Submitted to the Annals of Applied Statistics

B-BIND: BIOPHYSICAL BAYESIAN INFERENCE FOR NEURODEGENERATIVE DYNAMICS

BY ANAMIKA AGRAWAL[1,6], VICTORIA M. RACHLEFF[2,4], KYLE J. TRAVAGLINI[2], SHUBHABRATA MUKHERJEE[5], PAUL K. CRANE[5], MICHAEL HAWRYLYCZ[2], C. DIRK KEENE[4], ED LEIN[2], GONZALO E. MENA[3,a] AND MARIANO I. GABITTO[2,7,b]

[1]*Center for Data-Driven Discovery for Biology, Allen Institute*

[2]*Human Cell Types Department, Allen Institute*

[3]*Department of Statistics & Data Science, Carnegie Mellon University*

[4]*Department of Laboratory Medicine and Pathology, University of Washington*

[5]*Department of Medicine, University of Washington*

[6]*Department of Physiology and Biophysics, University of Washington*

[7]*Department of Statistics, University of Washington,* [a]*gmena@andrew.cmu.edu;* [b]*mariano.gabitto@alleninstitute.org*

Throughout an organism's life, a multitude of complex and interdependent biological systems transition through biophysical processes that serve as indicators of the underlying biological states. Inferring these latent, unobserved states is a goal of modern biology and neuroscience. However, in many experimental setups, we can at best obtain discrete snapshots of the system at different times and for different individuals. This challenge is particularly relevant in the study of Alzheimer's Disease (AD) progression, where we observe the aggregation of pathology in brain donors, but the underlying disease state is unknown. This paper proposes a biophysically motivated Bayesian framework (B-BIND: Biophysical Bayesian Inference for Neurodegenerative Dynamics), where the disease state is modeled and continuously inferred from observed quantifications of multiple AD pathological proteins. Inspired by biophysical models, we describe pathological burden as an exponential process. The progression of AD is modeled by assigning a latent score, termed pseudotime, to each pathological state, creating a pseudotemporal order of donors based on their pathological burden. We study the theoretical properties of the model using linearization to reveal convergence and identifiability properties. We provide Markov chain Monte Carlo estimation algorithms, illustrating the effectiveness of our approach with multiple simulation studies across various data conditions. Applying this methodology to data from the Seattle Alzheimer's Disease Brain Cell Atlas, we infer the pseudotime ordering of donors. Finally, we analyze the information within each pathological feature to refine the model, focusing on the most informative pathologies. This framework lays the groundwork for continuous pseudotime modeling in the analysis of neurodegenerative diseases.

1. Introduction.

Modeling how cellular processes change in time due to endogenous or exogenous effects, in the context of discrete observation, is a fundamental problem in the biological sciences. Central to this problem is the longstanding goal of inferring a longitudinal trajectory delineating the dynamic evolution of each system. A wealth of research has cast this problem as learning a latent unobserved variable -termed *pseudotime*- dictating the progression of events where relevant measured quantitites are expressed as functions of this

---

Keywords and phrases: Hierarchical Bayesian Inference, Noisy Matrix factorization, Identifiable, Alzheimer's Disease, Trajectory Inference.

latent variable (Campbell and Yau, 2018, 2016; Hou et al., 2023; Street et al., 2018; Reid and Wernisch, 2016). Pseudotime inference was first motivated by the famous metaphor of Conrad Waddington in which undifferentiated cells progress along a continuum developmental trajectory to reach their committed cell fates (Waddington, 1940, 1957). Inference of such models is challenging for two reasons. Firstly, the lack of a universal definition of time renders the problem unidentifiable unless constraints are imposed. Secondly, and more critically, the successful inference of the parameters governing the system's evolution requires the inference of the pseudotime variable as well, adding an additional layer of complexity. These are particular challenges in modern biological applications, where measurements describing progression can only be collected once per individual, providing only a cross-sectional snapshot of the system's state. Aligning cross-sections along the pseudotime axis poses an additional inferential challenge, potentially impacting our ability to recover underlying dynamics in comparison with scenarios having multiple time measurements per individual or when pseudotimes are known beforehand.

Accurate temporal modeling is essential in the study of neurodegenerative diseases, in which we observe the aggregation of pathology in brain donors, but where the underlying disease state and its progression is unknown. Here, we focus on the study of neurodegeneration in Alzheimer's Disease (AD) and consider biophysically-inspired models for protein pathology accumulation and the pseudo-temporal dynamics of diseases progression. Pseudotime dynamics of disease progression and relevant parameters are inferred by using data from the recently released Seattle AD Brain Cell Atlas consortium (SEA-AD, https://portal.brain-map.org/explore/seattle-alzheimers-disease). Crucially, these cross-sectional measurements encompass a diverse set of pathophysiological markers that provide information on different stages of disease progression.

1.1. *Relation to prior work.* Pseudotime or ordering methods are widely applied due to their practical significance across various domains. While both share the goal of inferring data's inherent order, pseudotime goes a step further than simple ordination by attributing significant importance to the pseudotime latent variable, quantifying the system's progression along an inferred dynamic trajectory. Here, we offer a concise overview of related methods. In the analysis of cellular development using single-cell RNA-seq data, pseudotime is classically derived by computing a low-dimensional representation of each cell, inferring principal curves, followed by computing suitable minimum spanning trees (Magwene, Lizardi and Kim, 2003; Trapnell et al., 2014; Street et al., 2018). In contrast to these non-parametric methods, another line of methods frames the problem from a model-based, parametric perspective. For example, in the context of ecological niches and species modeling (Hui et al., 2015; O'Hara and van der Veen, 2024; Hui et al., 2023; Popovic, Hui and Warton, 2022; van der Veen et al., 2023; Roberts, 2020; Hoegh and Roberts, 2020), measurements of relative abundances of species are sorted along sites of greater abundance. There, generalized latent variable models are used to represent ordination as latent variables, and no true ordination occurs, but instead, multivariate latent variable ordination is used to represent underlying gradients that influenced species composition. In Political Science, the related problem of sorting latent preferences of agents (such as judges) given observed votes appears under the name of 'ideal point' models (Gelman and Hill, 2006) and this latent individual ordering is inferred by stating random-effects variables. In the field of psychometrics, models of changes in ability level over time utilize Item Response Theory (Kim and Camilli, 2014; Crane et al., 2006) to approximate non-linear growth.

Model-based approaches facilitate the simultaneous inference of the pseudotimes as well as any other relevant parameters. The method by Campbell and Yau (2018) models pseudotime inference in the context of single-cell omics as a matrix factorization problem where parameters are inferred through Bayesian methods. Then, it is possible to characterize both the timing of events as well as the progression of each feature (genes). While their setup is like ours, their work focuses on modeling, and it doesn't study in-depth the interplay of the different variables (such as the number of samples and dimension of cross-sections), the effects of pooling or shrinkage, and it omits rigorous consistency and performance guarantees as we do here.

Finally, a growing line of research applies pseudotime-based analysis to the context of aging (Pierson et al., 2019) and disease progression analysis, notably, to the study of cancer progression (Gupta and Bar-Joseph, 2008; Huang et al., 2023) and neural degeneration, including AD (Mukherjee et al., 2020). However, most studies fail to address a fundamental degeneracy present in latent variable inference problems, wherein pseudotime solutions can be scaled or affine transformed while still producing a valid solution—a mathematical indeterminacy inherent in many proposed frameworks. Our work attempts to resolve such degeneracy by imposing relevant mathematical constraints and is unique in that it addresses both the theoretical and applied aspects. We expect that this approach will help better inform the deployment of pseudotime inference methods for disease progression analysis.

1.2. *Organization.* This paper is organized as follows. In Section 2, we review the study of neurodegeneration and protein aggregation caused by AD pathology, and describe the data on which we based our work. In Section 3, we introduce a biophysically-inspired model to describe AD neuropathology aggregation. Biophysical considerations justify the use of generalized linear models with an exponential link function tying the unobserved timing of events to observations. We supplement this likelihood specification with priors that constrain the parameter space and represent complex dependency structures across measurements. We then obtain a family of Hierarchical Bayesian models that can be inferred using modern computational tools. In Section 4 we provide a mathematical analysis of a simplified linear model under which we can make precise statements regarding the identifiability, consistency, and convergence rates of our model as the number of individuals and measurements increase. Central to this analysis is the observation that the underlying inferential target can be understood at the most fundamental level as an instance of noisy matrix factorization. These theoretical results are supplemented with extensive simulations in Section 5, that show that in most aspects the theoretical results presented in the linear case also manifest in the generalized-linear setup. Simulations also show a few distinctive phenomena of generalized linear models that are not well captured by theoretical results on the linear case. In section 6, we apply our method to the SEA-AD dataset, and building on previous sections and specialized biological knowledge, we demonstrate that pseudotime and progression parameters can be reliably estimated. In Section 6.2, we extend our modeling framework to inform future experimental design. Finally, in Section 7, we discuss the significance and limitations of our work, and sketch future directions.

2. SEA-AD dataset. The present work is motivated by the need to understand the impact of pathological protein aggregation on cellular vulnerability across individuals spanning the complete spectrum of AD pathology. Quantitative neuropathological analysis was performed on the middle temporal gyrus (MTG) (Fig. 1A), a region of the cerebral cortex organized in a layered structure (from layers 1 to 6), of $d = 84$ postmortem brain donors included in the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) study (Gabitto et al., 2023). Donors from the SEA-AD cohort were drawn from two ongoing longitudinal studies: the Adult Changes in Thought (ACT) and the University of Washington Alzheimer's Disease Research Center (ADRC). We note that the SEA-AD cohort consists of donors with age distribution skewed towards advanced age (average age at death 88, SD=8) (Fig. 1B), which suggests a priori that high AD pathology values might be overrepresented in the cohort. The publicly available quantitative neuropathology dataset (accessible at sea-ad.org) comprises $M = 35$ total phenotypic measurements derived from 6 immunohistochemical stains targeting pathological proteins (hyperphosphorylated tau (pTau) and amyloid beta (A-β), along with additional markers selected for examining cellular changes in neurons, astrocytes, and microglia. Table 1 presents key markers used in this study and their corresponding extracted features. Immunostains for each protein and cellular marker are processed by a machine learning algorithm (https://indicalab.com/halo-ai/) that creates masks for each protein pathology, segments positive areas, and quantifies the burden of each pathology (Fig. 1A). Next, we provide a brief overview of general AD pathophysiology and elucidate how these pertinent variables contribute to understanding AD progression.

AD is the most common cause of age-related dementia and is characterized neuropathologically by the stereotyped spatiotemporal progression of two pathological proteins, intracellular hyperphosphorylated tau (pTau) in the form of neurofibrillary tangles (NFTs) and amyloid beta (Aβ) in the form of extracellular plaques. In AD, pTau and Aβ follow opposite progression patterns throughout the brain. Neuropathologists have mapped the progression of these pathological proteins and established discrete staging scales for each of them. These staging protocols order donors based on the presence or absence of pathological proteins in defined brain regions; Braak stages (I-VI) semi-quantitatively assess pTau NFTs (Braak and Braak, 1991), while Thal Phasing (1-5) semiquantitatively assesses Aβ plaques (Thal et al., 2002). A higher score on any scale indicates the presence of the relevant pathology across a greater number of brain regions. The AD neuropathologic change score (ADNC) (Hyman et al., 2012), attempts to integrate both scores into a general scale of AD progression. The two hallmark AD pathological proteins spatially overlap for the first time in the temporal lobe of the cerebral cortex, specifically the inferior and middle temporal gyrus (MTG). The MTG is thought to represent a key transition zone which delineates normal aging and preclinical AD from advanced stages of AD associated with widespread neocortical spread of pathology (Braak and Braak, 1991).

TABLE 1
Measurements used named by markers and the measured quantity

Feature name	Biological measurement	Type of marker
Percent 6e10 positive area	Amyloid-beta immunoreactive area	Pathological peptide
Number of 6e10 positive objects per area	Number of amyloid-beta plaques	Pathological peptide
Average 6e10 positive object area	Size of amyloid-beta plaques	Pathological peptide
Number of AT8 positive cells per area	Number of pTau positive cells	Pathological peptide
Percent AT8 positive area	pTau immunoreactive area	Pathological peptide
Average AT8 positive cell area	pTau immunoreactive area	Pathological peptide
Number of NeuN strong positive cells per area	NeuN immunoreactive area (mature neurons)	Cellular
Percent NeuN positive area	NeuN immunoreactive area (mature neurons)	Cellular
Percent of Iba1 and 6e10 positive co-localized objects	Iba1 immunoreactive area (microglia)	Cellular-pathological colocalization
Number of Iba1 and 6e10 positive co-localized objects per area	Iba1 immunoreactive area (microglia)	Cellular-pathological colocalization
Average Iba1 positive process area per cell	Iba1 immunoreactive area (microglia)	Cellular-pathological colocalization
Percent GFAP positive area	GFAP immunoreactive area (astrocytes)	Cellular marker

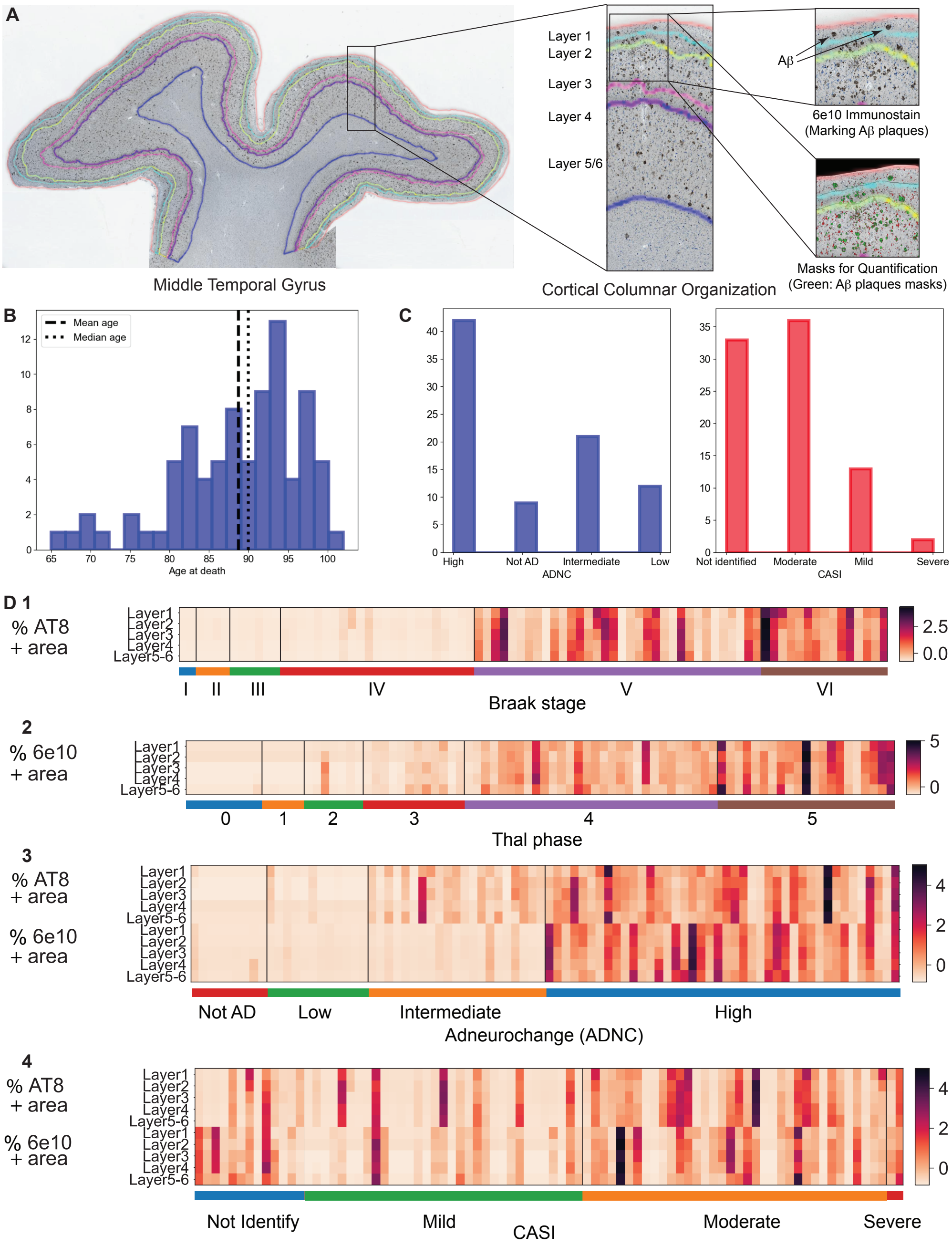


FIG 1. ***Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) data sets profiling neuropathology in AD brain donors.*** *(A) Left, image of the middle temporal gyrus of a representative brain donor in which pathology data was stained and imaged over* $L = 5$ *cortical layers of MTG (middle box). Right, amyloid-beta aggregates stained with 6e10 stain are segmented by using ML algorithms (green masks) and quantified. (B) Histogram depicting age distribution of donors at death, average age at death = 88, median age = 90. (C) Histogram depicting the distribution of ADNC and CASI scores for SEA-AD donors ordered on the x-axis. (D1-D4) Neuropathological measurements, percentage AT8 (measuring phosphorylated Tau pathology) and 6e10 (measuring amyloid* β *pathology) positive stain area, for each MTG cortical layer organizing donors by (1) BRAAK stages, (2) THAL phases, (3) ADNC scores, and (4) CASI scores.*

