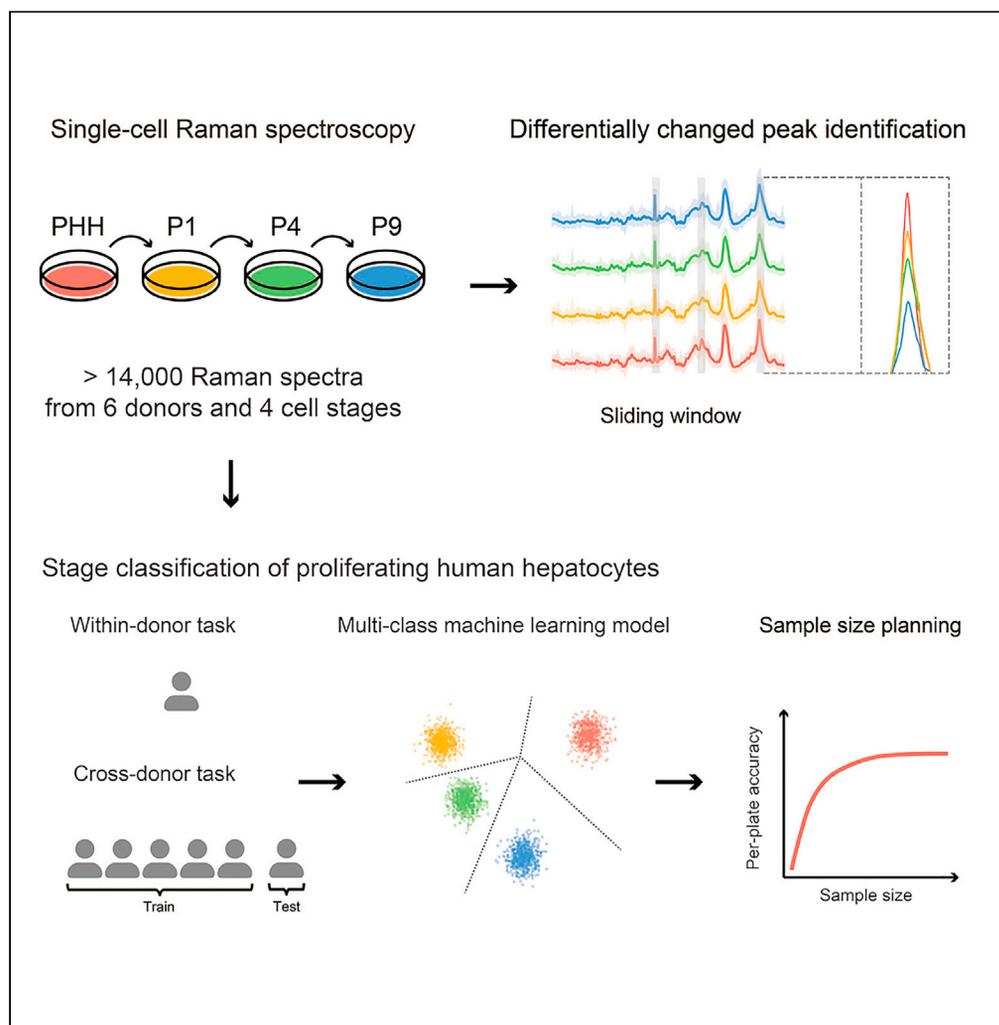


## Article

## Applying machine learning for multi-individual Raman spectroscopic data to identify different stages of proliferating human hepatocytes



Bihan Shen, Chen Ma, Lili Tang, Zhitao Wu, Zhaoliang Peng, Guoyu Pan, Hong Li

gypan@simm.ac.cn (G.P.)  
lihong01@sinh.ac.cn (H.L.)

**Highlights**

Repository of 14,000 Raman spectra from primary and proliferating hepatocytes

Potential biomarker identification by sliding window analysis

Accurate cell stage identification using machine learning models

Sample size planning for feasible in-process control in cell product manufacturing

## Article

## Applying machine learning for multi-individual Raman spectroscopic data to identify different stages of proliferating human hepatocytes

Bihan Shen,<sup>1,3,5</sup> Chen Ma,<sup>2,3,5</sup> Lili Tang,<sup>4</sup> Zhitao Wu,<sup>2,4</sup> Zhaoliang Peng,<sup>2,3</sup> Guoyu Pan,<sup>2,3,4,\*</sup> and Hong Li<sup>1,3,6,\*</sup>

## SUMMARY

Cell therapy using proliferating human hepatocytes (ProlIHs) is an effective treatment approach for advanced liver diseases. However, rapid and accurate identification of high-quality ProlIHs from different donors is challenging due to individual heterogeneity. Here, we developed a machine learning framework to integrate single-cell Raman spectroscopy from multiple donors and identify different stages of ProlIHs. A repository of more than 14,000 Raman spectra, consisting of primary human hepatocytes (PHHs) and different passages of ProlIHs from six donors, was generated. Using a sliding window algorithm, potential biomarkers distinguishing the different cell stages were identified through differential analysis. Leveraging machine learning models, accurate classification of cell stages was achieved in both within-donor and cross-donor prediction tasks. Furthermore, the study assessed the relationship between donor and cell numbers and its impact on prediction accuracy, facilitating improved quality control design. A similar workflow can also be extended to encompass other cell types.

## INTRODUCTION

Liver disease is a significant global health issue that leads to a serious economic burden and 2 million deaths annually worldwide.<sup>1,2</sup> Orthotopic liver transplantation is the only curative treatment for advanced liver failure, but the shortage of suitable donors and the long-term use of immunosuppressive drugs have impeded its clinical utility.<sup>3</sup> Hepatocyte transplantation provides an alternative strategy for treating some liver diseases with minimal invasiveness.<sup>4,5</sup> For example, acute liver failure, fulminant hepatitis, and liver-based metabolic disorders have been successfully treated by hepatocyte transplantation in preclinical and clinical research.<sup>6–13</sup> However, there are major bottlenecks for hepatocyte transplantation: limited sources, difficult expansion and maintenance, and poor quality of primary human hepatocytes (PHHs).

Recent studies have developed alternative sources of hepatocytes, including hepatic progenitor cells, directly differentiated embryonic stem cells (ESCs), induced pluripotent stem cells (iPSCs), and transdifferentiated adult cells.<sup>14–16</sup> Zhang et al. reported that proliferating human hepatocytes (ProlIHs) could proliferate continually for more than two months, resulting in a 10,000-fold expansion *in vitro*.<sup>17</sup> ProlIHs exhibit both hepatocyte-like and progenitor-like features and efficiently repopulate the damaged liver in FRG mice without tumor formation.<sup>18</sup> Harvesting high-quality ProlIHs via cell transplantation is crucial. Traditional techniques such as RT-qPCR and western blotting are time-consuming and require a large number of cells. While flow cytometry has the potential to accurately measure antigen quantities in individual cells, it requires specific labeling with antibodies or fluorescently labeled microbeads.<sup>19</sup> A faster and easier-to-operate method is needed to improve the quality control of ProlIHs.

Raman spectroscopy is a rapid, noninvasive, and label-free analytical technique.<sup>20</sup> The regions of the Raman spectrum can represent the states of carbohydrates, nucleic acids, proteins, lipids and other biomolecules.<sup>21</sup> Therefore, Raman spectroscopy is gaining popularity for the classification of biological samples, such as human pluripotent stem cell-derived neurons,<sup>22</sup> pathogens,<sup>23–26</sup> tumors,<sup>27–31</sup> and viruses.<sup>32,33</sup>

In our previous investigation, we successfully performed Raman profiling of PHHs and different passages of ProlIHs and found that Raman spectroscopy could track the biochemical changes occurring during the dedifferentiation process from PHHs to ProlIHs.<sup>34</sup> Considering the heterogeneity of individuals and cells, we expanded the number of human donors from one to six, expanded more passages of ProlIHs, collected Raman spectra of a larger number of cells, and conducted this research to construct a more accurate and robust machine learning model for classifying PHHs and ProlIHs.

