Supplementary Material

Appendix A: Proof of $h_{t_0}(t)$ properties

In Section 4 of the main text, we specified

$$G_{t_0}(t) = A\exp\left(-[h_{t_0}(t)]^2\right)$$

with i) $h_{t_0}(t_0) = 0$ and ii) $h_{t_0}(t)$ a monotonically increasing function to be a Gaussian-like transient and stated that a representation of the form

$$h_{t_0}(t) = (t - t_0) \times \text{softplus}\left(c + \int_0^t (\tau - t_0) f(\tau)\mathrm{d}\tau\right),$$

with $f(\tau)$ an integrable function with positive range, satisfied these properties. Below we formalise and prove this result.

Proof

Let $a < 0 < b$ and $t, t_0 \in (a, b)$. We would like to show that for any positive-range real function $f : (a, b) \mapsto \mathbb{R}_{>0}$ which is locally L_1 integrable, the function

$$h_{t_0}(t) = (t - t_0) \times \text{softplus}\left(c + \int_0^t (\tau - t_0) f(\tau)\mathrm{d}\tau\right) \tag{1}$$

satisfies i) $h(t_0) = 0$ and ii) $h_{t_0}(t)$ is monotonically increasing.

Condition i): Evaluating $h(t_0)$ gives

$$h_{t_0}(t_0) = \underbrace{(t_0 - t_0)}_{=0} \times \text{softplus}\Bigg(\underbrace{c + \int_0^{t_0} (\tau - t_0) f(\tau)\mathrm{d}\tau}_{:= g(t_0)}\Bigg),$$

implying that condition i) is satisfied iff $g(t_0)$ is finite. As c is finite, this is equivalent to the requirement that

$$I(t_0) = \int_0^{t_0} (\tau - t_0) f(\tau)\mathrm{d}\tau$$

is finite. As $t - t_0$ is absolutely continuous and $f(\tau)$ is locally L_1 integrable, integration by parts with $u(\tau) = \tau - t_0$ and $v(\tau) = \int_0^\tau f(x)\mathrm{d}x$ can be applied to obtain

$$I(t_0) = \underbrace{u(t_0)v(t_0)}_{=0} - \underbrace{u(0)v(0)}_{=0} - \int_0^{t_0} v(\tau)\mathrm{d}\tau,$$

which is finite as $v(\tau)$ is continuous due to its integral definition. ■

Condition ii): To prove that $h_{t_0}(t)$ is monotonically increasing, we would like to show that $\frac{\mathrm{d}}{\mathrm{d}t} h_{t_0}(t) > 0 \; \forall \, t, t_0 \in (a, b)$. Differentiating (1) gives

$$\begin{aligned}\frac{\mathrm{d}h_{t_0}}{\mathrm{d}t} &= \text{softplus}\left(c + \int_0^t (\tau - t_0) f(\tau)\mathrm{d}\tau\right) \\ &\quad + (t - t_0)^2 f(t) \text{softplus}'\left(c + \int_0^t (\tau - t_0) f(\tau)\mathrm{d}\tau\right).\end{aligned}$$

Noting that $f(t)$, softplus and $\text{softplus}'$ have positive range, the result follows. ■

Appendix B: Example transients

In Section 4 of the main text, we introduced the notion of modelling Gaussian-like transients within BasisDeVAE via functions of the form

$$G_{t_0}(t) = A \exp\left(-[h_{t_0}(t)]^2\right) \tag{2}$$

with

$$h_{t_0}(t) = (t - t_0) \times \text{softplus}\left(c + \int_0^t (\tau - t_0) f(\tau) \mathrm{d}\tau\right), \tag{3}$$

where $f(\tau)$ is a neural network with softplus output activation.

The motivation for this choice comes from the idea that under the functional form (3), $h_{t_0}(t)$ satisfies i) $h_{t_0}(t_0) = 0$ and ii) $h_{t_0}(t)$ is a monotonically increasing function (see Appendix A for proof). This makes $h_{t_0}(t)$ a natural generalisation of the function $(t - t_0)/(\sqrt{2}\sigma)$ which would feature within a parametric Gaussian representation of a transient, motivating our use of the term "Gaussian-like transient" for functions of the form (2). Examples of these $G_{t_0}(t)$ transients for different choices of $h_{t_0}(t)$ in the context of the Ernst et al. (2019) spermatogenesis data are shown in figure 1, where it can be seen that $G_{t_0}(t)$ allows a wide range of transient behaviours to be captured via the training of the neural network $f(\tau)$ (right panel).

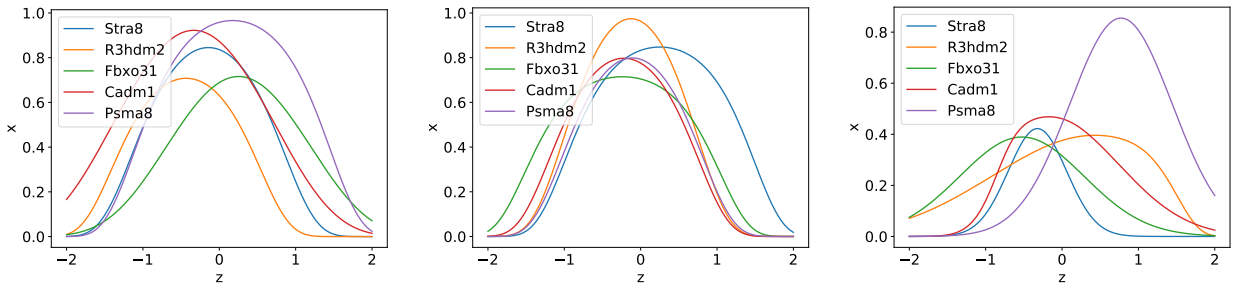

Figure 1: **Example $G_{t_0}(t)$ transients.** Left: $h_{t_0}(t)$ of the form (3) with $f(\tau)$ a random feature-dependent constant. Middle: $h_{t_0}(t)$ of the form (3) with $f(\tau)$ a Xavier-initialised neural network. Right: Same functional form as the middle panel, but after training on the Ernst et al. (2019) data via BasisDeVAE.

