Scientific article

Genomic analysis of individual cells by NGS and real-time PCR

Andreas Meier, Evelyn Fricke, Evelyn Fisch, Silke Baedker, Holger Wedler, and Christian Korfhage

QIAGEN GmbH, Qiagen Straße 1, 40724 Hilden, Germany

Abstract: Genomic analysis of single cells by next-generation sequencing (NGS) has become increasingly important in research fields such as the study of cancer, aging, pre-implantation diagnostics, and metagenomics. Here, we describe the analysis of the genome of single cells by NGS. Since the amount of DNA within a single cell is very limited, accurate whole genome amplification (WGA) of genomic DNA is required. We explore the reliability of QIAGEN's REPLI-g® Single Cell WGA technology and its application in NGS with regards to genome coverage, amplification of non-target sequences, sequence bias, and error rates. In contrast to PCR-based WGA methods, REPLI-g Single Cell WGA uses an optimized formulation of Phi 29 DNA polymerase with high proofreading activity and strong processivity, which ensures a reduced level of errors and bias in NGS of single-cell genomes.

Introduction

Cell heterogeneity plays a central role in biological phenomena during normal development or disease (e.g., cancer development or aging). As gene regulation is a fundamental process, the genomic analysis of single cells to dissect phenotypic variability is of key interest to scientists. Deep genome analysis of small biological samples using NGS, microarrays, or real-time PCR is often limited by the small amount of sample available (6 pg gDNA/human cell). Typically, 1 ng to 1 µg of DNA is necessary for genomic analysis, which requires the DNA of approximately 10^2 to 10^5 cells (mammalian genome) or 10^5 to 10^9 cells (bacterial genome). Thus, an average genome that is common to all of the hundreds to million cells from which the DNA has been isolated is analyzed and cell-to-cell genome variation is neglected. Genomic variations from cell to cell within a sample can be discovered only if the individual genome of a very small cell number down to one single cell is analyzed.

There are two primary reasons why genomes of single cells are analyzed.

Some samples comprise of a very small number of cells and cannot be used for whole genome analysis. One example is preimplantation genetic diagnostics (PGD) research, which involves analysis of a few cells, often dissected from the trophoblast. Trophoblast cells are used for the analysis of chromosomal aberrations (e.g., aneuploidies) or monogenetic diseases. Another example is the study of cancer. In recent years, it has become apparent that tumors comprise of multiple cell clones with varying genome structures. Cancer is caused by a few genomic changes that result in alterations to cell programming, leading to massive cell divisions (11, 18). More and more genomic changes are enriched during tumor growth, which means that the genomic structure of individual cells starts to vary from that of non-tumor cells. The spectrum of genomic changes is very broad and includes single base mutations, insertions, deletions, translocations, copy number variations, and chromosome loss or gain. ▶

Contents

Certain sample comprise of a mixture of different organisms that can only be reliably differentiated by genome analysis (e.g., mixed sample of microorganisms from the environment, biofilms, or wastewater). Generally, microorganisms are often differentiated by selective media. In most environmental samples, only 1% of the microbial species can be grown in pure culture (6, 8). However, identification of microorganisms can be performed by single-cell metagenomic analysis to determine the heterogeneity of the sample.

Consequently, accurate WGA is required for reliable genetic analysis (e.g., NGS) when amounts of genomic DNA are limited, as in the case of single cell DNA (4, 5, 8, 9). Incomplete or biased genome amplification with missing or underestimated sequence information is a common limitation when analyzing single cells, in particular, by single-cell sequencing. To overcome such bottlenecks, we have developed an easy-to-use single-cell WGA method. This method is based on isothermal Multiple Displacement Amplification (MDA) and comprises of an innovative lysis process and use of an optimized form of Phi 29 DNA polymerase. Overall, our new method results in effective lysis of cells, complete DNA denaturation, and reliable amplification of the whole genome of a single cell with high accuracy and minimal amplification bias. To demonstrate the robustness of this new method in single-cell WGA, we amplified DNA from a variety of human and bacterial single cells, and verified the resulting genome coverage by NGS and qPCR. In this article, we describe this optimized method and evaluate our experimental results with regards to cell-to-cell variations, GC content in comparison to genomic DNA, percentage of genome coverage with respective error rates, and genome-wide real-time PCR analysis.

