Title

Mapping enzyme catalysis with metabolomic biosensing

Names and affiliations

Linfeng Xu[1], Kai-Chun Chang[1], Emory M. Payne[2], Cyrus Modavi[1], Leqian Liu[1], Claire M. Palmer[3], Nannan Tao[4], Hal S. Alper[3,5], Robert T. Kennedy[2], Dale S. Cornett[6] and Adam R. Abate[1,7]*

1 Department of Bioengineering and Therapeutic sciences, University of California, San Francisco, San Francisco, CA, USA
2 Department of Chemistry, University of Michigan, Ann Arbor, MI, USA
3 Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX, USA
4 Bruker Nano Surfaces, San Jose, CA, USA
5 McKetta Department of Chemical Engineering, The University of Texas at Austin, Austin, TX, USA
6 Bruker Daltonics, Billerica, MA, USA
7 Chan Zuckerberg Biohub, San Francisco, CA, USA

* e-mail: adam@abatelab.org

Abstract:

Enzymes are represented across a vast space of protein sequences and structural forms and have activities that far exceed the best chemical catalysts; however, engineering them to have novel or enhanced activity is limited by technologies for sensing product formation. Here, we describe a general and scalable approach for characterizing enzyme activity that uses the metabolism of the host cell as a biosensor by which to infer product formation. Since different products consume different molecules in their synthesis, they perturb host metabolism in unique ways that can be measured by mass spectrometry. This provides a general way by which to sense product formation, to discover unexpected products and map the effects of mutagenesis.

Main:

Introduction

Enzyme engineering uses an iterative cycle in which libraries of gene variants are designed, synthesized into proteins, and tested for the activity of interest[1]. The success of these engineering campaigns, however, is dependent on technologies for conducting these steps. Moreover, while there have been significant advances in design and build, test remains a bottleneck[2]. For example, the dominant strategy is well plate screening, because it is simple and flexible, and allows a variety of measurement techniques to directly quantify the product of the enzyme[2,3]. Well plate screening, however, is severely limited in scalability, testing just hundreds of variants per cycle[3]; this is a major issue because the likelihood of identifying superior variants scales with the number screened. Thus, alternative approaches based on selections, flow cytometry, and droplet microfluidics are valuable because they afford much higher throughput, screening $>10^7$ variants per cycle[2,4]. However, a major constraint of these approaches is that they do not directly detect product formation, requiring a secondary assay to tether it to a detectable readout[4]. Identifying such assays can be challenging and, often, is not possible for enzymes of interest[2]. To enhance our ability to engineer enzymes through screening, a new method is needed that combines the generality of well plates with the scalability of microfluidics.

In this paper, we describe a screening approach that combines the scalability of microfluidics with the generalizability of mass spectrometry (MS) (**Figure 1**). Our approach leverages the background metabolism of the host as a biosensor with which to assess the activity of an enzyme embedded in it, and a novel microscale mass spectrometry technology to characterize metabolomic changes. Since the enzyme catalyzes molecules of central metabolism, its activity perturbs the host's metabolite profile[1,5], generating a signature that can be detected even if the enzyme product is not directly observed. Our approach thus provides a general way to map the catalysis of a mutated enzyme, to characterize the range of products it generates, and recover the sequences of variants with desired activities.
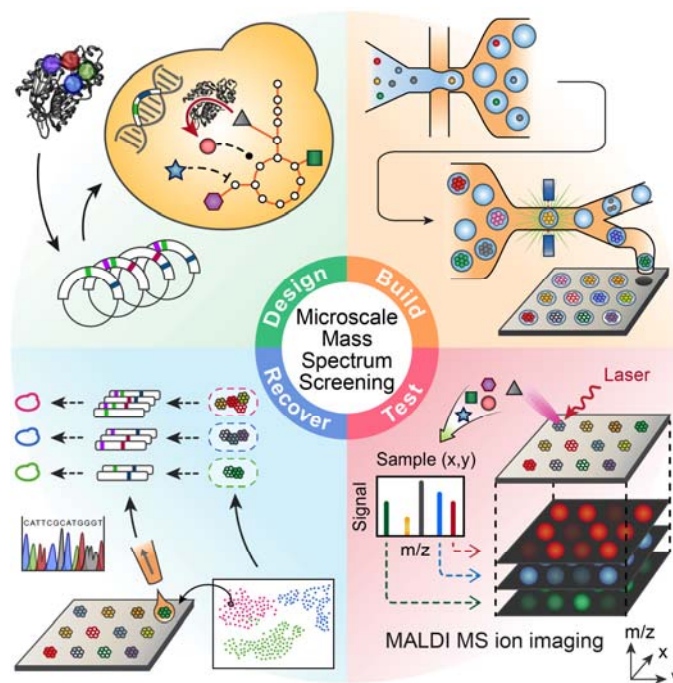

