Supporting Information

Direct Interrogation of DNA Content Distribution in Nanoparticles by a Novel Microfluidics-based Single-Particle Analysis

Cyrus W. Beh[a], Deng Pan[a], Jason Lee[a], Xuan Jiang[b,c], Kelvin J. Liu[a],

Hai-Quan Mao[a,b,c,d], Tza-Huei Wang[a,b,c,e*]*

[a]Department of Biomedical Engineering, Johns Hopkins School of Medicine,

720 Rutland Avenue, Baltimore, MD 21205, USA;

[b]Department of Materials Science and Engineering, Whiting School of Engineering,

Johns Hopkins University, Baltimore, MD 21218, USA;

[c]Institute for NanoBioTechnology, Johns Hopkins University, Baltimore, MD 21212, USA;

[d]Translational Tissue Engineering Center, Johns Hopkins School of Medicine,

400 North Broadway, Baltimore, MD 21287, USA.

[e]Department of Mechanical Engineering, Whiting School of Engineering,

Johns Hopkins University, Baltimore, MD 21218, USA;

Corresponding Authors:

Tza-Huei Wang, 108 Latrobe Hall, 3400 N. Charles Street, Baltimore, MD 21218; 410-516-7086 (tel), 410-516-7254 (fax), Email: thwang@jhu.edu;

Hai-Quan Mao, 113 Maryland Hall, 3400 N. Charles Street, Baltimore, MD 21218; 410-516-8792 (tel), 410-516-5293 (fax), Email: hmao@jhu.edu.

1. Materials and Methods

A. CICS Instrumentation

The instrumentation for CICS is identical to that previously reported.[1, 2] Briefly, two lasers (Ar-ion laser, 488 nm and HeNe laser, 633 nm, Melles Griot, Carlsbad CA) are combined using an optical fiber coupler. The beams are expanded into large illumination discs using shaping optics, which are then focused into a light sheet using a cylindrical lens. The light is then focused into a microfluidic channel using a 100× oil immersion microscope objective (UPlanFl, Olympus, Center Valley PA). The objective collects the emitted photons from the sample in an epifluorescence configuration. The excitation light and emission fluorescence are separated using a dichroic mirror (z488/633rpc, Chroma Technology, Bellows Falls, VT). A confocal aperture (600 × 150 μm, National Aperture, Salem, NH) is used for spatial filtration, yielding an effective detection volume of 7 × 2 μm on the microfluidic chip. Finally, dichroic mirrors and band-pass filters (Omega Optical, Brattleboro, VT) are used to spectrally separate the green and red fluorescence and to select the desired wavelengths for detection on the avalanche photodiode (APD SPCM-AQR13, Perkin Elmer, Vaudreuil, Quebec, Canada). Fluorescence data is collected and processed using custom software using the Labview interface as well as MATLAB scripts.

B. Streptavidin and Biotin-DNA-Cy5 Sample Preparation

Streptavidin (1.9 μM) was mixed with biotinylated and fluorescently labeled single-stranded DNA (21 nt, IDT, Coralville, IA) at various molar ratios in binding and washing (BW) buffer (1 M NaCl and 0.5× Tris-EDTA buffer, pH 7.5) and incubated overnight. The sample was diluted to 100 nM and incubated with an appropriate amount of streptavidin-coated beads (Dynabeads MyOne Streptavidin C1, Life Technologies, Carlsbad, CA) for 1 h in BW buffer to remove free biotinylated DNA. The sample was then run on an agarose gel (0.5%, TBE running buffer, 8 V/cm, 2 h) and

imaged with a Typhoon 9400 imager and processed with ImageQuant TL (GE Healthcare Life Sciences). Samples were further diluted to a final concentration of 20 pM streptavidin and tested on CICS.