<sup>1</sup>Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai 200031, China

<sup>2</sup>Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China

<sup>3</sup>University of Chinese Academy of Sciences, Beijing 100049, China

<sup>4</sup>School of Chinese Materia Medica, Nanjing University of Chinese Medicine, Nanjing 210023, China

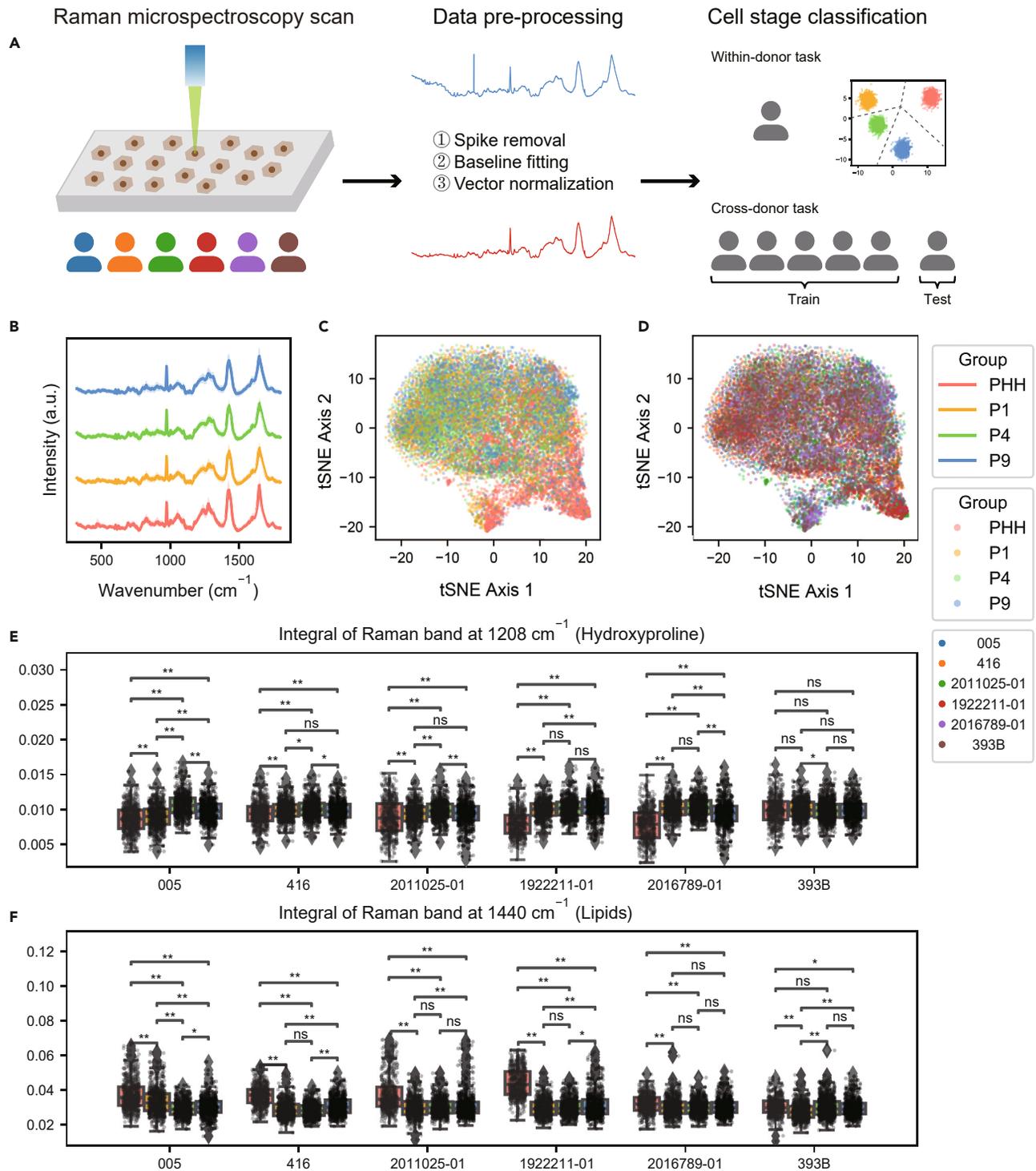
<sup>5</sup>These authors contributed equally

<sup>6</sup>Lead contact

\*Correspondence: [gypan@simm.ac.cn](mailto:gypan@simm.ac.cn) (G.P.), [lihong01@sinh.ac.cn](mailto:lihong01@sinh.ac.cn) (H.L.)

<https://doi.org/10.1016/j.isci.2024.109500>

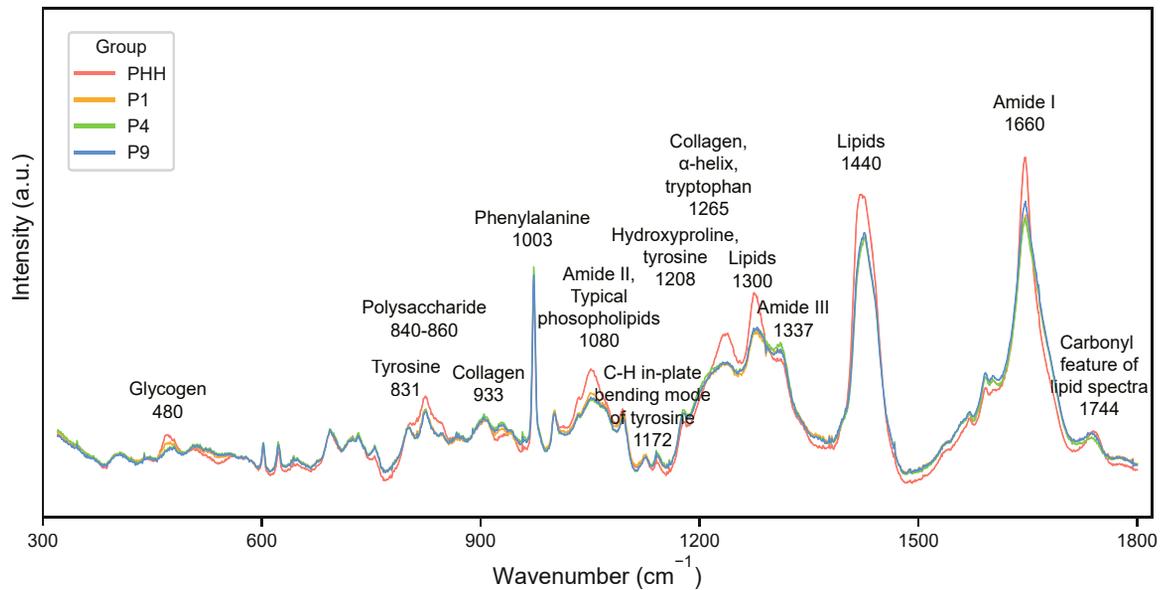




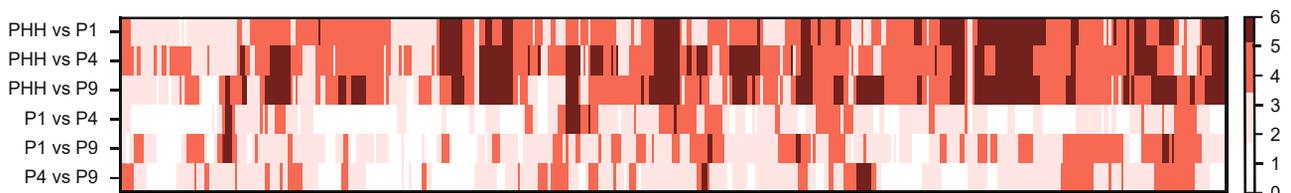
**Figure 1. Raman spectra of human primary hepatocytes and proliferating hepatocytes**

(A) Schematic illustration of the study, including Raman spectra measurements, data preprocessing, and classification model construction. (B) The average Raman spectra acquired from PHH, P1, P4, and P9 from six different donors. (C and D) Raman spectra were subjected to t-SNE visualization, where the colors represent different cell stages (C) and donor origins (D). (E and F) Raman band intensity comparisons between different cell stages among six donors were quantified for various biomolecular components, including  $1,208\text{ cm}^{-1}$  (hydroxyproline) (E) and  $1,440\text{ cm}^{-1}$  (lipids) (F). Student's t test with Bonferroni correction was used. Asterisks indicate the level of statistical significance: \* $p < 0.05$ , \*\* $p < 0.01$ , n.s., not significant.

