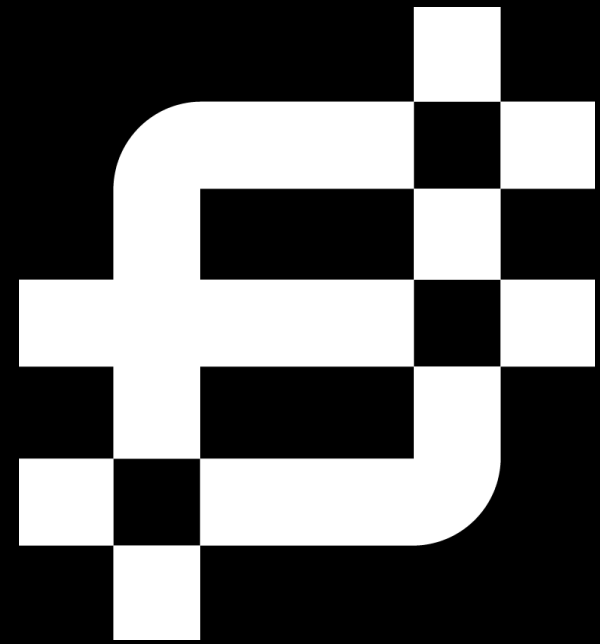# Single-Cell Analysis of Breast Cancer Single Nucleotide Variants

L. Suzanne Weaver, Peilin Chen, Jing Wang, David Wang, Keith Szulwach, Larry Xi, David Ruff, Ramesh Ramakrishnan

Fluidigm Corporation, South San Francisco, CA USA 94080

Introduction

Research linking genetic variants to disease has been challenging in the past, because traditional approaches are commonly limited to detecting high frequency variants within a bulk sample. This approach assumes that the sample is comprised of identical cells with the same variants. However, in the case of cancer genomics, somatic mutations accumulate over time and a single tumor can give rise to multiple sub-clones. Employing a multi-cell based DNA sequencing strategy to heterogeneous samples can mask low abundant variants and make it difficult to distinguish them from sequencing errors. Sequencing the genome of individual cells, however, circumvents the problem and provides deeper insight into the molecular heterogeneity within the cell population. For example, single-cell sequencing reveals whether multiple variants coexist in a subset of cells or are randomly distributed throughout the population. Since the amount of DNA present in every cell is very limited, a prerequisite to single-cell sequencing is DNA amplification. To map subpopulations of cells, large numbers of individual cells are needed, and the requisite amplification and library preparation are costly, laborious, and subject to handling errors using current techniques. We have developed a method for whole genome amplification (WGA) of single cells using an integrated microfluidic system to automate the capture, lysis, and DNA amplification from up to 96 individual cells in a single workflow. Amplified products harvested from the system are ready for downstream library preparation and targeted, whole exome, or whole genome sequencing. Whole genome sequencing results of a few cells indicate improvement in genome coverage, uniformity, and GC bias over state-of-the-art WGA methods. We also demonstrate targeted sequencing performance of single cells compared to bulk genomic DNA, and show single nucleotide variant (SNV) distributions in normal/disease-paired breast cancer cells for a panel of oncogenes.

Overview of the C_1™ Single-Cell Auto Prep System for Whole Genome Amplification (DNA Seq)

The C_1™ Single-Cell Auto Prep System is an integrated microfluidic system that provides a unique and simplified workflow for single-cell isolation, wash, live/dead cell staining, cell lysis, neutralization, and whole-genome amplification from up to 96 cells per run (Figure 1). Overall hands-on time from single cell to sequence ready libraries is five hours, with an overall runtime of <20 hours.

