Single-Cell Whole Genome Sequencing on the C1 System: a Performance Evaluation

Introduction

Cancer is a dynamic evolutionary process of which intratumor genetic and phenotypic diversity is an inherent feature. The genetic diversity among cells of a tumor provides a precursor for clonal selection and a source of therapeutic escape. It is therefore important to capture the mutational load of individual tumors or biopsies. Unfortunately, traditional total DNA or bulk analysis masks the fundamental features of subclonal variegation of genetic lesions and of clonal phylogeny. Single-cell genetic profiling via whole genome sequencing (WGS) provides a clear resolution to this problem but requires a sensitive and accurate method that can detect low-abundance variants and distinguish them from sequencing errors.

WGS at the single-cell level is challenging, since the amount of genomic DNA present in a single cell is very limited and, thus, requires amplification. Fluidigm developed a method for whole genome amplification (WGA) of single cells using the C1 System to automate capture, lysis, and DNA amplification for up to 96 individual cells in a single workflow. Amplified products harvested from the system are ready for downstream library preparation for targeted resequencing (TR), whole exome sequencing (WES), or WGS (Figure 1A). TR, WES, and WGS methods add value in terms of resulting sequencing data.

TR and WES provide sequencing information regarding specific regions of interests and protein coding genomic regions, respectively. However, WGS is capable of providing comprehensive unbiased genotype data throughout the genome, including both coding and noncoding gene regulatory regions. Ultimately, single-cell-resolution WGS will enable a more complete genomic view of how somatic mosaicism contributes to normal development and disease processes. In this technical note, we describe the performance of single-cell WGS, which shows exceptional genome coverage and uniformity across a broad GC range.

Experimental Design

We performed single-cell WGS and WES on primary ductal carcinoma cells (CRL-2338™/HCC1954) and normal B-lymphoblasts (CRL-2339™/HCC1954BL) derived from the same individual. The cells were subcultured and cryopreserved according to the manufacturer's recommendation (ATCC). One vial of each culture was thawed and loaded on the C1 IFC according to Fluidigm DNA sequencing procedures. Refer to the protocol, Using the C1 System to Generate Single-Cell Libraries for DNA Sequencing (PN 100-7135). The Fluidigm integrated single-cell DNA sequencing

workflow allows for simultaneous single-cell isolation, live/dead cell staining, lysis, neutralization, and WGA from up to 96 cells using the C1 System (Figure 1A). Typical WGA yields are 150–200 ng of amplified genomic DNA, enabling multiple downstream applications, including TR,WES, and/or WGS to identify and validate genetic variants in single cells. Hands-on time from single-cell to WGS-ready libraries is only 1 hour and 15 minutes, with a total time to completion of less than 12 hours (Figure 1B).

Figure 1

A

B

Process	Total Time to Complete	Hands-on Time
C1 Whole Genome Amplification	9.5 hr	1 hr
WGS Library Prep	1.5 hr	15 min
Exome Enrichment	1.5 days	4 hr

