

Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

Title

Single cell analysis: the new frontier in 'Omics'

Permalink

<https://escholarship.org/uc/item/9fx7d1cs>

Author

Wang, Daojing

Publication Date

2010-04-30

Peer reviewed

1 **Single cell analysis: the new frontier in ‘Omics’**

2 **Daojing Wang¹ and Steven Bodovitz²**

3 1. Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA

4 2. BioPerspectives, San Francisco, CA

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

6

7 **Cellular heterogeneity arising from stochastic expression of genes, proteins,**
8 **and metabolites is a fundamental principle of cell biology, but single cell analysis**
9 **has been beyond the capabilities of ‘Omics’ technologies. This is rapidly changing**
10 **with the recent examples of single cell genomics, transcriptomics, proteomics, and**
11 **metabolomics. The rate of change is expected to accelerate owing to emerging**
12 **technologies that range from micro/nanofluidics to microfabricated interfaces for**
13 **mass spectrometry to third- and fourth-generation automated DNA sequencers. As**
14 **described in this review, single cell analysis is the new frontier in Omics, and single**
15 **cell Omics has the potential to transform systems biology through new discoveries**
16 **derived from cellular heterogeneity.**

17

18 **Single cell analysis: needs and applications**

19 *Cellular heterogeneity*

20 Cellular heterogeneity within an isogenic cell population is a widespread event [1, 2].
21 Stochastic gene and protein expression at the single cell level has been clearly
22 demonstrated in different systems using a variety of techniques [3-5]. Therefore,
23 analyzing cell ensembles individually with high spatiotemporal resolutions will lead to a

1 more accurate representation of cell-to-cell variations instead of the stochastic average
2 masked by bulk measurements. Disconnect between single cell and average cell
3 measurements is exemplified in **Figure 1a**. Using an integrated microfluidic bioprocessor
4 for single cell gene expression analysis, Mathies group showed that siRNA knockdown
5 of GAPDH gene expression led to two distinct groups of individual Jurkat cells-partial
6 knockdown (~50%) and complete knockdown (~0%). The average result from 50 cells
7 (~21%) is not representative of any one individual cell [6].

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

21

22 *Potential applications of single cell analysis*

23 At the simplest level, SCA reduces biological noise. It provides fundamental

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

22 Advances in SCA have the potential to accelerate not only biological research, but
23 also diagnostics. Preimplantation genetic diagnosis (PGD), for example, is the analysis of

1 a single cell from a biopsy of an embryo after *in vitro* fertilization. PGD is used to test for
2 genetic diseases and chromosome aneuploidies. The most common molecular analysis
3 techniques are PCR and FISH (fluorescent *in situ* hybridization) [14], and newer
4 methods, such as comparative genomic hybridization (CGH) [15], are beginning to be
5 utilized. As more genetic diseases and predispositions are identified, multiplexed PGD
6 using single cell Omics technologies will enable early and accurate identification and
7 possible treatment of genetic abnormalities in human embryos.

8

9 **Single cell Omics: current state-of-the-art**

10 *Omics for single cell analysis*

11 'Omics' – the large-scale studies of genes (genomics and epigenomics), transcripts
12 (transcriptomics), proteins (proteomics), metabolites (metabolomics), lipids (lipidomics),
13 and interactions (interactomics) – is both the foundation and driving force for systems
14 biology [16]. Omics strives to identify, quantify, and characterize all of the components
15 in cellular systems with spatiotemporal resolution, and thereby dissect the intracellular
16 pathways and networks. Currently, major technologies for genomics, epigenomics, and
17 transcriptomics include DNA sequencing and microarrays (planar-, bead-, and fiber optic
18 arrays); for proteomics, mass spectrometry (MS) and protein arrays; and for
19 metabolomics, MS and NMR.

20 The abovementioned major technologies have undergone dramatic advances in
21 automation and miniaturization. The logical limit to the ever-increasing miniaturization
22 of Omics technologies is the comprehensive spatiotemporal analysis of genes, transcripts,
23 proteins, metabolites, and interactions in single cells and their subcellular compartments.

1 The movement of Omics into SCA represents a significant shift. Previous well-
2 established methods for SCA, such as imaging and flow cytometry, are limited to the
3 examination of a small number of genes, proteins or metabolites. As a result, these
4 methods can only be used to open narrow windows into the complexity and dynamics of
5 intracellular pathways. By contrast, single cell Omics has the potential to enable systems
6 biology at the level of single cells, ultimately representing the unique convergence of
7 technology and biology.

8

9 *Nucleic acid analysis*

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

23 Further improvement in the breadth of single cell mRNA analysis has been

1 achieved recently using either mRNA sequencing (mRNA-Seq) [24] or quantitative PCR
2 (qPCR) [25]. In the first study, the Surani and Lao groups together developed an assay
3 with mRNA-Seq [26] for single cell whole-transcriptome analysis; in this method, single
4 cell cDNA amplification [27] was improved from a previous version by increasing the
5 cDNA length up to 3 kb. Briefly, the mRNAs from a single cell were reverse-transcribed
6 into cDNAs, which were subsequently amplified by PCR. Finally, amplified cDNAs were
7 digitally counted with an ultrahigh-throughput DNA sequencer. The researchers first
8 verified reproducibility of the method, and then demonstrated its utility in dissecting
9 functional consequences of *Dicer1* and *Ago2* knockdowns, two of the critical genes for
10 miRNA synthesis and mRNA regulation, by comparing single mouse *Dicer1*^{-/-} and *Ago2*^{-/-}
11 oocytes versus wild-types (**Figure 2a**). The described method has the added advantage
12 over probe-based methods (such as qPCR) in that it is effective at locating splicing
13 variants, including *de novo* variants, but may face challenges in accurate quantitation due
14 to the required amplification process.

15 To improve mRNA quantitation from single cells, the Kambara group performed
16 direct qPCR from a cDNA pool without pre-amplification [25]. Gene expression for four
17 house-keeping genes (TBP, SDHA, B2M, and EEF1G) was quantitatively analyzed in 14
18 single cells, showing that the absolute amounts of each of the four genes differed from
19 cell to cell (**Figure 2b**). They further showed that the average number of cDNA
20 molecules obtained from single cells was comparable to the per-cell number measured
21 from the cell pools (10-1000 cells); however, much larger standard deviation was
22 observed for the single cells, as expected from the cellular heterogeneity. The research
23 group is currently developing higher-throughput methodology to analyze more individual

1 cells (personal communication). A larger sample size will enable better statistical analysis
2 of cell-to-cell variation. The method would also benefit from increased multiplexing of
3 the capture beads for the immobilized cDNA library to increase transcriptome coverage.

4 Single cell genomics has been achieved for microorganisms. Church and
5 colleagues have sequenced the genome of single cells of *Escherichia coli* and
6 *Prochlorococcus* through so-called polymerase cloning (“ploning”), an optimized version
7 of multiple displacement amplification (MDA) using random primers [28]. They obtained
8 ~65% sequence coverage at an average depth of 4-fold for a single *Prochlorococcus* cell
9 through whole-genome shotgun sequencing. However, the method may suffer
10 background contamination during the amplification of single cell DNA and difficulty in
11 assembling the genomes. Compared to microorganisms, which typically have a genome
12 size of several million base pairs containing a few thousand genes, mammalian genomes
13 are much more complex. For example, the human genome has 3 billion base pairs
14 containing estimated 20,000-25,000 genes. To the best of our knowledge, there has been
15 no report yet on single cell genomics (i.e., global analysis of chromosomal DNA) or
16 epigenomics (i.e., global analysis of DNA methylation etc) for mammalian cells.
17 However, this may change soon with the recent introduction of several next-generation
18 sequencing technologies (discussed in the next section), particularly those centered on
19 single molecule sequencing.

