Published in final edited form as:

Angew Chem Int Ed Engl. 2011 January 10; 50(2): 390–395. doi:10.1002/anie.201006089.

Single Cell Multiplex Gene Detection and Sequencing Using Microfluidically-Generated Agarose Emulsions**

Richard Novak[†],

Center for Exposure Biology, University of California, Berkeley, 307 Lewis Hall, Berkeley, California 94720 (USA). UC Berkeley/UC San Francisco Joint Graduate Group in Bioengineering, University of California, Berkeley, 307 Lewis Hall, Berkeley, California 94720 (USA)

Dr. Yong Zeng[†],

Center for Exposure Biology, University of California, Berkeley, 307 Lewis Hall, Berkeley, California 94720 (USA). Department of Chemistry, University of California, Berkeley, 307 Lewis Hall, Berkeley, California 94720 (USA)

Dr. Joe Shuga,

Center for Exposure Biology, University of California, Berkeley, 307 Lewis Hall, Berkeley, California 94720 (USA). School of Public Health, University of California, Berkeley, 211 Hildebrand Hall, Berkeley, California 94720 (USA)

Gautham Venugopalan,

UC Berkeley/UC San Francisco Joint Graduate Group in Bioengineering, University of California, Berkeley, 307 Lewis Hall, Berkeley, California 94720 (USA)

Prof. Daniel A. Fletcher,

UC Berkeley/UC San Francisco Joint Graduate Group in Bioengineering, University of California, Berkeley, 307 Lewis Hall, Berkeley, California 94720 (USA)

Prof. Martyn T. Smith, and

Center for Exposure Biology, University of California, Berkeley, 307 Lewis Hall, Berkeley, California 94720 (USA). School of Public Health, University of California, Berkeley, 211 Hildebrand Hall, Berkeley, California 94720 (USA)

Prof. Richard A. Mathies^{*}

Center for Exposure Biology, University of California, Berkeley, 307 Lewis Hall, Berkeley, California 94720 (USA). UC Berkeley/UC San Francisco Joint Graduate Group in Bioengineering, University of California, Berkeley, 307 Lewis Hall, Berkeley, California 94720 (USA). Department of Chemistry, University of California, Berkeley, 307 Lewis Hall, Berkeley, California 94720 (USA)

Genetic assays, such as polymerase chain reaction (PCR), typically report on multiple cells or mixtures of genomic DNA. As a result, they cannot properly characterize the genetic heterogeneity of a cell population or detect the co-occurrence of different mutations within a single cell, which are key to understanding the development, progression, and treatment of cancers.[1,2] In particular, since initial mutagenesis occurs inherently at the single-cell level, the detection and characterization of carcinogenesis will be dramatically facilitated by analytical techniques with single cell resolution.

**This work was supported by the trans-NIH Genes, Environment and Health Initiative, Biological Response Indicators of Environmental Systems Center Grant U54 ES016115-01 to MTS and RAM, and by Superfund Basic Research Program NIEHS Grant P42 ES004705 to MTS. RN is supported by an NSF Graduate Research Fellowship, and JS is supported by the Canary Foundation and ACS Early Detection Postdoctoral Fellowship.

^{*}Fax: (+1) 510 642-3599, ramathies@berkeley.edu.

[†]These authors contributed equally to this work.

Cytometric sorting, limiting dilution, and micromanipulation have been previously used to perform single-cell PCR assays in 96-well PCR plates, but these approaches are not ideal for large-scale screening applications.[3] Microfluidic technology offers fundamentally new capabilities for manipulating fluids, molecules, and cells that are very pertinent for developing high-throughput single-cell analysis methods.[4–7] Microfluidic droplet technology is particularly advantageous for single-cell/molecule analysis because it facilitates rapid statistical compartmentalization of targets for massively parallel pico- to nanoliter scale assays.[8–10] In particular, microfluidic emulsion PCR (ePCR) enables high-fidelity digital single-molecule counting due to its unique ability to ensure equal population sampling and amplification efficiencies across all reaction compartments.[4,11,12]

To date most single cell genomic analysis have been reported on bacterial samples.[13–15] For mammalian cells, droplet-based genetic analyses have predominantly implemented reverse transcriptase PCR for phenotypic profiling.[11,16] A difficulty in single cell PCR is the persistent technical challenge of integrating a robust and scalable DNA extraction method.[13,14] The relative lack of suitable single-cell genomic analysis technologies combined with the significant genetic heterogeneity associated with cancer, underscores the importance of developing new microdroplet methodologies that integrate robust single-cell genome preparation with multiplex PCR.

