Published online 7 January 2013

Nucleic Acids Research, 2013, Vol. 41, No. 6 ***e66***
doi:10.1093/nar/gks1352

Genome-wide copy number profiling of single cells in S-phase reveals DNA-replication domains

Niels Van der Aa[1], Jiqiu Cheng[2,3], Ligia Mateiu[1], Masoud Zamani Esteki[1], Parveen Kumar[1], Eftychia Dimitriadou[4], Evelyne Vanneste[4], Yves Moreau[2,3], Joris Robert Vermeesch[4] and Thierry Voet[1,*]

[1]Laboratory of Reproductive Genomics, Department of Human Genetics, KU Leuven, Leuven, 3000, Belgium, [2]ESAT-SCD, Department of Electrical Engineering, KU Leuven, Heverlee, 3001, Belgium, [3]SISTA, IBBT-KULeuven Future Health Department, KU Leuven, Heverlee, 3001, Belgium and [4]Laboratory for Cytogenetics and Genome Research, Department of Human Genetics, KU Leuven, Leuven, 3000, Belgium

Received August 1, 2012; Revised and Accepted December 5, 2012

ABSTRACT

Single-cell genomics is revolutionizing basic genome research and clinical genetic diagnosis. However, none of the current research or clinical methods for single-cell analysis distinguishes between the analysis of a cell in G1-, S- or G2/M-phase of the cell cycle. Here, we demonstrate by means of array comparative genomic hybridization that charting the DNA copy number landscape of a cell in S-phase requires conceptually different approaches to that of a cell in G1- or G2/M-phase. Remarkably, despite single-cell whole-genome amplification artifacts, the log2 intensity ratios of single S-phase cells oscillate according to early and late replication domains, which in turn leads to the detection of significantly more DNA imbalances when compared with a cell in G1- or G2/M-phase. Although these DNA imbalances may, on the one hand, be falsely interpreted as genuine structural aberrations in the S-phase cell's copy number profile and hence lead to misdiagnosis, on the other hand, the ability to detect replication domains genome wide in one cell has important applications in DNA-replication research. Genome-wide cell-type-specific early and late replicating domains have been identified by analyses of DNA from populations of cells, but cell-to-cell differences in DNA replication may be important in genome stability, disease aetiology and various other cellular processes.

INTRODUCTION

Methods to profile the genome of a single cell are paramount to study fundamental processes of genome maintenance (1), to dissect the cellular makeup of genetically heterogeneous tissues to understand phenotypes and diseases (2–5) and to enable the genetic diagnosis of rare cells in the clinic (6–12). Single-cell DNA-copy number profiling methods underpinned by array comparative genomic hybridization (aCGH), SNP-array or next-generation sequencing (NGS) analyses delivered new insight in DNA mutation during human gametogenesis (13–15), embryogenesis (1,16) and tumourigenesis (2,4,5), as well as in the aetiology of congenital and acquired genetic diseases (2,4,5,16). In the clinic, single-cell genomics is revolutionizing preimplantation genetic diagnosis (PGD) of human embryos following *in vitro* fertilization (8–12) and may in the future become important for diagnosis, prognosis and treatment of cancer by the analysis of circulating tumour cells isolated from the patient's blood stream (6,7).

The minute amount of DNA present in a cell must first be amplified to meet the DNA input requirements for hybridization onto microarrays or for the preparation of a next-generation sequencing library. However, to date, all available whole-genome amplification (WGA) methods result in a biased representation of the original single-cell genome including artifacts as allele drop out, preferential amplification (17), structural DNA anomalies (18) and nucleotide copying errors (4,5,13). Although the majority of current single-cell DNA copy-number analysis pipelines correct for allelic WGA bias, none of them consider the fact that the sensitivity and the specificity of DNA-copy

*To whom correspondence should be addressed. Tel: +32 16 330841; Fax: +32 16 346060; Email: Thierry.Voet@med.kuleuven.be

number profiling methods may be affected by the cell cycle status of the isolated cell (19–24). During S-phase the cell's genetic material is replicated progressively from multiple origins of DNA replication that should be fired only once during a cell's cycle. The DNA regions that replicate from a single replication origin, also known as replicons, typically range from 30–450 kb in the mammalian genome, although replicons with sizes <10 kb or >1 Mb have also been reported (25). These replicons are the building units of replication domains, which consist of loci with a similar replication timing. Although replication domains follow a cell type-specific time schedule (26–28), origin firing within domains occurs stochastically (29). Hence, a genetic snapshot of a diploid cell in S-phase will demonstrate alternating loci of copy number state 2, 3 or 4. The number of the loci, their size and copy number state is dynamic over the entire S-phase. Consequently, to warrant reliable interpretation and detection of structural DNA imbalances in single cells, it is imperative to investigate to what extent cell cycle status may introduce aberrations in DNA-copy number profiles of individual cells.

Although DNA-copy number profiles of individual cells in S-phase are hypothetically compromised by ongoing DNA replication, the ability to detect the newly synthesized DNA in a single S-phase cell will deliver novel understanding of DNA replication. Thus far, genome-wide studies of DNA replication are limited to the analyses of populations of cells (27,28,30,31). In 2004, Woodfine *et al*. (30) proved that aCGH analysis of DNA extracted from millions of S-phase cells versus differentially labelled DNA of many G1-phase cells on BAC arrays allowed the deduction of a DNA replication timing pattern of those cells. More recent studies, using *in vivo* pulse labelling of newly synthesized DNA with 5-bromo-2-deoxyuridine (BrdU) and subsequent oligo-array aCGH (28,31) or sequencing (27) analyses of the BrdU-labelled DNA fractions isolated from populations of early and late S-phase cells revealed genome-wide replication timing patterns at much higher resolution and uncovered the plasticity of replication patterns among cell types (27,28,31). Methods that enable the detection of DNA replication in a single cell genome wide will, however, allow the investigation of the variation in DNA-replication timing among cells belonging to the same type. This will lead to a better understanding of DNA replication and its association with other cellular processes such as cell fate transitions (31) and disease aetiology (26,32-37). The fact that aberrant DNA replication has been linked to genomic disorders (33,35) and cancer (26,32,36,38) emphasizes further the importance of understanding DNA replication at the single-cell level.

Here we applied aCGH to whole-genome amplified DNA of individual G1-, S- and G2/M-phase cells as well as to non-WGA DNA extracted from populations of G1-, S- and G2/M-cells, which served as controls. We demonstrate that individual replication domains can be detected genome wide at the single-cell level using aCGH and can be mistaken for genuine structural DNA imbalances in the cell. We propose a work flow to detect single cells in S-phase and to correct for DNA replication bias before copy number profiling.

MATERIALS AND METHODS

Cell culture, fluorescence-activated cell sorting (FACS) and DNA extraction

Epstein-Barr virus (EBV)-immortalized lymphoblastoid cells were cultured in DMEM/F12 medium (Gibco) with 10% fetal bovine serum (Thermo Scientific). Cells were stained with Vybrant DyeCycle Orange stain (Invitrogen) according to the manufacturer's protocol. In short, cells were washed with 1× PBS and incubated at a concentration of 10^6 cells/ml at 37°C for 30 min after adding Vybrant DyeCycle Orange at 2 µl/10^6 cells. Cells of three lymphoblastoid cell lines at $\sim 5 \times 10^6$ cells/ml were sorted using a FACS Vantage SE or a FACSAria III with FACSDiva software (BD Biosciences). Two female normal control cell lines and one cell line from a male carrying a 750 kb amplification on the p-arm of chromosome 4 (39) were used. Fractions of $1 \times 10^5 - 2.5 \times 10^6$ of S-phase, G1-phase and G2/M-phase cells were collected (Supplementary Figure S1a). Collected cells were lysed by overnight incubation at 50°C in lysis buffer (0.25 mM EDTA, 0.5% SDS and 0.1 mg/ml proteinase K in 1× PBS), and DNA was extracted using standard phenol/chloroform extraction and ethanol precipitation.

Isolation and whole-genome DNA amplification of single cells

Single EBV-immortalized lymphoblastoid cells in S-, G1- and G2/M-phase were collected from FACSed populations. The windows for the collected fractions were slightly narrowed (Supplementary Figure S1b), and cells were sorted at the 'Single Cell' precision setting using FACSDiva (BD Biosciences). Sorted cells were individually picked using a mouth system as described before (17), and single-cell whole-genome amplification was performed using the Sureplex amplification system (BlueGnome) according to the manufacturer's protocol.

BAC array hybridizations

For hybridization of multi-cell control samples, two array designs were used: Constitutional Chip 4.0 BAC array (Perkin Elmer) and 24sure array v3.0 (BlueGnome). For hybridization of single-cell samples, four array designs were used: two batches of 24sure array v2.0 (BlueGnome) and two batches of 24sure arrays v3.0 (BlueGnome). Differences between the batches of BlueGnome arrays entail differences in probes. Supplementary Table S1 lists the array used for each sample. For the hybridizations with multi-cell DNA, 1 µg (for Perkin Elmer arrays) or 150 ng (for BlueGnome arrays) of multi-cell DNA and an equal amount of commercial male reference DNA (Kreatech) was labelled for 2 h by random primer labelling (BioPrime aCGH Genomic Labeling System; Invitrogen) using respectively Cy5- and Cy3-dCTPs (GE Healthcare). For the hybridizations with single-cell amplified material, 150 ng of amplified DNA and 150 ng of SureRef male reference DNA (BlueGnome) were labelled for 4 h by random primer labelling (BioPrime aCGH Genomic Labeling System; Invitrogen) using

respectively Cy5- and Cy3-dCTPs (GE Healthcare). SureRef male reference DNA is provided by BlueGnome and represents a male reference DNA sample amplified by the Sureplex amplification system. Hybridization and washing were performed according to the manufacturer's protocol.