C. Determining Appropriate Plasmid DNA Labeling Density

Plasmid pVR1255 is a 6.4-kb pcDNA encoding luciferase driven by human cytomegalovirus promoter. It was labeled with Cy5 at varying densities using a commercial kit (Mirus Label-It) by diluting the dye. Labeling density was determined by comparing the absorbance at 280 and 650 to obtain the concentration of DNA and dye, respectively, as described by the manufacturer. Under Poissonian statistics, the coefficient of variation of the number of labels on each DNA molecule is given by $\frac{1}{\sqrt{\lambda}}$, where λ is the labeling density of Cy5 per DNA molecule. Therefore, increasing the labeling density can reduce the variation in the fluorescence signal from each DNA. Conversely, labeling density is limited by the onset of fluorescence quenching due to the close proximity of the dyes after DNA compaction, in some cases resulting in almost complete quenching. In order to verify the absence of significant quenching, unlabeled and labeled DNA (average 7.5, 20, 30 and 40 labels per DNA) mixed at different proportions were used to form the nanoparticles, from which we determined that an average labeling density of 7.5 Cy5 dyes per DNA molecule to be an appropriate level (Section S3).

D. Particle Preparation

Linear PEI (17 kDa) was mixed with 1 μg of DNA at a nitrogen/phosphorous (N/P) ratio of 10 in 20 mM pH 5.5 sodium acetate buffer, and incubated for 30 min before testing. To achieve different dye concentration in the polyplexes, labeled and unlabeled DNA were mixed to different proportions before adding linear PEI. In order to minimize nanoparticle aggregation, 60% glycerol was used as a thickening agent to slow down particle aggregation, and to prevent sedimentation.[3]

Each sample is freshly prepared immediately prior to testing. Particle size as determined by dynamic light scattering did not change significantly in glycerol, suggesting that particles remained largely unchanged (Table S2).

For PEI-*g*-PEG particles, PEI (9 kDa)-*g*-PEG (5 kDa, 8%) block copolymer was mixed with 1 μg of DNA in a 70% dimethylformamide (DMF) solution and crosslinked overnight with glutaraldehyde to fix the morphology. The preparation is then diluted to 10% DMF and dialyzed for 3 h in water prior to mixing with glycerol for CICS testing.

E. CICS Experiments

For the streptavidin/biotin experiments the streptavidin was diluted to 20 pM streptavidin in water, after excess biotinylated DNA was removed using streptavidin-coated magnetic beads. The sample contained 0.1% Triton X-100 (Sigma-Aldrich), 1% v/v antifade (SlowFade Gold, Life Technologies) and fluorescent microspheres diluted 10^6-fold as flow tracers (40 nm yellow-green carboxylated Fluospheres, Life Technologies). Samples were loaded into the microfluidic device with a 0.75-μm tall and 5-μm wide analytical channels using pressurized nitrogen. Flow tracers were excited with an argon laser (488 nm, 2.55 mW) and pressure was adjusted until the average burst width of the flow tracers in the green fluorescent channel was 2.5 ms. The argon laser shutter was then closed, the HeNe laser (633 nm, 2.55 mW) was switched on and fluorescence data collected in the Cy5 detection channel.

For the polyplex nanoparticle samples, the process was identical to the streptavidin samples, except that the polyplexes were diluted in 60% glycerol and the samples were loaded into 2-μm tall channels. Triton X-100 was not used because the channels were large enough to fill easily without a surfactant, and because the presence of detergents can potentially affect the stability of the nanoparticles. The pressure was adjusted until the average flow tracer burst width was around 5 ms before data collection. Data were also collected for the flow tracers at the end of each

experiment to ensure relatively stable flow rate (< 10% variation). In all instances, at least 2,500 peaks were collected over a 5-minute period for data processing and fitting.

S2. Data Fitting Methodology

Analysis of the fluorescence distributions for our samples revealed that they were poorly described by Gaussian approximation (Figure S1a). Instead, a lognormal distribution appeared to best describe our data. The lognormal properties of the fluorescence arise from the multiplicative combination of different sources of error in the system, including flow instability, detection volume inhomogeneity, detector noise and spatial variation in efficiency, focal drifts during measurement, as well as non-uniformity due to the samples such as labeling density and DNA incorporation into particles.[4-6] A significant source of error in our polymer/DNA nanoparticle samples is the variation in labeling density. On the other hand, since each biotinylated oligonucleotide contains only one Cy5, the low fluorescence of the streptavidin/oligo samples suffers from relatively higher error from the system due to the Poissonian nature of the photoexcitation and photodetection processes. Recognizing the parallels between the variations faced in CICS and TIRF instrumentations as well as the sort of variation we were observing in our DNA labeling vs the labeling of proteins with fluorophores, we found the algorithm proposed by Mutch et al. to be highly appropriate to our data and adapted it for our data.