To address these challenges, we have developed an agarose droplet-based platform that leverages emulsion generator array technology for high-throughput single cell genetic analysis.[4,12] Single cells were microfluidically encapsulated together with primer-functionalized beads in agarose gel droplets for subsequent SDS lysis and proteinase K digestion to release genomic DNA. Using the co-encapsulated primer beads and purified genomes, we demonstrate multi-locus single-cell sequencing of the control gene β -actin and across the chromosomal translocation t(14;18), a mutation associated with 85–90% of follicular lymphoma cases.[1,17] The coupling of our robust and high-throughput single cell DNA purification method with sequencing of multiple gene targets within single cells will enable detailed studies of mutation co-occurrence and synergy during carcinogenesis.

The underlying principle of our highly parallel cell digestion and DNA purification method is the microfluidic encapsulation of cells in agarose droplets (Figure 1) to maintain single genome fidelity during cell lysis and DNA purification as well as efficient multiplex emulsion PCR target amplification for subsequent analysis. Single lymphoblast cells were encapsulated along with primer-functionalized beads in 1.5% low-melt point agarose using a four-channel microfluidic emulsion generator array (MEGA) (Figure 2A).[4] Micropump actuation was optimized to account for the increased viscosity of molten agarose without modifications of the microfluidic design, demonstrating the versatility of the MEGA platform. Droplet generation at ~40 °C for 30 min resulted in the encapsulation of approximately 18,000 cells at up to 0.3 cells per droplet (cpd) on average. Figure 2B shows the generation of uniform 3-nL agarose droplets containing primer-functionalized beads. The inset highlights an example of cell and bead co-encapsulation in a single droplet.

Agarose droplet encapsulation enables reproducible single-cell DNA extraction and isolation by allowing us to adapt robust DNA purification protocols. An example of this process is depicted in Figure 2C, where a bead and a cell were co-encapsulated in one agarose droplet. DNA release was accomplished by overnight incubation of the gel droplets in SDS lysis buffer and proteinase K to lyse the cell and digest DNA-binding histone proteins. The void left by the cell in the agarose is occupied by brightly-fluorescent genomic DNA that exhibits minimal diffusion into the surrounding gel (Figure 2D), due to the relatively small pore size of 1.5% agarose (~130 nm).[18]. An incubation temperature of 52°C facilitated enzymatic protein digestion while preserving the integrity of agarose droplets. By staining with

propidium iodide, we were able to visualize single high-molecular weight DNA strands protruding from the relatively small agarose droplets (Figure S1D) indicating that the majority of nuclear proteins were removed by the combination of proteinase K and SDS. The agarose droplets were stable for at least one week when stored in ethanol at 4°C as determined by confocal imaging of DNA diffusion radii.

A key benefit of agarose encapsulation is the ability to perform mechanical manipulation of the isolated genomic DNA without mixing the genetic contents of different cells. Agarose droplets equilibrated with PCR mix were re-emulsified by mechanically agitating in dispersing oil to produce uniform nanoliter droplet “reactors” for massively parallel single cell PCR analysis. Excess PCR mix produces microfines (emulsions < 1µm diameter) that enhance emulsion stability during thermal cycling.[19] The agarose droplets melt during the hot start phase of PCR and remain liquid throughout the amplification process, maximizing reagent and amplicon diffusion rates. We varied the initial heating rate and tested various concentrations of Triton X-100 as well as combinations of Abil em90 and Span 80 detergents in oil. A slow temperature ramp profile (0.1 °C s⁻¹) resulted in improved short-term stability, while the addition of 4 mg mL⁻¹ BSA in PCR mix and 0.8% Triton X-100 in emulsion oil minimized droplet merger over the course of PCR thermal cycling (Figure 2E). Fluorescently-labeled amplicons bound to the primer-functionalized beads could be seen inside the agarose droplets following PCR (Figure 2F). The 34 µm crosslinked beads were selected due to their ability to amplify targets exceeding 1 kb in amounts of at least 100 amol per bead.[12] Beads in droplets without genomes did not fluoresce, indicating the absence of transfer of genomic targets among agarose droplets.

