ACS APPLIED MATERIALS & INTERFACES

www.acsami.org

Research Article

Scalable Nucleic Acid Storage and Retrieval Using Barcoded Microcapsules

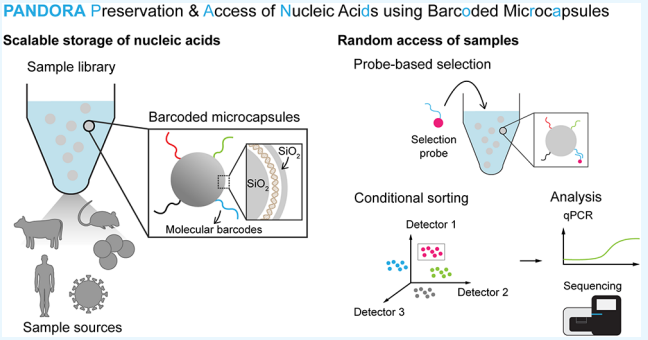
James L. Banal* and Mark Bathe*

Cite This: *ACS Appl. Mater. Interfaces* 2021, 13, 49729−49736

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Rapid advances in nucleic acid sequencing and synthesis technologies have spurred a major need to collect, store, and sequence the DNA and RNA from viral, bacterial, and mammalian sources and organisms. However, current approaches to storing nucleic acids rely on a low-temperature environment and require robotics for access, posing challenges for scalable and low-cost nucleic acid storage. Here, we present an alternative method for storing nucleic acids, termed Preservation and Access of Nucleic aciDs using barcOded micRocApsules (PANDORA). Nucleic acids spanning kilobases to gigabases and from different sources, including animals, bacteria, and viruses, are encapsulated into silica microcapsules to protect them from environmental denaturants at room temperature. Molecular barcodes attached to each microcapsule enable sample pooling and subsequent identification and retrieval using fluorescence-activated sorting. We demonstrate quantitative storage and rapid access to targeted nucleic acids from a pool emulating standard retrieval operations implemented in conventional storage systems, including recovery of 100,000−200,000 samples and Boolean logic selection using four unique barcodes. Quantitative polymerase chain reaction and short-read sequencing of the retrieved samples validated the sorting experiments and the integrity of the released nucleic acids. Our proposed approach offers a scalable long-term, room-temperature storage and retrieval of nucleic acids with high sample fidelity.



KEYWORDS: *Genetics, silica, assays, biopolymers, barcodes, sorting*

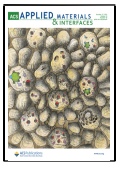
INTRODUCTION

Large-scale, low-cost, high-fidelity storage of nucleic acid samples is crucial to basic, translational, and clinical applications and research,[1,2] including forensics,[3] genetics,[4] and genomics.[5] Nucleic acid storage requires robust procedures to maintain sample quality, integrity, and function. Current storage temperature requirements for nucleic acids range from 4 to −196 °C,[2,4,6] to maintain negligible degradation. However, maintaining such a low temperature for extended periods requires significant energy cost. In addition, large-scale cryogenic storage of nucleic acid materials requires extensive robotics for access, stringent cold-chain management logistics,[1,2,7] and redundant copies of samples stored in mirror storage facilities to mitigate the risk of sample loss.[2] Finally, cold storage of nucleic acids in remote or low-resource areas will involve costly measures and complex cold-chain logistics to maintain the integrity and quality of the isolated sample during transport.[7]

Recently, alternative room temperature approaches have been developed to address the challenges of storing nucleic acids at low temperatures. A common method involves dehydration of nucleic acids in the presence of stabilizing matrices, such as trehalose,[2] a non-reducing disaccharide that protect cells under oxidative stress[8] and enables organisms to withstand dehydration and extreme temperatures during cryptobiosis,[9,10] and commercially available matrices, such as DNAstable and RNAstable.[11] More sophisticated storage strategies use hermetically sealed metal capsules, such as DNAshell and RNAshell, which provide an anhydrous and anoxic atmosphere for dehydrated nucleic acids.[7,12]

Although effective at protecting nucleic acids at room temperature, the preceding strategies are limited by the quantity of nucleic acids that can be stored, termed DNA loading[13] or generally nucleic acid loading to encompass both DNA and RNA, mainly because the storage vessels are much larger than the actual nucleic acid sample itself. As an alternative to conventional storage, silica encapsulation of DNA[14−16] and RNA[17] was developed to emulate the protective characteristics of fossils. Silicification provides an impervious shell around the nucleic acids protecting the encapsulated material from reactive

Received: August 6, 2021
Accepted: October 1, 2021
Published: October 15, 2021



Downloaded via MASSACHUSETTS INST OF TECHNOLOGY on May 2, 2022 at 18:10:32 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

© 2021 American Chemical Society

https://doi.org/10.1021/acsami.1c14985
ACS Appl. Mater. Interfaces 2021, 13, 49729−49736

oxygen species, nucleases, and high temperatures approaching 200 °C.[14,15] Furthermore, accelerated weathering experiments on nucleic acids encapsulated in silica suggest that encapsulation permits long-term storage of nucleic acids at room temperature exceeding 100 years, comparable to other existing storage systems.[14,16,17] Silica encapsulation of nucleic acids can be chemically reversed by adding a buffered oxide etch,[15] which dissolves the silica capsules but does not affect the integrity of nucleic acids, including messenger RNA.[17] Although effective at physically and chemically protecting nucleic acids, current demonstrations of nucleic acid encapsulation using silica still require that the samples are separated physically even after encapsulation to avoid sample mixing, which does not take full advantage of the high nucleic acid loading silica capsules can offer. Strategies to uniquely identify each silica-encapsulated sample from a pool are critical to allow for the pooling of a diverse set of samples to maximize the nucleic acid loading capacity silica capsules can offer while still allowing for high-fidelity retrieval of specific or subsets of samples from a pool.

We recently expanded on the silica encapsulation approach by introducing molecular barcodes on encapsulated nucleic acids.[18] Molecular barcodes on individual silica capsules provide identification of each sample, enabling sample pooling and allowing for selection and retrieval of samples from a pool using particle sorting approaches, such as fluorescence-activated sorting (FAS). In addition, we also observed that the functional characteristics of encapsulated recombinant plasmids encoding digital data are retained by direct transformation of the recovered plasmid into bacteria after an end-to-end storage, access, and retrieval cycle, which suggests that our approach can preserve the fidelity of nucleic acids while enabling dense storage of a heterogeneous set of samples.[18] Here, we sought to develop a general-purpose storage and retrieval system for nucleic acids, which we term Preservation and Access of Nucleic aciDs using barcOded micRocApsules (PANDORA). Specifically, we investigated if we can apply PANDORA to any nucleic acid, including animal, bacterial, and viral genomes, and perform similar retrieval operations standard for conventional storage systems using FAS (Figure 1). We additionally characterized the efficiency of encapsulation, the fidelity of nucleic acid sequence after de-encapsulation, and the cost for each sample type toward implementing a scalable, low-cost solution for nucleic acid sample preservation.

