Label-free biochemical imaging of neural organoids via deep learning-enhanced Raman microspectroscopy

Dimitar Georgiev[*]
Imperial College London, UK
University of Oxford, UK

Ruoxiao Xie[*, †]
Imperial College London, UK

Daniel Reumann
Imperial College London, UK
University of Oxford, UK

Xiaoyu Zhao
Imperial College London

Álvaro Fernández-Galiana
Imperial College London, UK
University of Oxford, UK

Mauricio Barahona[#]
Imperial College London, UK
m.barahona@imperial.ac.uk

Molly M. Stevens[#]
University of Oxford, UK
Imperial College London, UK
molly.stevens@dpag.ox.ac.uk

Abstract

Three-dimensional organoids have emerged as powerful models for studying human development, disease and drug response *in vitro*. Yet, their analysis remains constrained by standard imaging and characterisation techniques, which are invasive, require exogenous labelling and offer limited multiplexing. Here, we present a non-invasive, label-free imaging platform that integrates Raman microspectroscopy with deep learning-based hyperspectral unmixing for unsupervised, spatially resolved biochemical analysis of neural organoids. Our approach enables high-resolution mapping of cellular and subcellular structures in both cryosectioned and intact organoids, achieving improved imaging accuracy and robustness compared to conventional methods for hyperspectral analysis. This work establishes a versatile framework for high-content, label-free (bio)chemical phenotyping with broad applications in organoid research and beyond.

1 Introduction

The ability to image and accurately characterise the architecture and molecular composition of cells and tissues has been central to advancing biological and medical sciences. To this end, a broad array of techniques – ranging from histological staining and immunofluorescence to omics-based profiling – has been developed, each offering distinct insights across spatial and molecular scales [1–6]. Nonetheless, a fundamental challenge persists: most available methods are inherently invasive and/or require molecular labelling.

These constraints become particularly limiting as scientific research shifts toward complex, three-dimensional (3D) tissue models, such as organoids. Over the past decade, neural organoids derived from pluripotent stem cells have emerged as powerful biological models of the developing human brain, recapitulating key features of early developmental dynamics and organisation [7–10]. Compared to two-dimensional (2D) neural cell culture models, neural organoids exhibit enhanced complexity

[*]These authors contributed equally to this work.
[†]Present address: University of Liverpool, UK
[#]Corresponding authors.

Accepted at AI for Science workshop (NeurIPS 2025).

and three-dimensional structures, providing a more advanced representation of tissue architecture[7,10], cellular diversity[11,12] and functional network formation[13]. Consequently, neural organoids have spurred new opportunities for studying human brain development, disease and drug response *in vitro*[14–17]. Yet, despite growing interest, organoid research remains constrained by the limitations of conventional methods for tissue imaging and characterisation. There is a pressing need for techniques that can interrogate spatial organisation and dynamics in intact organoid systems under physiologically relevant settings, without prior knowledge of specific molecular targets or the introduction of exogenous labels.

Currently, immunofluorescence staining combined with fluorescence microscopy is widely used as a gold standard, reference technique for visualising the morphological organisation of 2D and 3D cellular models, owing to its high specificity and spatial resolution[18,19]. In this approach, fluorophore-conjugated antibodies are used to selectively bind to and visualise specific target molecules[20]. Despite its strengths, however, this technique has several inherent limitations. First, while protein labelling is generally well-established, detecting other biomolecular classes, such as carbohydrates, lipids and various small molecules (e.g. drugs and metabolites), often requires more specialised probes or labelling strategies, or is not possible altogether. Additionally, immunofluorescence staining involves multiple washing and permeabilisation steps, which can disrupt cellular architecture and compromise membrane integrity, potentially leading to artefactual results[21]. These steps may also displace small molecules such as drugs, metabolites and other transient compounds, obscuring their true spatial localisation *in situ* within cells and tissues. While other labelling strategies, such as fluorescent protein tagging, genetic modifications and dye-based staining, may help avoid these washing steps, they remain label-based (i.e. targets must be known and available), and issues with labelling non-proteins persist. Furthermore, regardless of the labelling strategy, the intrinsic spectral overlap of fluorescent markers constrains multiplexing, typically to four targets in standard fluorescence microscopy[22], or seven in more advanced techniques[23,24]. Moreover, genetic modifications used in live-cell imaging to express fluorescent protein reporters can be costly, time-consuming, and may induce poorly understood side effects, such as protein aggregation or disruption of ion transport dynamics. Enabling non-invasive, label-free analysis would thus be essential to expand the potential of neural organoids in neurodevelopmental research, disease modelling, drug screening and beyond.

Raman spectroscopy (RS), a technique within vibrational spectroscopy, offers a promising alternative for biochemical analysis. Unlike fluorescence-based imaging, RS provides molecular contrast through the analysis of the intrinsic inelastic light scattering of molecules, thereby enabling non-invasive, label-free chemical characterisation with minimal sample preparation[25–27]. Consequently, RS has become a ubiquitous analytical technique across the life and physical sciences[28–38], with increasing use in analysing biological specimens such as cells and tissues[39–44]. In the context of neurodevelopment, RS has been applied to profile neural cells at various stages of differentiation. Yet, most studies to date have focused on 2D cultures of individual cell types with RS measurements acquired at a single-cell level – e.g. distinguishing neural stem cells from neurons by identifying distinct cell-specific spectral signatures[45–49]. The application of RS to complex three-dimensional cultured tissues containing multiple cell types, such as neural organoids, remains challenging due to their high biochemical heterogeneity, which hinders signal unmixing and interpretation. Recently, Bruno et al. explored the use of supervised machine learning to classify RS spectra from cortical organoids at four predefined maturation time points (weeks 6, 12, 16 and 20)[50]. However, their approach focused on single-point prediction and did not explore the three-dimensional architecture, cellular interactions and functional complexity intrinsic to neurodevelopment. So far, the use of RS for spatially resolved analysis of neural organoid structure and maturation remains unexplored.

Our lab recently reported a technique based on RS for 3D biochemical imaging and drug distribution analysis in liver organoids[51]. The proposed methodology relied on the acquisition of 'tissue phantoms' – i.e. calibrated reference spectra from biomolecules of interest, which were used to guide the estimation of chemical concentration maps across the scanned area using ordinary least squares with non-negativity constraint[52]. This enabled the detection and visualisation of key biomolecules and drugs *in situ* in whole liver organoids. Nonetheless, the approach has two limitations. First, it assumes that target biomolecules are known, thereby restricting potential applications and undermining the label-free capabilities of RS. Second, the performance and robustness of ordinary least squares for chemical estimation can be compromised by complex mixtures, experimental noise and artefacts, and variations introduced during data calibration and preprocessing, leading to inaccurate qualitative and quantitative analysis. To address these limitations, we recently introduced a deep learning framework

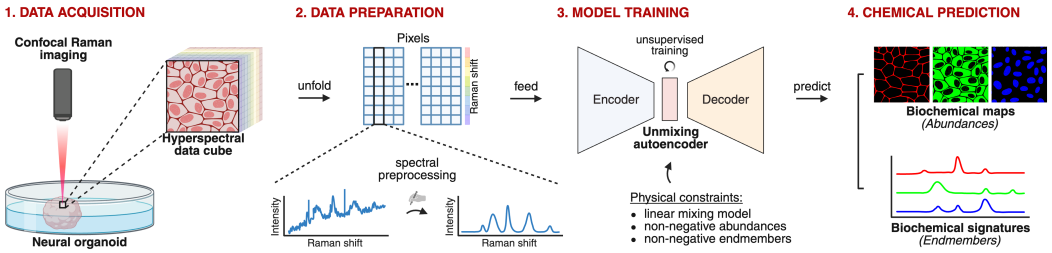

Figure 1: Schematic diagram of our organoid imaging pipeline combining Raman microspectroscopy and deep learning-enhanced hyperspectral analysis.

for RS chemometrics, using hyperspectral unmixing autoencoder neural networks, which enabled fully unsupervised, label-free (bio)chemical analysis with enhanced accuracy compared to established chemometric techniques[53].

Building on this work, we present here a pipeline for non-invasive, label-free biochemical imaging of neural organoids based on spontaneous Raman microspectroscopy and unsupervised autoencoder-based hyperspectral unmixing analysis. We show that our approach achieves improved organoid imaging compared to standard methods for RS analysis, as validated against fluorescence microscopy, and enables detailed morphochemical imaging and spatially resolved analysis in both cryosectioned and intact neural organoids with subcellular resolution.

2 Deep learning-enhanced Raman microspectroscopy

We begin by discussing the technical methodology behind our organoid imaging pipeline based on Raman microspectroscopy combined with autoencoder-based hyperspectral unmixing. Figure 1 provides an overview of our approach.

Raman measurements were collected from fixed neural organoid samples by performing point-wise imaging scans across predefined regions of interest. Volumetric scans were generated by acquiring sequential scans along the axis perpendicular to the imaging plane (e.g. sampling at different depths along the z-axis). Acquired data were preprocessed to remove non-specific signal contributions and spectral artefacts, such as cosmic spikes, autofluorescence baselines and dark noise (see Appendix A). This was done using RamanSPy – an open-source Python toolbox for RS data analysis we recently developed[54].

