Virtual microfluidics for digital quantification and single-cell sequencing

Liyi Xu1,2, Ilana I Brito1–3, Eric J Alm1–3 & Paul C Blainey1,2

We have developed hydrogel-based virtual microfluidics as a simple and robust alternative to complex engineered microfluidic systems for the compartmentalization of nucleic acid amplification reactions. We applied in-gel digital multiple displacement amplification (dMDA) to purified DNA templates, cultured bacterial cells and human microbiome samples in the virtual microfluidics system, and demonstrated whole-genome sequencing of single-cell MDA products with excellent coverage uniformity and markedly reduced chimerism compared with products of liquid MDA reactions.

Applications from microbial genome discovery to biomedicine are driving the broader application of high-throughput analyses of nucleic acids at the level of single molecules and single cells. Assays involving highly parallel clonal nucleic acid amplification are needed for accurate copy-number analysis with absolute quantification of DNA sequences and fragments in genomics and prenatal diagnostics. Such assays are also needed to overcome nonspecific background for the detection of rare sequence targets in microbial communities and blood-plasma-based diagnostics. In the burgeoning field of single cell analysis, high-throughput and high-fidelity whole-genome and whole-transcriptome amplification (WGA and WTA) reactions are needed to produce sufficient material for sequence library construction to support the discovery and validation of new genomes and the analysis of genomic and functional heterogeneity.

A variety of approaches have been explored for compartmentalization across a large number of discrete reactors, including SBS plates, high-density microfluidic arrays, lab-on-chip systems and microdroplet systems (Supplementary Table 1 and Fig. 1a). However, they require complex instrumentation and microfabricated consumables that hinder broad deployment. An ideal platform should resist external contaminants and cross-compartment mixing, exhibit high throughput in small reaction volumes, be stable under temperature change, and allow facile addition and removal of reagents and samples. Finally, it should generate high-quality amplified products and minimize biases and artifacts such as chimeric fragments—commonly formed in PCR, WGA and WTA—that can severely impact results.

Inspired by earlier work on culturing microbes in hydrogels, we tested dMDA of purified, diluted lambda phage DNA in the hydrogel format (Fig. 1b–d, Supplementary Fig. 1 and Online Methods). Our estimate of 10 pg MDA product per cluster (Supplementary Fig. 2) suggests that we approached the discovery and validation of new genomes and the analysis of genomic and functional heterogeneity.

MDA is a popular amplification method for single-cell genome sequencing. To evaluate the virtual microfluidics concept for WGA, we tested dMDA of purified, diluted lambda phage DNA in the hydrogel format (Fig. 1b–d, Supplementary Fig. 1 and Online Methods). Our estimate of 10 pg MDA product per cluster (Supplementary Fig. 2) suggests that we approached the discovery and validation of new genomes and the analysis of genomic and functional heterogeneity.

In the virtual microfluidics system, reactions can be controlled (Fig. 1d), and parameters to test how in situ single-molecule MDA reactions can be controlled (Fig. 1d and Supplementary Figs. 4 and 5), observing that the smaller pore sizes in higher-density gels limited the spread of DNA products.

We then applied digital MDA at the single-cell level using the virtual microfluidics system (see Online Methods). Individual log-phase Escherichia coli cells could be identified in the hydrogel by fluorescence microscopy (Fig. 1e). We lysed the embedded cells by heat treatment and carried out MDA on the denatured genomic DNA, observing the appearance of MDA clusters at the reaction endpoint (Fig. 1e).

1Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. 2The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. 3The Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

Correspondence should be addressed to P.C.B. (pblainey@mit.edu).