RESULTS

To demonstrate PANDORA as a general-use storage system for nucleic acid systems, we encapsulated five nucleic acid samples from various sources into monodisperse 6 μm non-fluorescent, silica capsules (Supporting Information Figure S1) using a sol–gel approach that we and others[14−18] have developed (Supporting Information Section S2). The five nucleic acid samples represent different lengths and sources of nucleic acids that are commercially available or easily accessible: 3 gigabase bovine genome, 2.5 gigabase murine genome, 2.8 megabase *S. aureus* genome, 30 kilobase SARS-CoV-2 genome, and total RNA from HEK293 cells. Using 1 mg of ammonium-functionalized silica particles for each nucleic acid sample, we encapsulated 1 μg of bovine and murine genomes, 2 μg of *S. aureus* genome, 10 μg of total RNA isolated from HEK293 cells, and 10,000,000 copies of SARS-CoV-2 genome in separate tubes (Supporting Information Section S2). Electrostatic interaction between the phosphate backbone of the nucleic acid strands and ammonium-functionalized silica core particles

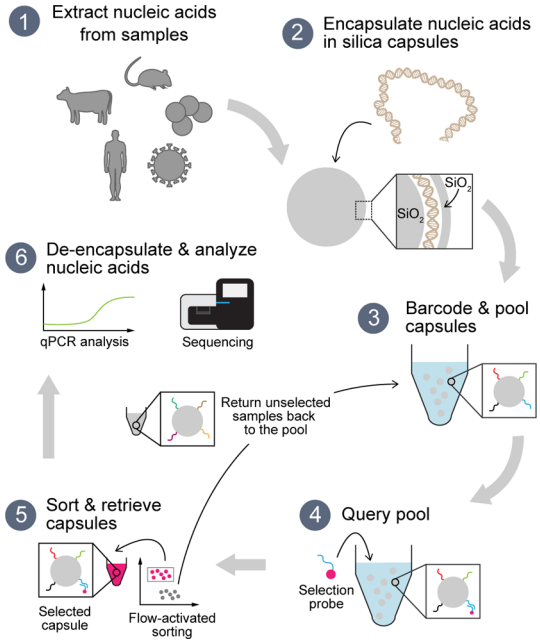


Figure 1. Overview of PANDORA workflow. Nucleic acids from any source are extracted and encapsulated in silica capsules. Encapsulated nucleic acids are barcoded and pooled. Samples of interest are selected by the addition of selection probes that hybridize to barcodes on the target sample capsules. Selected samples are then physically captured using FAS. Retrieved capsules are de-encapsulated to recover the stored nucleic acid which can be used for nucleic acid analyzes using quantitative polymerase chain reaction (qPCR) or sequencing.

encouraged the adsorption of the DNA and RNA samples on the surface of silica particles. Next, the addition and incubation of tetraethoxysilane and *N*-(3-(trimethoxysilyl)propyl)-*N*,*N*,*N*-trimethylammonium chloride facilitated the formation of an encapsulation shell around the adsorbed nucleic acid samples. After 4 days of encapsulation, an amino-modified silane precursor was added to modify the surface of the silica encapsulated nucleic acid samples, introducing a chemical handle to subsequently attach bifunctional linkers that would allow for the attachment of amino-modified single-stranded DNA (ssDNA). Each sample was barcoded using four amino-modified 25-nucleotide ssDNA selected from a library of 240,000 orthogonal primers[19] (Supporting Information Table S1), which function as sample labels. After barcoding, all samples were pooled together, thus forming the sample library. Upon attachment of DNA barcodes, the hydrodynamic diameter of the barcoded capsules increased relative to the silica core-only sample (Supporting Information Figure S2), consistent with the expected result if the ssDNA barcodes and poly(ethylene glycol) linker were successfully conjugated on the surface of the silica.[20]

A simple way to label samples using PANDORA is to use different barcodes for each sample. With four labels per sample and using the entire 240,000 orthogonal library of 25-nucleotide strands, PANDORA can uniquely label $>10^{20}$ samples. Alternatively, we can also use the barcodes to represent different categories and classes, a strategy we exploited to build a scalable filesystem for DNA data storage.[18] In this work, we used the barcodes to describe different ways samples can be labeled in conventional storage systems: general barcodes representing a class of samples, a barcode for unique identification, and a barcode representing temporal information (Supporting Information Table S1). For example, *Bos taurus*, *Mus musculus*, and

https://doi.org/10.1021/acsami.1c14985
ACS Appl. Mater. Interfaces 2021, 13, 49729−49736

