



Artifact-free quantification and sequencing of rare recombinant viruses by using drop-based microfluidics

Authors: Ye Tao, Assaf Rotem, Huidan Zhang, Shelley K. Cockrell, Stephan A. Koehler, Connie Chang, Lloyd W. Ung, Paul G. Cantalupo, Yukun Ren, Jeffrey S. Lin, Andrew B. Feldman, Christiane E. Wobus, James M. Pipas, David A. Weitz

NOTICE: This is a postprint of an article that originally appeared in Extremophiles on November 2015. DOI: <http://dx.doi.org/10.3389/fmicb.2015.01251>

[If there is a citation, include it on a line by itself, if grey literature, include the full date of this version (July 4, 2014). include a doi if possible]

Artifact-Free Quantification and Sequencing of Rare Recombinant Viruses by Using Drop-Based Microfluidics

Ye Tao,^[a, b] Assaf Rotem,^[a] Huidan Zhang,^[a, c] Shelley K. Cockrell,^[d] Stephan A. Koehler,^[a] Connie B. Chang,^[a, e] Lloyd W. Ung,^[a] Paul G. Cantalupo,^[d] Yukun Ren,^[b] Jeffrey S. Lin,^[f] Andrew B. Feldman,^[f, g] Christiane E. Wobus,^[h] James M. Pipas,^[d] and David A. Weitz*^[a, i]

Recombination is an important driver in the evolution of virus-es and thus is key to understanding viral epidemics and im-proving strategies to prevent future outbreaks. Characteriza-tion of rare recombinant subpopulations remains technically challenging because of artifacts such as artificial recombinants, known as chimeras, and amplification bias. To overcome this, we have developed a high-throughput microfluidic technique with a second verification step in order to amplify and se-quence single recombinant viruses with high fidelity in picoli-ter drops. We obtained the first artifact-free estimate of in vitro recombination rate between murine norovirus strains MNV-1 and WU20 co-infecting a cell ($P_{\text{rec}} = 3.3 \times 10^{-4} \pm 2 \times 10^{-5}$) for a 1205 nt region. Our approach represents a time- and cost-ef-fective improvement over current methods, and can be adapt-ed for genomic studies requiring artifact- and bias-free selec-tive amplification, such as microbial pathogens, or rare cancer cells.

Viruses pose a continuing threat to public health due to vari-ous compounding factors, including a lack of effective treat-ment, their rapid evolution, and the difficulty of characterizing rare but potentially lethal strains.^[1] Although recombinant mu-tations are very infrequent, they have the potential to generate dangerous, virulent strains, because the large-scale exchange of genetic material can provide adaptations to escape immune systems and resist treatment.^[2] To more effectively combat such lethal adaptations a better understanding of viral recombination is required;however,currentgenomic sequencing techniques involveamplification artifacts that obscure rare mu-tations, and amplification bias complicatesdetermining their frequencies of occurrence.^[3]

The major limitations for studyingribonucleic acid (RNA) viral recombinants arise from the unavoidable genomic am-plification step precedingthe sequencing.When amplifying rare recombinanttemplates in the presence of their parental genomes,template switching can occur during reversetran-scription(RT) andPCR, thereby resulting in aconcatenation of two parental fragments. These artificial recombinants are knownaschimeras, and they are indistinguishable from true biological recombinants (see Figure S1 in the Supporting Infor-mation).^[4] Chimeras can be avoided in emulsionPCR^[5] by dilut-ing the templateso that drops contain at most asingle tem-plate (Table S1).However,emulsion PCR cannot overcome am-plification bias and sequencing bias, which prevent sequencing methods from quantifying recombinant subpopulations.^[6] To overcome counting bias, on-chip digital PCR^[7] maybeused to count the number of true recombinants in asample;however, digitalPCR is limited in the number of samples it can test and does not have the capability to retrieve the contentofampli-fied samples for sequencing (Table S1). Thus, amethod that can both count and sequence rare recombinantswill improve our understanding of viral recombination, therebyleadingto better antiviral drugs and vaccines.