C

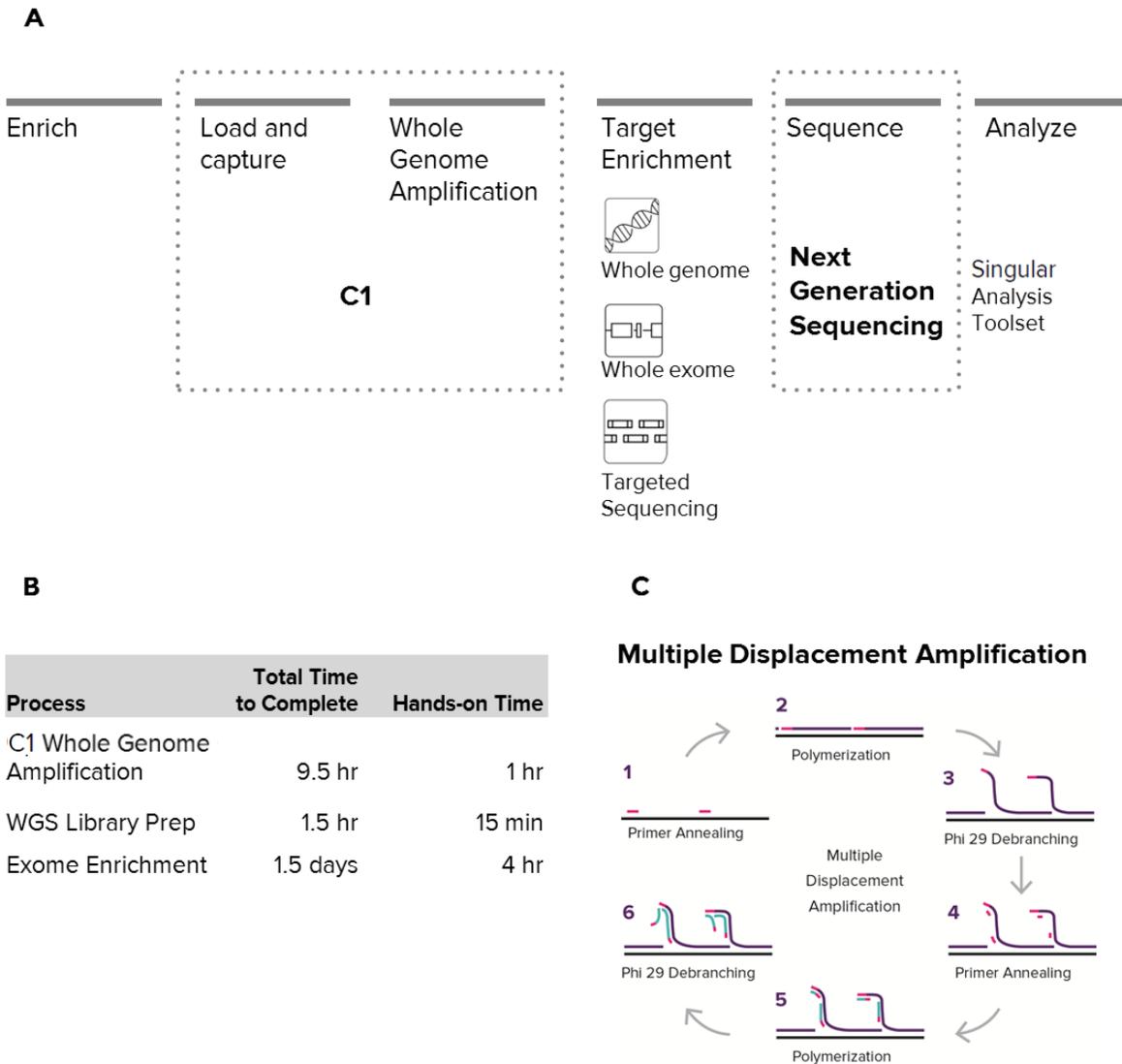**Multiple Displacement Amplification**

Figure 1: From cells to sequence-ready libraries in less than 24 hours. The C1 takes you from cells to sequence with less than 6 hours of hands-on time. A. 96 single-cell whole genome harvests are exported from the C1 for library preparation. Whole genome libraries are prepared using Illumina Nextera. Whole exome sequencing libraries can be further prepared using Illumina® Nextera™ Rapid Capture Exome. After sequencing, alignment with Bowtie2 and SNP calls are made with GATK; single-cell variant analysis is supported with Singular Analysis Toolset. B. Whole genome sequencing libraries can be ready to sequencing in less than 12 hours with only 1 hour and 15 minutes hands-on time. Whole exome sequencing libraries require an additional 1.5 days and 4 hours hands-on time. C. Fluidigm whole genome amplification is performed by multiple displacement amplification (MDA) using high-fidelity Φ29 DNA polymerase within the C1 IFC.
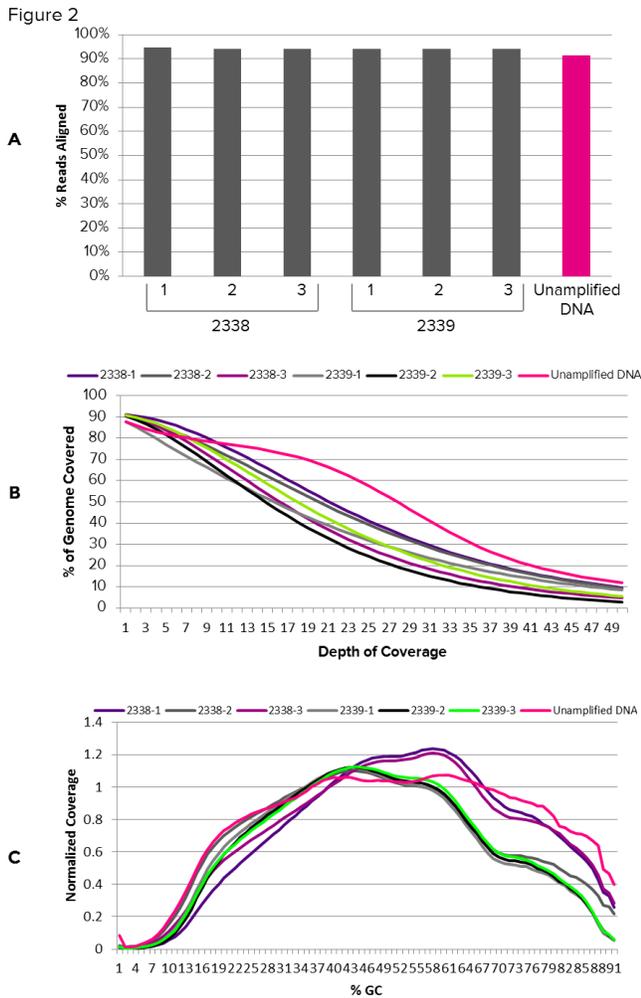
Figure 2

