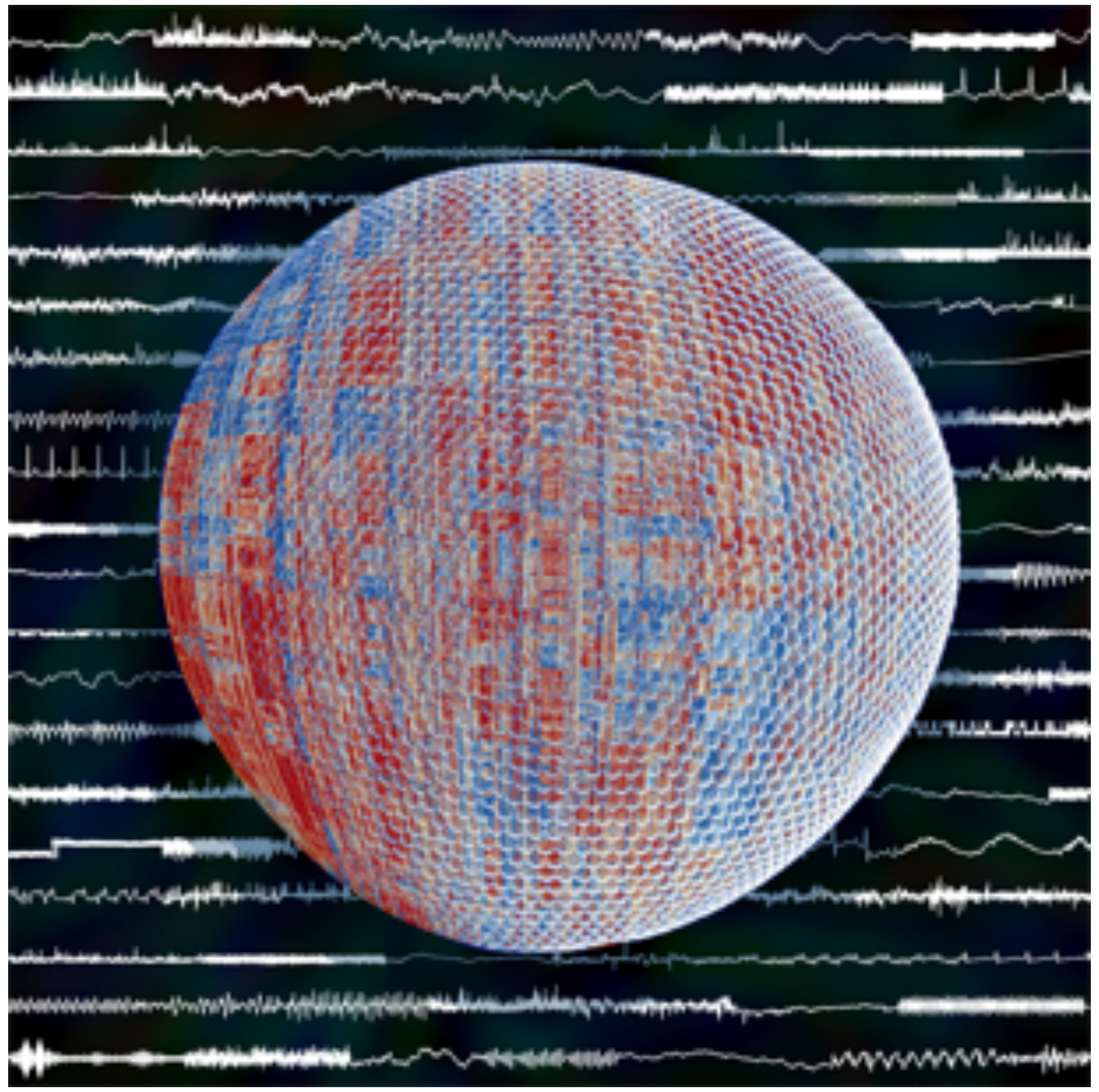
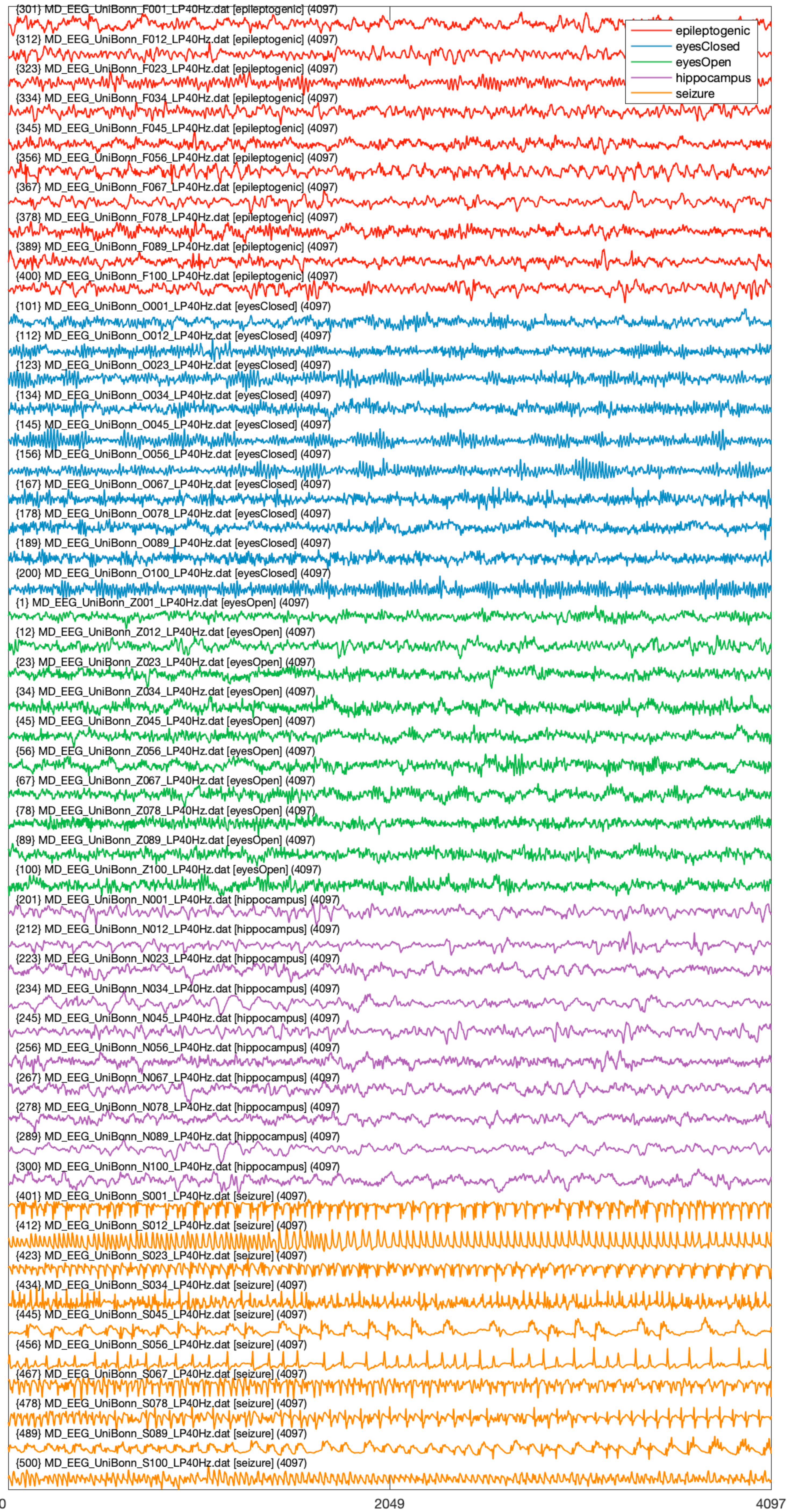
Visualizing and understanding complex neural time series



Advanced Statistical Methods and Dynamic Data Visualizations for Mental Health Studies, June 2021

Dr Ben Fulcher, Dynamics and Neural Systems Group, School of Physics, The University of Sydney.



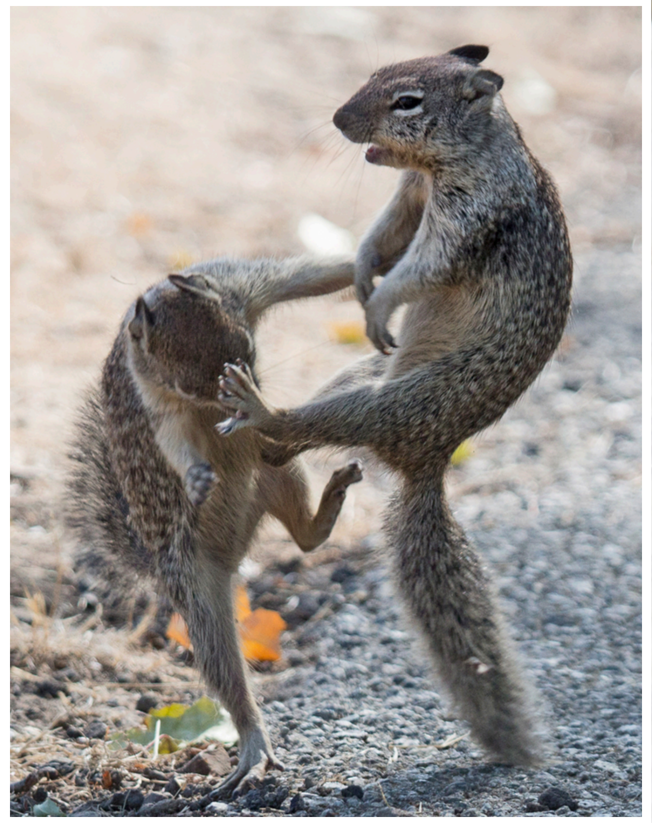
Setting

- Modern datasets often contain a large number of time series:
  - EEG, MEG, fMRI, calcium imaging, …
  - ECG, accelerometer, self-reports, …
- Common scientific questions:
  - Is there interesting structure in the dataset?
  - What time-series properties distinguish different classes of data (e.g., schizophrenia/control)?
  - How accurately can I classify conditions (‘biomarkers’)?

Interactive visualizations can help to understand complex time-series datasets.

Today

- **Feature-based time-series analysis** (representing time-series using properties) is **powerful**
  - We have developed a range of tools for doing this systematically.
- I'll give an overview of the tools available, focusing on **two key analyses**:
  - Finding structure in a dataset through a **low-dimensional projection**.
  - **Classifying** a dataset (and understanding why).
- I'll give a quick **demo** of an **interactive** low-dimensional projection of a time-series dataset (we'll try it together in the interactive session).



Time-series analysis is a **very** interdisciplinary field

We should learn from each other…

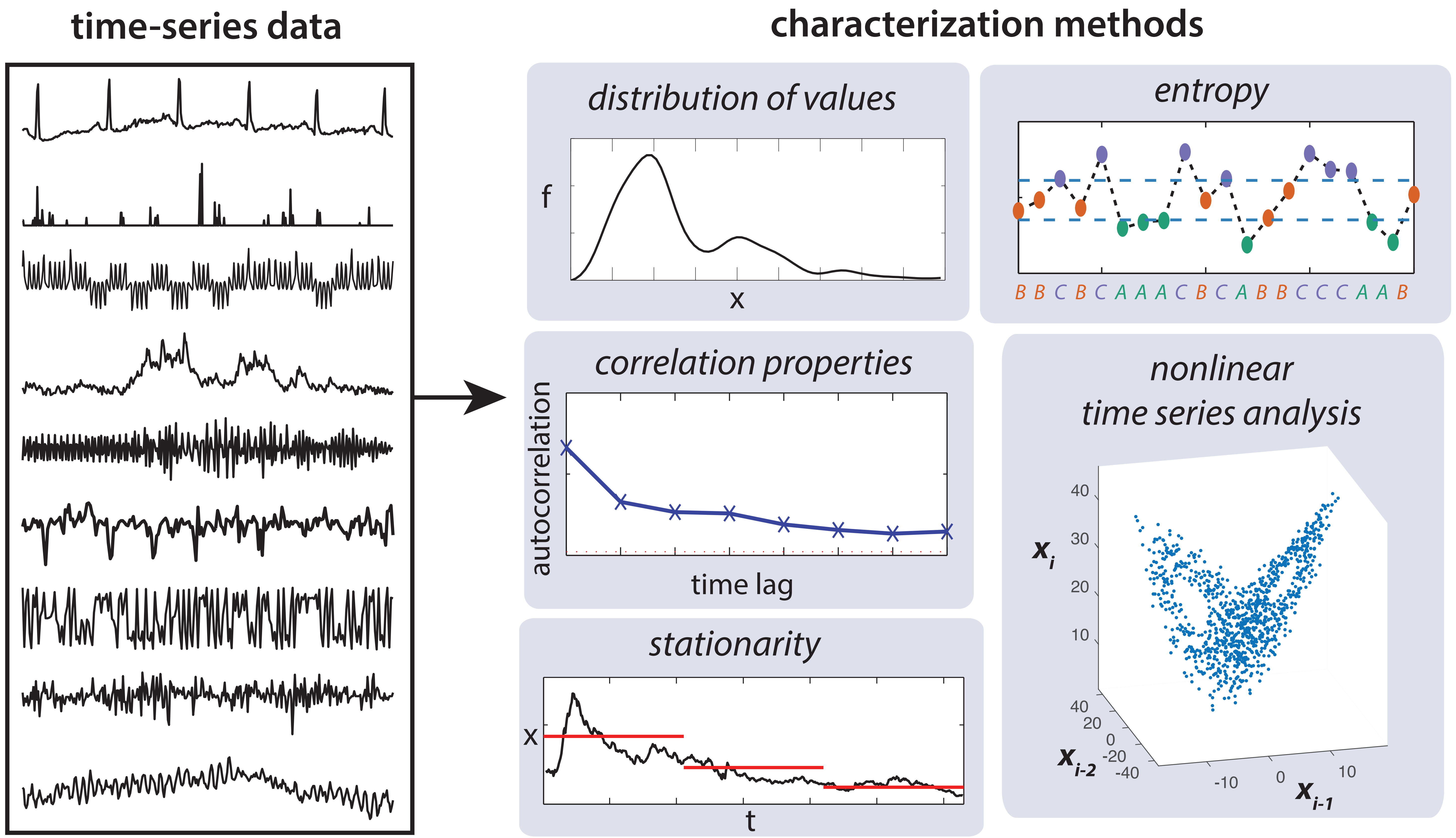
How can we reduce the barriers to meaningful interdisciplinary exchange?
(rapid knowledge transfer across fields)

Many of our measurements of the world are in the form of **time series**

Repeated measurements of some system over time: $(x_1, x_2, x_3, \ldots)$

Characterizing time series using *features*

How can I reduce complex time-varying patterns to informative summary statistics?



- We consider *features*, which map a time series onto a single real number
- These numbers are often interpretable:
  - ‘periodic’, ‘unpredictable’, ‘nonlinear’, ‘stationary’, ‘intermittent’, ‘bursty’, …
- Feature-based time series analysis involves representing time series as a set of features.

What feature(s) should I use?

Methods for time-series analysis have been developed across diverse scientific literature for decades
The *hctsa* feature set contains a sample of >7000 features

Static Distribution

- Quantiles
- Trimmed means
- Fits to standard distributions
- Outliers
- Moments
- Entropy
- Rank-orderings
- Standard deviation

Stationarity

- StatAv
- Sliding window measures
- Step detection
- Distribution comparisons

Correlation

- Linear autocorrelation
- Decay properties
- Additive noise titration
- Nonlinear autocorrelations
- Time reversal asymmetry
- Generalized self-correlation
- Recurrence structure
- Autocorrelation robustness
- Scaling and fluctuation analysis
- Permutation robustness
- Local extrema
- Seasonality tests
- Zero crossing rates

Basis Functions

- Wavelet transform
- Peaks of power spectrum
- Spectral measures
- Power in frequency bands

Information Theory

- Sample Entropy
- Lempel-Ziv Complexity
- Automutual information
- Information dynamics
- Approximate Entropy
- Tsallis entropies

(Phys) Nonlinear

- 2D embedding structure
- Taken's estimator
- Fractal dimension
- Correlation dimension
- Poincaré sections
- Surrogate data
- Nonlinear prediction error
- Lyapunov exponent estimate
- False nearest neighbors

Model Fitting

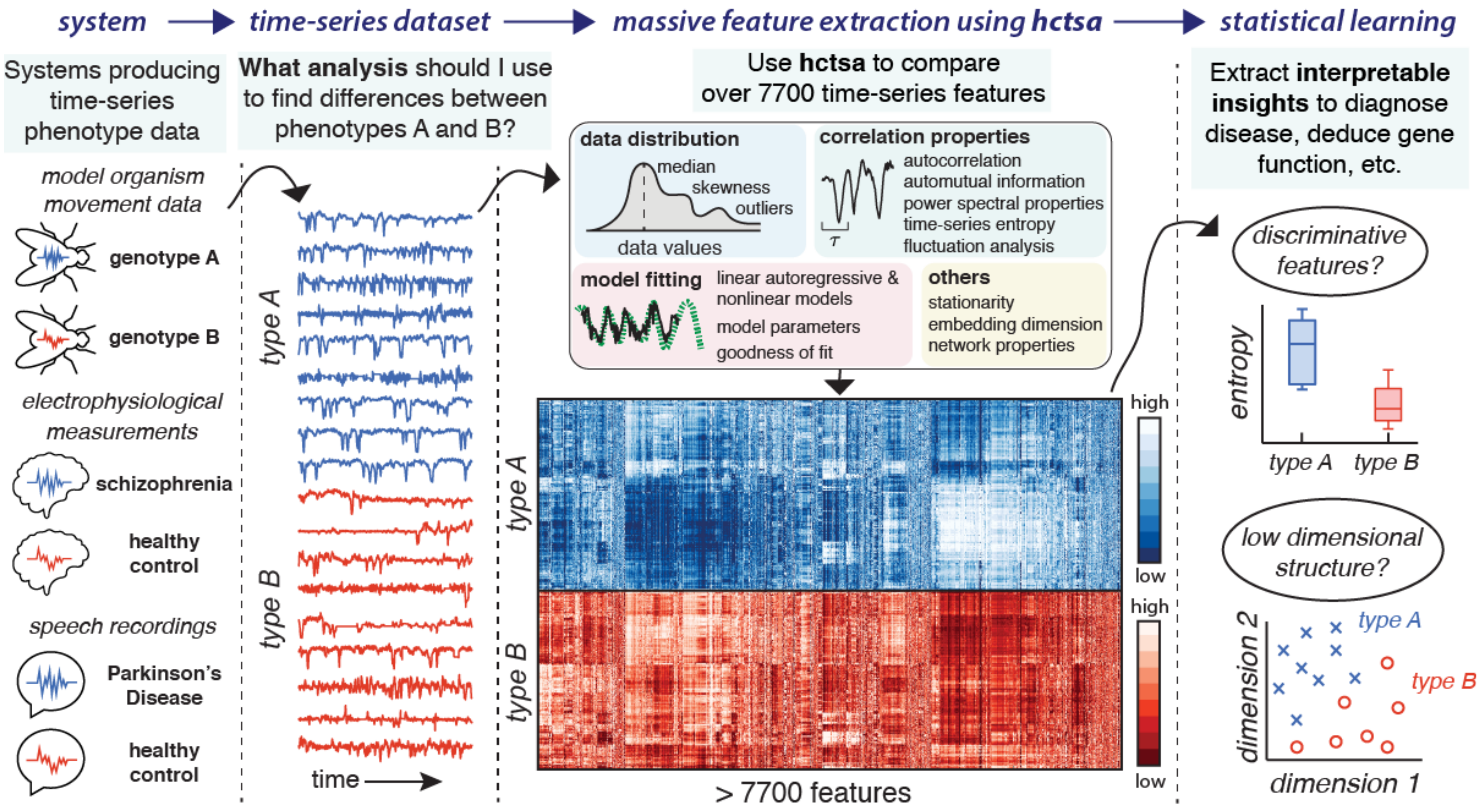
- Local prediction
- GARCH models
- Fourier fits
- AR models
- Exponential smoothing
- State space models
- Hidden Markov models
- Piecewise splines
- Biased walker simulations
- ARMA models
- Gaussian Processes

Others

- Transition matrices
- Local motifs
- Dynamical system coupling
- Visibility graph
- Stick angle distribution
- Extreme events
- Singular spectrum analysis
- Domain-specific techniques

The highly comparative approach

Compare the performance of a comprehensive library of scientific time-series methods: pick those that best suit your problem



Fulcher & Jones (2017). *hctsa*: A Computational Framework for Automated Time-Series Phenotyping Using Massive Feature Extraction. *Cell Systems*, **5**, 527–531.