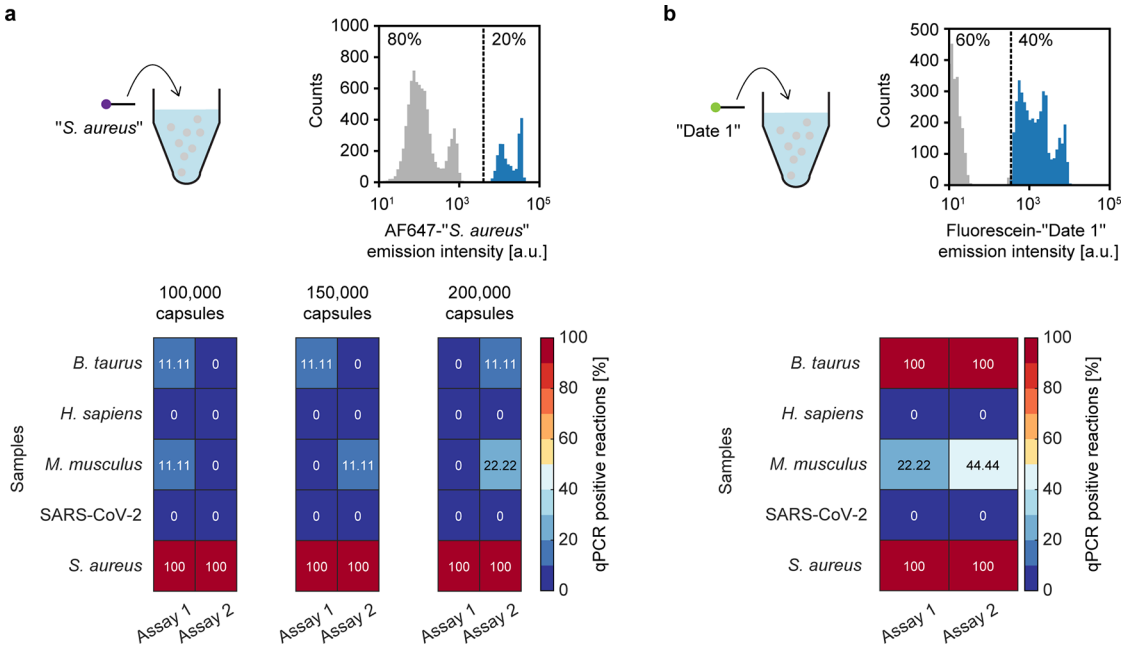


Figure 2. Single-barcode selection and retrieval. (a) Retrieval of discrete quantities of *S. aureus* capsules using "*S. aureus*" probe labeled with Alexa Fluor 647 (AF647) and using FAS (top, right). Retrieved samples were validated using qPCR panel assays (bottom). (b) Retrieval of *B. taurus* and *S. aureus* samples using "Date 1" probe labeled with fluorescein. FAS was used to sort sample capsules that have the "Date 1" barcode and subjected to qPCR analyses to validate the sorted capsules (bottom). Percentages denote the fraction of capsules that were sorted in each gate.

Homo sapiens are classified as eukaryotes, and each is labeled with unique identification labels, namely their species labels, and a date label. Analyses of the reaction post-encapsulation supernatants for each nucleic acid using quantitative polymerase reaction (qPCR) assay panels (Supporting Information Sections S5 and S6) did not show any amplification which suggests quantitative encapsulation efficiency of all nucleic acid samples (Supporting Information Section S7 and Figure S3).

With the barcodes attached to each sample and all samples pooled into one vessel, we attempted to perform sample selections and retrieval operations emulating typical processes performed in conventional storage systems. To select a sample from a pool, a 15-nucleotide dye-labeled probe strand complementary to the barcode of interest is added to the pool. The dye-labeled probe diffuses in a solution and hybridizes to the barcodes of interest that are displayed on the surface of the target capsule or a range of sample capsules. We then used FAS as an analytical method to assess the robustness of our selection through analyses of gated fractions and simultaneously retrieve the selected sample. In FAS, capsules in the solution are flowed through as droplets generated at a frequency of tens of kilohertz, passing through a laser beam. At a sufficiently low concentration of capsules in the solution, each droplet would only contain one capsule. If the droplet interrogated at the laser beam had a capsule with a fluorescent dye hybridized, the FAS device would record the detected fluorescence intensity at the appropriate detection channel. Sorting gates, which are user-specified ranges of particle scattering and fluorescence intensities on a specific laser−filter pair, are then drawn to sort the desired droplets that contain the target sample (Supporting Information Section S4).

Using FAS, we performed sample selection using single barcodes to demonstrate retrieval of single or multiple samples. In one example, we retrieved 100,000, 150,000, and 200,000 capsules of *S. aureus* samples from a pool using the "*S. aureus*" probe and using FAS to retrieve the selected capsules (Figure 2).

We observed that the fluorescence intensities of the selected capsules are consistent with our previous work, suggesting that the number accessible of DNA barcodes, approximately 10^8 per capsules,[18] are retained. Sorted capsules were then de-encapsulated using buffered oxide etch. After desalting the de-encapsulated nucleic acids, the samples were then subjected to probe-based qPCR assay panels that were designed for each sample (Supporting Information Sections S5 and S6). qPCR assay panels reveal that most of the samples retrieved in all cases were *S. aureus* samples with small contamination of *B. taurus* and *M. musculus* samples (Figure 2, Supporting Information Figure S4), which could arise from imperfect sorting by the FAS equipment. From the sorted capsules, we estimated the quantities of retrieved *S. aureus* samples to be 40 ± 20 pg for 100,000 capsules, 270 ± 80 pg for 150,000 capsules, and 340 ± 90 pg for 200,000 capsules (mean ± standard deviation; $n = 3$), equivalent to 1.3 ± 0.8 fg (mean ± standard deviation; $n = 9$) of *S. aureus* genome per capsule (Supporting Information Figures S4 and S5) or 1.1 ± 0.6 ng cm^{-2} (mean ± standard deviation; $n = 9$) in nucleic acid loading. Given that we did not observe any remaining nucleic acid in the supernatant after encapsulation (Supporting Information Figure S3), we are still below the nucleic acid loading capacity of the silica capsules. Increasing the ratio of nucleic acid stored per mass quantity of silica capsules would improve our nucleic acid loading capacity and reach the state-of-the-art nucleic acid loading of >30 ng cm^{-2} for silica capsules.[15,21] Higher nucleic acid loadings beyond the state-of-the-art can be realized through layer-by-layer encapsulation of nucleic acids.[21]

To further characterize the retrieval efficiency of the query and sorting process, we compared the *S. aureus* genomes per capsule obtained from the capsule retrieval process (Supporting Information Section S8 and Figure S4) to the average *S. aureus* genomes per capsule measured for unsorted samples using qPCR (Supporting Information Section S9 and Figure S5). We

https://doi.org/10.1021/acsami.1c14985
ACS Appl. Mater. Interfaces 2021, 13, 49729−49736

