Title
Finding a helix in a haystack: nucleic acid cytometry with droplet microfluidics

Journal
Lab on a Chip, 17(12)

ISSN
1473-0197

Authors
Clark, Iain C
Abate, Adam R

Publication Date
2017-06-13

DOI
10.1039/c7lc00241f

Peer reviewed

HHS Public Access

Author manuscript

Lab Chip. Author manuscript; available in PMC 2018 June 18.

Published in final edited form as:

Lab Chip. 2017 June 13; 17(12): 2032–2045. doi:10.1039/c7lc00241f.

Finding a helix in a haystack: nucleic acid cytometry with droplet microfluidics

Iain C. Clark and **Adam R. Abate***

Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, USA

Abstract

Nucleic acids encode the information of life, programming cellular functions and dictating many biological outcomes. Differentiating between cells based on their nucleic acid programs is, thus, a powerful way to unravel the genetic bases of many phenotypes. This is especially important considering that most cells exist in heterogeneous populations, requiring them to be isolated before they can be studied. Existing flow cytometry techniques, however, are unable to reliably recover specific cells based on nucleic acid content. Nucleic acid cytometry is a new field built on droplet microfluidics that allows robust identification, sorting, and sequencing of cells based on specific nucleic acid biomarkers. This review highlights applications that immediately benefit from the approach, biological questions that can be addressed for the first time with it, and considerations for building successful workflows.

Introduction

Across all of biology cells are categorized by their phenotypes. These characterizations are useful because phenotypes determine the roles of different cells in phenomena like nitrogen cycling (microbes), immune defense (T cells), or sensory stimulation (neurons). Often cells with interesting and complex phenotypes exist within heterogeneous populations where identifying and studying them can be challenging. Separating cells with particular phenotypes, therefore, is essential for developing a mechanistic understanding of their underlying biology. Fluorescence-activated cell sorting (FACS) was developed to accomplish this feat, characterizing large numbers of cells and isolating subpopulations based on specific protein biomarkers. However, it is only possible when the biomarkers exist, are known, and can be identified by fluorescence, which is often not the case.

Phenotypes are ultimately a manifestation of a genetic program expressed under specific environmental conditions. As a result, the clearest biomarker of a phenotype is often the presence of specific sequences encoded in the genome or transcriptome of the cell. Cancer cells with unique transcriptional regimes,[1–3] cells that harbor integrated provirus,[4] and uncultivable microbes, can be uniquely identified by their RNA or DNA content, but often nothing else. Given the centrality of nucleic acids to all living things, and the unique ability

* adam.abate@ucsf.edu; Tel: 415 476 9819.

of specific sequences to distinguish important phenotypes, a tool for isolating cells based only on nucleic acids is essential.

Nucleic acid cytometry is an approach for identifying and sorting nucleic acid sequences free in solution or packaged in cells or viruses. This is analogous to an experimental BLAST sequence search and represents a new and powerful technique for studying subpopulations of biological importance. The technology uses droplet microfluidics to encapsulate cellular material in water droplets suspended in oil, creating discrete reactors. Within each reactor, a soluble detection assay produces a signal that identifies the presence of a specific sequence in a cell, virus, or molecule. Positive droplets, identified by their fluorescence, are sorted and their contents released for further analysis. This technology is a valuable alternative to cell sorting based on surface markers, expanding sorting capabilities from proteins to nucleic acids. Different technical approaches to nucleic acid cytometry are possible using droplet microfluidics; the devices, detection assay, and sorting method may vary depending on the application. Indeed, many different methods for full and partial nucleic acid cytometry have been reported in the literature (Table 1). We do not attempt to review techniques for microfluidic device construction or design, as these topics have been covered extensively.[5–9] The goal of this review is to highlight applications in biology that benefit from this approach, not to review the different workflows and devices possible. We provide examples of how nucleic acid cytometry can be used in targeted comparative genomics, viral genome sequencing, non-coding RNA biology, and rare species enrichment. In addition, we discuss practical considerations for implementation, which draws on our experience developing workflows.

2 The workflow: digital, high-throughput, and flexible

The first step of nucleic acid cytometry is to partition a sample containing a mixture of nucleic acids into aqueous droplets, such that the droplets encapsulate one or no target molecules. A soluble assay is performed in all droplets, interrogating for specific sequences of interest. The simplest embodiment of this operation is digital droplet PCR (ddPCR). Droplets with a target have a higher effective concentration than the bulk sample, enabling reliable single copy detection. Upon amplification, positive droplets become fluorescent, allowing them to be counted, and providing an absolute measure of the starting target concentration.[28,29] The number of partitions that can be generated determines the sensitivity of ddPCR for quantifying rare sequences. With ultrahigh-throughput microfluidic techniques, millions of picoliter droplets can be generated in under an hour, providing over five decades of DNA quantitation and, thus, significantly higher sensitivity than common qPCR. Consequently, ddPCR is displacing qPCR in many instances, particularly in the clinic where absolute quantitation without a standard curve is valuable.

Droplet technologies are not limited to purified DNA: any nucleic acid can serve as a target, provided it can be detected via a fluorescence assay. For example, cells, viruses, mRNA, non-coding RNA, or spliced transcripts can all serve as templates for ddPCR, making the approach general. Sample preparation depends on the starting material, but the concept is the same: a heterogeneous sample of nucleic acid entities is partitioned into droplets and target molecules are detected with an assay. Any assay that produces a sequence specific

fluorescent signal can be used. Polymerase chain reaction or isothermal amplification techniques (loop mediated, recombinase polymerase, rolling circle, etc.) are usually required to yield sufficient DNA for detection. Amplification is coupled to a fluorescent signal using hydrolysis probes (TaqMan), molecular beacons, or intercalating dyes (SYBR). Designing the detection assay requires only knowledge of short segments of target sequence, and the DNA oligomer probes are constructed in a robust and inexpensive chemical synthesis process. This makes detection of specific cell types and viruses far simpler than with antibody labeling, which depends on the availability of specific and sensitive antibodies.

But microfluidic technologies have developed far beyond the capability of just partitioning samples into droplets – they can also reliably sort at kilohertz rates. This adds a new and valuable layer of capability to digital nucleic acid detection: the ability to isolate positive droplets and analyze their contents, thereby allowing correlation of a detected nucleic acid with whatever material is co-encapsulated in the droplet. This opens new avenues in biology, such as isolating genomic mutations that correlate with expression of a cancer gene, determining the host of an uncultivable virus, or measuring correlations between nucleic acids sequences existing within the same cell, even if they are on different chromosomes. Nucleic acid cytometry thus consists of two components: digital droplet detection of target nucleic acids and fluorescence activated droplet sorting (Fig. 1).

3 Cell-free nucleic acid cytometry

While current sequencing technologies have incredible capacity that grows each year, they are nevertheless unable to characterize all relevant nucleic acids in most biological samples. For example, modern high-throughput sequencers provide hundreds of billions of base pairs of sequence information, but a gram of soil can contain over a thousand times this amount.[30] Given the staggering diversity of most samples, the only reasonable approach to answering most questions is to focus sequencing on the relevant nucleic acids. Nucleic acid cytometry affords a powerful solution to this problem, allowing specific molecules to be isolated based on the presence of “keyword” subsequences. Molecules containing this keyword are enriched by sorting, discarding unwanted reads from abundant and uninteresting molecules, and providing far deeper coverage of new and interesting ones. Eastburnet al. (2015)[16] used this approach to enhance coverage of five target regions in the human genome, which improved variant calling with low input DNA. Additional applications of this approach include sequencing integrated provirus, studying genetic variation of specific chromosomal regions, targeted comparative genomics of bacteria, and enrichment of specific sequences from cell-free DNA in blood.

3.1 Targeted comparative genomics

Microbes catalyze a staggering array of chemical transformations in the environment including mercury methylation, degradation of xenobiotic compounds, and carbon, nitrogen, and sulfur cycling.[31–33] Identifying the genetic bases for these phenotypes is a first step towards developing a mechanistic understanding of them. Microbial transformations generally rely on multiple proteins that work together, and are often encoded in close proximity on the genome. Looking for conserved regions in bacteria with the same

phenotype is therefore useful. Often a single enzyme in the pathway has been purified and sequenced, providing a starting point. Sequences surrounding this are compared to find accessory genes that may also be important; when large numbers of genomes are available, a greater picture of the genetic diversity and ecological role can be inferred. For example, denitrification is encoded by, at a minimum, a series of catalytic enzymes for the sequential reduction of nitrate to nitrogen gas.[31] Surrounding these genes are detoxification systems for removing reactive nitrogen species, and sensors and transcription factors for regulating expression in the absence of oxygen. Finding these systems together in multiple organisms, especially distantly related ones, supports the hypothesis that they are functionally involved in nitrate respiration.

A common challenge, however, is that this type of comparative genomics requires genome sequences from organisms with the phenotype of interest. Bulk sequencing of an environmental sample may completely miss sequences from rare, but important organisms. While abundant sequences can provide much insight into an environment, they do not allow targeted sequencing of cells with a particular phenotype. For this reason, in microbiology labs the primary strategy is to isolate, culture, and sequence the microbes with the phenotype of interest. Enrichment by culturing, however, is inherently low throughput, and also suffers from cultivation bias, since the vast majority of microbes cannot be cultured. Phenotypes of microbes that cannot be cultured, cannot be studied. A cultivation-independent approach to finding and sequencing regions of interest would enable far more effective comparative genomics studies.

