Integration of flow cytometry and single cell sequencing

Andreyev*, **, Dmitry S., Zybailov Boris, L.**

*X-BIO Institute, University of Tyumen, Tyumen, Russia

**Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, AR.

Correspondence: dan14444@yahoo.com or BLZybaylov@uams.edu

Keywords: single cell sequencing, cytometry, DROP-SEQ, CITE-SEQ, REAP-SEQ

Abstract

Integrating cytometric analysis of cells, mitochondria and other polynucleotide-containing biological particles with high-throughput single particle sequencing would provide an ultimate bioanalytical tool, simultaneously assessing phenotype, functionality, genome and transcriptome of each particle in a large population. Here, we describe how such integration could be performed by adapting existing, well-established technologies.

There is a set of important problems in biology and biomedicine linked to heterogeneity in phenotypes and genetic makeup of cells or other biological particles with genetic content, such as mitochondria, nuclei, and exosomes, among others. Examples include studies of polyclonal seeding in metastatic cancers, carcinogenesis prevention in stem cells, microbiome-host interactions, drug resistance development, organism development and aging, mitochondrial heteroplasmy in neurodegeneration, carcinogenesis, infertility, inherited pathologies, and many others [1]. Thousands to millions of particles, such as cancerous cells in an organism or mitochondria in an oocyte, have to be individually analyzed.

High-throughput analysis of single cell genomes and transcriptomes has become possible only recently, with increased sequencing power and the development of barcoding technologies such as Drop-Seq [2]. In these, polynucleotides from each cell are labeled with a oligonucleotide “barcode” that is unique for that cell. To do this, multitudes of particles carrying barcodes unique for each cell are fabricated via a “split-and-pool” process [2]; so that each cell is paired with a barcoding particle in a droplet (Fig. 1-5), followed by lysis and polynucleotides ligation with “barcodes”.

Shortly after the appearance of single cell sequencing via barcoding technologies, a simultaneous phenotype and genome/transcriptome high-throughput analysis of single cells emerged via CITE-SEQ and REAP-SEQ, offering phenotypic insight (limited to antibody-antigen interactions) in addition to sequencing data [3,4]. However, antibody-based phenotype analysis is limited to exposed epitopes, while on-flow methods, such as cytometry and capillary electrophoresis are more comprehensive, for example enabling insight into membrane potential, cell stage, size and morphology [5,6]. These non-destructive methods could supply cells already assessed for phenotypic properties to a barcoding chip for genotype/transcriptome analysis. However, the randomness of split-and-pool barcode synthesis, currently used in Drop-Seq, prevents establishing one-to-one correspondence between phenotypes and sequences.

A possible way to achieve such one-to-one correspondence is to use identifiable particles carrying unique and *known* oligonucleotide barcodes. In contrast to the regular barcoding, such

non-random barcodes would tell us not only if a set of sequences originated from the same cell, but also which particular on-flow–analyzed cell they came from (Fig. 1).