Scanning was performed using a DNA-microarray scanner (Agilent Technologies) for 24sure slides and a Genepix 4000 B (Axon instruments—Molecular Devices) for Constitutional Chip 4.0 slides. Feature extraction was performed using GenePix Pro 6.0 software (Axon, Molecular Devices).

Data normalization and copy number calling

GenePix Pro 6.0 software (Axon—Molecular Devices) was used to normalize the data such that the ratio of the means for all features equals 1. All further analyses were performed using R (version 2.13.2-www.r-project.org). Data were loaded into R using the limma package (40-42). Log2 intensity ratios with a signal/noise ratio of less than 2 were excluded from the analyses. The remaining intensities were normalized within each array using autosomal median correction. Next, the log2 intensity ratios of replicate probes were averaged using the R-package snapCGH (43). A number of log2 signal corrections were applied in this study, including 'technical GC-bias correction' (abbreviated here as techGC), 'per sample GC-bias correction' (abbreviated here as p.s.GC) and 'channel clone' correction. Technical GC-bias correction was performed by subtracting from each probe's log2 intensity ratio the corresponding value from a mean locally weighted regression. This mean locally weighted regression was computed using (i) the log2 intensity ratios for all G1- and G2/M-phase samples (per batch) and (ii) the %GC content of the corresponding micro-array probe locus. Using the spline function in R, missing values in this mean loess regression curve were inferred before correction of the log2 intensity ratios. Per sample GC correction was performed by subtracting from each probe's log2 intensity ratio the corresponding value from a locally weighted regression between the sample's log2 intensity ratios and the local %GC content. Channel clone normalization was performed on the feature extracted data as described by Cheng *et al.* (22). Copy number calling was performed using Circular Binary Segmentation (CBS) (44) combined with the CGHcall package (45) in R, or BlueFuse software (BlueGnome) with default parameter settings. Furthermore, normalized log2 intensity ratio values were also segmented using Piecewise Constant Fitting (PCF), which fits a piecewise constant function to the data, controlling the number of change points by a penalty parameter γ(46). In this study, we used a γ value of 1. Integer DNA copy number (both before and after segmentation) was estimated as $2\exp(\log2\ \text{ratio}) \times \Psi$, with ploidy Ψ of the cell set to 2.

Comparison of log2 intensity ratios with predicted replication domains

The replication ratio data of C0202 lymphoblastoid cells (replicate 1—Ryba *et al.* 2010 (28); GEO accession number GSE20027) was used to compare with our array data. These replication ratios or 'replication factors' are normalized log2 ratios from oligonucleotide array hybridizations comparing newly replicated DNA of early to late S-phase cell populations. To compare data obtained from different platforms and to compare our data with the known replication domain data (28), we used genomic bins, where for each data set, the quantitative data values of the probes were assigned to bins spanning replication domains (28). Pearson correlation tests, Student's *t*-tests and Wilcoxon rank sum tests were performed using the stats package in R.

To quantify the replication domains that had a matching S-phase single-cell log2 profile, we narrowed the analysis to only those domains that were covered by five or more microarray probes to account for putative single-cell WGA bias. Subsequently, a single-cell log2 profile within a particular replication domain was considered matching to an early or a late replication state when the mean value of the single-cell log2 intensity ratios across the domain-specific probes surpassed the threshold determined for an early or a late replicating domain in the positive or negative direction, respectively. The threshold used to quantify the matches of single-cell log2 intensity profiles to early replicating domains was set at the mean value of the autosomal log2 intensity ratios across all G1- and G2/M-phase single cells plus the standard deviation (SD) across the autosomal log2 intensity ratios of all G1- and G2/M-phase single cells. Similarly, the threshold used to quantify the matches of single-cell log2 intensity profiles to late replicating domains was set at the mean value of the autosomal log2 intensity ratios across all G1- and G2/M-phase single cells minus the standard deviation (SD) across the autosomal log2 intensity ratios of all G1- and G2/M-phase single cells.

Principal component analysis (PCA) was performed using R.

Data visualization

Circosplots of log2 intensity ratios and DNA-copy number calls were drawn using circos 'http://circos.ca/' (47). The log2 intensity ratios per sample were converted to heat maps by (i) calculating the mean log2 intensity value across all probes per replication domain known from Ryba *et al.* (28) and colour coding of each value, or (ii) smoothing of consecutive log2 intensity ratios using a moving mean across three probes and colour coding of the mean values. Copy number calls were visualized as gains (green) or losses (red).

RESULTS

S-phase log2 intensity ratios demonstrate a characteristic relationship with GC content in the DNA that is independent of WGA or aCGH artifacts

Cell populations enriched for G1-, S- and G2/M-phase were collected for three EBV-transformed lymphoblastoid cell lines by FACS for DNA content (Supplementary Figure S1). Of these populations, 14 S-, 8 G1- and

8 G2/M-phase single cells were isolated, WGAed and analysed by aCGH. As a control, non-amplified genomic DNA samples extracted from millions of cells of different populations enriched for G1-, S- or G2/M-phase cells were analysed in a similar manner in parallel.

Because whole-genome amplification is known to be affected by local richness of the DNA in guanine and cytosine bases (21), we first aimed to correct the obtained log2 intensity ratios for putative bias towards %GC content before downstream DNA-copy number analyses. However, investigation of the S-phase single-cell log2 ratios in the context of local %GC by smoothed locally weighted regression (Methods) revealed a typical correlation that was distinct from that of G1- or G2/M-phase single cells (Figure 1a). For loci having a GC fraction lower than 0.4, S-phase cells demonstrated by far lower log2 intensity ratio values when compared with G1- or G2/M-phase single cells, and vice versa for loci that were more GC-rich. This suggests that S-phase single-cell log2 intensity ratios were not only influenced by GC-dependent WGA bias, but also by GC-dependent biological factors. Indeed, the pattern can be reconciled with the fact that loci demonstrating an early replication timing are typically also gene-dense (48) (Supplementary Figure S2). A similar difference in interdependence between local %GC content and log2 intensity ratios for S-phase and non-S-phase DNA samples was observed in the multi-cell non-WGA control experiments (Supplementary Figure S3).

Loci with different DNA-replication timing are apparent in single S-phase cells

To investigate whether DNA-replication domains can be detected in single S-phase cell aCGH data, all single-cell log2 intensity ratios were first corrected based on a mean regression curve calculated between local %GC and all G1- and G2/M-phase single-cell log2 intensity ratios. This approach ensured a correction for GC-dependent WGA- and putative aCGH-hybridization artifacts, but leaving biologically relevant GC bias in S-phase cell data intact.

A clear oscillating pattern of consecutive positive and negative log2 intensity ratios became apparent in the aCGH data of single S-phase cells that furthermore coincided well with regions of known early and late DNA replication (28), respectively (Supplementary Figure S4a). This log2 pattern was not detected in the aCGH data of non-S-phase cells (Supplementary Figure S4b). To further investigate the correlation of the pattern of the log2 intensity ratios in S-phase cells to DNA replication, we generated heat maps of the mean log2 intensity ratio value per replication domain for each single-cell or multi-cell sample. These heat maps prove a strong correlation between the S-phase single-cell log2 intensity ratio profile and the anticipated replication states, and oscillate according to the replication status anticipated for each domain (Figure 2a and b). Additionally, S-phase single-cell heat maps were concordant with the control S-phase multi-cell heat maps (Figure 2a and b),