Received 16 February; accepted 27 June; published online 1 August 2016; DOI:10.1038/NMETH.3955
Next, we tested the potential of virtual microfluidics to support single-cell shotgun genome sequencing (Fig. 2). We mixed log-phase *E. coli* (BL21) and *Staphylococcus aureus* subspp. *aureus* (RN6390/8325; *S. aureus*) strains at about 200,000 cells per mL and embedded the cells in a 300-µm-thick PEG hydrogel. The mixed-input approach powered the detection of cross-contamination between samples and contaminants (including *E. coli* DNA) from other sources. Embedded cells were lysed by enzymatic and heat treatment, and *MDA* reagents were introduced by diffusion into the gel. Gel punches (80 total, 60 nL each) were recovered manually in a grid pattern (Fig. 2a), and each was reamplified to an overall 109- to 1010-fold amplification in digital MDA in PEG hydrogel. Top, time-lapse diffusion to compartmentalize templates into discrete physical boundaries. Bottom, DNA cluster number increases with template concentration. FOV, 650- × 650-nm field of view. NTC, no–DNA template control. (c) Calibration curve illustrates linear relationship between template concentration and cluster number (n = 2 or 3 FOV at each concentration). (d) MDA cluster size is correlated with gel weight percentage and affected by acrydite-modified hexamer anchorage. (e) Hydrogel-encapsulated *E. coli* express GFP and exclude SYTOX Orange before lysis. SYTOX Orange staining reveals product clusters after MDA. Top images are from the same field of view. NTC, MDA control lacking *E. coli*.

The superior coverage performance in *S. aureus* may be attributable to the lower GC content of *S. aureus* (33%) compared with *E. coli* (51%), better accessibility (deproteination) of the genome after lysis and/or higher average genome equivalents per cell in *S. aureus* due to cell cycle dynamics. One *E. coli* outlier library showed extremely poor genome coverage. This library had low complexity (37% duplicate reads), which pointed to poor library quality rather than MDA as the cause for low genome coverage.

To rigorously evaluate sequence coverage distribution, we calculated the Gini index (a measure of inequity ranging from 0 to 1) for each of our single-cell data sets and previously published single-cell *E. coli* liquid MDA data sets for which raw read data were available and fold-amplification was known17 (Fig. 2d). Coverage uniformity in our single-cell punch samples compared favorably with published single-cell data sets at similar amplification gain (Supplementary Note 4).

We also analyzed chimeric reads (Supplementary Note 3), which occur with high frequency in MDA by a cross-priming mechanism33 and directly confound de novo assembly, analysis of rearrangements and mapped read counting. Our single-cell data sets contained about 0.5% chimeric reads, approximately five-fold lower than published short-read data sets produced using liquid single-cell MDA samples (Fig. 2e and Supplementary Table 4). Chimeric reads spanning more than 10 kb of the template are even less frequent (about 0.1%, Supplementary Fig. 8), raising the possibility of extracting trustworthy long-range information from single-cell MDA samples using long-read sequencing and other technologies. This dramatic reduction in chimeric reads can be explained by the local isolation of MDA intermediates (due to restricted diffusion) that prevents cross-priming. It may be that (at least) all of the chimeras we observed were generated during the liquid-phase secondary amplifications in our workflow. These results suggest that it is beneficial to run MDA in PEG hydrogels for all applications at all scales.

Next, we used virtual microfluidics for single-cell genome sequencing of human stool samples from the Fiji Community Microbiome Project (FijiCOMP; see Online Methods). The FijiCOMP
samples contain a vast uncharacterized diversity of microbial species that differ from the microbiome of Western subjects. We processed 421 hydrogel punch samples and compared the distribution of detected organisms with the distribution observed from shotgun metagenomic profiling, finding that both approaches recovered the same top microbial families (Fig. 3a and Supplementary Table 8). Interestingly, the second most abundant microbial family found in the single-cell data set, the Succinivibrionaceae, was not initially detected (but was later confirmed) in the shotgun data using standard taxonomic assignment methods such as MetaPhlAn34. This discrepancy highlights the importance of unbiased approaches like single-cell analysis for organisms that are poorly represented in reference databases.

