

Robust high-performance nanoliter-volume single-cell multiple displacement amplification on planar substrates

Kaston Leung^{a,1}, Anders Klaus^{a,1}, Bill K. Lin^a, Emma Laks^{b,c}, Justina Biele^{b,c}, Daniel Lai^b, Ali Bashashati^b, Yi-Fei Huang^{b,c}, Radhouane Aniba^{b,c}, Michelle Moksa^{a,d}, Adi Steif^{b,c}, Anne-Marie Mes-Masson^{e,f,g}, Martin Hirst^{a,d,h}, Sohrab P. Shah^{b,c,h}, Samuel Aparicio^{b,c,h}, and Carl L. Hansen^{a,i,2}

^aMichael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada V6T 1Z4; ^bDepartment of Molecular Oncology, BC Cancer Agency, Vancouver, BC, Canada V5Z 1L3; ^cDepartment of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada V6T 1Z4;

^dDepartment of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada V6T 1Z4; ^eCentre de Recherche du Centre Hospitalier de l'Université de Montréal, Montréal, QC, Canada H2X 0A9; ^fInstitut du Cancer de Montréal, Montréal, QC, Canada H2X 0A9; ^gDivision of Gynecologic Oncology, Department of Medicine, Université de Montréal, Montréal, QC, Canada H3C 3J7; ^hMichael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada V5Z 1L3; and ⁱDepartment of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

Edited by Paul C. Blainey, Massachusetts Institute of Technology, Cambridge, MA, and accepted by Editorial Board Member Douglas Koshland May 25, 2016 (received for review October 24, 2015)

The genomes of large numbers of single cells must be sequenced to further understanding of the biological significance of genomic heterogeneity in complex systems. Whole genome amplification (WGA) of single cells is generally the first step in such studies, but is prone to nonuniformity that can compromise genomic measurement accuracy. Despite recent advances, robust performance in high-throughput single-cell WGA remains elusive. Here, we introduce droplet multiple displacement amplification (MDA), a method that uses commercially available liquid dispensing to perform high-throughput single-cell MDA in nanoliter volumes. The performance of droplet MDA is characterized using a large dataset of 129 normal diploid cells, and is shown to exceed previously reported single-cell WGA methods in amplification uniformity, genome coverage, and/or robustness. We achieve up to 80% coverage of a single-cell genome at 5x sequencing depth, and demonstrate excellent single-nucleotide variant (SNV) detection using targeted sequencing of droplet MDA product to achieve a median allelic dropout of 15%, and using whole genome sequencing to achieve false and true positive rates of 9.66×10^{-6} and 68.8%, respectively, in a G1-phase cell. We further show that droplet MDA allows for the detection of copy number variants (CNVs) as small as 30 kb in single cells of an ovarian cancer cell line and as small as 9 Mb in two high-grade serous ovarian cancer samples using only 0.02x depth. Droplet MDA provides an accessible and scalable method for performing robust and accurate CNV and SNV measurements on large numbers of single cells.

single-cell sequencing | whole genome amplification | multiple displacement amplification | microdroplet | nanoliter volume

Genomic heterogeneity is now appreciated for its functional significance across oncology, development, neuroscience, and biotechnology. Rapid advances in cell handling, DNA amplification methods, and sequencing throughput have enabled genomic interrogation of human disease and physiology with single-cell resolution, including studies of heterogeneity in blood and solid tumors (1–4), circulating tumor cells (5), neurons (6), gametes (7, 8), and embryos (9). Despite the intense interest in single-cell genome sequencing, it remains widely inaccessible for two primary reasons: lack of robustness and high cost. In particular, whole genome amplification (WGA) is a critical step required to generate sufficient DNA mass for sequencing in most single-cell genomic studies, but it is prone to amplification bias, contamination, and poor coverage. Preferential amplification of some genomic regions results in distorted representation of the original template and compromises the accuracy of downstream measurements. High variation in single-cell genomic coverage and bias, both within experiments and between users, remains a fundamental problem.

In response to these challenges, there has been considerable effort and advancement in molecular strategies for amplifying low picogram amounts of genomic material. Of these, three WGA strategies have been most widely adopted, each with its relative strengths and weaknesses (10, 11): PCR-based protocols (1, 12), multiple displacement amplification (MDA) (2–4, 7), and combinations of displacement preamplification followed by PCR (8, 13–15). Of these strategies, MDA is most widely used when

Significance

The study of cell-to-cell genomic differences in complex multicellular systems such as cancer requires genome sequencing of large numbers of single cells. This in turn necessitates the uniform amplification of single-cell genomes with high reproducibility across large numbers of cells, which remains an outstanding challenge. Here, we introduce a method that uses commercially available liquid dispensing to perform inexpensive and high-throughput single-cell whole genome amplification (WGA) in nanoliter volumes. For the first time, to our knowledge, we demonstrate robust and highly uniform nanoliter-volume single-cell WGA across a large replicate set consisting of more than 100 single cells. Comparison with previous datasets shows that this method improves uniformity and achieves levels of genome coverage and genomic variant detection comparable or superior to existing methods.

Author contributions: K.L., A.K., S.P.S., S.A., and C.L.H. designed research; K.L., A.K., and B.K.L. performed research; E.L., J.B., Y.-F.H., R.A., M.M., A.S., A.-M.M.-M., and M.H. contributed new reagents/analytic tools; K.L., A.K., D.L., and A.B. analyzed data; and K.L., A.K., and C.L.H. wrote the paper.

Conflict of interest statement: K.L., A.K., S.A., S.P.S., and C.L.H. are coinventors on a patent application (PCT/CA2016/000031) that covers the methods and devices described in this paper, and have a potential financial interest in this work through the revenue-sharing policies of the University of British Columbia. Following submission of this manuscript, the aforementioned patent was exclusively licensed to AbCellera (www.abcellera.com), a University of British Columbia-based startup company in which K.L., A.K., and C.L.H. have an equity position.

This article is a PNAS Direct Submission. P.C.B. is a guest editor invited by the Editorial Board.

Data deposition: The sequences reported in this paper have been deposited in the Sequence Read Archive database (accession no. [SRP078069](https://www.ncbi.nlm.nih.gov/sra/SRP078069)). Supporting datasets are available at www.msl.ubc.ca/sites/default/files/pdata/kaston-leung/Leung-et-al-supplementary-dataset-1.zip, www.msl.ubc.ca/sites/default/files/pdata/kaston-leung/Leung-et-al-supplementary-dataset-2.zip, www.msl.ubc.ca/sites/default/files/pdata/kaston-leung/Leung-et-al-supplementary-dataset-3.zip, and www.msl.ubc.ca/sites/default/files/pdata/kaston-leung/Leung-et-al-supplementary-dataset-4.zip as described in the *SI Appendix*.

¹K.L. and A.K. contributed equally to this work.

²To whom correspondence should be addressed. Email: carl.lars.hansen@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1520964113/-DCSupplemental.

maximization of coverage breadth (fraction of reference genome covered at $\geq 1\times$) is the priority. Previous studies have shown that the implementation of MDA in nanoliter volumes can reduce amplification bias (7, 16, 17), minimize contamination and non-specific products, and provide economies of scale that can significantly reduce reagent costs. However, to date, this approach has been demonstrated on only a limited number of single cells (< 10). Although such datasets highlight the optimal performance of a method, these small sample sizes make it impossible to assess robustness, performance, or scalability, which are of utmost importance when evaluating a WGA method. Furthermore, nanoliter-volume MDA has been implemented exclusively in microfluidic devices that require specialized instrumentation and micro-fabrication, limiting accessibility.

To assess the robustness and performance of nanoliter-volume single-cell MDA, we adapted a commercially available non-contact liquid dispenser to systematically test amplification performance across large numbers of replicate reactions. For the first time, to our knowledge, we characterize the robustness and performance of a scalable nanoliter-volume single-cell MDA technology (referred to here as droplet MDA) across a large dataset composed of 129 normal diploid cells. Using both low-depth ($0.02\times$) and higher-depth ($6.5\times$) whole genome sequencing (WGS), we comprehensively compared the performance of droplet MDA to other recently reported state-of-the-art methods. This large single-cell dataset shows that our approach exceeds previously reported methods in one or more key performance metrics, including amplification bias, robustness, scalability, and coverage breadth. Our study further reveals that the observed variability in droplet MDA performance is driven by biological differences in cell state and cycle.

We then assessed the copy number variant (CNV) measurement capabilities enabled by droplet MDA using single cells from both a normal diploid cell line and an ovarian cancer cell line with CNVs on multiple length scales. We next evaluated single nucleotide variant (SNV) detection performance in single cells using both PCR-based targeted sequencing and deep WGS ($48\times$) of droplet MDA product. Finally, to demonstrate applicability to primary samples, we used droplet MDA and low-depth WGS to perform CNV profiling of single nuclei from two high-grade serous ovarian cancer specimens. Taken together, this dataset demonstrates that droplet MDA provides a scalable and accessible method of performing robust and accurate CNV and SNV measurements on large numbers of single cells.

Results

Nanoliter-Volume Single-Cell MDA Reaction Formulation. All cells and reagents were deposited onto a planar substrate using a commercially available piezoelectric noncontact liquid dispenser (Fig. 1 and *SI Appendix*, Fig. S1). Droplets containing single cells were obtained by dispensing cells at limiting dilution. Cells were stained with a fluorescent dye binding double-stranded DNA,

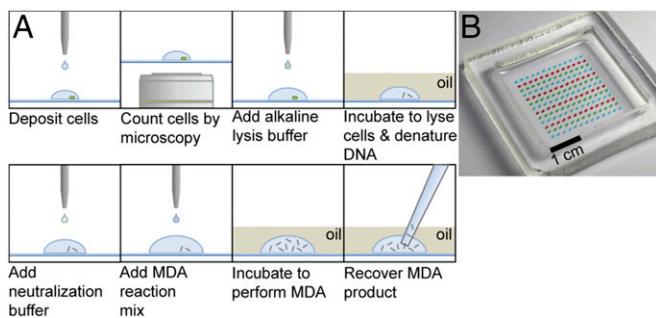


Fig. 1. Droplet MDA system. (A) Schematic of single-cell droplet MDA protocol. (B) Substrate with 100-nL droplets of food dye covered by light mineral oil.

and the number of cells in each droplet was then determined by fluorescence microscopy. We observed an average single-cell occupancy rate of 22% over all experiments described here. MDA reagents were then sequentially deposited onto each droplet. To minimize evaporation during heating of the substrate, the droplet array was covered with mineral oil before any heating steps (18). A total MDA reaction volume of 100 nL was used for this study. Additional details of methods and materials are available in *SI Appendix, Materials and Methods*.

Assessment of Amplification Bias from Low-Depth WGS. To evaluate the amplification bias of droplet MDA, we analyzed cells from the 184-hTERT mammary epithelial cell line, which has been well characterized and has a chromosomally stable and nearly diploid karyotype (19). To make assessment of a large number of replicates practical, we initially performed low-depth (mean, $0.02\times$) WGS of single-cell MDA reaction products.

We developed an analysis pipeline to allow for the fair comparison of sample datasets having unequal average sequencing depths using the HMMcopy software package (20) (*SI Appendix, Materials and Methods*). We first evaluated the effect of amplification time and total MDA yield on bias and coverage. Amplification bias has been shown to scale with reaction gain (ratio of output DNA mass to input DNA mass) (10), and reduction of MDA time has been used to reduce bias (4, 17). In contrast to previous results, we observed a trend of reduced bias with increasing single-cell MDA reaction time and yield (*SI Appendix, Table S1*), as determined by the SD of reads per 1-Mb bin (*SI Appendix, Fig. S2*). We posit that this effect is due to the lower yield of nanoliter-volume reactions and losses in subsequent sequencing library preparation, because each genomic region must be amplified above a threshold to be represented in the library. Based on this result, we performed all subsequent MDA reactions for 18–20 h. The typical yield of our 100 nL-volume single-cell MDA reactions was ~ 60 ng, representing roughly a 10,000-fold amplification of a normal diploid human genome.

Using this optimized protocol, we performed droplet MDA on 166 single 184-hTERT cells and nuclei in six separate experiments spanning an 8-mo period. Of these, 129 single-cell samples were selected at random for low depth (mean, $0.02\times$) WGS, of which FACS was used to obtain 10 nuclei in G1 phase, 11 nuclei in S phase, and 13 nuclei in G2 phase (*SI Appendix, Fig. S3*). The remaining 95 samples of the 129 samples sequenced were unsorted. Three reactions containing cell suspension fluid but no cells, which were processed on the same substrate directly adjacent to single-cell reactions, were also sequenced as stringent negative controls to detect potential cross-contamination between reactions or cell-free DNA in the suspension.

We assessed the bias of droplet MDA relative to the following published methods that have been shown to achieve high coverage breadth: multiple annealing and looping-based amplification cycles (MALBAC) on single cancer cell line cells (15), MDA of single sperm cells in custom microfluidic devices (referred to herein as “custom microfluidic MDA”) (7), microwell MDA of single neurons (MIDAS) (17), MDA of single cancer cell line cells in G2/M phase (nuc-seq) (4), WGA of genomic fragments partitioned into picoliter droplets performed on single endothelial cells (eWGA) (21), and MDA of single B-lymphoblast cell line cells in a commercially available microfluidic device (referred to herein as “commercial microfluidic MDA”) (22). We analyzed all publicly available single-cell WGS data in each study, using data from normal diploid cells for comparison wherever possible. We note that sperm cells are haploid and thus are expected to have reduced coverage and increased bias relative to samples with higher ploidy. Conversely, the nuc-seq method, which specifically selects single cells with at least four genome copies, is expected to benefit from increased ploidy.

With these caveats, we sought to perform a fair comparison of all methods using the aforementioned pipeline to reanalyze all raw datasets using the same analysis tools and parameters on

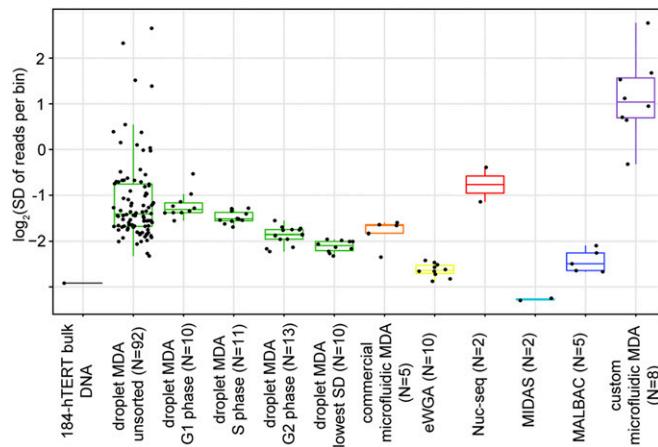


Fig. 2. Scatterplots and boxplots of SD of reads per 1-Mb bin comparing other published methods with all 184-hTERT single-cell droplet MDA samples sequenced to low depth. Unamplified bulk 184-hTERT gDNA is also included for comparison. Shown for the 184-hTERT single-cell droplet MDA samples are unsorted single cells (with three extreme outliers omitted), single cells FACS-sorted by cell phase (G1, S, and G2), and the 10 samples with the lowest SDs of all sorted and unsorted single cells.

equal quantities of aligned data. To reduce the confounding of bias measurements by true biological variation in single cell genomes, we performed CNV analysis on sequencing data from bulk DNA from nonnormal diploid cell lines where available, and omitted chromosomes with large-scale CNVs from our bias analysis (*SI Appendix*, Fig. S4 and Table S2). Although the omission of such chromosomes is necessary to perform as fair of a comparison as possible, we note that it often is not done in other comparisons in the literature. Chromosomes X and Y were omitted for the sperm cells.

We first computed basic alignment metrics using data from all samples randomly downsampled to the same number of total sequenced bases (*SI Appendix*, Table S3). Single-cell droplet MDA libraries ($n = 129$) had a mean duplicate rate of $0.11 \pm 0.025\%$, a mean alignment rate of $97.99 \pm 0.65\%$, and a mean coverage breadth of $0.71 \pm 0.11\%$. This was the highest observed alignment rate and the lowest duplicate rate of all methods compared. The mean coverage breadth of droplet MDA samples was higher than that of all other datasets, with the exception of those reported from MIDAS. Libraries generated from droplet MDA negative control reactions containing cell suspension fluid but no cells had negligible coverage breadth (mean, $0.065 \pm 0.025\%$; $n = 3$), showing that contamination, both from external sources and from other reactions on the same substrate, do not contribute appreciably to sequencing data.

To accommodate the low WGS depth to which all included single-cell droplet MDA samples were sequenced, we performed a comparison using a bin size of 1 Mb (mean of 100 reads per bin). We then assessed the bias on this genomic length scale by comparing the SD in reads per bin (Fig. 2). When considering the 92 unsorted single-cell droplet MDA samples (excluding three extreme outliers), we observed a median SD lower than that of both nuc-seq and custom microfluidic MDA. When including only the 10 droplet MDA samples (the largest number of samples available from all other methods compared) with the lowest SD, the SD values of droplet MDA samples were significantly ($P < 0.05$) lower than those of commercial microfluidic MDA ($P = 0.0376$, one-sided Wilcoxon rank-sum test), nuc-seq ($P = 0.0152$), and custom microfluidic MDA ($P = 2.29e-05$). These results are qualitatively reflected in the read depth plots for the sample with the lowest SD from each method (Fig. 3). The MIDAS samples were found to exhibit SDs below that of the unamplified bulk 184-hTERT DNA sample (0.152 vs. 0.19). This result, based on reported data for the MIDAS method, is

difficult to reconcile with the mechanism of MDA amplification, and we note that another group was unable to reproduce the MIDAS results (10). The SDs of all samples compared are listed in *SI Appendix*, Table S4.

Because visual inspection by fluorescence microscopy was used to identify cells in droplets, we hypothesize that many of the non-FACS-sorted samples included apoptotic cells or cellular debris, from which reduced amplification uniformity relative to intact and viable cells would be expected. This is supported by the observation that the SD of droplet MDA samples that were FACS-sorted before deposition into droplets, and gated to exclude such suboptimal samples, exhibited a much narrower distribution than unsorted samples (Fig. 2). The FACS-sorted samples also allowed us to test the effect of cell phase and ploidy on amplification bias. As noted elsewhere, amplification bias is dependent on cell phase (4). We observed that the median SD in reads per bin decreased from G1 phase to S phase to G2 phase (Fig. 2). Cells in G2 phase exhibited lower SD than the nuc-seq samples, which selects for cells in G2 phase ($P = 0.0095$, one-sided Wilcoxon rank-sum test). Within any given sorted cell population, the performance of the method displays low variability. These results strongly support the notion that the main contributor to MDA variability is biological, and that samples exhibiting the highest SD likely are cells with increased genomic variation owing to the biological state of the cell, and are not the result of variability in the performance of the method itself.

We also examined the GC content in each sample group (*SI Appendix*, Fig. S5) before correction for GC content bias. Interestingly, the curves for S phase 184-hTERT cells processed by droplet MDA exhibited distinct peaks at a GC content of ~0.4, perhaps suggesting that DNA replication in S phase begins in regions with this level of GC content. Notably, the curves of samples from all other methods except nuc-seq exhibited very similar peaks and were distinct from those of the G1 phase 184-hTERT cells processed by droplet MDA, which decreased monotonically.

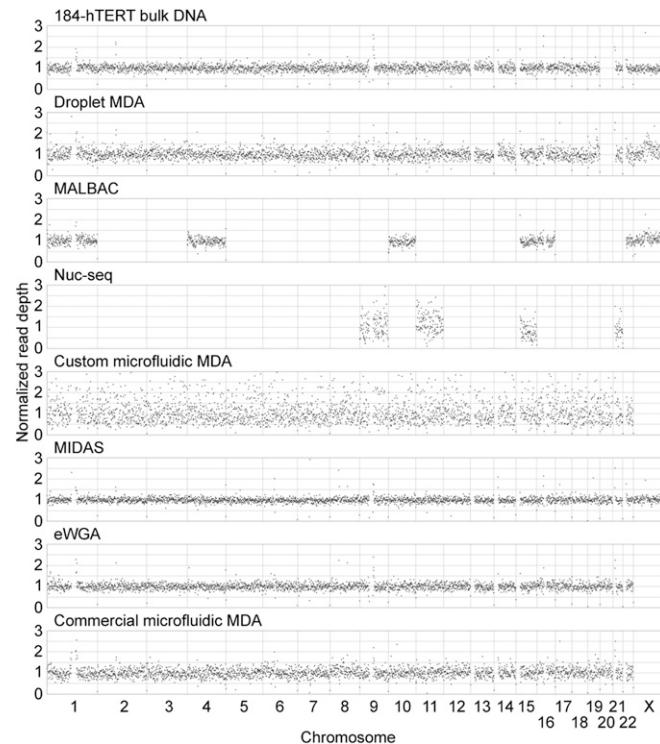


Fig. 3. Normalized read depth plots using 1-Mb bins for the sample from each method with the lowest SD in reads per bin.

Assessment of Amplification Bias from High-Depth WGS. We next performed higher-depth (mean, 6.5 \times) WGS on 15 unsorted 184-hTERT single-cell droplet MDA samples among those with the lowest SD in reads per 1-Mb bin and the two samples with the highest SD (determined from low-depth WGS data as described above), and the 10 sorted G1 phase single-cell samples. We performed a comparison with data from other methods by downsampling all datasets to the same depth (3.73 \times), defined as the number of aligned bases divided by reference size (taking into consideration omitted genomic regions). From these downsampled data, we generated Lorenz curves for each sample to depict uniformity of coverage. This downsampling step, often omitted in other published comparisons, is crucial because higher depth can result in higher coverage breadth and thus a shift in the X-intercept of the Lorenz curve. Lorenz curves for the two samples with the lowest SD in each sample group are shown in Fig. 4A, and those of all samples analyzed are shown in *SI Appendix, Fig. S6*. MIDAS samples were not included in this analysis, because available data were at a depth of only 0.2 \times . The most uniform unsorted droplet MDA and commercial microfluidic MDA samples had the highest uniformity of all samples analyzed, whereas the G1-phase droplet MDA samples were less uniform than those of all other methods using diploid cells, again underscoring the important role of cell state.

We then used the analysis pipeline described above to bin reads into 1 kb-bins (mean of 30 reads per bin), computed the SD in reads per bin, and plotted coverage breadth as a function of sequencing depth (0.5–5 \times) for the sample with the lowest SD from each method (Fig. 4B). As before, the commercial microfluidic MDA and droplet MDA samples achieved the highest coverage breadth of all samples at any given sequencing depth, covering ~84% and 80% of the reference genome, respectively, when sequenced to 5 \times depth.