After preprocessing, hyperspectral unmixing was performed to derive constituent spectral components (endmembers) and quantify their relative contributions (abundances) in each measurement[55,56]. This was achieved using a custom autoencoder neural network model designed for hyperspectral unmixing. The model builds upon our previously reported approach[53], with substantial extensions and optimisations across the entire pipeline, spanning preprocessing, network architecture, training strategy and endmember estimation (see Appendix A for more details). Briefly, the autoencoder architecture included: (i) an encoder responsible for abundance estimation, which comprised a multi-branch spectral convolutional block followed by a five-layer fully connected network, and (ii) a one-layer decoder responsible for endmember identification. Our model incorporated relevant physical constraints in the representation learning process, namely, constraining the learnt endmembers and abundances to be non-negative through the use of appropriate activation functions and weight clipping, and constraining the mixing process to linear mixing through the use of a single-layer linear decoder. Models were trained on individual scans or datasets in an unsupervised manner by minimising a reconstruction loss based on the weighted sum of the mean squared error and spectral angle divergence[57] between input spectra and output reconstructions. The number of endmembers to extract for each scan/dataset was determined using principal component analysis (PCA)[58,59].

After model training, learnt endmember signatures and their corresponding abundances were derived, generating multi-channel abundance images/volumes with channels rendering chemical maps of the distribution of learnt endmembers. The obtained endmembers and abundances were examined to filter out components linked to non-tissue signals, such as background and noise. The remaining tissue-specific endmembers were characterised via peak assignment, whereby prominent peaks were

linked to chemical bonds to identify essential biomolecular species, such as nucleic acids, proteins, lipids and carbohydrates. Peak identification was guided by established reference materials[40]. The corresponding abundance maps were visualised as morphochemical images showing the spatial distribution of the obtained chemical components to study organoids' morphology and biochemical composition.

Further technical details regarding data acquisition, preprocessing and unmixing analysis are presented in Appendix A.

3 Experimental validation against fluorescence microscopy

To validate that our method produces accurate morphochemical organoid reconstructions, we evaluated its performance against fluorescence microscopy and standard methods for hyperspectral RS analysis. To this end, defined tissue regions within two day-30 neural organoid cryosections were imaged using confocal Raman microspectroscopy, followed by immunostaining and fluorescence microscopy (Figure 2A). Raman imaging was performed first to avoid potential signal interference from staining molecules. To ensure morphological diversity, we selected areas containing both distinct neural rosette structures and surrounding neural tissue for downstream imaging (Figure 2B). Raman imaging was performed over areas of $600\,\mu\mathrm{m}\times600\,\mu\mathrm{m}$ and $400\,\mu\mathrm{m}\times400\,\mu\mathrm{m}$ from the two respective sections, with a pixel size of $2\,\mu\mathrm{m}$ (Figures 4C and 5C in Appendix B). Subsequently, the organoid sections were stained for nuclei (DAPI), cell membranes (DiO), neural progenitor cells (SOX2) and neurons (TUJ1), and fluorescence microscopy images were taken to validate and correlate with organoid morphology predicted by RS (Figures 4B and 5B in Appendix B).

Using our autoencoder-based imaging pipeline, we analysed the acquired RS data from the first organoid section to infer its biochemical composition. Our method extracted four tissue-specific endmembers, which we associated with contributions from lipids, nucleic acids and proteins (Figure 4D–E in Appendix B). Figure 2C shows a side-by-side comparison between the predicted abundance maps of the two endmembers related to nucleic acid-rich and lipid-rich species and the corresponding fluorescence microscopy images of nuclei (DAPI) and cell membranes (DiO), respectively. We observed that our RS imaging pipeline accurately reconstructed individual nuclei, aligning with the DAPI staining across the imaged tissue (Figure 2C (right), and Figure 4F in Appendix B). It also effectively captured important membrane features, including both fine cellular membranes and prominent structures such as the apical domain of the visible neural rosette. Qualitatively, we observed strong correspondence between the lipid-rich component obtained with RS and the DiO signal around the apical domain of the rosette and with the TUJ1 signal outside the rosette (Figure 4B in Appendix B).

Having shown that our method accurately reconstructs key cellular components in cryosectioned neural organoid samples, we used the data collected from the second organoid section to compare the unmixing performance of our developed approach against a range of standard chemometric methods for multivariate analysis. This included methods for dimensionality reduction (principal component analysis (PCA)[58,59], non-negative matrix factorisation (NMF)[60]), clustering (k-means clustering[61,62]), and hyperspectral unmixing (N-FINDR[63] and vertex component analysis (VCA)[64] for endmember identification, applied together with non-negative least squares (NNLS)[52] for abundance estimation). Our autoencoder-based pipeline derived six tissue-specific endmembers, which we characterised as lipid-, nucleic acid- and protein-rich species (Figure 5D in Appendix B). Visual examination of the estimated abundance maps revealed major cellular and subcellular features, including membranes, nuclei, cytoplasmic regions and apical domains of neural rosettes (Figure 5E in Appendix B). Compared to standard methods, our approach offered improved robustness to artefacts and noise and produced more discriminative and interpretable abundance maps, enabling better-delineated imaging reconstruction of organoid morphology (Figure 6 in Appendix B).

To quantitatively assess the performance of our autoencoder-based pipeline, we evaluated the image similarity between Raman-based reconstructions and fluorescence microscopy images in an overlapping organoid region. To this end, we spatially aligned both modalities and computed the Pearson's correlation coefficient (Pearson) and structural similarity index measure (SSIM)[65] between RS-derived abundance maps of nuclei and membranes and fluorescence images of DAPI and DiO, respectively. We benchmarked the performance of our method against two conventional chemometric approaches: (i) univariate analysis, and (ii) multivariate hyperspectral unmixing using VCA and

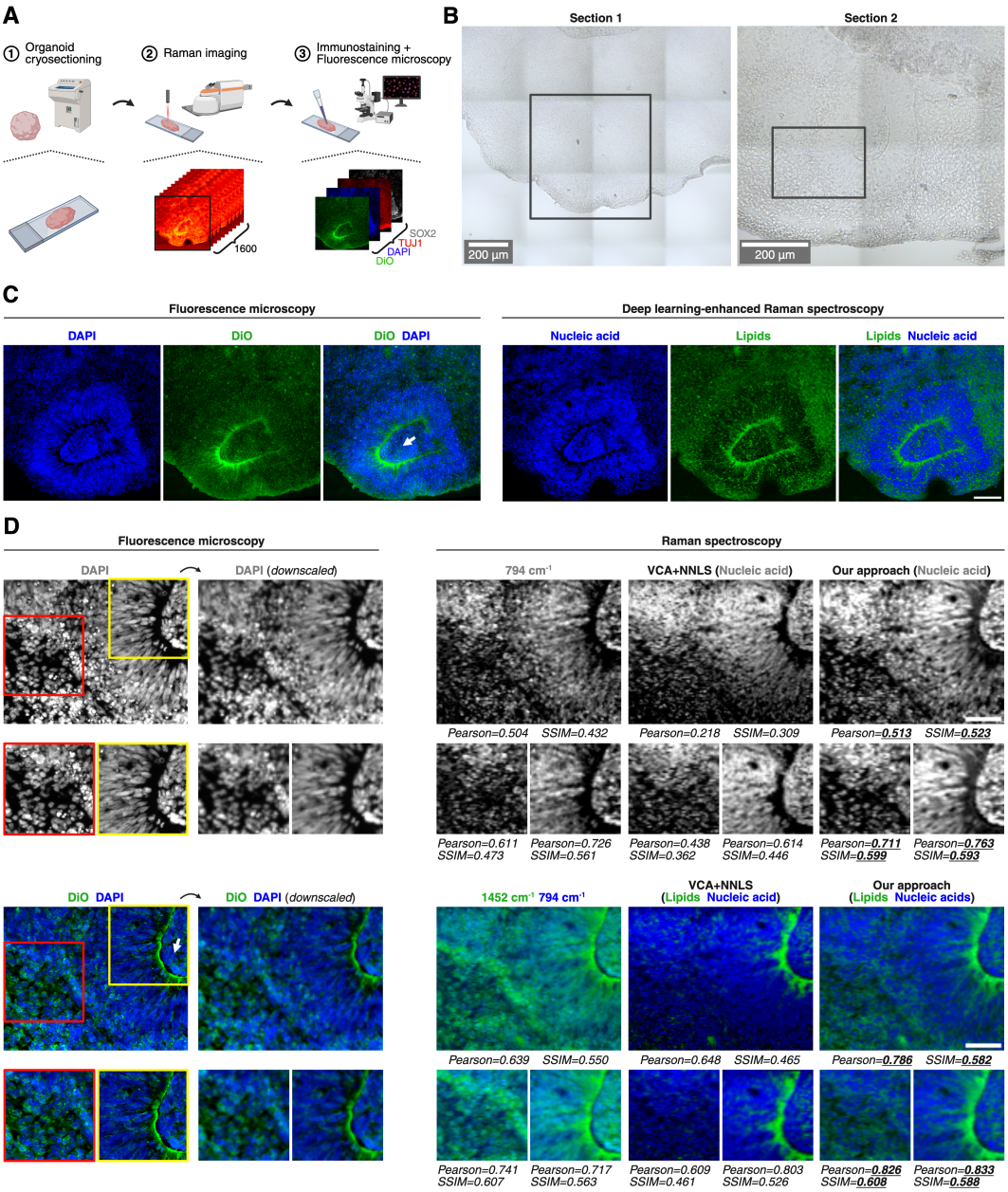

Figure 2: Deep learning-enhanced Raman microspectroscopy enables accurate morphochemical imaging of neural organoids as validated against fluorescence microscopy. **A,** Schematic diagram of our validation strategy, which involved the collection of paired Raman imaging and fluorescence microscopy data from neural organoid cryosections. **B,** Brightfield microscopy images of the two day-30 neural organoid cryosections imaged for validation. Marked areas show the regions used for validation in **C** and **D**, respectively. Scale bars: $200\,\mu\text{m}$. **C,** Imaging results on first organoid section. Comparison of fluorescence microscopy images of DAPI (nuclei) and DiO (membranes) (*left*), and RS-based abundance reconstructions of nucleic acid and lipids derived with our autoencoder pipeline (*right*). Arrow indicates the lumen of the present neural rosette. Scale bar: $100\,\mu\text{m}$. **D,** Imaging results on second organoid section. Comparison of fluorescence microscopy images of DAPI (nuclei) and DiO (membranes) (*left*), and RS-based abundance reconstructions of nucleic acid and lipids derived with three methods for Raman imaging, including univariate analysis, unmixing analysis with VCA and NNLS, and our autoencoder pipeline. Similarity metrics, namely Pearson's correlation coefficient (Pearson) and structural similarity index measure (SSIM), were performed with respect to the respective downscaled fluorescence images. The highest achieved similarity scores are indicated by underlined bold text. Arrow indicates the lumen of the present neural rosette. Scale bar: $50\,\mu\text{m}$.