We carried out de novo assembly of the single-cell data sets and assigned taxonomy to ribosomal gene sequences and 31 ‘single-copy’ bacterial marker genes (at the family level)35 (Supplementary Table 9). This enabled us to make crude assessments of genome coverage and sample purity in the FijiCOMP single-cell data sets, for which we lack strain-specific bona fide reference sequence. Of the 293 assemblies (up to 12 Mbp), we classified 117 as single-amplified genomes with assembly size greater than 100 kb and strong enrichment of sequences from a single taxonomy (Fig. 3b–d). Overall, the data quality observed from these human microbiome bacteria was consistent with the results of our studies with lab-cultured Gram-negative and Gram-positive samples and demonstrates the applicability of the hydrogel method to real-world samples, including lysis and amplification of a variety of naturally occurring microbes.

Virtual microfluidics offers a broadly applicable method for high-throughput digital assays and single-cell analysis that does not require special equipment or environmental controls. Our implementation is technically distinct from conventional microfluidic approaches and presents a distinct set of advantages and limitations. For sensitive applications, extra measures against contamination (such as use of a laminar flow cabinet, separate preamplification and postamplification laboratories and equipment) can further reduce contamination. The primary throughput limitation in this initial demonstration is the 60-μL punch volume, which restricts the number of subsamples that can be obtained from a single hydrogel. Throughput could be increased by using a thinner gel with more surface area and/or reducing punch size. The fraction of retrieved samples containing a single-cell WGA reaction product can be improved by using image data to guide product retrieval. The thin hydrogel format affords excellent physical access for imaging, equipment and reagents, which enables an assortment of subsampling approaches including punch and pickers, localized hydrogel dissolution and localized affinity tagging or barcoding. Barcoding approaches could conceivably enable retrieval of all amplified products en masse while allowing in silico demultiplexing to sort sequence reads according to cell of origin (Supplementary Note 5).

Besides reducing the production of chimeras in MDA, the physical characteristics of the hydrogel environment may enhance coverage extent and uniformity from WGA and WTA samples through the self-limiting reactivity within each virtual compartment, similar to a recently reported emulsion approach10. In addition, the straightforward addition and removal of reagents to and from product clusters en masse and excellent optical
access make the virtual microfluidics system ideally suited for rare-cell assays that incorporate in situ cell or product clustering (Supplementary Note 5). We expect that virtual microfluidics will find application as a low-cost digital assay platform and as a high-throughput platform for single-cell sample preparation.

METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. Raw sequencing data on E. coli and S. aureus are accessible at the NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA279815 with BioSample accession numbers SAMN03451478–SAMN03451501. FijiCOMP metagenomic reads can be found under BioProject accession number PRJNA217052 with the BioSample accession numbers SRX345831, SRX344363, SRX344765, SRX343094, SRX344442, SRX344605, SRX343839, SRX343780, SRX345901, SRX344600, SRX343866, SRX343411, SRX344189, SRX343430, SRX346966, SRX345329, SRX343800 and SRX344616. FijiCOMP virtual microfluidics 117 single cells are accessible with the BioSample accession numbers SAMN04461233–SAMN04461349.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS
The authors thank S. Kim, D. Feldman and A. Kulesza for advice on bioinformatics, microscopy and image analysis; N. Ranu and the Hung lab (Broad Institute) for bacterial samples; L. Griffith, G. Lagoudas, J. Borrajo and L. Morinishi for helpful discussions; and members of the Griffth lab (MIT), especially C. Chopko, J. Valdez and H. Lee, for hydrogel expertise. This work was supported in part by a Lawrence Summers Fellowship from the Broad Institute (L.X.), a Career Award at the Scientific Interface from the Burroughs Welcome Fund (P.C.B.) and grants from the Center for Microbiome Informatics and Therapeutics at MIT; a National Human Genome Research Institute, grant number U54HG003067, to the Broad Institute; the Center for Environmental Health Sciences at MIT and the Fijian Ministry of Health. The Broad Institute and MIT may seek to commercialize aspects of this work, and related applications for intellectual property have been filed.

