On-flow analysis and single-cell sequencing with deterministic barcoding

(Integration of Cytometry/FACS and Drop-Seq for simultaneous analysis of phenotype and genome/transcriptome of each cell/organelle in large populations)

Dmitry Andreyev, Boris Zybailov. Department of Biochemistry and Molecular Biology, UAMS, Little Rock, AR

UAMS UNIVERSITY OF ARKANSAS FOR MEDICAL SCIENCES

Introduction

With the rapid advancement of bioanalytical and data processing methods, the study of biological particles with genetic content (BPs) such as cell and organelle properties in bulk is increasingly becoming obsolete, being replaced by high-throughput analysis of individual BPs. Such a broad-ranging shift in biomedical research is driven by increasing biological evidence emphasizing the significance of BP-to-BP heterogeneity, e.g. cellular heterogeneity within a given population. On-flow methods e.g. flow cytometry and its derivatives [1] are the most versatile and the most high-throughput methods for assessing BP phenotypes, e.g. functional properties, molecular content, and morphology. High-throughput individual BP sequencing, e.g. Drop-Seq [2,3], now provides unprecedented insight into cellular heterogeneity of genotype, gene expression, extent of DNA damage, regulatory mutations, etc. However, these methods can either measure phenotype or genome/transcriptome, but not both for the same particle. The randomness of barcoding oligonucleotide sequences in Drop-Seq does not provide for identification of a particular BP of polynucleotide origin, it only establishes if polynucleotides originate from the same BP or not. This prevents matching of cytometry-measured phenotypic properties to genome/transcriptome of individual BPs. Perhaps, the closest technology that does match some phenotypic properties with transcriptomes of individual BPs is a recently reported approach termed CITE-SEQ [4] where antibody-oligonucleotide tagging enables limited phenotype identification using cell surface markers; however, this antibody-based platform offers only limited phenotypic insight, compared to cytometry. Here, we present a high throughput instrument that can unambiguously match Drop-Seq information for a specific BP with cytometric data for that same BP.

Approach: IBC. *Identifiable* barcoding carriers (IBCs, Fig. 1B, 2,3), serially paired with BPs in cytometer's outflow (Fig. 1A, 4, 5) in a regular Drop-Seq procedure (Fig. 1C) would maintain the desired links between cytometric and Drop-Seq data. IBCs are based on identifiable cores (e.g. beads or shells with regulated permeability (Fig. 6) [5], and appropriate deterministic oligonucleotide growing procedures. For cores identification, we suggest two approaches: (i) identification by a serial position in a microfluidic channel on chip, where serial number is additionally verified by reference beads (Fig $1B_1$, 2, 8) or (ii) identification by unique for each bead, on-flow detectable ID (Fig $1B_2$, 3).

Approach: synchronization. Linking a database of barcode sequences with cytometric data is prone to errors. For chip-based IBCs, spots for oligonucleotide synthesis could be empty, occupied by multiple beads, or beads may stick outside the spots. For in-drop pairing, the ratio of one particle to one bead per drop might be compromised. In a cytometer BPs may stick, switch order, or BP pair could be detected as a single event. Errors will manifest themselves as an index mismatch between barcode database and cytometric data, and has to be corrected. Pre-sequencing correction: fraction of fluorescently marked reference beads added into the IBCs (Fig.1B violet) and BPs (Fig.1B red) , to monitor cytometer (Fig.1A) and IBC-chip (Fig.$1B_1$) outflows, and in-droplet pairing (Fig.1D. 5F). Post-sequencing correction: known barcodes at reference beads are analyzed. Any mismatches between expected and actual bead sequence are used for correction of compromised data, e.g. for BPs between reference beads where number of BPs is compromised. Applied in parallel, these methods are expected to provide robust and reliable error correction.