NNLS. Univariate analysis was performed using the intensity profiles at $794\,\mathrm{cm}^{-1}$ and $1454\,\mathrm{cm}^{-1}$, which we found to be indicative of nuclear and lipid-rich membrane regions, respectively (Figure 5D in Appendix B).

As shown in Figure 2D, our autoencoder-based pipeline achieved the highest reconstruction accuracy across all evaluations – both quantitatively, with superior Pearson and SSIM scores, and qualitatively, with more interpretable and biologically relevant output. Nuclear regions were recovered with higher accuracy, with visually better-defined nuclei and lower non-specific signal. Membrane-associated structures were also more clearly delineated, enabling the visualisation of both large-scale features such as apical domains and individual cellular membranes.

Furthermore, our approach demonstrated improved robustness to experimental noise, maintaining consistent reconstruction quality under reduced denoising conditions. In comparison, the performance of conventional methods deteriorated markedly, yielding noisier, less interpretable spectral components and abundance profiles, or failing to converge to a solution during unmixing altogether (Figure 7 in Appendix B).

4 Biochemical imaging of intact neural organoids at subcellular resolution

After demonstrating that our method achieves accurate subcellular imaging of 2D cryosectioned neural organoid tissue, we performed higher-resolution Raman imaging of a tissue region within an intact 3D neural organoid. The ability to resolve and visualise subcellular structures in intact organoid tissues in their native environment is critical for interrogating neurodevelopmental processes, such as nuclear positioning, polarity establishment and lumen formation, and studying their effect on brain tissue organisation, architecture and function.

To this end, we acquired a larger 400×346 Raman imaging scan taken with $1\,\mu\mathrm{m}$ pixel size from an intact day-10 neural organoid sample (Figure 3A). Using our Raman imaging pipeline, we analysed the data and identified six tissue-specific components. The spectral endmembers and corresponding abundance maps are presented in Figure 3B and Figure 3C, respectively. Endmember 1 (red) exhibited a lipid-rich profile with prominent peaks at $1077\,\mathrm{cm}^{-1}$ (C–C and C–O stretching, phospholipids), $1305\,\mathrm{cm}^{-1}$, $1448\,\mathrm{cm}^{-1}$ and $1657\,\mathrm{cm}^{-1}$. Its abundance map outlined membrane-like structures, along with more confined bright spots resembling lipid droplets. Endmember 2 (grey) was characterised by vibrations associated with both protein and lipid species, with peaks at $1000\,\mathrm{cm}^{-1}$, $1124\,\mathrm{cm}^{-1}$ (C–C and C–N stretching), $1249\,\mathrm{cm}^{-1}$, $1311\,\mathrm{cm}^{-1}$, $1454\,\mathrm{cm}^{-1}$ and $1654\,\mathrm{cm}^{-1}$. Its abundance was observed in regions partially overlapping with membranes, as well as in isolated domains, possibly indicating cytoplasmic or extracellular matrix compartments. Endmember 3 (blue) was marked by strong nucleic acid bands at $790\,\mathrm{cm}^{-1}$ and $1095\,\mathrm{cm}^{-1}$, alongside protein-associated bands ($1005\,\mathrm{cm}^{-1}$, $1249\,\mathrm{cm}^{-1}$, $1328\,\mathrm{cm}^{-1}$, $1455\,\mathrm{cm}^{-1}$ and $1659\,\mathrm{cm}^{-1}$). As in previous analyses, the abundance map revealed well-defined spherical compartments consistent with individual cell nuclei. Endmember 4 (green) presented a distinct protein-rich signature with peaks at $1005\,\mathrm{cm}^{-1}$, $1251\,\mathrm{cm}^{-1}$, $1454\,\mathrm{cm}^{-1}$ and $1662\,\mathrm{cm}^{-1}$, along with vibrations associated with ring breathing modes in nucleic acid bases at $1573\,\mathrm{cm}^{-1}$. The abundance map highlighted cellular-level compartments resembling cytoplasmic regions, as well as small punctate structures within nuclei with high signal intensity, potentially corresponding to biomolecular condensates. Endmember 5 (yellow) displayed a mixed profile comprising bands linked to proteins ($1259\,\mathrm{cm}^{-1}$, $1326\,\mathrm{cm}^{-1}$, $1453\,\mathrm{cm}^{-1}$ and $1662\,\mathrm{cm}^{-1}$) and nucleic acid ($1095\,\mathrm{cm}^{-1}$). This component appeared more diffusely distributed, with substantial spatial overlap with endmember 3. Endmember 6 (purple) presented a noisier signature, with carbohydrate-associated bands at $493\,\mathrm{cm}^{-1}$ (glycogen), $850\,\mathrm{cm}^{-1}$ (C–O–C skeletal mode, glycogen) and $938\,\mathrm{cm}^{-1}$ (C–C stretching), along with peaks linked to proteins and nucleic acids at $754\,\mathrm{cm}^{-1}$ and $1103\,\mathrm{cm}^{-1}$. The abundance overlapped with endmember 4, particularly around nuclear condensates, potentially reflecting glycoproteins within cell nuclei.

To further inspect the spatial reconstructions produced by our imaging pipeline, we examined magnified fields from selected regions within the imaged organoid tissue (Figures 3D and 3E). The resulting reconstructions revealed consistent, well-defined cellular features, which delineated individual cells and subcellular structures resembling nuclei, cytoplasmic regions, cellular membranes and nuclear condensates.

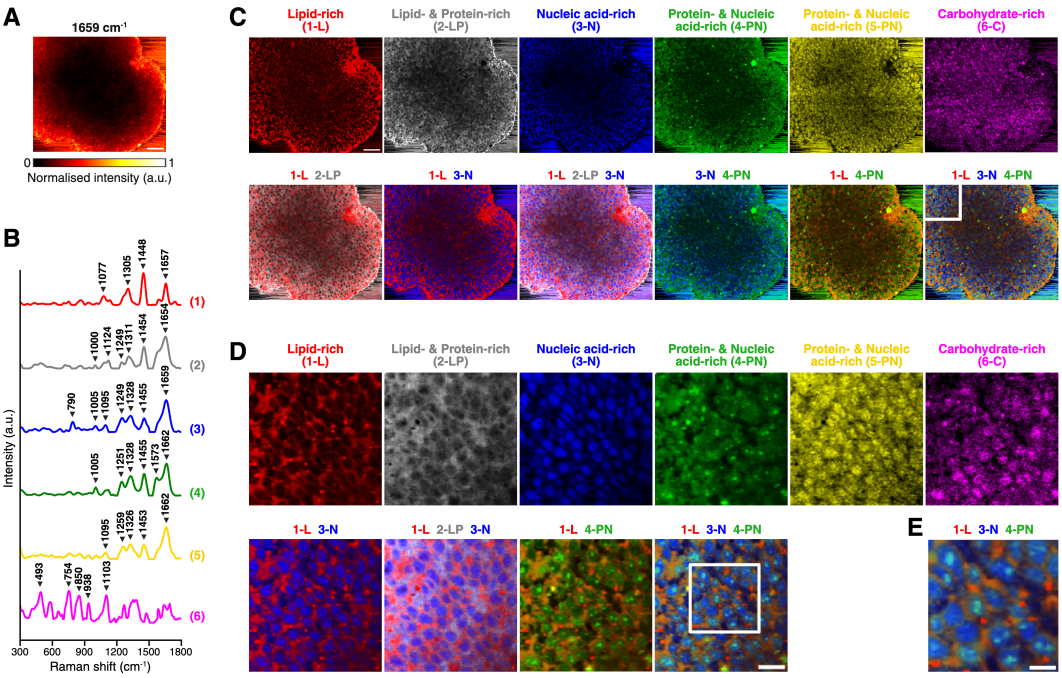

Figure 3: Biochemical imaging of intact neural organoids at subcellular resolution. **A,** Snapshot of collected Raman imaging data from an intact day-10 neural organoid sample, showing the intensity profile at the $1659\,\mathrm{cm}^{-1}$ band (proteins and lipids). Scale bar: $50\,\mu\mathrm{m}$. **B–E,** Results of hyperspectral unmixing analysis. **B,** Derived tissue-specific endmember signatures. Endmembers scaled for visualisation purposes (maximum intensity set to 1). **C,** Corresponding abundance maps. Scale bar: $50\,\mu\mathrm{m}$. **D,** Zoomed-in abundance images of a selected area (area marked in **C**, bottom right panel). Scale bar: $20\,\mu\mathrm{m}$. **E,** Further zoomed-in abundance image (area marked in **D**). Scale bar: $10\,\mu\mathrm{m}$.

