# Phenotypic Testing and Sorting of Artificial Cells: Building a Bridge from Spectroscopy to Genetics Postprint

**Authors:** Ma Bo, Xu Jian

**Date:** 2023-03-19T00:00:00+00:00

## Abstract

In recent years, genome sequencing, editing, and synthesis technologies have undergone rapid advancements, propelling unprecedented progress in the ability to "design" and "synthesize" genotypes, while simultaneously rendering phenotypic detection of artificial cells a critical bottleneck in synthetic biology development. For the rapid testing and evaluation of cellular functions, single-cell analysis technologies are of significant importance and hold great promise; however, an ideal solution must exhibit features including non-invasiveness to living cells, label-free operation, provision of panoramic phenotypes, capability to resolve complex functions, high speed and throughput, low cost, and integration with omics analyses. From this perspective, this article highlights progress in single-cell functional characterization, sorting, and omics technology systems based on label-free molecular spectroscopy, and discusses key challenges and future directions in this field. The synergistic exploitation of multiple spectroscopic techniques and multimodal imaging, combined with high-throughput spectroscopy-activated cell sorting technology and downstream single-cell omics technologies, is establishing and broadening an expansive bridge connecting spectroscopy and genetics. This bridge not only provides a novel solution for high-throughput, panoramic phenotypic detection and screening of cell factories, but will also promote "single-cell-precision spectroscopic phenome-functional genome" as a new type of biological big data to serve synthetic biotechnology driven by "data science".

## Full Text

# Phenotyping and Sorting of Synthetic Cells: Building a Bridge from Spectroscopy to Genetics

**MA Bo, XU Jian***
Single-Cell Center, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao 266101, China

## Abstract

Recent years have witnessed rapid advances in genome sequencing, editing, and synthesis technologies, which have propelled capabilities in genotype "design" and "manufacturing" to unprecedented heights. Concurrently, phenotyping of synthetic cells has emerged as a critical bottleneck in synthetic biology. For rapid testing and evaluation of cellular functions, single-cell analysis technologies hold significant promise. However, an ideal solution must feature non-invasive live-cell probing, label-free operation, landscape-like phenotyping capability, discrimination of complex functions, high speed and throughput with low cost, and seamless integration with downstream omics analysis via cell sorting. This article focuses on recent progress in label-free molecular spectroscopy-based single-cell functional characterization, sorting, and omics technologies, and discusses key challenges and future directions in this field. The synergistic application of multiple spectroscopic techniques and multimodal imaging, combined with high-throughput spectroscopy-activated cell sorting and downstream single-cell omics, is constructing and expanding a broad bridge connecting spectroscopy and genetics. This bridge not only provides novel solutions for high-throughput, landscape-like phenotyping and screening of cell factories, but also promotes single-cell-precision "spectroscopic phenome-functional genome" as a new type of biological big data, serving data science-driven synthetic biotechnology.

**Keywords:** synthetic biology, phenotyping, molecular spectroscopy, spectroscopy-activated cell sorting, single-cell phenome, single-cell functional genomics

## Introduction

The core mission of synthetic biology is to artificially design and construct novel biological systems with specific physiological functions, thereby establishing biomanufacturing routes for pharmaceuticals, functional materials, or energy substitutes, based on elucidating and modeling the fundamental principles of biological synthesis [1]. The leapfrog development of this technology depends on breakthroughs in three common technical components: "genotype design," "genotype synthesis," and "cellular phenotype testing" (the design-build-test cycle) [2]. In recent years, dramatic improvements in the throughput and cost of genome sequencing and synthesis, along with widespread application of genome editing technologies, have enabled the design and construction of mutants and

even artificial cells with unprecedented power [3-6]. However, the development of cellular phenotyping speed and throughput has lagged far behind, sometimes by several orders of magnitude. For example, screening for cells with desired phenotypic combinations from genotype mutant libraries often requires enormous manpower, funding, and time (the screening for microbial cell factories producing antimalarial drug precursors consumed approximately 150 person-years of effort [7,8]). Consequently, cellular phenotyping has become one of the "rate-limiting steps" in synthetic biotechnology development.

### 1.1 Status of Artificial Cell Phenotyping Technology

**1.1.1 Characteristics of Single-Cell Detection and Analysis Technology** For rapid testing and evaluation of cellular functions, an ideal single-cell analysis technology should possess six characteristics: (1) Live-cell detection. In many cases, the functions of genetic elements and modules only have biological significance when measured and investigated non-invasively in living cells (i.e., with minimal perturbation to cellular state). Additionally, live-cell detection means that cells can be directly subjected to subsequent cultivation or other operations after measurement. (2) Label-free operation. As mentioned earlier, from the perspective of mining natural biological elements, microbial cells in nature exhibit enormous genetic diversity and are often difficult to cultivate. Therefore, there is currently no universally applicable cell labeling method for cell types in complex microbiomes. From the viewpoint of gene function screening based on chassis cells such as *E. coli* and yeast, the phenotypes and functions that can be labeled using fluorescent probes remain extremely limited compared to the potentially measurable cellular functions. Consequently, label-free cell detection offers significant advantages. (3) Provision of landscape-like phenotypic information. Many cellular functions of interest to researchers are reflected or determined by multiple phenotypes. Measuring only a single phenotype or characterizing a single compound, protein, or gene often fails to probe the target function, whereas "landscape-like" analysis capable of simultaneously providing multiple phenotypes or even measuring the phenome offers important advantages. (4) Ability to resolve complex functions. Many important and even core cellular functions are reflected or determined collectively by multiple genes. Therefore, detection targeting a single protein or compound molecule often cannot distinguish and identify such functions. (5) Rapid, high-throughput, and low-cost operation. A single colony on a plate can contain up to $10^9$ cells, thus single-cell-precision phenotyping imposes higher demands on speed, throughput, and cost compared to colony-level analysis. (6) Integration with single-cell functional genomics analysis. Single-cell-precision genomics, transcriptomics, proteomics, metabolomics, and epigenomics represent one of the most rapidly advancing frontiers in life science methodology [9]. If single-cell phenotyping or phenome analysis can be directly coupled with these single-cell functional genomics approaches through cell sorting, it would enable the establishment of truly single-cell-precision "phenotype-genotype" models, thereby driving breakthroughs in synthetic biology at the individual cell, cell population, and even

cell community levels.

