# Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference

Alexandra L Young et al.

---

**Presenters: Beiqin Zeng & Zhongkang Guo**

# CONTENT

# 01 | Background

# Background-Neurodegenerative disorders

**Incurable**

**Widespread**

**Heterogeneous**

Frontotemporal dementia **(FTD)**

Alzheimer's disease **(AD)**

1 **Phenotypic heterogeneity**

Individuals belong to a range of disease subtypes

2 **Temporal heterogeneity**

Individuals are at different stages of a dynamic process

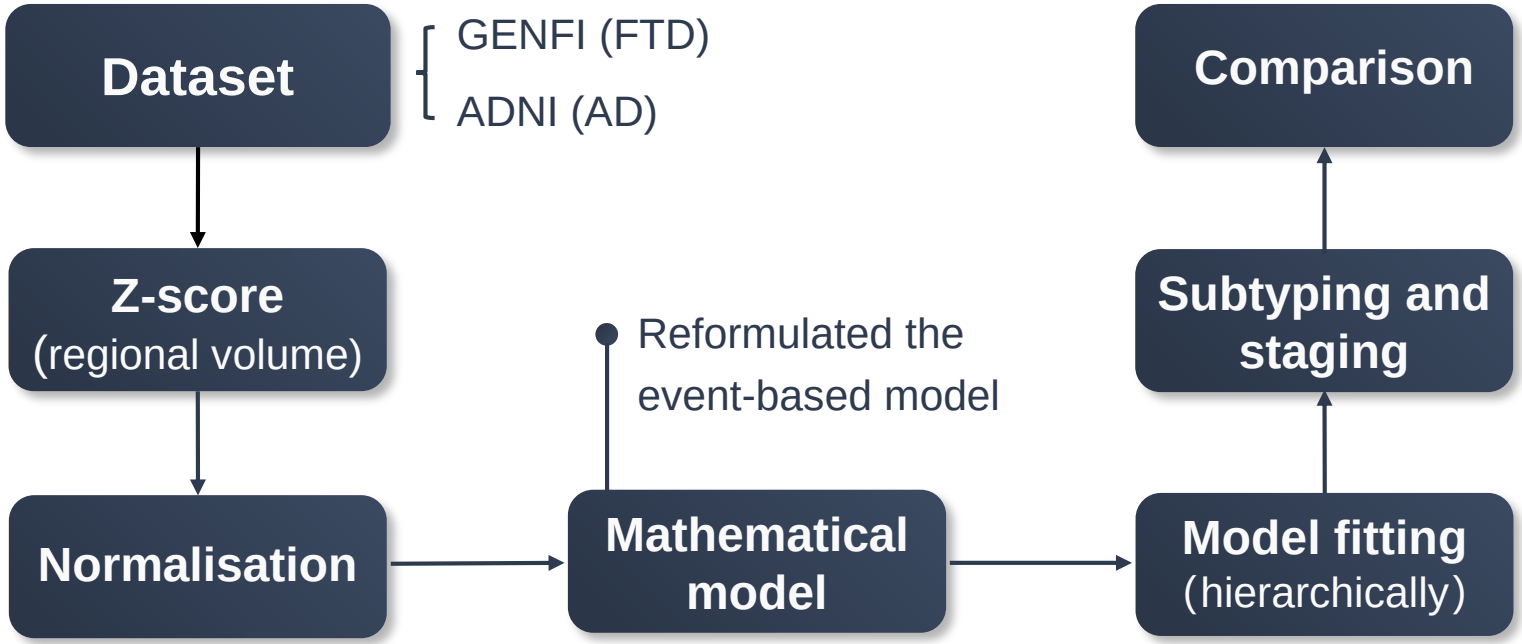
# Background-Previous studies

| | Temporal heterogeneity | Phenotypic heterogeneity |
|---|---|---|
| Stage-only models [1,2] | YES | NO |
| Subtype-only models [3] | NO | YES |