Experimental protocol

Cell culture

Bacillus subtilis 168 cells were grown in YT medium (Sigma Y2377) at 37°C overnight. The overnight culture was freshly diluted in YT medium at a concentration of 0.1 OD_{600}. After 2 h at 37°C, the OD of the cell culture was measured again at 600 nm. The culture was diluted in 0.9% NaCl solution/20% glycerol to various storage concentrations between 50 and 1000 cells/µl, or was pelleted for genomic DNA preparation. For purification of genomic DNA, we used the QIAamp® DNA Mini Kit, following the procedure as described in the kit handbook.

For HeLa cell culture, we cultivated 4 x 10^4 Hela S3 cells for 3 days in 15 ml DMEM (1% FCS; 0.1% non-essential amino acids; 0.1% Penicillin/Streptavidin). Following that, the cell medium was removed, cells were washed once with 10 ml of 1x PBS (pH 7.4), and then treated for 5 min at 37°C with 1 ml Trypsin/EDTA solution. The treatment was stopped by adding 10 ml DMEM. The cell suspension was transferred to a 50 ml polypropylene tube and was pelleted by centrifugation for 5 min at 300 x *g*. After resuspension in PBS (pH 7.4), the cell concentration was determined after mixing 10 µl cell suspension with 10 µl Trypan blue (0.4%) using the Neubauer cell counting chamber. On average, we determined a cell concentration of 4 x 10^5/ml. Cells were diluted to the appropriate concentration with 1x PBS, aliquoted, and frozen in liquid nitrogen.

Whole genome amplification

For WGA, cells were further diluted in PBS so that the intended cell number was suspended in a 4 µl volume. WGA was performed using the REPLI-g Single Cell Kit (QIAGEN) according to the protocol in the kit handbook. Briefly, 4 µl cell suspensions were incubated with 3 µl Buffer D2 for 10 min at 65°C. After stopping the reaction by adding 3 µl Stop Solution, a master mix comprising H20 sc, REPLI-g sc Reaction Buffer, and REPLI-g sc DNA Polymerase was added. The reaction was performed for 3 h at 30°C. The concentration of the DNA after WGA was determined using the Quant-iT™ PicoGreen® dsDNA Reagent (Life Technologies, cat. no. P7581), according to the manufacturer's instructions.

Real-time PCR

Real-time PCR was performed on 1 ng genomic DNA or WGA DNA per PCR reaction using RT² Profiler PCR Arrays (QIAGEN), as described in the kit handbook. RT² Profiler PCR Arrays were pipetted using the QIAgility® instrument (QIAGEN). Alternatively, the QuantiTect® SYBR® Green PCR Kit (QIAGEN) was used and reactions were performed using indicated primers sets as per the instructions in the kit handbook.

Sequencing

For DNA fragmentation, we used 2 µg genomic DNA or WGA DNA (without subsequent purification of WGA DNA) and the Covaris® S220 instrument. After fragmentation, DNA was purified using the MinElute® PCR Purification Kit (QIAGEN) according to the protocol in the kit handbook. Library preparation

was performed using either the GeneRead™ Library Prep Kit (I) (QIAGEN), or alternatively, the TrueSeq™ DNA Sample Prep Kit (Illumina), according to the manufacturer's instructions. The NGS library was purified using the GeneRead Size Selection Kit (QIAGEN) as per the instructions in the kit handbook. We performed 150 bp paired-end sequencing on an Illumina® MiSeq® instrument. For reference mapping, we used the complete genome data of *Bacillus subtilis* P168. Read alignment was performed using Bowtie2.