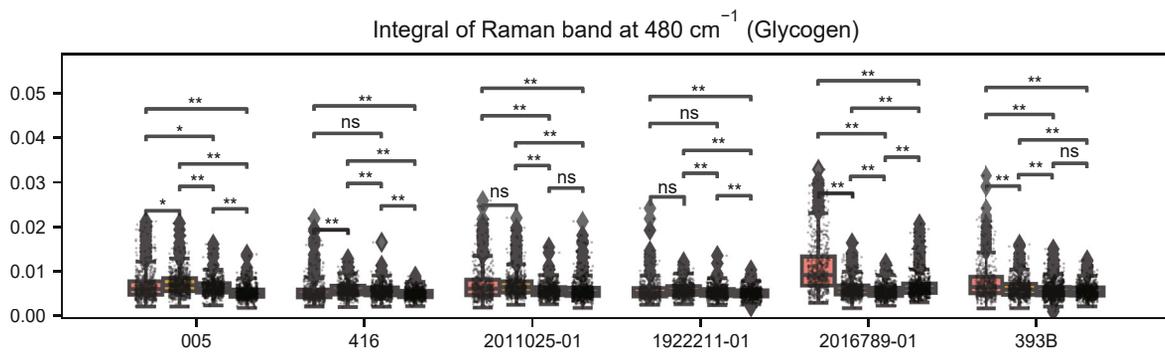
A



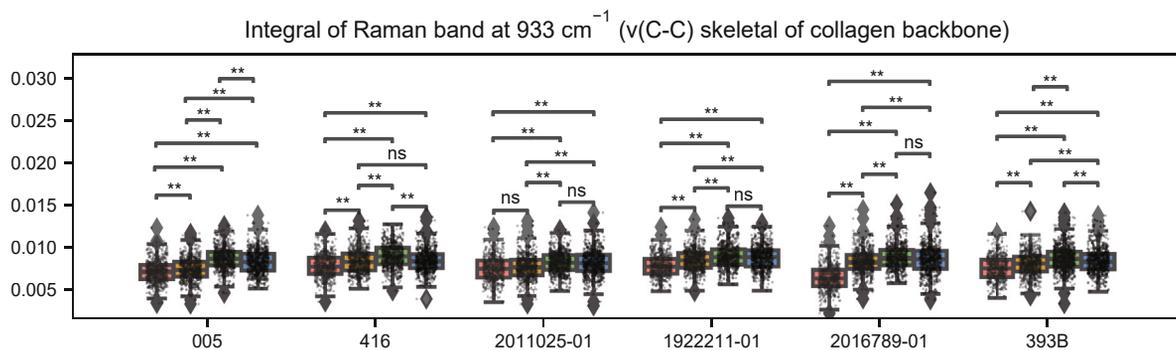
B



C



D



**Figure 2. Differentially changed Raman peaks between primary and proliferating hepatocytes**

(A) Average Raman spectra acquired from six donors, representing PHH, P1, P4, and P9. The spectra are annotated with Raman bands that are indicative of cell stage differences.

(B) Heatmap plots of the aggregated results of differentially changed peaks of any two cell stages, where the color indicates the number of donors that yielded a significant p value < 0.05 and an absolute effect size larger than 0.2.

(C and D) Raman band intensity comparisons between different cell stages among six donors were quantified for various biomolecular components, including 480 cm<sup>-1</sup> (glycogen) (C) and 933 cm<sup>-1</sup> (v(C-C) skeletal collagen backbone) (D). Student's t test with Bonferroni correction was used. Asterisks indicate the level of statistical significance: \*p < 0.05, \*\* p < 0.01, n.s., not significant.

In this study, we produced a large-scale repository of Raman spectra covering more than 14,000 PHHs and ProlIHs from six donors. Potential biomarkers of PHHs and ProlIHs were identified by differential analysis using a sliding window algorithm. Machine learning models were built to distinguish different cell stages in within-donor and cross-donor prediction tasks. Moreover, we estimated the relationship between the number of donors and cells and the prediction accuracy for better experimental design.

**RESULTS**

**Raman profiling of the ProlIHs**

To construct a workflow for determining whether PHHs are transformed into ProlIHs during cell culture, our study included three steps: Raman spectral measurements, data preprocessing, and classification model construction (Figure 1A). PHHs from six donors (Lot: 416, 393B, 005, 2011025-01, 2016789-01, and 1922211-01) were used to induce ProlIHs. PHHs, ProlIHs at passage 1 (P1), passage 4 (P4), and passage 9 (P9) were collected to derive Raman spectra. The fingerprint region ranging from 320 cm<sup>-1</sup> to 1,800 cm<sup>-1</sup> consisting of 2,214 Raman bands was used (Figure 1B), which covered the majority of the vibrational modes of biomolecules. At least 600 spectra were examined for PHH, P1, P4, and P9. After quality control, more than 14,000 spectra were obtained (Table 1). Further visualization of the Raman spectra by t-distributed stochastic neighbor embedding (t-SNE) is shown in Figures 1C and 1D. In terms of cell stage, PHH appeared to have greater homogeneity and discrepancy than other stages (Figure 1C). Moreover, the differences among donors were minimal (Figure 1D). This finding suggested that Raman spectra could distinguish different cell stages, especially PHHs from ProlIHs.

Our previous study of donor 005 confirmed that the Raman spectra change at 1,206 cm<sup>-1</sup> (hydroxyproline) and 1,440 cm<sup>-1</sup> (lipids) during the transition from PHH to ProlIHs. Here, we also assessed the aforementioned changes in other donors (Figures 1E and 1F). The intensity of the 1,206 cm<sup>-1</sup> hydroxyproline band significantly increased from PHH to P9 (Figure 1E). The Raman band intensity of lipids at 1,440 cm<sup>-1</sup> decreased significantly during the progression of ProlIHs proliferation (Figure 1F). This indicates that the differentially changed peaks are conserved among donors.

**Identification of potential biomarkers**

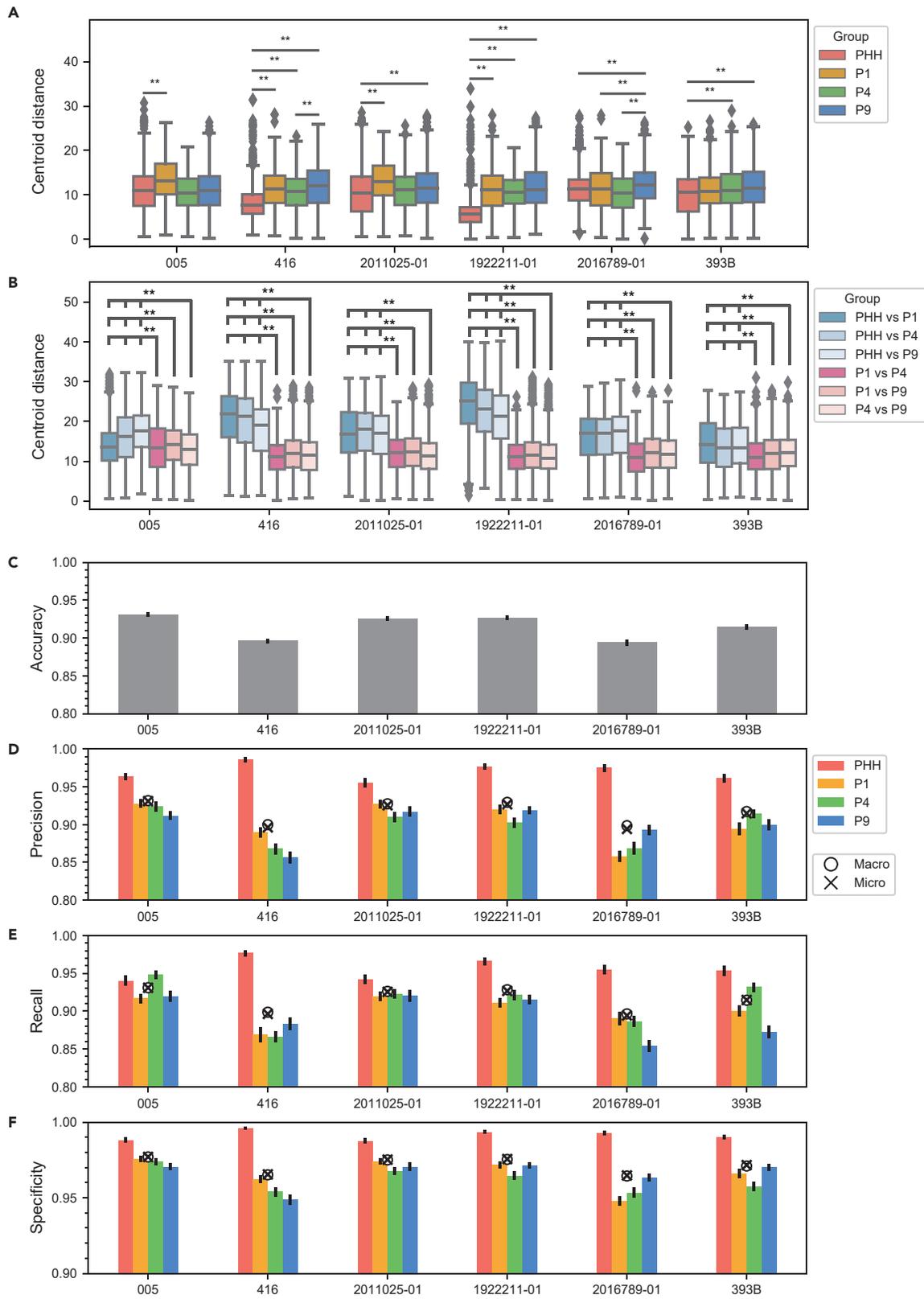
To automatically detect differentially changed peaks across cell stages, Raman spectra were processed by sliding windows with sum pooling, and spectra of any two cell stages were compared by Student's t test with Bonferroni correction. The aggregated results from different donors are shown in Figure 2B. After the assignment of the peaks, several differentially changed Raman bands with the corresponding biomolecules were identified (Figure 2A), including those at 480 cm<sup>-1</sup> (glycogen), 831 cm<sup>-1</sup> (tyrosine), 840–860 cm<sup>-1</sup> (polysaccharide), 933 cm<sup>-1</sup> (v(C-C) skeletal of collagen backbone), 1,003 cm<sup>-1</sup> (phenylalanine), 1,080 cm<sup>-1</sup> (amide II, typical phospholipids), 1,172 cm<sup>-1</sup> (C–H in plane bending mode of tyrosine), 1,206 cm<sup>-1</sup> (hydroxyproline), 1,265 cm<sup>-1</sup> (α-helix, collagen, tryptophan), 1,300 cm<sup>-1</sup> (lipids), 1,337 cm<sup>-1</sup> (amide III), 1,440 cm<sup>-1</sup> (CH2 and CH3 formation vibrations of lipids), 1,660 cm<sup>-1</sup> (amide I of proteins), and 1,744 cm<sup>-1</sup> (carbonyl feature of lipid spectra). Hepatocytes maintain glucose homeostasis through glycogen synthase and degradation. In particular, the integral of the peak at 480 cm<sup>-1</sup> decreased from P1 to P9, which implied that glycogen synthase was downregulated during cell dedifferentiation (Figure 2C). The band at 933 cm<sup>-1</sup> increased significantly from PHH to P9 (Figure 2D).