Nucleic acid cytometry provides an elegant solution, allowing any gene cluster in an environment to be recovered without the need to culture the host microbes. Using a specific conserved gene as a keyword search, intact molecules in which the target sequence is embedded can be recovered and sequenced. The resultant data is a metagenome focused around that keyword, in which thousands of variants can be characterized and compared (Fig. 2). When applied to environmental DNA, nucleic acid cytometry is a transformative tool for targeted comparative genomics without the need to isolate, culture, and sequence entire genomes. Longer read sequencing, or single drop sorting can be used to preserve single molecule resolution and facilitate the assembly of regions contiguous to detected sequence.

A valuable application of targeted comparative genomics is mining environmental biochemical libraries for novel gene or pathway variants. This is an especially interesting approach for tailoring microbial synthesis of natural products, including antibiotics.[34,35] Chemical synthesis relies on petroleum feedstocks, heavy metal catalysts, high temperatures, strong acids and bases, and harsh solvents. Many chemicals with complex structures have no established synthesis pathway or suffer from low yield. In these circumstances, microbial synthesis has the potential to greatly improve the economics of product formation. Enzymes have tailored substrate specificity, are amazingly efficient catalysts at room temperature, and can produce high purity products. Genome sequences available in databases can be mined to identify enzymes in useful pathways. However, the pathways in these databases are skewed towards abundant species. Nucleic acid cytometry addresses this limitation, allowing environmental sequences to be search experimentally using a keyword that targets a

conserved gene. The resulting data is likely to contain a more diverse set of sequences, increasing the chances of finding novel enzymes with the desired substrate, activity, or product.

3.2 Viral genome sequencing

Sequencing viruses is difficult due to their rarity and the need to sequence single genomes separately. Current methods to obtain single genomes rely on serial dilution in well plates, which is severely limited in throughput, providing just tens of genomes per plate. To obtain a comprehensive picture of viral diversity and evolution, orders-of-magnitude more viral genomes must be sequenced. Nucleic acid cytometry provides a powerful solution to this challenge. Tao et al. (2015)[17] used this approach to identify, sort, and sequence recombination sites after con-infection with murine norovirus strains MNV-1 and WU20. A recent demonstration of cell-free DNA sorting enriched for Lambda virus genomes in a background of ΦX174, and used double emulsions to sort using FACS.[36] A similar workflow can be applied to virus that establish latent reservoirs.

Many human viral infections – HIV, herpesviruses, Epstein–Barr, human cytomegalovirus, hepatitis B virus, and human papillomavirus – establish a silent reservoir that prevents eradication.[37] This reservoir can be a barrier to disease treatment and is the major reason an HIV cure does not exist.[38] Cells harboring latent virus can only be identified by the presence of the proviral sequence, and therefore elude both the immune system and our ability to characterize them at the molecular level. However, sequencing proviral genomes in patients with latent infection is an important part of understanding disease progression and prognosis.[39] For example, understanding sequence variation in patients can identify mutations that enable evasion of host immunity and help tailor treatments. Using viral sequences as the keywords by which to recover all DNA molecules with integrated provirus, regardless of where in the genome they reside or in which cell type, yields molecules that contain both full viral genomes and the junctions between virus and host. This allows insertion sites to be mapped and related questions to be addressed.[40] As such, nucleic acid cytometry on cell-free DNA is a powerful tool for characterizing latent viral reservoirs.

4 Whole cell nucleic acid cytometry

DNA or RNA markers are often the best indicators of important phenotypes like cancer[41] and pluripotency.[42,43] Genome-wide analysis of cells with these nucleic acid markers can provide critical insights into their underlying biology. However, this requires that a cell's DNA and RNA be co-encapsulated and preserved during the workflow. Experiments that use purified nucleic acids are not amenable to full genome or transcriptome sequencing of single cells because nucleic acids are fragmented and mixed during bulk purification. Whole cell analysis solves this issue by compartmentalizing all cell contents in the droplets. After encapsulation, single-cell resolution can be maintained by barcoding or sorting.

Barcoding of genomes[44] or transcriptomes[45–48] using droplet microfluidics uniquely tags cellular DNA or RNA with a nucleotide sequence. After sequencing, this tag is used to group reads originating from single cells. In silico clustering of cells based on expression allows subpopulations to be studied individually and in the context of the larger population.

This approach, although high throughput (>10,000 cells), is poorly suited to studying rare cells. Sequencing large numbers of cells reduces sequencing depth per cell, and means that the majority of reads do not contain relevant information. Barcoding also relies on the in silico identification of a cell through sequencing. This makes correlating RNA and DNA sequences within the same single cells difficult, unless the genome and transcriptome are barcoded together, which has never before been described in a high-throughput format.

Nucleic acid cytometry allows the study of cell populations too rare to detect with barcoding workflows. Sequencing of sorted cells yields reads that pertain only to cells of interest, greatly improving coverage of rare molecules. In addition, RNA or DNA can serve as the detection target, and the genome or transcriptome can be sequenced, providing flexibility in how cells are identified and studied. The development of this technology began with high-throughput sequence-based detection and counting.[20,21,23,24,27] Zhang et al. (2012)[23] and Zhu et al. (2012)[24] used agarose to encapsulate cells and capture PCR amplicons that were detectable with flow cytometry. Eastburn et al. (2013)[21] developed a workflow for detecting RNA from whole cells in droplets using RT-PCR. Analysis of drops has expanded from counting of positives to the sorting and analysis of their contents. Novak (2011)[25] developed a method for the PCR-based detection of mammalian cells. Positive sorting and Sanger sequencing of targets confirmed known genomic mutations. Detection of RNA using RT-PCR, followed by sorting from a background of leukocytes, enabled targeted sequencing.[18] Subsequent work showed that sorted cells could be pooled and RNA-sequenced to elucidate genome-wide transcriptional changes associated with cancer.[13] Given the flexibility of TaqMan assays, this approach can also be adapted to detect and sort based on alternatively spliced transcripts or the presence of non-coding RNA.

4.1 Finding cells with alternatively spliced transcripts

Alternative splicing of RNA generates a large number of protein isoforms that modify localization, activity, and protein–protein interactions.[49] As such, splicing greatly impacts the functional biology of a cell, and is associated with both normal cell lineage differentiation and disease.[50–52] For example, splice variants of the neuron-restrictive silencer factor (NRSF) are overexpressed in small cell lung cancer, leading to a stop codon and truncated isoform that has been proposed as a clinical biomarker.[53]Nucleic acid cytometry can find and study cells expressing this NRSF variant. TaqMan assays can distinguish alternatively splice variants using either exon junction spanning primers or probes.[54] Sorting cells with these exon junctions can elucidate the underlying genetic basis and functional consequences of splicing on a genome-wide scale. This is useful for studying splice variants associated with an important phenotype, but for which no mechanistic understanding has been established.

4.2 Finding cells expressing specific non-coding RNA

There is increasing evidence of the importance of non-coding RNA (ncRNA) in normal cell function and disease.[55,56] MicroRNAs (miRNA) post-transcriptionally silence protein-coding genes by promoting mRNA degradation and inhibiting the initiation of translation.[57] miRNA can target multiple mRNAs, making it difficult to understand the effect on cellular pathways on a genome-wide scale. The ability to perform transcriptomics on cells enriched

for a specific miRNA would facilitate an understanding of normal miRNA function, and cases where miRNA dysregulation causes disease. Nucleic acid cytometry enables this by sorting cells based on expression of target miRNAs. It is also applicable to long noncoding RNAs (lncRNA) that have been ubiquitously detected in sequencing data, but for which defining functional roles is often challenging. For example, although functional roles have been established for some lcRNA, including in X chromosome inactivation, allelic-specific expression, and regulation of pluripotency, the majority have no known function.[56] Comparing the transcriptomes of cells enriched for specific lcRNA to those without could significantly advance discovery of lcRNA function (Fig. 3).

4.3 Finding extracellular vesicles with specific cargo

Another area where nucleic acid cytometry benefits biology is in the study of extracellular vesicles (EV), which are classified based on their biogenesis and size as exosomes, microvesicles or apoptotic bodies. Exosomes can transmit proteins, DNA and RNA between cells; miRNA cargo has been shown to influence host transcription upon fusion, providing a mechanism of long-range cell-to-cell communication.[58] Exosome subpopulations have distinct cargo depending on their source and are therefore likely to impact recipients differently.[59] For example, exosomes from cancer may promote cell survival, proliferation, and tumorigenesis[60,61] while ones from B-cells and dendritic cells can promote T cell immune response.[62–64]

A current challenge is to unravel the heterogeneity of EVs; bulk purification and sequencing approaches lose the link between EV cargo and their originating cells. Sorting approaches for understanding this diversity, including the functional significance of minor subpopulations, rely on FACS of submicron particles. This approach has several limitations, including the technical challenges of detecting submicron particles using FACS,[65] and the inability to sort EVs based on miRNA content. miRNA is a major constituent of exosomes, and can play an important role in their mechanistic action after cargo delivery.[66,67] An interesting application of nucleic acid cytometry would be to separate exosomes based on the presence of cancer-specific miRNA biomarkers.[68] This has the significant advantage of single molecule detection and is not subject to sorting challenges based on exosome size; the drop maker used for exosome encapsulation uniquely determines the sorted drop size, resulting in easily detectable (>20 μm) fluorescent drops. Bulk exosome sequencing could first be used to find markers,[69] with subsequent droplet sorting allowing for marker-specific subpopulations to be sequenced separately, preserving the correlation between DNA and RNA cargo. In this manner, tumor specific mutations can be related to exosome RNA, allowing the molecular dissection of tumor heterogeneity without the need for biopsy (Fig. 4).

