

A lab-on-chip for biothreat detection using single-molecule DNA mapping

Electronic Supplemental Material

Robert H. Meltzer, Jeffrey R. Krogmeier, Lisa W. Kwok, Richard Allen, Bryan Crane, Joshua W. Griffis, Linda Knaian, Nanor Kojanian, Gene Malkin, Michelle Nahas, Vyacheslav Papkov, Saad Shaikh, Kedar Vyavahare, Qun Zhong, Yi Zhou, Jonathan W. Larson, and Rudolf Gilmanshin

Correspondence should be addressed to R.H.M. rmeltzer@usgenomics.com

U.S. Genomics, 12 Gill Street, Suite 4700, Woburn MA, USA

Air sampling detector work-flow

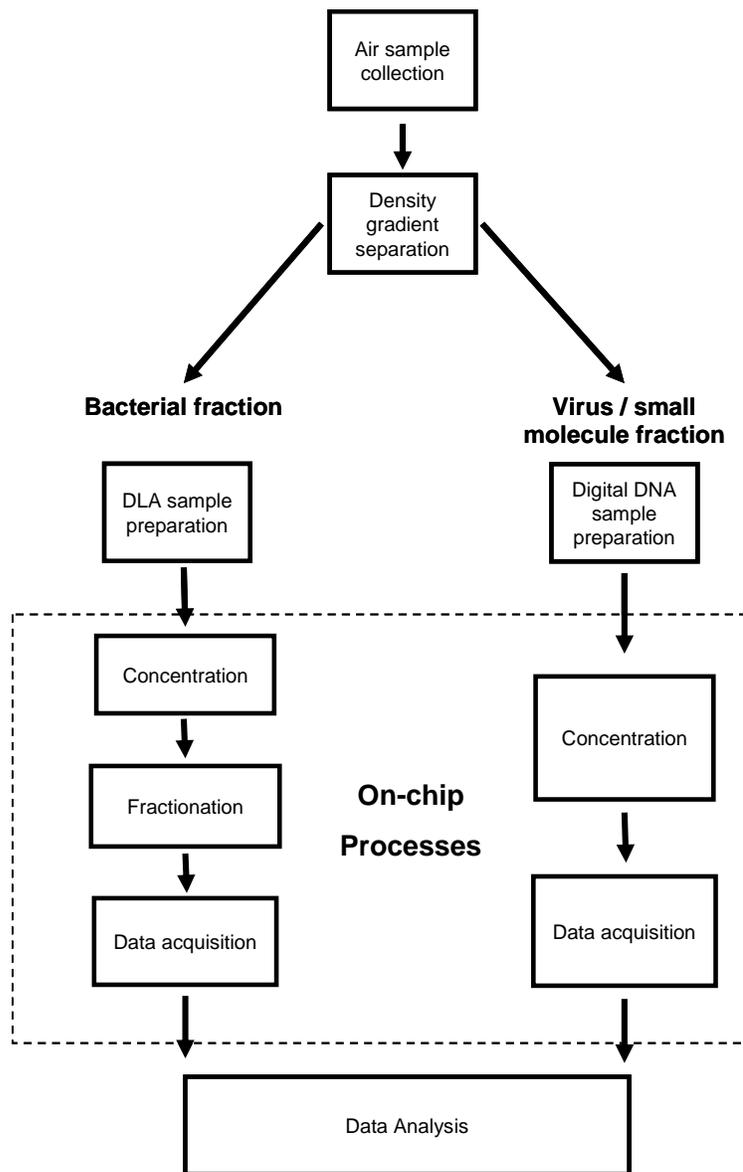


Figure S1. Work-flow for integrated air sampling detector. Processes within the dashed rectangle are accommodated on LOC-DLA and described in detail in this manuscript.

Bacterial culture and preparation

E. coli K12 MG1655 and *S. epidermidis* (ATCC 12228, NC_004461) were purchased from ATCC (American Type Culture Collection). For DNA sample preparation, a single colony of either bacterium was picked and cultured in 5 ml Luria-Bertani or trypticase soy broth, respectively. Samples were cultured overnight at 37° C with agitation. For detection of targets in complex biological background, 10^4 *E. coli* cells and 10^5 *S. epidermidis* cells were added to a frozen complex background mixture. This mixture consisted of *Brevibacterium epidermidis* 19.2%, *Burkholderia gladioli* 9.3%, *Bacillus muralis* 6.1%, *Corynebacterium ammoniagenes* 5.5%, *Flavobacterium johnsoniae* 8.6%, *Paracoccus denitrificans* 8.8%, *Rhizobium radiobacter* 12.4%, *Stenotrophomonas maltophilia* 15.4%, *Vibrio fischeri* 9.2% (by cell counting). The growth conditions of the background components are presented in Table S1. Aliquots of 1.9×10^6 cells (9.5 ng of DNA) of this mixture were prepared and frozen at -80° C for subsequent use. This mixture of bacteria was selected as a representative biological background as found in air samples, and consists primarily of 4 phyla; Actinobacteria, Bacteroides, Firmicutes, and Proteobacteria.

