Single cell analysis: the new frontier in 'omics'

Daojing Wang[1] and Steven Bodovitz[2]

[1] Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[2] BioPerspectives, San Francisco, California, USA

Cellular heterogeneity that arises from stochastic expression of genes, proteins and metabolites is a fundamental principle of cell biology, but single cell analysis has been beyond the capability of 'omics' technology. This is rapidly changing with the recent examples of single cell genomics, transcriptomics, proteomics and metabolomics. The rate of change is expected to accelerate owing to emerging technologies that range from micro/nanofluidics to microfabricated interfaces for mass spectrometry to third- and fourth-generation automated DNA sequencers. As described in this review, single cell analysis is the new frontier in omics, and single cell omics has the potential to transform systems biology through new discoveries derived from cellular heterogeneity.

Single cell analysis: needs and applications

Cellular heterogeneity

Cellular heterogeneity within an isogenic cell population is a widespread event [1,2]. Stochastic gene and protein expression at the single cell level has been clearly demonstrated in different systems using a variety of techniques [3–5]. Therefore, analyzing cell ensembles individually with high spatiotemporal resolutions will lead to a more accurate representation of cell-to-cell variations instead of the stochastic average masked by bulk measurements. The disconnect between single cell and average cell measurements is exemplified in Figure 1a. Using an integrated microfluidic bioprocessor for single cell gene expression analysis, the Mathies group has shown that siRNA knockdown of GAPDH gene expression led to two distinct groups of individual Jurkat cells with partial knockdown ($\sim$50%) and complete knockdown ($\sim$0%). The average result from 50 cells ($\sim$21%) was not representative of any one individual cell [6].

To understand fully the cellular specificity and complexity of tissue microenvironments under physiological conditions, it is necessary to measure molecular signatures with single cell resolution. A clear example is provided by the recent work from Kim and colleagues, who analyzed single cell gene expression profiles using high-resolution confocal microscopy and correlated them with known cell lineages in *Caenorhabditis elegans* [7]. The group generated expression profiles of 93 genes in 363 specific cells from L1 stage larvae. Cells were clustered into groups in a 2D scatter plot according to their correlation in gene expression (Figure 1b). Two features of the scatter plot stand out: (i) cells are diverse, but cluster with known fates such as muscles and neurons; and (ii) cells from homogeneous tissue (e.g. intestinal cells) cluster more tightly than those from heterogeneous tissue (e.g. neurons). However, even within the homogeneous tissue, individual cells show clear heterogeneity. Single cell analysis (SCA) will therefore be crucial for elucidating cellular diversity and heterogeneity.

Potential applications of SCA

At the simplest level, SCA reduces biological noise. It provides fundamental improvements in experimental design and data analysis for applications predicated on single cells. Stem cells, for example, hold great potential for regenerative medicine because they can self-renew and differentiate along different lineages. However, embryonic stem cells, adult stem cells and induced pluripotent stem cells are all heterogeneous populations [2,8,9]. SCA can target specific populations and therefore elucidate signaling pathways and networks for self-renewal and for differentiation. Cancer is a heterogeneous disease and dissecting cell-to-cell variations is extremely important in understanding tumor initiation, progression, metastasis and therapeutic responses. For example, cancer cells have recently been shown to harbor homoplasmic and heteroplasmic mutations in mitochondrial DNA (mtDNA) that are above and beyond the widespread heterogeneity of mtDNA in normal human cells [10]. SCA might functionally differentiate normal and cancer cells, and cancer cells at various developmental stages. The key therapeutic targets in tumors might be cancer stem cells [11], which represent a small percentage of the total mass of tumors, but could be responsible for tumor repopulation following treatment [12]. SCA has the potential to identify more accurately these cells and their unique susceptibility. Neurons are the basic unit of the nervous system. Their electrical properties have been well characterized by electrophysiology, but the molecular complexity of ion channels, neurotransmitters and neuropeptides at the single cell level and even the subcellular level is only beginning to be understood [13]. SCA can help elucidate neural communication in unprecedented detail, which could yield new strategies to understand and treat neurological disorders.

Advances in SCA have the potential to accelerate not only biological research, but also diagnostics. Preimplantation genetic diagnosis (PGD), for example, is the analysis of a single cell from a biopsy of an embryo after *in vitro*

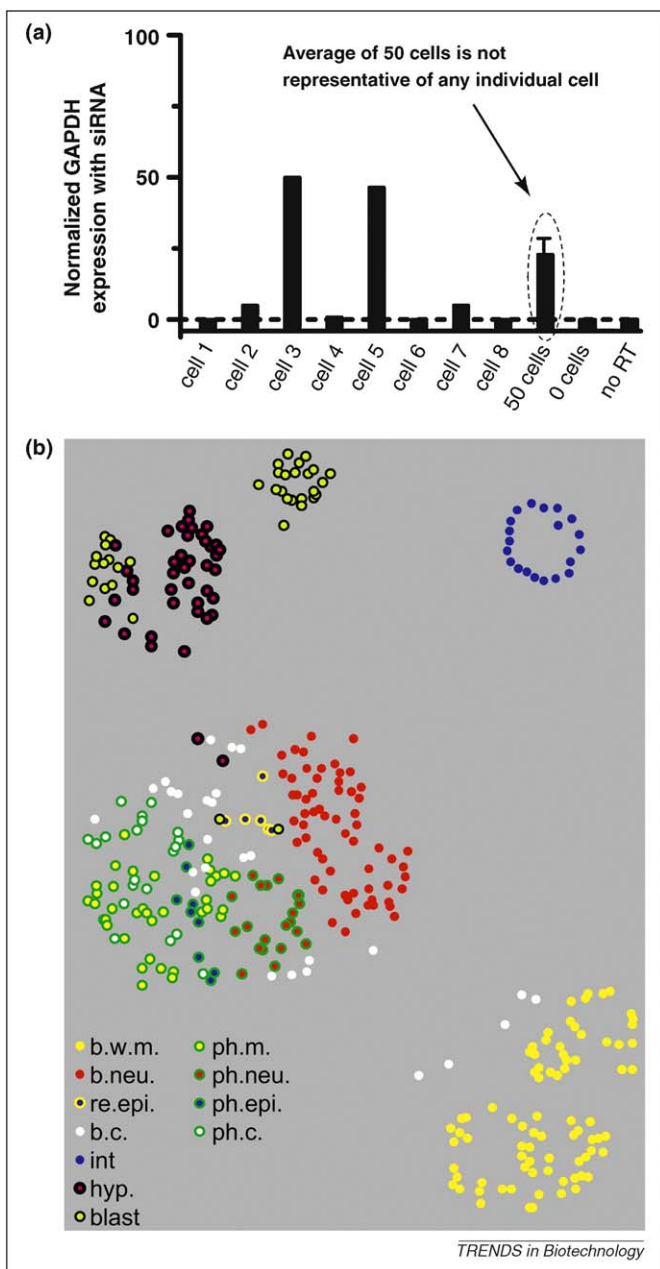*Corresponding author:* Wang, D. (djwang@lbl.gov).