**Subtype and Stage Inference (SuStaIn)**

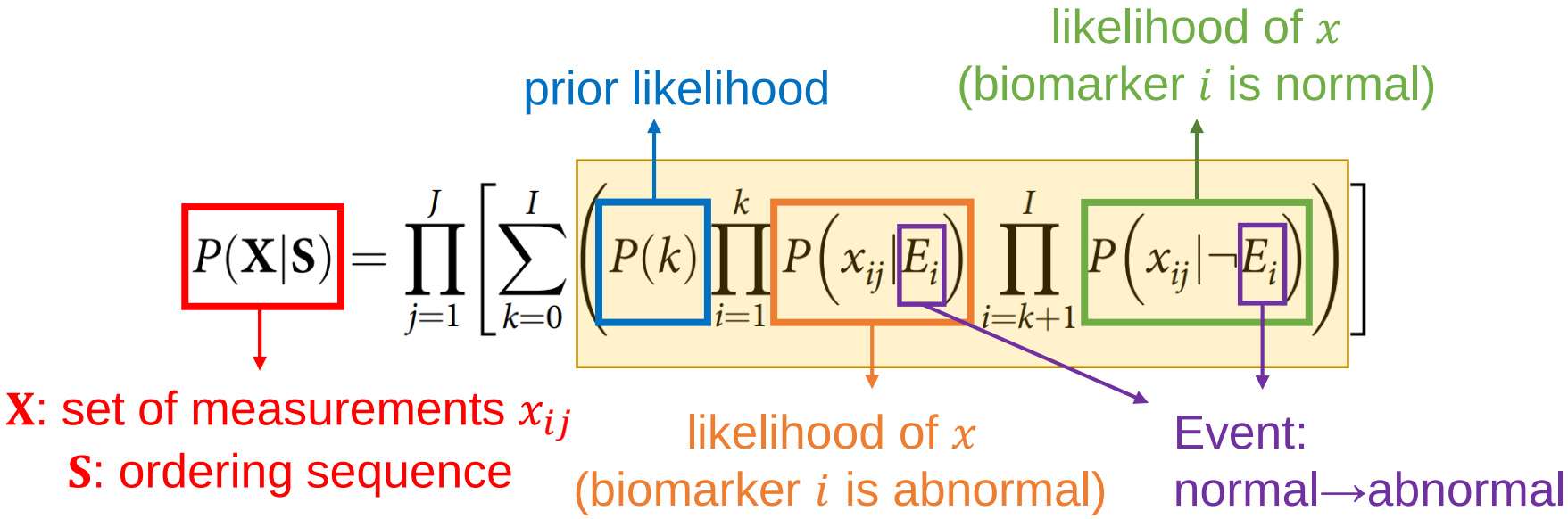
# 02 | Methods

# Methods-Flowchart



# Methods-Mathematical model

## Event-based model (instantaneous transition)



# Methods-Mathematical model

## Linear z-score model (continuous accumulation)

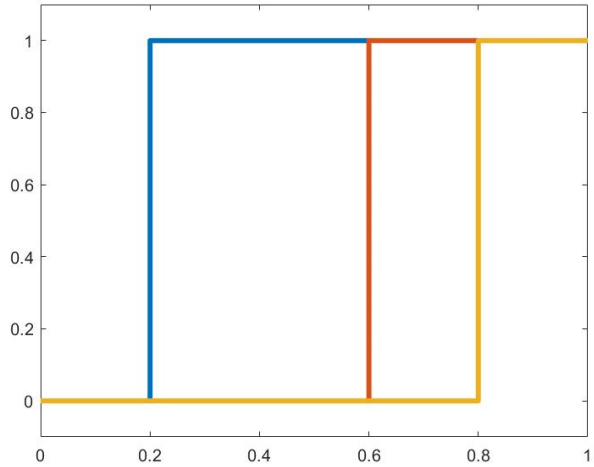
$$P(\mathbf{X}|\mathbf{S}) = \prod_{j=1}^{J}\left[\sum_{k=0}^{N}\left(\int_{t=\frac{k}{N+1}}^{t=\frac{k+1}{N+1}}\left(P(t)\prod_{i=1}^{I}P\left(x_{ij}|t\right)\right)\partial t\right)\right]$$