In addition to pathology measurements, the SEA-AD dataset encompasses a range of supplementary metadata pertaining to each brain donor, including clinical, cognitive, and demographic information. Established pathological assessments, previously described, such as Braak staging, Thal phasing, and ADNC are part of the repertoire of variables included, together with the last Cognitive Abilities Screening Instrument (CASI) (Teng et al., 1994a) test, administered during the donor's lifetime to evaluate cognitive function (Fig. 1C). We used these per-donor-metrics to provide initial descriptions of the data and performed data-entry quality controls. Illustrative quantitative neuropathology measurements, categorized by staging and cognitive criteria, are depicted in Fig. 1D1-D4, which depicts a notable increment in pathological burden as metrics increase in value. While the full dataset consists of $M = 35$ pathological markers, we use a subset of those in our model. We considered $M = 12$ markers (described in Table 1), removing noisy biomarkers that have low signal-to-noise ratio or that were sparse, i.e. data was available just for a few donors.

3. Model Description. We aim to elucidate the challenge of modeling the inherent variability in a biological process by using a latent variable that represents pseudotime. The latent variable is named pseudotime because it not only orders observations but also governs the progression of the biological process. While we specifically focus on AD pathology aggregation (see Section 2), this framework has broader applications, such as investigating aging-related phenotypes or studying different cell types in cellular development.

In modeling AD progression, we assume donors traverse a single disease trajectory spanning the entire disease spectrum, from no disease to the end stage, yet this trajectory remains unknown. Donors are randomly selected across the disease spectrum, driven in practice by donor availability, and we lack prior knowledge of the sampling distribution, $p(t)$. For each of the donors, we collect multiple measurements reflecting the state along the disease trajectory, including varying levels of pathological protein aggregates. Lastly, observations are corrupted by observational noise (Fig. 2). Notable previous work in aging studies has observed that to ensure identifiability of the link function between the latent pseudotime and the observational process (represented by $f(t)$ in Fig. 2) should be monotonic and injective (Pierson et al., 2019). However, in their work, time was a known variable, compared to our case in which nothing is known about latent pseudotime.

In the next sections we construct a model for cross-sectional SEA-AD donor data, $X^d_{m,l}$, comprising quantitative neuropathological measurements from $d = 1 \dots D = 84$ donors, encompassing $m = 1 \dots M = 12$ distinct protein measurements taken across $l = 1 \dots L = 5$ cortical layers in MTG. These measurements reflect the accumulation of various pathological proteins within each donor. We define a donor-specific one-dimensional latent variable, referred to as $t_d \in [0, 1]$. This variable represents the pseudo-temporal score that organizes donors according to their varying disease burdens, from low ($t_d = 1$) to high ($t_d = 1$). Inspired by biophysical principles of protein aggregation, and the cited aging-related studies (Pierson et al., 2019), we will find that a monotonic and injective exponential link function best accommodates our data. We will infer all model parameters through Bayesian inference.

3.1. *Deterministic Biophysical model of Pathology Aggregation.* Alzheimer's Disease is characterized by the proliferation of misfolded protein aggregates. As mentioned before, amyloid-Beta (Aβ) aggregates and Tau neurofibrillary tangles (τ-NFT) are two examples of well-studied protein aggregates that have been widely used as pathological markers to study the spread and severity of AD (Braak and Braak, 1991; Thal et al., 2002). Several mechanistic models, chiefly rooted in polymer kinetics, have been used to study the aggregation and seeding dynamics of protein aggregates (Vaquer-Alicea and Diamond, 2019; Davis and Sindi, 2016). In our work, we will leverage the Nucleated Polymerization Model (NPM), which is

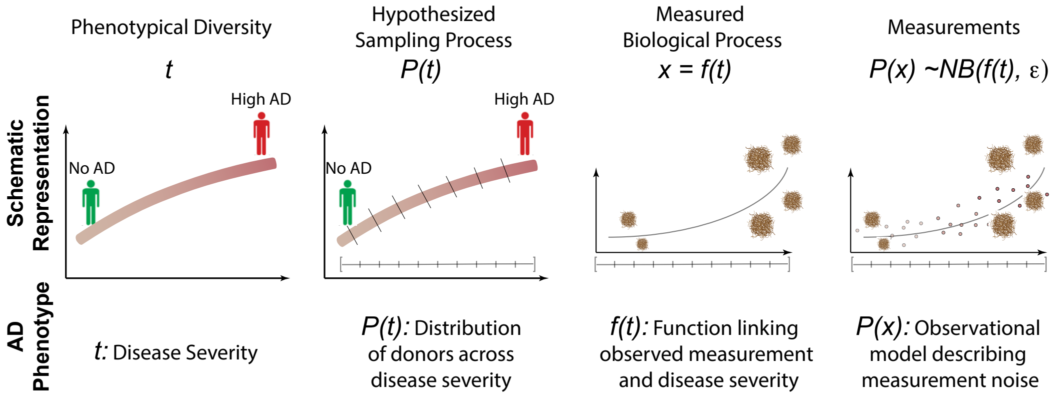

Phenotypical Diversity
t
Hypothesized Sampling Process
P(t)
Measured Biological Process
x = f(t)
Measurements
P(x) ~NB(f(t), ε)
Schematic Representation
No AD
High AD
No AD
High AD
AD Phenotype
t: Disease Severity
P(t): Distribution of donors across disease severity
f(t): Function linking observed measurement and disease severity
P(x): Observational model describing measurement noise

FIG 2. ***Schematic representing the modeling of a general biological process by a latent variable, t, termed pseudotime.*** *In this work, pseudotime represents a scale that describes the trajectory of Alzheimer's Disease along the entire disease spectrum. Donors are sampled along AD trajectory from distribution $p(t)$. Then, multiple observed biophysical processes x are measured from each donor. These measurements can be linked to the underlying pseudotime by a mean function, $f(t)$, and when observed, are corrupted by noise (described by the distribution $p(x|t)$ of observational noise.*

the most widely accepted model used to describe protein aggregate propagation dynamics (Masel, Jansen and Nowak, 1999), to describe the temporal progression of AD-associated pathology markers.

The NPM model describes a cycle of aggregate growth followed by multiplication, leading to the dominance and growth of toxic protein aggregates. The growth of such aggregates occurs through attachment with surrounding 'healthy' monomeric units, which is then followed by a fragmentation process that multiplies the number of 'seeds' capable of growing into many more misfolded aggregates (Fig. 3A). Once the monomers are incorporated, they become a part of the "unhealthy" or misfolded population. The misfolded aggregates also tend to be more stable than their healthy counterparts, such as in the case of prionic proteins (Tompa et al., 2001). In an aggregation model for a spatially limited domain (e.g. a single anatomical region of the brain), the important dynamical parameters to consider are the initial seeding levels, denoted by β_0, and the rate of multiplication through aggregation and division, denoted by β_1 ((Fig. 3B). In the 'early' time regime, i.e. when the available pool of healthy protein monomers is not limiting, we will describe the evolution of a pathological protein $X_{m,l}$ for measurement m in layer l with the following dynamics arising from the NPM biophysical model (Masel, Jansen and Nowak, 1999):

$$X_{m,l}(t) = e^{\beta_0^{m,l} + \beta_1^{m,l} t} \tag{1}$$

where t denotes the latent pseudotemporal variable describing the progression of AD severity. The details of the kinetic model of pathology propagation, and the simplifying assumptions used to arrive at such functional form, are included in Supplementary Material, Appendix Section A.

3.2. *Modeling the Generative Distribution of Observed Pathology.* The above biophysical argument suggests to consider a trajectory that is an exponential transformation of an affine function of the latent pseudotime variable. To account for observation noise, we couple this latent trajectory with a sampling model where observed counts $X_{m,l}^d \in \mathbf{N}_0$ describing pathological proteins follow a negative binomial distribution. Specifically, we consider a

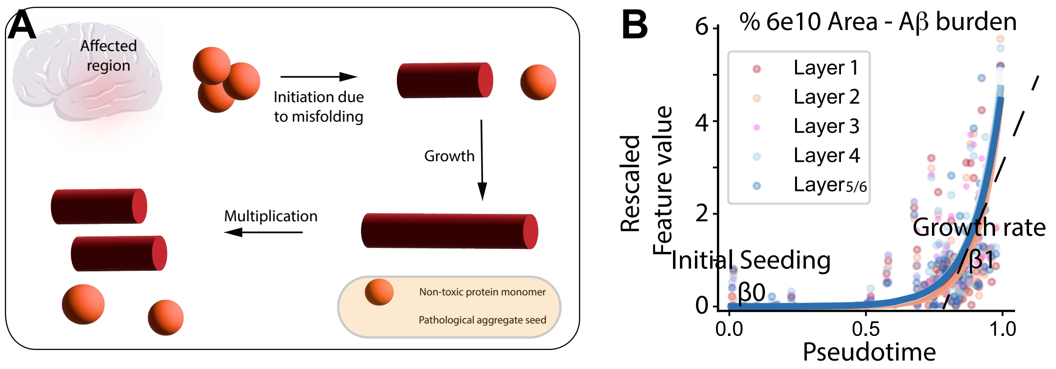


FIG 3. ***We assume that AD protein pathology aggregation can be described by an exponential process.*** *A) Schematic representing protein aggregation process. Due to protein misfolding, an initial seeding event is triggered which later leads to aggregate growth. Randomly, protein aggregates break down multiplying in number. The conjunction of growth and multiplication events can be described by exponential dynamics. B) Scatterplot depicting measurement for $A\beta$ burden, the percentage area stained positive by 6e10 antibody, in each cortical layer ordered according to our model. For each protein pathology, the model fits an initial seeding variable, β_0, and a growth aggregation rate β_1.*

count parameter $A_{m,l} \in \mathbf{N}^+$, and a probability of success in the (negative) logit scale $\rho \in \mathbf{R}$ such that the likelihood of $X^d_{m,l}$ given $A_{m,l}$ and $\rho = \beta_0^{m,l} + \beta_1^{m,l} t_d$ is given as

$$p(X^d_{m,l} = x | A_{m,l}, \beta_1^{m,l}, \beta_0^{m,l}, t_d) = \binom{x + A_{m,l} - 1}{x} \left(\frac{1}{1+e^{-\rho}}\right)^{x} \left(\frac{1}{1+e^{\rho}}\right)^{A_{m,l}}. \tag{2}$$

The mean and variance of the above distribution are given by $\mu = A_{m,l}e^{\rho}$, $\sigma^2 = \mu + \frac{\mu^2}{A_{m,l}}$. This implies that the conditional mean of $X^d_{m,l}$ given $\beta_0^{m,l}, \beta_1^{m,l}$ and t_d is exactly the quantity $X_{m,l}(t)$ defined in Eq. (1), up to the $A_{m,l}$ factor that can be absorbed into $\beta_0^{m,l}$. Consequently, the negative binomial model can be understood as providing a noisy version of the underlying dynamics. As customary, we understand A as an overdispersion parameter A (Gelman, Hill and Vehtari, 2021) modulating different levels of observation noise whereby in the limit $A \to \infty$ we recover the usual Poisson regression. Eq. (2) is one instance of a generalized linear model with a logarithmic link function. Still, other choices could be made if observations were of a different nature. For example, logistic regression could be used if observations were binary.

3.2.1. *Priors.* In addition to the likelihood in Eq. (2), we must specify priors for the model parameters $\Theta = \{t, \beta_0, \beta_1, A\}$. First, we specify a sequence of independent donor-specific pseudotimes following a Beta distribution with parameters t^a, t^b:

$$t_d \sim Beta(t^a, t^b). \tag{3}$$

In particular, $t_d \in [0,1]$. The decision to choose such a bounded interval is two-fold. At the conceptual level, biological systems are constrained in space and time. This implies in our application to AD that protein propagation and disease trajectory do not last indefinitely. When looking at a particular brain region, the accumulation of pathological proteins will range from 0 to a high number, with the last case implying that the pathological protein has covered the entire region. At the mathematical level, as we will later show in Section 4, choosing a bounded interval enables identifiability of the entire system, a desired property in our setup where our latent variable represents not only ordination but also pseudo time. The case $t^a = t^b = 1$ corresponds to a uniform prior. Although this prior may seem unbiased, it encodes strong beliefs about the phenotypical diversity of the sample; for example, that the average severity level is 0.5. This justifies the use of a more general $Beta$ prior to encode information on the sampling process. We will see in Section 4 that we can benefit from suitable choices of this prior.

We must also specify priors on $\beta_0^{m,l}, \beta_1^{m,l}$, parameters that represent the baseline level of pathology, and the exponential rate of pathological growth of each of the markers. We will consider two models, one with measurement and layer-independent priors on $\beta_1^{m,l}$ and the second with hierarchical priors pooling information across layers. In the former case, we are assuming that only the protein aggregation rate (through its parameter β_1) can be related across layers, and for each measurement, we will infer a covariance matrix across layers. Finally, we consider priors for the over-dispersion parameters $A^{m,l}$ since they are not revealed to us. In this case, we impose half-normal distributions.

In summary, the plate diagrams of our hierarchical and fully factorized models are depicted in Fig. 4, and we detail next the entire generative process of our model:

1. For each donor, d
    a) Choose $t_d \sim$ Beta(1, 1)
2. For each measurement, m
    a) **If** Hierarchical
        Then $\beta_1^m \sim$ Normal(0, 1)
            $L_\Sigma^m \sim$ LKJ(M, 1)
            $d_\sigma^m \sim \mathcal{H}$alfCauchy(1)
            $\Sigma^m = \sqrt{d_\sigma^m} L_\Sigma^m$
        Else Pass
    b) For each layer, l
        i. Choose $A^{m,l} \sim \mathcal{H}$alfNormal(0, 1)
        ii. Choose $\beta_0^{m,l} \sim$ Normal(0, 1)
        iii. **If** Hierarchical
            Then Choose $\beta_1^{m,l} \sim$ Normal(β_1^m, Σ_m)
            Else Choose $\beta_1^{m,l} \sim$ Normal(0, 1)

3.3. *Posterior Bayesian Inference through HMC sampling.*

To perform posterior inference from our models we used numpyro version 0.13.0 (Bingham et al. (2019); Phan, Pradhan and Jankowiak (2019)), a Python based probabilistic programming environment. Similar to the more traditional probabilistic language, Stan (Carpenter et al. (2017)), numpyro is equipped with effective posterior sampling algorithms such as Hamiltonian Monte Carlo (HMC, Duane et al. (1987); Neal et al. (2011)). In particular, we utilized the implementation of the No-U-Turn Sampler (NUTS) (Hoffman et al. (2014)) kernel to obtain posterior samples, and monitored convergence using R-hat and effective sample size metrics (Gelman et al. (2021)).

