ONTARIO NEURODEGENERATIVE DISEASE RESEARCH INITIATIVE

# ONDRI Data Curation: From REDCap to Release

Kelly Sunderland and Derek Beaton

2018FEB07

ONDRI

ONTARIO BRAIN INSTITUTE | INSTITUT ONTARIEN DU CERVEAU

# Outline

- Overview of ONDRI
- Overview of curation cycle
  - Pre-processing
  - Data Standards
  - Outliers
- ONDRI and Brain-CODE in the future

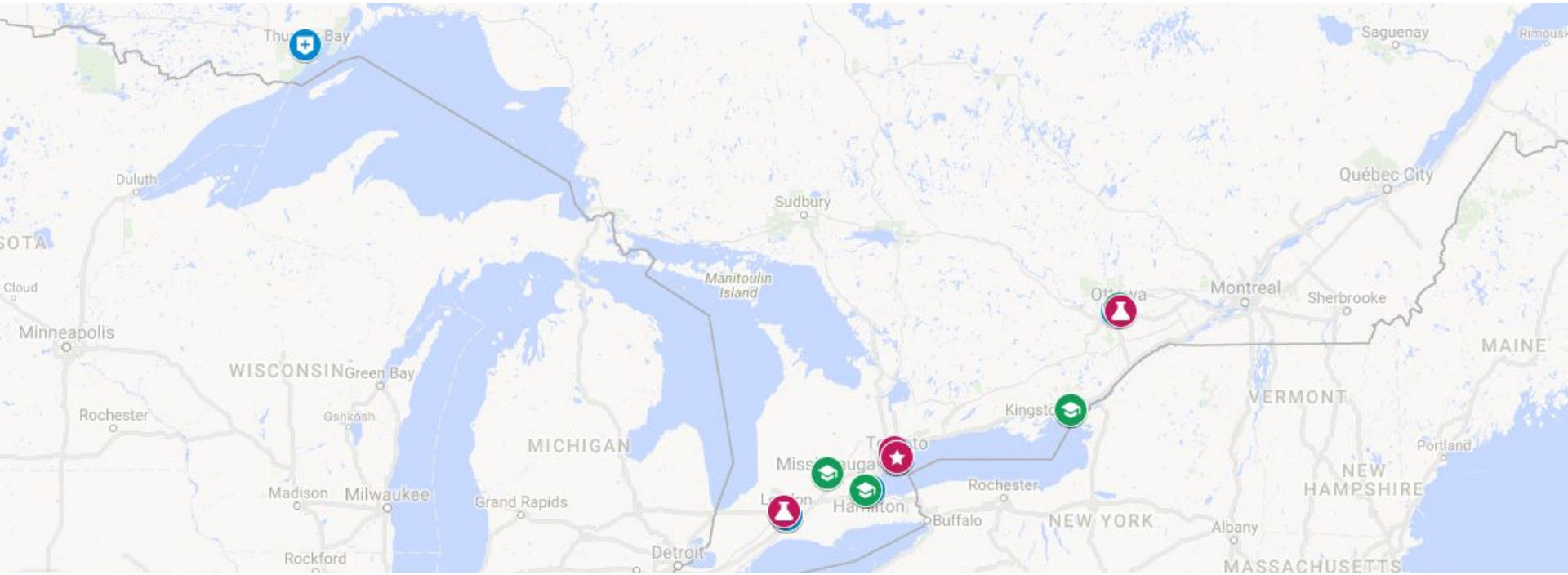
# Outline

- Overview of ONDRI
- Overview of curation cycle
  - Pre-processing
  - Data Standards
  - Outliers
- ONDRI and Brain-CODE in the future

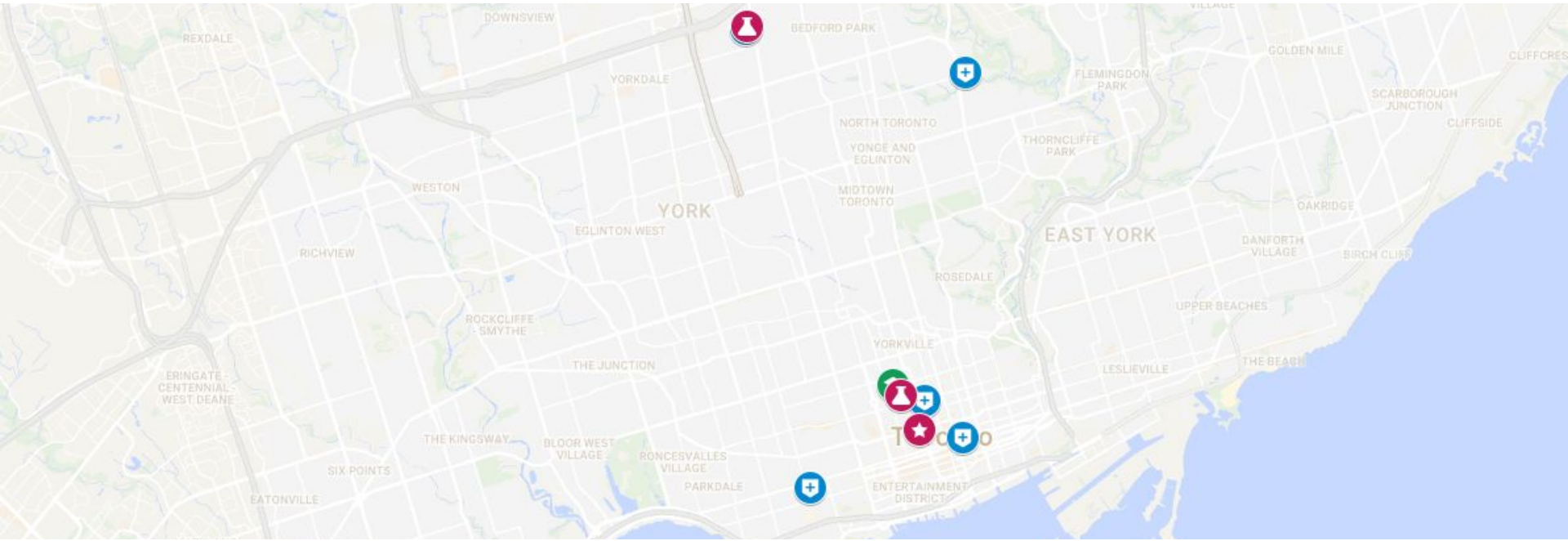
# ONDRI

- Longitudinal
- Multi-cohort
  - Alzheimer's & Mild Cognitive Impairment (AD/MCI)
  - Frontotemporal dementia (FTD)
  - Parkinson's Disease (PD)
  - Amyotrophic lateral sclerosis (ALS)
  - Vascular cognitive impairment (VCI)
- Ontario-wide multi-site, collaborative
- Better diagnosis and (eventually) treatment