**Within-donor classification using machine learning models**

According to t-SNE plots, PHHs displayed greater homogeneity and discrepancy than other stages. To verify this phenomenon, the centroid distance was used to quantify the intra- and inter-stage heterogeneity. As shown in Figure 3A, a smaller centroid distance was observed within

**Table 1. Summary of Raman spectra of PHHs and different stages of ProlIHs collected from six human donors**

Donor	PHH	P1	P4	P9	Total
005	607	619	604	604	2434
2011025-01	508	609	608	604	2329
2016789-01	522	615	603	615	2355
1922211-01	503	601	603	594	2301
416	540	614	609	605	2368
393B	528	626	850	607	2611



**Figure 3. Within-donor classification between primary and proliferating hepatocytes**

(A and B) Intra- (A) and inter-stage (B) heterogeneity quantified using centroid distance. Intra-stage refers to cells within the same cell group, e.g., PHHs. Inter-stage refers to cells between two different cell groups, e.g., PHH and P1. (C–F) Performance of classification models separately by cells from each donor: accuracy (C), precision (D), recall (E), and specificity (F). Error bars represent 95% confidence intervals.

the PHHs than within the other cell stages, indicating less intra-stage heterogeneity of the PHHs. Moreover, the heterogeneity between PHHs and ProlIHs was significantly greater than the heterogeneity between different stages of ProlIHs (Figure 3B). These results indicate that the growth rates and states of cells became more diverse during the transition from PHHs to ProlIHs.

Next, machine learning models were constructed to classify different cell stages derived from the same donor. Since this is a multi-class task, one-vs-rest (OvR) and one-vs-one (OvO) strategies with different base models, i.e., support vector machine (SVM), random forest and neural network, were applied and evaluated using nested cross-validation with 10 repetitions.<sup>35</sup> The performance of OvO was greater than that of OvR regardless of the base model, indicating that OvO could better capture subtle differences among cell stages (Table S1). The SVM with OvO strategy achieved the best performance and was chosen for subsequent analysis. Figures 3C–3F shows the performance of the OvO SVM models for each donor. The overall accuracy ranged from 0.89 to 0.93 among the models of different donors. The micro-average and macro-average of multiple classes for each evaluation metric were nearly the same since the number of cells in multiple datasets was roughly balanced. The macro-precision and micro-precision ranged from 0.90 to 0.93, the macro-recall and micro-recall ranged from 0.90 to 0.93, and the macro-specificity and micro-specificity ranged from 0.96 to 0.98. Notably, performance fluctuated among different cell stages, with PHHs consistently achieving superior precision, recall, and sensitivity compared to other cell stages.

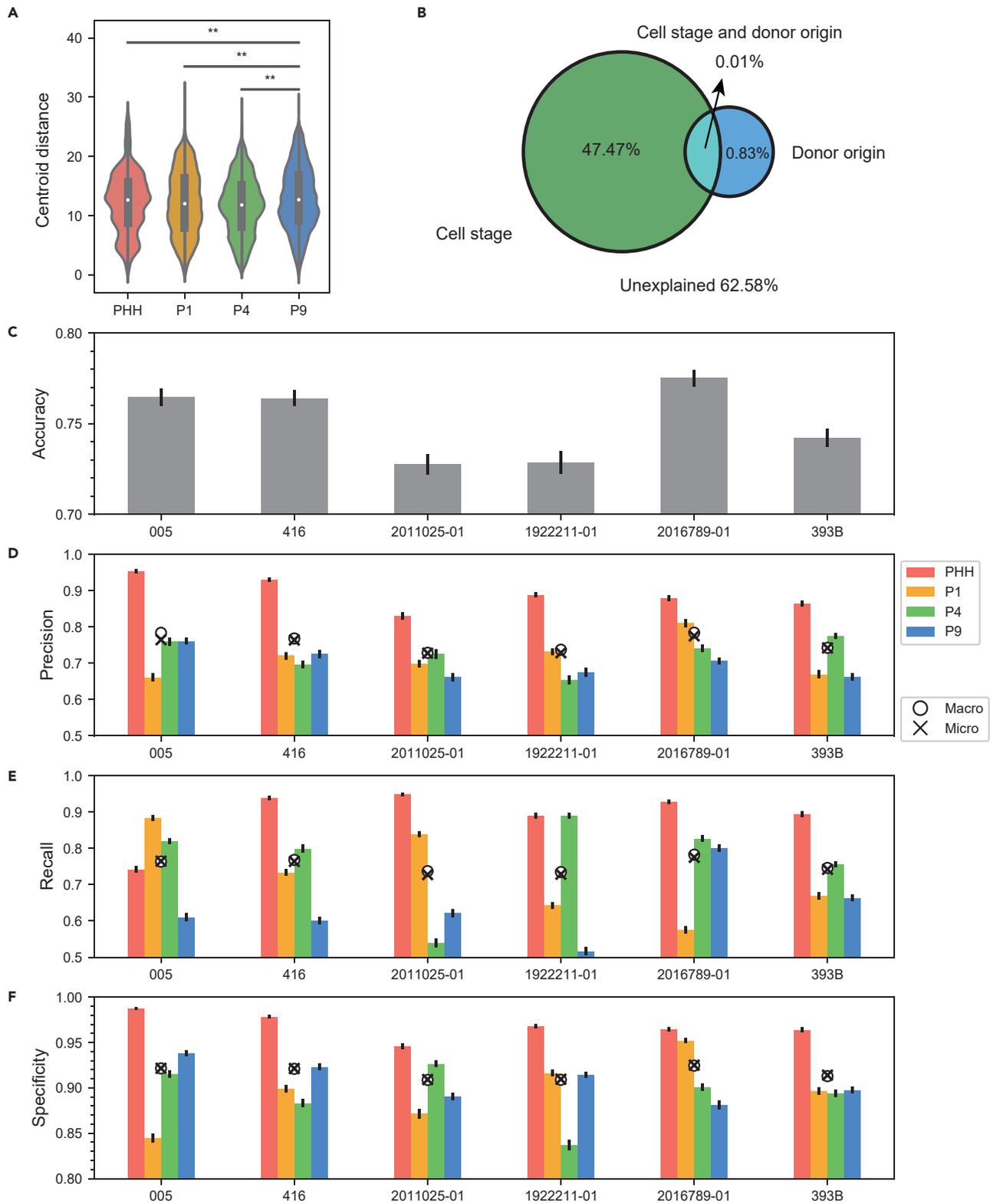
**Cross-donor classification using machine learning models**

A practical application scenario in cell culture is using Raman spectroscopy to determine cell stages for a new donor by classification models built from previously collected donors. Cross-donor classification is more challenging than within-donor classification, as the model should not only distinguish different cell stages but also diminish the batch effect and individual heterogeneity of donors. Differences among different donors were also quantified by centroid distance for different cell stages, and we found that P9 showed greater cross-donor heterogeneity (Figure 4A). As the cell culture time increases, the growth rate of cells may become more inconsistent, and cells collected on the same day may be a mixture of different passages. Moreover, we used variation partitioning to examine the contribution of cell stage and donor origin to the variance in the data (Figure 4B). Remarkably, cell stage accounted for 47.47% of the total variance, while donor origin accounted for only 0.83%. This laid the foundation for cross-donor classification, as the difference among cell stages outweighed donor origins. Therefore, we conducted cross-donor classification evaluations, in which one donor was used for independent testing and the rest were used for training. As shown in Figures 4C–4F, the accuracy of cross-donor classification decreased, ranging from 0.73 to 0.78. The precision, recall and specificity also decreased slightly. In terms of cell stage, PHH tended to achieve higher performance.