20

21 ***Protein and metabolite analysis***

22 Proof-of-principle SCA has been performed to characterize peptides using matrix-
23 assisted laser desorption ionization (MALDI) MS [29-31], or to characterize small

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

16 More recently, laser ablation electrospray ionization mass spectrometry (LAESI-
17 MS) was demonstrated for *in situ* metabolomic profiling of single cells [34] (**Figure 3b**).
18 Single cell ablation was achieved by delivering mid-IR laser pulses through a GeO₂-
19 based glass fiber (15 μm radius at the tip). When comparing two physically adjacent,
20 individual *Allium cepa cultivar* cells with differences in purple pigmentation, LAESI-MS
21 revealed similar essential metabolites for both variants, with colorless cells containing
22 anthocyanin and pigmented cells containing anthocyanidins, other flavonoids, and their
23 glucosides. One of the key advantages of this method is that single cells can be analyzed

1 *in situ*, although, as mentioned by the authors, further improvement in optics and ablation
2 geometry is needed for cell-by-cell analysis of biological tissue. Another challenge for
3 the LAESI-MS is the lack of front-end separation of cellular contents using either LC or
4 CE, because laser ablation is performed directly on whole cells. This hinders further
5 improvement in metabolome coverage and identification.

6 Previously, flow cytometry (i.e. fluorescence-activated cell sorting, or FACS) has
7 been used to profile phosphoprotein networks in single cancer cells [37] and monitor
8 global protein levels in single *Saccharomyces cerevisiae* [38]. In addition, fluorescence
9 imaging has been used to measure the levels and locations of nearly 1,000 different
10 endogenously tagged proteins in individual living cancer cells at high temporal resolution
11 [39]. Current technologies based on fluorescence require prior *in vivo* or *in vitro* labeling,
12 and are generally limited to 10-20 simultaneous measurements due to the spectral
13 overlap. For *in vitro* labeling, background arising from cross-reactivity and/or nonspecific
14 binding remains a challenge. For *in vitro* labeling, tagging with GFP/YFP and their
15 derivatives may alter folding or other characteristics (e.g. interactions with other proteins)
16 of the endogenous proteins. Furthermore, photobleaching of fluorophores and
17 autofluorescence of cells may interfere with kinetic measurements. A potential solution is
18 MS-based flow cytometry ('mass cytometry'), which has been described for a single cell
19 multi-target immunoassay [40]. However, the method still requires labeling of antibodies
20 with elemental tags (e.g. lanthanide) prior to antigen binding.

21 Mass spectrometry-based single cell proteomics has not been achieved yet despite
22 tremendous interest in the field. Recently, an LC-MS/MS-based method was described
23 for quantitative proteomic analysis of single pancreatic islets containing 2,000-4,000 cells

1 [41]. A combination of improved chromatographic methods, direct replicate
2 measurements, and high accuracy (1-2 ppm) and sensitivity (attomole to femtomole)
3 provided by a LTQ-Orbitrap mass spectrometer, allowed for detection of more than 6,000
4 proteins from nanogram quantities of protein mixtures. This ultrasensitive method
5 represented an important step toward single cell proteomics. However, due to the extreme
6 complexity and huge dynamic range of the proteome, new breakthroughs in sample
7 manipulation and detection are needed to achieve *in situ* single cell proteomics with deep
8 proteome coverage (discussed in the next section).

9

10 **Single cell Omics: future prospects**

11 Technologies for analyzing nucleic acids, proteins, and metabolites are evolving rapidly.
12 This section includes notable examples that have the potential to make major
13 contributions to single cell Omics.

14

15 ***New technologies for nucleic acid analysis***

16 Driven by personal genomics and the NIH-launched “\$1000 Genome” project, the
17 technologies for nucleic acid analysis, particularly the next-generation sequencing
18 technologies, have undergone dramatic advances [42-47]. **Figure 4** showcases some of
19 the technologies that may contribute to SCA.

20 Geiss and colleagues from NanoString Technologies, for example, have
21 developed the nCounter gene expression system (**Figure 4a**) for direct multiplexed
22 measurement of gene expression with color-coded probe pairs without amplification [42].
23 The technology utilizes molecular barcodes and single-molecule imaging to detect and

1 count hundreds of unique transcripts directly in a single reaction. The company
2 demonstrated that the nCounter system is more sensitive than microarrays and similar in
3 sensitivity to real-time PCR. The technology has the potential for single cell
4 transcriptomics.

5 In addition, multiple breakthroughs in single molecule sequencing technologies
6 could pave the way for single cell genomics, epigenomics, and transcriptomics. Turner
7 and Korlach together with colleagues from Pacific Biosciences have developed a
8 technology for single molecule and real-time DNA sequencing by a single DNA
9 polymerase (**Figure 4b**) [44]. They detected the enzymatic incorporation of
10 fluorescently-labeled deoxyribonucleoside triphosphates into the growing DNA strand
11 with zero-mode waveguide nanostructure arrays. They demonstrated a median accuracy
12 of 99.3%, with no systematic error beyond the fluorophore-dependent error rates, with
13 the consensus sequences generated from the single-molecule reads at 15-fold coverage.
14 Optimism for this technology is high: the company has raised approximately \$266
15 million to date through a combination of grants and venture capital according to the
16 company's website.

17 Traz and colleagues from Helicos Biosciences have developed a high-throughput
18 and amplification-free method for transcriptome quantification, named single-molecule
19 sequencing digital gene expression (smsDGE) (**Figure 4c**) [45]. The technology utilizes a
20 reverse-transcription and polyA-tailing sample preparation procedure followed by single-
21 molecule sequencing that generates a single read per transcript as published earlier [43].
22 They confirmed accurate quantification using spiked-in RNAs, and demonstrated
23 sequencing of the yeast transcriptome in a single run, yielding an average of 12 million

1 aligned reads per channel. While they have not used this technology for SCA, they
2 recently published a ChIP-Seq study in collaboration with the Bernstein group using a
3 small number of cells and only 50 pg of DNA [48].

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

15 **Table 1** summarizes the key technologies and other attributes of five major third-
16 generation DNA sequencing companies, including three mentioned above (Helicos,
17 Pacific Bio, and Oxford Nanopore) and two others (Complete Genomics and Ion
18 Torrent). In addition, Life Technologies is actively developing a single molecule
19 sequencing technology using quantum dots, and Cipriany et al. recently demonstrated a
20 method for single molecule epigenetic analysis using nanofluidics and multicolor
21 fluorescence microscopy to detect DNA methylation [49]. (For more information on
22 next-generation sequencing technologies, please refer to a recent review by Michael
23 Metzker [50]) Further improvements in detection sensitivity, sample manipulation and

1 recovery, and sequence assembly will enable the abovementioned technologies to
2 become suitable for direct single cell Omics applications.