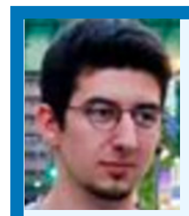
Our Feature Sets and Tools

hctsa

Compute >**7700** time-series features

Low-dimensional projections
Classification, …

https://github.com/benfulcher/hctsa



catch22

Compute **22** time-series features

Fast-coded in C

Rcatch22 and **catch22.jl**

https://github.com/chlubba/catch22

CompEngine

Drag-and-drop time-series data

Explore connections across diverse scientific data

https://comp-engine.org/

theft

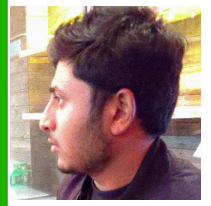
Feature computation, analysis, and visualization for feature-based time-series analysis

https://github.com/hendersontrent/theft

theft web portal

Drag-and-drop online access to theft functionality

https://dynamicsandneuralsystems.shinyapps.io/timeseriesfeaturevis/



CompEngine Features

Drag-and-drop time-series features

Explore connections across diverse features

https://www.comp-engine-features.org/

Multivariate Time-Series Features

Extension to multivariate datasets: measures of complex interactions

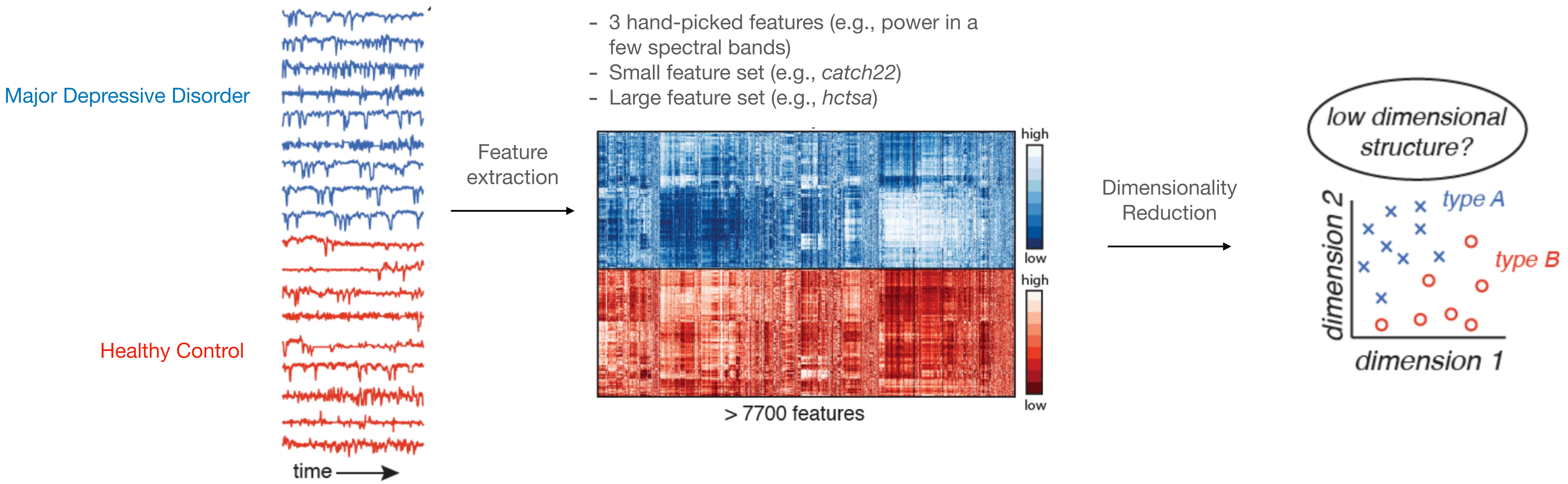
Others have done work in this space also:

https://github.com/benfulcher/hctsa/wiki/Related-time-series-resources

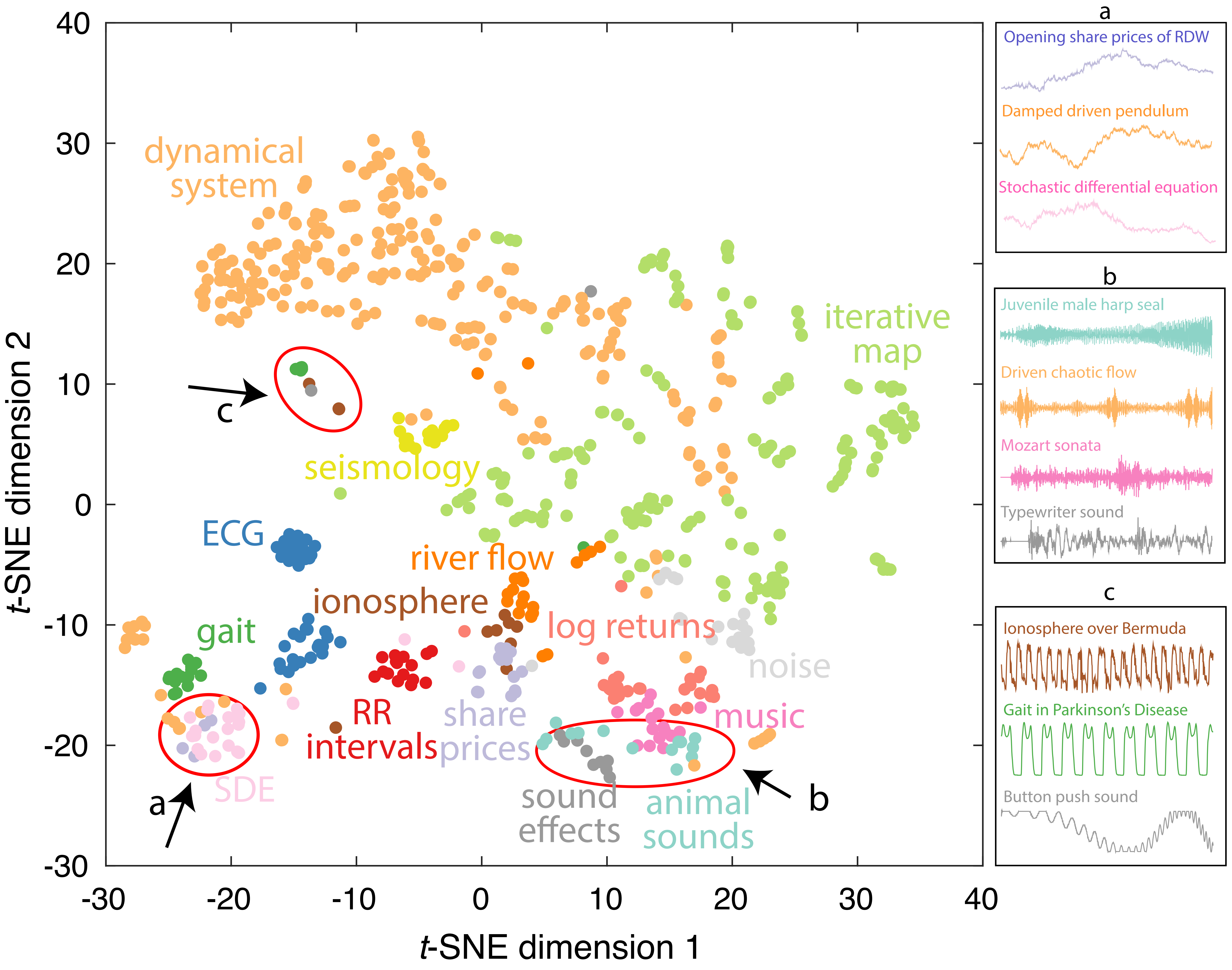
Low-Dimensional Feature-Space Projections

How are my time-series data structured?

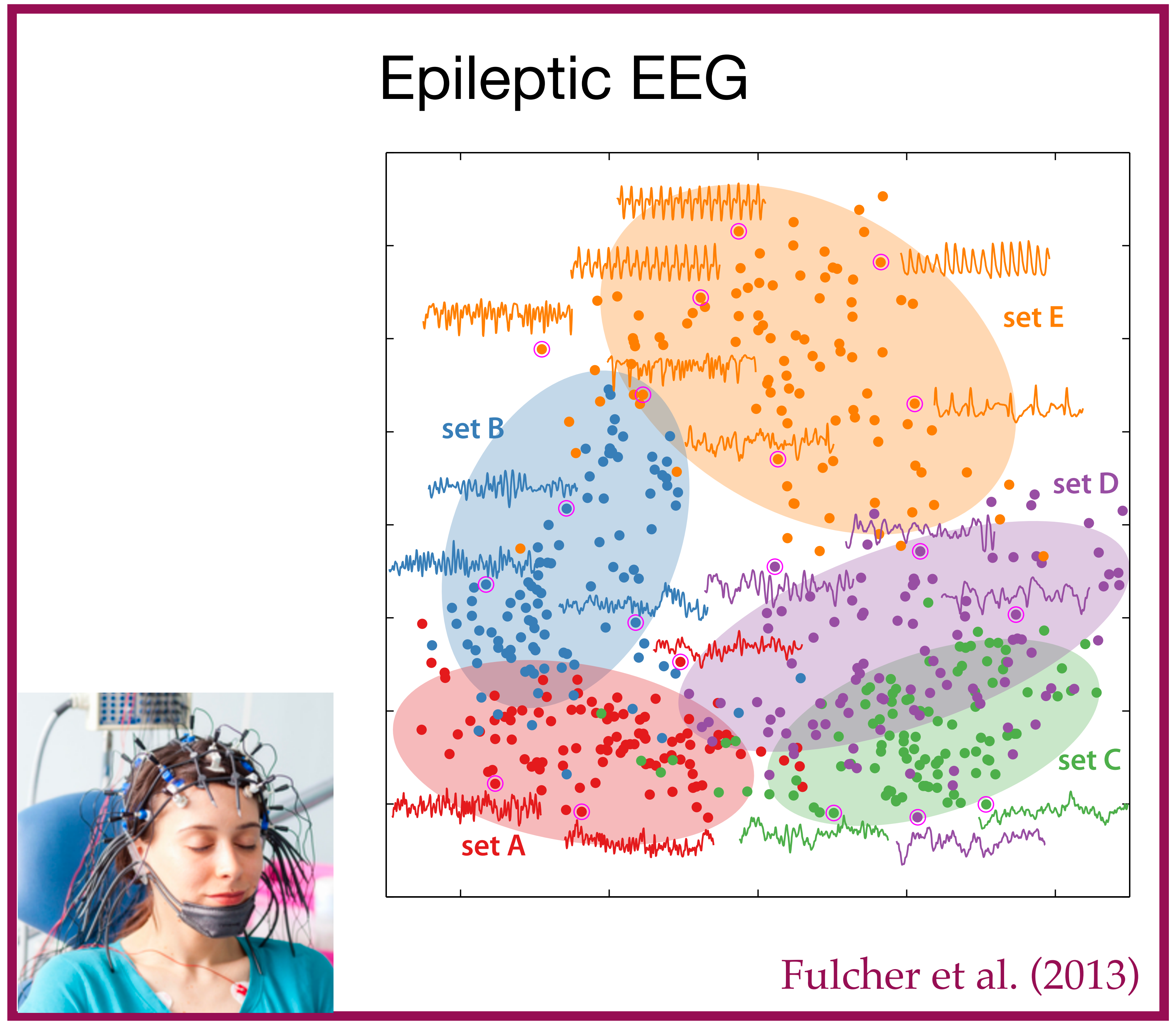
Represent each time series as a set of features (interpretable structural properties), and look for patterns in the low-dimensional feature space: ***time series with similar properties are close in the space.***

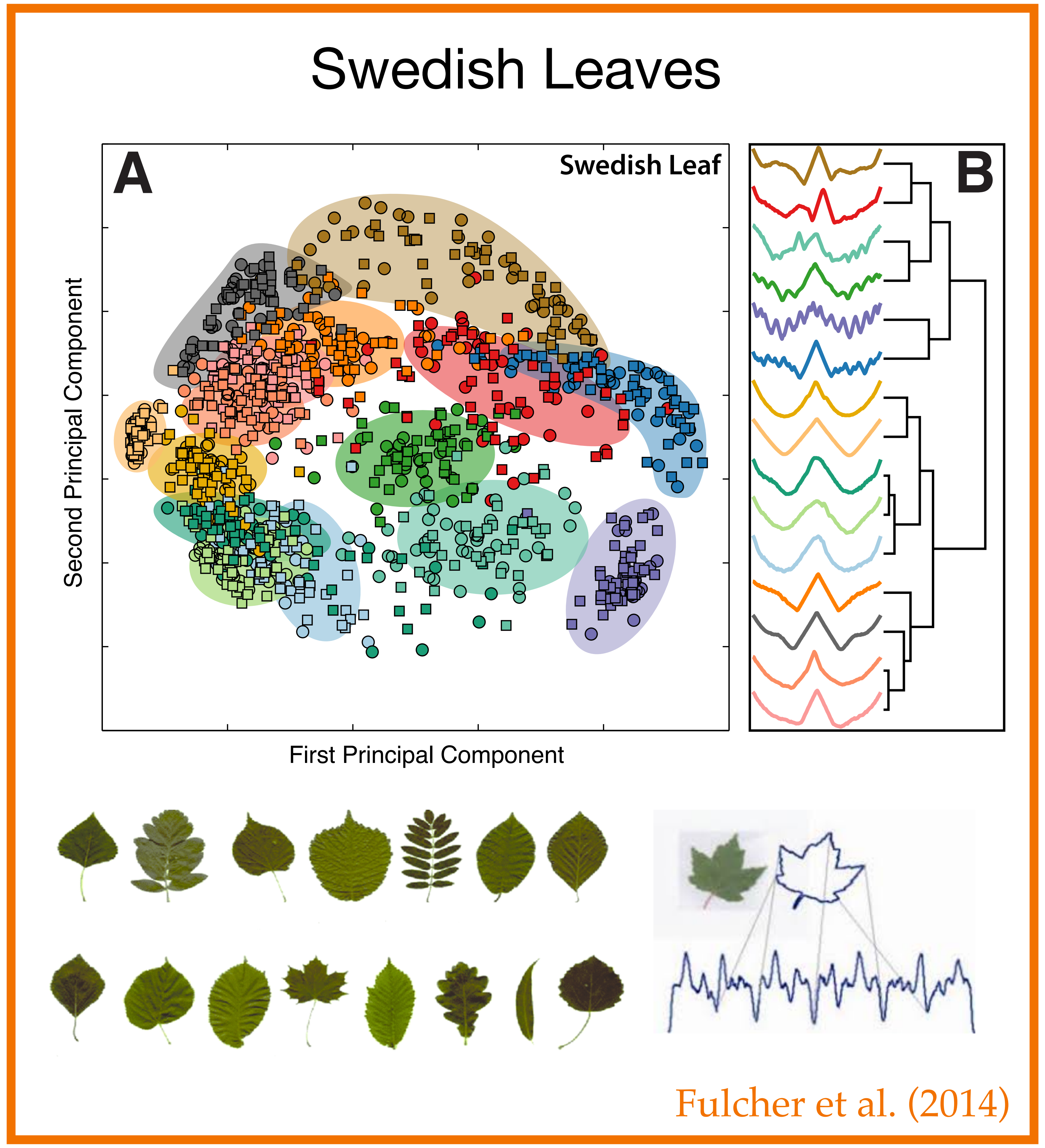


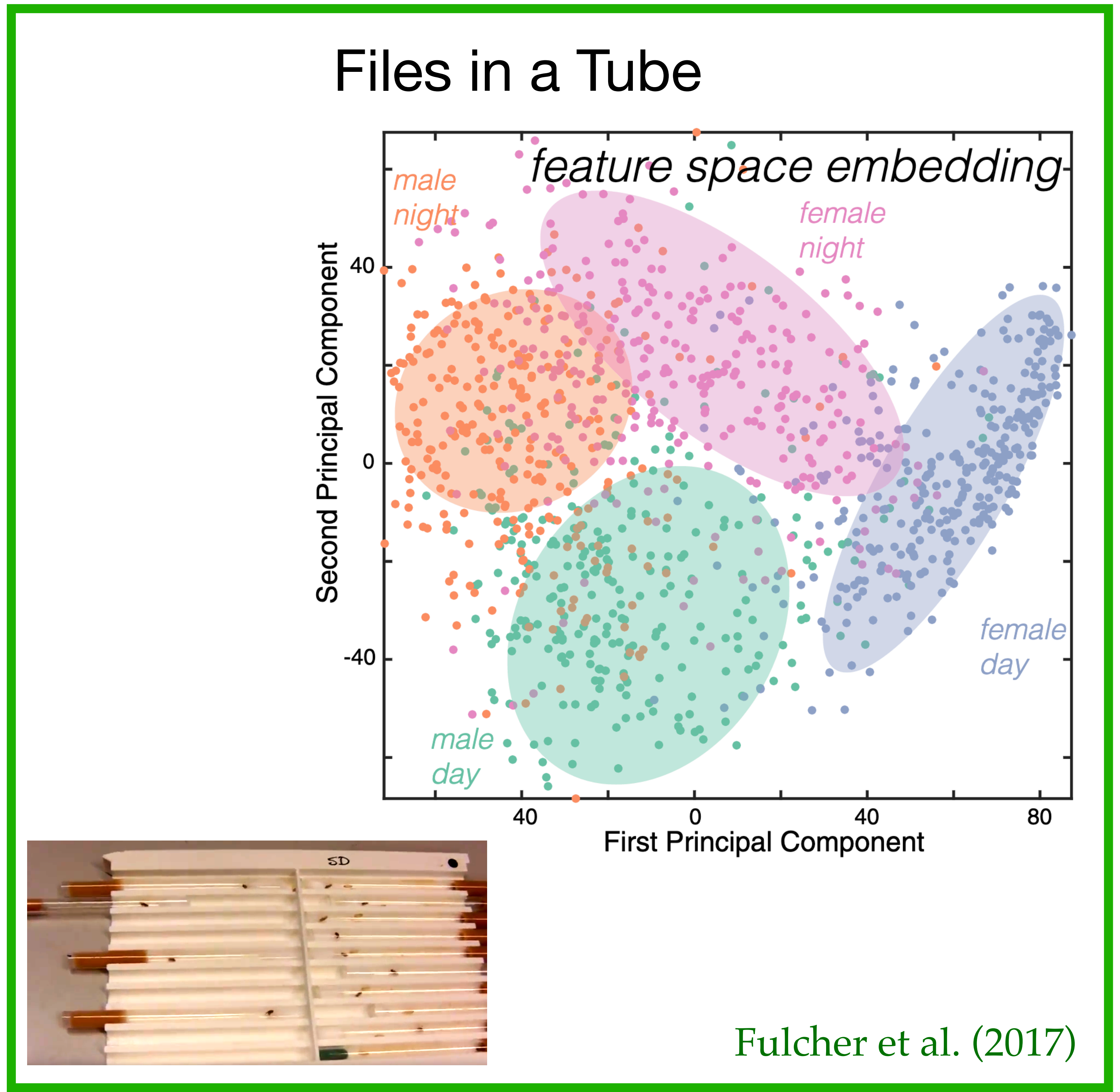
Low-dimensional feature-space projections