Data collected from the single molecule fluorescence detection yields the number of photons emitted by each molecule or particle. We have found that the photon count for each fluorescence event is directly proportional to the time each molecule/particle resides in the detection volume, as long as the residence time is significantly longer than the detector sampling time. Using a sampling time of 0.1 ms and mean dwell times of at least 2.5 ms in all cases, it is possible to account for small

differences in flow velocities across different samples by dividing the photon count by the residence time of flow tracer particles (~5 ms), yielding the corrected fluorescent burst size of each event.

The corrected burst sizes are then binned into histograms, and the resulting calibrating distribution (denoted D_{DNA} in our samples) and distributions of interest ($D_{particle}$) are then passed into a Matlab implementation of the fitting algorithm using the Simulated Annealing fitting method.[5] D_{DNA} is normalized by the total number of DNA fluorescence events to give $D_{DNA,1}$, with a total area of 1 and corresponds to the c = 1 basis sub-distribution. This represents the normalized fluorescence distribution that one can expect if the nanoparticles consisted entirely of particles containing a single fluorescent DNA molecule. Although such particles will be indistinguishable from free DNA molecules, barring self-quenching, the absence of a free DNA band when nanoparticles were run on an agarose gel indicate that all the DNA molecules were in fact incorporated into particles.

The fluorescence of each detected peak in D_{DNA} is then multiplied by a scaling factor c (= 2, 3, 4…) to derive the distribution of perfectly monodisperse nanoparticle preparations containing c fluorescent DNA molecules. The distributions for all the c values are then normalized and constitute the full set of basis distributions $\{D_{DNA,c}\}_{c=1\ldots C}$. The upper limit C is chosen such that it is around four times the average number of DNA per particle, determined by trial and error, in order to minimize computational load. Each basis distribution is then assigned a weight, a_c, and the sum of the weighted distributions, $\sum a_c \times D_{DNA,c}$ is fitted to the $D_{particle}$ by optimizing the weights, $\{a_c\}$. Upon convergence, the weight for each basis distribution indicates the number of particles in $D_{particles}$ containing c DNA molecules. For the streptavidin-biotinylated oligonucleotide system, D_{DNA} and $D_{particles}$ are replaced by D_{oligo} and $D_{conjugate}$, respectively.

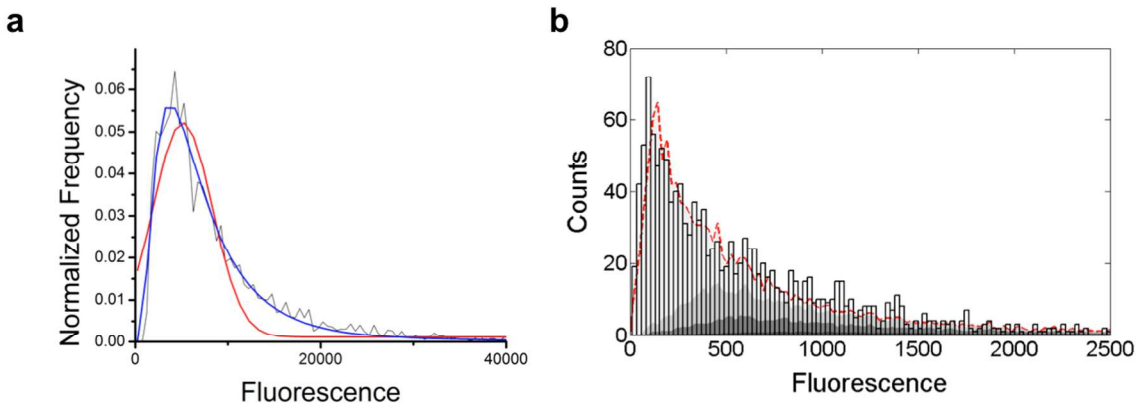

Figure S1. a. Distribution of labeled DNA fluorescence intensity (black) was found to be better fitted with a log-normal approximation (blue) than with a Gaussian one (red). The reason for this is that the convolution of multiple factors (fluorescent dye labeling density, thermal fluctuations, detector efficiencies, etc), each with a Gaussian or Gaussian-approximated variations, results in a log-normal distribution. b. A typical set of fitted data is shown with the stacked plots (shaded) representing the sum of the contribution to the signal of constituent basis distributions. The raw polyplex fluorescence distribution (bars) is well approximated by the re-convolved signal (red dashed line) constructed using the basis distributions generated from the DNA fluorescence distribution.