Cross-donor classification is usually more difficult than within-donor classification because of donor heterogeneity and the experimental environment. As expected, the per-cell performance of the cross-donor classification models decreased. In practical scenarios, the actual requirement is to infer the cell stage from randomly selected cells in the given plate using Raman spectra instead of accurately inferring the stage of each cell. Figure 5A shows the fractions of correctly and incorrectly predicted cells for the given plate, and there was a clear edge between them, indicating that our cross-donor classification model is applicable for determining PHH and ProlIHs passages for a new donor. Therefore, we defined the per-plate accuracy, which is the percentage of correctly classified cells, using the cross-donor classification model. When we exhausted the dataset, where all cells from five donors were used to train a model to predict all the cells from the left-out donor, the average per-plate accuracy was 0.76. To guide clinical use, we employed systematic sample size planning (SSP) to estimate the effects of sample size on the classification models. First, we examined the lower bound of the donor repository. The per-plate accuracy and model stability increased as the number of donors involved increased (Figure 5B). Among all cell stages, three donors would be the best choice for achieving optimal performance. Second, training cells were downsampled to estimate the minimum number of training cells. In progressive sampling, the classification accuracy of all cell stages showed a clear tendency of improvement. When the data size approached 9,600 cells (i.e., approximately 480 cells from four cell stages derived from five donors), the learning curve reached asymptotic behavior, which means that the performance of the model no longer increased significantly with more data (Figures 5F–5I). Finally, the minimum number of test cells was determined. The learning curve started to converge at approximately 200 cells, suggesting that 200 cells would be enough to gauge the overall stage of an unseen plate (Figure 5J–5M).

**DISCUSSION**

The present study provided a large-scale Raman spectra repository of different cell stages of ProlIHs obtained from six donors, resulting in a total of more than 14,000 spectra. Machine learning models were utilized for two typical tasks for cell stage classification: within-donor and cross-donor tasks. We found that OvO models outperformed OvR models regardless of the base model, suggesting that OvO would be more suitable for discerning subtle variations among cell stages. In the within-donor task, we achieved high performance in per-cell prediction (the accuracy, precision and recall were approximately 0.92, and the specificity was 0.97). In the cross-donor task, the data from the test donor were



**Figure 4. Cross-donor classification between primary and proliferating hepatocytes**

(A) Cross-donor heterogeneity was quantified by centroid distance.

(B) The proportion of data variance explained by cell stage and donor origin.

(C–F) Performance of classification models trained on cells from five donors and tested on cells from another donor: accuracy (C), precision (D), recall (E), and specificity (F). Error bars represent 95% confidence intervals.

held out from training process. However, due to the shortage of donors, there have been no thorough investigations on cross-donor prediction tasks in which we constructed models without any information from the test donor. It is more clinically relevant and more challenging because models should achieve high accuracy in identifying cell stages as well as eliminating noise from donor sources, experimental environments and so on. Although the performance on the cross-donor task decreased for per-cell prediction, the per-plate accuracy was still high. In practical applications, our cross-donor model can determine whether most cells within the plate reach the desired stage and whether cells can be harvested. Moreover, we provided the minimum number of training donors, training cells and testing cells for building classification models, which would be helpful for better experimental design.

In conclusion, our study not only provided abundant Raman spectral resources for PHHs and ProlIHs but also offered a complete analysis workflow, including data preprocessing, peak identification, within-donor classification, cross-donor classification and SSP. Our results will be helpful for the quality control of ProlIHs using Raman spectroscopy. The analysis framework can also be extended to encompass other cell types.

**Limitations of the study**

There are several challenges. Since only a few Raman peaks have been annotated to known biomolecules, differentially changed peaks and classification models between PHHs and ProlIHs lack interpretability.<sup>36</sup> The Raman signal is inherently weak, and experimental variability may produce high noise; therefore, standardized experimental protocols, denoising methods and computational analyses are necessary to maintain robust classification.<sup>20,21</sup>

**STAR★METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
- **METHOD DETAILS**
  - Raman Microspectroscopy
  - Peak identification
  - tSNE and data heterogeneity
  - Machine learning classification models
  - Within-donor and cross-donor classification pipeline
  - Sample size planning (SSP)
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

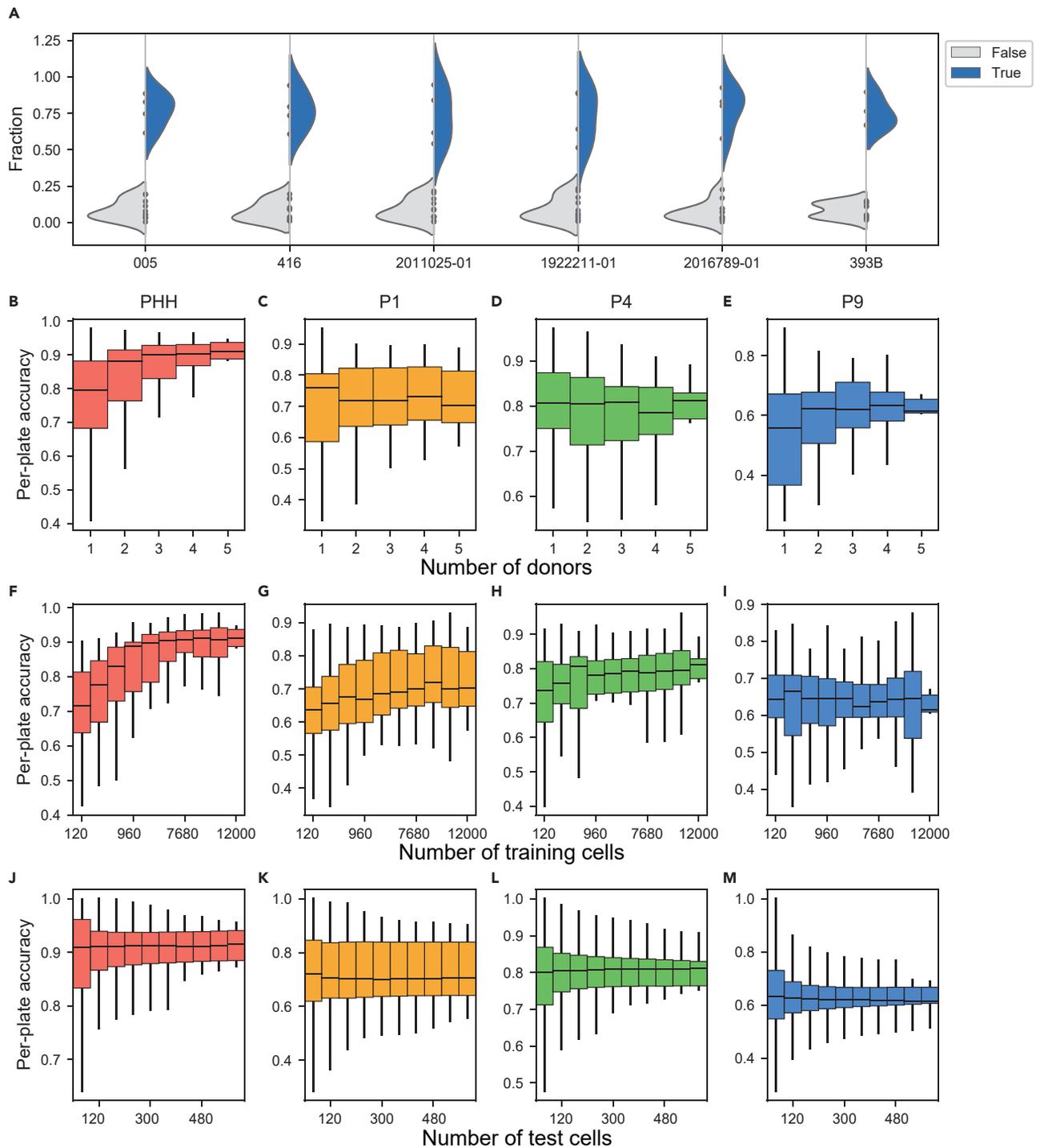
**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2024.109500>.