Figure 2: Fluidigm whole genome amplification is optimized to reduce amplification bias. A. Across the six single-cell WGS experiments (CRL-2338/HCC1954 cells 1–3 and CRL-2339/ HCC1954-BL cells 1–3), mapping rates were 94.16% ± 0.24% (mean ±95% CI). Unamplified genomic DNA (red) shows a mapping rate of 91.51%. B. On average, more than 90% of the genome was covered in single cells. Plotted is data from three single cells of each cell line; CRL- 2338/HCC1954 cells 1–3 and CRL-2339/HCC1954-BL cells 1–3. Unamplified genomic DNA (black) demonstrated a very similar coverage. C. Fluidigm WGA method improves coverage in GC-rich regions. Plotted is data from three single cells of each cell line; CRL-2338/ HCC1954 cells 1–3 and CRL-2339/HCC1954-BL cells 1–3 and unamplified genomic DNA (black). In single cells, GC-rich regions (more than 70% GC-rich) display on average 70% coverage of the unamplified genomic DNA.

Library preparation is performed using a Tn5 transposase-based protocol to simultaneously fragment DNA and ligate adapters, allowing indexing of up to 96 samples at a time (Nextera™ Rapid Capture Exome Kit, Illumina). Using the same library, exome enrichment can also be performed adding only four hours of hands-on time and allowing sequence-ready whole exome libraries to be prepared in approximately 2.5 days (Figure 1B). Whole genome amplification is performed by multiple displacement amplification (MDA) using high-fidelity Φ29 DNA polymerase (illustra™ GenomiPhi™ V2 DNA Amplification Kit, GE Healthcare) within the C1 IFC (Figure 1C). Cell viability was verified at capture using the LIVE/DEAD® Cell Viability Assay Kit (Life Technologies) (data not shown). After harvest, amplified DNA is used to generate a WGS library that is also compatible with downstream exome enrichment (Nextera Rapid Capture Exome Kit, Illumina). For the library preparation, we used 50 ng of amplified DNA from 50 individual cells of each line (CRL-2338/HCC1954 and CRL-2339/1954BL). Three WGS libraries of each cell line were directly sequenced to an average depth of 23X genome-wide. All 50 cells from each cell line were further processed through exome enrichment (Nextera Rapid Capture Exome, Illumina) and sequenced to an average depth of 27X genome-wide. Reads obtained were aligned to hg19 using Bowtie2 (Langmead 2012) followed by local realignment around INDELs, base-quality recalibration, and multisample variant calling using Genome Analysis Toolkit (McKenna et al. 2010). Downstream processing of raw variant calls and analysis of single-cell mutations were performed using the Fluidigm Singular Analysis Toolset 3.5. Refer to the

Figure 3

Singular Analysis Toolset User Guide (PN 100-5066).