Figure 1. Observation of cellular heterogeneity. **(a)** mRNA expression of GAPDH from individual Jurkat cells after siRNA knockdown. The levels fall into roughly two categories: 50% and 100% knockdowns (i.e. 50% and 0% expression remained). Importantly, the average GAPDH expression obtained from the measurement of 50 cells (21±4%) was not representative of any individual cell. Adapted with permission from Ref. [6]. **(b)** Clustering diagram of 363 cells in *C. elegans* according to similarities between gene expression profiles. Distance between cells in the x–y plane indicates levels of similarity. Colors indicate different tissue types. b.w.m., body wall muscle; b.neu., body neurons; re.epi., rectal epithelial cells; b.c., other body cells; int., intestine cells; hyp., hypodermal cells; blast, blast cells; ph.m., pharyngeal muscle; ph.neu., pharyngeal neurons; ph.epi., pharyngeal epithelial cells; ph.c., other pharyngeal cells. Adapted with permission from Ref. [7].

fertilization. PGD is used to test for genetic diseases and chromosome aneuploidies. The most common molecular analysis techniques are PCR and fluorescent *in situ* hybridization (FISH) [14], and newer methods, such as comparative genomic hybridization (CGH) [15], are beginning to be utilized. As more genetic diseases and predispositions are identified, multiplexed PGD using single cell omics tech-

nology could enable early and accurate identification and possible treatment of genetic abnormalities in human embryos.

Single cell omics: current state-of-the-art
Omics for single cell analysis
Omics is the large-scale study of genes (genomics and epigenomics), transcripts (transcriptomics), proteins (proteomics), metabolites (metabolomics), lipids (lipidomics) and interactions (interactomics), and is both the foundation and the driving force for systems biology [16]. Omics strives to identify, quantify and characterize all of the components in cellular systems with spatiotemporal resolution, and thereby dissect the intracellular pathways and networks. Currently, major technologies for genomics, epigenomics and transcriptomics include DNA sequencing and microarrays (planar-, bead- and fiber-optic arrays); for proteomics, mass spectrometry (MS) and protein arrays; and for metabolomics, MS and NMR.

The above-mentioned major technologies have undergone dramatic advances in automation and miniaturization. The logical limit to the ever-increasing miniaturization of omics technology is the comprehensive spatiotemporal analysis of genes, transcripts, proteins, metabolites and interactions in single cells and their subcellular compartments. The movement of omics into SCA represents a significant shift. Previous well-established methods for SCA, such as imaging and flow cytometry, are limited to the examination of a small number of genes, proteins or metabolites. As a result, these methods can only be used to open narrow windows into the complexity and dynamics of intracellular pathways. By contrast, single cell omics has the potential to enable systems biology at the level of single cells, ultimately representing the unique convergence of technology and biology.

Nucleic acid analysis
Nucleic acid analysis has already reached the single cell level [4,17–21]. Newer technologies, such as 'lab-on-a-chip' [22], have further advanced the field. For example, Mathies and colleagues have shown a hybrid (Polydimethylsiloxane (PDMS)/glass) microfabricated bioprocessor for single cell gene expression analysis [6] (as discussed above). In addition, Quake and colleagues have used PDMS structures and on-chip PCR to realize large-scale integration for applications in single cell mRNA isolation and analysis [23]. Quake's technology has been adopted by Fluidigm Corporation, thus demonstrating the commercial promise of miniaturization for single cell omics. The Fluidigm Dynamic Array^{TM} enables gene expression measurements in individual cells, and represents the first dedicated commercial product for SCA. This technology, however, needs greater multiplexing capability for global transcriptomic analysis. It currently can test up to 96 individual cells against 96 genes. The number of genes needs to increase by a minimum of two orders of magnitude for routine transcriptome profiling.

Further improvement in the breadth of single cell mRNA analysis has been achieved recently using mRNA sequencing (mRNA-Seq) [24] or quantitative PCR (qPCR) [25]. In the first study, the Surani and Lao groups together

developed an assay with mRNA-Seq [26] for single cell whole-transcriptome analysis. In this method, single cell cDNA amplification [27] was improved from a previous version by increasing the cDNA length up to 3 kb. Briefly, the mRNAs from a single cell were reverse-transcribed into cDNAs, which were subsequently amplified by PCR. Finally, amplified cDNAs were digitally counted with an ultra-high-throughput DNA sequencer. The researchers first verified reproducibility of the method, and then demonstrated its utility in dissecting functional consequences of *Dicer1* and *Ago2* knockdowns, two of the crucial genes for miRNA synthesis and mRNA regulation, by comparing single mouse *Dicer1*$^{-/-}$ and *Ago2*$^{-/-}$ oocytes versus wild-types (Figure 2a). The described method has the added advantage over probe-based methods (such as qPCR) in that it is effective at locating splicing variants, including *de novo* variants, but could face challenges in accurate quantitation as a result of the required amplification process.

To improve mRNA quantitation from single cells, the Kambara group has performed direct qPCR from a cDNA pool without pre-amplification [25]. Gene expression for four house-keeping genes (TBP, SDHA, B2 M and EEF1G) was quantitatively analyzed in 14 single cells, which showed that the absolute amounts of each of the four genes differed from cell to cell (Figure 2b). They further showed that the average number of cDNA molecules obtained from single cells was comparable to the per-cell number measured from the cell pools (10–1000 cells); however, a much larger standard deviation was observed for the single cells, as expected from the cellular heterogeneity. The research group is currently developing a higher-through-put methodology to analyze more individual cells (personal communication). A larger sample size will enable better statistical analysis of cell-to-cell variation. The method would also benefit from increased multiplexing of the capture beads for the immobilized cDNA library to increase transcriptome coverage.

Single cell genomics has been achieved for microorganisms. Church and colleagues have sequenced the genome of single cells of *Escherichia coli* and *Prochlorococcus* through so-called polymerase cloning ("ploning"), an optimized version of multiple displacement amplification (MDA) using random primers [28]. They obtained ~65% sequence coverage at an average depth of fourfold for a single *Prochlorococcus* cell through whole-genome shotgun sequencing. However, the method can suffer background contamination during the amplification of single cell DNA and difficulty in assembling the genomes. Compared to microorganisms, which typically have a genome size of several million base pairs and a few thousand genes, mammalian genomes are much more complex. For example, the human genome has three billion base pairs and an estimated 20 000–25 000 genes. To the best of our knowledge, there has been no report yet on single cell genomics (i.e. global analysis of chromosomal DNA) or epigenomics (i.e. global analysis of DNA methylation) for mammalian cells. However, this could change soon with the recent introduction of several next-generation sequencing technologies (discussed in the next section), particularly those centered on single-molecule sequencing.

Protein and metabolite analysis

Proof-of-principle SCA has been performed to characterize peptides using matrix-assisted laser desorption ionization (MALDI) MS [29–31], and to characterize small molecules using electrospray ionization (ESI) MS [32–34]. The Sweedler group previously has profiled peptides directly from single neurons and individual organelles using MALDI-MS [35,36]; however, the number of peptides detected was limited. Recently, the group was able to couple capillary electrophoresis with ESI-MS (CE-ESI-MS) and achieved reproducible metabolomic analysis of an isolated single R2 neuron and its different subcellular regions [33]. As shown in Figure 3a, different subcellular regions of a single R2 neuron (soma *vs.* neurite) have different relative amounts of several metabolites with the extracted ion of 146 m/z. However, this method requires offline manual isolation and lysis of single cells in a vial before the front-end CE separation followed by MS, thereby posting a significant challenge for *in situ* metabolomic analysis. In addition, CE sample injection is dictated by capillary inlet and outlet of the sampling interfaces; currently, only 0.1% of the total content of a single metacerebral cell, corresponding to 6 nl of the cell lysate, was injected (although more than 100 compounds were detected). Increasing the injected amount of cellular contents can improve the metabolome coverage from a single cell.