Results and discussion

The REPLI-g Single Cell method is a two-step process comprising of cell lysis/denaturation and DNA amplification. The process starts with a gentle alkaline lysis and denaturation step. During this step, the cell membrane is broken up, DNA is released from the cell, and all protein-DNA interactions (e.g., chromatin) are resolved to prevent inhibition of DNA amplification. In contrast to heat denaturation, the gentle alkaline denaturation provided by REPLI-g Single Cell Kit prevents DNA fragmentation or generation of abasic sites. After cell lysis and DNA denaturation, amplification reagents (including primer, dNTPs, and optimized Phi 29 DNA polymerase) were added and whole genome amplification was started at 30°C. The process of REPLI-g Single Cell WGA is based on MDA technology (3, 4, 13). Random primers are hybridized to the single-stranded DNA and are extended at 30°C by the proofreading Phi 29 DNA polymerase. The DNA polymerase migrates along the DNA strand and displaces the complementary strand. The displaced strand becomes a template itself for replication (Figure 1). Due to the enzyme's high processivity, MDA products deliver long DNA strands of up to 70 kb. In contrast to PCR-based WGA (15, 19), MDA results in WGA products with up to 1000-fold lower error rates (10^{-5} to 10^{-6} errors per nucleotide; [4, 13]). Especially at hairpin structures, *Taq* DNA Polymerase tends to delete short stretches of DNA sequences. Due to its strong strand displacement activity, Phi 29 DNA polymerase is able to resolve hairpin structures and displace the complementary strand (Figure 1) (2, 17). Consequently, in regions of highest *Taq* DNA Polymerase error rate (e.g., hairpin structures), Phi 29 DNA polymerase does not fail (12, 17).

To study the sensitivity and performance of REPLI-g Single Cell WGA, we used a dilution of intact cells of the Gram+ bacterium, *Bacillus subtilis* for WGA. On average, every WGA reaction

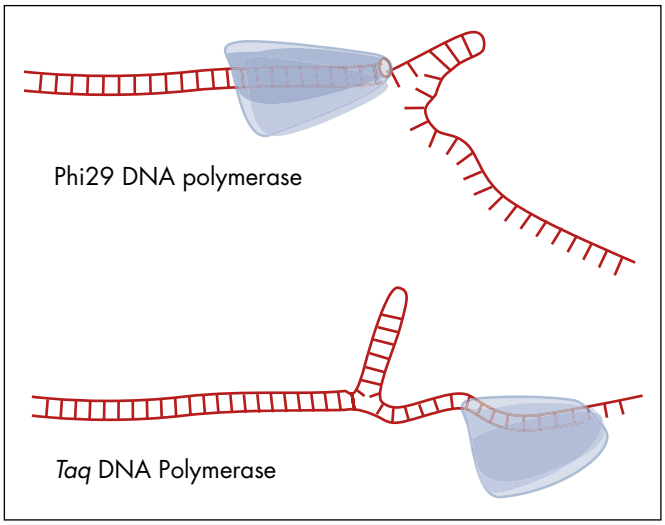

Figure 1. Phi 29 DNA polymerase can resolve hairpin structures and displace the complementary strand. Upon encountering secondary DNA structures, *Taq* DNA Polymerase may pause DNA synthesis, slip, or dissociate from the template. This can result in inaccurate DNA amplification, incomplete loci coverage, and short fragment sizes. REPLI-g Single Cell WGA utilizes Phi 29 DNA polymerase, which displaces secondary structures, enabling accurate and highly uniform amplification of the entire genome.

comprised of 2–5 cells (according to Poisson distribution) corresponding to 4.5 fg/cell. After amplification for 3 h at 30°C, approximately 30 µg of DNA was amplified, 2 µg of which was used (without additional purification after WGA) for NGS. We used the same process for sheared DNA and for NGS libraries prepared using the GeneRead Library Prep Kit (I) (QIAGEN) or the TrueSeq DNA Sample Prep Kit (Illumina) as that used for non-amplified genomic DNA isolated from bacteria. We analyzed the libraries generated from the WGA DNA amplified from a few bacterial cells by 150 bp pair-end sequencing using the Illumina MiSeq instrument. We used the read aligner, Bowtie2, to map qualified reads to the reference genome. We observed almost complete genome coverage even when using just a single cell. Up to 99.22% genome coverage was achieved, with coverage of >10 fold (Table 1). Using non-amplified DNA, up to 99.63% genome coverage was achieved, with coverage of >10 fold (Table 1). Up 99.22% of all reads generated from REPLI-g Single Cell WGA DNA could be mapped to the reference *Bacillus subtilis* genome.