Results

Since the amount of DNA in a single cell is limiting to detection of genetic variants, amplification of the genomic DNA is required. Thus, to assess the performance of a single-cell WGS method, we directly compared single-cell amplified genomic DNA to bulk genomic DNA without amplification. For this, we compared WGS reads from six single cells using the Fluidigm single-cell DNA sequencing protocol on the C1 to unamplified bulk WGS and evaluated the mapping rate, coverage uniformity, GC coverage, WGA-induced base error rate, single-nucleotide variant (SNV) accuracy, and allelic dropout rates of the method. The total amount of sequence generated per cell ranged from 67 to 96 gigabases (Gb). In addition, 133.6 Gb of sequence from an unamplified bulk genomic DNA was generated (Figure 2).

When comparing the mapping rate between amplified single-cell genomic DNA and unamplified bulk genomic DNA, read alignment rates were very similar at more than 94% (Figure 2A). More than 90% of the genome was covered on average, very similar to the coverage observed with unamplified genomic DNA (Figure 2B). GC-rich regions of the human genome are difficult to amplify. Thus, to further assess the performance of single-cell WGA, we compared amplified single-cell genomic DNA and unamplified bulk genomic DNA across regions of the genome with varied % GC content. We found that the distribution of coverage by amplified single-cell genomic DNA was nearly identical to that observed in unamplified bulk genomic DNA (Figure 2C), suggesting that very
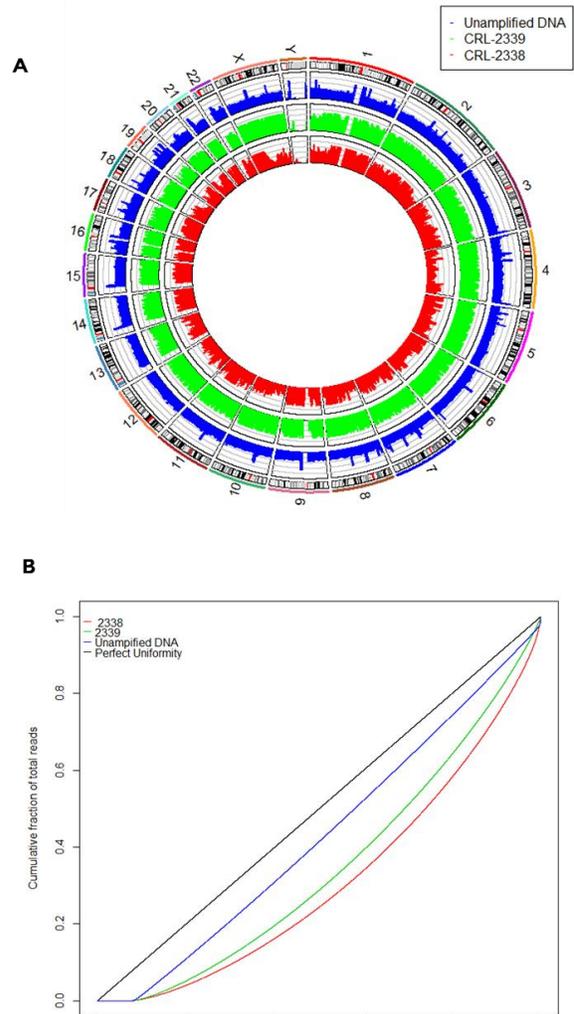

Figure 3: Fluidigm whole genome amplification is optimized to maximize coverage and uniformity. A. Unamplified (blue) and single-cell genomic DNA achieve similar coverage across the genome: CRL-2338/HCC1954 (red), CRL-2339/HCC1954BL − (green). Coverage is calculated as reads per 100 kb, normalized to the total number of aligned reads in each sample. B. A Lorenz curve of single-cell amplified DNA: CRL-2338/HCC1954 (red), CRL- 2339/HCC1954-BL (green) and unamplified bulk genomic DNA (blue). The Lorenz curve gives the cumulative fraction of reads as a function of the cumulative fraction of the genome. Perfect uniformity is represented by the black diagonal line, with nonuniformity indicated by deviation from the diagonal. Single-cell amplified genomic DNA displays a remarkable uniformity compared to the unamplified genome.

little GC bias was induced by Fluidigm WGA. Thus, Fluidigm single-cell DNA sequencing

method yields an exceptional performance in amplifying GC-rich regions.