More recently, laser ablation electrospray ionization mass spectrometry (LAESI-MS) has been demonstrated for *in situ* metabolomic profiling of single cells [34] (Figure 3b). Single cell ablation was achieved by delivering mid-IR laser pulses through a GeO_2-based glass fiber (15 μm radius at the tip). When comparing two physically adjacent, individual *Allium cepa cultivar* cells with differences in purple pigmentation, LAESI-MS revealed similar essential metabolites for both variants, with colorless cells that contained anthocyanin and pigmented cells that contained anthocyanidins, other flavonoids and their glucosides. One of the key advantages of this method is that single cells can be analyzed *in situ*, although, as mentioned by the authors, further improvement in optics and ablation geometry is needed for cell-by-cell analysis of biological tissue. Another challenge for the LAESI-MS is the lack of front-end separation of cellular contents using either LC or CE, because laser ablation is performed directly on whole cells. This hinders further improvement in metabolome coverage and identification.

Previously, flow cytometry (i.e. fluorescence-activated cell sorting) has been used to profile phosphoprotein networks in single cancer cells [37] and monitor global protein levels in single *Saccharomyces cerevisiae* cells [38]. In addition, fluorescence imaging has been used to measure the levels and locations of nearly 1000 different endogenously tagged proteins in individual living cancer cells at high temporal resolution [39]. Current technology based on fluorescence requires prior *in vivo* or *in vitro* labeling, and is generally limited to 10–20 simultaneous measurements because of the spectral overlap. For *in vitro* labeling, background noise that arises from cross-reactivity and/or nonspecific binding remains a challenge. For *in vivo* labeling, tagging with green or yellow fluorescent proteins (GFP or YFP) and their derivatives might alter
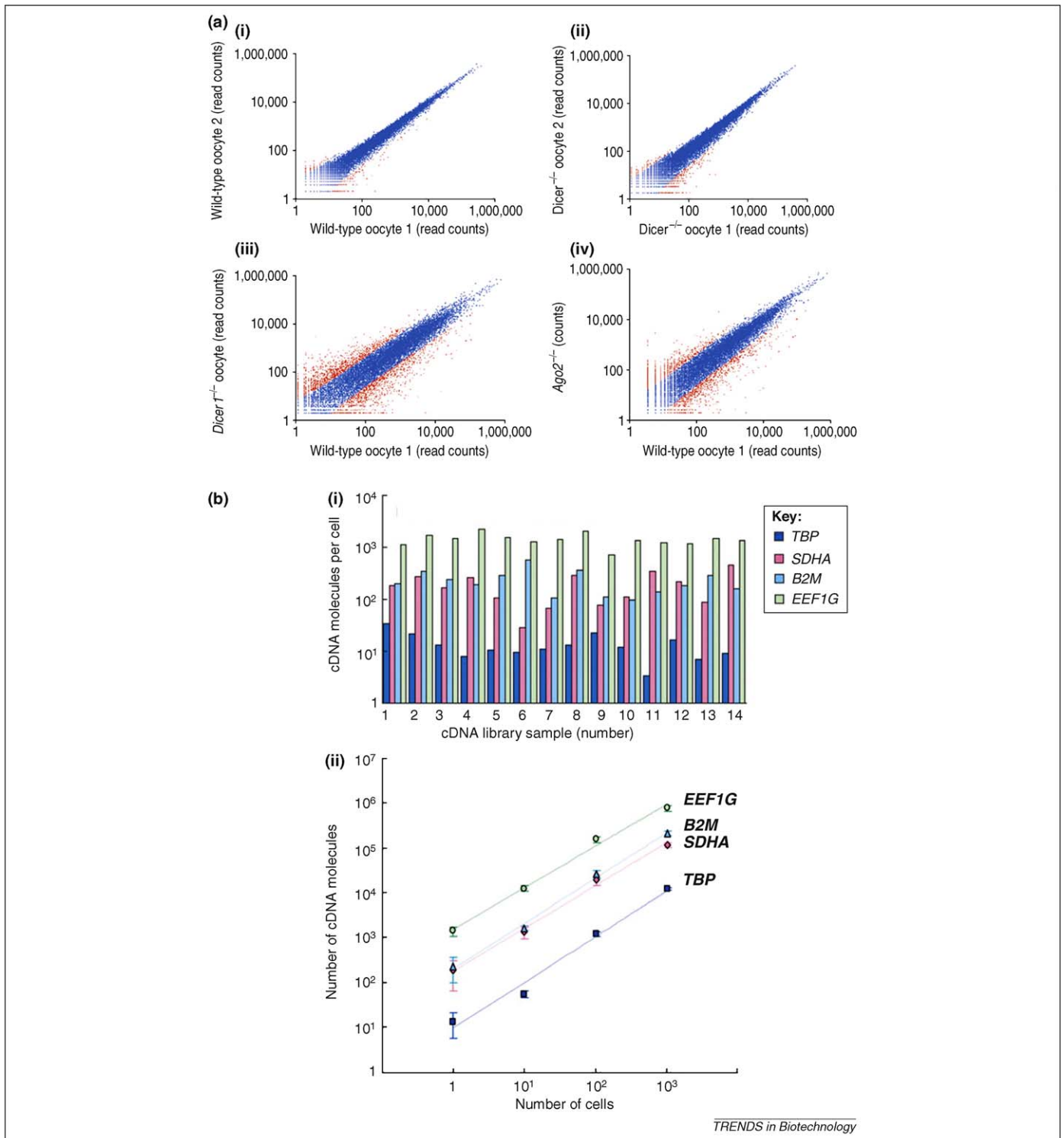
Figure 2. Single cell gene expression analysis. **(a)** Correlation plots of the quantile-normalized mRNA sequencing reads for mouse oocytes, showing **(i)** one wild-type oocyte versus another wild-type oocyte; **(ii)** one *Dicer1*⁻/⁻ oocyte versus another *Dicer1*⁻/⁻ oocyte; **(iii)** one wild-type oocyte versus one *Dicer1*⁻/⁻ oocyte; and **(iv)** one wild-type oocyte versus one *Ago2*⁻/⁻ oocyte. All of the reads with changes of greater than fourfold are plotted in red. Adapted with permission from Ref. [24]. **(b)** Number of cDNA molecules measured by qPCR in single human colon carcinoma HCT116 cells and cell pools, showing **(i)** cell-to-cell variation in expression of housekeeping genes among 14 single cells, and **(ii)** average gene expression for single cells as well as cell pools with error bars (single cell: mean ± SD, *n*=14; 10–1000 cells: mean ± SD, *n*=5). Adapted with permission from Ref. [25].

folding or other characteristics (e.g. interactions with other proteins) of the endogenous proteins. Furthermore, photobleaching of fluorophores and autofluorescence of cells might interfere with kinetic measurements. A potential solution is MS-based flow cytometry ("mass cytometry"), which has been described for a single cell multi-

target immunoassay [40]. However, the method still requires labeling of antibodies with elemental tags (e.g. lanthanide) before antigen binding.