To demonstrate highly parallel genotyping with single cell resolution, we performed a multiplex PCR assay of cancer cells harboring the t(14;18) translocation at various mutant (RL)-to-wild type (TK6) cell ratios. The Cy5-labeled control gene product, β-actin, enabled quantification of total cell frequency, while the FAM-labeled translocation t(14;18) product, spanning the bcl-2 and IgH genes across their breakpoint regions, determined mutation frequency. A representative flow cytometric profile of beads following multiplex PCR amplification obtained with 50% RL cells at an average cell frequency of 0.3 cpd (Figure 3A) demonstrated distinct populations of negative, Cy5-labeled β-actin only positive, and FAM-labeled t(14;18) and Cy5-labeled β-actin double positive beads. Beads containing t(14;18) only were never seen, further indicating the conservation of single-genome integrity during cell lysis or PCR. By maintaining a constant total cell density in the single cell regime (0.1 to 0.3 cpd on average) and varying the relative concentrations of mutant RL and wild-type TK6 cells, we generated a standard curve (Figure 3B) to confirm that amplification originated from single cells. In this stochastic regime, linearity (r = 0.993) of the measured mutant RL cell concentration with respect to the input in the 0% to 100% RL cell frequency range tested indicated successful genetic analysis of single cells. Importantly, in the subset of samples further tested, the ratio of total amplicon-positive bead frequency determined by flow cytometry to the microscopically observed cell encapsulation frequency indicated high PCR efficiency (113± 24 %).

Single-cell sequencing of both target gene loci was accomplished by reamplifying the template bound to single beads diluted to the stochastic limit in a standard 96-well plate and separating products from individual wells using gel electrophoresis. Gel analysis further confirmed amplification from single genomes, and frequencies of β-actin single positive to double positive events matched flow cytometry frequency results in the samples tested. Fluorescence-based sorting can be applied to amplicon-labeled beads to sequence only populations of interest and thereby increase efficiency. Size-separated products from single beads were excised from the gel for sequencing. Figure 4 demonstrates recovery of the expected t(14;18) and β-actin sequences (see also Figure S2 and S3). The random nucleotide

insertion sequence in t(14;18) matched the unique translocation “fingerprint” determined by sequencing of RL cells in bulk.[17]

The reamplification step enables integration of single cell PCR with standard molecular biology techniques to achieve sequencing of multiple genetic loci in individual cells at a rate required for meaningful population analysis. Using the single bead reamplification and size separation approach, it is possible to sequence multiple target genes from hundreds of single cells in a single experiment and to obtain meaningful statistical analysis of gene sequence variation in a cell population. Furthermore, the number of target loci per cell is limited only by the ability to excise individual bands from a gel, provided the multiplex PCR has been validated. Adaptation of our agarose cell encapsulation and emulsion PCR method to next-generation sequencing technologies has the potential to provide additional gains in throughput for single cell sequencing. Although alternative single cell-resolution genetic analysis approaches, such as fluorescence *in situ* hybridization[20,21] and *in situ* PCR, [22,23] are utilized for investigating cell mutation progression, the throughput offered by these methods is limited, and the lack of amplicon recovery prevents detection of unknown mutations. In the case of follicular lymphoma, identifying the breakpoint sequence along the IgH and bcl-2 genes can help elucidate the mechanisms of erroneous genetic recombination that cause the translocation t(14;18), while the random insertion sequence between the two chromosomes is a unique identifier of distinct mutation events.[17]