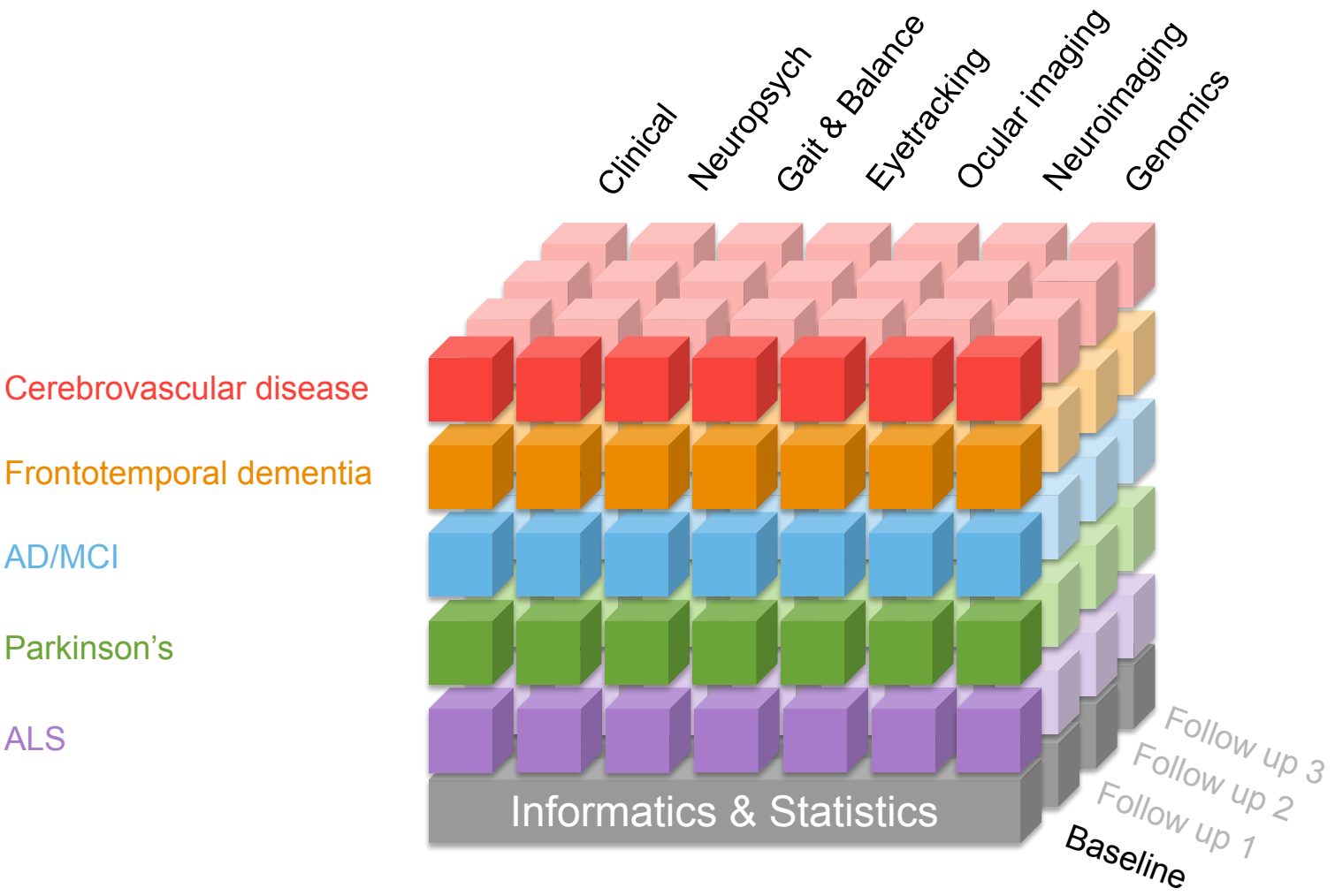
# ONDRI


Thunder Bay
Duluth
Sudbury
Saguenay
Québec City
Manitoulin Island
Minneapolis
WISCONSIN
Green Bay
Rochester
Oshkosh
MICHIGAN
Madison
Milwaukee
Grand Rapids
Rockford
Detroit
Buffalo
Rochester
Montreal
Sherbrooke
MAINE
VERMONT
NEW HAMPSHIRE
NEW YORK
Albany
Portland
MASSACHUSETTS

# ONDRI


DOWNSVIEW
REXDALE
BEDFORD PARK
YORKDALE
FLEMINGDON PARK
GOLDEN MILE
CLIFFCREST
SCARBOROUGH JUNCTION
CLIFFSIDE
NORTH TORONTO
YONGE AND EGLINTON
THORNCLIFFE PARK
WESTON
MIDTOWN TORONTO
YORK
OAKRIDGE
EGLINTON WEST
EAST YORK
RICHVIEW
DANFORTH VILLAGE
BIRCH CLIFF
ROSEDALE
UPPER BEACHES
ROCKCLIFFE-SMYTHE
YORKVILLE
THE JUNCTION
LESLIEVILLE
THE BEACH
ERINGATE-CENTENNIAL-WEST DEANE
THE KINGSWAY
BLOOR WEST VILLAGE
RONCESVALLES VILLAGE
SIX POINTS
PARKDALE
ENTERTAINMENT DISTRICT
EATONVILLE

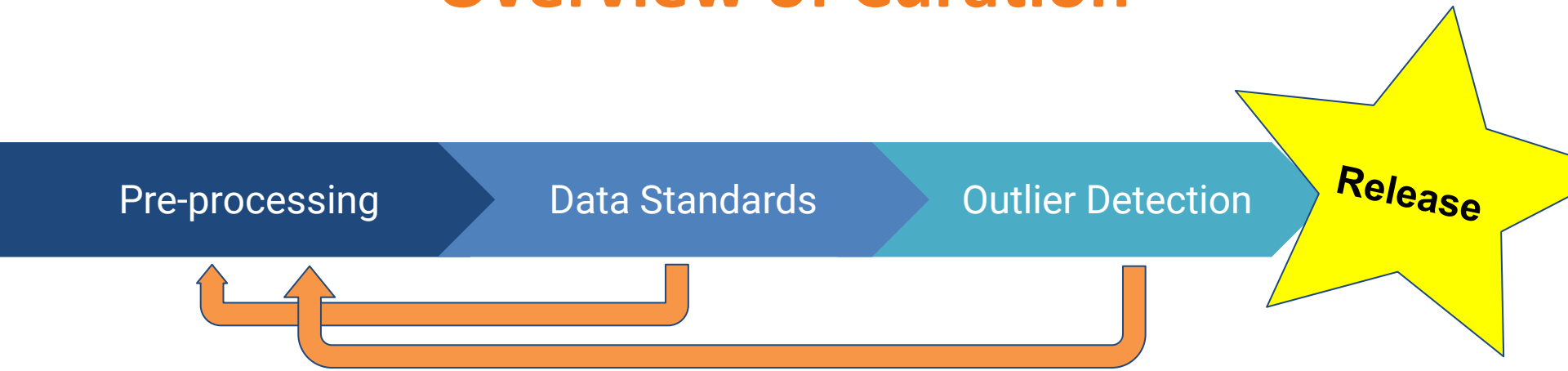
# ONDRI


Clinical
Neuropsych
Gait & Balance
Eyetracking
Ocular imaging
Neuroimaging
Genomics
Cerebrovascular disease
Frontotemporal dementia
AD/MCI
Parkinson's
ALS
Informatics & Statistics
Follow up 3
Follow up 2
Follow up 1
Baseline