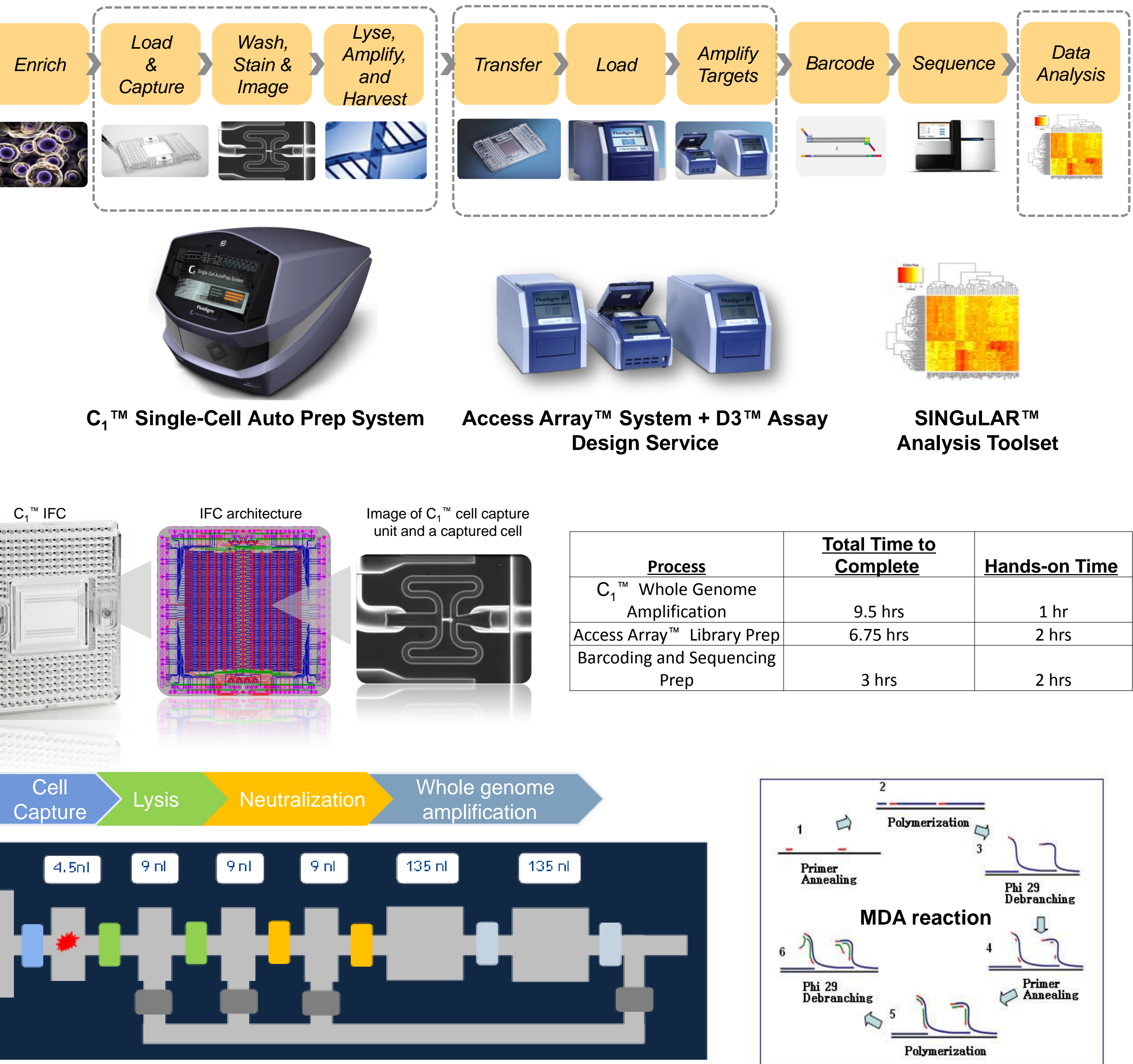| Process	Total Time to Complete	Hands-on Time
C_1™ Whole Genome Amplification	9.5 hrs	1 hr
Access Array™ Library Prep	6.75 hrs	2 hrs
Barcoding and Sequencing Prep	3 hrs	2 hrs

Figure 1
The Fluidigm integrated DNA Seq workflow allows for the simultaneous capture, lysis, neutralization, and whole genome amplification of DNA from up to 96 cells using the C_1™ Single-Cell Auto Prep System. Amplified DNA may then be used for multiple downstream applications, including targeted resequencing, whole exome sequencing, or whole genome sequencing to identify and interrogate genetic variants in single cells. Whole genome amplification is performed by Multiple Displacement Amplification (MDA) using high fidelity Phi29 DNA polymerase (GE illustra Genomiphi V2) within the C_1™ IFC. Coupling whole genome amplified DNA with the 48.48 Access Array™ System and D3™ Assay Design enables a streamlined targeted resequencing protocol. Downstream analyses of single-cell variants can be performed using the SINGuLAR™ Analysis Toolset.