For nanoparticle distributions that use a mixture of labeled and unlabeled DNA, the binomial distribution of labeled DNA needs to be taken into account, thereby complicating the analysis. Therefore, we ideally want to use only labeled DNA for the preparation of the nanoparticles. A typical histogram of experimental data fitted with our algorithm is shown in Figure S1b.

S3. Analysis of Gel Results and Comparison to Fitting Results

It is worth noting that the order of the bands on the gel is somewhat different than expected. In a typical electrophoresis experiment, one would expect the migration rate to decrease with more conjugated DNA, and hence mass. However, because streptavidin is significantly larger than ssDNA oligonucleotide, its presence contributes greatly to the mass, such that an increase in the number of bound ssDNA oligonucleotide results in higher electrophoretic mobility due to higher charge/mass ratio. We have been able to obtain different band orders by increasing the agarose concentration

from 0.5% to 2%, as well as perform the electrophoresis on PAGE gels of different concentration. At these higher gel concentrations, the mass increase due to the ssDNA becomes a significant contributing factor, yielding titration series of unclear trends. A similar study has been reported with streptavidin-coated quantum dots.[7]

The proportion of streptavidin in the each occupancy state, N, can be determined from the gel band intensity, I_N, by accounting for the number of bound ligands, *i.e.*

$$P(N) = \frac{I_N / N}{\sum_N (I_N / N)}$$

Streptavidin with 2 and 3 bound biotinylated DNA (referred to as S-2 and S-3 in this section) constituted 29.4% and 67.9% of the sample, as determined by our fitting strategy (Table S1, Column 5). The corresponding proportions, *P(N)*, as determined from the gel images were 15.9% and 77.3% (Table S1, Column 3). The inaccuracy of quantification using gel images notwithstanding, the difference can be easily accounted for by considering the incomplete labeling of the oligonucleotides. While the quantification using our method is strictly by means of fluorescence intensity, gels are able to distinguish between sizes, regardless of presence or absence of the fluorescent dye. Consider a distribution of 10^3 streptavidin molecules. Based on the band intensity, this corresponds to 159 S-2, 773 S-3 and 68 S-4 conjugates. When the DNA labeling is assumed to be around 95% (binomial distributions, n = 2, 3, and 4, p = 0.95), of the 773 S-3 molecules, 663 will have 3 fluorescent DNA molecules, while 103 and 5 will have 2 and 1 fluorescent DNA molecules, respectively. This is a reasonable assumption since the sample has been purified by IDT with HPLC after synthesis, and the labeling density is guaranteed to be above 90%. Doing the same for the S-2 and S-4 conjugates, we get the distribution of fluorescent ligands in the samples (*P(N)**, Table S1, Column 4). When we consider only fluorescent ligands instead of

total number of ligands on each streptavidin, the results from the gel experiment (S-2 and S-3 at 24.9% and 67.5%, respectively) and those from our data fitting become very similar (Figure 2).

Table S1. Gel band analysis accounting for the incomplete labeling of oligonucleotides.

Occupancy, N	Band Intensity, %	% of Streptavidin with N Ligands, *P(N)*	% of Streptavidin with N Fluorescent Ligands, *P(N)**	Fitted Estimate of *P(N)* by Current Method
1	0	0	2.0	2.1
2	10.8	15.9	24.9	29.4
3	79.0	77.3	67.5	67.9
4	9.2	6.8	5.5	0.6

S4. Considerations for DNA Labeling Density

For these experiments, an appropriate level of DNA labeling needed to be achieved. On one hand, a high DNA labeling density ensures high signal-to-noise ratio, making detection of DNA and particles easy, and also reduces the molecule-to-molecule DNA labeling variability, which can be described as a Poissonian process. On the other hand, because of the close proximity of the fluorophores once compaction by the polymer is achieved, significant – or even complete – fluorescence self-quenching can result. Therefore, it is necessary to determine the appropriate labeling density that can be used in these experiments.