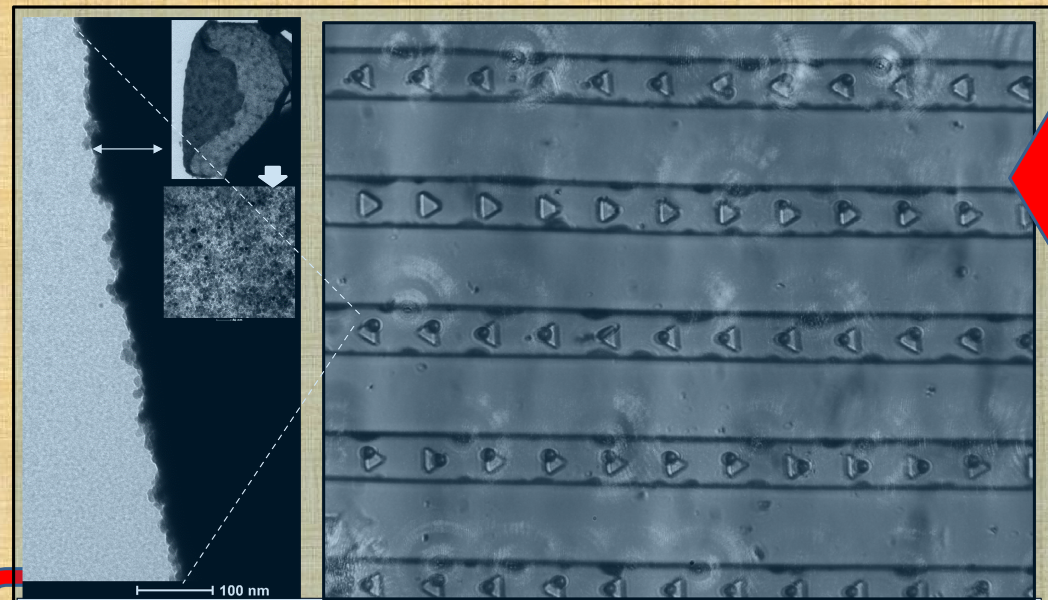


Fig. 6. **Magnetic beads in traps**. Right: Surface-magnetized IBC cores in triangular traps (some traps in the top two channels are empty). Left: surface of a magnetized bead.

Fig. 5. A core layout of on-chip instrument. A: EOF-pump channels. B: BP detection spot and sorter junction (i.e. FACS). C: BP accumulator. D: IBCs chip (see also Fig. 6 right). E: droplet generator. F: droplet composition detection spot.