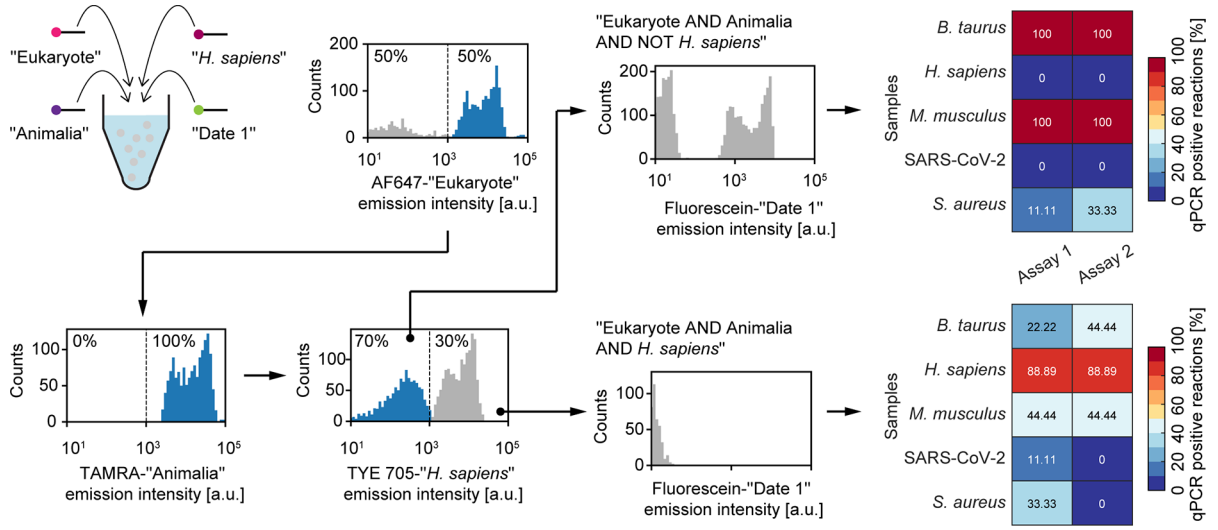


Figure 3. Multi-barcode selection and retrieval using Boolean logic. *B. taurus* and *M. musculus* sample capsules were retrieved from the pool using multi-barcode selection with FAS. "Eukaryote" probe labeled with AF647 and "Animalia" probe labeled with "TAMRA" selected *B. taurus*, *M. musculus*, and *H. sapiens* from the pool. The addition of the "*H. sapiens*" probe labeled with TYE 705 differentiated *H. sapiens* from *B. taurus* and *M. musculus* sample capsules, which corresponds to the Boolean logic "Eukaryote AND Animalia AND NOT *H. sapiens*". The addition of the "Date 1" probe internally validated the sorted populations. Because only *B. taurus* capsules should be positive for "Date 1", capsules that correspond to "Eukaryote AND Animalia AND NOT *H. sapiens*" would contain a subpopulation positive for "Date 1" while populations sorted in the "Eukaryote AND Animalia AND*H. sapiens*" selection query should not contain any subpopulations that are positive for "Date 1" barcodes. The sorted capsules were further validated with qPCR. Percentages denote the fraction of capsules that were sorted in each gate.

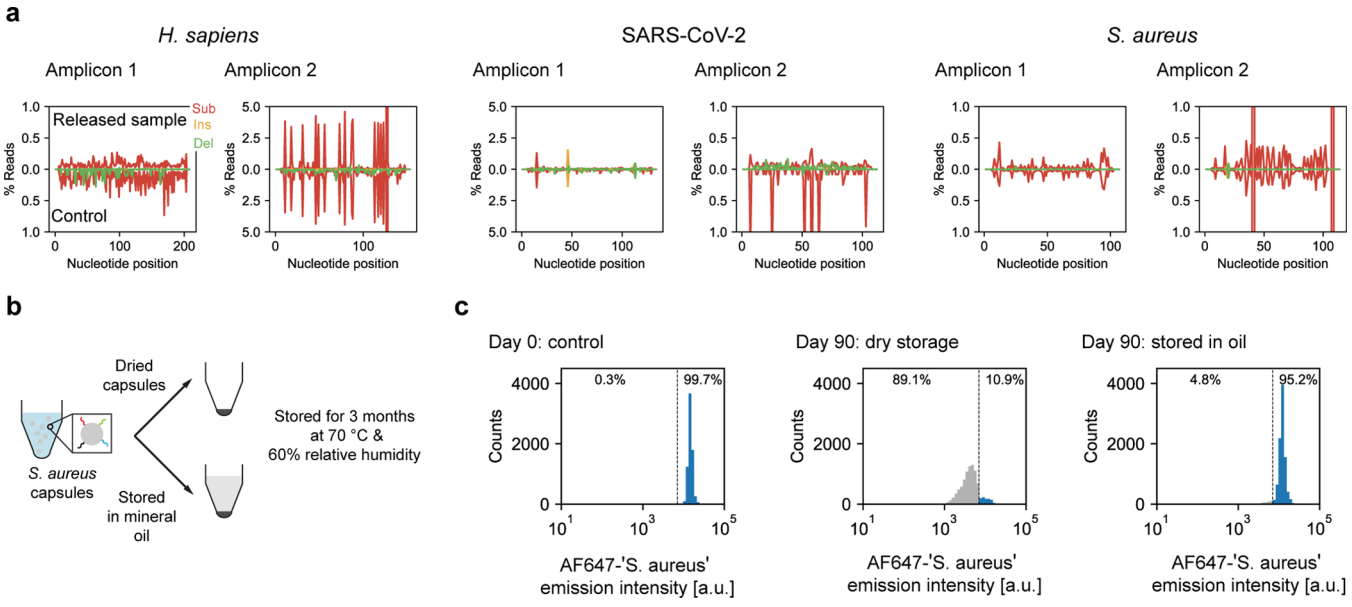


Figure 4. Error rates from sequencing and assessment of barcode stability. (a) Comparison of amplicon sequencing error rates between samples released from encapsulation (top) and unencapsulated controls (bottom). (b) Scheme for accelerated weather experiments to investigate storage approaches for PANDORA capsules. (c) We used FAS to quantify the extent of barcode degradation after 3-month storage at 70 °C and 60% relative humidity. We compared capsules that were stored as dry pellets to capsules that were stored in mineral oil. Sub, substitutions. Ins, insertions. Del, deletions. Percentages denote the fraction of capsules that belong in each gate.

were able to estimate a retrieval efficiency of 32%, which is consistent with the observed retrieval efficiencies of FAS for high-purity sorting with cells.[22] Depending on the sorting goal, the sorting precision strategy used can contribute to sample losses.[22] In the preceding example (Figure 2), a high-purity sorting strategy was prioritized over quantitative sorting of particles, resulting in a reduced recovery efficiency of capsules. The overall *S. aureus* genome retrieval yield, defined as fraction of *S. aureus* genome per capsule retrieved from sorting relative to the expected *S. aureus* genome per capsule assuming quantitative encapsulation, which is 2 μg mg^{-1}, is 0.30%.

In another example of single barcode retrieval, we attempted to retrieve a group of samples with a similar label instead of a single sample. In conventional systems, similar retrieval operations are performed using common labels, such as spatial or temporal information. To demonstrate a similar operation in PANDORA, we selected samples with the label "Date 1" using FAS and sorted 200,000 capsules in total (Figure 2). Validation

https://doi.org/10.1021/acsami.1c14985
ACS Appl. Mater. Interfaces 2021, 13, 49729−49736

using qPCR assay panels showed that the retrieved capsules contained the target *B. taurus* and *S. aureus* samples which contained the "Date 1" label (Figure 2, Supporting Information Figure S6).