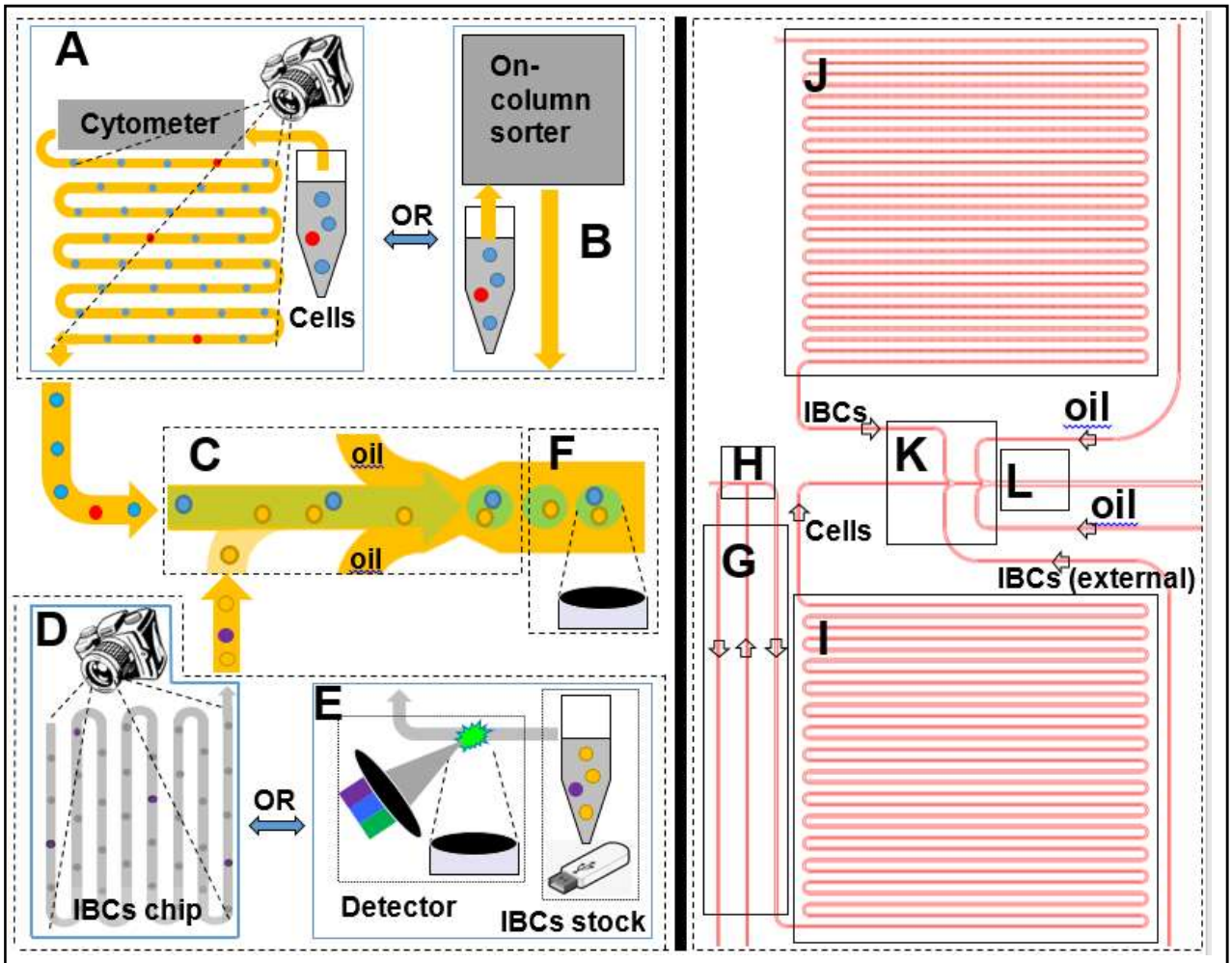

Figure 1. Integration of on-flow analysis and barcoding.
Left panel: **a block-chart of the instrument**. A cytometer (*A or B*) supplies stream of cells (*blue dots*) with reference cores (*red dots*) to the droplet generator (C). Simultaneously, IBCs are supplied to the same droplet generator from a serial position-encoding chip, where IBC identification is performed by a serial number (D) or from a pool of IBCs with unique IDs where identification is performed by a combination of fluorescent tags at the detector (E). Reference cores (*A,B red dots*, D,E *violet dots*) are used for error managing. After pairing cells with IBCs (C), droplet compositions are monitored (F) and processed/sequenced as in regular Drop-Seq systems.
Right panel: **layout example of an on-chip instrument embodiment**. Electroosmotic pumps (channels, separately accessed by electrodes and containing ionized groups on surface, G) are employed to propel and direct cells through detection spot and sorting junction (H). Cells of interest are equidistantly stored in cell accumulator (I) and when supplied to droplet generator (J). Simultaneously, IBCs are supplied to the droplet generator from a chip (K) or from an external source (i.e. left panel E*).* After pairing detector (L), cells are processed/sequenced as in regular Drop-Seq systems.

Possible designs of identifiable barcoding carriers (IBCs)

First, a method for a carrier identification is required. It could be identification by a serial number in the consecutive outflow (Fig.1D) or by a unique, on-flow detectable cores (Fig.1E). Second, a method for oligonucleotide synthesis with unique and known sequence for each carrier is also needed.