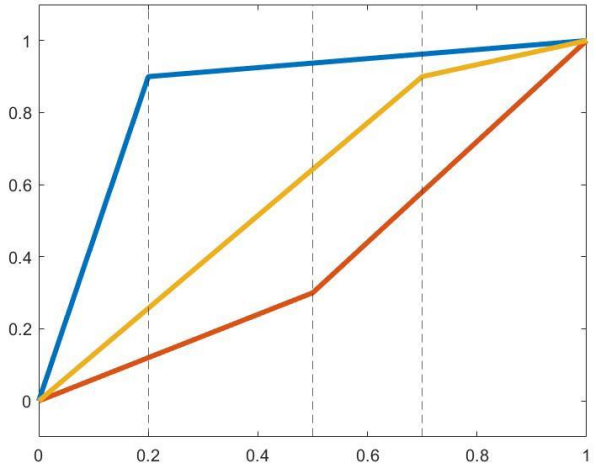
$$P\left(x_{ij}|t\right) = \mathrm{NormPDF}\left(x_{ij}, g_i(t), \sigma_i\right)$$

# Methods-Mathematical model

## Linear z-score model (continuous accumulation)

$$P\left(x_{ij}|t\right) = \text{NormPDF}\left(x_{ij}, g_i(t), \sigma_i\right)$$

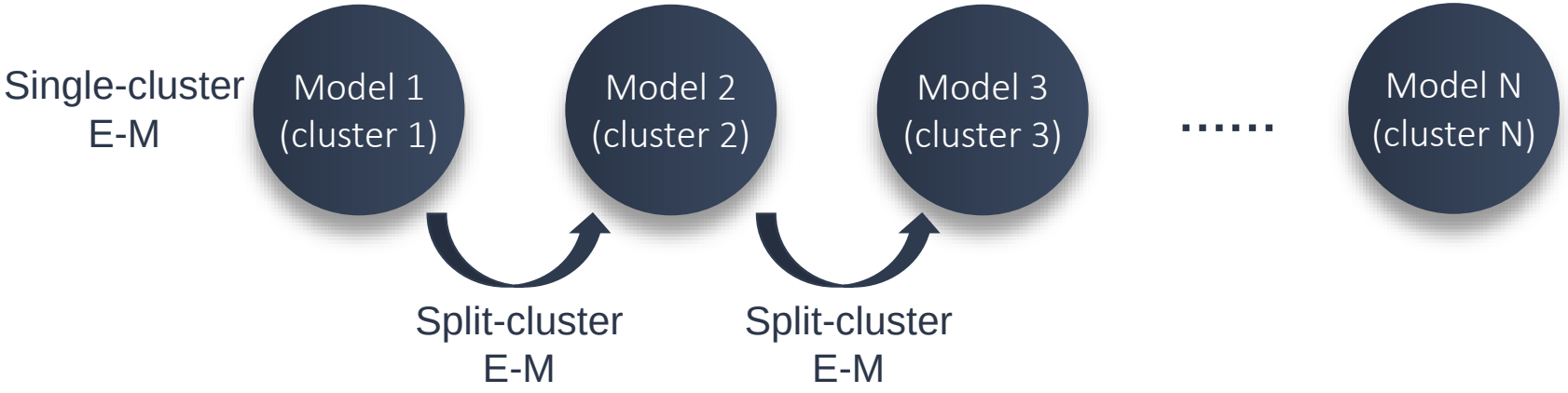


# Methods-Model fitting

~~Standard fitting~~: simultaneously optimize
- subtype membership
- subtype trajectory
- the posterior distribution