**ACKNOWLEDGMENTS**

This work was supported by the National Natural Science Foundation of China (T2122018, 32170680, and 82173878), the National Key R&D Program of China (2021YFF1200900), the Shanghai Municipal Science and Technology Major Project, the CAS Youth Innovation Promotion Association (Y2022076), the “Organ Reconstruction and Manufacturing” Strategic Priority Research Program of the Chinese Academy of Sciences (DA16020205), the Independent Deployment Program of the Institute of Pharmaceutical Innovation of the Chinese Academy of Sciences (LX211004), and the development fund for Shanghai talents.

We are thankful to Lijian Hui (Center for Excellence in Molecular Cell Science, CAS, China), Ruimin Huang (Shanghai Institute of Materia Medica, Chinese Academy of Sciences), Mengya Lin (School of Life Sciences and Biotechnology, Shanghai Jiao Tong University), Jiaying Chen (School of Chemistry and Chemical Engineering, South China University of Technology), and Zhikang Li (School of Chemical Science and Engineering, Tongji University) for greatly supporting this work.



**Figure 5. Sample size planning for clinical application**

(A) The fractions of correctly and incorrectly predicted cells for the given plate.

(B–M) Three sample size planning strategies were performed to show the changes in per-plate accuracy with the number of training donors (B–E), training cells (F–I), and testing cells (J–M). The minimal sample size was at the point where the rising curve was flat.

## AUTHOR CONTRIBUTIONS

B.S.: conceptualization, methodology, formal analysis, writing – original draft. C.M.: conceptualization, validation, investigation, data curation, writing – original draft. L.T.: investigation, writing – review & editing. Z.W.: resources. Z.P.: resources. G.P.: conceptualization, supervision, funding acquisition. H.L.: conceptualization, methodology, supervision, funding acquisition, writing – review & editing.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 23, 2023

Revised: February 28, 2024

Accepted: March 12, 2024

Published: March 13, 2024

## REFERENCES

- Asrani, S.K., Devarbhavi, H., Eaton, J., and Kamath, P.S. (2019). Burden of liver diseases in the world. *J. Hepatol.* 70, 151–171. <https://doi.org/10.1016/j.jhep.2018.09.014>.
- Sakellariopoulos, T., Vougas, K., Narang, S., Koinis, F., Kotsinas, A., Polyzos, A., Moss, T.J., Piha-Paul, S., Zhou, H., Kardala, E., et al. (2019). A Deep Learning Framework for Predicting Response to Therapy in Cancer. *Cell Rep.* 29, 3367–3373.e4. <https://doi.org/10.1016/j.celrep.2019.11.017>.
- Wang, F.S., Fan, J.G., Zhang, Z., Gao, B., and Wang, H.Y. (2014). The global burden of liver disease: the major impact of China. *Hepatology* 60, 2099–2108. <https://doi.org/10.1002/hep.27406>.
- Dwyer, B.J., Macmillan, M.T., Brennan, P.N., and Forbes, S.J. (2021). Cell therapy for advanced liver diseases: Repair or rebuild. *J. Hepatol.* 74, 185–199. <https://doi.org/10.1016/j.jhep.2020.09.014>.
- Enosawa, S., Horikawa, R., Yamamoto, A., Sakamoto, S., Shigeta, T., Nosaka, S., Fujimoto, J., Nakazawa, A., Tanoue, A., Nakamura, K., et al. (2014). Hepatocyte transplantation using a living donor reduced graft in a baby with ornithine transcarbamylase deficiency: a novel source of hepatocytes. *Liver Transpl.* 20, 391–393. <https://doi.org/10.1002/lt.23800>.
- Messina, A., Luce, E., Hussein, M., and Dubart-Kupperschmitt, A. (2020). Pluripotent-Stem-Cell-Derived Hepatic Cells: Hepatocytes and Organoids for Liver Therapy and Regeneration. *Cells* 9, 420. <https://doi.org/10.10390/cells9020420>.
- Fox, I.J., Chowdhury, J.R., Kaufman, S.S., Goertzen, T.C., Chowdhury, N.R., Warkentin, P.I., Dorko, K., Sauter, B.V., and Strom, S.C. (1998). Treatment of the Crigler-Najjar syndrome type I with hepatocyte transplantation. *N. Engl. J. Med.* 338, 1422–1426. <https://doi.org/10.1056/nejm199805143382004>.
- Unzu, C., Planet, E., Brandenberg, N., Fusil, F., Cassano, M., Perez-Vargas, J., Friedli, M., Cosset, F.L., Lutolf, M.P., Wildhaber, B.E., and Trono, D. (2019). Pharmacological Induction of a Progenitor State for the Efficient Expansion of Primary Human Hepatocytes. *Hepatology* 69, 2214–2231. <https://doi.org/10.1002/hep.30425>.
- Najimi, M., Khuu, D.N., Lysy, P.A., Jazouli, N., Abarca, J., Sempoux, C., and Sokal, E.M. (2007). Adult-derived human liver mesenchymal-like cells as a potential progenitor reservoir of hepatocytes? *Cell Transplant.* 16, 717–728. <https://doi.org/10.3727/000000007783465154>.
- Lee, S.Y., Kim, H.J., and Choi, D. (2015). Cell sources, liver support systems and liver tissue engineering: alternatives to liver transplantation. *Int. J. Stem Cells* 8, 36–47. <https://doi.org/10.15283/ijsc.2015.8.1.36>.
- Touboul, T., Hannan, N.R.F., Corbinau, S., Martinez, A., Martinet, C., Branchereau, S., Mainot, S., Strick-Marchand, H., Pedersen, R., Di Santo, J., et al. (2010). Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology* 51, 1754–1765. <https://doi.org/10.1002/hep.23506>.
- Si-Tayeb, K., Noto, F.K., Nagaoka, M., Li, J., Battle, M.A., Duris, C., North, P.E., Dalton, S., and Duncan, S.A. (2010). Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology* 51, 297–305. <https://doi.org/10.1002/hep.23354>.
- Yovchev, M.I., Xue, Y., Shafritz, D.A., Locker, J., and Oertel, M. (2014). Repopulation of the fibrotic/cirrhotic rat liver by transplanted hepatic stem/progenitor cells and mature hepatocytes. *Hepatology* 59, 284–295. <https://doi.org/10.1002/hep.26615>.
- Peng, W.C., Kraaijer, L.J., and Kluiver, T.A. (2021). Hepatocyte organoids and cell transplantation: What the future holds. *Exp. Mol. Med.* 53, 1512–1528. <https://doi.org/10.1038/s12276-021-00579-x>.
- Kakinuma, S., Nakauchi, H., and Watanabe, M. (2009). Hepatic stem/progenitor cells and stem-cell transplantation for the treatment of liver disease. *J. Gastroenterol.* 44, 167–172. <https://doi.org/10.1007/s00535-008-2297-z>.
- Lu, W.Y., Bird, T.G., Boulter, L., Tsuchiya, A., Cole, A.M., Hay, T., Guest, R.V., Wojtacha, D., Man, T.Y., Mackinnon, A., et al. (2015). Hepatic progenitor cells of biliary origin with liver repopulation capacity. *Nat. Cell Biol.* 17, 971–983. <https://doi.org/10.1038/ncb3203>.
- Zhang, K., Zhang, L., Liu, W., Ma, X., Cen, J., Sun, Z., Wang, C., Feng, S., Zhang, Z., Yue, L., et al. (2018). In Vitro Expansion of Primary Human Hepatocytes with Efficient Liver Repopulation Capacity. *Cell Stem Cell* 23, 806–819.e4. <https://doi.org/10.1016/j.stem.2018.10.018>.
- Qiao, S., Feng, S., Wu, Z., He, T., Ma, C., Peng, Z., Tian, E., and Pan, G. (2021). Functional Proliferating Human Hepatocytes: In Vitro Hepatocyte Model for Drug Metabolism, Excretion, and Toxicity. *Drug Metab. Dispos.* 49, 305–313. <https://doi.org/10.1124/dmd.120.000275>.
- Suhito, I.R., Han, Y., Min, J., Son, H., and Kim, T.H. (2018). In situ label-free monitoring of human adipose-derived mesenchymal stem cell differentiation into multiple lineages. *Biomaterials* 154, 223–233. <https://doi.org/10.1016/j.biomaterials.2017.11.005>.
- Guo, S., Popp, J., and Bocklitz, T. (2021). Chemometric analysis in Raman spectroscopy from experimental design to machine learning-based modeling. *Nat. Protoc.* 16, 5426–5459. <https://doi.org/10.1038/s41596-021-00620-3>.
- Butler, H.J., Ashton, L., Bird, B., Cinque, G., Curtis, K., Dorney, J., Esmonde-White, K., Fullwood, N.J., Gardner, B., Martin-Hirsch, P.L., et al. (2016). Using Raman spectroscopy to characterize biological materials. *Nat. Protoc.* 11, 664–687. <https://doi.org/10.1038/nprot.2016.036>.
- Hsu, C.C., Xu, J., Brinkhof, B., Wang, H., Cui, Z., Huang, W.E., and Ye, H. (2020). A single-cell Raman-based platform to identify developmental stages of human pluripotent stem cell-derived neurons. *Proc. Natl. Acad. Sci. USA* 117, 18412–18423. <https://doi.org/10.1073/pnas.2001906117>.
- Ho, C.S., Jean, N., Hogan, C.A., Blackmon, L., Jeffrey, S.S., Holodniy, M., Banaei, N., Saleh, A.A.E., Ermon, S., and Dionne, J. (2019). Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nat. Commun.* 10, 4927. <https://doi.org/10.1038/s41467-019-12898-9>.
- Yu, S., Li, X., Lu, W., Li, H., Fu, Y.V., and Liu, F. (2021). Analysis of Raman Spectra by Using Deep Learning Methods in the Identification of Marine Pathogens. *Anal. Chem.* 93, 11089–11098. <https://doi.org/10.1021/acs.analchem.1c00431>.
- Rho, E., Kim, M., Cho, S.H., Choi, B., Park, H., Jang, H., Jung, Y.S., and Jo, S. (2022). Separation-free bacterial identification in arbitrary media via deep neural network-based SERS analysis. *Biosens. Bioelectron.* 202, 113991. <https://doi.org/10.1016/j.bios.2022.113991>.
- Wang, L., Tang, J.W., Li, F., Usman, M., Wu, C.Y., Liu, Q.H., Kang, H.Q., Liu, W., and Gu, B. (2022). Identification of Bacterial Pathogens at Genus and Species Levels through Combination of Raman Spectrometry and Deep-Learning Algorithms. *Microbiol. Spectr.* 10, e0258022. <https://doi.org/10.1128/spectrum.02580-22>.