Microbe	Collection Type	Culture Broth/Agar	Growth Temp (°C)
<i>Bacillus muralis</i>	DSM 16288	Nutrient Agar	30
<i>Stenotrophomonas maltophilia</i>	ATCC 13637	Nutrient Agar	30
<i>Paracoccus denitrificans</i>	DSM 413	Tryptic Soy Agar	30
<i>Rhizobium radiobacter</i>	DSM 5172	Tryptic Soy Agar	30
<i>Burkholderia gladioli</i>	DSM 4285	Tryptic Soy Agar	30
<i>Flavobacterium johnsoniae</i>	ATCC 17061	Tryptic Soy Agar	30
<i>Vibrio fischeri</i>	ATCC 49387	Marine Agar	30
<i>Brevibacterium epidermidis</i>	DSM 20660	Coryne Broth	30
<i>Corynebacterium ammoniagenes</i>	DSM 20306	Coryne Broth	30

Table S1. Sources and growth conditions of the components of the model biological background of an air sample.

Fabrication of the lab-on-chip

The LOC-DLA device was fabricated in fused silica, with multiple feature depths in order to accommodate the required functionalities. Shallow features were deep reactive ion-etched (DRIE) to $1.0 \pm 0.1 \mu\text{m}$ or $2.0 \pm 0.2 \mu\text{m}$, while deep features were wet-etched to $20 \pm 2 \mu\text{m}$ in a fused silica wafer ($500 \pm 25 \mu\text{m}$ thickness). Holes (inner diameter 0.75 mm or 0.4 mm) were ultrasonically drilled through the wafer to enable access to the etched features. The chip was sealed by fusion bonding to a fused silica cover slip ($170 \pm 12.5 \mu\text{m}$ thickness).

The images in Figure S2 indicate the difference in the etching profile using DRIE and wet etching methods. 1, 2, and 3 μm deep features were generated using DRIE (Fig. S2a). The 3 μm deep features were generated by superposition of 1 μm and 2 μm etching steps. This method yields vertical walls. 20 μm deep features were generated using wet etching with a TiWAu mask (Fig. S2b). Again, complex features (i.e. 21 μm deep structures) were fabricated by superimposing DRIE and wet etch steps. Wet etching yields highly sloped walls, and therefore a trapezoidal channel cross section (Fig. S2c). This cross section was considered in the design and implementation of chip flushing and sample transfer protocols. The Navier-Stokes equations were implemented in COMSOL using a three-dimensional model of the trapezoidal 20 μm etch depth channel to simulate fluid velocity in HD flow (Fig. S2d). A significant velocity differential was observed, particularly at the acute angles near the interface between the etched wall and the top of the channel. These low-flow zones limit the rate of fluid exchange during chip fabrication, surface coating, and cleaning procedures.

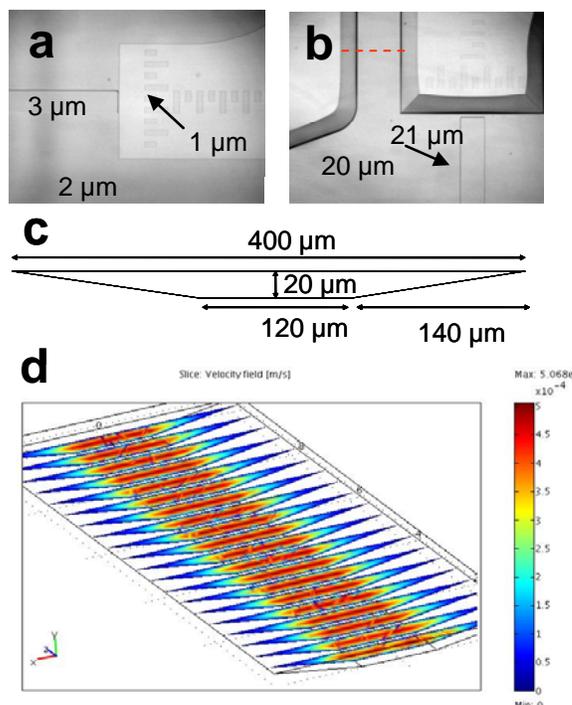


Figure S2. Etching profiles in fabricated microfluidic devices. a) Features generated with DRIE. b) Features generated with a combination of DRIE and wet etching. c) Cross-section of wet-etched 20 μm deep channel. Cross section marked in (b) with red dashed line d) Heat map of flow profile through trapezoidal cross-section 20 μm deep channel.