# Outline

- Overview of ONDRI
- Overview of curation cycle
  - Pre-processing
  - Data Standards
  - Outliers
- ONDRI and Brain-CODE in the future

# Overview of Curation




- Performed by Platforms
- Prepares data for analysis

- Platform formats dataset to meet data standards
- NIBS verifies adherence
- Returns to pre-processing if does not comply

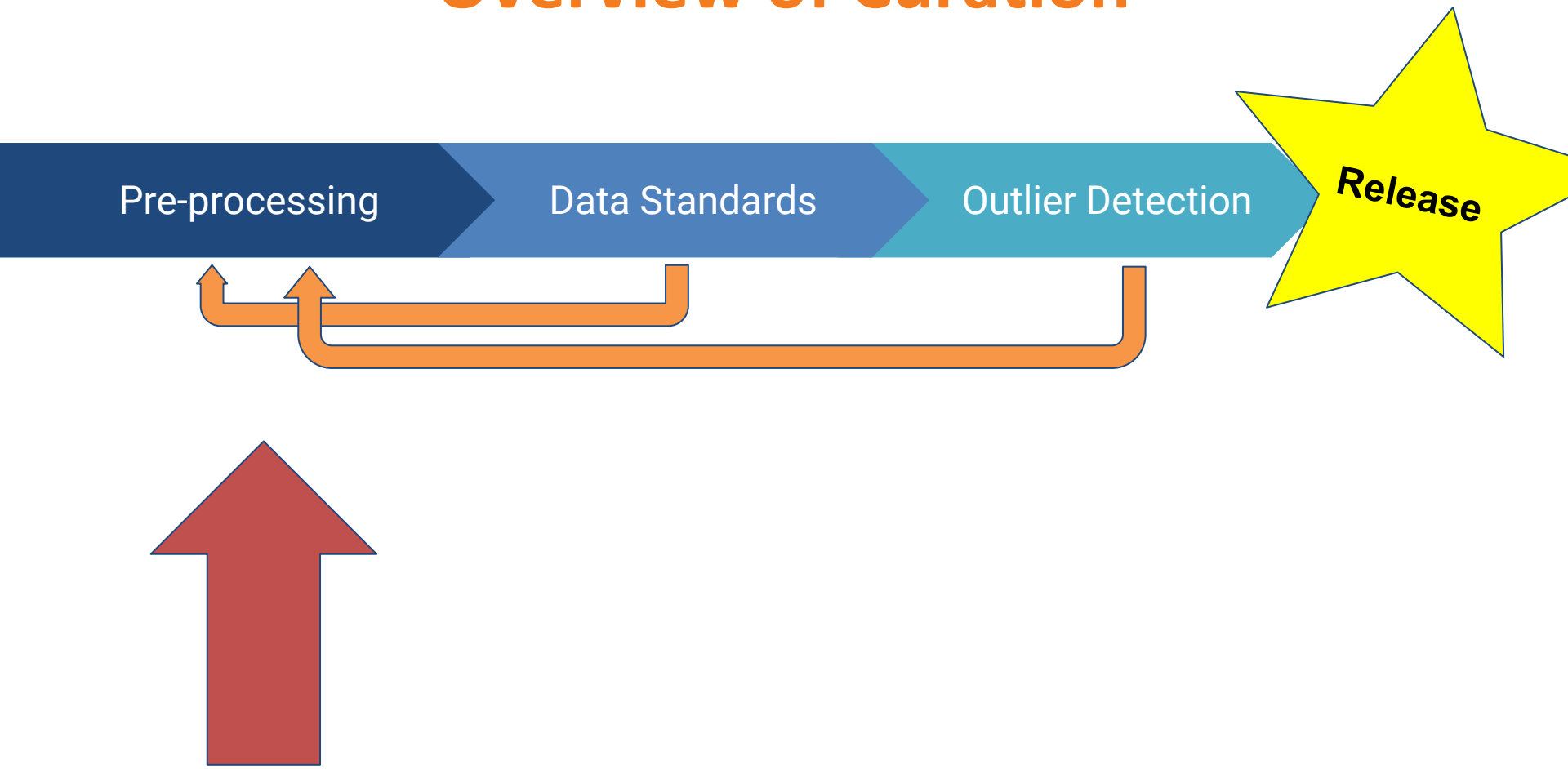
- Performed by NIBS
- Search for observations that appear distinct in relation to other participants
- Returns to pre-processing if errors are found

ONDRI

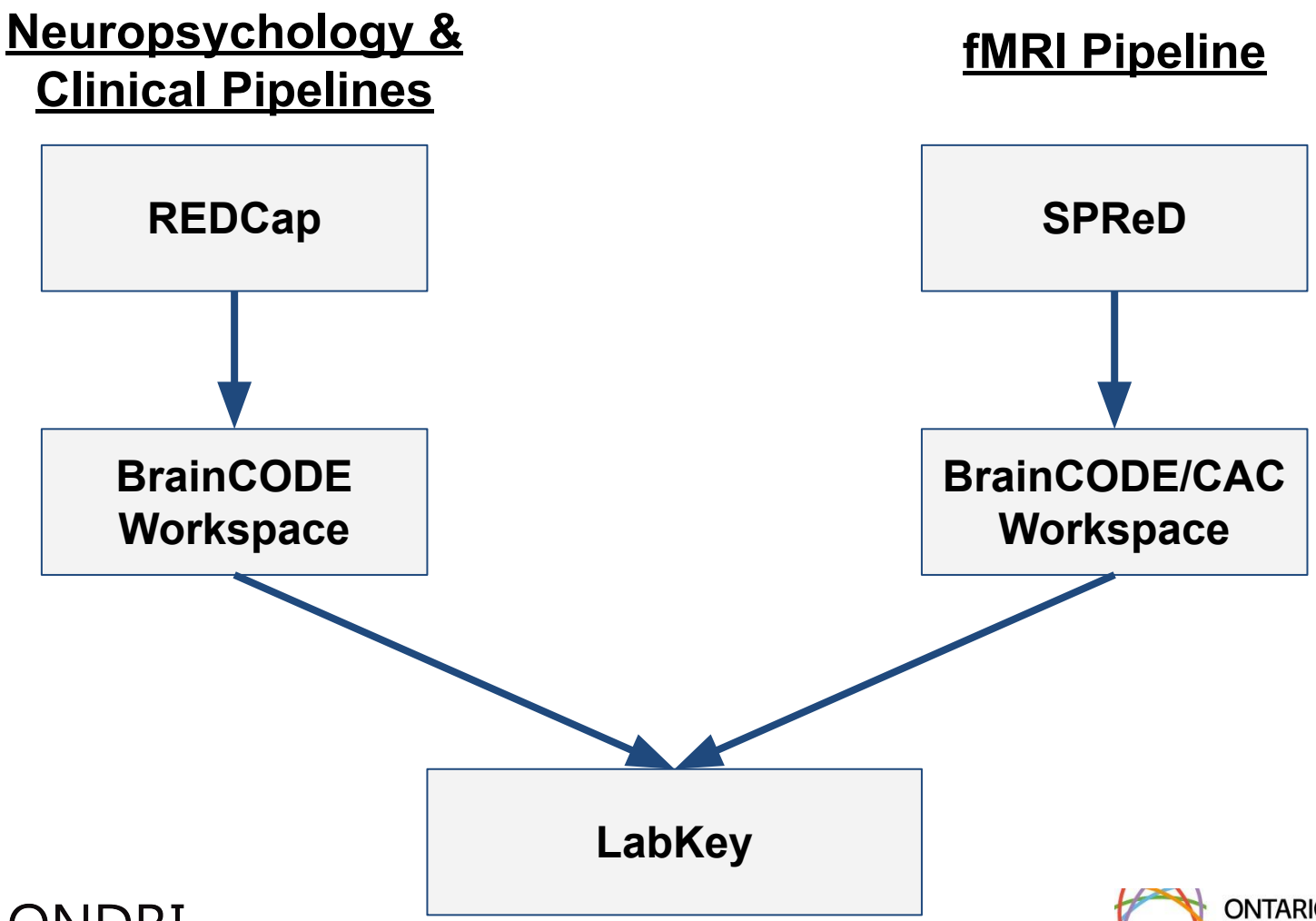
*NIBS = Neuroinformatics & Biostatistics Team

