Diagnostic optical sequencing

Lee E. Korshoj^{a,b} and Prashant Nagpal^{a,b,c,*}

^aDepartment of Chemical and Biological Engineering, ^bRenewable and Sustainable Energy Institute (RASEI), ^cMaterials Science and Engineering, University of Colorado Boulder, Boulder, CO 80303, USA

*Email: pnagpal@colorado.edu

Supplementary information available.

Abstract: Advances in precision medicine require high-throughput, inexpensive, point-of-care diagnostic methods with multi-omics capability for detecting a wide range of biomolecules and their molecular variants. Optical techniques have offered many promising advances towards such diagnostics. However, the inability to squeeze light with several hundred-nanometer wavelengths into angstrom-scale volume for single nucleotide measurements has hindered further progress. Recently, a block optical sequencing (BOS) method has been shown for determining relative nucleobase content in DNA k-mer blocks with Raman spectroscopy, and a block optical content scoring (BOCS) algorithm was developed for robust content-based genetic biomarker database searching. Here, we performed BOS measurements on positively-charged silver nanoparticles to achieve 93.3% accuracy for predicting nucleobase content in DNA k-mer blocks (where k=10), as well as measurements on RNA and chemically-modified nucleobases for extensions to transcriptomic and epigenetic studies. Our high-accuracy BOS measurements were then used with BOCS to correctly identify a β -lactamase gene from the MEGARes antibiotic resistance database and confirm the *Pseudomonas aeruginosa* pathogen of origin from <12 content measurements (<15% coverage) of the gene. These results prove the integration of BOS/BOCS as a diagnostic optical sequencing platform. With the versatile range of available plasmonic substrates offering simple data acquisition, varying resolution (single-molecule to ensemble), and multiplexing, this optical sequencing platform has potential as the rapid, cost-effective method needed for broad-spectrum biomarker detection.

Introduction

Precision medicine can become one step closer to realization with inexpensive, non-specific assays capable of broad-spectrum diagnostics. Ideally, a single point-of-care test could rapidly screen an array of genomic, transcriptomic, and epigenetic biomarkers to inform proper therapeutic treatment.^{1,2} One urgent application is to combat antibiotic-resistant infections, which continue to create an increasing financial strain on healthcare systems worldwide and lead to high rates of mortality for the millions affected each year.^{3,4} Clinicians carry the burden to rapidly assess the potential resistances of pathogens and administer the correct antibiotics from the onset of infection. Unfortunately, initial antibiotic regimens are incorrect as much as 50% of the time.^{5–7} As initial administration of ineffective antibiotic treatments has proven to result in higher mortality rates, where in some cases survival decreases by 8% per hour, it becomes clear that more effective diagnostics are needed.^{8–11} Current resistance diagnostics and pathogen profiling are often only performed after initial antibiotics fail. Most of these assays rely on cell cultures, PCR amplifications, and microarray analyses.^{12–19} These tests require hours to days to complete and carry significant costs, while many are specific for merely one or a few well-characterized strains or resistances. Therefore, a method to rapidly and affordably screen a wide range of drug resistance in clinically relevant microbial pathogens is vital for prescribing patients with appropriate and timely treatments that reduce mortality rates as well as the spread of further resistance. Such a method would not be limited to antibiotic resistance screening, but also applicable in the screening for cancers and other genetic diseases where early detection is critical for patient survival.^{20–22}

Here, we address the clinical diagnostic needs discussed above by demonstrating an optical sequencing platform capable of rapid, broad-spectrum diagnostics. Our diagnostic optical sequencing combines label-free Raman spectroscopy measurements of A-G-C-T (adenine-guanine-cytosine-thymine) content within DNA fragments (block optical sequencing – BOS²³) with a robust bioinformatics algorithm for genetic biomarker database searching (block optical content scoring – BOCS²⁴). BOS uses plasmonic substrates for surface-enhanced Raman spectroscopy (SERS) measurements of nucleotide content within unlabeled DNA k-mer blocks as an alternative to single-letter sequencing. Given the range of available plasmonic metallic nanoparticles and nanopatterned surfaces, BOS is flexible to many applicable substrates and can be made high-throughput with millions of simultaneous SERS measurements through multiplexing.^{25–30} BOCS uses a machine learning approach for mapping and scoring k-mer block reads from BOS to a genetic biomarker database. Initial simulations of BOCS showed accurate recognition of specific resistance, cancer, and other genetic disease genes with significantly less than full coverage of the genes.

We successfully coupled BOS measurements with the BOCS algorithm for characterization of a β -lactamase gene within the pathogen of origin. Specifically, we show that merely a few highly accurate BOS readings of DNA k-mer block content ($\ll$ full coverage of the gene) from silver nanoparticles can be used with the BOCS algorithm to identify the correct OXA β -lactamase (class D) gene from a comprehensive antibiotic resistance database and confirm the *Pseudomonas aeruginosa* pathogen from which it originates. While our initial BOS method

employed a multiplexed, silver-coated nanopyramid substrate for SERS, we utilized metallic nanoparticles here to demonstrate broader applicability and varying resolution (single-molecule versus ensemble). We also show extensions to transcriptomics and epigenomics. Ultimately, the results here demonstrate the use of combined BOS/BOCS as a diagnostic optical sequencing platform for inexpensive and rapid identification of broad-spectrum genetic, transcriptomic, and epigenomic biomarkers.