Manifold-bonded LOC-DLA

LOC-DLA chips were bonded to custom-designed polymethylmethacrylate (PMMA) manifolds (Eastern Plastics, Bristol, CT) using a UV-curable adhesive Dymax 140M (Dymax Corporation, Torrington, CT) (Fig. S3). The manifold incorporated features to allow for fluidic and electrical access to the chip, as well as precision alignment features to interface with the DNA stretching optics. The reverse face of the manifold features multiple press-fit PEEK tubes, threaded fittings, and threaded stoppers to allow for fluidic access and to control hydrodynamic siphoning in the device (Fig. S3b).

The manifolds also included 60 μm tall bosses around each access port to the chip. These features allowed the chip to be uniformly elevated above the manifold, thus permitting adhesive to fill the volume between manifold and microfluidic device through capillary action. The presence of bosses around each access port also protected against seepage of adhesive into the chip.

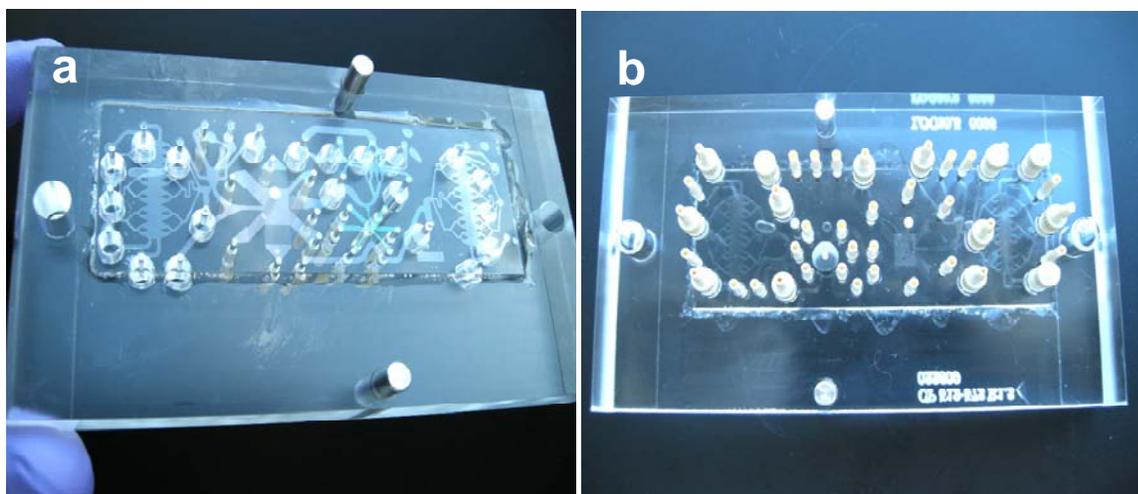


Figure S3. Manifold-bonded LOC-DLA device. a) LOC-DLA bonded to PMMA manifold. b) Reverse face of the manifold with fittings for fluidic access.

Fabrication of photopolymerized polyacrylamide gels

Photopolymerized polyacrylamide gel banks were fabricated *in situ* in LOC-DLA to serve as DNA concentrators. First, a solution of 30% acrylamide-bisacrylamide monomer mixture, 480 ppm hydroquinone, and 0.97% 2-hydroxy-2-methylpropiophenone, was injected into the microfluidic device. Pressure and vacuum were applied in sequences to the access ports to ensure that the solution uniformly filled the chip, paying close attention to the three areas in which gel banks were to be photopolymerized (Fig. 1a, features 2, 3, and 5). The fluid-filled microfluidic device was then mounted on a programmable XY scanning stage (Ludl Electronic Products Ltd., Hawthorne, NY) on an upright microscope (E800, Nikon Corporation, Tokyo, Japan) with a 100 W mercury arc lamp (Nikon Corporation, Tokyo, Japan) and a 335 nm excitation filter (Semrock, Inc., Rochester, NY). A square aperture (1 mm × 1 mm) was positioned in the back focal plane of the microscope and projected through a 4× (0.13 NA, 17.1 mm WD, Nikon Corporation, Tokyo, Japan) microscope objective lens. Using a low NA lens ensured that the light intensity was uniform across the aperture image.

To fabricate the photopolymerized gel, the aperture image was projected on the LOC-DLA concentrator feature and aligned to fiducial marks along the edges of each gel. A shutter (Ludl Electronic Products Ltd.) positioned in front of the light source was opened and the stage moved the chip at a fixed velocity past the aperture image with a uniform dwell time of the scanned aperture of 8 seconds, thus inducing the radical polymerization reaction. Once the polyacrylamide gels were polymerized, the microfluidic channels were rinsed with water followed by a thorough flush with TBE.