Fig. 1 | Metabolomic biosensing with microscale mass spectrometry provides a general strategy for screening enzymes. Enzyme variants are designed and transformed into yeast (design), then synthesized in the yeast where they consume molecules of central metabolism to generate product (build). Using printed droplet microfluidics, they are dispensed to a picoliter well array and subjected to MALDI-MS imaging to quantify cell metabolites (test). UMAP clusters cells according to metabolomic profile, where each cluster indicates a different enzyme phenotype. Desired mutants are extracted from the plate, sequenced, and confirmed in bulk cultures.
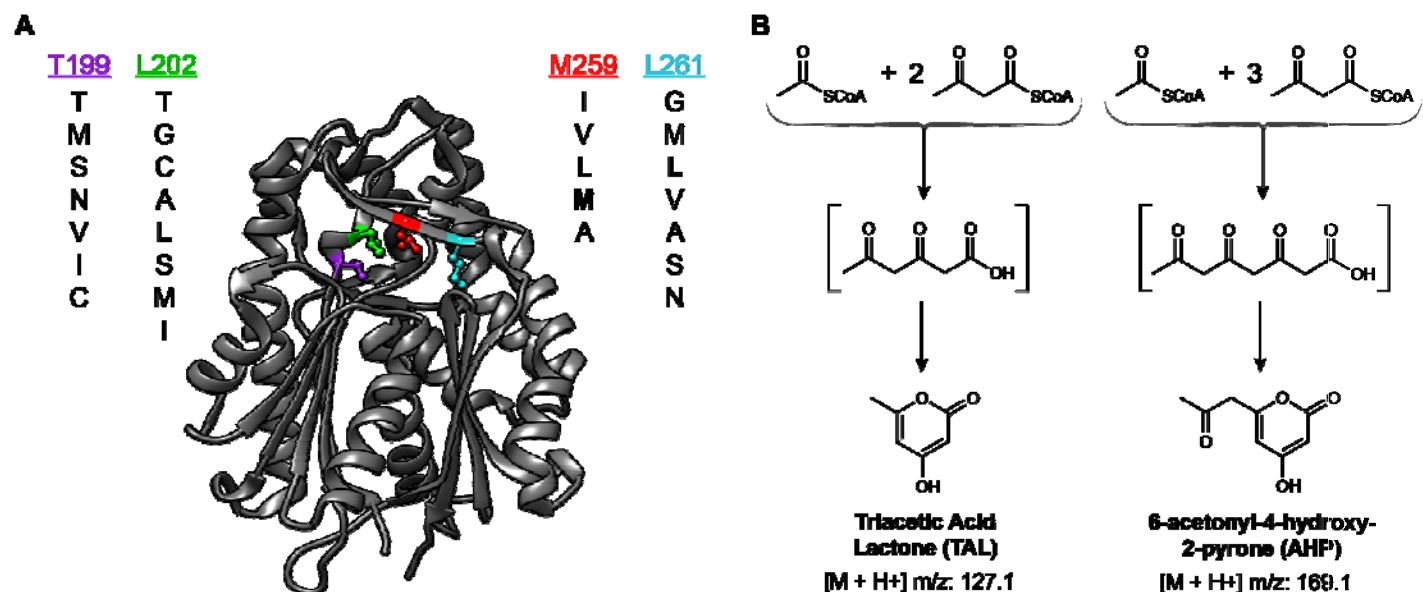
Results

Fig. 2 | Overview of the type 3 PKS variant library. A, Crystal structure of *Gerbera hybrida* G2PS1 (PDB ID: 1EE0) showing the location of the four residues selected for mutagenesis and identity of the residue mutants. All residues except T199 directly form the active site cavity. **B**, The smallest condensation/cyclization products expected from type III polyketide synthase activity: Triacetic Acid Lactone (TAL, the native product of G2PS1) from one acetyl-CoA and two Malonyl-CoA and 6-Acetonyl-4-Hydroxy-2-Pyrone (AHP) from one acetyl-CoA and three Malonyl-CoA. Higher order polyketides, not shown, are possible from additional condensations of Malonyl-CoA.