MS-based single cell proteomics has not been achieved yet despite great interest in the field. Recently, an LC-MS/MS-based method has been described for quantitative
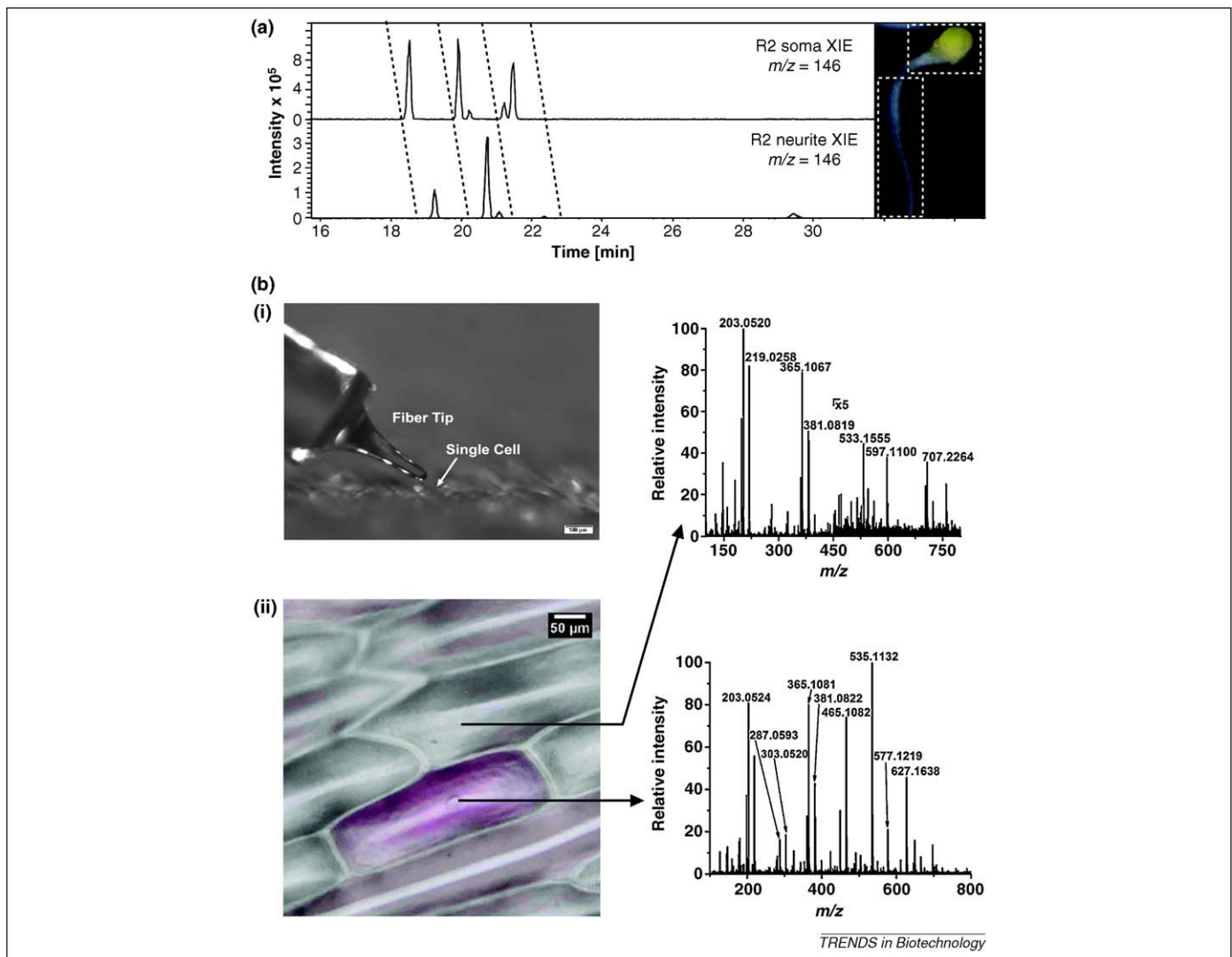
Figure 3. Single cell metabolome analysis. **(a)** Metabolomic profiling of a single *Aplysia* R2 neuron using CE-ESI-MS, showing the extracted ion electropherogram (XIE) obtained for 146 m/z from different subcellular regions (i.e. soma *vs.* neurite). Adapted with permission from Ref. [33]. **(b)** Metabolomic profiling using LAESI-MS, showing **(i)** etched optical fiber for laser ablation relative to the target single cell (scale bar = 100 μm), and **(ii)** optical image of neighboring individual colorless and pigmented epidermal cells of the purple *A. cepa cultivar* (scale bar = 50 μm). Also shown are the corresponding LAESI-MS spectra. Selected similar and different peaks (m/z) are indicated by arrows in the lower spectrum. Adapted with permission from Ref. [34].

proteomic analysis of single pancreatic islets that contain 2000–4000 cells [41]. A combination of improved chromatographic methods, direct replicate measurements, and high accuracy (1–2 ppm) and sensitivity (attomole to femtomole) provided by an LTQ-Orbitrap mass spectrometer, allowed for detection of >6000 proteins from nanogram quantities of protein mixtures. This ultrasensitive method represents an important step toward single cell proteomics. However, as a result of the extreme complexity and huge dynamic range of the proteome, new breakthroughs in sample manipulation and detection are needed to achieve *in situ* single cell proteomics with deep proteome coverage (discussed in the next section).

Single cell omics: future prospects
Technology for analyzing nucleic acids, proteins and metabolites is evolving rapidly. This section includes notable examples that have the potential to make major contributions to single cell omics.

New technologies for nucleic acid analysis
Driven by personal genomics and the National Institutes of Health-launched "$1000 Genome" project, the technologies for nucleic acid analysis, particularly next-generation sequencing technologies, has undergone dramatic advances [42–47]. Figure 4 showcases some of the technology that could contribute to SCA.

Geiss and colleagues from NanoString Technologies, for example, have developed the nCounter gene expression system (Figure 4a) for direct multiplexed measurement of gene expression with color-coded probe pairs without amplification [42]. The technology utilizes molecular barcodes and single-molecule imaging to detect and count hundreds of unique transcripts directly in a single reaction. The company has demonstrated that the nCounter system is more sensitive than microarrays and similar in sensitivity to real-time PCR. The technology has the potential for single cell transcriptomics.