Drop-based microfluidics (DBM) is arapidly emerging tech-nique whereby micron-sized aqueous drops immersed in an

inert carrier oil are used as vessels for biological experiments.^[8] The drops are surrounded by surfactants carefully chosen to prevent drop coalescence or surface adsorption of biological molecules. Individual drops can be filled, steered, split, combined, detected, and sorted in a microfluidic device at thousands per second,^[9] and their content is easily retrieved for downstream analysis by separating the emulsion. The small volume of these drops (~10 pL) is ideal for isolating single molecules and assaying them with minimal background noise. Hence DBM is ideal for developing a platform to amplify single templates for both counting and downstream sequencing.

Here we developed a DBM technique that addresses the limitations of current assays for characterizing and quantifying rare genomes: single templates are isolated, individually amplified in drops, and then placed in microwells for processing. By detecting, sorting, and consensus sequencing the amplicons, we unambiguously determined both recombination frequency and loci (Figure 1 A). We demonstrate our technique by accurate characterization of recombinant progeny from two RNA murine norovirus (MNV) strains, MNV-1 and WU20, by co-infect-

ing a murine RAW 264.7 macrophage culture. This platform decreases the artifacts produced during RT, PCR amplification, and sequencing to an arbitrarily low value.

To determine the concentration of the parental strains, we co-encapsulated viral templates with Taqman probe (Life Technologies) in 8.25 pL drops and performed off-chip amplification. This probe fluoresces only in the presence of viral templates (bright and dark drops in Figure 1 C). We injected the drops into a reading and sorting device (Figure 1 A 4), and obtained distributions of size and intensity (Figure 2 A). The average number of templates per drop (λ) decreases with the dilution factor (X) according to $\lambda = cV/X$, where c is the stock concentration and V is the volume of the drop. The number of templates in any drop (n) follows a Poisson distribution ($p_n = \lambda^n e^{-\lambda} / n!$).^[10] As λ approaches 0, this distribution is binary because the probability of a drop containing multiple templates is negligible. Accordingly, λ approaches the ratio of bright drops (which contain one template) to dark drops (no template): $\lambda = N_+ / N_-$. The data from the dilution series fit this linear relationship well, and we obtained viral stock concentra-