4.4 Targeted genome sequencing of environmental microbes

Sequencing of environmental samples has revealed massive phylogenetic diversity.[70] Many of the microbes identified using these techniques cannot be grown in the lab, and shotgun metagenomics fails to adequately reconstruct their genomes, either because they are exceedingly rare, or because assembly is difficult in complex samples. Although metagenomics can provide insight into the cultivation conditions required for different

phylogenetic subsets, the majority of microbes identified by sequencing remain uncultivable. Without cultivation, genomes from many taxonomic subsets cannot be purified, preventing deep sequencing and assembly.

As a proof of concept, Lim et al. (2015)[15] sorted E. coli based on the presence of tolA from population of E. coli ΔtolA. The sorted cells contained two point mutations in the outer membrane lipoprotein LpoA (K16A), which was confirmed by sequencing. This approach lays the groundwork for isolating unique subpopulations from environmental samples, providing a means of enriching genomes from taxonomic groups without the need to culture. For example, 16S sequencing of an environmental sample might reveal a novel bacterial clade. Using primers that target this group, cells are sorted, massively enriching their genomes for sequencing (Fig. 5). This approach greatly facilitates the genomics of novel uncultivable clades. Since this can be applied to enrich for organisms based on any sequence keyword, any gene can be used as a biomarker for sorting. For instance, a functional gene involved in antibiotic resistance could be used to sequence and understand the genomes of drug-resistant bacteria. Similarly, the presence of a viral sequence can be used to understand the host-range of phage or to genome sequence specific subsets of the viral community.[10]

5 Environmental virus sequencing

Viruses are the most abundant biotic unit on earth,[71,72] yet their genomic diversity is mostly unexplored due to technical limitations in cultivation, isolation, sequencing, and classification. Moreover, compared to bacterial sequences generated via metagenomics, of which 10% may be novel, over 60% of viral metagenomes have no sequence equivalents in current databases.[73] No single sequence, like ribosomal DNA in bacteria, is conserved in all viral genomes, which makes diversity estimates difficult and necessitates full genome sequencing. Viral metagenomics has greatly expanded the genomes in sequence databases and therefore diversity estimates,[74] but challenges remain in de novo assembly of genomes from samples with incredible sequence variability and uneven coverage.[75] Often it is desirable to complement metagenomics approaches with single virus sequencing, in which all reads originate from a single genome. This increases coverage, improves assembly, and facilitates the prediction of open reading frames. Isolation of single viral particles with FACS allows for full genome sequencing,[76,77] but is biased to abundant species or those that can be labeled with antibodies. This is inadequate given that most viruses are known only by the presence of contigs generated during environmental sequencing. Isolating viruses based on sequence keywords can drastically increase the number of genomes that can be studied. This approach has recently been demonstrated by enriching SV40 genomes from a heterogeneous sample[14] and T4 genomes from a background of ΦX174.[78]

Viruses represent a vast genetic reservoir of virulence factors with human health implications, and metabolic genes that can influence the biogeochemical cycling of nutrients by bacterial hosts.[79,80] A better understanding of the viral sequence space is needed to predict the origin of environmental pathogens and to understand the role of phage in the evolution and function of their microbial hosts. Using nucleic acid cytometry, it is possible to detect pathogenic factors or bacterial metabolic genes in virus, which would allow for a

greater understanding of how viral gene reservoirs transfer pathogenic islands and manipulate bacterial sulfur and nitrogen cycling.[74]

6 Building successful workflows

Applications of nucleic acid cytometry are numerous because any nucleic acid can be used to detect a subpopulation and any co-encapsulated nucleic acid can be analyzed. However, the design and execution of experiments is nontrivial, and a balance between the experimental goals and technical hurdles must be reached. Key decisions when planning an experiment include: the choice of starting material, sample preparation and handling, the detection method, sorting resolution and throughput, and downstream analysis. Each of these considerations is highly coupled, necessitating the development of an integrated biological and microfluidic workflow.

6.1 Sample selection: purified nucleic acids or whole cells?

The choice of starting material is largely governed by the nucleic acid detection target and the analysis goals. In many cases, a simplified workflow that encapsulates purified nucleic acids can be used. Using cell-free nucleic acids simplifies the workflow by avoiding cell handling and obviating the need for in-drop cell lysis. This approach is sufficient when small fragments (<1 Mbase) and their surrounding context are being studied.[16] When using purified nucleic acids as the starting material, careful quantification of concentration is required to achieve appropriate loading of targets per drop. As in digital droplet PCR, controlled fragmentation is used to evenly load droplets with nucleic acids and overcome sample viscosity effects. Using purified DNA or RNA is appealing for biological applications where the detected sequence is contiguous to the analyzed sequence, for example when a gene is detected and its adjoining operon is sequenced.

Whole cells are used when the detected sequence is distant from the analyzed sequence or when the detected sequence is a physically distinct molecule from the analyzed sequence. For example, when the detected sequence is on one chromosome and the analyzed sequence is on another, or when the detected sequence is DNA and the analyzed sequence is RNA. In such cases, encapsulating whole cells preserves the cellular link between these two molecules. In theory, whole cell experiments can be used to detect any nucleic acid, and determine any other measurable property of the detected cell. For example, cells can be sorted based on a genomic mutation and transcriptome sequenced or cells can be sorted based on a transcript and genome sequenced.

6.2 A cultured approach: tips for cell handling

Careful cell handling is an important part of successful experiments. Cells damaged by aggressive centrifugation and re-suspension, or by prolonged incubation at room temperature lyse and release nucleic acids into solution. This results in droplets that contain a detection target but no cell. This is more pronounced when the target is RNA, because RNA copy number can be high and is more readily released when membrane integrity is compromised. In more extreme cases, strands of precipitated DNA can clog the microfluidic device inlet.

To minimize lysis before encapsulation, it is best to reduce residence time in the syringe and keep cells cold during injection.

Cell staining is used to identify cell-containing drops during sorting. Eukaryotic cells are stained using cell permeant calcein dye that is hydrolyzed by intracellular esterases. After staining, cells are re-suspended in a density matching solution (OptiPrep) to prevent settling and maintain the desired encapsulation frequency. The concentration of OptiPrep is adjusted based on the cell's density.

In whole cell experiments, cells are lysed to release nucleic acids for detection. Lysis methods depend on cell type and detection target. While many virus and extracellular vesicles may not require lysis,[14] larger mammalian cells do.[21] Detergents, salts and proteases are commonly employed. There are two major considerations: both the lysis agent used and the released cellular material can inhibit detection or post-sort molecular biology. Cell lysis with Proteinase K followed by heat inactivation, dilution of cell lysate and drop splitting, has been used to lyse and mitigate inhibition.[21] Dilution followed by drop splitting limits the material available for later sequencing, which is especially important in single cell applications. Reverse transcription, even in the presence of inactivated cell lysate, is sub-optimal and may limit coverage in transcriptomic applications.[47]

An alternative to diluting cellular material is to use beads or hydrogels to capture nucleic acids, allowing for the removal of lysis buffers and inhibitors.[25,46] Hydrogels provide flexibility in cases where particularly harsh lysis is required, for example when using enzymes to break down peptidoglycan or release DNA from chromatin. Hydrogels are porous, allowing bulk biological reactions if nucleic acids remain captured. However, hydrogels also permit leakage of cellular material and the potential for cross-contamination.

In any droplet workflow, cross contamination can occur at several stages. Cell lysis in the syringe, co-encapsulation due to cell–cell adhesion or improper re-suspension, droplet merger during thermocycling or re-injection, and false positive sorts can all contribute to cross contamination. Cross contamination can be measured with two-population experiments, where the two cell types can be uniquely differentiated, such as mouse and human cells. Two cell experiments provide an estimate of the workflow's ability to isolate target nucleic acids. When RNA is detected using reverse transcription PCR, the presence of RNA positive drops that do not contain a calcein stained cell is an indicator of cross-contamination during cell handling or drop making.