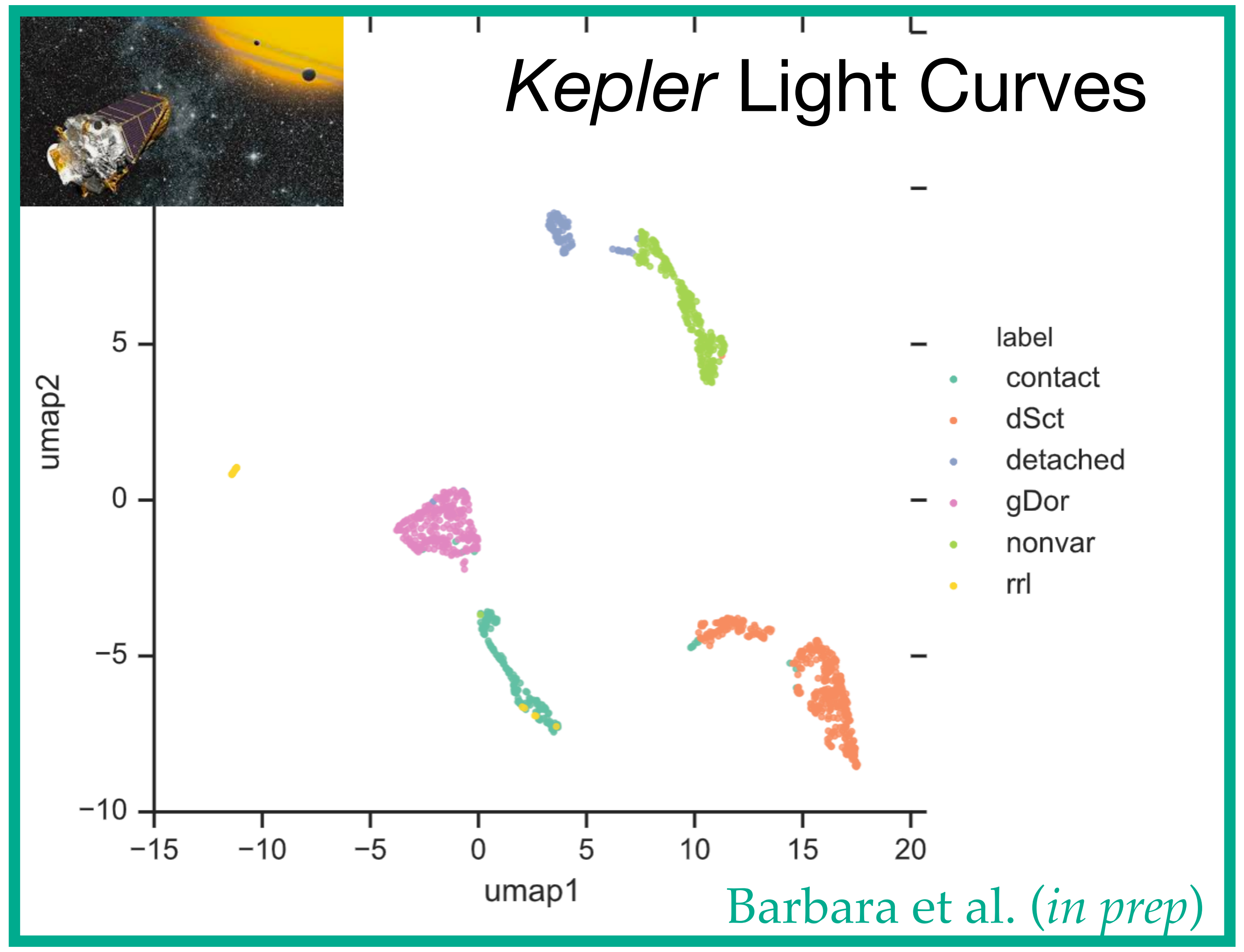


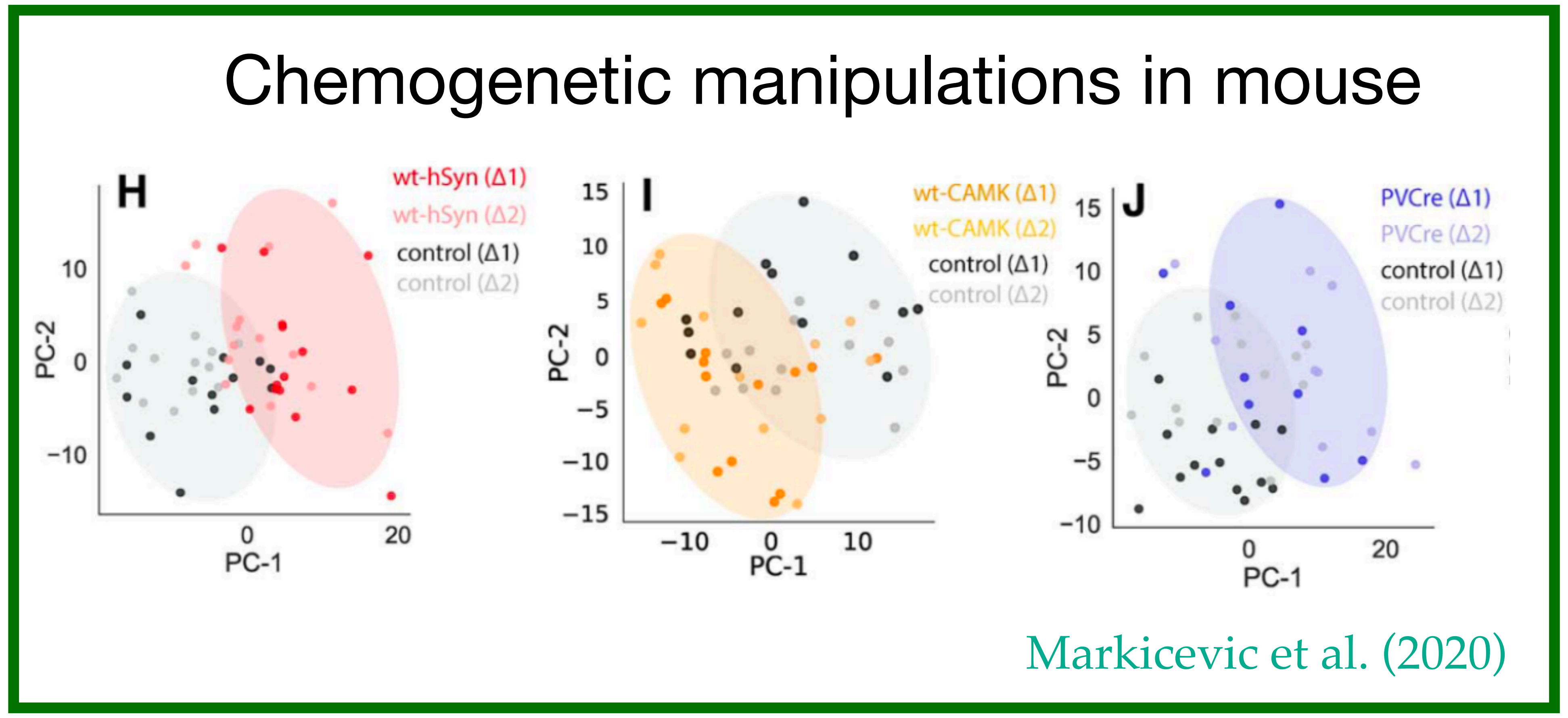
Fulcher et al. A self-organizing, living library of time-series data. *Scientific Data* **7**, 213 (2020).

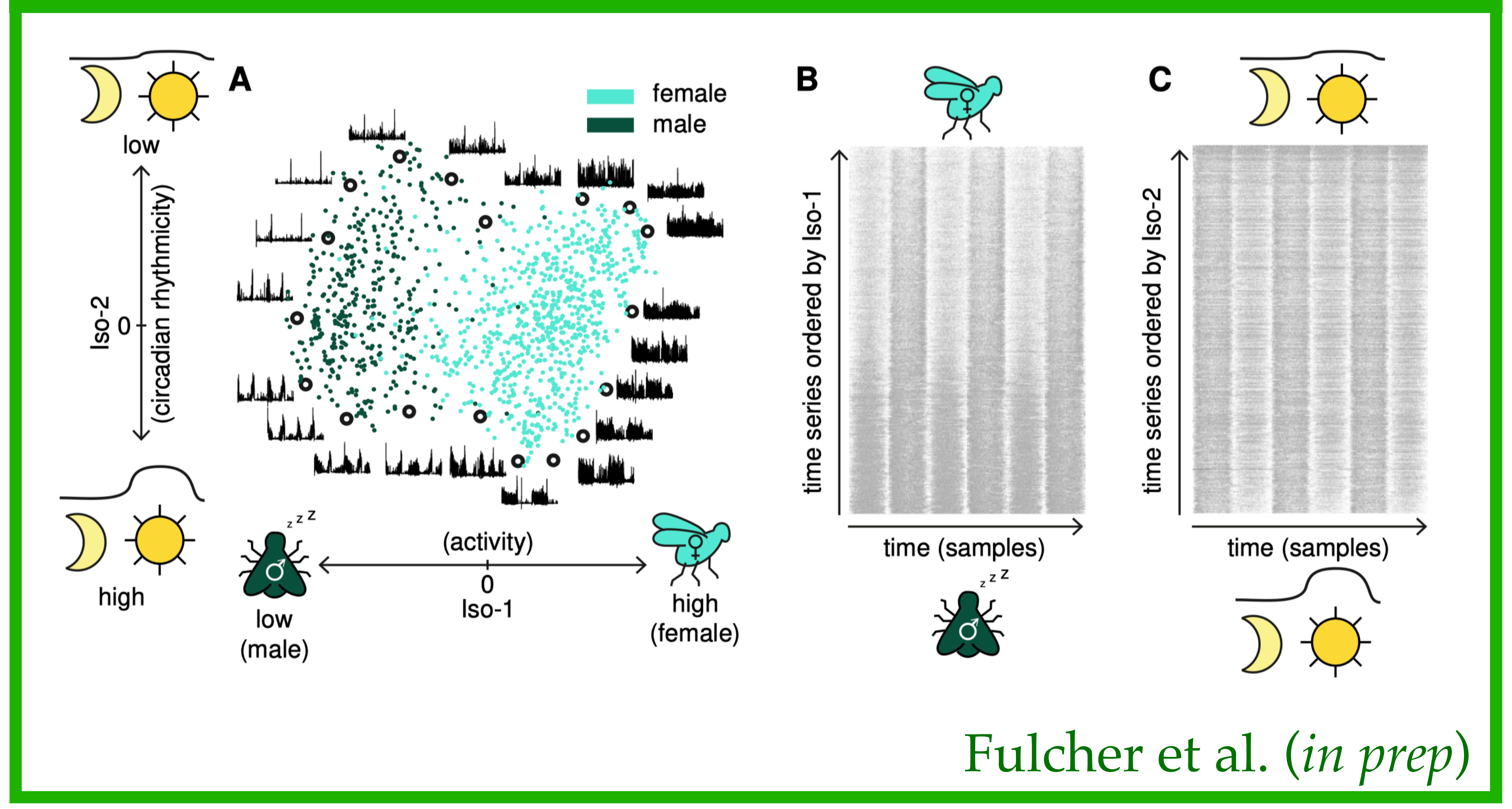

Epileptic EEG
set E
set B
set D
set C
set A
Fulcher et al. (2013)


Swedish Leaves
A
Swedish Leaf
B
Second Principal Component
First Principal Component
Fulcher et al. (2014)


Files in a Tube
feature space embedding
male night
female night
female day
male day
Second Principal Component
First Principal Component
40
0
-40
80
Fulcher et al. (2017)


Kepler Light Curves
label
contact
dSct
detached
gDor
nonvar
rrl
umap2
umap1
Barbara et al. (in prep)


Chemogenetic manipulations in mouse
H
wt-hSyn (Δ1)
wt-hSyn (Δ2)
control (Δ1)
control (Δ2)
I
wt-CAMK (Δ1)
wt-CAMK (Δ2)
control (Δ1)
control (Δ2)
J
PVCre (Δ1)
PVCre (Δ2)
control (Δ1)
control (Δ2)
PC-2
PC-1
Markicevic et al. (2020)


A
female
male
low
high
Iso-2 (circadian rhythmicity)
(activity)
Iso-1
low (male)
high (female)
B
time series ordered by Iso-1
time (samples)
C
time series ordered by Iso-2
time (samples)
Fulcher et al. (in prep)

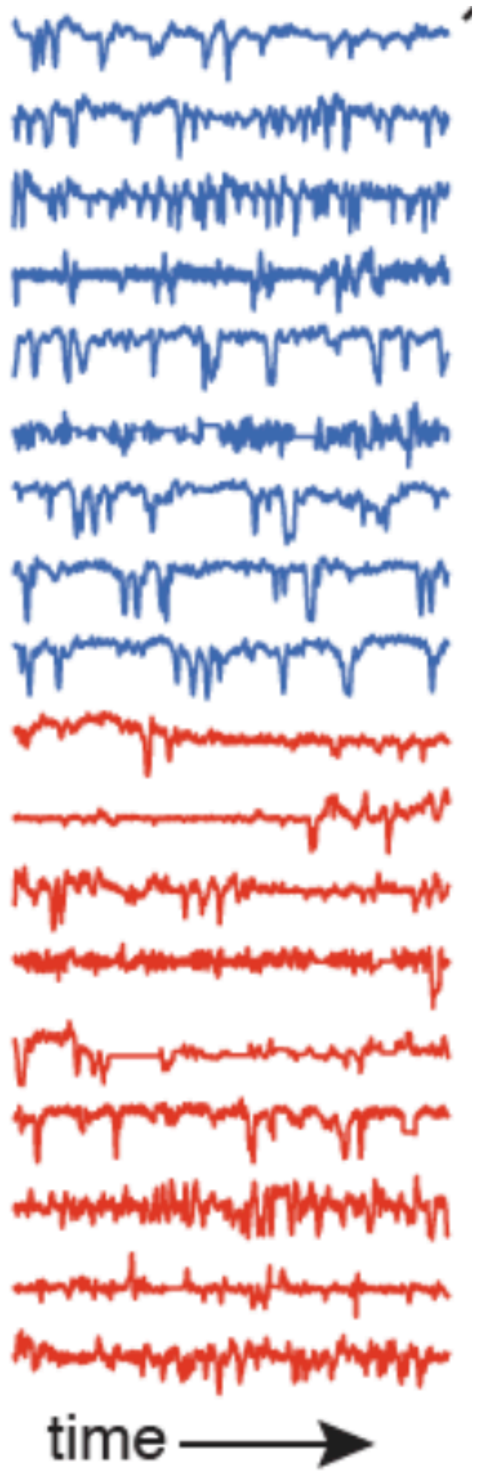
Classification

What types of features distinguish classes in my dataset?