Results

Performance of Whole-Genome Amplification (DNA Seq) using the C_1™ Single-Cell Auto Prep System
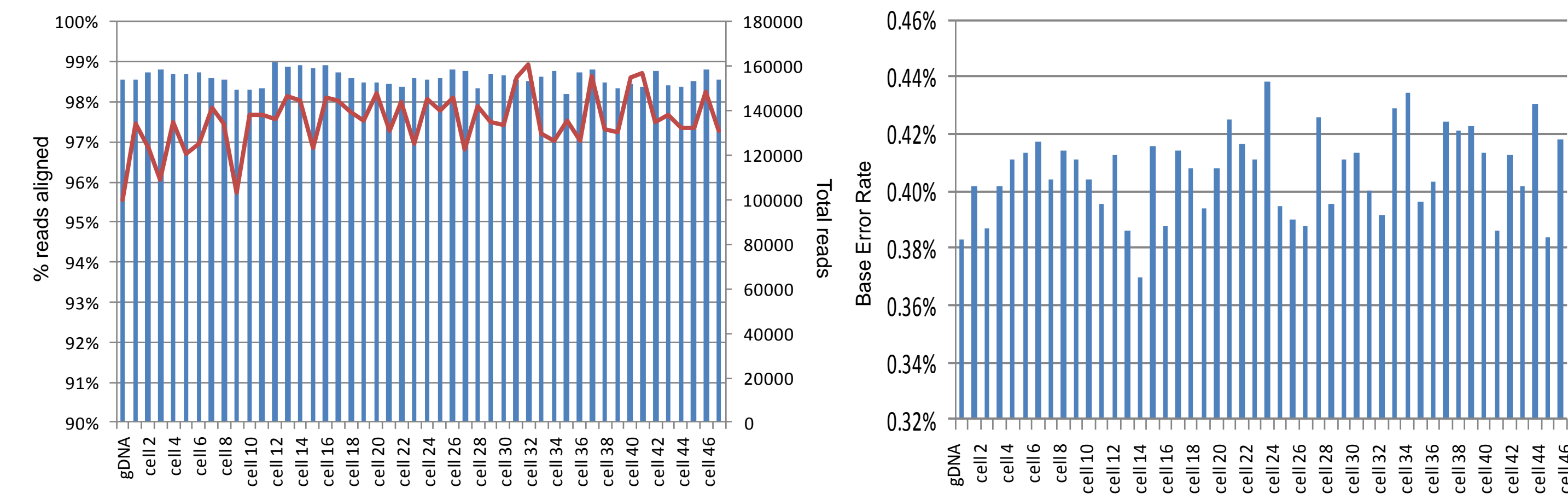
Figure 2
Single-cell whole genome amplified DNA using the DNA Seq protocol on the C_1™ Single-Cell Auto Prep System routinely yields >95% read alignment rates with <0.5% base error rates, comparable to that generated from unamplified bulk genomic DNA.
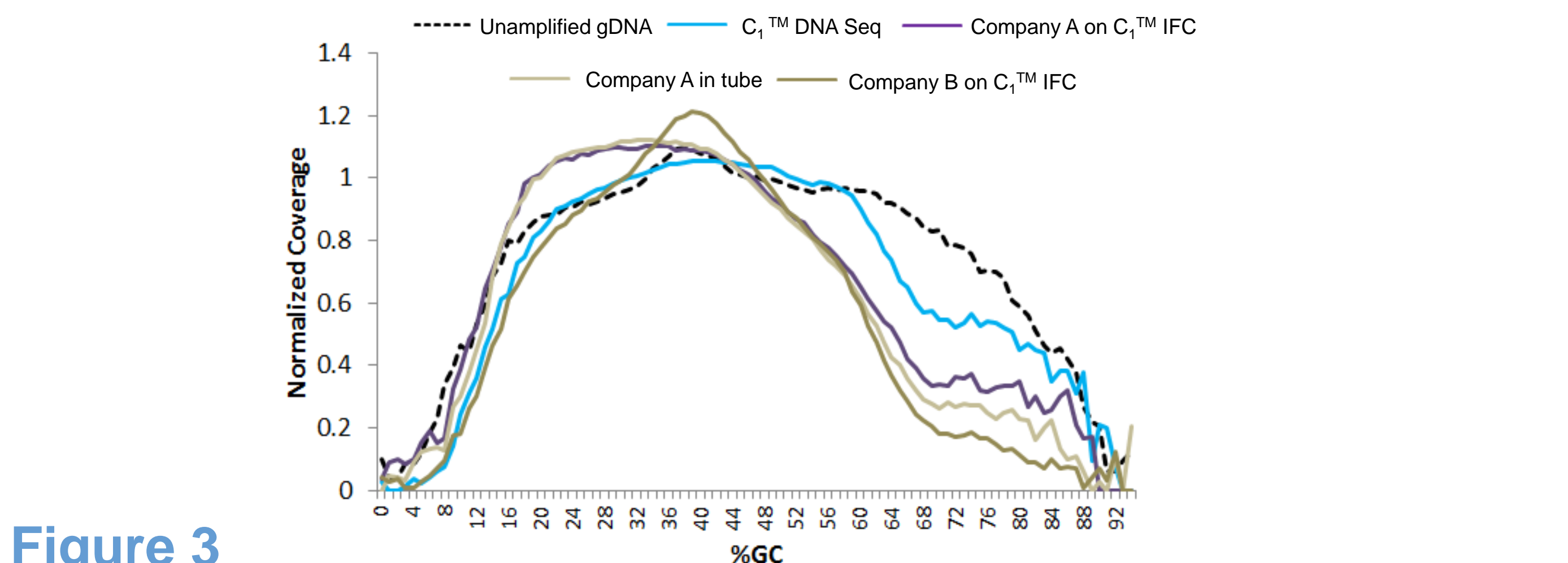
Figure 3
Single-cell whole genome amplification using the C_1™ Single-Cell Auto Prep System offers improvements in the representation of difficult to amplify GC rich regions of the human genome in comparison to current state-of-the art whole genome amplification methods.