To infer the posterior parameters of our models (Fig. 4) using the SEA-AD dataset, we sampled from the posterior with 5 HMC chains run sequentially for 5000 warm-up iterations, and then 1000 samples were collected in each chain. We randomly rotate the seed for the pseudo-random number generator in each chain. To ensure convergence, we discarded the chains that had an R-hat below 1.05 or had an effective sample size below 50. The results from these fits are described in Section 6. For experiments with synthetic data, we used different parameters to the ones described above (see Supplement Section on Proofs for details).

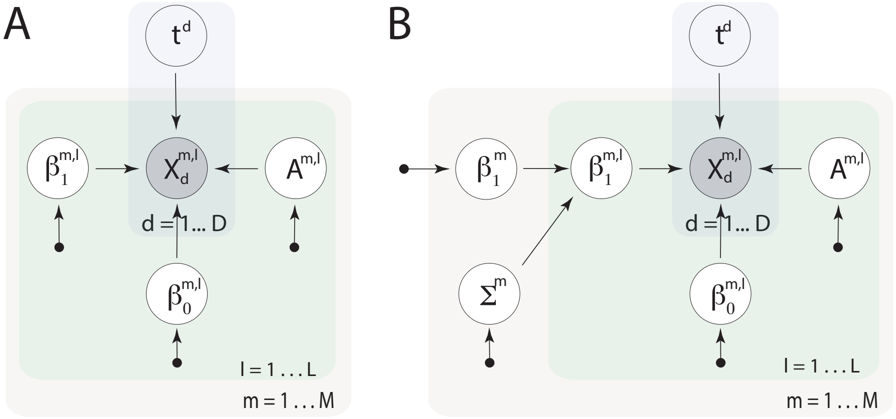


FIG 4. ***Plate diagram depicting a model in which parameters are inferred independently in each measurement and each layer (A), and model with hierarchical priors pooling information across layers in each measurement (B).*** *In both models, pathological measurements are depicted by* $X_d^{m,l}$*, where* $m = 1 \dots M$ *index each of the measurements and* $l = 1 \dots L$ *each layer. The latent pseudotime variable is denoted by* t_d*, where* $d = 1 \dots D$ *index the brain donor. Measurement and layer-specific initial-seeding, growth-rate, and total-counts parameters are denoted by* $\beta_0^{m,l}$*,* $\beta_1^{m,l}$*, and* $A^{m,l}$*. In the case of the hierarchical model in (B), we pool information across layers and infer covariance matrices* Σ_m*, assuming that each layer has its own initial seeding events but their growth rates are similar.*

4. Model analysis. Our problem departs from the classical regression setup in that we not only estimate the regression parameters β but also must infer the latent parameter t. Here, we delve into the inferential implications of the joint estimation of t and β; we study the fundamental statistical limitations to our problem and argue that our modeling choices are well-justified in that they enable efficient use of all available sources of information to counter these limitations.

Results in this section are statements regarding conditions for identifiability and the best possible convergence rates for the mean squared error (MSE) of estimators of β and t. Although we keep our Bayesian estimators in mind, this analysis is rather concerned with the properties of the problem itself than with our particular method. We will later show in Section 5 that our Bayesian estimators (understood as posterior means) exhibit the behavior of the best possible estimators.

Instead of the generalized linear model in Eq. (2), here we consider the simpler linear model where a rich theoretical understanding of statistical limits is readily available:

$$X_d^n = \beta_0^n + \beta_1^n t_d + \epsilon_d^n. \tag{4}$$

We understand X as a $D \times N$ matrix by vectorizing layer and feature information and treating them as a unique column 'feature' indexed by $n = 1, \dots, N$ where $N = L \times M$. Here, ϵ_d^n are i.i.d. Gaussian centered variables with variance ϕ^2. These noise variances play an analog role to the overdispersion parameter A in Eq. (2).

The key observation is that Eq. (4) corresponds to a rank-two noisy factorization problem, so that at the most elementary level, the estimation limits of β and t are the limits of this low-rank factorization. Although the theory is general enough to enable the analysis of (4), for simplicity we will neglect the β_0^n term and study the rank-one noise matrix factorization

$$X = \beta t^\top + \epsilon. \tag{5}$$

To apply existing statistical results, we divide the estimation of t and β into the estimation of the normalized parameters $\beta' = \beta / ||\beta||_2$, $t' = t/||t||_2$ and the individual signal strengths

$||t||_2^2$ and $||\beta||^2$. To state consistency results, we conceive limiting experiments consisting of sequences of true $\{t\}_d$, $\{\beta\}_n$ (for example, obtained by sampling from their population distributions), and of matrices X and ϵ that are revealed as either D or N (or both) get larger. For our asymptotic statements on D or N we must assume that $\bar{t}^2_\infty := \lim_D \bar{t}^2_D$ and $\bar{\beta}^2_\infty := \lim_N \overline{\beta}^2_N$ exist, where $\bar{t}^2_D := ||t||_2^2/D$ and $\bar{\beta}^2_N := ||\beta||_2^2/N$,respectively.

4.1. *Subspace estimation.* We first study the ability to estimate the normalized parameters t' and β'. We use the standard Frobenius metric for comparing subspaces U_1 and U_2(see Appendix for details), which in the one-dimensional case reduces to

$$D_F(U_1, U_2) := \min\{\big\|U_1 - U_2\|_2, \big\|U_1 + U_2\|_2\}.$$

To obtain estimation bounds, we turn to specialized, best-known risk analysis for subspace estimation Cai and Zhang (2018) based on variants of the well-known Wedin's $\sin\Theta$ perturbation theorem Wedin (1972). Let $\hat{\beta}$ and $\hat{t}$ be the first left and right singular vectors of X (in particular, they have unit norm).

PROPOSITION 1. *As a consequence of Theorems 3 and 4 in Cai and Zhang (2018), we have*

$$\frac{1}{N} E D_F(\hat{\beta}, \beta')^2 \lesssim \frac{\phi^2}{ND} \frac{\left(\overline{\beta}^2_N \bar{t}^2_D + \frac{\phi^2}{N}\right)}{\left(\overline{\beta}^2_N \bar{t}^2_D\right)^2} \wedge 1$$

$$\frac{1}{D} E D_F(\hat{t}, t')^2 \lesssim \frac{\phi^2}{ND} \frac{\left(\overline{\beta}^2_N \bar{t}^2_D + \frac{\phi^2}{D}\right)}{\left(\overline{\beta}^2_N \bar{t}^2_D\right)^2} \wedge 1$$

These estimators are rate-optimal in the sense that the risk of any other estimators of β' and t' have risk lower bounded by the right sides above (up to constants).

We can use Proposition 1 to understand how estimation performance for β' and t' improve as we increase either D or N (or both). We observe that for fixed N, the average MSE in the estimation of β' decreases at the rate D if $\bar{t}^2_D$ is assumed to remain bounded away from zero (we are excluding sparse non-informative sequences of donors), which is the same rate that would be achieved if D were known. Therefore, adding more donors is helpful in the estimation of β'. Likewise, if D is fixed, then the average MSE over t' decreases at the rate $1/N$ if $\overline{\beta}^2_N$ (i.e. we are not adding too many non-informative features), implying that adding more features help. Also, larger magnitudes $\overline{\beta}^2_N$ are beneficial for the estimation of t' and larger magnitudes of $\bar{t}^2_D$ are beneficial for the estimation of β'.

4.2. *Signal strength estimation.* Proposition 1 suggests that the spaces spanned by t' and β' can be estimated as easily as in the usual regression case (t known). We now turn to the problem of estimating the signal strengths $||\beta||$ and $||t||$. The ability to estimate these parameters consistently is critical since otherwise, we are at risk of inflating the significance of our estimators of β: for example, if $||\beta||$ is grossly overestimated, we may come to falsely detect many β_n as significant features explaining the severity of the pathology.

Estimating the signal strengths is more delicate since identifiability issues might be at play. The main result is that by imposing the constraint that severity levels belong to $[0, 1]$ (or any other compact interval defined beforehand), we can identify $||\beta||$ and $||t||$ if both N and D simultaneously grow to infinity.

PROPOSITION 2. *Let* t_d *be sampled from a distribution* $\mathbf{t}$ *supported on* $[0,1]$. *Suppose that both* t_∞ *and* β_∞ *exist. Then, there are estimators of* $\overline{t}_D^2$ *and* $\overline{\beta}_N^2$ *converging almost surely to* t_∞^2 *and* β_∞^2 *as* N *and* D *grow to infinity.*

The mathematical details are presented in the Appendix. At a high level, the proof consists of first showing that we can first identify the product $\beta_\infty \times t_\infty$. Then, we can use the information on the support of $\mathbf{t}$ to re-scale the principal vector $\hat{t}$ so that it will lie in $[0,1]$. This re-scaling enables the identifiability of each of the terms β_∞ and t_∞. The statement in Proposition (2) holds whether or not ϕ^2 is known: in the latter case, we can estimate it incurring an asymptotically negligible bias.

Proposition 2 gives the optimistic result that the entire system can be identified, but the main downside is that we are forced to consider simultaneous asymptotics in N and D. Unfortunately, as we will show, this heavy requirement cannot be relaxed, and the estimation of $||\beta||$ will be biased. In other words, the estimation of $||\beta||$ will be biased even if a cohort of infinite donors were available. We will comment on strategies to overcome this limitation. Even if these biases persist, the overall message of Proposition 2 is that if multiple informative features β are available, they should all be used to inform inferences.

5. Simulation studies.

The previous results provide us with tools to reason about inferential capabilities in our model. In this section, we complement these results with a comprehensive set of simulations. These simulations not only give an empirical validation to the theory in Section 4; more importantly, we use them to understand our procedure's behavior in cases not covered by the theory.

This gap arises in three aspects: First, results in Section 4 are concerned with the estimation problem and not with our particular estimator. Second, they concern a simplified model that doesn't incorporate the effect of a nonlinear link function and a non-Gaussian likelihood, and where coefficients β are not assumed to possess a hierarchical structure that we could possibly benefit from. From the experiments in this section, we can make the following claims

- Our Bayesian estimator exhibits the behavior anticipated by Propositions 1 and 2 in that simplified setup (Sections 5.1, 5.2)
- The use of suitable priors on t can be helpful to overcome the identifiability issues described in Section 4 (Section 5.1)
- Beyond the simplified linear model, the qualitative behavior of estimators still matches the phenomena described in the previous section, although some additional new phenomena emerge:
  - In non-linear/non-Gaussian models, hierarchical models for β are beneficial if such structure is present in the data. We don't observe a benefit from using hierarchical models in linear Gaussian models (Section 5.3)
  - Non-linearities in the link function imply that different features can be differentially informative at different levels of severity. This implies the need for careful choice of features and donor experimental design (Section 5.4)

Here, we only describe the main features of each experiment but defer a more comprehensive explanation of each setup and parameter and sampler values to the Appendix.

5.1. *Elementary simulations in the linear model.*

We start by studying the empirical performance of our Bayesian estimators in the setup of the linear model in Eq. (4). Unlike the full hierarchical model of Section 3, we consider a flat prior over β. We study different choices of true pseudotime distribution $\mathbf{t}$ and pseudotime priors $p(t)$. The most elementary choice on pseudotimes t_d consists of them being sampled equispaced in $[0,1]$. We also consider sampling from a $Beta(0.2, 0.2)$ distribution and from a "cubic" distribution; i.e., where $\mathbf{t}$ is proportional to t^3 on $[0, 1]$. As priors we consider the uniform distribution in $[0, 1]$ (representing the equispaced sampling) and the $Beta(0.2, 0.2)$.

We study the performance in estimation of β comparing against the regression estimator that would be obtained if the pseudotimes were known. If that was the case, $\beta = (\beta_0, \beta_1)$ could be inferred via least squares with closed-form expressions.

As shown in Fig. 5 and Supplementary Fig. S4, pseudotime inference signifies a sizable increase in the MSE of the posterior mean of β over the baseline where pseudotimes are known. As expected, the MSE decreases as the variance parameter ϕ^2 increases (here, assumed known). Proposition 2 anticipates that β may be biased even if D is arbitrarily large. This is consistent with Fig. 5A, where the MSE of β doesn't converge to the baseline, indicating this persisting bias. Although uncertainty quantification is not the main intent of our work, we show that similar behavior occurs when we compare the posterior variance of our estimator with the posterior variance of the usual linear regression estimator (Supplementary Fig. S4).

In Fig. 5A, we also illustrate the effect of the relationship between true pseudotime distribution $\mathbf{t}$ and priors $p(t)$. The MSE of β is the closest to the regression baseline if the prior matches the true distribution of the data. However, if there is a disagreement, the MSE of β can even grow with D. One pessimistic interpretation is that the prior has a too sizable influence on the outcomes even for large donor sample size D. While this is true, at the same time, there is no way to get around this problem because of the inherent lack of identifiability when $N < \infty$. More optimistically, these results show that it is possible to encode our understanding of how donor sampling occurred if that knowledge is available.

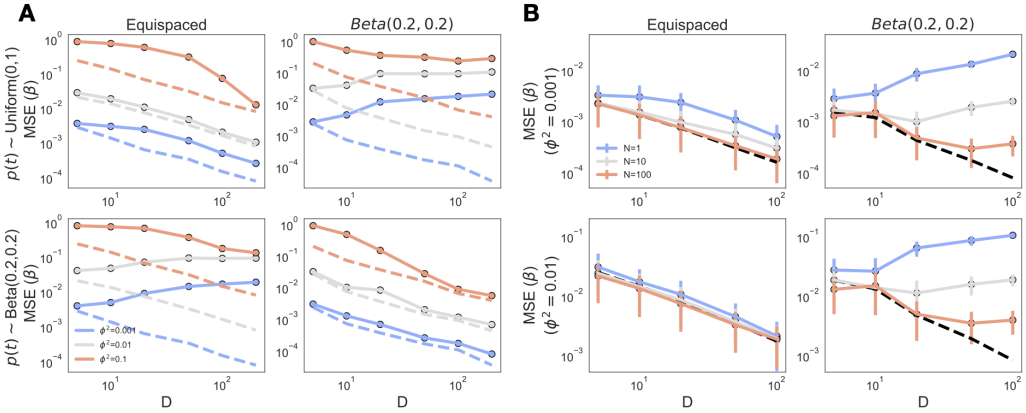


FIG 5. ***Effect of choices of pseudotime priors (A) number of features*** N ***(B) and in the linear model (Eq. (4)) on estimation error (MSE) of*** $\beta = (\beta_0, \beta_1)$. *(A): Priors* $p(t) \sim Uniform(0, 1)$ *and* $p(t) \sim Beta(0.1, 0.1)$ *are shown on different rows and population distribution of pseudotimes* $\mathbf{t}$ *on different columns. Colors indicate noise levels* ϕ^2. *We used a single feature for this experiment (*$N = 1$*). (B): Different rows correspond to different observation noise levels* ϕ^2, *and columns indicate the true distribution of times* $\mathbf{t}$. *Colors indicate the number of features* N. *Features were chosen to be equal, specifically, we used* $\beta_0^n = 0$ *and* $\beta_0^n = 1$. *In both A and B, dashed lines correspond to the baseline where pseudotimes are known.*

5.2. *Using multiple features.* Fig. 5B shows that we can benefit from including informative features: as we increase N, the average MSE for each feature decreases, converging to the linear regression (known pseudotimes) baseline. For many features (e.g., $N = 100$), divergences from this baseline are observed only for many donors D, where MSEs are already low. This phenomenon can be understood in the light of our theoretical results: By virtue of Proposition 1, a higher number of features (with the same β) adding more features will lead to a more precise estimation of the (normalized) vector of pseudotimes t. These more precise estimates will, in turn, lead to estimates $\hat{\beta}$ that are closer to the ones that would be obtained if pseudotimes were known.