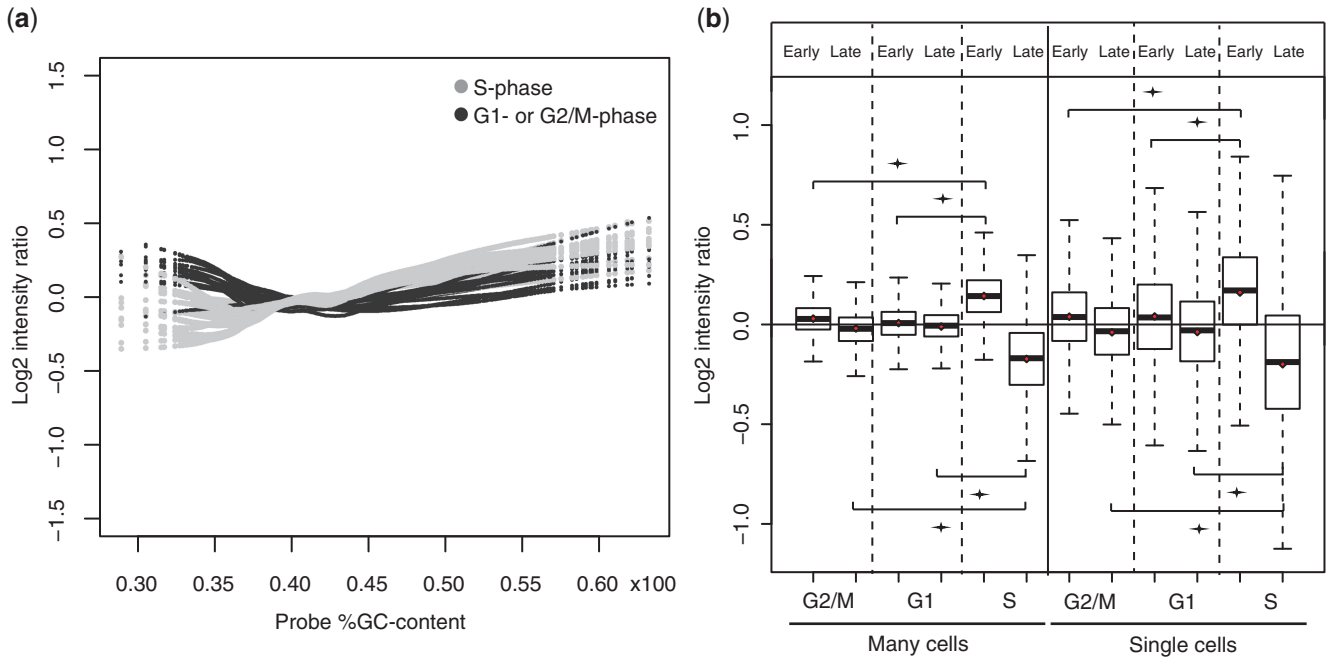

Figure 1. DNA replication is mirrored in single-cell aCGH log2 intensity ratios. (**a**) The correlation between log2 intensity ratio values and %GC content is different for S-phase single cells (grey lines) when compared with G1- and G2/M-phase cells (black lines). The X-axis depicts the %GC content per probe, the Y-axis the log2 intensity ratios per sample. Each line is a Loess fit using the data of a single-cell sample. (**b**) Boxplots for single cells (right) and multi-cell controls (left) depicting autosomal log2 intensity ratios that were pooled per cell cycle phase and per early or late DNA-replication domain. The annotation of the DNA-replication domains is described by Ryba *et al.* (28). Boxplots show the median of the log2 intensity ratios (central line), the quartiles (box and whiskers) and the mean of the log2 intensity ratios (diamonds). Relevant significant differences between G1-, G2/M- and S-phase cells are marked by a star (Student's *t*-test; $P < 0.05$).

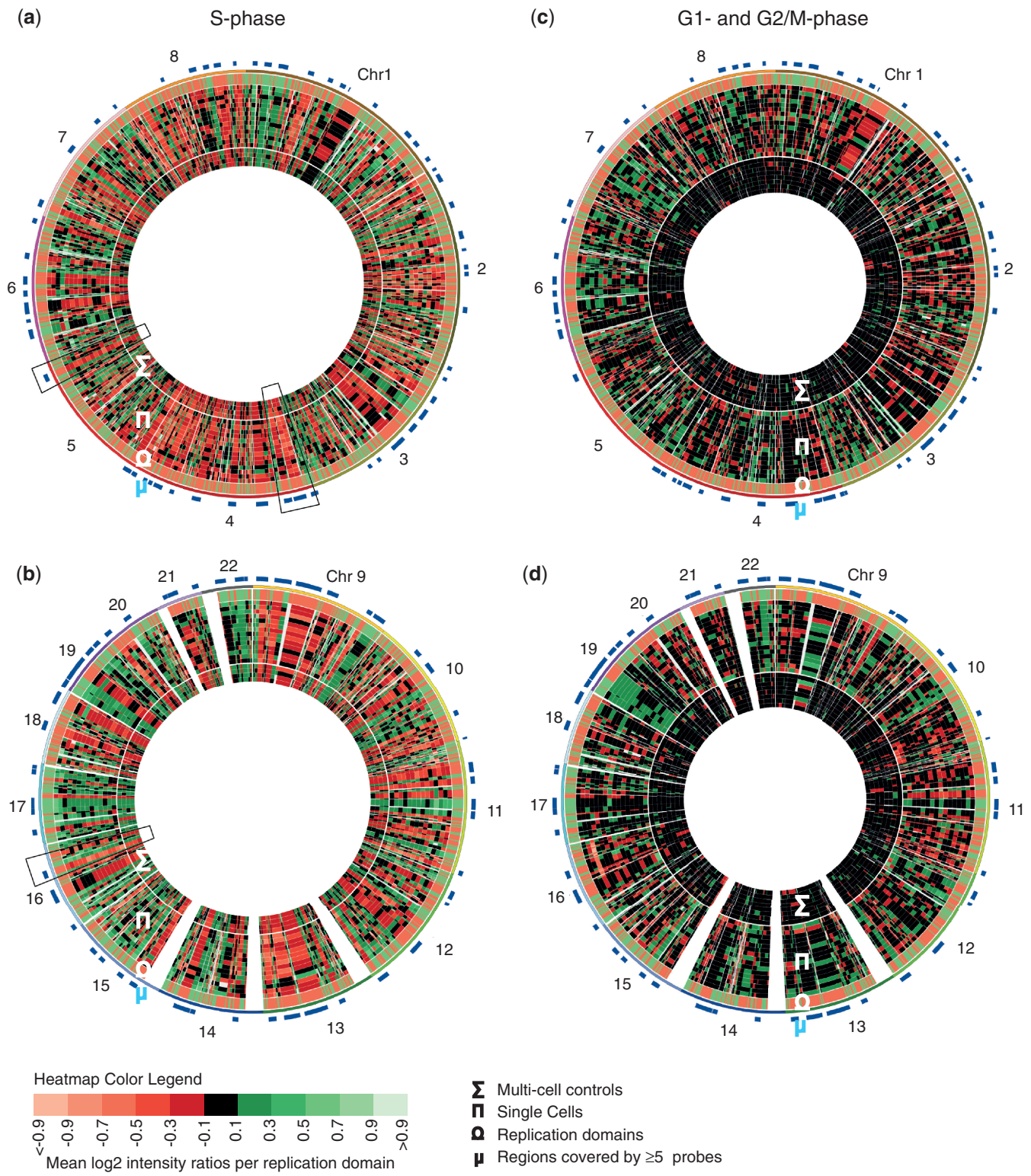

Figure 2. Log2 intensity ratios in S-phase cells oscillate according to predicted replication timing. For each single-cell and multi-cell control sample, a heat map of the mean log2 intensity ratio per replication domain across all autosomes is depicted in a circosplot. The colour-code legend of the heat map is depicted at the bottom of the figure. In each circosplot, the outermost circle depicts the predicted replication timing pattern as published by Ryba *et al.* (28) (early replicating domains in green; late replicating domains in red; replication domains covered by five or more microarray probes are marked by a blue bar on the outside). This is followed (from the outside to the inside of the circosplot) by the heat maps representing all single-cell log2 intensity ratio data (one cell per rim) and subsequently by the heat maps reflecting all multi-cell control log2 intensity ratio data (one multi-cell control per rim). (**a** and **b**) All S-phase single-cell and multi-cell samples. The top circosplot depicts the chromosomes 1 to 8 (a), bottom circosplot chromosomes 9 to 22 (b). The 14 S-phase single cell samples are shown in the following order (outside to inside): S1.3, S1.2, S3.1, S7.5, S7.6, S1.4, S7.7, S1.1, S7.1, S7.4, S4.1, S4.2, S7.2 and S7.3; followed by S-phase multi-cell control samples. (**c** and **d**) All G1- and G2/M-phase single-cell and multi- cell samples. Top circosplot depicts the chromosomes 1 to 8 (c), bottom circosplot chromosomes 9 to 22 (d). The 16 single-cell

(continued)

while the G1- and G2/M-phase single-cell as well as multi-cell heat maps showed no significant correlation with known replication domains (Figure 2c and d). Subsequent PCA of the mean log2 intensity ratio values per domain revealed that single-cell S-phase samples cluster separately from single-cell G1- and G2/M-phase samples, strongly suggesting that the pattern of log2 intensity ratios across single-cell S-phase samples is similar, and clearly distinct from the pattern of log2 intensity ratios in G1- and G2/M-phase single-cell samples (Figure 3). Furthermore, the Pearson correlation of the single-cell log2 intensity ratios to cell type- and locus-matching replication factor values obtained from analyses of multi-cell DNA samples by Ryba *et al.* (28) was significantly higher for single S-phase cells than for G1- and G2/M-phase cells (Methods, mean r of 0.51 and 0.24, respectively; Wilcoxon rank-sum test $P = 3.1 \times 10^{-5}$). Also, binning single-cell log2 intensity ratios according to known early and late replicating regions (28) revealed positive log2 intensity ratios for early replicating loci in S-phase cells that were significantly higher than the log2 ratios obtained for the same loci in non-S-phase cells (Figure 1b; student *t*-test; $P = 2.2 \times 10^{-16}$). Across late-replicating loci, significantly lower log2 ratios were detected in S-phase cells than in G1- and G2/M-phase cells (Figure 1b; student *t*-test; $P = 2.2 \times 10^{-16}$). In our multi-cell control aCGH experiments, similar observations were made (Figure 1b).