In addition, multiple breakthroughs in single molecule sequencing technology could pave the way for single cell
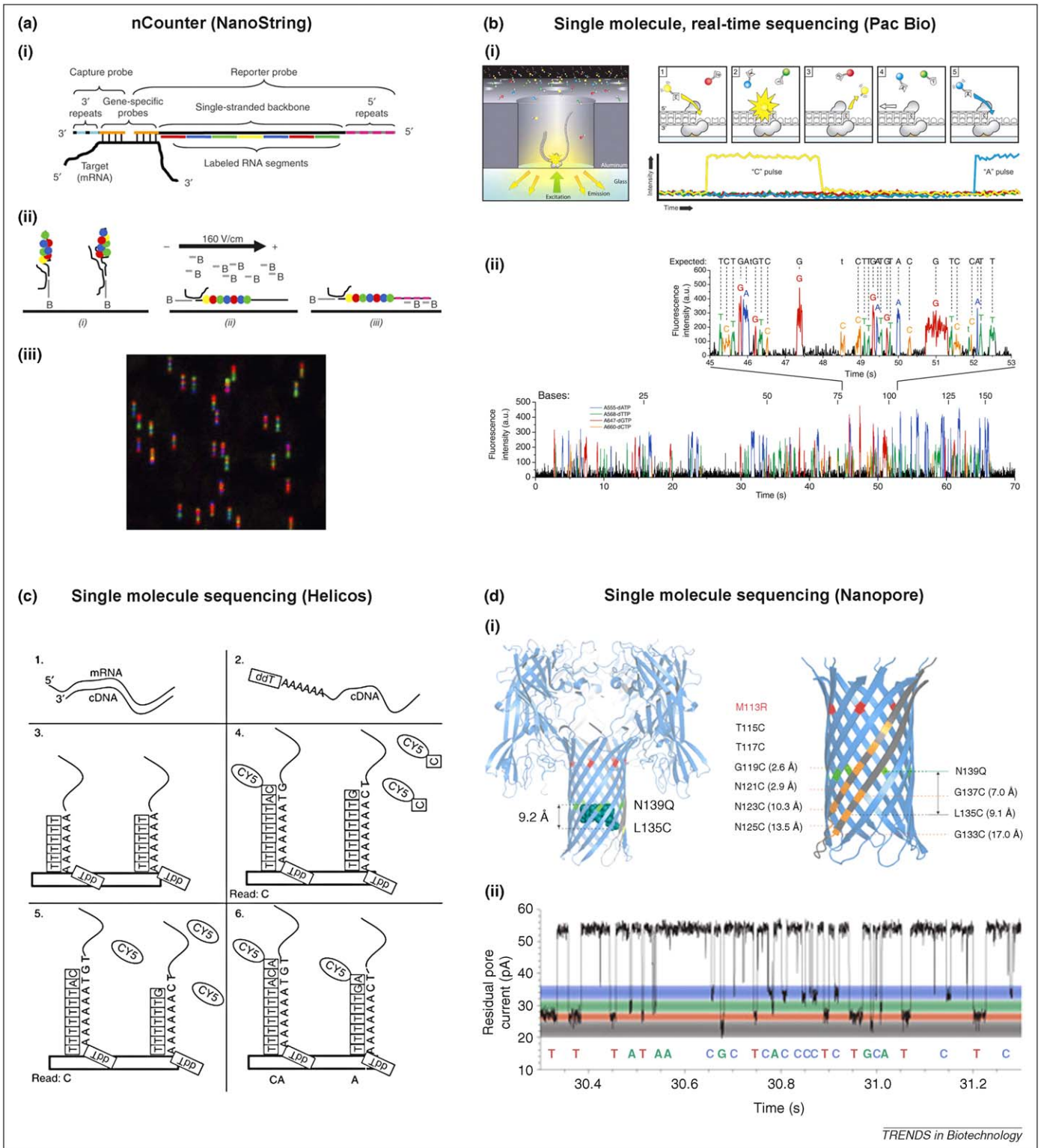
Figure 4. New technologies for nucleic acid analysis. **(a)** nCounter gene expression system from NanoString Technologies, showing **(i)** schematic representation of the hybridized complex (not to scale) and **(ii)** schematic representation of, from left to right, binding, electrophoresis and immobilization; and **(iii)** false-color image of immobilized reporter probes. Adapted with permission from Ref. [42]. **(b)** Single-molecule, real-time DNA sequencing system from Pacific Biosciences, showing **(i)** schematic representation of the experimental geometry with a single molecule of DNA template-bound DNA polymerase immobilized at the bottom of a zero-mode waveguide (ZMW) (left) and a schematic event sequence of the phospholinked dNTP incorporation cycle, with a corresponding expected time trace of detected fluorescence intensity from the ZMW (right); and **(ii)** total intensity output of all four dye-weighted channels with pulses colored corresponding to the least-squares fitting decisions of the algorithm and the entire read that proceeds through all 150 bases of the linear templates. Adapted with permission from Ref. [44]. **(c)** Single-molecule sequencing digital gene expression system from Helicos Biosciences, illustrating sample preparation and sequencing workflow: (1) preparation of the first-strand cDNA from mRNA; (2) addition of 3′ tail of dATP followed by dideoxy-TTP (ddT) blocking; (3) hybridization of tailed sample to poly-dT oligonucleotide covalently attached to the flow-cell channel surface; (4) sequencing of a single base by adding a Cy5-labeled nucleotide; (5) cleaving off the Cy5 dye label; and (6) adding and imaging of next nucleotide. Adapted with permission from Ref. [45]. **(d)** Single-molecule nanopore sequencing system from Oxford Nanopore Technologies, showing **(i)** nanopore structure of the WT-(M113R/N139Q)$_6$(M113R/N139Q/L135C)$_1$ mutant with the cyclodextrin covalently attached at position 135 (space-filling model) (left) and a close-up of the β barrel with arginine at position 113 and the location of cysteine in the mutants tested in the study (right); and **(ii)** single-channel recording from the nanopore that indicates discrimination of dGMP, dTMP, dAMP and dCMP, with colored bands representative of the residual current distribution for each nucleotide. Adapted with permission from Ref. [46].

genomics, epigenomics and transcriptomics. Turner and Korlach together with colleagues from Pacific Biosciences have developed a technology for single-molecule and real-time DNA sequencing by a single DNA polymerase (Figure 4b) [44]. They have detected the enzymatic incorporation of fluorescently labeled dNTPs into the growing DNA strand with zero-mode waveguide nanostructure arrays. They have demonstrated a median accuracy of 99.3%, with no systematic error beyond the fluorophore-dependent error rates, with the consensus sequences generated from the single-molecule reads at 15-fold coverage. Optimism for this technology is high: the company has raised approximately $266 million USD to date, through a combination of grants and venture capital, according to the company website.

Traz and colleagues from Helicos Biosciences have developed a high-throughput, amplification-free method for transcriptome quantification, named single-molecule sequencing digital gene expression (smsDGE) (Figure 4c) [45]. The technology utilizes a reverse-transcription and polyA-tailing sample preparation procedure followed by single-molecule sequencing that generates a single read per transcript as published earlier [43]. They have confirmed accurate quantification using spiked-in RNAs, and have demonstrated sequencing of the yeast transcriptome in a single run, which yielded an average of 12 million aligned reads per channel. Although they have not used this technology for SCA, they recently have published a ChIP-Seq study in collaboration with the Bernstein group using a small number of cells and only 50 pg of DNA [48].

Bayley and colleagues from Oxford Nanopore Technologies have developed a single-molecule sequencing technology using nanopores (Figure 4d) [46]. The nanopores are constructed from hemolysin mutants that are covalently attached to an adaptor molecule to detect the bases cleaved from the DNA by an exonuclease enzyme. The researchers have demonstrated the continuous identification of unlabelled nucleoside 5'-monophosphate molecules with an average 99.8% accuracy. Methylated cytosine has been distinguished from the four standard bases. At the present time, the company analyzes multiple copies of nucleic acids, but is theoretically capable of single-copy detection, although longer sequences would be required because of diffusion (personal communication). Moreover, the platform could be adapted for analysis of proteins, other polymers and small molecules, potentially accelerating multiple phases of SCA.