These results demonstrate the effectiveness of our imaging platform in intact organoid tissues, allowing us to resolve complex biochemical heterogeneity and interrogate key morphological features at cellular and subcellular scales in an unsupervised, label-free manner.

5 Discussion

We have developed here a method for non-invasive, label-free organoid imaging based on Raman microspectroscopy combined with unsupervised autoencoder-based hyperspectral analysis. We showed that our approach is effective at discerning key biochemical components in unlabelled Raman spectroscopy measurements, enabling the visualisation of structural and morphological features in both cryosectioned and intact neural organoids and at different scales and resolutions.

By eliminating the need for destructive preparation steps and external labels, our workflow offers simpler sample preparation that does not compromise organoid integrity, with potential for live-organoid analysis. This opens new avenues for research into early neurodevelopment, but also in developmental diseases and drug discovery, where our approach could allow studying disease aetiology in neural organoid disease models or visualising drug distribution *in situ*. We note that our pipeline can also be applied to other types of samples, including other types of organoids and tissues. Beyond imaging, it can also be used as a feature extraction technique in RS for downstream applications – e.g. organoid phenotyping, disease classification, drug response modelling.

Acknowledgments and Disclosure of Funding

The authors thank Dr Simon Vilms Pedersen for discussions, Kevin Giraldo Rodriguez for assistance with Raman imaging of an organoid cryosection, Ellena O'Keeffe for performing immunostaining of an organoid cryosection, and Dr Akemi Nogiwa Valdez for data management support. The authors

also thank the Oxford Organoid Hub (OOH) for providing imaging infrastructure. Cryosectioned organoids imaged by fluorescence microscopy and Raman microspectroscopy for correlative analysis were generated by D.R. in the Knoblich laboratory (Vienna, Austria). D.G. acknowledges support by UK Research and Innovation (UKRI Centre for Doctoral Training in AI for Healthcare, grant number EP/S023283/1). R.X. and M.M.S. acknowledge support from the Engineering and Physical Sciences Research Council (EP/P00114/1 and EP/T020792/1). D.R. and M.M.S. acknowledge support from UK Research and Innovation (UKRI) Innovate UK under the UK government's Horizon Europe guarantee (10145749) associated to a grant from the European Union's Horizon Europe research and innovation programme (101071203). M.M.S. and D.R. acknowledge support from the Advanced Research and Invention Agency (SCNI-PR01-P11). Á.F.-G. acknowledges support from the Schmidt Science Fellows, in partnership with the Rhodes Trust. M.B. acknowledges support from the Engineering and Physical Sciences Research Council (EP/N014529/1, funding the EPSRC Centre for Mathematics of Precision Healthcare at Imperial College London, and EP/T027258/1). M.M.S. acknowledges support from the Royal Academy of Engineering Chair in Emerging Technologies award (CiET2021\94). Figures were assembled in BioRender.

Competing Interests

M.M.S. invested in, consults for (or was on scientific advisory boards or boards of directors) and conducts sponsored research funded by companies related to the biomaterials field. The authors wish to acknowledge that X.Z. is affiliated with Supervision Medicine. The company has in no way been involved in the design, execution, or reporting of the research presented in this paper and has no ownership or rights to any data presented within this manuscript. All other authors declare they have no competing interests.

References

[1] John D Bancroft and Marilyn Gamble. *Theory and practice of histological techniques*. Elsevier health sciences, 2008.

[2] Paul W Harms, Timothy L Frankel, Myrto Moutafi, Arvind Rao, David L Rimm, Janis M Taube, Dafydd Thomas, May P Chan, and Liron Pantanowitz. Multiplex immunohistochemistry and immunofluorescence: a practical update for pathologists. *Modern Pathology*, 36(7):100197, 2023.

[3] Ben NG Giepmans, Stephen R Adams, Mark H Ellisman, and Roger Y Tsien. The fluorescent toolbox for assessing protein location and function. *Science*, 312(5771):217–224, 2006.

[4] Patrik L Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O Westholm, Mikael Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, 2016.

[5] Anjali Rao, Dalia Barkley, Gustavo S França, and Itai Yanai. Exploring tissue architecture using spatial transcriptomics. *Nature*, 596(7871):211–220, 2021.

[6] Yehudit Hasin, Marcus Seldin, and Aldons Lusis. Multi-omics approaches to disease. *Genome biology*, 18(1):83, 2017.

[7] Madeline A Lancaster, Magdalena Renner, Carol-Anne Martin, Daniel Wenzel, Louise S Bicknell, Matthew E Hurles, Tessa Homfray, Josef M Penninger, Andrew P Jackson, and Juergen A Knoblich. Cerebral organoids model human brain development and microcephaly. *Nature*, 501(7467):373–379, 2013.

[8] Madeline A Lancaster and Juergen A Knoblich. Organogenesis in a dish: modeling development and disease using organoid technologies. *Science*, 345(6194):1247125, 2014.

[9] Sergiu P Pașca. The rise of three-dimensional human brain cultures. *Nature*, 553(7689):437–445, 2018.

[10] Ilaria Chiaradia and Madeline A Lancaster. Brain organoids for the study of human neurobiology at the interface of in vitro and in vivo. *Nature Neuroscience*, 23(12):1496–1508, 2020.

[11] Giorgia Quadrato, Tuan Nguyen, Evan Z Macosko, John L Sherwood, Sung Min Yang, Daniel R Berger, Natalie Maria, Jorg Scholvin, Melissa Goldman, Justin P Kinney, et al. Cell diversity and network dynamics in photosensitive human brain organoids. *Nature*, 545(7652):48–53, 2017.

[12] Silvia Velasco, Amanda J Kedaigle, Sean K Simmons, Allison Nash, Marina Rocha, Giorgia Quadrato, Bruna Paulsen, Lan Nguyen, Xian Adiconis, Aviv Regev, et al. Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature*, 570(7762):523–527, 2019.

[13] Cleber A Trujillo, Richard Gao, Priscilla D Negraes, Jing Gu, Justin Buchanan, Sebastian Preissl, Allen Wang, Wei Wu, Gabriel G Haddad, Isaac A Chaim, et al. Complex oscillatory waves emerging from cortical organoids model early human brain network development. *Cell stem cell*, 25(4):558–569, 2019.

[14] Xuyu Qian, Hongjun Song, and Guo-li Ming. Brain organoids: advances, applications and challenges. *Development*, 146(8):dev166074, 2019.

[15] Elizabeth Di Lullo and Arnold R Kriegstein. The use of brain organoids to investigate neural development and disease. *Nature Reviews Neuroscience*, 18(10):573–584, 2017.

[16] Oliver L Eichmüller and Juergen A Knoblich. Human cerebral organoids—a new tool for clinical neurology research. *Nature Reviews Neurology*, 18(11):661–680, 2022.

[17] Zehra Yildirim, Kyle Swanson, Xuekun Wu, James Zou, and Joseph Wu. Next-Gen Therapeutics: Pioneering Drug Discovery with iPSCs, Genomics, AI, and Clinical Trials in a Dish. *Annual Review of Pharmacology and Toxicology*, 65, 2024.

[18] Clara Brémond Martin, Camille Simon Chane, Cédric Clouchoux, and Aymeric Histace. Recent trends and perspectives in cerebral organoids imaging and analysis. *Frontiers in Neuroscience*, 15:629067, 2021.

[19] Rashmiparvathi Keshara, Yung Hae Kim, and Anne Grapin-Botton. Organoid imaging: seeing development and function. *Annual review of cell and developmental biology*, 38(1):447–466, 2022.

[20] Kyuseok Im, Sergey Mareninov, M Fernando Palma Diaz, and William H Yong. *An introduction to performing immunofluorescence staining*. Springer, 2019.

[21] Ulrike Schnell, Freark Dijk, Klaas A Sjollema, and Ben NG Giepmans. Immunolabeling artifacts and the need for live-cell imaging. *Nature methods*, 9(2):152–158, 2012.

[22] Lixue Shi, Mian Wei, Yupeng Miao, Naixin Qian, Lingyan Shi, Ruth A Singer, Richard KP Benninger, and Wei Min. Highly-multiplexed volumetric mapping with Raman dye imaging and tissue clearing. *Nature biotechnology*, 40(3):364–373, 2022.

[23] Hiromichi Tsurui, Hiroyuki Nishimura, Susumu Hattori, Sachiko Hirose, Ko Okumura, and Toshikazu Shirai. Seven-color fluorescence imaging of tissue samples based on Fourier spectroscopy and singular value decomposition. *Journal of Histochemistry & Cytochemistry*, 48(5):653–662, 2000.

[24] Francesco Cutrale, Vikas Trivedi, Le A Trinh, Chi-Li Chiu, John M Choi, Marcela S Artiga, and Scott E Fraser. Hyperspectral phasor analysis enables multiplexed 5D in vivo imaging. *Nature methods*, 14(2):149–152, 2017.

[25] Norman Colthup. *Introduction to infrared and Raman spectroscopy*. Elsevier, 2012.

[26] Richard L McCreery. *Raman spectroscopy for chemical analysis*. John Wiley & Sons, 2005.

[27] Ewen Smith and Geoffrey Dent. *Modern Raman spectroscopy: a practical approach*. John Wiley & Sons, 2019.

[28] Holly J Butler, Lorna Ashton, Benjamin Bird, Gianfelice Cinque, Kelly Curtis, Jennifer Dorney, Karen Esmonde-White, Nigel J Fullwood, Benjamin Gardner, Pierre L Martin-Hirsch, et al. Using Raman spectroscopy to characterize biological materials. *Nature protocols*, 11(4):664–687, 2016.