Furthermore, we demonstrate that individual DNA-replication domains can be detected in single S-phase cells when compared with G1- or G2/M-phase cell log2 intensity ratio profiles as clear patches of log2 intensity ratios deviating unidirectionally from the zero-axis (Supplementary Figure S5). For instance, multiple consecutive probes with negative log2 intensity ratios in all single S-phase cells on the short arm of chromosome 4 as well as on the long arms of chromosomes 5 and 16 pinpointed known late replicating domains in EBV-transformed lymphoblastoid cells (Supplementary Figure S5). In the vast majority of G1- and G2/M-phase single cells, those regions did not show these clear patches of log2 intensity ratios deviating from the zero axis,

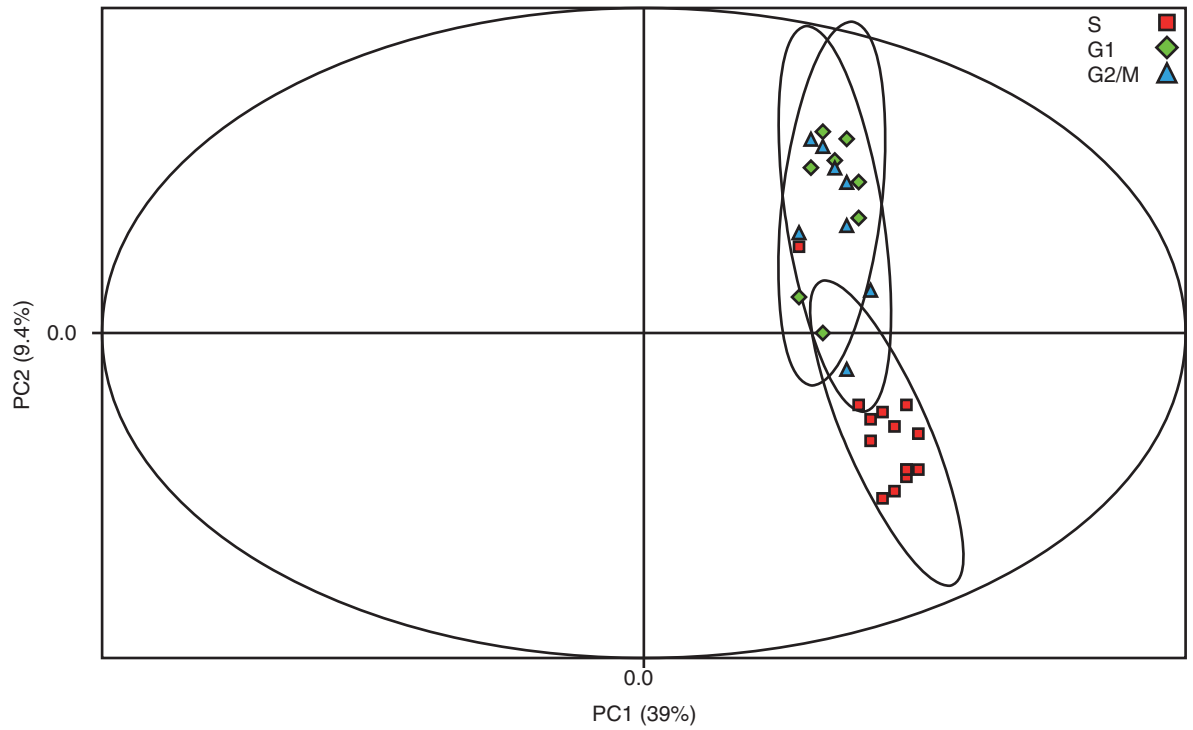

Figure 3. PCA of the single-cell mean log2 intensity ratio values per replication domain. The S-phase single-cell data are clearly separated from G1-, G2/M-phase single-cell data for all but one sample (cell S3.1 clusters aberrantly). Apart from an increased standard deviation across autosomal log2 ratios, cell S3.1 did not show the typical log2 intensity ratio behaviour expected following DNA replication as the other S-phase cells did. This suggests that either this cell is a G1- or G2/M-phase cell sorted wrongly into the S-phase cell population, or more likely that the whole-genome amplification product was of bad quality.

Figure 2. Continued
samples are shown in the following order (outside to inside): G1.1, G1.2, G1.3, G1.4, G3.1, G4.1, G7.1, G7.2, M1.1, M1.2, M1.3, M3.1, M3.2, M4.1, M7.1 and M7.2; followed by G1- and G2/M-phase multi-cell control samples. The oscillating pattern of consecutive positive and negative log2 intensity ratios genome wide orchestrated by early and late DNA replication in S-phase single cells can be clearly observed, and is concordant with the pattern in multi-cell controls. In contrast, the G1-phase and G2/M-phase single cells do not demonstrate such genome wide oscillation of log2 intensity ratios in accordance with the replication timing pattern. Similar observations can be made for the multi-cell controls. Three specific regions for which the log2 intensity plots are shown in Supplementary Figure S5 at high resolution are marked by a black box.

rather they showed log2 intensity ratios rippling around zero (Supplementary Figure S5). Furthermore, when we looked specifically at those replication domains covered by five or more microarray probes ($n = 121$, marked with blue bars in Figure 2), we found that the vast majority of the domains had a matching log2 intensity ratio profile across multiple S-phase cells (Supplementary Figure S6) in accordance with the progression of the cells in S-phase (see below). A single-cell log2 profile within a particular replication domain was considered matching to an early or a late replication state when the mean value of the single-cell log2 intensity ratios across the domain-specific probes surpassed the threshold determined for an early or a late replicating domain in the positive or negative direction, respectively (Material and Methods). In contrast, for all but two domains, the log2 intensity profiles of G1- and G2/M-phase cells almost never matched with the anticipated DNA-replicative status (Supplementary Figure S6). One of these two DNA-replication domains with matching log2 intensity ratios in G1- and G2/M-phase single cells likely represents a recurrent single-cell WGA artifact, as it has no match in the G1- or G2/M-phase multi-cell sample analyses. The other region shows a match in three of the eight multi-cell sample analyses, and thus presumably is a recurrent artifact of the labelling or hybridization aCGH process.

Ranking of individual S-phase cells according to replication status

Since the S-phase cells were isolated by FACS using DNA-content cut-offs that capture cells in a variety of DNA-replication stages ranging from early to mid to late S-phase, we investigated whether the 14 S-phase single cells could be ranked in their replication status based on their log2 intensity profiles. To achieve this, we quantified the fraction of early and late replication domains in the cells that had a matching S-phase single-cell log2 profile (Material and Methods). To account for putative whole-genome amplification bias, again only domains covered by five or more microarray probes were considered (Figure 2). It is expected that, given the normalization of the log2 intensity ratios we apply, cells residing in early S-phase demonstrate an initial overrepresentation of log2 intensity ratios matching early replicating regions while cells beyond mid S-phase show an initial overrepresentation of log2 intensity ratios matching late replicating domains. The fraction of early replication domains that had matching log2 intensity ratio profiles varied from 8.7% to 65.2% across the S-phase single cells, and 10% to 74.7% for the late replication domains (Figure 4a). Using the fraction of early and late replication domains demonstrating matching log2 intensity ratio profiles per cell, the individual S-phase cells could be ranked from early to mid to late S-phase (left to right in Figure 4a). This ranking based on replication domains covered by five or more microarray probes was in addition corroborated by the anticipated bimodal distributions of all single-cell S-phase autosomal log2 intensity ratios (Figure 4b). The fractions of early and late replication domains with matching log2 intensity ratios was significantly lower across the G1- and G2/M-phase single cells (Figure 4a right panel, Wilcoxon rank-sum test $P = 1 \times 10^{-5}$ for early domains and $P = 6 \times 10^{-5}$ for late domains). Last but not least, lenient PCF segmentation of the single-cell log2 intensity ratios and copy number transformation further confirmed the order of ranking of the individual S-phase cells (see below; Figure 5a).

DNA replication introduces (pseudo) false positive copy number variations

As a result of the oscillation in log2 intensity ratios according to DNA replication in S-phase cells, also the standard deviation (SD) of the log2 intensity ratios across autosomal probes was found to be significantly higher in aCGH analyses of single S-phase cells than in hybridizations of WGA-DNA of G1- and G2/M-phase cells (the mean SD for S-, G1- and G2/M-phase single-cell data is respectively 0.35, 0.25 and 0.20; Wilcoxon rank-sum test $P < 0.05$; Supplementary Figure S7). Since duplicate probes on the array show similar log2 intensity ratios, it is highly unlikely that the increment in SD in single-cell S-phase log2 values is a technical hybridization artifact (Supplementary Figure S7). These higher SD values and the log2 intensity ratio oscillations typical for S-phase cells may drastically affect the sensitivity and the specificity of algorithms that interpret single-cell data for structural DNA imbalances (see below). Similar observations were made in the multi-cell control aCGH experiments (the mean SD for S-, G1- and G2/M-phase multi-cell data is 0.21, 0.09 and 0.10, respectively; Wilcoxon rank-sum test $P < 0.05$; Supplementary Figure S7). Using lenient PCF segmentation and subsequent integer DNA-copy number transformation of the single-cell log2 intensity ratios (preprocessed as above), a number of interesting observations could be made: (i) In the S-phase single cells, many DNA-copy number gains and losses were revealed, often coinciding with known early and late replication domains or with a concatenation of (large) DNA-replication domains having a similar replication timing (Figure 5a, Supplementary Figure S8a; Materials and Methods). The latter concatenation of replication domains is most likely due to segmentation or smoothing parameters applied on the log2 intensity ratios by the copy number typing algorithm, thereby skipping intervening smaller domains of opposite replication timing that are covered by only one or few probes (Supplementary Figure S8). (ii) The S-phase cells categorized as being early S-phase cells demonstrated more DNA-copy number gains coinciding with early replicating domains, whereas S-phase cells ranked as being late S-phase cells contained more false positive deletions coinciding with late replicating domains (Figure 5a, Supplementary Figure S8a; the order of the S-phase single cells on Figure 5a from the outside to the inside of the circosplot is according to their early to mid to late rank in S-phase). These copy number profiles are thus in accordance with the previous ranking of the S-phase single cells according to their progression status in the