Fitting hierarchically

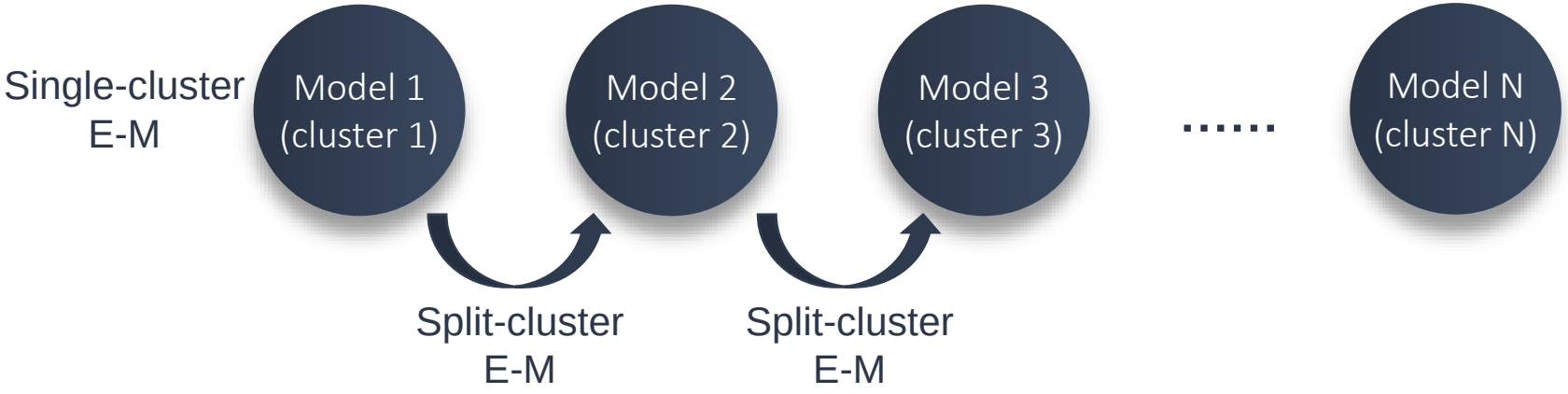


# Methods-Model fitting

**GREEDY?**

**good convergence**

(within a 0.001% of the maximum likelihood)



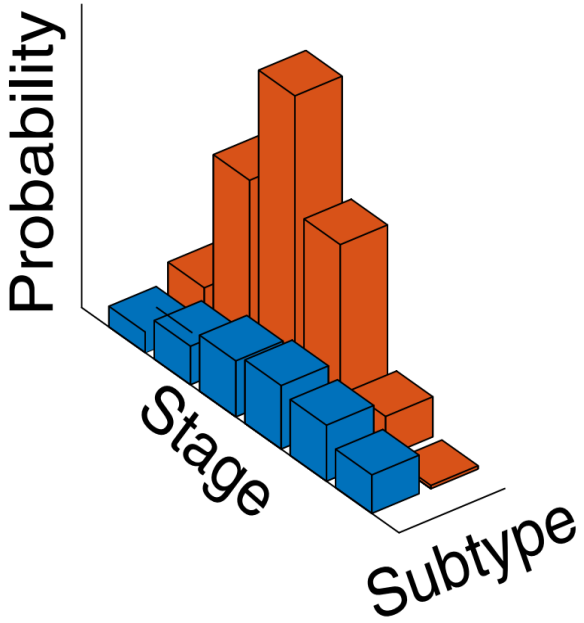
# Methods-Patient subtyping and staging

**Subtyping**

Evaluate the likelihood of each subtype (by integrating over disease stage) and choose the subtype with the highest likelihood.

**Staging**

Evaluate the likelihood of each stage of the most probable subtype and choose the stage with the highest likelihood.



# 03 | Results

# Results-An overview of the SuStaIn Modelling Technique

**Unsupervised learning**

- **Input**: snapshots of biomarker measurements (unknown subtype & stage)
- **Output**: the probability the subject belongs to each subtype and stage.