Single-cell analysis does not rely on cell cultivation but directly analyzes each individual cell's function, thus overcoming the challenge that most microbial cells in the environment remain difficult to cultivate. This holds great significance for mining biological elements, modules, or chassis cells from human and environmental microbiomes.

### 1.1.2 Types of Single-Cell Spectroscopic Detection Technologies

Metabolites are the final products of gene expression in cells and typically the most direct carriers of cellular phenotype and function. Therefore, detection of the metabolome, including identification of metabolic states, represents one of the most direct and effective approaches for cellular function testing [10]. Currently, in artificial cell testing workflows, gas chromatography-mass spectrometry (GC-MS), liquid chromatography-mass spectrometry (HPLC-MS), or nuclear magnetic resonance (NMR) are commonly used for systematic analysis of cell "populations" or "communities" to achieve cellular function detection [11]. However, metabolites from single cells, especially single microbial cells, are extremely minute and cannot be amplified like nucleic acids. Consequently, single-cell metabolomics analysis remains challenging due to sensitivity limitations of current mass spectrometry, chromatography, and NMR detection [12]. Moreover, these methods typically require cell lysis to extract or prepare intracellular metabolites, making it difficult to directly interface with subsequent cultivation or genomic and transcriptomic analysis of the same single cell. In contrast, single-cell spectroscopy collects molecular spectra from individual cells at specific times, spatial locations, and states, capable of reflecting and displaying characteristics of intracellular metabolite(s). Furthermore, the measurement process can be non-invasive, non-destructive, and even maintain live-cell status. Therefore, spectroscopy-activated cell sorting can isolate cells with specific spectral signatures and directly couple them with single-cell cultivation and various destructive single-cell functional genomics analyses. Thus, single-cell spectroscopy technology offers an effective means to overcome these challenges.

#### (1) Cell Function Detection Based on Fluorescence Spectroscopy

Fluorescence spectroscopy offers high sensitivity, excellent specificity, and high analytical throughput, making it one of the most widely applied single-cell phenotyping technologies [13]. However, most cells lack autofluorescence, necessitating the design of fluorescent probes for cell labeling. Common fluorescent probes include small-molecule fluorescent dyes and fluorescent proteins. Some small-molecule fluorescent dyes can penetrate cell membranes to directly label target molecules inside cells; for instance, small-molecule enzymatic fluorescent probes enable detection and imaging of intracellular enzymes [14]. Additionally, small-molecule fluorescent probes can be conjugated to antibodies for specific labeling of cell surface antigens, thereby achieving cell function detection. Fluorescent proteins (such as green fluorescent protein) constitute another important

class of fluorescent probes, often used as reporter genes fused with target genes to monitor target gene expression in cells, thus characterizing the dynamic processes and functional heterogeneity of gene expression. Fluorescent proteins can also detect intracellular metabolic small molecules. For example, a series of high-performance genetically encoded fluorescent probes for NADPH (iNap) have enabled high spatiotemporal resolution detection and imaging of NADPH metabolism in live animals, live cells, and various subcellular structures [15,16]; genetically encoded dopamine fluorescent probes and novel acetylcholine fluorescent probes can detect dynamic changes of endogenous dopamine and acetylcholine in fruit flies, zebrafish, and mice [17,18]. Currently, cell function detection based on fluorescent probes has been widely applied in research areas including gene expression regulation, protein spatial localization and trafficking, protein folding, signal transduction, protease activity analysis, biomolecular interactions, and dynamic monitoring of cellular metabolic processes.

Despite these advantages, fluorescence-based cell function detection faces fundamental limitations that restrict its broad application in biological element mining and cell phenotyping. These limitations include the requirement for prior knowledge of biomarkers, the need for cell labeling or even genetic manipulation, and the ability to typically detect only a few labeled molecules simultaneously. For the vast majority of microbial cells in nature, biomarkers are often unknown, and no universal cell labeling methods exist, resulting in a relatively narrow applicable cell type range for fluorescence detection. Even for common chassis cells such as *E. coli* and yeast, most phenotypes constituting the cellular phenome—including metabolic activity of various substrates, metabolite profiles, environmental stress responses, and cell-cell interactions—are difficult to measure directly using a universal fluorescent labeling approach, often requiring construction of independent cellular fluorescent sensors. Introducing fluorescence-based intracellular sensors into cells typically requires design and engineering of transcription factors to convert specific metabolite concentrations into cellular fluorescence intensity [19]; this prerequisite of DNA transformation and genetic manipulation limits applicable cell types and application scenarios. Most importantly, excellent target molecule specificity and simultaneous multi-target detection may be mutually exclusive. Due to interference among multicolor fluorescence, simultaneous detection of multiple gene functional phenotypes cannot yet be widely applied [20], making "landscape-like" measurement of the phenome difficult to achieve.

