abundant taxa. Targeted gene surveys detect microbial diversity orders of magnitude higher than that seen in metagenomics studies, and metagenome-assembled genomes detect even less diversity.

The explosion of microbial sequencing data in recent years has shown that there are major microbial lineages for which no cultivated representatives are available[4,5]. Without a cultured representative of novel lineages to study, assigning function to individual genes requires additional innovation. Improved metagenome-assembled genomes will support these efforts. For example, metagenomics has been used to discover isolation and enrichment strategies for uncultivated microorganisms[9]. Metagenome-assembled genomes will also benefit the study of applied research problems, such as assessment of the antibiotic-resistance potential of environmental, agricultural and clinical samples[10]. Although shotgun metagenomics data can determine the resistance potential of the community as a whole, high-quality genomes can reveal which microbial genome has the resistance genes. Knowing whether resistance genes occur in the genomes of pathogens or in commensal microorganisms is of importance to health agencies.

The repurposing of read-cloud sequencing for microbial ecology is likely to open up exciting new possibilities in metagenomics. Characterizing the unculturable microbial members of the biosphere could yield insights into geobiology, climatology and microbial ecology as well as plant, animal and human health.

COMPETING INTERESTS
The authors declare no competing interests.

1. Bishara, A. *et al. Nat. Biotechnol.* **36**, 1067–1075 (2018).
2. Amann, R.I. *et al. Appl. Environ. Microbiol.* **56**, 1919–1925 (1990).
3. Parks, D.H. *et al. Nat. Microbiol.* **2**, 1533–1542 (2017).
4. Rinke, C. *et al. Nature* **499**, 431–437 (2013).
5. Rinke, C. *et al. Nat. Protoc.* **9**, 1038–1048 (2014).
6. Volodymyr, K., Snyder, M.P. & Batzoglou, S. *Bioinformatics* **32**, i216–i224 (2016).
7. Genome 10K Community of Scientists. *J. Hered.* **100**, 659–674 (2009).
8. Zheng, G.X.Y. *et al. Nat. Biotechnol.* **34**, 303–311 (2016).
9. He, X. *et al. Proc. Natl. Acad. Sci. USA* **112**, 244–249 (2015).
10. Su, J.-Q. *et al. Microbiome* **5**, 84 (2017).

Proteomics goes parallel

Ben C Collins & Ruedi Aebersold

Massively parallel sequencing of peptides could signal a new era of high-throughput proteomics.

Proteomics has yet to attain the power of genomics and transcriptomics. The impressive performance of technologies for nucleic acid sequencing rests on massively parallel measurements of short oligonucleotides, using fluorescence as a readout. In this issue, Swaminathan *et al.*[1] demonstrate that parallel fluorescence sequencing is also achievable for peptides. Their innovative method combines elements of classic protein chemistry with features of the optical systems used in nucleic acid sequencing. Although further optimization of the approach is needed, the study fascinates with the prospect of a generally accessible, reliable and truly universal proteomic technology.

Proteins are indispensable to living systems in their roles as chemical catalysts, structural components and mediators of physiological processes. The ability to accurately identify and quantify proteins would greatly contribute to the understanding of biology. Today, proteomes are frequently predicted or inferred from transcriptomes. It is well documented that the dependency between protein and mRNA levels is complex, and that predicting one from the other is imprecise and unreliable[2]. Why then are necessarily imprecise predictions from mRNA preferred over direct protein measurements in many instances? The answer lies in the state and accessibility of the respective measurement techniques: whereas essentially complete transcriptome analysis is readily available to biologists through core facilities and commercial providers, proteome analysis is still most effectively performed by expert labs and cannot easily reach the throughput, robustness and reproducibility of transcriptome analysis.

The first generation of DNA sequencers, which produced groundbreaking genome maps, was based on sequential sequencing of isolated DNA segments—an intrinsically slow and expensive process even with automation. Widely accessible genomic analysis became possible only with the development of methods that sequence millions of nucleic acid segments in parallel[3], allowing complete genomic maps to be generated at high throughput and coverage and at low cost. These commercially well-supported techniques have transformed biomedical research and become a mainstay of experimental biology.

Although 'top–down' proteomics approaches are emerging[4], proteins have traditionally been quantified and sequenced using 'bottom–up' methods. As in genomics, these methods analyze constituent segments—in this case, peptides generated by enzymatic cleavage of proteins. In the 1950s, Pehr Edman invented a cyclic process of chemical reactions, known as Edman degradation[5], to determine the amino acid sequence of peptides. It consists of the coupling of phenyl isothiocyanate to accessible amino groups followed by release of the derivatized N-terminal amino acid from the peptide chain, generating a new N terminus. The released amino acid is identified, and the process is repeated to establish the peptide sequence. The Edman process is slow and requires large amounts of highly purified peptides. Yet, essentially all protein sequences known until the early 1990s were determined with this process.