(straightforward extension to real-valued labels: regression)

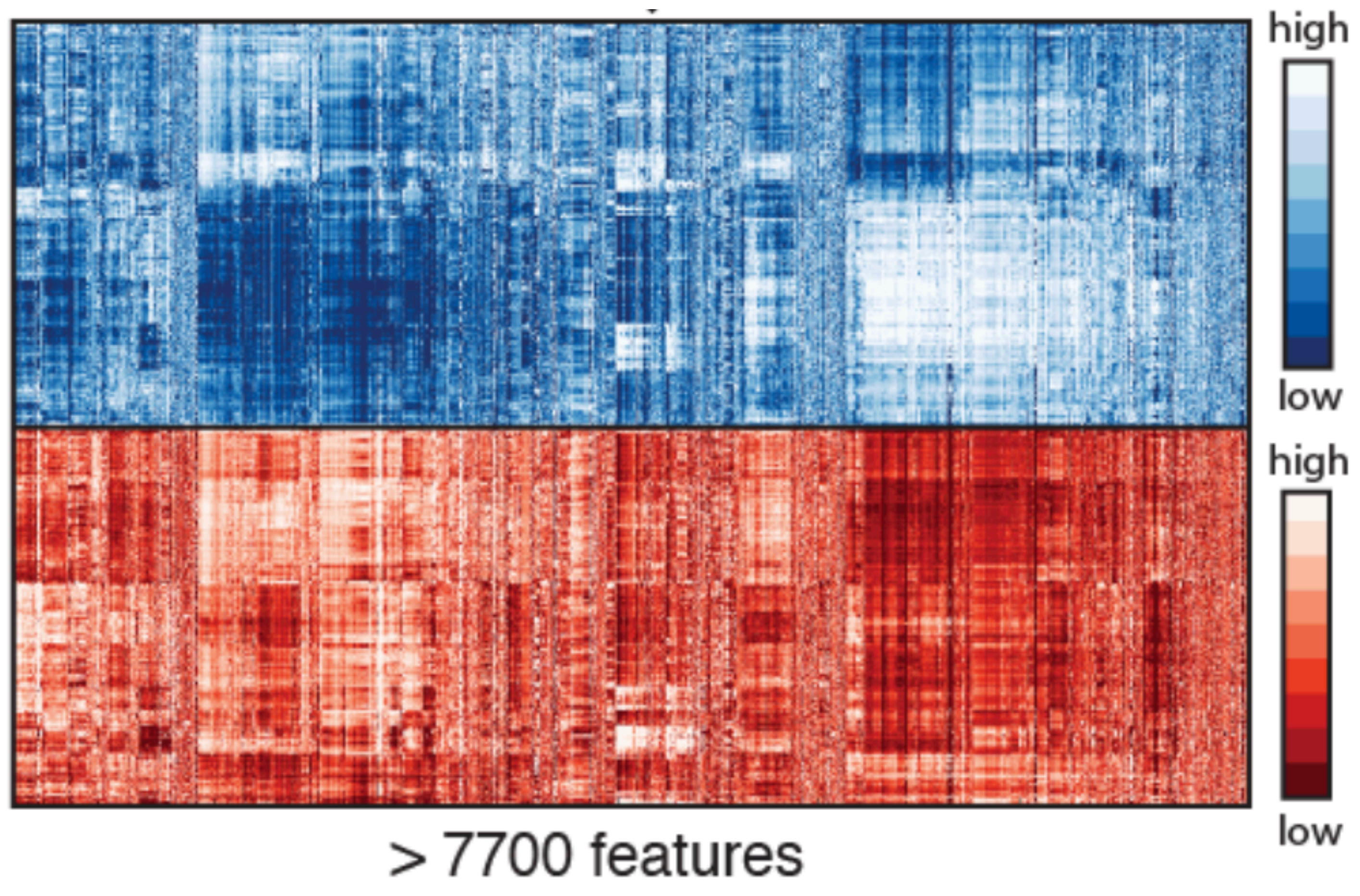
Major Depressive Disorder

Healthy Control



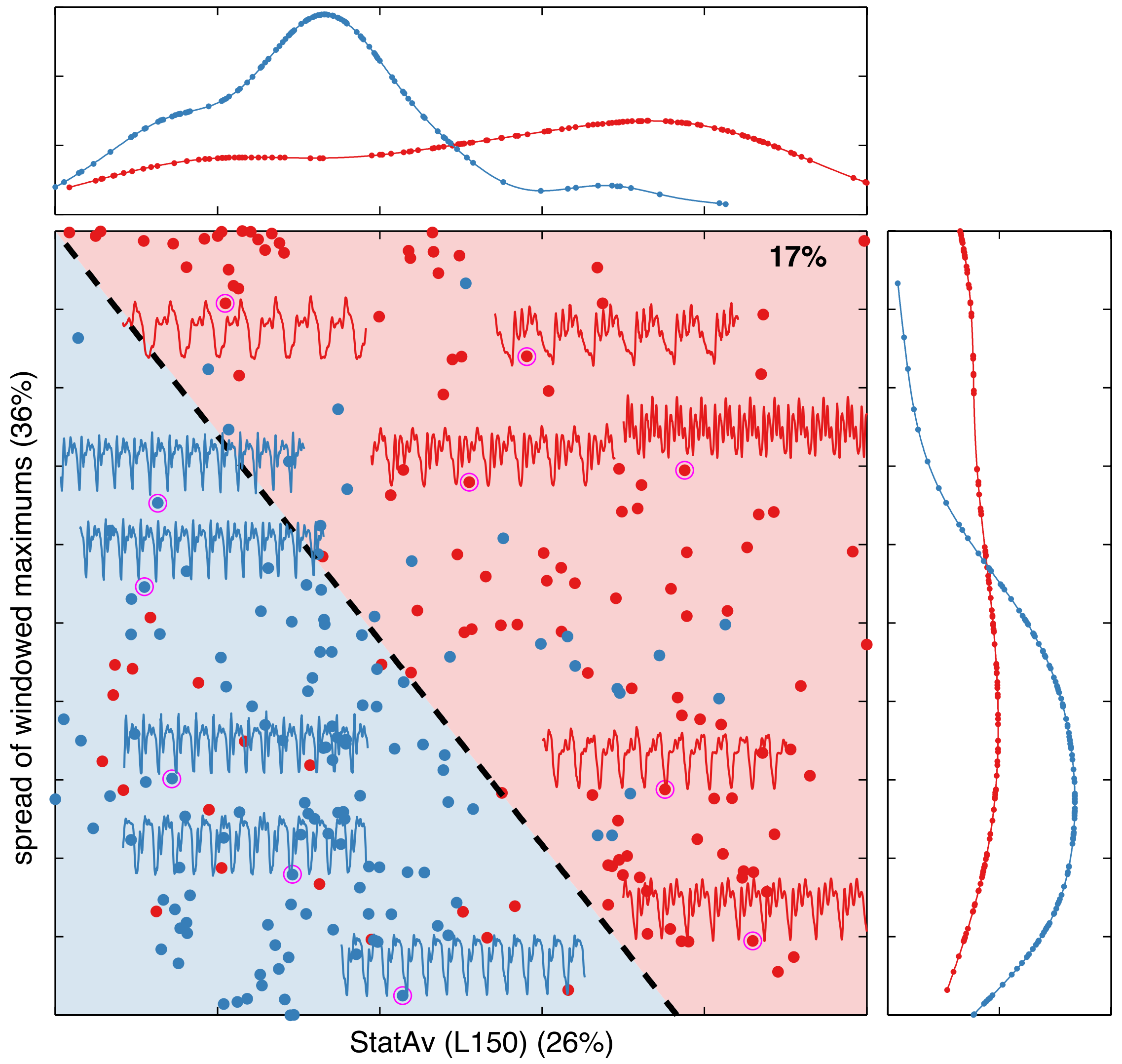
Feature extraction

- 3 hand-picked features (e.g., power in a few spectral bands)
- Small feature set (e.g., *catch22*)
- Large feature set (e.g., *hctsa*)



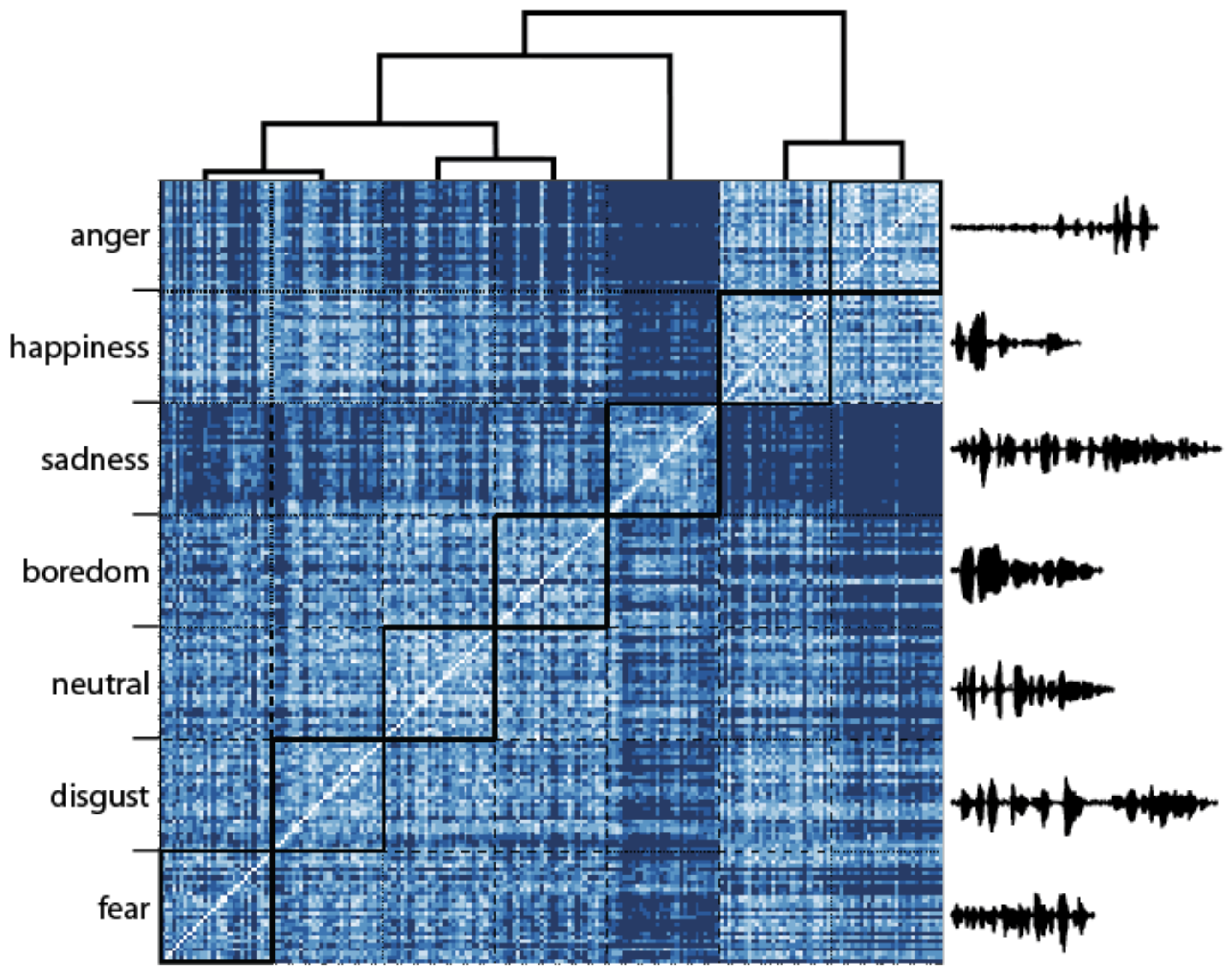
Train classifier

Classifying Parkinsonian Speech



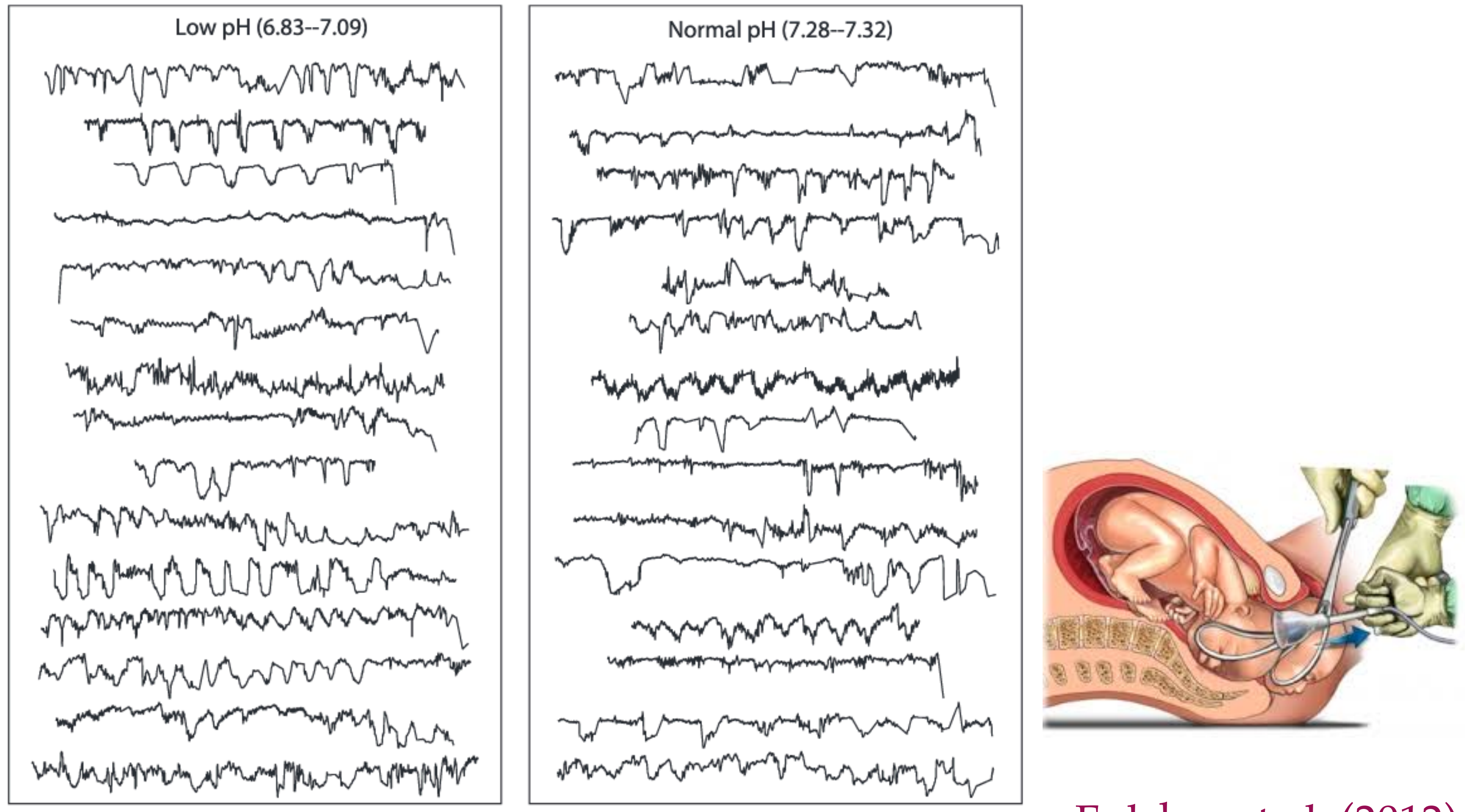
Fulcher et al. (2013)

Classifying Emotions from Speech



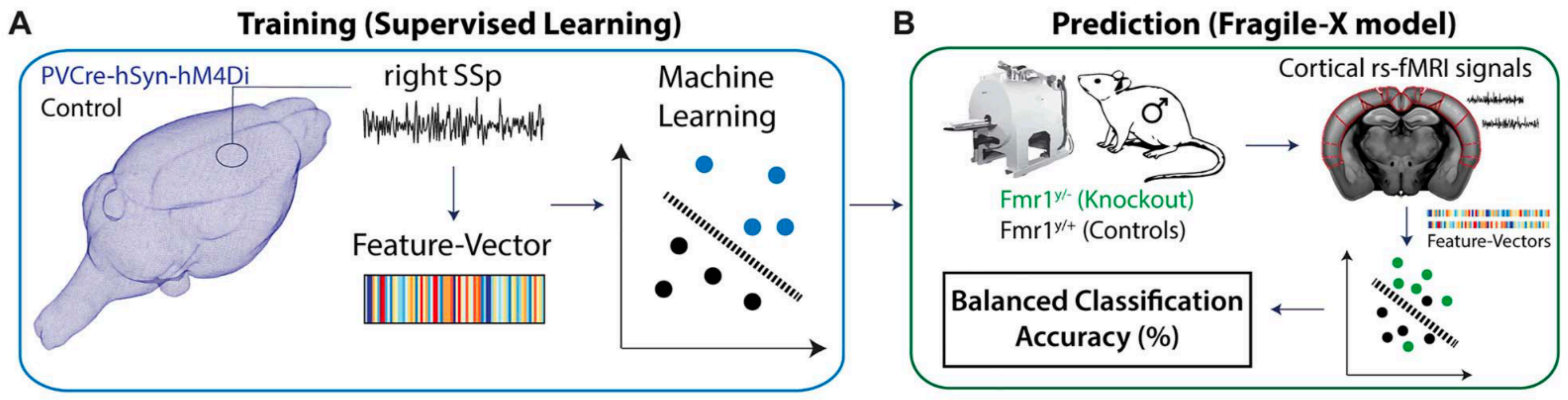
Fulcher et al. (2013)

Data-Driven Labour Interventions



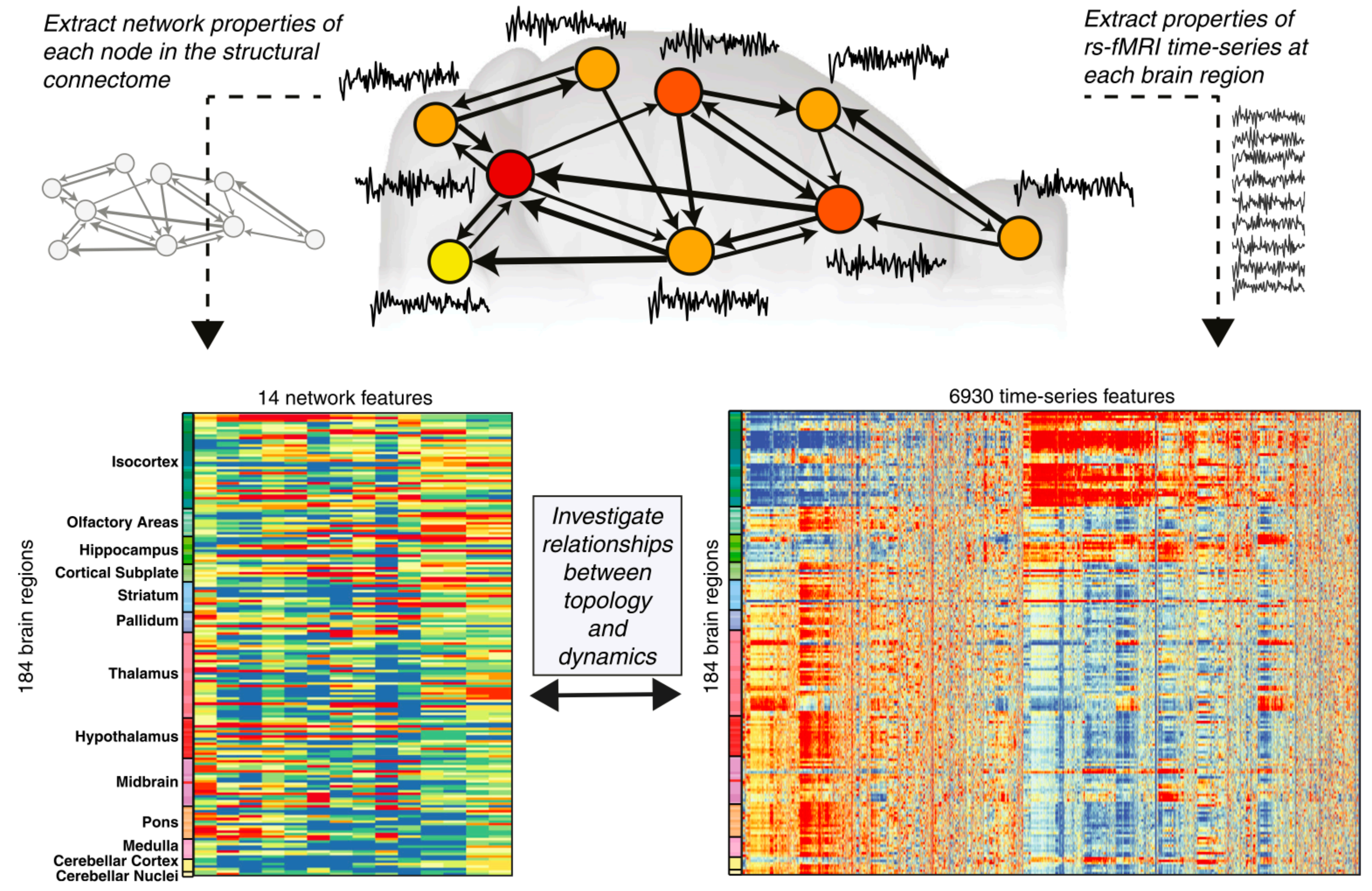
Fulcher et al. (2012)

Chemogenetic manipulations for mouse fMRI



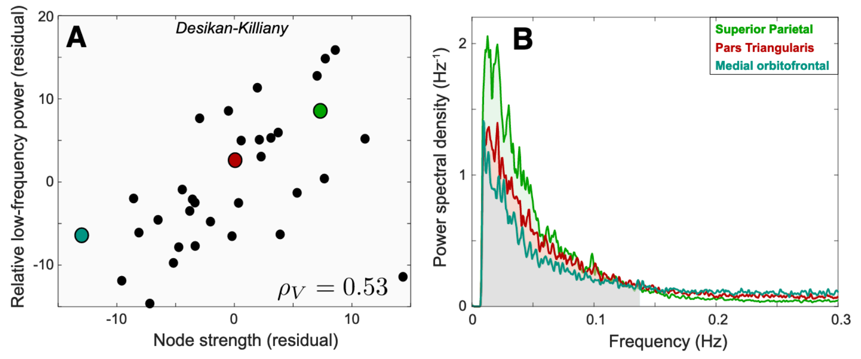
Markicevic et al. (2020).