Table 1 summarizes the key technologies and other attributes of five major third-generation DNA sequencing companies, including the three mentioned above (Helicos,

Pacific Bio and Oxford Nanopore) and two others (Complete Genomics and Ion Torrent). In addition, Life Technologies is actively developing a single-molecule sequencing technology using quantum dots, and Cipriany *et al*. recently have demonstrated a method for single-molecule epigenetic analysis using nanofluidics and multi-color fluorescence microscopy to detect DNA methylation [49]. (For more information on next-generation sequencing technologies, please refer to a recent review [50].) Further improvements in detection sensitivity, sample manipulation and recovery, and sequence assembly will enable the above-mentioned technologies to become suitable for direct single cell omics applications.

New technologies for protein and metabolite analysis
The holy grail of proteomics and metabolomics is to decode the proteome and metabolome at the single cell level. This is challenging because the amount of proteins and metabolites is extremely limited in a single cell (i.e. average 1×10^5 molecules for proteins). Furthermore, there are no amplification methods for proteins and metabolites comparable to those for nucleic acids. We believe that the most prevalent bottleneck for SCA of proteins and metabolites *in situ* is the marriage between extremely efficient sample manipulation and highly sensitive detection. Along this line, we believe micro/nanofluidics interfaced with MS will be hotly pursued for single cell proteomics and metabolomics. Figure 5 showcases some promising technologies that could contribute to this endeavor.

To achieve single cell omics, it is crucial to select and manipulate single cells efficiently. Laser capture microdissection has been used routinely to isolate a small number of cells from tissue [51]. In addition, micropipette-based micromanipulators have been used to manipulate single cells. However, both of these technologies are difficult to implement in a micro/nanofluidic device, particularly for living single cells. Newer technology based on droplet encapsulation could circumvent this challenge. For example, Chiu and coworkers have developed femtoliter- and picoliter-sized droplets as nanolabs for manipulating single cells and subcellular compartments [52]. Figure 5a shows one of their PDMS microfluidic devices for thermoelectric manipulation of single cells [53]. In this device, the T-channel is used for droplet generation and freezing above where a thermoelectric cooler is placed. The same group earlier has combined optical trapping, microfluidic encapsulation and rapid laser photolysis, and demonstrated selective encapsulation of single cells and analysis of subcellular organelles through fluorescence spectroscopy [54].

Table 1. A comparison of representative third-generation DNA sequencing companies

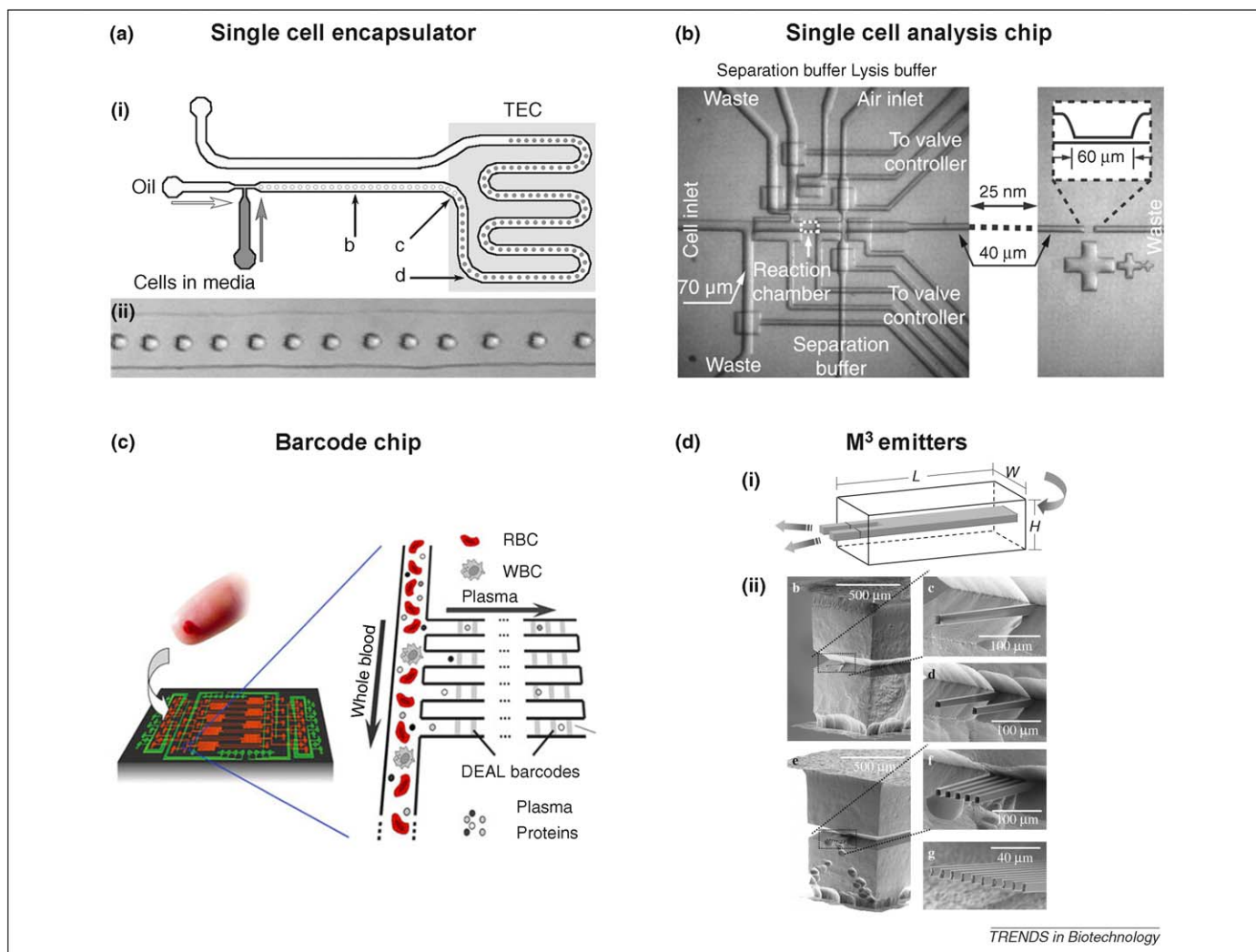|	Helicos	Pacific Biosciences	Oxford Nanopore	Complete Genomics	Ion Torrent
Key technology	Amplification-free sequencing	Zero-mode waveguide nanostructure arrays	Protein nanopores	Self-assembling DNA nanoarrays	Chemical sensitive field-effect transistor arrays
Single-molecule detection	Yes	Yes	Yes	No	Undisclosed
Commercialization	Launched in 2008	Launched in March 2010	Undisclosed estimated launch	Sequencing services	Launched in March 2010
Funding to date (millions USD)[a]	$115	$266	$64	$45	$23
Refs	[43]	[44]	[46]	[47]	[66]

[a]Company website data.