Results and discussion

BOS measurements with positively-charged silver nanoparticles

Due to the distinct biochemical nature of the nucleobases in DNA, each has unique interactions with light photons and a unique Raman spectrum.^{31–34} With SERS measurements, surface plasmon polaritons squeeze light photons with several hundred-nanometer wavelengths into molecular length scales, where they can interact with the nucleotides in DNA.^{25,35,36} Despite the advances in nano-optics and three-dimensional plasmonic nanofocusing, the nanometer scale mode volumes prevent characterization of photon interactions with single nucleotides. While this inhibits single-letter sequencing of DNA, it allows for determination of A-G-C-T content within k-mer blocks. By measuring content instead of single-letter sequences, which is more than sufficient for biomarker detection, the BOS system remedies the massive data analysis problem present with other sequencing platforms by incorporating inherent lossy data compression, and has the potential for extremely high-throughput measurements with millions of simultaneous multiplexed reads.^{36–38}

In this study, we collected BOS measurements from ssDNA k-mer blocks with positively-charged, spermine-coated silver nanoparticles (Ag NPs) as the plasmonic substrate (Figure 1a). Recent work has shown strong, reproducible SERS signal from a range of substrates like single DNA molecules on nanopyramid substrates,²³ and ensemble measurement from ~25 nm cationic nanoparticles for ssDNA, dsDNA, and RNA.^{29,39–41} The Ag NPs remain stable in colloidal solution due to electrostatic repulsion of the positively-charged ligands, and show no significant background Raman signal (Figure S1). SERS signal is only achieved upon aggregation with the addition of negatively-charged nucleic acids, which is DNA k-mer blocks for our measurements. As seen from the extinction spectrum in Figure 1a, a strong localized surface plasmon resonance (LSPR) peak is observed at ~392 nm for blank Ag NPs, and a large red shift is observed with the addition of DNA due to aggregation. When added to the Ag NPs, the DNA strands attach electrostatically to the nanoparticle surfaces and at interparticle hot spots leading to strong SERS excitation with a 532 nm Raman laser. In the analysis of all SERS measurements shown here, we perform a consistent signal processing (cosmic ray removal, smoothing, shift correction – Figure S2) and normalization (baseline subtraction, normalization to standard peak – Figure S3).

For sequencing applications it is essential to first know the specific Raman signal, or signature, for each of the four nucleobases A, G, C, and T. To get these signatures, we performed SERS measurements on homologous 10-mer DNA sequences (i.e., poly(N)₁₀, where N is A, G, C, and T). Shown in Figure 1b, there exists a complex pattern of Raman peak features for each

nucleobase over the range of 350 – 2000 cm^{-1} shift (assignment of specific Raman peak modes provided in Table S1). For BOS, we identified a single distinctive Raman peak for each nucleobase as the “signature peak” which is later used to determine content in unknown sequence blocks. For purines, we selected the ring breathing modes at $\sim 740 \text{ cm}^{-1}$ for A and $\sim 690 \text{ cm}^{-1}$ for G.^{29,31,34,40,42} For pyrimidines, we selected the ring bending modes at $\sim 600 \text{ cm}^{-1}$ for C and $\sim 460 \text{ cm}^{-1}$ for T.^{29,32,33,40,42} Within each SERS measurement, the PO_2^- stretching mode peak at 1089 cm^{-1} due to the phosphate backbone is used as an internal standard for normalizing the relative peak intensities, as is consistent with other studies employing nanoparticle substrates.^{27,40} All signature peaks and the PO_2^- normalization peak are highlighted in Figure 1b. In the SERS signal from a mixed sequence DNA block, the four signature peaks are present with relative intensities (normalized to the PO_2^- peak) corresponding to their respective content, shown for a 10-mer DNA block with content A:1, G:4, C:2, and T:3 in Figure 1c. These relative intensities in signature peak locations can therefore be used to deconvolute the signal from an unknown mixed sequence.

While the focus in this study is put on DNA, it is important to note that impactful extensions exist for transcriptomics and epigenomics by applying BOS to RNA and chemically modified nucleobases. As shown in Figure 1d for a homologous RNA sequence of repeating adenine, A, the signature peaks are nearly identical between RNA and DNA. This is evident by the similar $\sim 740 \text{ cm}^{-1}$ ring breathing mode peak for A, which is consistent with recent work by other groups.⁴¹ Additionally, small perturbations to the Raman spectrum can be seen due to modified nucleobases like the modification of cytosine, C, to 5-methylcytosine, 5mC (also measured from homologous sequences), which plays an important role in gene regulation.^{43,44} Seen in the highlighted regions of Figure 1e, we observe shifts and intensity changes to the signature peak of C at $\sim 600 \text{ cm}^{-1}$ and the strong ring breathing mode for pyrimidines at $\sim 800 \text{ cm}^{-1}$, which are consistent with previous studies.²⁶ This opens the opportunity to directly apply our diagnostic optical sequencing platform to transcriptomic and epigenomic studies.

