HIGHLIGHTS

- Drop-seq enables highly parallel analysis of individual cells by RNA-seq
- Drop-seq encapsulates cells in nanoliter droplets together with DNA-barcoded beads
- Systematic evaluation of Drop-seq library quality using species mixing experiments
- Drop-seq analysis of 44,808 cells identifies 39 cell populations in the retina

CHEMGENES CORPORATION

FOR A COMPLETE PRODUCT LISTING, VISIT US AT
WWW.CHEMGENES.COM

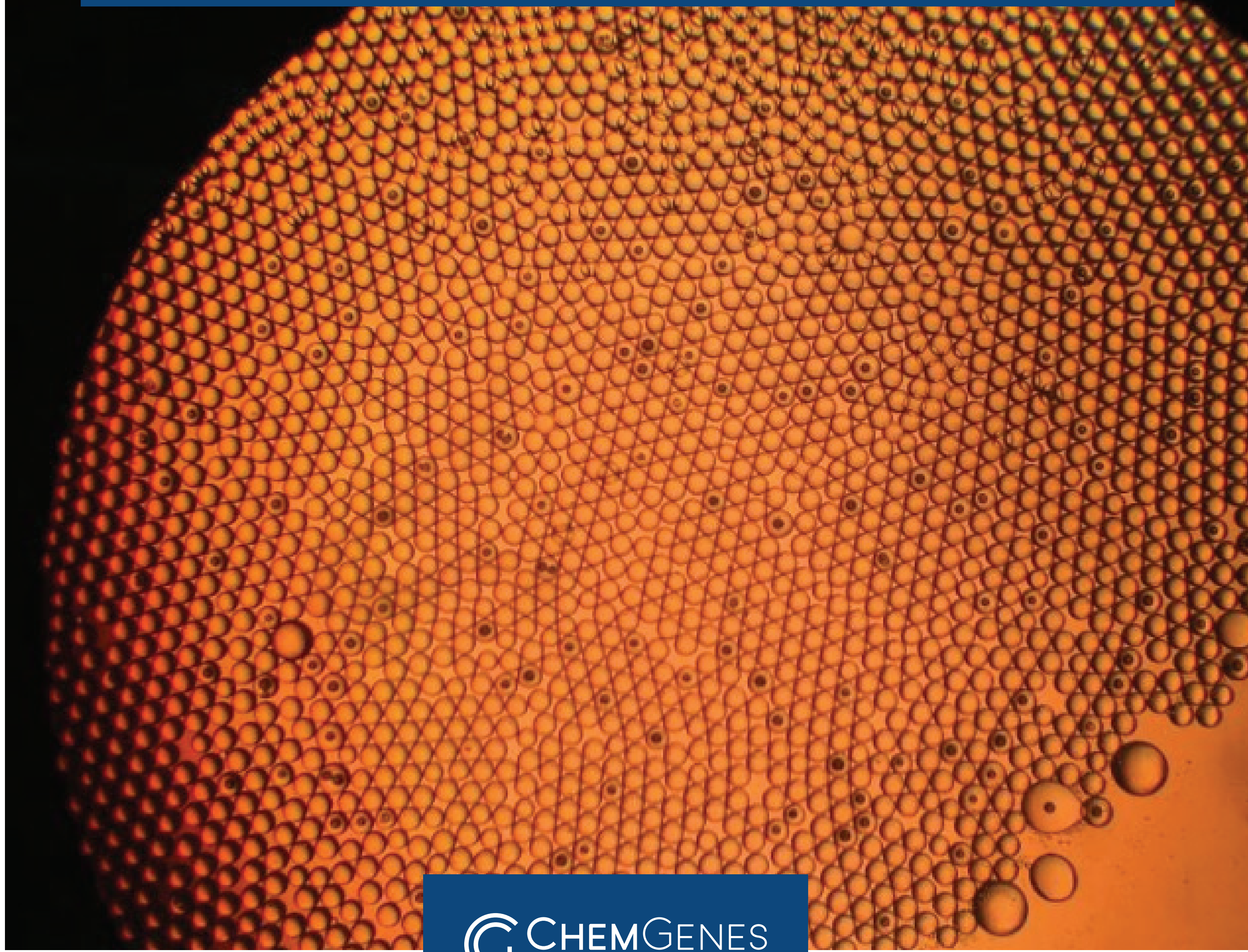
DROP-SEQ

HIGHLY PARALLEL GENOME-WIDE EXPRESSIONPROFILING OF INDIVIDUAL CELLS USING NANOLITER DROPLETS

CHEMGENES CORPORATION

DROP-SEQ

HIGHLY PARALLEL GENOME-WIDE EXPRESSIONPROFILING OF INDIVIDUAL CELLS USING NANOLITER DROPLETS



CHEMGENES CORPORATION

HIGHLIGHTS

- Drop-seq enables highly parallel analysis of individual cells by RNA-seq
- Drop-seq encapsulates cells in nanoliter droplets together with DNA-barcoded beads
- Systematic evaluation of Drop-seq library quality using species mixing experiments
- Drop-seq analysis of 44,808 cells identifies 39 cell populations in the retina

CHEMGENES CORPORATION

FOR A COMPLETE PRODUCT LISTING, VISIT US AT
WWW.CHEMGENES.COM

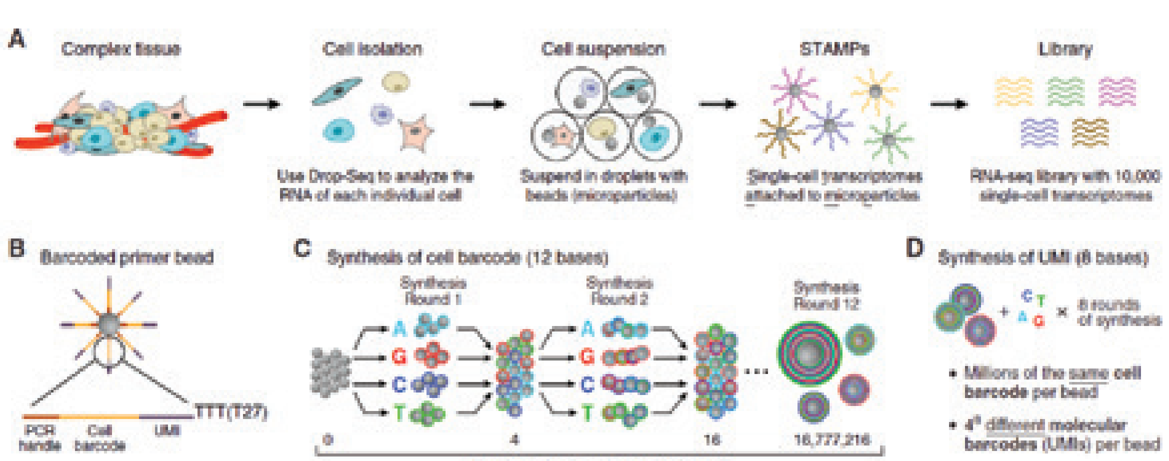
Drop-seq is a technology that allows biologists to analyze genome-wide gene expression in thousands of individual cells in a single experiment. This work is described in Macosko et al., Cell, 2015

Capturing single cells along with sets of uniquely barcoded primer beads together in tiny droplets enables large-scale, highly parallel single-cell transcriptomics. Applying this analysis to cells in mouse retinal tissue revealed transcriptionally distinct cell populations along with molecular markers of each type.

SUMMARY

Cells, the basic units of biological structure and function, vary broadly in type and state. Singlecell genomics can characterize cell identity and function, but limitations of ease and scale have prevented its broad application. Here we describe Drop-seq, a strategy for quickly profiling thousands of individual cells by separating them into nanoliter-sized aqueous droplets, associating a different barcode with each cell's RNAs, and sequencing them all together. Drop-seq analyzes mRNA transcripts from thousands of individual cells simultaneously while remembering transcripts' cell of origin. We analyzed transcriptomes from 44,808 mouse retinal cells and identified 39 transcriptionally distinct cell populations, creating a molecular atlas of gene expression for known retinal cell classes and novel candidate cell subtypes. Drop-seq will accelerate biological discovery by enabling routine transcriptional profiling at single cell resolution.