ONTARIO BRAIN INSTITUTE
INSTITUT ONTARIEN DU CERVEAU

# Overview of Curation



# Pre-processing: BrainCODE in Use

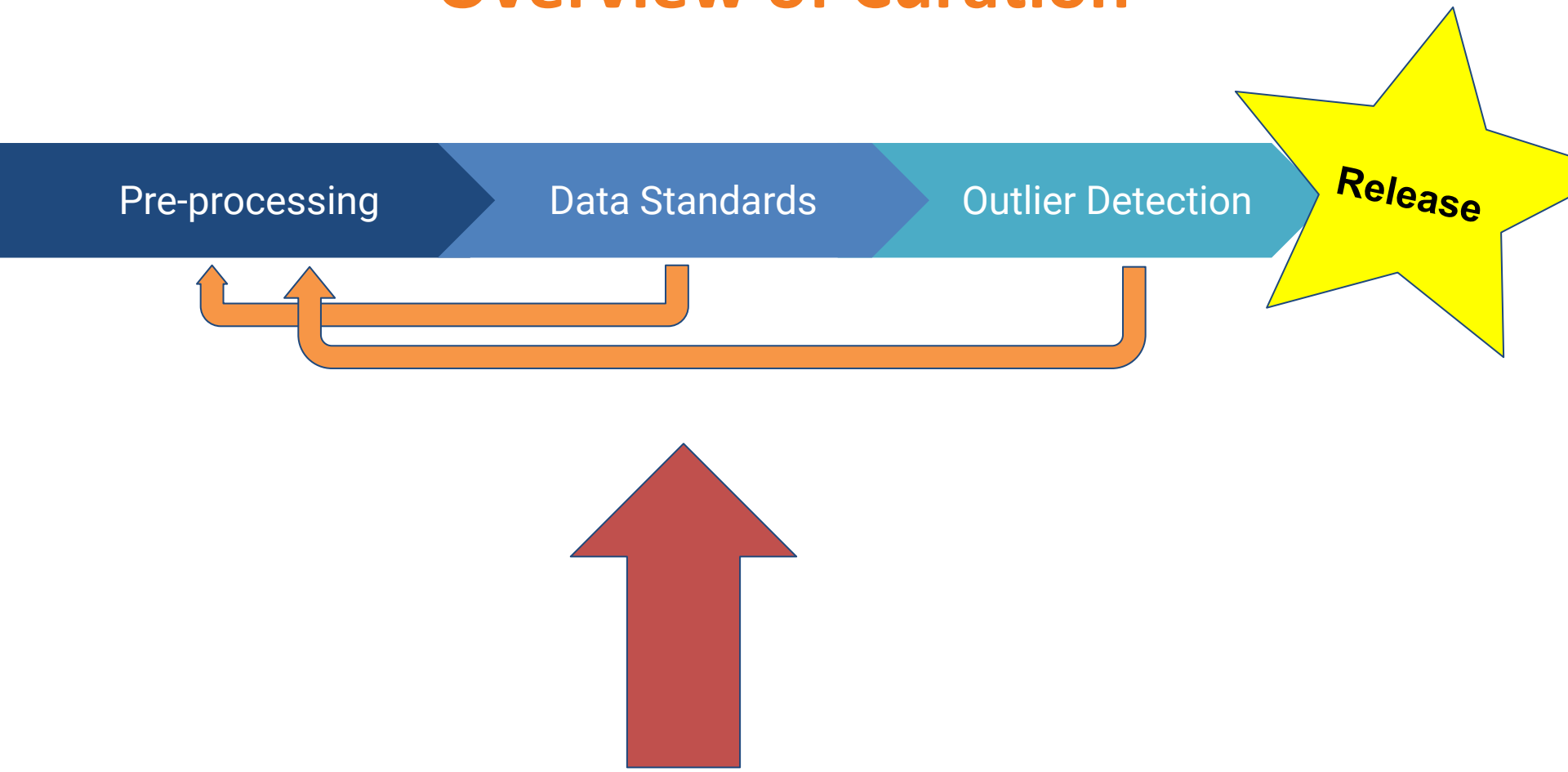



ONDRI



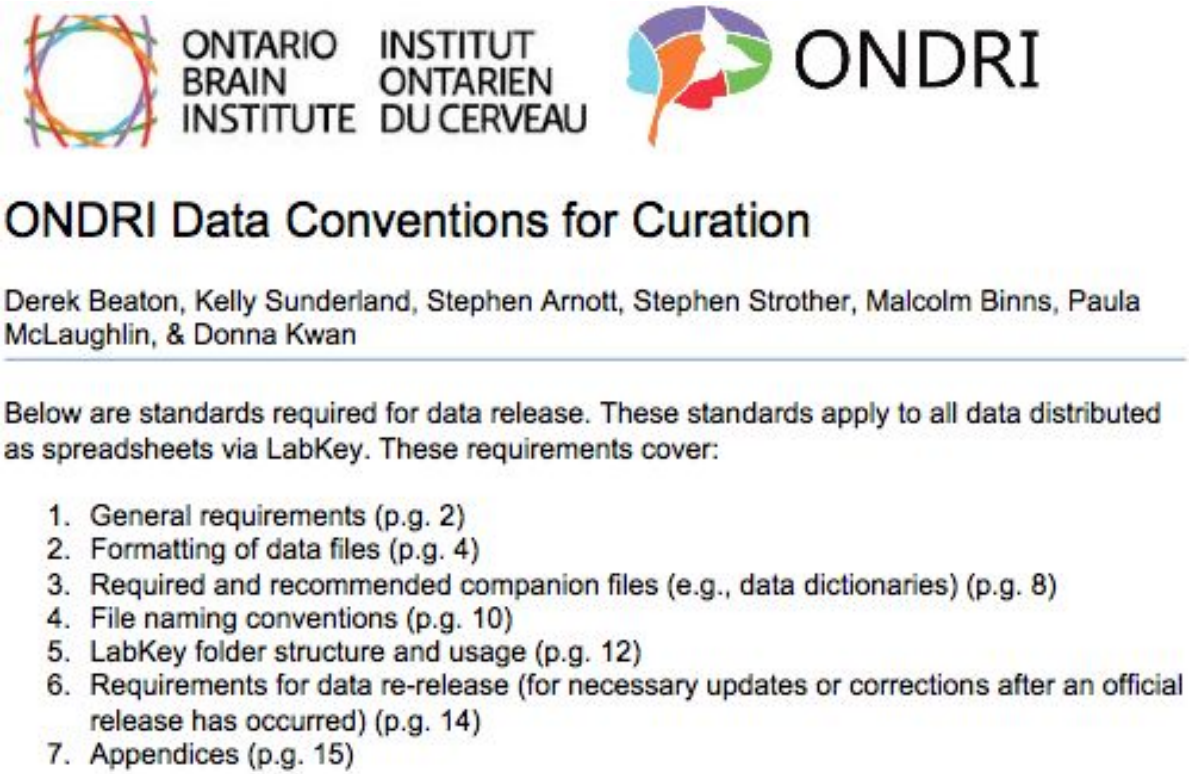
ONTARIO BRAIN INSTITUTE INSTITUT ONTARIEN DU CERVEAU

# Overview of Curation



# Data Standards

- File format & Required set of files
- Subject & date formats
- Visit, site, missing codes



ONTARIO BRAIN INSTITUTE | INSTITUT ONTARIEN DU CERVEAU

ONDRI

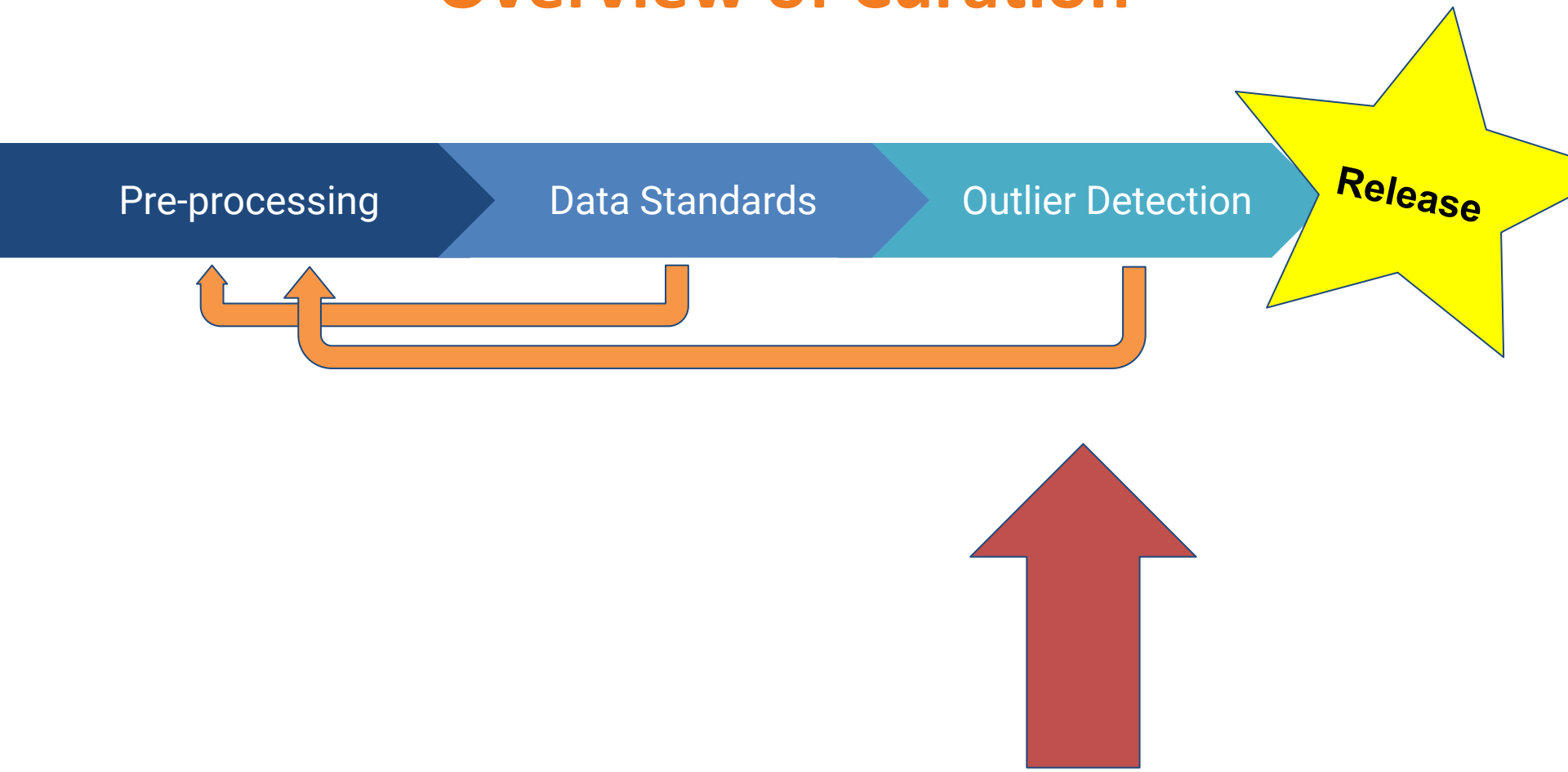
## ONDRI Data Conventions for Curation

Derek Beaton, Kelly Sunderland, Stephen Arnott, Stephen Strother, Malcolm Binns, Paula McLaughlin, & Donna Kwan

Below are standards required for data release. These standards apply to all data distributed as spreadsheets via LabKey. These requirements cover:

1. General requirements (p.g. 2)
2. Formatting of data files (p.g. 4)
3. Required and recommended companion files (e.g., data dictionaries) (p.g. 8)
4. File naming conventions (p.g. 10)
5. LabKey folder structure and usage (p.g. 12)
6. Requirements for data re-release (for necessary updates or corrections after an official release has occurred) (p.g. 14)
7. Appendices (p.g. 15)