Besides single barcode selection, sample selection involving multiple barcodes and Boolean logic operations can refine search queries. For PANDORA, we can leverage different dyes attached to the various query probes used for selection and exploit the ability of state-of-the-art FAS devices to monitor more than ten channels simultaneously, which thereby allows for database-like querying and selection operations akin to conventional storage systems for nucleic acids using Boolean logic.[18] To demonstrate multiple gating conditions using different probes, we added "Eukaryote," "Animalia," "*H. sapiens*," and "Date 1" probes to the sample pool wherein each probe is labeled with a different fluorophore: Alexa Fluor 647, tetramethylrhodamine, TYE 705, and fluorescein, respectively (Figure 3). Unlike our previous work, where the silica core particles already contained fluorescein as an internal fluorescent marker,[18] the silica core particles in this work are non-fluorescent, enabling the use of fluorescein as an additional dye channel for sample selection using Boolean logical search. The "Eukaryote" and "Animalia" labels selected *B. taurus*, *M. musculus*, and *H. sapiens* samples, while the addition of the "*H. sapiens*" probe differentiated the *H. sapiens* samples from the *B. taurus* and *M. musculus* samples. The probe "Date 1" was only used as a fiducial marker to internally cross-validate the populations selected using the "Eukaryote", "Animalia", and "*H. sapiens*" probes and gating conditions. Sample populations that only satisfy the query "Eukaryote AND Animalia AND NOT *H. sapiens*" should be positive for the "Date 1" probes due to the presence of *B. taurus* in the selected sample subset. In contrast, the "Eukaryote AND Animalia AND *H. sapiens*" query, which should give only *H. sapiens* samples that do not contain the "Date 1" barcode, should not show any subpopulations that contain capsules that are positive for the "Date 1" subselection (Figure 3). Subjecting the de-encapsulated sorted population to the qPCR assay panel validated the retrieved samples (Figure 3, Supporting Information Figures S7 and S8).

To further test the application of PANDORA for downstream nucleic acid analyses, including monitoring the cellular transcriptome or single nucleotide mutations at the genome level, we performed short-read sequencing of amplicons generated from de-encapsulated *S. aureus*, *H. sapiens*, and SARS-CoV-2 samples. *S. aureus* amplicons were generated by PCR amplification of different regions of *S. aureus* genome. *H. sapiens* and SARS-CoV-2 amplicons were generated by first generating complementary DNA strands using reverse transcription followed by PCR amplification. Analyses of sequencing data showed that de-encapsulated nucleic acids have similar error rates (insertion, deletions, and substitutions) compared to unencapsulated nucleic acid controls (Figure 4a, Supporting Information Section S13), which suggests that the de-encapsulation step did not cause degradation of the stored nucleic acid. Furthermore, the error rates observed for the de-encapsulated samples are comparable to reported error rates for synthetic DNA strands that have been exposed to elevated temperatures using various storage strategies. However, as noted previously,[23] it can be argued that the relatively low error rates could be caused by the use of an inevitable PCR amplification step to amplify strands to reach the minimum required quantities required for sequencing (500 ng) and to attach sequencing adapters, which could bias amplification toward more intact strands.[23]

Besides, the internal nucleic acids stored in PANDORA capsules, the integrity of single-stranded barcodes affixed on the surface of each capsule must be retained for retrieval. Unlike the internal nucleic acids, the barcodes are exposed to the environment and require secondary protection to eliminate the exposure of the ssDNA barcodes to denaturants, including water, which can cause degradation via depurination when nucleic acids are exposed over long periods of time.[24] Storage in hydrophobic media,[25] such as silicone oil which has high thermal and chemical stability over hydrocarbon oils, can be used as a strategy to mitigate barcode degradation. We investigated the protective ability of silicone oil by measuring using FAS the number of barcodes that remain on the silica surface after storage in an incubator with a set temperature of 70 °C and 60% relative humidity for three months (Supporting Information Section S14). As a control, we also stored dried pellets of capsules in the same incubator for comparison (Figure 4b). After 3-month incubation, the capsules were reconstituted in buffer (40 mM Tris, 20 mM acetate, 2 mM EDTA, 1.0% Tween-20, 1.0% sodium dodecyl sulfate, and 500 mM NaCl), hybridized with a fluorescent probe, and sorted through a FAS device for barcode degradation analysis. Samples that were stored dry showed a significant fraction of capsules with lower fluorescence intensities (Figure 4c, middle) than capsules that were not exposed to elevated temperatures and humidity for three months (Figure 4c, left). The decrease in capsule population with high intensities is attributed to decreased number of barcodes to which the fluorescent probes can hybridize. Given that we anticipate 10^8 barcodes per capsule,[18] we expect that not all barcodes are degraded, and all capsules will still have some remaining number of barcodes, albeit with reduced quantities. In contrast, capsules stored in silicone oil showed that 95% of the capsules retained similar intensities compared to the control capsules, suggesting that immersing the sample library in silicone oil can protect the single-stranded DNA barcodes (Figure 4c, right). Besides immersing the sample library in oil, conventional room-temperature storage strategies, such as protective matrices and commercially available hermetically sealed capsules, can also be applied to PANDORA with the added benefit of increasing the nucleic acid loading of each conventional room-temperature storage strategy because each vessel now contains a more diverse library of nucleic acids instead of simply having a single sample of the same type.

DISCUSSION

We present here PANDORA as an alternative room-temperature storage platform for nucleic acids. PANDORA enables the storage of nucleic acids in microcapsules instead of millimeter-scale tubes or vessels that improve nucleic acid loading for storage compared to conventional storage systems. Silica encapsulation offers protection of nucleic acids against physical, chemical, and biochemical denaturants at room temperature. Molecular barcoding enables individual identification of samples and sample pooling, which significantly reduces the footprint for storage of nucleic acids. Selection and retrieval of samples using molecular barcodes provide solution-phase strategies to retrieve samples from a pool using single labels or database-like search operations involving multiple labels and Boolean logic. Importantly, the labeling scheme we have developed in this work can be further expanded to select a range of samples based on common label types, for example, time ranges, or wildcard

https://doi.org/10.1021/acsami.1c14985
ACS Appl. Mater. Interfaces 2021, 13, 49729−49736

searching—types of queries also used to select samples in conventional storage systems—using Boolean logic operations that we have demonstrated here.