Next, we used read depths from the 1-kb bins to analyze the characteristic length scale of coverage bias by computing the power spectra of read density variations. The mean power spectra of all samples from each sample group and the power spectrum of the sample with the lowest SD in reads per 1-kb bin from each sample group are shown in Fig. 4C and D, respectively. As observed in other studies, the amplitudes of MDA samples generally showed a downward inflection point between 10⁴ and 10⁵ bp, corresponding to the mean fragment length of the MDA product (10, 15, 23). Whereas the SD in reads per 1-Mb bin of droplet MDA samples was higher compared with those of eWGA and MALBAC (Fig. 2), the mean power spectra of the droplet MDA

samples dropped below that of eWGA and MALBAC at length scales smaller than 10⁴–10⁵ bp, explaining the superior performance of the droplet MDA samples in the Lorenz curve and coverage breadth vs. depth comparisons.

CNV Detection. Because the 184-hTERT cell line is normal diploid and genetically stable, we first assessed CNV detection using droplet MDA by comparing the concordance of copy number calls between single-cell samples and unamplified bulk DNA as has been done elsewhere (14, 24). We again used the HMMcopy software package, which takes in normalized binned read depth, groups contiguous bins into segments predicted to have equal copy number, and assigns a copy number to bins in each segment using a hidden Markov model (20). We binned the high-depth WGS data from the 15 samples sequenced to high-depth and bulk DNA into bins of 1 Mb, 100 kb, and 10 kb (mean of 100 reads per bin for all bin sizes), assigned a copy number to each bin using HMMcopy, and computed the binwise concordance (fraction of genomic bins that have an identical copy number state for two samples) between each single-cell sample and bulk DNA (*SI Appendix, Fig. S7*). For 1-Mb, 100-kb, and 10-kb bin sizes, the median concordance rate was 100% (of 2,487 bins), 93.1% (of 24,162 bins), and 83.4% (of 227,172 bins), respectively. Using 1-Mb bins (mean of 100 reads per bin), the median concordance rate for all 129 184-hTERT single-cell samples sequenced to low depth is 100% (*SI Appendix, Fig. S8*).

We next performed droplet MDA on 30 single cells from the TOV2295 cell line (25), derived from a high-grade serous ovarian cancer, which is genetically unstable and has CNVs on multiple length scales. Low-depth (mean, 0.02 \times) WGS was performed on all 30 samples, and higher-depth (mean, 6.5 \times) WGS was performed on the eight samples with the highest coverage breadth calculated from the low-depth WGS data. Using the high-depth data, reads were binned into 1-Mb, 100-kb, and 10-kb bins (mean of 100 reads per bin for all bin sizes). Read depth plots using 1-Mb bins from four representative single-cell samples qualitatively matched that of bulk DNA (Fig. 5), with many of the same large-scale variations discernible in both. We again used HMMcopy to find segments of contiguous bins with the same copy number. Because the TOV2295 cell line is genetically unstable, unlike the 184-hTERT cell line, there is likely to be true biological variation between the copy number profiles of individual single cells and bulk DNA. However, we were able to detect segments with identical genomic location, size, and copy number in two or more of the single-cell and bulk samples, as well as segments in two or

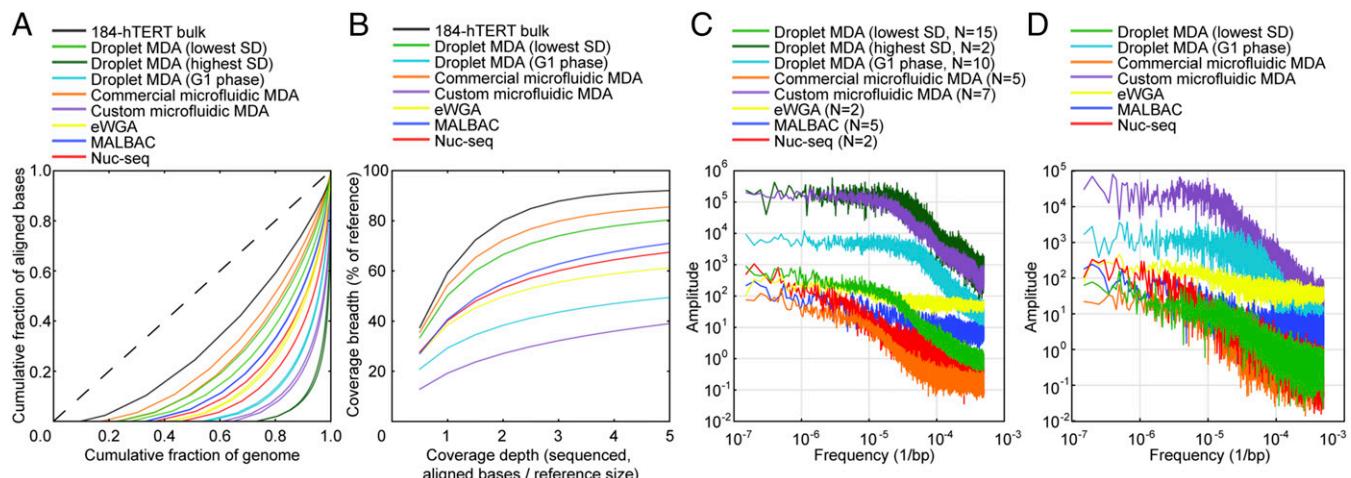


Fig. 4. Analysis of amplification bias from high-depth WGS. (A) Lorenz curves depicting uniformity of coverage for the two samples with the lowest SD in reads per 1-Mb bin from each amplification method. (B) Coverage breadth as a function of sequencing depth for the sample with the lowest SD in reads per 1-kb bin from each method. (C) Mean power spectra of 1-kb binned read depth for all samples analyzed from each amplification method. (D) Power spectra of 1-kb binned read depth for the sample with the lowest SD in reads per bin from each method.

more single-cell samples but not in the bulk sample (*SI Appendix*, Tables S5 and S6). We reason that these segments represent true CNV events and cell-to-cell heterogeneity, because it is highly unlikely that such identical events would be present by chance. Using a 10-kb bin size, the smallest such segment was 30 kb in length.

SNV Detection by Targeted Sequencing and WGS. To evaluate the effectiveness of droplet MDA for nucleotide-level measurements, we then performed targeted sequencing on 29 randomly selected 184-hTERT single cell droplet MDA samples that underwent WGS. Identification of SNVs by targeted sequencing can be used to infer clonal structure and evolutionary history (2, 26) at relatively low cost. For each sample, a portion of the MDA product was used as template for a targeted sequencing library preparation pipeline targeting 39 known heterozygous loci identified in a previous study (27). For each sample, a binomial exact test was used to call the presence or absence of an allele at each locus covered by at least 50 reads, as described previously (26) (*SI Appendix*, Fig. S9). The median allele dropout (ADO) across all samples was calculated to be 15%, comparable with the best results that have been reported previously (2). Because we performed both targeted sequencing and WGS on these samples, we were also able to determine whether there was any correlation between targeted sequencing performance and bias measured by WGS. As expected, we observed that single-cell samples with lower SD in reads per bin generally also had lower ADO (*SI Appendix*, Fig. S10). When including only the quartile of samples with the lowest SD in reads per 1-Mb bin, the median ADO was reduced to 8%.

We next evaluated the performance of SNV discovery using WGS data. We performed deep WGS (48x depth) of a single G1 phase 184-hTERT single-cell droplet MDA sample, selected to have average uniformity as defined by the median of SD in reads per 1-Mb bin of the 10 G1 phase single-cell samples analyzed. SNV false-positive rate (FPR) and true-positive rate (TPR) were evaluated at sites from the dbSNP database using the mutationSeq software tool (28), applying various coverage depth and SNV-calling probability thresholds (*SI Appendix*, Table S7; details in *SI Appendix*, Materials and Methods). We found that a

large fraction of false-positive SNVs occurred at the ends of single-cell sequencing “read islands” (*SI Appendix*, Figs. S11 and S12). Increasing the minimum single-cell coverage depth threshold removed many of these false positives at the expense of reducing the set of SNP sites considered. Because FPR and TPR have an inverse correlation (*SI Appendix*, Fig. S13), reporting of both metrics is crucial for the evaluation of SNV detection performance. For a single-cell coverage depth ≥ 15 and a mutationSeq SNV-calling probability threshold of 90%, the FPR and TPR were 9.66×10^{-6} and 68.8%, respectively. This FPR is comparable to that obtained by commercial microfluidic MDA using whole exome sequencing data from a cell in unknown cell phase (22).

Demonstration on High-Grade Serous Ovarian Cancers. As a demonstration of applicability to primary samples, we applied droplet MDA to 39 single nuclei isolated from two high-grade serous ovarian cancer tumor specimens that were previously characterized in bulk using SNP genotyping arrays (29), performed low-depth (mean, $0.02\times$) WGS, and binned reads into 1-Mb bins. For each specimen, read depth plots of four single nuclei samples are shown in *SI Appendix*, Fig. S14. In both specimens, read depth profiles from three of the single nuclei shown closely matched that of bulk DNA, whereas a fourth nucleus appeared to have a normal diploid genome, likely from noncancerous tissue collected with the tumor. Using the binned read depths, we again found copy number segments with identical genomic location, length, and copy number common to two or more samples (single nuclei and bulk) for each specimen. We found 59 and 125 such segments ranging in length from 9 to 250 Mb and from 10 to 250 Mb for specimens 1 and 2 respectively (*SI Appendix*, Tables S8 and S9).

Discussion

We have implemented a new single-cell MDA method that exploits the improved reaction performance and reduced reagent consumption of nanoliter-scale volumes using a commercially available liquid dispensing system. This approach preserves the programmable and multistep nanoliter-volume processing of our previously reported microfluidic droplet-based device (13), but without the need for any specialized microfabrication. In contrast to closed-format microfluidic devices that require multiple sizes of trapping structures to isolate single cells of various diameters (22), the droplet MDA method is able to isolate cells of different sizes without any modification. Unlike similar systems that formulate nanoliter-volume single-cell reactions on an open array (17, 18), droplet MDA avoids the possibility of cross-contamination between reactions by using noncontact dispensing to place reactions in spatially distinct locations separated by oil. Depending on the speed of the dispenser used, this approach is also rapid and highly scalable. In this study, we formulated 154 100 nL-reactions per $3\text{ cm} \times 3\text{ cm}$ substrate in ~ 4 min.

Here, we have, for the first time to our knowledge, evaluated a single-cell nanoliter-volume WGA method across a large dataset comprising 129 normal diploid cells of known genotype. In contrast, previous studies have presented data from 10 or fewer replicates. In total, we analyzed 219 single-cell droplet MDA samples for this study. Through random selection of a large fraction of all samples that were processed, this dataset allows for evaluation of the method’s robustness and the variability in performance across samples. We believe that such efforts are essential to establish the reproducibility needed to drive such applications as single-cell tumor sequencing, where biological insight depends on a method that is consistent across large numbers of cells.

We have shown that the biological state of the processed cells has a major influence on the observed amplification performance. In particular, cells sorted to exclude apoptotic cells and cellular debris exhibit a much narrower distribution in amplification bias than unsorted cells. Furthermore, our experiments on cells sorted by cell phase demonstrate that ploidy is an important

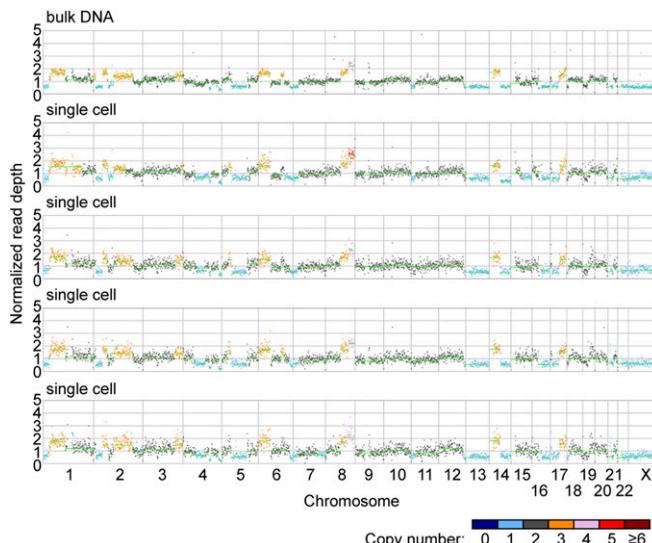


Fig. 5. Normalized read depth plots using 1-Mb bins for TOV2295 bulk DNA and four single-cell samples. Horizontal green lines indicate segments of contiguous bins inferred to have the same copy number, where the read depth of each segment is equal to the median read depth of the bins in that segment. Inferred copy numbers are indicated by the color of the data point for each bin.

determinant of single-cell WGA performance, including both GC content bias and uniformity. Within each sorted cell population, the performance of droplet MDA was highly robust, displaying low variability.

Our analyses of low-depth and high-depth WGS data indicate that the performance of droplet MDA compares favorably to that of other methods considered, with our least-biased samples having comparably low bias on multiple length scales relative to other methods. Nonetheless, we emphasize that comparing our method with previous reports is complicated by differences in sample type, quality, and size. Our data demonstrate a clear effect of cell cycle on amplification uniformity, suggesting that direct comparison of methods should be done using datasets from normal diploid cells sorted for G1 phase. Apart from the nuc-seq protocol, which focuses on isolation of G2 phase cells, previous reports do not control for cell cycle. This factor makes it difficult to evaluate performance in cases where data are present from only a small number of cells, and where the rationale for selecting single cells for analysis is not made explicit. Similarly, the absence of large datasets makes it impossible to evaluate robustness and variability of any given method.

The uniform coverage enabled by droplet MDA allows for recovery of a high fraction of the genome with relatively little sequencing effort and also enables accurate CNV calling. We identified copy number segments as small as 30 kb using 6.5 \times WGS in a cancer cell line, which to our knowledge is the most sensitive CNV detection yet demonstrated from an MDA-amplified single-cell genome, and as small as 9 Mb in primary tumor specimens using only 0.02 \times WGS. In both tumor specimens, droplet MDA analysis was able to clearly distinguish populations of cells with diploid and low aberration content from those with multiple CNVs and chromosomal structure aberrations using very low-depth WGS. A significant advantage of MDA is that its high coverage breadth and large mean fragment length allow for

application of both WGS and targeted sequencing to the same single cell. We demonstrated this by performing both on 29 single cells, achieving excellent SNV measurement performance, as quantified by low ADO. We also evaluated the performance of SNV discovery using deep WGS data from our median performing G1-phase cell and observed a false discovery rate that compares favorably with that reported in other methods from single cells in unknown cell phase.

These results demonstrate that droplet MDA provides an easily accessible way for researchers to exploit the benefits of nanoliter-volume processing for robust and cost-effective interrogation of single-cell genomic variation at both the copy number and single-nucleotide levels. We anticipate that this will be an increasingly useful tool as interest in single-cell studies continues to expand across numerous disciplines in the biological sciences.

Materials and Methods

More detailed information on the materials and methods used in this study is provided in *SI Appendix, Materials and Methods*. Ethical approval was obtained from the University of British Columbia Ethics Board. Women undergoing debulking surgery (primary or recurrent) for carcinoma of ovarian/peritoneal/fallopian tube origin were approached for informed consent for the banking of tumor tissue. Tissue was obtained from tumor sites in women histologically diagnosed with high-grade serous ovarian cancer and was collected before adjuvant therapy and frozen in cryovials. Consistent with the practice at University of British Columbia and the British Columbia Cancer Agency, all patients with high-grade serous cancer are referred to the hereditary cancer clinic and offered genetic testing for *BRCA1* and *BRCA2* mutations.

ACKNOWLEDGMENTS. We thank Hans Zahn for assistance with operation of the piezoelectric dispenser and Ramunas Stepanauskas for valuable discussions regarding MDA. Funding support was provided by Genome British Columbia, Genome Canada, the National Science and Engineering Research Council of Canada, and the Canadian Institutes of Health Research.

1. Navin N, et al. (2011) Tumour evolution inferred by single-cell sequencing. *Nature* 472(7341):90–94.
2. Gawad C, Koh W, Quake SR (2014) Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *Proc Natl Acad Sci USA* 111(50):17947–17952.
3. Hou Y, et al. (2012) Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* 148(5):873–885.
4. Wang Y, et al. (2014) Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 512(7513):155–160.
5. Lohr JG, et al. (2014) Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat Biotechnol* 32(5):479–484.
6. McConnell MJ, et al. (2013) Mosaic copy number variation in human neurons. *Science* 342(6158):632–637.
7. Wang J, Fan HC, Behr B, Quake SR (2012) Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* 150(2):402–412.
8. Lu S, et al. (2012) Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. *Science* 338(6114):1627–1630.
9. Fiorentino F, et al. (2014) Development and validation of a next-generation sequencing-based protocol for 24-chromosome aneuploidy screening of embryos. *Fertil Steril* 101(5):1375–1382.
10. de Bourcy CFA, et al. (2014) A quantitative comparison of single-cell whole genome amplification methods. *PLoS One* 9(8):e105585.
11. Hou Y, et al. (2015) Comparison of variations detection between whole-genome amplification methods used in single-cell resequencing. *Gigascience* 4(1):37.
12. Baslan T, et al. (2015) Optimizing sparse sequencing of single cells for highly multiplex copy number profiling. *Genome Res* 25(5):714–724.
13. Leung K, et al. (2012) A programmable droplet-based microfluidic device applied to multiparameter analysis of single microbes and microbial communities. *Proc Natl Acad Sci USA* 109(20):7665–7670.
14. Voet T, et al. (2013) Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Res* 41(12):6119–6138.
15. Zong C, Lu S, Chapman AR, Xie XS (2012) Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* 338(6114):1622–1626.
16. Marcy Y, et al. (2007) Nanoliter reactors improve multiple displacement amplification of genomes from single cells. *PLoS Genet* 3(9):1702–1708.
17. Gole J, et al. (2013) Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells. *Nat Biotechnol* 31(12):1126–1132.
18. Zhu Y, et al. (2015) Printing 2-dimensional droplet array for single-cell reverse-transcription quantitative PCR assay with a microfluidic robot. *Sci Rep* 5:9551.
19. Raouf A, et al. (2005) Genomic instability of human mammary epithelial cells over-expressing a truncated form of EMSY. *J Natl Cancer Inst* 97(17):1302–1306.
20. Ha G, et al. (2012) Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. *Genome Res* 22(10):1995–2007.
21. Fu Y, et al. (2015) Uniform and accurate single-cell sequencing based on emulsion whole-genome amplification. *Proc Natl Acad Sci USA* 112(38):11923–11928.
22. Szulwach KE, et al. (2015) Single-cell genetic analysis using automated microfluidics to resolve somatic mosaicism. *PLoS One* 10(8):e0135007.
23. Zhang C-Z, et al. (2015) Calibrating genomic and allelic coverage bias in single-cell sequencing. *Nat Commun* 6:6822.
24. Macaulay IC, et al. (2015) G&T-seq: Parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods* 12(6):519–522.
25. Létourneau IJ, et al. (2012) Derivation and characterization of matched cell lines from primary and recurrent serous ovarian cancer. *BMC Cancer* 12(1):379.
26. Eirew P, et al. (2015) Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* 518(7539):422–426.
27. Roth A, et al. (2014) PyClone: Statistical inference of clonal population structure in cancer. *Nat Methods* 11(4):396–398.
28. Ding J, et al. (2012) Feature-based classifiers for somatic mutation detection in tumour-normal paired sequencing data. *Bioinformatics* 28(2):167–175.
29. Bashashati A, et al. (2013) Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J Pathol* 231(1):21–34.

Supplementary Information

Robust high-performance nanoliter-volume single-cell multiple displacement amplification on planar substrates

Kaston Leung, Anders Klaus, Bill K. Lin, Emma Laks, Justina Biele, Daniel Lai, Ali Bashashati, Yi-Fei Huang, Radhouane Aniba, Michelle Moksa, Adi Steif, Anne-Marie Mes-Masson, Martin Hirst, Sohrab P. Shah, Samuel Aparicio, Carl L. Hansen

SI Materials and methods

Nanoliter-volume reaction formulation

All reagents and cell suspensions were dispensed onto the substrate using a non-contact piezoelectric spotter (S3, Scienion). A PDC 70 nozzle with a type 4 coating was used for all cell suspensions and reagents with the exception of the alkaline lysis solution for which an uncoated PDC 50 nozzle was used (all nozzles supplied by Scienion). Piezoelectric voltage and pulse width settings were adjusted for each dispensed reagent in order to achieve stable droplet formation. Nanoliter-volumes of reagents were assembled by dispensing droplets with volumes ranging from 100-400 pL at a frequency of 500 Hz, which coalesced on the substrate. The substrate was placed on a water-cooled chuck cooled to approximately 6°C during all dispensing steps to minimize evaporation without inducing condensation on the substrate. Prior to any heating steps, the droplet array was covered by approximately 2 mL of light mineral oil using a pipette, in order to minimize evaporation in the aqueous reaction droplets. All oil was then removed by pipette prior to any subsequent reagent addition.

Fabrication of substrate:

For robust addition of reagents to nanoliter-volume droplets on the substrate, the non-contact liquid dispenser used requires removal of all oil covering the droplets prior to deposition of additional reagents. The hydrophobicity of the substrate was thus adjusted such that deposited aqueous droplets remained immobilized during addition and removal of oil, and did not excessively spread out, causing possible cross-contamination with other neighbouring droplets on the substrate. Glass slides (75 mm x 50 mm) were spin coated with a 30 µm-layer of polydimethylsiloxane (PDMS) (Momentive). PDMS barriers, which keep the mineral oil on the reaction array (Fig. 1B), were fabricated by pouring liquid PDMS into a polymethylmethacrylate mold, which was then degassed in a vacuum chamber to remove air bubbles that may interfere with bonding of barriers to the slides. The PDMS base to curing agent ratio was 10:1 (by mass). The PDMS-coated slides and barriers were baked for 24 hrs at 80°C, then stored for future use at room temperature. Immediately prior to an experiment, barriers were bonded to the slides by plasma oxidation and the assembled substrates were baked for 20 min at 80°C. This plasma treatment served to both bond the barrier to the slide and to decrease the hydrophobicity of the PDMS surface. Substrates were then DNA decontaminated by exposure to 4 J/cm² of UV radiation using a UV oven.

Cell culture

184-hTERT cells were cultured at 37°C in 5% CO₂ in MEBM Mammary Epithelial Cell Basal Medium (Lonza) with 5 µg/mL transferrin (Sigma) and 2.5 µg/mL isoproterenol (Sigma). The media was further supplemented with MEGM Mammary Epithelial Cell Growth Medium Singlequots (Lonza) of 2 mL bovine pituitary extract, 0.5 mL hEGF, 0.5 mL insulin, and 0.5 mL hydrocortisone.

TOV2295 cells were cultured at 37°C in 5% CO₂ in a 1:1 mixture of MCDB 105 / M199 (Sigma) supplemented with 10% fetal bovine serum.