Electrical and pneumatic interface

Electrical and pneumatic controls interface with LOC-DLA using a “saddle” and “manifold” interface (Fig. S4). The manifold, as described in the text, consists of a PMMA mount directly bonded to the fused silica microfluidic chip. This component allows for access to the drilled holes on the chip through a series of either press-fit or screw-in PEEK fittings. The manifold also allows for precise alignment of the chip with the detection optics. The saddle consists of a series of buffer-filled wells with platinum electrodes leading to the high-voltage power supply and air-filled channels leading to pneumatics control (Fig S4a). Buffer-filled Tygon tubes provide electrical and pneumatic communication between the saddle and manifold. Through this arrangement, pneumatic and electrical control can be achieved simultaneously through multiple ports on the microfluidic device. Keeping the platinum electrodes remote from the chip limits the possibility of electrolytically generated bubbles to occlude the access ports or etched channels on the chip. The saddle also allows for fluid reservoirs to be filled to a uniform height, thus limiting hydrodynamic siphoning on LOC-DLA.

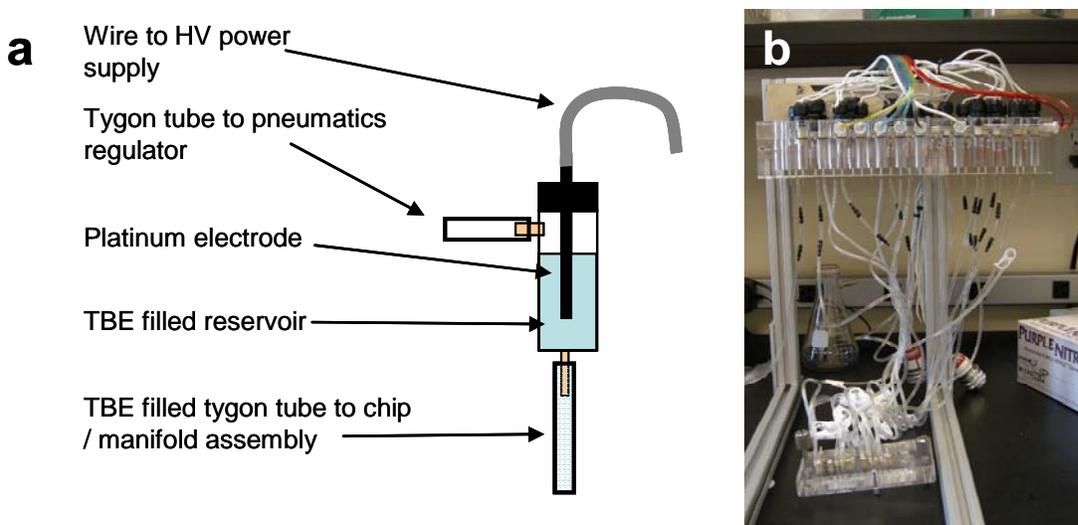
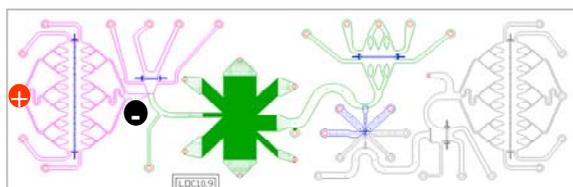


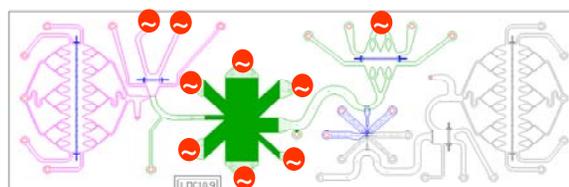
Figure S4. Remote electrical and pneumatic interface. a) Single unit of combined electrical and pneumatic interface. 23 similar wells are bundled into the integrated saddle shown in b). This interface provides pneumatic and electrical control of the LOC-DLA chip bonded to a manifold also shown in b).

LOC-DLA operation protocol

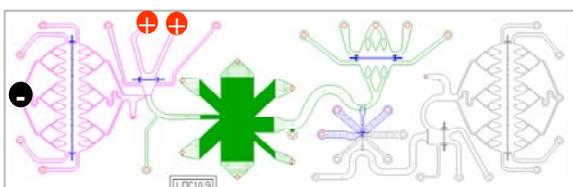
LOC-DLA operation requires several sequential electrical and pneumatic states. Here, a schematic representation of this sequence is shown for the DLA portion of LOC-DLA (Fig. S5). The positions of active positive and negative electrodes are indicated as red or black circles respectively.