[29] Kosuke Dodo, Katsumasa Fujita, and Mikiko Sodeoka. Raman spectroscopy for chemical biology research. *Journal of the American Chemical Society*, 144(43):19651–19667, 2022.

[30] Dustin W Shipp, Faris Sinjab, and Ioan Notingher. Raman spectroscopy: techniques and applications in the life sciences. *Advances in Optics and Photonics*, 9(2):315–428, 2017.

[31] D Cialla-May, X-S Zheng, K Weber, and J Popp. Recent progress in surface-enhanced Raman spectroscopy for biological and biomedical applications: from cells to clinics. *Chemical Society Reviews*, 46(13):3945–3961, 2017.

[32] Kenny Kong, Catherine Kendall, Nicholas Stone, and Ioan Notingher. Raman spectroscopy for medical diagnostics—From in-vitro biofluid assays to in-vivo cancer detection. *Advanced drug delivery reviews*, 89:121–134, 2015.

[33] Katherine JI Ember, Marieke A Hoeve, Sarah L McAughtrie, Mads S Bergholt, Benjamin J Dwyer, Molly M Stevens, Karen Faulds, Stuart J Forbes, and Colin J Campbell. Raman spectroscopy and regenerative medicine: a review. *NPJ Regenerative medicine*, 2(1):12, 2017.

[34] Isaac Pence and Anita Mahadevan-Jansen. Clinical instrumentation and applications of Raman spectroscopy. *Chemical Society Reviews*, 45(7):1958–1979, 2016.

[35] Álvaro Fernández-Galiana, Olga Bibikova, Simon Vilms Pedersen, and Molly M Stevens. Fundamentals and Applications of Raman-Based Techniques for the Design and Development of Active Biomedical Materials. *Advanced Materials*, page 2210807, 2023.

[36] Amrit Paudel, Dhara Raijada, and Jukka Rantanen. Raman spectroscopy in pharmaceutical product design. *Advanced drug delivery reviews*, 89:3–20, 2015.

[37] Lynn R Terry, Sage Sanders, Rebecca H Potoff, Jacob W Kruel, Manan Jain, and Huiyuan Guo. Applications of surface-enhanced Raman spectroscopy in environmental detection. *Analytical Science Advances*, 3(3-4):113–145, 2022.

[38] Shelby R Khandasammy, Marisia A Fikiet, Ewelina Mistek, Yasmine Ahmed, Lenka Halámková, Justin Bueno, and Igor K Lednev. Bloodstains, paintings, and drugs: Raman spectroscopy applications in forensic science. *Forensic Chemistry*, 8:111–133, 2018.

[39] Zanyar Movasaghi, Shazza Rehman, and Ihtesham U Rehman. Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews*, 42(5):493–541, 2007.

[40] Abdullah Chandra Sekhar Talari, Zanyar Movasaghi, Shazza Rehman, and Ihtesham Ur Rehman. Raman spectroscopy of biological tissues. *Applied spectroscopy reviews*, 50(1):46–111, 2015.

[41] Christoph Krafft, Benjamin Dietzek, and Jürgen Popp. Raman and CARS microspectroscopy of cells and tissues. *Analyst*, 134(6):1046–1057, 2009.

[42] Thomas Huser and James Chan. Raman spectroscopy for physiological investigations of tissues and cells. *Advanced drug delivery reviews*, 89:57–70, 2015.

[43] Giuseppe Pezzotti. Raman spectroscopy in cell biology and microbiology. *Journal of Raman Spectroscopy*, 52(12):2348–2443, 2021.

[44] Rachael Smith, Karen L Wright, and Lorna Ashton. Raman spectroscopy: an evolving technique for live cell studies. *Analyst*, 141(12):3590–3600, 2016.

[45] Chia-Chen Hsu, Jiabao Xu, Bas Brinkhof, Hui Wang, Zhanfeng Cui, Wei E Huang, and Hua Ye. A single-cell Raman-based platform to identify developmental stages of human pluripotent stem cell-derived neurons. *Proceedings of the National Academy of Sciences*, 117(31):18412–18423, 2020.

[46] Adrian Ghita, Flavius C Pascut, Melissa Mather, Virginie Sottile, and Ioan Notingher. Cytoplasmic RNA in undifferentiated neural stem cells: a potential label-free Raman spectral marker for assessing the undifferentiated status. *Analytical chemistry*, 84(7):3155–3162, 2012.

[47] Adrian Ghita, Flavius C Pascut, Virginie Sottile, Chris Denning, and Ioan Notingher. Applications of Raman micro-spectroscopy to stem cell technology: label-free molecular discrimination and monitoring cell differentiation. *EPJ Techniques and Instrumentation*, 2:1–14, 2015.

[48] Junnan Geng, Wei Zhang, Cheng Chen, Han Zhang, Anhong Zhou, and Yu Huang. Tracking the differentiation status of human neural stem cells through label-free Raman spectroscopy and machine learning-based analysis. *Analytical Chemistry*, 93(30):10453–10461, 2021.

[49] Arno Germond, Yulia Panina, Mikio Shiga, Hirohiko Niioka, and Tomonobu M Watanabe. Following embryonic stem cells, their differentiated progeny, and cell-state changes during iPS reprogramming by Raman spectroscopy. *Analytical Chemistry*, 92(22):14915–14923, 2020.

[50] Giulia Bruno, Michal Lipinski, Koseki J Kobayashi-Kirschvink, Christian Tentellino, Peter TC So, Jeon Woong Kang, and Francesco De Angelis. Label-free detection of biochemical changes during cortical organoid maturation via Raman spectroscopy and machine learning. *Analytical Chemistry*, 97(9):5029–5037, 2025.

[51] Vernon LaLone, Aleksandra Aizenshtadt, John Goertz, Frøydis Sved Skottvoll, Marco Barbero Mota, Junji You, Xiaoyu Zhao, Henriette Engen Berg, Justyna Stokowiec, Minzhi Yu, et al. Quantitative chemometric phenotyping of three-dimensional liver organoids by Raman spectral imaging. *Cell Reports Methods*, 3(4), 2023.

[52] Charles L Lawson and Richard J Hanson. *Solving least squares problems*. SIAM, 1995.

[53] Dimitar Georgiev, Álvaro Fernández-Galiana, Simon Vilms Pedersen, Georgios Papadopoulos, Ruoxiao Xie, Molly M Stevens, and Mauricio Barahona. Hyperspectral unmixing for Raman spectroscopy via physics-constrained autoencoders. *Proceedings of the National Academy of Sciences*, 121(45):e2407439121, 2024.

[54] Dimitar Georgiev, Simon Vilms Pedersen, Ruoxiao Xie, Alvaro Fernández-Galiana, Molly M Stevens, and Mauricio Barahona. RamanSPy: An open-source Python package for integrative Raman spectroscopy data analysis. *Analytical Chemistry*, 96(21):8492–8500, 2024.

[55] Nirmal Keshava and John F Mustard. Spectral unmixing. *IEEE signal processing magazine*, 19(1):44–57, 2002.

[56] Xiaoli Li, Ruiqing Zhou, Yifei Xu, Xuan Wei, and Yong He. Spectral unmixing combined with Raman imaging, a preferable analytic technique for molecule visualization. *Applied Spectroscopy Reviews*, 52(5):417–438, 2017.

[57] Fred A Kruse, AB Lefkoff, JW Boardman, KB Heidebrecht, AT Shapiro, PJ Barloon, and AFH Goetz. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, 44(2-3):145–163, 1993.

[58] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572, 1901.

[59] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

[60] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[61] Stuart Lloyd. Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2):129–137, 1982.

[62] James MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, volume 5, pages 281–298. University of California press, 1967.

[63] Michael E Winter. N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *Imaging Spectrometry V*, volume 3753, pages 266–275. SPIE, 1999.

[64] José MP Nascimento and José MB Dias. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE transactions on Geoscience and Remote Sensing*, 43(4):898–910, 2005.

[65] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[66] Darren A Whitaker and Kevin Hayes. A simple algorithm for despiking Raman spectra. *Chemometrics and Intelligent Laboratory Systems*, 179:82–84, 2018.

[67] Paul HC Eilers. A perfect smoother. *Analytical chemistry*, 75(14):3631–3636, 2003.

[68] Paul HC Eilers and Hans FM Boelens. Baseline correction with asymmetric least squares smoothing. *Leiden University Medical Centre Report*, 1(1):5, 2005.

[69] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[70] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015.

[71] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[72] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[73] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[74] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: a system for large-scale machine learning. In *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, OSDI'16, page 265–283, USA, 2016. USENIX Association.

[75] Abraham Savitzky and Marcel JE Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.

A Extended methods

A.1 Raman microspectroscopy

Raman spectroscopy measurements were performed using a confocal Raman microscope (alpha300 R, WITec GmbH) equipped with a $532\,\mathrm{nm}$ solid-state laser (WITec GmbH). Laser power at the sample plane was set to about $48\,\mathrm{mW}$ for the two organoid cryosections or $32\,\mathrm{mW}$ for the intact sample. Backscattered light was collected through a water immersion objective – $20\times$ (Zeiss W N-Achroplan, N.A.$=0.5$) for organoid cryosections or $63\times$ (Zeiss W Plan-Apochromat, N.A.$=1$) for intact sample, and delivered via a $10\,\mu\mathrm{m}$ single-mode silica fiber to an imaging spectrograph (Newton, Andor Technology Ltd) with a 600 groove/mm grating and a thermoelectrically cooled CCD detector ($-60°$C). The system provided a spectral resolution of $\sim$ 2–3 cm^{-1} over the wavenumber range 0–3670 cm^{-1}. Spectra were acquired using an integration time of 0.5 seconds.