Calibrating the BOS system with DNA block standards

To fully deconvolute the A-G-C-T content of an unknown mixed sequence DNA k-mer block, it is necessary to know the full range of intensity values for signature peaks of each nucleobase. Therefore, we used custom DNA k-mer blocks with known content as standards for generating content calibrations for BOS. The 14 calibration blocks are provided in Table 1. These 14 ssDNA 10-mer calibration blocks span the range of 0 – 1 fractional content for each of the four nucleobases. Blocks Cal_1, Cal_2,

Table 1. Calibration blocks (ssDNA 10-mers)

Block Name	Sequence	Content			
		A	G	C	T
Cal_1	AAAAAAAAAA	1.0	0.0	0.0	0.0
Cal_2	GGGGGGGGGG	0.0	1.0	0.0	0.0
Cal_3	CCCCCCCCCC	0.0	0.0	1.0	0.0
Cal_4	TTTTTTTTTT	0.0	0.0	0.0	1.0
Cal_5	AAAGAAAACA	0.8	0.1	0.1	0.0
Cal_6	GGGTGGGAGG	0.1	0.8	0.0	0.1
Cal_7	CCTCCACCC	0.1	0.0	0.8	0.1
Cal_8	TGTTTCTTT	0.0	0.1	0.1	0.8
Cal_9	AGAATAGAAT	0.6	0.2	0.0	0.2
Cal_10	CGGAGGAGCG	0.2	0.6	0.2	0.0
Cal_11	CGCTCCGCCT	0.0	0.2	0.6	0.2
Cal_12	CTTCTATTAT	0.2	0.0	0.2	0.6
Cal_13	AACGCATCCA	0.4	0.1	0.4	0.1
Cal_14	GTGCGATTGT	0.1	0.4	0.1	0.4

Cal_3, and Cal_4 provided the Raman signatures shown in Figure 1b, as they are of content = 1. Together, SERS measurements on the set of 14 calibration blocks were used to generate the BOS calibrations shown in Figure 2 (all SERS spectra of the calibration blocks are provided in Figure S4). On the left of Figure 2, SERS spectra are plotted for increasing fractional content of a particular nucleobase (lighter to darker shades) in the zoomed-in region of the signature peaks. A direct, linear correlation is observed between fractional content and the signature peak normalized

intensity for each of the nucleobases, seen in the fitted data points on the right of Figure 2 (data points and variance are from five technical replicates of each calibration block). The trends are all linear although the range of measured intensity values can vary significantly. Intensities for the adenine, A, signature peak at $\sim 740\text{ cm}^{-1}$ range from 0 to >4 while the intensities for the thymine, T, signature peak at $\sim 460\text{ cm}^{-1}$ range from 0 to 0.3. Linear fits to these trends, with intercept locked at zero, provide the finalized correlations which can be used to determine content in unknown DNA k-mer blocks.

Content identification within gene blocks

We applied the BOS calibrations towards identifying content within k-mer blocks from an actual gene sequence, for subsequent integration with the BOCS algorithm. The 15 gene blocks are provided in Table 2. These 15 ssDNA 10-mer gene blocks are from an OXA β -lactamase (class D) gene found in *P. aeruginosa* (full sequence and mapped blocks are provided in Figure S5). From SERS measurements on the 15 gene blocks (all SERS spectra of the gene blocks are provided in Figure S6), we measured the normalized intensity for signature peaks (averaged from three

Table 2. Gene blocks (ssDNA 10-mers)

Block Name	Sequence	Content			
		A	G	C	T
Gen_1	CCCACTTTCT	0.1	0.0	0.5	0.4
Gen_2	ACGAGGTTCT	0.2	0.3	0.2	0.3
Gen_3	GCGCAGGGAG	0.2	0.6	0.2	0.0
Gen_4	GATCAGCGCG	0.2	0.4	0.3	0.1
Gen_5	CCCCTCCTCT	0.0	0.0	0.7	0.3
Gen_6	GGTGCGAAC	0.2	0.5	0.2	0.1
Gen_7	AAGCGCAACG	0.4	0.3	0.3	0.0
Gen_8	CTTCGTCCTC	0.0	0.1	0.5	0.4
Gen_9	AGCGGCTCTA	0.2	0.3	0.3	0.2
Gen_10	GGTGGGTGGG	0.0	0.8	0.0	0.2
Gen_11	GACCGGGAGC	0.2	0.5	0.3	0.0
Gen_12	GCCAGGTTGT	0.1	0.4	0.2	0.3
Gen_13	GCCAATGTCT	0.2	0.2	0.3	0.3
Gen_14	AAGCCCCAGC	0.3	0.2	0.5	0.0
Gen_15	CCGTGCGCGC	0.0	0.4	0.5	0.1

technical replicates). Figure 3 details the process of using these measurements for predicting content within the blocks. Using the Gen_4 block as an example, the measured normalized intensity for each of the four signature peaks is used to predict a content for each nucleobase A, G, C, and T by multiplying the measured intensity by the slope of the linear calibrations. This initial raw predicted content is then normalized such that the sum is equal to one. Knowing that each nucleobase must be present in integer quantities, the final predicted content can be determined via rounding. The final predicted content for the Gen_4 block agrees with the actual content given in Table 2.