To do this, we formed particles using unlabeled and labeled DNA pre-mixed at different ratios, and tested them on the CICS set-up as described in the main text. The lowest average DNA labeling density in the experiments is 7.5, with less than 0.05% of the DNA containing no dye label, and a theoretical labeling CV of around 36%. DNA molecules with average Cy5 labels of 20, 30 and 40 were also prepared. In the absence of self-quenching, $\overline{F}$ for each DNA labeling density should

increase linearly as the proportion of DNA labeled with Cy5 is increased. While this is true when the proportion of labeled DNA is low (up to ~ 40–50%), in some cases there is a distinct drop-off in $\overline{F}$ at higher proportions of labeled DNA. By extrapolating from the initial linear region, the expected $\overline{F}$ in particles prepared with only labeled DNA, $\overline{F_{(100)}}^*$ can be estimated. For each labeling density, $\overline{F}\ /\overline{F_{(100)}}^*$ allows the extent of self-quenching between labeling densities to be qualitatively compared (Figure S2a). We found that the proportion of labeled DNA corresponding to the onset of this drop-off *decreases* as the DNA labeling density *increases*, a result of the decreasing mean dye-dye distance (Section S4).

Furthermore, by dividing the mean fluorescence of the particles by that of the constituent DNA and also accounting for the ratio of labeled to unlabeled DNA used in each preparation, we can calculate the total DNA content, $\overline{n_{DT}}$ in each case, which should remain fairly constant. One exception is when the number of DNA molecules in each particle is so low that a significant proportion of the particles consist wholly of unlabeled DNA, which will skew the estimated total DNA content in a manner described by a binomial process as described in the main text.

For the samples prepared with an average of 7.5 Cy5 labels on each DNA molecule, the value of $\overline{n_{DT}}$ initially drops slightly as the proportion of labeled DNA increases, as a result of the binomial process described above (Figure S2b). Further increases in the proportion of labeled DNA do not result in significant changes in the value of $\overline{n_{DT}}$, as the number converges to around 5. On the other hand, the results show a clear drop in the estimated $\overline{n_{DT}}$ at labeling densities of 30 and 40 per DNA. The larger $\overline{n_{DT}}$ (before the onset of the drop-off) in preparations using higher DNA labeling density may be a result of the presence of hydrophobic Cy5 moieties that may have a tendency to aggregate. Such aggregation of labeled molecules has previously been reported literature [8]. In agreement with this hypothesis, we measured the size of freshly prepared particles

on using DLS and determined that the nanoparticles have a number-average size of around 70 nm. However, the DNA particles with higher DNA labeling densities started to aggregate when left to stand, resulting in erroneous size measurements and unstable count rates. Even for particles prepared with DNA with 7.5 Cy5 labels, the size of the particles increased to ~200 nm after two days. This aggregation behavior is well reported in literature [3, 9]. Taking all these factors into consideration, a labeling density of 7.5 per DNA molecule is chosen for the method.