The storage and retrieval workflow we have demonstrated for PANDORA uses existing chemistries and equipment commonly found in academic and commercial laboratories, including biobanks, and can be compatible with an automated workflow involving liquid handling devices and large-scale shaking incubators. Besides compatibility with existing workflows, cost is an important consideration. We estimate that the cost of storing nucleic acids using PANDORA would be $3.40 per sample and an additional $1.00 for every instance that the sample library is queried. The storage cost is only incurred at the outset instead of an ongoing recurring fee for electricity when using freezers, which can accumulate to $2−$22 over 10 years depending on the site operation.[2] Although cost-competitive, the current use of silica encapsulation for PANDORA introduces significant storage latency, limiting the application of PANDORA to the archival storage of nucleic acids. Archival nucleic acid samples would encompass genomes of endangered species[26,27] or nucleic acid samples from a large cohort of individuals or patients for retrospective analysis,[28] which might include genomes or RNA transcriptomes.

The workflow from storage to retrieval for PANDORA is straightforward, although it requires buffered hydrofluoric acid and a desalting step to remove hexafluorosilicate salts that formed during de-encapsulation. For routine PCR applications as shown in Figure 2, it only takes less than 5 min to retrieve 100 pg quantities of nucleic acids even when the purity of sorted samples is prioritized over speed, making PANDORA already suitable for storing nucleic acids that only require downstream PCR amplification for analyses, for example, environmental surveillance of emerging pathogens[29−33] through retrospective analyses. For sequencing applications involving megabase or gigabase genomes, 1 μg quantities of genomes are minimally required for sequencing, which is equivalent to pulling out at least 1 billion capsules from the pool assuming 1 fg of genome per capsule. In this specific use case, magnetic-based sorting would be an ideal approach rather than FAS to select large quantities of capsules from a pool.

Besides single nucleic acid sample retrieval, PANDORA could also be used to store and retrieve a highly heterogeneous collection of nucleic acids derived from single cells. The preponderance of single-cell sequencing workflows using droplets has enabled the dissection of phenotypic heterogeneity of individual cells that are typically masked by ensemble-averaged measurements.[34−38] PANDORA could be used together with single-cell genomic workflows to partition and store nucleic acids from different tissues or organisms and across a gamut of temporal progression, ranging from short, for example, cell cycles, to long, for example, cell or whole organism development and timescales. Because latency associated with sequencing would be the rate-limiting step in single-cell genomics analysis, long-term storage of segregated nucleic acids would be required. PANDORA offers the ability to permanently encapsulate extracted nucleic acids, unlike water-in-oil droplets, while retaining the separation of individual single-cell nucleic acid partitions, which is challenging to implement using conventional storage technologies. Besides storage, retrieval of specific subsets of nucleic acid partitioned by PANDORA enables sample triaging for nucleic acid analysis through Boolean logical selection of sample subsets, for example, selecting tissue subtypes or cell cycle lineages using molecular barcodes, while also allowing for repeated access to libraries of extracted nucleic acid samples to perform sequencing or gene expression analyses as the need arises. Notwithstanding this additional important potential application area, sophisticated automated liquid handling would be required for its implementation, which may only be warranted when storage of many tens or hundreds of millions of individual nucleic acid samples need to be stored.

CONCLUSIONS AND OUTLOOK

We have demonstrated that PANDORA can be used as a general-purpose nucleic acid storage and platform for any nucleic acids. PANDORA reduces the storage footprint of nucleic acids from traditional storage vessels, such as PCR tubes, to micron-scale silica particles. Sample pooling with PANDORA further reduces the storage by also eliminating the need to store individual samples separately. Sample access using PANDORA is achieved using molecular-scale physical extraction methods, such as FAS, enabling sample selection using Boolean logical search queries. Although powerful, further optimization will be required to transition PANDORA as a general-use, room-temperature platform for the storage and access of nucleic acids. Advances in microfluidic encapsulation to seamlessly perform the encapsulation and barcoding steps in droplets similar to strategies used for single-cell sequencing[35,39] would enable the implementation of PANDORA outside of a laboratory setting. Replacement of silica with alternative encapsulants, such as synthetic polymers, will accelerate the encapsulation process from days to minutes, reducing the storage latency and enabling the use of more benign reagents than the buffered oxide etch for de-encapsulation, which may allow for direct analysis or use of the released nucleic acid without requiring a desalting step. Implementation of PANDORA as a real-time analysis tool would require parallelization of FAS to rapidly sort through a library of samples or devising strategies that use physical extraction of samples instead, such as magnetic-activated sorting[22,40] or nucleic-acid functionalized surfaces,[20,41] to sort and retrieve samples with high purity rapidly.

The versatility of the PANDORA platform to store nucleic acids from any origin or length, coupled with a scalable molecular barcoding system for identification and computation, provides opportunities for the large-scale collection and analysis of samples at an unprecedented scale, which has important implications for patient care, epidemiology, and public health.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsami.1c14985.

General methods for encapsulation, barcoding, fluorescence-activated sorting, de-encapsulation, and qPCR validation; SEM images; DLS and zeta potential characterization; qPCR assay panel design details; raw qPCR curves for all sorting experiments; determination of average *S. aureus* genome per capsule; short-read amplicon sequencing validation; and accelerated weathering experiments (PDF)

AUTHOR INFORMATION

Corresponding Authors

James L. Banal − *Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge,*

https://doi.org/10.1021/acsami.1c14985
ACS Appl. Mater. Interfaces 2021, 13, 49729−49736

Massachusetts 02139, United States; orcid.org/0000-0002-2364-4824; Email: jbanal@mit.edu

Mark Bathe − *Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States; Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, United States;* orcid.org/0000-0002-6199-6855; Email: mark.bathe@mit.edu

Complete contact information is available at: https://pubs.acs.org/10.1021/acsami.1c14985

Author Contributions

J.L.B. and M.B. conceived the encapsulation and labeling system. J.L.B. performed the experiments and analyzed the data. M.B. supervised the project. J.L.B and M.B. wrote the manuscript.

Funding

J.L.B. and M.B. gratefully acknowledge support from the Office of Naval Research (N00014-17-1-2609 and N00014-21-1-4013) and National Science Foundation (CCF-1956054).

Notes

The authors declare the following competing financial interest(s): Mark Bathe is the founder of Cache DNA LLC that is commercializing the technology disclosed in this work. The Massachusetts Institute of Technology have filed patents on the storage, access, and retrieval of biopolymers (US application numbers 16/012583 and 16/097594) on behalf of the inventors James L. Banal and Mark Bathe.