Individual cells are the building blocks of tissues, organs, and organisms. Each tissue contains cells of many types, and cells of each type can switch among biological states. In most biological systems, our knowledge of cellular diversity is incomplete; for example, the cell-type complexity of the brain is unknown and widely debated (Luo et al., 2008; Petilla Interneuron Nomenclature Group, et al., 2008). To understand how complex tissues work, it will be important to learn the functional capacities and responses of each cell type. A major determinant of each cell's function is its transcriptional program. Recent advances now enable mRNA-seq analysis of individual cells (Tang et al., 2009). However, methods of preparing cells for profiling have been applicable in practice to just hundreds (Hashimshony et al., 2012; Picelli et al., 2013) or (with automation) a few thousand cells (Jaitin et al., 2014), typically after first separating the cells by flow sorting (Shalek et al., 2013) or microfluidics (Shalek et al., 2014) and then amplifying each cell's transcriptome separately. Fast, scalable approaches are needed to characterize complex tissues with many cell types and states, under diverse conditions and perturbations. Here, we describe Drop-seq, a method to analyze mRNA expression in thousands of individual cells by encapsulating cells in tiny droplets for parallel analysis. Droplets—nanoliterscale aqueous compartments formed by precisely combining aqueous and oil flows in a microfluidic device (Thorsen et al., 2001; Umbanhowar et al., 2000)—have been used as tiny reaction chambers for PCR (Hindson et al., 2011; Vogelstein and Kinzler, 1999) and reverse transcription (Beer et al., 2008). We sought here to use droplets to compartmentalize cells into nanoliter-sized reaction chambers for analysis of all of their RNAs. A basic challenge of using droplets for transcriptomics is to retain a molecular memory of the identity of the cell from which each mRNA transcript was isolated. To accomplish this, we developed a molecular barcoding strategy to remember the cell-of-origin of each mRNA. We critically evaluated Drop-seq, then used it to profile cell states along the cell cycle.Wethen applied it to a complex neural tissue, mouse retina, and from 44,808 cell profiles identified 39 distinct populations, each corresponding to one or a group of closely related cell types. Our results demonstrate how large-scale single-cell analysis can help deepen our understanding of the biology of complex tissues and cell populations.