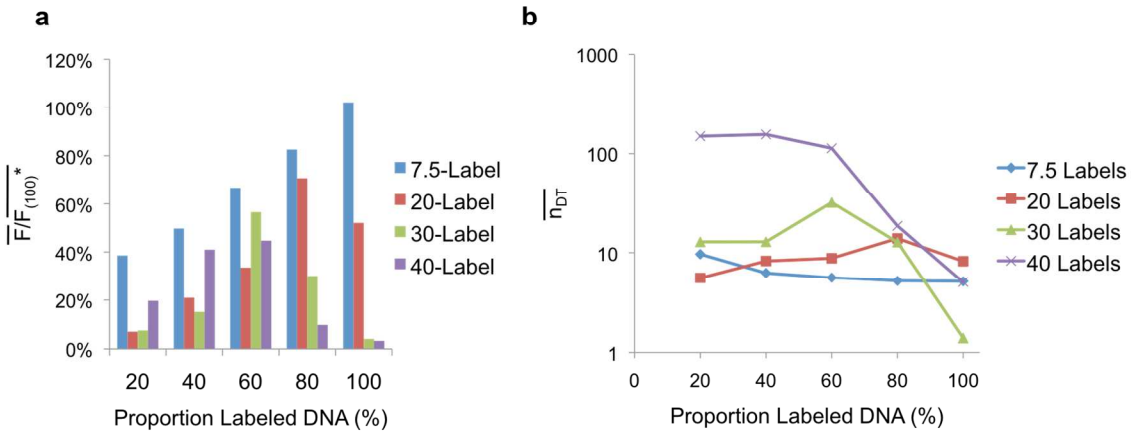

Figure S2. a. The mean fluorescence of particles at the various labeled DNA proportions were normalized to the expected fluorescence of particles formed using 100% labeled DNA in the absence of quenching ($F_{(100)}$*by extrapolation of the linear region) for each DNA labeling density (7.5, 20, 30, 40). At a labeling density of 7.5 per DNA molecule, no appreciable quenching of the signal was observed, evidenced by the linearity of n_{DL} for all the labeled DNA proportions tested. On the other hand, at higher densities, there is a distinct drop-off in the fluorescence. Furthermore, the onset of this drop-off occurs at a lower proportion of labeled DNA as the labeling density increases, illustrating the increasing severity of quenching . b. The total DNA content in each case, n_{DT}, can be calculated by accounting for the proportion of labeled DNA used in the preparation of the particles. The higher n_{DT} with increasing DNA labeling density shows that the dyes affect the particle formation process, most likely via enhanced aggregation. This further argues for a low labeling density.

S5. Dynamic Light Scattering (DLS) Data and Average Intensity

Nanoparticles were characterized using DLS and it was found that the PEG-PEI particles were larger than those without PEG (Table S2). This may be in part due to the PEG polymer corona that

forms around the nanoparticle.[10] The size of the particles do not appear to be significantly different in water or in 60% glycerol, although collapse of PEG corona may account for the smaller-than-expected size of the PEG-PEI particles in glycerol. This suggests that the presence of glycerol does not significantly alter the particle properties.

Table S2. Z-average diameter of nanoparticles in water and glycerol

Particle/Solution	**Z-Average Size (nm)**
Linear PEI/DNA nanoparticles in Water	114.9 ± 1.5
Linear PEI/DNA nanoparticles in 60% Glycerol	111.1 ± 1.8
PEI-*g*-PEG/DNA nanoparticles in Water	143.5 ± 0.8
PEI-*g*-PEG/DNA nanoparticles in 60% Glycerol	111.2 ± 4.3

A further piece of evidence is the proportion of labeled DNA corresponding to a drop-off in the fluorescence intensity. Using a particle diameter of 70 nm (number-average diameter) for calculations, and a Wigner-Seitz radius approximation for dye-dye distance, it is found that at the onset of the intensity drop-off, the dye-dye distance is approximately 7.5-7.8 nm for labeling densities of 20, 30 and 40. This is a reasonable distance for FRET processes, which explains the onset of Cy5 self-quenching. For example, some FRET processes contributing to fluorescence quenching has been estimated to have a FRET distance of 8.3–8.4 nm.[11] The higher density can also promote the formation of Cy5 aggregates that have been postulated to cause significant fluorescence quenching.[11-14] On the other hand, for the DNA with 7.5 labels, even when 100% of the DNA used was labeled the dye-dye distance was around 10.5 nm, too long for appreciable FRET. Together with the aggregation issues observed for higher labeling densities, these results argue for a lower labeling density.

S6. Variation in DNA Content Estimate

As illustrated in Figure S3a, the technical variation of our method is fairly low. The four sets of results are collected from the same preparation of the particles, which were then analyzed using our algorithm separately. However, the sample variation of the particle preparation is, as expected, very high. Figure S3b shows two sets of samples prepared under ostensibly identical conditions. Although the average DNA content in each case (4.8 and 5.0, respectively) is very similar, they have very different DNA distribution profiles. This batch-to-batch variation was a main motivation for developing a method capable to determining the subpopulations present in particle preparations, which we believe contribute to differences in gene delivery efficiency.