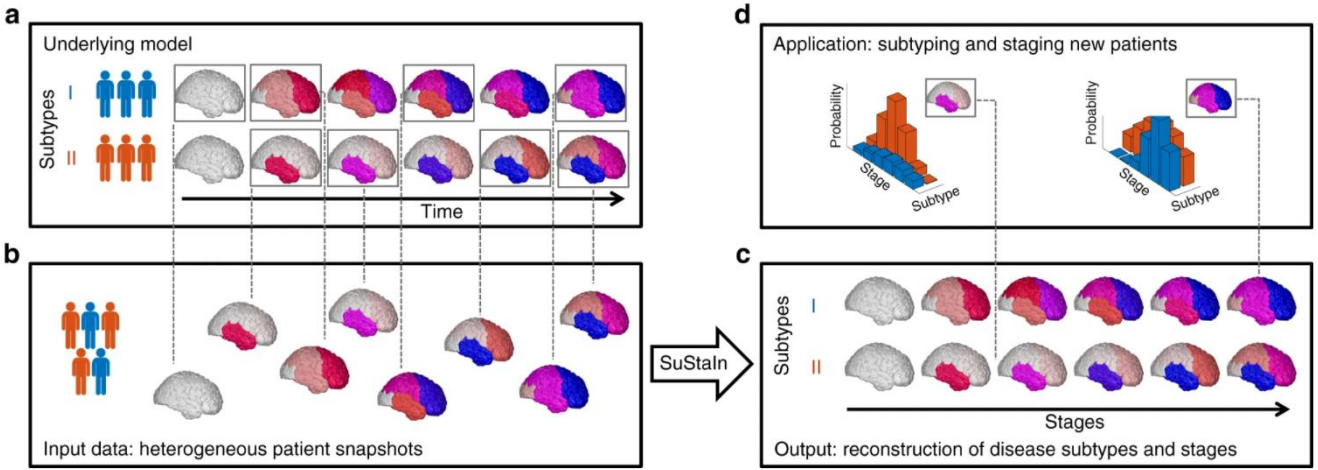



Fig 1 A conceptual overview of SuStaIn modelling technique

# Results–SuStaIn reveals within-genotype phenotypes in FTD

- Four subtypes in FTD
- Reproducibility: a high average similarity between cross-validation folds of >93%
- C9orf72 genotype has two distinct within-genotype phenotypes

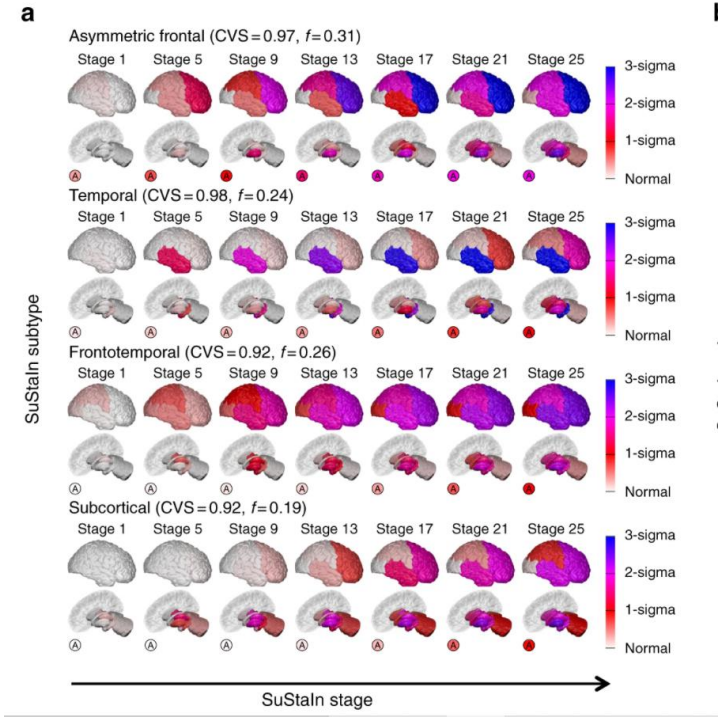



Fig 2 The progression pattern of four subtypes that SuStaIn identifies

# Results–SuStaIn reveals within-genotype phenotypes in FTD

- Four subtypes in FTD
- Reproducibility: a high average similarity between cross-validation folds of >93%
- C9orf72 genotype has two distinct within-genotype phenotypes