27. Wu, M., Wang, S., Pan, S., Terentis, A.C., Strasswimmer, J., and Zhu, X. (2021). Deep learning data augmentation for Raman spectroscopy cancer tissue classification. *Sci. Rep.* **11**, 23842. <https://doi.org/10.1038/s41598-021-02687-0>.
28. Shu, C., Yan, H., Zheng, W., Lin, K., James, A., Selvarajan, S., Lim, C.M., and Huang, Z. (2021). Deep Learning-Guided Fiberoptic Raman Spectroscopy Enables Real-Time In Vivo Diagnosis and Assessment of Nasopharyngeal Carcinoma and Post-treatment Efficacy during Endoscopy. *Anal. Chem.* **93**, 10898–10906. <https://doi.org/10.1021/acs.analchem.1c01559>.
29. Hollon, T.C., Pandian, B., Adapa, A.R., Urias, E., Save, A.V., Khalsa, S.S.S., Eichberg, D.G., D'Amico, R.S., Farooq, Z.U., Lewis, S., et al. (2020). Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat. Med.* **26**, 52–58. <https://doi.org/10.1038/s41591-019-0715-9>.
30. Jin, Z., Yue, Q., Duan, W., Sui, A., Zhao, B., Deng, Y., Zhai, Y., Zhang, Y., Sun, T., Zhang, G.P., et al. (2022). Intelligent SERS Navigation System Guiding Brain Tumor Surgery by Intraoperatively Delineating the Metabolic Acidosis. *Adv. Sci.* **9**, e2104935. <https://doi.org/10.1002/advs.202104935>.
31. Xie, Y., Su, X., Wen, Y., Zheng, C., and Li, M. (2022). Artificial Intelligent Label-Free SERS Profiling of Serum Exosomes for Breast Cancer Diagnosis and Postoperative Assessment. *Nano Lett.* **22**, 7910–7918. <https://doi.org/10.1021/acs.nanolett.2c02928>.
32. Pezzotti, G., Boschetto, F., Ohgitani, E., Fujita, Y., Shin-Ya, M., Adachi, T., Yamamoto, T., Kanamura, N., Marin, E., Zhu, W., et al. (2022). Raman Molecular Fingerprints of SARS-CoV-2 British Variant and the Concept of Raman Barcode. *Adv. Sci.* **9**, e2103287. <https://doi.org/10.1002/advs.202103287>.
33. Yang, Y., Xu, B., Murray, J., Haverstick, J., Chen, X., Tripp, R.A., and Zhao, Y. (2022). Rapid and quantitative detection of respiratory viruses using surface-enhanced Raman spectroscopy and machine learning. *Biosens. Bioelectron.* **217**, 114721. <https://doi.org/10.1016/j.bios.2022.114721>.
34. Ma, C., Zhang, L., He, T., Cao, H., Ren, X., Ma, C., Yang, J., Huang, R., and Pan, G. (2021). Single cell Raman spectroscopy to identify different stages of proliferating human hepatocytes for cell therapy. *Stem Cell Res. Ther.* **12**, 555. <https://doi.org/10.1186/s13287-021-02619-9>.
35. Galar, M., Fernández, A., Barrenechea, E., Bustince, H., and Herrera, F. (2011). An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recogn.* **44**, 1761–1776. <https://doi.org/10.1016/j.patcog.2011.01.017>.
36. Dodo, K., Fujita, K., and Sodeoka, M. (2022). Raman Spectroscopy for Chemical Biology Research. *J. Am. Chem. Soc.* **144**, 19651–19667. <https://doi.org/10.1021/jacs.2c05359>.
37. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **17**, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
38. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830.
39. Borcard, D., Legendre, P., and Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology* **73**, 1045–1055.
40. Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **28**, 1–26. <https://doi.org/10.18637/jss.v028.i05>.
41. Liland, K.H., Almøy, T., and Mevik, B.-H. (2010). Optimal Choice of Baseline Correction for Multivariate Calibration of Spectra. *Appl. Spectrosc.* **64**, 1007–1016. <https://doi.org/10.1366/000370210792434350>.
42. Ali, N., Girus, S., Rösch, P., Popp, J., and Bocklitz, T. (2018). Sample-Size Planning for Multivariate Data: A Raman-Spectroscopy-Based Example. *Anal. Chem.* **90**, 12485–12492. <https://doi.org/10.1021/acs.analchem.8b02167>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Biological samples</b>		
Primary human hepatocytes	Corning	416; 393B
Primary human hepatocytes	Novabiosis	005; 2011025-01; 2016789-01; 1922211-01
<b>Chemicals, peptides, and recombinant proteins</b>		
N-acetyl-cysteine	Sigma-Aldrich	Cat#A9165-100G
Nicotinamide	Solarbio	Cat#N8070
Recombinant humanFGF10	Peprotech	Cat#100-26
Recombinant human EGF	Peprotech	Cat#AF-100-15
Recombinant human HGF	Peprotech	Cat#100-39
Human [Leu15]-gastrin I	Sigma-Aldrich	Cat#G9145
A83-01	Tocris Bioscience	Cat#2939
Rho kinase inhibitor Y-27632	Selleck	Cat#s1049
Wnt3a protein	Stemimmune LLC	Cat#MST-WT3a-0010
Fetal bovine serum	Ausbian	Cat#VS500T
<b>Software and algorithms</b>		
Scripts for analysis	This paper	<a href="https://github.com/LiHongCSBLab/Raman_ProlIHH">https://github.com/LiHongCSBLab/Raman_ProlIHH</a>
Python v3.8.3	Python Software Foundation	<a href="https://www.python.org/">https://www.python.org/</a>
scipy v1.5.2	Virtanen et al. <sup>37</sup>	<a href="https://scipy.org/">https://scipy.org/</a>
sklearn v0.24.2	Pedregosa et al. <sup>38</sup>	<a href="https://scikit-learn.org">https://scikit-learn.org</a>
R v4.0.2	R Foundation for Statistical Computing	<a href="https://www.R-project.org/">https://www.R-project.org/</a>
Vegan R package v 2.6	Borcard et al. <sup>39</sup>	<a href="https://cran.r-project.org/web/packages/vegan/index.html">https://cran.r-project.org/web/packages/vegan/index.html</a>
Caret R package v6.0.88	Kuhn <sup>40</sup>	<a href="https://cran.r-project.org/web/packages/caret/index.html">https://cran.r-project.org/web/packages/caret/index.html</a>
Baseline R package v1.3.1	Liland et al. <sup>41</sup>	<a href="https://cran.r-project.org/web/packages/baseline/index.html">https://cran.r-project.org/web/packages/baseline/index.html</a>

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by lead contact, Hong Li ([lihong01@sinh.ac.cn](mailto:lihong01@sinh.ac.cn)).

#### Materials availability

No new reagents were generated from this paper.