Predicted content for all 15 gene blocks is provided in Figure 4, where the predictions are compared to the actual content. It is important to note that the accuracies reported in BOS are different than those from traditional single-letter sequencing. Since reads are of A-G-C-T content and not letter-by-letter sequences, one misidentification results in a double error. This is because the content of the incorrect nucleobase and substituted nucleobase are both affected. A confusion matrix analysis on the single nucleobase level throughout all 15 blocks shows that the majority of errors result from guanine, G, content being under identified ($\sim 10\%$ of G bases throughout the gene blocks). In total, errors in the predicted content were present in only four of the blocks – three

which resulted from a single nucleobase swap (80% accuracy for the block) and one which double-swapped nucleobases (60% accuracy for the block). Content in the other 11 blocks was predicted with 100% accuracy. Overall, the content was predicted at an average accuracy of 93.3%.

MDR pathogen profiling with BOS and BOCS

For full integration into a diagnostic method, the high-accuracy BOS reads were coupled with the BOCS algorithm for genetic biomarker detection. With the combined BOS/BOCS platform, we set out to demonstrate detection of a *P. aeruginosa* infection with drug-resistant β -lactamase gene. *P. aeruginosa* is a clinical multidrug-resistant (MDR) pathogen of critical importance due to its prevalence for causing bloodstream, urinary, and pulmonary infections in hospital settings, especially for immunocompromised patients in intensive care settings.^{45,46} Due to the multiple mechanisms of inherent and acquired resistance of this organism, patients infected with *P. aeruginosa* have limited therapeutic options. It is therefore imperative to have more early-stage, rapid diagnostic techniques in place to screen for *P. aeruginosa* so that effective antibiotic regimens can be prescribed from the onset of infection.

The BOCS algorithm was developed to specifically utilize DNA k-mer block content reads from the BOS system.²⁴ It operates analogously to probability-based sequence analyzers such as those employed for peptide identification from mass spectrometry data,^{47–49} and alignment programs used for mapping next-generation sequencing reads to reference genomes.^{50,51} In a similar fashion, the BOCS algorithm relies on probabilistic content alignments to database sequences of genetic biomarkers. Outlined in Figure 5a, the algorithm cycles through each of the logged k-mer block content reads from BOS and performs a content-based alignment with each gene sequence in a database, translating through the gene sequence one nucleotide at a time. The alignment tracks the number of match locations – where the k-mer block content matches the content of the k-length gene segment. The number of match locations is the fundamental parameter in a set of six probability factors that act as machine learning elements in the calculation of an overall content score. Genes in the database are probabilistically ranked, and identified, based on the content score as it compounds with more blocks being analyzed. The algorithm also incorporates logic elements such as a penalty score given to genes in the database where no matches are found during alignment, thresholding to eliminate low-ranking genes in the database that may skew the content scoring, and entropy screening to eliminate BOS reads that have a maximal number of permutations based on the content. Thorough simulations of BOCS with antibiotic resistance, cancer, and other genetic disease databases proved very robust, even under the pressures of variable k-mer block lengths, high error rates, and in the presence of blocks comprised of multiple genes.

We ran our measured gene blocks with predicted content at 93.3% accuracy through the BOCS algorithm against the MEGARes antibiotic resistance database⁵² comprised of ~4000 known resistance genes, including the OXA β -lactamase (class D) gene of our measured gene blocks. This analysis demonstrates the ability of the BOS/BOCS platform to diagnose antibiotic resistances from unknown samples with no prior knowledge of pathogen or strain. The table of gene blocks and their predicted content which was provided to the BOCS algorithm is shown in the lower portion of Figure 5. Note that three of the blocks were eliminated with the entropy screening functionality of BOCS because these blocks had predicted content with a maximum number of permutations/entropy (>25000 permutations of the 2-2-3-3 content within the 10-mer

Figure 5. MDR pathogen profiling with integrated BOS and BOCS. (a) Overview of the BOCS algorithm integration with BOS measurements. Starting with a log of measured content within DNA k-mer blocks from BOS (B1...Bn as shown) and a gene database (excerpts from the MEGARes antibiotic resistance database are shown), the blocks are individually aligned to each gene in the database based on content. This alignment consists of finding all match locations for the k-mer block content within a gene via translating through the gene one nucleotide at a time and looking at fragments of length k. For each block, a content score is calculated based on the number of matches for the k-mer block and various probability factors. As more blocks are analyzed, content scores are compounded and genes in the database are ranked and eliminated. The BOCS algorithm was run for the 15 10-mer gene blocks in Table 2 from an OXA β -lactamase gene (with the predicted content at 93.3% accuracy). Note that only 12 of the 15 blocks were used, as three were eliminated with entropy screening. Two cases were studied: 1. Identifying the gene from the MEGARes antibiotic resistance database of ~4000 resistance genes (b), and 2. Identifying the gene within the *P. aeruginosa* genome (c). Both cases demonstrate robust identification of the correct OXA resistance gene from content score ranking, requiring merely a few content measurements. Additionally, >90% of genes in both databases were eliminated after a single block was analyzed by the BOCS algorithm. The following settings were used when running BOCS: penalty score – 0.1, thresholding multiplier – 0.1, entropy screening – ‘on’ (eliminated only the blocks with permutations > 25000).

block), and therefore do not benefit the scoring and ranking. Figure 5b plots the content score as consecutive blocks have been analyzed (1 through 12 total blocks). It can be seen that the correct OXA gene from the MEGARes database is identified based on its top content score ranking well within the 12 blocks that are shown. The OXA gene begins to separate itself in ranking from the other database genes after merely the fifth block was analyzed, and it becomes easily identifiable after the twelfth block. The fifth block corresponds to only 0.063 or 6.3% coverage of the gene (i.e., 50 nucleotides analyzed in the five 10-mer blocks, out of 789 total nucleotides in the resistance gene sequence), while the twelfth block corresponds to 0.152 or 15.2% coverage of the gene. Also shown in Figure 5b is the specificity, or how significantly the gene database can be narrowed as consecutive blocks have been analyzed. We see that >90% of the total MEGARes genes in the database can be eliminated after merely the first block is analyzed.