**(2) Cell Function Detection Based on Raman Spectroscopy**

Raman spectroscopy is a label-free scattering spectrum arising from frequency-shifted scattering of incident light caused by molecular bonds excited to virtual energy states that have not yet returned to their original states [21]. Each single-cell Raman spectrum consists of over 1,500 Raman peaks corresponding to different types of chemical bonds, reflecting multidimensional information about the composition and content of chemical substances within a specific cell. Metabolite composition can rapidly respond to cellular physiological states, en-

vironmental changes, and genetic backgrounds. Therefore, each single-cell Raman spectrum can be regarded as a high-resolution "photograph" with over 1,500 "pixels" for characterizing the state, phenotype, and function of an individual cell. Each Raman peak or combination of peaks in a single-cell Raman spectrum potentially represents a phenotype. Consequently, just as a portrait photograph can simultaneously reveal numerous features of a human face, single-cell Raman spectroscopy can potentially characterize and reflect multiple cellular phenotypes simultaneously, enabling landscape-like phenotypic analysis [22]. Moreover, water molecules do not produce strong Raman signals and thus do not interfere with cellular Raman detection. Therefore, Raman imaging can be performed under optimal physiological conditions for cells, making Raman spectroscopy widely applicable for characterizing living systems.

For cell populations, accurate description of functional phenotypes and their heterogeneity at single-cell resolution is crucial, a need that Ramanome technology can fulfill. The term "Ramanome" refers to a collection of randomly sampled single-cell Raman spectra under specific conditions and time points, conceptually analogous to taking a "group photo" of a cell population (ramanome) or community (meta-ramanome) [22]. The Ramanome is equivalent to a metabolome that can be rapidly and inexpensively measured and monitored at single-cell precision, with its changes reflecting and characterizing the landscape-like, nearly infinite "states" and "functions" of the cellular system. For instance, Ramanome can detect substrate metabolic activity [23,24], quantify product content [25,26], characterize environmental stress responses (such as cellular stress states [27], drug susceptibility and resistance [28,29]), and cell-cell metabolic interactions [30]. These features fully demonstrate the complementarity of Ramanome with transcriptomics, proteomics, metabolomics, and other approaches, positioning it to become a universal tool and a novel phenomics data type for defining, characterizing, and monitoring cellular functions and their heterogeneity.

**(3) Cell Function Detection Based on Infrared Spectroscopy**

Infrared imaging typically employs polychromatic light in the near-infrared region as the excitation source to excite molecular energy level transitions in samples, thereby detecting their infrared absorption spectra. Since the 1990s, infrared spectroscopy has been used to distinguish normal from cancerous tissues [31,32]. Although infrared microspectroscopy can theoretically achieve diffraction-limited resolution, most current infrared microspectrometers are equipped with traditional globar light sources with temperatures of only 1,000–1,500 K and relatively weak brightness. Consequently, when the measurement slit is adjusted to less than 10 m (comparable to cell size), the light flux decreases significantly, severely affecting the sensitivity of single-cell infrared spectroscopy. Therefore, current micro-infrared spectroscopy can mostly only image components with high infrared absorption in cells, such as imaging the spatial distribution and content of carotenoids in *Haematococcus pluvialis* cells [33,34].

Synchrotron radiation, as a novel infrared light source, offers excellent properties

including broad spectral range (10–10,000 $cm^{-1}$), high brightness (2–3 orders of magnitude higher than conventional sources), low divergence, and temporal structure. Its high brightness characteristic is particularly suitable for single-cell micro-infrared spectroscopic imaging studies [35]. It can not only directly obtain intracellular biochemical information but also identify and detect special cellular phenotypes with subcellular imaging resolution. Compared with conventional infrared spectroscopy, synchrotron radiation infrared spectroscopy possesses more powerful capabilities for single-cell phenotyping. However, during cellular infrared spectroscopic imaging, testing often needs to be performed in aqueous solution to maintain cell viability. Water molecules exhibit significant infrared absorption at 3,000–3,750 $cm^{-1}$ and 1,600–1,700 $cm^{-1}$, making elimination of this interference critical for cellular infrared imaging. Microfluidic chip technology can precisely manipulate fluids, with microchannels accurately controlling the thickness of the extracellular water layer to maximally eliminate interference from the aqueous environment on cellular infrared spectra. The emergence of a series of microfluidic chip devices suitable for cellular infrared spectroscopic imaging has laid the platform foundation for high-throughput single-cell infrared imaging [32,36].

**(4) Cell Function Detection Based on Terahertz Spectroscopy**

Terahertz (THz) radiation lies between the microwave and infrared regions of the electromagnetic spectrum, with frequencies ranging from 0.3–3$×10^{12}$ Hz. When probing inside cells, THz is not limited by scattering and thus possesses excellent biological tissue penetration capability, offering promise for in situ and even in vivo cellular phenotyping. A coupled cavity resonator system capable of rapidly measuring flowing cells at 10 GHz frequency has enabled THz spectroscopic measurement of flowing mouse myoblast cells [37]. THz spectroscopic measurement signals directly correlate with cell volume, providing a rapid and accurate method for cell size distribution analysis, with potential applications including detection of circulating tumor cells in blood samples (which are generally larger than white blood cells). Although THz spectroscopic imaging holds significant potential for in situ or in vivo cellular analysis, the cellular phenotypes that can be detected based on published work remain limited and restricted, with practical applications still distant.

### 1.2 Single-Cell Phenotyping and Sorting Based on Spectroscopy

Building upon single-cell functional identification and characterization based on various spectroscopic technologies, spectroscopy-activated cell sorting enables isolation of single cells with specific functions, followed by determination of their corresponding genotypes and even single-cell functional genomes including transcriptomes, proteomes, metabolomes, and epigenomes, thereby establishing "phenotype-genotype" relationships at single-cell precision.