We conclude from these results that no side products were generated during WGA as >99% of all reads were mapped to the reference genome, and the extent of genome coverage was similar to non-amplified genomic DNA. ▶

Table 1. NGS runs using non-amplified genomic DNA and REPLI-g Single Cell amplified DNA

Parameter	Non-amplified gDNA	REPLI-g amplified DNA
Number of *Bacillus subtilis* cells used for NGS	4.5 x 10^8 cells	~ 2–5 cells
Genome coverage (coverage >10-fold)	99.96%	99.22%
Error rate	0.21%	0.23%
Indel rate	0.01%	0.01%
Chimera rate	0.18%	0.26%

To identify bias during amplification or artifacts generated by the MDA process, we computed error rates, chimeras, and GC content compared to the expected ratio by Picard metrics. We determined GC bias by fragment counts of 100 bp bins characterized by the fraction of G and C bases (GC%) according to the reference genome. The resulting GC curve was compared to the predicted curve for each bin based on the bin's GC%. Fragment counts were normalized to the expected number of bins giving the fraction of normalized coverage (Figure 2).

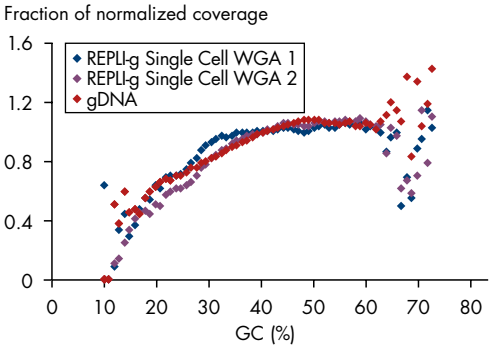

Figure 2. The REPLI-g Single Cell Kit ensures minimal bias, even for samples with high GC content. GC% bias was compared in NGS runs using non-amplified genomic DNA of approximately 150,000 *Bacillus subtilis* cells (red) and two independent REPLI-g Single Cell WGA DNA sets starting from a few cells of *Bacillus subtilis* (purple and blue squares). The x-axis describes the GC content (%) of 100 bp bins of the *Bacillus subtilis* genome and the y-axis shows the number of reads obtained in NGS, related to the number of expected reads. The purple and blue curves indicate the sequencing result of the genome from a few cells amplified by the REPLI-g Single Cell Kit. The consistency of all three curves shows that the REPLI-g Single Cell WGA method results in unbiased amplification, even for samples with high GC content.

Only minor differences were detected between curves obtained from genomic DNA sequencing and sequencing of WGA DNA. This indicates that only a minor GC bias is generated during REPLI-g amplification from a few bacterial cells (representing femtograms of DNA). Additionally, the %GC distributions of the sequencing replicates were correlated using the coefficient of determination, R^2. There was only minor variation in %GC distribution between replicate libraries from REPLI-g Single Cell WGA products, independent of whether the library was prepared using the GeneRead Library Prep Kit (I) from QIAGEN or using the Illumina's TrueSeq DNA Sample Prep Kit. The replicates had pairwise correlation values of 0.9977 (+/–0.0017), indicating low variability of independent REPLI-g Single Cell WGA reactions. Based on the data obtained, we can conclude that REPLI-g Single Cell WGA results in high reproducibility with regards to low GC bias during WGA.

To determine if amplification artifacts were produced, we determined the fraction of high-quality errors, insertion, deletions, or chimeras generated by the WGA method (Table 1). Due to the proofreading activity of Phi 29 DNA polymerase, a significant increase in error rate was not observed. While non-amplified DNA results in an error rate of about 0.21%, the error rate of REPLI-g amplified DNA generated from just a few cells was 0.23%. No difference with regards to insertion or deletions of bases was detected between the non-amplified DNA sample and the WGA DNA sample. Chimeric sequences were assumed to be generated primarily during ligation or the PCR step during NGS library amplification. For example, chimeric products are generated during PCR by incomplete extension. The single-stranded truncation product may hybridize nonspecifically to a foreign DNA strand and be extended to a double-stranded product, which in turn is amplified during the following PCR cycles (10, 12, 14). However, a study by Lasken and Stockwell assumes that MDA tends also to generate chimeric DNA rearrangements in amplified DNA (7). We could not find a significant increase in chimera rate since the chimera rate of amplified DNA is 0.26%, which is similar to what we detected for non-amplified DNA. Summarizing the data, REPLI-g Single Cell WGA does not significantly introduce sequence artifacts such as base exchanges, insertions, deletions, or chimeric sequences.