In the previous experiment, we assumed that all features were equal. We also show that the same phenomenon still holds if features are not identical but are instead sampled from a distribution. To this end, we consider the scenario where $\beta_0^n = 0$ and β_1^n are sampled from a Gaussian distribution with mean $\beta_1 = 1$ and variance $\sigma^2 \in \{0, 10\}$. Results in Fig. 6 show that MSE for β is at a comparable scale independent of whether β_1^n is random or not ($\sigma^2 = 0$ v.s. $\sigma^2 = 10$).

However, unlike in the previous experiment, the added randomness in β_1^n is beneficial for pseudotime estimation as we observe that time estimates in the $\sigma^2 = 10$ case are consistently closer to true values. This may seem paradoxical as the $\sigma^2 = 10$ corresponds to a noisier regime, and we may reasonably expect noisier results. Again, we can understand this phenomenon in the light of Proposition 1: larger values of σ^2 translate into larger values of $||\beta||^2$, which in turn lead to a lower upper bound on the discrepancy between true and estimated pseudotimes, up to a normalization constant. This experiment highlights the relevance of feature diversity (i.e., variability in β) in tasks where pseudo time is the primary inferential goal. We will come back to this point when we discuss non-linearities.

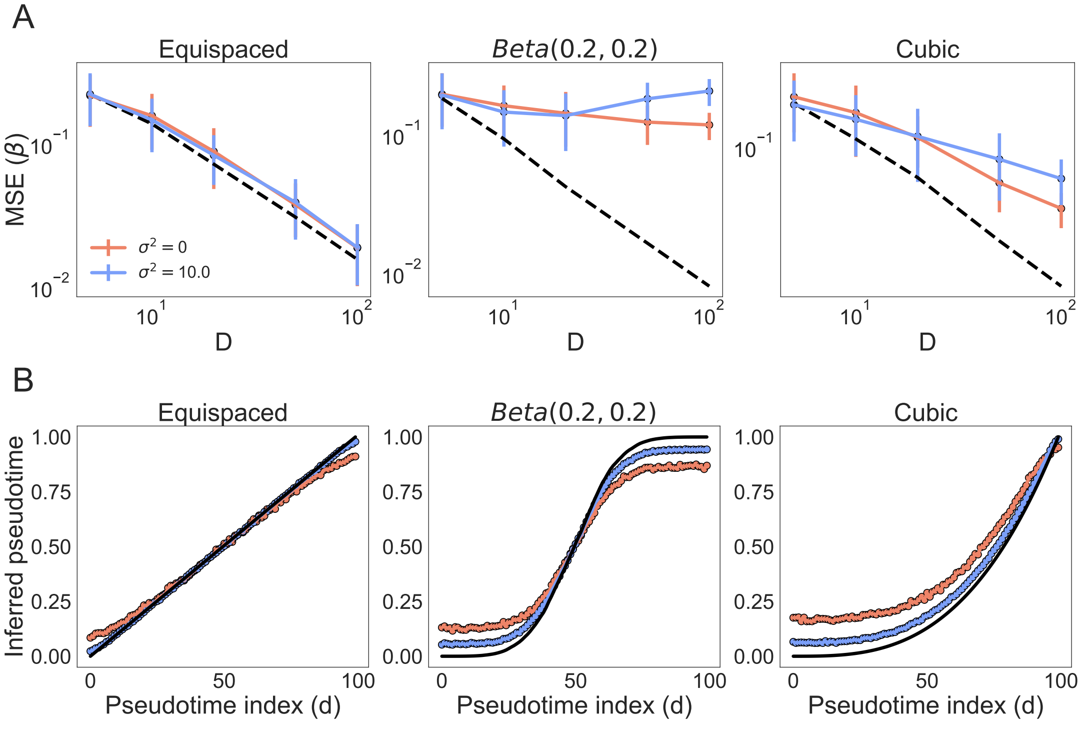


FIG 6. ***Effect of variability in β on estimation σ^2 (colors) in linear model (Eq. (4)) with $N = 10$ features on parameter and pseudotime estimation*** *(A) MSE of $\beta = (\beta_0, \beta_1)$. Dashed lines correspond to the baseline where the t_d's are known). (B) comparison of pseudotime estimates (averages over experiments)*

5.3. *Shrinkage in Gaussian and non-Gaussian models with hierarchical priors.* While previous experiments address the multiple features case and the effect of variability in feature coefficients β on our estimates, none of the Bayesian models in those experiments explicitly model the sampling structure in coefficients. One natural question is whether the James-Stein phenomenon would manifest here, i.e., whether the MSE of β or t can be lowered using shrinkage estimators. These estimators, in turn, are obtained by imposing a suitable hierarchical structure on β. This question is not obvious since it depends on the interplay of β and the "nuisance" parameter t. In this experiment, we not only consider the Gaussian-linear model but include the more realistic non-linear negative binomial model with logarithmic link function (Eq. (2)).

As in the previous experiment, we assume that β are sampled i.i.d. from a superpopulation; i.e. $\beta_0^n \sim \mathcal{N}(\beta_0, \sigma^2)$, $\beta_1^n \sim \mathcal{N}(\beta_1, \sigma^2)$. We consider Bayesian estimators based on four models: i) the one from previous experiments; i.e. a fully factorized independence prior on β (Fig 4A), and two hierarchical models (similar, but simpler than Fig 4B) where each β_0^n, β_1^n are sampled from a Gaussian distribution centered at β_0, β_1 and with variance σ^2_{model}, where σ^2_{model} is either assumed known and equal to the true variance σ^2 or treated as an additional parameter for which we perform Bayesian inference, using a half-normal hyperprior.

Fig. 7 shows that whether we benefit from a hierarchical model depends on the link function and the target parameter. As shown by Fig. 7B,C the hierarchical model is beneficial for the estimation of β in both the Gaussian and negative binomial case, although benefits tend to be more significant in the negative binomial model. More interestingly, regarding pseudotime estimation, benefits of shrinkage are observed only for the negative binomial model (Fig. 7A,C).

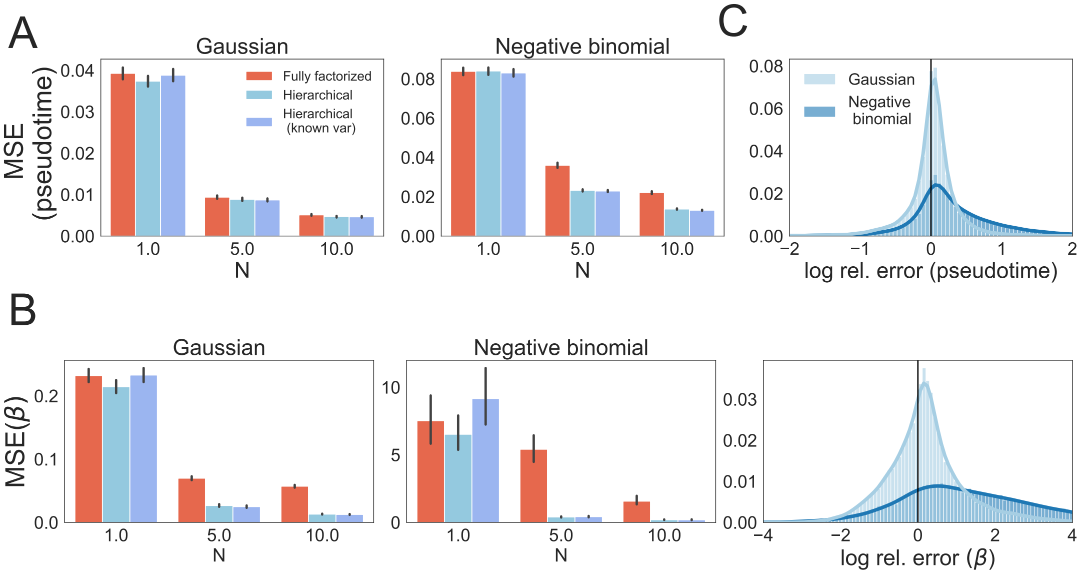


FIG 7. ***Estimation error of different hierarchical models for the linear-Gaussian model (Eq.** (4)**) and the negative binomial model (Eq.** (2)**).*** *(A) MSE in pseudotime estimation. (B) MSE for β estimation. (C) Histograms of relative error (log-scale) for the fully factorized model compared to the hierarchical model each sample is an experiment). Each sample is an experiment*

These results suggest that our framework can benefit from shared structure in β in two ways. First, through the "classical" James-Stein phenomenon whereby MSE on β is decreased by a hierarchical model on β. Second, at least in the negative binomial case, observe a less obvious "crossed' shrinkage effect whereby estimation of pseudotime is benefited by the structure that we imposed on β. This may, in turn, explain a more significant benefit in β estimation in this case.

Additionally, Figs. 7A, B show that regardless of the observation model (Gaussian or negative binomial) and the fitted model (hierarchical or fully factorized), our estimators produce lower pseudotime MSE (average over different t_d's) and β MSE for larger values of N. Also, Supplementary Fig. S9 shows that a larger number of donors D has little impact on the average pseudotime MSE, but does help decreasing the β MSE. Moreover, Supplementary Fig. S9 shows that benefits of the hierarchical model in pseudotime estimation for the negative binomial model are constrained to a sufficiently small number of donors compared to the number of features. Additional results are shown in Supplementary Figs. S7 and S8.

5.4. *Combining features with stage-specific information in non-linear models.* One main feature distinguishing linear and non-linear models is that the signal-to-noise ratio SNR depends nonlinearly on times, implying that individual features can be preferentially informative about particular levels of AD pathology severity. Indeed, if we define (see Hastie et al. (2009))

$$SNR = \frac{g^{-1}(Var(\rho))}{Var(X|\rho)},$$

where ρ is the linear response ($\rho = \beta_0^n + \beta_1^n t_d$ in Eq. (2)) and g is the link function (logarithmic in Eq. (2)), we observe that if $\beta_1^n > 0$, the SNR can be much larger at high compared to low severity levels, with an opposite behavior if $\beta_1^n < 0$.

We illustrate this phenomenon with an example of a negative binomial model with two features and an overdispersion parameter $A = 100$. The first feature has parameters $\beta_0^1 = 4, \beta_1^1 = -4$, so the mean counts $E(X_t^1)$ are large at low severity levels but decrease to eventually stabilize at high severity (see Supplementary Fig. S10 for examples of observations). The second feature has $\beta_0^2 = -4$ and $\beta_1^2 \in \{2, 4, 6\}$ so the behavior is inverted, and the overall strength of this feature can be smaller ($\beta_1^2 = 2$), equal ($\beta_1^2 = 4$) or larger ($\beta_1^2 = 6$) than the first feature.

We compare estimators based on each individual feature with the one based on both. Results are shown in Fig. 8. We observe that including both features leads to a lower MSE (on t and β) than any individual feature (Fig. 8A), suggesting that that the cross-talk of information between different features is beneficial for the estimation of each of them as they are uniquely informative about different aspects of the disease progression.

We can better understand this phenomenon by looking at pseudotime sequences inferred by each model. As shown in Fig. 8B, the model that only uses the first feature accurately infers lower levels of severity but struggles in later-pseudotime inference, and the converse is true for the model that only uses the second feature. In contrast, the model that combines both accurately infers pseudotimes along the entire range. Fig. 8B also shows that this consensus will be achieved even if there is asymmetry (one feature stronger than the other) as long as there is enough signal in the system (i.e., if both β_1^1, β_1^2 are sufficiently large). These observations are also supported by Fig. 8C, where we report MSE on pseudotime inference by classifying each donor as being on an 'early' ($t \leq 0.5$) or 'late' ($t \geq 0.5$) stage. Models with a single feature have the lowest MSE at the stages that the feature is informative about, and the overall pseudotime MSE of models containing the second feature decreases with the strength of this feature.

5.5. *Additional experiments.* In the Appendix (see Section "Supplementary experiments and experimental details", we present more experiments showing that the phenomena described in linear models (Sections 5.1, 5.2, 4) also hold if we use Negative Binominal likelihood (Eq. (2)). Additionally, we present supplementary experiments to illustrate the performance of the rate-optimal SVD-based estimators described in Sections 4 (Supplementary

Figs. S1 and S2),as an alternative to the Bayesian estimators whose performance was studied in this section . In this context, we also show that the benefit of shrinkage estimators can be observed beyond the situations described here (see Supplementary Fig. S3).

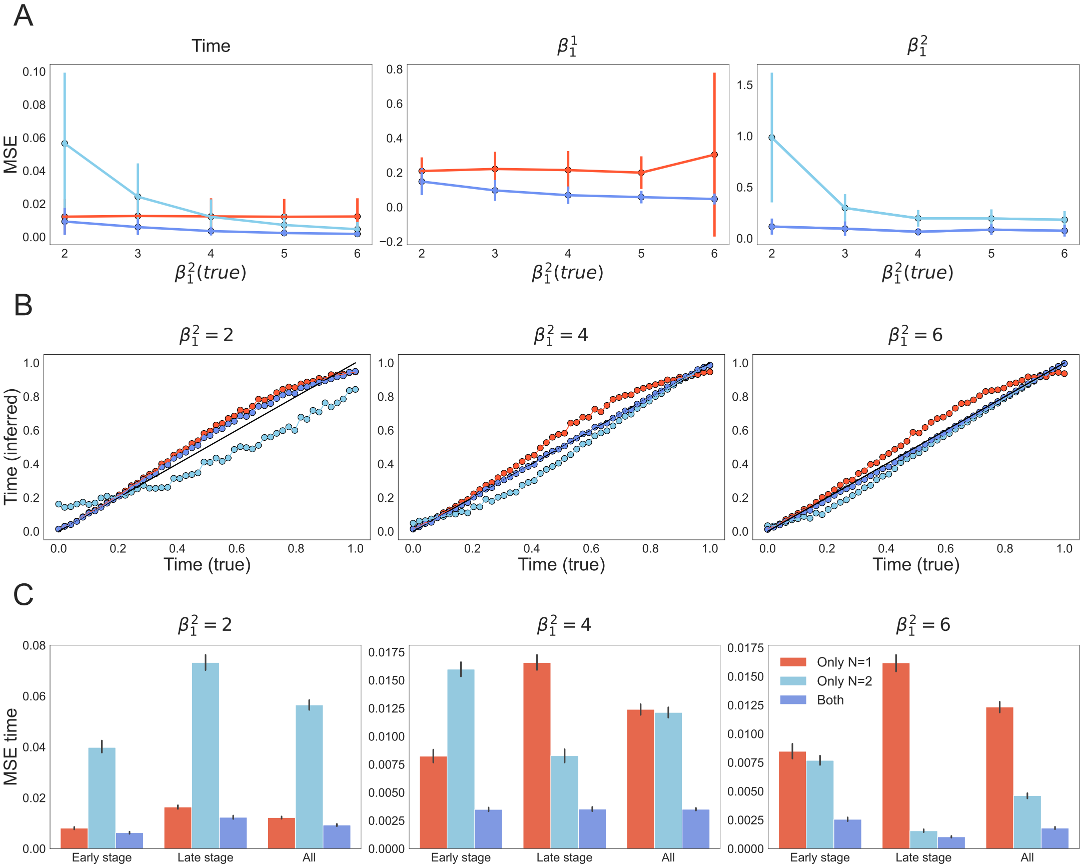


FIG 8. ***Inferring multiple features in the negative binomial non-linear model (Eq.** (2)**)** (A): Average MSE for pseudotimes, and MSE for the strength of the first and second features (β_1^1 and β_1^2) as a function of the strength of the second feature β_1^2 in models that only use individual features as well as the one that uses both (different colors). (B) Comparison of true and inferred pseudotimes (average over experiments) for different strengths of the second feature β_1^2. (C) Average MSE of inferred pseudotimes for the early and late stages, as well as their combination, for different strengths of the second feature β_1^2.*

6. Inferring latent pseudotime in the SEA-AD cohort.

Finally, we apply our generative models (Fig. 4) to the problem of describing the progression of pathology aggregation in AD, and infer latent pseudotime for each donor in the SEA-AD cohort. As a reminder, our dataset consists of $D = 84$ donors in which $M = 12$ pathological markers are measured in each of $L = 5$ cortical layers of the MTG (as detailed in Section 2). After sampling from the models' posteriors (see Section 3.3), posterior pseudotime estimates, t_d, span the entire progression of disease (Fig. 9A), with an increased representation of later times, in accordance with the cohort distribution that is biased towards older ages and higher disease metrics (such as Braak, or ADNC, see Fig. 1B,C). Growth-rate parameter β_1 and the parameter quantifying the initial seed levels of pathology β_0, revealed that canonical AD protein markers follow increasing exponential dynamics from small values, while, as expected, a marker quantifying the number of neuronal cells per unit area decreases as AD progresses (Fig. 9B,C).