**1.2.1 Fluorescence-Activated Cell Sorting** For decades, fluorescence flow sorting has served as a mainstream cell analysis and sorting technology. Com-

mercial fluorescence-activated cell sorting (FACS) instruments achieve detection and sorting throughputs of tens of thousands of cells per second with high automation and intelligence, but their cost has remained prohibitively high, representing a major barrier to widespread application. In recent years, the introduction of microfluidic technology has substantially reduced FACS costs while offering flexibility and precision advantages. Microfluidics-based FACS employs surface acoustic wave three-dimensional cell focusing technology, enabling precise cell focusing under low sheath flow conditions and avoiding cell damage from high flow velocities and shear forces in conventional FACS [38].

More recently developed fluorescence-activated droplet sorting (FADS) technology based on droplet microfluidics addresses the challenge of detecting secreted proteins or extracellular metabolic small molecules that conventional FACS struggles with, while achieving sorting throughputs up to 30 kHz [39,40]. After phenotypic detection and sorting, individual cells remain encapsulated in individual droplets, maintaining their independence and enabling seamless integration with downstream single-cell cultivation, sequencing, and other omics studies. Using the FADS platform, researchers have achieved toxic phenotyping and sorting of mammalian U937 cells against drug libraries [41] and screened for mutants with high horseradish peroxidase activity from directed evolution yeast mutant libraries (containing approximately $10^8$ mutants) [42]. Other reports describe integration of dual-channel detection systems on FADS platforms to screen directed evolution mutant libraries (approximately $10^7$ mutants) for esterases with high enantioselectivity preferentially producing ibuprofen enantiomers [43]. Recently, by coupling FACS with artificial intelligence, multi-parameter intelligent screening of *Chlamydomonas* mutant libraries has been demonstrated on a high-throughput, high-resolution image-processing-based single-cell phenotyping and sorting platform (intelligent image-activated cell sorting, IACS) [44]. In summary, FACS has accelerated the "testing" component of synthetic biology, enabling it to match the throughput of "design" and "synthesis" and promoting the development of synthetic biology. However, as previously mentioned, FACS is fundamentally limited by its requirement for fluorescent probes and cell labeling, as well as its limited number of simultaneously detectable phenotypes. There is an urgent need to develop cell sorting technologies based on label-free spectroscopic recognition, landscape-like phenotypic analysis, and universal applicability to all cells in nature.

**1.2.2 Raman-Activated Cell Sorting** As previously mentioned, Raman spectroscopy is a non-destructive, label-free single-cell phenotyping tool with broad application prospects in cell function screening based on microbiomes or engineered cell libraries [22]. How to sort target phenotypic cells at high throughput after Raman identification—namely, Raman-activated cell sorting (RACS) [45,46]—and apply them to downstream cultivation and omics analysis is equally important for constructing cell phenotyping platforms. In recent years, a series of Raman spectroscopy-based single-cell sorting technologies and

core devices have emerged, including Raman optical tweezers sorting [47-50], Raman-activated ejection sorting (RACE) [23], Raman-activated cell sorting with optical tweezers and droplet encapsulation (RAGE), Raman-activated microfluidic sorting (RAMS) [51,52], and Raman-activated droplet sorting (RADS) [53]. These new tools effectively couple single-cell Raman phenotyping measurement with genotypic analysis (or cultivation), providing important means for dissecting phenotype-genotype relationships at the single-cell level.

Raman-activated ejection sorting (RACE) [23] operates on the principle of sputtering a thin film onto a Raman measurement substrate, spotting samples onto this film, and encapsulating a receiving microwell array. After Raman detection of cells, specific cells can be ejected by applying a pulsed laser at the target cell location on the substrate chip; the cell is then stripped and propelled into a collection microwell, enabling flexible collection of single or multiple cells per microwell. Affected by spot size, direct ejection is generally used for small cell separation. For large cells and cell clusters, the target cell region can first be laser-cut (RAMD) before ejection. This method is relatively flexible and simple to operate with high sorting accuracy. However, laser radiation during Raman analysis and sorting can cause certain damage to cellular physiological activity [54], and in the dry-chip mode of RACE, heat dissipation is relatively difficult, potentially making radiation-induced cell activity damage more significant. Additionally, due to possible intracellular nucleic acid damage caused by laser radiation, the sequencing coverage of bacterial single cells after ejection generally does not exceed 20%, making genome assembly relatively difficult [23,55].

Raman-activated cell sorting with optical tweezers and droplet encapsulation (RAGE) is a recently developed Raman cell sorting technology created by our team that couples Raman optical tweezers with droplet-based single-cell encapsulation and export. RAGE overcomes the problem that single optical tweezer force cannot achieve target cell detachment from the focal plane for export. By coupling droplet microfluidic technology, it accomplishes precise sorting and rapid export of target single cells. Meanwhile, Raman detection is performed in aqueous phase, maximizing maintenance of cellular physiological activity and enabling precise matching of each cell with its corresponding Raman spectral phenotype to achieve "what you see is what you get." Furthermore, sorted single cells are already encapsulated in water-in-oil microdroplets, allowing direct coupling with subsequent single-cell cultivation and omics analysis. Our data demonstrate that bacterial cell viability is essentially unaffected after sorting through the RAGE system, enabling direct coupling with downstream single-cell cultivation. In single-cell nucleic acid amplification and sequencing directly coupled with RAGE, possibly due to the protective effect of liquid-phase Raman detection on target cells, the whole-genome sequencing coverage of target *E. coli* single cells has been substantially improved (reaching approximately 95%).