6.3 Detectomics: assay selection for downstream omics

Sequence specific single molecule nucleic acid detection is difficult, and amplification of a target sequence using polymerase chain reaction or isothermal techniques is generally used to yield a detectable signal. If RNA is the target, a reverse transcription step is added prior to amplification. Detection of amplified material relies on probes that bind the amplicon, or dyes that intercalate DNA or detect a byproduct of amplification. TaqMan hydrolysis assays have been widely adopted because they add an additional layer of specificity; primer dimers and non-specific amplification are much less likely to produce a false positive signal than intercalating dyes. Validated TaqMan qPCR assays are widely reported in the literature and

are available from several manufacturers. In general, TaqMan assays are straightforward to design and use, but our experience has been that robust assays in bulk do not always translate to droplet format. Often several assays must be screened for adequate signal to noise separation. Multiplexing of TaqMan assays is common in droplets, and allows for detection of multiple targets.[81] Other detection strategies, including FRET and molecular beacons have also been successfully implemented,[82–85] but may require more careful design and experimentation to achieve a specific and sensitive signal.

It is important to consider downstream analysis during the selection of a detection method. In particular, thermocycling at high temperature in the presence of divalent cations damages RNA, which may disrupt RNA sequencing applications. For single-cell work, preserving RNA quality during detection is important. RNAses can significantly degrade RNA; it is best to lyse and inactivate cellular material quickly after encapsulation. DNA is also fragmented during thermocycling; in single cell sorting applications this decreases the length of material that can be amplified after sorting. For example, if a 10 kilobase provirus is the target, DNA fragmentation could necessitate stitching smaller amplicons to cover the full region. Isothermal methods including loop-mediated isothermal amplification,[86] and recombinase polymerase amplification[87] are better suited in such cases, especially when processing single cells. A minor issue is drop coalescence during thermocycling, which can make sorting unreliable. Several tactics can reduce coalescence. Avoid handling tubes with gloves, which tend to carry and transfer charge. Reduce temperature and cycles as much as possible and use FC-40 oil instead of HFE-7500 during thermocycling. PCR additives like PEG 6k and Tween-20 improve droplet stability, but should be evaluated with respect to detection performance. Droplet filters that selectively retain large drops can also be beneficial when drops are re-injected for sorting.

6.4 Droplet sorting: the good and the bad

Some of the first applications of droplet sorting were in enzyme evolution, which used bulk homoginization to form double-emulsions that could be FACS sorted.[88,89] However, emulsions generated by bulk techniques are polydisperse, which translates to increased assay variability and bias in sequencing results. Because of this, microfluidics is now often used to generate extremely uniform oil-in-water emulsions.[90–92] Microfluidic sorting can isolate droplets based on fluorescence, absorbance, or electrochemical properties. Solenoid valves,[93] acoustic waves,[94,95] piezoelectric actuation,[96] thermocapillary valves,[97] and dielectrophoresis (DEP)[98] have all been used to actively sort droplets, but DEP is the most common due to its speed and the simplicity of the requisite devices, particularly when using conductive liquid electrodes.[99]

A system for DEP sorting of fluorescent droplets is depicted in (Fig. 6). Lasers are aligned onto the microfluidic channel where they excite fluorophores in droplets. Fluorescence is collected using photomultiplier tubes, which convert photons into electrical signals that are processed with a computer. When the fluorescence reaches a threshold value, a high voltage AC signal is applied to an on-chip electrode, creating a DEP force that pulls the droplet into a collection channel.

Dielectrophoretic droplet sorters generally run at less than 1 kHz, although careful design has pushed this limit to 30 kHz.[100] Sort speeds are constrained by the ability to accurately detect positive droplets, to apply a force strong enough to deflect a droplet's path without splitting it, and to selectively apply a sort force without influencing neighbors. The ability to detect positive droplets depends on their signal to noise ratio, which requires careful assay optimization. At fast speeds, the droplets spend significantly less time in the excitation window, making accurate fluorescence measurements challenging. In addition, the decision to sort relies on real-time analysis of photomultiplier tube signals, which can produce millions of digital data points per second. Analyzing this data in real time and basing sorting decisions on it requires intricately programmed high speed electronics, usually field-programmable gate arrays.[100]

The number of negative droplets that are incorrectly sorted (false positives) and the number of positive droplets that fail to sort (false negatives) are highly dependent on the emulsion quality and sorter design. Several common factors can greatly influence sorting error. False positives increase with aperiodic droplet spacing, which leads to drops entering the sort junction in close succession and interfering with each other's sorting. Controlled spacing of emulsions is generally the best approach to prevent aperiodicity and improve sorting efficiency.[100] Faster spacer oil can increase the distance between drops and assuage aperiodic effects, but also makes the duration of dielectrophoretic pulsing shorter. Voltage can be increased to a point, but eventually the high fields tear droplets apart, electrocoalesce droplets upstream, and even cause a device to fail due to arcing between positive and negative electrodes.

Droplet monodispersity is paramount for accurate sorting with current devices. A polydisperse emulsion results in false positive and false negative sorts. Gating on signal width can reject moderately sized coalesced droplets. Typically, a histogram of droplet durations shows multiple peaks for single, double, and triple-coalesced droplets. However droplets above a certain size will split at the sort junction, with half entering the positive channel. Another common failure with polydisperse emulsions occurs when small droplets catch up to larger ones in the reinjection channel. When the larger droplet enters the waste channel, it causes a local increase in pressure that forces the smaller trailing drop to take the alternative flow path into the sort channel. Adding links between the waste and sort outlets equalizes pressure fluctuations,[90,100] but cannot entirely prevent this effect. False negatives are mostly a result of applying an inadequate dielectrophoretic force by using the wrong frequency, amplitude, or pulse width. At faster speeds, it is beneficial to detect upstream of the electrode to allow longer sorting pulses.

Droplet sorters are customizable, making them ideal for advanced applications. However, they are also difficult and expensive to build, and require experience to operate well. This represents a barrier to broad adoption of nucleic acid cytometry. It is possible to couple the advantages of droplets with the ease of use and availability of fluorescence-activated cell sorting (FACS) instruments using double emulsions.[36,78,101] These water-in-oil-in-water (W/O/W) emulsions can be loaded onto FACS instruments directly. FACS is a mature technology and instruments are often located in core facilities where they are calibrated daily and well maintained. However FACS sorting of double emulsions also has challenges.

Shearing and drop destruction at larger sizes can occur; generally <30 μm double-emulsions are used to avoid this problem.[102] FACS is also susceptible to sorting errors, although built-in systems reject doublets. In contrast, microfluidic sorters don't exclude droplets that enter the sort junction in tandem. However, microfluidic sorters have several unique advantages over FACS. The sorting junction is directly visible under the microscope at all times, and sorting events can be confirmed by triggering a high-speed camera. Whereas shearing and clogging are hidden during FACS, the operator can observe and correct such anomalies during microfluidic sorting.

6.5 Pooled versus single-droplet experiments

Sorted droplets can be recovered individually or pooled together. Which mode to use depends on the experimental goals and has implications for handling and downstream processing. Pooling discards single-droplet resolution, but has several clear advantages. It eliminates the need for excessive post-sort amplification of material, is more tolerant of degradation, and is generally more compatible with standard downstream molecular biology techniques. This was demonstrated in a recent example, where standard RNA-sequencing was applied to a sorted pooled subpopulation.[13] Single drop sorting, while providing single-cell resolution, has several challenges. Positive drops must be reliably sorted, and their contents released and amplified. Unlike bulk collection strategies that can tolerate false negatives or incomplete breaking of drops, these problems compound quickly in single drop sorting. In addition, single drops containing single cells have only picograms of DNA and RNA that must be amplified. The human genome, with roughly 3.2×10^9 base pairs, has only 6.6 picograms of DNA per diploid cell. Concentrations this low cannot be visualized on gels or made into libraries for sequencing, necessitating amplification. While this challenge is not unique to droplets, it does mean that single droplet workflows are much more sensitive to loss of material and contamination. Automation of droplet collection into plates using a mechanical stage is an additional engineering problem that is necessary for reproducible single drop collection.

6.6 Downstream analysis: to the biology

After sorting, droplets must be broken to release their contents. If many droplets are sorted into one collection chamber, 1H,1H,2H,2H-perfluoro-1-octanol (PFO) is widely used to release their contents. This is followed by efficient aqueous extraction, but often requires cleanup to remove PFO, which is inhibitory to molecular biology. Methods such as electrocoalescence or gently heating in presence of an aqueous overlay are more compatible with preservation of limited nucleic acids from single cells. If the detection step amplifies nucleic acids, the amplicon represents a potentially large fraction of the recovered nucleic acids and can interfere with downstream reactions. Amplicons can be removed using biotinylated primers and streptavidin beads or by using dUTP in place of dTTP during amplification with subsequent uracil-DNA-glycosylase digestion.[14,16] Apart from issues of nucleic acid preservation and release of droplet contents, sorted material is amenable to virtually any type of molecular analysis, including epigenetic. For the human transcriptome, which is small compared to the genome, it is possible to sequence the entire transcriptome at a reasonable cost. For human genome sequencing, either targeted amplification of multiple genomic locations or whole genome amplification can generate enough material for

sequencing from single cells. However, the choice between these two methods is generally driven by cost and the objectives of the study, with researchers sequencing large numbers of small regions or small numbers of whole genomes.[103]

7 Conclusions

Cellular phenotypes are diverse and do not exist in isolation; in most cases cells are part of a heterogeneous population. Therefore, to understand unique subpopulations – cancer cells in the blood, immune cells that harbor latent provirus, antibiotic resistant bacteria in the environment – requires a robust isolation technique. Detection and sorting based on surface markers using FACS does not reliably isolate cells based on nucleic acid markers, especially single copy genomic targets, micro RNA, or alternatively spliced transcripts. Because important cells are often only identifiable based on their genomes or transcriptomes, a technology is needed that uniquely identifies and sorts based on specific DNA or RNA sequences. Droplet microfluidics provides a tool for detecting nucleic acids, whether they be in cells, virus, or free in solution, and for sorting this material for downstream analysis. We have presented examples of how this approach can be applied in biology, including in the discovery of ncRNA function, extracellular vesicle sequencing, targeted bacterial metagenomics and environmental viral sequencing. These are just a few of the many areas that can benefit from sorting based on nucleic acid sequence keywords.