3

4 *New technologies for protein and metabolite analysis*

5 The “holy grail” of proteomics and metabolomics is to decode the proteome and
6 metabolome at the single cell level. This is challenging due to the fact that the amount of
7 proteins and metabolites is extremely limited in a single cell (i.e. average 1×10^5
8 molecules for proteins). Furthermore, there are no amplification methods for proteins and
9 metabolites comparable to those for nucleic acids. We believe the most prevalent
10 bottleneck for SCA of proteins and metabolites *in situ* is the marriage between extremely
11 efficient sample manipulation and highly sensitive detection. Along this line, we believe
12 micro/nanofluidics interfaced with mass spectrometry will be hotly pursued for single-
13 cell proteomics and metabolomics. **Figure 5** showcases some promising technologies that
14 may contribute to this endeavor.

15 To achieve single cell Omics, it is critical to efficiently select and manipulate
16 single cells. Laser capture microdissection has been used routinely to isolate a small
17 number of cells from tissue [51]. In addition, micropipette-based micromanipulators have
18 been used to manipulate single cells. However, these technologies are difficult to
19 implement in a micro/nanofluidic device, particularly for living single cells. Newer
20 technologies based on droplet encapsulation may circumvent this challenge. For example,
21 Chiu and coworkers have developed femtoliter- and picoliter-sized droplets as nanolabs
22 for manipulating single cells and subcellular compartments [52]. **Figure 5a** shows one of
23 their PDMS microfluidic devices for thermoelectric manipulation of single cells [53]. In

1 this device, the T-channel was used for droplet generation and freezing (above where a
2 thermo-electric cooler (TEC) was placed). The same group earlier combined optical
3 trapping, microfluidic encapsulation, and rapid laser photolysis, and demonstrated
4 selective encapsulation of single cells and analysis of subcellular organelles through
5 fluorescence spectroscopy [54]. More recently, high-throughput hydrodynamic
6 encapsulation of single cells into picoliter-volume droplets on PDMS chips has been
7 developed by others [55, 56]. Furthermore, online analysis of the contents in single
8 droplets by MS has been demonstrated [57], suggesting promising potential of single cell
9 encapsulation-MS analysis.

10 Micro/nanofluidic devices have been fabricated to directly analyze proteins from
11 single cells or extremely small amounts of biomaterials. Zare and colleagues have
12 demonstrated chemical cytometry as well as protein counting in a single cell using PDMS
13 microfluidic devices [58, 59]. The analysis chips (**Figure 5b**) can manipulate, lyse, label,
14 separate, and quantify the protein contents of a single cell using single molecule
15 fluorescence counting [59]. Heath and colleagues have developed an integrated PDMS-
16 based barcode chip (**Figure 5c**) for rapid multiplexed analysis of proteins in a microliter
17 of blood [60]. The device achieved on-chip blood separation and rapid detection of a
18 panel of plasma proteins using fluorescence-based surface immunoassays. Although no
19 SCA has been demonstrated yet using the device, its efficient separation and sensitive
20 detection of a very small sample amount may enable its application in SCA. Han and
21 colleagues have developed a Si-based microfabricated nanofluidic device, an anisotropic
22 nanofilter array (ANA), for continuous-flow separation of DNA and proteins [61]. The
23 device may serve as a key component for sample preparation and separation in an

1 integrated bioanalytical system. However, all three of these methods (Zare, Heath, and
2 Han) require labeling of proteins using fluorescent antibody binding, which limits
3 analysis to a small number of previously characterized proteins.

4 Mass spectrometry, by contrast, can identify and quantify thousands of proteins
5 and their posttranslational modifications (e.g. phosphorylation and acetylation, etc) per
6 experiment. Wang and colleagues have developed Si/SiO₂-based microfabricated
7 monolithic multinozzle emitters for mass spectrometry (M³ emitters) (**Figure 5d**) [62].
8 Si/SiO₂-based materials were selected because hydrophobic polymers, such as PDMS,
9 have inherently undesirable properties for the ESI-MS application, including strong
10 affinity to proteins and incompatibility with certain organic solvents. In their design, each
11 M³ emitter consists of a parallel silica nozzle array protruding out from a hollow silicon
12 sliver. Once integrated with a mass spectrometer, M³ emitters achieve sensitivity and
13 stability in peptide and protein detection comparable to those of commercial silica-based
14 capillary nanoelectrospray. The additional advantages of these emitters are ease of
15 fabrication and potential to be integrated with other micro/nanofluidic structures. These
16 M³ emitters may serve as a critical component in a fully integrated silicon/silica-based
17 lab-on-a-chip for single cell proteomics and metabolomics.

18 Key challenges still remain to implement high-quality interfaces between
19 micro/nanofluidic chips and mass spectrometers [63], particularly for SCA. This requires
20 further reduction in the sampling size (picoliter to femtoliter, i.e. for single cells or
21 subcellular components), sample manipulation (i.e. efficient recovery of proteins and
22 peptides), integration of multiple processes (i.e. lysis, separation, and detection), and
23 dramatic improvement of the mass spectrometers (sensitivity, accuracy, resolution, and

1 dynamic range). If coupled with abovementioned on-chip droplet encapsulation of single
2 cells, MS-chip technologies may open new doors to high-throughput multiplexed SCA.
3 Therefore, breakthroughs in micro/nanofabrication methods and mass spectrometry
4 technologies are clearly needed to develop an integrated micro/nanofluidics device for
5 single cell proteomics and metabolomics.

6

7 ***Challenges and opportunities***

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

1 While these challenges are significant, opportunities abound. At the simplest
2 level, reducing cellular heterogeneity reduces noise, which is a goal of any analytical
3 system. At a higher level, though, the heterogeneity itself becomes informative. As
4 shown in **Figure 1a**, identifying the cells with complete GAPDH knockdown would
5 significantly reduce the noise in analyses of GAPDH functions. Furthermore, isolating
6 cells with both complete and partial knockdowns allows for an examination of RNA
7 interference mechanisms and the causes of the disparate, but apparently quantized,
8 responses. Detailed analysis of single cells using Omics technologies are expected to
9 elucidate the underlying cellular pathways and networks.

10 In summary, single cell analysis is the new frontier in ‘Omics’. Single cell Omics
11 represents the unique convergence of technology and biology and will transform systems
12 biology by converting cellular heterogeneity from a source of noise to a source of new
13 discoveries.

1 **Acknowledgements**

2 D.W. acknowledges the funding supports from the National Institutes of Health (NIH
3 R21GM077870), the US Department of Energy (DOE DE-AC02-05CH11231), the
4 National Aeronautics and Space Administration (NASA), the California Breast Cancer
5 Research Program (CBCRP), and the Lawrence Berkeley National Laboratory (LBNL
6 Discovery LDRD). The opinions in this review are those of the authors and do not
7 necessarily represent the official positions or policies of the abovementioned funding
8 agencies. S.B. discloses that he is the Principal of BioPerspectives.

