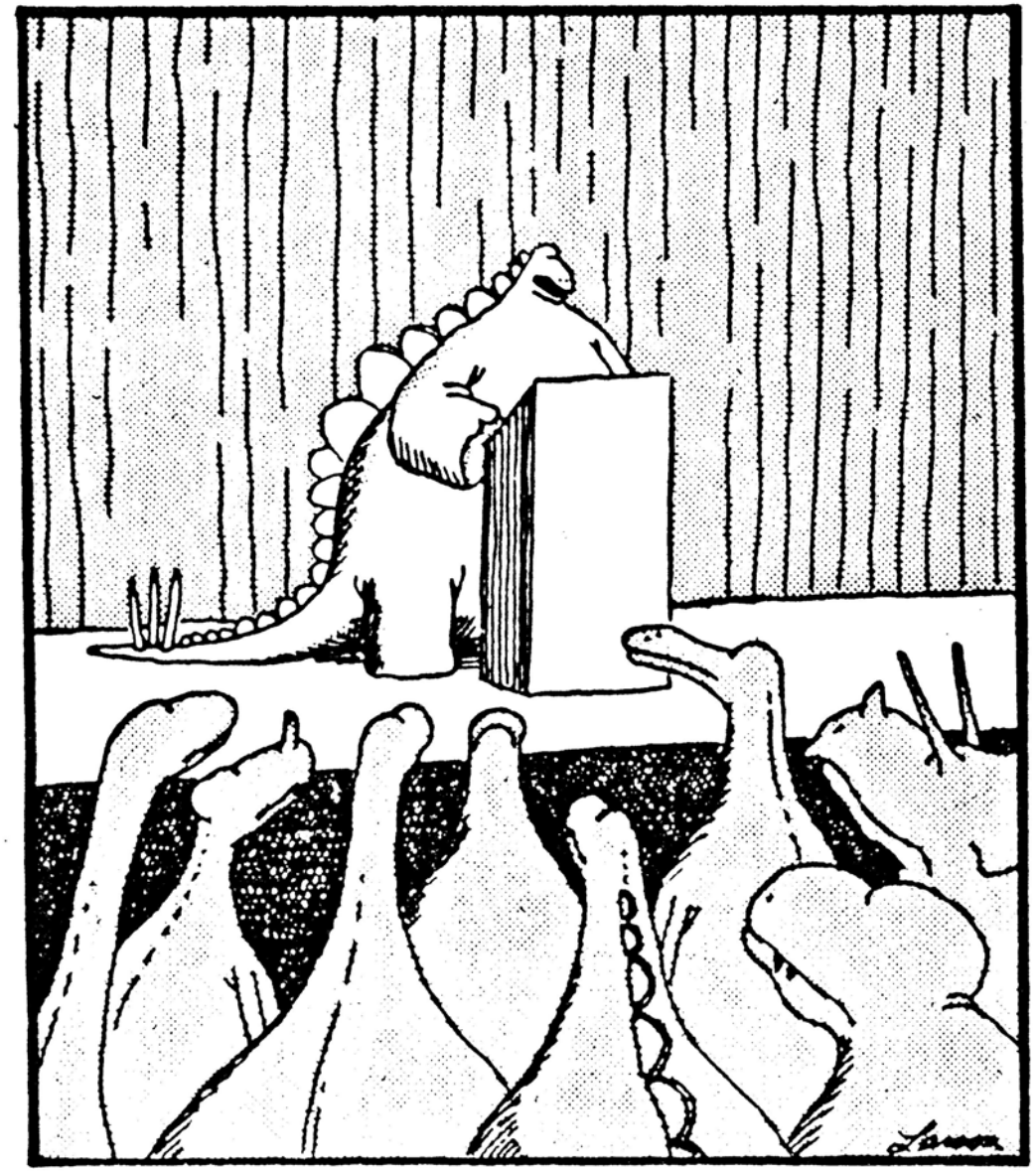
BRIDGING THE TRANSLATIONAL GAP: AN INTEGRATIVE SYSTEMS MODELING APPROACH FOR PRECISION BRAIN HEALTH

Magali Haas, MD, PhD

CEO & President

600

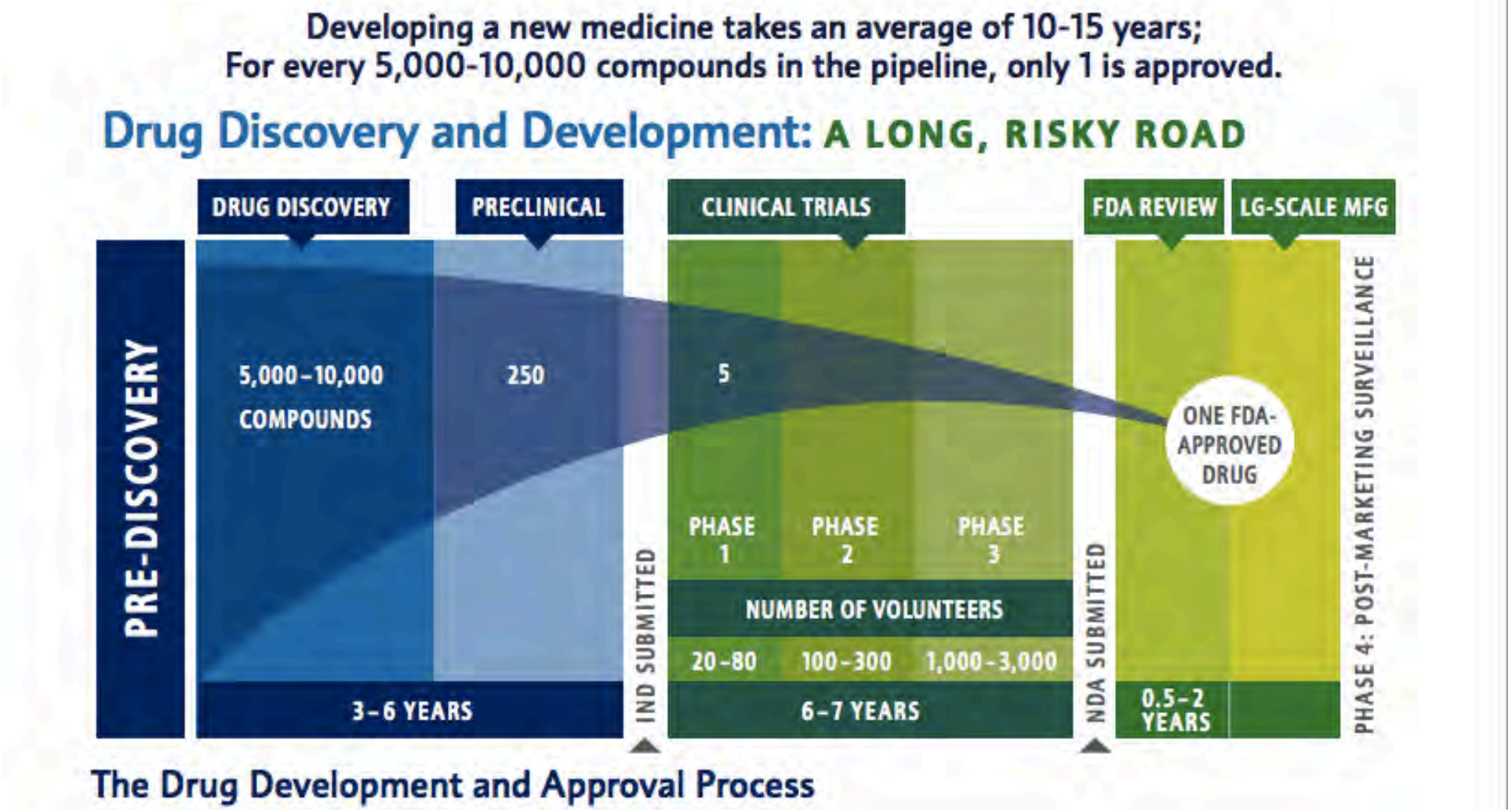
0



"The picture's pretty bleak, gentlemen. ... The world's climates are changing, the mammals are taking over, and we all have a brain about the size of a walnut."

Risk of R&D Drug Development

$1.2-3 billion, including the cost of failures



Reference: Tufts Center for the Study of Drug Development

The Reasons We Claim to Fail in CNS

What we say….

"The brain is too complex"
"We have no or poor targets"
"The population is heterogeneous"
"We have poor outcome measures"
"Low probability of success"
"Limited budgets"

ORION BIONETWORKS

The Reasons We Claim to Fail in CNS

What I hear…

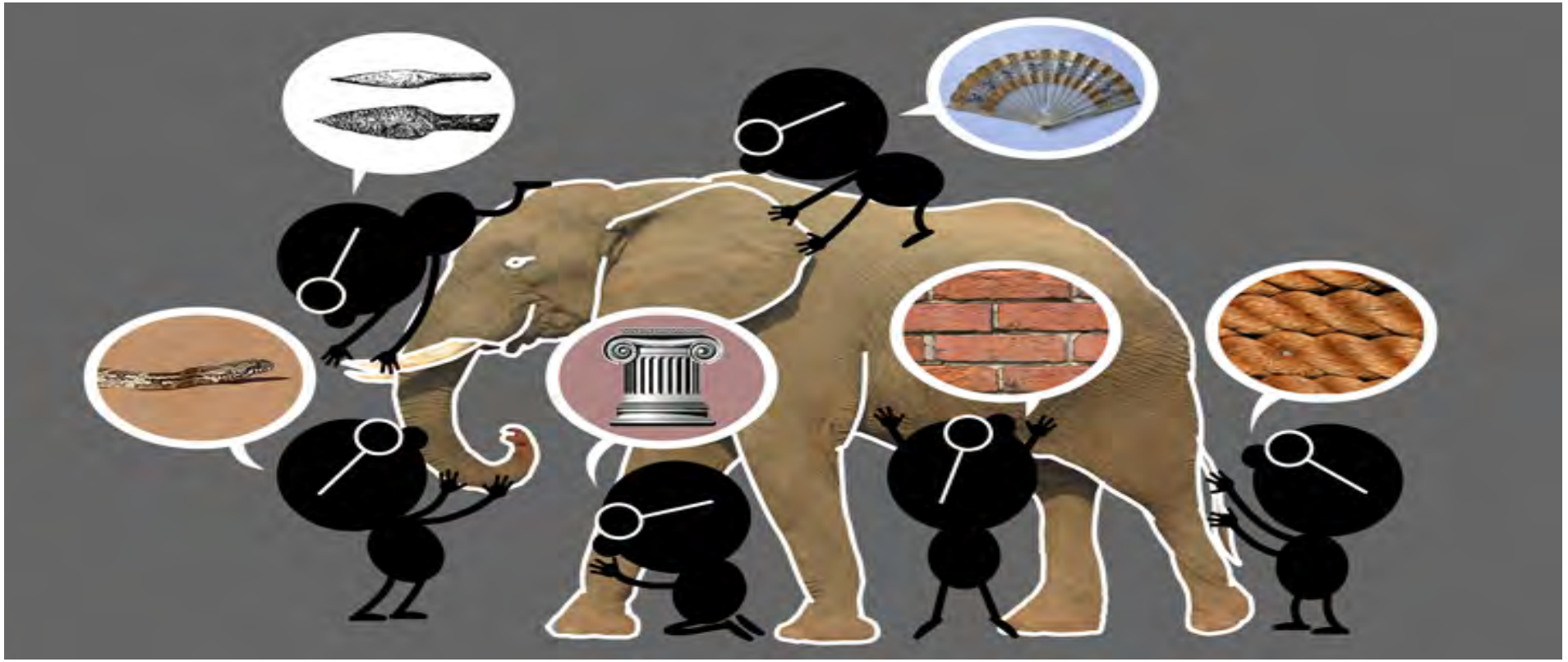
We need a mechanistic understanding of disease that embraces complex pathways

We need to phenotype populations deeply and longitudinally

We need to quantify traits and outcomes to measure impact

ORION BIONETWORKS

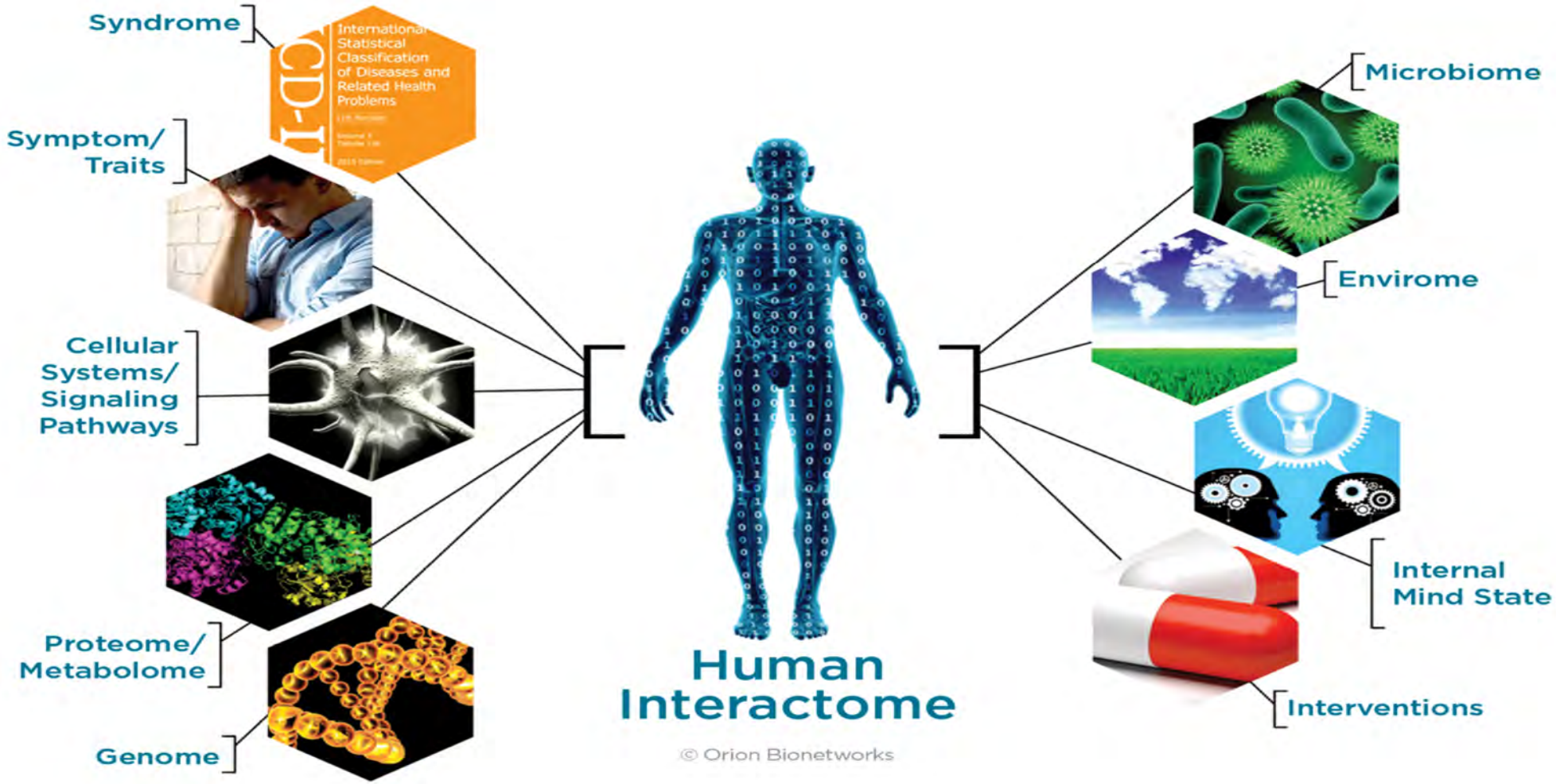
Medicine today is built on hundreds of years of individual observations…



Our approach has been LINEAR & LUCKY….

From Reductionist to Systems Modeling

Systematic Collection and Integration of Multi-modal Data



ORION BIONETWORKS

What If?

IMAGINE

WHAT IF WE COULD BUILD COMPUTER SIMULATIONS THAT PREDICT THE OUTCOME OF BRAIN DISEASE THE SAME WAY WE CAN ALREADY DO FOR WEATHER PREDICTION?

ORION BIONETWORKS

Star Date: September 2012

Lessons Learned

"Failure is simply the opportunity to begin again, this time more intelligently."
Henry Ford

- Not all data is 'created equal'.
- Reproducibility … crisis.
- Best practices & SOPs?
- Data platforms
- Operating models & Incentives

FORCED US TO REVISIT FUNDAMENTALS!