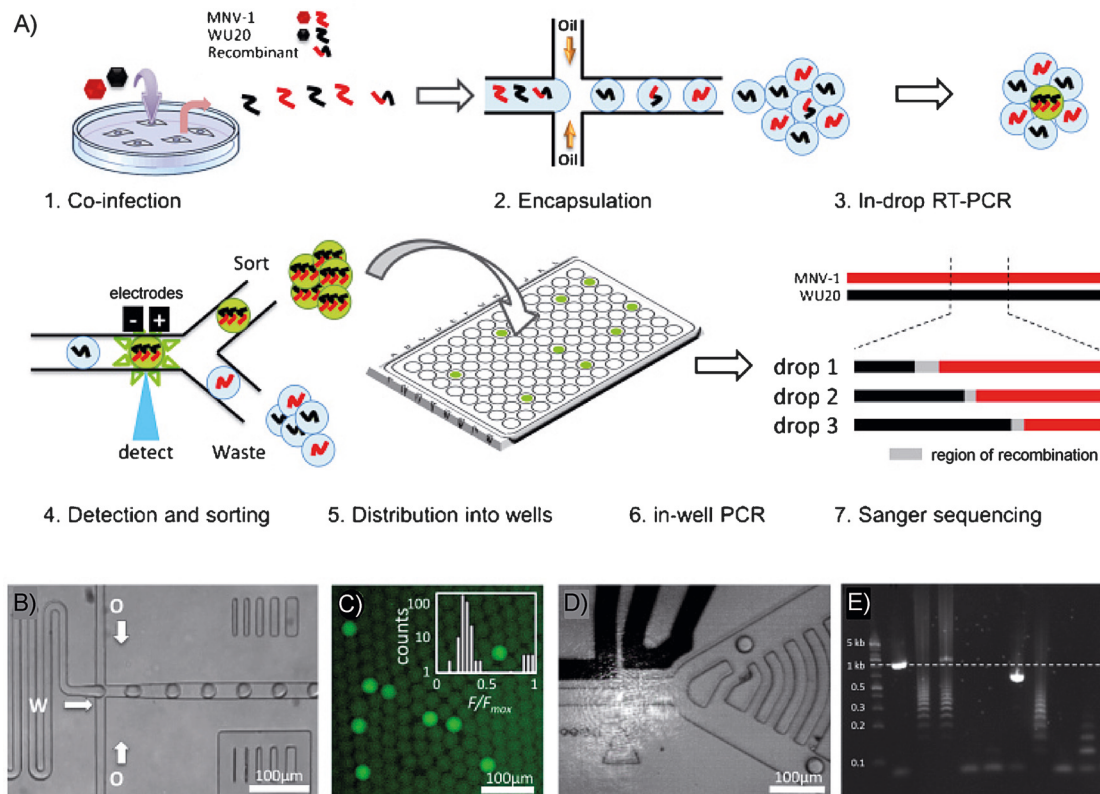


Figure 1. Method overview. A) Schematic of experimental flow. 1) Viral recombinants are generated by co-infecting a RAW 264.7 cell culture with two MNV parental strains, MNV-1 and WU20; 2) Single viral genomes from cell lysate are mixed with a one-step RT-PCR solution and a DNA-intercalating dye; viral solution is injected into a microfluidic drop-maker device; the oil streams then pinch off this viral solution for encapsulation in picoliter water-in-oil drops. 3) Collection of the generated drops and thermocycling off-chip to amplify the recombination fragments. 4) Re-injection of drops into a microfluidic fluorescent drop reader and sorter; drops containing amplicons from the recombinant genomes (bright) are sorted. 5) Distribution of sorted bright drops into a 96-well plate. 6) Second-round PCR, extraction of amplicons of the correct length from the gel, and submission to Sanger sequencing. 7) The region of template switching between parental genomes is determined by comparing the recovered sequence against the two reference genomes. B) Image of the microfluidic drop-maker device. Two oil streams (O) pinch off the flowing viral solution (W), for encapsulation in picoliter water-in-oil drops (see also Figure S2). C) Fluorescence image of drops after RT-PCR (inset: distribution of average fluorescence of each drop). D) Microfluidic fluorescent drop reader and sorter: electrodes are black, laser spot is visible above the triangle mark. E) Gel electrophoresis of the amplicons from nine wells of the plate. Dashed line marks the expected amplicon length (1205 bp). Amplicons of the correct length are extracted from the gel, and sent for Sanger sequencing.

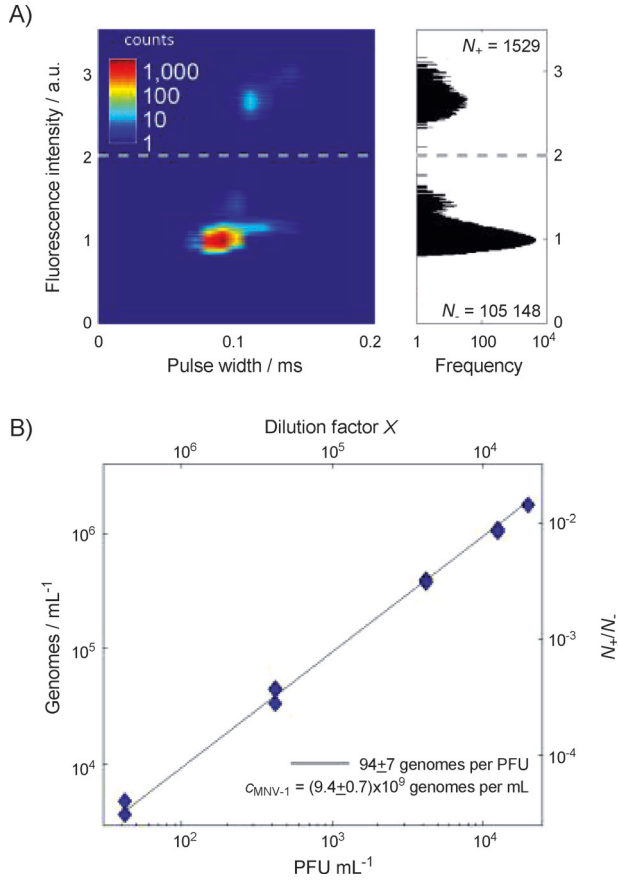


Figure 2. Quantification of the viral concentration of a MNV-1 stock. A) Heat map showing the peak and width of fluorescence pulses from drops passing through the laser beam of the microfluidic drop sorter. Drops with fluorescent peaks exceeding F_0 (----) contain viral RNA. Right: distribution of the peak fluorescence, where N_+ and N_- are the number of bright and dark drops, respectively (threshold F_0). B) Dependence of the fraction of bright to dark drops (N_+/N_-) on the dilution factor X . The concentration of plaque-forming units (PFU mL^{-1}) is based on the initial stock concentration (10^8 PFU mL^{-1}) and the concentration of MNV-1 genomes, $N_+/N_-/V$, where V is the drop volume (8.25 pL). A linear fit provides both the stock concentration of MNV-1 genomes ($c_{\text{MNV-1}}$) and the ratio of genomes to PFU in the stock.