Extending diagnostic applications further, we ran our measured gene blocks through the BOCS algorithm again after substituting the MEGARes database for the *P. aeruginosa* reference genome PAO1⁵³ containing the OXA β -lactamase (class D) gene. This analysis indicates the ability of the BOS/BOCS platform to confirm pathogens and specific strains responsible for the infection. It also shows the robustness of the BOCS algorithm in identifying specific genes in the background of an entire microbial genome. Figure 5c plots the content score from BOCS as consecutive gene blocks have been analyzed. Just as with the MEGARes resistance database, the correct OXA gene from the *P. aeruginosa* genome database is identified based on its top content score ranking within the 12 blocks. Additionally, high specificity is seen as >90% of the total genes in the database can be eliminated after merely the first block is analyzed. These results demonstrate the potential for using our BOS/BOCS platform as a diagnostic optical sequencing technique for profiling MDR pathogens. The results shown here for a single β -lactamase gene in *P. aeruginosa* can be extended to other resistance genes from pathogenic microbial strains without any changes in experimental setup, ultimately providing the broad-spectrum detection needed for directing appropriate and timely treatments in a clinical setting.

Conclusion

Sequencing methods like those from Illumina and Oxford Nanopore Technologies are transforming clinical diagnostics through rapid, cost-effective characterization of patient samples.⁵⁴ For antibiotic resistance screening, current steps for pathogen characterization continue to rely on culture-based assays that can take days to weeks for results to be finalized. Next-generation sequencing has proven to significantly improve these metrics. The majority of next-generation techniques sequence a fragmented DNA sample and then piece together the fragments with alignment algorithms and a reference sequence.^{37,51} Clinical applications of Illumina sequencing suggest its use as an effective method of regular infection screening for control of MDR pathogens in hospitals.⁵⁵ For this, 75-100 $\times$ gene coverage is routinely needed for accurate results, which can require several days of runtime and analysis. The MinION system from Oxford Nanopore Technologies has shown the ability to identify antibiotic resistance genes and classify pathogens down to the species level.⁵⁶⁻⁵⁸ With these studies, runtimes were 6-18 hours and gene

coverage was $>10\text{-}14\times$, with optimized experimental protocols. Despite the progress being made with these sequencing methods, limitations of the platforms in both the measurements and algorithmic analysis continue to necessitate high coverages for accurate results, leading to long runtimes and higher costs.

Results from our BOS/BOCS platform demonstrate the significant potential of an optical sequencing technique to improve on the limitations seen in these current state-of-the-art sequencing methods. An optical technique could lead to improvements in time, cost, and complexity. On the measurement side, the label-free optical detection within BOS simplifies the sample preparation and gives the potential for multiplexing data acquisition, which would dramatically decrease runtimes. On the analysis side, the BOCS algorithm exceeds the performance of the alignment algorithms used by next-generation sequencers by requiring lower coverages. We showed that $\ll 1$ gene coverage (i.e., not even seeing the entire gene) is often all that is required for identification of genetic biomarkers from a database. This low coverage cuts down on costs associated with reagents as amplification of the sample could be significantly decreased or eliminated, and runtimes reduced.

In conclusion, we have coupled Raman measurements of A-G-C-T content in DNA k-mer blocks with a genetic biomarker database searching algorithm for integration as a diagnostic optical sequencing platform. With block optical sequencing (BOS), we employed positively-charged silver nanoparticles for reproducible SERS measurements from DNA blocks. Identification of four Raman ring breathing and bending modes, each specific to one of the four nucleobases, were observed to have intensities that are linearly correlated with increasing nucleobase content within a DNA block. A set of standard blocks with known nucleobase content were used to fit linear calibrations that led to 93.3% accuracy in predicting the relative A-G-C-T content of a set of blocks from a β -lactamase gene. With the block optical content scoring (BOCS) algorithm, we used our gene block content measurements to correctly identify β -lactamase resistance in the *P. aeruginosa* pathogen. Early detection of this opportunistic, nosocomial pathogen and its resistance to β -lactam antibiotics is crucial for providing patients, who are often immunocompromised, with effective antibiotic regimens. The powerful BOCS algorithm detected the specific OXA β -lactamase (class D) gene from the MEGARes antibiotic resistance database of ~ 4000 genes without prior knowledge of pathogen or strain, and at merely $<15\%$ coverage of the gene. This can dramatically improve diagnosis times and costs compared to the culture-based detection strategies currently in place. Additionally, the nearly identical SERS signals seen between RNA and DNA provide evidence for directly translating the BOS/BOCS platform to the transcriptome. And detectable SERS variations due to chemically modified nucleobases such as 5-methylcytosine indicate the potential for epigenomic analyses. These results have demonstrated the use of the BOS/BOCS platform as a diagnostic optical sequencing technique, with extensions in genetics, transcriptomics, and epigenomics. Given the versatility of simple optical measurements and applicability to a wide range of biomarker databases, this platform has the potential for truly rapid and inexpensive, broad-spectrum diagnostics for personalized medicine.