Encapsulation enables robust parallel cell lysis and DNA purification of cell types that are difficult to screen using other single-cell PCR methods. The method can be adapted in future applications to encapsulate live single cells for cell culture and subsequent analysis of clonal populations.[24] Although a similar approach has been demonstrated recently for screening and amplifying alginate-encapsulated *E. coli* colonies containing plasmid libraries,[24] our approach enables robust genome purification of mammalian cells, uses 1 % of the reagent volume for emulsion PCR, does not require droplet sorting, and maintains single-cell segregation throughout all downstream analyses by incorporating primer-functionalized beads as amplicon substrates. Furthermore, single-cell sequencing techniques generally involve cell types that are amenable to lysis during PCR and sorting of single cells into 96-well PCR plates.[25,26] However, PCR amplification directly from single cells has been hindered by the lack of a DNA purification step that would remove histones and other nuclear components that inhibit polymerase activity. [25,27] In our approach, the incorporation of a highly parallel single-cell lysis and DNA purification step resulted in near 100% amplification efficiency. Finally, the ability to purify genomic DNA from single cells in a supporting matrix opens up the possibility of single cell epigenetic analysis, such as methylation-specific PCR of gene regulatory sequences following bisulfite conversion.

Cancer is an evolving disease driven by genetic instability, which results in a constant clonal divergence of the tumor cell population. Accumulation of mutations in the genes coding for cellular pathways plays an important role in carcinogenesis, metastasis, and therapeutic resistance.[28–30] Significant cellular heterogeneity may therefore exist in tumors; this heterogeneity dynamically changes at different stages of disease progression.[21] We have developed a robust agarose microdroplet method for detecting and sequencing multiple genetic targets from single cells in a scalable manner. While mutations in oncogenes such as TP53, CDKN2A, CDKN2B, and bcl-6 have been demonstrated to individually correlate with poor prognosis in hematopoietic cancers,[31–35] the ability to sequence multiple targets within single cells in high throughput will facilitate investigation of synergistic effects of mutation co-occurrence and their impact on disease progression and treatment. Furthermore, screening of large cell populations will uncover potential tumor heterogeneity at the single-nucleotide level that is otherwise obscured by the ensemble average. By overcoming the many problems associated with single-cell genetic analysis,[36,37] including non-uniform

cell lysis, DNA release, and amplification as well as low throughput, our single-cell agarose encapsulation technology provides the throughput, robustness, and resolution required for probing the stochastic mechanisms of carcinogenesis, progression, and response to chemotherapy.