Matrix-assisted laser desorption ionization (MALDI) MS has been used for single cell metabolomics[6] and to identify enzyme products from microbial colonies[7]. Here, we combine this approach with printed droplet microfluidics (PDM) to prepare, print, and screen all mutants from a semi-rationally varied 4-position library of the *Gerbera hybrida* G2PS1 type-3 polyketide synthase (t3PKS)[8], comprising 1960 codon shuffled members (**Figure 2A**) (**methods**). This enzyme is responsible for biosynthesis of triacetic acid lactone (TAL) through condensation of a starter acetyl-CoA unit with two malonyl-CoA molecules and subsequent cyclization of the triketide chain [8,9]. TAL has been used as a platform precursor for synthesis of high value chemicals commonly derived from fossil fuels. Mutations in the active site of these enzymes can alter the kinetics and spectrum of polyketide products formed[8,10], potentially accessing novel products (**Figure 2B, Figure S1**). We synthesize and transform our library into *Yarrowia lipolytica*, encapsulating and culturing single cells in 300 pL droplets to generate isogenic colonies (**Figure S2**). Culture expansion produces additional material compared to a single cell, boosting MS signal and providing accurate metabolomic data.

To perform microscale MS (µMS), we use a high-density plate comprising a glass slide etched with 10,000 wells, having 80 µm diameters and rounded bottoms (**Figure. 3A**); this shape concentrates the material to the center (**Figure S3**), enabling accurate µMS quantitation. Higher capacity slides with the dimensions of a MALDI plate can accommodate 100,000 wells (**Figure S4**). To maximize throughput, all wells must be loaded with one colony, which we accomplish with printed droplet microfluidics (PDM)[11] in ~30 min (**Figure S5**). During printing, we scan the colonies, dispensing only ones falling within a narrow cell density range (**methods**)[11].

Using this system, we print 9,000 library members and 1,000 reference strains to designated positions. Once loaded, the plate is dried, spray-coated with matrix, and subjected to MALDI-MS imaging (**methods**), providing data on all detectable metabolites from m/z 30 to 630. The results are reported as an 'image' in which each pixel comprises a 600-dimensional vector of signal amplitude for each mass-to-charge ratio (m/z) (**Figure S6-S8, Table S1**). Thus, the data can be thought of as representing a 600 'color' image, in which each slice reports the amplitude of the metabolite corresponding to the respective m/z. Importantly, MALDI's soft ionization preserves DNA[12], allowing PCR recovery of enzyme genes after MS analysis (**Figure S9**).
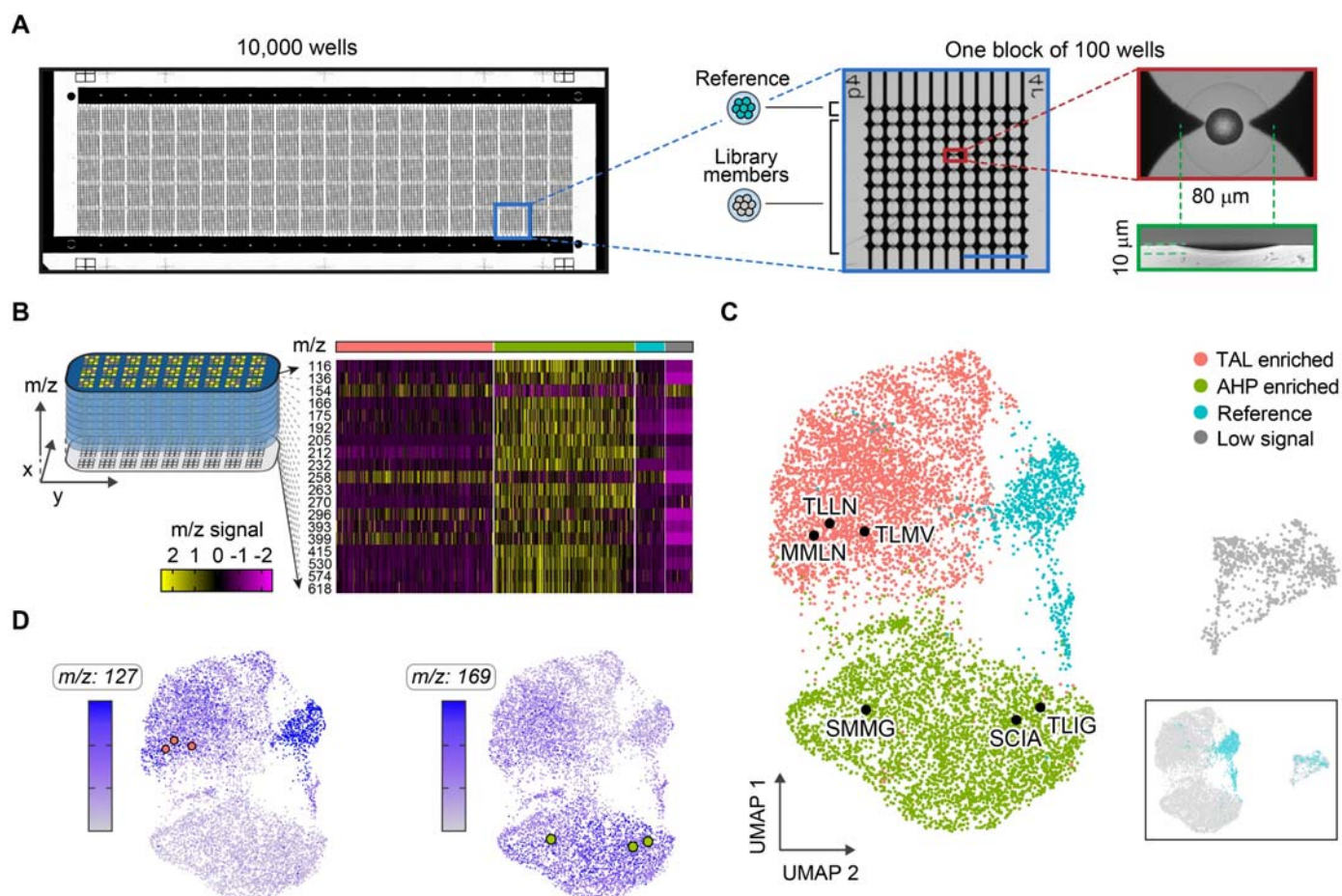
3