■ ACKNOWLEDGMENTS

We thank R. Falkovich for providing the HEK293 cells for this work and D. Mankus from the Koch Institute Nanotechnology Materials Core Facility for assistance in electron microscopy imaging.

■ REFERENCES

(1) Wan, E.; Akana, M.; Pons, J.; Chen, J.; Musone, S.; Kwok, P. Y.; Liao, W. Green Technologies for Room Temperature Nucleic Acid Storage. *Curr. Issues Mol. Biol.* **2010**, *12*, 135−142.

(2) Muller, R.; Betsou, F.; Barnes, M. G.; Harding, K.; Bonnet, J.; Kofanova, O.; Crowe, J. H.; International Society for, B. Preservation of Biospecimens at Ambient Temperature: Special Focus on Nucleic Acids and Opportunities for the Biobanking Community. *Biopreserv. Biobanking* **2016**, *14*, 89−98.

(3) Lee, S. B.; Clabaugh, K. C.; Silva, B.; Odigie, K. O.; Coble, M. D.; Loreille, O.; Scheible, M.; Fourney, R. M.; Stevens, J.; Carmody, G. R.; Parsons, T. J.; Pozder, A.; Eisenberg, A. J.; Budowle, B.; Ahmad, T.; Miller, R. W.; Crouse, C. A. Assessing a Novel Room Temperature DNA Storage Medium for Forensic Biological Samples. *Forensic Sci. Int.: Genet.* **2012**, *6*, 31−40.

(4) Fabre, A.-L.; Colotte, M.; Luis, A.; Tuffet, S.; Bonnet, J. An Efficient Method for Long-Term Room Temperature Storage of RNA. *Eur. J. Hum. Genet.* **2014**, *22*, 379−385.

(5) Tribioli, C.; Lufkin, T. Long-Term Room Temperature Storage of High-Quality Embryonic Stem Cell Genomic DNA Extracted with a Simple and Rapid Procedure. *J. Biomol. Tech.* **2006**, *17*, 249−251.

(6) Miernyk, K. M.; DeByle, C. K.; Rudolph, K. M. Evaluation of Two Matrices for Long-Term, Ambient Storage of Bacterial DNA. *Biopreserv. Biobanking* **2017**, *15*, 529−534.

(7) Clermont, D.; Santoni, S.; Saker, S.; Gomard, M.; Gardais, E.; Bizet, C. Assessment of DNA encapsulation, a new room-temperature DNA storage method. *Biopreserv. Biobanking* **2014**, *12*, 176−183.

(8) Benaroudj, N.; Lee, D. H.; Goldberg, A. L. Trehalose Accumulation During Cellular Stress Protects Cells and Cellular Proteins From Damage By Oxygen Radicals. *J. Biol. Chem.* **2001**, *276*, 24261−24267.

(9) Ramløv, H.; Westh, P. Survival of the Cryptobiotic Eutardigrade Adorybiotus coronifer During Cooling to −196 °C: Effect of Cooling Rate, Trehalose Level, and Short-Term Acclimation. *Cryobiology* **1992**, *29*, 125−130.

(10) Watanabe, M.; Kikawada, T.; Minagawa, N.; Yukuhiro, F.; Okuda, T. Mechanism Allowing an Insect to Survive Complete Dehydration and Extreme Temperatures. *J. Exp. Biol.* **2002**, *205*, 2799−2802.

(11) Clement, O.; Whitney, S.; Muller-Cohn, J.; Muller, R. Following Nature's Lead: Generating Compounds for Stabilizing Biomolecules. *Biopreserv. Biobanking* **2012**, *10*, 395−402.

(12) Liu, X.; Li, Q.; Wang, X.; Zhou, X.; He, X.; Liao, Q.; Zhu, F.; Cheng, L.; Zhang, Y. Evaluation of DNA/RNAshells for Room Temperature Nucleic Acids Storage. *Biopreserv. Biobanking* **2015**, *13*, 49−55.

(13) Kohll, A. X.; Antkowiak, P. L.; Chen, W. D.; Nguyen, B. H.; Stark, W. J.; Ceze, L.; Strauss, K.; Grass, R. N. Stabilizing Synthetic DNA for Long-Term Data Storage with Earth Alkaline Salts. *Chem. Commun.* **2020**, *56*, 3613−3616.

(14) Paunescu, D.; Fuhrer, R.; Grass, R. N. Protection and Deprotection of DNA-High-Temperature Stability of Nucleic Acid Barcodes for Polymer Labeling. *Angew. Chem., Int. Ed.* **2013**, *52*, 4269−4272.

(15) Paunescu, D.; Puddu, M.; Soellner, J. O. B.; Stoessel, P. R.; Grass, R. N. Reversible DNA encapsulation in silica to produce ROS-resistant and heat-resistant synthetic DNA 'fossils'. *Nat. Protoc.* **2013**, *8*, 2440−2448.

(16) Grass, R. N.; Heckel, R.; Puddu, M.; Paunescu, D.; Stark, W. J. Robust Chemical Preservation of Digital Information on DNA in Silica with Error-Correcting Codes. *Angew. Chem., Int. Ed.* **2015**, *54*, 2552−2555.

(17) Puddu, M.; Stark, W. J.; Grass, R. N. Silica Microcapsules for Long-Term, Robust, and Reliable Room Temperature RNA Preservation. *Adv. Healthcare Mater.* **2015**, *4*, 1332−1338.

(18) Banal, J. L.; Shepherd, T. R.; Berleant, J.; Huang, H.; Reyes, M.; Ackerman, C. M.; Blainey, P. C.; Bathe, M. Random Access DNA Memory using Boolean Search in an Archival File Storage System. *Nat. Mater.* **2021**, *20*, 1272−1280.

(19) Xu, Q.; Schlabach, M. R.; Hannon, G. J.; Elledge, S. J. Design of 240,000 Orthogonal 25mer DNA Barcode Probes. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2289−2294.

(20) Leidner, A.; Weigel, S.; Bauer, J.; Reiber, J.; Angelin, A.; Grösche, M.; Scharnweber, T.; Niemeyer, C. M. Biopebbles: DNA-Functionalized Core-Shell Silica Nanospheres for Cellular Uptake and Cell Guidance Studies. *Adv. Funct. Mater.* **2018**, *28*, 1707572.

(21) Chen, W. D.; Kohll, A. X.; Nguyen, B. H.; Koch, J.; Heckel, R.; Stark, W. J.; Ceze, L.; Strauss, K.; Grass, R. N. Combining Data Longevity with High Storage Capacity-Layer-by-Layer DNA Encapsulated in Magnetic Nanoparticles. *Adv. Funct. Mater.* **2019**, *29*, 1901672.