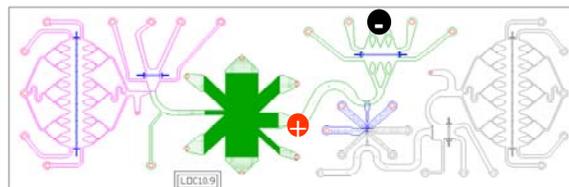
1) DNA transfer to gel 1: 25 μ A, 45 min



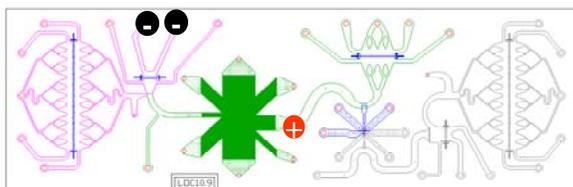
4) Prism operation: 80 V/cm 30° 450 ms, 0 V/cm 100 ms, 80 V/cm -90° 280 ms, the cycle repeats continuously for 45 minutes. Retained DNA accumulates simultaneously at Gel 3.



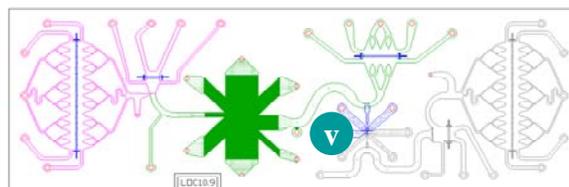
2) DNA transfer to gel 2: 3 μ A, 45 min



5) Gel 3 elution: 0.4 μ A, 4 min



3) DNA transfer toward prism: 1 μ A, 3 min



6) DLA acquisition: 10 psi vacuum 60 min

Figure S5. Typical operation protocol for LOC-DLA. Electrode positions, transfer currents, and transfer durations are shown for each of the sequential stages of LOC-DLA operation.

Detection of DNA fragments by template-based classification of optical traces of single molecules

Template-based fragment classification is a novel application of the DLA detection technology, and allows for sensitive detection of DNA fragments in the presence of a large excess of non-target DNA. For this classification method, traces of individual molecules are first identified in the raw data and exported by the *GeneEngineer* software. Similarly to data analysis by the clustering software,¹ poorly stretched or overlapping molecules are identified by their backbone traces and excluded from the data set.

The subsequent classification is based on a calculation of the likelihood that each of the individual traces could originate from one of the species from the targets database. This database contains a collection of optical patterns, each of which is produced in average by molecules of specific restriction fragments from every considered target organism in the DLA length range. The average template patterns are generated either by theoretical calculation, based on known sequences and binding probabilities, or experimentally produced by the DLA analysis and clustering of non-sequenced samples.¹

The likelihood calculation algorithm is based on a statistical model of the expected distribution of photons measured along a target DNA restriction fragment. The assumed log-normal distribution accounts for experimental noise and stochastic events at the single molecule level.

In order to simplify comparison of molecules to templates, all optical traces (both measured ones for single molecules and average ones for templates) are interpolated to be represented by the same number of intervals (200). Hence, measured optical traces are represented by 200 values of photon counts t_i ($i=1\dots 200$). The database template average intensity μ_i for each interval is used to calculate the probability $p(t_i, \mu_i)$ of observing intensity t_i in an interval i with mean μ_i . Assuming that the intervals are independent, we present the probability of observing a specific trace originating from a given target template as the product of partial probabilities from intervals: $P_{trace} = \prod_i p(t_i, \mu_i)$. The full probability of observing a specific trace also includes the Gaussian term G_L , modeling the length distribution of stretched molecules: $P(m | T) = G_L \cdot P_{trace}$, where $P(m | T)$ – is the probability that trace m originates from a

target fragment T . Since the probabilities P have very small values, we introduce the *distances* from templates to traces, where a *distance* is a negative logarithm of probability P : $D = -\log(P(m | T))$.

The key step of the classification process is the calculation of distances D from each single molecule trace to each database fragment. After calculating the distances, each of the measured molecules is assigned to a DNA fragment (“template”) to which it has the shortest distance. As a result of this process, several fragments from the database have one or more molecules associated with them. We assume that some of the single molecules may be misclassified due to various reasons: incomplete stretching or tagging, presence of nonspecific tags, shot noise in tag fluorescence channel, lack of proper templates in the database, etc. Therefore, the fragments that have molecules associated with them are merely the potential candidates for detection, and we perform post-classification analysis evaluating each of these groups of molecules.

Confidence of classification and identification for each molecule is correlated with the difference of distances $\Delta D = D_B - D_T$, where D_T is the distance from the template to which the molecule has been classified (“target”) and D_B is the distance to the next closest template (“background”). The difference in distances ΔD corresponds to relative likelihood (or *log-likelihood*) that the molecule has originated from the fragment T rather than some other fragment B:

$$\Delta D = D_B - D_T \propto \log \left[\frac{P(m_i | T)}{P(m_i | B)} \right] \quad (1)$$

For ambiguous molecules, the ratio of probabilities is close to 1 and log-likelihood is close to 0. The log-likelihood value increases with the confidence in classification of a molecule. We characterize each resultant group of classified molecules by two parameters: their quantity expressed as a fraction of the total number of molecules submitted for classification (after initial filtering), and the average log-likelihood, which is the average of ΔD of all molecules in the group.