ORION BIONETWORKS

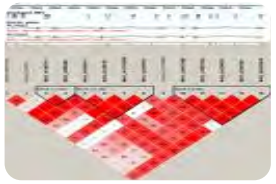
About Cohen Veterans Bioscience (2016-present)

We are a national, nonpartisan 501(c)(3) research organization dedicated to fast-tracking the development of diagnostic tests and personalized therapeutics for the millions of veterans and civilians who suffer the devastating effects of trauma-related and other brain disorders.

2017 TRAUMA SCORECARD	PTSD	mTBI
# of FDA Approached Diagnostics	0	2
Qualified Biomarkers	0	1?
Diagnostic Pipeline	?	1+
# of FDA Approached Therapeutics Overall	2	0
# of Therapeutics Approved in Past 15 Years	0	0
Therapeutic Pipeline	>20	>20

Lesson #1: We need better data

Published Literature & Repositories



Clinical Research Studies – Large Cohorts

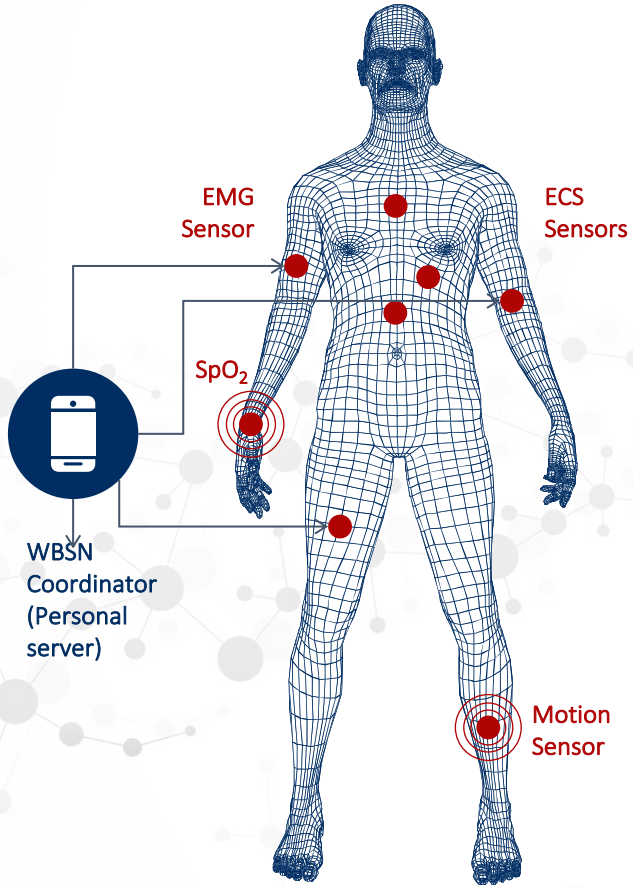
Clinician's Office – Electronic Health Records

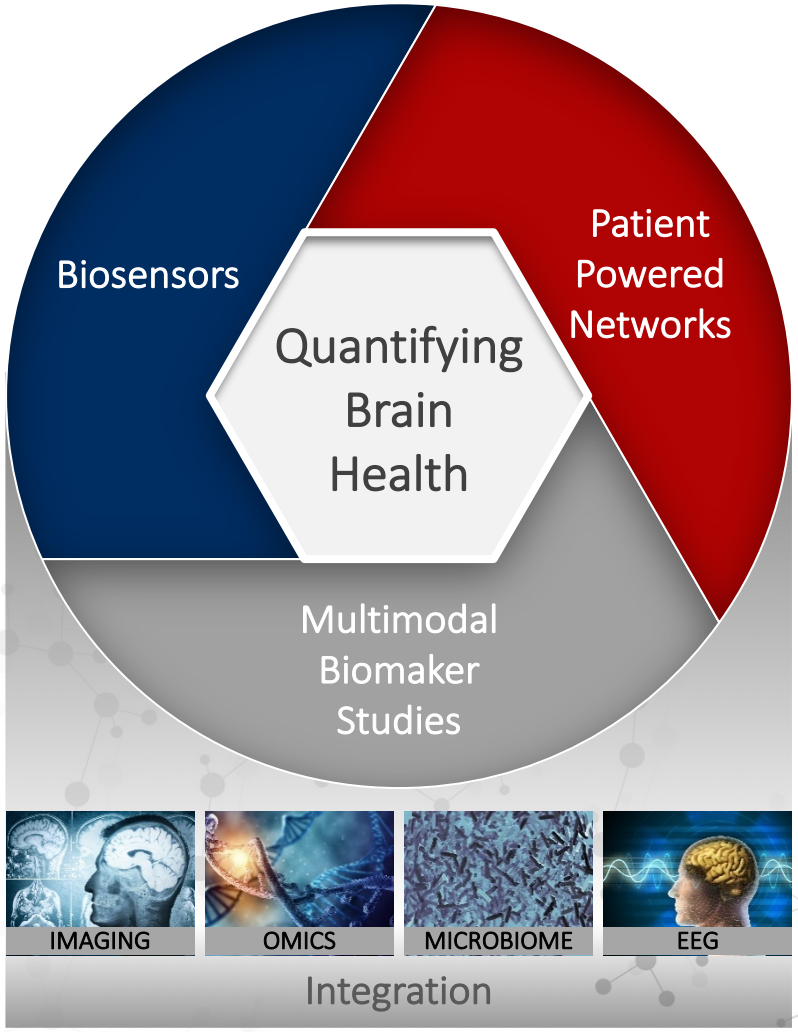
Limitations for Precision Medicine
Small studies/under-powered
Non-standardized
Low-reproducibility
Few longitudinal cohorts exist
Typically collect a limited dataset
Non-standardized practice
Intermittent and incomplete data
Superficial (generally little imaging, genomic, deep phenotype)
Variable quality/methodology
EHR developed for Reimbursement not Research

NOTE:
Data for "Systems Modeling" needs to meet stringent requirements often not met in traditional study programs

- Deep phenotyping
- Missing data
- Annotation
- Etc.

Nanotechnology:
Wireless
Body Area
Networks





Mobile& Social Media
Real-time
Real-world

Cohen Veterans Bioscience

Reproducibility *is* an issue

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

PlosMedicine, 2005

COMMENT

NIH plans to enhance reproducibility

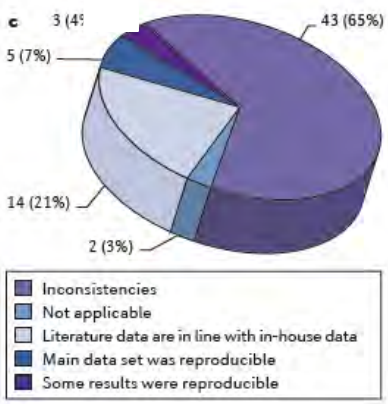
Francis S. Collins and **Lawrence A. Tabak** discuss initiatives that the US National Institutes of Health is exploring to restore the self-correcting nature of preclinical research.

A growing chorus of concern, from scientists and laypeople, contends that the complex system for ensuring the reproducibility of biomedical research is failing and is in need of restructuring[1,2]. As leaders of the US National Institutes of Health (NIH), we share this concern and here explore some of the significant interventions that we are planning.

Science has long been regarded as 'self-correcting', given that it is founded on the replication of prior work. Over the long term, that principle remains true. In the shorter term, however, the checks and balances that once ensured scientific fidelity have been hobbled. This has compromised the ability of today's researchers to reproduce others' findings.

Let's be clear: with rare exceptions, we have no evidence to suggest that irreproducibility is about scientific misconduct. In 2011, the Office of Research Integrity of the US Department of Health and Human Services pursued only 12 such cases[3]. Even if this represents only a fraction of the actual problem, such papers are vastly

2005 — 2011 — 2014 — 2015



% Studies in in-house projects reproduced at Bayer

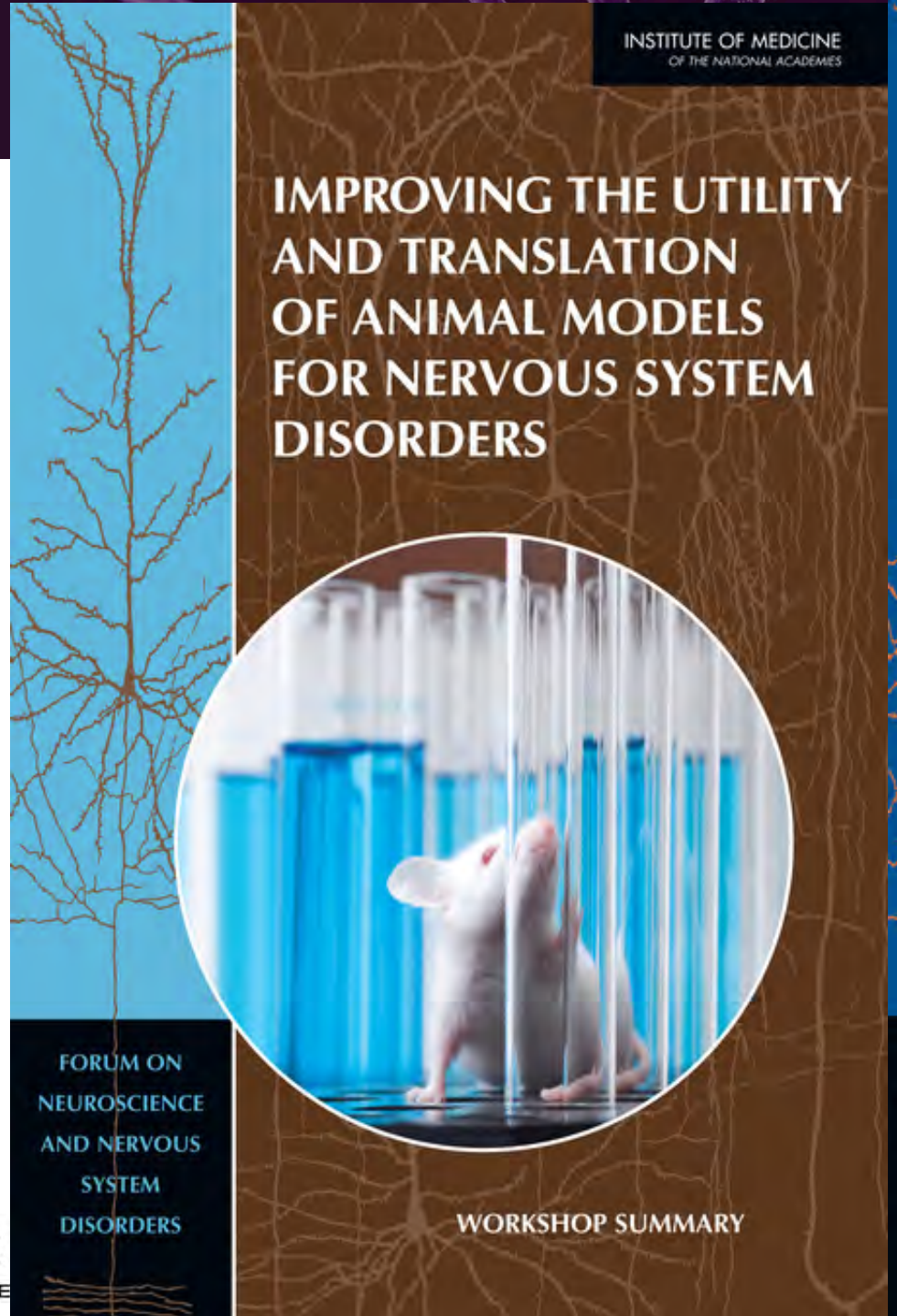
(Oncology, Women's Health, Cardiovascular)

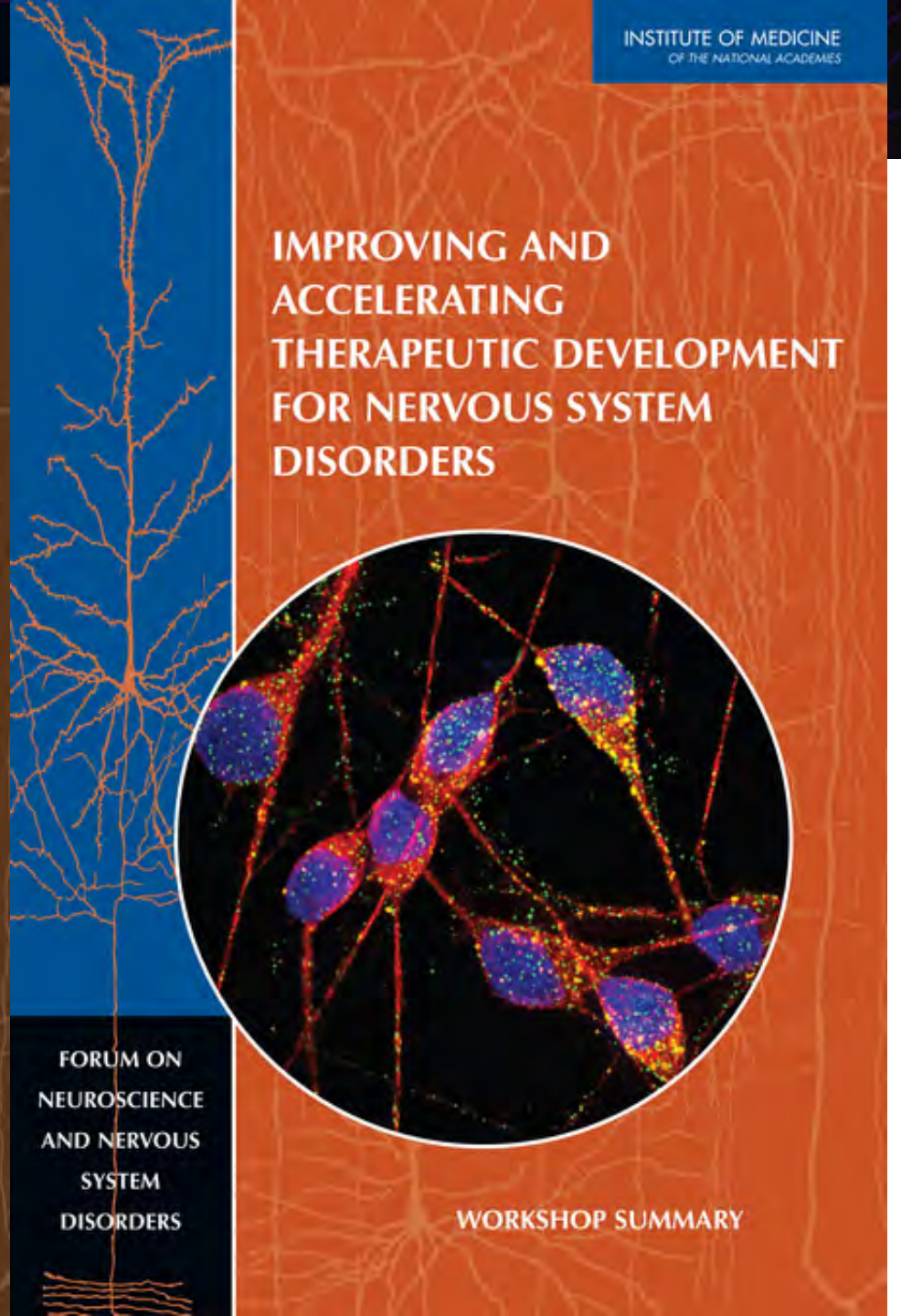
Prinz etal, Nat Rev Drug Dis, 2011

Launch of Multiple Initiatives:

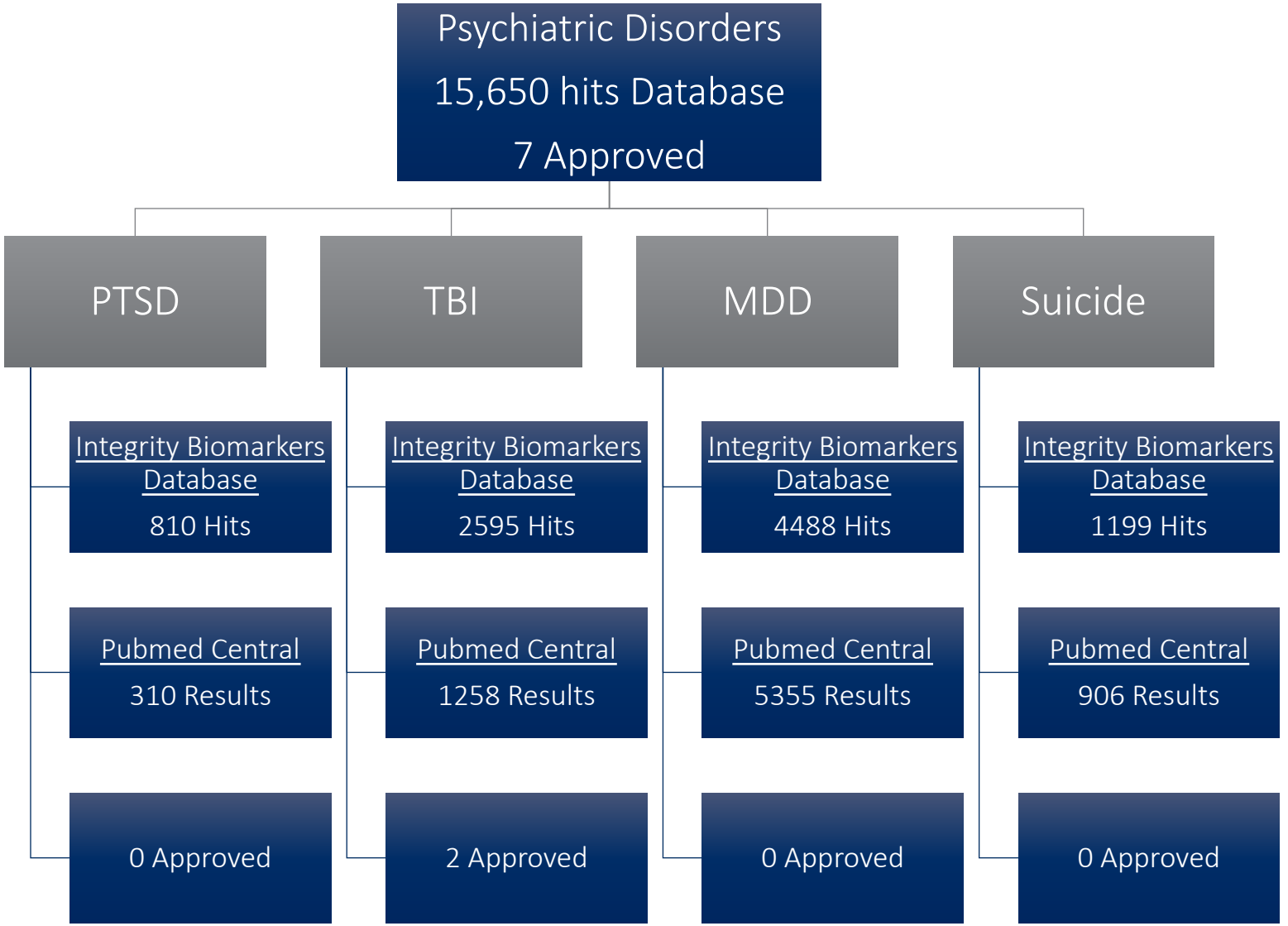
- Society for Neuroscience
- Wellcome Trust
- Preclinical Data Network → EQUIPD
- FASEB

….. and more….


INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES
IMPROVING THE UTILITY AND TRANSLATION OF ANIMAL MODELS FOR NERVOUS SYSTEM DISORDERS
FORUM ON NEUROSCIENCE AND NERVOUS SYSTEM DISORDERS
WORKSHOP SUMMARY


INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES
IMPROVING AND ACCELERATING THERAPEUTIC DEVELOPMENT FOR NERVOUS SYSTEM DISORDERS
FORUM ON NEUROSCIENCE AND NERVOUS SYSTEM DISORDERS
WORKSHOP SUMMARY

Biomarkers Discovered in Literature and Patents



The Devil's We Know

- **Study Design** - Hypothesis-driven/Candidate versus Unbiased/data-driven approaches to discovery
- **Statistical Power** – many biomarker studies have been underpowered/small studies
- **Reproducibility** – most studies have not been independently replicated
- **Methodology** – pre-analytic variables affect data (needle size, time to freeze, etc)
- **Standards** – across labs, across MRIs, across batches
- **Assay Performance** – often overlooked – garbage in → garbage out

Operating Models & Incentives: Examples

Grant-making (e.g. NIH)

- Bottom-up projects
- Siloed
- Crowd-sourced innovation
- Central funding

Solution-Driven (e.g.DARPA/IMEC)

- Top-down programs
- High-touch
- Low through-put
- Central funding

Patient-Driven (e.g. Advocacy)

- Influence-driven programming
- Broad reach
- Leveraged funding

Think Tank (e.g. Milken)

- Convener
- Thought leadership
- No direct program management
- Leveraged funding

CVB Strategic Roadmap

RESEARCH → PRACTICE

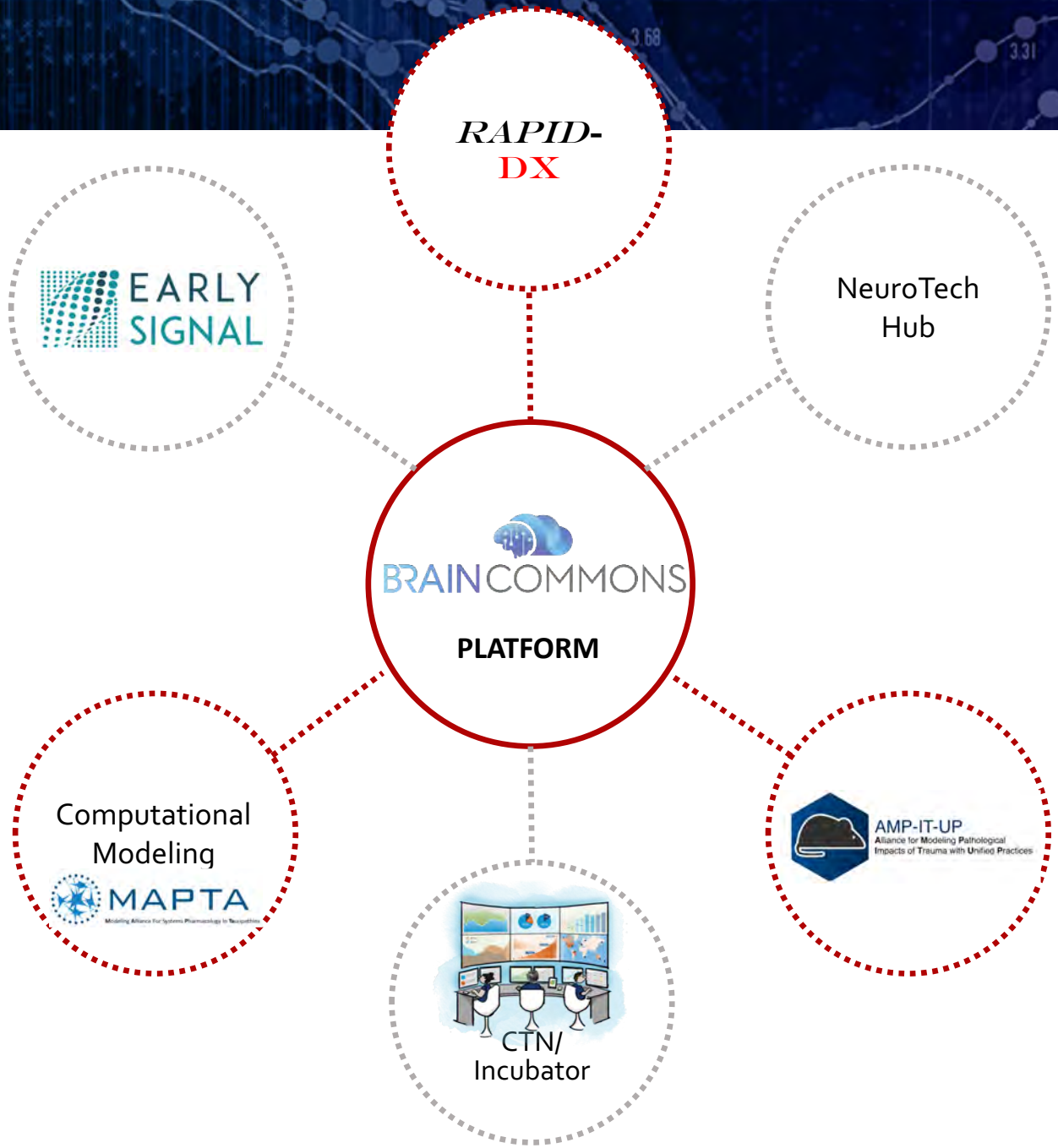
Continuum – from enhancing research capacity resources to funding strategic research and implementing promising practices

Bio-repositories	Predictive Models	Biomarkers & MOA	Data Science	mHealth & Outcomes	Therapeutics & Diagnostics	Adoption
Blood Imaging EEG Brain Banking Stem Cells	Animal Models Computer Models Stem Cell Models	Large, deep phenotype cohort studies Discovery & Replication Pathways & Circuits	Integrated analysis Machine Learning Bio-informatics	Wearables & mHealth Telehealth Virtual Trials Outcome Tools	Incubator Clinical Trials Accelerator Bridge Funding	Best Practice Guidelines Awareness Education/ Training

KNOWLEDGE ENGINEERING

A Platform Approach

Our Approach is
to **Build Enabling Platforms**
with **Strategic Partners ,**
incentivizing a **Team Science**
approach
to fast-track solutions.



Genetics of PTSD: Status November 2015

ORIGINAL ARTICLE

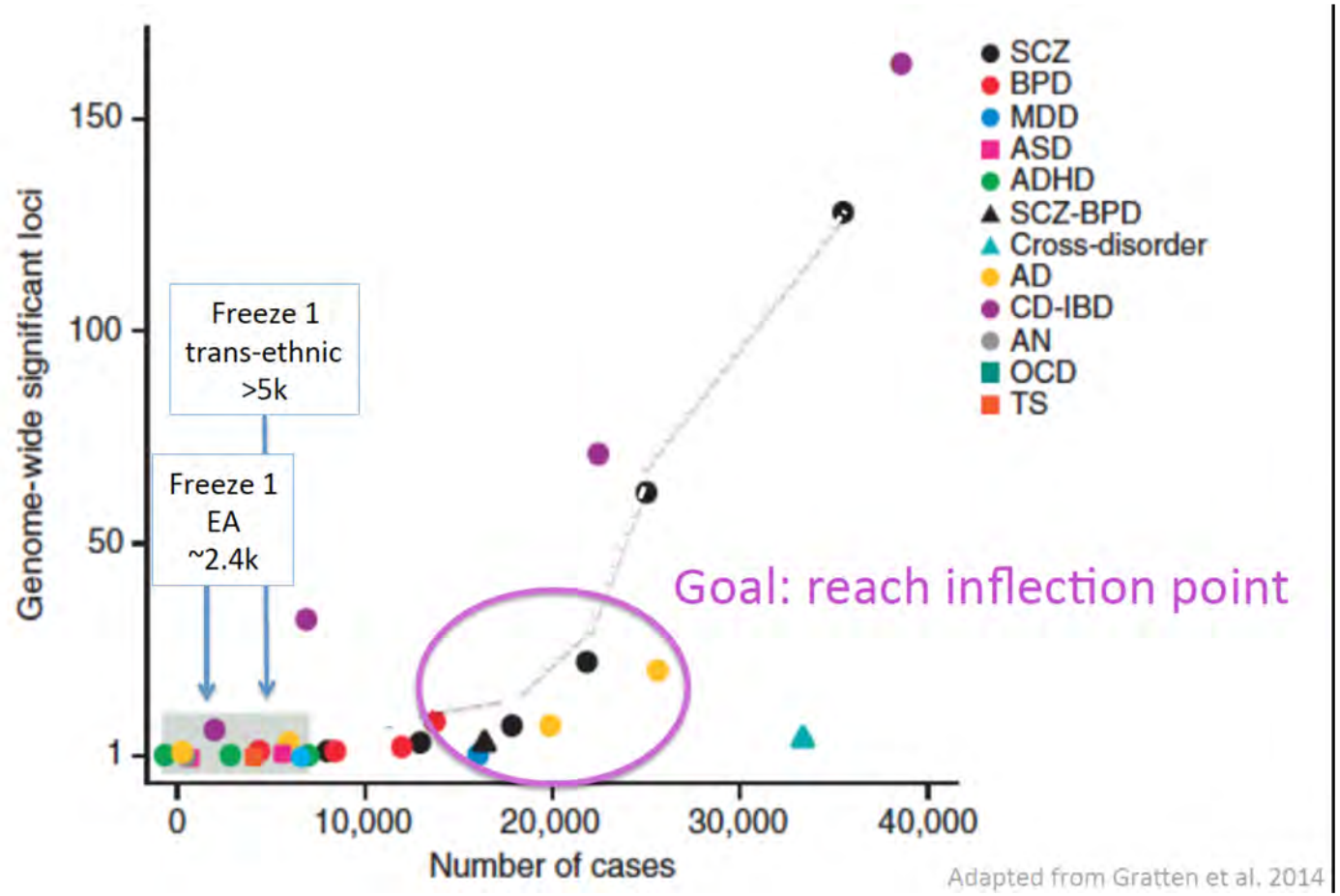
Largest GWAS of PTSD ($N = 20070$) yields genetic overlap with schizophrenia and sex differences in heritability

LE Duncan[1,2,3], A Ratanatharathorn[4], AE Aiello[5], LM Almli[6], AB Amstadter[7], AE Ashley-Koch[8], DG Baker[9,10], JC Beckham[11,12], LJ Bierut[13], J Bisson[14], B Bradley[15,16], C-Y Chen[3,17,18], S Dalvie[19], LA Farrer[20], S Galea[21], ME Garrett[8], JE Gelernter[22], G Guffanti[18,23], MA Hauser[8], EO Johnson[24], RC Kessler[25], NA Kimbrel[11,12], A King[26], N Koen[27,28], HR Kranzler[29], MW Logue[30,31], AX Maihofer[32], AR Martin[2,3], MW Miller[30,33], RA Morey[12,34], NR Nugent[35,36], JP Rice[37], S Ripke[2,3,38], AL Roberts[39], NL Saccone[40], JW Smoller[2,17], DJ Stein[27,28], MB Stein[32,41,42], JA Sumner[43], M Uddin[44], RJ Ursano[45], DE Wildman[46], R Yehuda[47,48], H Zhao[49], MJ Daly[2,3], I Liberzon[26,50], KJ Ressler[18,23], CM Nievergelt[9,10] and KC Koenen[2,17,51]

The Psychiatric Genomics Consortium-Posttraumatic Stress Disorder group (PGC-PTSD) combined genome-wide case–control molecular genetic data across 11 multiethnic studies to quantify PTSD heritability, to examine potential shared genetic risk with schizophrenia, bipolar disorder, and major depressive disorder and to identify risk loci for PTSD. Examining 20 730 individuals, we report a molecular genetics-based heritability estimate (h^2_{SNP}) for European-American females of 29% that is similar to h^2_{SNP} for schizophrenia and is substantially higher than h^2_{SNP} in European-American males (estimate not distinguishable from zero). We found strong evidence of overlapping genetic risk between PTSD and schizophrenia along with more modest evidence of overlap with bipolar and major depressive disorder. No single-nucleotide polymorphisms (SNPs) exceeded genome-wide significance in the transethnic (overall) meta-analysis and we do not replicate previously reported associations. Still, SNP-level summary statistics made available here afford the best-available molecular genetic index of PTSD—for both European- and African-American individuals—and can be used in polygenic risk prediction and genetic correlation studies of diverse phenotypes. Publication of summary statistics for ~10 000 African Americans contributes to the broader goal of increased ancestral diversity in genomic data resources. In sum, the results demonstrate genetic influences on the development of PTSD, identify shared genetic risk between PTSD and other psychiatric disorders and highlight the importance of multiethnic/racial samples. As has been the case with schizophrenia and other complex genetic disorders, larger sample sizes are needed to identify specific risk loci.