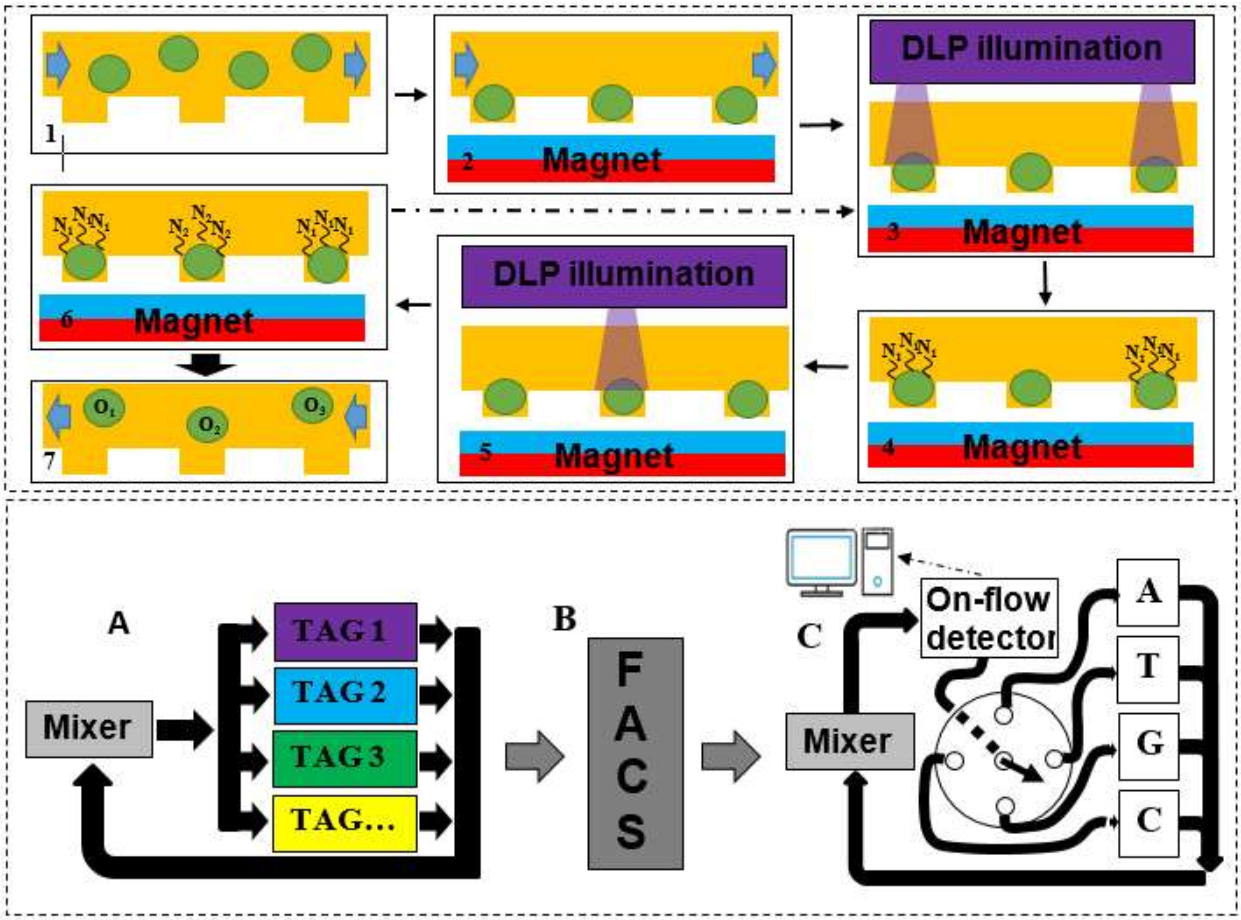

Figure 2. IBC fabrication approaches.

Top: **Processing of IBC chip with magnetic immobilization, side view.** (1,2) Magnetic beads with surface anchors protected by photo-cleavable groups (*green*) are immobilized in designated spots of a microfluidic channel (*yellow*) by external field. Chip with meandering channel (i.e. Fig.1D,J) is processed as a regular "gene chip": (3) spot-selective optical de-protection via digital light processing, (4) attachment of a nucleotide with another photo-cleavable protective group, (5,6) repetition until all desired oligonucleotides are completed. The processed chip could be stored in the same field until use, when (7) the magnet is removed and IBCs are serially supplied to the instrument.