Preparation of cells and nuclei

Cultured, frozen 184-hTERT and TOV2295 cells were thawed at room temperature. 150 µL of cells was washed with 1100 µL PBS in a 1.5 mL tube. The washed cells were pelleted by centrifugation at 2500 rpm and 4°C for 5 min and the supernatant was discarded. After a second wash and spin, the cells were resuspended in PBS with 5% glycerol. The glycerol serves to reduce evaporation of the nanoliter-volume droplets on the substrate. Additional PBS with 5% glycerol was added as necessary to achieve a concentration of approximately 1 cell per 8 nL, which was confirmed by haemocytometer. SYTO9 nucleic acid stain (Life Technologies) was added to the suspension to a final concentration of 15 µM.

Unsorted 184-hTERT nuclei were prepared by thawing cultured frozen 184-hTERT cells. 300 µL of cells was mixed with 300 µL of EZ lysis buffer (Sigma) in a 1.5 mL tube and incubated for 5 minutes on ice. 500 µL of PBS were added to the lysed cells and nuclei were pelleted by centrifugation at 2500 rpm and 4°C for 5 min. The supernatant was discarded and the nuclei were resuspended in PBS with 5% glycerol. After filtering using a 70 µm filter, additional PBS with 5% glycerol was added as necessary to achieve a concentration of approximately 1 cell per 8 nL, which was confirmed by haemocytometer.

184-hTERT nuclei sorted by cell cycle were prepared as above but resuspended in 1 mL PBS. 2 µL of propidium iodide (PI) stain (Life Technologies) was added and nuclei were filtered through a second 70 µm filter prior to sorting. G1, S, and G2 populations were differentiated by DNA content, as determined by the relative fluorescence of PI stain, and sorted into separate 1.5 mL tubes using a FACSaria III instrument (BD Biosciences). After sorting, nuclei were diluted as necessary with PBS with 5% glycerol to achieve a concentration of approximately 1 nucleus per 8 nL.

For primary samples, three 40 µm wide cryosections were mechanically homogenized in a tube containing 1 mL EZ lysis buffer (Sigma) using a laboratory paddle-blender. The homogenized tissue was then filtered with a 70 µm filter and washed with PBS as described above. The washed nuclei were resuspended in 1 mL PBS and 2 µL of propidium iodide stain (Life Technologies) was added. Nuclei were filtered through a second 70 µm filter and sorted and diluted as described above.

The SYTO9 stain, PBS, and glycerol were all DNA decontaminated prior to use by exposure to 4 J/cm² of UV radiation using a UV oven.

Cell deposition and identification

Prepared cell/nuclei suspensions were deposited onto the substrate in 8 nL droplets. The number of cells deposited at each position was visually determined by manual

fluorescent/brightfield microscopy using a fluorescent microscope (Leica) equipped with filters for excitation and emission wavelengths of SYTO9.

Cell lysis and multiple displacement amplification

To lyse cells and nuclei and denature double-stranded genomic DNA, 6 nL of alkaline lysis buffer consisting of 400 mM KOH, 10 mM EDTA, and 100 mM DTT (pH ~13) was added to each droplet. The droplet array was then covered with light mineral oil, and reactions were incubated at 65°C for 10 min by placing the substrate onto a flatbed thermocycler (Bio-Rad). The oil was then removed and 6 nL of neutralization buffer consisting of 1 M TrisHCl (pH ~4) was deposited on each droplet to neutralize the alkaline buffer. 80 nL of REPLI-g Single Cell Master mix (REPLI-g Single Cell Kit, Qiagen) supplemented with 0.1% Tween 20 surfactant and 5% glycerol was then added to each droplet and the array was again covered with light mineral oil. The MDA reactions (100 nL total volume) were then incubated at 30°C for the specified incubation time, after which amplification was terminated by heating to 65°C for 15 min. The mineral oil, alkaline lysis buffer, neutralization buffer, water, glycerol, and Tween 20 were all DNA decontaminated prior to use by exposure to 4 J/cm² of UV radiation using a UV oven.

Extraction of amplified DNA

After amplification, mineral oil was removed and approximately 250 nL of inert blue dye (Bio-Rad) was added to all reactions chosen for further processing in order to aid with visualization of the droplets to make manual extraction by pipette easier. The array was then covered with light mineral oil and reaction droplets were extracted by pipette and transferred into 30 µL of 0.1 mM EDTA TE buffer (Teknova) in 96-well plates. MDA yield was assessed by measuring the DNA concentration of 2 µL of diluted MDA product using the Qubit dsDNA HS Assay Kit (ThermoFisher).

Whole genome library preparation and sequencing

5 ng of MDA product was used for sequencing library construction. DNA was fragmented by Covaris shearing, end-repaired, and A-tailed. Illumina adaptors were ligated to the fragments and the library was amplified by limited-cycle PCR, during which samples were barcoded. Samples were sequenced using a MiSeq (Illumina), generating 75 bp paired-end reads. Samples chosen for additional high-depth sequencing were further sequenced on a HiSeq 2500 (Illumina), generating 125 bp paired-end reads.

Alignment and analysis of whole genome sequencing data

Sequencing reads were trimmed of adaptor and barcode sequences by Illumina software on the sequencing instrument. Unaligned sequencing data (fastq files) from other published studies was downloaded from the NCBI online database. All sequencing data was aligned to the GRCh37-lite reference genome using Bowtie 2 (version 2.2.5) with minimum and maximum fragment lengths set to 80 and 500 respectively. Aligned data was sorted and PCR duplicates

were marked using Picard Tools (version 1.119). SAMtools (version 1.2) was used to index aligned, sorted data.

Statistics shown in SI Appendix, Table S3 were calculated considering the entire reference genome using Bedtools (version 2.23.0) after downsampling all samples to the same number of total sequenced bases. In samples for which bulk data was available, Control-FREEC (1) was used to identify regions of the genome containing large scale CNVs. These regions were omitted from subsequent analyses in the analysis of all single cell samples (SI Appendix, Table S2).

Comparison of all samples using binned reads was performed as follows. The HMMcopy (2) readcounter function was first used to determine the number of aligned reads, excluding duplicate reads, falling within fixed-width bins across the genome for each sample. The mean number of reads per bin of the sample with the fewest reads was then found, and a binomial distribution was then used to randomly downsample binned reads of all other samples, resulting in equal mean numbers of reads per bin across all samples. This ensured that the same quantity of aligned data was compared for all samples. For the 184-hTERT samples binned into 10 kb and 100 kb bins, HMMcopy functions in R were then used to correct downsampled binned reads for biases due to the GC content and mappability of each bin. For all other sample datasets, correction for GC content was performed but correction for mappability was not. Instead, only bins with a mappability score ≥ 0.85 were considered.

Power spectra plots were generated by concatenating the 1 kb binned read depths for each sample and considering only the number of bins of the sample with the smallest reference. Binned read depths were then entered into the R function periodogram in the “Genecycle” package (<http://CRAN.R-project.org/package=GeneCycle>), which estimates power spectral density by generating a smoothed periodogram using the command “periodogram([binned read depths], method = "smooth")”. Lorenz curves were generated from the high depth sequencing data by downsampling all samples to the same depth, defined as the number of aligned bases divided by reference size (taking into consideration omitted genomic regions in each cell type). To generate breadth vs. depth curves, each sample was downsampled to between 0.5x and 5x sequencing depth relative to its reference at increments of 0.5x and Bedtools was used to calculate coverage breadth at each depth.

CNVs were called using the HMMcopy software package, which takes in normalized binned read depth, groups contiguous bins into segments predicted to have equal copy number, and assigns a copy number to bins in each segment using a Hidden Markov Model (2). For all sample datasets except for the 184-hTERT samples binned into 10 kb and 100 kb bins, the following custom HMMcopy parameters were used for CNV-calling. Seven copy number states were used; the m values were set to 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0 for copy number states 0 to 6 respectively; the mu values were set to 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0 for copy number states 0 to 6 respectively; the kappa values were set to 25, 50, 800, 50, 25, 25, 25 for copy number states 0 to 6 respectively; the e value was set to 0.995; and the S value was set to 35. To find the concordance between copy number states of bins of single-cell samples and bulk DNA in

184-hTERT cells (SI Appendix, Fig. S7 and S8), only bins with a mappability score above 0.85 were considered.

Due to low numbers of aligned reads, the G1-phase 184-hTERT single-cell sample with sample name “G1_phase_S61” was omitted from the analysis using low depth WGS data, and the G1-phase 184-hTERT single-cell sample with sample name “G1_phase_S66” was omitted from the analysis using high depth WGS data. There is thus no 1 Mb-bin read depth plot for G1_phase_S61 and no 100 kb-bin and 10 kb-bin read depth plots for G1_phase_S66 in the supplementary datasets. One of the custom microfluidic single-sperm samples was omitted from the generation of the mean power spectra plots because it had an extremely irregular power spectrum. Its inclusion would not have lowered the amplitude of the mean power spectrum for that sample group.

Targeted sequencing library preparation and data analysis

A panel of 48 primer pairs were designed and synthesized as previously described (3) and split into two pools of 24. Two 24-plex PCRs were assembled for each sample in 96-well plates and consisted of 50 nM forward primers, 50 nM reverse primers, approximately 5 ng MDA product, and 2x SsoFast EvaGreen PCR mix (Bio-Rad). Reactions were cycled on a Chromo 4 thermocycler (Bio-Rad) with cycling conditions of 98°C for 3 min, followed by 18 cycles of [98°C for 15 sec, 60°C for 4 min, 72°C for 30 sec] and the two 24-plexes were combined.

The multiplex product for each sample was used as template for 48 single-plex PCRs which were carried out using 48.48 Access Array integrated fluidic circuits according to the manufacturer’s recommended protocol (Fluidigm). Amplicon sequencing libraries were generated as previously described (3) and sequenced using a MiSeq according to the manufacturer’s protocol (Illumina).

Reads were trimmed of adaptor and barcode sequences by Illumina software and aligned to the GRCh37-lite reference genome using Bowtie 2. A binomial exact test was used to call the presence or absence of an allele at each locus covered by at least 50 reads, as previously described (4). Only the 39 target positions consistently called as heterozygous in the bulk DNA sample were considered for the single-cell samples. Allele dropout for these positions was calculated using the formula:

$$1 - \left(\frac{\text{homozygous calls} + 2 \times \text{heterozygous calls}}{\text{heterozygous calls in bulk}} \right)$$

Calculation of SNV false and true positive rates

SNV discovery performance was evaluated for the G1-phase 184-hTERT single-cell droplet MDA sample with the median SD in reads per 1 MB-bin of the 10 G1-phase single cell samples sequenced to low depth. The WGS library for this single-cell sample was whole genome sequenced to approximately 48x depth and compared with a previously sequenced bulk genomic DNA sample sequenced to the same depth as follows. SNV sites from the

dbSNP database were first screened for sites at which the bulk sample had a coverage depth of at least 50 reads where all reads either matched the reference (referred to here as “reference-matching sites”) or all reads did not match the reference (referred to here as “non-reference-matching sites”). The mutationSeq software tool (5) was then used to determine mismatches between the single cell and reference (SNVs) at both reference-matching and non-reference-matching sites for various minimum thresholds of coverage depth in the single cell and various minimum SNV-calling probabilities in the mutationSeq tool. The false positive rate (FPR) was then defined as the number of SNVs identified in the single cell at reference-matching sites having a specified minimum single-cell coverage depth divided by the total number of reference-matching sites having the specified minimum single-cell coverage depth (i.e. the rate of identifying SNVs in the single cell that were not identified in the bulk sample). The true positive rate (TPR) was defined as the number of SNVs identified in the single cell at non-reference-matching sites having a specified minimum single-cell coverage depth divided by the total number of non-reference-matching sites having the specified minimum single-cell coverage depth (i.e. the rate of identifying SNVs in the single cell that were also identified in the bulk sample). A table showing output from mutationSeq for the various coverage depth and SNV-calling probabilities used is shown in SI Appendix, Table S7. A plot showing the inverse correlation between FPR and TPR for various parameters is shown in SI Appendix, Fig. S13.

Proximity of SNV false positives to ends of read islands were computed as follows. SNVs were identified as described above using a mutationSeq SNV-calling probability threshold of 0.9 and a minimum single-cell coverage depth of 4. For each SNV site identified as a false positive, the coverage depth of the 101 positions up and downstream of the site in the reference genome was noted. The proximity of each site to the end of a read island was then found by computing the absolute value of the difference between the number of upstream positions with coverage depth ≥ 1 and the number of downstream positions with coverage depth ≥ 1 . Thus, an SNV site with an absolute difference value of x is $(101 - x)$ bases from the end of a read island.

References

1. Boeva V, et al. (2012) Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data.
2. Ha G, et al. (2012) Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. *Genome Research* 22(10):1995-2007.
3. Roth A, et al. (2014) PyClone: statistical inference of clonal population structure in cancer. *Nat Meth* 11(4):396-398.
4. Eirew P, et al. (2015) Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* 518(7539):422-426.
5. Ding J, et al. (2012) Feature-based classifiers for somatic mutation detection in tumour–normal paired sequencing data. *Bioinformatics* 28(2):167-175.

Supplementary Figures

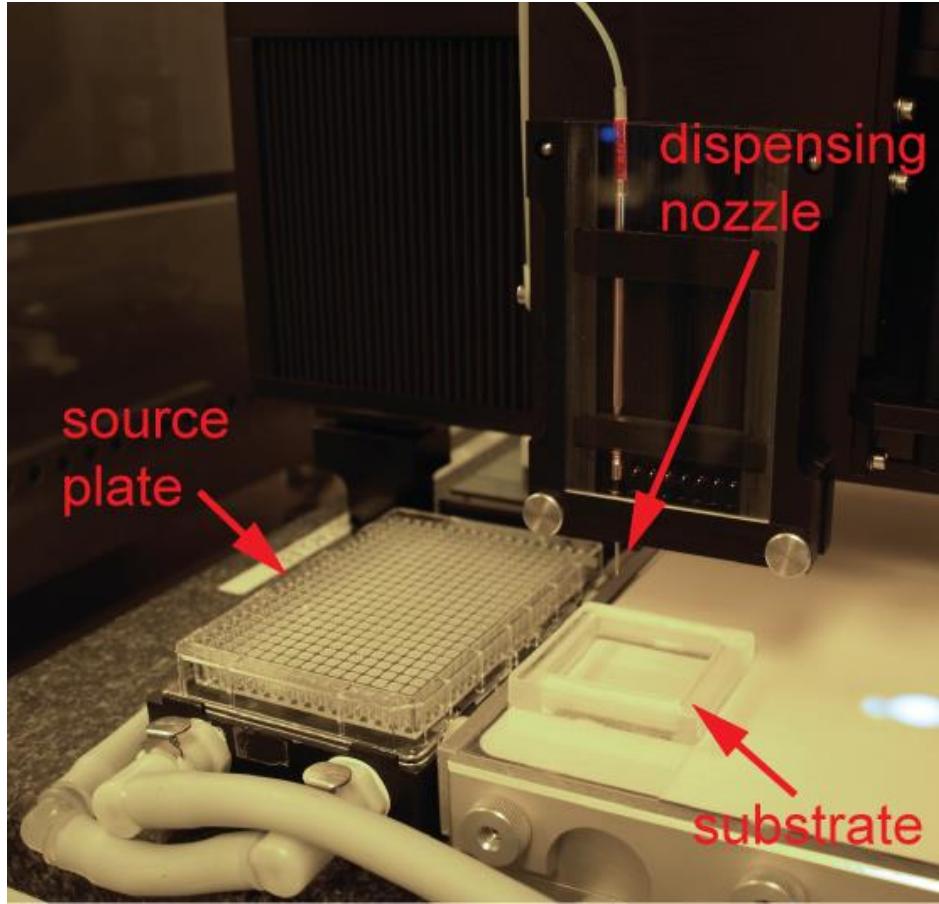


Fig. S1. Experimental setup for droplet MDA. Shown is the substrate below the piezoelectric dispensing nozzle with reagent source plate at left.

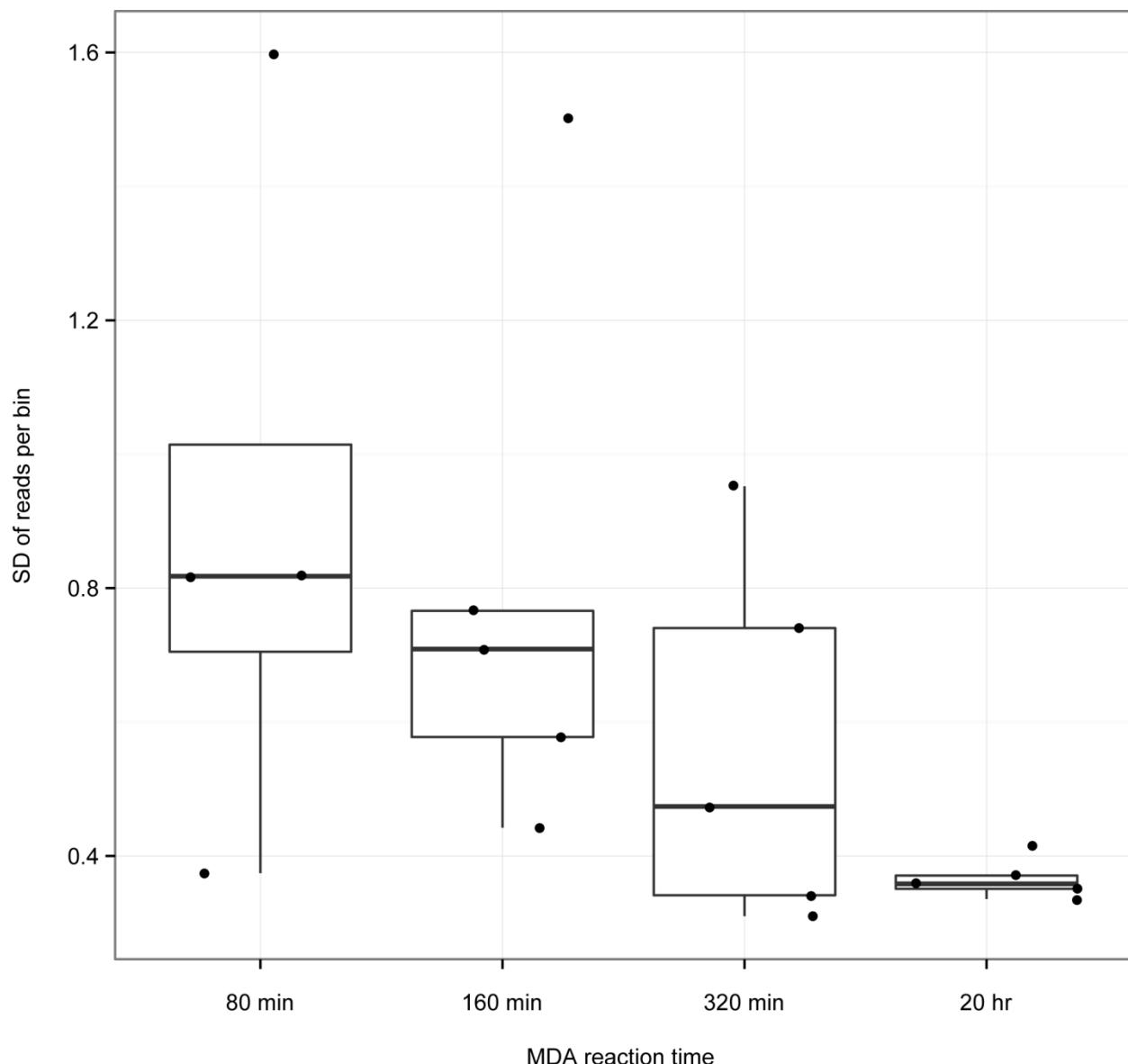


Fig. S2. Scatter and box plots of SD in reads per 1-Mb bin for different MDA incubation times. Upper whisker ranges from 75th percentile to sample with largest value within 1.5x interquartile range above 75th percentile, and lower whisker ranges from 25th percentile to sample with smallest value within 1.5x interquartile range below 25th percentile.

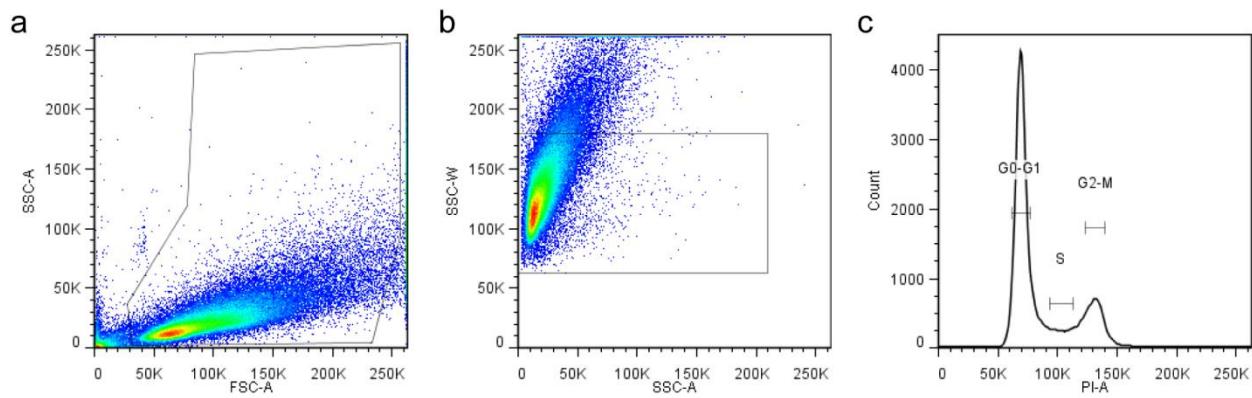


Fig. S3. FACS profiles for the gating strategy of 184-hTERT nuclei by DNA content for ploidy analysis. Nuclei were gated to exclude (A) debris and (B) doublets or clumped nuclei. (C) Nuclei in G0/G1, S, and G2/M phases were collected by flow sorting using propidium iodide staining.