The first two Raman cell sorting technologies are completed with cells in static or relatively static states, and their throughput cannot yet meet the demands of high-throughput cell phenotypic function sorting. Developing flow-based Raman

single-cell sorting technology is an inevitable trend. However, limited by the inherently weak Raman signal, the first bottleneck to address is Raman acquisition of cells in high-speed flow. We developed a Raman-activated microfluidic sorting (RAMS) technology based on dielectrophoretic single-cell trap-and-release, which solved this bottleneck and first achieved Raman flow cytometry sorting [51]. The RAMS system integrates a dielectrophoresis-based single-cell trap-and-release unit that can capture single cells in high-speed flow to complete Raman signal acquisition. Building upon this, we further established Raman-activated droplet sorting (RADS) technology, improving Raman cell sorting throughput and system stability [53]. By employing dielectrophoretic droplet sorting technology, the RADS system currently achieves the highest reported full-spectrum sorting throughput among RACS systems, reaching hundreds of cells per minute. Meanwhile, sorting accuracy for astaxanthin content in *Haematococcus pluvialis* exceeds 95%, with post-sorting cell viability reaching 93% [53]. Notably, the current sorting throughput is primarily limited by Raman detection rather than the system itself; coupling with high-sensitivity Raman detection technologies (such as stimulated Raman) [56] could enable ultra-high-throughput sorting.

The establishment and expansion of the above RACS technology family provides new and powerful technical means for artificial cell phenotyping and screening, and offers novel solutions for applications that current FACS cannot easily address [57]. However, these unit technologies alone cannot directly meet the practical needs of artificial cell phenotypic screening. There is an urgent need to construct complete cell screening workflows and equipment systems based on these key single-cell Raman sorting technologies, tailored to the practical demands of synthetic biotechnology.

#### 1.2.3 Other Label-Free Cell Function Detection and Sorting Methods

In addition to differences in spectral properties, variations in cell phenotype and function can also lead to changes in cellular physical properties. Therefore, cell phenotypes can be identified and sorted through measurement of single-cell physical parameters. For example, single-cell electrical impedance can reflect cell status and is directly related to various phenotypes including cell size and growth stage, and has been used to sort normal human cells from tumor cells [58]. Similarly, single-cell mechanical parameters differ between normal and tumor cells, with parameters such as cell deformability showing potential for tumor diagnosis [59].

The aforementioned physical properties based on cellular electrical impedance and mechanical deformability have also been applied to cell function sorting. Different single-cell manipulation force fields such as optical tweezers, electric field forces (dielectrophoresis, etc.), acoustic waves (surface acoustic waves), hydrodynamic forces, and magnetic fields can be conveniently integrated into microfluidic chip systems to achieve high-throughput, high-precision single-cell function sorting [60]. Overall, to accommodate the diverse demands of cell phenotypic sorting, single-cell function sorting technologies based on new con-

cepts or principles continue to emerge, advancing toward goals of high signal specificity, high detection dynamic range, label-free operation, landscape-like analysis, high throughput, and artificial intelligence integration [36,53,56,61].

## 2 Bottlenecks and Development Directions of Artificial Cell Phenotyping Methodology

Addressing challenges of live-cell non-invasiveness, label-free operation, landscape-like phenotyping, complex function resolution, rapid high-throughput and low-cost performance, and integration with omics analysis, single-cell spectroscopic imaging and sorting possess important distinctive advantages. This bridge connecting spectroscopy and genetics, linking single-cell phenomes with single-cell functional genomes, is rapidly extending and broadening. However, realizing its full potential requires efforts across many fronts while presenting tremendous opportunities.

**(1) Spectroscopy-Based Single-Cell Phenome Measurement.** On one hand, parallel or combined use of multiple fluorescent probes can expand the number of target molecules simultaneously detectable in single cells; however, due to interference among multicolor fluorescence, simultaneous detection of multiple gene functional phenotypes remains a significant challenge [22]. On the other hand, although Ramanome based on single-cell Raman spectroscopy can measure numerous phenotypes such as substrate metabolic activity, product profiles, and environmental stress responses without probes, "landscape-like" phenome analysis measuring these phenotypes simultaneously at single-cell precision still requires demonstration in specific applications. Meanwhile, phenotypic measurement at the single-cell level inevitably carries noise introduced by stochastic gene expression. Therefore, quantification and tracing of this noise are critical for inferring population- or community-level phenotypes from single-cell spectroscopic phenotypes and for distinguishing changes in single-cell state from genotypic changes.

**(2) Single-Cell Spectroscopic Imaging Balancing "Target Molecule Specificity" and "Landscape-Like Phenotyping."** Achieving high-specificity detection of target molecules through fluorescent probe design represents a hallmark of fluorescence spectroscopy, while Raman spectroscopy can directly analyze multiple phenotypes in virtually any cell in nature. The complementary advantages of these two approaches, their coupled use, and demonstration of compatibility between "target molecule specificity" and "landscape-like phenotyping" will greatly expand the application domains of single-cell science and technology. For example, recently proposed bioorthogonal labeled stimulated Raman scattering microscopy for live-cell imaging, based on stimulated Raman scattering of alkyne single-bond labeling, breaks the size limit of imaging reporter groups; moreover, alkyne reporter groups have virtually no Raman background interference, achieving "bioorthogonality" in Raman spectra. The combination of alkyne metabolic labeling of biomolecules with stimulated Raman microscopic imaging has enabled specific Raman

imaging of lipids, nucleic acids, proteins, and carbohydrates in live cells [62,63]. Meanwhile, combined use of fluorescence and Raman spectroscopy has found some applications in tumor detection [64]. Furthermore, parallel measurement and sorting of spectroscopic parameters alongside electrical and mechanical parameters at the single-cell level will push applications of single-cell phenome in synthetic biology to new dimensions and make important contributions to high-throughput phenotypic monitoring and molecular breeding in animals, plants, and microorganisms [65-67]. In spectroscopy-based single-cell phenome analysis, beyond innovations in multimodal detection and sorting principles and core devices, artificial intelligence and big data technologies will also play indispensable roles [44].