A.2 Spectral preprocessing of Raman imaging data

Acquired Raman spectroscopy measurements were exported as MATLAB files using the WITec Project FIVE Suite software and loaded in Python using RamanSPy[54]. Following this, each scan was individually preprocessed using RamanSPy as follows: (1) spectral cropping whereby wavenumbers less than 300 cm^{-1} were removed; (2) cosmic rays removal, using the algorithm reported by Whitaker and Hayes[66], with a kernel size $m = 3$ and z-value threshold $\tau = 6$; (3) denoising with a third-order Whittaker smoother[67] with $\lambda = 10^3$; (4) baseline correction with asymmetric least squares[68] with $\lambda = 10^5$; and (5) global vector normalisation, where intensity values were divided by the ℓ^2-norm of the spectrum with the highest norm in the given scan. Additionally, wavenumber calibration was performed on our cryosection scans by calculating the median position of the laser peak (Rayleigh scattering) and setting it to zero.

A.3 Hyperspectral unmixing analysis

Chemometric analysis was performed using a custom autoencoder neural network model designed for blind hyperspectral unmixing. In this context, Raman measurements $\mathbf{x}_i \in \mathbb{R}^b$ are treated as mixtures $\mathbf{x}_i = \mathcal{F}(M, \alpha_i)$ of a set of n unknown endmember components $M \in \mathbb{R}^{n \times b}$ based on their relative abundances $\alpha_\mathbf{i} \in \mathbb{R}^n$ in a given measurement $\mathbf{x}_i$.

To decompose a set of Raman measurements $\mathcal{X} \in \mathbb{R}^{m \times b}$ (i.e. a given scan) into endmember components M and their relative abundances $A \in \mathbb{R}^{m \times n}$, we train an autoencoder model $\mathcal{A}$ consisting of an encoder module $\mathcal{E}$ and a decoder module $\mathcal{D}$. The encoder was responsible for mapping input spectra $\mathbf{x}$ into latent space representations $\mathbf{z} = \mathcal{E}(\mathbf{x})$, and the decoder was responsible for mapping these latent representations into reconstructions of the original input $\widehat{\mathbf{x}} = \mathcal{D}(\mathbf{z}) = \mathcal{D}(\mathcal{E}(\mathbf{x})) = \mathcal{A}(\mathbf{x})$. The model was trained in an unsupervised manner by minimising the reconstruction error between the input $\mathbf{x}$ and the output $\widehat{\mathbf{x}}$. During this process, the model was guided to learn the endmember components M and their relative abundances A through the introduction of related physical constraints. Below, we provide additional details about the developed architecture and training procedure. For more information about hyperspectral unmixing, the reader is pointed to previous works by Keshava and Mustard[55] and Li et al.[56].

The encoder $\mathcal{E}$ comprised two separate blocks applied sequentially. The first part was a multi-branch convolutional block comprising four parallel convolutional layers with kernel sizes of 5, 10, 15 and 20, designed to capture patterns at multiple spectral scales. Each convolutional layer contained 32 filters with ReLU activation, He initialisation[69] and 'same' padding. Batch normalisation[70] and dropout with a rate of 0.2[71] were applied to each convolutional layer to improve training stability and generalisation. The outputs of the four convolutional layers were merged channel-wise through a fully connected layer to yield an output of dimension matching that of the input spectrum. The rationale behind this was to transform intensity values into representations that capture local spectral features (e.g. peak shape, width, local neighbourhood) and thus promote better generalisability. The second part of the encoder was a fully connected dimensionality reduction block, applied to learn patterns between the learnt spectral features. This block comprised a series of fully connected layers of sizes 256, 128, 64 and 32 with He initialisation and ReLU activation. Batch normalisation and dropout (rate of 0.5) were also applied at each fully connected layer. The block was followed by a final fully connected layer (Xavier uniform initialisation[72]) that reduced the final 32 features to a latent space of size n. The number n was treated as a hyperparameter that encodes the number of endmembers to extract, with latent representations treated as abundance fractions. To improve interpretation, non-negativity was enforced in the latent space using a 'softly-rectified' hyperbolic tangent function $f(x) = \frac{1}{\gamma} log(1 + e^{\gamma * tanh(x)})$, with $\gamma = 10$, as we previously reported[53]. This ensured that latent values were constrained to the range $(0, 1)$ in accordance with the abundance non-negativity constraint, but does not enforce abundances to sum to one (abundance sum-to-one constraint).

The decoder $\mathcal{D}$ consisted of a single fully connected layer (Xavier uniform initialisation) mapping the n-dimensional latent space representations back to the original spectral dimension b. In this layer, we used a linear activation and no bias term. Under this setup, the decoder mimics linear unmixing, where endmember signatures are encoded in the learnt weight matrix of the decoder layer[53]. A non-negative kernel constraint was used to enforce the non-negativity of endmember signatures using weight clipping during training. To accelerate model training, the weight matrix in the linear decoder

layer was initialised before training using endmember signatures derived with vertex component analysis[64].

Autoencoder models were trained for 5 epochs using the Adam optimiser[73] with a batch size of 32. The learning rate was set to 0.001. An exponential learning rate decay was used to ensure stable convergence, applying a factor of 0.9 to the learning rate after each epoch. The training loss $\mathcal{L}$ was defined as a weighted sum of the mean squared error (MSE) and spectral angle divergence (SAD) between input spectra $\mathbf{x}$ and output reconstructions $\widehat{\mathbf{x}}$:

$$\mathcal{L}(\mathbf{x}, \widehat{\mathbf{x}}) = \text{SAD}(\mathbf{x}, \widehat{\mathbf{x}}) + \lambda\, \text{MSE}(\mathbf{x}, \widehat{\mathbf{x}}), \tag{1}$$

where

$$\text{MSE}(\mathbf{x}, \widehat{\mathbf{x}}) = \frac{1}{b}||\mathbf{x} - \widehat{\mathbf{x}}||_2, \tag{2}$$

with b denoting the dimension of $\mathbf{x}$ (i.e. the number of spectral bands), and

$$\text{SAD}(\mathbf{x}, \widehat{\mathbf{x}}) = \arccos\left(\frac{\mathbf{x} \cdot \widehat{\mathbf{x}}}{\|\mathbf{x}\|_2\, \|\widehat{\mathbf{x}}\|_2}\right). \tag{3}$$

The weighting factor λ was set to 1000.

The number of endmembers to derive was estimated individually for each scan of interest using principal component analysis. Specifically, we computed the cumulative explained variance as a function of the number of principal components and manually selected the number of components at the point where the increase in explained variance visually plateaued.

Image similarity analysis between RS abundance images and fluorescence microscopy images was performed using Pearson's Correlation Coefficient (PCC) and Structural Similarity Index Measure (SSIM). Fluorescence images were first aligned to the corresponding RS images and then downscaled to match the RS image dimensions prior to analysis.

PCC was calculated as:

$$\text{PCC} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}}, \tag{4}$$

where x_i and y_i represent pixel intensities of the RS and fluorescence microscopy images, respectively, and $\bar{x}$, $\bar{y}$ are their mean intensities.

The Structural Similarity Index Measure (SSIM) was computed as:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{5}$$

where μ_x, μ_y denote means; σ_x^2, σ_y^2 variances; σ_{xy} covariance; and C_1, C_2 constants.

Autoencoder model implementation, training, and evaluation steps were carried out using TensorFlow[74]. Analysis with baseline chemometric methods was performed using RamanSPy[54].

B Extended results

(see the following four pages)

