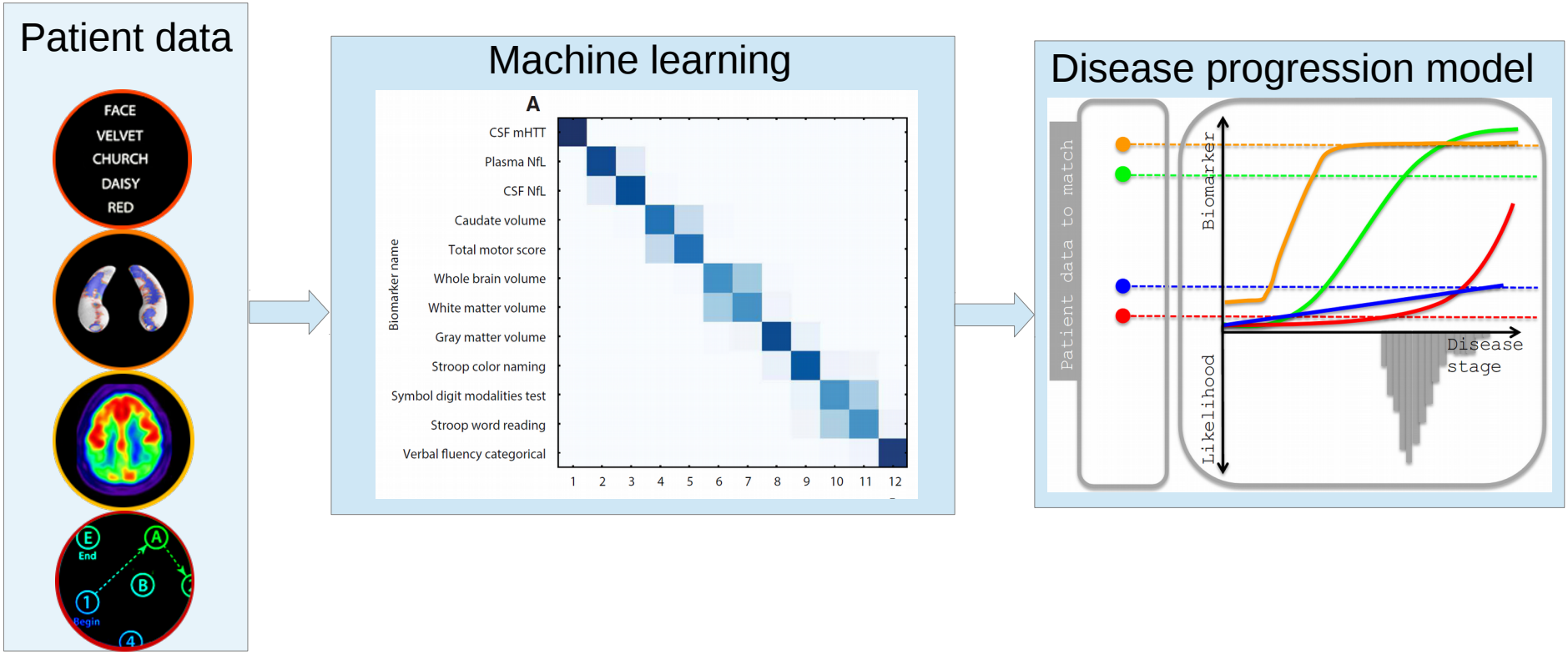
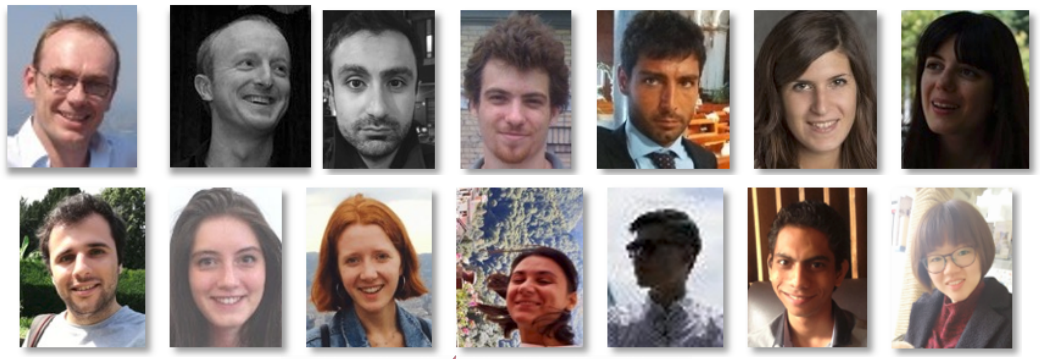
Data driven models of disease progression

Peter Wijeratne
MRC Senior Research Fellow
UCL Centre for Medical Image Computing



Acknowledgements

pond



HUNTINGTON'S DISEASE CENTRE

And all the participants of the Huntington's disease studies used here.

UK Research and Innovation

MRC Medical Research Council

CHDI FOUNDATION

EPSRC Engineering and Physical Sciences Research Council

Centre for Medical Image Computing (CMIC)

Maths, physics and engineering scientists at the interface of basic and biomedical sciences