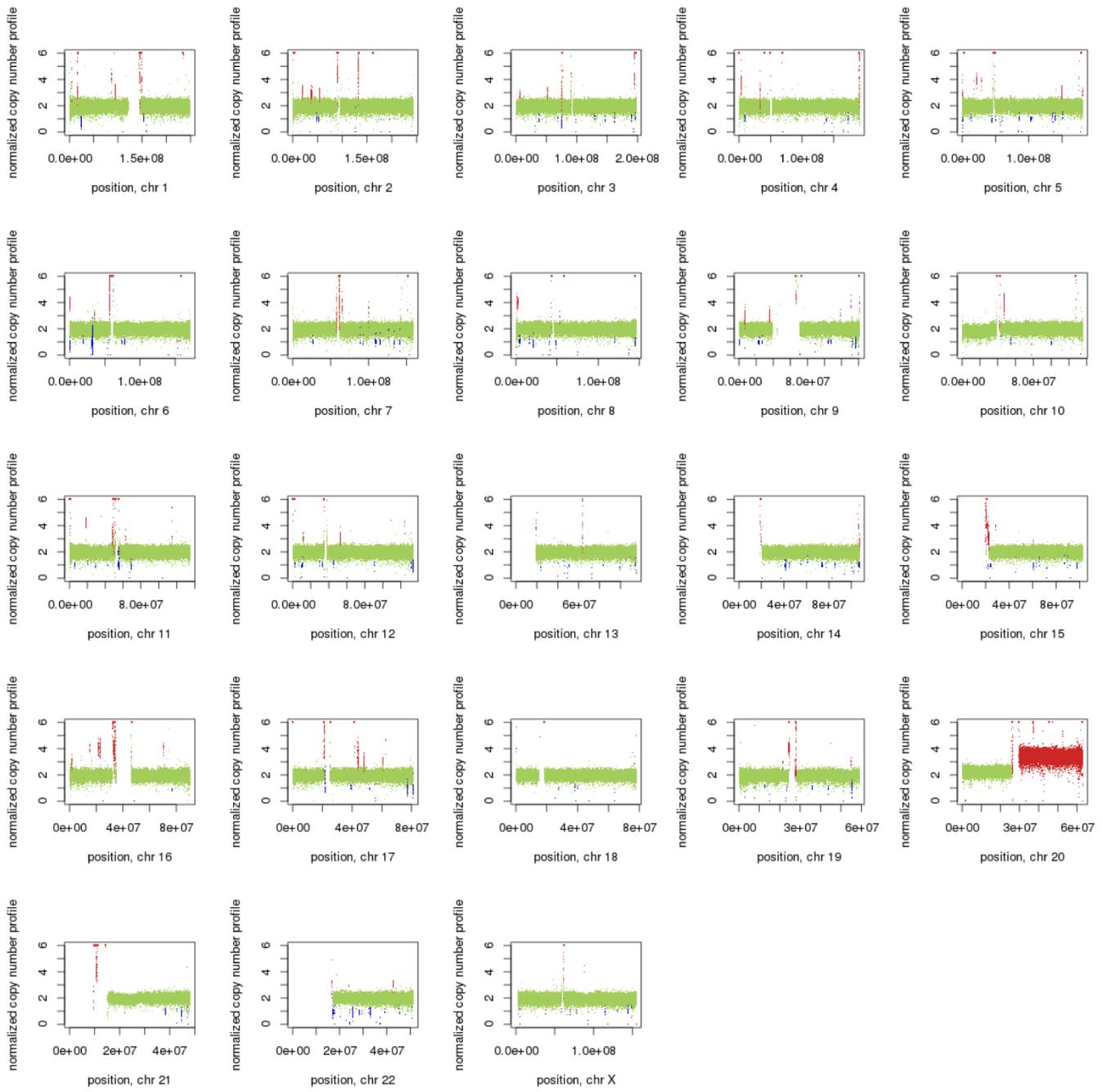


Fig. S4A. CNVs identified in 184-hTERT cell line (used in droplet MDA) bulk DNA by Control-FREEC.

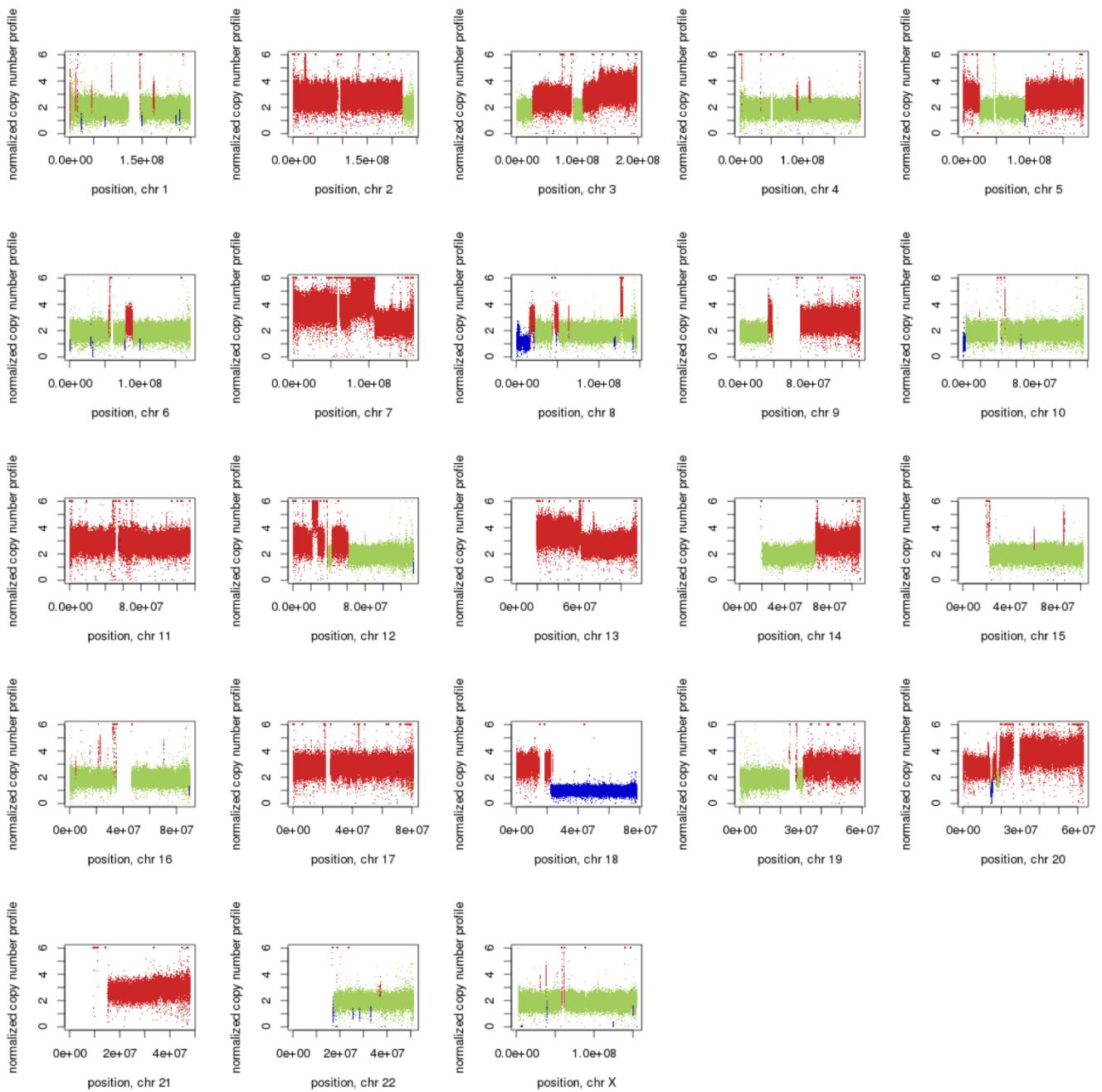


Fig. S4B. CNVs identified in SW480 cancer cell line (used in MALBAC) by Control-FREEC.

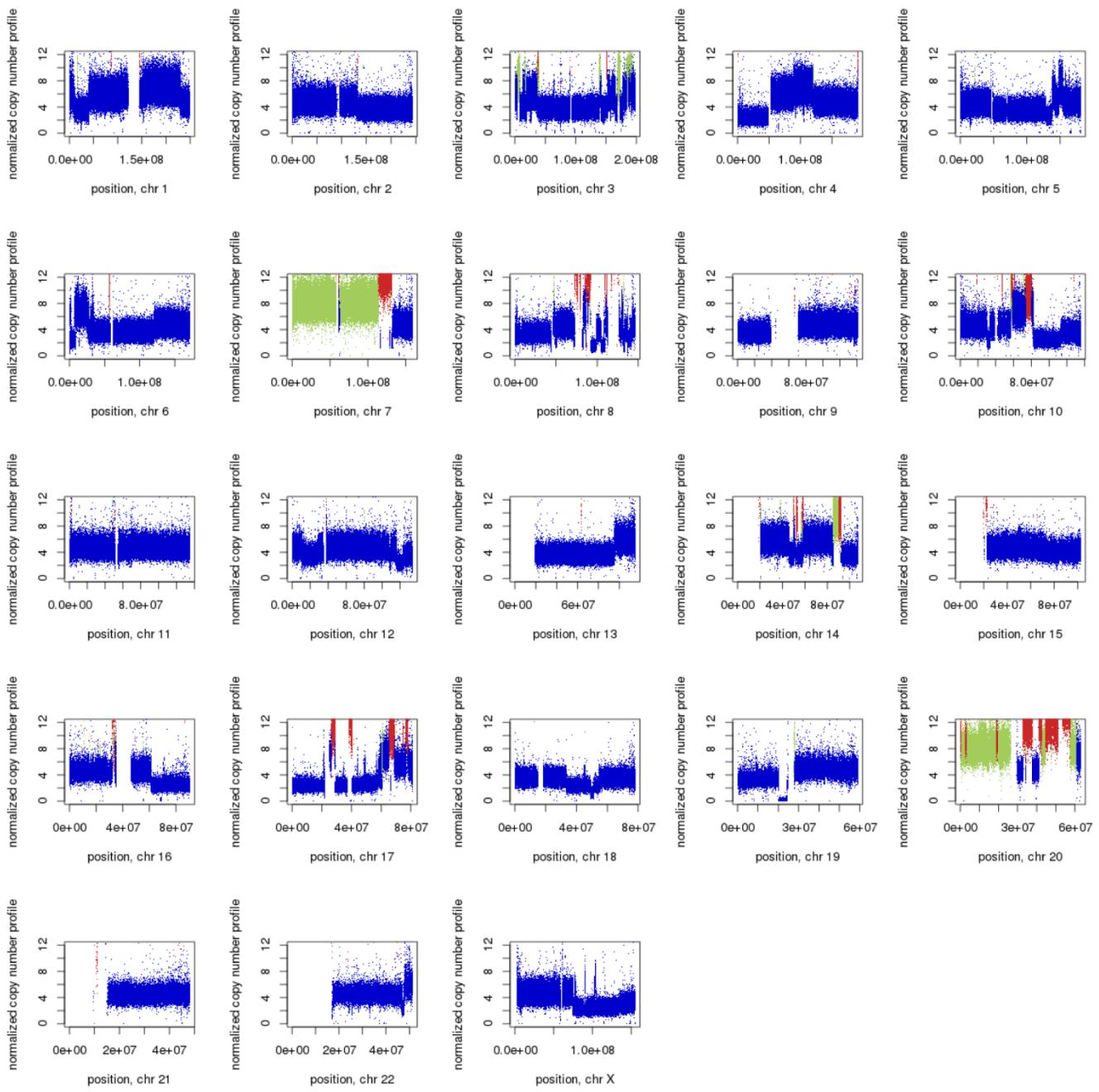


Fig. S4C. CNVs identified in SK-BR-3 breast cancer cell line (used in nuc-seq) by Control-FREEC.

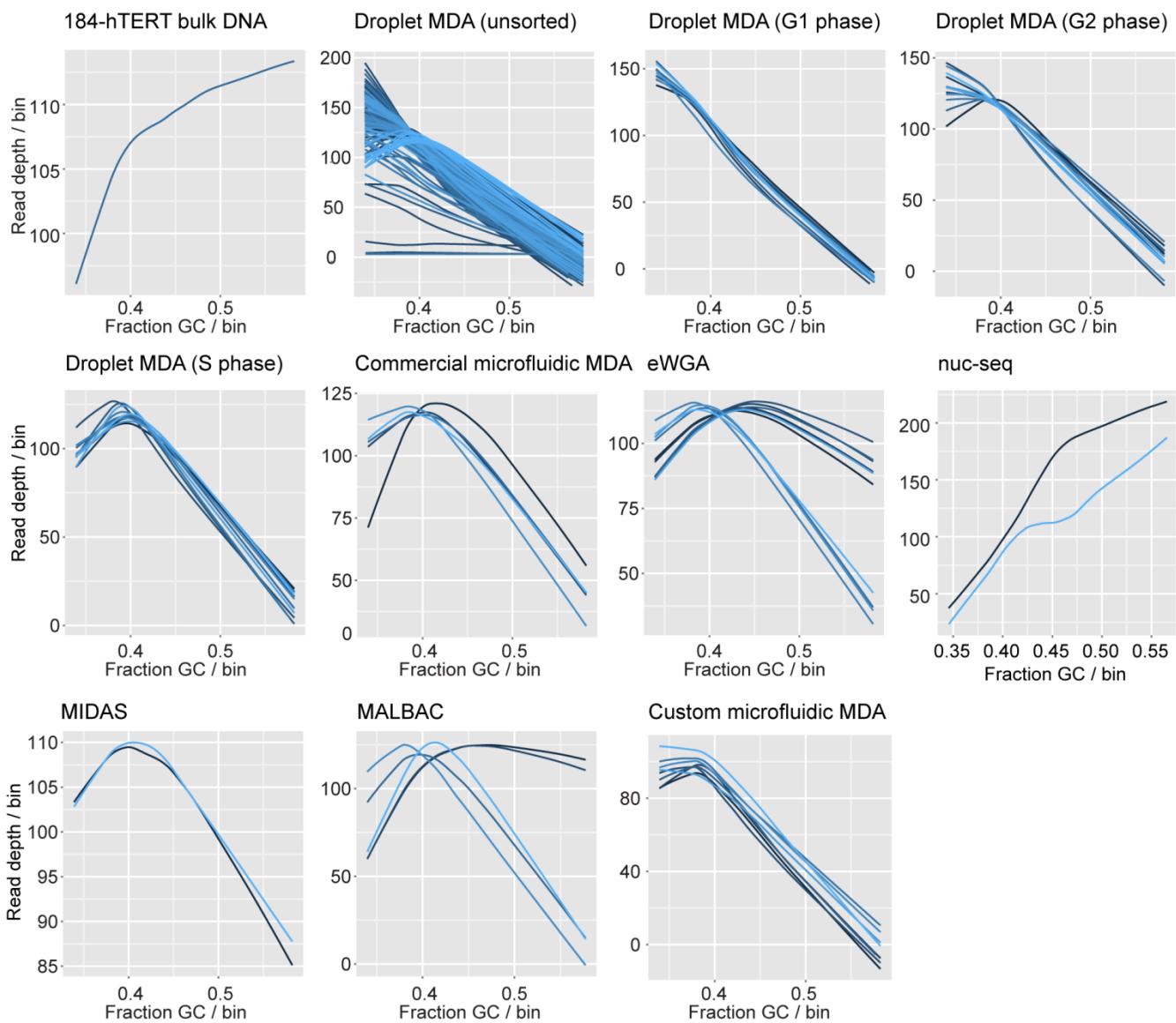


Fig S5. LOWESS regressions of read depth per bin vs. GC-content per bin for each sample group prior to correction for GC-content biases. One curve is generated for each sample. All droplet MDA samples were performed on single 184-hTERT cells.

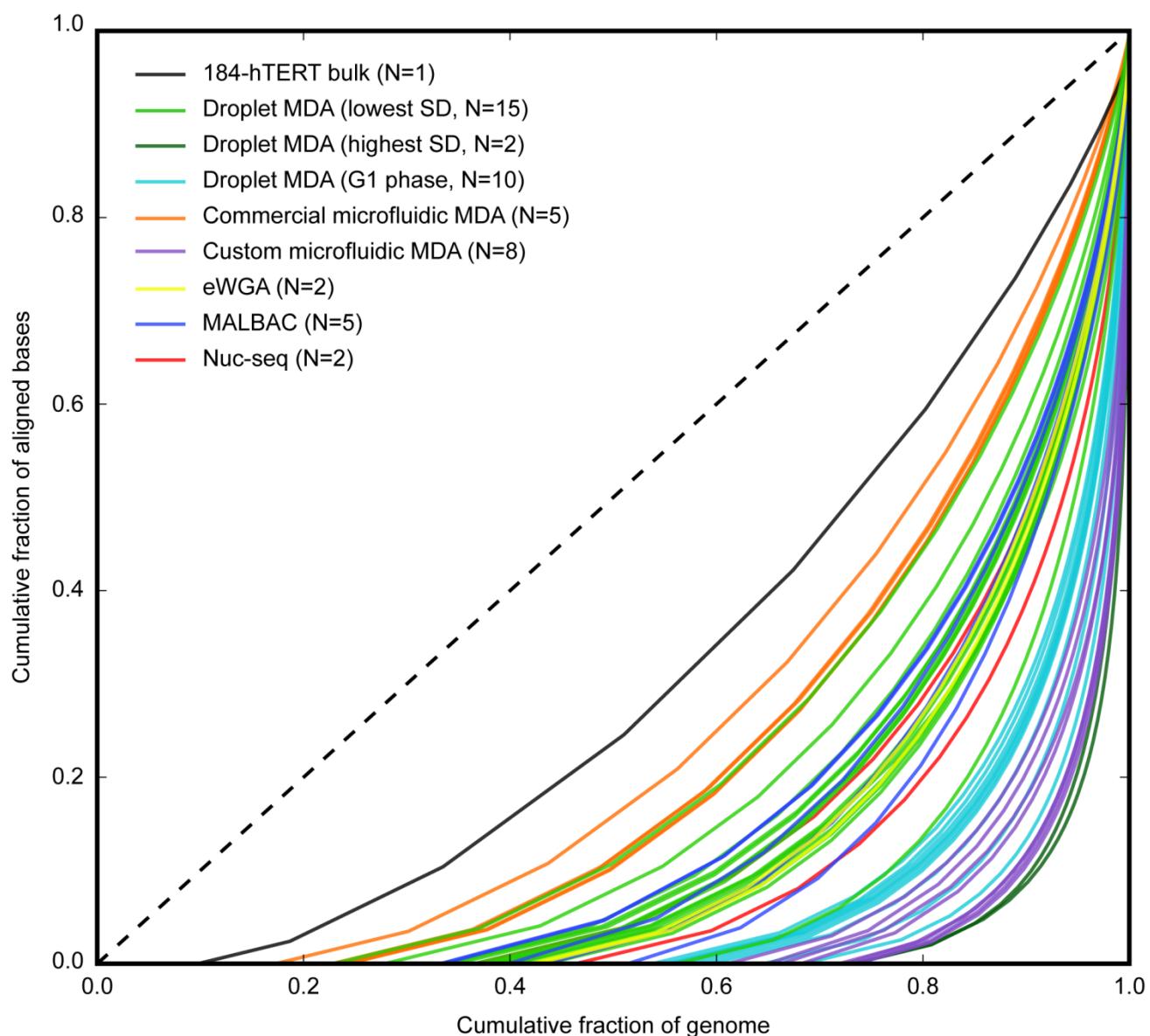


Fig. S6. Lorenz curves depicting uniformity of coverage for all samples analyzed from each amplification method.

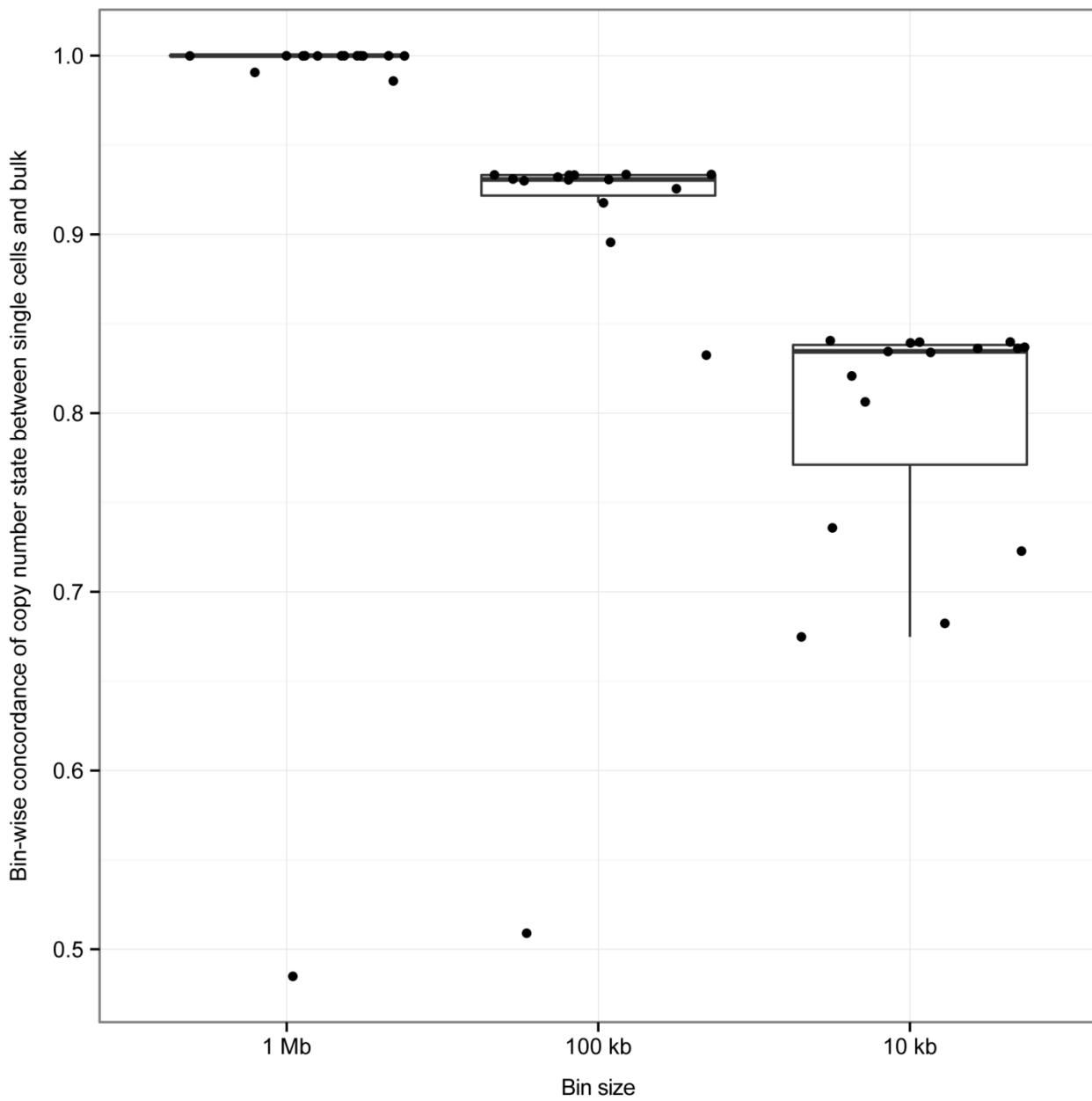


Fig. S7. Scatter and box plots of bin-wise concordance of copy number states between 15 184-hTERT single-cell droplet MDA samples sequenced to high depth and bulk DNA for different bin sizes. For 1 Mb, 100 kb, and 10 kb bin sizes, the median concordance rate is 100% (of 2,487 bins), 93.1% (of 24,162 bins), and 83.4% (of 227,172 bins) respectively. Upper whisker ranges from 75th percentile to sample with largest value within 1.5x interquartile range above 75th percentile, and lower whisker ranges from 25th percentile to sample with smallest value within 1.5x interquartile range below 25th percentile.