Figure 5. New technologies for protein and metabolite analysis. **(a)** A single-cell encapsulator, showing **(i)** the channel systems, and **(ii)** a string of droplets generated. Adapted with permission from Ref. [53]. **(b)** An SCA chip, showing the cell manipulation section on the left and the molecule counting section on the right. Adapted with permission from Ref. [59]. **(c)** An integrated blood barcode chip, comprising channels that harness the Zweifach–Fung effect for plasma separation from a finger prick of blood as well as multiple DNA-encoded antibody DEAL barcode arrays patterned on the surface of the plasma-skimming channel. Adapted with permission from Ref. [60]. **(d)** Microfabricated monolithic multinozzle emitters (M^3 emitters), showing **(i)** a schematic view of a nanoelectrospray emitter with two protruding nozzles, and **(ii)** scanning electron microscopy images and corresponding magnified views of the M^3 emitters with different nozzle numbers (1–10) and dimensions. Adapted with permission from Ref. [62].

More recently, high-throughput hydrodynamic encapsulation of single cells into picoliter-volume droplets on PDMS chips has been developed [55,56]. Furthermore, online analysis of the contents in single droplets by MS has been demonstrated [57], which suggests the potential of single cell encapsulation–MS analysis.

Micro/nanofluidic devices have been fabricated to analyze directly proteins from single cells or extremely small amounts of biomaterials. Zare and colleagues have demonstrated chemical cytometry as well as protein counting in a single cell using PDMS microfluidic devices [58,59]. The analysis chips (Figure 5b) can manipulate, lyse, label, separate and quantify the protein contents of a single cell using single-molecule fluorescence counting [59]. Heath and colleagues have developed an integrated PDMS-based barcode chip (Figure 5c) for rapid multiplexed analysis of proteins in a microliter of blood [60]. The device achieves on-chip blood separation and rapid detection of a panel of plasma proteins using fluorescence-based surface immunoassays. Although no SCA has been demonstrated yet using the device, its efficient separation and sensitive

detection of a very small sample could enable its application in SCA. Han and colleagues have developed a Si-based microfabricated nanofluidic device, an anisotropic nanofilter array, for continuous-flow separation of DNA and proteins [61]. The device could serve as a key component for sample preparation and separation in an integrated bioanalytical system. However, all three of these methods (Zare, Heath and Han) require labeling of proteins using fluorescent antibody binding, which limits analysis to a small number of previously characterized proteins.

MS, by contrast, can identify and quantify thousands of proteins and their posttranslational modifications (e.g. phosphorylation and acetylation). Wang and colleagues have developed Si/SiO$_2$-based microfabricated monolithic multinozzle emitters for MS (M^3 emitters), each of which consists of a parallel silica nozzle array that protrudes out from a hollow silicon sliver (Figure 5d) [62]. Si/SiO$_2$-based materials were selected because hydrophobic polymers, such as PDMS, have inherently undesirable properties for the ESI-MS application, including strong affinity to

proteins and incompatibility with certain organic solvents. Once integrated with a mass spectrometer, M^3 emitters achieve sensitivity and stability in peptide and protein detection comparable to those of commercial silica-based capillary nanoelectrospray. The additional advantages of these emitters are ease of fabrication and potential to be integrated with other micro/nanofluidic structures. These M^3 emitters could form a crucial component in a fully integrated Si/SiO_2-based lab-on-a-chip for single cell proteomics and metabolomics.

Key challenges still remain to implement high-quality interfaces between micro/nanofluidic chips and mass spectrometers [63], particularly for SCA. This requires further reduction in the sampling size (i.e. picoliter to femtoliter when studying single cells or subcellular components); sample manipulation (i.e. efficient recovery of proteins and peptides); integration of multiple processes (i.e. lysis, separation and detection); and dramatic improvement of mass spectrometers (sensitivity, accuracy, resolution and dynamic range). If coupled with the above-mentioned on-chip droplet encapsulation of single cells, MS-chip technology could open new doors to high-throughput multiplexed SCA. Therefore, breakthroughs in micro/nanofabrication methods and MS technologies are clearly needed to develop an integrated micro/nanofluidics device for single cell proteomics and metabolomics.

Challenges and opportunities

Single cell 'omics' presents unprecedented challenges and opportunities, both technologically and biologically. There is still a great need for further improvement in miniaturization, integration and detection sensitivity. There is also a great need for automation, throughput and bioinformatics to study multiple individual cells to achieve statistical significance. New software packages for statistics and bioinformatics [64], for example, might need to be developed for the large amount of data generated from multiple single cells. All of these technical challenges will add costs to existing omics methods in the short term, but miniaturization and automation are expected to enable the development of new affordable high-throughput methods. Moreover, most current omics-based SCA methods require cell lysis, which means that genes, proteins and metabolites are studied at specific timepoints outside of their physiological environments. This can be mitigated through live-cell imaging and other *in situ* techniques that can be performed on single cells once the key markers and pathways are discovered through SCA [65]. Furthermore, studying individual cells is inherently reductionist, which means that tissue-level properties, such as interactions between cells and their extracellular environment, need to be incorporated into experimental designs and data analyses [7].

While these challenges are significant, opportunities abound. At the simplest level, reducing cellular heterogeneity reduces noise, which is a goal of any analytical system. At a higher level, the heterogeneity itself becomes informative. As shown in Figure 1a, identifying the cells with complete GAPDH knockdown significantly reduces the noise in analysis of GAPDH functions. Furthermore, isolating cells with complete and partial knockdowns allows

for an examination of RNA interference mechanisms and the causes of the disparate, but apparently quantized, responses. Detailed analyses of single cells using omics technologies are expected to elucidate the underlying cellular pathways and networks.

In summary, SCA is the new frontier in omics. Single cell omics represents the unique convergence of technology and biology and will transform systems biology by converting cellular heterogeneity from a source of noise to a source of new discoveries.

References

1 Irish, J.M. *et al.* (2006) Mapping normal and cancer cell signalling networks: towards single-cell proteomics. *Nat. Rev. Cancer* 6, 146–155
2 Graf, T. and Stadtfeld, M. (2008) Heterogeneity of embryonic and adult stem cells. *Cell Stem Cell* 3, 480–483
3 Cai, L. *et al.* (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* 440, 358–362
4 Rosenfeld, N. *et al.* (2005) Gene regulation at the single-cell level. *Science* 307, 1962–1965
5 Weinberger, L.S. *et al.* (2005) Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* 122, 169–182
6 Toriello, N.M. *et al.* (2008) Integrated microfluidic bioprocessor for single-cell gene expression analysis. *Proc. Natl. Acad. Sci. U. S. A.* 105, 20173–20178
7 Liu, X. *et al.* (2009) Analysis of cell fate from single-cell gene expression profiles in C. elegans. *Cell* 139, 623–633
8 Takahashi, K. *et al.* (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131, 861–872
9 Chan, E.M. *et al.* (2009) Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. *Nat. Biotechnol.* 27, 1033–1037
10 He, Y. *et al.* (2010) Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature* 2010 Mar 3. [Epub ahead of print]
11 Clarke, M.F. *et al.* (2006) Cancer stem cells—perspectives on current status and future directions: AACR Workshop on cancer stem cells. *Cancer Res.* 66, 9339–9344
12 Bao, S. *et al.* (2006) Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* 444, 756–760
13 O'Dowd, D.K. and Smith, M.A. (1996) Single-cell analysis of gene expression in the nervous system. Measurements at the edge of chaos. *Mol. Neurobiol.* 13, 199–211
14 Munne, S. *et al.* (1994) Chromosome abnormalities in human arrested preimplantation embryos: a multiple-probe FISH study. *Am. J. Hum. Genet.* 55, 150–159
15 Pinkel, D. *et al.* (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* 20, 207–211
16 Hood, L. *et al.* (2004) Systems biology and new technologies enable predictive and preventative medicine. *Science* 306, 640–643
17 Elowitz, M.B. *et al.* (2002) Stochastic gene expression in a single cell. *Science* 297, 1183–1186
18 Levsky, J.M. *et al.* (2002) Single-cell gene expression profiling. *Science* 297, 836–840
19 Shav-Tal, Y. *et al.* (2004) Dynamics of single mRNPs in nuclei of living cells. *Science* 304, 1797–1800
20 Capodieci, P. *et al.* (2005) Gene expression profiling in single cells within tissue. *Nat. Methods* 2, 663–665