Molecular Psychiatry advance online publication, 25 April 2017; doi:10.1038/mp.2017.77

Voluntary Consortium



Cohen Veterans Bioscience

Post-CVB-Stanley Partnership (2016-2017)

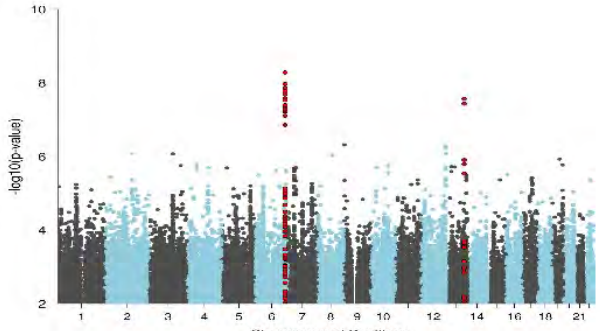
~21,000 cases, 59,000 controls = ~80,000 samples

Integrated 56 studies from around the Globe

SIGNIFICANT RESULTS (1.5 Years Post-Project Launch)

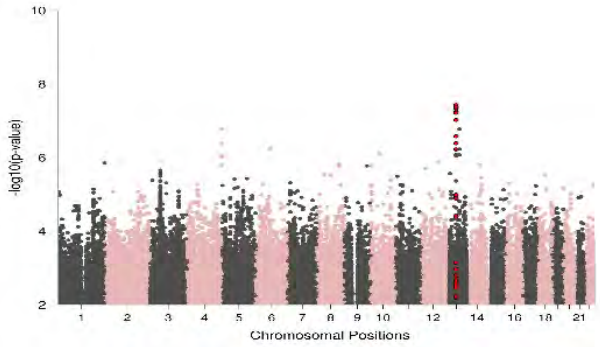
GWAS in European ancestry:

- N cases: 12,813; N controls: 35,640; N total: 48,453;
- N studies: 50
- GWAS hits: 2



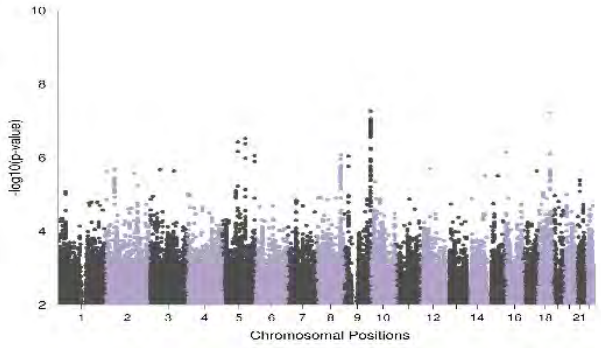
GWAS in African ancestry:

- N cases: 4,289; N controls: 10,500; N total: 14,7893;
- N studies: 30
- GWAS hits: 1



GWAS in Latino ancestry:

- N cases: 1,981; N controls:3,722; N total: 5,703;
- N studies: 6
- GWAS hits: suggestive 2



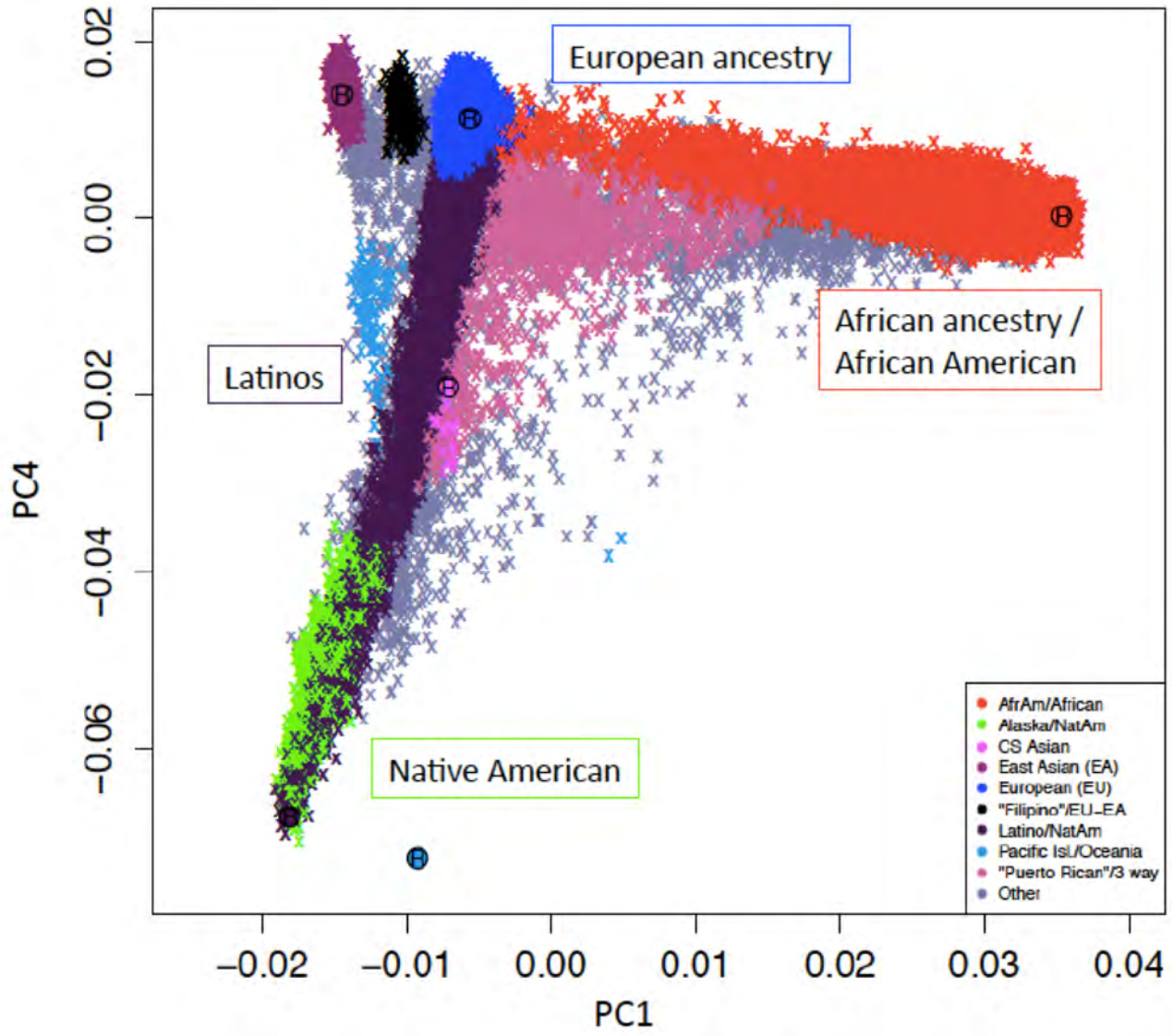
Next Steps (2018)

Integrate across diverse ancestries
Expand sample size
- UK biobank (N~150,00)
- Million Veterans Program

Interrogate “hits”

HOT OFF THE PRESS!

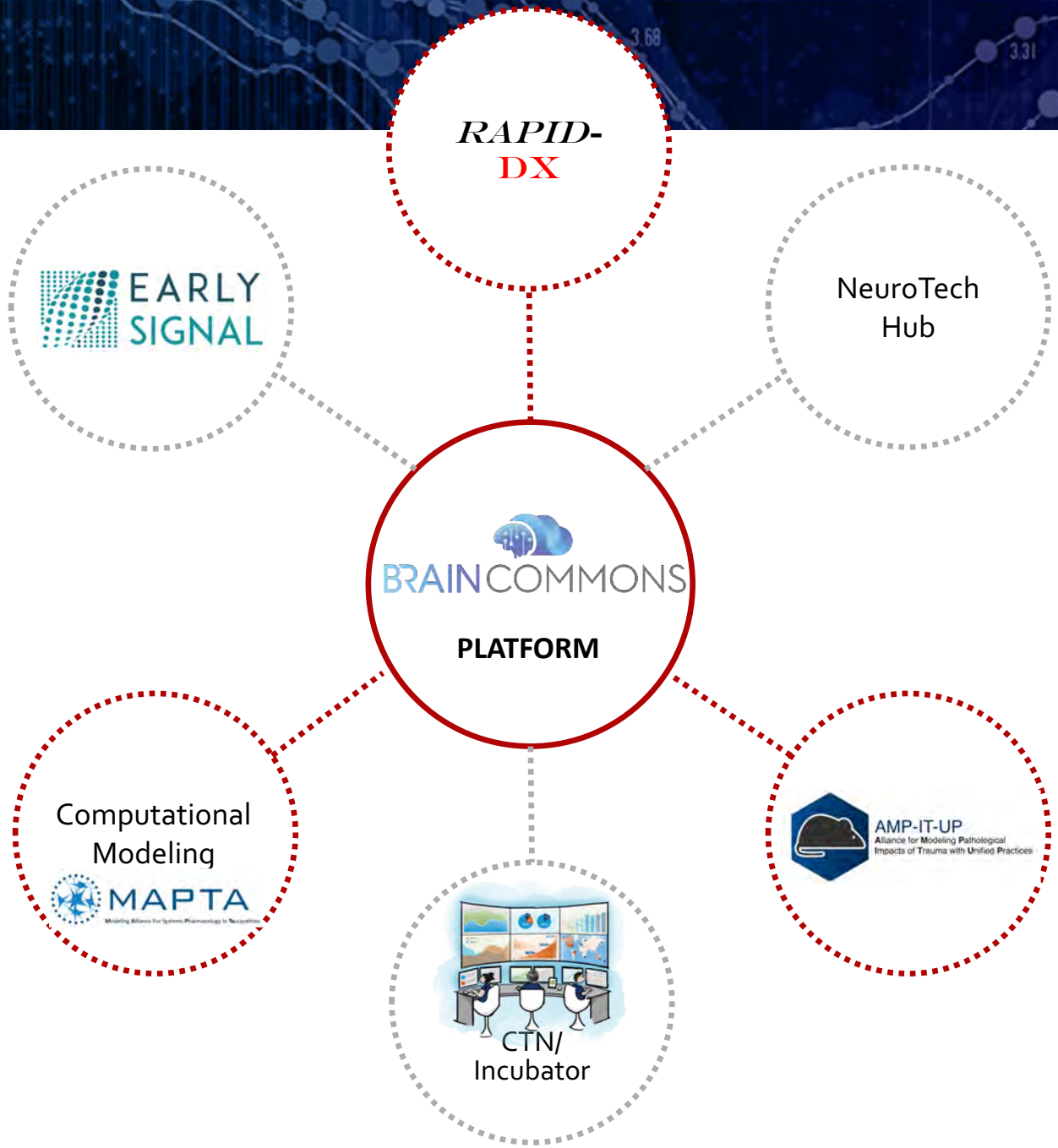
Manuscript just submitted
30K cases & 170K controls
6 SNPs identified!



Cohen Veterans Bioscience

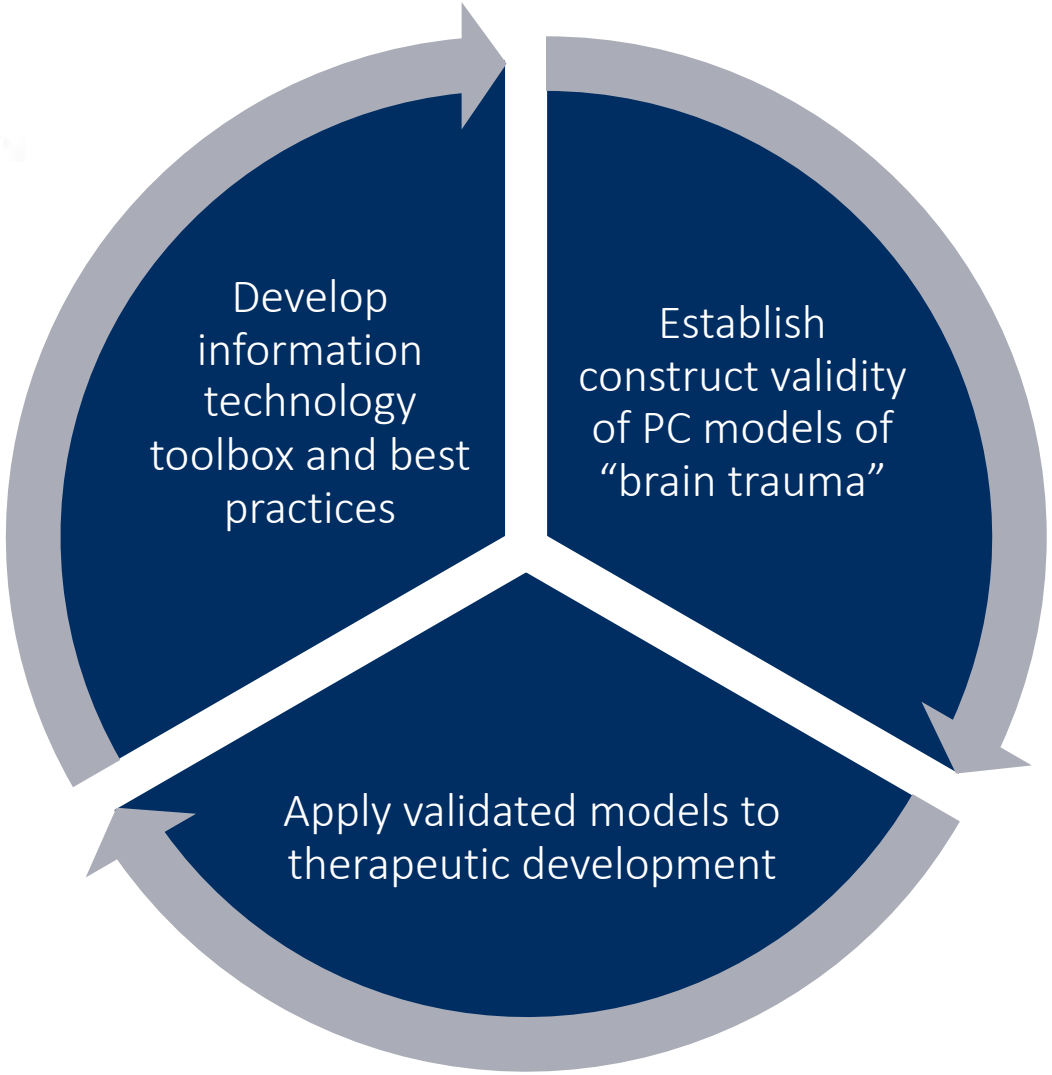
A Platform Approach

Our Approach is
to **Build Enabling Platforms**
with **Strategic Partners ,**
incentivizing a **Team Science**
approach
to fast-track solutions.



AMP-IT-UP

Alliance for **M**odeling **P**athological
Impacts of **T**rauma with **U**nified **P**ractices

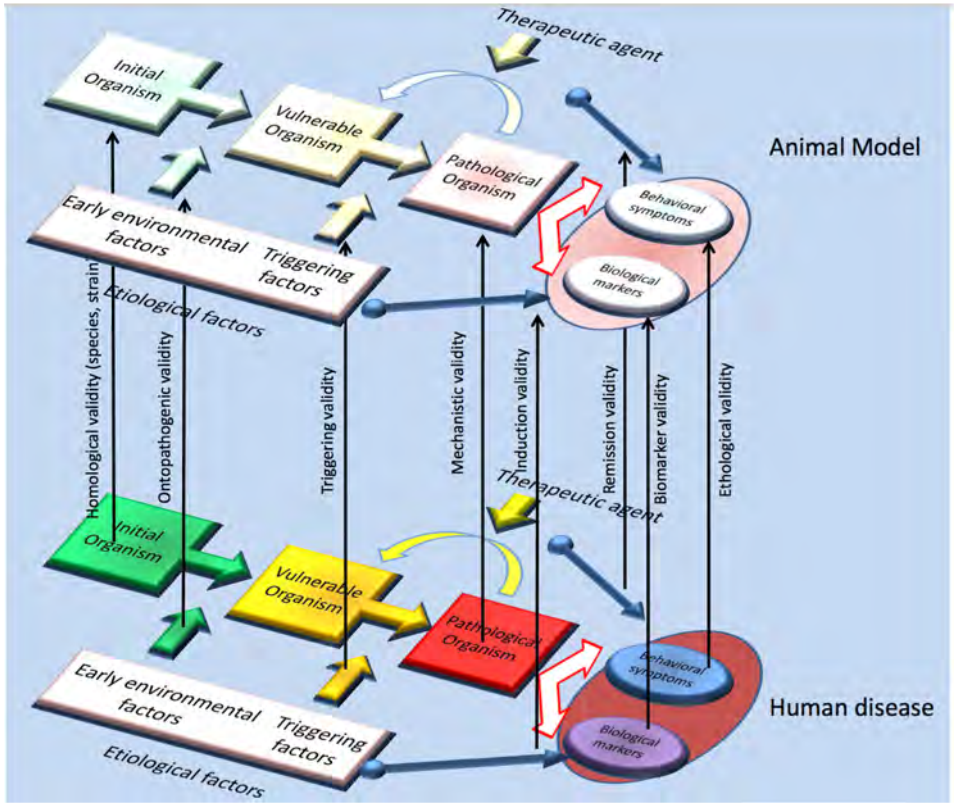


AMP-IT-UP, a collaborative working group, launched to jump-start the field of PTSD & TBI animal modeling to:

- Harmonize constructs in humans and preclinical animal and computational models to facilitate the development of translatable preclinical model systems;
- Develop best practice standards for reproducibility and robustness of PTSD model development and study conduct;
- Explore innovative nano- and imaging technologies to expand the armamentarium for preclinical research;
- Explore the use of computational modeling approaches to provide a complementary modeling approach to animal models;
- Centrally collect and synthesizing all available information about models;
- Identify gaps requiring additional study;
- Support study programs to address those gaps.