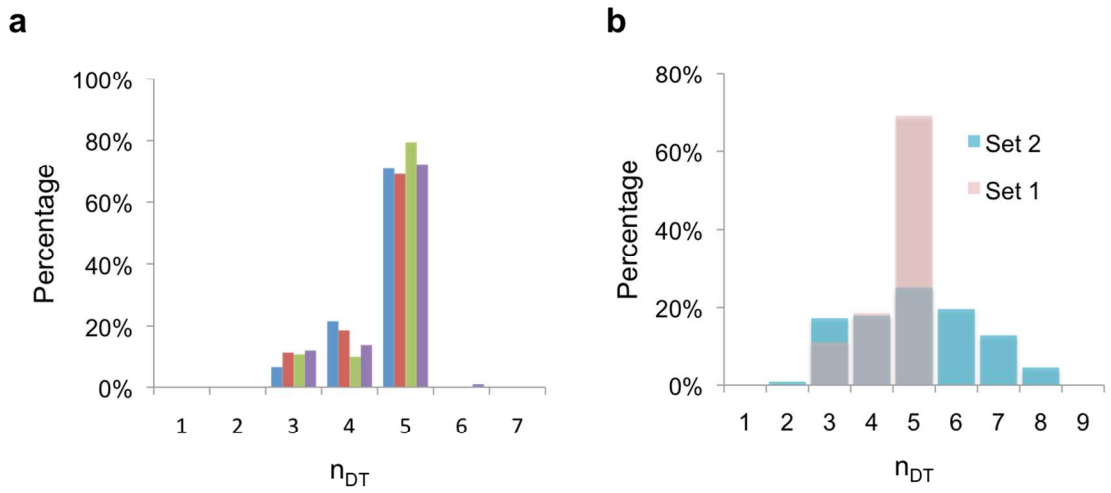

Figure S3 a. We tested the technical variation of our method by collecting four sets of data from the same particle preparation (n > 3000 in all cases) and processing them using our algorithm. We found that all the fitted distribution were similar, showing an large subpopulation containing 5 DNA per particle, as well as smaller sub populations containing 3 and 4 DNA molecules. b. On the other hand, the variation between sample preparations were very high. Despite the similar average DNA content in two separate preparations (expected n_{DT} of 4.8 and 5.0 for sets 1 and 2, respectively), the distributions of the subpopulations were very different. This batch-to-batch variation motivates this work, which provides a way to probe a hitherto neglected parameter.

Additional References:

1. Liu, K. J.; Brock, M. V.; Shih, I.; Wang, T. H. *J. Am. Chem. Soc.* **2010,** 132, 5793-5798.
2. Liu, K. J.; Wang, T.-H. *Biophys. J.* **2008,** 95, 2964-2975.
3. Sharma, V. K.; Thomas, M.; Klibanov, A. M. *Biotechnol. Bioeng.* **2005,** 90, 614-620.
4. Limpert, E.; Stahel, W. A.; Abbt, M. *Bioscience* **2001,** 51, 341-352.
5. Mutch, S. A.; Fujimoto, B. S.; Kuyper, C. L.; Kuo, J. S.; Bajjalieh, S. M.; Chiu, D. T. *Biophys. J.* **2007,** 92, 2926-2943.
6. Mutch, S. A.; Gadd, J. C.; Fujimoto, B. S.; Kensel-Hammes, P.; Schiro, P. G.; Bajjalieh, S. M.; Chiu, D. T. *Nat. Protocols* **2011,** 6, 1953-1968.
7. Zhang, Y.; Liu, K. J.; Wang, T.-L.; Shih, I.-M.; Wang, T.-H. *ACS Nano* **2011,** 6, 858-864.
8. Brakmann, S.; Löbermann, S. *Angew. Chem. Int. Ed.* **2001,** 40, 1427-1429.
9. Lai, E.; van Zanten, J. H. *Biophys. J.* **2001,** 80, 864-873.
10. Cruz, L. J.; Tacken, P. J.; Fokkink, R.; Figdor, C. G. *Biomaterials* **2011,** 32, 6791-6803.
11. Schobel, U.; Egelhaaf, H.-J.; Brecht, A.; Oelkrug, D.; Gauglitz, G. *Bioconjug. Chem.* **1999,** 10, 1107-1114.
12. Kang, J.; Kaczmarek, O.; Liebscher, J.; Dähne, L. *Int. J. Polym. Sci.* **2010,** 2010, 1-7.
13. Berlier, J. E.; Rothe, A.; Buller, G.; Bradford, J.; Gray, D. R.; Filanoski, B. J.; Telford, W. G.; Yue, S.; Liu, J.; Cheung, C.-Y.; Chang, W.; Hirsch, J. D.; Beechem Rosaria P. Haugland, J. M.; Haugland, R. P. *J. Histochem. Cytochem.* **2003,** 51, 1699-1712.
14. Gruber, H. J.; Hahn, C. D.; Kada, G.; Riener, C. K.; Harms, G. S.; Ahrer, W.; Dax, T. G.; Knaus, H.-G. *Bioconjug. Chem.* **2000,** 11, 696-704.