CMIC

Great Ormond Street Hospital

University College London Hospital

Moorfield's Eye Hospital

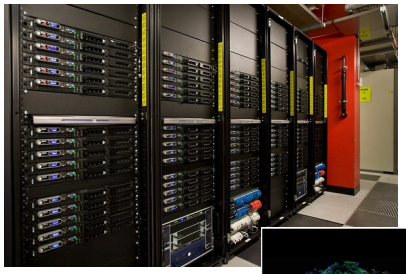
Royal Free Hospital

Royal National Orthopaedic Hospital

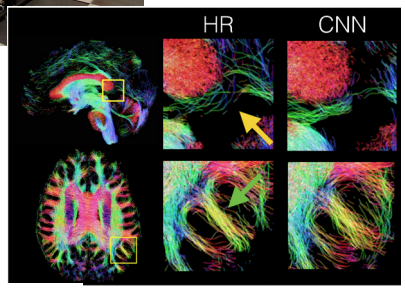
POND: bridging the gap

Basic sciences

cmic
Centre for Medical Image Computing

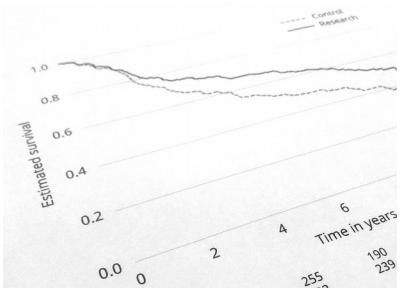


Cluster computing



Imaging + machine learning

UCL EPSRC CDT in Medical Imaging



Statistical methods

pond
Progression Of Neurodegenerative Disease

Clinical sciences

HUNTINGTON'S DISEASE CENTRE

Leonard Wolfson Experimental Neurology Centre

Eisai

Advanced imaging

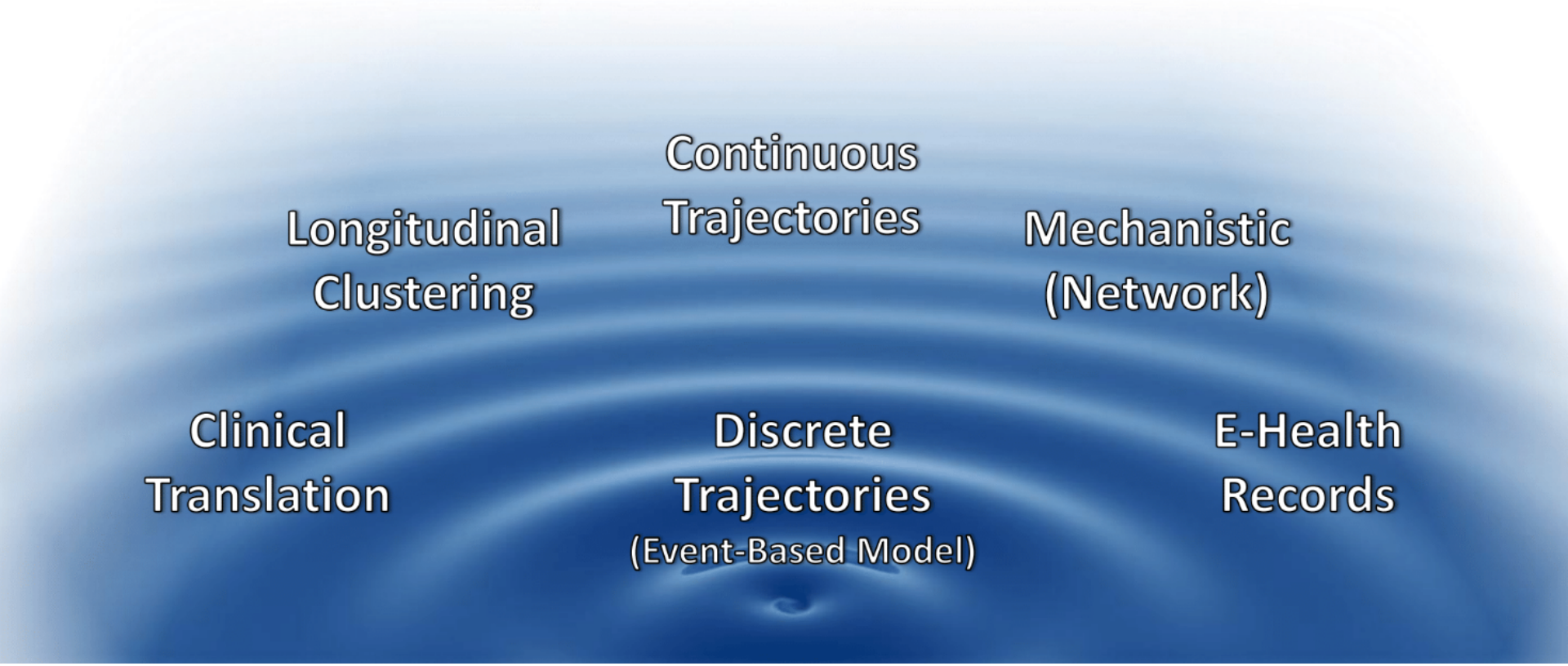
wellcome centre human neuroimaging