References

1. d'Amore F, Chan E, Iqbal J, Geng H, Young K, Xiao L, Hess MM, Sanger WG, Smith L, Wiuf C, et al. *Clin Canc Res*. 2008; 14:7180–7187.
2. Riethdorf S, Wikman H, Pantel K. *Int J Canc*. 2008; 123:1991–2006.
3. Vogelstein B, Kinzler KW. *Proc Natl Acad Sci USA*. 1999; 96:9236–9241. [PubMed: 10430926]
4. Zeng Y, Novak R, Shuga J, Smith MT, Mathies RA. *Anal Chem*. 2010; 82:3183–3190. [PubMed: 20192178]
5. Andersson H, Berg AVD. *Curr Opin Biotech*. 2004; 15:44–49. [PubMed: 15102465]
6. El-Ali J, Sorger PK, Jensen KF. *Nature*. 2006; 442:403–411. [PubMed: 16871208]
7. Toriello NM, Douglas ES, Thaitrong N, Hsiao SC, Francis MB, Bertozzi CR, Mathies RA. *Proc Natl Acad Sci USA*. 2008; 105:20173–20178. [PubMed: 19075237]
8. Huebner A, Sharma S, Srisa-Art M, Hollfelder F, Edel JB, deMello AJ. *Lab Chip*. 2008; 8:1244–1254. [PubMed: 18651063]
9. Schaerli Y, Hollfelder F. *Mol BioSyst*. 2009; 5:1392. [PubMed: 20023716]
10. Clausell-Tormos J, Lieber D, Baret J, El-Harrak A, Miller OJ, Frenz L, Blouwolff J, Humphry KJ, Koster S, Duan H. *Chem & Biol*. 2008; 15:427–437. [PubMed: 18482695]
11. Beer N, Wheeler E, Lee-Houghton L, Watkins N, Nasarabadi S, Hebert N, Leung P, Arnold D, Bailey C, Colston B. *Anal Chem*. 2008; 80:1854–1858. [PubMed: 18278951]
12. Kumaresan P, Yang CJ, Cronier SA, Blazej RG, Mathies RA. *Anal Chem*. 2008; 80:3522–3529. [PubMed: 18410131]
13. Chao T, Ros A. *J R Soc Interface*. 2008; 5:S139–S150. [PubMed: 18682362]
14. Zare RN, Kim S. *Annu Rev Biomed Eng*. 2010; 12:187–201. [PubMed: 20433347]
15. Ottesen EA, Hong JW, Quake SR, Leadbetter JR. *Science*. 2006; 314:1464–1467. [PubMed: 17138901]
16. Kiss MM, Ortoleva-Donnelly L, Beer NR, Warner J, Bailey CG, Colston BW, Rothberg JM, Link DR, Leamon JH. *Anal Chem*. 2008; 80:8975–8981. [PubMed: 19551929]
17. McHale CM, Lan Q, Corso C, Li G, Zhang L, Vermeulen R, Curry JD, Shen M, Turakulov R, Higuchi R, et al. *J Natl Cancer Inst Monogr*. 2008; 2008:74–77. [PubMed: 18648008]
18. Xiong J, Narayanan J, Liu X, Chong TK, Chen SB, Chung T. *J Phys Chem B*. 2005; 109:5638–5643. [PubMed: 16851608]
19. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y, Chen Z, et al. *Nature*. 2005; 437:376–380. [PubMed: 16056220]
20. Hicks J, Muthuswamy L, Krasnitz A, Navin N, Riggs M, Grubor V, Esposito D, Alexander J, Troge J, Wigler M, et al. *Cold Spring Harbor Symp Quant Biol*. 2005; 70:51–63. [PubMed: 16869738]
21. Navin N, Krasnitz A, Rodgers L, Cook K, Meth J, Kendall J, Riggs M, Eberling Y, Troge J, Grubor V, et al. *Genome Res*. 2010; 20:68–80. [PubMed: 19903760]
22. Leppanen PM, Koponen J, Turunen MP, Pakkanen T, Yla-Herttuala S. *J Gene Med*. 2001; 3:173–178. [PubMed: 11318116]
23. Mulrooney PM, Michalak TI. *J Virol*. 2003; 77:970–979. [PubMed: 12502813]
24. Walser M, Pellaux R, Meyer A, Bechtold M, Vanderschuren H, Reinhardt R, Magyar J, Panke S, Held M. *Nucl Acids Res*. 2009; 37:e57. [PubMed: 19282448]
25. Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N. *Proc Natl Acad Sci USA*. 1992; 89:5847–5851. [PubMed: 1631067]
26. Ausubel, FM.; Brent, R.; Kingston, RE.; Moore, DD.; Seidman, J.; Smith, JA.; Struhl, K., editors. *Curr Protoc Mol Biol*. John Wiley & Sons, Inc; Hoboken, NJ, USA: 2001.

27. Goldenberger D, Perschil I, Ritzler M, Altwegg M. *Genome Res.* 1995; 4:368–370.
28. Hanahan D, Weinberg RA. *Cell.* 2000; 100:57–70. [PubMed: 10647931]
29. Vermeulen L, Todaro M, de Sousa Mello F, Sprick MR, Kemper K, Perez Alea M, Richel DJ, Stassi G, Medema JP. *Proc Natl Acad Sci USA.* 2008; 105:13427–13432. [PubMed: 18765800]
30. Navin NE, Hicks J. *Mol Oncol.* 2010; 4:267–283. [PubMed: 20537601]
31. Robles AI, Harris CC. *Cold Spring Harbor Perspect Biol.* 2010; 210.1101/cshperspect.a001016
32. Vitolo U, Botto B, Capello D, Vivenza D, Zagonel V, Gloghini A, Novero D, Parvis G, Calvi R, Ariatti C, et al. *Leukemia.* 2002; 16:268–275. [PubMed: 11840294]
33. Young KH, Leroy K, Moller MB, Colleoni GWB, Sanchez-Beato M, Kerbaux FR, Haioun C, Eickhoff JC, Young AH, Gaulard P, et al. *Blood.* 2008; 112:3088–3098. [PubMed: 18559976]
34. Lossos IS, Levy R. *Semin Canc Biol.* 2003; 13:191–202.
35. O'Shea D, O'Riain C, Taylor C, Waters R, Carlotti E, MacDougall F, Gribben J, Rosenwald A, Ott G, Rimsza LM, et al. *Blood.* 2008; 112:3126–3129. [PubMed: 18628487]
36. Brown RB, Audet J. *J R Soc Interface.* 2008; 5:131–138.
37. Huang H, Bu Y, Zhou G. *World J Gastroenterol.* 2006; 12:3814–3820. [PubMed: 16804964]

Figure 1.