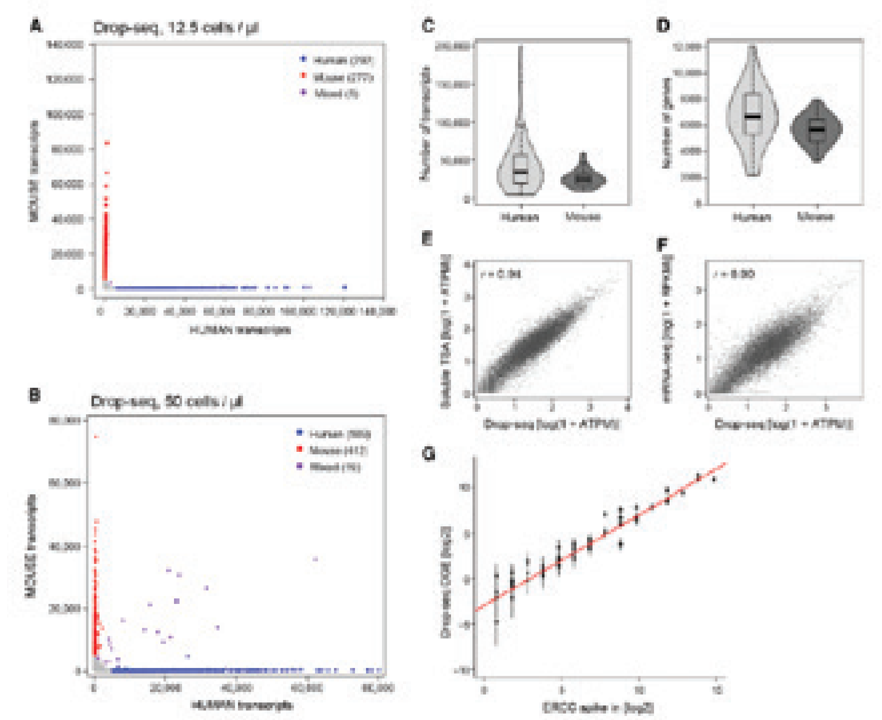


[A] Drop-Seq barcoding schematic. A complex tissue is dissociated into individual cells, which are then encapsulated in droplets together with icroparticles (gray circles) that deliver barcoded primers. Each cell is lysed within a droplet; its mRNAs bind to the primers on its companion microparticle. The mRNAs are reverse-transcribed into cDNAs, generating a set of beads called ''single-cell transcriptomes attached to microparticles'' (STAMPs). The barcoded STAMPs can then be amplified in pools for high-throughput mRNA-seq to analyze any desired number of individual cells.

[B] Sequence of primers on the microparticle. The primers on all beads contain a common sequence (''PCR handle'') to enable PCR amplification after STAMP formation. Each microparticle contains more than 108 individual primers that share the same ''cell barcode'' (C) but have different unique molecular identifiers (UMIs), enabling mRNA transcripts to be digitally counted (D). A 30-bp oligo dT sequence is present at the end of all primer sequences for capture of mRNAs.

[C] Split-and-pool synthesis of the cell barcode. To generate the cell barcode, the pool of microparticles is repeatedly split into four equally sized oligonucleotide synthesis reactions, to which one of the four DNA bases is added, and then pooled together after each cycle, in a total of 12 split-pool cycles. The barcode synthesized on any individual bead reflects that bead's unique path through the series of synthesis reactions. The result is a pool of microparticles, each possessing one of 412 (16,777,216) possible sequences on its entire complement of primers (see also Figure S1).

[D] Synthesis of a unique molecular identifier (UMI). Following the completion of the ''split-and-pool'' synthesis cycles, all microparticles are together subjected to eight rounds of degenerate synthesis with all four DNA bases available during each cycle, such that each individual primer receives one of 48 (65,536) possible sequences (UMIs).



CRITICAL EVALUATION OF DROP-SEQ USING SPECIES-MIXING EXPERIMENTS

[A and B] Drop-seq analysis of mixutres of mouse and human cells. Mixtures of human (HEK) and mouse (3T3) cells were analyzed by Drop-seq at the concentrations shown. The scatter plot shows the number of human and mouse transcripts associating to each STAMP. Blue dots indicate STAMPs that were designated from these data as human-specifiic (average of 99% human transcripts); red dots indicate STAMPs that were mouse-specific (average 99%). At the lower cell concentration, one STAMP barcode (of 570) associated with a mixture of human and mouse transcripts (A, purple). At the higher cell concentration, about 1.9% of STAMP barcodes associated with mouse-human mixtures (B). Data for other cell concentrations and a different single-cell analysis platform are in Figures S3B and S3C.

[C and D] Sensitivity analysis of Drop-seq at high read-depth. Violin plots show the distribution of the number of transcripts (C, scored by UMIs) and genes (D) detected per cell for 54 HEK (human) STAMPs (blue) and 28 3T3 (mouse) STAMPs (green) that were sequenced to a mean read depth of 737,240 high-quality aligned reads per cell.

[E and F] Correlation between gene expression measurements in Drop-seq and non-single-cell RNA-seq methods. Comparison of Drop-seq gene expression measurements (averaged across 550 STAMPs) to measurements from bulk RNA analyzed by: (E) an in-solution template switch amplification (TSA) procedure similar to Smart-seq2 (Picelli et al., 2013) (Supplemental Experimental Procedures); and (F) Illumina TruSeq mRNA-seq. All comparisons involve RNA derived from the same cell culture flask (3T3 cells). All expression counts were converted to average transcripts per million (ATPM) and plotted as log (1+ATPM).

[G] Quantitation of Drop-seq capture efficiency by ERCC spike-ins. Drop-seq was performed with ERCC control synthetic RNA at an estimated concentration of 100,000 ERCC RNA molecules per droplet. 84 beads were sequenced at a mean depth of 2.4 million reads, aligned to the ERCC reference sequences, and UMIs counted for each ERCC species, after applying a stringent down-correction for potential sequencing errors (Table S1 and Supplemental Experimental Procedures). For each ERCC RNA species above an average concentration of one molecule per droplet, the predicted number of molecules per droplet was plotted in log space (x-axis), versus the actual number of molecules detected per droplet by Drop-seq, also in log space (y-axis). Error bars indicate SD. The intercept of a regression line, constrained to have a slope of 1 and fitted to the seven highest points, was used to estimate a conversion factor (0.128). A second estimation, using the average number of detected transcripts divided by the number of ERCC molecules used (100,000), yielded a conversion factor of 0.125.