Clinical trials

POND: capabilities

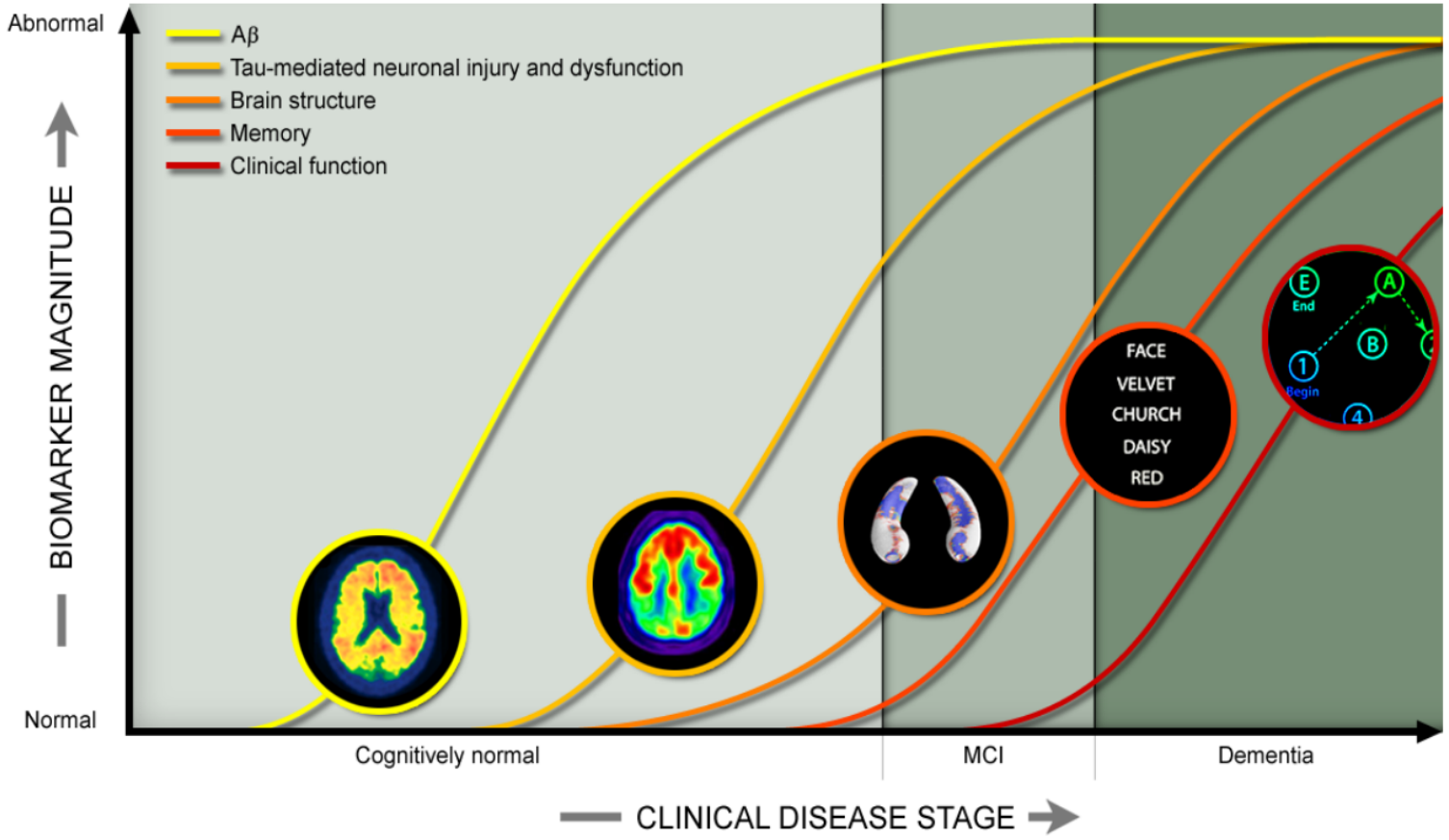
pond

Progression Of Neurodegenerative Disease



High level: Disease progression modelling

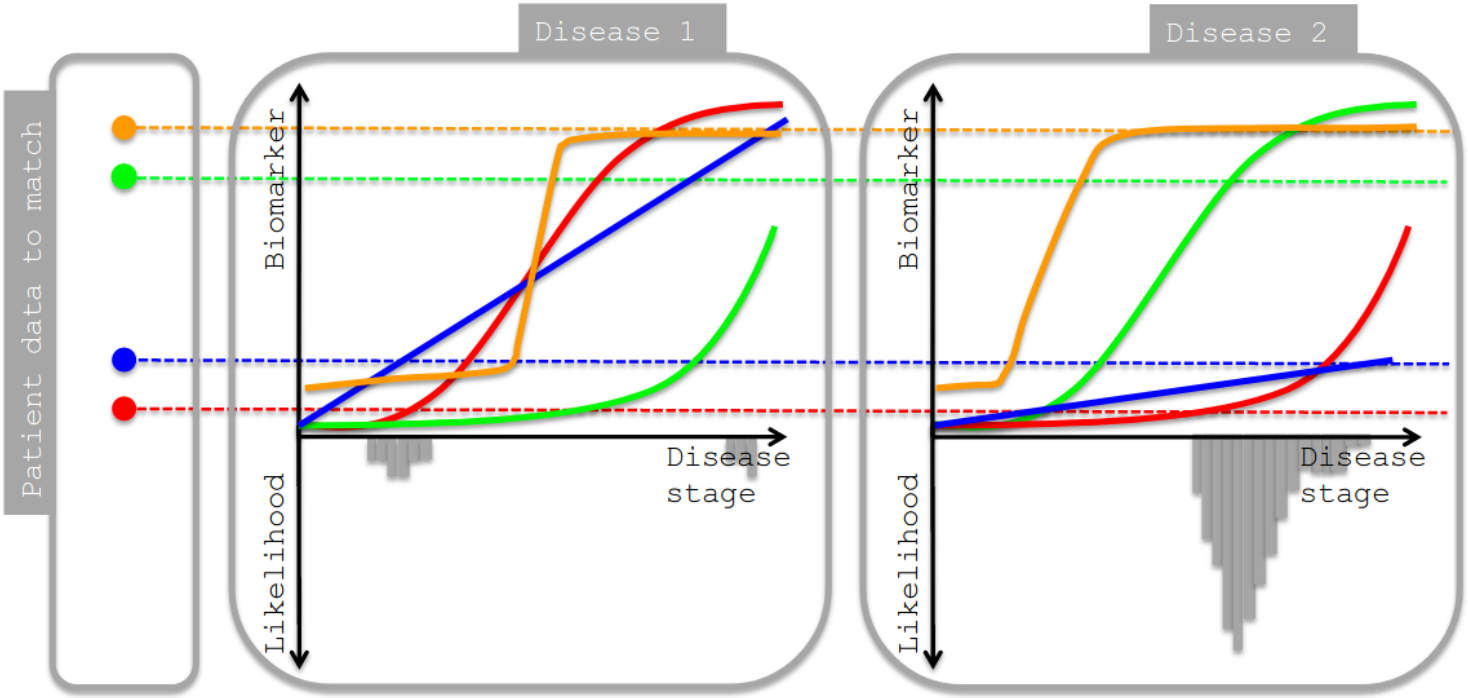
http://adni.loni.usc.edu/study-design/#background-container



A picture of how components of a disease progresses over time

High level: Disease progression modelling

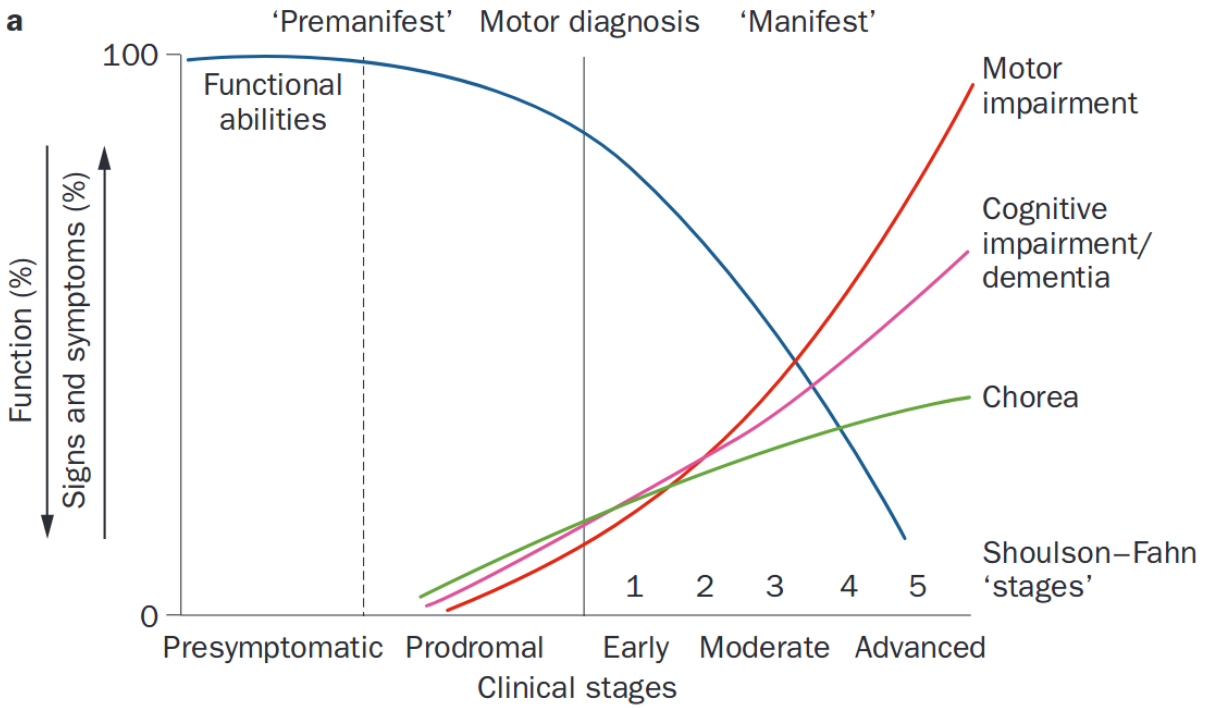
Disease progression models learn patterns of disease-related changes from data



- Can use models to infer temporal ordering of changes
- Can also stage and stratify patients →clinical trial design

Real application: neurological disease

Can we estimate where a patient is along their disease path?