1 **References**

- 2 1. Irish, J.M., *et al.* (2006) Mapping normal and cancer cell signalling networks:
3 towards single-cell proteomics. *Nat Rev Cancer* 6, 146-155
- 4 2. Graf, T., and Stadtfeld, M. (2008) Heterogeneity of embryonic and adult stem
5 cells. *Cell Stem Cell* 3, 480-483
- 6 3. Cai, L., *et al.* (2006) Stochastic protein expression in individual cells at the single
7 molecule level. *Nature* 440, 358-362
- 8 4. Rosenfeld, N., *et al.* (2005) Gene regulation at the single-cell level. *Science* 307,
9 1962-1965
- 10 5. Weinberger, L.S., *et al.* (2005) Stochastic gene expression in a lentiviral positive-
11 feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* 122, 169-182
- 12 6. Toriello, N.M., *et al.* (2008) Integrated microfluidic bioprocessor for single-cell
13 gene expression analysis. *Proc Natl Acad Sci U S A* 105, 20173-20178
- 14 7. Liu, X., *et al.* (2009) Analysis of cell fate from single-cell gene expression
15 profiles in *C. elegans*. *Cell* 139, 623-633
- 16 8. Takahashi, K., *et al.* (2007) Induction of pluripotent stem cells from adult human
17 fibroblasts by defined factors. *Cell* 131, 861-872
- 18 9. Chan, E.M., *et al.* (2009) Live cell imaging distinguishes bona fide human iPSC
19 cells from partially reprogrammed cells. *Nat Biotechnol* 27, 1033-1037
- 20 10. He, Y., *et al.* (2010) Heteroplasmic mitochondrial DNA mutations in normal and
21 tumour cells. *Nature* 2010 Mar 3. [Epub ahead of print]
- 22 11. Clarke, M.F., *et al.* (2006) Cancer stem cells--perspectives on current status and
23 future directions: AACR Workshop on cancer stem cells. *Cancer Res* 66, 9339-9344
- 24 12. Bao, S., *et al.* (2006) Glioma stem cells promote radioresistance by preferential
25 activation of the DNA damage response. *Nature* 444, 756-760
- 26 13. O'Dowd, D.K., and Smith, M.A. (1996) Single-cell analysis of gene expression in
27 the nervous system. Measurements at the edge of chaos. *Mol Neurobiol* 13, 199-211
- 28 14. Munne, S., *et al.* (1994) Chromosome abnormalities in human arrested
29 preimplantation embryos: a multiple-probe FISH study. *Am J Hum Genet* 55, 150-159
- 30 15. Pinkel, D., *et al.* (1998) High resolution analysis of DNA copy number variation
31 using comparative genomic hybridization to microarrays. *Nat Genet* 20, 207-211
- 32 16. Hood, L., *et al.* (2004) Systems biology and new technologies enable predictive
33 and preventative medicine. *Science* 306, 640-643
- 34 17. Elowitz, M.B., *et al.* (2002) Stochastic gene expression in a single cell. *Science*
35 297, 1183-1186
- 36 18. Levsky, J.M., *et al.* (2002) Single-cell gene expression profiling. *Science* 297,
37 836-840
- 38 19. Shav-Tal, Y., *et al.* (2004) Dynamics of single mRNPs in nuclei of living cells.
39 *Science* 304, 1797-1800
- 40 20. Capodiceci, P., *et al.* (2005) Gene expression profiling in single cells within tissue.
41 *Nat Methods* 2, 663-665
- 42 21. Bengtsson, M., *et al.* (2005) Gene expression profiling in single cells from the
43 pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome*
44 *Res* 15, 1388-1392
- 45 22. Whitesides, G.M. (2006) The origins and the future of microfluidics. *Nature* 442,
46 368-373

- 1 23. Marcus, J.S., *et al.* (2006) Microfluidic single-cell mRNA isolation and analysis.
2 *Anal Chem* 78, 3084-3089
- 3 24. Tang, F., *et al.* (2009) mRNA-Seq whole-transcriptome analysis of a single cell.
4 *Nat Methods* 6, 377-382
- 5 25. Taniguchi, K., *et al.* (2009) Quantitative analysis of gene expression in a single
6 cell by qPCR. *Nat Methods* 6, 503-506
- 7 26. Wang, Z., *et al.* (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat*
8 *Rev Genet* 10, 57-63
- 9 27. Kurimoto, K., *et al.* (2007) Global single-cell cDNA amplification to provide a
10 template for representative high-density oligonucleotide microarray analysis. *Nat Protoc*
11 2, 739-752
- 12 28. Zhang, K., *et al.* (2006) Sequencing genomes from single cells by polymerase
13 cloning. *Nat Biotechnol* 24, 680-686
- 14 29. Whittall, R.M., *et al.* (1998) Nanoliter chemistry combined with mass
15 spectrometry for peptide mapping of proteins from single mammalian cell lysates. *Anal*
16 *Chem* 70, 5344-5347
- 17 30. Jo, K., *et al.* (2007) Mass spectrometric imaging of peptide release from neuronal
18 cells within microfluidic devices. *Lab Chip* 7, 1454-1460
- 19 31. Rubakhin, S.S., and Sweedler, J.V. (2007) Characterizing peptides in individual
20 mammalian cells using mass spectrometry. *Nat Protoc* 2, 1987-1997
- 21 32. Mizuno, H., *et al.* (2008) Live single-cell video-mass spectrometry for cellular
22 and subcellular molecular detection and cell classification. *J Mass Spectrom* 43, 1692-
23 1700
- 24 33. Lapainis, T., *et al.* (2009) Capillary electrophoresis with electrospray ionization
25 mass spectrometric detection for single-cell metabolomics. *Anal Chem* 81, 5858-5864
- 26 34. Shrestha, B., and Vertes, A. (2009) In situ metabolic profiling of single cells by
27 laser ablation electrospray ionization mass spectrometry. *Anal Chem* 81, 8265-8271
- 28 35. Li, L., *et al.* (1999) In situ sequencing of peptides from biological tissues and
29 single cells using MALDI-PSD/CID analysis. *Anal Chem* 71, 5451-5458
- 30 36. Rubakhin, S.S., *et al.* (2000) Measuring the peptides in individual organelles with
31 mass spectrometry. *Nat Biotechnol* 18, 172-175
- 32 37. Irish, J.M., *et al.* (2004) Single cell profiling of potentiated phospho-protein
33 networks in cancer cells. *Cell* 118, 217-228
- 34 38. Newman, J.R., *et al.* (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals
35 the architecture of biological noise. *Nature* 441, 840-846
- 36 39. Cohen, A.A., *et al.* (2008) Dynamic proteomics of individual cancer cells in
37 response to a drug. *Science* 322, 1511-1516
- 38 40. Bandura, D.R., *et al.* (2009) Mass Cytometry: Technique for Real Time Single
39 Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight
40 Mass Spectrometry. *Anal Chem* 81, 6813-6822
- 41 41. Waanders, L.F., *et al.* (2009) Quantitative proteomic analysis of single pancreatic
42 islets. *Proc Natl Acad Sci U S A* 106, 18902-18907
- 43 42. Geiss, G.K., *et al.* (2008) Direct multiplexed measurement of gene expression
44 with color-coded probe pairs. *Nat Biotechnol* 26, 317-325
- 45 43. Harris, T.D., *et al.* (2008) Single-molecule DNA sequencing of a viral genome.
46 *Science* 320, 106-109

