Deriving accurate molecular indicators of protein synthesis through Raman-based sparse classification:
Supplementary Information

N. Pavillon\textsuperscript{1}, N. I. Smith\textsuperscript{1,2}
\textsuperscript{1}Biophotonics Laboratory, Immunology Frontier Research Center (IFReC),
\textsuperscript{2}Open and Transdisciplinary Research Institute (OTRI),
Osaka University, Yamadaoka 3-1, Suita, 565-0871, Suita, Osaka, Japan

Figure S1: Selection of the regularization parameter $\lambda$ in Lasso through cross-validation, based on binomial deviance. The selected value (right dashed line) corresponds to the $\lambda$ value that increases deviance compared to the average minimum (left dashed line) by less than one standard deviation.
Figure S2: Representative Raman spectra from single cells, for both control and LPS experimental conditions. (A) Raw spectra as recorded by the camera, and (B) Pre-processed spectra as used for classification.

Figure S3: Amount of parameters for Lasso and PCA/LDA in function of sample size. While the limitation to 90% of variance implies that the amount of variables in PCA/LDA rapidly saturates, it keeps increasing in the case of Lasso. The line represents the average of 10 runs with different random selection of training subsets, the shaded regions represent the standard deviation.
Figure S4: (A) Coefficients of classification models for variance-limited PCA/LDA and PCA/Lasso. The limit on variance implies that selected coefficients are concentrated on low-order PCs, while Lasso selects coefficients in the whole range. An offset of ±0.5 has been added for visibility. (B) Coefficients in the low-order PCs within the variance limit, where multiple coefficients are zero (shown in green) in the case of PCA/Lasso.

Figure S5: IL-6 secretion measured by enzyme-linked immunosorbent assay (ELISA MAX, Biolegend) in the culture medium of Raw264 cells exposed to LPS (50 ng/mL) and increasing dosage of CHX, represented by the median of triplicate measurements, with the error bar showing the standard deviation. Colors represent 2 independent experiments, and the dashed bar represents the limit of detection of the kit. The manufacturer’s protocol was followed, and absorbance was measured at 450 nm.
Figure S6: PCA scores for components 1 to 12 on LPS/LPS+CHX data, displayed as density plots. Lines show the limits containing 50%, 75% and 95% of the respective populations.
Figure S7: Loading vectors corresponding to scores in Fig. S5. The respective contribution to variance is shown for each vector.
| Band [cm\(^{-1}\)] | Possible assignment | Vibrational mode | Reported shift [cm\(^{-1}\)] |
|-------------------|--------------------|-----------------|-----------------------------|
| **Positive (LPS+CHX)** | | | |
| 710–727 | Adenine | Ring breathing mode \(^1\) | 728 |
| 865–885 | Ribose | Ribose-base linkage \(^1\) | 868 |
| | Tryptophan | Indole ring \(^2\) | 875–880 |
| 1226–1237 | Uracil | Ring stretching \(^3\) | 1234 |
| | Protein secondary structure | Amide III \(^4\), (rand coils \(^4\)) | 1231, 1230–1295 |
| 1050–1113 | DNA/RNA | PO\(_2\) stretching \(^4\) | 1092 |
| | Phospholipids | PO\(_2\) symmetric stretch \(^2\) | 1080 |
| | Protein primary structure | C-N stretch, chain C-C stretch \(^4\) | 1066, 1080 |
| | | | |
| 1483–1493 | Guanine | CN stretching \(^1\) | 1485 |
| | Guanine, Adenine | Ring mode \(^5\) | 1486 |
| 1367–1386 | Adenine | C-H bending \(^1\) | 1378 |
| | Guanine | Ring stretch \(^2\) | 1370 |
| 773–781 | Uracil, Cytosine | Ring breathing \(^2\) | 782 |
| 1018–1020 | Ribose | CO stretch \(^4\) | 1017 |
| 3037–3046 | | | |
| **Negative (LPS)** | | | |
| 1399–1423 | Amino-acids | Carboxyl groups side chain \(^6\) | 1404 |
| | Adenine, Guanine | N=H bending \(^1\) | 1419 |
| 913–940 | \(\alpha\)-helix | Backbone CC stretch \(^7,4\) | 933, 937 |
| | Ribose-phosphate | C-O, C-C stretching \(^2\) | 914–925 |
| 1542–1550 | Tryptophan | Amide II, indole ring \(^7,2\) | 1551 |
| | | Amide II, CN stretch \(^5\) | 1548 |
| 801–815 | Lipids | PO\(_2\) stretching \(^4\) | 811 |
| | A-form helix | PO\(_2\) stretching \(^2\) | 813–816 |
| 1028.5 | | | 1027 |
| 1175–1186 | Tyrosine, Phenylalanine | Side chain/Indole ring \(^7\) | 1181 |
| | Tyrosine | CH in plane bend \(^4\) | 1176 |
| 1156–1164 | Ribose | Ribose-phosphate \(^1\) | 1162 |
| | | Backbone CC, CN \(^6\) | 1060 |
| 1194–1205 | Proline, Tyrosine \(^8\) | - | 1194, 1200 |
| | Pyruvate, Coenzyme A \(^8\) | - | 1197, 1203 |
| 1736.5 | | CO ester stretch \(^3\) | 1735 |
| 1792–1797 | | | |
| 1042 | Ribose | Backbone ribose \(^4\) | 1043 |
| 642–645 | Uracil | Ribose-uracil link bending \(^4\) | 647 |
| | Tyrosine | C-C twist \(^4\) | 645 |
| 1317–1328 | Guanine | Ring stretch \(^2\) | 1325 |
| | Amide III | NH, CH stretch \(^2,7\) | 1300–1340, 1327 |