Landscape of Available "PTSD" Models

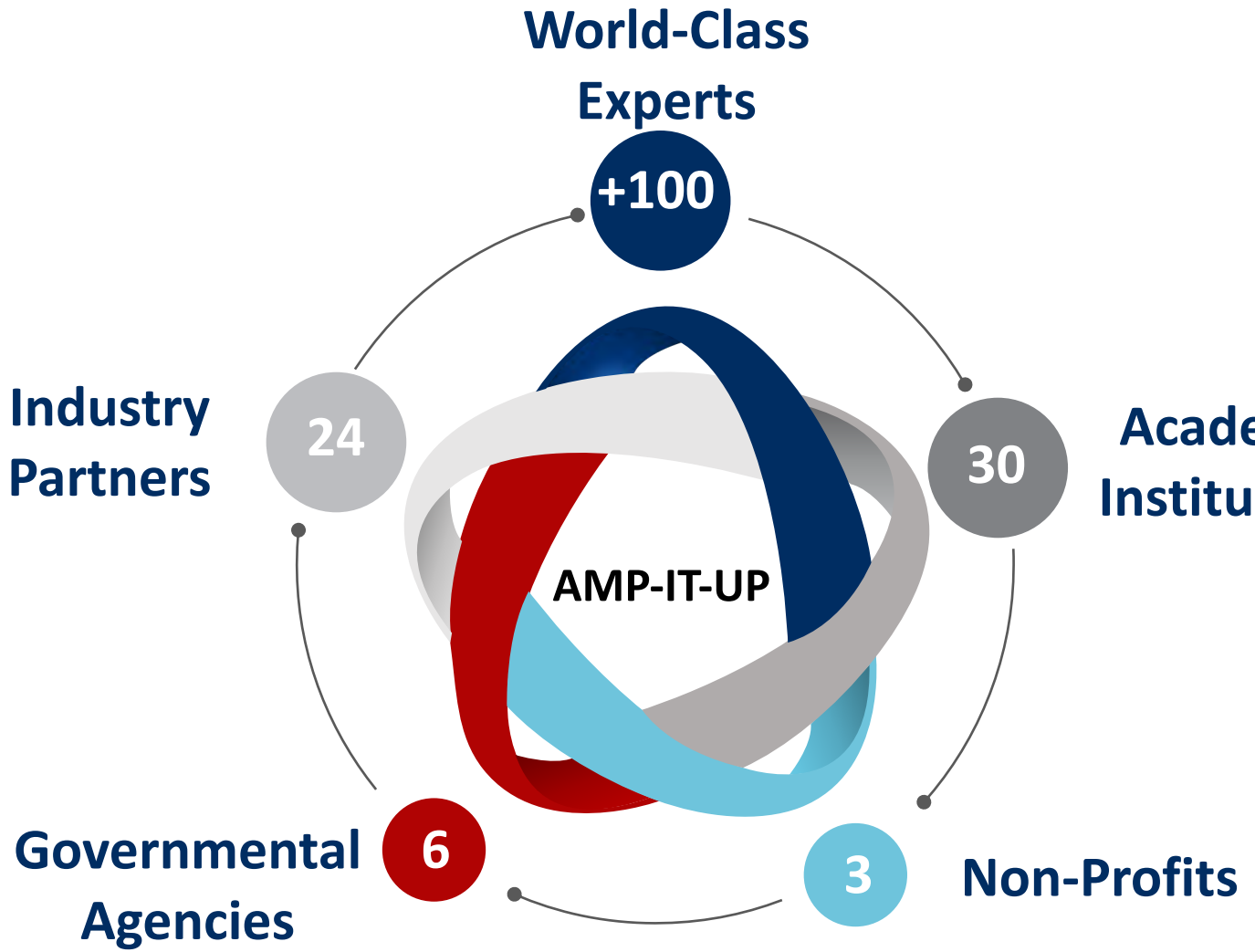
	Model	Stressor Type
1	Immobilization	Processive
2	Chronic Social Defeat Stress / Resident - Intruder	Psychosocial, Pain
3	Social Structure	Social
4	Predator/Predator Odor	Innate
5	Single Prolonged Stress	Acute, Severe, Life Threatening
6	Early Life Stress	Developmental
7	Environmental Stress	Home Cage Disruption/Enrichment Poverty
8	Chronic Unpredictable Stress	Environmental, chronic
9	Sleep Deprivation	Physical
10	Shock stressors	Physical, Pain
11	"Sequester Pain" or "Rat Party"	Psychosocial
12	Hypoxia	Physical
13	Underwater Trauma	Physical, Life Threatening
14	Genetic models and/or susceptible lines	



Belzung, C., & Lemoine, M. (2011). Criteria of validity for animal models of psychiatric disorders: focus on anxiety disorders and depression. *Biology of Mood & Anxiety Disorders*, *1*(1), 9.

AMP-IT-UP + PDN Membership

AMP-IT-UP
Alliance for Modeling Pathological Impacts of Trauma with Unified Practices



Cohen Veterans Bioscience

PROJECT OUTLINES AND OBJECTIVES

OBJECTIVES

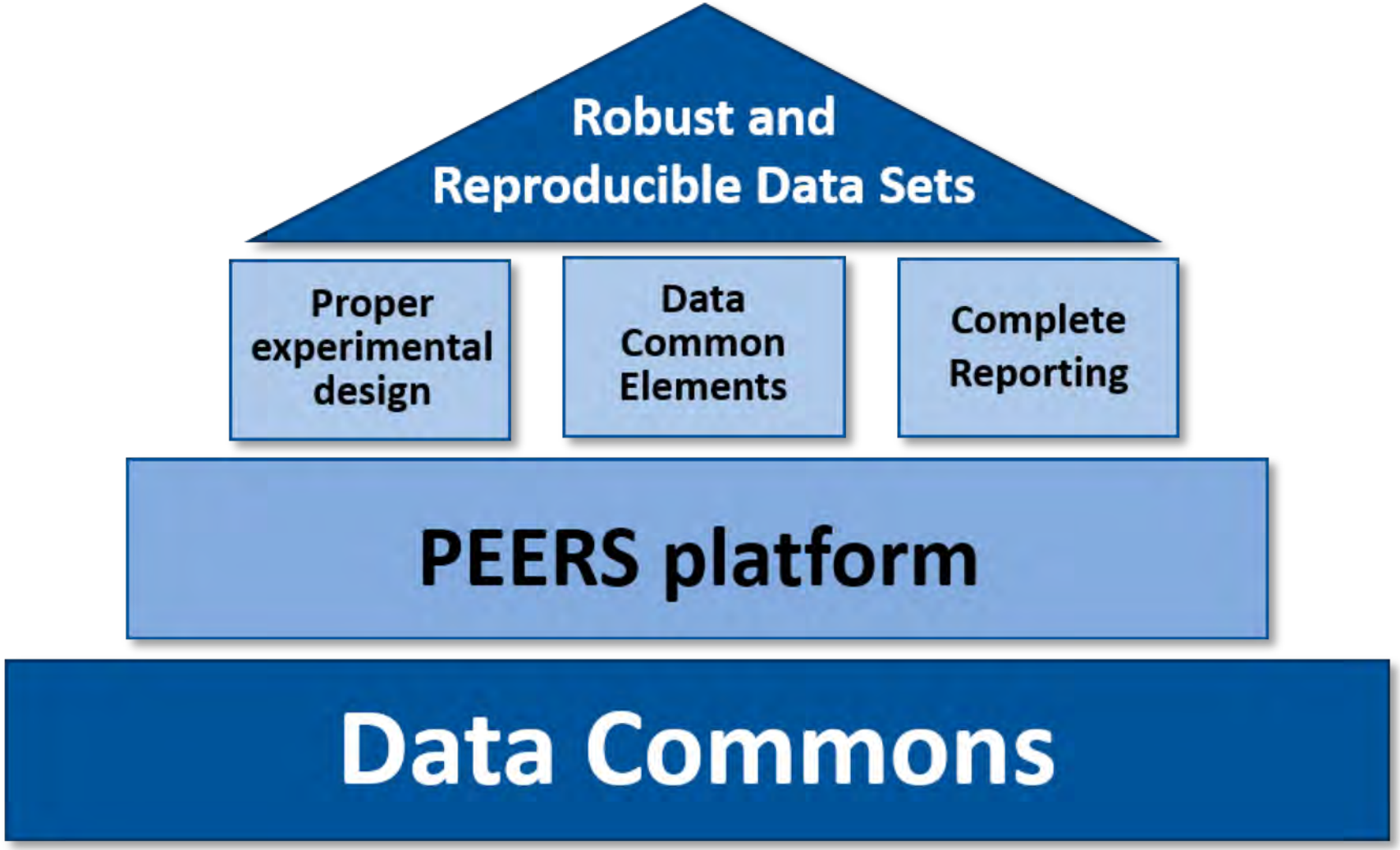
The EQIPD consortium will

- Define those variables in study design and data analysis that influence outcome in pre-clinical neuroscience (focus on Alzheimer's disease and psychosis) and (neuro-)safety studies conducted in industry
- Establish whether these are the same variables which influence outcome in academia
- Define the components which will make up the EQIPD quality management system
- Define consensus quality management recommendations for non-regulated research and development
- Validate the feasibility of the quality management system in prospective studies
- Deliver an online educational platform providing certified education and training in the principles and application of quality and rigour

Helpful references

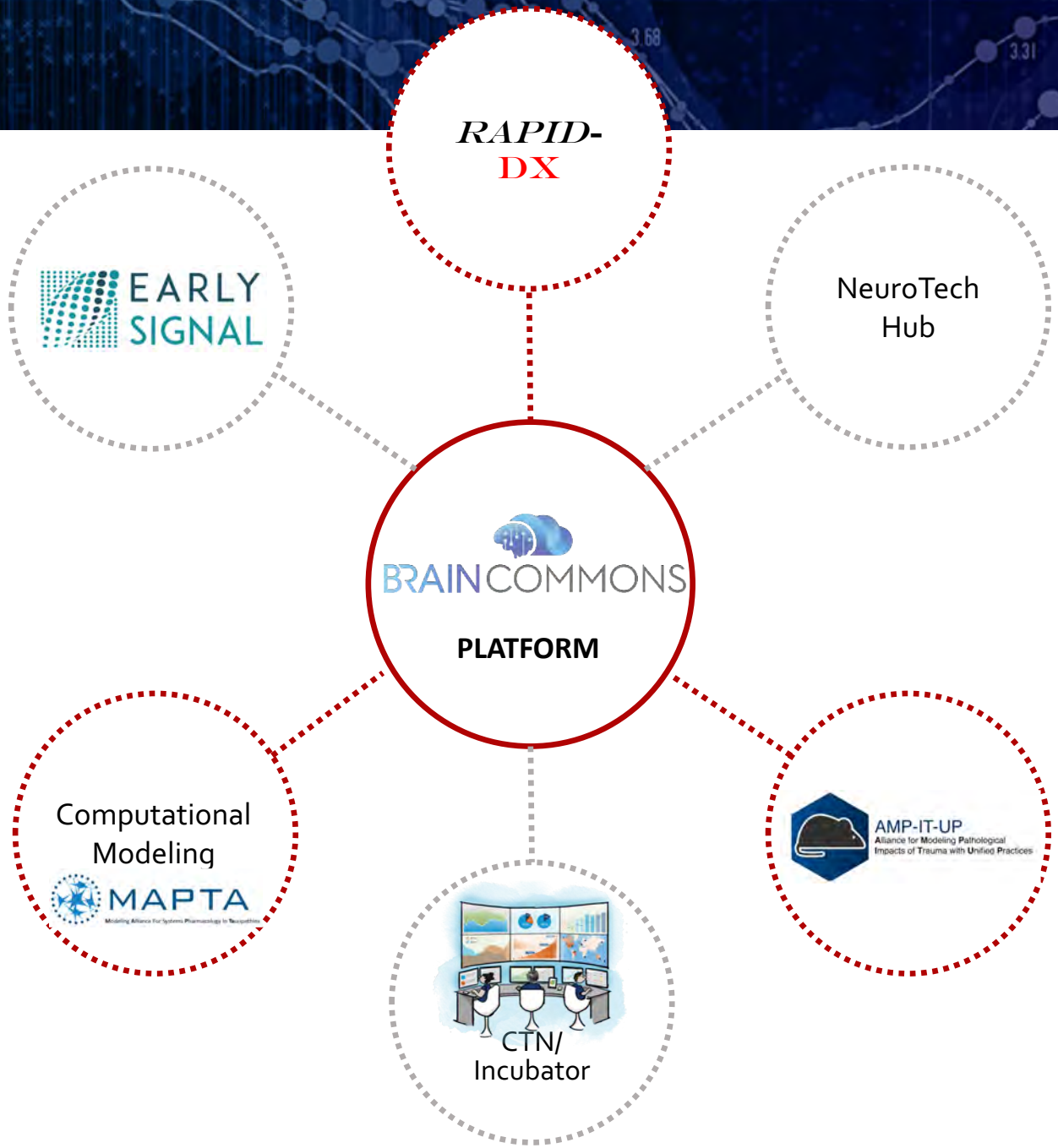
- Members
- Publications
- Work Packages
- Contact

PEERS


Robust and
Reproducible Data Sets
Proper experimental design
Data Common Elements
Complete Reporting
PEERS platform
Data Commons

A Platform Approach

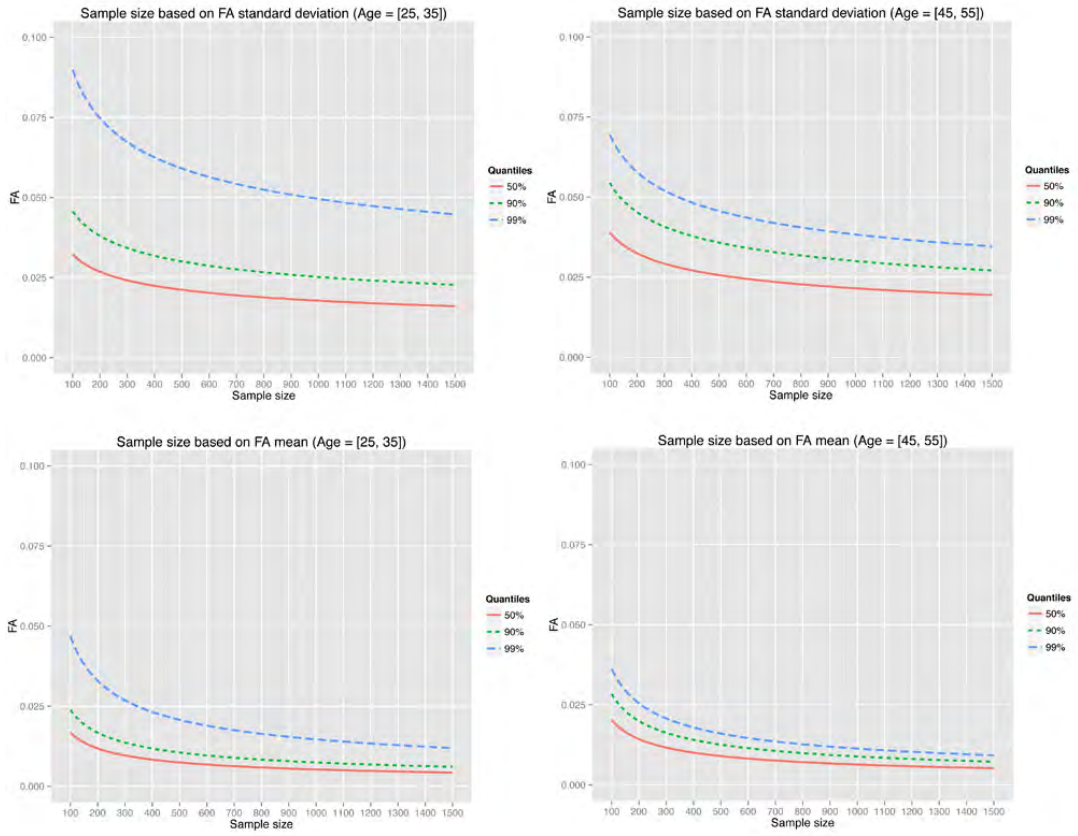
Our Approach is
to **Build Enabling Platforms**
with **Strategic Partners ,**
incentivizing a **Team Science**
approach
to fast-track solutions.