Figure S6 presents the scatter plot of average log-likelihood vs. fraction of observed molecules. In this specific example we have introduced digitally randomized templates that are known to not match sample targets. These serve as null templates that allow us to model noise in

the experiment and analysis. This, in turn allows for correction in the log-likelihood scale in order to set the threshold for positive detection above the background of mis-classified molecules. The p-values shown in Fig.5a in the main part of the paper have been calculated using the distribution of log-likelihood for null templates.

Finally, for each fragment we can calculate the product of the average log-likelihood and relative quantity. We call this value the total log-likelihood of detection; it is presented on the Fig. 5b in the paper. Since both the quantity of detected molecules and the average log-likelihood are higher for targets truly present in the sample, their product highlights the species identified in the mixture.

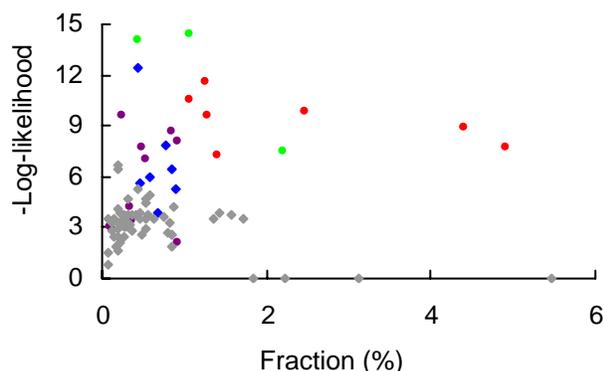


Figure S6. Molecule classification. Scattering plot of average log-likelihood for each fragment vs. relative quantity of molecules classified as such fragments. The plot corresponds to the same data as shown on Fig. 5a and Fig. 5b in the paper. It is demonstrated that only relevant targets (*E. coli* (red), *S. epidermidis* (green), *V. fischeri* (purple) and *F. jonsoniae* (blue) exhibit fragments outside the cloud of background noise and null templates (grey) in the left lower corner of the graph.

1. White, E.J., Fridrikh, S.V., Chennagiri, N., Cameron, D.B., Gauvin, G.P. & Gilmanshin, R. *Clin. Chem.* (2009) **55**, 2121-2129.

Electric field distribution at DNA concentrators optimized with serpentine channels

The primary and final concentrator geometries of LOC-DLA were carefully designed to achieve uniform distribution of DNA at the gel surface. This was achieved by incorporating serpentine channels at the DNA injector in an effort to normalize the path length for all streams of DNA from the injector to the gel surface. This also had the effect of making the electric field distribution at the gel surface more uniform. To illustrate this point, the primary concentrator geometry from LOC-DLA was modeled with and without serpentine channels using COMSOL Multiphysics software (Fig. S7). Although the use of serpentine channels results in higher electric field intensities in the channels (Fig. S7b), the streamline path lengths are more uniform, and electric field intensity at the surface of the gel is more uniform than observed in the model without serpentine channels (Figs. S7b and S7c). This results in more uniform accumulation of DNA at the gel face and faster DNA transfer times.