Experimental

Materials

All custom DNA oligomers (10-mer calibration and gene blocks, Tables 1 and 2) and the custom homologous RNA oligomer (7-mer poly(rA)₇) were ordered from Integrated DNA Technologies, Inc. (IDT). The custom homologous 5-methylcytosine oligomer (5-mer poly(5mC)₅) was ordered from Gene Link. All oligomers were stored at -20°C or -80°C prior to use. Spermine tetrahydrochloride (99%), polyethyleneimine (PEI, branched, M.W. 10,000, 99%), and silver nitrate (99.9+%) were purchased from Alfa Aesar, and sodium borohydride from Fisher Chemical. All water used for synthesis, dilutions, and SERS measurements was ultrapure deionized (DI) water (Barnstead Thermolyne NANOpure Diamond purification system, water resistivity > 18 MΩ-cm).

Ag NP synthesis

The synthesis protocol was adapted from van Lierop et al.³⁹ Prior to synthesis, all glass vials were left to soak in PEI solution (0.4% v/v) overnight followed by extensive rinsing with ultrapure DI water. For Ag NPs, 40 µL silver nitrate solution (0.5 M) and 14 µL spermine tetrahydrochloride solution (0.1 M) were mixed with 20 mL ultrapure DI water and stirred for 20-30 min in dark. After 20-30 min, 500 µL of sodium borohydride solution (0.01 M) was spiked into the mixture (with continued stirring for 5-10 min). Ag NP colloids were allowed to sit overnight in dark (at room temperature), and the sediment at the bottom of the vial was then discarded.

Sample preparation

Samples for SERS measurements were prepared by mixing 5 µL of DNA oligomer solution (100 µM in ultrapure DI water) with 495 µL of Ag NP colloidal solution, for a final DNA concentration of ~1 µM. The DNA-Ag NP mixture was allowed to equilibrate for 2-3 h, followed by a quick sonication prior to measuring.

UV-vis characterization

Ag NPs and their interaction with DNA were characterized with a VWR UV-1600PC spectrophotometer using 1 cm path length glass cuvettes. An LSPR peak at ~392 nm was observed in the extinction spectrum for the blank Ag NP colloidal solution, with significant red shift seen upon the addition of DNA (Figure 1a).

SERS measurements

SERS measurements were collected with a 532 nm 5 mW laser (Changchun New Industries Optoelectronics Tech Co., Ltd.) focused on the colloidal sample (500 µL volume) through an Olympus BX51 microscope with 5X 0.10 NA objective, and spectra were collected with a Princeton Instruments imaging spectrometer. For the calibration and gene blocks, five and three technical replicates were measured respectively under the same experimental conditions (30 s exposure time, 10 accumulations).

SERS signal processing

Raw SERS spectra were processed by (1) Removing cosmic ray spikes in the spectrum, (2) 12-point average smoothing (± 6 in each left/right direction), (3) Correcting small inconsistencies in the 1089 cm^{-1} PO_2^- normalization peak. See Figure S2 for more details.

SERS signal normalization

Processed SERS spectra were normalized by (1) Finding all baseline points along the spectrum between which linear fits are constructed, (2) Subtracting the linear fits between the baseline points, (3) Normalizing by the 1089 cm^{-1} PO_2^- peak intensity. See Figure S3 for more details.

BOCS algorithm

The BOCS algorithm, with details on the mapping and scoring aspects, was extensively described previously by Korshoj and Nagpal.²⁴ The full code and databases to run simulations of BOCS can be found online at <https://github.com/lkorshoj/Block-Optical-Content-Scoring>.

Conflicts of interest

There are no conflicts to declare.

Acknowledgments

This work was supported by W.M. Keck Foundation, and partial support through National Science Foundation Soft Materials MRSEC at the University of Colorado through NSF Award DMR 1420736. L.E.K. acknowledges financial support from National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1650115.

References

- 1 M. A. Hamburg and F. S. Collins, *N. Engl. J. Med.*, 2010, **363**, 301–304.
- 2 M. U. Ahmed, I. Saaem, P. C. Wu and A. S. Brown, *Crit. Rev. Biotechnol.*, 2014, **34**, 180–196.
- 3 L. Ventola, *Pharm. Ther.*, 2015, **40**, 277–283.
- 4 T. U. Berendonk, C. M. Manaia, C. Merlin, D. Fatta-Kassinos, E. Cytryn, F. Walsh, H. Bürgmann, H. Sørum, M. Norström, M. N. Pons, N. Kreuzinger, P. Huovinen, S. Stefani, T. Schwartz, V. Kisand, F. Baquero and J. L. Martinez, *Nat. Rev. Microbiol.*, 2015, **13**, 310–317.
- 5 C. Kang, S. Kim, W. B. Park, K. Lee, H. Kim, E. Kim, M. Oh and K. Choe, *Antimicrob. Agents Chemother.*, 2005, **49**, 760–766.
- 6 T. H. McConville, S. B. Sullivan, A. Gomez-Simmonds, S. Whittier and A.-C. Uhlemann, *PLoS One*, 2017, **12**, e0186195.
- 7 O. Zarkotou, S. Pournaras, P. Tselioti, V. Dragoumanos, V. Pitiriga, K. Ranellou, A. Prekates, K. Themeli-Digalaki and A. Tsakris, *Clin. Microbiol. Infect.*, 2011, **17**, 1798–