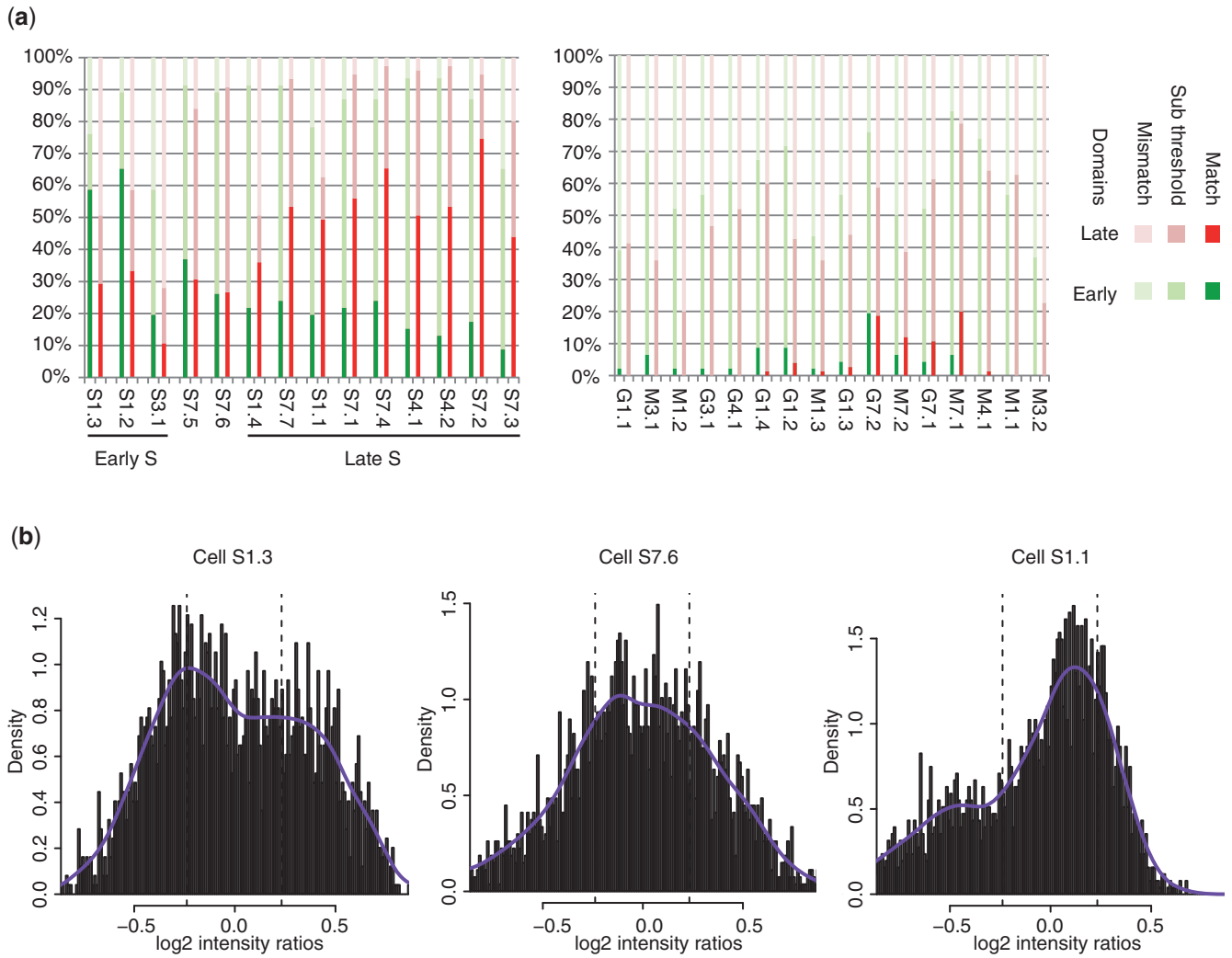

Figure 4. Single S-phase cells ranked according to their progression in S-phase. (**a**) Barplots showing per cell the fraction of early and late replicating domains covered by five or more probes ($n = 121$) for which (i) the single-cell mean log2 intensity ratios surpassed the threshold that was assigned for the expected replication timing (indicated as 'Match' in the legend; see also Materials and Methods), (ii) the single-cell mean log2 intensity ratios did not surpass the threshold but were still above or below the zero axis in accordance with the expected replication timing (indicated as 'Sub threshold' in the legend) and (iii) the single-cell mean log2 intensity ratios mismatched the expected replication timing (indicated as 'Mismatch' in the legend). Fractions of early replicating domains are depicted in green colours, those for late replicating regions in red colours. The left panel shows all S-phase single-cell samples ordered into groups of early-, mid- and late-S-phase cells (left to right), based on the fraction of early and late replication domains having matching mean log2 intensity ratios in the S-phase cells. The right panel shows the data for all single-cell G1- and G2/M-phase samples. (**b**) Three single-cell S-phase log2 distribution plots of an early-S-phase cell, a mid-S-phase cell and a late-S-phase cell from left to right. The density plot across all log2 intensity ratios is shown in purple. Cell S1.3 (left panel) shows a bimodal distribution of the log2 intensity values with the highest peak left and a lower peak on the right side of the high peak, suggesting the cell is in early S-phase. Cell S7.6 (middle panel) shows a bimodal distribution with two peaks of approximately the same height, suggestive for a cell in mid-S-phase. Cell S1.1 (right panel) shows a low peak to the left of the high peak which is close to zero, suggesting this cell is at a late stage in S-phase.

S-phase (Figure 4). Using more stringent CBS segmentation and CGHcall for copy number detection on the same preprocessed single-cell log2 intensity ratios revealed less but larger copy number aberrations of which still many coincided well with (concatenated) DNA-replication domains (Figure 5c, Supplementary Figure S8b). These observations are in stark contrast with the DNA-imbalance profiles detected for G1- and G2/M-phase single cells, in which the copy number aberrations called following lenient PCF or CBS segmentation are not only fewer in number, but also show only stochastic overlap with the known replication domains (Figure 5b and d). Similar observations could be made in the computed copy number landscapes of the multi-cell control samples.

To investigate further to what extent DNA replication influences single-cell DNA-copy number calling, we analysed our samples using a variety of algorithmic approaches and different preprocessing methods (Figure 1, Supplementary Figure S7 and S9 show the average autosomal SD and the log2 intensity ratios in early and late replicating domains following all preprocessing methods used). Table 1 lists the number of putative structural DNA imbalances identified in the autosomes by the different preprocessing and copy-number calling methods, while