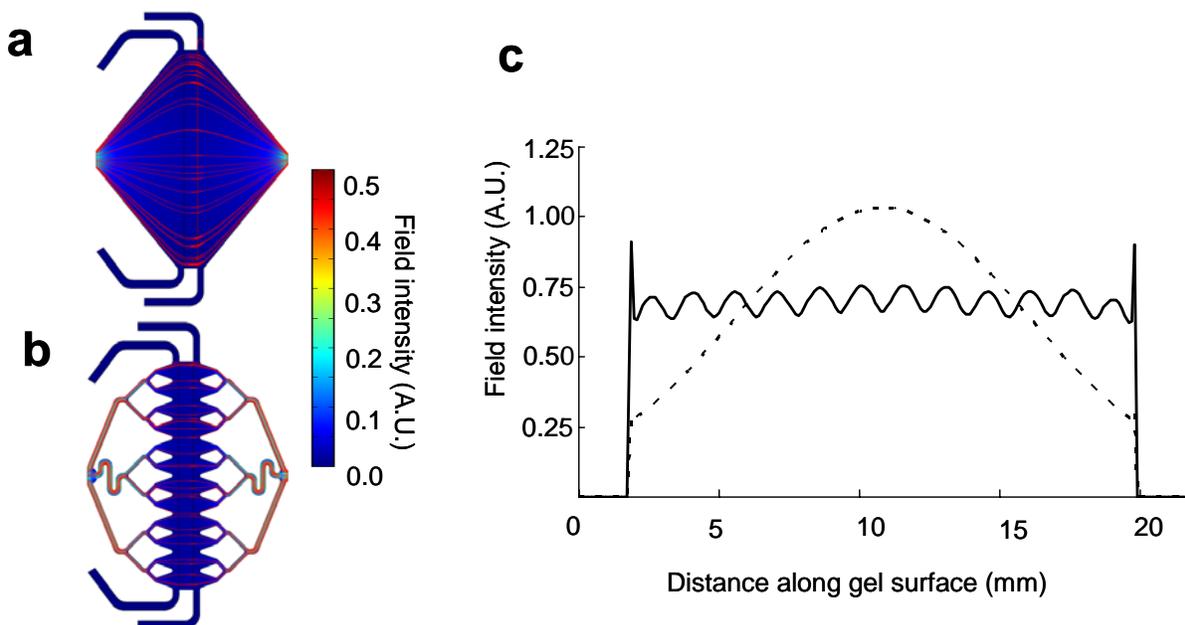


Figure S7. Electric field distribution at gel concentrator. Electric field distribution was calculated in COMSOL Multiphysics for primary gel concentrators without (a) and with (b) serpentine channel geometries. A heat map of electric field intensity with stream lines is presented. The electric field distribution at the gel surface (c) is more uniform in the model with serpentine channels (solid line) than the model without (dashed line).

Control of ion depletion in LOC-DLA gel concentrators

During sustained concentration protocols, the current passing through polymer gel concentrators has been observed to decrease over time. This is likely due to an ion-depletion effect, in which cations selectively pass through the membrane. In order to combat this effect, the gel concentrators on LOC-DLA include cross-flow channels to allow for active buffer replenishment from a large off-device reservoir (Fig. S8a). DNA entering from the opposite port does not interact with these cross-flows, and is thus protected from flow-induced shear. Side ports on the DNA side of the gel, used for fabrication and gel cleaning, are blocked during operation to avoid cross flow. Fig. S8b demonstrates the effect of ion depletion during concentration steps. With no backside flush applied, the observed current across the gel concentrator decreases from 25 to ~ 7.5 μA at an applied voltage of ~ 1 kV. Upon initiation of backside cross flow, the initial observed current is restored and remains relatively constant over 120 minutes.

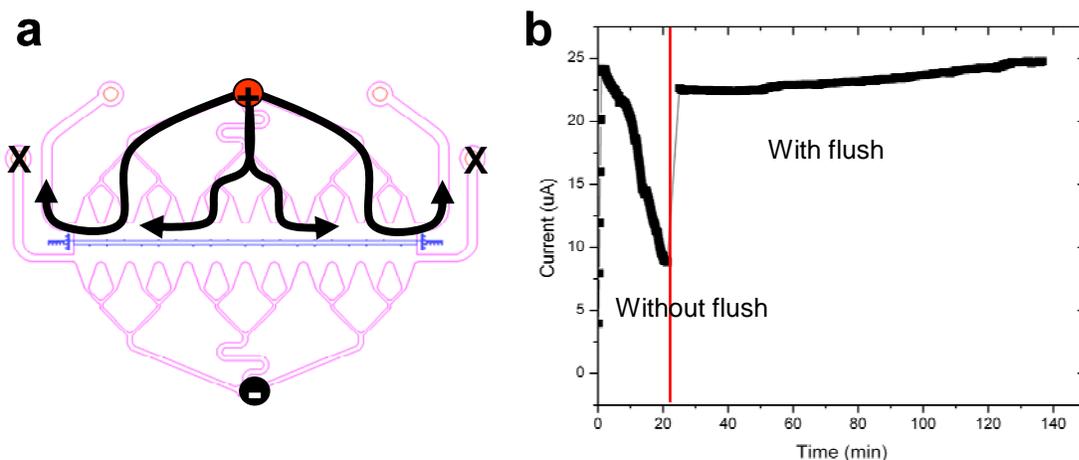


Figure S8. Buffer flushing stabilizes current through polymer gel concentrators. (a) Schematic for flow of buffer across back side of gel concentrator feature. (b) Effect of ion depletion on electric current through concentrator at constant voltage. Application of backside flushing re-establishes initial conductivity.

Improved molecule retention due to decreased fragment overlap with DNA filtration

The efficiency of DLA is determined by the ability to rapidly collect the lengths and bisPNA tag signatures of DNA fragments in the 150 kb-300 kb range. It is critical that fragments not overlap in order to assign each tag pattern correctly. This condition of non-overlap can be achieved by using dilute DNA samples but that would increase the time required for identification of rare targets in a complex mixture. To minimize overlap while also minimizing the experiment time, we predict the optimal non-overlapped read rate in terms of experimentally controllable factors such as the DNA velocity in the funnel and the fragment length distribution.