Consistent with our predictions from simulation studies, we observed that for $T = 84$ donors, both the hierarchical model and the fully-factorized model result in similar posterior distributions of model parameters. This is due to the size of the dataset in which a sufficient number of donors compensate for the extra information provided by the layer and measurement covariance incorporated by the hierarchical model. Quantitative metrics comparing both models include the expected log posterior predictive distribution (Vehtari, Gelman and Gabry, 2017) and the Watanabe Information Criteria (WAIC, Watanabe and Opper (2010)). These metrics indicate that the hierarchical model better fits the data but the gain at this data regime is marginal (Supplementary Figures S11 and S12).

Next, we evaluated our model through correlations of our posterior pseudotime estimates with established neuropathological assessments, gaining the ability to independently validate our estimates with orthogonal metadata not used to fit the model. We observe that the pseudotime parameter is highly correlated with neuropathological assessments Braak, Thal, and ADNC (positive correlation) and cognitive scores CASI (negative correlation, Teng et al. (1994b)) (Fig. 9 E,F), These orthogonal metrics describe the brain-wide progression of multiple pathological proteins or quantify cognitive decline (see Section 2).

The definition of the Braak stages asserts that Braak stage IV (or higher) is identified by the presence of pTau pathology in the MTG, as described by Braak et al. (2006). This relationship aligns our inferred pseudotime, which is based on pTau measurements, with the Braak staging system. Nevertheless, this correlation is only moderate (the Spearman correlation between the inferred pseudotime and a binarized Braak score—0 for stages below 4 and 1 for stages 4 and above—is 0.35) when compared to the correlation using all Braak stages (correlation >0.65). This suggests that pseudotime is effectively ordering donors across the entire range of Braak stages.

6.1. *Exploring the importance of pathological markers as model regressors.*

We next investigated the contribution of each feature to our latent pseudotime. As our problem does not map to a regression problem, in which variable importance has been extensively studied, we resorted to multiple metrics to quantify each pathological marker contribution. To compare the effect of each pathological marker as an explanatory variable, making an analogy to regression problems, we first look at the growth-rate parameter β_1. Parameter values differ considerably across all $M = 12$ features, with six having the highest absolute value (Fig. 10A). However, in our case, parameter magnitude could be dependent on the signal-to-noise ratio in each feature and be loosely related to pseudotime inference. To find a different way of quantifying associations between pseudotime and each measurement, we utilized calculation of the mutual information (MI), understanding MI as a means to identify feature importance in nonlinear models (Beraha et al., 2019). MI identifies the 7 features as the most relevant ones to inform pseudotime inference (Fig. 10B).

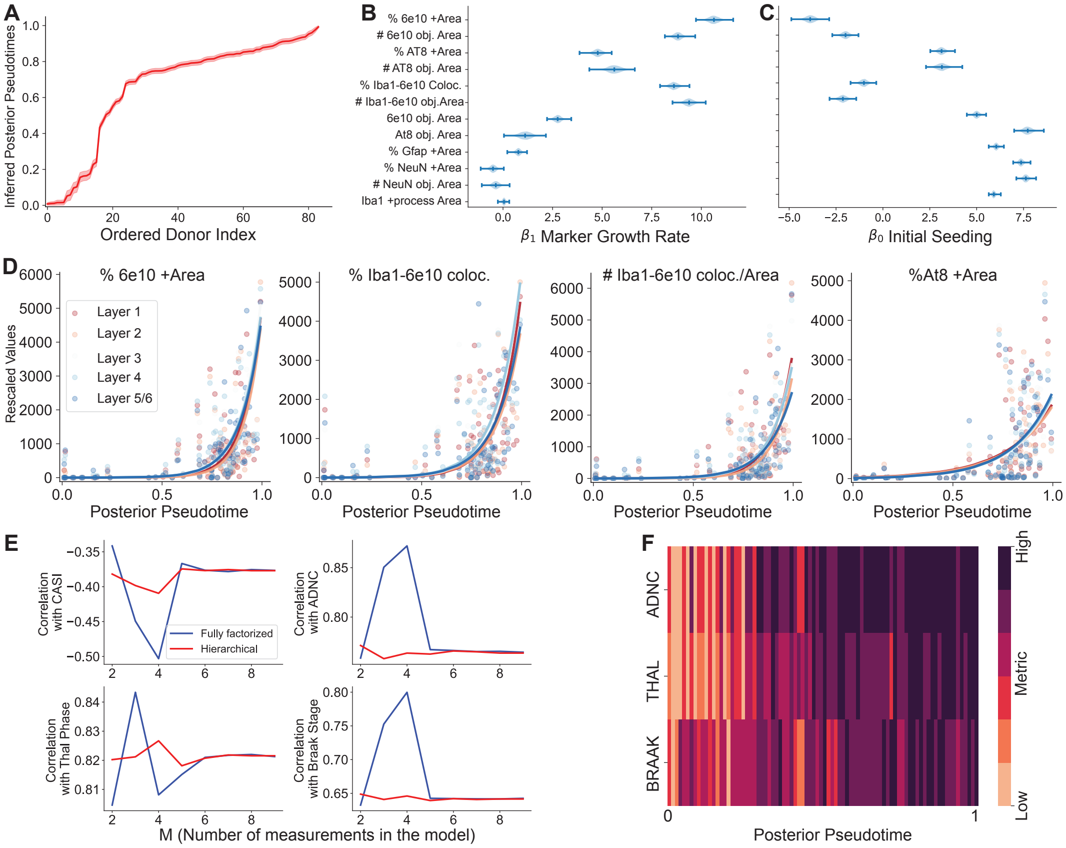


FIG 9. ***Posterior distributions of model parameters for SEAD-AD data set of pathological proteins.*** *(A) Inferred pseudotimes t_d for donors sorted in ascending order. B) posterior distribution of β_1, growth-rate and (C) β_0, initial seeding model parameters. In each case, error bars denote the distribution of posterior means inferred for each layer. (D) Posterior means (solid lines) plotted in conjunction with the data for 4 exemplary pathological markers. Color-codes indicate observations for each cortical layer.* ***Inferred pseudoprogression correlates with existing metrics for pathology progression.*** *(E) Spearman correlation between the associated progression metric and inferred pseudoprogression score td, as the number of measurements to infer td increase . Both models (fully factorized and hierarchical) show high correlations with these metrics. However only a few measurements are required to maximize the correlations. (F) All relevant metrics increase monotonically with the inferred pseudoprogression score inferred from our model*

As an additional feature importance computational experiment, we calculated pseudotime posterior estimates for models using each feature in isolation. Then, we calculated the error in the posterior pseudotime parameter t_d inferred from only one feature compared to all features, $(|t_{d,M=1..12} - t_{d,M_i}|$ (Fig. 10 C) and the standard deviation in the posterior distribution of each pseudotime parameter t_d, to show posterior uncertainty on pseudotime estimation (Fig. 10D). Our previous simulation studies informed us that features with high growth-rate parameter β_1 would have a high-dynamic range of pathology and observations with high pathological values. This in turn is more informative for obtaining identifiable estimates of t_d. Our computational results are in accordance with our simulation studies, revealing that the first seven pathological variables (order according to decreasing MI) have small pseudotime error and variance (Fig. 10C-D). Taken together all these computational experiments suggest that the seven most informative variables are a more compact subset able to create reliable pseudotime posterior estimates.

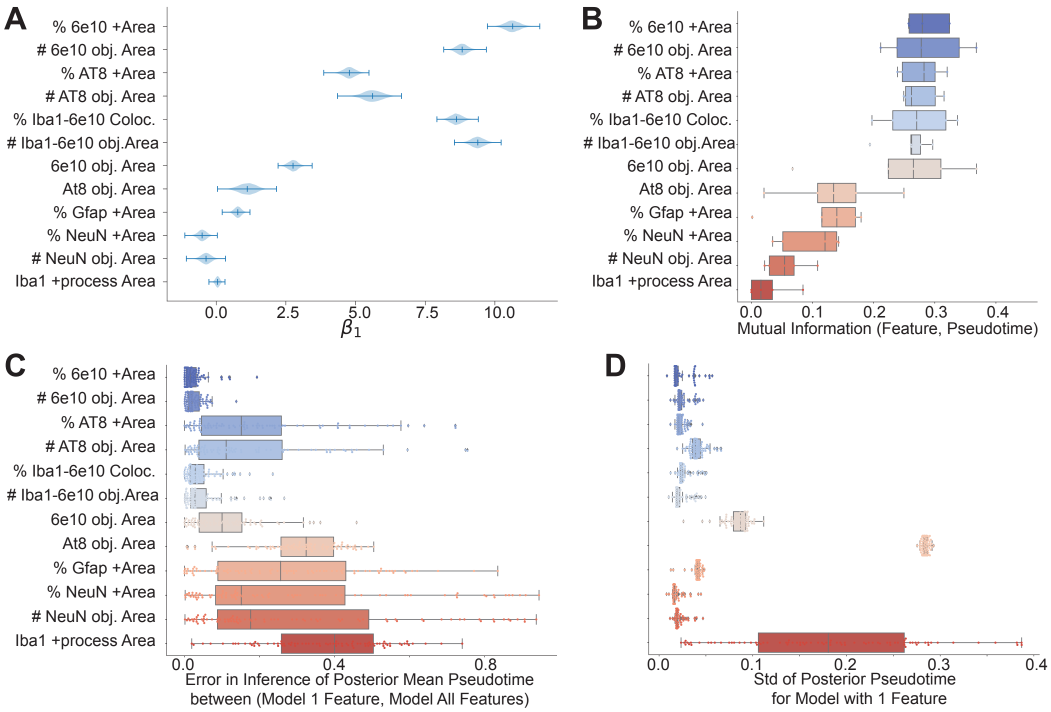


FIG 10. ***Multiple metrics describing the importance of each pathological feature to inform latent pseudotime inference.*** *(A) Growth-rate parameter, β_1, for each pathological feature, ordered according to decreasing mutual information (MI). (B) MI between each pathological feature and posterior pseudotime. (C) Error in pseudotime inference by comparing posterior times inferred using one feature vs all $M = 12$ features. (D) The standard deviation of posterior estimate, t_d, for each single-feature model.*

6.2. *Informing Future Experimental Design.* This section is driven by the aim of leveraging our models to guide the design of future experiments. While this aim could be framed as a Bayesian optimization problem, we opt for a qualitative approach guided by several principles. Firstly, the process of collecting, processing, and analyzing neuropathological measurements, as presented here, is labor-intensive. Our objective is to utilize the insights gained from the preceding section to identify a concise set of highly informative variables. Although reducing the feature set could be in detriment of the quality of fit (as described in our Simulated Studies Section 5), this will streamline the time required for future experiments. Secondly, the selection criteria for brain donors in the SEA-AD cohort resulted in a bias towards older donors and those at advanced pathological stages. Most donors are of old age, mean age = 88 years old, and exhibit an inferred pseudotime greater than 0.5. In the future, we would like to diversify the SEA-AD cohort by including donors that uniformly span the entire pseudotime spectrum (from 0 to 1). A balanced cohort, representative of all disease stages, facilitates downstream analysis of multiple disease epochs. Moreover, pseudotime being an inferred parameter poses challenges in its use for donor preselection and cohort balancing. To address this, we aim to enhance the model by incorporating other observable variables such as donor metadata or staging information. This approach will enable us to strategically sample donors and mitigate biases.

To achieve these goals, we investigate a model incorporating $M = 7$ pathological features, chosen for their high standardized beta coefficient, mutual information, and low error and standard deviation in time inference (see Fig. 10). These features constitute a minimal set as desired. To suggest future brain donors based on easily measurable variables, we augment the model with categorical staging metrics (Braak, Thal, and ADNC). Let m represent a staging metric, d denote a donor, and k signify a one-hot encoded stage level, then the vector $s_k^{d,m}$ represents:

$$s_k^{d,m} = \begin{cases} 1, \text{if } k = \text{donor's stage} \\ 0, \text{if } k \neq \text{donor's stage} \end{cases} \tag{6}$$

For example, Thal phases range from phase 0 to 5, with increasing phases indicating an increasing number of brain regions containing amyloid-beta pathological burden. A donor exhibiting phase-2 Thal phase would be represented as $s^{d,Thal} = [0, 0, 1, 0, 0, 0]$. We incorporate staging metrics into our formalism and model them as:

$$\begin{aligned} p_k^{d,m} &= \alpha_{0,m}^k + \alpha_{1,m}^k t_d \\ s^{d,m} &= \text{DirichletMultinomial}(c^{d,m} = e^{p^{d,m}}, \text{count} = 1) \end{aligned} \tag{7}$$

Assuming that pseudotime $t_d = 0$ aligns with all staging metrics being zero and that their values are monotonically higher, we assign prior parameters α_0 and α_1 for each stage k.

$$\begin{aligned} \alpha_0^k &\sim \mathcal{N}(-10k, 5) \\ \alpha_1^k &\sim \mathcal{N}(2k - 10, 5) \end{aligned} \tag{8}$$

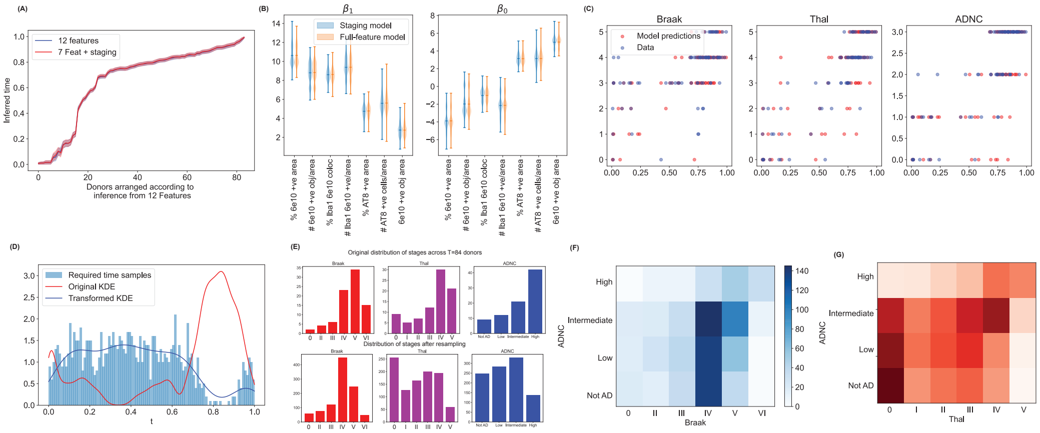


FIG 11. ***Model augmented with staging metrics is able to effectively represent experimental data.*** *(A) Comparison of posterior time estimates predicted by models spanning all features and features plus staging metrics. (B) Posterior β_0 and β_1 distributions (concatenating values from all layers and samples) in both models (C) Posterior predictive estimates reveal effective modeling of staging metric (Braak, Thal and ADNC).* ***Augmented Model incorporating staging information informs experimental design to balance SEA-AD cohort across disease stages*** *(D) To balance SEA-AD cohort, we used kernel density estimates to describe the distribution of inferred pseudotime values. Next, we empirically calculate a pdf corresponding to the least probable time values. $N = 1000$ samples are drawn from this distribution and the resulting histogram is displayed (E) Pseudotime samples from (D) are then converted to the associated Braak, Thal and ADNC stage. Top plot shows the distribution of stages in the donor dataset, bottom plot shows the distribution of stages corresponding to the resampled times. (F,G) Heatmaps showing the Braak/ADNC and Thal/ADNC joint histograms. The associated distribution of stages can aid experimentalists in improving donor sampling to balance their cohort.*

Next, we perform Bayesian posterior inference and compare posterior estimates of the augmented model (Augmented, $M = 7$ + Staging metrics) against our model that includes all features (All Feats, $M = 12$) in Fig. 11 and Supplementary Fig. S15. Both models resulted in similar posterior estimates, revealed by comparing posterior pseudotime estimates for each donor and dynamic parameters (β_0, β_1) across measurements common to both settings (Fig. 11A-B). In both cases, neuropathological data used to infer posterior parameters results in concordant estimates and relatively small posterior standard deviations. We observed that staging information does not significantly impact pseudotime posterior estimates, both estimates concentrating around the same mean, when most informative neuropathological features are selected. Although speculative, this may be indicative of the fact that the staging information may be too coarse to provide any additional information on disease progression on top of the one already provided by neuropathological markers. At the same time, the augmented model is able to effectively model staging metrics, observed when comparing staging metrics' posterior predictive distributions and observed data (Fig. 11C). Taken together, these results highlights the ability of our augmented model together with the set of informative features to accurately model pseudotime and AD staging metrics in concert.