We tested REPLI-g Single Cell WGA reactions also on the more complex human genome. For this application, HeLa cells were diluted from 1000 to 3 cells per reaction. Cells were lysed and DNA was amplified using reagents from the REPLI-g Single Cell Kit. The yield was slightly higher (>25 µg/WGA) for high cell numbers (>10 cells) compared to samples comprising of just 3 cells (20 µg/WGA). After WGA, DNA was diluted and 1 ng of WGA DNA was used for RT² Profiler PCR Arrays (QIAGEN). The WGA results for samples containing 3 cells are shown in Figure 3. All genes analyzed by RT² Profiler PCR Arrays are well represented after WGA of 3 cells. The reproducibility is high, as indicated by the C_T value error bars.

Since RT² Profiler PCR Arrays are designed to cover exonic sequences, GC bias of the REPLI-g Single Cell WGA method was tested by correlating the C_T value during real-time PCR with the exonic GC content of the gene. A correlation of the C_T value and GC content would indicate if the REPLI-g Single Cell WGA method introduced GC bias during amplification of the human genome. Since no correlation between the C_T value of WGA DNA and the GC content of the exonic gene sequence was observed (Figure 4), we conclude that GC bias is not introduced by REPLI-g amplification of the complex human genome from single cells. ▶

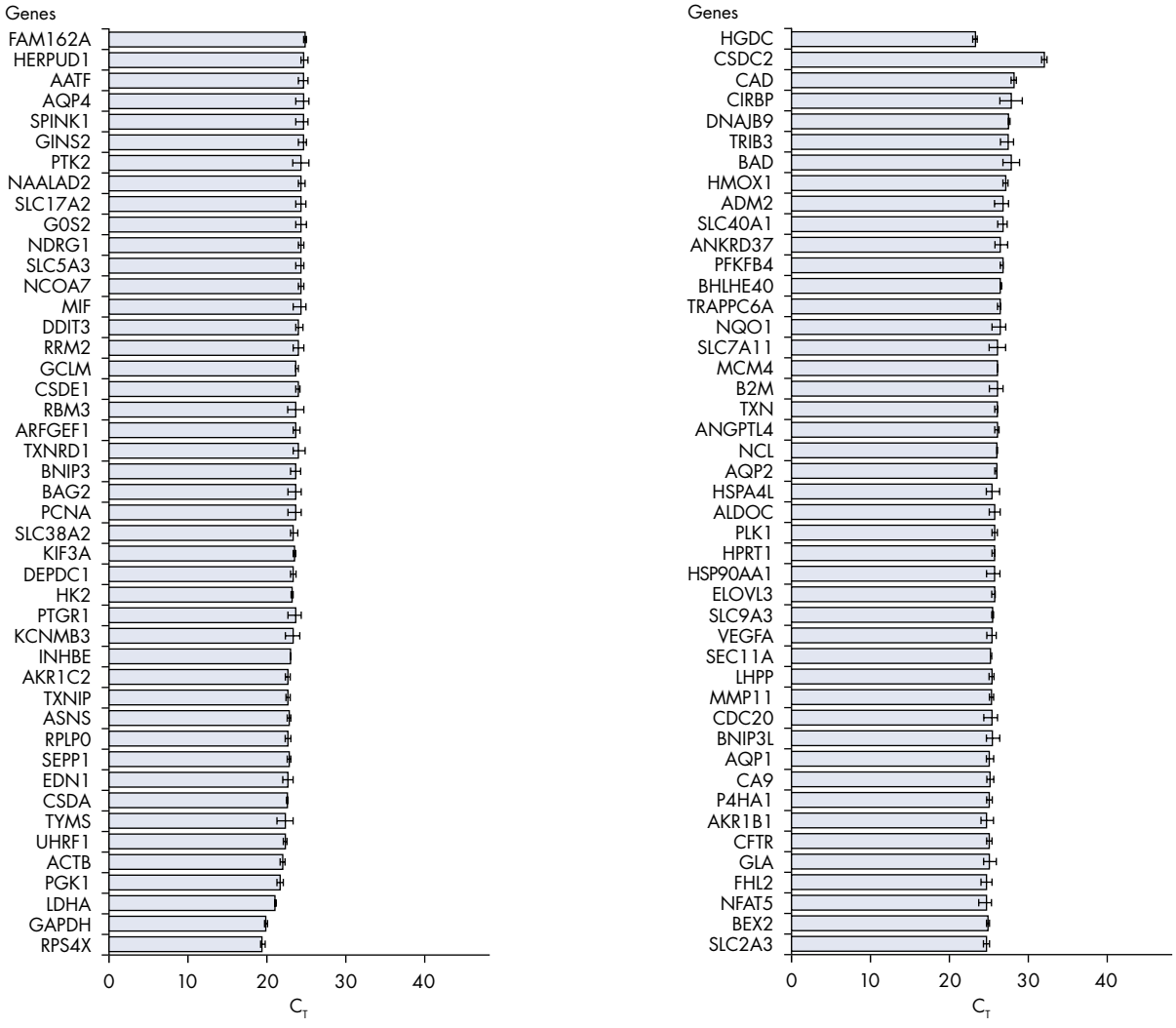

Figure 3. Complete genome coverage. HeLa cells (3 cells on average) were used for individual REPLI-g Single Cell reactions. After WGA, real-time PCR analysis of 88 loci across the entire genome was performed using 10 ng of WGA DNA for each primer assay. Low and consistent C_T values were observed in real-time PCR for all loci with no dropout from any marker, indicating that the DNA was successfully amplified from all areas of the genome and is highly suited for single-cell genomics.