AUTHOR CONTRIBUTIONS P.C.B. and L.X. conceived the concept for this study. L.X. designed and implemented experiments and conducted data analysis. I.L.B. and E.J.A. provided Fiji microbiome samples. I.L.B. conducted analysis of gut microbiota data. L.X., P.C.B. and I.L.B. wrote and all authors approved the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

![Figure 3](https://example.com/figure3.png)

Figure 3 | Fiji microbiome project (FijiCOMP) single-cell whole-genome sequencing. (a–c) Results for 117 single-cell data sets from five donor individuals. (a) Distribution of top ten microbial families from single-cell assemblies and metagenome shotgun sequencing. Samples were weighted according to the number of single cells analyzed (Supplementary Table 8). (b) De novo assemblies from single-cell sequencing data ranged from 100 kbp to 2 Mbp. Line indicates the mean assembly size. (c) The total number of AMPHORA genes is nearly equal (dotted line) to the number of AMPHORA genes from the top phylogenetic family for each sample, supporting the assertion that each data set arises from an individual bacterial cell. A Gaussian-distributed random jitter (μ = 0, σ² = 0.1) was applied to enhance visualization. (d) Overview of FijiCOMP single-cell hydrogel samples.

| Accession family | Abundance (%) |
|------------------|---------------|
| Acidaminococcaceae | 33.3 |
| Bacteroidaceae     | 33.3 |
| Clostridaceae      | 33.3 |
| Enterobacteriaceae | 33.3 |
| Eubacteriaceae     | 33.3 |
| Lachnospiraceae    | 33.3 |
| Succinivibrionaceae| 33.3 |
| Prevotellaceae     | 33.3 |

| Accession number | Sample categorizations | Number of punches |
|------------------|------------------------|-------------------|
| PRJNA217052      | Low read counts; no assembly | 16 |
| PRJNA217052      | Laboratory contamination (e.g., col. P. aeruginosa) | 29 |
| PRJNA217052      | Human cell sequences amplified | 3 |
| PRJNA217052      | No phylogenetic markers | 80 |
| PRJNA217052      | Enrichment of multiple taxonomies from assemblies | 108 |
| PRJNA217052      | Assembly <100 kb | 68 |

| Accession family | Total AMPHORA genes |
|------------------|----------------------|
| Acidaminococcaceae | 33.3 |
| Bacteroidaceae     | 33.3 |
| Clostridaceae      | 33.3 |
| Enterobacteriaceae | 33.3 |
| Eubacteriaceae     | 33.3 |
| Lachnospiraceae    | 33.3 |
| Succinivibrionaceae| 33.3 |
| Prevotellaceae     | 33.3 |