Structure-function coupling in mouse



Sethi et al. (2017)

…and human



Fallon et al. (2020)

https://github.com/benfulcher/hctsa/wiki/Publications-using-hctsa

Distinguishing types of energy use in buildings

Liu et al. (2019). A hybrid model for appliance classification based on time series features. *Energy and Buildings*, **196**, 112-123.

Miller, C. (2019). What's in the box?! Towards explainable machine learning applied to non-residential building smart meter classification. *Energy and Buildings*, **199**, 523-536.

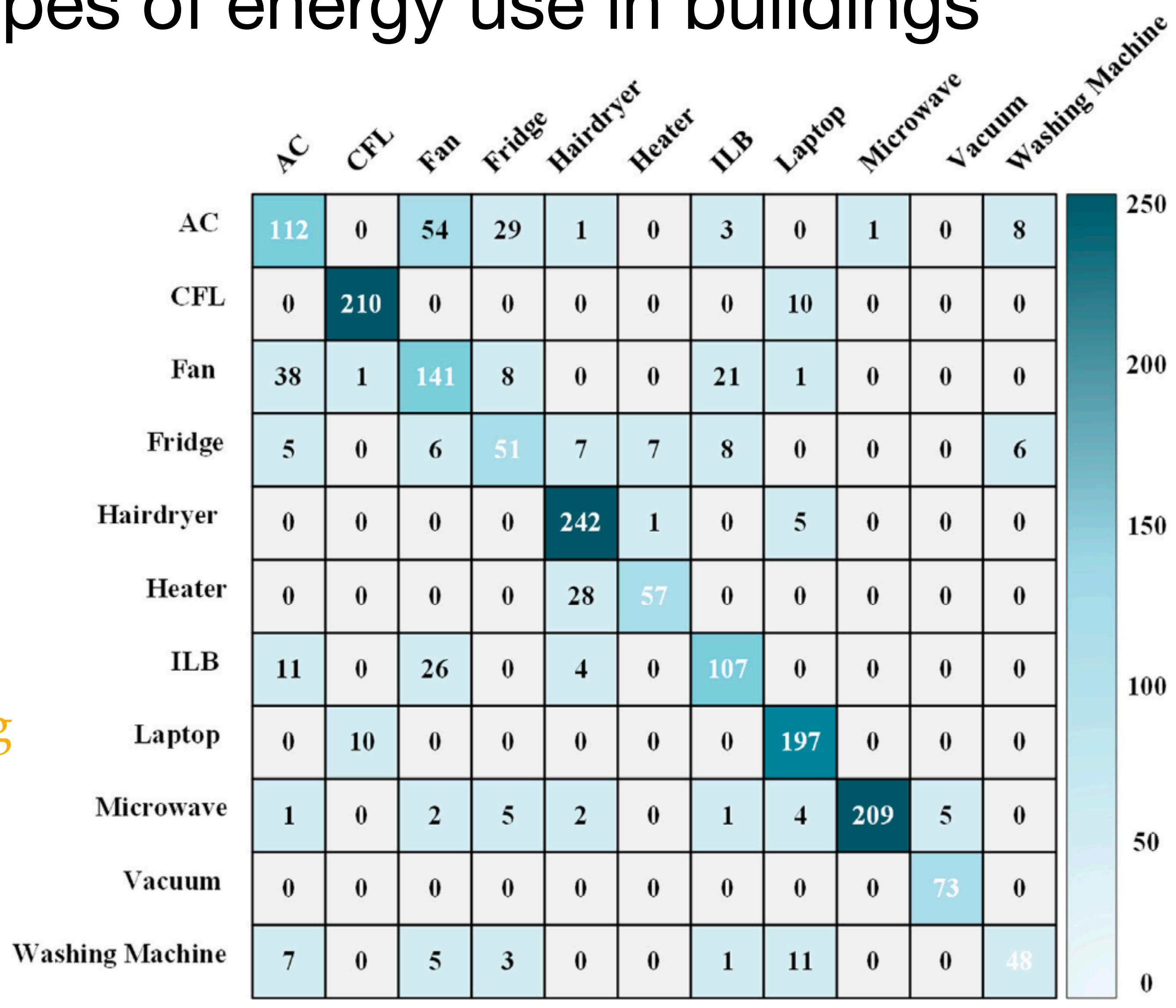
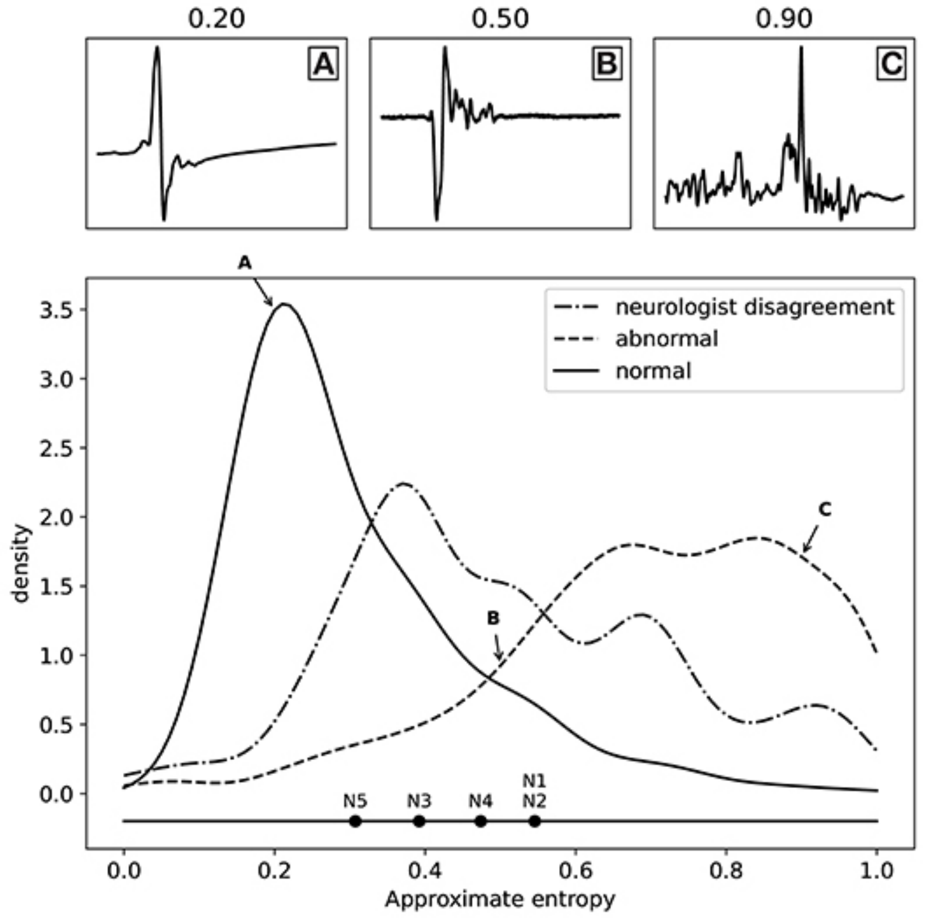


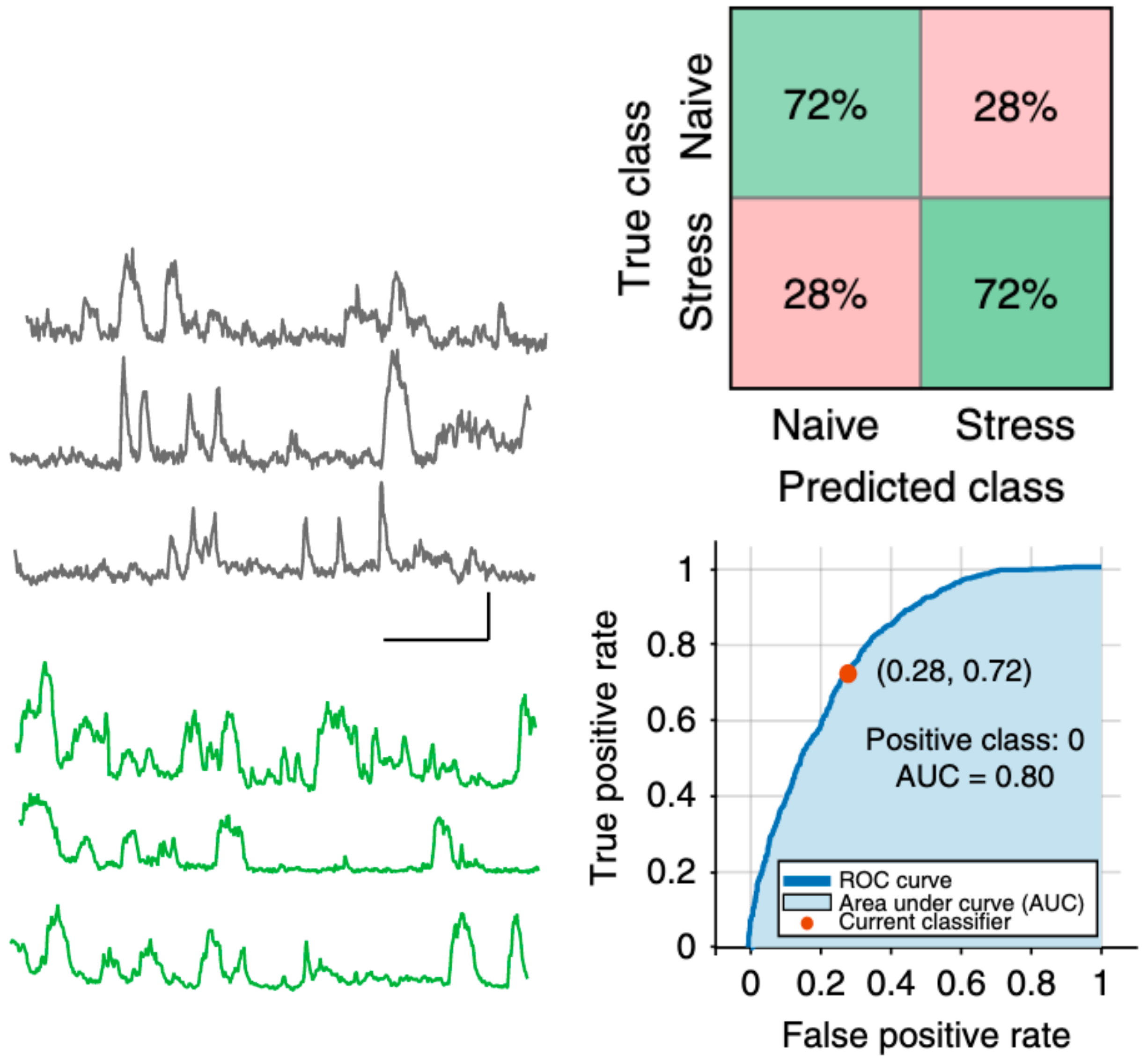
Fig. 7. Confusion matrix of the proposed model.

Distinguish multiple sclerosis MEPs



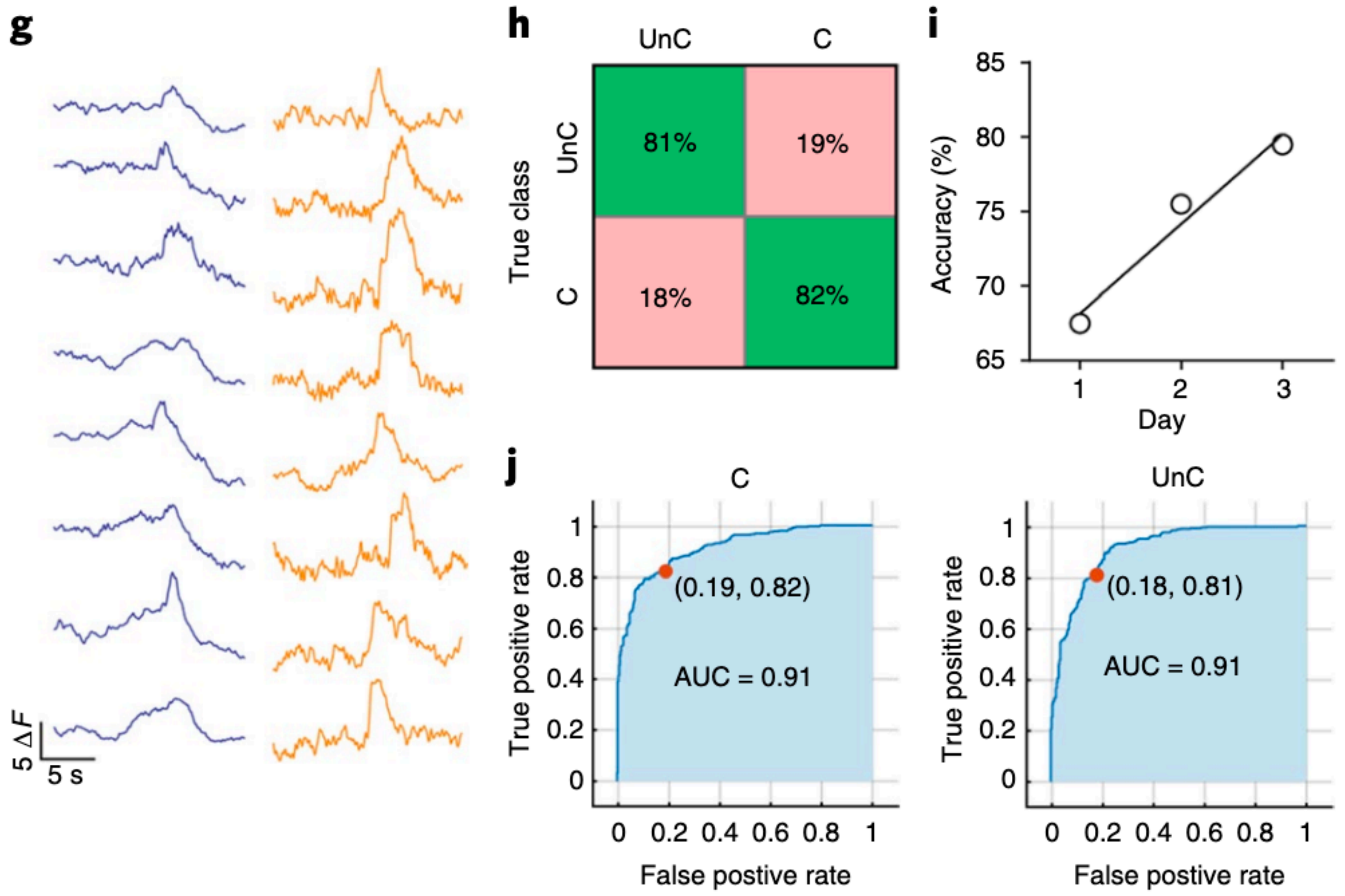
Yperman et al. Deciphering the Morphology of Motor Evoked Potentials. *Front. Neuroinform.* **14**:28 (2020).

Assess stress-induced changes in astrocyte calcium dynamics



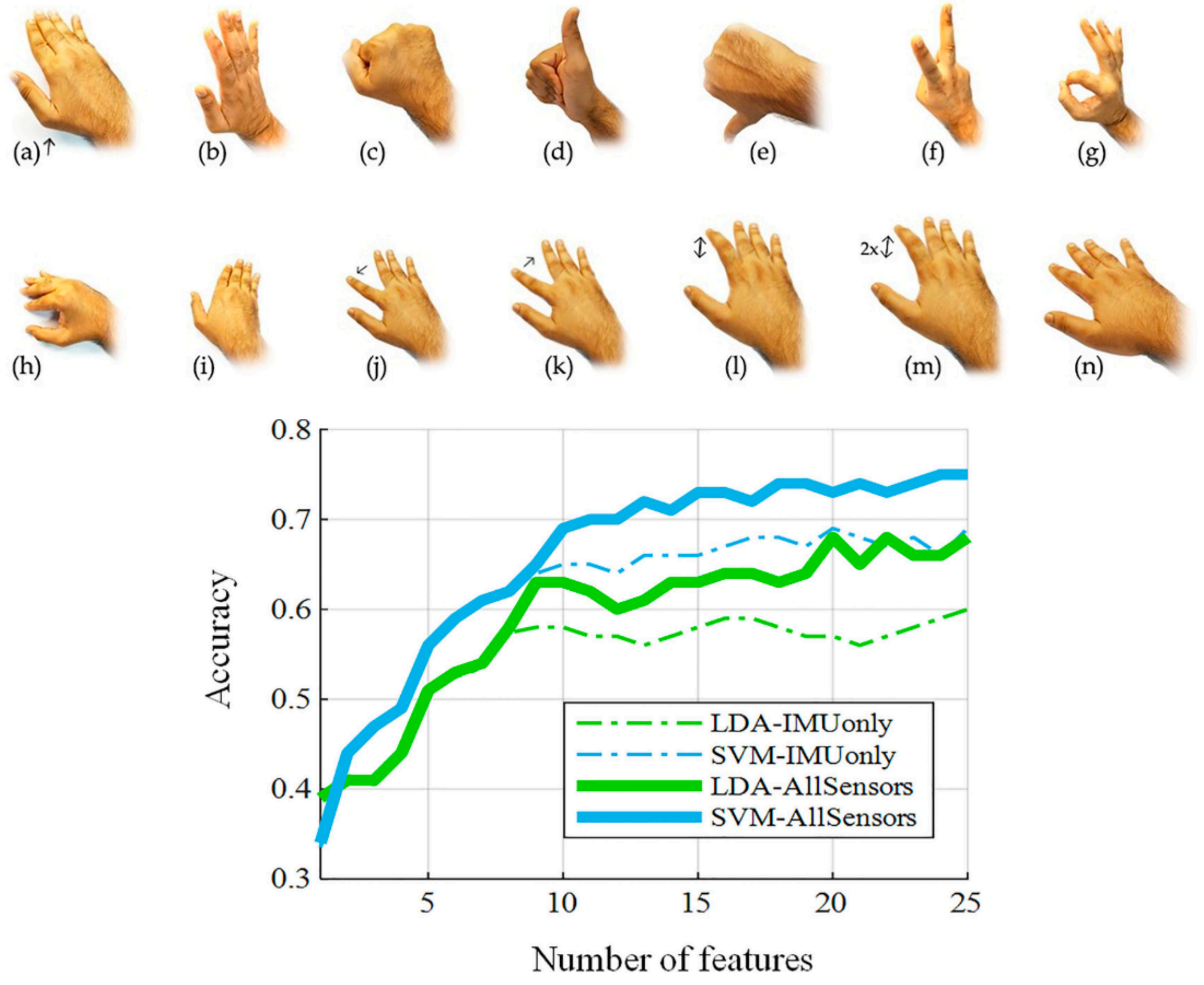
Murphy-Royal et al. Stress gates an astrocytic energy reservoir to impair synaptic plasticity. *Nat Commun* **11,** 2014 (2020).