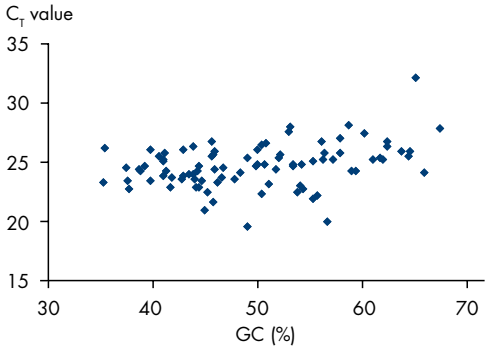

Figure 4. Minimal bias. GC bias after REPLI-g WGA of 3 human cells was calculated from the experiment shown in figure 3. GC content was calculated from exon regions of the gene and was plotted against the C_T value obtained in real-time PCR after WGA.

Conclusion

Starting from a few bacterial cells, genomic DNA was amplified in a simple isothermal whole genome amplification process using the REPLI-g Single Cell Kit (QIAGEN). WGA DNA was analyzed by NGS to obtain deep information about the completeness, coverage, bias and errors during the whole genome amplification process. In an alternative approach, genomic DNA from single cells of human origin was also amplified.

It is obvious that the lower the number of genomes as starting material, the higher the amplification factor, and therefore the higher the likelihood of amplification bias and misincorporation of bases. To obtain the highest amplification rates and evaluate the rates of deviation (e.g., bias, errors, deletion, insertions, etc.) introduced by the amplification process, we chose bacterial cells as a target system. Starting from bacterial cells, whole genome amplification generates approximately 30 µg DNA, accounting for an amplification factor of >10^9. We conclude from NGS analysis of WGA DNA that:

- REPLI-g Single Cell WGA results in high reproducibility
- Genome coverage is similar in extent to non-amplified genomic DNA
- Negligible GC bias is introduced during REPLI-g Single Cell WGA
- REPLI-g Single Cell WGA does not significantly introduce sequence artifacts such as base exchanges, insertions, deletions, or chimeric sequences
- No artifacts are generated during WGA since >99% of all reads could be mapped to the reference genome

