## Supplementary Information

### Sequencing ultra-rare targets with compound nucleic acid cytometry

Chen Sun[1], Kai-Chun Chang[1] and Adam R. Abate[1,2,3*]

[1] Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94158, USA
[2] California Institute for Quantitative Biosciences, University of California San Francisco, San Francisco, CA 94158, USA
[3] Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

[*] Corresponding author: adam@abatelab.org

### Table of Contents

**Derivation of equations for enrichment power**

We define the enrichment power as the ratio of target purity in the sample after enrichment to before,

$$\text{Enrichment power} = \frac{\text{Target purity after enrichment}}{\text{Target purity before enrichment}} \qquad (1)$$

where $N_T$ is the number of target molecules, $N_O$ is the total number of off-target molecules, $D$ is the total number of droplets and $f$ is the false positive rate of the assay.
In the mixed DNA sample, the total number of molecules is $N_T + N_O$ and thus

$$\text{Target purity before enrichment} = \frac{N_T}{N_T + N_O} \qquad (2)$$

We encapsulate these $N_T + N_O$molecules into $D$ droplets, such that each drop contains $\frac{N_T + N_O}{D}$ molecules. The DNA mixture is partitioned at limiting dilution such that individual droplets rarely contain more than one target expected by Poisson statistics, thus $N_T$ drops are PCR positive. We also expect $fD$ false positive drops. Because the false positive rate is generally small and the target is ultra-rare, the instances of false positive and true positive in the same droplet are infrequent. In together, $N_T + fD$ drops are sorted and the number of molecules in the sorted sample is $\frac{N_T + N_O}{D}(N_T + fD)$, of which $N_T$ molecules are the targeted ones. Therefore,

$$\text{Target purity after enrichment} = \frac{N_T}{\frac{N_T + N_O}{D}(N_T + fD)} \qquad (3)$$

and to simplify the enrichment power to

$$\text{Enrichment power} = \frac{\frac{N_T}{\frac{N_T + N_O}{D}(N_T + fD)}}{\frac{N_T}{N_T + N_O}} = \frac{1}{\frac{N_T}{D} + f} \qquad (4)$$

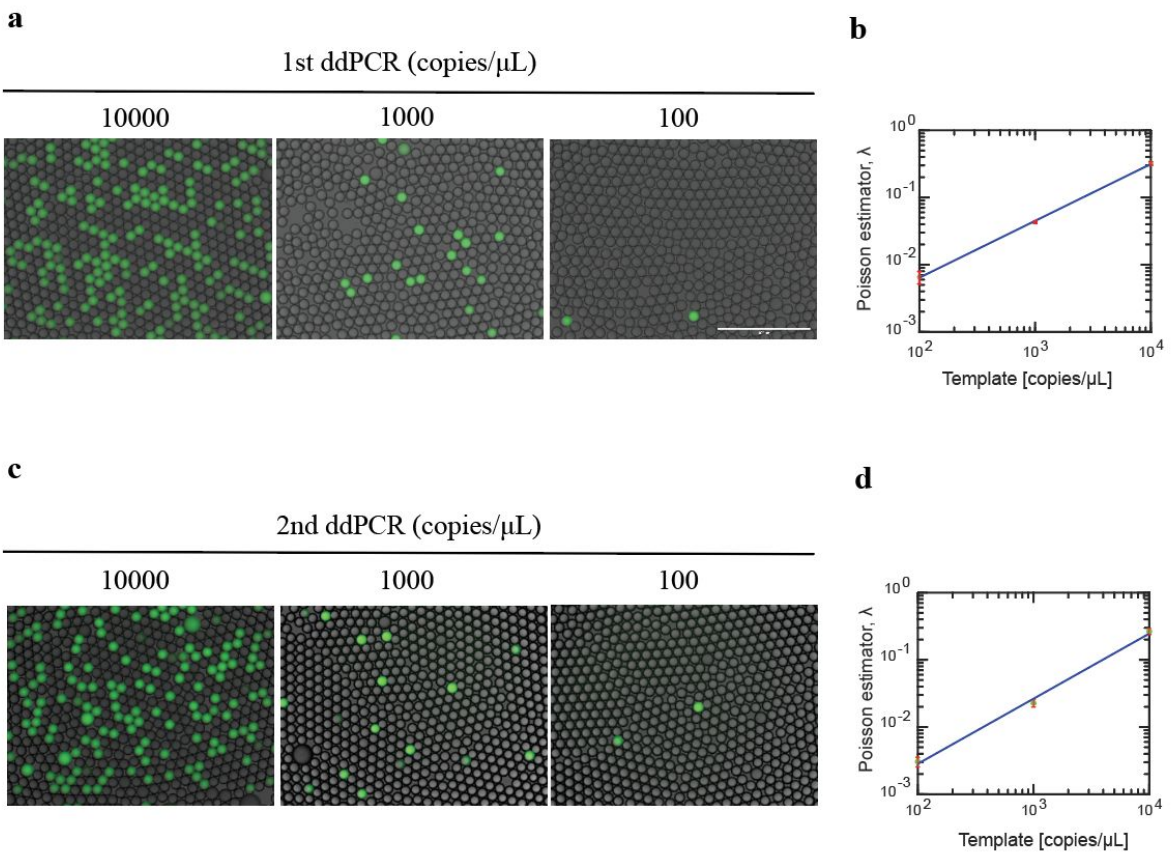


**Figure S1.** TaqMan sets reliably detect ΦX 174 DNA in compound NAC. Fluorescence images of droplets after PCR amplification with TaqMan sets used in **(a)** 1st NAC and **(c)** 2nd NAC for ΦX 174 DNA templates at varying concentrations. 5 ng/µL Lambda DNA was added to each reaction to confirm the specificity of the TaqMan sets. The template copy number per droplet estimated by assuming a Poisson distribution scales with the input template concentrations during ddPCR in **(b)** 1st NAC, and **(d)** 2nd NAC.