Consider a sample of DNA with a distribution of fragment lengths. The fragment distribution is characterized by c_i , the number of fragments of length L_i produced by the restriction enzyme. From such a distribution we can calculate $\langle L \rangle$, the average length of the distribution and δ_i , the fractional concentration of the length of interest L_i relative to the digest distribution by the following:

$$\langle L \rangle = \frac{\sum_{j=1}^n c_j L_j}{\sum_{j=1}^n c_j} \quad (2)$$

$$\delta_i = \frac{c_i}{\sum_{j=1}^n c_j} \quad (3)$$

Knowing these distribution parameters and using simple counting arguments, we can derive that the rate of reading non-overlapped strands within the length range i is given by:

$$k_i = -\frac{v\delta_i}{\langle L \rangle} \ln(1 - occ) \left((1 - occ)^{\left(1 + \frac{L_i}{\langle L \rangle}\right)} \right) \quad (4)$$

Where v is the read rate, δ_i is the fraction of digested lengths within the length range L_i , and $\langle L \rangle$ is the average length of the distribution. The occupancy, occ represents the fraction of the detector time that is occupied with DNA (both overlapped and non-overlapped) and is an experimentally controllable variable.

By taking the first derivative of (4) with respect to the occupancy we find there is one maximum, $occ_{\max} = 1 - \exp(-1/(1 + L_i / \langle L \rangle))$. Assuming the device is run at this maximal occupancy then (3) becomes

$$k(v, \delta_i, \langle L \rangle, L_i, occ_{\max}) = \frac{v\delta_i}{e(1 + L_i / \langle L \rangle)} \quad (5)$$

Using this expression the fragment distribution is shown to have a large effect on the maximal read rate achievable.

Experimentally, the DNA prism component of the LOC-DLA device is used to shift the fragment distribution towards longer fragments in the analyzed DNA samples. This allows for the sample to be read at higher occupancy without suffering from overlapped DNA fragments.

Restriction digest of *S. epidermidis* and *E. coli*

E. coli and *S. epidermidis* were used as model organisms to demonstrate fractionation and detection in LOC-DLA. Here we present the theoretical fragment distribution using published sequences of *E. coli* K12 MG1655 and *S. Epidermidis* ATCC1228, NC_004461 using the *Apal* restriction locus (GGGCCC) (Fig. S9). *E. coli* has 7 fragments between 150 and 300 kb (between red vertical lines), representing 30.2% genome coverage. *S. Epidermidis* has 3 fragments in the same range representing 26.4% genome coverage.

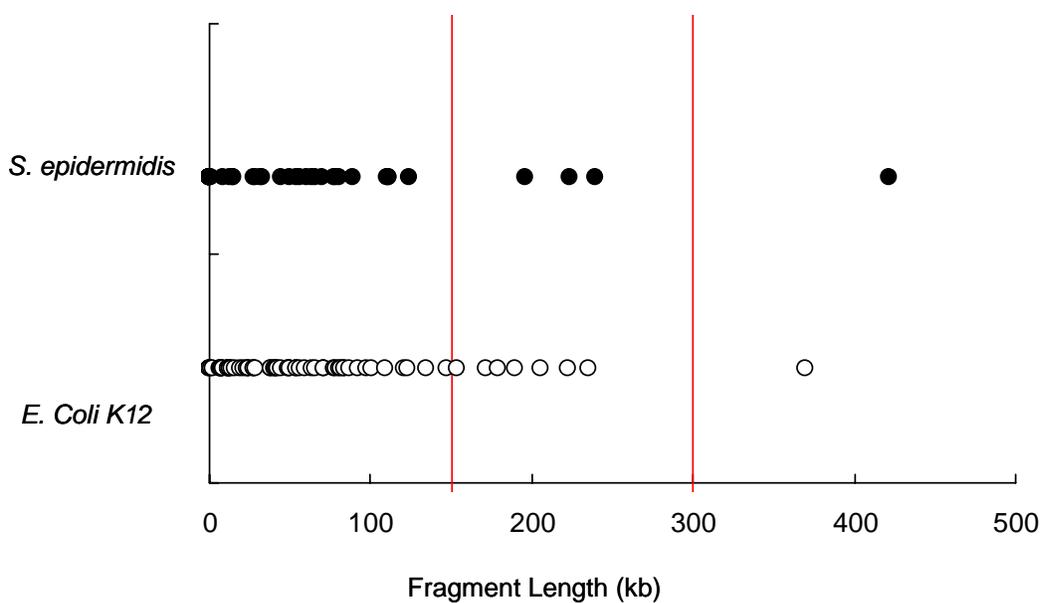


Figure S9. Restriction map of *E. coli* and *S. epidermidis*. Restriction maps of *E. coli* K12 MG1655 (open circles) and *S. Epidermidis* ATCC1228, NC_004461 (filled circles) using the *Apal* cognate sequence (GGGCCC).