FIGURE 1

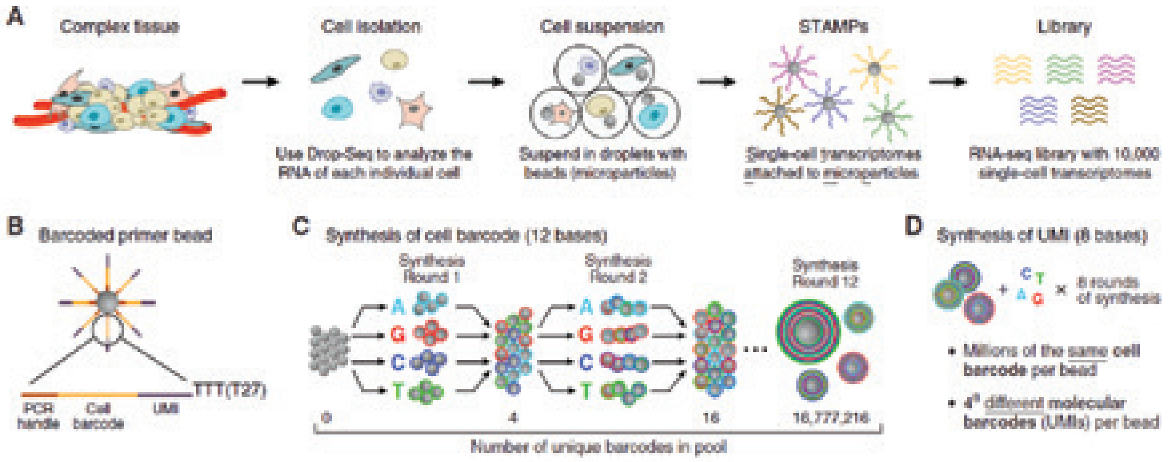