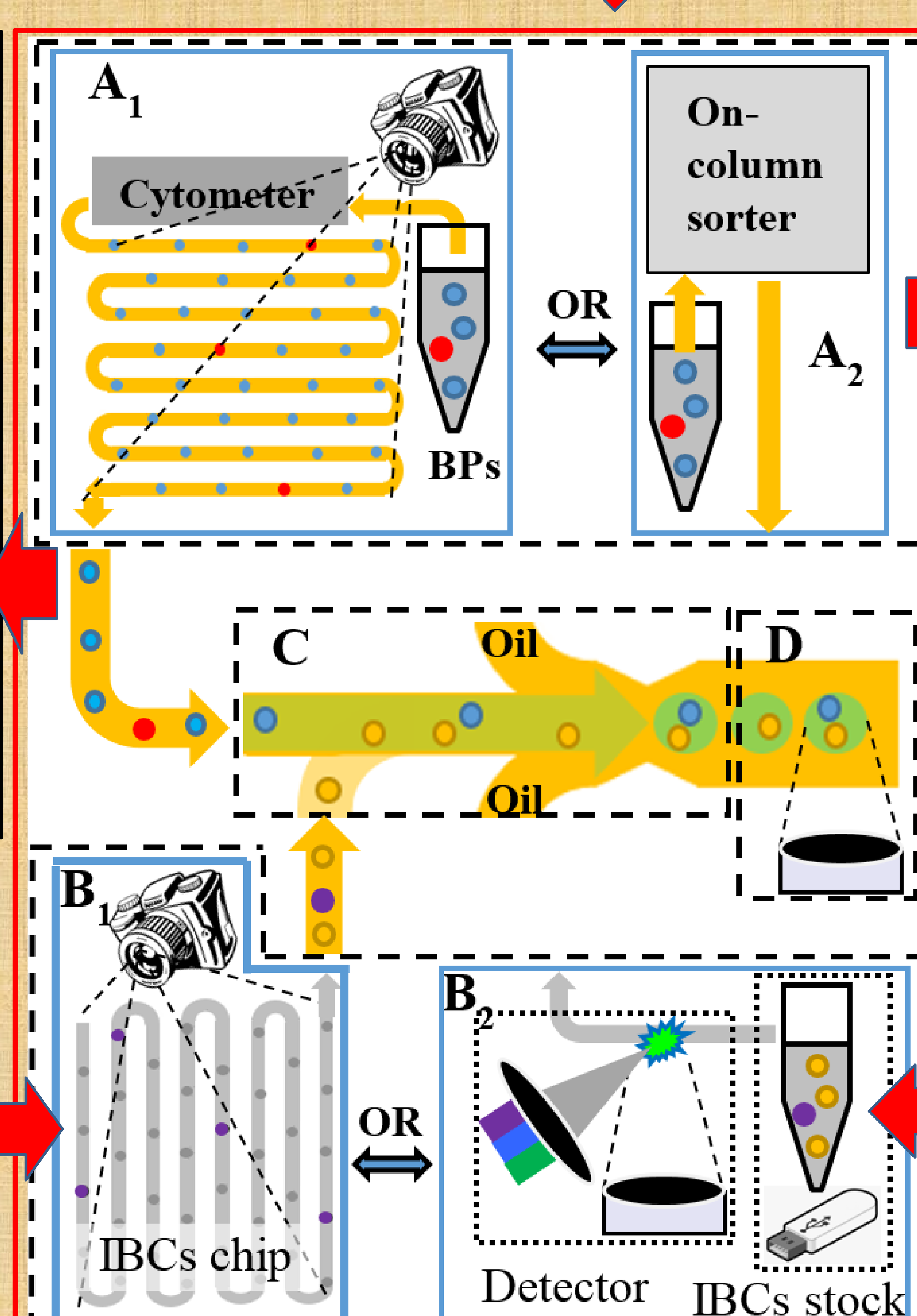


Fig. 1. The instrument block-chart. On-flow analysis module (either A_1 or A_2, see also Fig. 4,5) supplies stream of analyzed BPs with reference beads (red) to droplet generator (C). Simultaneously, IBCs with reference beads are supplied to the same droplet generator from serial position-encoding chip (B_1, see also Fig. 2) or unique ID pool (B_2, see also Fig. 3). After pairing BPs with IBCs (IBC in excess), droplets compositions is monitored (D) and processed/sequenced as in regular Drop-Seq systems.

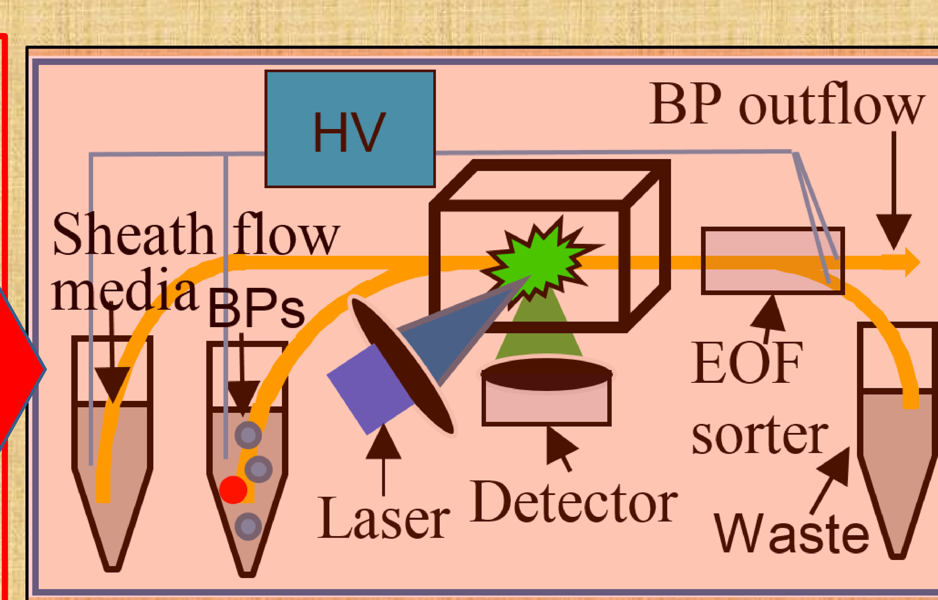


Fig. 4. IBC-saving EOF-based sorter example. Most cytometers could directly supply BPs for pairing only at expense of IBCs overuse because BPs in outflow are diluted with sheath fluid. On-column detection or augmentation of a cytometer with a BP accumulator would solve the problem. Filed-driven, e.g. electroosmotic systems [1] with feedback could supply BP's on demand, eliminating IBC waste. Droplets-based sorters would require serial, ordered collection of droplets and benefit from the accumulator, but on-column sorters has no limitations.

Fig. 7 Silica shells with regulated permeability for use as IBC cores. A: half-dissolved polystyrene templates, B: hollow shells.