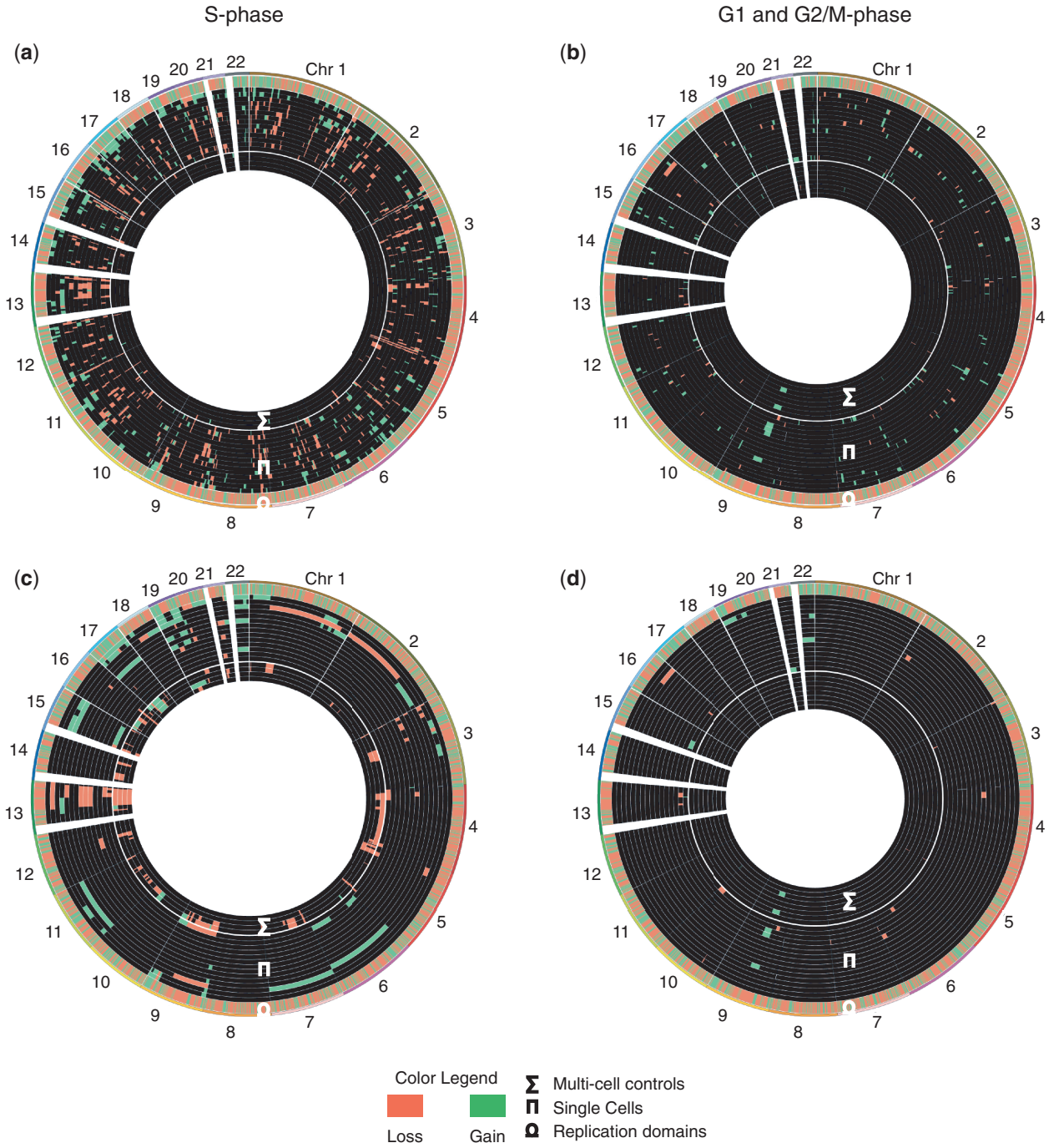

Figure 5. Copy number aberrations called in S-phase samples are concordant with replication domains. Circosplots depict the loci called as DNA copy number gains (green) or losses (red) across the autosomes for different copy number analysis methods. The left circosplots contain all S-phase single-cell and multi-cell samples (panels a, c), the right circosplots contain all G1- and G2/M-phase samples (panels b, d). In each circosplot, the outermost circle depicts the predicted replication timing pattern as published by Ryba *et al.* (28) (early replicating domains in green; late replicating domains in red). This is followed (from the outside to the inside of the circosplot) by the copy number heat maps of all single cells (one cell per rim, ranked according to S-phase progression as in Figure 4a) and subsequently by the copy number heat maps of all multi-cell control samples (one multi-cell control per rim). Importantly, (i) the type and size of copy number aberrations called in the S-phase single cells are often concordant with one particular replication domain or with a concatenation of (larger) replication domains having similar replication timing. (ii) Cells ranked to be late in S-phase show more DNA losses while cells ranked to be in early S-phase show more DNA-copy number gains (from the outside to the inside of the circosplot: the S-phase cells are ranked from early to mid to late S-phase stage, respecting the ranking proposed in Figure 4a). (**a, b**) Piecewise constant fitting and integer DNA-copy number transformation of log2 intensity ratios normalized according to their autosomal median value and corrected for GC bias according to a mean Loess regression curve between %GC and G1/G2/M-phase log2 values (techSD). (**c, d**) CBS segmentation and CGHcall (Methods) after log2 intensity ratios were normalized according to their autosomal median value and corrected for GC bias according to a mean Loess regression curve between %GC and G1/G2/M-phase log2 values (techSD). (a, c) Specifically, the 14 S-phase single cell samples are shown in the following order (outside to inside): S1.3, S1.2, S3.1, S7.5, S7.6, S1.4, S7.7, S1.1, S7.1, S7.4, S4.1, S4.2, S7.2 and S7.3; followed by S-phase multi-cell control samples. (b, d) The 16 G1- and G2/M-phase single-cell samples are shown in the following order (outside to inside): G1.1, G1.2, G1.3, G1.4, G3.1, G4.1, G7.1, G7.2, M1.1, M1.2, M1.3, M3.1, M3.2, M4.1, M7.1 and M7.2; followed by G1- and G2/M-phase multi-cell control samples.

Table 1. Number of autosomal DNA-copy number variants detected using different log2 preprocessing and copy-number detection methods

Method	S-phase single cells	G/M-phase single cells	S-phase multi-cell control	G/M-phase multi-cell control
BlueFuse[a]	12 (5.42)[g]	5 (3.62)	10 (2)[h]	0 (0)[h]
Median[b]	9.28 (4.03)[g]	2.40 (2.03)	26 (22.51)[g]	0.75 (0.83)
p.s.GC[c]	7 (5.46)[g]	1.19 (1.01)	11.75 (8.79)[g]	0.75 (0.83)
techGC[d]	7.71 (5.92)[g]	1.81 (1.70)	27 (16.39)[g]	0.75 (0.83)
Channel Clone[e]	0.36 (1.29)	0.25 (0.75)	0.75 (0.83)	0.75 (0.83)
PCF[f]	67.2 (23.3)[g]	13.25 (15.19)	8.25 (7.79)	1.75 (1.71)

For each method, the mean number of copy number aberrations detected over all experiments is given. The standard deviation between the number of copy number aberrations detected for individual samples is given between parentheses.
Copy number variants were called using
[a]BlueFuse commercial software (BlueGnome) with default parameters for both preprocessing of log2 intensity values and copy number calling.
[b]CBS segmentation and CGHcall (Methods) after log2 intensity ratios were normalized according to their autosomal median value.
[c]CBS segmentation and CGHcall (Methods) after log2 intensity ratios were normalized according to their autosomal median value and their local %GC content.
[d]CBS segmentation and CGHcall (Methods) after log2 intensity ratios were normalized according to their autosomal median value and corrected for GC bias according to a mean Loess regression curve between %GC and G1/G2/M-phase log2 values.
[e]CBS segmentation and CGHcall (Methods) after 'Channel clone' preprocessing as described by Cheng *et al.* (22).
[f]PCF segmentation with γ-value = 1 after log2 intensity ratios were normalized according to their autosomal median value and corrected for GC bias according to a mean Loess regression curve between %GC and G1/G2/M-phase log2 values.
[g]The count in the S-phase samples is significantly higher than the count in the G1- and G2/M-phase samples (Wilcoxon ranked-sum test $P < 0.05$).
[h]Only the multi-cell control DNA samples hybridized to the BlueGnome platform can be analysed by the BlueFuse software.

Figure 5 and Supplementary Figure S10 depict all detected single-cell and multi-cell DNA copy number landscapes genome wide. Importantly, (i) despite variation between individual single-cell samples, significantly more false positive copy-number aberrations were called in the aCGH data of single S-phase cells than in cells that were caught in the other cell cycle phases (Wilcoxon rank sum $P < 0.05$ for each method, except the channel clone method), (ii) many of the copy number aberrations called in S-phase single-cell and multi-cell analyses demonstrated concordance with known replication domains or with a concatenation of (larger) domains having similar replicative timings most likely due to data segmentation or smoothing parameters of the copy number typing algorithm thereby skipping intervening smaller domains of opposite replication timing (Figure 5, Supplementary Figure S8, Supplementary Figure S10), and (iii) different preprocessing methods prior to copy number calling had a significant effect on the amount, size and type of (pseudo) false positive copy number aberrations called in the cells and their respective overlap with known replication domains (Figure 5, Supplementary Figure S8, Supplementary Figure S10, Table 1). Similar results were obtained with the control multi-cell DNA samples (Figure 5, Supplementary Figure S10, Table 1). For example, the commercial BlueFuse software, used extensively in IVF clinics for PGD or screening purposes, called many small false positive copy number aberrations in single S-phase cells. However, a channel clone normalization (22), entailing both a GC correction and a batch wave correction, removed most small false positive structural DNA imbalances detected by the other preprocessing and copy number detection approaches in the S-phase single-cell and multi-cell samples. In one G1-phase cell a genuine *de novo* deletion of the q-arm of chromosome 16 was detected across all copy number analyses (Supplementary Figure S10, Supplementary Figure S11, Figure 5).

DISCUSSION

Single-cell genomics has important applications in basic genome research (1,2,4,5,16) and methods for detecting structural DNA imbalances in a cell are becoming routinely applied in the clinic for the purpose of genetic diagnosis (8-12). Understanding all origins of false copy number annotation in single-cell genome analyses is thus of utmost importance. Current studies do not distinguish between the analyses of cells in G1-, S- or G2/M-phase and target cells randomly for analysis (1,4,5,7,8,11,16, 19–21). However, here we demonstrate that charting the genetic landscape of a cell in S-phase requires conceptually different approaches to the analysis of a cell in G1- or G2/M-phase. The variability in DNA-replication status for consecutive loci across the genome of S-phase cells 'tricks' methods for single-cell copy number profiling, leading to false-positive calls for structural DNA imbalances.