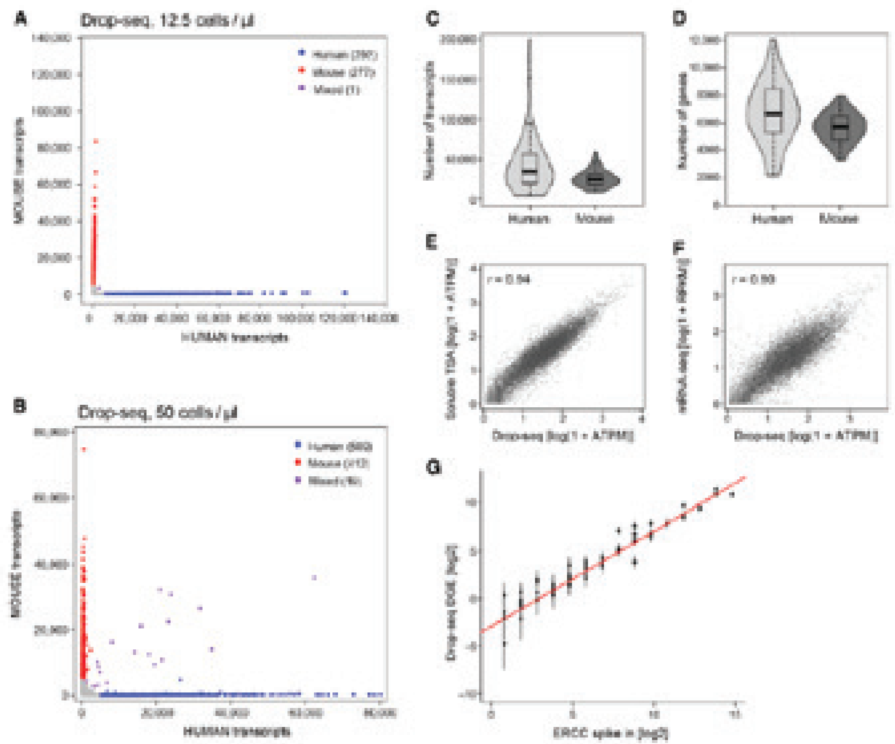
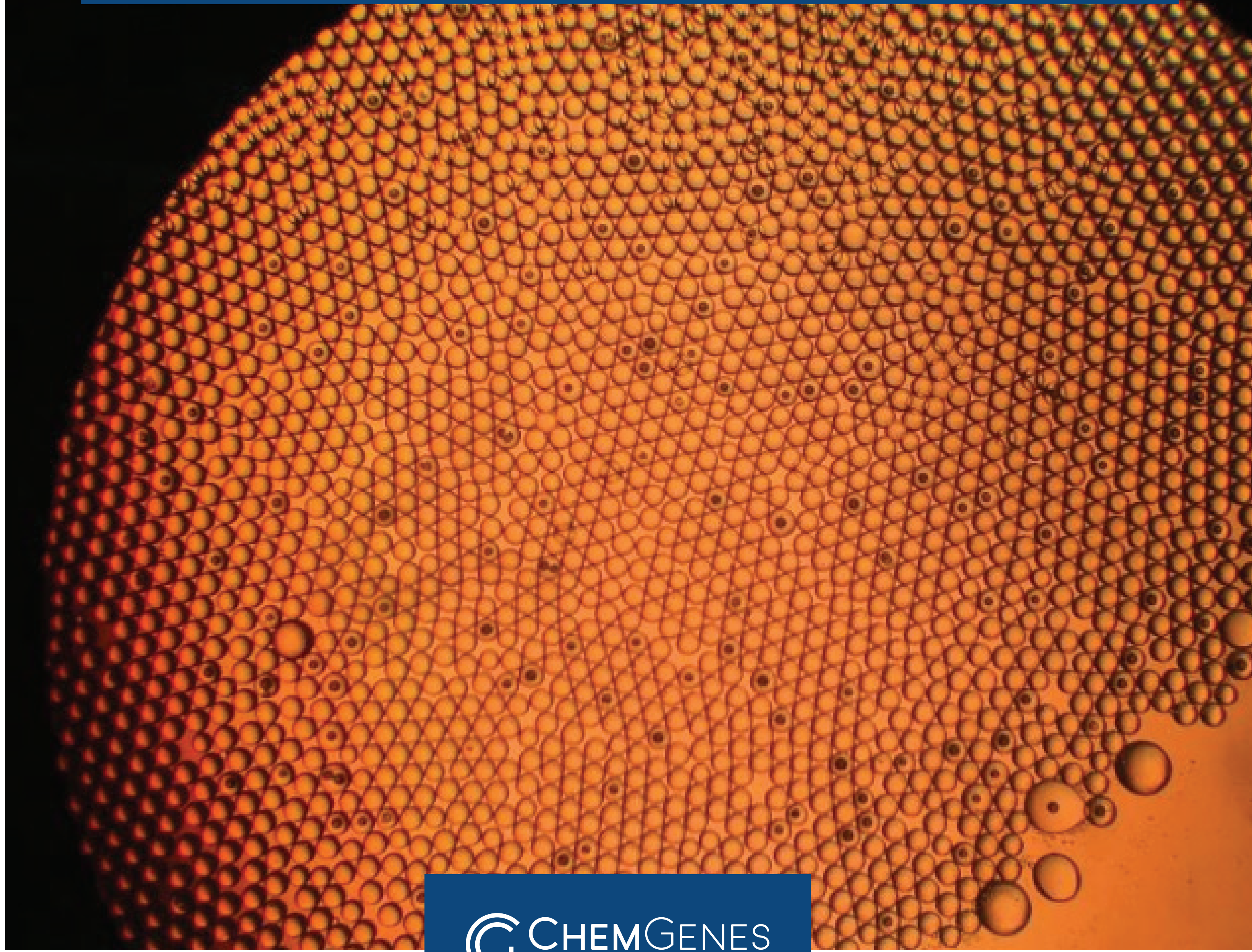