In the 1990s, mass spectrometry (MS) became the method of choice for protein sequencing, leaving Edman degradation in the realm of science history. MS techniques to infer protein identity and quantity from measurements of the mass-to-charge ratio and fragmentation pattern of peptide segments have become highly sophisticated, powerful and versatile, and are thus widely used[6]. Emulating the path of genomics, these techniques have progressed from manual sequencing of specific oligomers to automated, sequential sequencing of peptides at high throughput, and from there to parallel sequencing of multiple peptides by means of data-independent analyses[7,8], exemplified by SWATH-MS[9]. Although their throughput, accuracy and reproducibility are remarkable, the goal of achieving routine, complete proteome

Ben C. Collins and Ruedi Aebersold are in the Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland. Ruedi Aebersold is also in the Faculty of Science, University of Zurich, Zurich, Switzerland.
e-mail: aebersold@imsb.biol.ethz.ch or collins@imsb.biol.ethz.ch

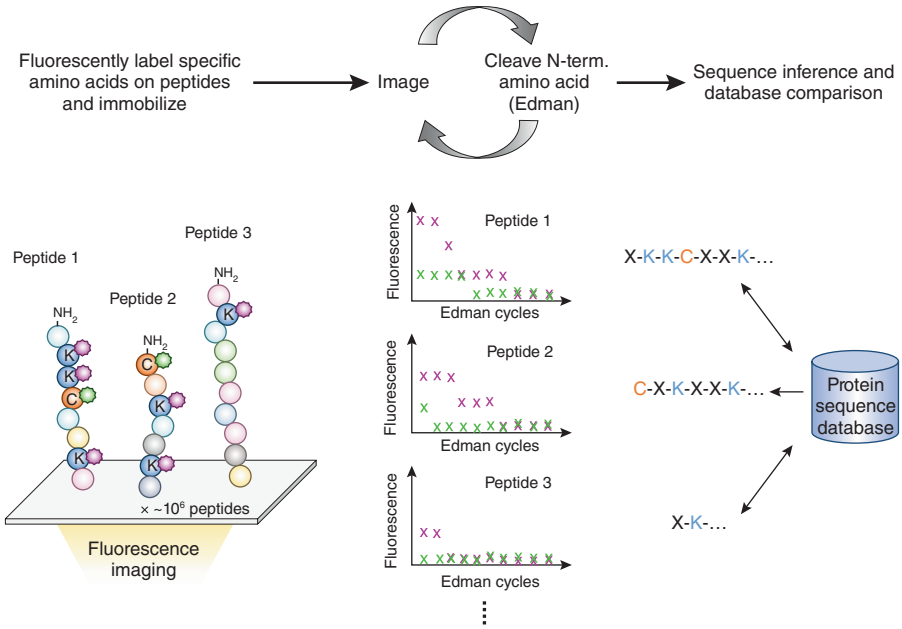

Figure 1 Peptide fluorosequencing as described by Swaminathan *et al.*[1] Complex peptide mixtures, most likely derived from enzymatic or chemical cleavage of protein extracts, are labeled with different fluorophores for each amino acid residue (left). In this case, we depict a two-color scheme where lysine and cysteine residues are labeled with distinct fluorophores. The labeled peptides are immobilized at their C termini using amide linkage to aminosilanes on a glass coverslip. The peptides are then subjected to iterative cycles of cleavage of the N-terminal amino acid residue by Edman degradation and fluorescence imaging (center). The fluorescence intensity at each location (i.e., peptide) is tracked as a function of Edman cycles. The pattern of drops in fluorescence intensity is interpreted to provide a partial sequence annotation for each peptide, which can be matched and scored against a protein sequence database to infer the most likely set of proteins present in the sample (right).

quantification of large sample cohorts, akin to genomic analyses, has remained elusive.

It is conceivable that continued advances within the current framework of data-independent-acquisition MS will eventually achieve a performance on par with that of genomics. But it is also possible that a full account of the complexity and depth of proteomes will require disruptive new technologies. Although nanopore sequencing of proteins has shown promise[10], the peptide fluorosequencing method of Swaminathan *et al.*[1] appears to be the most advanced example of such a disruptive approach with a clear path to routine use. It is a marriage across the ages—between the largely forgotten Edman degradation chemistry and the principles of massively parallel-in-space fluorescence imaging developed for next-generation DNA sequencing (**Fig. 1**).