(22) Sutermaster, B. A.; Darling, E. M. Considerations for High-Yield, High-Throughput Cell Enrichment: Fluorescence Versus Magnetic Sorting. *Sci. Rep.* **2019**, *9*, 227.

(23) Organick, L.; Nguyen, B. H.; McAmis, R.; Chen, W. D.; Kohll, A. X.; Ang, S. D.; Grass, R. N.; Ceze, L.; Strauss, K. An Empirical Comparison of Preservation Methods for Synthetic DNA Data Storage. *Small Methods* **2021**, *5*, 2001094.

(24) An, R.; Jia, Y.; Wan, B.; Zhang, Y.; Dong, P.; Li, J.; Liang, X. Non-Enzymatic Depurination of Nucleic Acids: Factors and Mechanisms. *PLoS One* **2015**, *9*, No. e115950.

(25) Puddu, M.; Paunescu, D.; Stark, W. J.; Grass, R. N. Magnetically Recoverable, Thermostable, Hydrophobic DNA/Silica Encapsulates and Their Application as Invisible Oil Tags. *ACS Nano* **2014**, *8*, 2677−2685.

(26) Ryder, O. A.; McLaren, A.; Brenner, S.; Zhang, Y.-P.; Benirschke, K. ECOLOGY:DNA Banks for Endangered Animal Species. *Science* **2000**, *288*, 275−277.

(27) Brandies, P.; Peel, E.; Hogg, C. J.; Belov, K. The Value of Reference Genomes in the Conservation of Threatened Species. *Genes* **2019**, *10*, 846.
(28) Giannopoulou, E.; Katsila, T.; Mitropoulou, C.; Tsermpini, E.-E.; Patrinos, G. P. Integrating Next-Generation Sequencing in the Clinical Pharmacogenomics Workflow. *Front. Pharmacol.* **2019**, *10*, 384.
(29) Laverick, M. A.; Wyn-Jones, A. P.; Carter, M. J. Quantitative RT-PCR for the Enumeration of Noroviruses (Norwalk-Like Viruses) in Water and Sewage. *Lett. Appl. Microbiol.* **2004**, *39*, 127−136.
(30) Aw, T. G.; Gin, K. Y. H. Environmental Surveillance and Molecular Characterization of Human Enteric Viruses in Tropical Urban Wastewaters. *J. Appl. Microbiol.* **2010**, *109*, 716−730.
(31) Bulman, S. R.; McDougal, R. L.; Hill, K.; Lear, G. Opportunities and Limitations for DNA Metabarcoding in Australasian Plant-Pathogen Biosecurity. *Australas. Plant Pathol.* **2018**, *47*, 467−474.
(32) Sato, Y.; Mizuyama, M.; Sato, M.; Minamoto, T.; Kimura, R.; Toma, C. Environmental DNA Metabarcoding to Detect Pathogenic Leptospira and Associated Organisms in Leptospirosis-Endemic Areas of Japan. *Sci. Rep.* **2019**, *9*, 6575.
(33) Lodder, W.; de Roda Husman, A. M. SARS-CoV-2 in Wastewater: Potential Health Risk, But Also Data Source. *Lancet Gastroenterol. Hepatol.* **2020**, *5*, 533−534.
(34) Klein, A. M.; Mazutis, L.; Akartuna, I.; Tallapragada, N.; Veres, A.; Li, V.; Peshkin, L.; Weitz, D. A.; Kirschner, M. W. Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* **2015**, *161*, 1187−1201.
(35) Macosko, E. Z.; Basu, A.; Satija, R.; Nemesh, J.; Shekhar, K.; Goldman, M.; Tirosh, I.; Bialas, A. R.; Kamitaki, N.; Martersteck, E. M.; Trombetta, J. J.; Weitz, D. A.; Sanes, J. R.; Shalek, A. K.; Regev, A.; McCarroll, S. A. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **2015**, *161*, 1202−1214.
(36) Lan, F.; Demaree, B.; Ahmed, N.; Abate, A. R. Single-Cell Genome Sequencing at Ultra-High-Throughput with Microfluidic Droplet Barcoding. *Nat. Biotechnol.* **2017**, *35*, 640−646.
(37) Zheng, G. X. Y.; Terry, J. M.; Belgrader, P.; Ryvkin, P.; Bent, Z. W.; Wilson, R.; Ziraldo, S. B.; Wheeler, T. D.; McDermott, G. P.; Zhu, J.; Gregory, M. T.; Shuga, J.; Montesclaros, L.; Underwood, J. G.; Masquelier, D. A.; Nishimura, S. Y.; Schnall-Levin, M.; Wyatt, P. W.; Hindson, C. M.; Bharadwaj, R.; Wong, A.; Ness, K. D.; Beppu, L. W.; Deeg, H. J.; McFarland, C.; Loeb, K. R.; Valente, W. J.; Ericson, N. G.; Stevens, E. A.; Radich, J. P.; Mikkelsen, T. S.; Hindson, B. J.; Bielas, J. H. Massively Parallel Digital Transcriptional Profiling of Single Cells. *Nat. Commun.* **2017**, *8*, 14049.
(38) Lareau, C. A.; Duarte, F. M.; Chew, J. G.; Kartha, V. K.; Burkett, Z. D.; Kohlway, A. S.; Pokholok, D.; Aryee, M. J.; Steemers, F. J.; Lebofsky, R.; Buenrostro, J. D. Droplet-Based Combinatorial Indexing for Massive-Scale Single-Cell Chromatin Accessibility. *Nat. Biotechnol.* **2019**, *37*, 916−924.
(39) Liu, L.; Dalal, C. K.; Heineike, B. M.; Abate, A. R. High Throughput Gene Expression Profiling of Yeast Colonies with Microgel-Culture Drop-Seq. *Lab Chip* **2019**, *19*, 1838−1849.
(40) Adams, J. D.; Kim, U.; Soh, H. T. Multitarget Magnetic Activated Cell Sorter. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18165−18170.
(41) Hu, Y.; Grösche, M.; Sheshachala, S.; Oelschlaeger, C.; Willenbacher, N.; Rabe, K. S.; Niemeyer, C. M. Bottom-Up Assembly of DNA-Silica Nanocomposites into Micrometer-Sized Hollow Spheres. *Angew. Chem., Int. Ed.* **2019**, *58*, 17269−17272.

https://doi.org/10.1021/acsami.1c14985
ACS Appl. Mater. Interfaces 2021, 13, 49729−49736