**(3) Standardization, Instrumentation, and Intelligence of the Complete Single-Cell "Imaging-Sorting-Sequencing-Cultivation-Big Data" Workflow.** First, to some extent, for label-free spectroscopies such as Raman and infrared, single-cell spectral signal quality positively correlates with excitation light energy delivered to cells. Therefore, spectral acquisition may damage cells and their nucleic acids, resulting in low post-sorting single-cell cultivation viability, low efficiency of single-cell genome amplification, and hindering throughput improvement of the spectral acquisition-sorting workflow. Although our preliminary data show that flow-based Raman detection and microdroplet encapsulation can protect cell viability after laser irradiation [53] and improve quality of nucleic acid amplification and sequencing after Raman sorting, innovative solutions are still needed to substantially increase measurement and sorting throughput while ensuring cell viability and signal quality. On the other hand, methodologies for single-cell Raman and infrared spectroscopic measurement and analysis are still being optimized, with considerable distance from standardization in experimental workflows, computational analysis, and data aspects. Therefore, urgent collaborative efforts across the field are needed to establish spectroscopic acquisition and analysis technology and equipment standards for specific cell types and application scenarios, laying the foundation for large-scale collection and sharing of molecular spectroscopy-based single-cell "phenome-genome" big data in the future. Additionally, single-cell spectroscopic testing devices operate on different principles, making them suitable for different cell types and states. For example, due to the presence of photosynthetic pigments, photosynthetic cells often require a "quenching" process before full Raman spectrum measurement, potentially affecting analysis and sorting speed. Therefore, constructing a fully automated synthetic biofoundry platform also requires consideration of different requirements in working principles, operational processes, and analytical throughput among spectroscopic detection principles and cell properties, among various phenotyping devices, and between genotype design/synthesis and cell phenotyping testing stages. Building upon this foundation, aided by big data and cloud computing, novel equipment and technology service networks targeting specific single-cell testing, sorting, sequencing, and cultivation needs will continuously emerge to support the scaling and intelligence of synthetic biofoundry platforms.

In summary, non-invasive single-cell functional characterization based on spectroscopy remains a challenging yet promising endeavor. The synergistic alliance among multiple spectroscopic methods, even multimodal imaging combining optical, electrical, and acoustic approaches, coupled with high-throughput spectroscopy-activated cell sorting and downstream single-cell omics technologies, will undoubtedly construct a broad bridge connecting spectroscopy and genetics. This will provide next-generation solutions for rapid and automated characterization, screening, and mechanism elucidation of cell factories. Meanwhile, based on this bridge, single-cell spectroscopic phenome and corresponding single-cell genomic and transcriptomic data will gradually accumulate. This single-cell-precision "phenome-genome" as a new type of biological big data will spawn a series of novel, common, and systematic experimental instruments, computational tools, and even big data networks, accelerating data science-driven design and construction of cell factories and thereby transforming research paradigms and methodologies in synthetic biology.

Therefore, we should seize current opportunities and, based on close domestic and international collaboration, jointly plan ambitious, visionary, creative, and competitive programs for single-cell phenome analysis and sorting methodology and equipment development. This will contribute a series of "Made in China" and "Created in China" new methods, new tools, and new instruments to synthetic biology and its vast application domains.

## References


1. Check E. Synthetic biology: designs on life. Nature, 2005, 438(7067): 417-418.
2. Khalil AS, Collins JJ. Synthetic biology: applications come of age. Nature Reviews Genetics, 2010, 11(5): 367-379.
3. Gibson DG, Glass JI, Lartigue C, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. Science, 2010, 329(5987): 52-56.
4. Annaluru N, Muller H, Mitchell LA, et al. Total synthesis of a functional designer eukaryotic chromosome. Science, 2014, 344(6179): 55-58.
5. Ostrov N, Landon M, Guell M, et al. Design, synthesis, and testing toward a 57-codon genome. Science, 2016, 353(6301): 819-822.
6. Richardson SM, Mitchell LA, Stracquadanio G, et al. Design of a synthetic yeast genome. Science, 2017, 355(6329): 1040-1044.
7. Ro DK, Paradise EM, Ouellet M, et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature, 2006, 440(7086): 940-943.
8. Kwok R. Five hard truths for synthetic biology. Nature, 2010, 463(7279): 288-290.
9. Wang D, Bodovitz S. Single cell analysis: the new frontier in 'omics'. Trends in Biotechnology, 2010, 28(6): 281-290.
10. Fiehn O. Metabolomics—the link between genotypes and phenotypes.