We assessed base coverage uniformity in single-cell WGA and found that amplified genomic DNA is remarkably uniform across all chromosomes (Figure 3A) and faithfully represents the human genome at single-cell level with a high reproducibility among single cells (Figure 3B).

The Φ29 DNA polymerase enzyme used in the MDA reaction of Fluidigm single-cell DNA sequencing method is reputed for its high fidelity at low temperature and for its ability to easily displace DNA during DNA replication (Figure 1C). To quantify the rate of errors induced by Φ29 during single-cell WGA, we compared error rates in amplified single-cell genomic DNA to the observed error rates in unamplified bulk genomic DNA. Amplified single-cell genomic DNA demonstrated a base error rate of ≤0.1% in all cells analyzed, confirming the remarkable high fidelity of Φ29 DNA polymerase in amplifying DNA from individual cells (Figure 4A).

To assess the accuracy in identifying SNVs in single cells, we established the concordance of SNV calling in single cells compared to 11,685 high-confidence SNVs identified in bulk genomic DNA. In six single-cells, we found that the SNV concordance was 92.96% ±2.73% (mean ±95% CI) (Figure 4B). Accurate identification of heterozygous SNVs in single cells can be influenced by preferential amplification of one allele. To quantify such biases in our system, we measured the allelic dropout rate in each single cell at 6,404 high-confidence heterozygous SNVs identified in bulk DNA. From six single cells, we found that the dropout rate was on average 12.56% ±4.92%
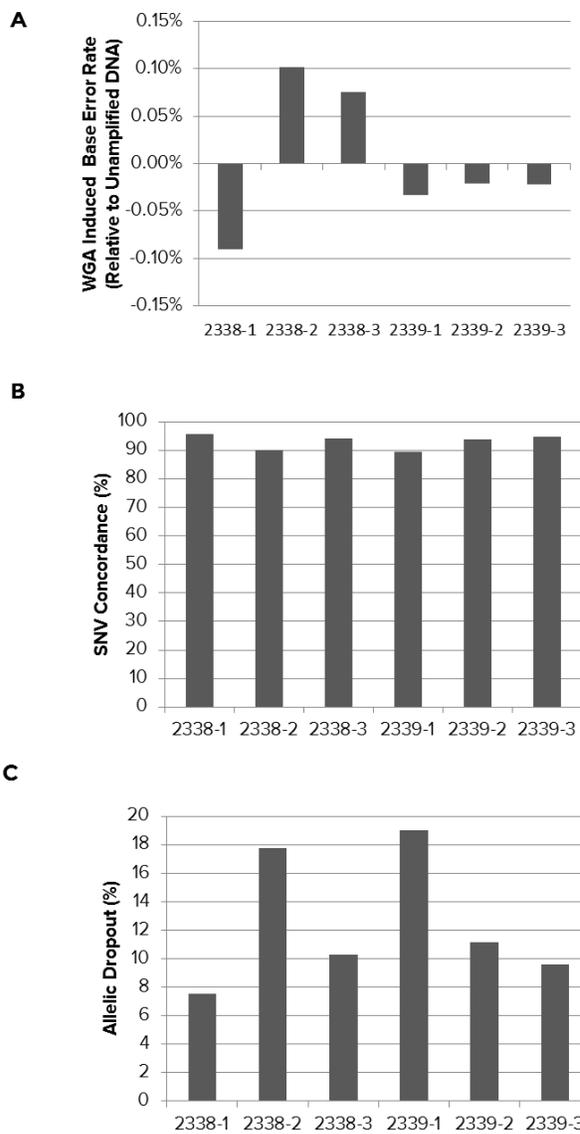

Figure 4: WGA-induced base error rate, SNV concordance, and allelic dropout rates in single cells. A. The error rate in each of the six single-cell WGS experiments (CRL-2338/HCC1954 cells 1–3 and CRL-2339/HCC1954-BL cells 1–3) was ≤0.1% relative to unamplified bulk genomic DNA. B. C1 single-cell DNA sequencing SNV detection yields 92.96% ±2.73% (mean ±95% CI) concordance across the six single-cell WGS experiments (CRL-2338/HCC1954 cells 1–3 and CRL-2339/HCC1954-BL cells 1–3). C. C1 single-cell DNA sequencing allelic dropout yields 12.56% ±4.92% (mean ±95% CI) allelic dropout rates across the six single-cell WGS experiments (CRL-2338/HCC1954 cells 1–3 and CRL-2339/HCC1954-BL cells 1–3).