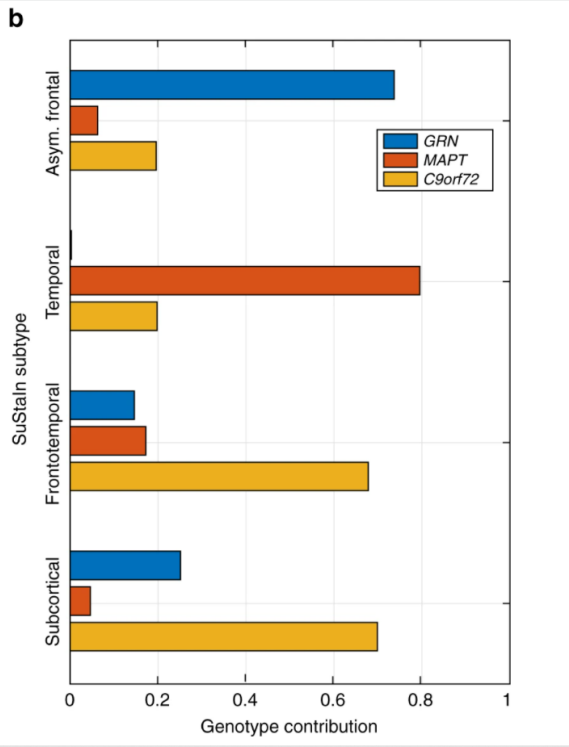



Fig 3 The contribution of each genotype to each of the SuStaIn subtypes

# Results–SuStaIn identifies three subtype progression patterns in AD

- Three subtypes in AD
- Reproducibility: an average similarity between cross-validation folds of >92%