Bottom: **Processing of IBC cores with on-flow detectable IDs**. (A) IDs generation: cores are combinatorially labeled by fluorescent tags. (B) Duplicate removal: all cores with the same ID (combination of tags) but one are removed using a cell sorter. (C) ID-specific barcode generation by tracked mixing synthesis: cores from a mixer are randomly directed to one of four chambers where A, T, G, or C nucleotide is added to them, and the on-flow detector records which nucleotide was added to each given core. Then, the core returns to the mixer and the cycle repeats. The resulting mixture of IBCs with a corresponding database of ID-barcode links could be stored until use, when IBCs subsequently pass through the on-flow detector to the instrument.

For IBCs identified by a serial number in the consecutive outflow of cores, such as beads or shells [7], can be reversibly immobilized in a microfluidic channel on surface of a chip at pre-defined spots using magnetic, hydrodynamic [8], and dielectrophoretic [9] methods. Magnetic immobilization is arguably the simplest approach (Fig.2, *top*) while others could provide controlled, one-by-one core release. Oligonucleotides with pre-defined, unique sequence can be synthesized on each core-containing spot of the chip, as employed in a photolithographic "gene chip" fabrication [10] (Fig.2, *top*). When needed, IBCs could be released from the surface of the chip into a microfluidic channel and serially eluted to the droplet generator (Fig.1C).

For IBCs on-flow identified by intrinsic IDs, the intrinsic IDs could be optical (including fluorescent) labels [11], magnetic, or RFID cores [12]. Commercial cytometric bead arrays [13] contain two fluorophores at ten concentrations, giving 2^{10}=1024 IDs, which might not be sufficient for high-throughput analysis. However, increasing the number of fluorophores to four would provide 4^{10}=1,048,576 IDs; eight fluorophores would provide 8^{10}= 10,737,418,824 IDs. Large increase in the number of IDs available for IBCs with use of few more fluorophores in "split-and-pool" synthesis (Fig. 2*A*) means that the throughput of an instrument employing such IBCs with combinatorial fluorescent IDs, is limited only by sequencing power, not by the number of IDs. Fluorescence-based IDs would require a cytometer-like detector with the matching number of channels. Commercially-available cytometers with adapted software are suited for this task.

After preparing a sufficient number of cores with different IDs, duplicates with the same ID need to be removed. Any cell sorter with adapted software (Fig.2B) can be used, to read a core's fluorescence in each channel, compare intensities with values from previous cores in a data array, sort out duplicate cores if the ID was already seen or record data and collect the core if the ID is new, and repeat the process until all cores are sorted or the ID data array is completed.

Having a set of cores with unique IDs, the corresponding unique oligonucleotide barcodes on each core have to be created. Among possible ways to do this [14], we suggest tracked mixing synthesis as an optimal approach (Fig.2C). Being relatively simple, it does not pre-define a sequence for a given ID, but unique sequences for each ID will be known after the synthesis. Resulting set of IBCs with a complimentary database, defining links between each core ID and barcode sequence, could be supplied to the droplet generator via the on-flow detector.

On-flow analyzer properties

Most cytometers could directly supply cells to barcoding chip (Fig.1C) at the expense of complimentary IBC overuse because cells in outflow are diluted with sheath fluid, and too many IBC-containing droplets will not contain cells. On-column detection (Fig.1B,H) or augmentation of a cytometer with a cell accumulator (Fig.1A,I) would solve the dilution problem and reduce IBC overuse. Field-driven systems, such as electroosmotic systems [5], with a feedback loop could supply cells on demand, eliminating the IBC waste by placing one cell and one IBC into each droptet.

Droplet-based sorters would require serial, ordered collection of droplets and would benefit from a cell accumulator (Fig.1A) to mitigate associated dilution and errors in order. In contrast, flow-manipulating sorters with on-column detection ((Fig.1B,H) would not have these limitations.

Synchronization and error correction using optical and gene tags

Linking a database of barcode sequences with cytometric data is prone to errors. In IBC-generating chips, spots for oligonucleotide synthesis could be empty or occupied by multiple cores, or cores may stick elsewhere. In cell/IBC pairing droplets, the one-to-one ratio could be compromised. In a cytometer, multiple cells could be detected as a single event, or cells may stick or switch elution order. These errors would manifest themselves in mismatching indexes in the barcode database and cytometric data, and they would have to be corrected. Two possible correction approaches could be employed.