- 1803.
- 8 A. Kumar, D. Roberts, K. E. Wood, B. Light, J. E. Parrillo, S. Sharma, R. Suppes, D. Feinstein, S. Zanotti, L. Taiberg, D. Gurka, A. Kumar and M. Cheang, *Crit. Care Med.*, 2006, **34**, 1589–1596.
- 9 T. P. Lodise, P. S. Mckinnon, L. Swiderski and M. J. Rybak, *Clin. Infect. Dis.*, 2003, **36**, 1418–1423.
- 10 S. Vergara-Lopez, M. C. Dominguez, M. C. Conejo, A. Pascual and J. Rodriguez-Bano, *J. Hosp. Infect.*, 2015, **89**, 123–131.
- 11 D. J. Diekema and M. A. Pfaller, *Clin. Infect. Dis.*, 2013, **56**, 1614–1620.
- 12 S. Kalsi, M. Valiadi, M. N. Tsaloglou, L. Parry-Jones, A. Jacobs, R. Watson, C. Turner, R. Amos, B. Hadwen, J. Buse, C. Brown, M. Sutton and H. Morgan, *Lab Chip*, 2015, **15**, 3065–3075.
- 13 B. Strommenger, C. Kettlitz, G. Werner and W. Witte, *J. Clin. Microbiol.*, 2003, **41**, 4089–4094.
- 14 M. Doumith, M. Day, H. Ciesielczuk, R. Hope, A. Underwood, R. Reynolds, J. Wain, D. M. Livermore and N. Woodford, *J. Clin. Microbiol.*, 2015, **53**, 160–166.
- 15 L. B. Harrison and N. D. Hanson, *Antimicrob. Agents Chemother.*, 2017, **61**, 1–8.
- 16 M. H. Kabir, D. Meunier, K. L. Hopkins, C. G. Giske and N. Woodford, *J. Antimicrob. Chemother.*, 2016, **71**, 1213–1216.
- 17 P. Bogaerts, S. Oueslati, D. Meunier, C. Nonhoff, S. Yunus, M. Massart, O. Denis, N. Woodford, K. L. Hopkins, T. Naas, L. Dortet, T.-D. Huang and Y. Glupczynski, *Sci. Rep.*, 2017, **7**, 9937.
- 18 V. Perreten, L. Vorlet-fawer, P. Slickers, R. Ehricht, P. Kuhnert and J. Frey, *J Clin Microbiol*, 2005, **43**, 2291–2302.
- 19 C. Strauss, A. Endimiani and V. Perreten, *J. Microbiol. Methods*, 2015, **108**, 25–30.
- 20 I. E. Tothill, *Semin. Cell Dev. Biol.*, 2009, **20**, 55–62.
- 21 W. A. Gahl, T. C. Markello, C. Toro, K. F. Fajardo, M. Sincan, F. Gill, H. Carlson-Donohoe, A. Gropman, T. M. Pierson, G. Golas, L. Wolfe, C. Groden, R. Godfrey, M. Nehrebecky, C. Wahl, D. M. D. Landis, S. Yang, A. Madeo, J. C. Mullikin, C. F. Boerkoel, C. J. Tifft and D. Adams, *Genet. Med.*, 2012, **14**, 51–59.
- 22 R. B. Ramoni, J. J. Mulvihill, D. R. Adams, P. Allard, E. A. Ashley, J. A. Bernstein, W. A. Gahl, R. Hamid, J. Loscalzo, A. T. McCray, V. Shashi, C. J. Tifft and A. L. Wise, *Am. J. Hum. Genet.*, 2017, **100**, 185–192.
- 23 D. M. Sagar, L. E. Korshoj, K. B. Hanson, P. P. Chowdhury, P. B. Otoupal, A. Chatterjee and P. Nagpal, *Small*, 2018, **14**, 1703165.
- 24 L. E. Korshoj and P. Nagpal, *Comput. Biol. Med.*, 2019, **110**, 196–206.
- 25 P. Nagpal, N. C. Lindquist, S.-H. Oh and D. J. Norris, *Science*, 2009, **325**, 594–597.
- 26 J. Morla-Folch, H. N. Xie, P. Gisbert-Quilis, S. G. De Pedro, N. Pazos-Perez, R. A. Alvarez-Puebla and L. Guerrini, *Angew. Chemie - Int. Ed.*, 2015, **127**, 13854–13858.
- 27 L. Xu, Z. Lei, J. Li, C. Zong, C. J. Yang and B. Ren, *J. Am. Chem. Soc.*, 2015, **137**, 5149–