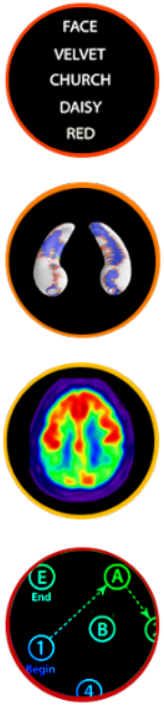


Patient stage is a latent variable – it generates the observed measurements, but is not measured directly (unlike in physics events, where we know time)

→ Infer using statistical and machine learning methods

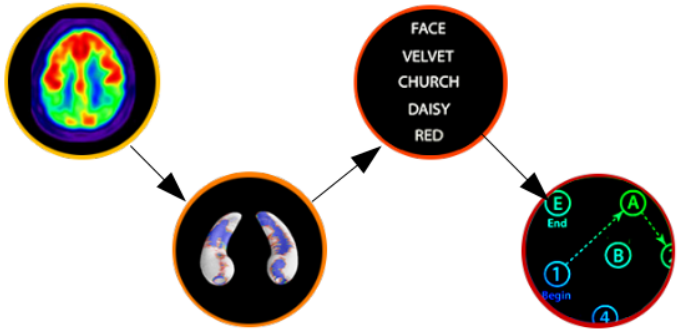
Real application: multi-modal data

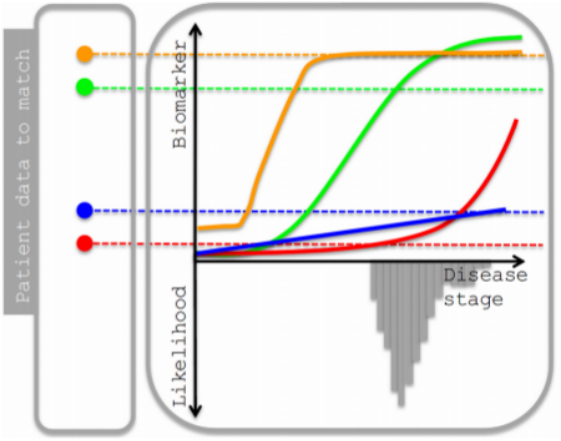
Disease progression models learn patterns of disease-related changes from data



Patient data

Machine learning





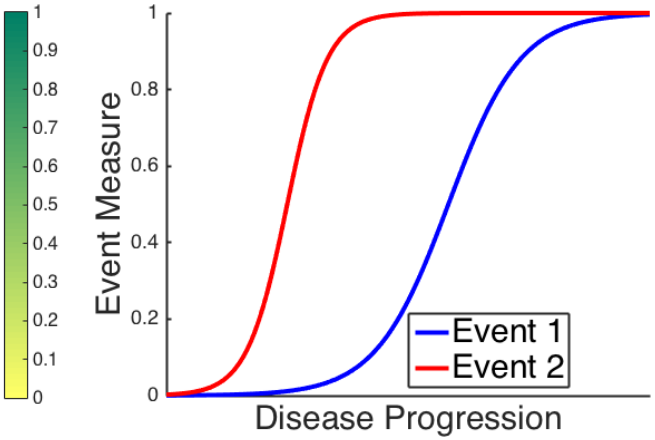
Disease progression model

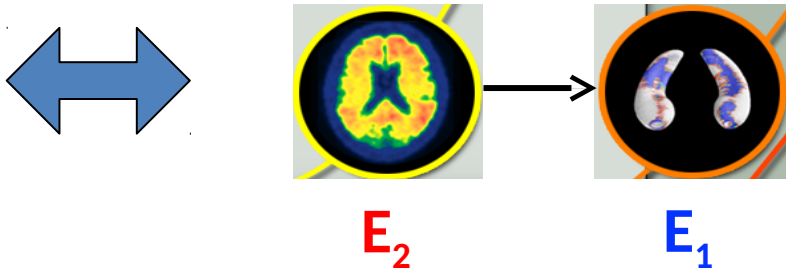
- Can use models to infer temporal ordering of changes
- Can also stage and stratify patients →clinical trial design

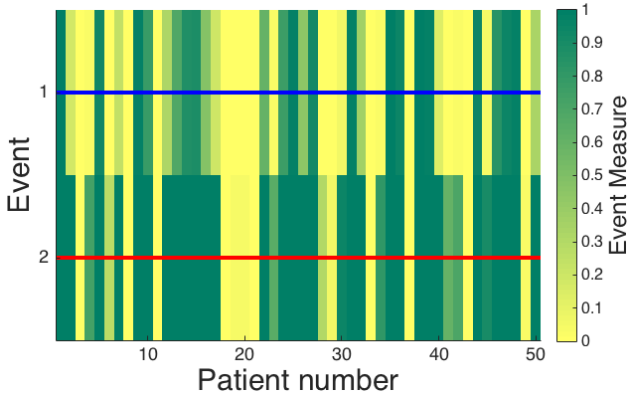
Example: Event-based model (EBM)

EBM estimates ordering of **binary events** from data – normal or abnormal

Data can be cross-sectional and any combination of types (imaging, clinical, genetic...)







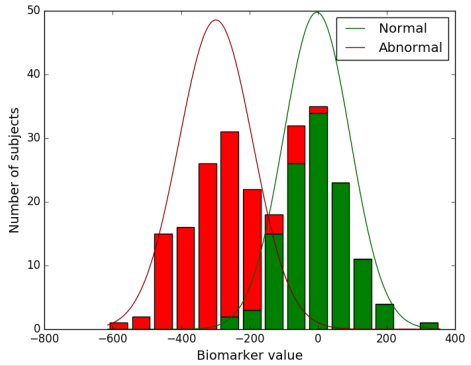
<u>Simple example: 2 event measures</u>

More patients have greater abnormality in Event 2 than Event 1

→Event 2 **measurably abnormal** before Event 1

Example: Event-based model (EBM)

More formally: EBM is a generative model of observed data from unknown sequence



data

uniform prior

$$P(X|S)=\prod_{j=1}^{N}\left[\sum_{k=0}^{Z}\left(P(k)\prod_{i=1}^{k}P(x_{ij}|E_i)\prod_{i=k+1}^{Z}P(x_{ij}|\neg E_i)\right)\right]$$