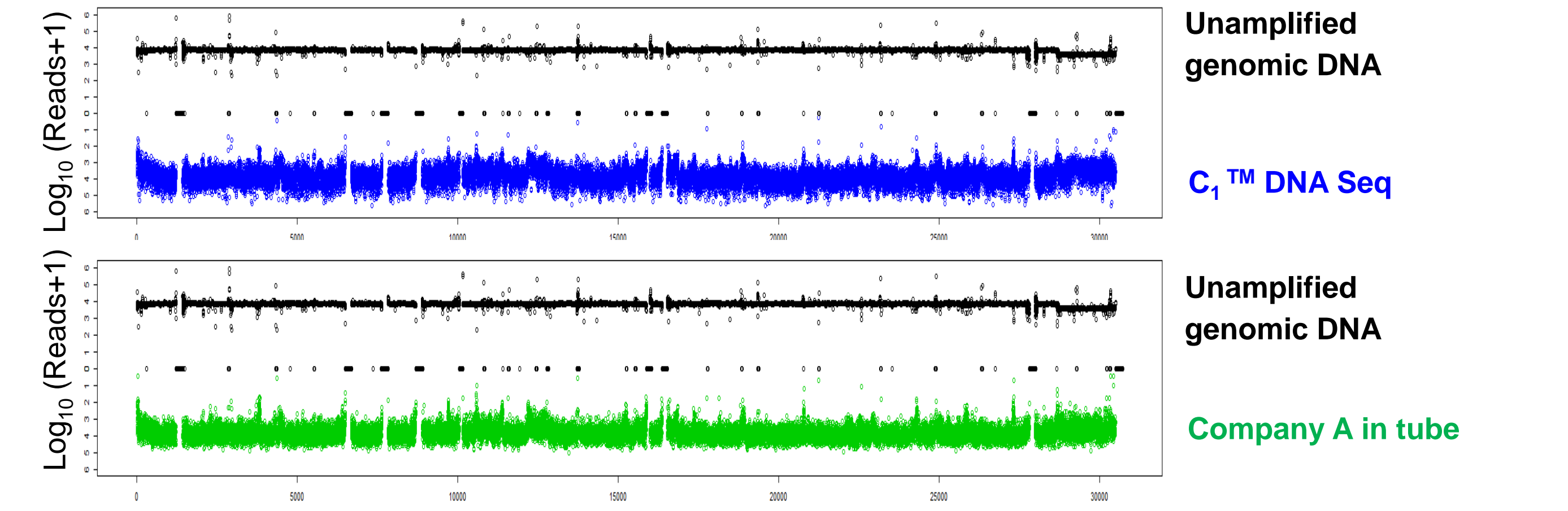
Figure 4
Uniformity of sequencing reads derived from C_1™ DNA Seq and in tube WGA compared to unamplified gDNA. Normalized read densities were determined in 100kb bins and plotted across human chr1 (hg19).
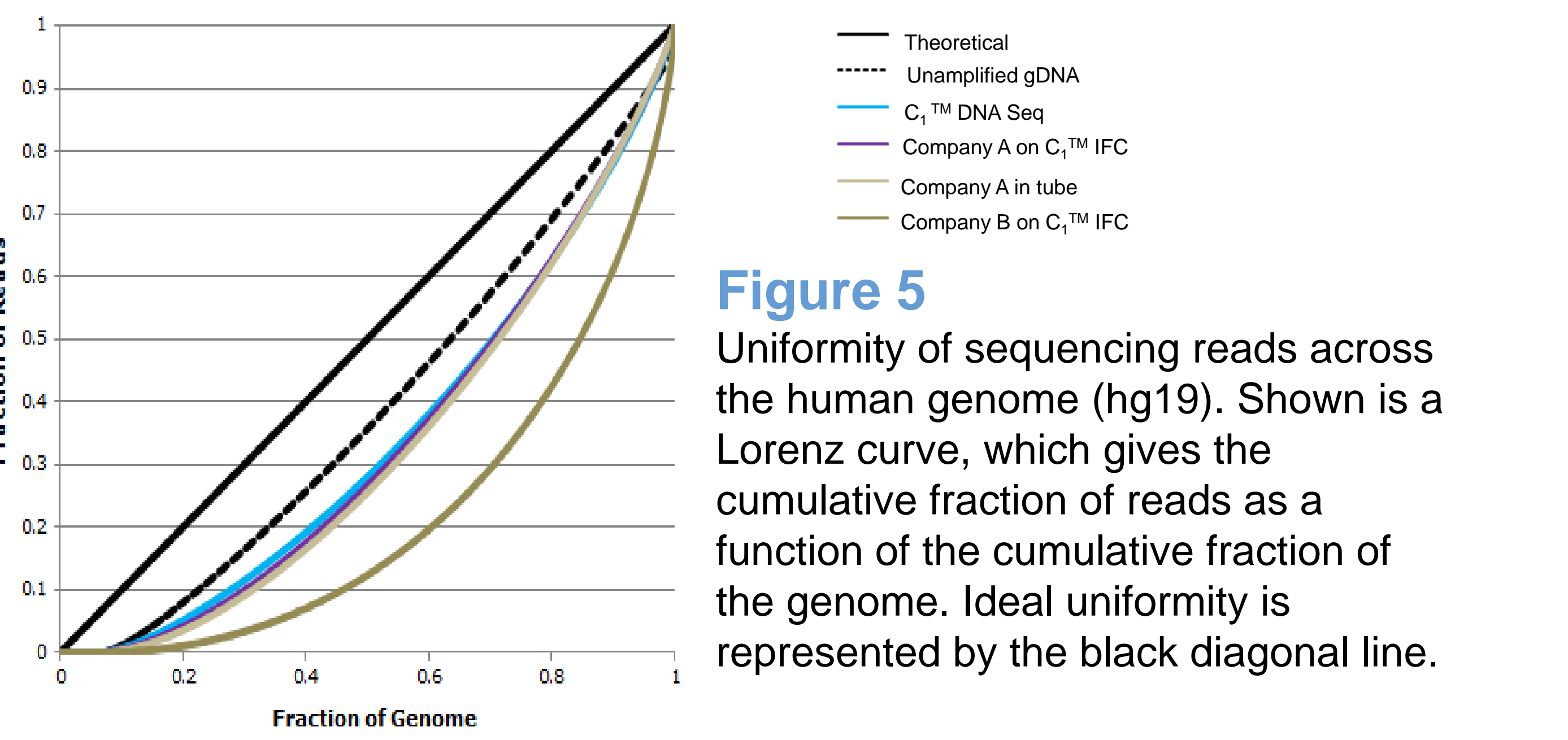
Figure 5
Uniformity of sequencing reads across the human genome (hg19). Shown is a Lorenz curve, which gives the cumulative fraction of reads as a function of the cumulative fraction of the genome. Ideal uniformity is represented by the black diagonal line.