Finally, we utilize our augmented model to guide experimental design, aiming to achieve balance (equal representation of donors across pseudotime) in the SEA-AD cohort. To accomplish this, we first estimate the distribution of posterior pseudotimes, $p(t)$, by 1) aggregating mean posterior pseudotime estimates for each donor, 2) constructing a histogram, and then 3) applying kernel density estimation. Next, we approximate the unnormalized distribution of under-sampled pseudotimes in the SEA-AD cohort by computing $p_{under}(t) = 1 - p(t)$. This approximation enables us to generate samples of under-sampled pseudotimes ($n_{samples} = 1000$, see Fig. 11 D). These values can then be employed to compute posterior predictive staging metrics based on the posterior estimates of staging parameters (see Fig. 11E-G). These predictions inform the staging metrics of suggested new donors that should be included in the study to achieve cohort balance. Given the current distribution of our cohort, it is apparent that sampling from early and middle stages would be advantageous.

7. Discussion.

We introduced a Bayesian framework to model Alzheimer's Disease progression, incorporating a latent pseudotime variable to depict disease evolution. We modeled observations utilizing biophysically-inspired assumptions that posit pathological aggregation follows exponential dynamics. Our inference method is useful for datasets restricted to cross-sectional data, consolidating observations and inferring a single and common disease trajectory. We demonstrated the utility of our framework using simulation studies and real datasets originating from the SEA-AD consortium, profiling AD neuropathological proteins in a brain region. Applying our model to the SEA-AD cohort dataset, consisting of 84 donors, allowed us to infer donor pseudotime and pathological protein dynamics. Pseudotime estimates are biased towards late values, in accordance with our observations that the SEA-AD cohort is biased towards older donors and high pathology. Posterior predictive checks validate the goodness of our fits. In addition, comparisons with existing neuropathological staging-based metrics reveal that pseudotime ordering aligns well with brain-wide neuropathological assessments, indicating that our model is able to map AD trajectory when detailed information from only one brain region is used (the Middle Temporal Gyrus).

Next, we assembled a collection of metrics to assess the impact of each pathological protein on pseudotime inference, identifying the most informative features. These features exhibited high growth rates, mutual information between their values and pseudotime, and individually exhibited precise predictions of pseudotime with low prediction error. Incorporating these highly informative variables alongside neuropathological assessments, we enhanced our model to deduce the latter. When integrated with staging data, our enhanced model can guide the identification of neuropathological stages for future donors to be included in the existing SEA-AD cohort, ensuring a balanced representation of donors across all pseudotimes and enhancing downstream analysis across the disease spectrum. While our work assumes that the pathology markers align donors along the same pseudotime trajectory, which is true for the MTG region of the brain, this assumption may not hold true for other brain regions. This would require approaches that can overcome low-data regimes to successfully leverage all regions in a combined overarching multi-regional pseudotime.

Our results demonstrated the success of our methodology, supported by our theoretical explorations. In our work, we derived theoretical results on a simplified linear model to understand the best possible performance. A stimulating recent line of work has provided algorithmic (Kidzinski et al., 2022) and statistical (Chen, Li and Zhang, 2020; Du, Wasserman and Roeder, 2023) guarantees for matrix factorization problems in generalized-linear setups, a setup akin to ours. However, these results by themselves are not sufficient to explain the richness of phenomena manifesting in our setup. Future studies should study the simplified model and extend Proposition 2 by deriving explicit rate bounds for the signal strength estimates, and also extend our theoretical results to a generalized linear model case. We made contributions in this direction through extensive simulations and showed that whether or not we benefit from hierarchical models for pseudotime estimation depends on the type of model and error distribution, and benefits are only observed beyond linear models with Gaussian errors. Future studies should develop a theoretical framework to explain this phenomenon.

Acknowledgments. The authors would like to thank Sivaraman Balakrishnan for useful theoretical discussions, Bob Carpenter for insightful discussion on modeling and inference approaches, Iris Stone for manuscript feedback, and the entire SEA-AD team for creating an open source resource for the study of Alzheimer's Disease.

Funding. A.A. was supported by the Shanahan Foundation Fellowship. G.M was supported by NSF Grant 2412895. M.I.G. was supported by National Institutes on Aging (NIA) Grant U19AG060909 and R01102-01-071-20. This work used Bridges-2 at Pittsburgh Supercomputing Center through allocation MTH230027 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services and Support (ACCESS) program, which is supported by NSF Grant 2138259, 2138286, 2138307, 2137603, and 213829.

REFERENCES

Beraha, M., Metelli, A. M., Papini, M., Tirinzoni, A. and Restelli, M. (2019). Feature Selection via Mutual Information: New Theoretical Insights. In *2019 International Joint Conference on Neural Networks (IJCNN)* 1-9. https://doi.org/10.1109/IJCNN.2019.8852410

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P. A., Horsfall, P. and Goodman, N. D. (2019). Pyro: Deep Universal Probabilistic Programming. *J. Mach. Learn. Res.* **20** 28:1–28:6.

Braak, H. and Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta neuropathologica* **82** 239–259.

Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H. and Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta neuropathologica* **112** 389–404.

Cai, T. T. and Zhang, A. (2018). Rate-Optimal Perturbation Bounds for Singular Subspaces with Applications to High-Dimensional Statistics. *The Annals of Statistics* **46** 60–89.

Campbell, K. R. and Yau, C. (2016). Order under uncertainty: robust differential expression analysis using probabilistic models for pseudotime inference. *PLoS computational biology* **12** e1005212.

Campbell, K. R. and Yau, C. (2018). Uncovering pseudotemporal trajectories with covariates from single cell and bulk expression data. *Nature communications* **9** 2442.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P. and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software* **76**.

Chen, Y., Li, X. and Zhang, S. (2020). Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications. *Journal of the American Statistical Association* **115** 1756–1770.

Crane, P. K., Gibbons, L. E., Jolley, L. and van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Medical care* **44** S115–S123.

Davis, J. K. and Sindi, S. S. (2016). A mathematical model of the dynamics of prion aggregates with chaperone-mediated fragmentation. *Journal of mathematical biology* **72** 1555–1578.

Du, J.-H., Wasserman, L. and Roeder, K. (2023). Simultaneous inference for generalized linear models with unmeasured confounders. *arXiv preprint arXiv:2309.07261*.

Duane, S., Kennedy, A. D., Pendleton, B. J. and Roweth, D. (1987). Hybrid monte carlo. *Physics letters B* **195** 216–222.

Gabitto, M. I., Travaglini, K. J., Rachleff, V. M., Kaplan, E. S., Long, B., Ariza, J., Ding, Y., Mahoney, J. T., Dee, N., Goldy, J. et al. (2023). Integrated multimodal cell atlas of Alzheimer's disease. *Research Square*.

Gelman, A. and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Gelman, A., Hill, J. and Vehtari, A. (2021). *Regression and other stories*. Cambridge University Press.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2021). Bayesian Data Analysis Third edition (with errors fixed as of 6 April 2021). *Issue: April*.

Gupta, A. and Bar-Joseph, Z. (2008). Extracting dynamics from static cancer expression data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **5** 172–182.

Hastie, T., Tibshirani, R., Friedman, J. H. and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* **2**. Springer.

Hoegh, A. and Roberts, D. W. (2020). Evaluating and presenting uncertainty in model-based unconstrained ordination. *Ecology and Evolution* **10** 59–69.

Hoffman, M. D., Gelman, A. et al. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15** 1593–1623.

Hou, W., Ji, Z., Chen, Z., Wherry, E. J., Hicks, S. C. and Ji, H. (2023). A statistical framework for differential pseudotime analysis with multiple single-cell RNA-seq samples. *Nature communications* **14** 7286.

Huang, K., Zhang, Y., Gong, H., Qiao, Z., Wang, T., Zhao, W., Huang, L. and Zhou, X. (2023). Inferring evolutionary trajectories from cross-sectional transcriptomic data to mirror lung adenocarcinoma progression. *PLOS Computational Biology* **19** e1011122.

Hui, F. K. C., Taskinen, S., Pledger, S., Foster, S. D. and Warton, D. I. (2015). Model-based approaches to unconstrained ordination. *Methods in Ecology and Evolution* **6** 399-411. https://doi.org/10.1111/2041-210X.12236

Hui, F. K., Warton, D. I., Foster, S. D. and Haak, C. R. (2023). Spatiotemporal joint species distribution modelling: A basis function approach. *Methods in Ecology and Evolution* **14** 2150–2164.

Hyman, B., Phelps, C., Beach, T., Bigio, E., Cairns, N., Carrillo, M., Dickson, D., Duyckaerts, C., Frosch, M., Masliah, E., Mirra, S., Nelson, P., Schneider, J., Thal, D., Thies, B., Trojanowski, J., Vinters, H. and Montine, T. (2012). National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement.* **8**.

Kidzinski, L., Hui, F. K., Warton, D. I. and Hastie, T. J. (2022). Generalized Matrix Factorization: efficient algorithms for fitting generalized linear latent variable models to large data arrays. *Journal of Machine Learning Research* **23** 1–29.

Kim, S. and Camilli, G. (2014). An item response theory approach to longitudinal analysis with application to summer setback in preschool language/literacy. *Large-scale assessments in Education* **2** 1–17.

Magwene, P. M., Lizardi, P. and Kim, J. (2003). Reconstructing the temporal ordering of biological samples using microarray data. *Bioinformatics* **19** 842–850.

Masel, J., Jansen, V. A. and Nowak, M. A. (1999). Quantifying the kinetic parameters of prion replication. *Biophysical chemistry* **77** 139–152.

Mukherjee, S., Heath, L., Preuss, C., Jayadev, S., Garden, G. A., Greenwood, A. K., Sieberts, S. K., De Jager, P. L., Ertekin-Taner, N., Carter, G. W. et al. (2020). Molecular estimation of neurodegeneration pseudotime in older brains. *Nature communications* **11** 5781.

Neal, R. M. et al. (2011). MCMC using Hamiltonian dynamics. *Handbook of markov chain monte carlo* **2** 2.

O'Hara, R. B. and van der Veen, B. (2024). Hierarchical Ordination, A unifying framework for drivers of community processes. *bioRxiv* 2024–01.

Phan, D., Pradhan, N. and Jankowiak, M. (2019). Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro. *arXiv preprint arXiv:1912.11554*.

Pierson, E., Koh, P., Hashimoto, T., Koller, D., Leskovec, J., Eriksson, N. and Liang, P. (2019). Inferring Multidimensional Rates of aging from Cross-sectional Data. *Proc. Mach. Learn. Res.* **89**.

Popovic, G. C., Hui, F. K. and Warton, D. I. (2022). Fast model-based ordination with copulas. *Methods in Ecology and Evolution* **13** 194–202.

Reid, J. E. and Wernisch, L. (2016). Pseudotime estimation: deconfounding single cell time series. *Bioinformatics* **32** 2973–2980.

Roberts, D. W. (2020). Comparison of distance-based and model-based ordinations. *Ecology* **101** e02908.

Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E. and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics* **19** 1–16.

Teng, E., Hasegawa, K., Homma, A., Imai, Y., Larson, E., Graves, A., Sugimoto, K., Yamaguchi, T., Sasaki, H., Chiu, D. and et al. (1994a). The Cognitive Abilities Screening Instrument (CASI): a practical test for cross-cultural epidemiological studies of dementia. *Int Psychogeriatr.* **6**.

Teng, E. L., Hasegawa, K., Homma, A., Imai, Y., Larson, E., Graves, A., Sugimoto, K., Yamaguchi, T., Sasaki, H., Chiu, D. et al. (1994b). The Cognitive Abilities Screening Instrument (CASI): a practical test for cross-cultural epidemiological studies of dementia. *International psychogeriatrics* **6** 45–58.

Thal, D. R., Rüb, U., Orantes, M. and Braak, H. (2002). Phases of Aβ-deposition in the human brain and its relevance for the development of AD. *Neurology* **58** 1791–1800.

Tompa, P., Tusnády, G., Cserző, M. and Simon, I. (2001). Prion protein: evolution caught en route. *Proceedings of the National Academy of Sciences* **98** 4431–4436.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S. and Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology* **32** 381–386.

van der Veen, B., Hui, F. K., Hovstad, K. A. and O'Hara, R. B. (2023). Concurrent ordination: Simultaneous unconstrained and constrained latent variable modelling. *Methods in Ecology and Evolution* **14** 683–695.

Vaquer-Alicea, J. and Diamond, M. I. (2019). Propagation of protein aggregation in neurodegenerative diseases. *Annual review of biochemistry* **88** 785–810.