Assess the stress controllability of neurons



Daviu et al. CRH neurons encode stress controllability and regulate defensive behavior selection. *Nat Neurosci* **23**, 398–410 (2020).

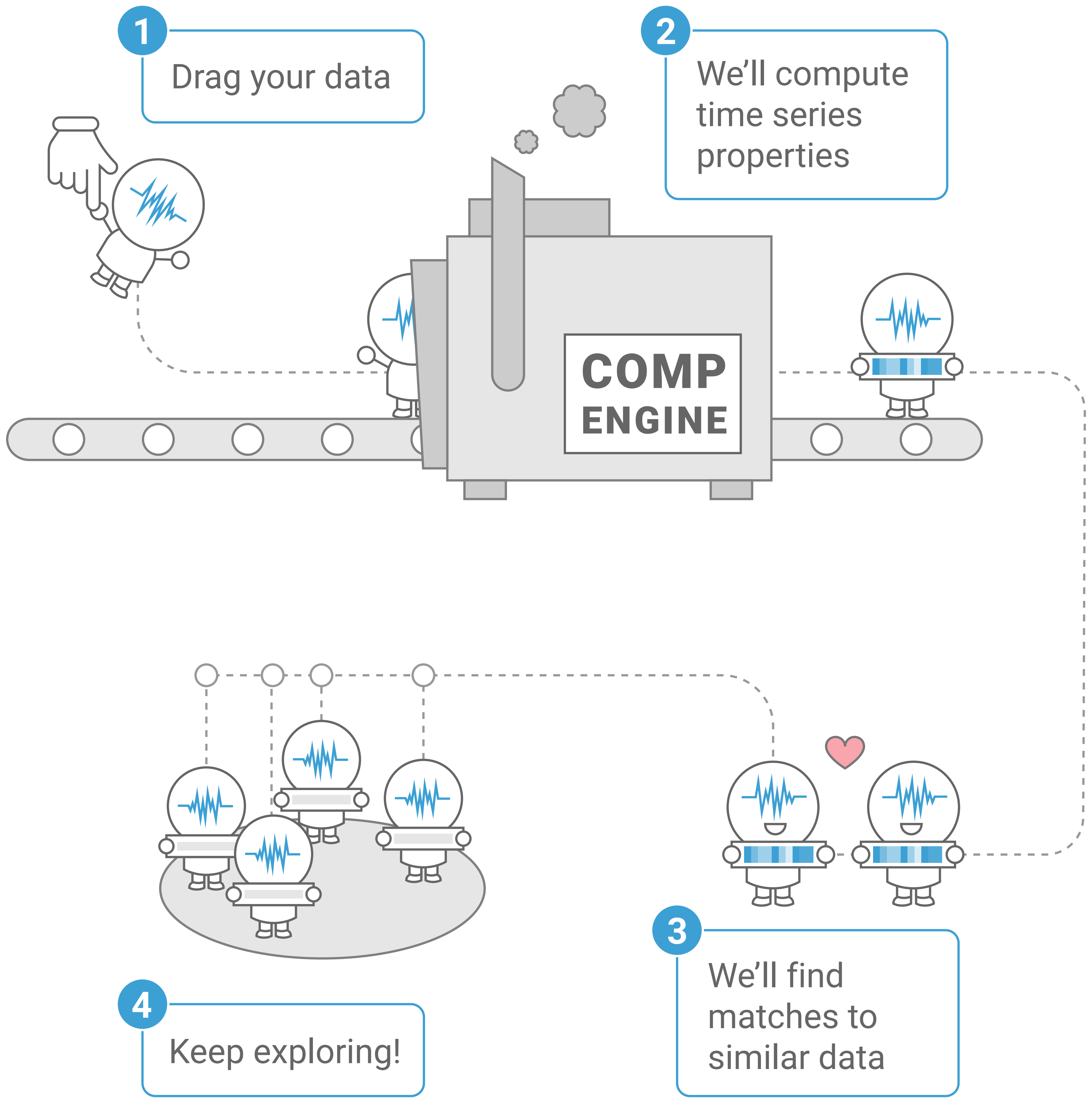
Hand-gesture recognition



Siddiqui et al. Multimodal hand gesture recognition using single IMU and acoustic measurements at wrist. PLoS ONE, **15**, e0227039 (2020).

Finding Connections

Are other scientists studying similar data to me?



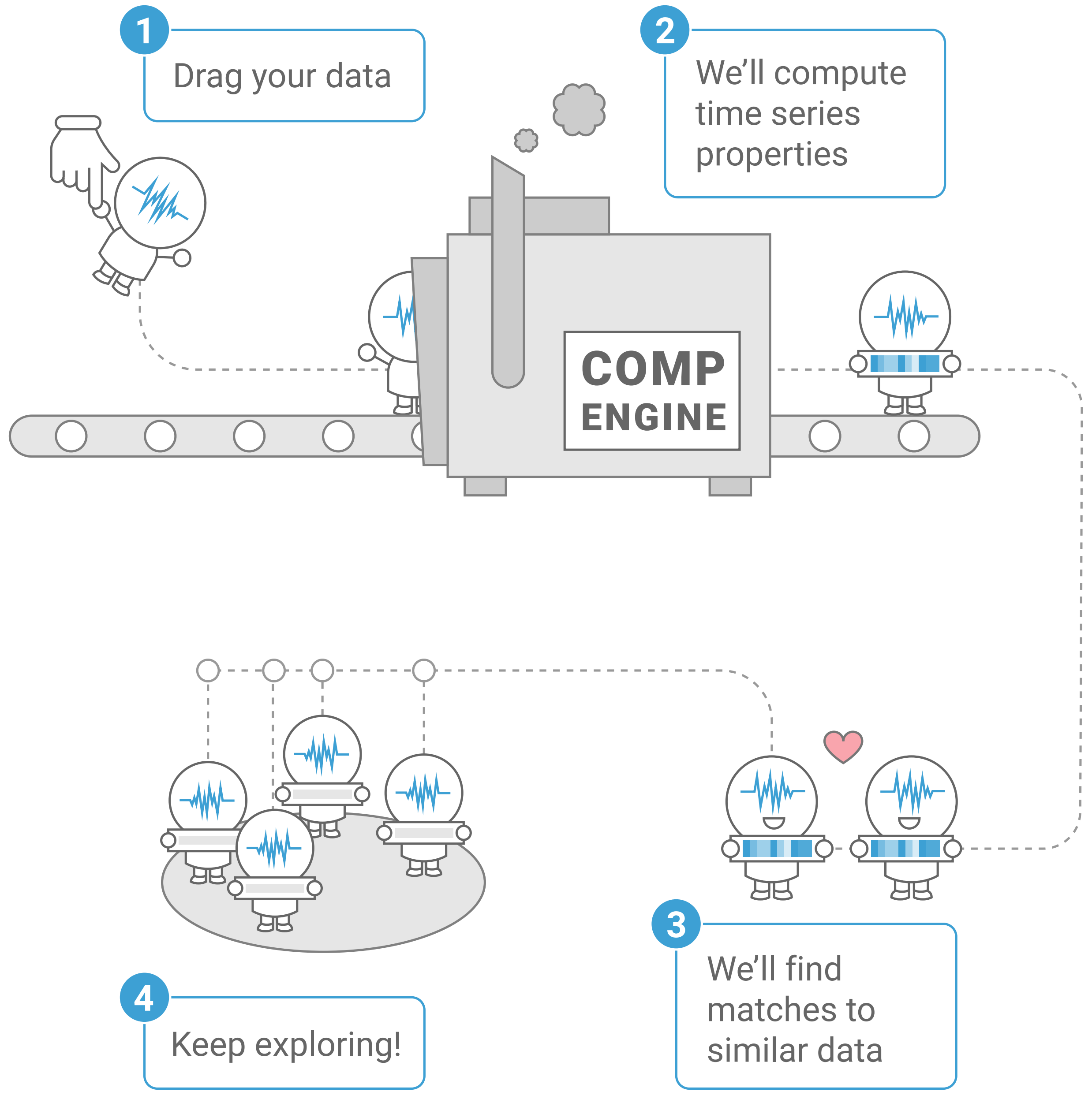
- *CompEngine Time Series* is a self-organizing database of interdisciplinary time-series data
- Connects diverse scientists through the structure of their data
- Bulk download functionality, and API for custom time-series data download: facilitates comprehensive empirical phenotyping of time-series analysis algorithms

www.comp-engine.org

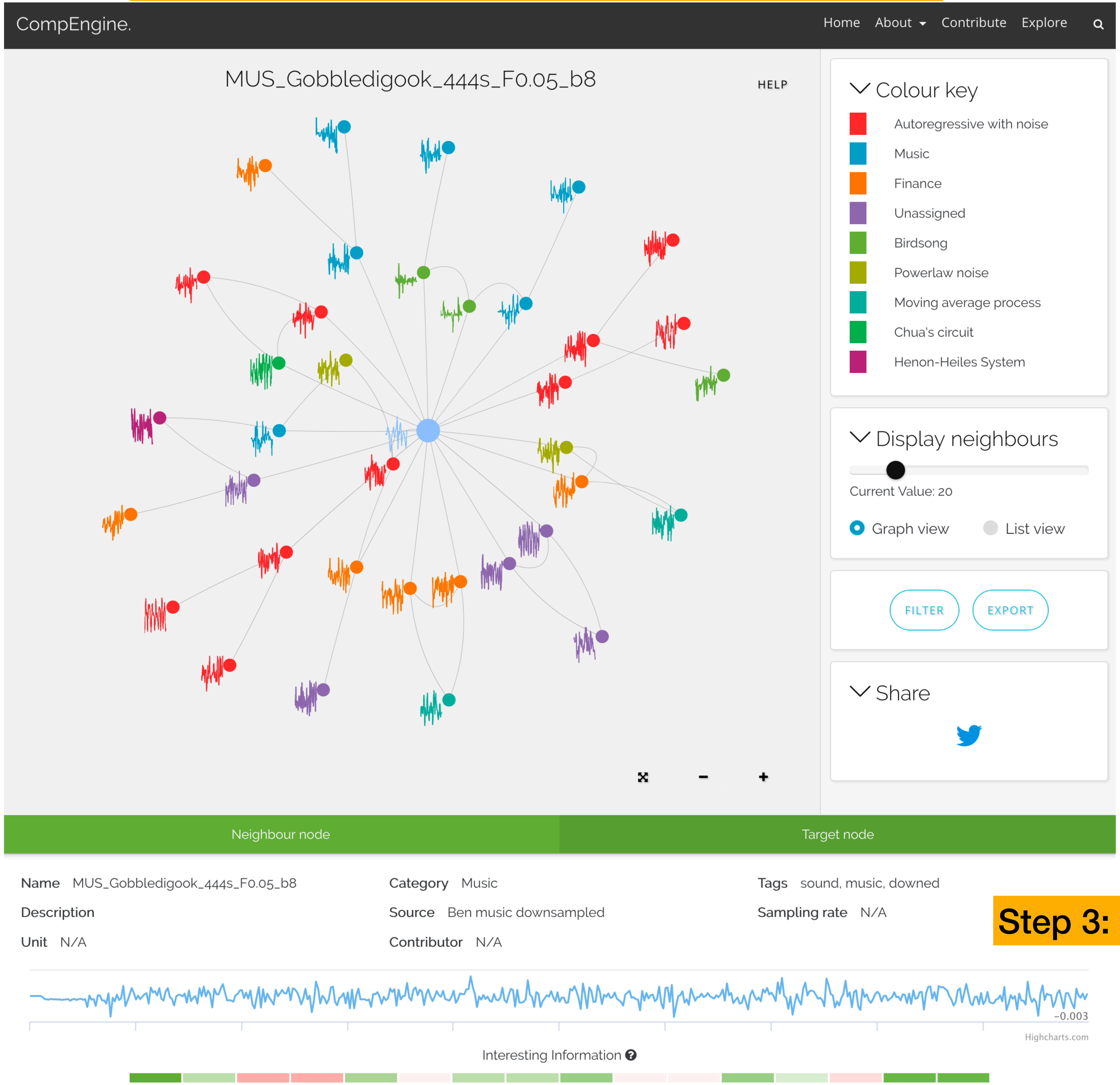
Fulcher et al. A Self-Organizing, Living Library of Time-Series Data. *Scientific Data* **7**: 213 (2020)

Demo

Step 1: Drag on your data



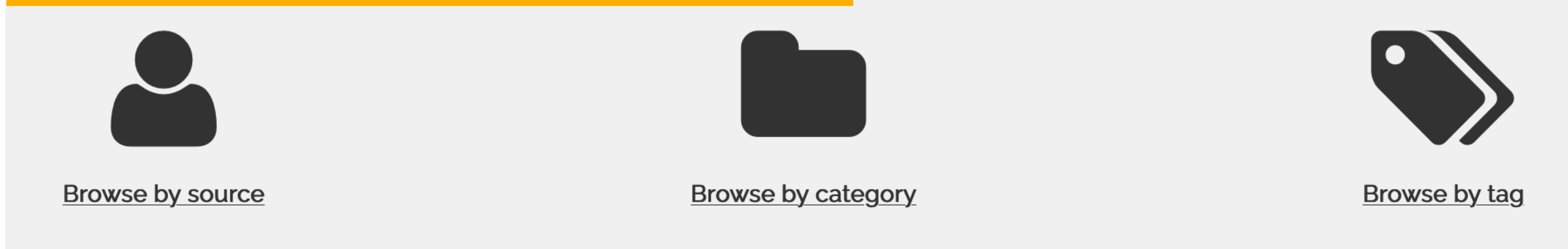
Step 2: Interactively Explore Similar Scientific Data