3. Wang, J., Fan, H.C., Behr, B. & Quake, S.R. Cell 150, 402–412 (2012).
4. Nugget, J.F., Cohen, S. & Fay, C.A. Clin. Chem. 61, 79–88 (2015).
5. Sykes, P.J. et al. Biotechniques 13, 444–448 (1992).
6. Vogelstein, B. & Kinzler, K.W. Proc. Natl. Acad. Sci. USA 96, 9236–9241 (1999).
7. Blainey, P.C. & Quake, S.R. Nucleic Acids Res. 39, e19 (2011).
8. Raghunathan, A. et al. Appl. Environ. Microbiol. 71, 3342–3347 (2005).
9. Zhang, K. et al. Nat. Biotechnol. 24, 680–686 (2006).
10. Fu, Y. et al. Proc. Natl. Acad. Sci. USA 112, 11923–11928 (2015).
11. Pamp, S.J., Harrington, E.D., Quake, S.R., Relman, D.A. & Blainey, P.C. Genome Res. 22, 1107–1115 (2012).
12. Dodsworth, J.A. et al. Nat. Comm. 4, 1854 (2013).
13. Hess, M. et al. Science 331, 463–467 (2011).
14. Love, K.R., Bagh, S., Choi, J. & Love, J.C. Trends Biotechnol. 31, 280–286 (2013).
15. Thorsen, T., Maerkl, S.J. & Quake, S.R. Science 298, 580–584 (2002).
16. Landry, Z.C., Giovanonni, S.J., Quake, S.R. & Blainey, P.C. Methods Enzymol. 531, 61–90 (2013).
17. de Bourcy, C.F.A. et al. PLoS ONE 9, e105585 (2014).
18. Marcy, Y. et al. PLoS Genet. 3, 1702–1708 (2007).
19. Thorsen, T., Roberts, R.W., Arnold, F.H. & Quake, S.R. Phys. Rev. Lett. 86, 4163–4166 (2001).
20. Mazutis, L. et al. Nat. Protoc. 8, 870–891 (2013).
21. Hindson, C.M. et al. Nat. Methods 10, 1003–1005 (2013).
22. Morinishi, L.S. & Blainey, P.J. Vis. Exp. 103, e52925 (2015).
23. Blainey, P.C. FEMS Microbiol. Rev. 37, 407–427 (2013).
24. Podar, M., Keller, M. & Hugenholtz, P. in Uncultivated Microorganisms, Vol. 10 (ed. Epstein, S.S.) 241–256 (Springer, Berlin Heidelberg, 2009).
25. Mitra, R.D. & Church, G.M. Nucleic Acids Res. 27, e34 (1999).
26. Allen, L.Z. et al. PLoS ONE 6, e17722 (2011).
27. Rauber, G.P., Lutolf, M.P. & Hubbell, J.A. Biophys. J. 89, 1374–1388 (2005).
28. Wu, Y., Joseph, S. & Aluru, N.R. J. Phys. Chem. B 113, 3512–3520 (2009).
29. Phelps, E.A. et al. Adv. Mater. 24, 64–70 (2011).
30. Dean, F.B. et al. Proc. Natl. Acad. Sci. USA 99, 5261–5266 (2002).
31. Marcy, Y. et al. Proc. Natl. Acad. Sci. USA 104, 11889–11894 (2007).
32. Woyke, T. et al. PLoS ONE 6, e26161 (2011).
33. Lasken, R.S. & Stockwell, T.B. BMC Biotechnol. 7, 19 (2007).
34. Segata, N. et al. Nat. Methods 9, 811–814 (2012).
35. Wu, M. & Scott, A.J. Bioinformatics 28, 1033–1034 (2012).
ONLINE METHODS

Cross-linked PEG hydrogel formation. Hydrogel components, including four-arm PEG acrylate (molecular weight (MW) 10,000) and HS-PEG-SH (MW 3,400), were obtained from Laysan Bio. For every 25 µL of 10% (wt/v) cross-linked hydrogel, 1.6 mg of four-arm PEG acrylate and 1.1 mg of HS-PEG-SH were dissolved in pH 7.4 PBS (Invitrogen). The 25 µL hydrogel mixture was briefly vortexed and centrifuged to ensure mixing, and it was allowed to sit on bench for 10 min while the hydrogel components cross-linked through the reaction between the thiol and acrylate groups.

In-gel PCR. The primers (Supplementary Table 3) used for PCR on purified λ DNA (48 kbp, New England BioLabs (NEB)) were ordered through Integrated DNA Technologies (IDT). A 25 µL hydrogel PCR reaction consisted of 2 U of VentR(exo-) polymerase (NEB), 1× Thermopol Reaction Buffer (NEB), 0.4 mM dNTP (NEB), 1 µM Primers, 5% DMSO (Sigma), 0.5 mg/mL BSA (NEB), 1.6 mg four-arm PEG acrylate in PBS, 1.1 mg HS-PEG-SH in PBS and λ DNA template (NEB) of various concentrations. The 25 µL of the above components were loaded in a 9 mm x 9 mm frame-seal chamber (Bio-Rad). The following thermal protocol was run on an MJ Research PTC-100 twin tower thermal cycler: 30 °C for 30 min (gel polymerization), 98 °C for 3 min, 98 °C for 30 s, 57 °C for 30 s, 57 °C for 1 min for 40 to 60 cycles, 72 °C for 5 min and hold at 4 °C. The gel was stained with 500 nM SYTOX Orange nucleic acid dye (Invitrogen).