Workflow diagram demonstrating the use of agarose emulsion droplets for genetic analysis and multi-locus sequencing of single mammalian cells. (A) Schematic of the 4-layer glass-PDMS microfluidic emulsion generator array used for agarose droplet formation. A 3-valve micropump actuates an agarose-cell-bead suspension from the sample inlet toward cross-channel nozzles where droplet generation occurs. (B) The genomes of single cells are released in the gelled droplets using a standard SDS lysis/proteinase K digestion protocol. (C) The agarose droplets are equilibrated in PCR buffer containing fluorescent forward primers, emulsified with oil by mechanical agitation, and thermally cycled. (D) Following multiplex PCR amplification, primer beads are released by breaking the emulsion and melting the agarose. The fluorescent beads are then rapidly quantified using flow cytometry or further PCR-amplified for sequencing of target genes.

Figure 2.

Microfluidic agarose encapsulation and genetic analysis of single mammalian cells. (A) Microfluidic emulsion generator array (MEGA) device showing the pneumatic micropump for fluid actuation and the fluidic channels. (B) Agarose emulsion generation showing single cells encapsulated along with primer-labeled beads in uniform 3 nL droplets. Inset highlights a cell (circled) and a bead in one droplet. (C) Differential interference contrast (DIC) image of a cell (arrow) at the edge of an agarose droplet after protease digestion. (D) Projection of confocal micrograph of Figure 2C showing acridine orange stained DNA confined in the agarose after release from the cell (arrow). (E) Following PCR amplification of the chromosomal translocation t(14;18) and β -actin, agarose droplets containing beads (arrows) remain intact and do not merge. (F) Epifluorescence micrograph of Figure 2E demonstrates primer-labeled beads (arrows) becoming fluorescent if a t(14;18)⁺ RL cell is present in the same droplet but otherwise remaining dark. Scale bars are 100 μ m in B, E, and F and 10 μ m in C and D.

Figure 3.

High-throughput digital genetic analysis of cancer cells. (A) A representative flow cytometry plot and gated populations of beads from a sample containing 50% RL lymphoblast cells ($N=1,206$) at a frequency of 0.3 cells per droplet. The observed fraction of total positive events (26.3%) compares favorably with an expected frequency of 25.9% based on the Poisson distribution. (B) Standard curve of $t(14;18)^+$ RL cell detection from mixtures of mutant RL and wild-type TK6 cells. Experimental quantification by multiplex emulsion PCR and flow cytometry exhibits a linear response between the 10% and 100% mutant cell frequencies tested.

Figure 4.

Multiplex single-cell sequencing. Sequencing of two loci from DNA products isolated from single cells was accomplished by reamplifying individual amplicon-labeled beads in separate PCR reactions and separating the two products using gel electrophoresis. This representative image of a section of the gel from a 50% RL cell sample shows amplified t(14;18) and β -actin from single beads (lanes 3 and 6), β -actin-only reactions originating from TK6 cells (lanes 1, 2 and 4), reactions from negative beads (lane 5), and a 100 bp DNA ladder. A lack of t(14;18) amplification without β -actin indicates true single cell analysis. Both products were excised and sequenced to obtain sequence data from two genes colocalized in a single cell. The breakpoint locations on chromosomes 14 and 18 as well as the unique insertion sequence confirmed the recovery of the expected RL lymphoblast sequence. By integrating single cell emulsion PCR with a standard sequencing technique, it is possible to sequence multiple genes from hundreds of single cells in the span of 2–3 days, opening up the possibility for population-level analyses of genetic heterogeneity.