References

1. Blanco, L., Bernad, A., Lázaro, J.M., Martín, G., Garmendia, C., and Salas, M. (1989) Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. J. Biol. Chem. **264**, 8935.
2. Canceill, D., Viguera, E., and Ehrlich, S.D. (1999) Replication slippage of different DNA polymerases is inversely related to their strand displacement efficiency. J. Biol. Chem. **274**, 27481.
3. Dean, F.B. et al. (2002) Comprehensive human genome amplification using multiple displacement amplification. Proc. Natl. Acad. Sci. USA **99**, 5261.
4. Hosono, S. et. al. (2003) Unbiased whole-genome amplification directly from clinical samples. Genome Res. **13**, 954.
5. Hou, Y. et al. (2012) Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell **148**, 873.
6. Kalisky, T., Blainey, P., and Quake, S.R. (2011) Genomic analysis at the single-cell level. Annu. Rev. Genet. **45**, 431.
7. Lasken, R.S. and Stockwell, T.B. (2007) Mechanism of chimera formation during the Multiple Displacement Amplification reaction. BMC Biotechnology **7**, 19.
8. Marcy, Y. et. al. (2007) Dissecting biological ''dark matter'' with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. Proc. Natl. Acad. Sci. USA **104**, 11889.
9. Melnikov, A. et. al. (2011) Hybrid selection for sequencing pathogen genomes from clinical samples. Genome Biol. **201**, R73.
10. Meyerhans, A., Vartanian, J. P., and Wain-Hobson, S. (1990) DNA recombination during PCR. Nucleic Acids Res., **18**, 1687.
11. Navin, N. et. al. (2010) Inferring tumor progression from genomic heterogeneity. Genome Res. **20**, 68.
12. Odelberg, S.J., Weiss, R.B., Hata, A., and White R. (1995) Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase. Nucleic Acids Res. **23**, 2049.
13. Paez, J.G. et. al. (2004) Genome coverage and sequence fidelity of ϕ29 polymerase-based multiple strand displacement whole genome amplification. Nucleic Acids Research **32**, e71
14. Paabo, S., Irwin, D.M. and Wilson, A.C. (1990) DNA damage promotes jumping between templates during enzymatic amplification. J. Biol. Chem., **265**, 4718.
15. Telenius, H., Carter, N.P., Bebb, C.E., Nordenskjöld, M., Ponder, B.A., and Tunnacliffe, A. (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics **13**, 718.
16. Van Ooyen, S., Loeffert, D., and Korfhage, C. (2011) Overcoming constraints of genomic DNA isolated from paraffin-embedded tissue. www.qiagen.com/Products/Catalog/Sample-Technologies/DNA-Sample-Technologies/Genomic-DNA/REPLI-g-FFPE-Kit#resources
17. Viguera, E., Canceill, D., and Ehrlich, S.D. (2001) Replication slippage involves DNA polymerase pausing and dissociation. EMBO J. **20**, 2587.
18. Weinberg, R. (2007). The Biology of Cancer. New York: Garland Science.
19. Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., and Arnheim, N. (1992) Whole genome amplification from a single cell: implications for genetic analysis. Proc. Natl. Acad. Sci. USA **89**, 5847.

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit handbook or user manual. QIAGEN kit handbooks and user manuals are available at www.qiagen.com or can be requested from QIAGEN Technical Services or your local distributor.

Visit www.qiagen.com/single-cell-analysis for more information.

www.qiagen.com

Australia ■ 1-800-243-800
Austria ■ 0800-281011
Belgium ■ 0800-79612
Brazil ■ 0800-557779
Canada ■ 800-572-9613
China ■ 800-988-0325
Denmark ■ 80-885945
Finland ■ 0800-914416
France ■ 01-60-920-930
Germany ■ 02103-29-12000
Hong Kong ■ 800 933 965
India ■ 1-800-102-4114
Ireland ■ 1800 555 049
Italy ■ 800-787980
Japan ■ 03-6890-7300
Korea (South) ■ 080-000-7145
Luxembourg ■ 8002 2076
Malaysia ■ 603-7981-5510
Mexico ■ 01-800-7742-436
The Netherlands ■ 0800-0229592
Norway ■ 800-18859
Singapore ■ 1800-742-4368
Spain ■ 91-630-7050
Sweden ■ 020-790282
Switzerland ■ 055-254-22-11
Taiwan ■ 0800-665-957
UK ■ 0808-234-3665
USA ■ 800-426-8157