sequence

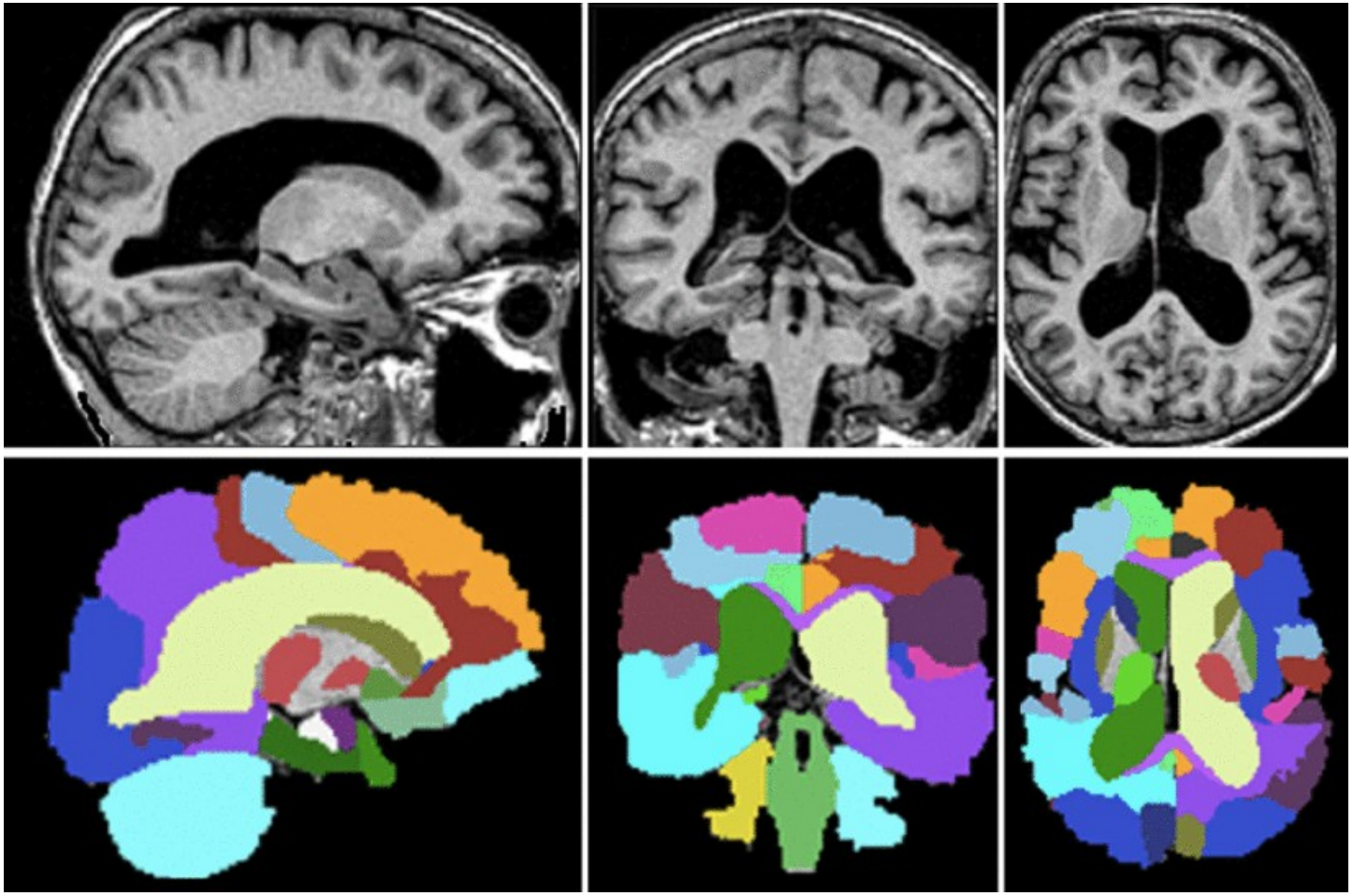
prob. Abnormal

prob. Normal

- The EBM needs likelihood distributions for normal and abnormal subjects

→Learn directly from data

Example: imaging data



Extract regional brain volumes using Geodesic Information Flows*

→ Reduces inter-subject variability by using spatially variant graphs to connect morphologically similar subjects

* MJ Cardoso *et al*. Geodesic Information Flows: Spatially-Variant Graphs and Their Application to Segmentation and Fusion. IEEE Transactions on Medical Imaging, 34 (2015), pp. 1976-1988, doi: 10.1109/TMI.2015.2418298

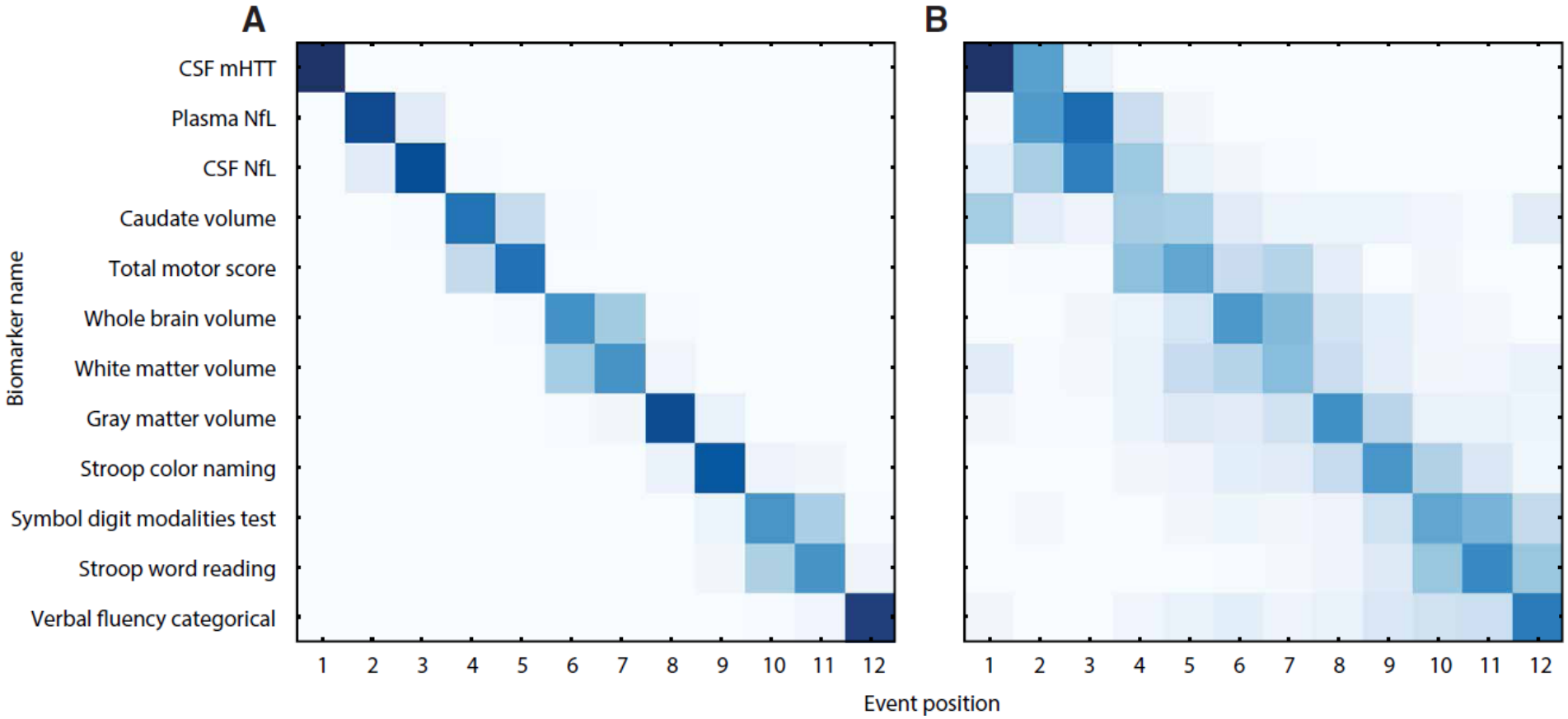
EBM in Huntington's disease

SCIENCE TRANSLATIONAL MEDICINE | RESEARCH ARTICLE

HUNTINGTON'S DISEASE

Evaluation of mutant huntingtin and neurofilament proteins as potential markers in Huntington's disease