In-gel MDA. A 25 µL hydrogel MDA reaction consisted of 0.5 µL of REPLI-g sc Polymerase (Qiagen), 1× phi 29 buffer (NEB), 50 µM Random Hexamers (IDT; including two phosphorothioate bonds at 3’ terminus), 2.5% DMSO, 0.4 mM dNTP, 0.5 mg/mL BSA, 500 nM SYTOX Orange (Invitrogen) and denatured λ DNA. λ DNA was denatured (using alkaline buffer “D1”, Qiagen) and neutralized (buffer “N1”, Qiagen) according to Qiagen REPLI-G sc kit protocol before hydrogel encapsulation. All MDA and gel components, except polymerase and SYTOX Orange dye, were UV treated for 30 min using the Stratalinker UV crosslinking instrument (Stratagene) to render contaminating background DNA incompetent for MDA. The 25 µL MDA reaction mixture was loaded in a 9 mm x 9 mm frame-seal chamber (Bio-Rad, about 300 µm in height). The gel was sealed in the chamber with a plastic cover and maintained at 30 °C for 8 h or longer in the MJ Research PTC-100 twin tower thermal cycler. After the reaction, we imaged the gel using Nikon ECLIPSE Ti inverted microscope or Nikon ultra-fast laser scanning confocal microscope (MIT Koch Institute Microscopy Core Facility).

In-gel real-time dMDA. MDA hydrogel reactions were set up as described above and conducted at room temperature for 6 h on a Nikon ECLIPSE Ti Epi-Fuorescence Microscope excited with a Lumencor Spectra X light engine (Lumencor) with fluorescent emissions filtered through an SpGold filter (Semrock). MATLAB was used to capture time-lapse image stacks through a Nikon 20x 0.4 NA objective and Hamamatsu C11440 camera with 15 min intervals, 100 ms exposure time, and 10% Lumencor excitation power. All samples were stained with 500 nM SYTOX Orange. Each E. coli MDA cluster or mammalian cell image stack was cropped and processed as described below.

In-gel single-microbe MDA-cultured E. coli and S. aureus. Antibiotic resistant Staphylococcus aureus subsp. aureus (GFP) NCTC 8325 and Escherichia coli (RFP) BL21 strains were obtained as cryogenic stocks. For each culture, the frozen stock was incubated in 5 mL LB broth and cultured at 37 °C overnight. 10 µL of 25 mg/mL chloramphenicol was added to S. aureus culture and 5 µL of 50 mg/mL ampicillin was added to the E. coli culture. 50 µL and 20 µL of each overnight culture were added to fresh 5 mL LB broth with the respective antibiotic concentration. After 2 h incubation (to achieve exponential growth phase), 1 mL of each culture (O.D. 600 nm = 0.2) were centrifuged for 2 min at >10 krpm and the pellet was washed with 500 µL PBST (1% Tween-20) twice. The equal-ratio mixture of microbes was diluted to 206,000 cells/mL, and 1 µL of each was encapsulated in the same hydrogel sample to produce an average of less than 1 microbe per 500 µm diameter view. In addition to hydrogel MDA reaction mix described above, lysozyme (Sigma, final concentration 2.5 mg/mL) and lysostaphin (Sigma, final concentration 0.1 mg/mL) were added to the mix. The hydrogel was left at RT to let crosslink for 20 min and crosslinked hydrogels were incubated at 37 °C for 1 h for microbe lysis and heated to 95 °C for 5 min to denature genomic DNA before rapid quenching on ice. 1 µL of REPLI-g sc Polymerase (Qiagen) diluted in 2 µL water was then added on top of the hydrogel and allowed to diffuse into the gel. Next, the gel chamber was resealed and MDA was conducted for 10 h. After the MDA reaction, the sample was heated to 65 °C for 5 min to deactivate phi29 polymerase.