References

1. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RCT, Gaasenbeek M, Angelo M, Reich M, Pinkus GS, Ray TS, Koval MA, Last KW, Norton A, Lister TA, Mesirov J, Neuberg DS, Lander ES, Aster JC, Golub TR. Nat. Med. 2002; 8:68–74. [PubMed: 11786909]
2. Huang E, Ishida S, Pittman J, Dressman H, Bild A, Kloos M, D'Amico M, Pestell RG, West M, Nevins JR. Nat. Genet. 2003; 34:226–230. [PubMed: 12754511]
3. Adler AS, Lin M, Horlings H, Nuyten DSA, van de Vijver MJ, Chang HY. Nat. Genet. 2006; 38:421–430. [PubMed: 16518402]
4. Morissette G, Flamand L. J. Virol. 2010; 84:12100–12109. [PubMed: 20844040]
5. Kintses B, van Vliet LD, Devenish SR, Hollfelder F. Curr. Opin. Chem. Biol. 2010; 14:548–555. [PubMed: 20869904]
6. Theberge AB, Courtois F, Schaerli Y, Fischlechner M, Abell C, Hollfelder F, Huck WTS. Angew. Chem., Int. Ed. 2010; 49:5846–5868.
7. Shembekar N, Chaipan C, Utharala R, Merten CA. Lab Chip. 2016; 16:1314–1331. [PubMed: 27025767]
8. Guo MT, Rotem A, Heyman JA, Weitz DA. Lab Chip. 2012; 12:2146–2155. [PubMed: 22318506]
9. Solvas, XCi, deMello, A. Chem. Commun. 2011; 47:1936–1942. RSC.
10. Lim SW, Lance ST, Stedman KM, Abate AR. J. Virol. Methods. 2017; 242:14–21. [PubMed: 28042018]
11. Sukovich DJ, Lance ST, Abate AR. Sci. Rep. 2016; 7:1–9.
12. Lance ST, Sukovich DJ, Stedman KM, Abate AR. Virol. J. 2016; 13:1–9. [PubMed: 26728778]
13. Pellegrino M, Sciambi A, Yates JL, Mast JD, Silver C, Eastburn DJ. BMC Genomics. 2016; 17:361. [PubMed: 27189161]
14. Han HS, Cantalupo PG, Rotem A, Cockrell SK, Carbonnaux M, Pipas JM, Weitz DA. Angew. Chem. 2015; 127:14191–14194.
15. Lim SW, Tran TM, Abate AR. PLoS One. 2015; 10:e0113549. Search. [PubMed: 25629401]

16. Eastburn DJ, Huang Y, Pellegrino M, Sciambi A, Ptáček LJ, Abate AR. Nucleic Acids Res. 2015; 43:e86. [PubMed: 25873629]
17. Tao Y, Rotem A, Zhang H, Cockrell SK, Koehler SA, Chang CB, Ung LW, Cantalupo PG, Ren Y, Lin JS, Feldman AB, Wobus CE, Pipas JM, Weitz DA. ChemBioChem. 2015; 16:2167–2171. [PubMed: 26247541]
18. Eastburn DJ, Sciambi A, Abate AR. Nucleic Acids Res. 2014; 42:e128. [PubMed: 25030902]
19. Konry T, Lerner A, Yarmush ML, Smolina IV. Technology. 2013; 01:88–96.
20. Eastburn DJ, Sciambi A, Abate AR. PLoS One. 2013; 8:e62961. [PubMed: 23658657]
21. Eastburn DJ, Sciambi A, Abate AR. Anal. Chem. 2013; 85:8016–8021. [PubMed: 23885761]
22. Leung K, Zahn H, Leaver T. Proc. Natl. Acad. Sci. U. S. A. 2012; 109:7665–7670. [PubMed: 22547789]
23. Zhang H, Jenkins G, Zou Y, Zhu Z, Yang CJ. Anal. Chem. 2012; 84:3599–3606. [PubMed: 22455457]
24. Zhu Z, Zhang W, Leng X, Zhang M, Guan Z, Lu J, Yang CJ. Lab Chip. 2012; 12:3907–3913. [PubMed: 22836582]
25. Novak R, Zeng Y, Shuga J, Venugopalan G, Fletcher DA, Smith MT, Mathies RA. Angew. Chem. 2011; 123:410–415.
26. Walser M, Pellaux R, Meyer A, Bechtold M, Vanderschuren H, Reinhardt R, Magyar J, Panke S, Held M. Nucleic Acids Res. 2009; 37:e57. [PubMed: 19282448]
27. Kumaresan P, Yang CJ, Cronier SA, Blazej RG, Mathies RA. Anal. Chem. 2008; 80:3522–3529. [PubMed: 18410131]
28. Sykes PJ, Neoh SH, Brisco MJ, Hughes E, Condon J, Morley AA. BioTechniques. 1992; 13:444–449. [PubMed: 1389177]
29. Vogelstein B, Kinzler KW. Proc. Natl. Acad. Sci. U. S. A. 1999; 96:9236–9241. [PubMed: 10430926]
30. Trevors JT. Antonie Van Leeuwenhoek. 2010; 97:99–106. [PubMed: 19921459]
31. Rodionov DA, Dubchak IL, Arkin AP, Alm EJ, Gelfand MS. PLoS Comput. Biol. 2005; 1:e55. [PubMed: 16261196]
32. Pereira IAC, Ramos AR, Grein F, Marques MC, da Silva SM, Venceslau SS. Front. Microbiol. 2011; 2:69. [PubMed: 21747791]
33. Parks JM, Johs A, Podar M, Bridou R, Hurt RA, Smith SD, Tomanicek SJ, Qian Y, Brown SD, Brandt CC, Palumbo AV, Smith JC, Wall JD, Elias DA, Liang L. Science. 2013; 339:1332–1335. [PubMed: 23393089]
34. Mak WS, Tran S, Marcheschi R, Bertolani S, Thompson J, Baker D, Liao JC, Siegel JB. Nat. Commun. 2015; 6:10005. [PubMed: 26598135]
35. Tang X, Li J, Millán-Aguiñaga N, Zhang JJ, O'Neill EC, Ugalde JA, Jensen PR, Mantovani SM, Moore BS. ACS Chem. Biol. 2015; 10:2841–2849. [PubMed: 26458099]
36. Sukovich DJ, Lance ST, Abate AR. Sci. Rep. 2016; 7:1–9.
37. Lieberman PM. Cell Host Microbe. 2016; 19:619–628. [PubMed: 27173930]
38. Siliciano RF. Nat. Med. 2014; 20:480–481. [PubMed: 24804757]
39. Simen BB, Simons JF, Hullsiek KH, Novak RM, MacArthur RD, Baxter JD, Huang C, Lubeski C, Turenchalk GS, Braverman MS, Desany B, Rothberg JM, Egholm M, Kozal MJ, Beirn Terry. Community Programs for Clinical Research on AIDS. J. Infect. Dis. 2009; 199:693–701. [PubMed: 19210162]
40. Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, Spindler J, Ferris AL, Mellors JW, Kearney MF, Coffin JM, Hughes SH. Science. 2014; 345:179–183. [PubMed: 24968937]
41. Liu ET. Curr. Opin. Genet. Dev. 2003; 13:97–103. [PubMed: 12573442]
42. Chen L, Daley GQ. Hum. Mol. Genet. 2008; 17:R23–R27. [PubMed: 18632692]
43. Ivanova NB, Dimos JT, Schaniel C, Hackney JA, Moore KA, Lemischka IR. Science. 2002; 298:601–604. [PubMed: 12228721]
44. Lan F, Haliburton JR, Yuan A, Abate AR. Nat. Commun. 2016; 7:11784. [PubMed: 27353563]

45. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. Cell. 2015; 161:1187–1201. [PubMed: 26000487]
46. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. Cell. 2015; 161:1202–1214. [PubMed: 26000488]
47. Rotem A, Ram O, Shoresh N, Sperling RA, Schnall-Levin M, Zhang H, Basu A, Bernstein BE, Weitz DA. PLoS One. 2015; 10:e0116328. [PubMed: 26000628]
48. Rotem A, Ram O, Shoresh N, Sperling RA, Goren A, Weitz DA, Bernstein BE. Nat. Biotechnol. 2015; 33:1165–1172. [PubMed: 26458175]
49. Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, Stamm S. Gene. 2013; 514:1–30. [PubMed: 22909801]
50. Tazi J, Bakkour N, Stamm S. Biochim. Biophys. Acta. 2009; 1792:14–26. [PubMed: 18992329]
51. Skotheim RI, Nees M. Int. J. Biochem. Cell Biol. 2007; 39:1432–1449. [PubMed: 17416541]
52. Oltean S, Bates DO. Oncogene. 2014; 33:5311–5318. [PubMed: 24336324]
53. Coulson JM, Edgson JL, Woll PJ, Quinn JP. Cancer Res. 2000; 60:1840–1844. [PubMed: 10766169]
54. Vandenbroucke II, Vandesompele J, De Paepe A, Messiaen L. Nucleic Acids Res. 2001; 29:e68. [PubMed: 11433044]
55. Esteller M. Nat. Rev. Genet. 2011; 12:861–874. [PubMed: 22094949]
56. Kung JTY, Colognori D, Lee JT. Genetics. 2013; 193:651–669. [PubMed: 23463798]
57. He L, Hannon GJ. Nat. Rev. Genet. 2004; 5:522–531. [PubMed: 15211354]
58. Valadi H, Ekström K, Bossios A, Sjöstrand M, Lee JJ, Lötvall JO. Nat. Cell Biol. 2007; 9:654–659. [PubMed: 17486113]
59. Willms E, Johansson HJ, Mäger I, Lee Y, Blomberg KEM, Sadik M, Alaarg A, Smith CIE, Lehtiö J, El Andaloussi S, Wood MJA, Vader P. Sci. Rep. 2016; 6:22519. [PubMed: 26931825]
60. Silva M, Melo SA. Curr. Genomics. 2015; 16:295–303. [PubMed: 27047249]
61. Xu W, Yang Z, Lu N. J. Exp. Clin. Cancer Res. 2016; 35:156. [PubMed: 27686593]
62. Raposo G, Nijman HW, Stoorvogel W, Liejendekker R, Harding CV, Melief CJ, Geuze HJ. J. Exp. Med. 1996; 183:1161–1172. [PubMed: 8642258]
63. de Candia P, De Rosa V, Casiraghi M, Matarese G. J. Biol. Chem. 2016; 291:7221–7228. [PubMed: 26887954]
64. Zitvogel I, Regnault A, Lozier A, Wolfers J, Flament C, Tenza D, Ricciardi-Castagnoli P, Raposo G, Amigorena S. Nat. Med. 1998; 4:594–600. [PubMed: 9585234]
65. Kormelink TG, Arkesteijn GJA, Nauwelaers FA, van den Engh G, Nolte't Hoen ENM, Wauben MHM. Cytometry, Part A. 2016; 89:135–147.
66. Halkein J, Tabruyn SP, Ricke-Hoch M, Haghikia A, Nguyen N-Q-N, Scherr M, Castermans K, Malvaux L, Lambert V, Thiry M, Sliwa K, Noel A, Martial JA, Hilfiker-Kleiner D, Struman I. J. Clin. Invest. 2013; 123:2143–2154. [PubMed: 23619365]
67. Squadrito ML, Baer C, Burdet F, Maderna C, Gilfillan GD, Lyle R, Ibberson M, De Palma M. Cell Rep. 2014; 8:1432–1446. [PubMed: 25159140]
68. Park YH, Shin HW, Jung AR, Kwon OS, Choi Y-J, Park J, Lee JY. Sci. Rep. 2016; 6:30386. [PubMed: 27503267]
69. San Lucas FA, Allenson K, Bernard V, Castillo J, Kim DU, Ellis K, Ehli EA, Davies GE, Petersen JL, Li D, Wolff R, Katz M, Varadhachary G, Wistuba I, Maitra A, Alvarez H. Ann. Oncol. 2016; 27:635–641. [PubMed: 26681674]
70. Castelle CJ, Hug LA, Wrighton KC, Thomas BC, Williams KH, Wu D, Tringe SG, Singer SW, Eisen JA, Banfield JF. Nat. Commun. 2013; 4:2120. Search. [PubMed: 23979677]
71. Bergh Ø, BØrsheim KY, Bratbak G, Heldal M. Nature. 1989; 340:467–468. [PubMed: 2755508]
72. Wigington CH, Sonderegger D, Brussaard CPD, Buchan A, Finke JF, Fuhrman JA, Lennon JT, Middelboe M, Suttle CA, Stock C, Wilson WH, Wommack KE, Wilhelm SW, Weitz JS. Nat. Microbiol. 2016; 1:15024. [PubMed: 27572161]
73. Edwards RA, Rohwer F. Nat. Rev. Microbiol. 2005; 3:504–510. [PubMed: 15886693]

74. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaqué D, Coordinators TO, Bork P, Acinas SG, Wincker P, Sullivan MB. Nature. 2016; 537:689–693. [PubMed: 27654921]
75. Rose R, Constantinides B, Tapinos A, Robertson DL, Prosperi M. Virus Evol. 2016; 2:vew022. [PubMed: 29492275]
76. Allen LZ, Ishoey T, Novotny MA, McLean JS, Lasken RS, Williamson SJ. PLoS One. 2011; 6:e17722. CAS. [PubMed: 21436882]
77. Martínez JM, Swan BK, Wilson WH. ISME J. 2014; 8:1079–1088. [PubMed: 24304671]
78. Lance ST, Sukovich DJ, Stedman KM, Abate AR. Virol. J. 2016; 13:1–9. [PubMed: 26728778]
79. Cantalupo PG, Calgua B, Zhao G, Hundesa A, Wier AD, Katz JP, Grabe M, Hendrix RW, Girones R, Wang D, Pipas JM. mBio. 2011; 2:e00180. [PubMed: 21972239]
80. Rohwer F, Thurber RV. Nature. 2009; 459:207–212. [PubMed: 19444207]
81. Pekin D, Skhiri Y, Baret J-C, Le Corre D, Mazutis L, Ben Salem C, Millot F, El Harrak A, Hutchison JB, Larson JW, Link DR, Laurent-Puig P, Griffiths AD, Taly V. Lab Chip. 2011; 11:2156–2166. RSC. [PubMed: 21594292]
82. Hsieh AT-H, Pan PJ-H, Lee AP. Microfluid. Nanofluid. 2009; 6:391.
83. Srisa-Art M, deMello AJ, Edel JB. Anal. Chem. 2007; 79:6682–6689. [PubMed: 17676925]
84. Rane TD, Zec HC, Puleo C, Lee AP, Wang T-H. Lab Chip. 2012; 12:3341–3347. [PubMed: 22842841]
85. Zanoli LM, Licciardello M, D'Agata R, Lantano C, Calabretta A, Corradini R, Marchelli R, Spoto G. Anal. Bioanal. Chem. 2012; 405:615–624. [PubMed: 22212864]
86. Rane TD, Chen L, Zec HC, Wang T-H. Lab Chip. 2015; 15:776–782. [PubMed: 25431886]
87. Li Z, Liu Y, Wei Q, Liu Y, Liu W, Zhang X, Yu Y. PLoS One. 2016; 11:e0153359. [PubMed: 27074005]
88. Bernath K, Hai M, Mastrobattista E, Griffiths AD, Magdassi S, Tawfik DS. Anal. Biochem. 2004; 325:151–157. [PubMed: 14715296]
89. Mastrobattista E, Taly V, Chanudet E, Treacy P, Kelly BT, Griffiths AD. Chem. Biol. 2005; 12:1291–1300. [PubMed: 16356846]
90. Agresti JJ, Antipov E, Abate AR, Ahn K, Rowat AC, Baret J-C, Marquez M, Klibanov AM, Griffiths AD, Weitz DA. Proc. Natl. Acad. Sci. U. S. A. 2010; 107:4004–4009. [PubMed: 20142500]
91. Mazutis L, Gilbert J, Ung WL, Weitz DA, Griffiths AD, Heyman JA. Nat. Protoc. 2013; 8:870–891. [PubMed: 23558786]
92. Gielen F, Hours R, Emond S, Fischlechner M, Schell U, Hollfelder F. Proc. Natl. Acad. Sci. U. S. A. 2016; 113:E7383–E7389. [PubMed: 27821774]
93. Cao Z, Chen F, Bao N, He H, Xu P, Jana S, Jung S, Lian H, Lu C. Lab Chip. 2013; 13:171–178. RSC. [PubMed: 23160342]
94. Franke T, Braunmüller S, Schmid L, Wixforth A, Weitz DA. Lab Chip. 2010; 10:789–794. RSC. [PubMed: 20221569]
95. Johansson L, Nikolajeff F, Johansson S, Thorslund S. Anal. Chem. 2009; 81:5188–5196. [PubMed: 19492800]
96. Shemesh J, Bransky A, Khoury M, Levenberg S. Biomed. Microdevices. 2010; 12:907–914. [PubMed: 20559875]
97. Baroud CN, Delville J-P, Gallaire F, Wunenburger R. Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys. 2007; 75:046302.
98. Ahn K, Kerbage C, Hunt TP, Westervelt RM, Link DR, Weitz DA. Appl. Phys. Lett. 2006; 88:024104.
99. Sciambi A, Abate AR. Lab Chip. 2014; 14:2605–2609. [PubMed: 24671446]
100. Sciambi A, Abate AR. Lab Chip. 2014; 15:47–51.
101. Yan J, Bauer W-A, Fischlechner M, Hollfelder F, Kaminski C, Huck W. Micromachines. 2013; 4:402–413.