#### Data and code availability

- All data reported in this paper will be shared by the [lead contact](#) upon request.
- All original code has been deposited at [https://github.com/LiHongCSBLab/Raman\\_ProlIHH](https://github.com/LiHongCSBLab/Raman_ProlIHH) and is publicly available as of the date of publication. DOIs are listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Primary human hepatocytes (PHHs) were purchased from Corning (Lot: 416 and 393B) and Novabiosis (Lot: 005, 2011025-01, 2016789-01, and 1922211-01). PHHs were cultured in OptiCulture Hepatocyte Media (Novabiosis).

Proliferating human hepatocytes (ProlIHs) were induced according to previously reported methods.<sup>17,34</sup> Briefly, to generate ProlIHs at passage 1 (P1), PHHs were treated with HM for 7 days under hypoxic cell culture conditions (2% O<sub>2</sub>).

PHHs, ProlIHs P1, P4, and P9 were seeded on 8-well Chamber Raman Scattering Microslides for 24 h. The cells were washed with PBS and then fixed in 4% paraformaldehyde for 15 min. To collect the Raman spectra, the cells were washed with sterile water and evaporated naturally.

## METHOD DETAILS

### Raman Microspectroscopy

Raman spectroscopy was performed using a WITec alpha 300 R Raman Imaging Microscope as described previously.<sup>34</sup> In brief, the spectra were collected with an excitation wavelength of 532 nm and 100 × air objective (N.A. 0.9). To increase the signal-to-noise ratio of the spectra, a 1800 lines/mm grating was used with a 10 s integration time and 9 mW laser power. The fingerprint region of the Raman spectra (from 300 to 1800 cm<sup>-1</sup>) was measured. A single cell was randomly subjected to 10 Raman spectra detections. At least 60 cells per donor/batch were collected. All spectra underwent normalization procedures, including cosmic ray subtraction, baseline correction,<sup>41</sup> and area normalization.<sup>34</sup>

### Peak identification

To identify differentially changed peaks among cell stages, Raman spectra were processed by a sliding window with sum pooling. The window size was 20 points, and the stride was 5 points. The differentially changed peaks were reported with Student's t-test with Bonferroni correction p value < 0.05 and effect size > 0.2.

### tSNE and data heterogeneity

To measure the variance among different cells, the centroid distance was calculated. Due to the curse of dimensionality, t-SNE was used to project Raman spectral features in two-dimensional space for distance calculations.

$$\text{Centroid distance} = \frac{1}{n} \sum_{i=1}^n \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2},$$

where  $(x_i, y_i)$  is the coordinate of the t-SNE-embedded Raman spectrum of sample  $i$ , and  $(\bar{x}, \bar{y})$  is the coordinate of the centroid. The intra-stage heterogeneity is defined as intra-centroid distance of cells from the same stage, which is expressed as  $\frac{1}{n_s} \sum_{i=1}^{n_s} \sqrt{(x_i - \bar{x}_s)^2 + (y_i - \bar{y}_s)^2}$ , where  $n_s$  is the number of cells in the given stage, and  $(\bar{x}_s, \bar{y}_s)$  is calculated as  $(\frac{1}{n_s} \sum_{i=1}^{n_s} x_i, \frac{1}{n_s} \sum_{i=1}^{n_s} y_i)$ . Inter-stage heterogeneity is measured by intra-centroid distance of cells from different stages, which is calculated as  $\frac{1}{n_{s_A} + n_{s_B}} (\sum_{i=1}^{n_{s_A}} \sqrt{(x_i - \bar{x}_{s_B})^2 + (y_i - \bar{y}_{s_B})^2} + \sum_{j=1}^{n_{s_B}} \sqrt{(x_j - \bar{x}_{s_A})^2 + (y_j - \bar{y}_{s_A})^2})$ , where  $n_{s_A}$  and  $n_{s_B}$  represent the number of cells in stages A and B, respectively.  $(\bar{x}_{s_A}, \bar{y}_{s_A})$  and  $(\bar{x}_{s_B}, \bar{y}_{s_B})$  are the centroids of cells within stages A and B, respectively. Cross-donor heterogeneity is defined as the inter-centroid distance of cells from the same stage but from different donors, which is expressed as  $\frac{1}{n} \sum_{k \in D} \sum_{i \notin D_k} \sqrt{(x_i - \bar{x}_{D_k})^2 + (y_i - \bar{y}_{D_k})^2}$ , where  $n$  is the number of summation items,  $D = \{D_k\}$  ( $k = 1, 2, \dots$ ), and  $(\bar{x}_{D_k}, \bar{y}_{D_k})$  is the centroid of cells from donor  $k$ .<sup>37</sup>

To illustrate the portion of dataset variance explained by donors and cell stages, variation partitioning was conducted using the 'varpart()' function from the 'vegan' R package.<sup>39</sup> Donors and cell stages served as explanatory variables, while the intra-centroid distances of the whole dataset served as response variables.

### Machine learning classification models

Identification of the developmental stage of ProlIHs is a multiclass classification task. A support vector machine (SVM) aims at finding a hyperplane to separate the positive and negative samples in a high-dimensional space. Since SVMs are designed for binary-class tasks, two typical strategies are usually used to split a multiclass task into multiple binary-class tasks: one-vs-rest (OvR) and one-vs-one (OvO). OvR requires one model to be trained for each class, which results in  $N$  models built for  $N$  classes. OvO involves one model to be created for each pair of classes, resulting in  $N(N - 1)/2$  models for  $N$  classes. Finally, the class label is assigned the highest prediction score or the most votes among the models from OvR or OvO, respectively. SVM models were implemented by the "caret" package in R 4.0.2.<sup>40</sup>

For each class, the performances were measured by precision, recall and specificity. Macro- and micro-averages were adopted for representing performance across all classes. The overall accuracy rate was also evaluated.

$$\text{Accuracy} = \frac{\sum(TN_i + TP_i)}{\sum(TN_i + TP_i + FN_i + FP_i)}$$

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i}$$

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i}$$

$$\text{Specificity}_i = \frac{TN_i}{TN_i + FP_i}$$

$$\text{Macro - precision} = \frac{\sum \text{Precision}_i}{N}$$

$$\text{Micro - precision} = \frac{\sum TP_i}{\sum(TP_i + FP_i)}$$

$$\text{Macro - recall} = \frac{\sum \text{Recall}_i}{N}$$

$$\text{Micro - recall} = \frac{\sum TP_i}{\sum(TP_i + FN_i)}$$

$$\text{Macro - specificity} = \frac{\sum \text{Specificity}_i}{N}$$

$$\text{Micro - specificity} = \frac{\sum TN_i}{\sum(TN_i + FP_i)}$$

where  $TP_i$ ,  $FP_i$ ,  $TN_i$  and  $FN_i$  are the number of true positives, false positives, true negatives and false negatives for class  $i$ , respectively, and  $N$  is the number of classes. Metrics were implemented by scikit-learn in Python 3.8.<sup>38</sup>

### Within-donor and cross-donor classification pipeline

In the within-donor classification task, nested cross-validation was undertaken to evaluate the model performance using Raman spectra of cells from the given donor. The inner fourfold loop is responsible for tuning the hyperparameters (i.e., the cost and the smoothing parameter) via a grid search, and the outer fivefold loop is responsible for model evaluation (Figure S1). The procedures were repeated 10 times for reliability and fidelity. In the cross-donor classification task, data from the given donor were withheld as the independent test set. Fivefold cross-validation was applied to find the best hyperparameter configuration using the remaining data from the donor repository. The performance was finally assessed by data from the withheld donor (Figure S2).

### Sample size planning (SSP)

SSP is meant to determine the minimal number of measurements needed for a statistical analysis.<sup>20,42</sup> It is important for biological analysis of Raman spectra to retain both efficiency and reliability. Learning curve (LC)-based methods are commonly used. LC measures the improvement of models with increasing data size. In this study, there were three types of SSPs for cross-donor tasks with per-plate accuracy as the model metric. First, the SSP of the number of training donors was used to determine how many donors were needed in the training set to obtain good generalizability to an unseen donor. With  $k$  training donors and  $m$  donors in total, there were  $m \times C_{m-1}^k$  possible combinations to iterate. Second, a SSP of the number of training cells was carried out. The number of training cells sampled from five donors ranged from approximately 120 to 10800. Each sampling process was repeated 10 times. Third, the number of test cells in the SSP was determined. The number of test cells sampled from the test donor ranged from approximately 25 to 2300. Each sampling process was repeated 10 times.

## QUANTIFICATION AND STATISTICAL ANALYSIS

Model performances are presented as the mean of each iteration, and error bars represent 95% confidence intervals. Asterisks indicate the level of statistical significance: \* $p < 0.05$ , \*\* $p < 0.01$ , n.s., not significant.