Normative Neuroimaging Library

- In 2017 CVB and partners launched a comprehensive effort to establish a neuroimaging library (>3000 subjects) to inform currently available FDA approved tools for interrogating advanced imaging.
- Addresses a critical current need for FDA approved tools that are presently being used for clinical care.
- Addresses a rapidly expanding need for tools making their way into clinical care within the next 3-5 years. Upon completion, these tools will require high quality normative data to function properly

Normative Neuroimaging Library: Precision analysis to determine library size

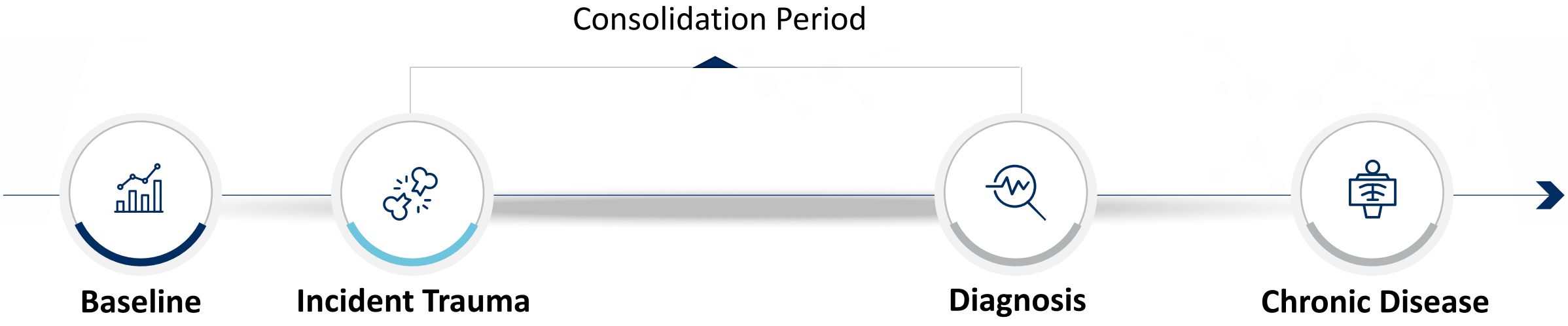


- Include 500 subjects for each of the following age groups:
  - 18-25; 26-35; 36-45; 46-55; 56-65
- Implement consensus recommendations of 2014 ACR Montreal panel
- Include both DOD and civilian sites
  - Current sites: Lackland AFB, SAMMC, Baylor College of Medicine, University of Virginia

TARGET = 3000 brains scanned

>500 individuals scanned since 2017!

Biomarkers & Diagnostics for PTSD & TBI



Risk Management	Screening/ Detection	Diagnosis	Prognosis	Prediction	Monitoring
Identify factors to assess susceptibility of disease	Indicate the presence of disease; early detection	Definitive diagnosis and general typing	Assess disease aggressiveness & likelihood of recurrence	Predict efficacy or response to different treatments	Monitor disease recurrence & therapeutic response

Cohen Veterans Bioscience

Research Alliance for PTSD/TBI Innovation and Discovery Diagnostics (RAPID-Dx)

U.S. Department of Veterans Affairs

WOUNDED WARRIOR PROJECT

Launched February 2018

Current Focus

Objective #1	•Pursue infrastructure development by forming public-private partnerships to share data across cohorts, evaluating bio-assays to select best-in-class platforms, implementing SOPs in samples collection, handling, and analytical approaches, and consolidating data to the Brain Commons for deployment of large-scale analytics.
Objective #2	•Promote biomarker discovery by using robust assays from platform evaluations, identifying priority COUs/TDPs and pursuing these in appropriate RAPID-Dx cohorts, implementing statistical analyses plans to ensure statistical power, robustness, and replicability.
Objective #3	•Conduct biomarker replication by building on discovery studies to validate assays and clinical contexts of use, which can be complemented by seeking guidance from regulatory partners.
Objective #4	•Pending Objective 1-3, replicated biomarkers can be developed further for clinical implementation through pursuit of regulatory approval, attainment of wide-spread reimbursement, and adoption into clinical practice.

37

Cohen Veterans Bioscience

Consensus Workshop - Priority Diagnostic Opportunities

Deconstruct population into coherent stratified subtypes
[with common biological constructs using large cohort phenotype & biomarker data (imaging, fluid, EEG)]

Pre-exposure risk diagnostic
[post-trauma trajectory]

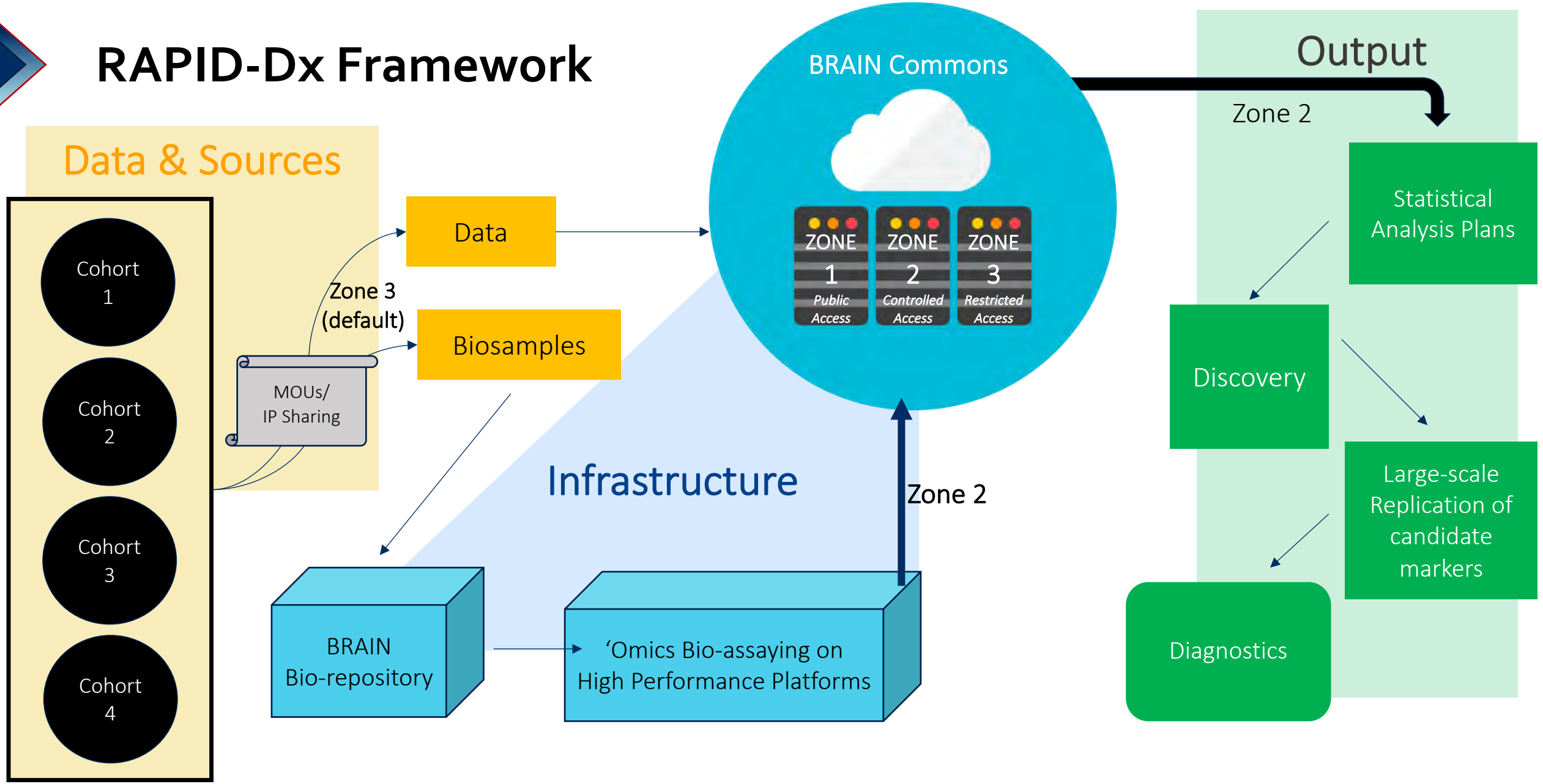
Acute post-exposure prognostic
[supports transition from PCP/ER to psychiatrist/treatment]

Co-diagnostic for therapeutic Response and monitoring

For chronic disease, provides an "index" of severity based on biotype domain

Cohen Veterans Bioscience

RAPID-Dx Framework



PTSD/TBI Biotyping

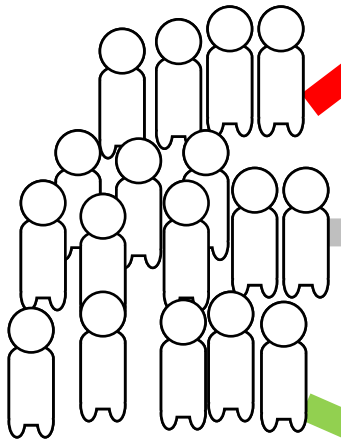
Controls

Class discovery

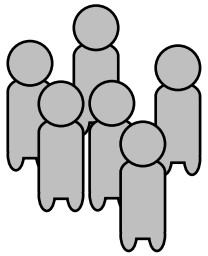
Class prediction

Biological Mechanisms/ Treatments

Patients

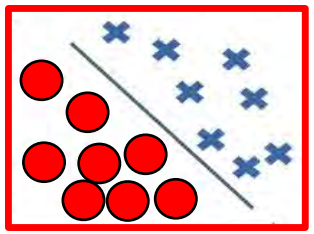


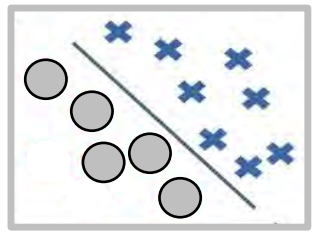
Cluster 1

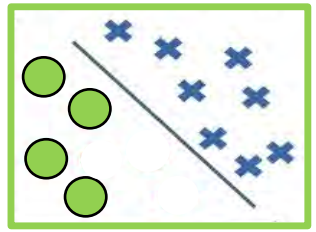


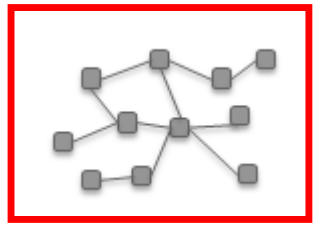
Cluster 2

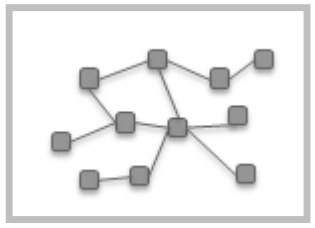
Cluster 3

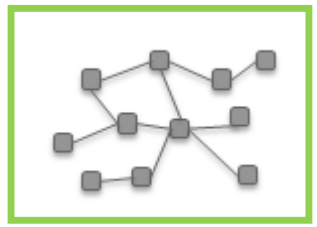








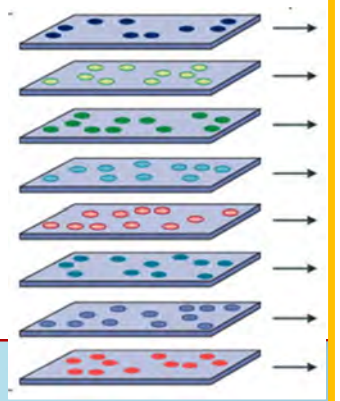




Biotype 1

Biotype 2

Biotype 3



- Imaging
- Physiology
- Immunology
- Metabolomics
- Lipidomics
- Proteomics
- Transcriptomics
- (Epi)Genetics

- PCA
- Factor analysis
- Clustering
- Outlier detection
- …

- Linear regression
- Logistic regression
- Random Forest
- SVM
- …

Cohen Veterans Bioscience

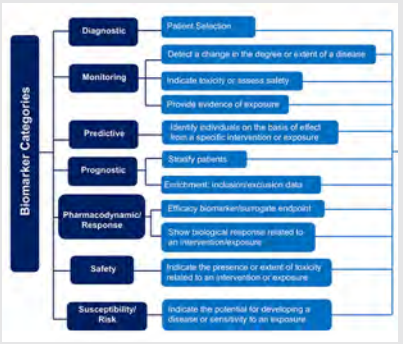
Data Harmonization

- Data models
- Clinical/Behavioral constructs

Analytic

- Sample size
- Pipeline
- Power

Regulatory/ Context of Use



SAP

Modality

- Behavioral
- Genomic
- Humoral
- Neuroimaging

Prior Knowledge

- Literature maps
- Meta-analyses
- Previous SAP results

RAPID-Dx Inflammation Bake-off – Technical Performance and Clinical Dynamic Range

Classes of Samples for Analysis

Technical Samples

- Healthy controls (BioVT)
- HC Stim Pools (BIoVT)
- Recombinant cytokines (BD)

Clinical Samples

- Healthy Controls (BioVT)
- PTSD/TBI (TRACTS, Precision Med)
- PD (MJFF)

Goal: To evaluate both the platforms analytical and technical performance

Goal: To understand likely measurements from individuals w/ PTSD, PD vs. HCs

Although this is not a case-control matched study and lacks statistical power for discovery, these samples will contribute to understanding around whether a given assay offers sensitivity and linear range necessary for assessing endogenous cytokines in our RAPID-Dx cohorts

RAPID-Dx Inflammation Bake-off – Target Analytes

Target Analytes

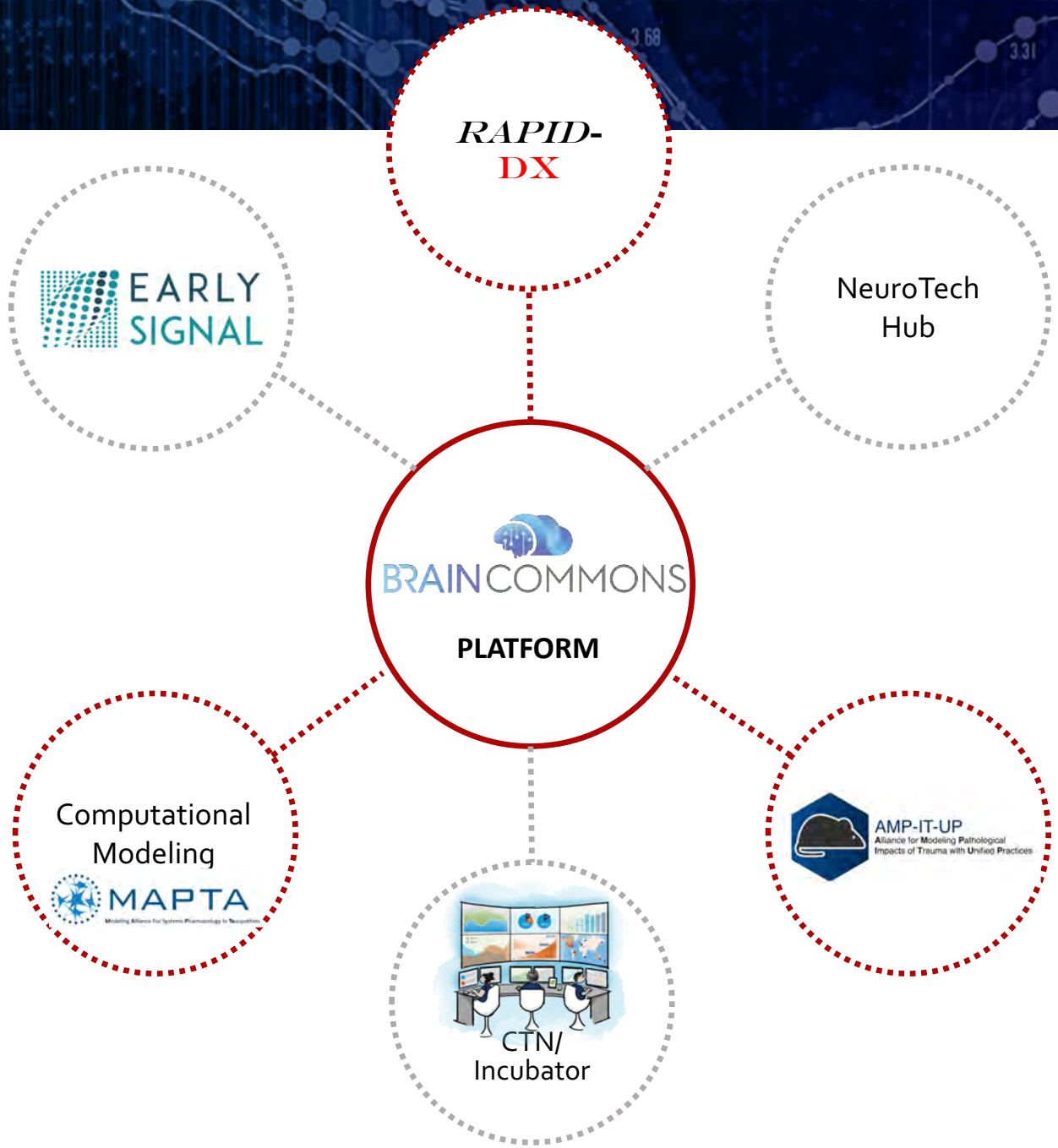
1. IL-1beta
2. IL-2
3. IL-4
4. IL-6
5. IL-8
6. IL-10
7. IL-12
8. IL-12/IL-23
9. IFN-gamma
10. TNF-alpha
11. MCP-1 (cytokine also known as CCL2)
12. Fractalkine (chemokine CX3C)

Cohen Veterans Bioscience

THE MICHAEL J. FOX FOUNDATION FOR PARKINSON'S RESEARCH

A Platform Approach

Our Approach is
to **Build Enabling Platforms**
with **Strategic Partners ,**
incentivizing a **Team Science**
approach
to fast-track solutions.