Plant Molecular Biology, 2002, 48(1-2): 155-171.
11. Villas-Bôas SG, Mas S, Åkesson M, et al. Mass spectrometry in metabolome analysis. Mass Spectrometry Reviews, 2005, 24(5): 613-646.
12. Zenobi R. Single-cell metabolomics: analytical and biological perspectives. Science, 2013, 342(6163): 1243259.
13. Zhang Y, Wang D. Single-cell analysis: the new frontier in 'omics'. Trends in Biotechnology, 2010, 28(6): 281-290.
14. Liu H, Chen L, Xu C, et al. Recent progresses in small-molecule enzymatic fluorescent probes for cancer imaging. Chemical Society Reviews, 2018, (47): 7140-7180.
15. Zhao Y, Yang Y. Profiling metabolic states with genetically encoded fluorescent biosensors for NADH. Current Opinion in Biotechnology, 2015, 31: 86-92.
16. Tao R, Zhao Y, Chu H, et al. Genetically encoded fluorescent sensors reveal dynamic regulation of NADPH metabolism. Nature Methods, 2017, 14(7): 720-728.
17. Jing M, Zhang P, Wang G, et al. A genetically-encoded fluorescent acetylcholine indicator for in vitro and in vivo studies. Nature Biotechnology, 2018, 36(8): 726-737.
18. Sun F, Zeng J, Jing M, et al. A genetically-encoded fluorescent sensor enables rapid and specific detection of dopamine in flies, fish, and mice. Cell, 2018, 174(2): 481.
19. Wang T, Guan C, Guo J, et al. Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. Nature Communications, 2018, 9: 2475.
20. Mardis ER. Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics, 2008, 9: 387-402.
21. Raman CV, Krishnan KS. A new type of secondary radiation. Nature, 1928, 121: 501-502.
22. Xu J, Ma B, Su X, et al. Emerging trends for microbiome analysis: from single-cell functional imaging to microbiome big data. Engineering, 2017, 3(1): 66-70.
23. Jing X, Gou H, Gong Y, et al. Raman-activated cell sorting and metagenomic sequencing revealing carbon-fixing bacteria in the ocean. Environmental Microbiology, 2018, 4: 29727057.
24. Wang Y, Song Y, Tao Y, et al. Raman-activated cell sorting and metagenomic sequencing revealing carbon-fixing bacteria in the ocean. Environmental Microbiology, 2018, 4: 29727057.
25. Ji Y, He Y, Cui Y, et al. Raman spectroscopy provides a rapid, noninvasive method for quantitation of starch in live, unicellular microalgae. Biotechnology Journal, 2014, 9(12): 1512-1518.
26. Liu J, Huang Q. Screening of astaxanthin-hyperproducing Haematococcus pluvialis using Fourier Transform Infrared (FT-IR) and Raman microspectroscopy. Applied Spectroscopy, 2016, 70(10): 1639-1648.
27. Wang T, Ji Y, Wang Y, et al. Quantitative dynamics of triacylglycerol accumulation in microalgae populations at single-cell resolution revealed

by Raman microspectroscopy. Biotechnology for Biofuels, 2014, 7: 58.
28. Teng L, Wang X, Wang X, et al. Label-free, rapid and quantitative phenotyping of stress response in E. coli via ramanome. Scientific Reports, 2016, 6: 34359.
29. Tao Y, Wang Y, Huang S, et al. Metabolic-activity-based assessment of antimicrobial effects by D2O-labeled single-cell Raman microspectroscopy. Analytical Chemistry, 2017, 89(7): 4108-4115.
30. Wang X, Ren L, Su Y, et al. Raman-Activated Droplet Sorting (RADS) for label-free high-throughput screening of microalgal single-cells. Analytical Chemistry, 2017, 89(22): 12569-12577.
31. Diem M, Mazur A, Lenau K, et al. Molecular pathology via IR and Raman spectral imaging. Journal of Biophotonics, 2013, 6(11-12): 855-886.
32. Sabbatini S, Conti C, Orilisi G, et al. Infrared spectroscopy as a new tool for studying single living cells: Is there a niche. Biomedical Spectroscopy and Imaging, 2017, 6(3-4): 85-99.
33. Liu J, Huang Q. Screening of astaxanthin-hyperproducing Haematococcus pluvialis using Fourier Transform Infrared (FT-IR) and Raman microspectroscopy. Applied Spectroscopy, 2016, 70(10): 1639-1648.
34. Liu J, Song L, Huang Q. Rapid screening astaxanthin-hyperproducing Haematococcus pluvialis mutants through near infrared spectroscopy. Letters in Applied Microbiology, 2016, 62(2): 185-191.
35. Petibois C, Cestelli-Guidi M, Piccinini M, et al. Synchrotron radiation FTIR imaging in minutes: a first step towards real-time cell imaging. Analytical and Bioanalytical Chemistry, 2010, 397(6): 2123-2129.
36. Li F, Wang X, Wang X, et al. Microfluidic Raman-activated cell sorting. Analyst, 2015, 140(18): 6178-6183.
37. Chen J, Chen Y, Zhao Z, et al. Terahertz spectroscopy of live cells in microfluidic devices. Analytical Chemistry, 2017, 89(7): 4108-4115.
38. Ding X, Lin S, Kiraly B, et al. On-chip manipulation of single microparticles, cells, and organisms using surface acoustic waves. PNAS, 2012, 109(28): 11105-11109.
39. Sciambi A, Abate AR. Accurate microfluidic sorting of droplets at 30 kHz. Lab on a Chip, 2015, 15(1): 47-51.
40. Qiao Y, Zhao X, Zhu J, et al. Fluorescence-activated droplet sorting of lipolytic microorganisms using a compact optical system. Lab on a Chip, 2017, 18(1): 190-196.
41. Brouzes E, Medkova M, Savenelli N, et al. Droplet microfluidic technology for single-cell high-throughput screening. PNAS, 2009, 106(34): 14195-14200.
42. Agresti JJ, Antipov E, Abate AR, et al. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. PNAS, 2010, 107(9): 4004-4009.
43. Ma F, Chung MT, Yao Y, et al. Efficient molecular evolution to generate enantioselective enzymes using a dual-channel microfluidic droplet screening platform. Nature Communications, 2018, 9(1): 1030.
44. Nitta N, Sugimura T, Isozaki A, et al. Intelligent image-activated cell