In-gel single-microbe MDA human gut microbiome samples. We received ethics approvals for human subjects research from the Columbia University IRB, Massachusetts Institute IRB, Broad Institute IRB, and two research ethics committees in Fiji: HRERC at CMNHS, FNU and FNRHerc at MoHFiij Ministry of Health. Fiji Community Microbiome Project (FijiCOMP) study participants from five agrarian villages within the Fiji Islands provided stool samples stored in 20% glycerol within 30 min of voiding and frozen at ~80 °C (five participants: M1.20, W2.21, WL26, W2.33 and M2.41) (Supplementary Table 8). 10 µL of thawed cells were resuspended in 500 µL PBST (0.1%). Samples were sonicated for 20 s and filtered through 35 µm nylon mesh and 5 µm membrane (Pall Corp.) to collect filtrate with a 500 µL PBST wash. Samples were further diluted 1:500 to ~1,200-fold in PBST to reach the final concentration of ~30 cells per µL. The diluted cell samples (2 µl) then underwent alkaline lysis (1.5 µL D2 buffer) for 15 min at room temperature, after which the solution was neutralized (1.5 µL stop solution). Hydrogel monomer mix (1.3 mg four-arm PEG acrylate and 0.9 mg SH-PEG-SH) and MDA master mix were pipetted gently down the wall of each sample tube. MDA master mix includes 1× phi 29 buffer (NEB), 50 µM random hexamers with two phosphorothioate bonds at 3’ terminus, 2.5% DMSO, 0.4 mM dNTP, 0.5 mg/mL BSA, 500 nM SYTOX Orange (Invitrogen) and 1 µL REPLI-g Sc Polymerase (Qiagen). Only gentle tapping was used to ensure reagent mixing in order to not disrupt the lysed and denatured microbial genomes. 25 µL of each microbial suspension was added into a frame-seal chamber, and the sealed chamber was incubated at 30 °C for 12 h, followed by 65 °C for 5 min.
Image acquisition and analysis. Z-stack images were taken by Nikon ultra-fast laser scanning confocal microscope with pinhole = 1.2, HV = 112, offset = 0, laser wavelength = 561 nm, laser power = 1.3 to 1.5, using a 20× objective on Galvano mode. Acquisition speed was 1 frame per s and z step size was 0.95 µm. On the inverted microscope, z-stack images were taken with the exposure time 100 ms, Lumencor excitation power 10%, binning size 2 and z step size 10 µm. Both z-stacks were first processed into max intensity projections in FIJI (FIJI is just imagej). Max projection tif files were then loaded into MATLAB. Background was obtained by applying a Gaussian filter of hsize 200 and sigma 50. All max projections were background subtracted and thresholded at 2–2.5× s.d. above the mean intensity. Cluster count, cluster area (radius) and cluster mean intensity were obtained with the bwconncomp and regionprops functions.

For whole-gel (25 µL, 9 mm × 9 mm) microbe density approximation, we imaged the gel with a 4x objective in a 5 × 5 grid with a 31% overlap. The 25 images were stitched using the FIJI stitching function. Fluorescent DNA clusters were counted, and only gels with the appropriate cluster range and dispersion (60 to 80 per gel) were selected for hydrogel cluster retrieval. Images of the sampled locations were acquired but not used to guide sampling, sample preparation or data analysis in this case.

MDA product cluster retrieval. In order to identify and retrieve a regular array of punches (not guided by cluster image data), we produced a tape stencil to guide the punch tool (Adhesive Applications High Tack Silicone Film Tape). We laser cut the double-sided tape with a 9 × 9 array of 500 µm diameter circles that has a center-to-center distance of 947 µm (Full Spectrum Laser LLC MLE-40). The tape stencil was applied on top of the frame seal plastic cover. The gel was peeled off the glass slide by allowing it to adhere to the plastic cover. The gel was then punched with a 1 mm diameter steel punch (MilliTek) and the microsamples collected in a 96 well lobind twinteck plate (Eppendorf). The steel punch was cleaned with bleach and 70% ethanol after each use.