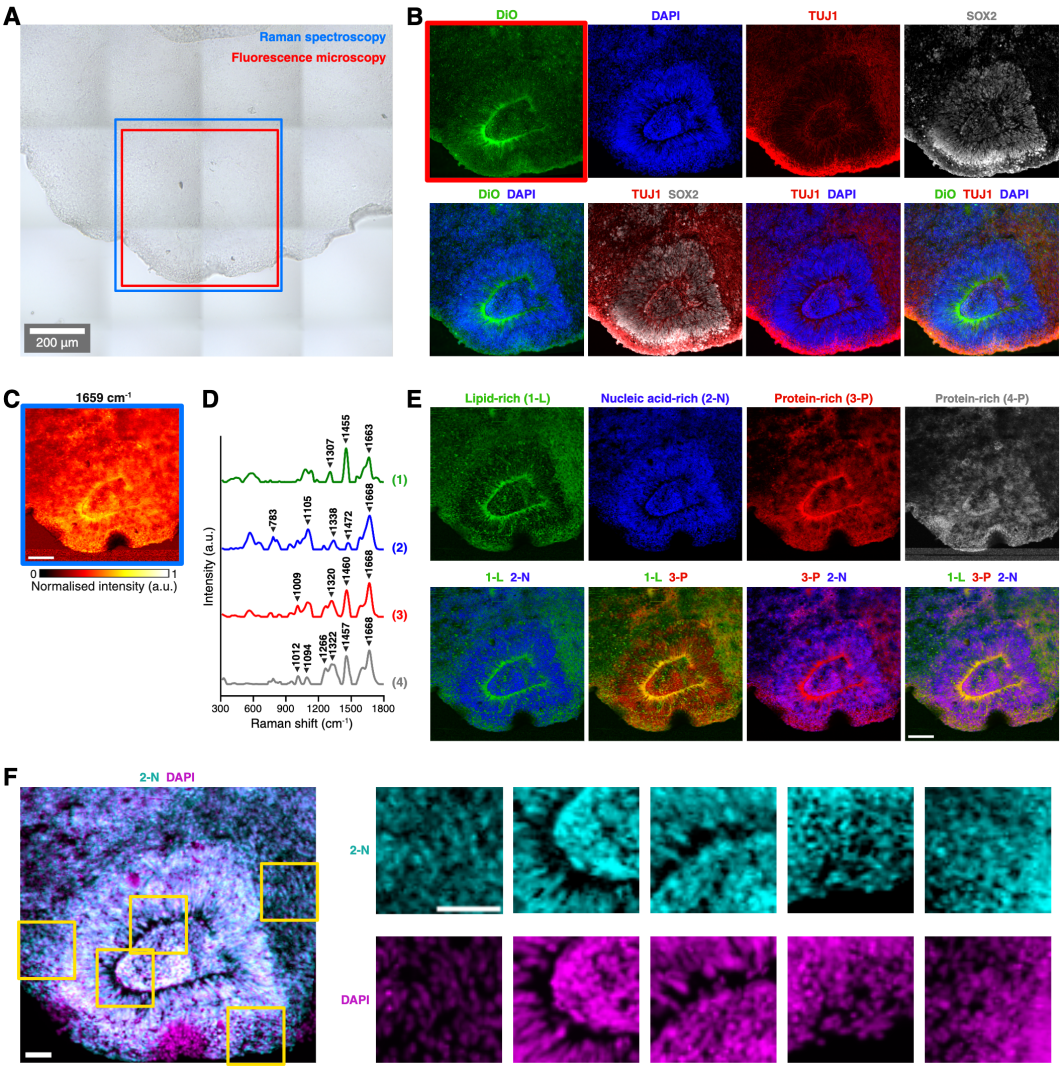

Figure 4: Extended results for Raman imaging scan of a first day-30 neural organoid cryosection. **A,** Brightfield image of a sectioned neural organoid tissue region imaged using Raman spectroscopy (blue) and fluorescence microscopy (red). Scale bar: $200\,\mu\mathrm{m}$. **B,** Fluorescence microscopy images with markers for nuclei (DAPI), membranes (DiO), neural progenitors (SOX2) and neurons (TUJ1). **C,** Raman intensity profile at the $1659\,\mathrm{cm}^{-1}$ band (linked to proteins and lipids). Image size: 300×300 pixels. Step size: $2\,\mu\mathrm{m}$. Scale bar: $100\,\mu\mathrm{m}$. **D–E,** Result of hyperspectral unmixing analysis. **D,** Derived tissue-specific endmember signatures. Endmembers scaled for visualisation purposes (maximum intensity set to 1). **E,** Corresponding abundance maps. Scale bar: $100\,\mu\mathrm{m}$. Endmember 1 (green) presented a lipid-rich spectral signature with distinct peaks at $1307\,\mathrm{cm}^{-1}$, $1455\,\mathrm{cm}^{-1}$, and $1663\,\mathrm{cm}^{-1}$. Its abundance was consistent with membrane structures, including cellular membranes and the apical domain of the visible neural rosette. The signal of this endmember showed a high degree of visual overlap with DiO and TOJ1. Endmember 2 (blue) displayed a nucleic acid-rich profile with a prominent band at $783\,\mathrm{cm}^{-1}$, alongside vibrations linked to proteins ($1338\,\mathrm{cm}^{-1}$, $1472\,\mathrm{cm}^{-1}$ and $1668\,\mathrm{cm}^{-1}$). Strong glass signal was also observed around $570\,\mathrm{cm}^{-1}$ and $1100\,\mathrm{cm}^{-1}$. The abundance map revealed well-defined, spherical compartments, closely resembling cell nuclei, with increased density within the present neural rosette structure. Endmembers 3 (red) and 4 (grey) shared similar spectral signatures, dominated by protein-associated peaks around $\sim 1010\,\mathrm{cm}^{-1}$, $\sim 1100\,\mathrm{cm}^{-1}$, $\sim 1265\,\mathrm{cm}^{-1}$, $\sim 1320\,\mathrm{cm}^{-1}$, $\sim 1460\,\mathrm{cm}^{-1}$, and $\sim 1670\,\mathrm{cm}^{-1}$. Both endmembers were generally broadly distributed across the sample. Endmembers 3 showed increased signal at the apical domain of the present rosette, potentially suggesting structural proteins. The abundance of endmember 4 was increased within the rosette structure. **F,** Overlaid images of DAPI and endmember associated with nucleic acids (2-N). Scale bars: $25\,\mu\mathrm{m}$.

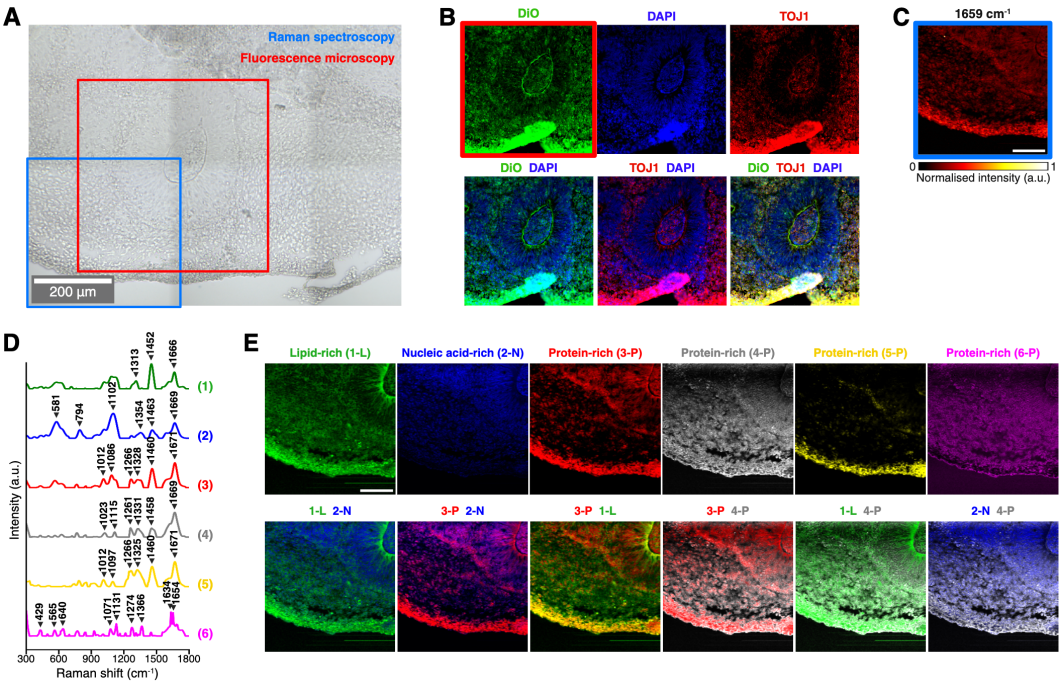

Figure 5: Extended results for Raman imaging scan of a second day-30 neural organoid cryosection. **A,** Brightfield image of a sectioned neural organoid tissue imaged using Raman spectroscopy (blue) and fluorescence microscopy (red). Scale bar: $200\,\mu\mathrm{m}$. **B,** Fluorescence microscopy images with markers for nuclei (DAPI), membranes (DiO) and neurons (TUJ1). Due to partial tissue detachment during the staining process, a localised increase in fluorescence intensity was observed at the tissue edge. This peripheral region was identified as an artefact and was therefore excluded from subsequent analyses. **C,** Raman intensity profile at the $1659\,\mathrm{cm}^{-1}$ band (linked to proteins and lipids). Image size: 200×200 pixels. Step size: $2\,\mu\mathrm{m}$. Scale bar: $100\,\mu\mathrm{m}$. **D–E,** Result of hyperspectral unmixing analysis. **D,** Derived tissue-specific endmember signatures. Endmembers scaled for visualisation purposes (maximum intensity set to 1). **E,** Corresponding abundance maps. Scale bar: $100\,\mu\mathrm{m}$. Endmember 1 (green) exhibited a lipid-rich profile with prominent peaks at $1313\,\mathrm{cm}^{-1}$, $1452\,\mathrm{cm}^{-1}$, and $1666\,\mathrm{cm}^{-1}$. The corresponding abundance map presented morphological architecture consistent with membrane-like structures, including individual cellular membranes and the apical domain of the present neural rosette. Endmember 2 (blue) displayed a strong nucleic acid band at $794\,\mathrm{cm}^{-1}$, alongside several vibrations linked to protein species ($1354\,\mathrm{cm}^{-1}$, $1463\,\mathrm{cm}^{-1}$ and $1669\,\mathrm{cm}^{-1}$), as well as bands attributed to the glass slide at ($581\,\mathrm{cm}^{-1}$ and $1102\,\mathrm{cm}^{-1}$). The abundance map revealed well-defined, spherical compartments, closely resembling cell nuclei. Endmembers 3 (red), 4 (grey) and 5 (yellow) exhibited similar spectral signatures, dominated by protein-associated peaks around $\sim 1015\,\mathrm{cm}^{-1}$, $\sim 1100\,\mathrm{cm}^{-1}$, $\sim 1265\,\mathrm{cm}^{-1}$, $\sim 1330\,\mathrm{cm}^{-1}$, $\sim 1460\,\mathrm{cm}^{-1}$, and $\sim 1670\,\mathrm{cm}^{-1}$. Endmember 4 showed a broad spatial distribution, likely corresponding to cytoplasmic or general background tissue signal. Endmembers 3 and 5, which showed a more pronounced $1460\,\mathrm{cm}^{-1}$ band, and an increased signal near the edge of the tissue section, likely due to tissue curling, as well as the basal domains of the present neural rosettes. Endmember 6 (magenta) presented a noisier signature, with prominent protein-associated peaks at $1131\,\mathrm{cm}^{-1}$ (C–N stretching), $1274\,\mathrm{cm}^{-1}$, and $1366\,\mathrm{cm}^{-1}$ (tryptophan). Its spatial distribution showed substantial overlap with that of Endmember 4.