- 5154.
- 28 S. Lal, N. K. Grady, J. Kundu, C. S. Levin, J. B. Lassiter and N. J. Halas, *Chem. Soc. Rev.*, 2008, **37**, 898–911.
- 29 L. Guerrini, Ž. Krpetić, D. Van Lierop, R. A. Alvarez-Puebla and D. Graham, *Angew. Chemie - Int. Ed.*, 2015, **127**, 1160–1164.
- 30 A. Barhoumi, D. Zhang, F. Tam and N. J. Halas, *J. Am. Chem. Soc.*, 2008, **130**, 5523–5529.
- 31 H. Deng, V. A. Bloomfield, J. M. Benevides and G. J. Thomas, Jr., *Biopolymers*, 1999, **50**, 656–666.
- 32 M. Shanmugasundaram and M. Puranik, *J. Raman Spectrosc.*, 2009, **40**, 1726–1748.
- 33 J. S. Singh, *J. Mol. Struct.*, 2008, **876**, 127–133.
- 34 J. M. Benevides, S. A. Overman and G. J. Thomas, Jr., *J. Raman Spectrosc.*, 2005, **36**, 279–299.
- 35 N. C. Lindquist, P. Nagpal, A. Lesuffleur, D. J. Norris and S. H. Oh, *Nano Lett.*, 2010, **10**, 1369–1373.
- 36 N. C. Lindquist, P. Nagpal, K. M. McPeak, D. J. Norris and S.-H. Oh, *Reports Prog. Phys.*, 2012, **75**, 036501.
- 37 S. Goodwin, J. D. McPherson and W. R. McCombie, *Nat. Rev. Genet.*, 2016, **17**, 333–351.
- 38 J. Bao and M. G. Bawendi, *Nature*, 2015, **523**, 67–70.
- 39 D. van Lierop, K. Zeljka, L. Guerrini, I. A. Larmour, J. A. Dougan, K. Faulds and D. Graham, *Chem. Commun.*, 2012, **48**, 8192–8194.
- 40 J. Morla-Folch, R. A. Alvarez-Puebla and L. Guerrini, *J. Phys. Chem. Lett.*, 2016, **7**, 3037–3041.
- 41 J. Morla-folch, H. Xie, R. A. Alvarez-puebla and L. Guerrini, *ACS Nano*, 2016, **10**, 2834–2842.
- 42 J. Morla-Folch, P. Gisbert-Quilis, M. Masetti, E. Garcia-Rico, R. A. Alvarez-Puebla and L. Guerrini, *Angew. Chemie - Int. Ed.*, 2017, **56**, 2381–2385.
- 43 M. Esteller, *Hum. Mol. Genet.*, 2007, **16**, R50–R59.
- 44 S. Sharma, T. K. Kelly and P. A. Jones, *Carcinogenesis*, 2010, **31**, 27–36.
- 45 D. M. Livermore, *Antimicrob. Resist.*, 2002, **34**, 634–640.
- 46 A. Potron, L. Poirel and P. Nordmann, *Int. J. Antimicrob. Agents*, 2015, **45**, 568–585.
- 47 L. Käll, J. D. Storey, M. J. MacCoss and W. S. Noble, *J. Proteome Res.*, 2008, **7**, 29–34.
- 48 D. N. Perkins, D. J. C. Pappin, D. M. Creasy and J. S. Cottrell, *Electrophoresis*, 1999, **20**, 3551–3567.
- 49 A. I. Nesvizhskii, O. Vitek and R. Aebersold, *Nat. Methods*, 2007, **4**, 787–797.
- 50 H. Li and N. Homer, *Brief. Bioinform.*, 2010, **11**, 473–483.
- 51 K. Reinert, B. Langmead, D. Weese and D. J. Evers, *Annu. Rev. Genomics Hum. Genet.*, 2015, **16**, 133–151.
- 52 S. M. Lakin, C. Dean, N. R. Noyes, A. Dettenwanger, A. S. Ross, E. Doster, P. Rovira, Z. Abdo, K. L. Jones, J. Ruiz, K. E. Belk, P. S. Morley and C. Boucher, *Nucleic Acids Res.*,

- 2017, **45**, D574–D580.
- 53 C. K. Stover, X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrenner, M. J. Hickey, F. S. L. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. W. Hancock, S. Lory and M. V Olson, *Nature*, 2000, **406**, 959–964.
- 54 X. Didelot, R. Bowden, D. J. Wilson, T. E. A. Peto and D. W. Crook, *Nat. Rev. Genet.*, 2012, **13**, 601–612.
- 55 S. R. Leopold, R. V Goering, A. Witten, D. Harmsen and A. Mellmann, *J. Clin. Microbiol.*, 2014, **52**, 2365–2370.
- 56 A. Kilianski, J. L. Haas, E. J. Corriveau, A. T. Liem, K. L. Willis, D. R. Kadavy, C. N. Rosenzweig and S. S. Minot, *Gigascience*, 2015, **4**, 1–8.
- 57 P. M. Ashton, S. Nair, T. Dallman, S. Rubino, W. Rabsch, S. Mwaigwisya, J. Wain and J. O’Grady, *Nat. Biotechnol.*, 2015, **33**, 296–300.
- 58 K. Schmidt, S. Mwaigwisya, L. C. Crossman, M. Doumith, D. Munroe, C. Pires, A. M. Khan, N. Woodford, N. J. Saunders, J. Wain, J. O’Grady and D. M. Livermore, *J. Antimicrob. Chemother.*, 2017, **72**, 104–114.