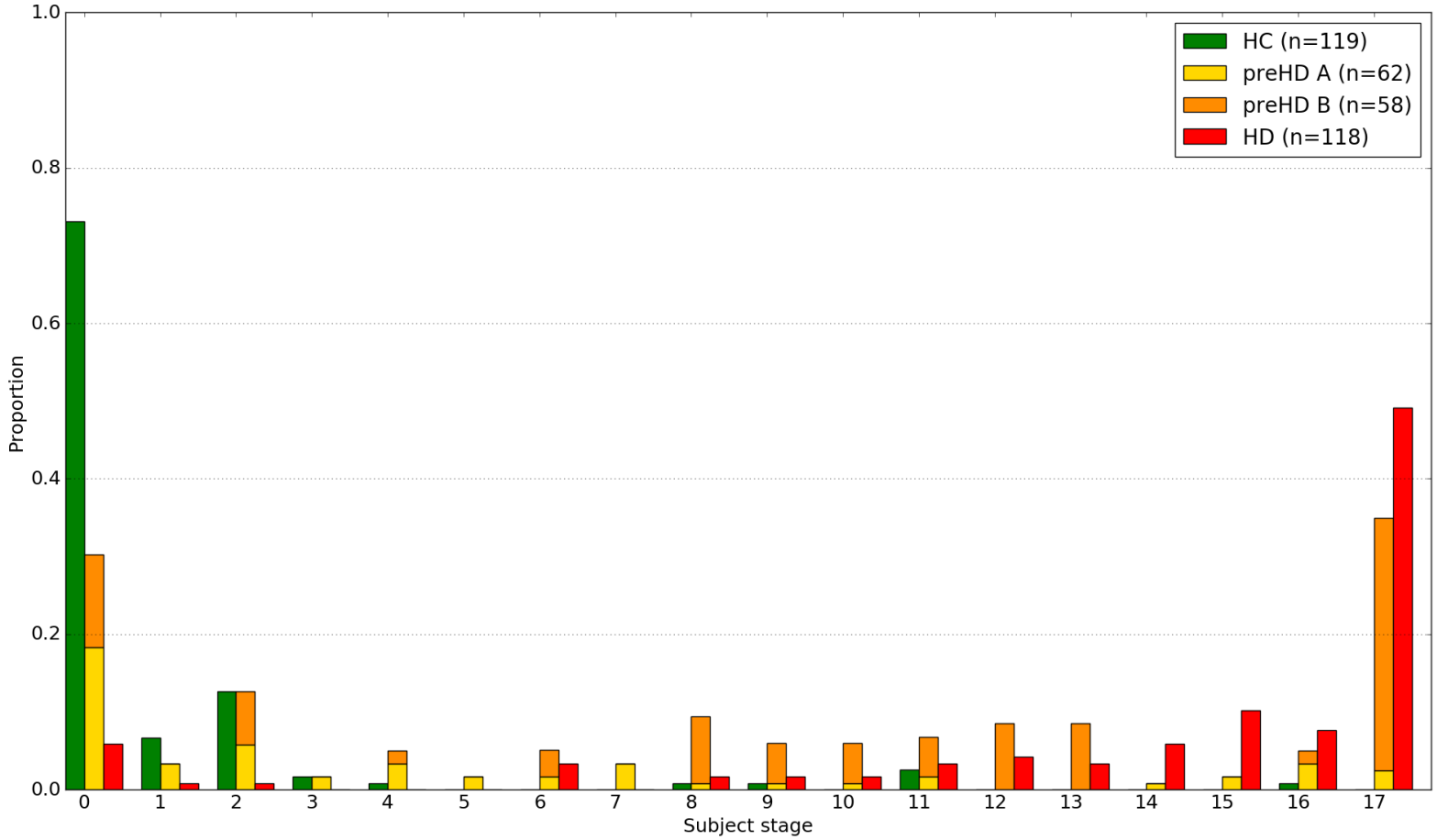
Lauren M. Byrne[1*†], Filipe B. Rodrigues[1†], Eileanor B. Johnson[1], Peter A. Wijeratne[2], Enrico De Vita[3,4], Daniel C. Alexander[2,5], Giuseppe Palermo[6], Christian Czech[6], Scott Schobel[6], Rachael I. Scahill[1], Amanda Heslegrave[7], Henrik Zetterberg[7,8,9,10], Edward J. Wild[1*]



- Finds that biofluid markers change before imaging and clinical markers

Simplest way is to take the stage that maximises the likelihood for each patient

$$argmax_k P(X_j|\bar{S},k) = argmax_k P(k) \prod_{i=1}^{k} P(x_{ij}|E_i) \prod_{i=k+1}^{I} P(x_{ij}|\neg E_i)$$



+ estimate sequence from data, instead of a priori

+ no a priori biomarker thresholds – learned from data

+ Bayesian → characterise uncertainty

+ naturally extends to any type of dynamic biomarker

+ only needs cross-sectional data

- assumes events are binary

- assumes single event sequence over sample

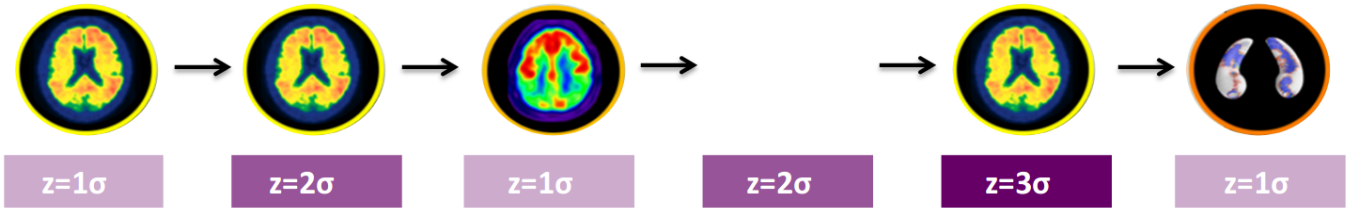
→ SuStaIn

- assumes measurements are independent

- assumes no covariance between event measures

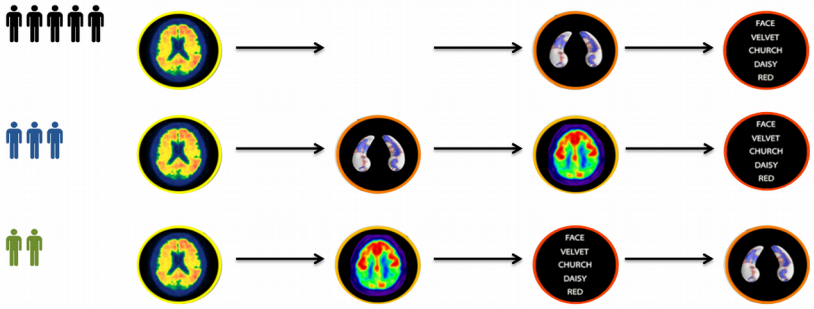
- requires a prior labelling

1. Continuous generalisation of EBM: instead of instantaneous abnormality, markers are a linear combination of z-scores



"Z-score model"

2. Total model is mixture of linear z-score models: grouped into clusters with distinct progression patterns



"Algorithm"

1. Continuous generalisation of EBM: instead of instantaneous abnormality, markers are a linear combination of z-scores

$$P(\mathbf{X}|\mathbf{S}) = \prod_{j=1}^{J}\left[\sum_{k=0}^{N}\left(\int_{t=\frac{k}{N+1}}^{t=\frac{k+1}{N+1}}\left(P(t)\prod_{i=1}^{I}P\left(x_{ij}|t\right)\right)\partial t\right)\right]$$