tions: $c_{\text{MNV-1}} = 9.4 \times 10^9$ and $c_{\text{WU20}} = 4.7 \times 10^{10}$ genome copies per mL (Figure 2B). Assuming a conversion factor of 1 PFU (plaque-forming unit) per 100 viral genomes,^[11] these concentrations are consistent with their infectivity titers (10^8 and $5 \times 10^8 \text{ PFU mL}^{-1}$, respectively).

We modified the above assay to determine the concentration of recombinants resulting from a macrophage cell-culture co-infected with a mixture of both parental viruses, each with a multiplicity of infection of 2. We use a one-step differential RT-PCR cocktail^[12] where each of the two primers is specific to only one parental virus, in order to selectively amplify those recombinants where template switching occurs in a selected 1205 bp region (Figure 3A; the top row of the gels confirms that our cocktail amplified recombinants from the co-infection without amplifying parental viruses in the viral mixture). For our drop-based assay, the differential RT-PCR cocktail was co-

encapsulated with viral templates, and in-drop RT-PCR was performed off-chip. In order to register amplicons, the cocktail was supplemented with a DNA-intercalating dye, EvaGreen (Biotium, Hayward, CA). We observed a broad distribution of fluorescent drops, as expected when using a non-specific DNA probe that registers both specific and non-specific target amplification (Figure 3B).

To increase the fidelity of our recombinant assay and exclude false bright drops containing non-specific amplicons we used a gel-verification procedure. Drops whose fluorescence exceeded the threshold ($F_0 = 9.5$; Figure 3B) were sorted in a microfluidic device. The sorted drops were mixed with a sufficient volume of buffer drops and distributed into many micro-wells, such that at most one fluorescent drop was in each well. A second round of differential in-well PCR was performed to produce enough amplicons for gel electrophoresis, in order to select drops containing amplicons of the right length (Figure 1E). The fluorescence threshold for sorting was set sufficiently low to minimize loss of drops containing recombinants (absence of a visible target band for drops with $F < F_0$; Figure 3A, lower gels). Discarding recombinants of the wrong length is justified because only potentially viable viruses are considered,^[13] however, amplicons of the right length are not guaranteed to be true recombinants as they could be the result of in-drop chimera generation.

Because in-drop chimera generation requires the presence of multiple templates in the same drop, chimera generation can be sufficiently repressed by diluting the template such that the probability of any drop containing multiple templates is negligible. We assayed a dilution series of a recombinant-free mixture of both parental viruses, where target amplicons can only result from in-drop chimera generation. For $\lambda \geq 0.3$, the fraction of gel-verified drops exhibits a quadratic trend, $R_{\text{chimera}} \propto \lambda^2$, which is expected in the limit of small λ , and is shown by the filled blue squares and solid blue line in Figure 3C. We found no chimera drops for $\lambda = 0.02$ and 0.1 , thus establishing a safe range for chimera-free RT-PCR.

We determined the recombination frequency (the fraction of recombinant genomes produced in one virus replication cycle) in co-infected cells from a dilution series experiment with a two-step recombinant assay. To estimate the recombinant frequency we made two assumptions: the viral yield of the infection far exceeds that of the initial inoculum, and almost all of the cells are infected by both strains due to the high viral loading. The recombination frequency (P_{rec}) is thus the ratio of the fraction of recombinant drops (R_{rec}) to the concentration of parental virus, λ (solid triangles in Figure 3C). To correct for chimeras, we estimated the chimera frequency (R_{chimera}) from the ratio of the fraction of gel-verified chimera drops (blue squares in Figure 3C) to that of drops containing both parental viruses ($(1 - e^{-(\lambda/2)^2})^2$) and obtained $P_{\text{chimera}} = R_{\text{chimera}} / (1 - e^{-(\lambda/2)^2}) = 1.2 \times 10^{-4} \pm 1 \times 10^{-5}$ (squares in Figure 3D). Accordingly the corrected recombinant frequency is $P_{\text{rec}} = (R_{\text{rec}} - R_{\text{chimera}}) / \lambda = 3.3 \times 10^{-4} \pm 2 \times 10^{-5}$ (triangles in Figure 3D).