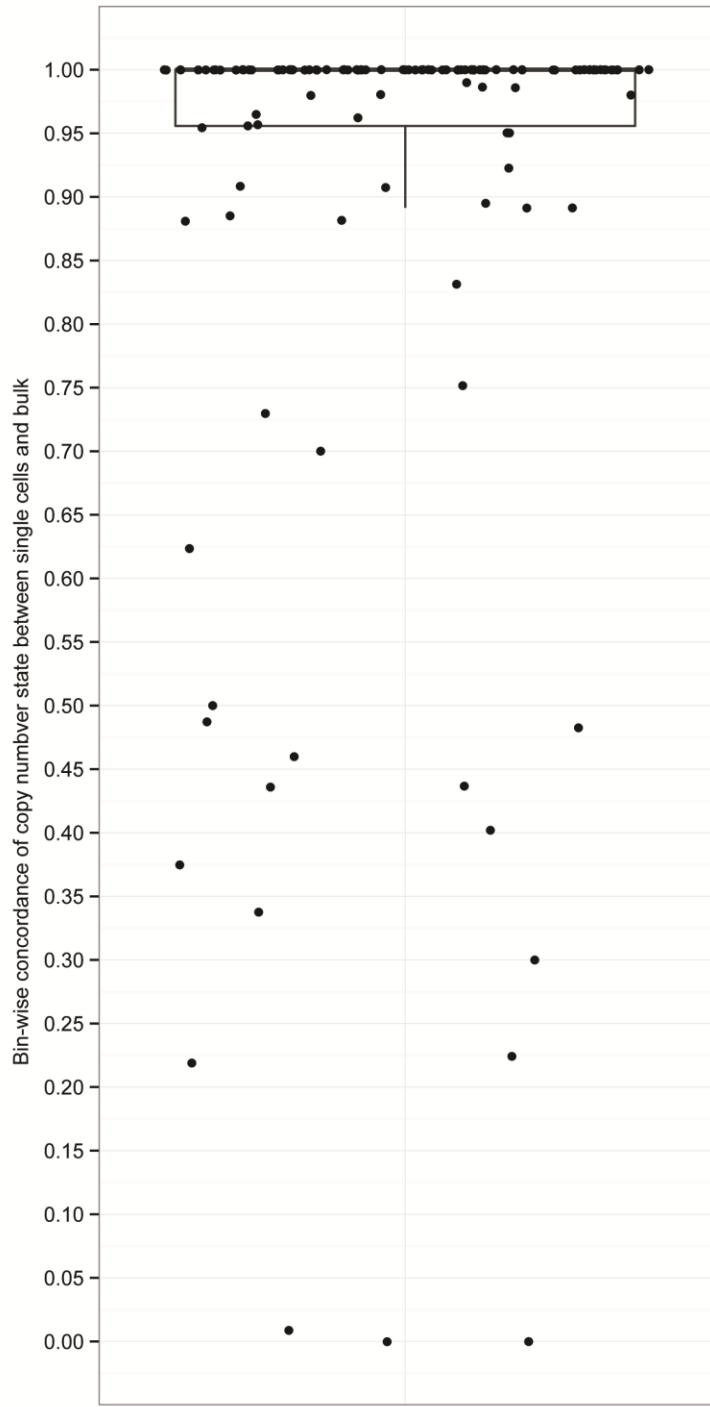


Fig. S8. Scatter and box plots of bin-wise concordance of copy number states between each of the 129 184-hTERT single-cell droplet MDA samples and bulk DNA using 1-Mb bins. The median concordance rate is 100% (of 227,172 bins). Upper whisker ranges from 75th percentile to sample with largest value within $1.5 \times$ interquartile range above 75th percentile, and lower whisker ranges from 25th percentile to sample with smallest value within $1.5 \times$ interquartile range below 25th percentile.

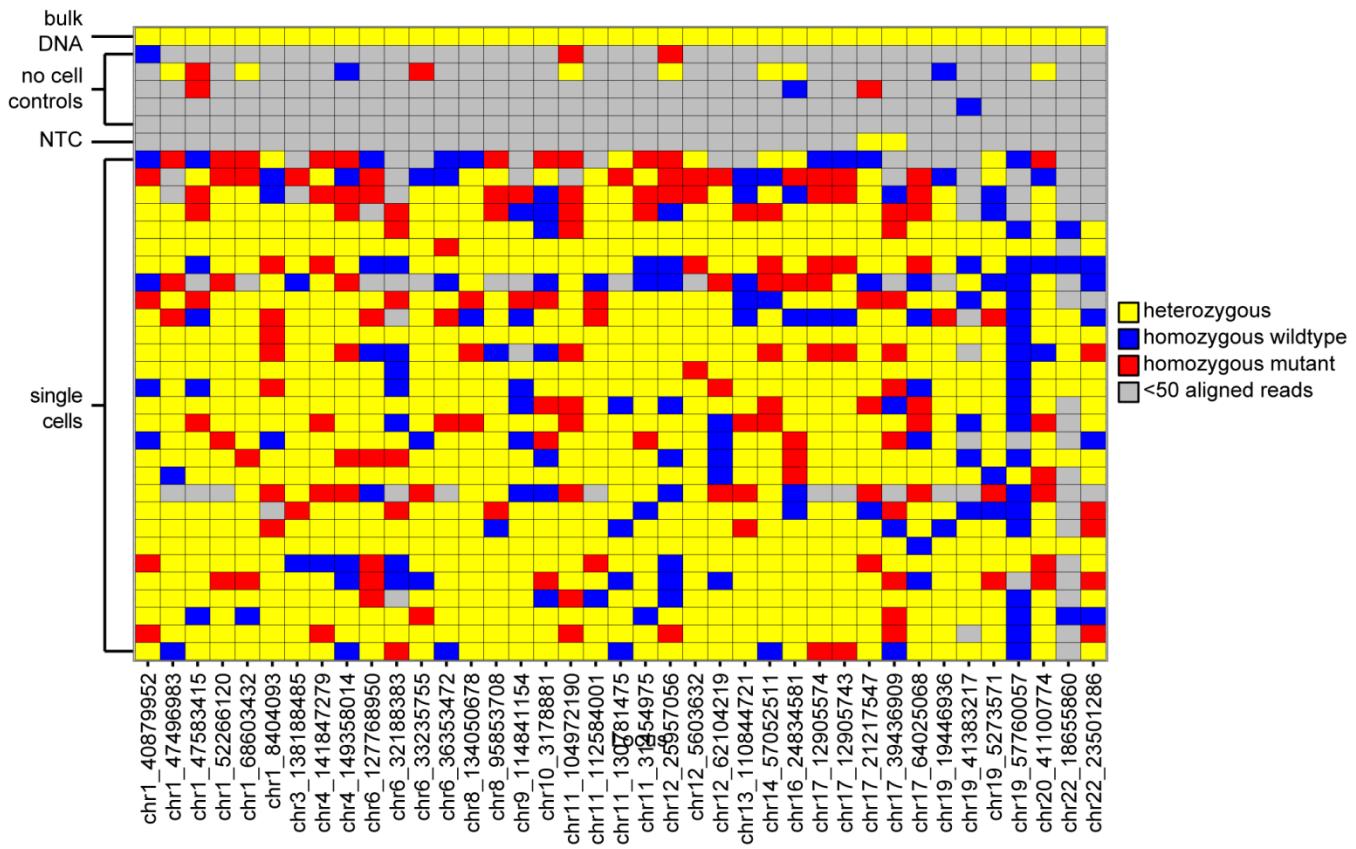


Fig. S9. Zygosity at 39 heterozygous loci determined by targeted sequencing of 184-hTERT single-cell droplet MDA samples. Shown are bulk 184-hTERT gDNA, no-cell control reactions (droplet MDA performed on cell suspension fluid but no cells), NTC (no-template control - water as template for PCR reaction), and 184-hTERT single-cell droplet MDA samples.

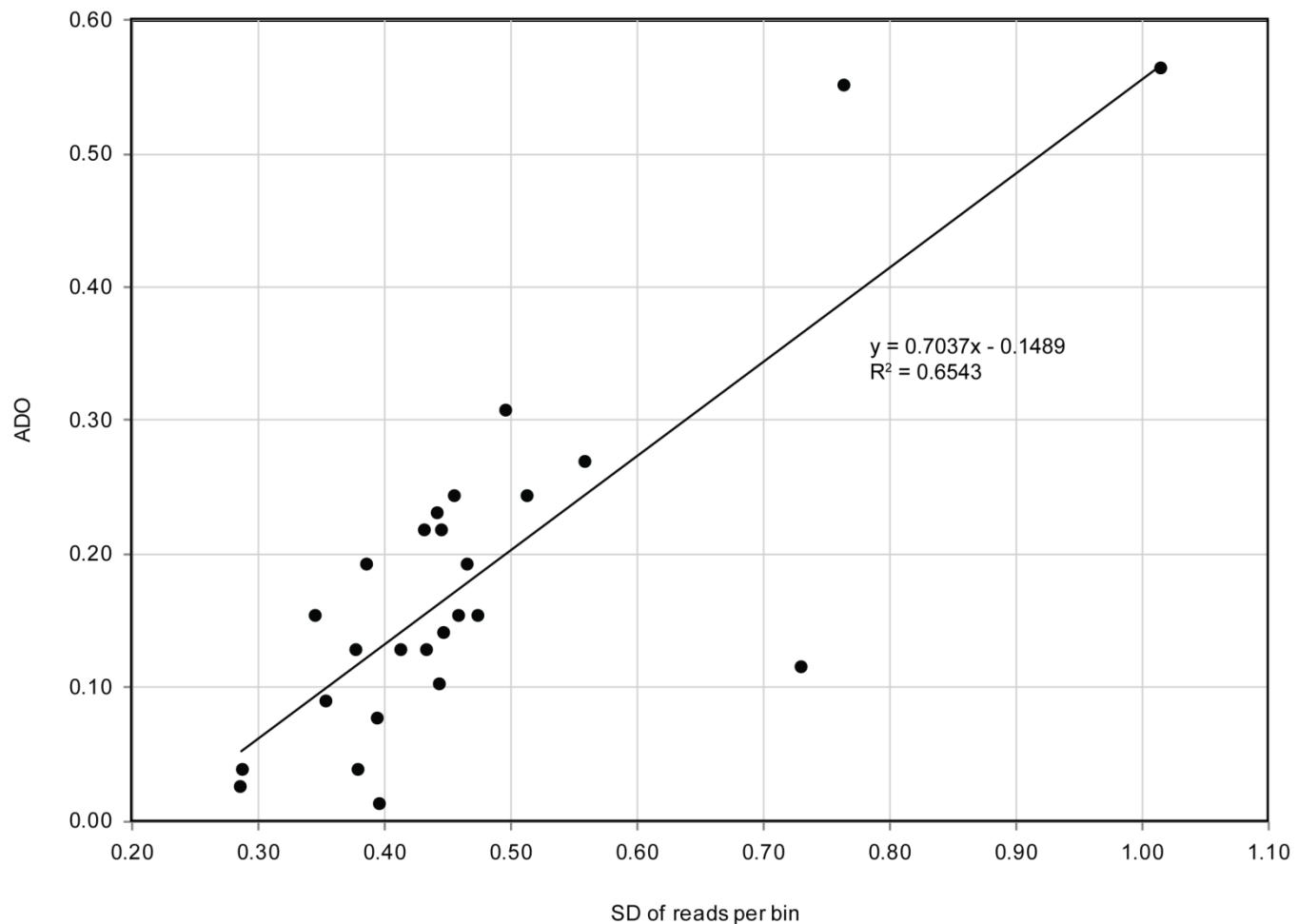


Fig. S10. Correlation between SD of reads per 1-Mb bin and allelic dropout (ADO) in 29 184-hTERT single-cell droplet MDA samples. Three outliers are not shown.

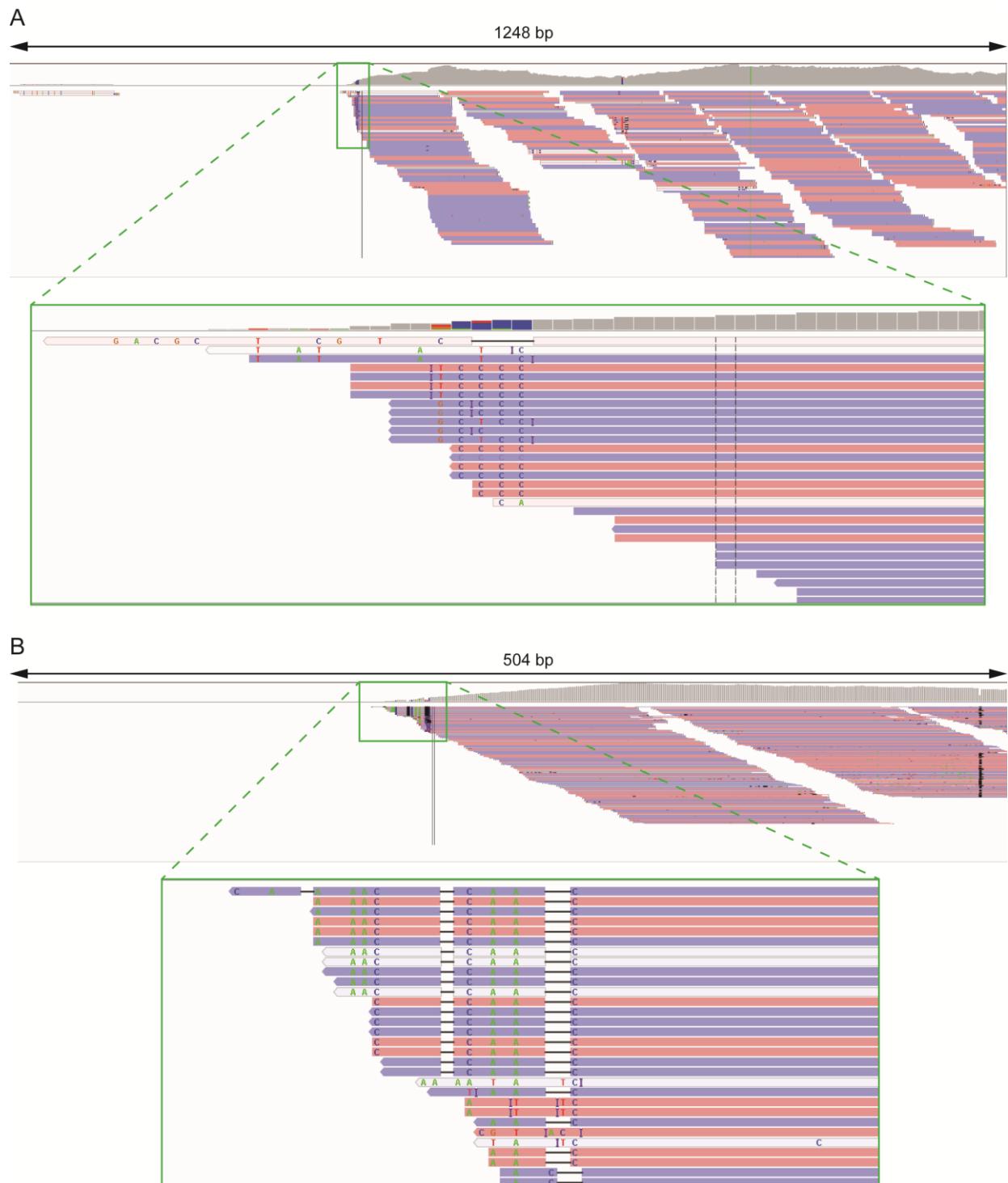


Fig. S11. Two representative IGV plots of the deeply sequenced G1-phase single-cell droplet MDA sample showing SNV false positives clustered at the ends of read islands. Mismatches between the single cell and the reference are indicated by displaying the differing base on the read (colours denote different nucleotides). All mismatches shown are not mismatches between the bulk sample and reference and are thus called as false positives. Reads are coloured by pair status where the first and second reads in a pair are shown in blue and red respectively. White reads indicate a mapping quality close to 0.

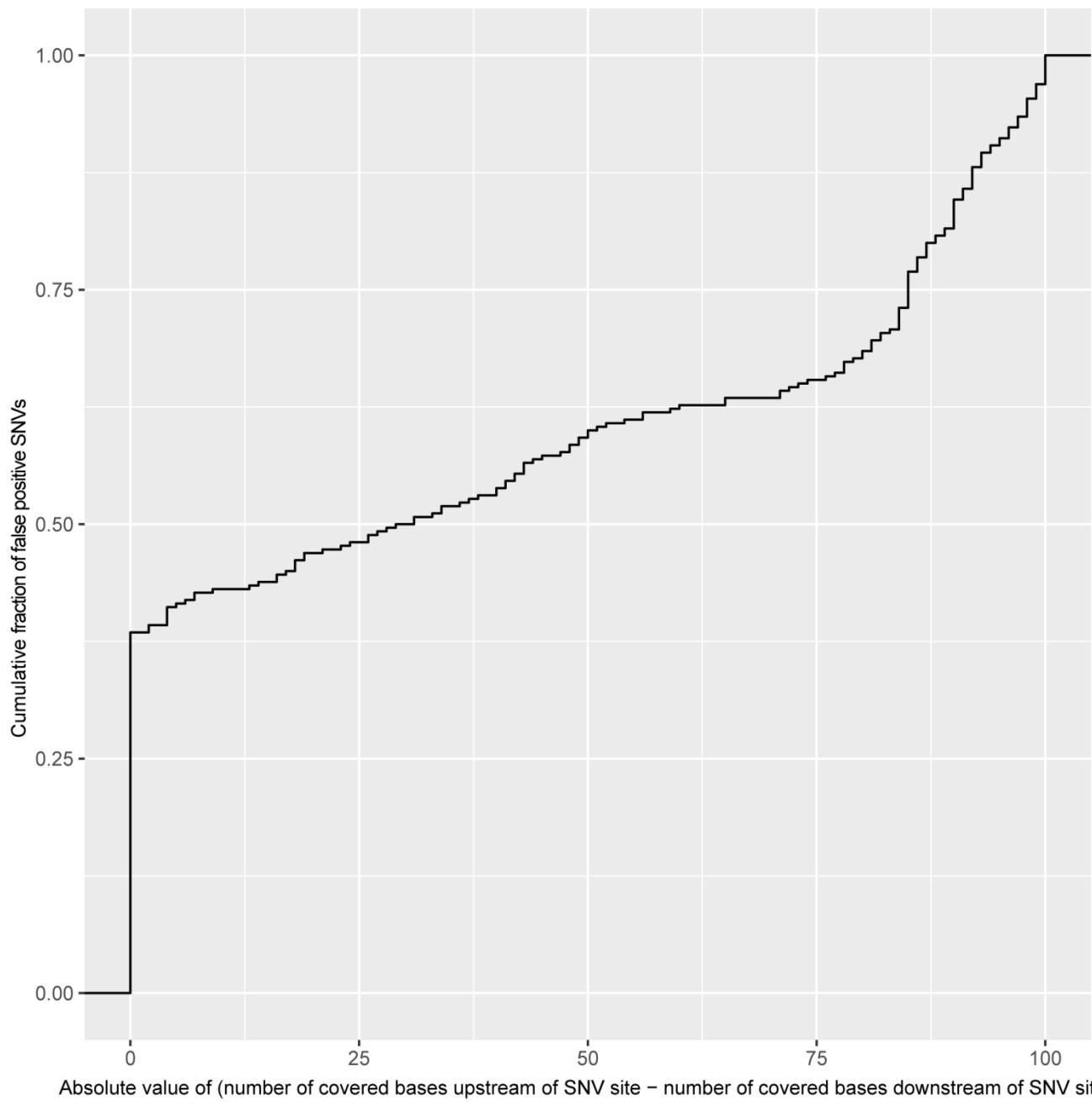


Fig. S12. Proximity of SNV false positives to ends of read islands for SNV false positives identified using a mutationSeq SNV-calling probability threshold of 0.9 and a minimum single-cell coverage depth of 4. For each SNV false positive site, the distance from the end of a read island was computed as described in SI Appendix. The plot shows that over 50% of false positive SNVs are within 75 bases of a read island end.

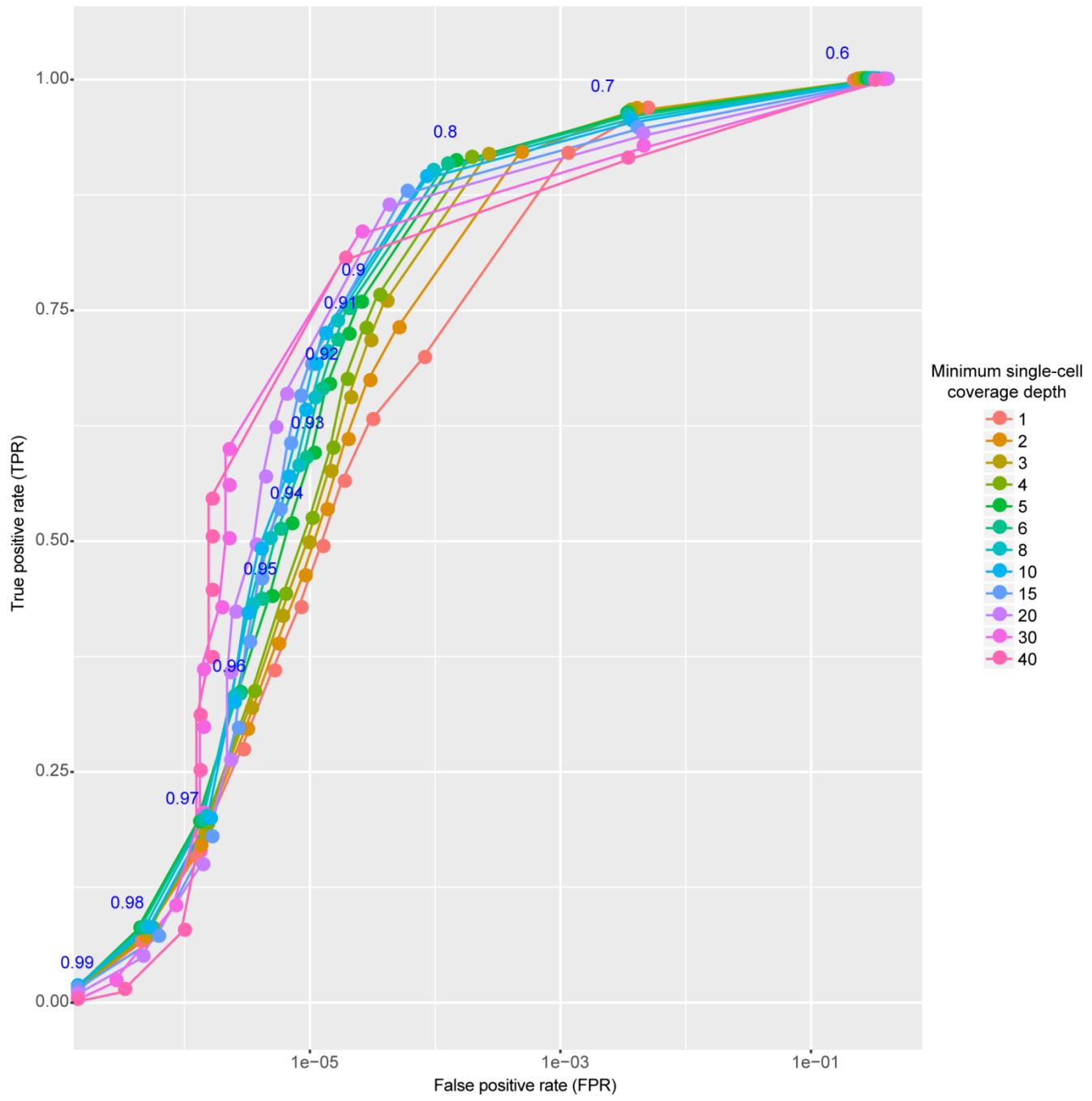


Fig. S13. Relationship between SNV true (TPR) and false (FPR) positive rates for different minimum single-cell coverage depth thresholds (indicated by line color) and mutationSeq SNV-calling probability thresholds (indicated by numbers in blue font). TPR and FPR display an inverse correlation.