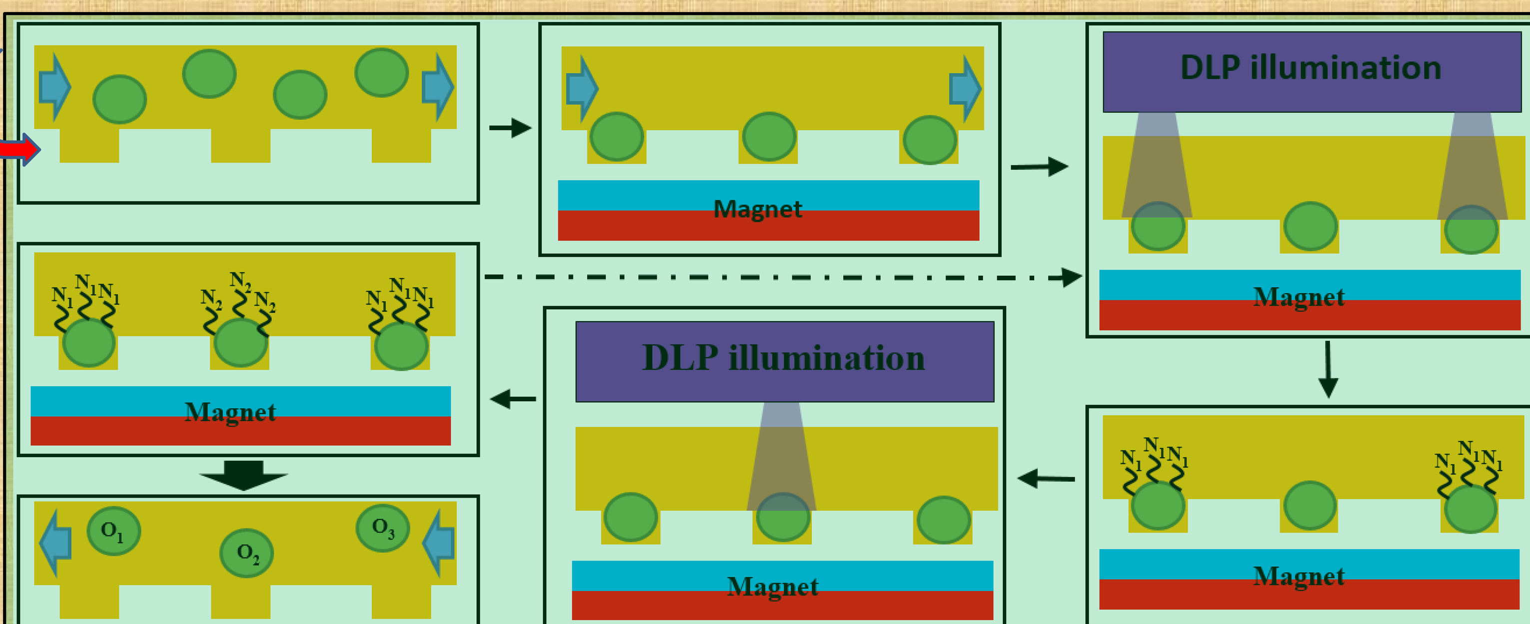
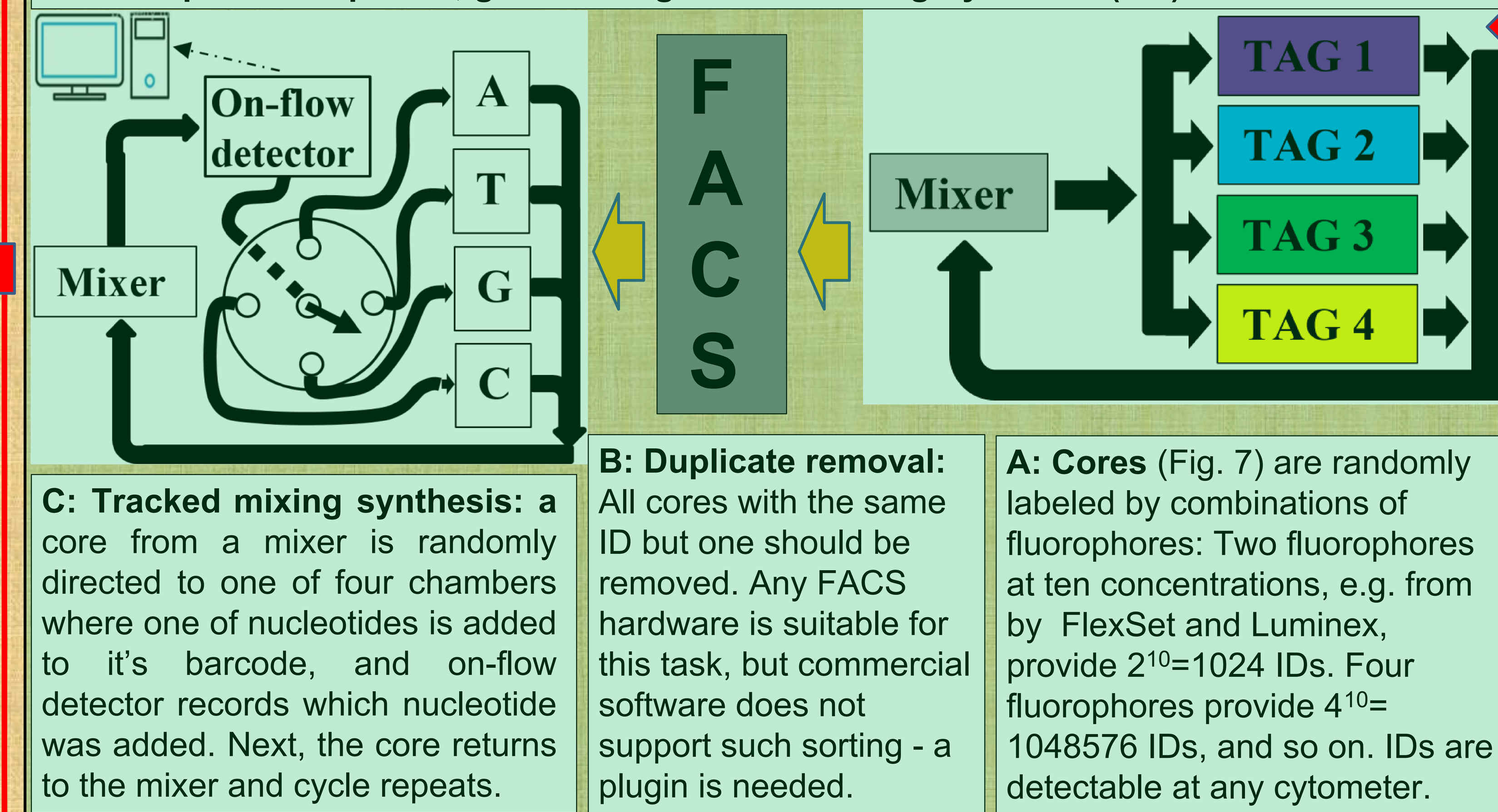