What is a Data Commons?

Data commons co-locate **data** with **cloud computing** infrastructure and commonly used **software services, tools & apps** for managing, analyzing and sharing data to create an **interoperable resource** for the research community.*

*Robert L. Grossman, Allison Heath, Mark Murphy, Maria Patterson and Walt Wells, A Case for Data Commons Towards Data Science as a Service, IEEE Computing in Science and Engineer, 2016. Source of image: The CDIS, GDC, & OCC data commons infrastructure at a University of Chicago data center.

TRANSLATIONAL RESEARCH NEEDS A NEXT-GENERATION DATA COMMONS

DATA SCALABILITY

Volume, speed and complexity of data growing beyond current capacity

High-throughput molecular analyses, neuroimaging and sensor technologies are generating petabyte datasets daily

Relational database/non-cloud architectures are static and have limited scalability

COMPUTE

Predictive modeling analytics need to be performed in the cloud where data are stored and data standards will need to be adopted

It is no longer practical to move large datasets (time, expense, security)

Integrating multiple datatypes into computational models is not trivial and requires high compute speeds for machine learning / AI approaches

To enable data integration across cohorts / clouds we need to utilize Common data models / Standards / Ontologies / APIs

USABILITY

Data repositories are under-utilized: users are challenged by uploading & accessing data; limited bioinformatics training; lack of incentives

Data Curation support for legacy datasets; ETLs for future

Training – online courses and hackathons to educate users

Visualization Tools – need a range of tools for researchers to manage and explore data

Transparency & Reproducibility - need data provenance tools – Jupyter notebooks, analytic scripts, DOIs, annotation to make design & analysis of BM & PC studies open

SUSTAINABILITY

Many platforms funded by short-term government grants; no sustainable biz model built in to update/improve platform, support data storage, support compute

BRAIN COMMONS – SCALABLE, SECURE, SUSTAINABLE

GEN 3 BIONIMBUS PLATFORM

Scalable, interoperable, big data cloud-based platform

USER INTERFACE & PIPELINES

User-friendly web-interface, visualization tools & suite of analytics pipelines

COGNITIVE COMMUNITY

Social media-based cognitive network that links researchers, analysts, patients and clinicians to knowledge.

SUSTAINABLE BUSINESS MODEL

Membership, fee-based & XAP-store revenue models for long-term sustainability of platform

10 Platform Requirements

Data Volume
No. 1

- Petabyte Scale
- (FAIR)
- Data Driven Discovery

Data Variety
No. 2

- Unstructured
- Heterogeneous

No Constraints on Raw Data Type

Data Velocity
No. 3

- High Throughput Approaches

Data flows - streaming, ingest - processing

Data Value
No. 4

- Complex Analytics

Best in class analytics and bioinformatics tools, workflows, pipelines

Data Volume
No. 5

- Globally - Cloud

Durability, redundancy, survivability, longevity, platform sustainability, recoverability, reproducibility

Computation
No. 6

- Workflows, Pipelines

Finding, computing style, configurability, complex analytics, locality of reference. "One size does not fit all."

Interoperability
No. 7

- Heterogeneous Data

Seamless , APIs, restful services, global , concurrency, standards, open source.

Privacy & Security
No. 8

- Global Compliance

HIPPA HITECH, EU-GDPR, FISMA, NIST, NIH-BD2K, FAIR, GA4GH, VA, DOD, FDA, BIDS, FedRAMP, etc.

Scalability
No. 9

- Governance

Identity management, accounting and auditability, reusability, raw and curated data

Sustainability
No. 10

- Reusability

Retainability, protectability, survivability, funded, affordable

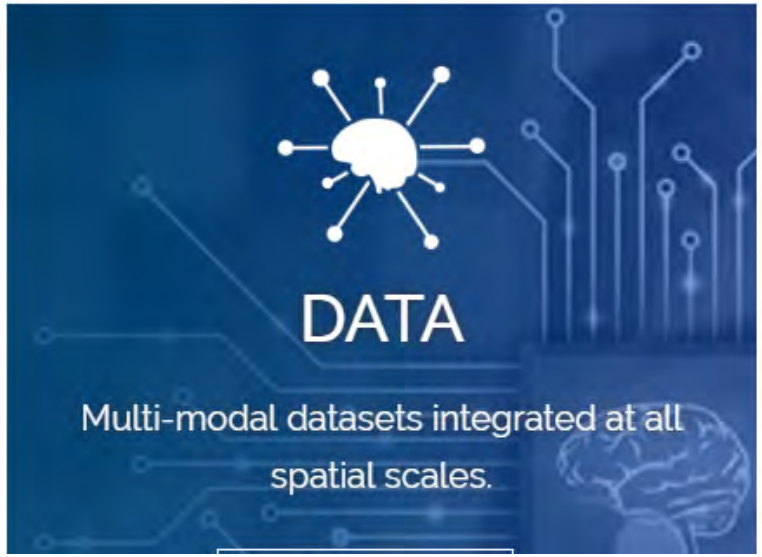
Platforms

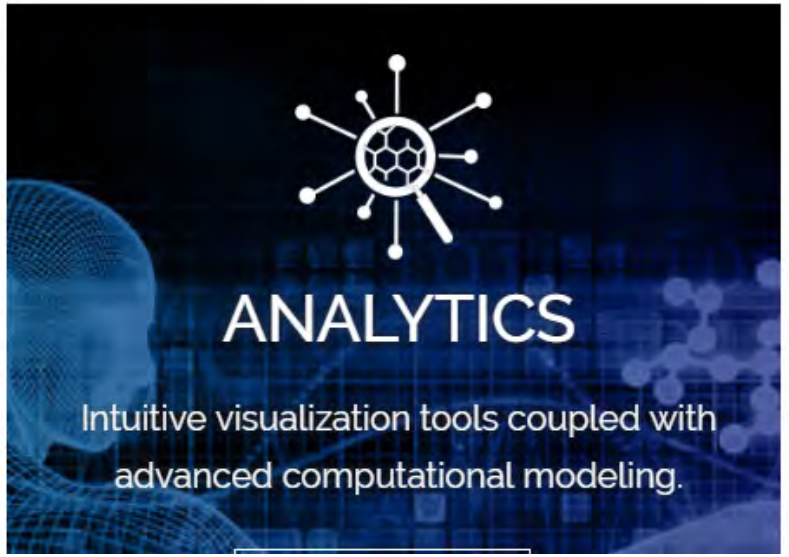
90+

PLATFORMS AND SOFTWARE TECHNOLOGIES EVALUATED

#1. NCI Genomic Data Commons (GDC)	#2. tranSMART Knowledge Management Platform	#3. Informatics for Integrating Biology and the Bedside (i2b2)	#4. Ontario Brain Institute (Brain-CODE)	#5. EU EPILEPSIAE Database
#6. IEEG.org – International Epilepsy Electrophysiology	#7. NSF Cloud Platforms - Computing in the Cloud	#8. NIMH Data Archive - National Institute of Mental Health	#9. MIT "SuperCloud"	#10. HPI Hasso Pattner Institute - Univ of Potsdam
#11. EMC – Pivotal - Large Scale Hadoop Testbed	#12. Perkin Elmer – "Signals"	#13. PMI (Precision Medicine Initiative) New York Genome Center + IBM	#14. The Open Cloud Consortium – Open Science Data Cloud	#15. CG HUB from The Cancer Genome Atlas (TCGA)
#16. Cancer Genome Collaboratory - (Canada)	#17. Blackflynn	#18. "Genome Bridge" – The Broad	#19. IBM Watson Health & IBM Watson Health Cloud	#20. MVP - Million Veterans Program (GenISIS)
#21. Intel PCCSB - Intel Parallel Computing Center Structural Biology	#22. Collaborative Cancer Cloud - Intel	#23. LONI Laboratory of Neuro Imaging - IDA Image and Data Archive (USC)	#24. European Open Science Cloud	#25. ICGC Data Portal

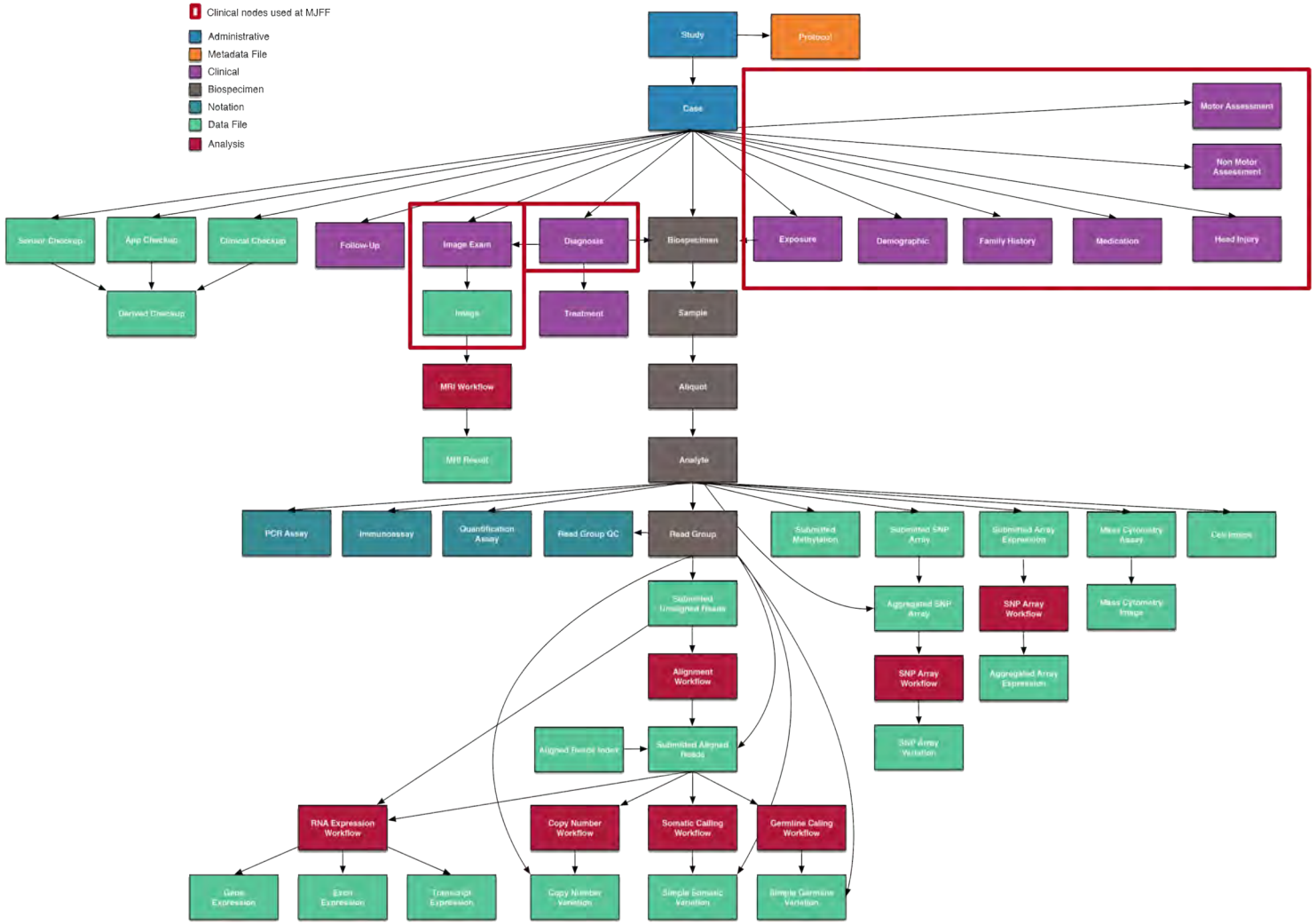
3 Pillars





Team Science Principles
Community Driven to foster collaboration and innovations
Accelerate translational research into the clinic

Data: Data Model



DATA MODEL

- Graph Based
- Flexible, scalable
- Unique for brain disorders
- Each contributed dataset it mapped to the global data model
- Allows for 'control' of access at the level of nodes for a given dataset
- Linked to CDISC & other relevant standards

DATA TYPES

- Clinical data
- Imaging data
- Genomic data
- Biospecimen data
- Sensor/Wearable data (streaming)
- Preclinical (under construction)

Unified graph-based data model, capable of supporting heterogeneity of brain data

Data: Zones

ZONE 1
Public Access

- Similar to PubMed access
- Open to public without qualification
- Metadata describing the data available in Brain Commons will also be Public Data

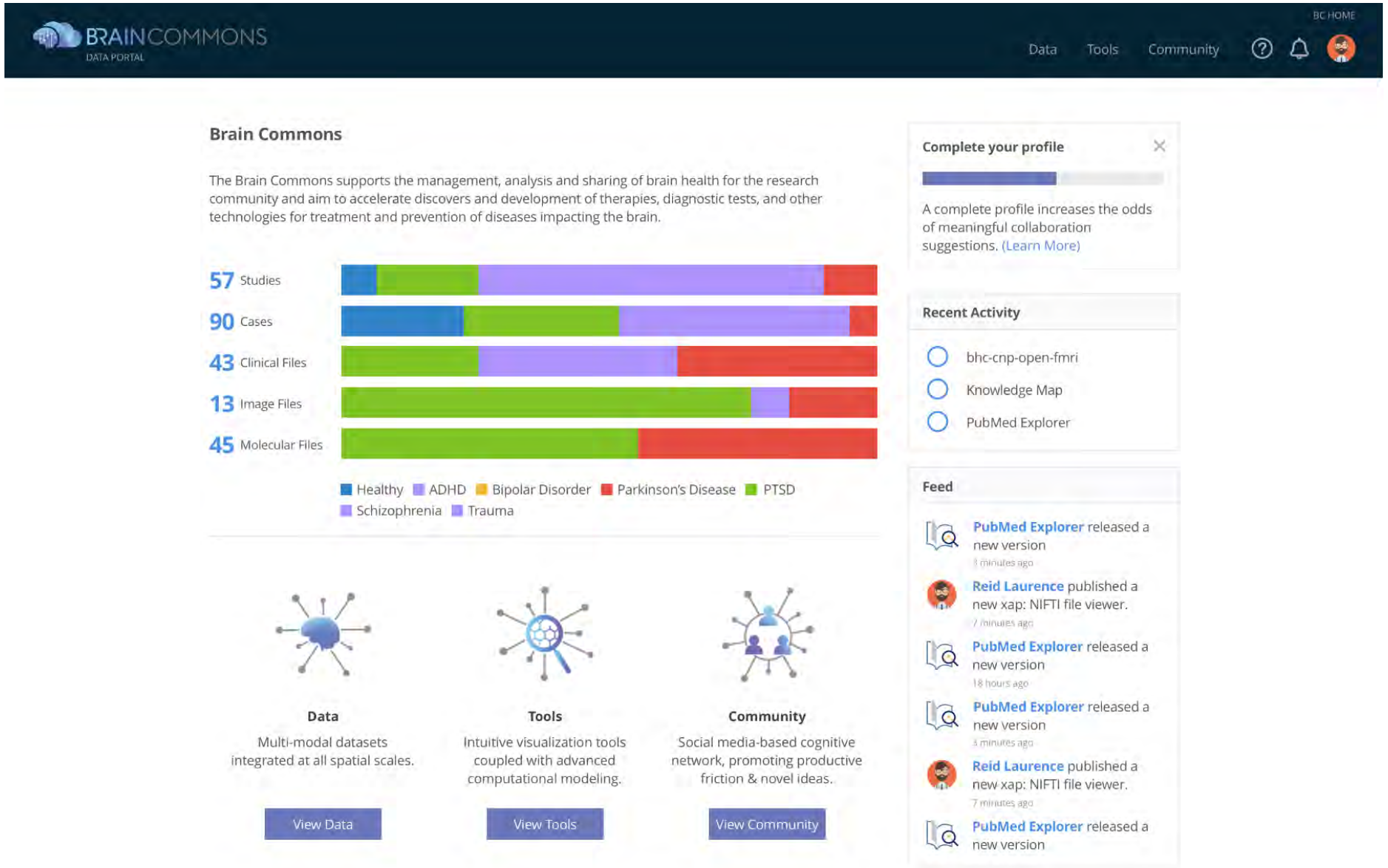
ZONE 2
Controlled-Access

- Similar to dbGAP access
- Qualified researchers with Data User Agreement (DUA)