- 1 44. Eid, J., *et al.* (2009) Real-time DNA sequencing from single polymerase
2 molecules. *Science* 323, 133-138
- 3 45. Lipson, D., *et al.* (2009) Quantification of the yeast transcriptome by single-
4 molecule sequencing. *Nat Biotechnol* 27, 652-658
- 5 46. Clarke, J., *et al.* (2009) Continuous base identification for single-molecule
6 nanopore DNA sequencing. *Nat Nanotechnol* 4, 265-270
- 7 47. Drmanac, R., *et al.* (2010) Human genome sequencing using unchained base
8 reads on self-assembling DNA nanoarrays. *Science* 327, 78-81
- 9 48. Goren, A., *et al.* (2010) Chromatin profiling by directly sequencing small
10 quantities of immunoprecipitated DNA. *Nat Methods* 7, 47-49
- 11 49. Cipriany, B.R., *et al.* (2010) Single Molecule Epigenetic Analysis in a
12 Nanofluidic Channel. *Anal Chem* Feb 25. [Epub ahead of print]
- 13 50. Metzker, M.L. (2010) Sequencing technologies - the next generation. *Nat Rev*
14 *Genet* 11, 31-46
- 15 51. Emmert-Buck, M.R., *et al.* (1996) Laser capture microdissection. *Science* 274,
16 998-1001
- 17 52. Chiu, D.T., and Lorenz, R.M. (2009) Chemistry and biology in femtoliter and
18 picoliter volume droplets. *Acc Chem Res* 42, 649-658
- 19 53. Sgro, A.E., *et al.* (2007) Thermoelectric manipulation of aqueous droplets in
20 microfluidic devices. *Anal Chem* 79, 4845-4851
- 21 54. He, M., *et al.* (2005) Selective encapsulation of single cells and subcellular
22 organelles into picoliter- and femtoliter-volume droplets. *Anal Chem* 77, 1539-1544
- 23 55. Chabert, M., and Viovy, J.L. (2008) Microfluidic high-throughput encapsulation
24 and hydrodynamic self-sorting of single cells. *Proc Natl Acad Sci U S A* 105, 3191-3196
- 25 56. Edd, J.F., *et al.* (2008) Controlled encapsulation of single-cells into monodisperse
26 picolitre drops. *Lab Chip* 8, 1262-1264
- 27 57. Fidalgo, L.M., *et al.* (2009) Coupling microdroplet microreactors with mass
28 spectrometry: reading the contents of single droplets online. *Angew Chem Int Ed Engl* 48,
29 3665-3668
- 30 58. Wu, H., *et al.* (2004) Chemical cytometry on a picoliter-scale integrated
31 microfluidic chip. *Proc Natl Acad Sci U S A* 101, 12809-12813
- 32 59. Huang, B., *et al.* (2007) Counting low-copy number proteins in a single cell.
33 *Science* 315, 81-84
- 34 60. Fan, R., *et al.* (2008) Integrated barcode chips for rapid, multiplexed analysis of
35 proteins in microliter quantities of blood. *Nat Biotechnol* 26, 1373-1378
- 36 61. Fu, J., *et al.* (2007) A patterned anisotropic nanofluidic sieving structure for
37 continuous-flow separation of DNA and proteins. *Nat Nanotechnol* 2, 121-128
- 38 62. Kim, W., *et al.* (2007) Microfabricated monolithic multinozzle emitters for
39 nanoelectrospray mass spectrometry. *Anal Chem* 79, 3703-3707
- 40 63. Sikanen, T., *et al.* (2009) Microchip technology in mass spectrometry. *Mass*
41 *Spectrom Rev* 2009 Jun 9. [Epub ahead of print]
- 42 64. Pop, M., and Salzberg, S.L. (2008) Bioinformatics challenges of new sequencing
43 technology. *Trends Genet* 24, 142-149
- 44 65. Yu, J., *et al.* (2006) Probing gene expression in live cells, one protein molecule at
45 a time. *Science* 311, 1600-1603

1 **Figure Legends**

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

14
15 **Figure 2.** Single cell gene expression analysis. (a) Correlation plots of the quantile-
16 normalized mRNA sequencing reads for mouse oocytes, showing (i) one wild-type
17 oocyte versus another wild-type oocyte; (ii) one *Dicer1*^{-/-} oocyte versus another *Dicer1*^{-/-}
18 oocyte; (iii) one wild-type oocyte versus one *Dicer1*^{-/-} oocyte; and (iv) one wild-type
19 oocyte versus one *Ago2*^{-/-} oocyte. All of the reads with changes of greater than fourfold
20 are plotted in red. Adapted with permission from Ref. [24]. (b) Number of cDNA
21 molecules measured by qPCR in single human colon carcinoma HCT116 cells and cell
22 pools, showing (i) cell-to-cell variation in expression of “housekeeping” genes among 14
23 single cells, and (ii) average gene expression for single cells as well as cell pools with

1 error bars (single cell: mean \pm s.d., n=14; 10–1,000 cells: mean \pm s.d., n=5). Adapted
2 with permission from Ref. [25].

3

4 **Figure 3.** Single cell metabolome analysis. (a) Metabolomic profiling of a single *Aplysia*
5 R2 neuron using CE-ESI-MS, showing the extracted ion electropherogram (XIE)
6 obtained for 146 m/z from different subcellular regions, i.e. soma versus neurite. Adapted
7 with permission from Ref. [33]. (b) Metabolomic profiling using LAESI-MS, showing (i)
8 etched optical fiber for laser ablation relative to the target single cell (scale bar =
9 100 μ m), and (ii) optical image of neighboring individual colorless and pigmented
10 epidermal cells of the purple *A. cepa cultivar* (scale bar = 50 μ m). Also shown (ii-a and
11 ii-b) are the corresponding LAESI-MS spectra. Selected similar and different peaks (m/z)
12 are indicated by arrows (ii-b). Adapted with permission from Ref. [34].

13

14 **Figure 4.** New technologies for nucleic acid analysis. (a) nCounter gene expression
15 system from NanoString Technologies, showing (i) schematic representation of the
16 hybridized complex (not to scale); (ii) schematic representation of, from left to right,
17 binding, electrophoresis, and immobilization; and (iii) false-color image of immobilized
18 reporter probes. Adapted with permission from Ref. [42]. (b) Single-molecule, real-time
19 DNA sequencing system from Pacific Biosciences, showing (i, left) schematic
20 representation of the experimental geometry with a single molecule of DNA template-
21 bound DNA polymerase immobilized at the bottom of a zero-mode waveguide (ZMW),
22 (i, right) schematic event sequence of the phospholinked dNTP incorporation cycle, with
23 a corresponding expected time trace of detected fluorescence intensity from the ZMW;

1 (ii, top) total intensity output of all four dye-weighted channels with pulses colored
2 corresponding to the least-squares fitting decisions of the algorithm, and (ii, bottom) the
3 entire read that proceeds through all 150 bases of the linear templates. Adapted with
4 permission from Ref. [44]. (c) Single-molecule sequencing digital gene expression
5 system from Helicos Biosciences, illustrating sample preparation and sequencing
6 workflow: (1) preparation of the first-strand cDNA from mRNA, (2) addition of 3' tail of
7 dATP followed by dideoxy-TTP (ddT) blocking, (3) hybridization of tailed sample to
8 poly-dT oligonucleotide covalently attached to the flow-cell channel surface, (4)
9 sequencing of a single base by adding a Cy5-labeled nucleotide, (5) cleaving off the Cy5
10 dye label, and (6) adding and imaging of next nucleotide. Adapted with permission from
11 Ref. [45]. (d) Single-molecule nanopore sequencing system from Oxford Nanopore
12 Technologies, showing (i, left) nanopore structure of the WT-
13 (M113R/N139Q)₆(M113R/N139Q/L135C)₁ mutant with the cyclodextrin covalently
14 attached at position 135 (space-filling model), and (i, right) close-up of the β barrel with
15 the arginines at position 113 and the location of the cysteines in the mutants tested in the
16 study; and (ii) single-channel recording from the nanopore that indicates discrimination
17 of dGMP, dTMP, dAMP, and dCMP, with colored bands representative of the residual
18 current distribution for each nucleotide. Adapted with permission from Ref. [46].