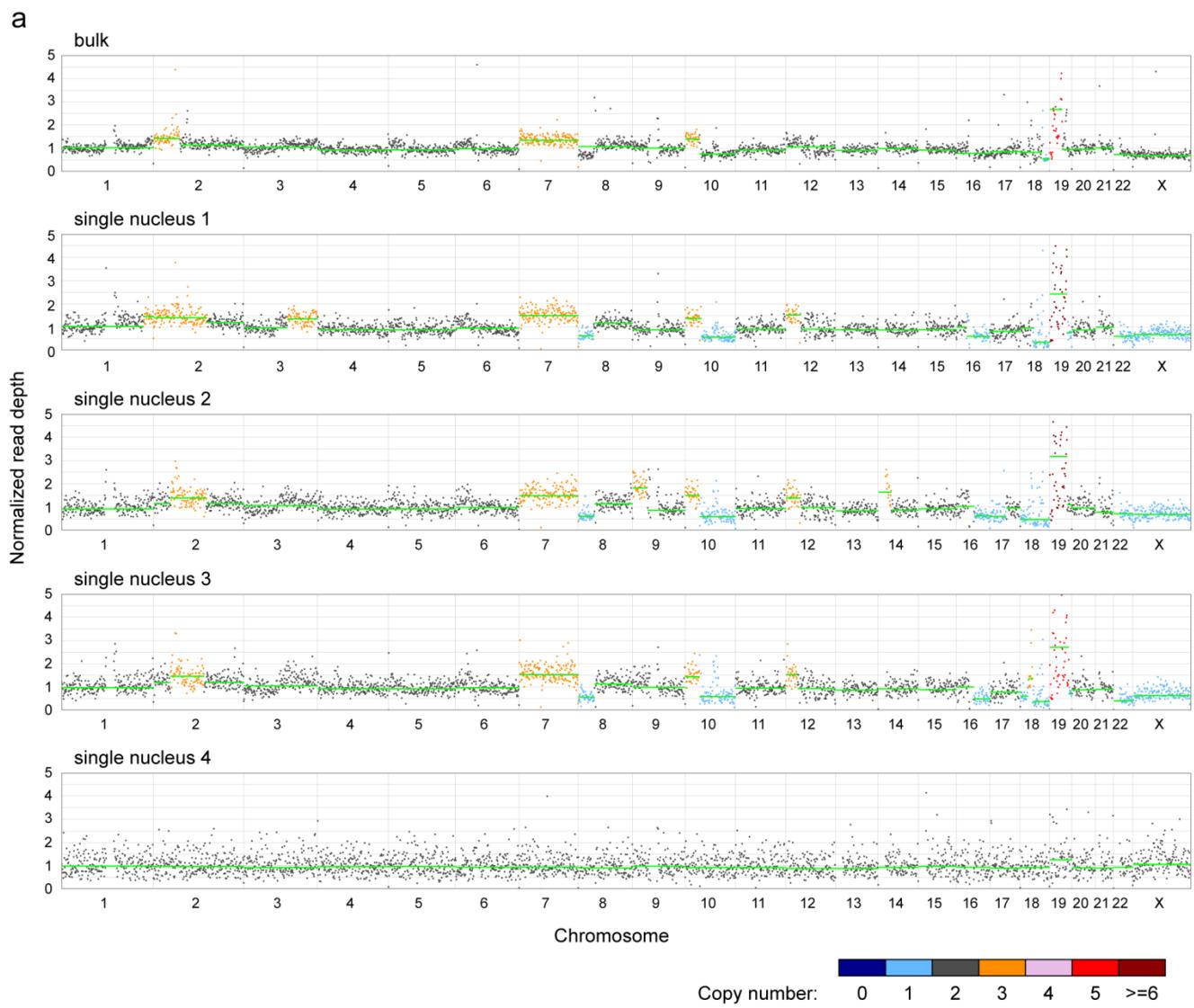


Fig. S14A. Normalized read depth using 1-Mb bins for bulk DNA and 4 single-nuclei samples from a high-grade serous ovarian cancer specimen 1. Horizontal green lines indicate segments of contiguous bins inferred to have the same copy number, where the read depth of each segment is equal to the median read depth of the bins in that segment. Inferred copy numbers are indicated by the color of the datapoint for each bin (see legend). Read depth profiles from nuclei 1-3 closely match that of bulk DNA. Nucleus 4 appears to be normal diploid.

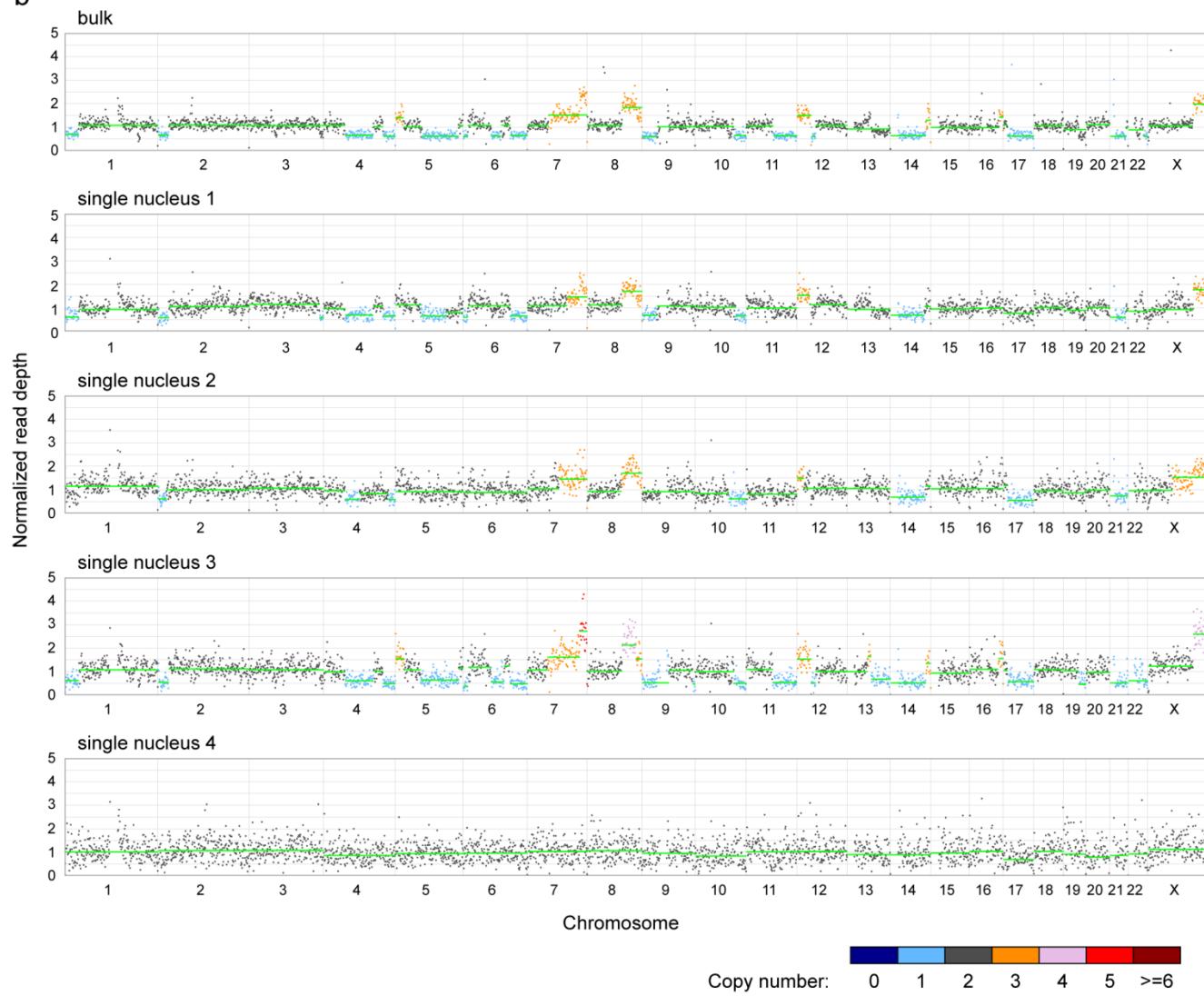
b

Fig. S14B. Normalized read depth using 1-Mb bins for bulk DNA and 4 single-nuclei samples from a high-grade serous ovarian cancer specimen 2. Horizontal green lines indicate segments of contiguous bins inferred to have the same copy number, where the read depth of each segment is equal to the median read depth of the bins in that segment. Inferred copy numbers are indicated by the color of the datapoint for each bin (see legend). Read depth profiles from nuclei 1-3 closely match that of bulk DNA. Nucleus 4 appears to be normal diploid.

Supplementary Tables

Table S1. Yields (mean \pm SD) of different droplet MDA reaction times performed on single 184-hTERT cells

MDA reaction time	Yield (ng)
80 min	5.21 \pm 2.88 (n=4)
160 min	44.35 \pm 10.19 (n=4)
320 min	69 \pm 5.7 (n=5)
20 hrs	68.16 \pm 3.57 (n=5)

Table S2. Chromosomes considered in each cell type for analysis of bias.

Cell type	Chromosomes considered
184-hTERT (droplet MDA)	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,21,22,X
SW480 (MALBAC)	1,4,10,15,16,22,X
SK-BR-3 (nuc-seq)	9,11,15,21
Sperm (microfluidic MDA)	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22
Neuron (MIDAS)	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,X
Human umbilical vein endothelial cell (eWGA)	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22
GM12752 (commercial microfluidic MDA)	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22

Table S3: Alignment metrics for all samples included in comparison of amplification bias.

Alignment metrics were computed after downsampling all samples to the same approximate number of total sequenced bases.

Sample group	Sample	Number of reads	Number of bases	Percent duplicate reads	Percent mapped reads	Coverage breadth (%)
bulk 184-hTERT gDNA	S47	311696	26332294	0.111968071	97.97976233	0.7368
no cell / nucleus droplet MDA	S24	311632	26324262	6.639882939	96.93901782	0.0359
	G4_S6	360200	26231462	23.86729595	97.02054414	0.0825
	G5_S7	368958	26337526	37.01695044	97.31134709	0.0765
	run10_11_S19	310802	26254854	0.17663979	98.0949286	0.7569
Droplet MDA on single unsorted 184-hTERT cells	run10_11_S20	310740	26250936	0.148033726	98.00090107	0.7676
	run10_11_S21	310992	26270858	0.102896538	98.21281576	0.764
	run10_11_S22	311674	26329589	0.134114491	98.20806355	0.7653
	run10_11_S23	310762	26251138	0.176340737	97.90514928	0.7491
	run12_S10	311844	26339248	0.093957235	98.52522415	0.7504
	run12_S11	311390	26299992	0.143228749	98.30309258	0.4208
	run12_S12	311580	26315239	0.10494897	98.42223506	0.744
	run12_S13	311172	26281863	0.094802874	98.4214518	0.7548
	run12_S14	311878	26339842	0.120880601	98.363142	0.7421
	run12_S15	311226	26286641	0.094786425	98.44357477	0.7583
	run12_S16	312088	26359911	0.531260414	98.17359206	0.3204
	run12_S17	311586	26317081	0.080876548	98.32502102	0.7458
	run12_S18	310690	26242282	0.083040973	98.43187743	0.7487
	run12_S19	310866	26255905	0.092644419	98.50996893	0.7597
	run12_S1	311466	26306294	0.131314493	97.28573905	0.4824
	run12_S20	310636	26239410	0.758443967	98.35627551	0.0846
	run12_S21	311126	26279404	0.094174065	98.49450062	0.76
	run12_S22	310872	26257089	0.119341723	98.45145269	0.7625
	run12_S23	312038	26355373	0.110883931	98.50531025	0.761
	run12_S24	311354	26299132	0.123653462	98.50363252	0.7594
	run12_S25	311134	26280875	0.101885361	98.52475139	0.7563
	run12_S26	311944	26348394	0.086233427	98.52633806	0.7678
	run12_S27	311310	26293899	0.097330635	98.47226238	0.7568
	run12_S28	312078	26360058	0.099013708	98.61348765	0.7657
	run12_S29	311996	26350502	0.081731817	98.57402018	0.7606
	run12_S2	311340	26297755	0.086721912	97.62735273	0.6479
	run12_S30	310956	26265582	0.069463204	98.49946616	0.7653
	run12_S31	312046	26354819	0.094537344	98.64122597	0.7695
	run12_S32	311276	26291155	0.093486167	98.52446061	0.7597
	run12_S33	312410	26388532	0.062097884	98.67609872	0.7647
	run12_S34	311766	26332390	0.098150536	98.50560998	0.7614
	run12_S35	311222	26286827	0.103141809	98.62863165	0.7596

run12_S36	311806	26337459	0.087875153	98.61484385	0.7644
run12_S37	310594	26233830	0.146493493	97.76492785	0.7343
run12_S38	311418	26303576	0.083167961	97.82960523	0.75
run12_S39	311348	26297852	0.086398499	97.81209451	0.7112
run12_S3	311048	26272038	0.195468224	97.99098531	0.4671
run12_S40	310688	26243044	0.077891647	97.79714698	0.7247
run12_S41	311298	26290050	0.103437863	97.74910215	0.7393
run12_S42	312408	26387503	0.090266574	97.97060255	0.7526
run12_S43	311906	26339446	0.100030137	97.38671266	0.7064
run12_S44	311896	26343110	0.086887937	97.9470721	0.7543
run12_S45	310848	26253533	0.101335701	97.91119776	0.7407
run12_S46	311054	26272825	0.113485118	97.81388441	0.7514
run12_S47	311066	26272806	0.089691577	97.91716227	0.7124
run12_S48	310958	26264741	0.096476051	97.92383537	0.7534
run12_S49	310850	26255142	0.091362393	98.00965096	0.7548
run12_S4	310802	26251475	0.132882028	98.18083539	0.6625
run12_S50	310920	26260694	0.101955487	97.77306059	0.7439
run12_S5	311274	26293197	0.124649023	98.08785829	0.6123
run12_S6	311598	26316612	0.112966065	98.0089731	0.5911
run12_S7	311450	26301852	0.084122652	97.88087976	0.5779
run12_S8	311220	26275580	0.077437183	98.37317653	0.7341
run12_S9	311144	26278443	0.809592986	98.17544288	0.076
run6_7_S10	311774	26323768	0.080827779	98.11465998	0.7007
run6_7_S11	311032	26261449	0.066874148	98.25677101	0.6874
run6_7_S12	311250	26273348	0.06875502	98.07036145	0.6741
run6_7_S13	311282	26283656	0.08513181	98.20259443	0.7027
run6_7_S14	310656	26234207	0.084659559	98.11624433	0.6334
run6_7_S15	311774	26328969	0.077620328	98.46298922	0.7183
run6_7_S16	311306	26292597	0.09315593	98.7266548	0.7245
run6_7_S17	312050	26355628	0.069219676	98.7226406	0.7321
run6_7_S18	311720	26325900	0.103618632	98.57981522	0.7205
run6_7_S19	311046	26270035	0.105129145	98.54362377	0.7183
run6_7_S20	310870	26258405	0.10132853	98.66182005	0.726
run6_7_S21	311970	26346214	0.063788185	98.49536814	0.7162
run6_7_S22	310580	26225153	0.08564621	98.59198918	0.6943
run6_7_S23	311390	26291859	0.072898937	98.57445647	0.7059
run6_7_S24	311834	26329342	0.077926076	98.58033441	0.7125
run6_7_S25	310916	26262206	0.075261485	98.71251399	0.7514
run6_7_S26	310802	26253259	0.085585035	98.48263525	0.7345
run6_7_S27	311092	26277977	0.10704229	98.65827472	0.7507
run6_7_S28	311590	26319163	0.092429154	98.62511634	0.746
run6_7_S29	310778	26251768	0.092027106	98.61798454	0.7513
run6_7_S30	311108	26279852	0.084215128	98.66734382	0.7537

	run6_7_S31	310658	26240408	0.083693322	98.51219025	0.7429
	run6_7_S32	311072	26277875	0.076831087	98.71315965	0.7519
	run6_7_S33	310856	26259406	0.09136063	98.77210027	0.753
	run6_7_S34	311528	26314314	0.138671323	97.75782594	0.4582
	run6_7_S35	311140	26277034	0.064279745	98.06132288	0.6272
	run6_7_S36	312022	26354477	0.093583145	97.95879778	0.6235
	run6_7_S37	311412	26305971	0.083490681	98.6069901	0.7473
	run6_7_S38	311592	26320383	0.069000488	98.61292973	0.7559
	run6_7_S39	310902	26263034	0.081376125	98.54970377	0.7532
	run6_7_S40	311616	26322816	0.091779626	98.57549035	0.7563
	run6_7_S41	311354	26300904	0.098601592	98.75672065	0.747
	run6_7_S42	311978	26351824	0.109302579	98.58131022	0.7489
	run6_7_S43	311322	26298068	0.064884589	98.66729624	0.7548
	run6_7_S44	311536	26316076	0.070938832	98.72952083	0.7535
	run6_7_S45	311392	26303097	0.070650498	98.62038845	0.7471
	run6_7_S46	312062	26358726	0.092609802	98.74031442	0.7491
	run6_7_S7	311660	26321118	0.105563755	98.161458	0.71
	run6_7_S8	311082	26268869	0.101259475	98.03363743	0.6688
	run6_7_S9	311874	26335936	0.095551409	98.20568563	0.6812
Droplet MDA on single 184-hTERT nuclei in G1 phase	S59	312000	26356323	0.065064103	97.56987179	0.7384
	S60	311850	26341905	0.085297419	97.77392977	0.732
	S62	311402	26304229	0.083814491	96.63907104	0.6918
	S63	311554	26316930	0.082489713	97.71628674	0.7325
	S64	311930	26350540	0.101945949	96.92911871	0.7125
	S65	311114	26279703	0.067820799	97.49738038	0.7263
	S66	311366	26300966	0.137137645	96.78802438	0.648
	S67	311192	26287427	0.072623975	97.3855369	0.73
	S68	311374	26301972	0.072260369	97.45579271	0.742
	S69	311672	26327740	0.069945327	97.51469494	0.7389
Droplet MDA on single 184-hTERT nuclei in S phase	S70	312348	26387255	0.069793948	97.14517141	0.7469
	S71	311740	26335844	0.059985886	96.78353756	0.7415
	S72	311818	26341793	0.059008781	97.2400567	0.7493
	S74	311576	26319040	0.076385858	97.33130922	0.751
	S75	310980	26270329	0.110618046	97.08727249	0.7456
	S76	311194	26288317	0.062019191	97.17346735	0.7429
	S77	310412	26221270	0.065396956	97.08999652	0.7334
	S78	310770	26253784	0.089777006	96.8095376	0.7538
	S79	310704	26244404	0.079818734	96.90380555	0.736
	S80	310780	26253172	0.058562327	97.25111011	0.7568
Droplet MDA on single 184-hTERT nuclei in G2 phase	S81	312016	26357343	0.080765089	97.20847649	0.7452
	S82	311822	26338443	0.081136033	97.13618667	0.7428
	S83	311468	26310462	0.07095432	97.19168582	0.7545
	S84	311180	26285495	0.07326949	96.05598046	0.7215

	S85	311938	26350291	0.064115305	97.54758959	0.7528
	S86	310922	26264778	0.055962589	96.26851751	0.7485
	S87	310728	26246833	0.095260163	97.19593986	0.7469
	S88	311408	26305210	0.070646868	97.26949854	0.7529
	S89	311678	26330701	0.088873774	97.22887082	0.7494
	S90	311272	26292506	0.105374078	97.35536765	0.7537
	S91	311234	26288674	0.077755001	97.11888804	0.7513
	S92	311554	26317737	0.100785097	96.57170186	0.7518
	S93	311352	26297965	0.073871374	96.75158663	0.7221
	S94	311564	26317999	0.084733795	96.14043985	0.7305
Nuc-seq single cell samples	SRX151835	262658	26265800	0.20559054	85.34748608	0.7095
	SRX151836	263244	26324400	0.173223321	65.96693562	0.549
MIDAS single-cell samples	SRX316149	527284	26364200	37.09006911	87.69543548	0.6972
	SRX316150	525435	26271750	26.90703893	92.41028862	0.7503
Custom microfluidic MDA single cell samples	SRX151616	264738	26209062	18.96478783	94.16706329	0.4818
	SRX151727	264360	26171640	9.274852474	90.68278106	0.5194
	SRX151728	265322	26266878	16.19541538	94.7897272	0.5267
	SRX151729	265112	26246088	19.68752829	93.85957633	0.5264
	SRX151846	253172	26299372	20.18785648	80.14906862	0.3839
	SRX151850	254702	26300802	21.37910185	81.93418191	0.3942
	SRX151852	251864	26233524	21.48778706	85.37464663	0.3935
	SRX151853	253884	26224004	16.08648044	81.77199036	0.4361
MALBAC single cell samples	SRX202787	291290	26216100	6.813484843	84.80311717	0.6429
	SRX202978	292504	26325360	6.779736345	85.26174001	0.6716
	SRX204744	262628	26262800	19.3939717	93.17094902	0.7043
	SRX204745	262786	26278600	21.81356693	91.96342271	0.6992
	SRX205035	260934	26354334	7.946453893	58.96855143	0.4565
eWGA single cell samples	SRS829072	261491	26410591	1.930085548	0.842847364	0.6885
	SRS829074	261538	26415338	0.906942777	0.860027988	0.7011
	SRS829075	259718	26231518	0.162483925	0.881644707	0.7069
	SRS829076	260255	26285755	0.544850243	0.870926591	0.703
	SRS829077	260704	26331104	0.18680189	0.910411808	0.7309
	SRS829078	260351	26295451	0.142115836	0.840891719	0.6689
	SRS829079	260028	26262828	0.2422816	0.859368991	0.6767
	SRS829080	260892	26350092	0.289008479	0.833360164	0.6524
	SRS829082	260402	26300602	0.95774994	0.849713136	0.6556
	SRS829073	515697	26300547	0.60267948	0.954380188	0.7606
Commercial microfluidic MDA single cell samples	SRR2135902	260265	26286765	0.2893205	73.07398229	0.6114
	SRR2135903	260098	26269898	0.305654023	73.9382848	0.6182
	SRR2135904	260731	26333831	0.28266681	74.95656443	0.6287
	SRR2135905	260245	26284745	0.273972603	67.28621107	0.5622
	SRR2135906	260297	26289997	0.32501335	75.47224901	0.6314

Table S4. Standard deviation (SD) in reads per 1 MB bin of each sample compared

The read depth plots (1-Mb bin size, mean of 100 reads/bin) in PDF format for each sample are available at <http://www.msl.ubc.ca/sites/default/files/pdata/kaston-leung/Leung-et-al-supplementary-dataset-1.zip>. Filenames correspond to sample names below. Droplet MDA samples using FACS-sorted 184-hTERT cells have the cell phase followed by the word “phase” in the sample name.