The first step of the new method is to generate an array of sequencing substrates by fluorescently labeling peptides at specific amino acid side chains and immobilizing them at their C termini in the flow cell of a sequencing system. The immobilized peptides are then subjected to Edman degradation steps in parallel, and after each step the ensemble of immobilized substrates is imaged. In contrast to classic Edman degradation, in which the phenylthiohydantoin–amino acid conjugates eliminated at each step are identified, the stepwise degradation serves simply as a register to measure the decrease in fluorescence intensity caused by elimination of a labeled amino acid. The sequence of each immobilized substrate is inferred by relating the constraints derived from the observed fluorescence patterns to a protein sequence database using a sophisticated software tool developed for this purpose.

The study demonstrates the first steps toward feasibility of peptide fluorosequencing. Specifically, the authors (i) describe an imaging system compatible with the harsh conditions associated with Edman degradation chemistry, (ii) determine the precise position of fluorescently labeled lysine or cysteine residues in model peptides, (iii) characterize sources of error and inefficiencies in the system, (iv) simulate the potential to identify proteins from more complex proteomes and provide a computational framework to infer peptide sequences from the observed fluorescence patterns, and (v) localize a particular phosphorylated serine residue from a peptide containing multiple serine residues.

The peptide fluorosequencing method of Swaminathan *et al.*[1] is exciting because it highlights a clear path toward peptide, and conceivably protein, sequencing at very high throughput and reproducibility and potentially low cost. A substantial advantage of the system is that it capitalizes on a collection of well-characterized processes from other strategies (Edman chemistry, massively parallel DNA sequencing and MS-based computational strategies for sequence database searching) that may speed maturation from proof of concept to a routinely applicable method. Furthermore, the data generated by the method should bear some resemblance to the data produced by its massively parallel antecedents in the world of genomics and transcriptomics. This could accelerate the adoption of peptide fluorosequencing by the broader biological community, in contrast to MS-based proteomics technologies, whose uptake has arguably been slowed by their technical and computational difficulty.

As Swaminathan *et al.*[1] note, several technical and conceptual challenges must be overcome before the method can reach its full potential. The issues are mainly rooted in the nature of Edman chemistry and the complexity of the human proteome, and include the following: (i) even at the yield per degradation step shown in the paper (91–97%), the length of achievable peptide sequences is limited; (ii) because the sequencing yield is sequence dependent, challenging sequences, such as proline-rich stretches, may obscure the sharpness of the fluorescence patterns; (iii) the number of functional groups accessible to fluorescent labeling is limited to the chemically reactive groups in peptides, predominantly amino, carboxyl and sulfhydryl groups, thus capping the information content of the fluorescence patterns; (iv) modified residues will generally not be recognized unless they are specifically fluorescently labeled, and a specific labeling chemistry is known for only a small subset of modifications; and (v) the large dynamic range of the human cellular proteome (~10^7), along with the high number of peptides generated per protein by enzymatic digestion (~10^2) and the large number of open reading frames expressed per cell (~10^4), constitute an enormous analytical challenge, even disregarding proteoform diversity. For peptide fluorosequencing, meeting these challenges requires a level of substrate multiplexing that has not yet been achieved.

Although the system implemented by the authors is limited to the analysis of relatively simple sample mixtures, the path forward seems well laid out and is certainly one worth taking.

COMPETING INTERESTS
The authors declare no competing interests.

1. Swaminathan, J. *et al. Nat. Biotechnol.* **36**, 1076–1082 (2018).
2. Liu, Y., Beyer, A. & Aebersold, R. *Cell* **165**, 535–550 (2016).
3. Goodwin, S., McPherson, J.D. & McCombie, W.R. *Nat. Rev. Genet.* **17**, 333–351 (2016).
4. Toby, T.K., Fornelli, L. & Kelleher, N.L. *Annu. Rev. Anal. Chem.* **9**, 499–519 (2016).
5. Edman, P. *Acta Chem. Scand.* **4**, 283–293 (1950).
6. Aebersold, R. & Mann, M. *Nature* **537**, 347–355 (2016).
7. Venable, J.D., Dong, M.Q., Wohlschlegel, J., Dillin, A. & Yates, J.R. *Nat. Methods* **1**, 39–45 (2004).
8. Purvine, S., Eppel, J.T., Yi, E.C. & Goodlett, D.R. *Proteomics* **3**, 847–850 (2003).
9. Gillet, L.C. *et al. Mol. Cell. Proteomics* **11**, O111.016717 (2012).
10. Robertson, J.W.F. & Reiner, J.E. *Proteomics* **18**, e1800026 (2018).

An avatar for precision cancer therapy

Shumei Kato & Razelle Kurzrock

Screening patient-derived tumor cell cultures against a drug library is a promising adjunct to clinical decision-making.