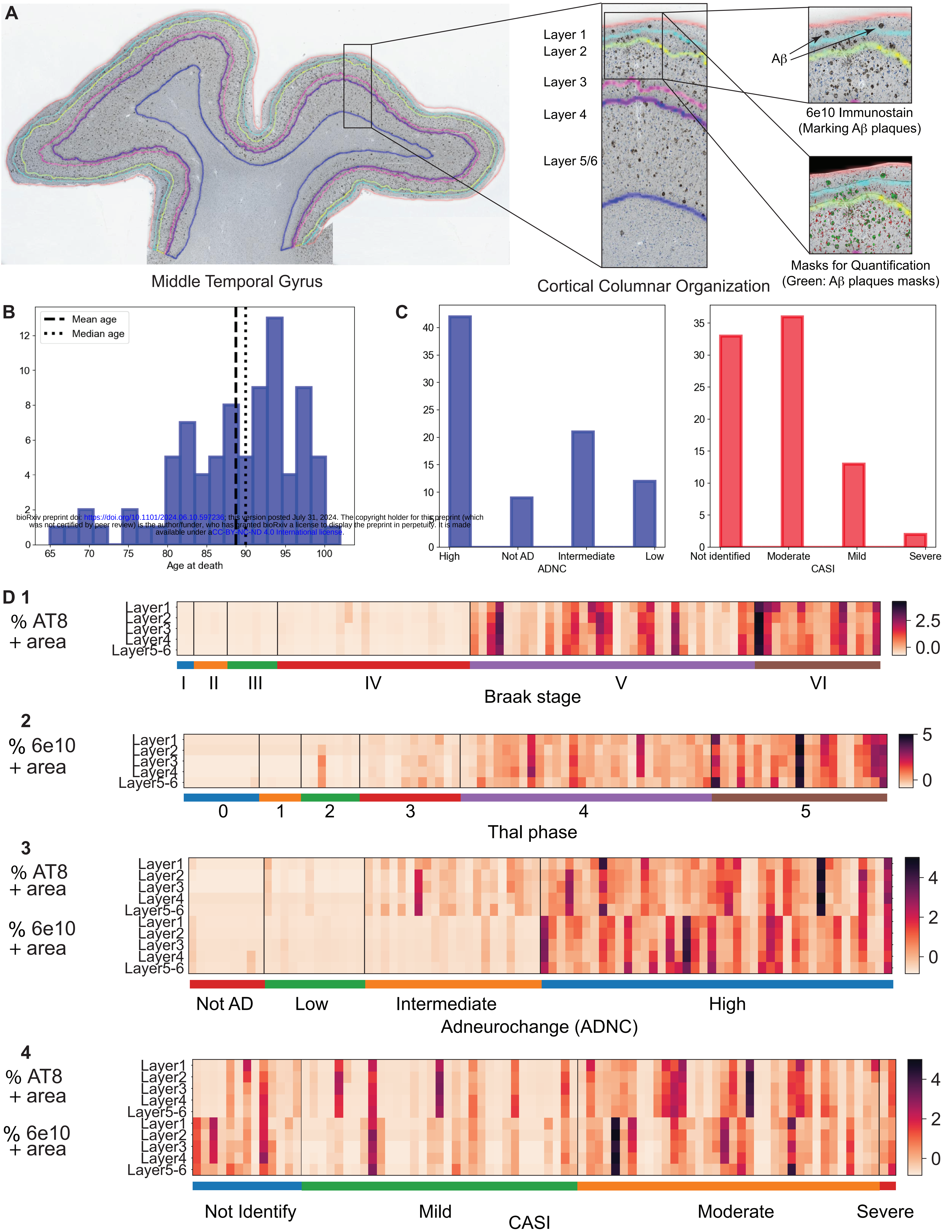
Vehtari, A., Gelman, A. and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing* **27** 1413–1432.

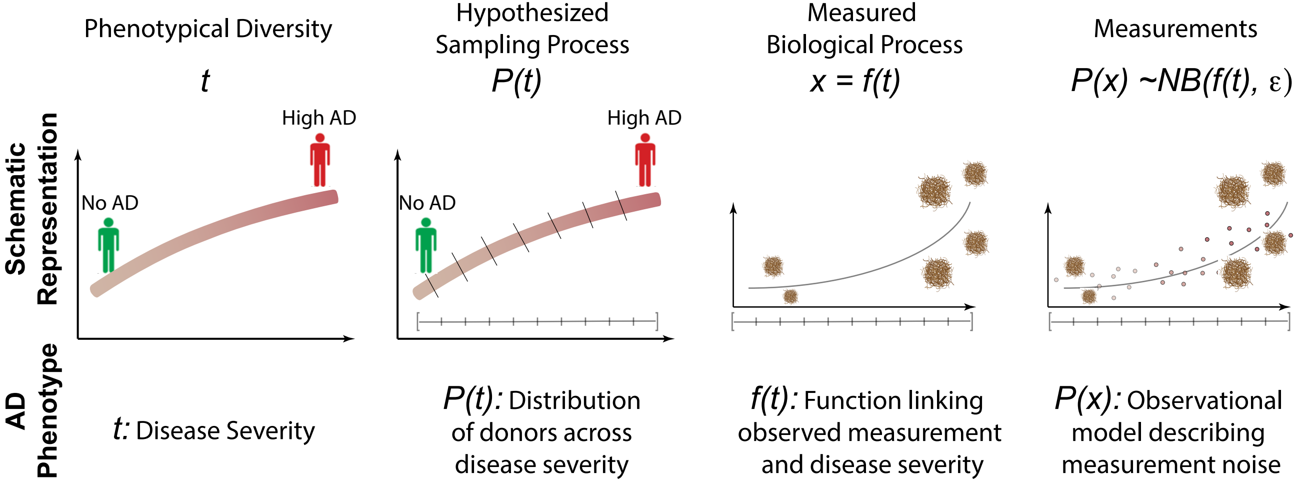
Waddington, C. H. (1940). *Organisers and Genes*. University Press; Cambridge.

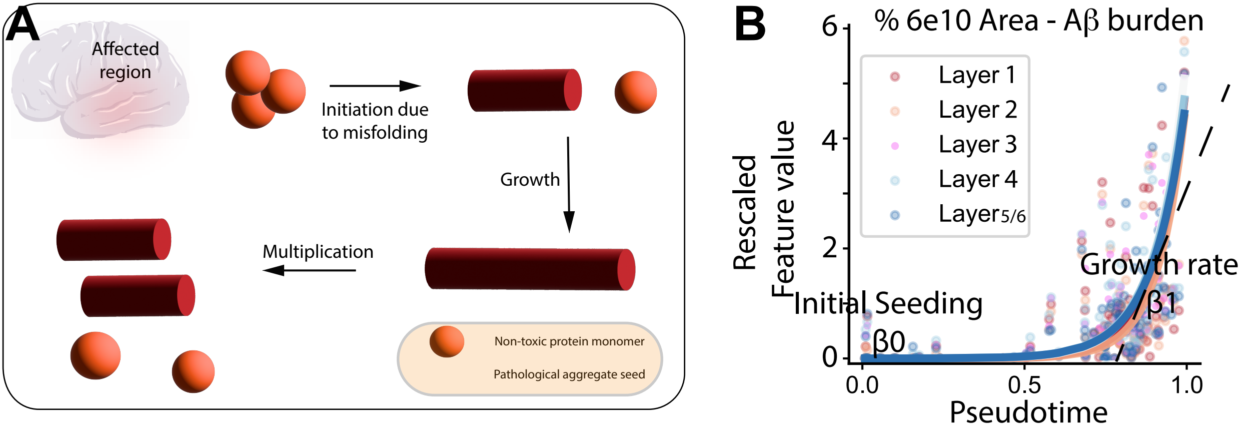
Waddington, C. H. (1957). *The strategy of genes*. Taylor and Francis Group.

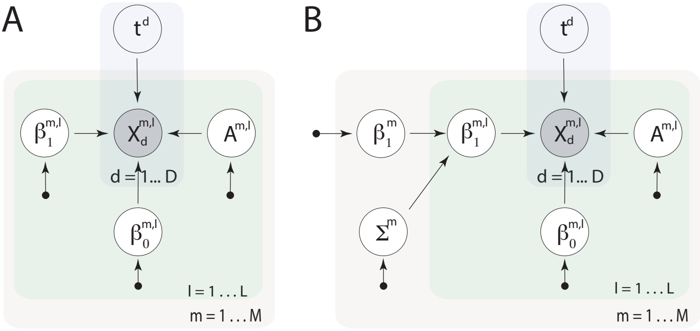
Watanabe, S. and Opper, M. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research* **11**.

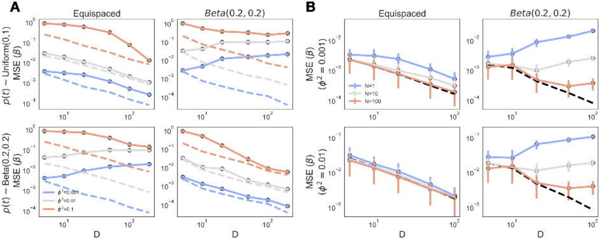
Wedin, P.-Å. (1972). Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics* **12** 99–111.

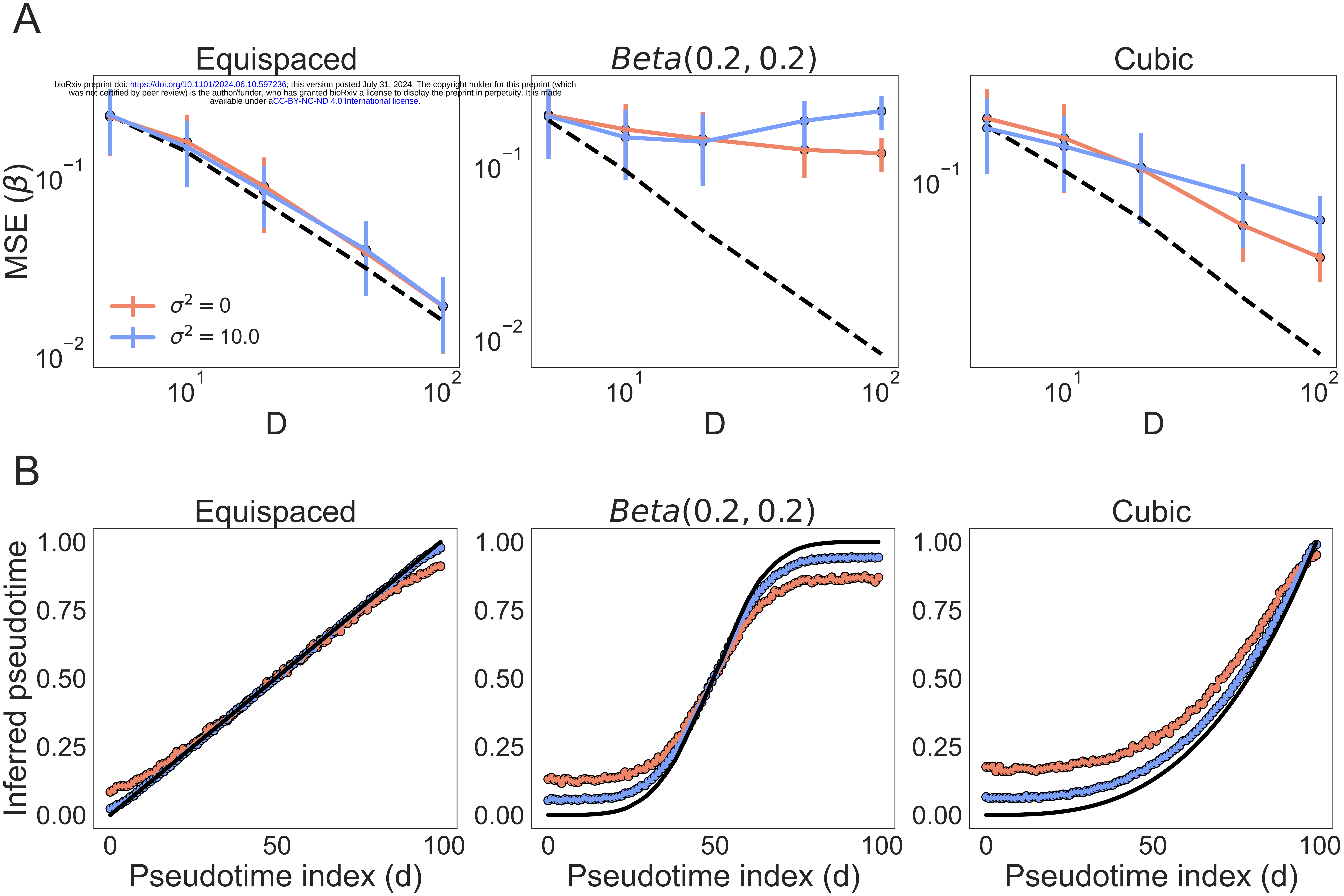

A
Layer 1
Layer 2
Layer 3
Layer 4
Layer 5/6
Aβ
6e10 Immunostain
(Marking Aβ plaques)
Masks for Quantification
(Green: Aβ plaques masks)
Middle Temporal Gyrus
Cortical Columnar Organization
B
Mean age
Median age
0
2
4
6
8
10
12
65
70
75
80
85
90
95
100
Age at death
C
0
5
10
15
20
25
30
35
40
High
Not AD
Intermediate
Low
ADNC
0
5
10
15
20
25
30
35
Not identified
Moderate
Mild
Severe
CASI
D 1
% AT8
+ area
Layer1
Layer2
Layer3
Layer4
Layer5-6
2.5
0.0
I
II
III
IV
V
VI
Braak stage
2
% 6e10
+ area
Layer1
Layer2
Layer3
Layer4
Layer5-6
5
0
0
1
2
3
4
5
Thal phase
3
% AT8
+ area
% 6e10
+ area
Layer1
Layer2
Layer3
Layer4
Layer5-6
Layer1
Layer2
Layer3
Layer4
Layer5-6
4
2
0
Not AD
Low
Intermediate
High
Adneurochange (ADNC)
4
% AT8
+ area
% 6e10
+ area
Layer1
Layer2
Layer3
Layer4
Layer5-6
Layer1
Layer2
Layer3
Layer4
Layer5-6
4
2
0
Not Identify
Mild
Moderate
Severe
CASI

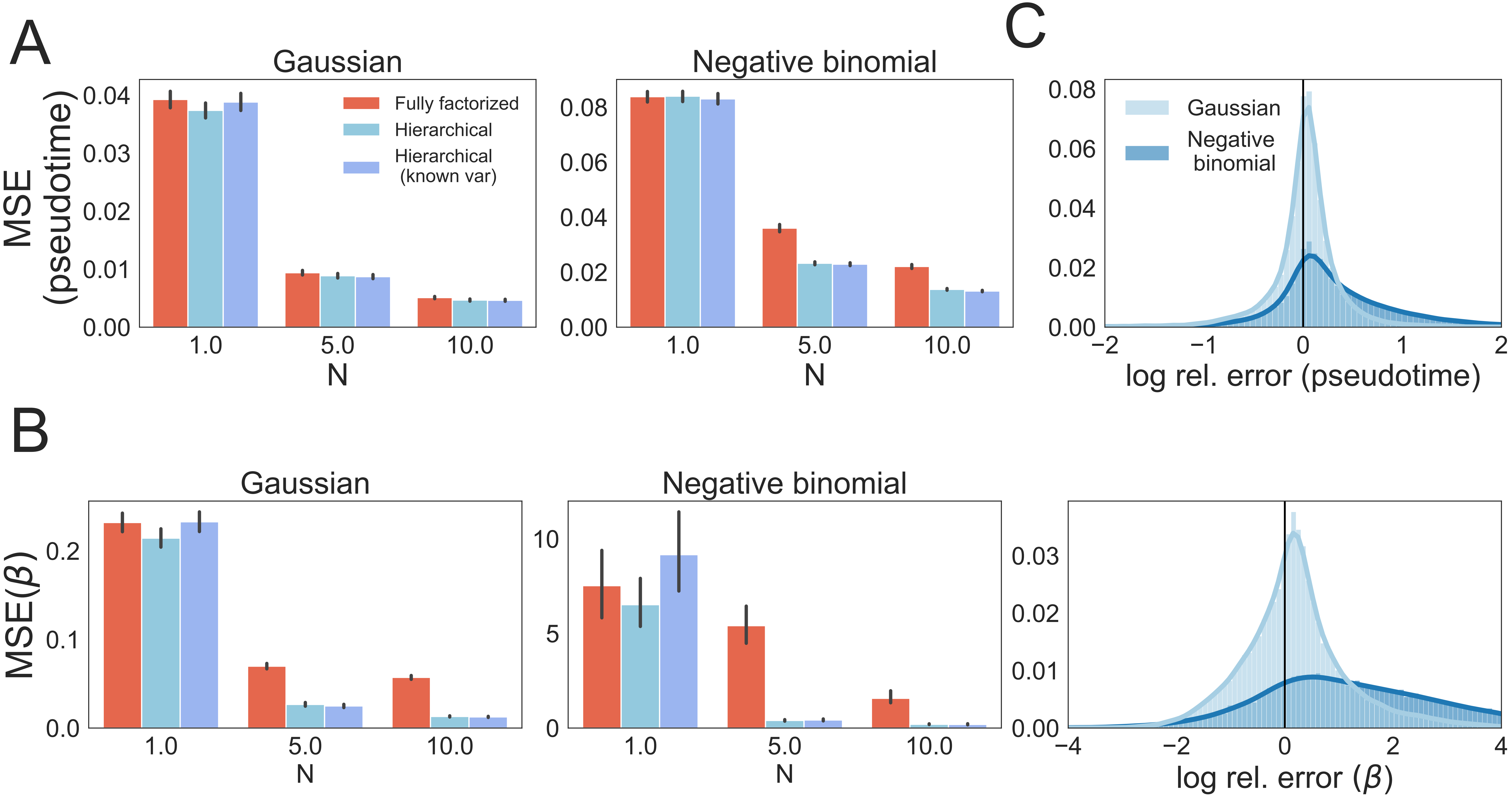

Phenotypical Diversity
t
Hypothesized Sampling Process
P(t)
Measured Biological Process
x = f(t)
Measurements
P(x) ~NB(f(t), ε)
Schematic Representation
High AD
No AD
High AD
No AD
AD Phenotype
t: Disease Severity
P(t): Distribution of donors across disease severity
f(t): Function linking observed measurement and disease severity
P(x): Observational model describing measurement noise