Table S1: Possible band assignments for the separation vectors of LPS versus LPS+CHX. Regions are listed in the order of decreasing strength.
References

[1] A. J. Hobro, D. M. Standley, S. Ahmad, and N. I. Smith. “Deconstructing RNA: optical measurement of composition and structure”. *Phys. Chem. Chem. Phys.* **15**(31), 13,199 (2013).

[2] D. I. Ellis, D. P. Cowcher, L. Ashton, S. O’Hagan, and R. Goodacre. “Illuminating disease and enlightening biomedicine: Raman spectroscopy as a diagnostic tool”. *Analyst* **138**, 3871–3884 (2013).

[3] K. Maquelin, C. Kirschner, L.-P. Choo-Smith, N. van den Braak, H. P. Endtz, D. Naumann, and G. J. Puppels. “Identification of medically relevant microorganisms by vibrational spectroscopy”. *J. Microbiol. Methods* **51**(3), 255–271 (2002).

[4] I. Notingher, S. Verrier, S. Haque, J. M. Polak, and L. L. Hench. “Spectroscopic study of human lung epithelial cells (A549) in culture: Living cells versus dead cells”. *Biopolymers* **72**(4), 230–240 (2003).

[5] H. Byrne, G. Sockalingum, and N. Stone. *Biomedical Applications of Synchrotron Infrared Microspectroscopy*, chap. Raman Microscopy: Complement or Competitor?, pp. 105–142. RSC (2011).

[6] M. Ogawa, S. Nakamura, Y. Horimoto, H. An, T. Tsuchiya, and S. Nakai. “Raman Spectroscopic Study of Changes in Fish Actomyosin during Setting”. *J. Agric. Food Chem.* **47**(8), 3309–3318 (1999).

[7] L. Ashton and E. W. Blanch. “pH-induced conformational transitions in α-lactalbumin investigated with two-dimensional Raman correlation variance plots and moving windows”. *J. Mol. Struct.* **974**(1), 132–138 (2010).

[8] J. De Gelder, K. De Gussem, P. Vandenabeele, and L. Moens. “Reference database of Raman spectra of biological molecules”. *J. Raman Spectrosc.* **38**(9), 1133–1147 (2007).