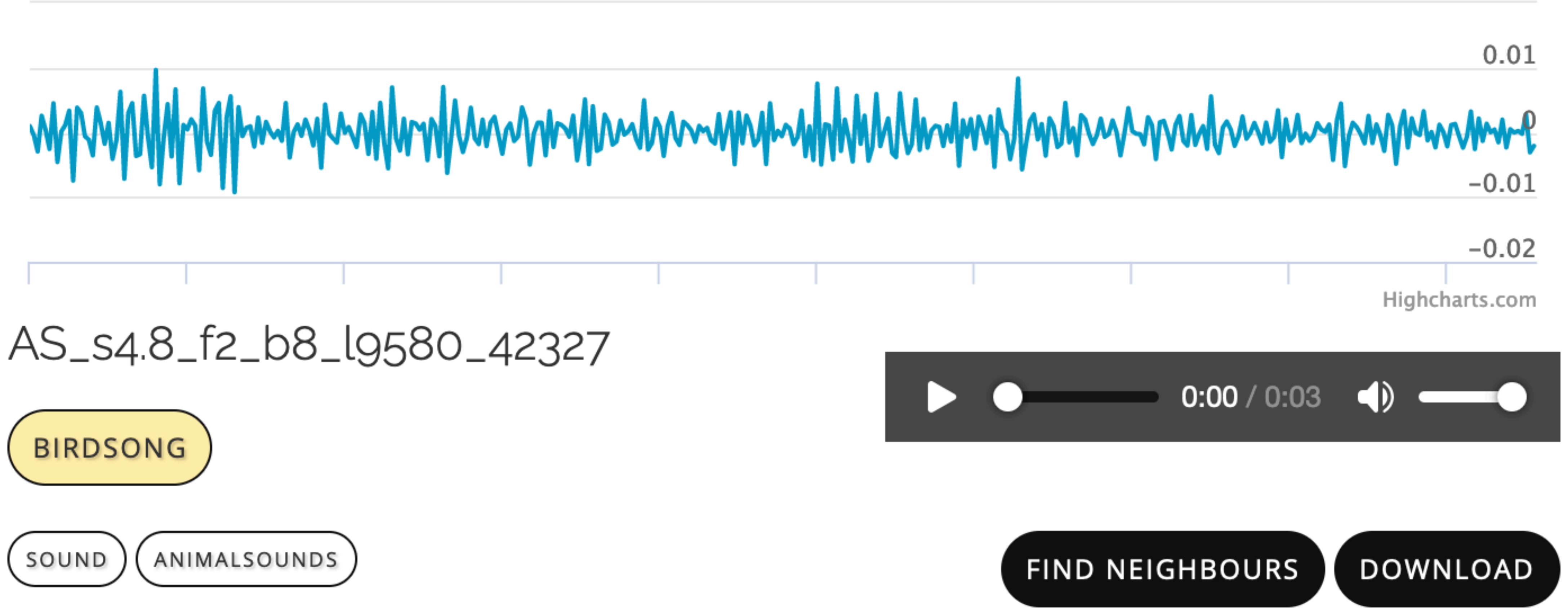
Step 3: Contribute your data

If you don’t have data on-hand, you can still explore

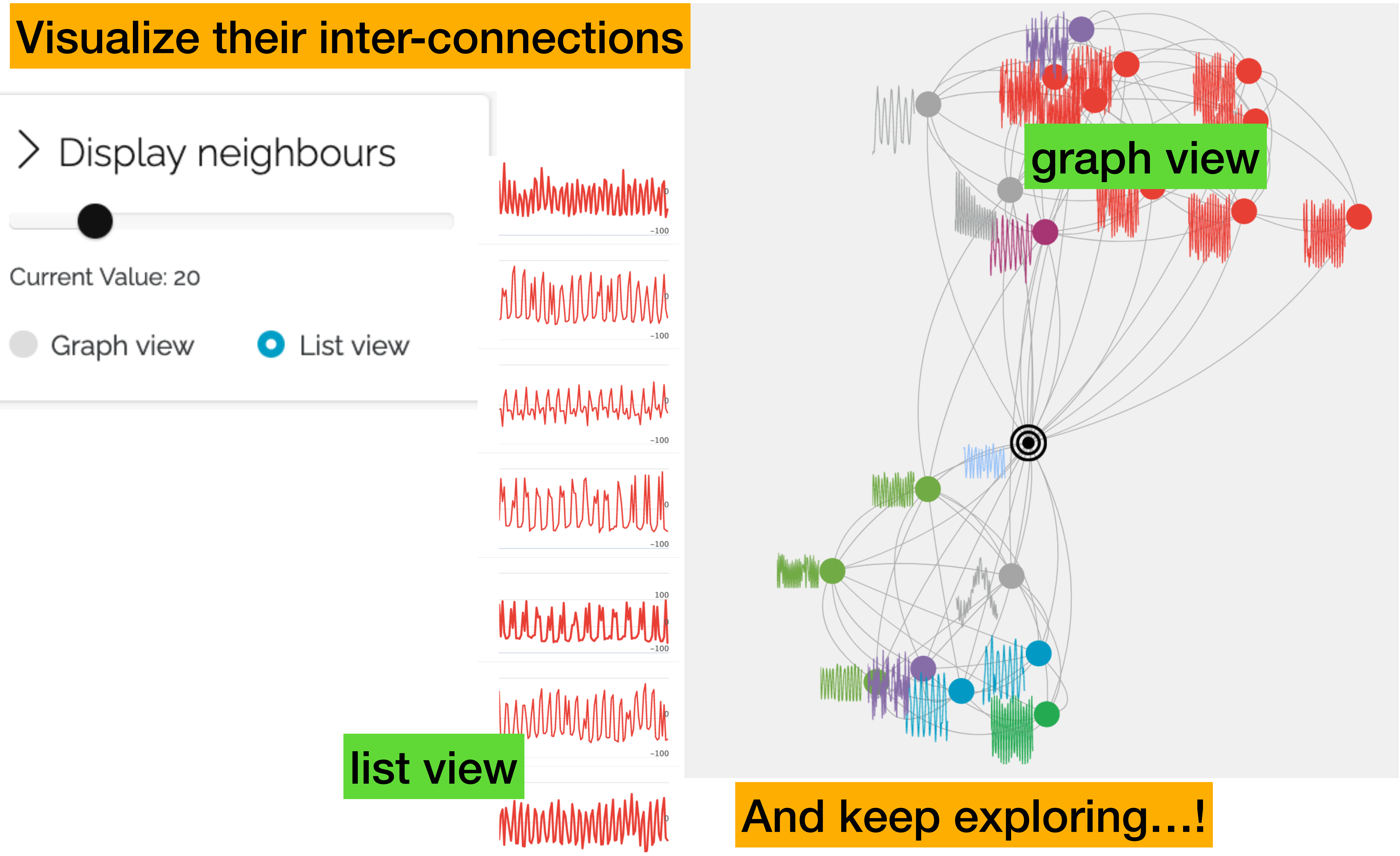
Browse the full time-series library



Interactively Explore Scientific Time Series

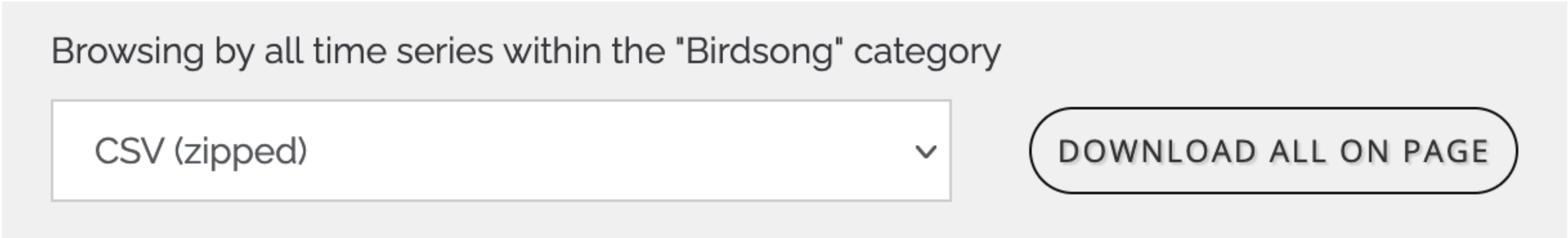


Visualize their inter-connections



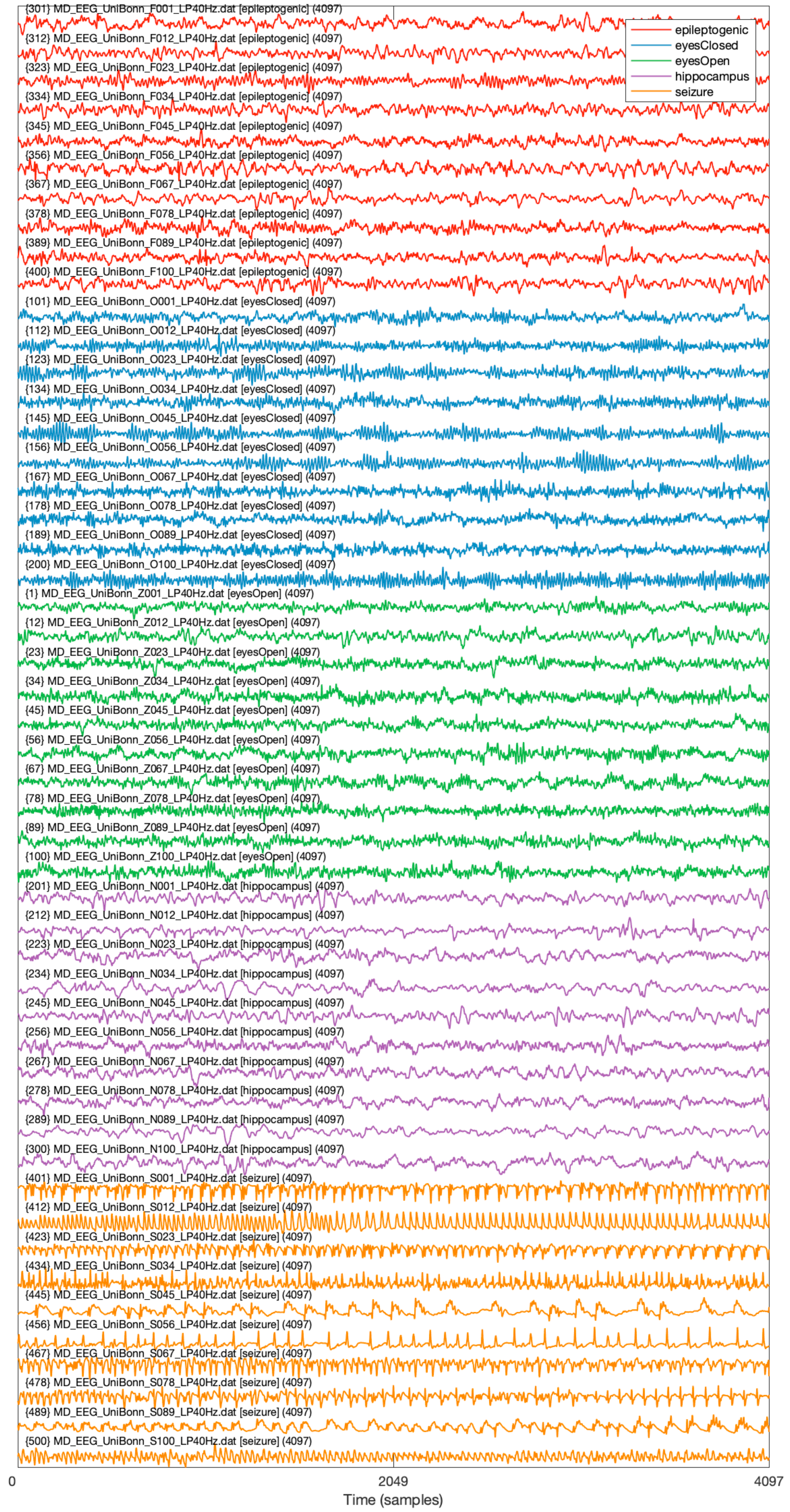
And keep exploring…!

Download any/all data you find:



Demo

100 examples of each of 5 classes of EEG

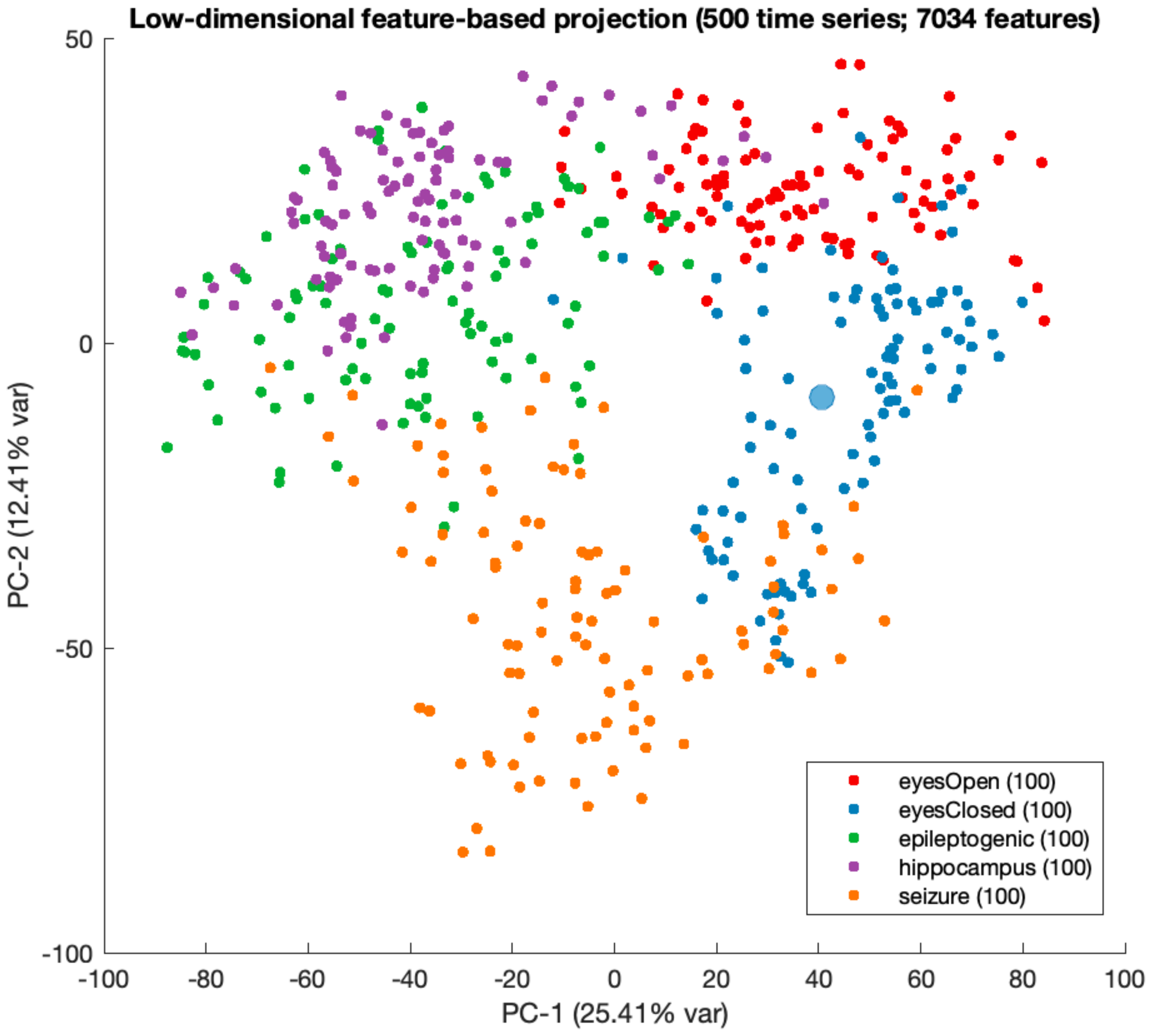


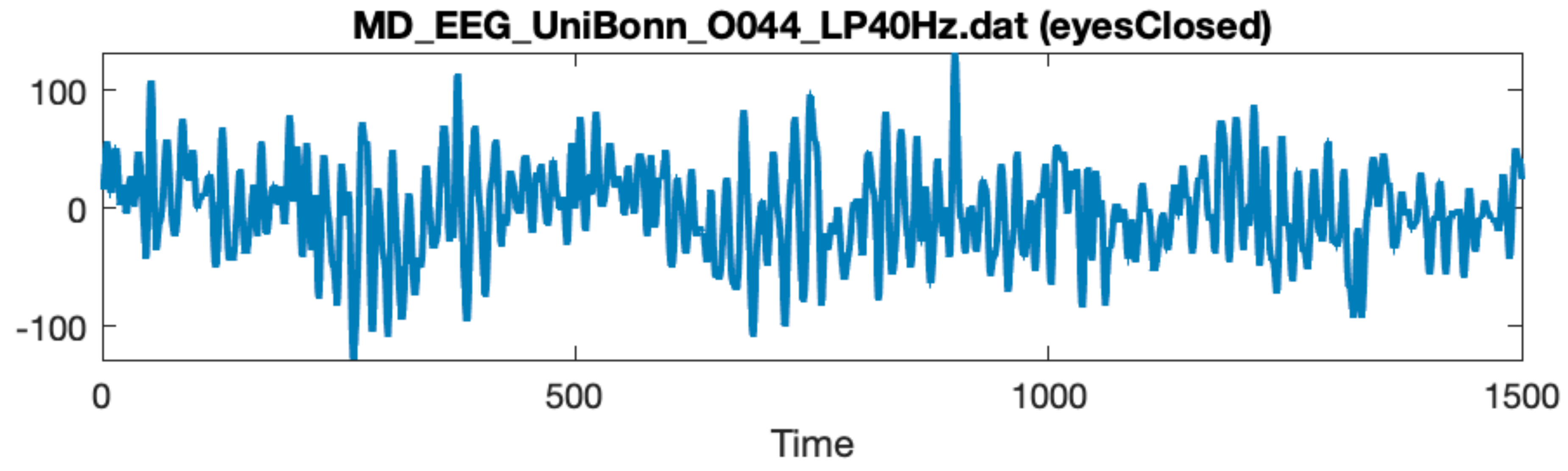
hctsa + catch22

https://github.com/benfulcher/hctsa

https://github.com/chlubba/catch22

Interactive visualization



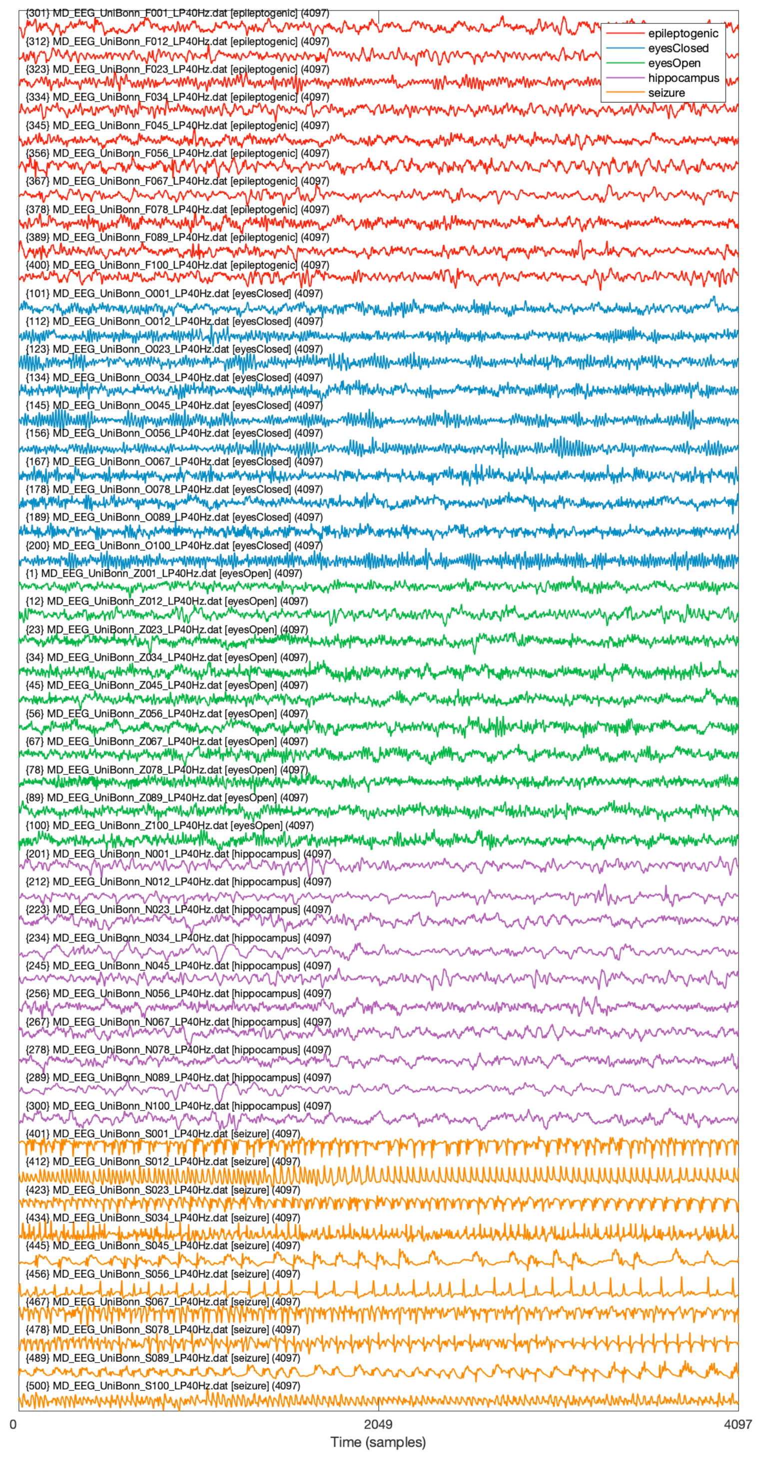


Demo

Load in a dataset → **Compute time-series features** → **Interact with your low-dimensional data visualization**