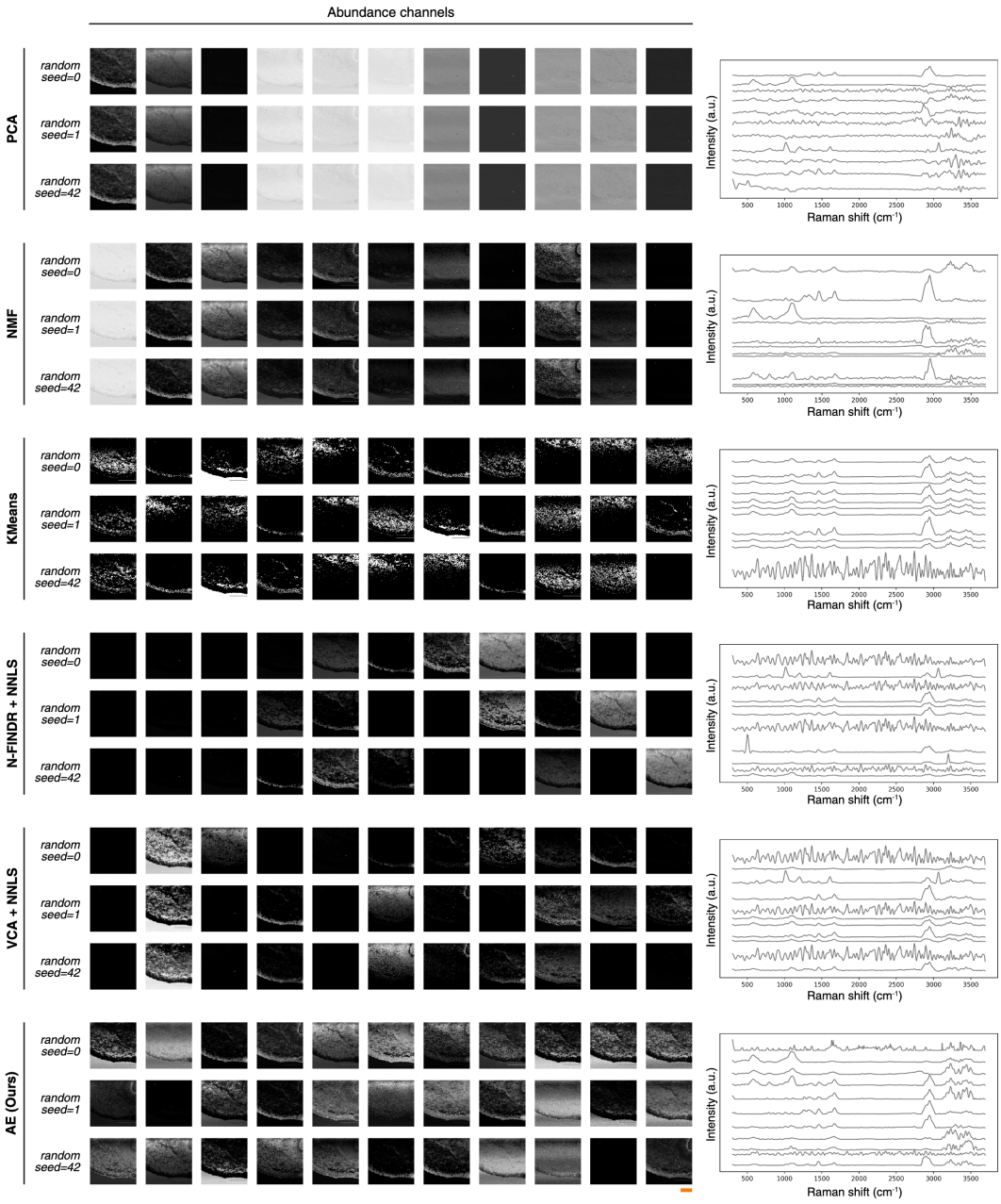

Figure 6: Comparative analysis of different chemometric methods on Raman imaging scan of a day-30 neural organoid cryosection. Methods include: principal component analysis (PCA), non-negative matrix factorisation (NMF), k-means clustering, N-FINDR + non-negative least squares (NNLS), vertex component analysis (VCA) + NNLS, our autoencoder (AE) pipeline. Abundance images were normalised for visualisation purposes. Endmember signatures displayed on the right correspond to the analyses performed with a random seed set to 42. Scale bar: $100\,\mu\mathrm{m}$ (bottom right, marked in orange).

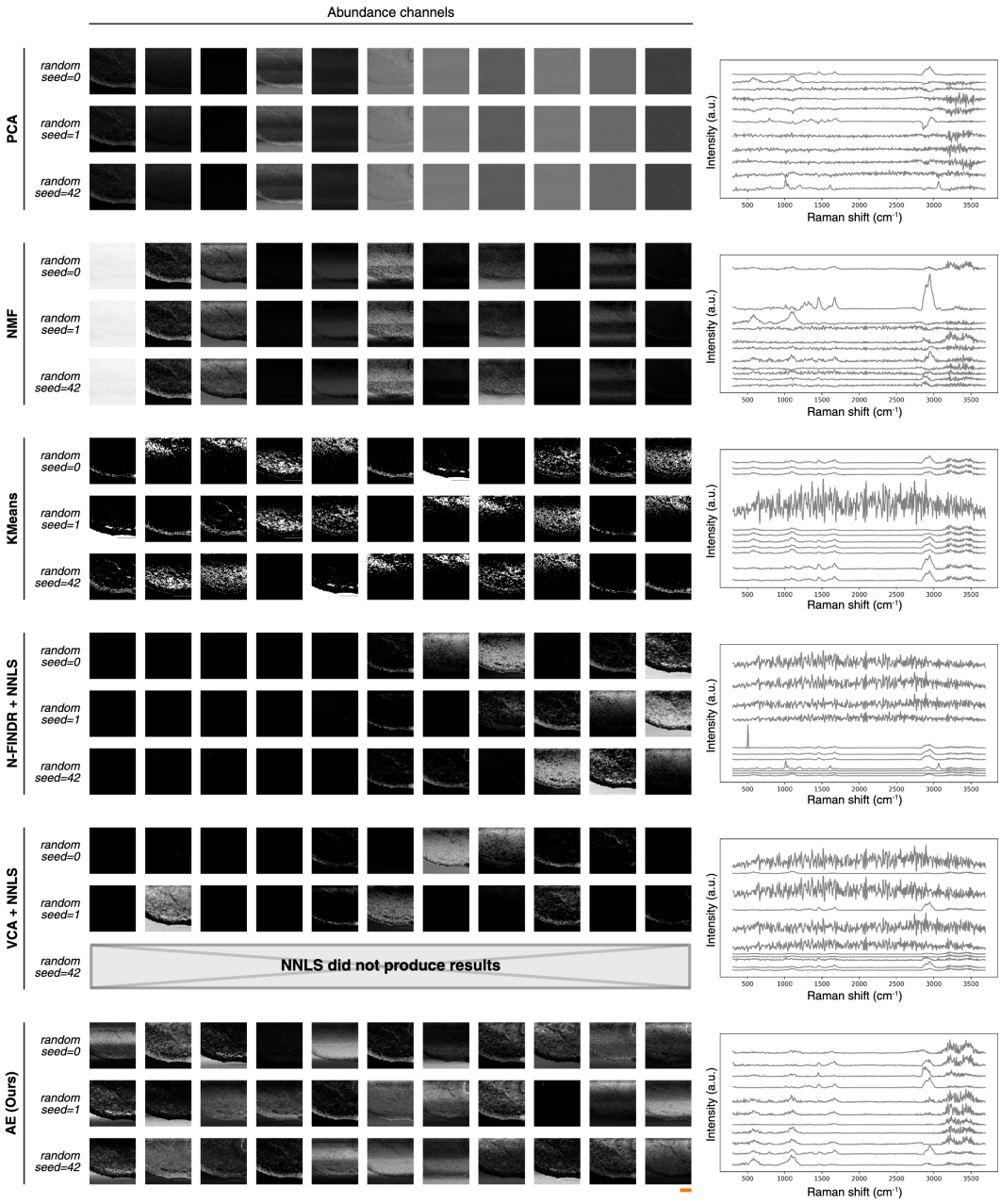

Figure 7: Comparative analysis of different chemometric methods on Raman imaging scan of a day-30 neural organoid cryosection in the presence of higher degrees of experimental noise. Higher noise levels were simulated by applying a more conservative denoising procedure with a second-order Savitzky-Golay filter with a window size of 7[75]. Methods include: principal component analysis (PCA), non-negative matrix factorisation (NMF), k-means clustering, N-FINDR + non-negative least squares (NNLS), vertex component analysis (VCA) + NNLS, our autoencoder (AE) pipeline. One of the analyses with VCA+NNLS failed to converge, with the NNLS implementation terminating after exceeding the maximum number of iterations. Abundance images were normalised for visualisation purposes. Endmember signatures displayed on the right correspond to the analyses performed with a random seed set to 42. Scale bar: $100\,\mu\mathrm{m}$ (bottom right, marked in orange).