Single Nucleotide Variant Identification and Characterization in Single Cells
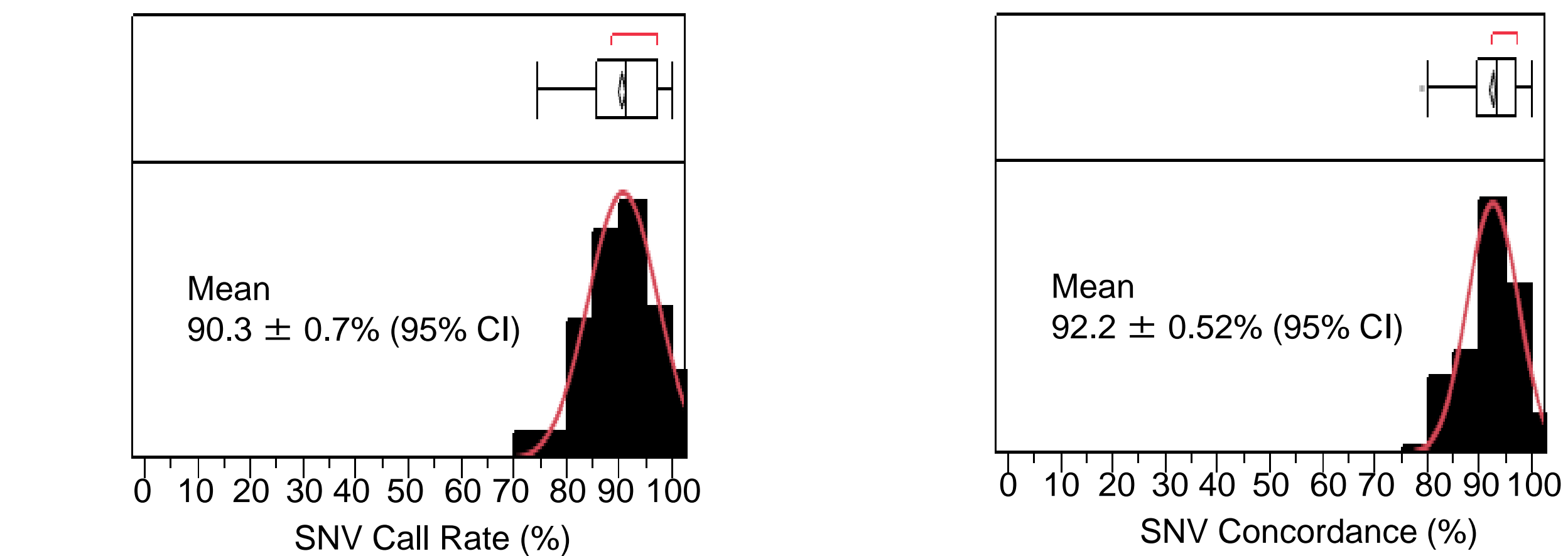
Figure 6
C_1™ DNA Seq SNV detection yields >90% sensitivity and >92% concordance, with less than 10% allelic dropout, in single-cell whole-genome amplified DNA as compared to bulk genomic DNA (369 single GM12752 cells, 27 SNVs). DNA amplified with C_1™ DNA Seq was used for targeted resequencing of a panel of oncogenes (EGFR, MET, BRCA1, BRCA2, and TP53) on the 48.48 Access Array™ System and Illumina MiSeq platform.