102. Zinchenko A, Devenish SRA, Kintses B, Colin P-Y, Fischlechner M, Hollfelder F. Anal. Chem. 2014; 86:2526–2533. [PubMed: 24517505]

103. Gawad C, Koh W, Quake SR. Nat. Rev. Genet. 2016; 17:175–188. [PubMed: 26806412]

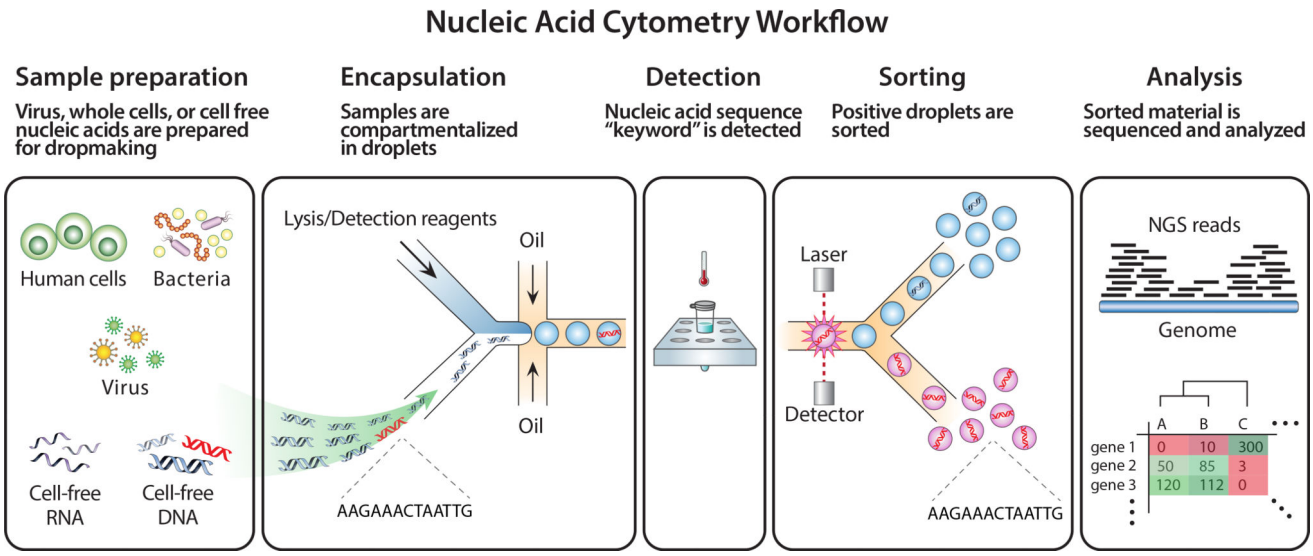

Figure 1.

Nucleic acid cytometry is a general method for isolating nucleic acids based on the presence of a keyword sequence. The workflow encapsulates material, detects specific DNA or RNA sequences, and sorts droplets containing those sequences. Sorted droplets are amenable to numerous downstream analysis techniques.

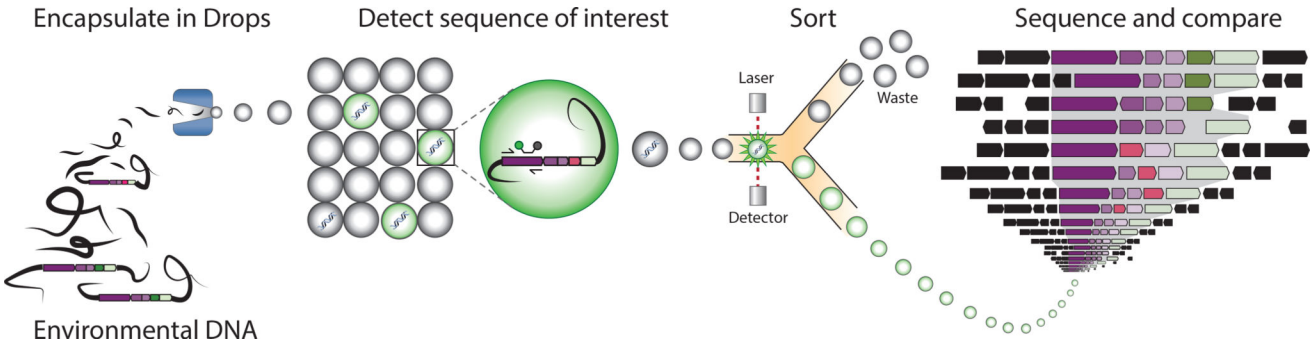

Figure 2.

Targeted comparative genomics. An example workflow: cell-free nucleic acid cytometry isolates bacterial operons based on a conserved gene sequence. The enriched material is sequenced, generating reads targeted to a specific region of interest. Bacterial operons are denoted as colored arrows, with regions of synteny highlighted in grey.

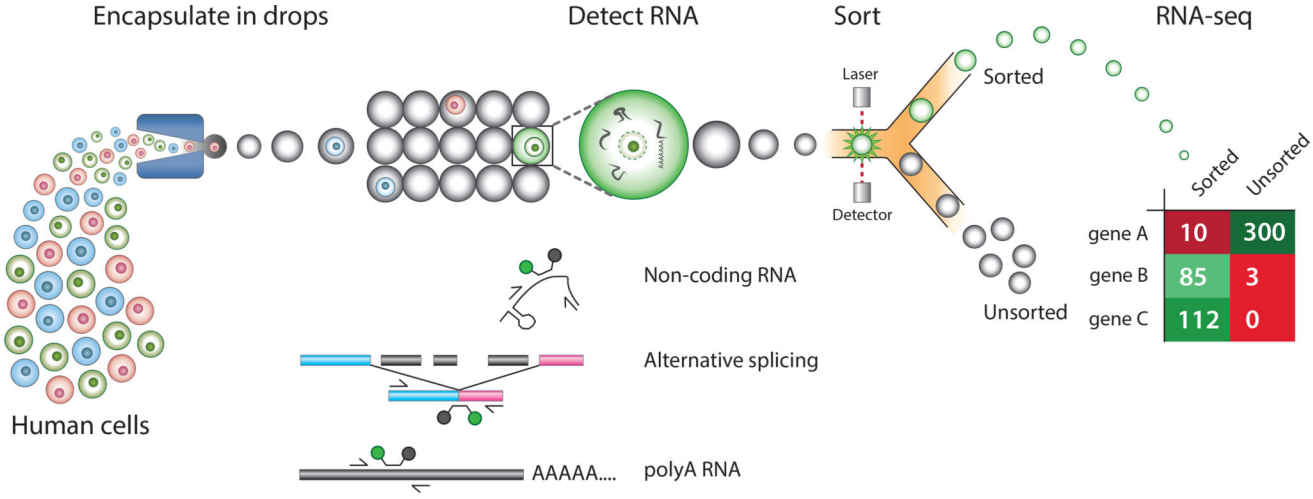

Figure 3.

Sorting and sequencing cells based on specific RNA expression. Whole cell nucleic acid cytometry encapsulates single cells in droplets and detects RNA using reverse transcription PCR. Sorting and RNA-seq on positive drops is used to understand the transcriptional landscape of cells with disease-specific transcription, alternative splicing, or non-coding RNA.

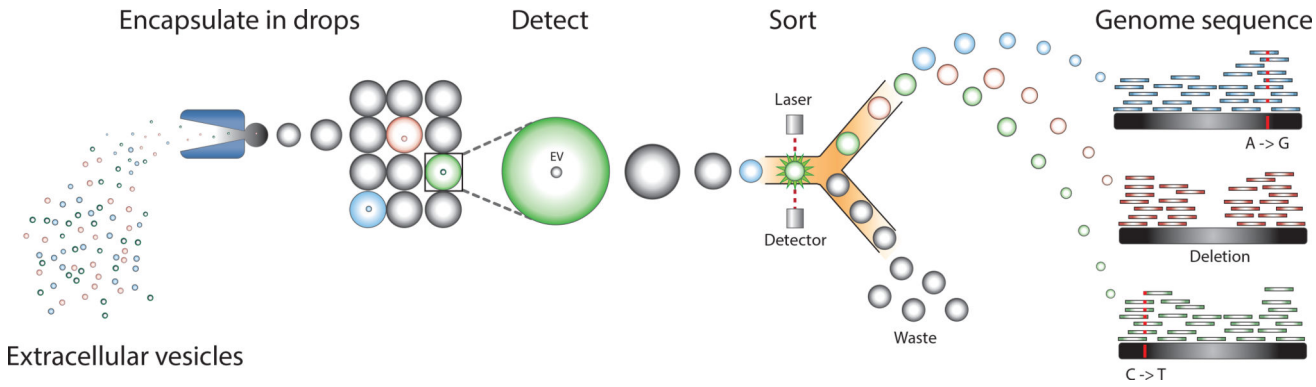

Figure 4.