For pre-sequencing correction, fluorescently marked reference cores could added into IBCs identified by serial number (but not IBCs with intrinsic IDs), and to analyzed cells. Images of the IBC-producing chip and the cell accumulator (Fig.1A,D) would reveal the elution order of these reference IBCs among regular cells and IBCs (Fig.1A,D, red and violet dots, respectively). These data will be compared with data from on-flow detectors (Fig.1A,B,E,F).

For post-sequencing correction, known barcode sequences of paired reference IBCs (one from the cell sample and another from the IBC supply) will be compared with pairing data (Fig.1F) and order of elution expected from optical analysis (Fig.1 A,B,E).

Applied in parallel, these methods would reveal if IBCs and cells switched positions in elution order or were lost, and they would reveal junk particles reads and pairing mismatches. Compromised data will be repaired for confirmed cases of switched positions. Alternatively, if the correction data is insufficient for repair, the compromised data for cells between reference IBCs (Fig.1A,B, blue dots between red dots) will be eliminated.

Concluding remarks

We anticipate that the recent single-cell sequencing technologies based on in-drop barcoding are ready to be augmented with the full power of on-flow detection methods, such as cytometry, enabling a transformative analytical tool for comprehensive assessment of heterogeneity in genetic makeup and phenotypic properties of cells, organelles, exosomes, and other biological entities. Here we suggested pathways to such an analytical tool using the concept of identifiable barcoding carriers.

Acknowledgement

The project was supported by Center for Translational Pediatric Research (CPTR) NIH Center of Biomedical Research Excellence award P20GM121293 (to Alan Tackett) and in part by the Arkansas Biosciences Institute, the major research component of the Arkansas Tobacco Settlement Proceeds Act of 2000.

References

1. Leites, E.P. and V.A. Morais, Mitochondrial quality control pathways: PINK1 acts as a gatekeeper. Biochem Biophys Res Commun, 2017.
2. Lan, F., et al., Single-cell genome sequencing at ultra-high-throughput with microfluidic droplet barcoding. Nat Biotechnol, 2017. 35(7): p. 640-646.
3. Stoeckius, M., et al., Simultaneous epitope and transcriptome measurement in single cells. Nat Methods, 2017. 14(9): p. 865-868.
4. Peterson, V.M., et al., Multiplexed quantification of proteins and transcripts in single cells. Nature Biotechnology, 2017. 35: 936–939
5. Andreyev, D. and E.A. Arriaga, Simultaneous laser-induced fluorescence and scattering detection of individual particles separated by capillary electrophoresis. Anal Chem, 2007. 79(14): p. 5474-8.
6. McGrath, K.E., Archer, D., Characterizing cell morphology using imaging flow cytometry Science , 2015. 349(6251): p. 999
7. Andreyev, D.S. and E.A. Arriaga, Fabrication of perforated sub-micron silica shells. Scr Mater, 2007. 57(10): p. 957-959.
8. Kimmerling, R.J., et al., A microfluidic platform enabling single-cell RNA-seq of multigenerational lineages. Nat Commun, 2016. 7: p. 10220.
9. Rosenthal, A. and J. Voldman, Dielectrophoretic traps for single-particle patterning. Biophys J, 2005. 88(3): p. 2193-205.
10. Nuwaysir, E.F., et al., Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. Genome Res, 2002. 12(11): p. 1749-55.
11. Sukhanova, A. and I. Nabiev, Fluorescent nanocrystal-encoded microbeads for multiplexed cancer imaging and diagnosis. Crit Rev Oncol Hematol, 2008. 68(1): p. 39-59.
12. Wood, D.K., et al., A feasible approach to all-electronic digital labeling and readout for cell identification. Lab Chip, 2007. 7(4): p. 469-74.
13. Morgan, E., et al., Cytometric bead array: a multiplexed assay platform with applications in various areas of biology. Clinical Immunology, 2004. 110(3): p. 252-266.
14. Battersby, Bronwyn et al. Patent EP02792214A. Device and methods for directed synthesis of chemical libraries, E.P. Office, Editor. 2002.