Fig. 3 | Mapping g2ps1 mutant catalysis through metabolic biosensing. a, Overview of slide (75 mm x 25 mm) for single variant µMS comprising 10,000 wells grouped in 100 blocks of 100 wells (magnified blue box, scale bar 1 mm. Each circular well is paired with electrodes for printed droplet trapping (red box) and has a rounded profile that concentrates desiccated metabolites to the center (green). **b,** µMS generates a spatial image of the substrate in which each pixel contains a full m/z spectrum. Well spectra are extracted and clustered into four groups using UMAP, **c.** Mutants of interest are recovered, sequenced, and overlaid on the clusters as black dots. Inset shows the reference strain based on known print locations. **d,** Heat maps for m/z 127 and 189 indicate heigh production of TAL in the upper and reference clusters, and high production of AHP in the lower cluster.

A unique and valuable property of µMS is that it provides information on many molecules in the host cells, enabling discovery of unexpected enzyme activities. A general way to identify activities is to plot alterations in the cell metabolome as a uniform manifold approximation and projection (UMAP), a dimensionality reduction technique used in single cell sequencing[13,14] and MS imaging[15] that projects high dimensional data onto a plane while preserving cluster information. To obtain the clearest UMAP clustering, we apply an algorithm to select the best m/z peaks for inclusion, comprising ~60% of the total m/z intensity (**Figure 3B, Figure S9, S10, Movie S1, S2**) (**methods**). From this, we observe four clusters (**Figure 3C**) which, visually, resemble UMAP clusters of single cell RNAseq profiles, except they correspond to metabolite profiles[7,16]. The compact blue cluster is the 1000 control reference strains, which we confirm by substrate location (**Figure 3C**, *inset*). The gray island corresponds to wells with few cells that were mis-printed (**Figure S11**). This leaves two large clusters (red and green) which, presumably, correspond to cells with distinct metabolomic profiles. Since only the embedded enzyme varies between these cells, this implies the codon shuffled library exhibits two major activities that perturb cell metabolism in distinct ways. Mapping the m/z 127 data onto the clusters shows that the upper red island represents productive TAL mutants (**Figure 3D**, *left UMAP*). In addition to the desired product, some mutants generate another product dominantly produced in the lower green cluster. Upon careful inspection of the MS data, we determine that a molecule with the m/z value of 169 has differential abundance between these clusters (**Figure S12, Table S1, S2**). Using HPLC with tandem MS (HPLC-MS/MS), we confirm that the mass corresponds to the reported alternative product of this enzyme,

4

tetraketide 6-acetonyl-4-hydroxy-2-pyrone (AHP)[10,17]. When we overlay the AHP amplitude (m/z 169) on the UMAP, it is predominantly found in the green cluster (**Figure 3D,** *right UMAP*).