sorting. Cell, 2018, 175(1): 266-276.
45. Song Y, Yin H, Huang WE. Raman activated cell sorting. Current Opinion in Chemical Biology, 2016, 33: 1-8.
46. Zhang K, Song Y, Yin H, et al. Raman-activated cell sorting and metagenomic sequencing revealing carbon-fixing bacteria in the ocean. Environmental Microbiology, 2018, 4: 29727057.
47. Eriksson E, Enger J, Nordlander B, et al. A microfluidic system in combination with optical tweezers for analyzing rapid and reversible cytological alterations in single cells upon environmental changes. Lab on a Chip, 2007, 7(1): 71-76.
48. Lau AY, Lee LP, Chan JW. An integrated optofluidic platform for Raman-activated cell sorting. Lab on a Chip, 2008, 8(7): 1116-1120.
49. Xie C, Chen D, Li YQ. Raman sorting and identification of single living micro-organisms with optical tweezers. Optics Letters, 2005, 30(14): 1800-1802.
50. Huang WE, Ward AD, Whiteley AS. Raman tweezers sorting of single microbial cells. Environmental Microbiology Reports, 2009, 1(1): 44-49.
51. Zhang P, Ren L, Zhang X, et al. Raman-activated cell sorting based on dielectrophoretic single-cell trap and release. Analytical Chemistry, 2015, 87(4): 2282-2289.
52. McIlvenna D, Huang WE, Davison P, et al. Continuous cell sorting in a flow based on single cell resonance Raman spectra. Lab on a Chip, 2016, 16(8): 1420-1429.
53. Wang X, Ren L, Su Y, et al. Raman-Activated Droplet Sorting (RADS) for label-free high-throughput screening of microalgal single-cells. Analytical Chemistry, 2017, 89(22): 12569-12577.
54. Song Y, Kaster AK, Vollmers J, et al. Single-cell genomics based on Raman sorting reveals novel carotenoid-containing bacteria in the Red Sea. Microbial Biotechnology, 2017, 10(1): 125-137.
55. Lee KS, Palatinszky M, Pereira FC, et al. Raman-activated cell sorting and metagenomic sequencing revealing carbon-fixing bacteria in the ocean. Environmental Microbiology, 2018, 4: 29727057.
56. Cheng JX, Xie XS. Vibrational spectroscopic imaging of living systems: an emerging platform for biology and medicine. Science, 2015, 350(6264): 1052-1057.
57. Zhang K, Song Y, Yin H, et al. Raman-activated cell sorting and metagenomic sequencing revealing carbon-fixing bacteria in the ocean. Environmental Microbiology, 2018, 4: 29727057.
58. Chawla K, Burgel SC, Schmidt GW, et al. Integrating impedance-based growth-rate monitoring into a microfluidic cell culture platform for live-cell microscopy. Microsystems & Nanoengineering, 2018, 4(1): 1-8.
59. Islam M, Brink H, Blanche S, et al. Microfluidic sorting of cells by viability based on differences in cell stiffness. Scientific Reports, 2017, 7(1): 1997.
60. Shields CW 4th, Reyes CD, Lopez GP. Microfluidic cell sorting: a review of the advances in the separation of cells from debulking to rare cell isolation. Lab on a Chip, 2015, 15(5): 1230-1249.

61. Saeys Y, Gassen SV, Lambrecht BN. Computational flow cytometry: helping to make sense of high-dimensional immunology data. Nature Reviews Immunology, 2016, 16: 449-462.
62. Wei L, Yu Y, Shen Y, et al. Vibrational imaging of newly synthesized proteins in live cells by stimulated Raman scattering microscopy. PNAS, 2013, 110(28): 11226-11231.
63. Hong S, Chen T, Zhu Y, et al. Live-cell stimulated Raman scattering imaging of alkyne-tagged biomolecules. Angewandte Chemie International Edition, 2014, 53(23): 5827-5831.
64. Kong K, Rowlands CJ, Varma S, et al. Diagnosis of tumors during tissue-conserving surgery with integrated autofluorescence and Raman scattering microscopy. PNAS, 2013, 110(38): 15189-15194.
65. Guo Q, Yang W, Wu F, et al. High-throughput crop phenotyping: an accelerator for breeding and precision agriculture. Bulletin of Chinese Academy of Sciences, 2018, 33(9): 940-946.
66. Rozman J, Klingenspor M, Angelis MH. A review of standardized metabolic phenotyping of animal models. Mammalian Genome, 2014, 25(9-10): 497-507.
67. Bochner BR. Global phenotypic characterization of bacteria. FEMS Microbiology Reviews, 2009, 33(1): 191-205.

**MA Bo**
Lead of the Microfluidic Systems Group at the Single-Cell Center, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences (CAS-QIBEBT). He obtained his Ph.D. in Analytical Chemistry from the Dalian Institute of Chemical Physics, Chinese Academy of Sciences in 2008. From May 2008 to July 2012, he conducted postdoctoral research at the Crump Institute for Molecular Imaging at UCLA and Rice University. He joined CAS-QIBEBT in 2012. Since 2003, he has been dedicated to fundamental and applied research on microfluidic chip technology in analytical chemistry and life sciences. His current research focuses on microfluidics-based high-throughput single-cell analysis technology and instrumentation, as well as high-throughput screening, breeding, and directed evolution of industrial enzymes, strains, and microalgae. E-mail: mabo@qibebt.ac.cn

**XU Jian***
Lead of the Functional Genomics Group and Director of the Single-Cell Center, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences (CAS-QIBEBT). He obtained his M.S. in Computer Science and Ph.D. in Biochemistry from Washington University in St. Louis in 2003. From 2004 to 2008, he served as Team Lead for Genome Assembly and Analysis at the Genome Institute, Washington University. He joined CAS-QIBEBT as a full-time faculty member in 2008. His research primarily focuses on developing synthetic biology technologies for industrial microalgae, Ramanome technology, microbiome data science tools, and single-cell analysis instrument series, serving applications in photosynthetic biofuels, microecological health, marine monitoring, and biosafety. He has published over 100 peer-reviewed papers

cited nearly 8,000 times and serves as Senior Editor of *mSystems*. E-mail: xujian@qibebt.ac.cn

*Corresponding author

*Note: Figure translations are in progress. See original paper for figures.*