We show that S-phase log2 intensity ratios demonstrate a typical interdependency to the fraction of guanine and cytosine bases present in the respective loci interrogated by the microarray. Besides a %GC-dependent WGA bias apparent from the single-cell G1- or G2/M-phase log2 intensity ratio data, single-cell S-phase aCGH data in addition show a %GC-dependent bias that is biologically determined and in line with the fact that gene-dense regions tend to double their DNA earlier than gene-deserts. Simple correction of the S-phase log2 intensities ratios to the %GC bias is however not sufficient to remove all false-positive copy number calls in a single S-phase cell. The majority of current methods for single-cell copy number variation detection, based on aCGH (19–21) or sequencing (2), are able to interpret the data only in the context of the well-known whole-genome amplification artifacts such as allele drop out, preferential amplification and amplification artifacts due to GC-richness of the locus. However, none of the methods incorporates

strategies to correct the log2 intensity ratios for DNA-replication bias and hence to reduce false-positive calls for structural DNA imbalances in cells. In a population of cells, G0- or G1-cells are usually the predominant class, and thus S-phase cells will usually not interfere with copy number calling in multi-cell populations. However, once individual cells are randomly selected from these populations for single-cell genome analysis without prior knowledge on the cell's G1-, S-, G2/M-phase status, our results call for caution in the interpretation of the data. In addition, fast dividing cells as human blastomeres or some tumour cells are expected to reside more often in S-phase, and thus chances are higher to isolate a cell in S-phase from such a population. Replication domains for EBV-transformed lymphoblastoid cells based on the data from Ryba *et al.* (28) range between 2 kb and 25 Mb, with an average size of 1.8 Mb. Hence, depending on which stage in S-phase the cell is in, such individual replication domains or a concatenation of multiple domains can easily be misinterpreted as bona fide structural DNA imbalances. Our results indicate that a channel clone correction (22) before copy number variant detection, which adjusts the log2 intensity ratios to recurrent bias between cells of an analysed batch, removes most of the (pseudo) false positive structural DNA imbalances that are called in S-phase single-cell and multi-cell samples by the other preprocessing and copy number detection approaches evaluated here.

In contrast, if one wants to study DNA replication in the cell, it is best to correct the log2 intensities ratios for the technical WGA GC bias (observed across G1- and G2/M-phase cells) and to further investigate these values using e.g. a moving average or a lenient PCF or CBS segmentation of the log 2 intensity ratio values. DNA replication is a fundamental process of life; however, many aspects of its modus operandi, certainly at a single-cell resolution, as well as its molecular links to other cellular processes remain elusive (26,49). Additionally, disturbances in DNA replication have been found to lie at the heart of disease-causing structural DNA anomalies in the germline as well as in somatic tissues, through mechanisms as FOSTES (50), mmBIR (51) and micronuclei-mediated chromothrypsis (52). Methods to study DNA-replication timing in single cells have already been used to investigate cell-to-cell variation (53) and offered important novel insight in the efficiency, coordination and heterogeneity of replication at the level of individual replication origins (54). However, these single-cell methods underpinned by FISH (53) or DNA-combing (54,55) have important limitations. FISH is a locus-specific method, lacking in resolution and sensitivity (53). DNA-combing methods do offer better resolution, allowing for the study of stochasticity of individual origins, yet they are also locus-specific and it is not possible to trace individual DNA fibres back to the cell of origin (54,55). Here we show for the first time that DNA-replication domains can be detected in single cells on a genome-wide scale using an aCGH approach despite it is known that WGA artifacts such as allele dropout and preferential amplification can distort single-cell aCGH signals (17). Furthermore, based on the fractions of early and late replicating domains that show matching single-cell log2 intensity ratios, we were able to order S-phase single cells according to their specific progression in S-phase. Although this cell-to-cell variation in replication timing could be tracked, yet higher resolution analyses of individual cells undergoing DNA replication using single-cell deep-sequencing approaches hold the potential to study the stochasticity of DNA-replication timing between individual cells of the same cell type and may prove an indispensable method for studying the mechanisms of replication timing (31,56).

In conclusion, we deliver proof-of-concept for the detection of replication domains in single cells genome wide using aCGH. Hence, studies that analyse the genome of randomly selected single cells for copy number aberrations must take into account the false positive copy number aberrations due to DNA replication. We provide a work flow to detect individual cells in S-phase and to correct for DNA-replication bias prior to copy number profiling.

DATA DEPOSITION

Data are available at the Gene Expression Omnibus (GEO) through accession number GSE38761.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1–11.

FUNDING

Research Foundation Flanders (FWO) [G.A093.11 and 1.1.H28.12 to T.V. and N.V.d.A.]; KULeuven SymBioSys [PFV/10/016 to Y.M., J.R.V. and T.V.]; Agency for Innovation by Science and Technology (IWT) [SBO-60848 to J.R.V.]; KULeuven GOA MaNet (to Y.M.); Flemish government Hercules III PacBioRS (to Y.M.); EU-RTD CHeartED [FP7-HEALTH to Y.M.]; COST NGS Data analysis network [Action BM1006 to Y.M.]. Funding for open access charge: FWO [G.A093.11 to T.V.].

Conflict of interest statement. None declared.

REFERENCES

1. Vanneste,E., Voet,T., Le Caignec,C., Ampe,M., Konings,P., Melotte,C., Debrock,S., Amyere,M., Vikkula,M., Schuit,F. *et al.* (2009) Chromosome instability is common in human cleavage-stage embryos. *Nat. Med.*, **15**, 577–583.
2. Navin,N., Kendall,J., Troge,J., Andrews,P., Rodgers,L., McIndoo,J., Cook,K., Stepansky,A., Levy,D., Esposito,D. *et al.* (2011) Tumour evolution inferred by single-cell sequencing. *Nature*, **472**, 90–94.
3. Baillie,J.K., Barnett,M.W., Upton,K.R., Gerhardt,D.J., Richmond,T.A., De Sapio,F., Brennan,P.M., Rizzu,P., Smith,S., Fell,M. *et al.* (2011) Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*, **479**, 534–537.
4. Hou,Y., Song,L., Zhu,P., Zhang,B., Tao,Y., Xu,X., Li,F., Wu,K., Liang,J., Shao,D. *et al.* (2012) Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell*, **148**, 873–885.