# Overview of Curation



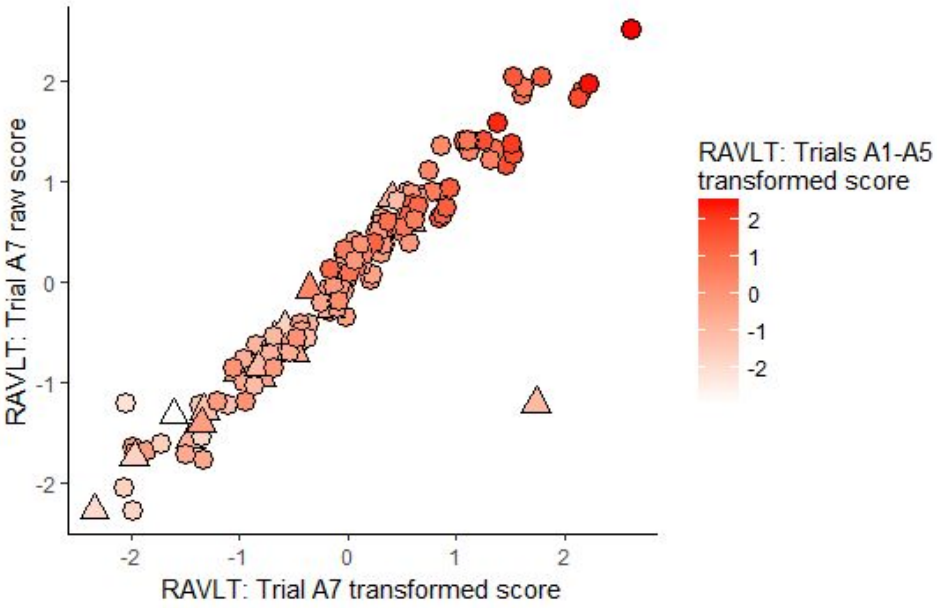
# Outlier Detection

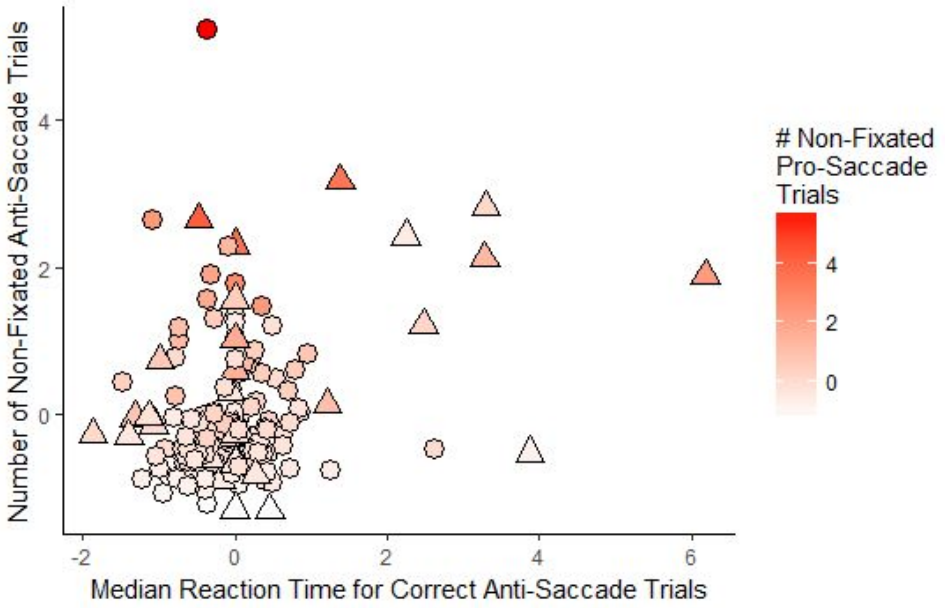
Three methodological components and papers:

1. **Sunderland et al.**, (in prep). Multivariate Outlier Detection is Superior to Univariate Approaches for Data Quality Evaluation in Large Studies
2. **Beaton et al.**, (in prep a). Generalization of the minimum covariance determinant algorithm for categorical and mixed data types
3. **Beaton et al.**, (in prep b). Robust outlier detection for low and high dimensional neuroimaging data with principal components analysis and split-half resampling

Software available: https://github.com/derekbeaton/ours

# Outlier Detection




Sunderland et al., (in prep)