For a detailed analysis of recombinants we used Sanger sequencing with amplicons from nine gel-verified drops isolated in microwells; we obtained unambiguous traces, similar to

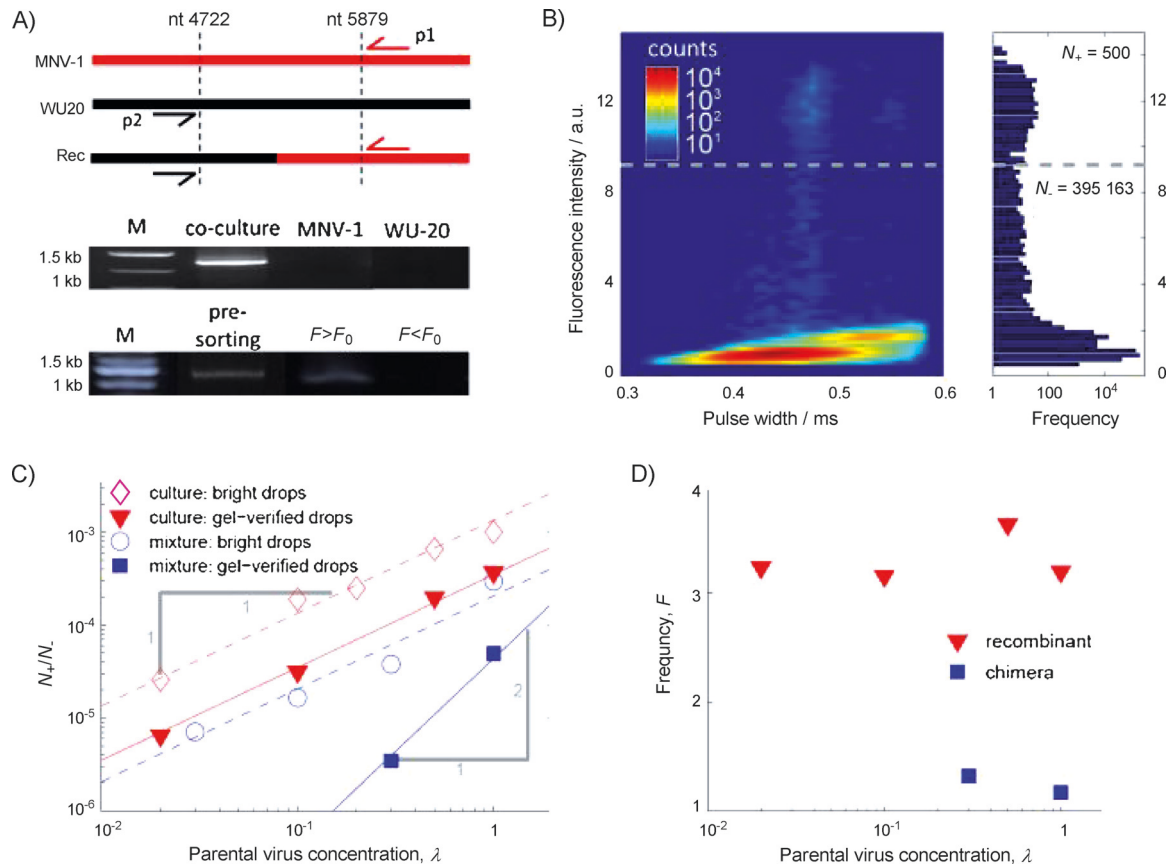


Figure 3. Detection of recombinant viruses. A) Top: primer design for amplifying recombinants where template switching occurs (between bp 4722 and 5879). Middle: gel electrophoresis of bulk RT-PCR of lysates from the co-infected cell culture and the parental stock solutions (MNV-1, WU20). M: DNA ladder. Bottom: gel electrophoresis of unsorted drops (pre-sorting), 500 bright drops with $F > F_0$ and the remaining ~350,000 drops with $F < F_0$. B) Heat map showing the peak and width of fluorescence pulses. The threshold for sorting is indicated by the dashed line. Right: distribution of the peak fluorescence of all pulses detected. N_+ : bright drops with $F > F_0$; N_- : remaining dark drops (parental virus concentration, $\lambda = 1$). C) Dependence of the fraction of bright and gel-verified bright drops (N_+/N_-) on the parental virus concentration (λ) for both the lysate from co-infected cell culture and from a mixture of parental viral genomes. D) Dependence of in-drop chimera (blue) and recombinant (red) frequency on the parental virus concentration.

those for clone-based sequencing.^[14,15] The sequences align with WU20 genome at the 5'-end, and with MNV-1 at the 3'-end; in between is the region of recombination where template switching occurred. We identified four unique regions in the sequenced drops (Figure S3). For example, the region of recombination of Drop-1 spans nucleotides 4950–4968 and for Drop-9 it spans nucleotides 5240–5255 (Figure 4A and B). Interestingly, Drop-9 also includes a synonymous transversion at nucleotide 5221. Using this method, the first recombinant (from Drop-1 sequence) was also detected from in vivo samples by examining the feces of mice that were co-infected with both parental strains.^[12]