5. Xu,X., Hou,Y., Yin,X., Bao,L., Tang,A., Song,L., Li,F., Tsang,S., Wu,K., Wu,H. *et al.* (2012) Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell*, **148**, 886–895.
6. Navin,N. and Hicks,J. (2011) Future medical applications of single-cell sequencing in cancer. *Genome Med.*, **3**, 31.
7. Mathiesen,R.R., Fjelldal,R., Liestol,K., Due,E.U., Geigl,J.B., Riethdorf,S., Borgen,E., Rye,I.H., Schneider,I.J., Obenauf,A.C. *et al.* (2012) High-resolution analyses of copy number changes in disseminated tumor cells of patients with breast cancer. *Int. J. Cancer*, 10.1002/ijc.26444.
8. Vanneste,E., Melotte,C., Voet,T., Robberecht,C., Debrock,S., Pexsters,A., Staessen,C., Tomassetti,C., Legius,E., D'Hooghe,T. *et al.* (2011) PGD for a complex chromosomal rearrangement by array comparative genomic hybridization. *Hum. Reprod.*, **26**, 941–949.
9. Alfarawati,S., Fragouli,E., Colls,P. and Wells,D. (2011) First births after preimplantation genetic diagnosis of structural chromosome abnormalities using comparative genomic hybridization and microarray analysis. *Hum. Reprod.*, **26**, 1560–1574.
10. Treff,N.R., Tao,X., Schillings,W.J., Bergh,P.A., Scott,R.T. Jr and Levy,B. (2011) Use of single nucleotide polymorphism microarrays to distinguish between balanced and normal chromosomes in embryos from a translocation carrier. *Fertil. Steril.*, **96**, e58–e65.
11. Johnson,D.S., Gemelos,G., Baner,J., Ryan,A., Cinnioglu,C., Banjevic,M., Ross,R., Alper,M., Barrett,B., Frederick,J. *et al.* (2010) Preclinical validation of a microarray method for full molecular karyotyping of blastomeres in a 24-h protocol. *Hum. Reprod.*, **25**, 1066–1075.
12. van Uum,C.M., Stevens,S.J., Dreesen,J.C., Drusedau,M., Smeets,H.J., Hollanders-Crombach,B., Die-Smulders,C.E., Geraedts,J.P., Engelen,J.J. and Coonen,E. (2012) SNP array-based copy number and genotype analyses for preimplantation genetic diagnosis of human unbalanced translocations. *Eur. J. Hum. Genet.*, **20**, 938–944, 10.1038/ejhg.2012.27.
13. Wang,J., Fan,H.C., Behr,B. and Quake,S.R. (2012) Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell*, **150**, 402–412.
14. Magli,M.C., Grugnetti,C., Castelletti,E., Paviglianiti,B., Ferraretti,A.P., Geraedts,J. and Gianaroli,L. (2012) Five chromosome segregation in polar bodies and the corresponding oocyte. *Reprod. Biomed. Online*, **24**, 331–338.
15. Handyside,A.H., Montag,M., Magli,M.C., Repping,S., Harper,J., Schmutzler,A., Vesela,K., Gianaroli,L. and Geraedts,J. (2012) Multiple meiotic errors caused by predivision of chromatids in women of advanced maternal age undergoing in vitro fertilisation. *Eur. J. Hum. Genet.*, **20**, 742–747.
16. Voet,T., Vanneste,E., Van der Aa,N., Melotte,C., Jackmaert,S., Vandendael,T., Declercq,M., Debrock,S., Fryns,J.P., Moreau,Y. *et al.* (2011) Breakage-fusion-bridge cycles leading to inv dup del occur in human cleavage stage embryos. *Hum. Mutat.*, **32**, 783–793.
17. Spits,C., Le Caignec,C., De Rycke,M., Van Haute,L., Van Steirteghem,A., Liebaers,I. and Sermon,K. (2006) Whole-genome multiple displacement amplification from single cells. *Nat. Protoc.*, **1**, 1965–1970.
18. Lasken,R.S. and Stockwell,T.B. (2007) Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC Biotechnol.*, **7**, 19.
19. Fiegler,H., Geigl,J.B., Langer,S., Rigler,D., Porter,K., Unger,K., Carter,N.P. and Speicher,M.R. (2007) High resolution array-CGH analysis of single cells. *Nucleic Acids Res.*, **35**, e15.
20. Geigl,J.B., Obenauf,A.C., Waldispuehl-Geigl,J., Hoffmann,E.M., Auer,M., Hormann,M., Fischer,M., Trajanoski,Z., Schenk,M.A., Baumbusch,L.O. *et al.* (2009) Identification of small gains and losses in single cells after whole genome amplification on tiling oligo arrays. *Nucleic Acids Res.*, **37**, e105.
21. Iwamoto,K., Bundo,M., Ueda,J., Nakano,Y., Ukai,W., Hashimoto,E., Saito,T. and Kato,T. (2007) Detection of chromosomal structural alterations in single cells by SNP arrays: a systematic survey of amplification bias and optimized workflow. *PloS One*, **2**, e1306.
22. Cheng,J., Vanneste,E., Konings,P., Voet,T., Vermeesch,J.R. and Moreau,Y. (2011) Single-cell copy number variation detection. *Genome Biol.*, **12**, R80.
23. Ampe,M., Verbeke,G., Vanneste,E. and Vermeesch,J.R. (2010) Analysis of array CGH data for the detection of single-cell chromosomal imbalances. *Online J. Bioinform.*, **11**, 224–244.
24. Konings,P., Vanneste,E., Jackmaert,S., Ampe,M., Verbeke,G., Moreau,Y., Vermeesch,J.R. and Voet,T. (2012) Microarray analysis of copy number variation in single cells. *Nat. Protoc.*, **7**, 281–310.
25. Ligasova,A., Raska,I. and Koberna,K. (2009) Organization of human replicon: singles or zipping couples? *J. Struct. Biol.*, **165**, 204–213.
26. Méndez,J. (2009) Temporal regulation of DNA replication in mammalian cells. *Crit. Rev. Biochem. Mol. Biol.*, **44**, 343–351.
27. Hansen,R.S., Thomas,S., Sandstrom,R., Canfield,T.K., Thurman,R.E., Weaver,M., Dorschner,M.O., Gartler,S.M. and Stamatoyannopoulos,J.A. (2010) Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl. Acad. Sci. USA*, **107**, 139–144.
28. Ryba,T., Hiratani,I., Lu,J., Itoh,M., Kulik,M., Zhang,J., Schulz,T.C., Robins,A.J., Dalton,S. and Gilbert,D.M. (2010) Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.*, **20**, 761–770.
29. Bechhoefer,J. and Rhind,N. (2012) Replication timing and its emergence from stochastic processes. *Trends Genet.*, 10.1016/j.tig.2012.03.011.
30. Woodfine,K., Fiegler,H., Beare,D.M., Collins,J.E., McCann,O.T., Young,B.D., Debernardi,S., Mott,R., Dunham,I. and Carter,N.P. (2004) Replication timing of the human genome. *Hum. Mol. Genet.*, **13**, 191–202.
31. Hiratani,I., Ryba,T., Itoh,M., Rathjen,J., Kulik,M., Papp,B., Fussner,E., Bazett-Jones,D.P., Plath,K., Dalton,S. *et al.* (2010) Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. *Genome Res.*, **20**, 155–169.
32. Ryba,T., Battaglia,D., Chang,B.H., Shirley,J.W., Buckley,Q., Pope,B.D., Devidas,M., Druker,B.J. and Gilbert,D.M. (2012) Abnormal developmental control of replication timing domains in pediatric acute lymphoblastic leukemia. *Genome Res.*, 10.1101/gr.138511.112.
33. State,M.W., Greally,J.M., Cuker,A., Bowers,P.N., Henegariu,O., Morgan,T.M., Gunel,M., DiLuna,M., King,R.A., Nelson,C. *et al.* (2003) Epigenetic abnormalities associated with a chromosome 18(q21-q22) inversion and a Gilles de la Tourette syndrome phenotype. *Proc. Natl. Acad. Sci. USA*, **100**, 4684–4689.
34. Smith,L., Plug,A. and Thayer,M. (2001) Delayed replication timing leads to delayed mitotic chromosome condensation and chromosomal instability of chromosome translocations. *Proc. Natl. Acad. Sci. USA*, **98**, 13300–13305.
35. D'Antoni,S., Mattina,T., Di Mare,P., Federico,C., Motta,S. and Saccone,S. (2004) Altered replication timing of the HIRA/Tuple1 locus in the DiGeorge and Velocardiofacial syndromes. *Gene*, **333**, 111–119.
36. Amiel,A., Elis,A., Blumenthal,D., Gaber,E., Fejgin,M.D., Dubinsky,R. and Lishner,M. (2001) Modified order of allelic replication in lymphoma patients at different disease stages. *Cancer Genet. Cytogenet.*, **125**, 156–160.
37. Hansen,R.S., Canfield,T.K., Fjeld,A.D., Mumm,S., Laird,C.D. and Gartler,S.M. (1997) A variable domain of delayed replication in FRAXA fragile X chromosomes: X inactivation-like spread of late replication. *Proc. Natl. Acad. Sci. USA*, **94**, 4587–4592.
38. Blow,J.J. and Gillespie,P.J. (2008) Replication licensing and cancer—a fatal entanglement? *Nat. Rev. Cancer*, **8**, 799–806.
39. Balikova,I., Martens,K., Melotte,C., Amyere,M., Van Vooren,S., Moreau,Y., Vetrie,D., Fiegler,H., Carter,N.P., Liehr,T. *et al.* (2008) Autosomal-dominant microtia linked to five tandem copies of a copy-number-variable region at chromosome 4p16. *Am. J. Hum. Genet.*, **82**, 181–187.
40. Ritchie,M.E., Silver,J., Oshlack,A., Holmes,M., Diyagama,D., Holloway,A. and Smyth,G.K. (2007) A comparison of background correction methods for two-colour microarrays. *Bioinformatics*, **23**, 2700–2707.

41. Smyth,G.K. and Speed,T. (2003) Normalization of cDNA microarray data. *Methods*, **31**, 265–273.
42. Smyth,G.K., Michaud,J. and Scott,H.S. (2005) Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics*, **21**, 2067–2075.
43. Smith,M.L., Marioni,J.C., Hardcastle,T.J. and Thorne,N.P. (2006) snapCGH: Segmentation, Normalization and Processing of aCGH Data Users' Guide. *Bioconductor*.
44. Olshen,A.B., Venkatraman,E.S., Lucito,R. and Wigler,M. (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*, **5**, 557–572.
45. van de Wiel,M.A., Kim,K.I., Vosse,S.J., van Wieringen,W.N., Wilting,S.M. and Ylstra,B. (2007) CGHcall: calling aberrations for array CGH tumor profiles. *Bioinformatics*, **23**, 892–894.
46. Nilsen,G., Liestol,K., Van Loo,P., Vollan,H.K., Eide,M.B., Rueda,O.M., Chin,S., Russel,R., Baumbusch,L.O., Caldas,C. *et al*. (2012) Copy number: Efficient algorithms for single- and multi-track copy number segmentation. *BMC Genomics*, **13**, 591.
47. Krzywinski,M., Schein,J., Birol,I., Connors,J., Gascoyne,R., Horsman,D., Jones,S.J. and Marra,M.A. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.*, **19**, 1639–1645.
48. Gilbert,N., Boyle,S., Fiegler,H., Woodfine,K., Carter,N.P. and Bickmore,W.A. (2004) Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell*, **118**, 555–566.
49. Farkash-Amar,S. and Simon,I. (2010) Genome-wide analysis of the replication program in mammals. *Chromosome Res.*, **18**, 115–125.
50. Lee,J.A., Carvalho,C.M. and Lupski,J.R. (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*, **131**, 1235–1247.
51. Zhang,F., Khajavi,M., Connolly,A.M., Towne,C.F., Batish,S.D. and Lupski,J.R. (2009) The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat. Genet.*, **41**, 849–853.
52. Crasta,K., Ganem,N.J., Dagher,R., Lantermann,A.B., Ivanova,E.V., Pan,Y., Nezi,L., Protopopov,A., Chowdhury,D. and Pellman,D. (2012) DNA breaks and chromosome pulverization from errors in mitosis. *Nature*, **482**, 53–58.
53. Azuara,V., Brown,K.E., Williams,R.R., Webb,N., Dillon,N., Festenstein,R., Buckle,V., Merkenschlager,M. and Fisher,A.G. (2003) Heritable gene silencing in lymphocytes delays chromatid resolution without affecting the timing of DNA replication. *Nat. Cell. Biol.*, **5**, 668–674.
54. Tuduri,S., Tourriere,H. and Pasero,P. (2010) Defining replication origin efficiency using DNA fiber assays. *Chromosome Res.*, **18**, 91–102.
55. Schultz,S.S., Desbordes,S.C., Du,Z., Kosiyatrakul,S., Lipchina,I., Studer,L. and Schildkraut,C.L. (2010) Single-molecule analysis reveals changes in the DNA replication program for the POU5F1 locus upon human embryonic stem cell differentiation. *Mol. Cell. Biol.*, **30**, 4521–4534.
56. Gilbert,D.M. (2010) Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat. Rev. Genet.*, **11**, 673–684.