Fig. 2. Processing of IBC chip (such as shown on Fig.1B1), side view: Magnetic beads (Fig 7 left) with surface anchors protected by photo-cleavable groups (green) are immobilized in designated spots of a microfluidic channel by external field, and chip with the channel is processed as a regular "gene chip" [6]: spot-selective optical de-protection via digital light processing (DLP), attachment of a nucleotide with another photo-cleavable protective group, repetition until all desired oligonucleotides completed. The processed chip could be stored in the same field until use, when magnet is removed and IBCs are serially supplied to the instrument.

Fig. 3: Fabrication of IBCs with unique optical IDs (right) and oligonucleotide barcodes with ID-specific sequence, grown using Tracked Mixing Synthesis (left).



C: Tracked mixing synthesis: a core from a mixer is randomly directed to one of four chambers where one of nucleotides is added to it's barcode, and on-flow detector records which nucleotide was added. Next, the core returns to the mixer and cycle repeats.

B: Duplicate removal: All cores with the same ID but one should be removed. Any FACS hardware is suitable for this task, but commercial software does not support such sorting - a plugin is needed.

A: Cores (Fig. 7) are randomly labeled by combinations of fluorophores: Two fluorophores at ten concentrations, e.g. from by FlexSet and Luminex, provide 2^{10}=1024 IDs. Four fluorophores provide 4^{10}= 1048576 IDs, and so on. IDs are detectable at any cytometer.