Sample group	Sample name	SD
Commercial microfluidic MDA	SRR2135904_dups_marked_HMMcopy_output	0.197029
Commercial microfluidic MDA	SRR2135905_dups_marked_HMMcopy_output	0.280707
Commercial microfluidic MDA	SRR2135906_dups_marked_HMMcopy_output	0.315723
Commercial microfluidic MDA	SRR2135902_dups_marked_HMMcopy_output	0.319212
Commercial microfluidic MDA	SRR2135903_dups_marked_HMMcopy_output	0.32958
eWGA	HU_eWGA02_dups_marked_HMMcopy_output	0.136329
eWGA	HU_eWGA01_dups_marked_HMMcopy_output	0.141469
eWGA	HU_eWGA03_dups_marked_HMMcopy_output	0.151841
eWGA	HU_eWGA04_dups_marked_HMMcopy_output	0.15881
eWGA	HU_eWGA10_dups_marked_HMMcopy_output	0.158951
eWGA	HU_eWGA05_dups_marked_HMMcopy_output	0.163036
eWGA	HU_eWGA07_dups_marked_HMMcopy_output	0.168824
eWGA	HU_eWGA09_dups_marked_HMMcopy_output	0.175103
eWGA	HU_eWGA08_dups_marked_HMMcopy_output	0.181379
eWGA	HU_eWGA06_dups_marked_HMMcopy_output	0.186964
droplet MDA	run6_7_S25_dups_marked_HMMcopy_output	0.199985
droplet MDA	run6_7_S27_dups_marked_HMMcopy_output	0.208113
droplet MDA	G2_phase_S83_dups_marked_HMMcopy_output	0.214665
droplet MDA	G2_phase_S88_dups_marked_HMMcopy_output	0.223435
droplet MDA	G2_phase_S92_dups_marked_HMMcopy_output	0.228672
droplet MDA	run6_7_S29_dups_marked_HMMcopy_output	0.239656
droplet MDA	run12_S31_dups_marked_HMMcopy_output	0.249223
droplet MDA	run6_7_S33_dups_marked_HMMcopy_output	0.249419
droplet MDA	run6_7_S20_dups_marked_HMMcopy_output	0.253358
droplet MDA	G2_phase_S91_dups_marked_HMMcopy_output	0.257745
droplet MDA	G2_phase_S89_dups_marked_HMMcopy_output	0.259306
droplet MDA	run12_S44_dups_marked_HMMcopy_output	0.263407
droplet MDA	run10_11_S20_dups_marked_HMMcopy_output	0.265229
droplet MDA	run6_7_S30_dups_marked_HMMcopy_output	0.267807
droplet MDA	run12_S30_dups_marked_HMMcopy_output	0.270473
droplet MDA	G2_phase_S84_dups_marked_HMMcopy_output	0.274831
droplet MDA	run6_7_S32_dups_marked_HMMcopy_output	0.276521
droplet MDA	run6_7_S17_dups_marked_HMMcopy_output	0.276877
droplet MDA	run10_11_S22_dups_marked_HMMcopy_output	0.279024
droplet MDA	G2_phase_S90_dups_marked_HMMcopy_output	0.280475

droplet MDA	run6_7_S40_dups_marked_HMMcopy_output	0.282981
droplet MDA	run6_7_S19_dups_marked_HMMcopy_output	0.283525
droplet MDA	run10_11_S21_dups_marked_HMMcopy_output	0.285037
droplet MDA	G2_phase_S94_dups_marked_HMMcopy_output	0.289073
droplet MDA	G2_phase_S82_dups_marked_HMMcopy_output	0.292599
droplet MDA	G2_phase_S86_dups_marked_HMMcopy_output	0.295147
droplet MDA	run12_S46_dups_marked_HMMcopy_output	0.295243
droplet MDA	run12_S28_dups_marked_HMMcopy_output	0.296028
droplet MDA	run6_7_S39_dups_marked_HMMcopy_output	0.296851
droplet MDA	run6_7_S31_dups_marked_HMMcopy_output	0.302624
droplet MDA	G2_phase_S85_dups_marked_HMMcopy_output	0.303269
droplet MDA	run6_7_S21_dups_marked_HMMcopy_output	0.307248
droplet MDA	S_phase_S74_dups_marked_HMMcopy_output	0.308415
droplet MDA	run12_S38_dups_marked_HMMcopy_output	0.308552
droplet MDA	run6_7_S23_dups_marked_HMMcopy_output	0.311338
droplet MDA	run10_11_S19_dups_marked_HMMcopy_output	0.311761
droplet MDA	G2_phase_S87_dups_marked_HMMcopy_output	0.311856
droplet MDA	run6_7_S16_dups_marked_HMMcopy_output	0.312961
droplet MDA	run6_7_S44_dups_marked_HMMcopy_output	0.313487
droplet MDA	run6_7_S43_dups_marked_HMMcopy_output	0.31426
droplet MDA	run6_7_S38_dups_marked_HMMcopy_output	0.315707
droplet MDA	run12_S48_dups_marked_HMMcopy_output	0.318421
droplet MDA	run6_7_S45_dups_marked_HMMcopy_output	0.320322
droplet MDA	S_phase_S72_dups_marked_HMMcopy_output	0.323834
droplet MDA	run6_7_S46_dups_marked_HMMcopy_output	0.330495
droplet MDA	run12_S49_dups_marked_HMMcopy_output	0.330521
droplet MDA	G1_phase_S69_dups_marked_HMMcopy_output	0.337603
droplet MDA	run6_7_S42_dups_marked_HMMcopy_output	0.337797
droplet MDA	run12_S34_dups_marked_HMMcopy_output	0.338873
droplet MDA	S_phase_S81_dups_marked_HMMcopy_output	0.339155
droplet MDA	run12_S35_dups_marked_HMMcopy_output	0.339442
droplet MDA	run6_7_S18_dups_marked_HMMcopy_output	0.34024
droplet MDA	S_phase_S80_dups_marked_HMMcopy_output	0.340468
droplet MDA	G2_phase_S93_dups_marked_HMMcopy_output	0.343496
droplet MDA	run12_S36_dups_marked_HMMcopy_output	0.344307
droplet MDA	S_phase_S76_dups_marked_HMMcopy_output	0.345312
droplet MDA	run6_7_S41_dups_marked_HMMcopy_output	0.346408
droplet MDA	S_phase_S79_dups_marked_HMMcopy_output	0.348022
droplet MDA	S_phase_S75_dups_marked_HMMcopy_output	0.348739
droplet MDA	run12_S29_dups_marked_HMMcopy_output	0.354658
droplet MDA	run6_7_S15_dups_marked_HMMcopy_output	0.362664
droplet MDA	run12_S32_dups_marked_HMMcopy_output	0.363184
droplet MDA	run12_S42_dups_marked_HMMcopy_output	0.365935

droplet MDA	run12_S15_dups_marked_HMMcopy_output	0.366407
droplet MDA	run12_S22_dups_marked_HMMcopy_output	0.366789
droplet MDA	run10_11_S23_dups_marked_HMMcopy_output	0.36866
droplet MDA	run12_S19_dups_marked_HMMcopy_output	0.372992
droplet MDA	S_phase_S78_dups_marked_HMMcopy_output	0.375068
droplet MDA	run12_S33_dups_marked_HMMcopy_output	0.378104
droplet MDA	G1_phase_S59_dups_marked_HMMcopy_output	0.381883
droplet MDA	run6_7_S24_dups_marked_HMMcopy_output	0.382281
droplet MDA	G1_phase_S63_dups_marked_HMMcopy_output	0.383074
droplet MDA	G1_phase_S68_dups_marked_HMMcopy_output	0.384254
droplet MDA	run6_7_S11_dups_marked_HMMcopy_output	0.387841
droplet MDA	run6_7_S26_dups_marked_HMMcopy_output	0.388665
droplet MDA	run12_S26_dups_marked_HMMcopy_output	0.392207
droplet MDA	S_phase_S77_dups_marked_HMMcopy_output	0.3957
droplet MDA	run12_S50_dups_marked_HMMcopy_output	0.395716
droplet MDA	G1_phase_S65_dups_marked_HMMcopy_output	0.397334
droplet MDA	S_phase_S70_dups_marked_HMMcopy_output	0.406541
droplet MDA	S_phase_S71_dups_marked_HMMcopy_output	0.410918
droplet MDA	G1_phase_S64_dups_marked_HMMcopy_output	0.411967
droplet MDA	run6_7_S7_dups_marked_HMMcopy_output	0.414015
droplet MDA	run12_S37_dups_marked_HMMcopy_output	0.414218
droplet MDA	run12_S21_dups_marked_HMMcopy_output	0.426286
droplet MDA	G1_phase_S60_dups_marked_HMMcopy_output	0.427752
droplet MDA	run12_S45_dups_marked_HMMcopy_output	0.428024
droplet MDA	run12_S13_dups_marked_HMMcopy_output	0.429465
droplet MDA	run6_7_S10_dups_marked_HMMcopy_output	0.439764
droplet MDA	run12_S24_dups_marked_HMMcopy_output	0.443019
droplet MDA	run12_S23_dups_marked_HMMcopy_output	0.446136
droplet MDA	G1_phase_S67_dups_marked_HMMcopy_output	0.447831
droplet MDA	run12_S27_dups_marked_HMMcopy_output	0.46768
droplet MDA	run6_7_S13_dups_marked_HMMcopy_output	0.476721
droplet MDA	run12_S41_dups_marked_HMMcopy_output	0.495628
droplet MDA	G1_phase_S62_dups_marked_HMMcopy_output	0.50873
droplet MDA	run12_S8_dups_marked_HMMcopy_output	0.509551
droplet MDA	run6_7_S37_dups_marked_HMMcopy_output	0.515212
droplet MDA	run12_S25_dups_marked_HMMcopy_output	0.533379
droplet MDA	run12_S12_dups_marked_HMMcopy_output	0.568361
droplet MDA	run12_S10_dups_marked_HMMcopy_output	0.578369
droplet MDA	run6_7_S12_dups_marked_HMMcopy_output	0.59036
droplet MDA	run12_S14_dups_marked_HMMcopy_output	0.595722
droplet MDA	run12_S40_dups_marked_HMMcopy_output	0.610564
droplet MDA	run12_S18_dups_marked_HMMcopy_output	0.611944
droplet MDA	run6_7_S28_dups_marked_HMMcopy_output	0.61367

droplet MDA		run6_7_S9_dups_marked_HMMcopy_output	0.62569
droplet MDA		run12_S17_dups_marked_HMMcopy_output	0.626448
droplet MDA		run6_7_S8_dups_marked_HMMcopy_output	0.627933
droplet MDA		run6_7_S22_dups_marked_HMMcopy_output	0.647547
droplet MDA		run12_S39_dups_marked_HMMcopy_output	0.683476
droplet MDA		G1_phase_S66_dups_marked_HMMcopy_output	0.691754
droplet MDA		run12_S47_dups_marked_HMMcopy_output	0.713661
droplet MDA		run12_S43_dups_marked_HMMcopy_output	0.737676
droplet MDA		run12_S2_dups_marked_HMMcopy_output	0.979808
droplet MDA		run6_7_S14_dups_marked_HMMcopy_output	0.989635
droplet MDA		run12_S4_dups_marked_HMMcopy_output	1.023491
droplet MDA		run6_7_S36_dups_marked_HMMcopy_output	1.072972
droplet MDA		run6_7_S35_dups_marked_HMMcopy_output	1.112805
droplet MDA		run12_S6_dups_marked_HMMcopy_output	1.296385
droplet MDA		run12_S7_dups_marked_HMMcopy_output	1.308626
droplet MDA		run12_S5_dups_marked_HMMcopy_output	1.458919
droplet MDA		run6_7_S34_dups_marked_HMMcopy_output	2.616446
droplet MDA		run12_S1_dups_marked_HMMcopy_output	2.860532
droplet MDA		run12_S3_dups_marked_HMMcopy_output	5.014445
droplet MDA		run12_S11_dups_marked_HMMcopy_output	6.296878
184-hTERT genomic DNA	unamplified	bulk S47_dups_marked_HMMcopy_output	0.132559
MALBAC		SRX202978_dups_marked_HMMcopy_output	0.157532
MALBAC		SRX202787_dups_marked_HMMcopy_output	0.160337
MALBAC		SRX204744_dups_marked_HMMcopy_output	0.177799
MALBAC		SRX205035_dups_marked_HMMcopy_output	0.209644
MALBAC		SRX204745_dups_marked_HMMcopy_output	0.234382
nuc-seq		SRR504599_dups_marked_HMMcopy_output	0.452437
nuc-seq		SRR504607_dups_marked_HMMcopy_output	0.762318
custom microfluidic MDA		SRX151853_dups_marked_HMMcopy_output	0.800449
custom microfluidic MDA		SRX151727_dups_marked_HMMcopy_output	1.5626
custom microfluidic MDA		SRX151728_dups_marked_HMMcopy_output	1.632294
custom microfluidic MDA		SRX151846_dups_marked_HMMcopy_output	1.932129
custom microfluidic MDA		SRX151850_dups_marked_HMMcopy_output	2.174643
custom microfluidic MDA		SRX151852_dups_marked_HMMcopy_output	2.887877
custom microfluidic MDA		SRX151616_dups_marked_HMMcopy_output	3.195293
custom microfluidic MDA		SRX151729_dups_marked_HMMcopy_output	6.822982
MIDAS		SRX316150_1_dups_marked_HMMcopy_output	0.101893
MIDAS		SRX316149_1_dups_marked_HMMcopy_output	0.105317

Table S5. Size ranges of copy number segments found in two or more TOV2295 single cells and bulk DNA.

Bin size	Segments found in 2 or more samples including single cells and bulk DNA		Segments found in 2 or more single cells but not in bulk DNA	
	Number of segments	Size range	Number of segments	Size range
1 Mb	71	7 Mb - 157 Mb	37	7 Mb - 126 Mb
100 kb	56	1 Mb - 115 Mb	43	1 Mb – 58.6 Mb
10 kb	61	30 kb – 19.1 Mb	59	30 kb – 19.1 Mb

Table S6. Copy number segments identified in two or more TOV2295 single-cell and bulk DNA samples using a 10-kb bin size.

“State” values correspond to the following copy numbers:

state	copy number
1	0 copies
2	1 copy
3	2 copies
4	3 copies
5	4 copies
6	5 copies
7	≥ 6 copies

chromosome	start	end	state	size	Number of samples
1	144950001	145410000	7	4.60E+05	2
2	259790001	260630000	2	8.40E+05	3
2	260630001	260730000	1	1.00E+05	2
2	339530001	341120000	4	1.59E+06	2
2	339530001	341180000	5	1.65E+06	2
2	381550001	383010000	4	1.46E+06	2
2	445790001	446180000	4	3.90E+05	2
2	482980001	483410000	2	4.30E+05	2
3	492460001	493170000	2	7.10E+05	2
3	593820001	593910000	2	9.00E+04	2
3	617410001	617540000	5	1.30E+05	2
3	634890001	634990000	5	1.00E+05	2
3	666700001	666870000	3	1.70E+05	2
4	725270001	725320000	1	5.00E+04	4
4	805950001	806180000	2	2.30E+05	2
4	881030001	881170000	4	1.40E+05	3
4	881030001	881180000	4	1.50E+05	2
4	881170001	881650000	2	4.80E+05	4
4	881180001	881650000	1	4.70E+05	3
5	903110001	903220000	7	1.10E+05	2
5	1014350001	1016680000	2	2.33E+06	2
5	1039340001	1039400000	2	6.00E+04	2
6	1120180001	1120650000	2	4.70E+05	2
6	1120200001	1120650000	2	4.50E+05	2
6	1120650001	1121310000	4	6.60E+05	2
6	1166680001	1166920000	2	2.40E+05	2
7	1278050001	1278270000	4	2.20E+05	2
7	1291280001	1291600000	4	3.20E+05	2
7	1291320001	1291600000	4	2.80E+05	2

7	1314820001	1315270000	3	4.50E+05	2
8	1436190001	1439830000	2	3.64E+06	2
9	1576770001	1576900000	4	1.30E+05	2
9	1634920001	1635300000	2	3.80E+05	2
9	1654910001	1655060000	4	1.50E+05	2
9	1658200001	1658460000	2	2.60E+05	2
9	1680200001	1680420000	2	2.20E+05	2
10	1715820001	1.72E+09	3	1.18E+06	2
10	1719060001	1723230000	1	4.17E+06	2
10	1813930001	1814170000	4	2.40E+05	2
11	1867540001	1871040000	1	3.50E+06	3
11	1876550001	1876760000	2	2.10E+05	2
12	2042310001	2042590000	2	2.80E+05	2
14	2200000001	2219140000	2	1.91E+07	2
14	2219140001	2219320000	1	1.80E+05	2
16	2432390001	2432600000	5	2.10E+05	2
17	2500250001	2500400000	1	1.50E+05	2
17	2536440001	2536730000	4	2.90E+05	2
17	2564410001	2564790000	2	3.80E+05	2
17	2578750001	2578860000	7	1.10E+05	2
18	2600780001	2600840000	7	6.00E+04	2
18	2609360001	2609450000	2	9.00E+04	2
18	2651020001	2651050000	7	3.00E+04	2
20	2737770001	2737880000	5	1.10E+05	2
20	2738890001	2739100000	2	2.10E+05	2
20	2741800001	2741960000	4	1.60E+05	2
21	2781690001	2792410000	3	1.07E+07	3
21	2781690001	2792410000	2	1.07E+07	2
21	2792410001	2792550000	1	1.40E+05	3
22	2829820001	2846330000	2	1.65E+07	4
22	2846330001	2846900000	1	5.70E+05	2
X	2927630001	2927750000	4	1.20E+05	2

Table S7. SNV false and true positive rates for various mutationSeq parameters.

Table headings are defined as follows.