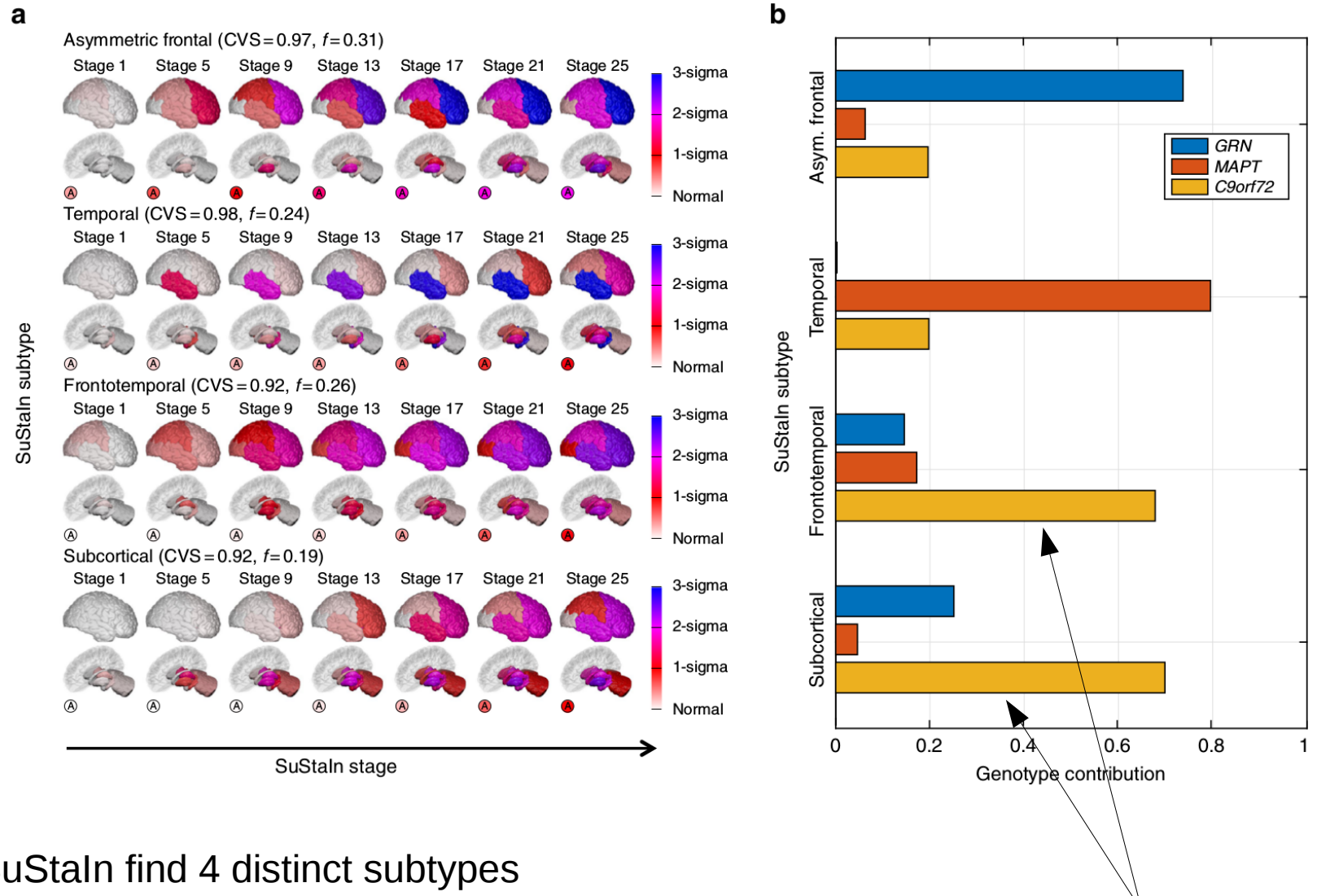
“Z-score model”

2. Total model is mixture of linear z-score models: grouped into clusters with distinct progression patterns

$$P(\mathbf{X}|\mathbf{M}) = \sum_{c=1}^{C} f_c P(\mathbf{X}|\mathbf{S}_c)$$

“Algorithm”

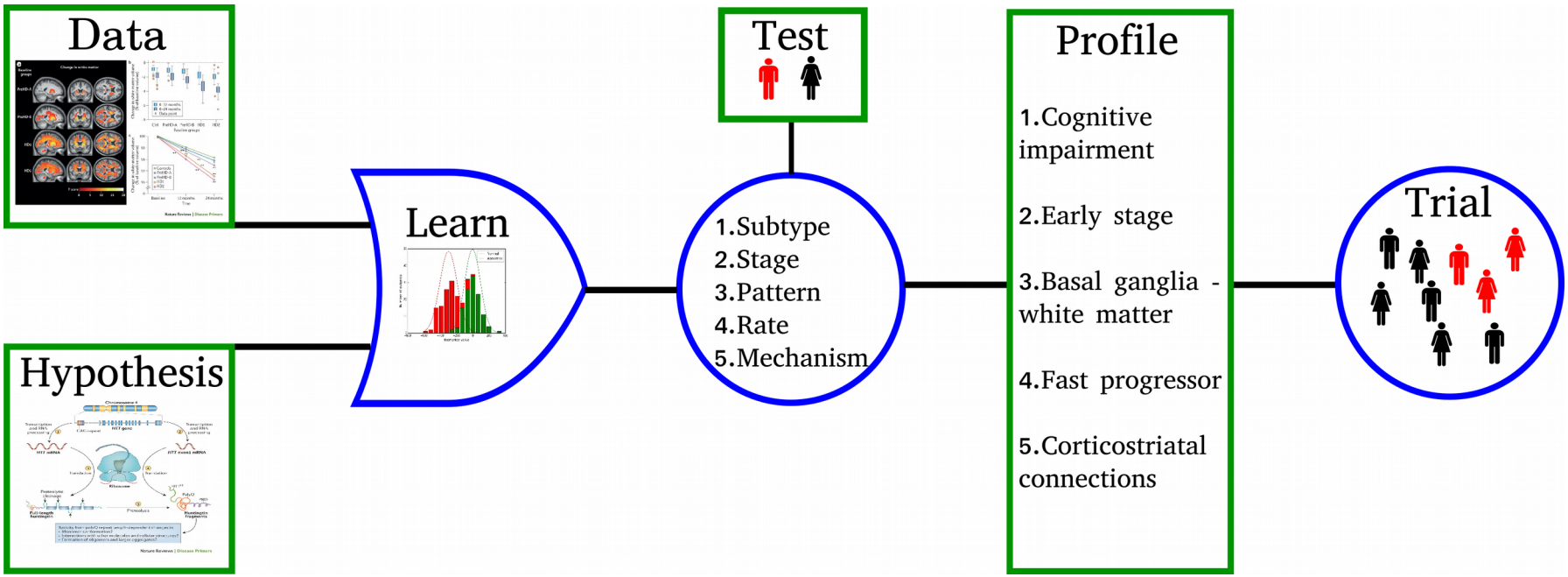
Results: SuStaIn in GENFI



- SuStaIn find 4 distinct subtypes
- Subtypes show genetic dependency: within-geneotype phenotypes
- Maintained in CV (>93% similarity)

Big vision: computational models for clinical trials

Patient data + machine learning = personalised profiles for clinical trial design



Model can be used for both prospective and retrospective analysis

→ Save money and time
→ Optimise trial design

Summary: disease progression models

- Event-based model (EBM) can be used to uncover a sequence of events across a population

- Subtype and stage inference (SuStaIn) can be used to uncover multiple sequences of events across a population

- Provides clinically meaningful prognostic information –

    - Most likely disease stage
    - Most likely next stage
    - Underlying sequence of changes
    - Data-driven subtypes

- Can easily extend model to include any types of dynamic marker – different imaging modalities (e.g. DWI, PET), biofluids

UCL-Japan collaboration – call to action

"Machine learning and computational modelling in the clinic"

How can EBM and SuStaIn be applied to data at Juntendo Hospital?