Appendix C: OASIS features

In Section 5.2 of the main text, we presented the results of experiments using data $\mathbf{X} \in \mathbb{R}^{2047 \times 13}$ from the OASIS-3 dataset (LaMontagne et al., 2019). Here we describe the meaning of each feature and how it is obtained from the raw OASIS-3 data.

We began by extracting MRI sessions with associated FreeSurfer segmentations and downloading these segmentations. By default, the FreeSurfer software provides highly granular cortical and subcortical segmentation. We apply the groupings described in the appendix of Klein & Tourville (2012) and in the FreeSurfer wiki to obtain the volumes of the following 13 interpretable regions associated with cognitive decline (also used in e.g. Young et al. (2018)): frontal lobe, temporal lobe, parietal lobe, occipital lobe, cingulate, insula, accumbens, amygdala, caudate, hippocampus, pallidum, putamen, thalamus. We then divide each regional measurement by the corresponding patient's total intracranial volume to obtain normalised volumes (Whitwell et al., 2001). Finally, we standardise each feature of the data by subtracting its mean and dividing by its standard deviation.

Appendix D: Single-cell BasisVAE cluster interpretability

In Section 5.3 of the main text, we fit BasisVAE1 and BasisVAE2 (defined in the main text) and BasisDeVAE to single-cell spermatogenesis data (Ernst et al., 2019). Using our derivative-based specification within BasisDeVAE allowed us to specify the meaning of each feature cluster, namely monotonically increasing/decreasing and Gaussian-like transient, making BasisDeVAE cluster behaviours immediately *interpretable*. Figure 2 shows that the interpretability of BasisVAE1 and

BasisVAE2 clusterings is less clear. In particular, it shows the behaviour of inferred BasisVAE1 and BasisVAE2 components with $\lambda = 1, \delta = \pm 1.5$. In the case of BasisVAE1, we see that changing δ leads to qualitatively different behaviour in components $k = 1$ and $k = 2$, meaning e.g. that both monotonic gene upregulation and monotonic gene downregulation could be attributed to component $k = 2$. Meanwhile, for BasisVAE2, we see a distinct qualitative similarity between each of the components, limiting the interpretability of a feature's cluster assignment.

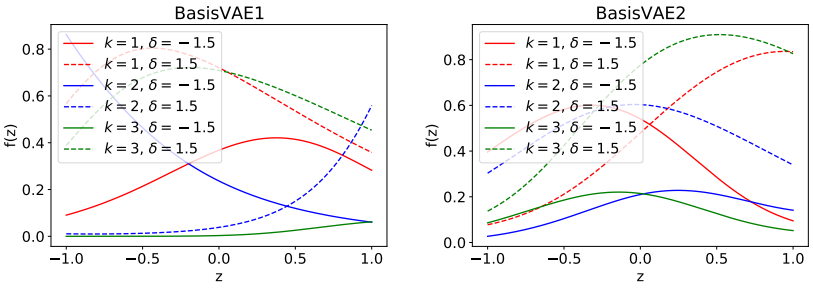

Figure 2: **Dependence of BasisVAE cluster predictions on translation parameter δ.** Variation of the appearance of BasisVAE1 (top) and BasisVAE2 (bottom) components with $\delta = 1.5$ (solid lines) and $\delta = -1.5$ (dashed). If the clustering were optimal, each plot would display qualitative similarity between same-colour lines and distinct behaviour between different-colour lines. This is not seen, supporting the observations made during synthetic data experiments that BasisVAE's basis functions can i) demonstrate multiple qualitatively different behaviours within a single basis function and ii) collapse such that each cluster's basis function has a very similar appearance (see Figure 4 of the main text).

References

Ernst, C., Eling, N., Martinez-Jimenez, C. P., Marioni, J. C., and Odom, D. T. Staged developmental mapping and x chromosome transcriptional dynamics during mouse spermatogenesis. *Nature Communications*, 10(1):1251, 2019. doi: 10.1038/s41467-019-09182-1. URL https://doi.org/10.1038/s41467-019-09182-1.

Klein, A. and Tourville, J. 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in Neuroscience*, 6:171, 2012. ISSN 1662-453X. doi: 10.3389/fnins.2012.00171. URL https://www.frontiersin.org/article/10.3389/fnins.2012.00171.

LaMontagne, P. J., Benzinger, T. L., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A. G., Raichle, M. E., Cruchaga, C., and Marcus, D. Oasis-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. *medRxiv*, 2019. doi: 10.1101/2019.12.13.19014902. URL https://www.medrxiv.org/content/early/2019/12/15/2019.12.13.19014902.

Whitwell, J. L., Crum, W. R., Watt, H. C., and Fox, N. C. Normalization of cerebral volumes by use of intracranial volume: Implications for longitudinal quantitative mr imaging. *American Journal of Neuroradiology*, 22(8):1483–1489, 2001. ISSN 0195-6108. URL http://www.ajnr.org/content/22/8/1483.

Young, A. L. et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. *Nature Communications*, 9(1):4273, Oct 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-05892-0. URL https://doi.org/10.1038/s41467-018-05892-0.