Our drop-based microfluidics technique combines two novel steps to achieve bias- and artifact-free characterization of rare RNA recombinants in the presence of a large background of genomic material. Single-step in-drop differential RT-PCR faithfully replicates low-abundant target templates with negligible chimera generation. Subsequent screening isolates

single drops containing recombinant amplicons for consensus sequencing to provide bias-free genomic data. To our knowl-

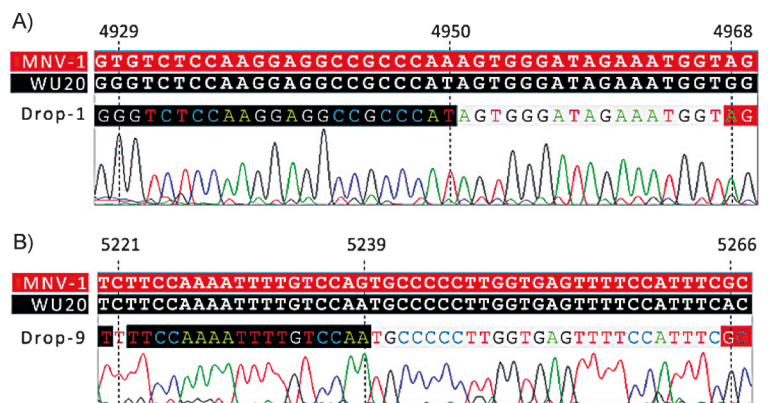


Figure 4. Two recombinant sequences obtained by Sanger sequencing. Recombination regions (gray background) exhibit template switching. The preceding WU20 parental sequence is highlighted with a black background; the downstream MNV-1 parental sequence is marked with a red background.

edge, this is the first use of in-drop single-step RT-PCR with a non-specific fluorescent probe; this is essential for assays such as our search for recombinants where the target sequence is unknown. The continuous isolation of single templates for amplification, detection, and sequencings represents an improvement over all methods for characterizing rare recombinant templates (Table S1) and provides detailed information on both the frequency and diversity of recombinants.^[15,16] Moreover, our technique can be adapted for studies of non-specific rare genomic events, such as lateral gene transfer in bacteria, homologous recombination in parasites, and V(D)J recombination in the mammalian immune system.^[17]

Experimental Section

Experimental details can be found in the Supporting Information.

Acknowledgements

Financial support for this work was from the U.S. Defense Advanced Research Projects Agency (contract number HR0011-11-C-0093). Funding for open access charge: Defense Advanced Research Projects Agency. We thank Nicholas S. Giacobbi for helpful discussions.