1. Subtypes in Parkinson's
2. Subtypes in epilepsy
3. Subtypes in natural ageing

The system could eventually be used to

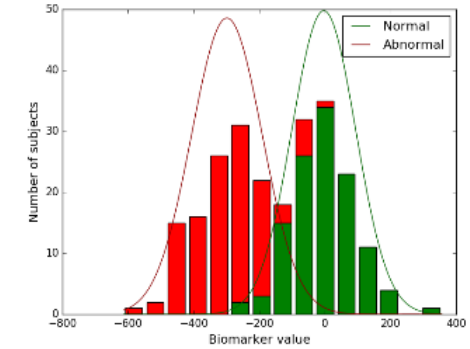
1. Provide fine-grained stratification for clinical trials
2. Aid in treatment planning in a clinical setting

Question – what data would you like to model?

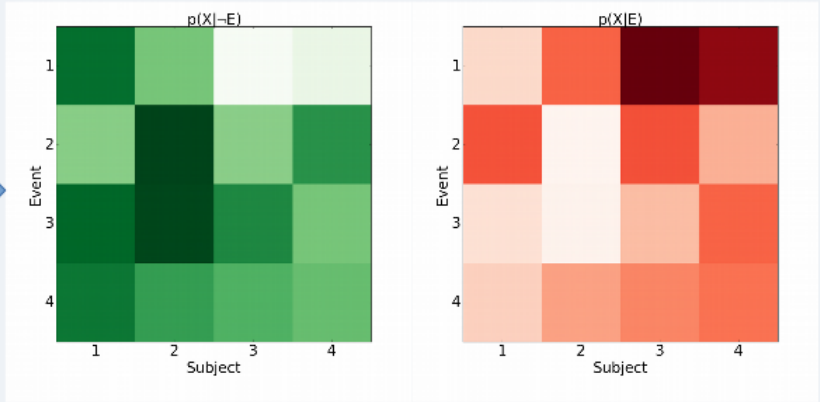
Back-up: Event-based model (EBM)

Event-based model

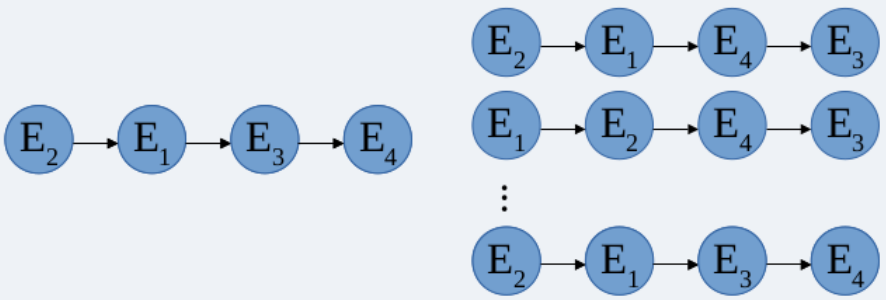
1. Fit mixture models to biomarkers

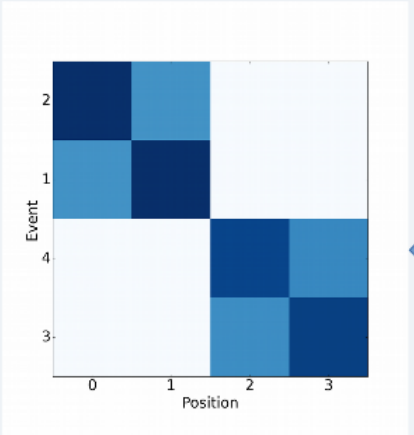


2. Calculate likelihoods of normality (event not occurred) and abnormality (event occurred)



3. Estimate most likely sequence by Markov Chain Monte Carlo sampling



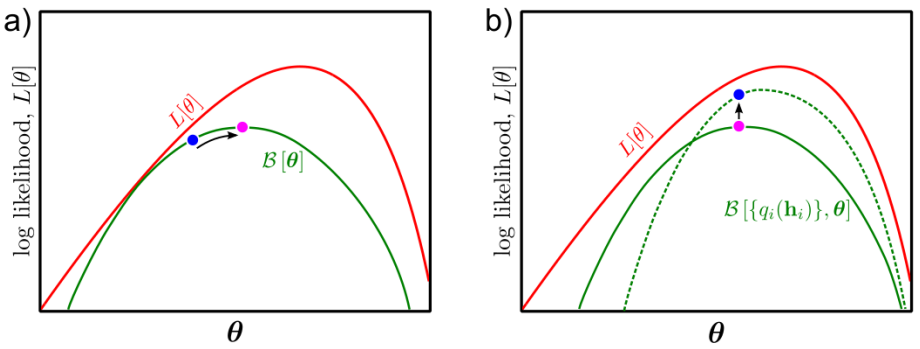


Positional variance diagram

Back-up: EBM parameter estimation

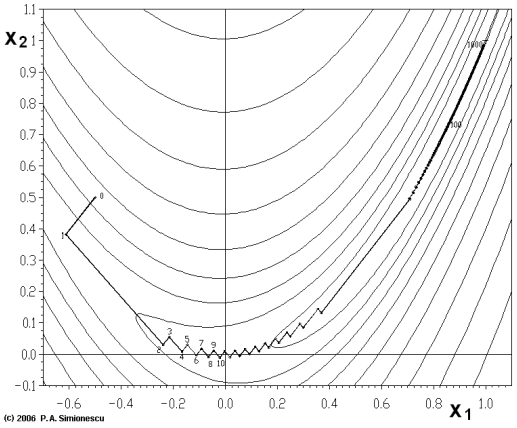
1. Mixture model fitting
– Expectation Maximisation

Prince, SJD. Cambridge University Press. 2012

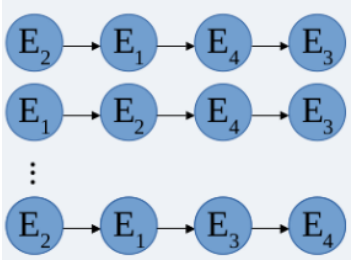


2. Latent variable (sequence) fitting
– Gradient Ascent

wikipedia.org/wiki/gradient_descent



3. Uncertainty estimation
– Markov Chain Monte Carlo



$$a = p(X \mid S')/p(X \mid S_t)$$