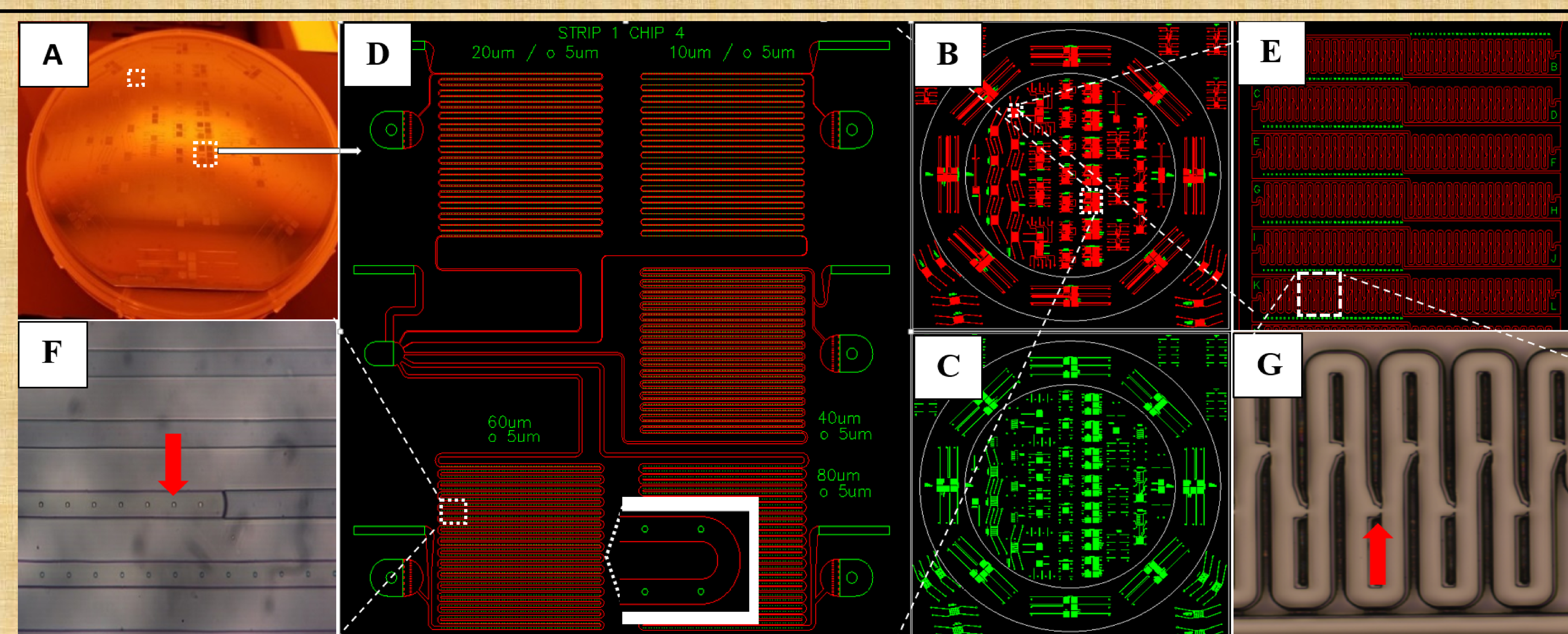


Fig. 8. Fabrication and evaluation of IBC chips. A: Image of SU-8 master for 59 IDC chips. **B,C:** Layouts of 5 µm layer containing traps, and for 40 µm layer containing channels. Layout made suitable for making both 105 mm (inner circle) and 130 mm (outer circle) 2-layer masters, with extra chips at corners. **D, E:** Expanded view of chips with magnetic and hydrodynamic traps. **F:** Image of a region of a PDMS/glass chip, half-filled with running buffer. Traps are visible only in air-filled half of the channel due to higher difference in refractive index. **G:** Image of chip master with hydrodynamic traps. Arrow marks a trap.

References

[1] Andreyev, D.; Arriaga, E. Simultaneous Laser-Induced Fluorescence and Scattering Detection of Individual Particles Separated by Capillary Electrophoresis. Analytical Chemistry (2007), 79(14), 5474-5478.
[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] Zheng, G.X., et al., *Massively parallel digital transcriptional profiling of single cells.* Nat Commun, 2017. **8**: p. 14049.
[4] Stoeckius, M., et al., *Simultaneous epitope and transcriptome measurement in single cells.* Nat Methods, 2017. **14**(9): p. 865-868
[5] Andreyev, D.S. and E.A. Arriaga, *Fabrication of perforated sub-micron silica shells.* Scr Mater, 2007. **57**(10): p. 957-959.
[6] Nuwaysir, E.F., et al., *Gene expression analysis using oligonucleotide arrays produced by maskless photolithography.* Genome Res, 2002. 12(11): p. 1749-55.

Acknowledgements

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