Keywords: drop-based microfluidics • error-free genomic amplification • RT-PCR • sequence determination • viruses

- [1] a) S. K. Gire, A. Goba, K. G. Andersen, R. S. G. Sealfon, D. J. Park, L. Kanneh, S. Jalloh, M. Momoh, M. Fullah, G. Dudas, S. Wohl, L. M. Moses, N. L. Yozwiak, S. Winnicki, C. B. Matranga, C. M. Malboeuf, J. Qu, A. D. Gladden, S. F. Schaffner, X. Yang, et al., *Science* **2014**, *345*, 1369–1372; b) X. Qiu, G. Wong, J. Audet, A. Bello, L. Fernando, J. B. Alimonti, H. Fausther-Bovendo, H. Wei, J. Aviles, E. Hiatt, A. Johnson, J. Morton, K. Swope, O. Bohorov, N. Bohorova, C. Goodman, D. Kim, M. H. Pauly, J. Velasco, J. Pettitt, et al., *Nature* **2014**, *514*, 47–53; c) M.-F. Yuen, C.-L. Lai, *J. Gastroenterol. Hepatol.* **2011**, *26*, 138–143.
- [2] a) T. T.-K. Lam, H. Zhu, D. K. Smith, Y. Guan, E. C. Holmes, O. G. Pybus, *The Journal of general virology* **2012**, *93*, 817–822; b) D. Khatchikian, M. Orlich, R. Rott, *Nature* **1989**, *340*, 156–157; c) E. Simon-Loriere, E. C. Holmes, *Nat. Rev. Microbiol.* **2011**, *9*, 617–626.
- [3] a) C. Cases-González, M. Arribas, E. Domingo, E. Lázaro, *J. Mol. Biol.* **2008**, *384*, 1120–1129; b) L. Cabanillas, M. Arribas, E. Lázaro, *BMC Evol. Biol.* **2013**, *13*, 11; c) S. C. Manrubia, C. Escarmis, E. Domingo, E. Lázaro, *Gene* **2005**, *347*, 273–282; d) J. W. Drake, J. J. Holland, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13910–13913.
- [4] a) F. Di Giallonardo, O. Zagordi, Y. Duport, C. Leemann, B. Joos, M. Kunzli-Gontarczyk, R. Bruggmann, N. Beerenwinkel, H. F. Gunthard, K. J. Metzner, *PLoS One* **2013**, *8*, e74249; b) M. Mild, C. Hedskog, J. Jernberg, J. Albert, *PLoS One* **2011**, *6*, e22741; c) W. Shao, V. F. Boltz, J. E. Spindler, M. F. Kearney, F. Maldarelli, J. W. Mellors, C. Stewart, N. Volfovsky, A. Levitsky, R. M. Stephens, J. M. Coffin, *Retrovirology* **2013**, *10*, 18; d) R. P. Smyth, T. E. Schlub, A. Grimm, V. Venturi, A. Chopra, S. Mallal, M. P. Davenport, J. Mak, *Gene* **2010**, *469*, 45–51; e) A. Routh, P. Ordoukhanian, J. E. Johnson, *J. Mol. Biol.* **2012**, *424*, 257–269.
- [5] D. Dressman, H. Yan, G. Traverso, K. W. Kinzler, B. Vogelstein, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8817–8822.
- [6] a) M. A. Quail, M. Smith, P. Coupland, T. D. Otto, S. R. Harris, T. R. Connor, A. Bertoni, H. P. Swerdlow, Y. Gu, *BMC Genomics* **2012**, *13*, 341; b) L. Liu, Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu, M. Law, *J. Biomed. Biotechnol.* **2012**, *2012*, 251364; c) D. C. Koboldt, K. M. Steinberg, D. E. Larson, R. K. Wilson, E. R. Mardis, *Cell* **2013**, *155*, 27–38.
- [7] a) E. A. Ottesen, J. W. Hong, S. R. Quake, J. R. Leadbetter, *Science* **2006**, *314*, 1464–1467; b) H. C. Fan, Y. J. Blumenfeld, Y. Y. El-Sayed, J. Chueh, S. R. Quake, *Am. J. Obstet. Gynecol.* **2009**, *200*, 543.e1–543.e7.