19

20 **Figure 5.** New technologies for protein and metabolite analysis. (a) A single-cell
21 encapsulator, showing (i) the channel systems, and (ii) a string of droplets generated.
22 Adapted with permission from Ref. [53]. (b) A single-cell analysis chip, showing the
23 cell-manipulation section on the left and the molecule-counting section on the right.

1 Adapted with permission from Ref. [59]. (c) An integrated blood barcode chip,
2 comprising channels that harness the Zweifach-Fung effect for plasma separation from a
3 finger prick of blood as well as multiple DNA-encoded antibody DEAL barcode arrays
4 patterned on the surface of the plasma-skimming channel. Adapted with permission from
5 Ref. [60]. (d) Microfabricated monolithic multinozzle emitters (M^3 emitters), showing (i)
6 a schematic view of a nanoelectrospray emitter with two protruding nozzles, and (ii)
7 SEM images and corresponding magnified views of the M^3 emitters with different nozzle
8 numbers (1-10) and dimensions. Adapted with permission from Ref. [62].

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

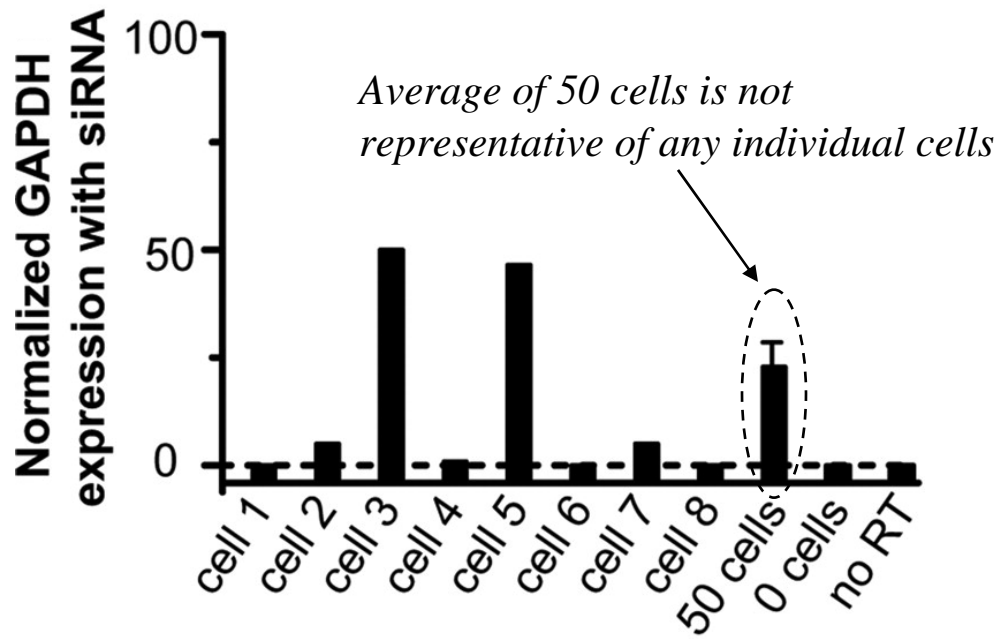
2

	Helicos	Pacific Biosciences	Oxford Nanopore	Complete Genomics	Ion Torrent
Key technology	Amplification-free sequencing, Ref. [43]	Zero-mode waveguide nanostructure arrays, Ref. [44]	Protein nanopores, Ref. [46]	Self-assembling DNA nanoarrays, Ref. [47]	Chemical sensitive field-effect transistor arrays ^a
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)^b	~US\$115	US\$266	£41.3	>US\$45	US\$23

3 a. Patent application WO/2008/076406

4 b: Company websites

(a)



(b)

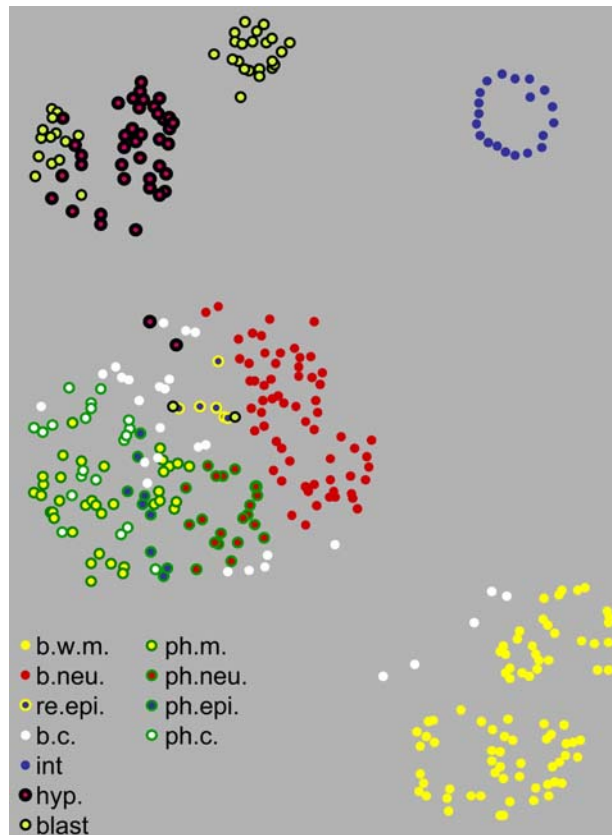
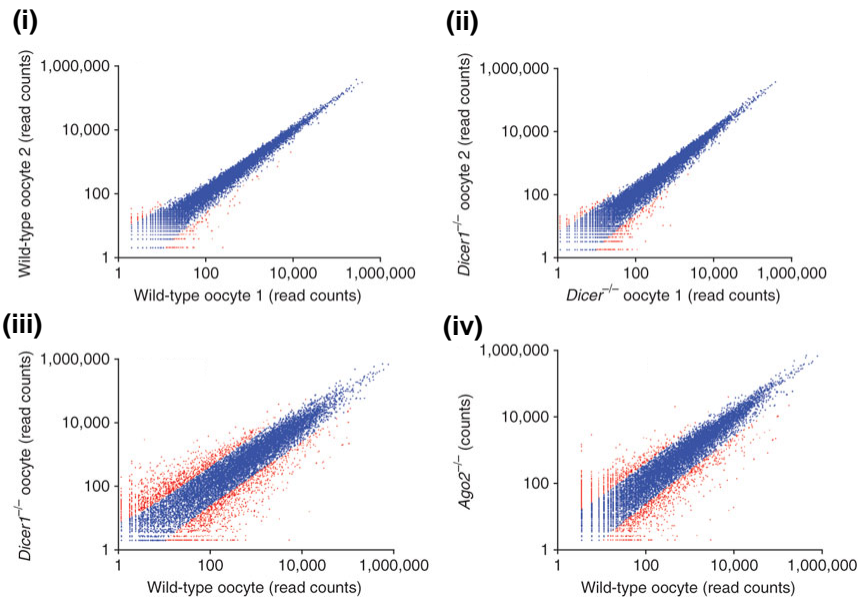


Figure 1

(a)



(b)