To characterize how these disparate activities relate to enzyme sequence, we select and sequence ~50 mutants from each cluster (**Table S3, S4**). Plotting the mutants versus normalized TAL and AHP shows colonies with extreme activity and, generally, that mutants efficient at producing one product are inefficient at producing the other (**Figure 4A**). To validate these results, we re-transform and test several mutants in bulk, finding good agreement (**Figures 4B**, **Figure S13, Table S3**). Interestingly, the TLLL mutant based on consensus design[18] of TAL enriched members from the UMAP also shows high TAL production, though it retains higher AHP side-activity than the TLLN mutant (**Figure S14**). These results demonstrate that generation of these products can be inferred from clustering the host cells' metabolomic profiles, even though the peaks associated with these molecules (m/z 127 and 169) are not included in the clustering. In this way, the host cell's metabolite profile affords a biosensor by which to infer changes in the embedded enzyme's activity.
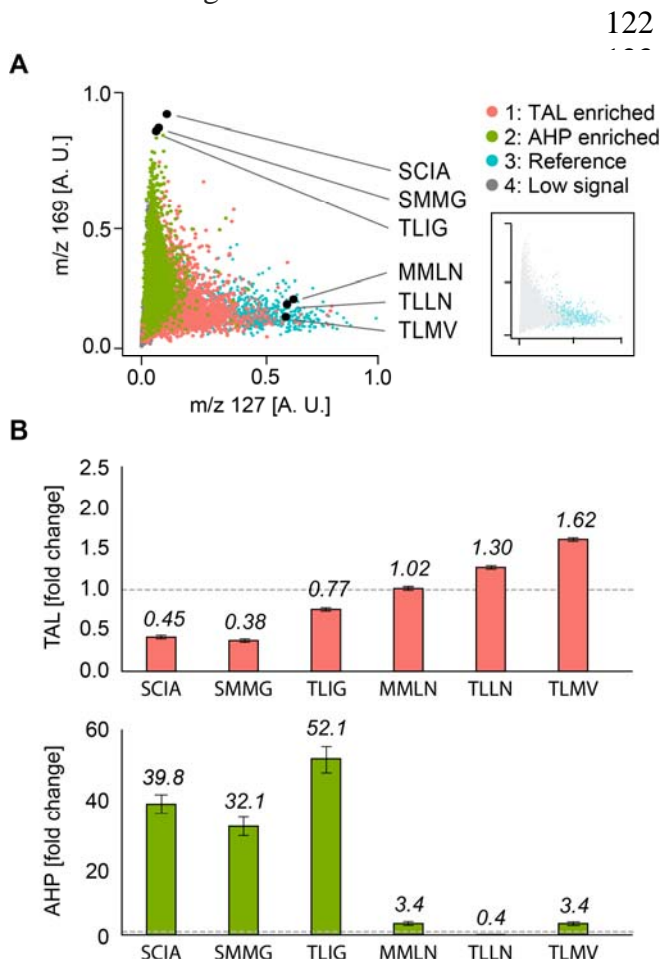
Discussion

Metabolomic biosensing allows inference of product formation even without direct product detection because the enzyme's activity perturbs the host cell's metabolite profile. Provided such perturbations are distinct for each activity, all activities generated by a mutant library can be detected, similar to the "compressed sensing" of smell sensation[19]. Such indirect sensing has beneficial features for enzyme engineering. In addition to providing a universal readout of product formation, the metabolites used to generate the UMAP can be selected to achieve the best cluster differentiation, even if the direct enzyme products are not included because they are difficult to detect due to their chemical or ionization properties. Furthermore, it should allow detection of all possible products, since each activity should perturb the host metabolite profile in unique ways and, thus, generate distinct clusters in the UMAP. By subsampling the clusters, unexpected products can be discovered. While we have utilized this analysis approach with yeast cells and MALDI-MS, it should apply to other bioproduction systems, including cell-free extracts, bacteria, and mammalian cells. Other readout modalities that provide sensitive, multiplexed detection of metabolites should be applicable, including other forms of mass spectrometry[20]

Fig. 4 | μMS screen of g2ps1 mutant library for increased production of TAL or AHP. a, Normalized m/z intensity for m/z 127 (TAL) versus 169 (AHP). Data points are colored based on clusters identified from UMAP analysis in Figure 2. Inset shows reference strain samples. The highest producers for TAL and AHP were isolated, sequenced, and confirmed in bulk analysis (black circles). **b,** TAL and AHP for selected bulk cultures as fold change over wild type (TLML).

and multiparametric spectroscopy[21]. Lastly, while we have focused on an enzyme library for its simplicity, the approach should apply to any engineering that perturbs host cell metabolism, including genetic circuits and biosynthetic pathways.