Secondary liquid MDA and PCR screening, cultured E. coli and S. aureus. The retrieved hydrogel punch (approximately 0.06 µL of hydrogel and 10 pg of DNA if a cluster was captured) was dissolved and denatured in 1 µL of 1 M KOH with 0.1 mM EDTA and 0.1 M DTT at 72 °C for 10 min before neutralization in 1 µL stop solution (Qiagen REPLI-g single cell kit). The neutralized product was added to 12.5 µL REPLI-g sc reaction mix with 1 µL of phi29 polymerase. The secondary MDA reaction was incubated for 10 h before polymerase deactivation at 65 °C for 5 min. The DNA products from MDA were cleaned by the SPRI procedure in 1:8:1 beads to DNA volume (Beckman Coulter). Each sample was analyzed for the presence of S. aureus and E. coli marker loci by four sets of primers in standard qPCR reactions with Jumpstart Taq 2× ready mix (Sigma-Aldrich), 1× Evagreen (Biotium), 1× ROX (Invitrogen) and 1 µM primers in Stratagene M3005. Both melting curve analysis and agarose gel electrophoresis (not shown) were used to support the qPCR results.

Secondary in-gel MDA, human gut microbiome samples. Hydrogel punches (approximately 0.24 µL of hydrogel and 10 pg of DNA if a cluster was captured) were dissolved and denatured in 1 µL of 400 mM KOH with 0.1 mM EDTA and 0.1 M DTT at 72 °C for 10 min before neutralization in 1 µL stop solution (Qiagen REPLI-g single cell kit). The neutralized product was added to 8 µL hydrogel and MDA master mix to reach a final volume of 10 µL for second-round MDA reaction in hydrogel. The MDA reaction was incubated for 10 h at 30 °C before polymerase deactivation at 65 °C for 5 min. The 10 µL gel was dissolved with 10 µL 400 mM KOH for 5 min at 72 °C and then neutralized with 6.6 µL 2.5% acetic acid.

WGS library construction and sequencing. We quantified the purified MDA products using the Qubit/Quanti-IT HS assay (Thermo Fisher Scientific) and normalized samples to 5 ng/µL. All SPRI procedures were conducted on the Bravo robotic system (Agilent Technologies). 5 ng of purified DNA was then added to 1 µL of 5× tagmentation DNA buffer, 2 µL H2O and 1 µL Nextera Tagmentation DNA enzyme (Illumina). The mixture was first incubated at 58 °C for 10 min. With the addition of 0.5 µL of 1% SDS, it was then incubated at 68 °C for 10 min, 4 °C for 3 min and 25 °C for 3 min to stop the tagmentation reaction. Another SPRI cleanup was carried out, followed by PCR library barcoding using Index primer N7 and S5 (Illumina) with the thermal protocol: 72 °C for 3 min, 98 °C for 30 s, 12 cycles of 98 °C for 10 s, 60 °C for 30 s, 72 °C for 30 s and a 5 min final extension at 72 °C. Samples were barcoded uniquely in the PCR step using standardized custom sample barcodes (Broad Institute Genomics platform). The PCR products were purified with SPRI twice with 1:1 beads to DNA volume and library quantification was carried out with the Quanti-It assay (Thermo Fisher Scientific) and the KAPA library quantification kit (KAPA Biosystems). Library normalization and pooling were conducted on the Janus Mini Varispan workstation (PerkinElmer). For E. coli and S. aureus samples, an average of 0.7 million paired-end reads were allocated for each sample in a MiSeq 500 cycle v2 run (Illumina). For stool samples, about 1 M reads (>50×) were allocated to each sample on HiSeq 2500 2× 101/125 runs (Illumina). See Supplementary Notes and Figures for all details regarding sequencing analysis.

Code availability. All custom MATLAB functions, python code and shell scripts are included in the Supplementary Software zip file.