- [8] a) A. Rotem, O. Ram, N. Shores, R. A. Sperling, M. Schnall-Levin, H. Zhang, A. Basu, B. E. Bernstein, D. A. Weitz, *PLoS One* **2015**, *10*, e0116328; b) A. E. Fischer, S. K. Wu, J. B. G. Proeschner, A. Rotem, C. B. Chang, H. Zhang, Y. Tao, T. S. Mehoke, P. M. Thielen, A. O. Kolawole, T. J. Smith, C. E. Wobus, D. A. Weitz, J. S. Lin, A. B. Feldman, J. T. Wolfe, *J. Virol. Methods* **2014**, *213*, 111–117; c) M. T. Guo, A. Rotem, J. A. Heyman, D. A. Weitz, *Lab Chip* **2012**, *12*, 2146–2155.
- [9] a) D. F. Jarosz, J. C. S. Brown, G. A. Walker, M. S. Datta, W. L. Ung, A. K. Lancaster, A. Rotem, A. Chang, G. A. Newby, D. A. Weitz, L. F. Bisson, S. Lindquist, *Cell* **2014**, *158*, 1083–1093; b) L. Mazutis, J. Gilbert, W. L. Ung, D. A. Weitz, A. D. Griffiths, J. A. Heyman, *Nat. Protoc.* **2013**, *8*, 870–891; c) J. J. Agresti, E. Antipov, A. R. Abate, K. Ahn, A. C. Rowat, J.-C. Baret, M. Marquez, A. M. Klibanov, A. D. Griffiths, D. A. Weitz, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4004–4009; d) A. R. Abate, T. Hung, P. Mary, J. J. Agresti, D. A. Weitz, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 19163–19166; e) D. R. Link, E. Grasland-Mongrain, A. Duri, F. Sarrazin, Z. Cheng, G. Cristobal, M. Marquez, D. A. Weitz, *Angew. Chem. Int. Ed.* **2006**, *45*, 2556–2560; *Angew. Chem.* **2006**, *118*, 2618–2622.
- [10] L. Mazutis, A. F. Araghi, O. J. Miller, J.-C. Baret, L. Frenz, A. Janoshazi, V. Taly, B. J. Miller, J. B. Hutchison, D. Link, A. D. Griffiths, M. Ryckelynck, *Anal. Chem.* **2009**, *81*, 4813–4821.
- [11] L. Baert, C. E. Wobus, E. Van Coillie, L. B. Thackray, J. Debevere, M. Uytendaele, *Appl. Environ. Microbiol.* **2008**, *74*, 543–546.
- [12] H. Zhang, S. K. Cockrell, A. O. Kolawole, A. Rotem, A. W. R. Serohijos, C. B. Chang, Y. Tao, T. S. Mehoke, Y. Han, J. S. Lin, N. S. Giacobbi, A. B. Feldman, E. Shakhnovich, D. A. Weitz, C. E. Wobus, J. M. Pipas, *J. Virol.* **2015**, *89*, 7722–7734.
- [13] K. M. Spann, P. L. Collins, M. N. Teng, *J. Virol.* **2003**, *77*, 11201–11211.
- [14] F. M. Ausubel, *Current Protocols in Molecular Biology*, Wiley, New York, **1994**, unit 16.1.
- [15] E. Mathijs, B. Muylkens, A. Mauroy, D. Ziant, T. Delwiche, E. Thiry, *J. Gen. Virol.* **2010**, *91*, 2723–2733.
- [16] a) P. McInerney, P. Adams, M. Z. Hadi, *Mol. Biol. Int.* **2014**, *2014*, 287430; b) S. G. Acinas, R. Sarma-Rupavtarm, V. Klepac-Ceraj, M. F. Polz, *Appl. Environ. Microbiol.* **2005**, *71*, 8966–8969.
- [17] a) J. G. Lawrence, A. C. Retchless, *Methods Mol. Biol.* **2009**, *532*, 29–53; b) C. M. Thomas, K. M. Nielsen, *Nat. Rev. Microbiol.* **2005**, *3*, 711–721; c) C. Chen, L. A. Fenk, M. de Bono, *Nucleic Acids Res.* **2013**, *41*, e193; d) C. Zhang, B. Xiao, Y. Jiang, Y. Zhao, Z. Li, H. Gao, Y. Ling, J. Wei, S. Li, M. Lu, X.-z. Su, H. Cui, J. Yuan, *mBio* **2014**, *5*, e01414-14; e) E. V. Koonin, M. Krupovic, *Nat. Rev. Genet.* **2015**, *16*, 184–192; f) M. Rashkovan, C. Vadnais, J. Ross, M. Gigoux, W.-K. Suh, W. Gu, C. Kosan, T. Moroy, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E5411–5419.