(mean ±95% CI) (Figure 4C). Taken together, these results suggest that Fluidigm single-cell WGS provides high-quality and

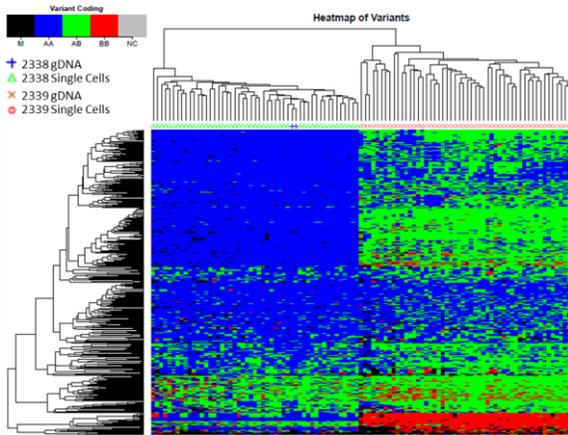accurate genotype information at the single-cell level.

Figure 5: Variants identified by single-cell WGS resolve tumor and normal cells derived from the same individual. Heat map representation of 474 SNVs identified within exon coding regions by WGS that were examined across a population of 50 normal B-lymphoblasts (CRL-2339/HCC1954-BL) and 50 tumor cells (CRL-2338/HCC1954). The identified variants clearly distinguish populations of tumor and normal cells derived from the same individual.

Finally, we determined whether the variants identified in single cells were informative for distinguishing tumor and normal cells derived from the same individual. To do so, we used the single-cell WGS data to identify 474 variants within protein coding regions and subsequently examined the distribution of these variants across an additional 100 single-cell WES libraries generated from control and tumor cell lines (50 each) (Figure 5). Overall, SNV concordance between WGS and WES was 94.29% ±0.5% (mean ±95% CI), confirming that both methods produced high-quality genotype profiles from single cells. With Fluidigm Singular Analysis Toolset, we could clearly discriminate the two populations with very high confidence using hierarchical clustering (Figure 5), demonstrating the utility of single-cell WGS

toward accurately identifying tumor and normal cells from the same individual.

Conclusion

The Fluidigm single-cell DNA sequencing protocol on the C1 System generates WGA DNA suitable for WGS, WES, and TR from up to 96 individual cells in a single workflow. In this technical note, we have shown that the C1 single-cell WGA offers unbiased amplification and allows genomes to be represented faithfully with high reproducibility and uniformity across single cells. Finally, Fluidigm single-cell DNA sequencing provides high-quality and accurate genotype information. Using Fluidigm Singular Analysis Toolset, the identified SNV profiles can easily and clearly distinguish tumor from normal cells derived from the same individual. Thus, C1 single-cell DNA sequencing provides the performance necessary to routinely analyze cancer and normal cells from the same individual and can very well become the method of choice to study the genetic diversity among diseased and normal single-cell populations.

References

Langmead B. and Salzberg S., "Fast gapped-read alignment with Bowtie 2." Nature Methods, 9 (2012): 357–9.

McKenna A., Hanna M., Banks E. et al. "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Research 20 (2010): 1,297–303.

Corporate headquarters
7000 Shoreline Court, Suite 100
South San Francisco, CA 94080 USA
Toll-free: 1+ 866 359 4354 | Fax: 650 871 7152
fluidigm.com

Sales
North America | +1 650 266 6170 | info-us@fluidigm.com
Europe/EMEA | +33 1 60 92 42 40 | info-europe@fluidigm.com
Japan | +81 (0)3 3662 2150 | info-japan@fluidigm.com
China | +86 (0)21 3255 8368 | info-china@fluidigm.com Asia | +1 650 266
6000 | info-asia@fluidigm.com
Latin America | +1 650 266 6000 | info-latinamerica@fluidigm.com

For Research Use Only. Not for use in diagnostic procedures.