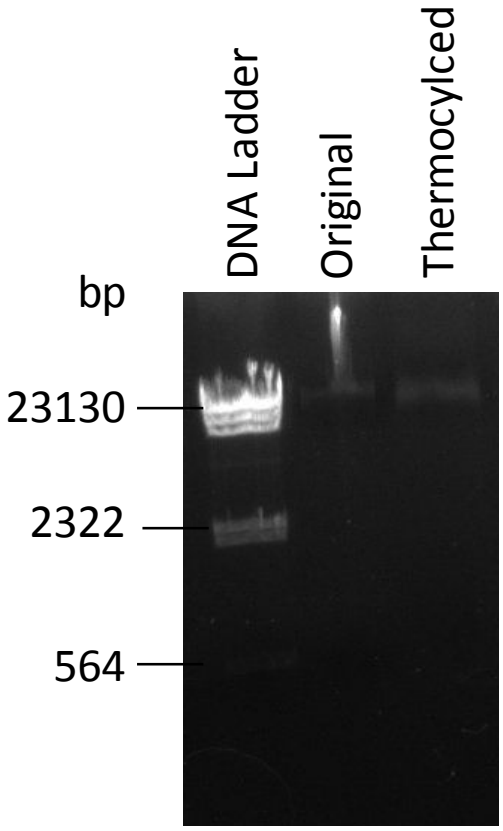


**Figure S2.** Gel analysis showing minimal DNA fragmentation during thermo cycling using 86 ºC denaturation. Lane 1, DNA ladders; Lane 2, DNA extracted from a human cell line; Lane 3, DNA went through 35 cycles of thermal cycling with 86 ºC denaturation.

**Table S1: $C_t$ values from the qPCR plots for single and double enrichment**

| | Single enrichment | | | Double enrichment | | |
|---|---|---|---|---|---|---|
| | unsort $C_t$ | single sort $C_t$ | $\Delta C_t$ | unsort $C_t$ | double sort $C_t$ | $\Delta C_t$ |
| ΦX 174 | 33.18 | 28.655 | -4.525 | 33.180 | 31.193 | -1.987 |
| Lambda | 15.055 | 17.733 | 2.68 | 15.055 | 27.020 | 11.965 |

**Table S2: Primers and probes for compound enrichment of ΦX 174 virus DNA**

| Primer Name | Sequence (5→3') |
|---|---|
| 1st NAC | |
| ΦX 174 probe1 | /56-FAM/TAAATCGAA/ZEN/GTGGACTGCTGGCGG/3IABkFQ/ |
| ΦX 174 F1 | GCAGGAATTACTACTGCTTGTTTAC |
| ΦX 174 R1 | GAATCGTTAGTTGATGGCGAAAG |
| 2nd NAC | |
| ΦX 174 probe2 | /56-FAM/CGTATGCAG/ZEN/GGCGTTGAGTTC/3IABkFQ/ |
| ΦX 174 F2 | GCAGATGGATAACCGCATCA |
| ΦX 174 R2 | CCTTATGGCCGTCAACATACA |
| Post sorting qPCR | |
| ΦX 174 probe3 | /56-FAM/ATGGAACTG/ZEN/ACCAAACGT/3IABkFQ/ |
| ΦX 174 F3 | GCGCTCTAATCTCTGGGCAT |
| ΦX 174 R3 | CAAAGAAACGCGGCACAGAA |
| lambda probe3 | /5Cy5/TGAGGTGCT/TAO/TTATGACTCTGCCGC/3IAbRQSp/ |
| lambda F | GCGAGTATCCGTACCATTCAG |
| *lambda* R | TCCCTTTCGGCATACCATTT |

**Table S3: Primers and probes for compound enrichment of HIV provirus DNA**

| Primer Name | Fluorophore, quencher | Sequence (5→3') |
|---|---|---|
| 1st NAC | | |
| *pol* probe | FAM, ZEN-3'IBFQ | AAGCCAGGAATGGATGGCC |
| *pol* F | | /5Biosg/ GCACTTTAAATTTTCCCATTAGYCCTA |
| *pol* R | | /5Biosg/ CAAATTTCTACTAATGCTTTTATTTTTTC |
| 2nd NAC | | |
| LTR probe | FAM, ZEN-3'IBFQ | CTGGTAACTAGAGATCCCT |
| LTR F | | GCCTCAATAAAGCTTGCCTTGA |
| LTR R | | GCTAGAGATTTTCCACACTGACTARA |
| *tat probe* | Cy5, TAO-3'IBRQSp | CTATGGCAGGAAGAAGCGGAGACAGC |
| *tat* F | | TGTAAAAAGTGTTGCYTTCATTG |
| *tat* R | | ACTACTTACTGCTTTGRTAGAG |