21 Bengtsson, M. *et al.* (2005) Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Res.* 15, 1388–1392

22 Whitesides, G.M. (2006) The origins and the future of microfluidics. *Nature* 442, 368–373

23 Marcus, J.S. *et al.* (2006) Microfluidic single-cell mRNA isolation and analysis. *Anal. Chem.* 78, 3084–3089

24 Tang, F. *et al.* (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, 377–382

25 Taniguchi, K. *et al.* (2009) Quantitative analysis of gene expression in a single cell by qPCR. *Nat. Methods* 6, 503–506

26 Wang, Z. *et al.* (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63

27 Kurimoto, K. *et al.* (2007) Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. *Nat. Protoc.* 2, 739–752

28 Zhang, K. *et al.* (2006) Sequencing genomes from single cells by polymerase cloning. *Nat. Biotechnol.* 24, 680–686

29 Whittal, R.M. *et al.* (1998) Nanoliter chemistry combined with mass spectrometry for peptide mapping of proteins from single mammalian cell lysates. *Anal. Chem.* 70, 5344–5347

30 Jo, K. *et al.* (2007) Mass spectrometric imaging of peptide release from neuronal cells within microfluidic devices. *Lab Chip* 7, 1454–1460

31 Rubakhin, S.S. and Sweedler, J.V. (2007) Characterizing peptides in individual mammalian cells using mass spectrometry. *Nat. Protoc.* 2, 1987–1997

32 Mizuno, H. *et al.* (2008) Live single-cell video-mass spectrometry for cellular and subcellular molecular detection and cell classification. *J. Mass Spectrom.* 43, 1692–1700

33 Lapainis, T. *et al.* (2009) Capillary electrophoresis with electrospray ionization mass spectrometric detection for single-cell metabolomics. *Anal. Chem.* 81, 5858–5864

34 Shrestha, B. and Vertes, A. (2009) *In situ* metabolic profiling of single cells by laser ablation electrospray ionization mass spectrometry. *Anal. Chem.* 81, 8265–8271

35 Li, L. *et al.* (1999) *In situ* sequencing of peptides from biological tissues and single cells using MALDI-PSD/CID analysis. *Anal. Chem.* 71, 5451–5458

36 Rubakhin, S.S. *et al.* (2000) Measuring the peptides in individual organelles with mass spectrometry. *Nat. Biotechnol.* 18, 172–175

37 Irish, J.M. *et al.* (2004) Single cell profiling of potentiated phospho-protein networks in cancer cells. *Cell* 118, 217–228

38 Newman, J.R. *et al.* (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441, 840–846

39 Cohen, A.A. *et al.* (2008) Dynamic proteomics of individual cancer cells in response to a drug. *Science* 322, 1511–1516

40 Bandura, D.R. *et al.* (2009) Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal. Chem.* 81, 6813–6822

41 Waanders, L.F. *et al.* (2009) Quantitative proteomic analysis of single pancreatic islets. *Proc. Natl. Acad. Sci. U. S. A.* 106, 18902–18907

42 Geiss, G.K. *et al.* (2008) Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* 26, 317–325

43 Harris, T.D. *et al.* (2008) Single-molecule DNA sequencing of a viral genome. *Science* 320, 106–109

44 Eid, J. *et al.* (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323, 133–138

45 Lipson, D. *et al.* (2009) Quantification of the yeast transcriptome by single-molecule sequencing. *Nat. Biotechnol.* 27, 652–658

46 Clarke, J. *et al.* (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* 4, 265–270

47 Drmanac, R. *et al.* (2010) Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 327, 78–81

48 Goren, A. *et al.* (2010) Chromatin profiling by directly sequencing small quantities of immunoprecipitated DNA. *Nat. Methods* 7, 47–49

49 Cipriany, B.R. *et al.* (2010) Single molecule epigenetic analysis in a nanofluidic channel. *Anal. Chem.* Feb 25. [Epub ahead of print]

50 Metzker, M.L. (2010) Sequencing technologies - the next generation. *Nat. Rev. Genet.* 11, 31–46

51 Emmert-Buck, M.R. *et al.* (1996) Laser capture microdissection. *Science* 274, 998–1001

52 Chiu, D.T. and Lorenz, R.M. (2009) Chemistry and biology in femtoliter and picoliter volume droplets. *Acc. Chem. Res.* 42, 649–658

53 Sgro, A.E. *et al.* (2007) Thermoelectric manipulation of aqueous droplets in microfluidic devices. *Anal. Chem.* 79, 4845–4851

54 He, M. *et al.* (2005) Selective encapsulation of single cells and subcellular organelles into picoliter- and femtoliter-volume droplets. *Anal. Chem.* 77, 1539–1544

55 Chabert, M. and Viovy, J.L. (2008) Microfluidic high-throughput encapsulation and hydrodynamic self-sorting of single cells. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3191–3196

56 Edd, J.F. *et al.* (2008) Controlled encapsulation of single-cells into monodisperse picolitre drops. *Lab Chip* 8, 1262–1264

57 Fidalgo, L.M. *et al.* (2009) Coupling microdroplet microreactors with mass spectrometry: reading the contents of single droplets online. *Angew. Chem. Int. Ed. Engl.* 48, 3665–3668

58 Wu, H. *et al.* (2004) Chemical cytometry on a picoliter-scale integrated microfluidic chip. *Proc. Natl. Acad. Sci. U. S. A.* 101, 12809–12813

59 Huang, B. *et al.* (2007) Counting low-copy number proteins in a single cell. *Science* 315, 81–84

60 Fan, R. *et al.* (2008) Integrated barcode chips for rapid, multiplexed analysis of proteins in microliter quantities of blood. *Nat. Biotechnol.* 26, 1373–1378

61 Fu, J. *et al.* (2007) A patterned anisotropic nanofluidic sieving structure for continuous-flow separation of DNA and proteins. *Nat. Nanotechnol.* 2, 121–128

62 Kim, W. *et al.* (2007) Microfabricated monolithic multinozzle emitters for nanoelectrospray mass spectrometry. *Anal. Chem.* 79, 3703–3707

63 Sikanen, T. *et al.* (2009) Microchip technology in mass spectrometry. *Mass Spectrom. Rev.* 2009 Jun 9. [Epub ahead of print]

64 Pop, M. and Salzberg, S.L. (2008) Bioinformatics challenges of new sequencing technology. *Trends Genet.* 24, 142–149

65 Yu, J. *et al.* (2006) Probing gene expression in live cells, one protein molecule at a time. *Science* 311, 1600–1603

66 Rothberg, J. *et al.* Ion Torrent Systems Inc. Methods and Apparatus for Measuring Analytes using Large Scale FET Arrays, WO/2008/076406.