Figure 7
Alleles from heterozygous loci (green) identified in bulk genomic DNA can be variably distributed across a population of single cells. The distribution of SNVs identified in bulk genomic DNA among single cells (369 single GM12752 cells, 27 SNVs) using the SINGuLAR™ Analysis Toolset.

Figure 8
Single-cell SNVs are uniquely distributed among breast cancer (CRL-2338) and normal lymphoblasts (CRL-2339) derived from the same individual. Forty-five single cells and 3 bulk genomic DNA controls from both CRL-2338 and CRL-2339 were isolated and subjected to C_1™ DNA Seq whole genome amplification followed by targeted re-sequencing of a panel of oncogenes (EGFR, MET, BRCA1, BRCA2, and TP53) on the 48.48 Access Array™ System and Illumina MiSeq platform.

Conclusions

- The DNA Seq protocol on the C_1™ Single-Cell Auto Prep System generates whole genome amplified DNA suitable for targeted resequencing, whole exome sequencing, and whole genome sequencing from up to 96 individual cells in a single workflow.

- Whole genome sequencing of DNA template amplified with C_1™ DNA Seq demonstrated >95% read alignment and <0.5% base error rates, comparable to what was observed for unamplified genomic DNA.

- C_1™ DNA Seq whole genome amplification offers improvements in genomic coverage, uniformity, and GC bias over state-of-the-art WGA methods.

- C_1™ DNA Seq SNV detection yields >90% sensitivity and >92% concordance, with less than 10% allelic dropout, in single-cell whole-genome amplified DNA as compared to bulk genomic DNA.

- Application of C_1™ DNA Seq to breast cancer cells and normal lymphoblasts from the same individual revealed a unique distribution of single-nucleotide variants among diseased and normal single-cell populations.