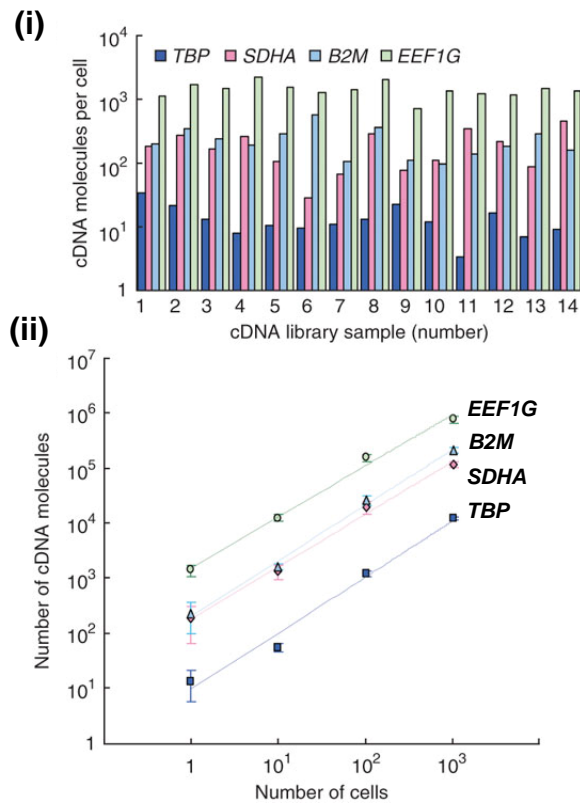
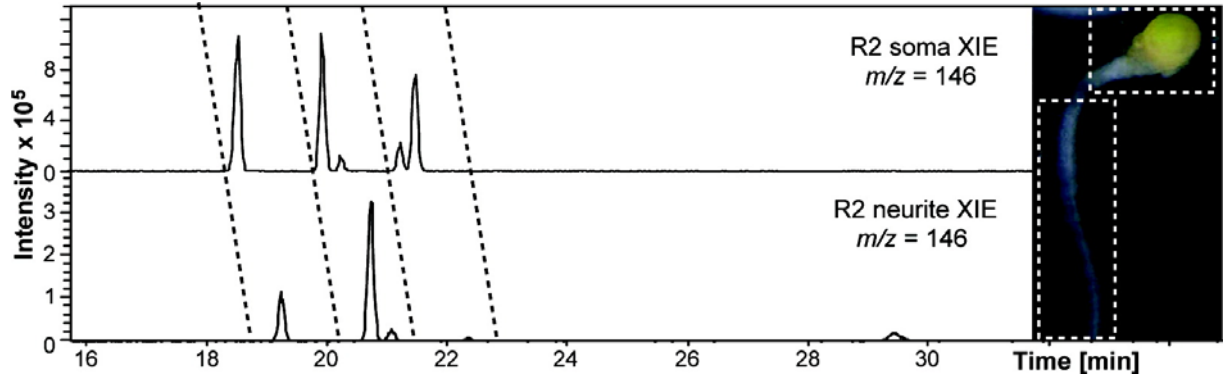


Figure 2

(a)



(b)