ZONE 3
Restricted-Use Only

- Individual institutions/users approved by funding organization
- Access prescribed at data cohort level

Portal


BRAINCOMMONS
DATA PORTAL
BC HOME
Data
Tools
Community
Brain Commons
The Brain Commons supports the management, analysis and sharing of brain health for the research community and aim to accelerate discovers and development of therapies, diagnostic tests, and other technologies for treatment and prevention of diseases impacting the brain.
57 Studies
90 Cases
43 Clinical Files
13 Image Files
45 Molecular Files
Healthy
ADHD
Bipolar Disorder
Parkinson's Disease
PTSD
Schizophrenia
Trauma
Data
Multi-modal datasets integrated at all spatial scales.
View Data
Tools
Intuitive visualization tools coupled with advanced computational modeling.
View Tools
Community
Social media-based cognitive network, promoting productive friction & novel ideas.
View Community
Complete your profile
A complete profile increases the odds of meaningful collaboration suggestions. (Learn More)
Recent Activity
bhc-cnp-open-fmri
Knowledge Map
PubMed Explorer
Feed
PubMed Explorer released a new version
3 minutes ago
Reid Laurence published a new xap: NIFTI file viewer.
7 minutes ago
PubMed Explorer released a new version
18 hours ago
PubMed Explorer released a new version
3 minutes ago
Reid Laurence published a new xap: NIFTI file viewer.
7 minutes ago
PubMed Explorer released a new version

Analytics

- Exposing the value of the data to the Brain Commons community
- Encourage analysis of the data ON the commons

CASUAL USER
Querying PubMed, latest research articles and trends

BIOINFORMATICIAN
Analysis of high dimensional data, linking statistical analysis to the functional biology

MACHINE LEARNING ENGINEER
Algorithm development

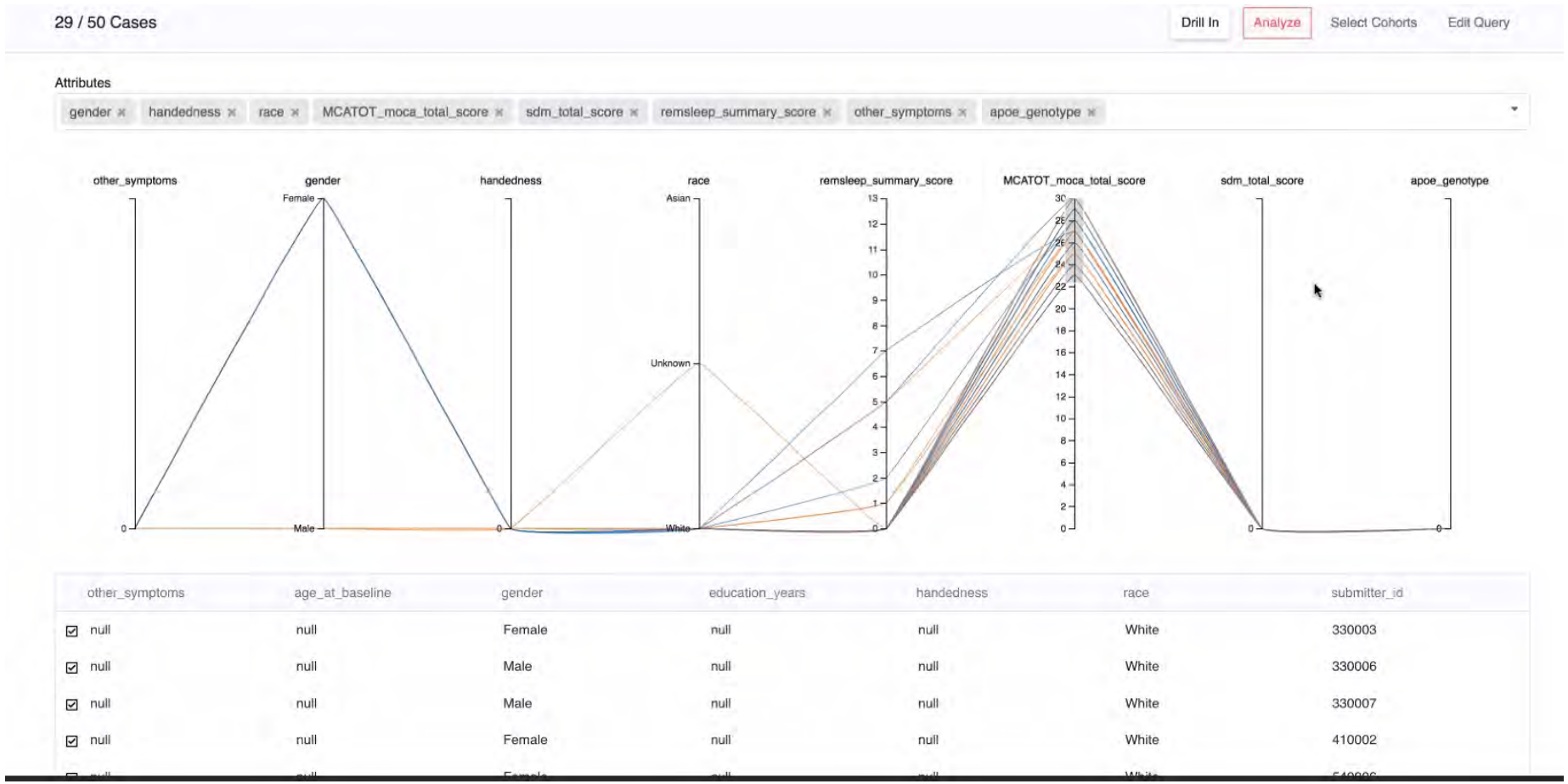
PRE-CONFIGURED TOOLS

NOTEBOOKS & DEVELOPMENT STUDIO

BIOLOGIST
Data exploration, visualization, basic statistics

DATA SCIENTIST
Applying advanced statistical modelling, biomarker discovery

Cohort Explorer: Faceted Search



Cohort Comparison

Step 1. Select cohort from saved search, or ‘live’ from faceted search

Step 2. Perform basic statistical analysis on selection

Select 2 case sets

You can create and save case, gene and mutation sets of interest from the Exploration Page

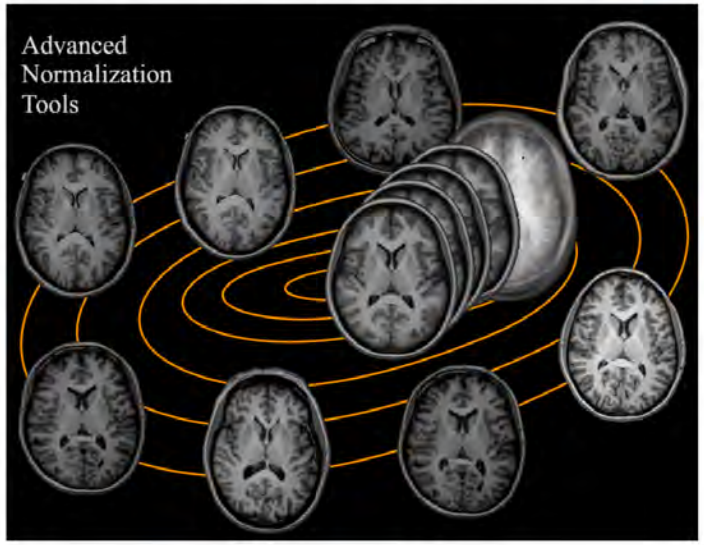
	Type	Name	Items
☐	Cases	test	183
☐	Cases	Kidney	2,089
☐	Cases	Colon	1,838

Notebooks

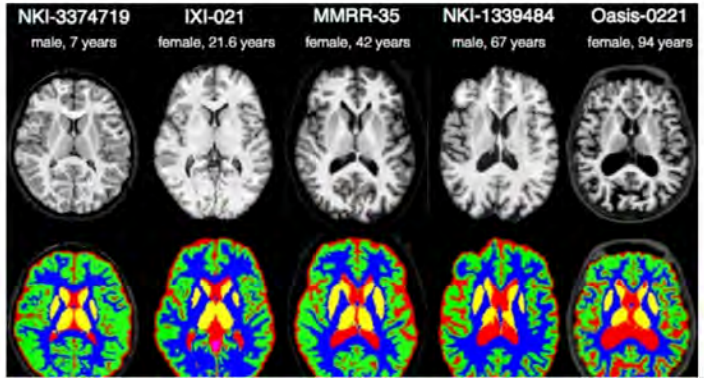
Advanced Normalization Tools

build unknown

ANTs computes high-dimensional mappings to capture the statistics of brain structure and function. See the FAQ page.



ANTs allows one to organize, visualize and statistically explore large biomedical image sets.



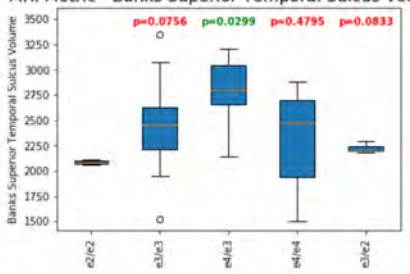
MRI Volume comparison (Wilcoxon non-parametric test)

Compare MRI Volume measures (banks superior temporal sulcus, caudal anterior cingulate and caudal middle frontal) for the PPMI cohort by genotype.

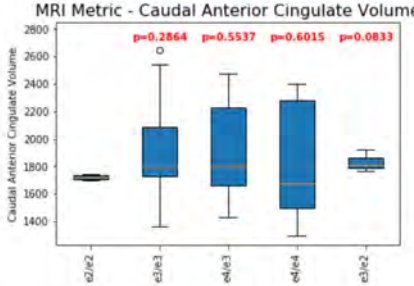
A **Wilcoxon rank-sum statistical test (pvalue < 0.05)** is applied to determine statistically significant differences against a baseline genotype passed as parameter:

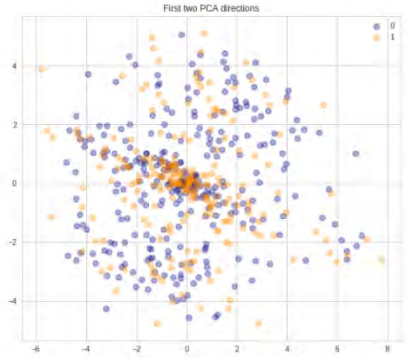
In [24]:
```
data_bank = bc.get_mri_subfield_by_genotype(project_id, 'apoe_genotype', 'Banks Superior Temporal Sulcus Volume', 'e2/e2')
data_cauant = bc.get_mri_subfield_by_genotype(project_id, 'apoe_genotype', 'Caudal Anterior Cingulate Volume', 'e2/e2')
data_caumid = bc.get_mri_subfield_by_genotype(project_id, 'apoe_genotype', 'Caudal Middle Frontal Volume', 'e2/e2')
```

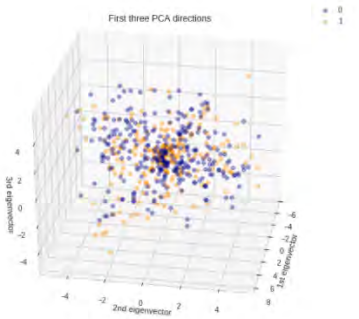
MRI Metric - Banks Superior Temporal Sulcus Volume



MRI Metric - Caudal Anterior Cingulate Volume







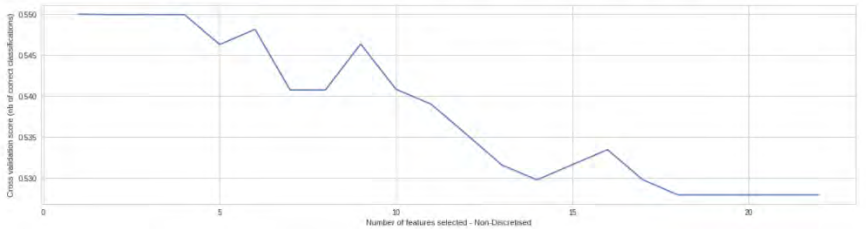
Recursive Feature Elimination

Feature ranking with recursive feature elimination and cross-validated selection of the best number of features.

In [29]:
```
# Calculating RFE for non-discretised dataset, and graphing the Importance for each feature, per dataset
selector1 = RFECV(LogisticRegression(), step=1, cv=5, n_jobs=-1)
selector1 = selector1.fit(df.drop('class', axis=1).values, df['class'].values)
print("Feature Ranking For Non-Discretised: %s" % selector1.ranking_)
print("Optimal number of features : %d" % selector1.n_features_)
# Plot number of features VS. cross-validation scores
plt.style.use('seaborn-whitegrid')
plt.figure(figsize=(20,5))
plt.xlabel("Number of features selected - Non-Discretised")
plt.ylabel("Cross validation score (nb of correct classifications)")
plt.plot(range(1, len(selector1.grid_scores_) + 1), selector1.grid_scores_);

# Feature space could be subsetted like so:
df_con_enc = df[df.columns[np.insert(selector1.support_, 0, True)]]
```

```
Feature Ranking For Non-Discretised: [ 8 15 20  6  4  3 19 13  2 16 11 14  7 17 12 10 18  1 22  5  9 21]
Optimal number of features : 1
```



Partnerships

THE MICHAEL J. FOX FOUNDATION
FOR PARKINSON'S RESEARCH

THE MICHAEL J. FOX FOUNDATION PARTNERS WITH COHEN VETERANS BIOSCIENCE ON BRAIN COMMONS, HARNESSING THE POWER OF BIG DATA FOR BRAIN DISEASES

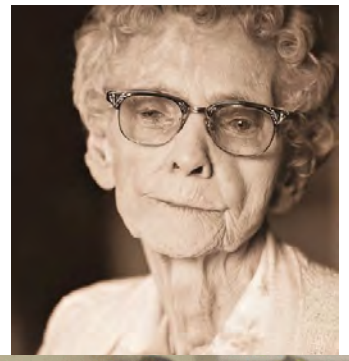
"Data sharing is critical to research discoveries," says MJFF CEO Todd Sherer, PhD. "By including data across neurodegenerative diseases, BRAIN Commons facilitates research collaboration, data exploration and reproducibility. The insights it creates will accelerate the development of new therapies for Parkinson's and similar diseases."

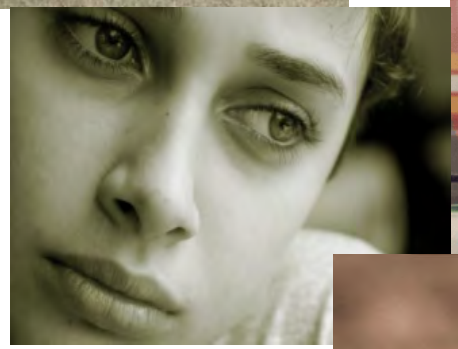
Partnerships

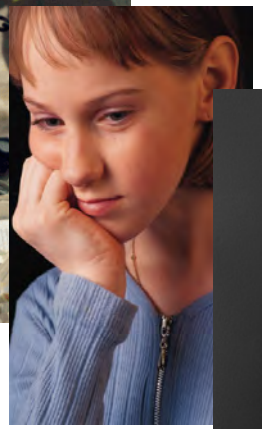
The Gen3 platform, multiple geographically distributed data commons can interoperate in different ways:

- through data peering
- through a FAIR-based set of APIs for applications
- through scattering queries/analyses and gathering the results
- through a controlled and monitored query/analysis gateway

PATIENTS ARE WAITING!







What It Takes

- Change the Conversation
- Build the Translational Toolbox
- Strategic Alliances
- Roadmap
- Leadership & Engagement
- It Takes You!

Proposal: Form a Trans-Atlantic e-Brain Consortium?

- Coalition of the willing
- Partner across “platforms”
- Establish a federated inter-operable data-sharing framework
- Network of networks

Cohen Veterans Bioscience

THANK YOU FOR YOUR ATTENTION!

www.cohenveteransbioscience.org