FIGURE 2



DROP–SEQ

HIGHLY PARALLEL GENOME-WIDE EXPRESSIONPROFILING OF INDIVIDUAL CELLS USING NANOLITER DROPLETS



CHEMGENES CORPORATION

FOR A COMPLETE PRODUCT LISTING, VISIT US AT
WWW.CHEMGENES.COM

CHEMGENES CORPORATION

Drop-seq is a technology that allows biologists to analyze genome-wide gene expression in thousands of individual cells in a single experiment. This work is described in Macosko et al., Cell, 2015

Capturing single cells along with sets of uniquely barcoded primer beads together in tiny droplets enables large-scale, highly parallel single-cell transcriptomics. Applying this analysis to cells in mouse retinal tissue revealed transcriptionally distinct cell populations along with molecular markers of each type.

SUMMARY

Cells, the basic units of biological structure and function, vary broadly in type and state. Singlecell genomics can characterize cell identity and function, but limitations of ease and scale have prevented its broad application. Here we describe Drop-seq, a strategy for quickly profiling thousands of individual cells by separating them into nanoliter-sized aqueous droplets, associating a different barcode with each cell's RNAs, and sequencing them all together. Drop-seq analyzes mRNA transcripts from thousands of individual cells simultaneously while remembering transcripts' cell of origin. We analyzed transcriptomes from 44,808 mouse retinal cells and identified 39 transcriptionally distinct cell populations, creating a molecular atlas of gene expression for known retinal cell classes and novel candidate cell subtypes. Drop-seq will accelerate biological discovery by enabling routine transcriptional profiling at single cell resolution.

Individual cells are the building blocks of tissues, organs, and organisms. Each tissue contains cells of many types, and cells of each type can switch among biological states. In most biological systems, our knowledge of cellular diversity is incomplete; for example, the cell-type complexity of the brain is unknown and widely debated (Luo et al., 2008; Petilla Interneuron Nomenclature Group, et al., 2008). To understand how complex tissues work, it will be important to learn the functional capacities and responses of each cell type. A major determinant of each cell's function is its transcriptional program. Recent advances now enable mRNA-seq analysis of individual cells (Tang et al., 2009). However, methods of preparing cells for profiling have been applicable in practice to just hundreds (Hashimshony et al., 2012; Picelli et al., 2013) or (with automation) a few thousand cells (Jaitin et al., 2014), typically after first separating the cells by flow sorting (Shalek et al., 2013) or microfluidics (Shalek et al., 2014) and then amplifying each cell's transcriptome separately. Fast, scalable approaches are needed to characterize complex tissues with many cell types and states, under diverse conditions and perturbations. Here, we describe Drop-seq, a method to analyze mRNA expression in thousands of individual cells by encapsulating cells in tiny droplets for parallel analysis. Droplets—nanoliterscale aqueous compartments formed by precisely combining aqueous and oil flows in a microfluidic device (Thorsen et al., 2001; Umbanhowar et al., 2000)—have been used as tiny reaction chambers for PCR (Hindson et al., 2011; Vogelstein and Kinzler, 1999) and reverse transcription (Beer et al., 2008). We sought here to use droplets to compartmentalize cells into nanoliter-sized reaction chambers for analysis of all of their RNAs. A basic challenge of using droplets for transcriptomics is to retain a molecular memory of the identity of the cell from which each mRNA transcript was isolated. To accomplish this, we developed a molecular barcoding strategy to remember the cell-of-origin of each mRNA. We critically evaluated Drop-seq, then used it to profile cell states along the cell cycle.Wethen applied it to a complex neural tissue, mouse retina, and from 44,808 cell profiles identified 39 distinct populations, each corresponding to one or a group of closely related cell types. Our results demonstrate how large-scale single-cell analysis can help deepen our understanding of the biology of complex tissues and cell populations.

REF: FIGURE 1 (BACK)

[A] Drop-Seq barcoding schematic. A complex tissue is dissociated into individual cells, which are then encapsulated in droplets together with icroparticles (gray circles) that deliver barcoded primers. Each cell is lysed within a droplet; its mRNAs bind to the primers on its companion microparticle. The mRNAs are reverse-transcribed into cDNAs, generating a set of beads called ''single-cell transcriptomes attached to microparticles'' (STAMPs). The barcoded STAMPs can then be amplified in pools for high-throughput mRNA-seq to analyze any desired number of individual cells.