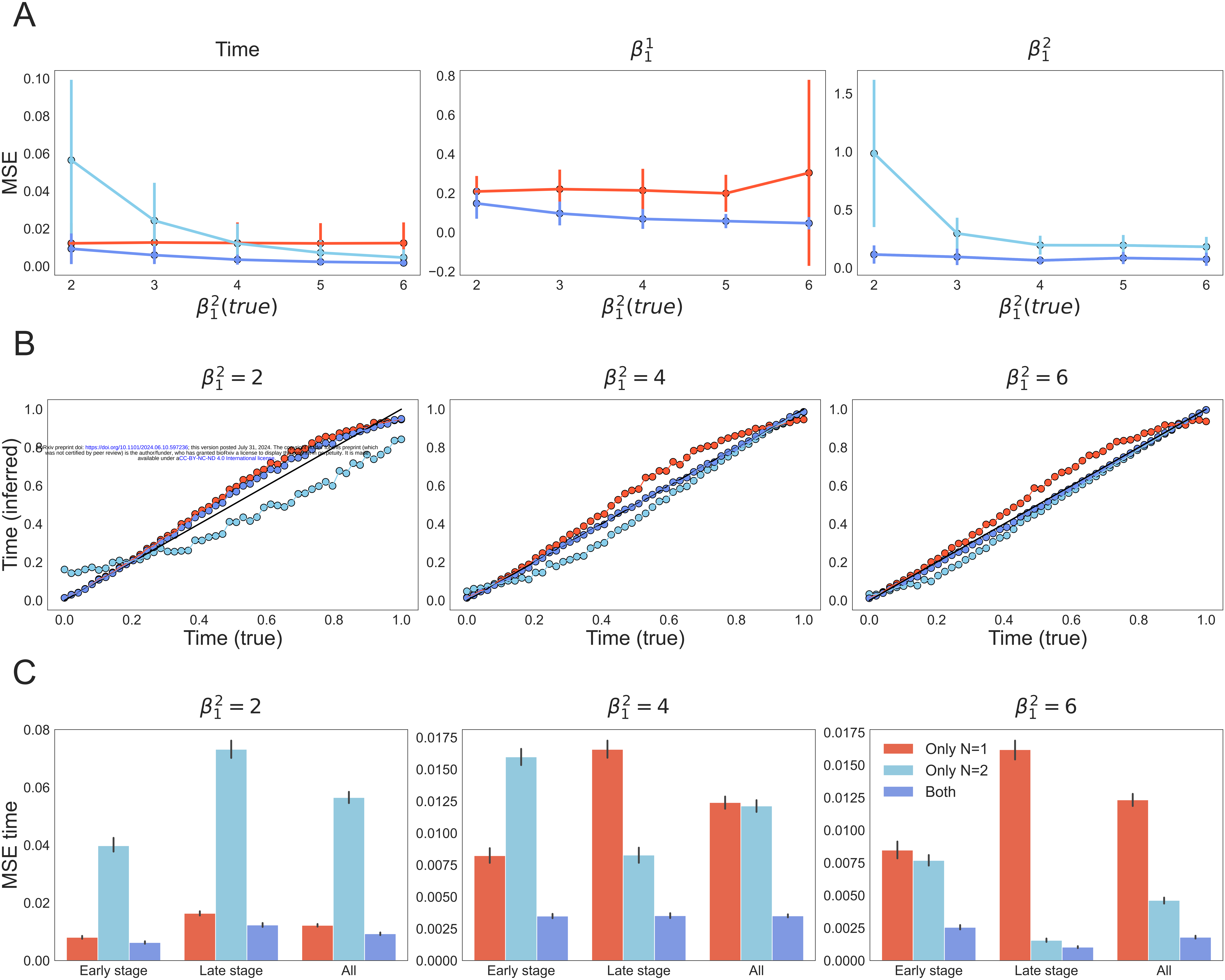

A
Affected region
Initiation due to misfolding
Growth
Multiplication
Non-toxic protein monomer
Pathological aggregate seed
B
% 6e10 Area - Aβ burden
Rescaled Feature value
6
4
2
0
Layer 1
Layer 2
Layer 3
Layer 4
Layer5/6
Initial Seeding
β0
Growth rate
/β1
0.0
0.5
1.0
Pseudotime

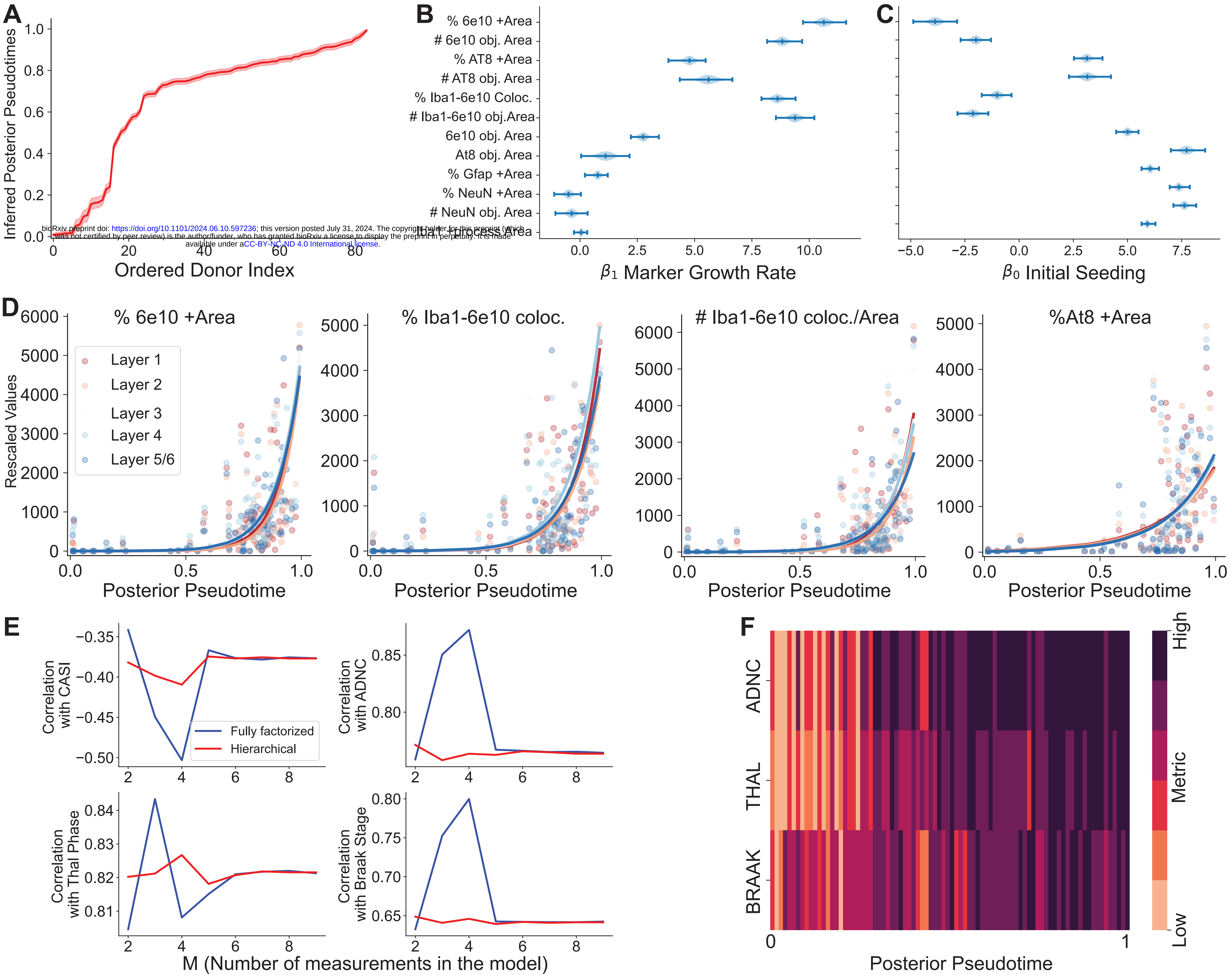

A
t^d
$\beta_1^{m,l}$
$X_d^{m,l}$
$A^{m,l}$
d = 1... D
$\beta_0^{m,l}$
l = 1 . . . L
m = 1 . . . M
B
t^d
β_1^m
$\beta_1^{m,l}$
$X_d^{m,l}$
$A^{m,l}$
d = 1... D
Σ^m
$\beta_0^{m,l}$
l = 1 . . . L
m = 1 . . . M


A
Equispaced
$Beta(0.2, 0.2)$
$p(t) \sim$ Uniform(0,1)
MSE (β)
$p(t) \sim$ Beta(0.2,0.2)
MSE (β)
ϕ^2=0.001
ϕ^2=0.01
ϕ^2=0.1
D
D
B
Equispaced
$Beta(0.2, 0.2)$
MSE (β)
($\phi^2 = 0.001$)
MSE (β)
($\phi^2 = 0.01$)
N=1
N=10
N=100
D
D


A
Equispaced
$Beta(0.2, 0.2)$
Cubic
MSE (β)
10^{-1}
10^{-2}
10^{1}
10^{2}
D
$\sigma^2 = 0$
$\sigma^2 = 10.0$
B
Equispaced
$Beta(0.2, 0.2)$
Cubic
Inferred pseudotime
1.00
0.75
0.50
0.25
0.00
0
50
100
Pseudotime index (d)


A
Gaussian
Negative binomial
C
MSE (pseudotime)
0.04
0.03
0.02
0.01
0.00
Fully factorized
Hierarchical
Hierarchical (known var)
1.0
5.0
10.0
N
0.08
0.06
0.04
0.02
0.00
Gaussian
Negative binomial
0.08
0.06
0.04
0.02
0.00
−2
−1
0
1
2
log rel. error (pseudotime)
B
Gaussian
Negative binomial
MSE(β)
0.2
0.1
0.0
1.0
5.0
10.0
N
10
5
0
0.03
0.02
0.01
0.00
−4
−2
0
2
4
log rel. error (β)


A
Time
MSE
0.10
0.08
0.06
0.04
0.02
0.00
2
3
4
5
6
$\beta_1^2(true)$
β_1^1
0.8
0.6
0.4
0.2
0.0
−0.2
$\beta_1^2(true)$
β_1^2
1.5
1.0
0.5
0.0
$\beta_1^2(true)$
B
$\beta_1^2 = 2$
$\beta_1^2 = 4$
$\beta_1^2 = 6$
Time (inferred)
1.0
0.8
0.6
0.4
0.2
0.0
Time (true)
C
$\beta_1^2 = 2$
$\beta_1^2 = 4$
$\beta_1^2 = 6$
MSE time
0.08
0.06
0.04
0.02
0.00
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
Early stage
Late stage
All
Only N=1
Only N=2
Both


A
Inferred Posterior Pseudotimes
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
Ordered Donor Index
B
% 6e10 +Area
# 6e10 obj. Area
% AT8 +Area
# AT8 obj. Area
% Iba1-6e10 Coloc.
# Iba1-6e10 obj.Area
6e10 obj. Area
At8 obj. Area
% Gfap +Area
% NeuN +Area
# NeuN obj. Area
Iba1 +process Area
0.0
2.5
5.0
7.5
10.0
β_1 Marker Growth Rate
C
−5.0
−2.5
0.0
2.5
5.0
7.5
β_0 Initial Seeding
D
% 6e10 +Area
% Iba1-6e10 coloc.
# Iba1-6e10 coloc./Area
%At8 +Area
Rescaled Values
Layer 1
Layer 2
Layer 3
Layer 4
Layer 5/6
0
1000
2000
3000
4000
5000
6000
0.0
0.5
1.0
Posterior Pseudotime
E
Correlation with CASI
Correlation with ADNC
Correlation with Thal Phase
Correlation with Braak Stage
Fully factorized
Hierarchical
−0.35
−0.40
−0.45
−0.50
0.85
0.80
0.84
0.83
0.82
0.81
0.80
0.75
0.70
0.65
2
4
6
8
M (Number of measurements in the model)
F
ADNC
THAL
BRAAK
0
1
Posterior Pseudotime
High
Metric
Low

A
% 6e10 +Area
# 6e10 obj. Area
% AT8 +Area
# AT8 obj. Area
% Iba1-6e10 Coloc.
# Iba1-6e10 obj.Area
6e10 obj. Area
At8 obj. Area
% Gfap +Area
% NeuN +Area
# NeuN obj. Area
Iba1 +process Area
0.0
2.5
5.0
7.5
10.0
β_1
B
% 6e10 +Area
# 6e10 obj. Area
% AT8 +Area
# AT8 obj. Area
% Iba1-6e10 Coloc.
# Iba1-6e10 obj.Area
6e10 obj. Area
At8 obj. Area
% Gfap +Area
% NeuN +Area
# NeuN obj. Area
Iba1 +process Area
0.0
0.1
0.2
0.3
0.4
Mutual Information (Feature, Pseudotime)
C
% 6e10 +Area
# 6e10 obj. Area
% AT8 +Area
# AT8 obj. Area
% Iba1-6e10 Coloc.
# Iba1-6e10 obj.Area
6e10 obj. Area
At8 obj. Area
% Gfap +Area
% NeuN +Area
# NeuN obj. Area
Iba1 +process Area
0.0
0.4
0.8
Error in Inference of Posterior Mean Pseudotime
between (Model 1 Feature, Model All Features)
D
0.0
0.1
0.2
0.3
0.4
Std of Posterior Pseudotime
for Model with 1 Feature

(A)
12 features
7 Feat + staging
Inferred time
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
Donors arranged according to inference from 12 Features
(B)
β_1
β_0
Staging model
Full-feature model
% 6e10 +ve area
# 6e10 +ve obj/area
% Iba1 6e10 coloc
# Iba1 6e10 +ve/area
% AT8 +ve area
# AT8 +ve cells/area
6e10 +ve obj area
(C)
Braak
Thal
ADNC
Model predictions
Data
0.00
0.25
0.50
0.75
1.00
(D)
Required time samples
Original KDE
Transformed KDE
t
(E)
Original distribution of stages across T=84 donors
Braak
Thal
ADNC
Distribution of stages after resampling
0
I
II
III
IV
V
VI
Not AD
Low
Intermediate
High
(F)
ADNC
Braak
High
Intermediate
Low
Not AD
0
II
III
IV
V
VI
(G)
ADNC
Thal
0
I
II
III
IV
V