Precision oncology aims to match patients to therapies on the basis of the genomic alterations in their tumors. This approach of combining molecular diagnostics with therapeutics has not only transformed the standard-of-care management for certain malignancies[1–3], but is also integral to treatment selection in pan-cancer, precision medicine clinical trials[4,5]. In clinical practice, however, such factors as intra- and intertumor molecular heterogeneity and complexity have sometimes led to disappointingly low response rates, with responses that can be short lived. Therefore, the adoption of a more personalized *N*-of-one strategy that also examines the functional effects of genomic alterations may be necessary to enhance the efficacy of drug selection. In a recent issue of *Nature Genetics*, Lee *et al.*[6] demonstrate the feasibility of using drug screening of patient-derived cell cultures (PDCs) to guide treatment choice for individual patients. They show that PDCs faithfully represent the molecular landscapes of the original diverse cancer types. Moreover, they exploit the PDC models to uncover new mechanisms of drug response and resistance for multiple targeted agents, and they illustrate how PDC screens can provide evidence for repurposing agents against additional cancers.

Lee *et al.*[6] derived a large number of tumor-sphere-forming PDCs (obtained directly from surgical specimens or malignant ascites) cultured in serum-free medium across 14 cancer types from 462 patients. The PDCs were dissociated into single cells and seeded into 384-well plates (500 cells per well) and treated with a 60-drug library targeting major oncogenic signaling molecules (**Fig. 1**). After six days of incubation, cell viability was assessed using an ATP monitoring system based on firefly luciferase. Comprehensive genomic and transcriptomic profiling demonstrated that the PDCs retained the molecular characteristics of their parental tumor tissues.

Previous studies have reported related approaches using patient-derived organoids (PDOs; self-organized, three-dimensional tissue cultures)[7], patient-derived xenografts (PDXs; tumor fragments engrafted into immunocompromised mice)[8] or patient-derived tumor cell cultures[9]. These models often (but not always) similarly recapitulate the molecular profiles of the parent tissue, as well as patient responses. The PDCs described by Lee *et al.*[6] differ in some ways from previous tumor cell cultures[9], which were first grown on fibroblast monolayers and which used immunofluorescent indicators to verify cellular origin. In comparing PDCs and PDOs, PDOs have the advantage of reflecting three-dimensional architecture and may include stromal cells, which may yield a more realistic recapitulation of cell-to-cell interactions than PDCs, since the latter are grown in monolayers; even so, PDOs have the disadvantage of being more complicated to develop and maintain than PDCs.

An important benefit of the PDC strategy is that it provides a rapid and facile readout of the functional effect on drug response that results from a complex array of genomic alterations in individual patients. PDCs deliver a faster timeline than PDX animal models, which require 6 to 7 weeks to become established versus 2 to 3 weeks for PDCs, and they are more amenable to large-scale, high-throughput drug screening. In the oncology clinic, screening for drug sensitivity in a timely fashion is critical because patients may not be able to wait several weeks for a treatment to be selected. Furthermore, patient tumors may evolve during longer time windows. There are also downsides to PDCs compared with PDX models in that PDCs do not include the tumor microenvironment, which can influence important factors such as angiogenesis. Finally, many current model systems, including PDCs, PDOs and PDXs, cannot adequately recapitulate the immune system.

The 60 different targeted agents tested by Lee *et al.*[6] are commonly used in the clinical setting. They include inhibitors of receptor tyrosine kinases, such as platelet-derived growth factor receptor (PDGFR), vascular endothelial growth factor receptor (VEGFR), and phosphoinositide 3-kinase (PI3K)–AKT–mammalian target of rapamycin (mTOR) (PAM), as well as histone deacetylase inhibitors and more. All in all, the authors studied 27,720 drug–PDC combinations (60 drugs × 462 PDCs), which revealed diverse patterns of drug sensitivities. From this matrix, they first noted that certain cancer types are more vulnerable to certain classes of inhibitor. For example, PDCs from patients with colorectal cancer or glioblastoma were more resistant to PAM pathway inhibitors; gastric cancer PDCs were more sensitive to these inhibitors. The authors also evaluated why gastric cancer PDCs were more sensitive to modulators of the PAM pathway using an independent database (The Cancer Genome Atlas) and found the PAM pathway to be more active in gastric cancer.

Next the investigators compared drug sensitivity and gene profiling results to determine whether specific gene markers in PDCs predict sensitivity versus resistance. This analysis led to robust, clinically relevant discoveries. Among several important observations, the small-molecule drug ibrutinib, which is approved by the European Medicine Agency (Amsterdam) and the US Food and Drug Administration (FDA) for patients with chronic lymphocytic leukemia, may also be efficacious for a subgroup of glioblastoma

Shumei Kato and Razelle Kurzrock are at the Center for Personalized Cancer Therapy and Division of Hematology and Oncology, Moores Cancer Center, University of California, San Diego, La Jolla, California, USA.
e-mail: rkurzrock@ucsd.edu