Author Contributions

Competing Interests

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/XXXXXXXXX.

References

1. Nielsen, J. Metabolic engineering. *Appl Microbiol Biotechnol* **55**, 263–283 (2001).

2. Petzold, C. J., Chan, L. J. G., Nhan, M. & Adams, P. D. Analytics for metabolic engineering. *Front. Bioeng. Biotechnol.* **3**, (2015).

3. Wunder, F., Kalthof, B., Muller, T. & Huser, J. Functional Cell-Based Assays in Microliter Volumes for Ultra-High Throughput Screening. *CCHTS* **11**, 495–504 (2008).

4. van Tatenhove-Pel, R. J. *et al.* Microdroplet screening and selection for improved microbial production of extracellular compounds. *Current Opinion in Biotechnology* **61**, 72–81 (2020).

5. Linster, C. L., Van Schaftingen, E. & Hanson, A. D. Metabolite damage and its repair or pre-emption. *Nat Chem Biol* **9**, 72–80 (2013).

6. Krismer, J. *et al.* A MALDI-MS Methodology for Studying Metabolic Heterogeneity of Single Cells in a Population. *Methods Mol Biol* **2064**, 113–124 (2020).

7. Si, T. *et al.* Profiling of Microbial Colonies for High-Throughput Engineering of Multistep Enzymatic Reactions via Optically Guided Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **139**, 12466–12473 (2017).

8. Jez, J. M. *et al.* Structural control of polyketide formation in plant-specific polyketide synthases. *Chem Biol* **7**, 919–930 (2000).

9. Markham, K. A. *et al.* Rewiring Yarrowia lipolytica toward triacetic acid lactone for materials generation. *PNAS* **115**, 2096–2101 (2018).

10. Abe, I., Oguro, S., Utsumi, Y., Sano, Y. & Noguchi, H. Engineered Biosynthesis of Plant Polyketides: Chain Length Control in an Octaketide-Producing Plant Type III Polyketide Synthase. *J. Am. Chem. Soc.* **127**, 12709–12716 (2005).

11. Cole, R. H. *et al.* Printed droplet microfluidics for on demand dispensing of picoliter droplets and cells. *PNAS* **114**, 8728–8733 (2017).

12. Kazdal, D. *et al.* Digital PCR After MALDI-Mass Spectrometry Imaging to Combine Proteomic Mapping and Identification of Activating Mutations in Pulmonary Adenocarcinoma. *Proteomics Clin Appl* **13**, e1800034 (2019).

13. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]* (2018).

14. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* **37**, 38–44 (2019).

15. Smets, T. *et al.* Evaluation of Distance Metrics and Spatial Autocorrelation in Uniform Manifold Approximation and Projection Applied to Mass Spectrometry Imaging Data. *Anal. Chem.* **91**, 5706–5714 (2019).

16. McCloskey, D., Xu, J., Schrübbers, L., Christensen, H. B. & Herrgård, M. J. RapidRIP quantifies the intracellular metabolome of 7 industrial strains of E. coli. *Metabolic Engineering* **47**, 383–392 (2018).

17. Jindaprasert, A., Springob, K., Schmidt, J., De-Eknamkul, W. & Kutchan, T. M. Pyrone polyketides synthesized by a type III polyketide synthase from Drosophyllum lusitanicum. *Phytochemistry* **69**, 3043–3053 (2008).

18. Porebski, B. T. & Buckle, A. M. Consensus protein design. *Protein Engineering, Design and Selection* **29**, 245–251 (2016).

19. Qin, S., Li, Q., Tang, C. & Tu, Y. Optimal compressed sensing strategies for an array of nonlinear olfactory receptor neurons with and without spontaneous activity. *PNAS* **116**, 20286–20295 (2019).

20. Dettmer, K., Aronov, P. A. & Hammock, B. D. Mass Spectrometry-based Metabolomics. *Mass Spectrom Rev* **26**, 51–78 (2007).

21. Liu, Z. *et al.* Mapping metabolic changes by noninvasive, multiparametric, high-resolution imaging using endogenous contrast. *Science Advances* **4**, eaap9302 (2018).