[B] Sequence of primers on the microparticle. The primers on all beads contain a common sequence (''PCR handle'') to enable PCR amplification after STAMP formation. Each microparticle contains more than 108 individual primers that share the same ''cell barcode'' (C) but have different unique molecular identifiers (UMIs), enabling mRNA transcripts to be digitally counted (D). A 30-bp oligo dT sequence is present at the end of all primer sequences for capture of mRNAs.

[C] Split-and-pool synthesis of the cell barcode. To generate the cell barcode, the pool of microparticles is repeatedly split into four equally sized oligonucleotide synthesis reactions, to which one of the four DNA bases is added, and then pooled together after each cycle, in a total of 12 split-pool cycles. The barcode synthesized on any individual bead reflects that bead's unique path through the series of synthesis reactions. The result is a pool of microparticles, each possessing one of 412 (16,777,216) possible sequences on its entire complement of primers (see also Figure S1).

[D] Synthesis of a unique molecular identifier (UMI). Following the completion of the ''split-and-pool'' synthesis cycles, all microparticles are together subjected to eight rounds of degenerate synthesis with all four DNA bases available during each cycle, such that each individual primer receives one of 48 (65,536) possible sequences (UMIs).

CRITICAL EVALUATION OF DROP-SEQ USING SPECIES-MIXING EXPERIMENT

REF: FIGURE 2 (BACK)

[A and B] Drop-seq analysis of mixutres of mouse and human cells. Mixtures of human (HEK) and mouse (3T3) cells were analyzed by Drop-seq at the concentrations shown. The scatter plot shows the number of human and mouse transcripts associating to each STAMP. Blue dots indicate STAMPs that were designated from these data as human-specifiic (average of 99% human transcripts); red dots indicate STAMPs that were mouse-specific (average 99%). At the lower cell concentration, one STAMP barcode (of 570) associated with a mixture of human and mouse transcripts (A, purple). At the higher cell concentration, about 1.9% of STAMP barcodes associated with mouse-human mixtures (B). Data for other cell concentrations and a different single-cell analysis platform are in Figures S3B and S3C.

[C and D] Sensitivity analysis of Drop-seq at high read-depth. Violin plots show the distribution of the number of transcripts (C, scored by UMIs) and genes (D) detected per cell for 54 HEK (human) STAMPs (blue) and 28 3T3 (mouse) STAMPs (green) that were sequenced to a mean read depth of 737,240 high-quality aligned reads per cell.

[E and F] Correlation between gene expression measurements in Drop-seq and non-single-cell RNA-seq methods. Comparison of Drop-seq gene expression measurements (averaged across 550 STAMPs) to measurements from bulk RNA analyzed by: (E) an in-solution template switch amplification (TSA) procedure similar to Smart-seq2 (Picelli et al., 2013) (Supplemental Experimental Procedures); and (F) Illumina TruSeq mRNA-seq. All comparisons involve RNA derived from the same cell culture flask (3T3 cells). All expression counts were converted to average transcripts per million (ATPM) and plotted as log (1+ATPM).

[G] Quantitation of Drop-seq capture efficiency by ERCC spike-ins. Drop-seq was performed with ERCC control synthetic RNA at an estimated concentration of 100,000 ERCC RNA molecules per droplet. 84 beads were sequenced at a mean depth of 2.4 million reads, aligned to the ERCC reference sequences, and UMIs counted for each ERCC species, after applying a stringent down-correction for potential sequencing errors (Table S1 and Supplemental Experimental Procedures). For each ERCC RNA species above an average concentration of one molecule per droplet, the predicted number of molecules per droplet was plotted in log space (x-axis), versus the actual number of molecules detected per droplet by Drop-seq, also in log space (y-axis). Error bars indicate SD. The intercept of a regression line, constrained to have a slope of 1 and fitted to the seven highest points, was used to estimate a conversion factor (0.128). A second estimation, using the average number of detected transcripts divided by the number of ERCC molecules used (100,000), yielded a conversion factor of 0.125.