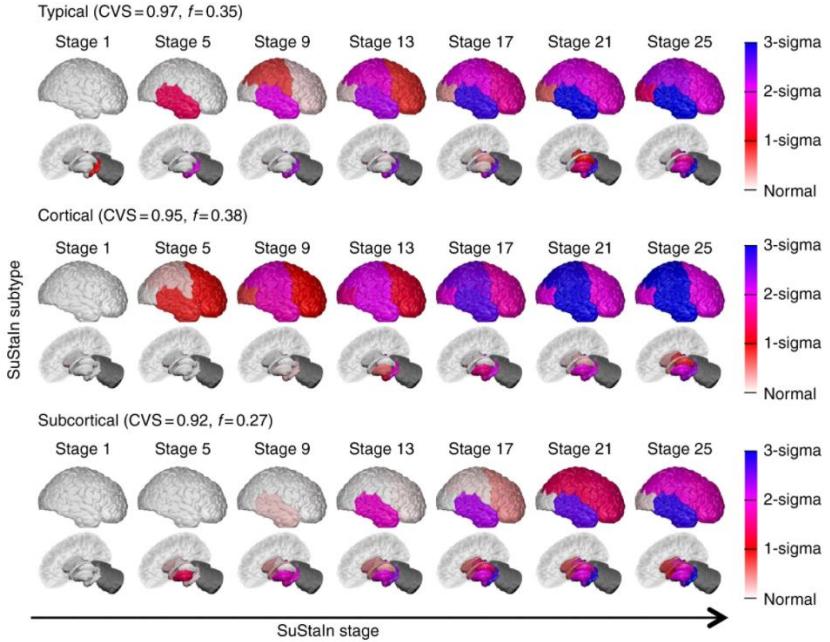



Fig 5 SuStaIn modelling of sporadic Alzheimer's disease using ADNI data

# Results–AD subtypes are reproducible in an independent data set

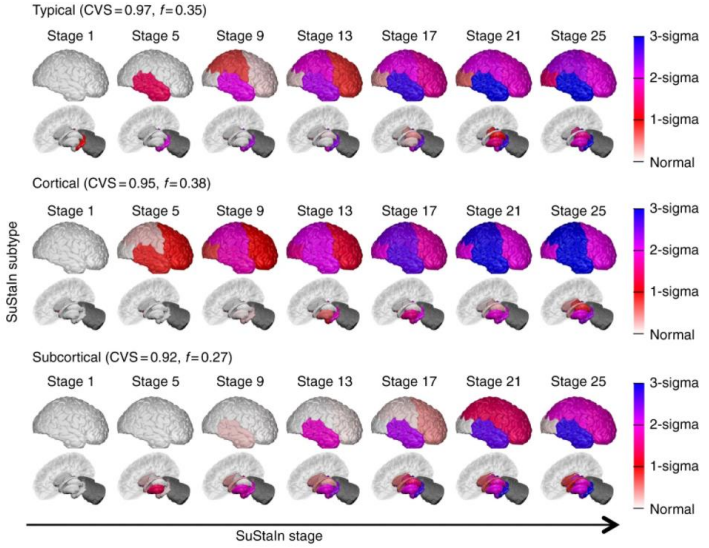



Fig 5 SuStaIn modelling using ADNI dataset

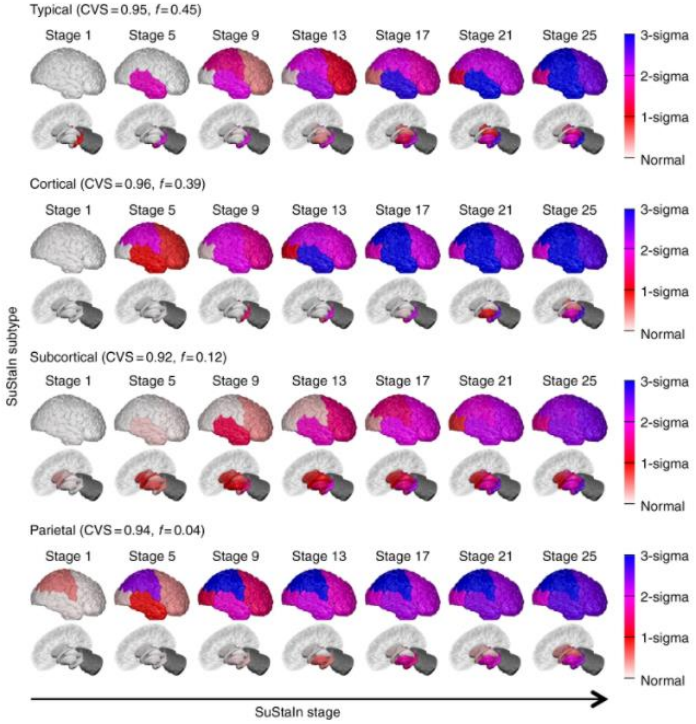



Fig 6 SuStaIn modelling using an independent dataset

# Results–Disease Subtyping and Staging

- These figures can assess how reliably SuStaIn assigns patients to subtypes

- **Figure 7(a):** the strength of assignment to the subtypes in **FTD** increases as the diseases progress (**88%** participants strongly assigned)

- **Figure 7(b):** the strength of assignment to the subtypes in **AD** increases as the diseases progress (**78%** participants strongly assigned)