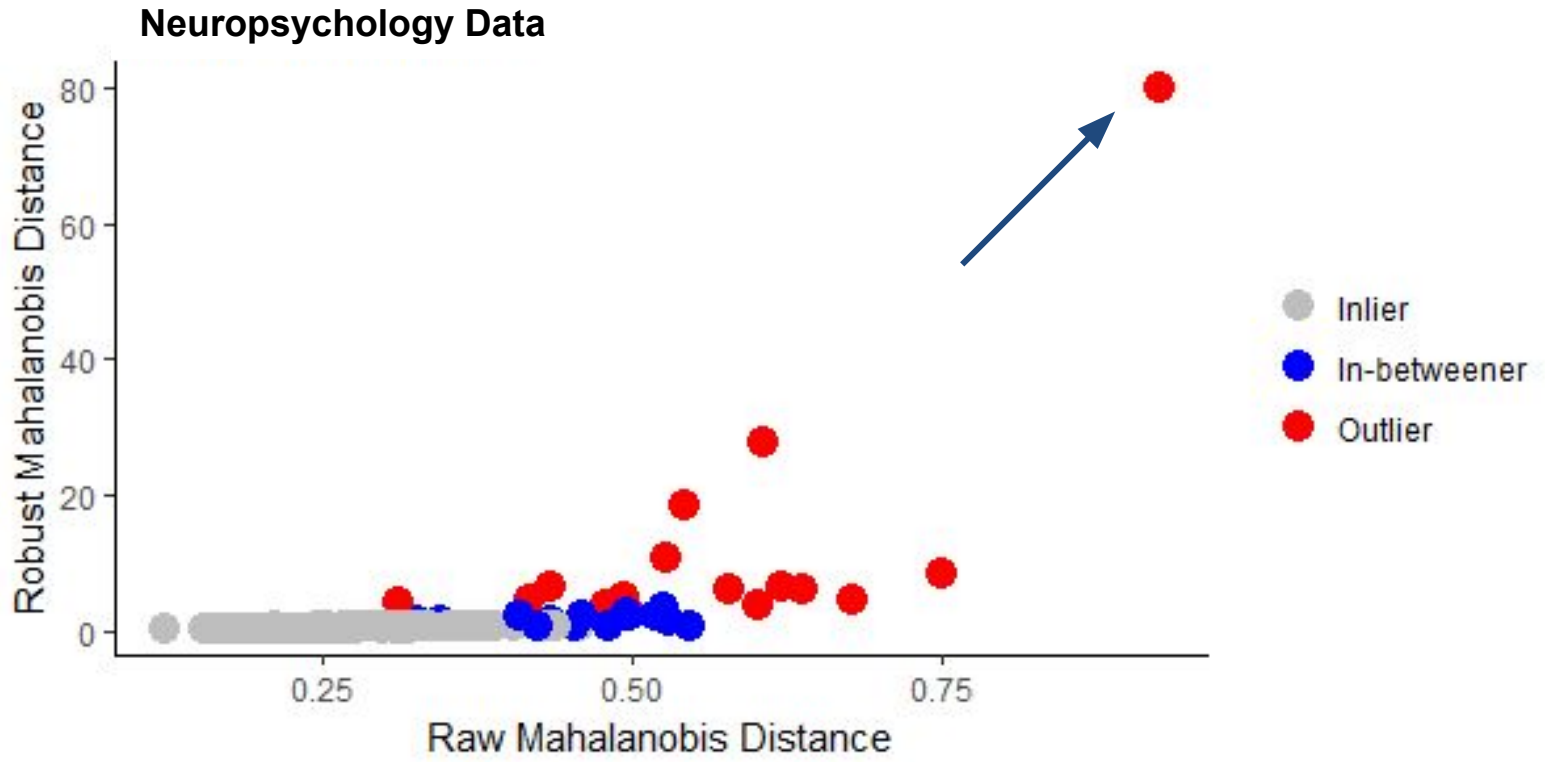
ONDRI



ONTARIO BRAIN INSTITUTE
INSTITUT ONTARIEN DU CERVEAU

# Outlier Detection

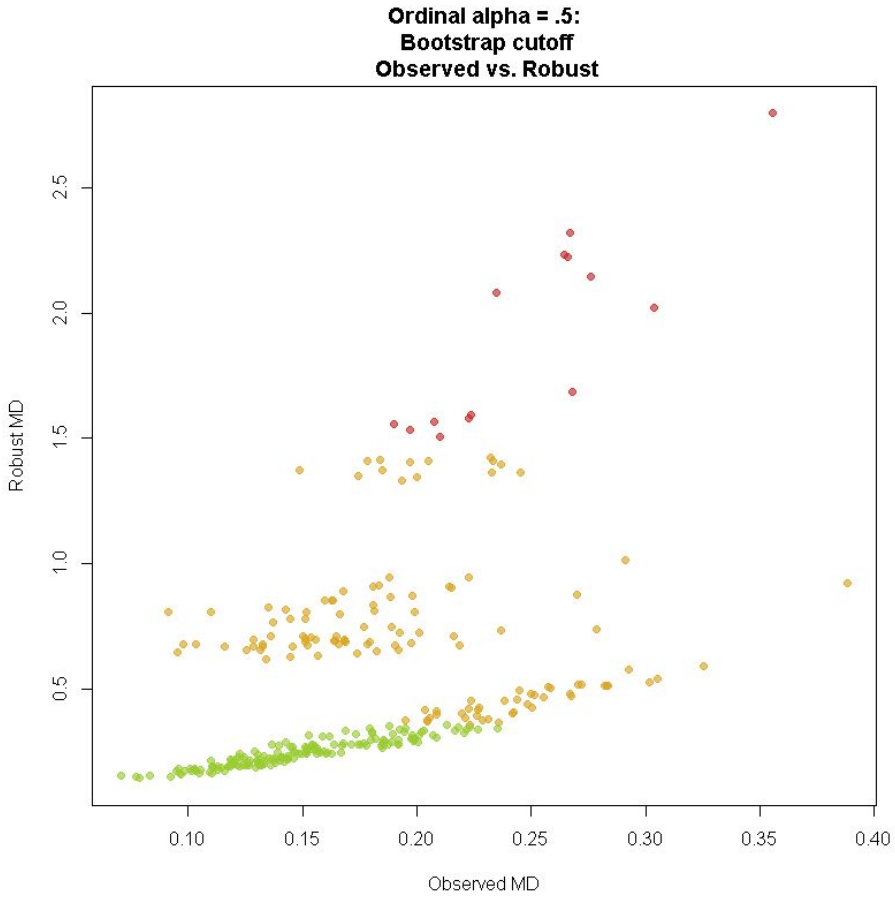
We also identified some truly unique participants!



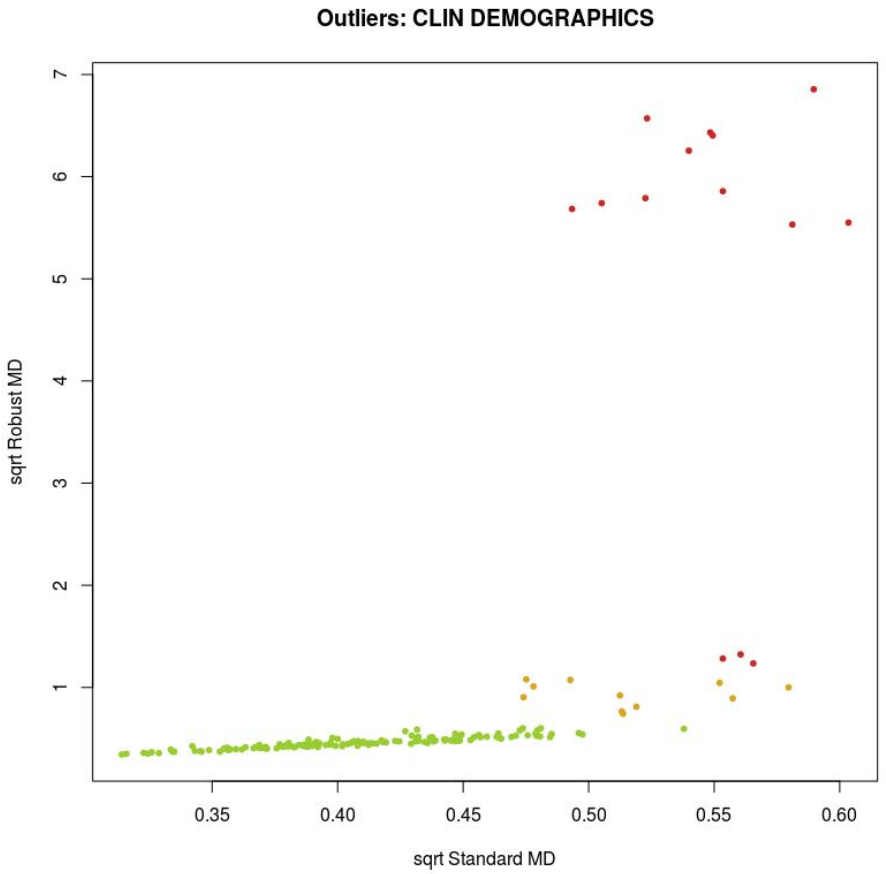
# Outlier Detection

- Proven very effective in ONDRI for identifying:
  - Errors
  - Anomalies
  - Robust subsamples and subspaces
- In use across projects
  - e.g., CAN-BIND
- **Limitation**: Only works for
  - Quantitative data
  - More observations (N) than variables (p)

# Categorical, Ordinal, and Mixed




Ordinal data (survey on autobiographical memory)




Categorical data (Demographics)