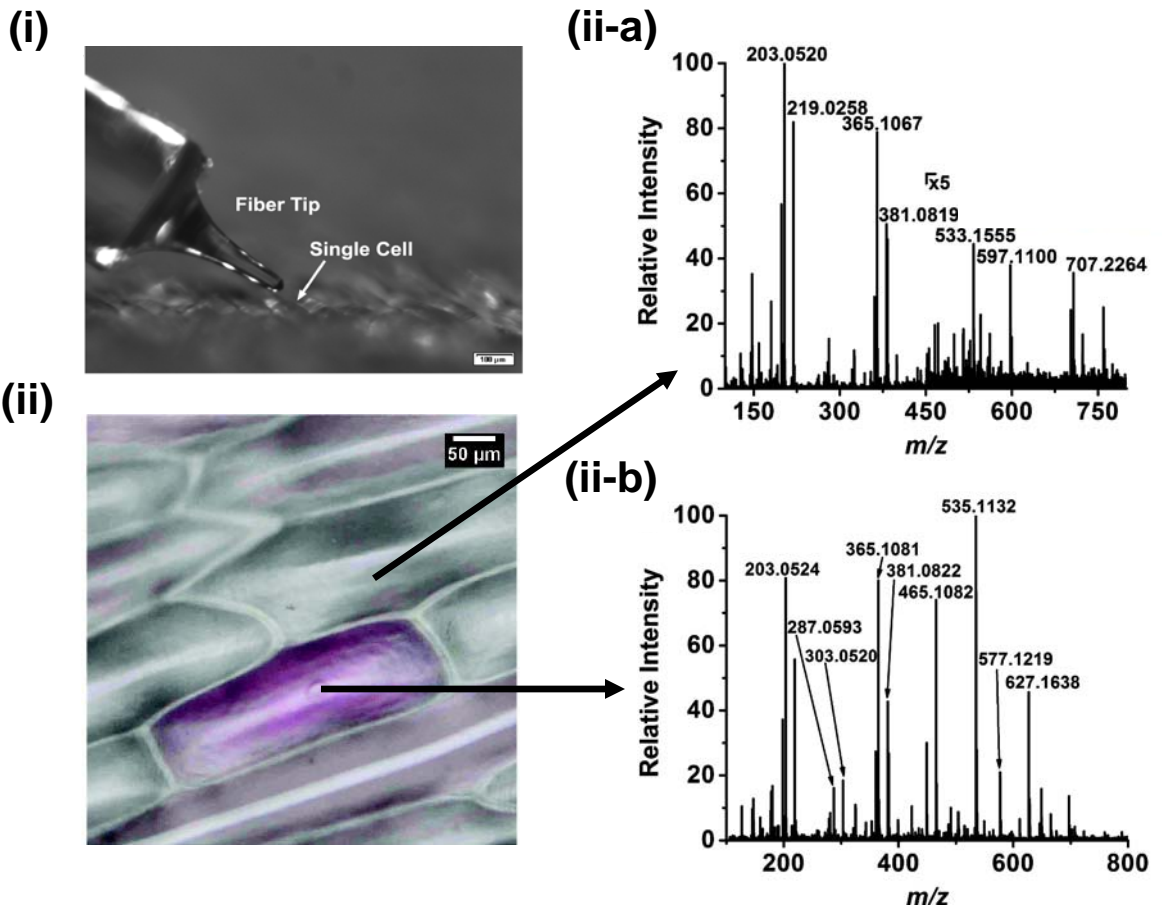


Figure 3

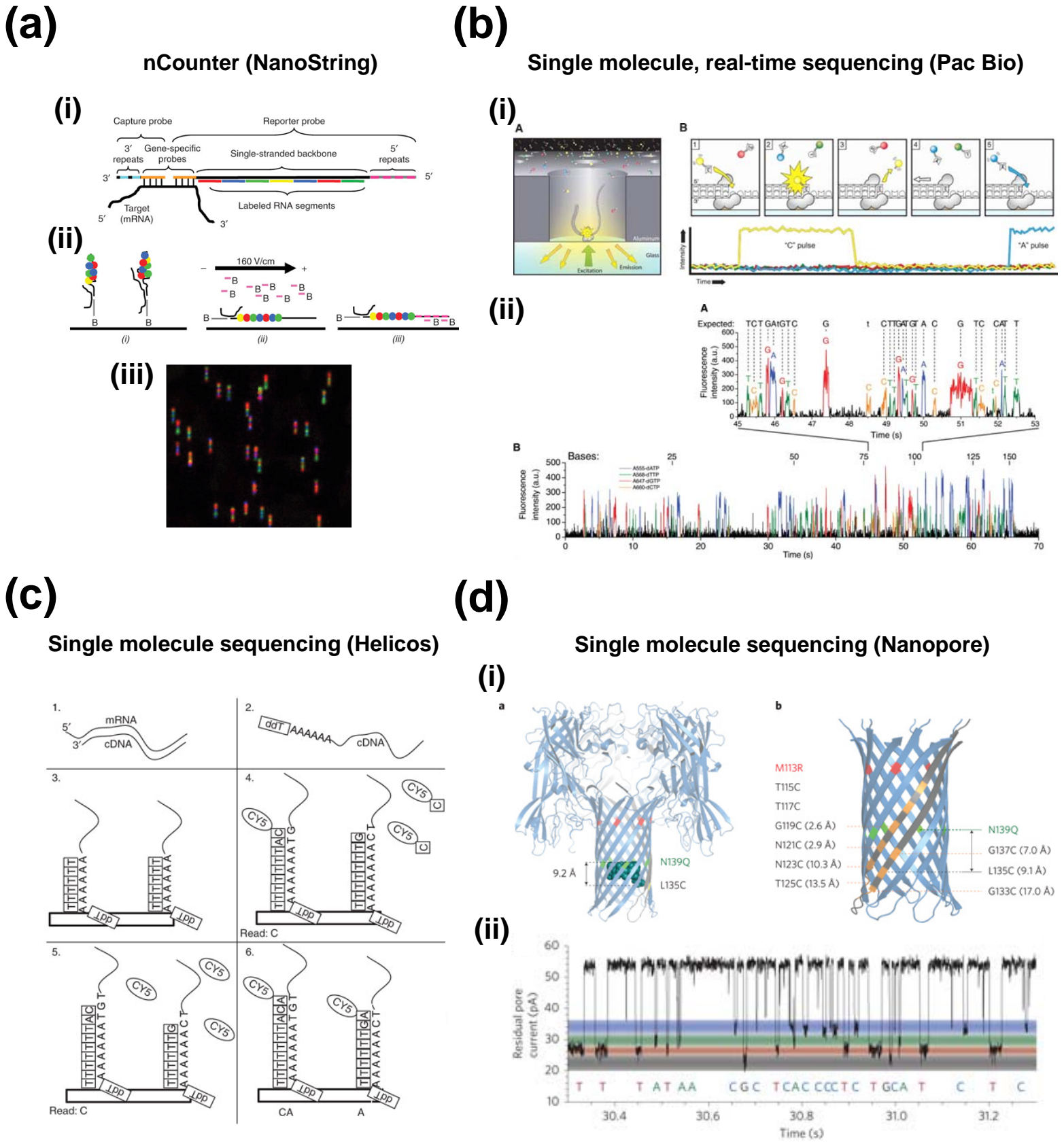
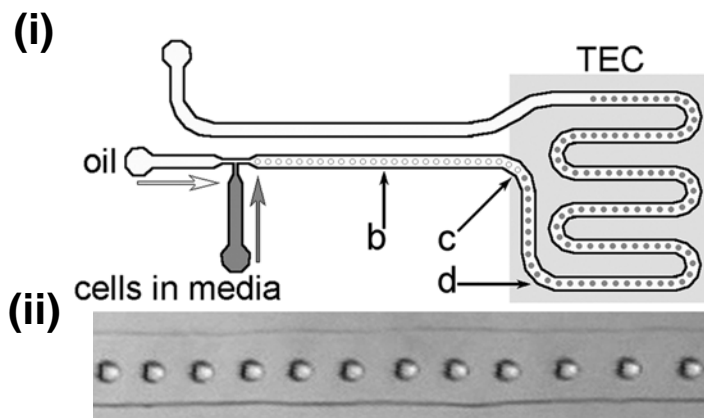


Figure 4

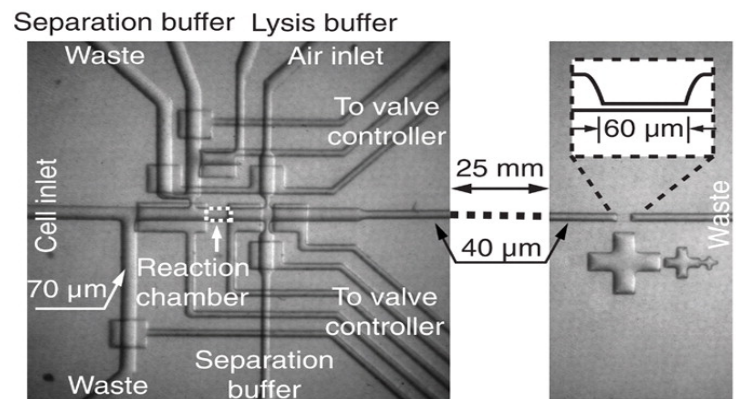
(a)

Single-cell encapsulator



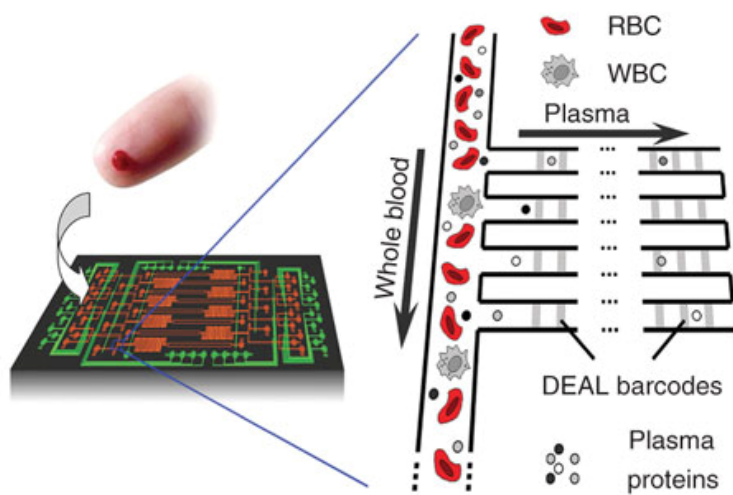
(b)

Single-cell analysis chip



(c)

Barcode chip



(d)

M³ emitters

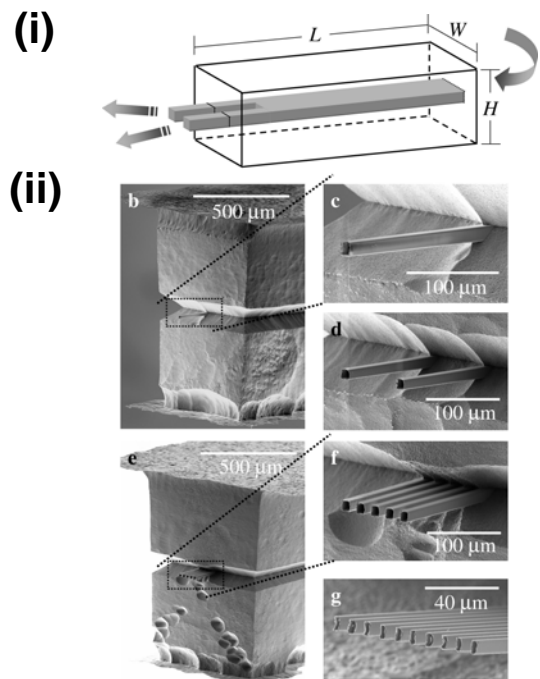


Figure 5