- mutationSeq SNV prob: mutationSeq SNV-calling probability threshold
- minimum single-cell coverage depth: Minimum single-cell coverage depth for which a reference-matching or non-reference-matching site was considered
- reference-matching sites: Number of dbSNP SNV sites at which the bulk sample had a coverage depth of at least 50 reads where all reads matched the reference
- SNVs at reference-matching sites: Number of reference-matching sites with specified minimum single-cell coverage depth at which mutationSeq has called an SNV (mismatch between single-cell and reference)
- Non-reference-matching sites: Number of dbSNP SNV sites at which the bulk sample had a coverage depth of at least 50 reads where all reads did not match the reference
- SNVs at non-reference-matching sites: Number of non-reference-matching sites with specified minimum single-cell coverage depth at which mutationSeq has called an SNV
- False positive rate: SNVs at reference-matching sites divided by reference-matching sites
- True positive rate: SNVs at non-reference-matching sites divided by non-reference-matching sites

mutationSeq SNV prob.	minimum single- cell coverage depth	reference- matching sites	SNVs at reference- matching sites	non- reference- matching sites	SNVs at non- reference- matching sites	false positive rate	true positive rate
0.6	1	9419807	1926375	118645	118134	2.04503E-01	0.99569
0.6	2	8732730	1914058	109728	109431	2.19182E-01	0.99729
0.6	3	8120725	1903354	101784	101557	2.34382E-01	0.99777
0.6	4	7690746	1896294	96247	96050	2.46568E-01	0.99795
0.6	5	7295308	1889028	91251	91063	2.58937E-01	0.99794
0.6	6	6976342	1886359	87266	87086	2.70394E-01	0.99794
0.6	8	6433972	1882315	80422	80252	2.92559E-01	0.99789
0.6	10	5996999	1880006	74938	74771	3.13491E-01	0.99777
0.6	15	5178555	1826657	64626	64465	3.52735E-01	0.99751
0.6	20	4590165	1737401	57350	57193	3.78505E-01	0.99726
0.6	30	3773656	1329513	47175	47020	3.52314E-01	0.99671
0.6	40	3217070	965606	40207	40053	3.00151E-01	0.99617
0.7	1	9419807	43754	118645	114591	4.64489E-03	0.96583
0.7	2	8732730	33032	109728	105935	3.78255E-03	0.96543
0.7	3	8120725	27644	101784	98095	3.40413E-03	0.96376
0.7	4	7690746	25351	96247	92617	3.29630E-03	0.96228
0.7	5	7295308	23096	91251	87647	3.16587E-03	0.96050
0.7	6	6976342	22371	87266	83674	3.20669E-03	0.95884
0.7	8	6433972	21260	80422	76848	3.30434E-03	0.95556

0.7	10	5996999	20622	74938	71374	3.43872E-03	0.95244
0.7	15	5178555	19841	64626	61072	3.83138E-03	0.94501
0.7	20	4590165	19467	57350	53800	4.24102E-03	0.93810
0.7	30	3773656	16166	47175	43627	4.28391E-03	0.92479
0.7	40	3217070	10370	40207	36661	3.22343E-03	0.91181
0.8	1	9419807	10095	118645	108795	1.07168E-03	0.91698
0.8	2	8732730	3998	109728	100713	4.57818E-04	0.91784
0.8	3	8120725	2027	101784	93205	2.49608E-04	0.91571
0.8	4	7690746	1406	96247	87839	1.82817E-04	0.91264
0.8	5	7295308	1000	91251	82941	1.37074E-04	0.90893
0.8	6	6976342	820	87266	79016	1.17540E-04	0.90546
0.8	8	6433972	580	80422	72254	9.01465E-05	0.89844
0.8	10	5996999	479	74938	66822	7.98733E-05	0.89170
0.8	15	5178555	290	64626	56600	5.60002E-05	0.87581
0.8	20	4590165	184	57350	49370	4.00857E-05	0.86085
0.8	30	3773656	92	47175	39238	2.43795E-05	0.83175
0.8	40	3217070	58	40207	32299	1.80288E-05	0.80332
0.9	1	9419807	724	118645	82543	7.68593E-05	0.69571
0.9	2	8732730	420	109728	79869	4.80949E-05	0.72788
0.9	3	8120725	313	101784	77010	3.85434E-05	0.75660
0.9	4	7690746	260	96247	73453	3.38069E-05	0.76317
0.9	5	7295308	176	91251	68951	2.41251E-05	0.75562
0.9	6	6976342	134	87266	65347	1.92078E-05	0.74883
0.9	8	6433972	100	80422	59133	1.55425E-05	0.73528
0.9	10	5996999	75	74938	54086	1.25063E-05	0.72174
0.9	15	5178555	50	64626	44469	9.65520E-06	0.68810
0.9	20	4590165	28	57350	37616	6.10000E-06	0.65590
0.9	30	3773656	8	47175	28119	2.11996E-06	0.59606
0.9	40	3217070	5	40207	21807	1.55421E-06	0.54237
0.91	1	9419807	280	118645	74569	2.97246E-05	0.62851
0.91	2	8732730	246	109728	73599	2.81699E-05	0.67074
0.91	3	8120725	233	101784	72664	2.86920E-05	0.71390
0.91	4	7690746	202	96247	70012	2.62653E-05	0.72742
0.91	5	7295308	140	91251	65790	1.91904E-05	0.72098
0.91	6	6976342	109	87266	62365	1.56242E-05	0.71465
0.91	8	6433972	83	80422	56472	1.29003E-05	0.70220
0.91	10	5996999	63	74938	51598	1.05053E-05	0.68854
0.91	15	5178555	41	64626	42263	7.91727E-06	0.65396
0.91	20	4590165	23	57350	35554	5.01071E-06	0.61995
0.91	30	3773656	8	47175	26279	2.11996E-06	0.55705
0.91	40	3217070	5	40207	20159	1.55421E-06	0.50138
0.92	1	9419807	166	118645	66627	1.76224E-05	0.56157
0.92	2	8732730	165	109728	66585	1.88944E-05	0.60682
0.92	3	8120725	161	101784	66386	1.98258E-05	0.65222
0.92	4	7690746	143	96247	64675	1.85938E-05	0.67197

0.92	5	7295308	98	91251	60834	1.34333E-05	0.66667
0.92	6	6976342	82	87266	57727	1.17540E-05	0.66151
0.92	8	6433972	66	80422	52420	1.02580E-05	0.65181
0.92	10	5996999	52	74938	47852	8.67100E-06	0.63855
0.92	15	5178555	34	64626	38939	6.56554E-06	0.60253
0.92	20	4590165	19	57350	32476	4.13928E-06	0.56628
0.92	30	3773656	8	47175	23557	2.11996E-06	0.49935
0.92	40	3217070	5	40207	17831	1.55421E-06	0.44348
0.93	1	9419807	112	118645	58252	1.18898E-05	0.49098
0.93	2	8732730	112	109728	58252	1.28253E-05	0.53088
0.93	3	8120725	112	101784	58251	1.37919E-05	0.57230
0.93	4	7690746	110	96247	57503	1.43029E-05	0.59745
0.93	5	7295308	74	91251	54026	1.01435E-05	0.59206
0.93	6	6976342	61	87266	51227	8.74384E-06	0.58702
0.93	8	6433972	49	80422	46511	7.61582E-06	0.57834
0.93	10	5996999	38	74938	42450	6.33650E-06	0.56647
0.93	15	5178555	28	64626	34318	5.40691E-06	0.53102
0.93	20	4590165	16	57350	28250	3.48571E-06	0.49259
0.93	30	3773656	7	47175	20030	1.85497E-06	0.42459
0.93	40	3217070	5	40207	14905	1.55421E-06	0.37071
0.94	1	9419807	75	118645	50393	7.96195E-06	0.42474
0.94	2	8732730	75	109728	50393	8.58838E-06	0.45925
0.94	3	8120725	75	101784	50393	9.23563E-06	0.49510
0.94	4	7690746	75	96247	50177	9.75198E-06	0.52134
0.94	5	7295308	49	91251	47052	6.71665E-06	0.51563
0.94	6	6976342	38	87266	44477	5.44698E-06	0.50967
0.94	8	6433972	29	80422	40186	4.50732E-06	0.49969
0.94	10	5996999	23	74938	36584	3.83525E-06	0.48819
0.94	15	5178555	20	64626	29471	3.86208E-06	0.45602
0.94	20	4590165	11	57350	24081	2.39643E-06	0.41990
0.94	30	3773656	5	47175	16858	1.32498E-06	0.35735
0.94	40	3217070	4	40207	12386	1.24337E-06	0.30806
0.95	1	9419807	46	118645	42279	4.88333E-06	0.35635
0.95	2	8732730	46	109728	42279	5.26754E-06	0.38531
0.95	3	8120725	46	101784	42279	5.66452E-06	0.41538
0.95	4	7690746	46	96247	42279	5.98121E-06	0.43928
0.95	5	7295308	34	91251	39857	4.66053E-06	0.43678
0.95	6	6976342	27	87266	37864	3.87022E-06	0.43389
0.95	8	6433972	21	80422	34434	3.26392E-06	0.42817
0.95	10	5996999	18	74938	31377	3.00150E-06	0.41871
0.95	15	5178555	16	64626	25035	3.08966E-06	0.38738
0.95	20	4590165	10	57350	20319	2.17857E-06	0.35430
0.95	30	3773656	5	47175	13920	1.32498E-06	0.29507
0.95	40	3217070	4	40207	9978	1.24337E-06	0.24817
0.96	1	9419807	26	118645	32133	2.76014E-06	0.27083

0.96	2	8732730	26	109728	32133	2.97730E-06	0.29284
0.96	3	8120725	26	101784	32133	3.20168E-06	0.31570
0.96	4	7690746	26	96247	32133	3.38069E-06	0.33386
0.96	5	7295308	19	91251	30372	2.60441E-06	0.33284
0.96	6	6976342	18	87266	28904	2.58015E-06	0.33122
0.96	8	6433972	15	80422	26432	2.33137E-06	0.32867
0.96	10	5996999	14	74938	24135	2.33450E-06	0.32207
0.96	15	5178555	13	64626	18997	2.51035E-06	0.29395
0.96	20	4590165	10	57350	14874	2.17857E-06	0.25935
0.96	30	3773656	5	47175	9542	1.32498E-06	0.20227
0.96	40	3217070	4	40207	6470	1.24337E-06	0.16092
0.97	1	9419807	11	118645	18341	1.16775E-06	0.15459
0.97	2	8732730	11	109728	18341	1.25963E-06	0.16715
0.97	3	8120725	11	101784	18341	1.35456E-06	0.18020
0.97	4	7690746	11	96247	18341	1.43029E-06	0.19056
0.97	5	7295308	9	91251	17589	1.23367E-06	0.19275
0.97	6	6976342	9	87266	16962	1.29007E-06	0.19437
0.97	8	6433972	9	80422	15978	1.39882E-06	0.19868
0.97	10	5996999	9	74938	14709	1.50075E-06	0.19628
0.97	15	5178555	8	64626	11412	1.54483E-06	0.17659
0.97	20	4590165	6	57350	8392	1.30714E-06	0.14633
0.97	30	3773656	3	47175	4803	7.94985E-07	0.10181
0.97	40	3217070	3	40207	3027	9.32526E-07	0.07529
0.98	1	9419807	4	118645	7466	4.24637E-07	0.06293
0.98	2	8732730	4	109728	7466	4.58047E-07	0.06804
0.98	3	8120725	4	101784	7466	4.92567E-07	0.07335
0.98	4	7690746	4	96247	7466	5.20106E-07	0.07757
0.98	5	7295308	3	91251	7097	4.11223E-07	0.07777
0.98	6	6976342	3	87266	6784	4.30025E-07	0.07774
0.98	8	6433972	3	80422	6329	4.66275E-07	0.07870
0.98	10	5996999	3	74938	5853	5.00250E-07	0.07810
0.98	15	5178555	3	64626	4445	5.79312E-07	0.06878
0.98	20	4590165	2	57350	2698	4.35714E-07	0.04704
0.98	30	3773656	1	47175	983	2.64995E-07	0.02084
0.98	40	3217070	1	40207	451	3.10842E-07	0.01122
0.99	1	9419807	0	118645	1223	0	0.01031
0.99	2	8732730	0	109728	1223	0	0.01115
0.99	3	8120725	0	101784	1223	0	0.01202
0.99	4	7690746	0	96247	1223	0	0.01271
0.99	5	7295308	0	91251	1221	0	0.01338
0.99	6	6976342	0	87266	1220	0	0.01398
0.99	8	6433972	0	80422	1217	0	0.01513
0.99	10	5996999	0	74938	1122	0	0.01497
0.99	15	5178555	0	64626	813	0	0.01258
0.99	20	4590165	0	57350	402	0	0.00701

0.99	30	3773656	0	47175	68	0	0.00144
0.99	40	3217070	0	40207	16	0	0.00040

Table S8. Copy number segments identified in two or more single-cell and bulk DNA samples from ovarian cancer specimen 1 using a 1-Mb bin size.

“State” values correspond to the following copy numbers:

state	copy number
1	0 copies
2	1 copy
3	2 copies
4	3 copies
5	4 copies
6	5 copies
7	≥ 6 copies

chromosome	start	end	state	size	Number of samples
1	1	250000000	3	2.50E+08	9
2	250000001	494000000	3	2.44E+08	2
3	494000001	693000000	3	1.99E+08	11
4	693000001	885000000	3	1.92E+08	16
5	885000001	1066000000	3	1.81E+08	16
6	1066000001	1238000000	3	1.72E+08	13
7	1238000001	1398000000	3	1.60E+08	2
7	1238000001	1398000000	4	1.60E+08	12
8	1398000001	1439000000	2	4.10E+07	2
8	1398000001	1441000000	2	4.30E+07	5
8	1398000001	1445000000	2	4.70E+07	2
8	1398000001	1545000000	3	1.47E+08	2
8	1439000001	1545000000	3	1.06E+08	2
8	1441000001	1545000000	3	1.04E+08	4
9	1545000001	1584000000	4	3.90E+07	2
9	1545000001	1687000000	3	1.42E+08	12
9	1584000001	1687000000	3	1.03E+08	2
10	1687000001	1725000000	4	3.80E+07	8
10	1687000001	1731000000	4	4.40E+07	2
10	1687000001	1823000000	3	1.36E+08	2
10	1725000001	1823000000	2	9.80E+07	7
10	1731000001	1823000000	2	9.20E+07	2
11	1823000001	1959000000	3	1.36E+08	14
12	1959000001	1993000000	4	3.40E+07	2
12	1959000001	1999000000	4	4.00E+07	3
12	1959000001	2093000000	2	1.34E+08	2
12	1959000001	2093000000	3	1.34E+08	4

12	1999000001	2093000000	3	9.40E+07	3
13	2093000001	2209000000	3	1.16E+08	12
14	2209000001	2317000000	3	1.08E+08	13
15	2317000001	2420000000	3	1.03E+08	15
16	2420000001	2447000000	3	2.70E+07	3
16	2420000001	2467000000	3	4.70E+07	4
16	2420000001	2468000000	3	4.80E+07	2
16	2420000001	2511000000	2	9.10E+07	3
16	2420000001	2511000000	3	9.10E+07	3
16	2447000001	2511000000	2	6.40E+07	3
16	2467000001	2511000000	2	4.40E+07	4
16	2468000001	2511000000	2	4.30E+07	2
17	2511000001	2552000000	2	4.10E+07	2
17	2511000001	2593000000	2	8.20E+07	3
17	2511000001	2593000000	3	8.20E+07	7
18	2593000001	2614000000	2	2.10E+07	2
18	2593000001	2672000000	2	7.90E+07	12
18	2626000001	2672000000	2	4.60E+07	2
18	2655000001	2672000000	2	1.70E+07	2
19	2672000001	2719000000	7	4.70E+07	8
19	2672000001	2732000000	3	6.00E+07	4
19	2719000001	2732000000	2	1.30E+07	3
19	2719000001	2732000000	3	1.30E+07	7
19	2723000001	2732000000	3	9.00E+06	2
20	2732000001	2796000000	3	6.40E+07	17
21	2796000001	2845000000	2	4.90E+07	5
21	2796000001	2845000000	3	4.90E+07	14
22	2845000001	2897000000	1	5.20E+07	2
22	2845000001	2897000000	2	5.20E+07	13
22	2845000001	2897000000	3	5.20E+07	3
X	2898000001	3054000000	2	1.56E+08	12
X	2898000001	3054000000	3	1.56E+08	3

Table S9. Copy number segments identified in two or more single-cell and bulk DNA samples from ovarian cancer specimen 2 using a 1-Mb bin size.

“State” values correspond to the following copy numbers:

state	copy number
1	0 copies
2	1 copy
3	2 copies
4	3 copies
5	4 copies
6	5 copies
7	≥ 6 copies

chromosome	start	end	state	size	Number of samples
1	1	36000000	2	3.60E+07	3
1	1	37000000	2	3.70E+07	2
1	1	38000000	2	3.80E+07	5
1	1	250000000	3	2.50E+08	4
1	36000001	250000000	3	2.14E+08	2
1	37000001	250000000	3	2.13E+08	2
1	38000001	250000000	3	2.12E+08	4
2	250000001	278000000	2	2.80E+07	9
2	250000001	494000000	3	2.44E+08	3
2	278000001	494000000	3	2.16E+08	5
3	494000001	682000000	3	1.88E+08	4
3	494000001	693000000	3	1.99E+08	8
3	682000001	693000000	2	1.10E+07	4
4	693000001	750000000	3	5.70E+07	8
4	693000001	885000000	3	1.92E+08	3
4	750000001	825000000	2	7.50E+07	2
4	750000001	826000000	2	7.60E+07	3
4	750000001	885000000	2	1.35E+08	2
4	826000001	851000000	3	2.50E+07	3
4	851000001	885000000	2	3.40E+07	5
5	885000001	906000000	4	2.10E+07	3
5	885000001	908000000	4	2.30E+07	2
5	885000001	950000000	3	6.50E+07	3
5	885000001	951000000	3	6.60E+07	3
5	885000001	1066000000	3	1.81E+08	3
5	950000001	1066000000	2	1.16E+08	3
5	951000001	1066000000	2	1.15E+08	2

5	1053000001	1066000000	3	1.30E+07	2
6	1066000001	1078000000	2	1.20E+07	2
6	1066000001	1079000000	2	1.30E+07	7
6	1066000001	1080000000	2	1.40E+07	4
6	1066000001	1238000000	3	1.72E+08	2
6	1079000001	1140000000	3	6.10E+07	3
6	1080000001	1140000000	4	6.00E+07	2
6	1140000001	1175000000	2	3.50E+07	3
6	1140000001	1176000000	2	3.60E+07	2
6	1140000001	1238000000	2	9.80E+07	3
6	1191000001	1238000000	2	4.70E+07	4
6	1192000001	1238000000	2	4.60E+07	3
6	1194000001	1238000000	2	4.40E+07	2
7	1238000001	1291000000	3	5.30E+07	2
7	1238000001	1293000000	3	5.50E+07	2
7	1238000001	1294000000	3	5.60E+07	5
7	1238000001	1308000000	3	7.00E+07	2
7	1294000001	1375000000	4	8.10E+07	2
7	1294000001	1377000000	4	8.30E+07	2
7	1375000001	1398000000	5	2.30E+07	2
7	1376000001	1398000000	5	2.20E+07	2
7	1377000001	1398000000	5	2.10E+07	2
8	1398000001	1489000000	3	9.10E+07	3
8	1398000001	1491000000	3	9.30E+07	11
8	1489000001	1545000000	5	5.60E+07	2
8	1491000001	1545000000	4	5.40E+07	6
9	1545000001	1616000000	2	7.10E+07	9
9	1545000001	1618000000	2	7.30E+07	2
9	1545000001	1687000000	3	1.42E+08	4
9	1616000001	1687000000	3	7.10E+07	7
9	1618000001	1687000000	3	6.90E+07	2
9	1664000001	1687000000	4	2.30E+07	2
10	1687000001	1775000000	3	8.80E+07	5
10	1687000001	1777000000	3	9.00E+07	2
10	1687000001	1790000000	3	1.03E+08	3
10	1687000001	1794000000	3	1.07E+08	3
10	1687000001	1823000000	3	1.36E+08	2
10	1775000001	1823000000	2	4.80E+07	6
10	1777000001	1823000000	2	4.60E+07	2
10	1790000001	1823000000	2	3.30E+07	3
10	1794000001	1823000000	2	2.90E+07	3
11	1823000001	1890000000	3	6.70E+07	2
11	1823000001	1894000000	3	7.10E+07	4

11	1823000001	1959000000	3	1.36E+08	4
11	1888000001	1959000000	2	7.10E+07	2
11	1890000001	1959000000	2	6.90E+07	2
11	1894000001	1959000000	2	6.50E+07	4
12	1959000001	1993000000	4	3.40E+07	6
12	1959000001	1997000000	4	3.80E+07	4
12	1959000001	2093000000	3	1.34E+08	2
12	1993000001	2009000000	2	1.60E+07	2
12	1997000001	2010000000	2	1.30E+07	2
12	2008000001	2093000000	3	8.50E+07	2
12	2009000001	2093000000	3	8.40E+07	3
12	2010000001	2093000000	3	8.30E+07	3
13	2093000001	2147000000	3	5.40E+07	2
13	2093000001	2156000000	3	6.30E+07	10
13	2093000001	2209000000	3	1.16E+08	4
13	2147000001	2157000000	4	1.00E+07	2
13	2156000001	2209000000	2	5.30E+07	10
13	2157000001	2209000000	2	5.20E+07	4
14	2209000001	2302000000	2	9.30E+07	6
14	2209000001	2303000000	2	9.40E+07	3
14	2209000001	2317000000	3	1.08E+08	2
14	2296000001	2317000000	3	2.10E+07	2
14	2302000001	2317000000	3	1.50E+07	4
14	2302000001	2317000000	4	1.50E+07	2
14	2303000001	2317000000	3	1.40E+07	2
15	2317000001	2420000000	3	1.03E+08	18
16	2420000001	2495000000	3	7.50E+07	2
16	2420000001	2496000000	3	7.60E+07	2
16	2420000001	2497000000	3	7.70E+07	4
16	2420000001	2511000000	3	9.10E+07	5
16	2495000001	2511000000	5	1.60E+07	2
16	2496000001	2511000000	4	1.50E+07	2
16	2497000001	2511000000	4	1.40E+07	4
17	2511000001	2521000000	3	1.00E+07	2
17	2511000001	2522000000	3	1.10E+07	7
17	2511000001	2593000000	2	8.20E+07	4
17	2511000001	2593000000	3	8.20E+07	3
17	2521000001	2593000000	2	7.20E+07	2
17	2522000001	2593000000	2	7.10E+07	7
18	2593000001	2672000000	3	7.90E+07	13
19	2672000001	2711000000	3	3.90E+07	4
19	2672000001	2732000000	3	6.00E+07	13
19	2711000001	2732000000	2	2.10E+07	4

20	2732000001	2796000000	3	6.40E+07	17
21	2796000001	2837000000	2	4.10E+07	2
21	2796000001	2838000000	2	4.20E+07	2
21	2796000001	2845000000	2	4.90E+07	13
21	2796000001	2845000000	3	4.90E+07	2
22	2845000001	2897000000	2	5.20E+07	10
22	2845000001	2897000000	3	5.20E+07	8
X	2898000001	3017000000	3	1.19E+08	2
X	2898000001	3018000000	3	1.20E+08	8
X	2898000001	3018000000	4	1.20E+08	2
X	3017000001	3054000000	5	3.70E+07	2
X	3018000001	3054000000	5	3.60E+07	10

Descriptions of Supplementary datasets

Supplementary Dataset 1. Normalized read depth plots for a bin size of 1 Mb for all samples compared are available at <http://www.msl.ubc.ca/sites/default/files/pdata/kaston-leung/Leung-et-al-supplementary-dataset-1.zip>. The names and SD in reads per 1-Mb bin of each sample can be found in Table S4. Inferred copy numbers are indicated by the color of the datapoint for each bin as indicated in <http://www.msl.ubc.ca/sites/default/files/pdata/kaston-leung/CNV-legend-1.pdf>.

Supplementary Dataset 2. Normalized read depth plots for 100-kb and 10-kb bins for all single 184-hTERT droplet MDA samples sequenced to high depth are available at <http://www.msl.ubc.ca/sites/default/files/pdata/kaston-leung/Leung-et-al-supplementary-dataset-2.zip>. The names and SD in reads per 1-Mb bin of each sample can be found in Table S4. Inferred copy numbers are indicated by the color of the datapoint for each bin as indicated in <http://www.msl.ubc.ca/sites/default/files/pdata/kaston-leung/CNV-legend-2.pdf>.

Supplementary Dataset 3. Normalized read depth plots for 1 Mb bins for all 30 TOV2295 single-cell droplet MDA samples sequenced to low depth, and read depth plots for 100-kb and 10-kb bins for the 8 single-cell droplet MDA samples sequenced to high depth are available at <http://www.msl.ubc.ca/sites/default/files/pdata/kaston-leung/Leung-et-al-supplementary-dataset-3.zip>. Inferred copy numbers are indicated by the color of the datapoint for each bin as indicated in <http://www.msl.ubc.ca/sites/default/files/pdata/kaston-leung/CNV-legend-1.pdf>.

Supplementary Dataset 4. Normalized read depth plots for 1-Mb bins for all single-nuclei droplet MDA samples from two high-grade serous ovarian cancer tumour specimens are available at <http://www.msl.ubc.ca/sites/default/files/pdata/kaston-leung/Leung-et-al-supplementary-dataset-4.zip>. Inferred copy numbers are indicated by the color of the datapoint for each bin as indicated in <http://www.msl.ubc.ca/sites/default/files/pdata/kaston-leung/CNV-legend-1.pdf>.