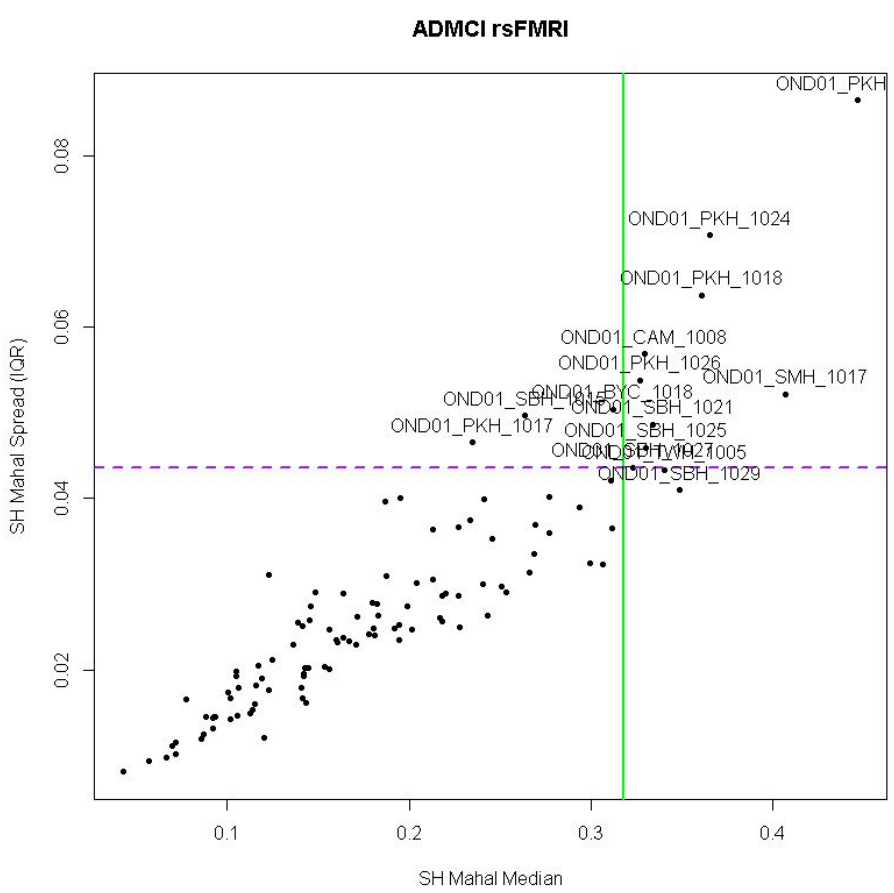
Beaton et al., (in prep a)

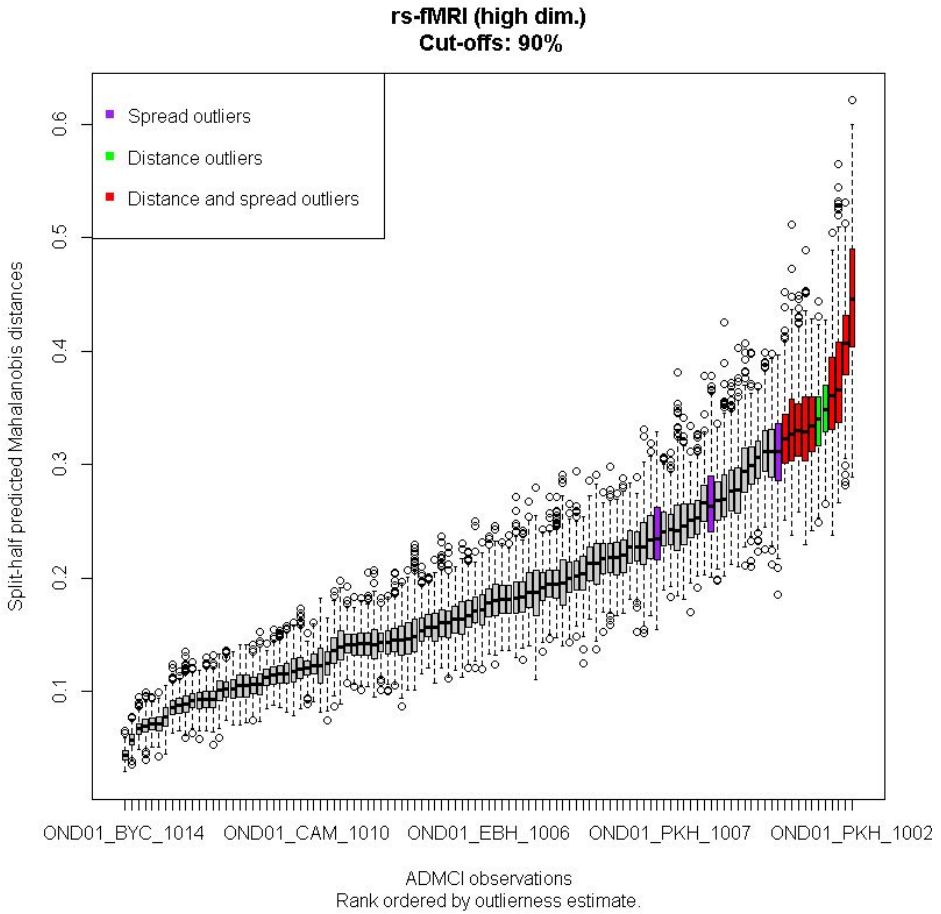
ONDRI



ONTARIO BRAIN INSTITUTE
INSTITUT ONTARIEN DU CERVEAU

# High dimensional: rs-fMRI




Beaton et al., (in prep b)

ONDRI



ONTARIO BRAIN INSTITUTE
INSTITUT ONTARIEN DU CERVEAU

# Outline

- Overview of ONDRI
- Overview of curation cycle
  - Pre-processing
  - Data Standards
  - Outliers
- ONDRI and Brain-CODE in the future

# What we're doing now

- Curation, preprocessing, and release
  - REDCap
  - SPReD/XNAT
  - Brain-CODE workspaces
  - LabKey
- Still requires some local steps
  - Curation for some platforms
  - Standards
  - Outliers

# Looking ahead

- Better/simpler integration of tools
  - Especially for curation purposes
- Get all platforms into a BrainCODE pipeline
  - Emphasis on reproducibility
- Standards & Outliers
  - Automate
  - Push from LabKey to Brain-CODE Workspace
    - And then back!
- Archives & Revisions
  - LabKey's Postgres SQL

# And Beyond!

- Never download data again
  - And never be out of sync of latest data!

- Perform
  - Analytics
  - Reports
  - Manuscripts

# Acknowledgements


- Neuroinformatics & Biostatistics:
  - Stephen Strother, Malcolm Binns, Stephen Arnott, Abiramy Uthirakumaran
- Platform Curators:
  - Paula McLaughlin (NPSY), Donna Kwan (CLIN), Chris Scott (NIMG), Julia Fraser & Ben Cornish (GAIT), Elena Leontieva (SDOCT), Brian Coe (EYTK), Allison Dilliott (GNMC)
- InDoc
  - Especially Kristen, Mojib, Sara, Anthony
- ONDRI Leads & Executives:
  - Rick Swartz, Mario Masellis, Doug Munoz, Natalie Rashkovan, Peter Kleinstiver, Mike Strong