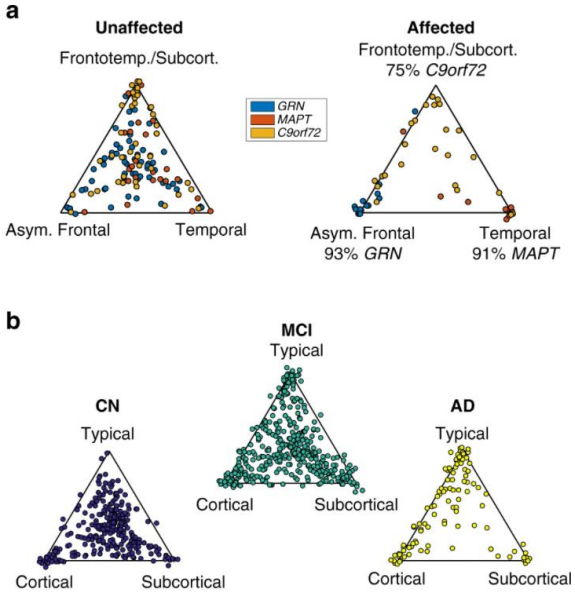



Fig 7 The assignability of the disease subtypes estimated by SuStaIn

# Results–Disease Subtyping and Staging

- **Figures 8** shows the reliability of the SuStaIn stages in each disease

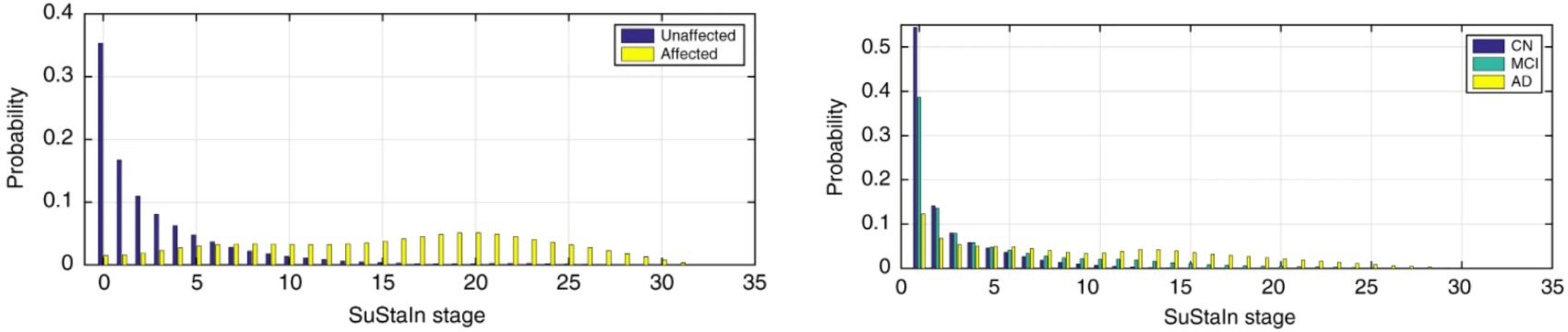



Fig 8 The probability subjects from each of the diagnostic groups belong to each of the SuStaIn stages

# Results–Disease Subtyping and Staging

- This experiment demonstrates the ability of SuStaIn to identify subtypes in a data set with a known ground truth.

**Table 1 Ability of subtypes to distinguish between different genotypes in symptomatic mutation carriers in GENFI using the SuStaIn subtypes in Fig. 2a**

| | *GRN* | *MAPT* | *C9orf72* |
|---|---|---|---|
| Asymmetric frontal (threshold $p > 0.65$) | 93% (13) | 9% (1) | 4% (1) |
| Temporal (threshold $p > 0.35$) | 0% (0) | 91% (10) | 21% (5) |
| Frontotemporal | 0% (0) | 0% (0) | 42% (10) |
| Subcortical | 7% (1) | 0% (0) | 33% (8) |
| Accuracy | 93% (13/14) | 91% (10/11) | 75% (18/24) |

Each entry is the percentage (number) of participants of a particular genotype assigned to that subtype. The final row indicates the percentage (fraction) of participants assigned to the correct subtype from each genotype. The results show that SuStaIn can accurately discriminate genotypes, validating the ability of SuStaIn to identify distinct phenotypes that align with known genetic groups

**Table 2 As Table 1, but for subtypes obtained from a subtypes-only model that accounts for heterogeneity in disease subtype but not disease stage, i.e. the subtypes in Fig. 6**

| | *GRN* | *MAPT* | *C9orf72* |
|---|---|---|---|
| Severe frontal (threshold $p > 0.99$) | 57% (8) | 9% (1) | 4% (1) |
| Severe temporal (threshold $p > 0.99$) | 0% (0) | 64% (7) | 8% (2) |
| Mild frontotemporal | 43% (6) | 27% (3) | 88% (21) |
| Accuracy | 57% (8/14) | 64% (7/11) | 88% (21/24) |

The results show that SuStaIn (Table 1) provides much better discrimination of the different genotypes than the subtypes-only model shown here, demonstrating the added utility of a model that accounts for heterogeneity in disease stage

# 04 | Discussion

# Discussion–Highlights

1. SuStaIn is the first tool to disentangle and characterise the temporal and phenotypic heterogeneity of neurodegenerative diseases.

2. SuStaIn further uncovers two distinct within-genotype phenotypes for carriers of a mutation in the C9orf72 gene, while finding the MAPT and GRN mutation groups are more homogeneous.

3. SuStaIn shows potential as a patient stratification tool in AD.

# Discussion–Limitations

1

## Limitations (from the paper)

1. Assumption: Biomarker variance is independent.
   Reality: Biomarkers tend to co-vary due to shared biological processes.
2. Assumption: The cohort is correctly diagnosed.
   Reality: the proportion of misdiagnosis in AD is non-negligible.

2

## Limitations (from our opinion)

1. Reproducibility in AD has no quantitative criteria.
2. The authors extract the z-score of brain volume from MRI images as features based on experience. Deep learning methods may help use more information from MRI images.

# THANK YOU