Sorting and sequencing of extracellular vesicles. EVs are sorted based on the presence of miRNA and genome sequenced, revealing the genotype of their cellular origin.

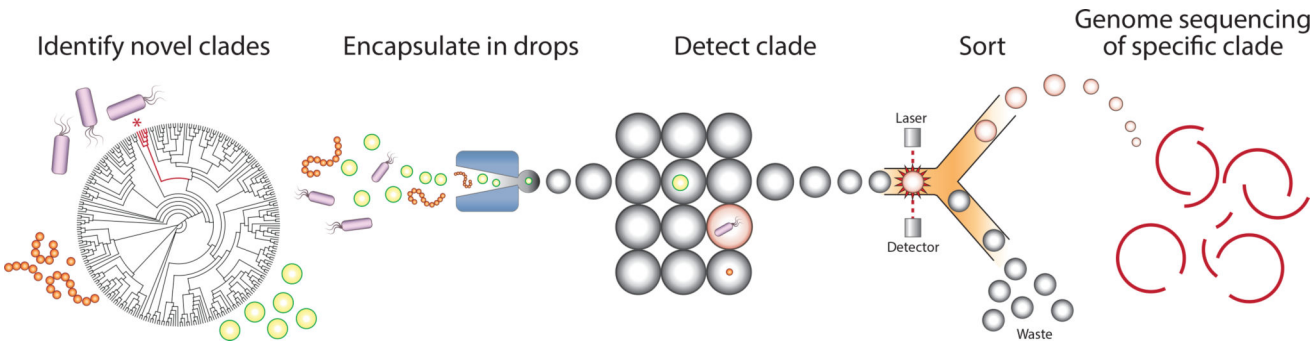

Figure 5.

Targeted sequencing of microbial genomes. Novel microbial clades are chosen for enrichment based on phylogenetic affiliation. Sorting and sequencing is used to understand the genomics of these uncultivable clades.

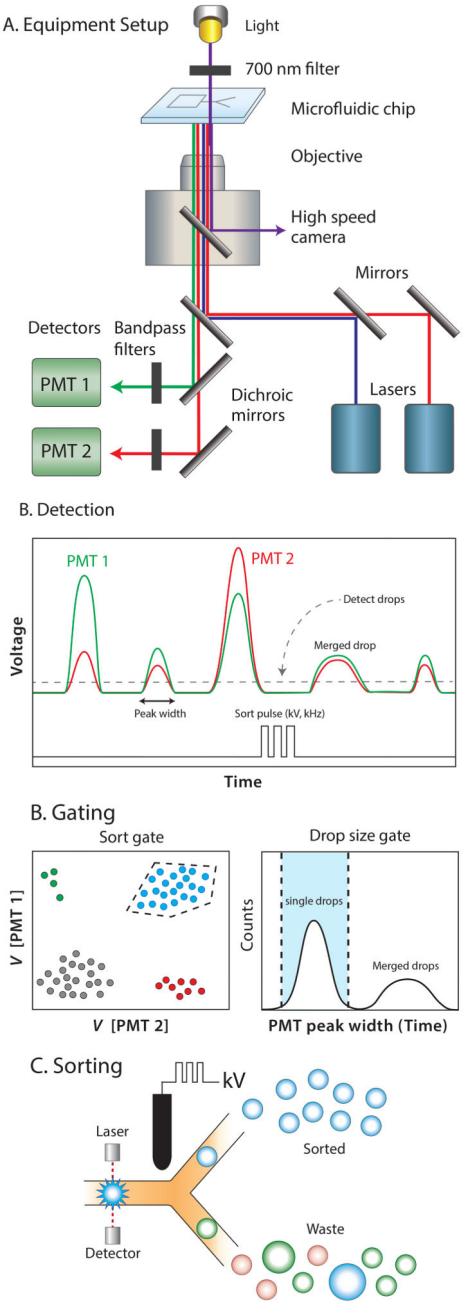

Figure 6.

Microfluidic droplet sorting allows for the isolation drops. A) Droplet sorters use an inverted microscope to align lasers onto a microfluidic channel. B) Drops containing fluorophores are excited as they pass through the laser line. Fluorescent signals are collected in real time using photomultiplier tubes. C) Gating of positive drops is based on the intensity, shape, and width of the detected signal. Large coalesced drops produce wide signals and can be discarded. D) Sorting is achieved using a high voltage AC signal that generates a dialectrophoretic force on water droplets.

Table 1

Droplet workflows that demonstrate components of nucleic acid cytometry

Sample type	Objective	Cell lysis	Detection	Sorting method	Analysis	Ref.
E. coli with T4 or ΦX174	Viral host enrichment	Thermocycling	TaqMan PCR	Droplet sorting	qPCR of sorted drops	Lim, 2017 (ref. 10)
Lambda DNA, ΦX174 DNA	Enrichment of rare molecules	None (purified DNA)	PCR multiplexed TaqMan or intercalating dye	FACS of double emulsions	qPCR of sorted double emulsions	Sukovich, 2017 (ref. 11)
Bacteriophage T4 and ΦX174	Sequence specific enrichment of virus	Thermocycling	PCR with intercalating dye	FACS of double emulsions	qPCR of sorted double emulsions	Lance, 2016 (ref. 12)
Human PC3 prostate cancer cells Raji B-lymphocyte cells	Sorting of cells based on specific sequences followed by transcriptome analysis	Not reported	TaqMan RT-PCR	Droplet sorting	Single cell qPCR, RNA-seq on pooled sort	Pellegrino, 2016 (ref. 13)
SV40 viral particles	Whole genome sequencing of rare virus	Thermocycling	PCR with intercalating dye	Droplet sorting	Whole genome amplification followed by Illumina sequencing	Han, 2015 (ref. 14)
E. coli ΔtolA, *E. coli* LpoA K168A	Rare microbe sequencing without culturing	Thermocycling	TaqMan PCR	Droplet sorting	Sanger sequencing of lpoA mutant	Lim, 2015 (ref. 15)
DNA from lymphoblast cells	Enrichment and sequencing of rare molecules	None (purified DNA)	TaqMan PCR	Droplet sorting	Illumina sequencing	Eastburn, 2015 (ref. 16)
Viral genomic RNA	Sequence viral recombinants	None (purified RNA)	TaqMan RT-PCR	Droplet sorting	Sanger sequencing of amplicons	Tao, 2015 (ref. 17)
Human DU145 prostate cancer cells Raji B-lymphocyte cells	High-throughput detection and sorting of cells based on specific sequences	Tween-20 proteinase K	TaqMan PCR and RT-PCR	Droplet sorting	SNP analysis, Illumina sequencing, qRT-PCR	Eastburn, 2014 (ref. 18)
BC-1 (CRL-2230) B-cell with Epstein–Barr virus	SNP and CNV detection	Cells fixed	Rolling circle amplification of genomic target	None	Analysis of fluorescent droplet images	Konry, 2013 (ref. 19)
Human PC3 prostate cancer, Raji B-lymphocyte RNA	High-throughput multiplex RT-PCR	Purified RNA	TaqMan RT-PCR	None	Microscopy	Eastburn, 2013 (ref. 20)
Human PC3 prostate cancer, Raji B-lymphocyte cells	High-throughput multiplex RT-PCR on single cells	Tween-20 proteinase K	TaqMan RT-PCR	None	Droplet cytometry	Eastburn, 2013 (ref. 21)
S. typhimurium SL1344, *E. coli* K12, environmental samples	Single cell analysis of environmental samples (qPCR, WGA)	Thermocycling	PCR with intercalating dye	Well recovery	qPCR, WGA, Illumina sequencing	Leung, 2012 (ref. 22)
Human cells (Kato III and MDA-MB-231)	Single cell digital RT-PCR	1% Triton X-100	Agarose-capture of PCR amplicon and intercalating dye	None	Flow cytometry	Zhang, 2012 (ref. 23)

Sample type	Objective	Cell lysis	Detection	Sorting method	Analysis	Ref.
E. coli O157:H7, *E. coli* K12 cells	Sequence specific detection of rare pathogens	Thermocycling	Agarose-capture of multiplex PCR amplicons	None	Flow cytometry	Zhu, 2012 (ref. 24)
Lymphoblast cells	High throughput sequencing of multiple targets in a single-cell	SDS, proteinase K in agarose hydrogels	Bead capture of fluorescent PCR amplicon	FACS of beads	Sanger sequenced b-actin gene and chromosomal translocation t(14;18)	Novak, 2011 (ref. 25)
E. coli	Screen for sequence motifs in library	Heat 96 °C, 10 minutes	PCR with intercalating dye	Complex object parametric analyzer and sorter	Sanger sequencing of plasmid insert	Walser, 2009 (ref. 26)
Human lymphocytes, *E. coli* K12	Production of Sanger and pyrosequencing compatible amplicons from single cells	Heat 95 °C, 10 minutes	Bead capture of fluorescent PCR amplicon	FACS of beads	Sanger sequencing	Kumaresan, 2008 (ref. 27)