TS_Init

100 examples of each of 5 classes of EEG



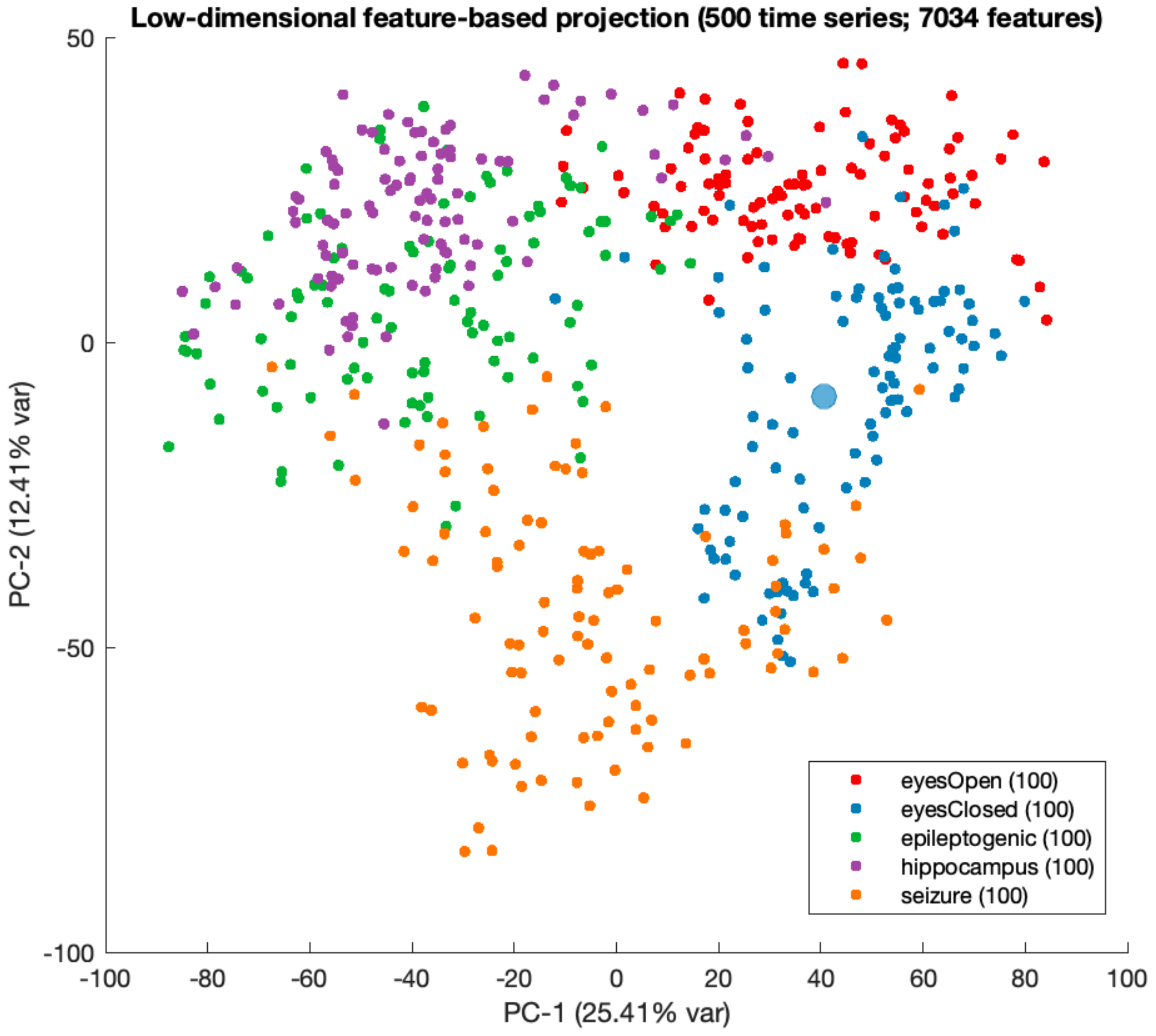
TS_Compute

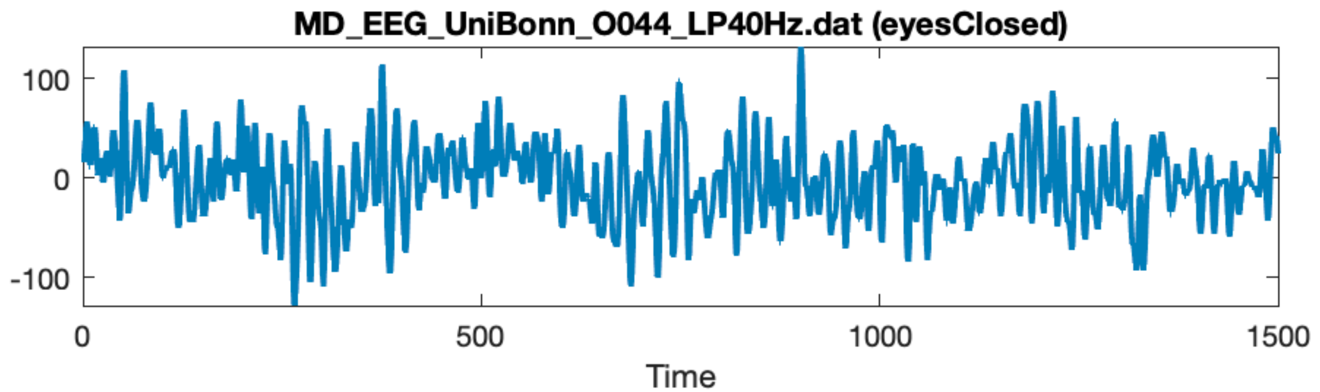
catch22 (22 features) for speed
hctsa (>7k) for comprehensiveness

TS_Normalize

Put all features on a similar scale

TS_LowDimInspect





https://github.com/benfulcher/hctsaTutorial_BonnEEG

1 Prepare Dataset:

`INP_Bonn_EEG.mat`

`labels`	500 x 1 cell	strings uniquely identify each time series
`timeSeriesData`	500 x 1 cell	vectors of time-series data
`keywords`	500 x 1 cell	class labels

2

Initialize (default *hctsa* feature set): `TS_Init('INP_Bonn_EEG.mat')`

Initialize (catch22 feature set): `TS_Init('INP_Bonn_EEG.mat','INP_mops_catch22.txt','INP_ops_catch22.txt',true,'HCTSA_catch22.mat')`

Generates: `HCTSA.mat`

`TS_DataMat`	500 (time series) x 22 (features) matrix [empty]
`TimeSeries`	500-row table with information about time series
`Operations`	22-row table with information about operations/features

3

Compute all features (without parallelization): `TS_Compute(false);`

(very fast for *catch22*)

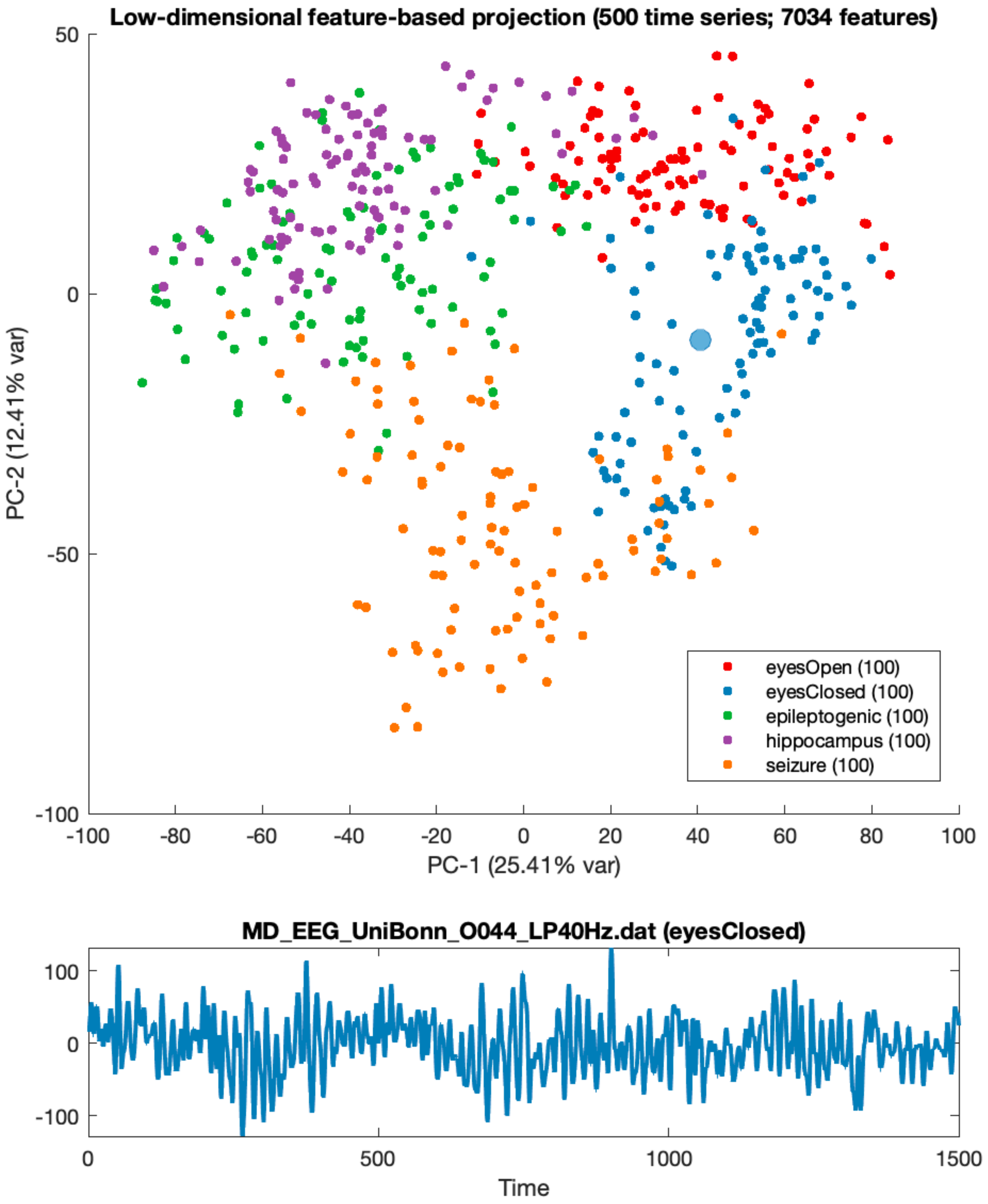
4

Normalize features to a similar scale (and filter poor performers): `TS_Normalize();`

5

Visualize! Analyze! E.g., Play with a low-dimensional representation!: `TS_LowDimInspect();`

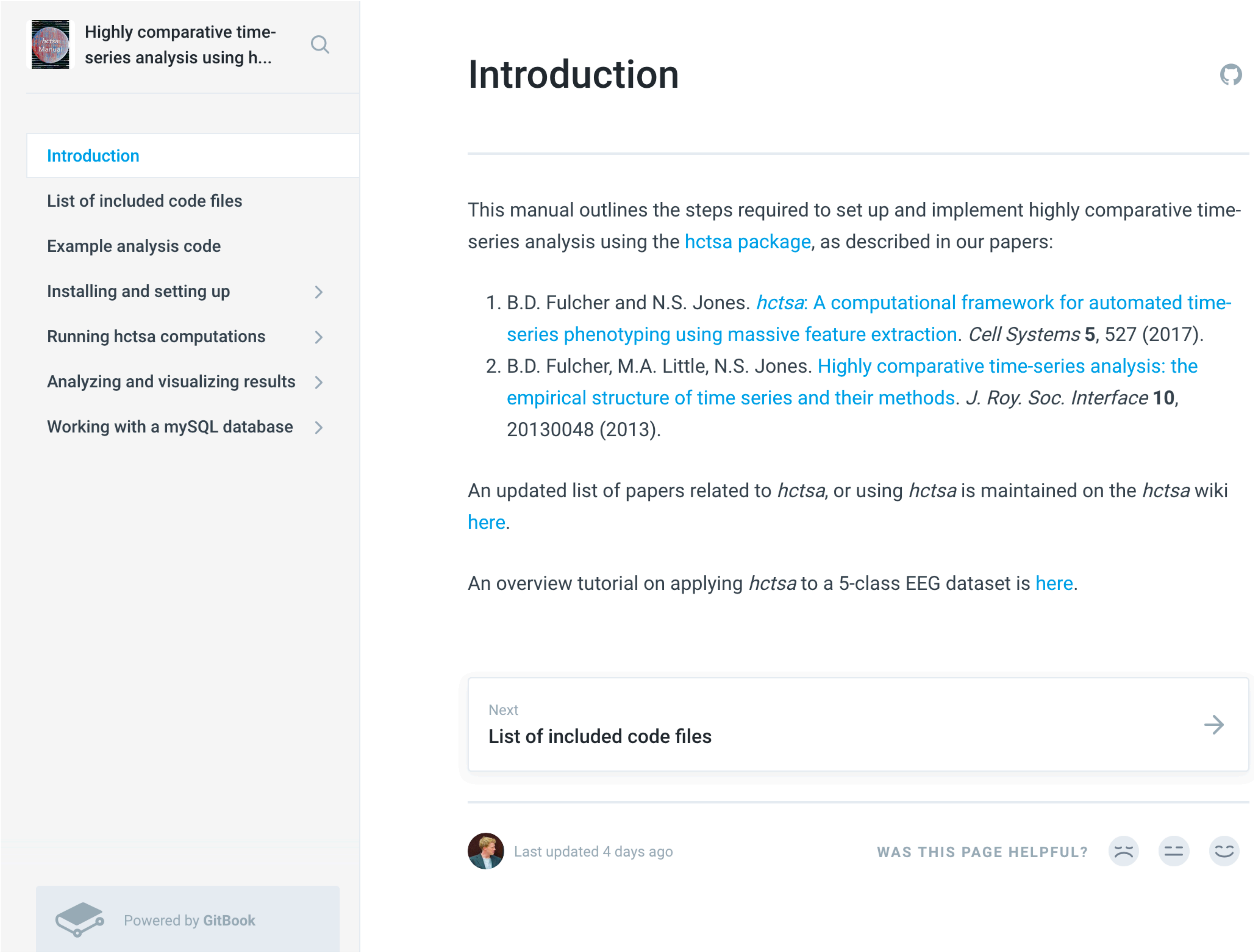
(Many other visualizations: see https://github.com/benfulcher/hctsaTutorial_BonnEEG)



Low-dimensional feature-based projection (500 time series; 7034 features)
PC-2 (12.41% var)
PC-1 (25.41% var)
50
0
-50
-100
-100
-80
-60
-40
-20
0
20
40
60
80
100
eyesOpen (100)
eyesClosed (100)
epileptogenic (100)
hippocampus (100)
seizure (100)
MD_EEG_UniBonn_O044_LP40Hz.dat (eyesClosed)
100
0
-100
0
500
1000
1500
Time

Going Further

Comprehensive documentation on GitBook + wiki



Highly comparative time-series analysis using h...

Introduction

List of included code files

Example analysis code

Installing and setting up

Running hctsa computations

Analyzing and visualizing results

Working with a mySQL database

Powered by GitBook

Introduction

This manual outlines the steps required to set up and implement highly comparative time-series analysis using the hctsa package, as described in our papers:

1. B.D. Fulcher and N.S. Jones. *hctsa*: A computational framework for automated time-series phenotyping using massive feature extraction. *Cell Systems* **5**, 527 (2017).
2. B.D. Fulcher, M.A. Little, N.S. Jones. Highly comparative time-series analysis: the empirical structure of time series and their methods. *J. Roy. Soc. Interface* **10**, 20130048 (2013).

An updated list of papers related to *hctsa*, or using *hctsa* is maintained on the *hctsa* wiki here.

An overview tutorial on applying *hctsa* to a 5-class EEG dataset is here.

Next

List of included code files

Last updated 4 days ago

WAS THIS PAGE HELPFUL?

https://hctsa-users.gitbook.io/hctsa-manual/

https://github.com/benfulcher/hctsa/wiki

Work through the full suite of *hctsa* functionality for this dataset:

https://github.com/benfulcher/hctsaTutorial_BonnEEG)

Work through other *hctsa* analyses for fly and worm phenotyping (open code and pre-computed data):

https://github.com/benfulcher/hctsa_phenotypingFly

https://github.com/benfulcher/hctsa_phenotypingWorm

FYI:

*Using reduced feature sets (like catch22), there is similar functionality in **theft** or through a drag-and-drop online interface*

theft

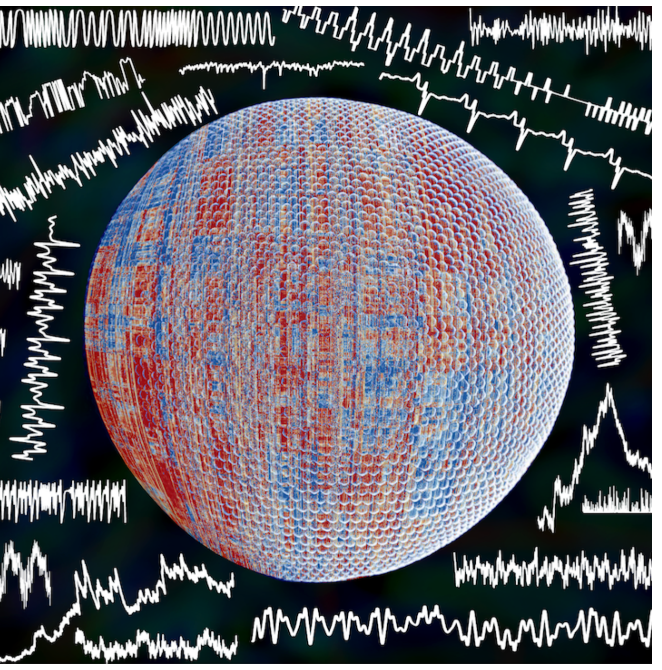
Feature computation, analysis, and visualization for feature-based time-series analysis

https://github.com/hendersontrent/theft

theft web portal

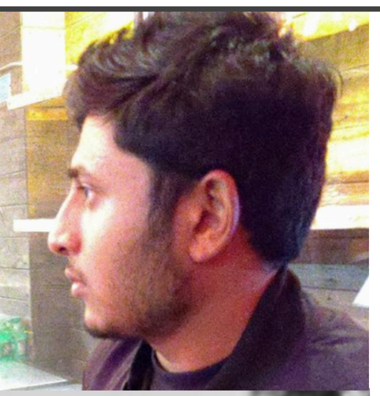
Drag-and-drop online access to theft functionality

https://dynamicsandneuralsystems.shinyapps.io/timeseriesfeaturevis/



Acknowledgements

ben.fulcher@sydney.edu.au
@bendfulcher
@compTimeSeries
benfulcher
www.benfulcher.com
www.comp-engine.org/
github.com/benfulcher/hctsa
github.com/chlubba/catch22



Diptanshu Mittal

Google Summer of Code

Oliver Cliff

Trent Henderson

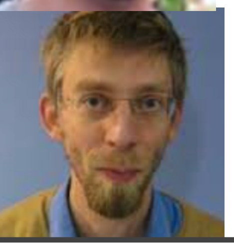
THE UNIVERSITY OF SYDNEY

https://dynamicsandneuralsystems.github.io/

Sarab Sethi

Carl Lubba

Imperial College London



Nick Jones

Selected References:

- Fulcher et al. (2013). Highly comparative time-series analysis: the empirical structure of time series and their methods. *J. Roy. Soc. Interface*.
- Fulcher & Jones (2014). Highly comparative feature-based time-series classification. *IEEE Trans. Knowl. Data Eng*.
- Fulcher & Jones (2017). *hctsa*: A Computational Framework for Automated Time-Series Phenotyping Using Massive Feature Extraction. *Cell Systems*.
- Sethi et al. (2017). Structural connectome topology relates to regional BOLD signal dynamics in the mouse brain. *Chaos*.
- Fulcher (2018). Feature-based time-series analysis, *Feature Engineering,* CRC Press.
- Lubba et al. (2020). catch22: CAnonical Time-series CHaracteristics. *Data Mining and Knowledge Discovery*.
- Fulcher et al. (2020). CompEngine: a self-organizing, living library of time-series data. *Scientific Data*.
- Markicevic et al. (2020). Cortical Excitation:Inhibition imbalance causes abnormal brain network dynamics as observed in neurodevelopmental disorders. *Cerebral Cortex*.
- Fallon et al. (2020). Timescales of spontaneous fMRI fluctuations relate to structural connectivity in the brain. *Network Neuroscience*.
- Shafiei et al. (2020). Topographic gradients of intrinsic dynamics across neocortex. *eLife*.

https://github.com/benfulcher/hctsa/wiki/Publications-using-hctsa