Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq

Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq

Iain C. Macaulay[1], Mabel J. Teng[2-3], Wilfried Haerty[1], Parveen Kumar[5], Chris P. Ponting[2,4], Thierry Voet[2,5]

Affiliations:

[1]The Genome Analysis Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK

[2]Sanger Institute–EBI Single–Cell Genomics Centre, Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK

[3]Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK

[4]MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU

[5]Department of Human Genetics, University of Leuven, KU Leuven, Leuven, 3000 Belgium

Keywords: Single-cell, genomics, transcriptomics, sequencing, multi-omics

Key references:

G&T-seq: parallel sequencing of single-cell genomes and transcriptomes.

Macaulay IC et al., Nat Methods. 2015 Jun;12(6):519-22.

Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity.

Angermueller C et al., Nat Methods. 2016 Mar;13(3):229-32

ABSTRACT:

Parallel sequencing of a single cell's genome and transcriptome provides a powerful tool to dissect genetic variation and its relationship to gene expression. Here we present a detailed protocol for G&T–seq, a method for separating and sequencing genomic DNA and full-length polyA(+) mRNA from single cells in parallel. We describe step-by-step the isolation and lysis of single cells, the physical separation of polyA(+) mRNA from genomic DNA using a modified oligo–dT bead capture and the respective whole-transcriptome and whole-genome amplifications, and finally the library preparation and sequence analyses of these amplification products. The method allows the detection of thousands of transcripts in parallel with the genetic variants captured by the DNA-seq data of the same single cell. G&T-seq differs from other currently available methods for parallel DNA and RNA sequencing from single cells as it involves physical separation of the DNA and RNA and does not require bespoke microfluidics platforms. The process can be implemented manually or with automation. When performed manually, paired genome and transcriptome sequencing libraries from 8 cells can be produced in approximately 3 days by researchers experienced in molecular laboratory work. For users with experience in the programming and operation of liquid handling robots, paired DNA and RNA libraries from 96 single cells can be produced in the same time frame. Analysis and integration of single cell G&T-seq data requires a high level of bioinformatic ability and familiarity with a wide range of informatics tools.

INTRODUCTION

The study of the genomes or transcriptomes of single cells continues to highlight the extent, nature and role of the cellular heterogeneity that arises in organisms in health and disease [1-3]. Advances in whole-genome amplification (WGA) have allowed diverse aspects of single-cell genomes to be analysed, including DNA copy number variants (CNV) [4,5], structural variants (SV) [6-8] and single nucleotide variants (SNV) [5,9-11]. WGA is currently performed by Multiple Displacement Amplification (MDA), Polymerase Chain Reaction (PCR), or a combination of displacement pre-amplification and PCR (DA-PCR). Each WGA method has its characteristic amplification artefacts, offering different resolution across the whole spectrum of genetic variants.[12-14] MDA is often the method of choice for genotyping or discovery of SNVs in single cells as it offers the widest breath of coverage across the whole genome with high fidelity stemming from the strong proof-reading capacity of phi29 polymerase [5,6,9-11,15,16]. In contrast, PCR- and DA-PCR based WGA (e.g. PicoPLEX [17] or MALBAC [18]) have less amplification bias and lower fidelity and therefore they are usually better suited for single-cell DNA copy number profiling [12-

[14]. In parallel, there are a variety of methods capable of exploring a single cell transcriptome by sequencing. Using whole-transcriptome amplification (WTA) of reverse transcribed mRNA-molecules, an increasing diversity of methods are capable of exploring the single-cell transcriptome by sequencing[1,2,19]. These methods allow either high-throughput tag sequencing of the 3' or 5' ends of mRNA [20-22] or more medium-throughput sequencing of full-length transcripts [23-27]. For instance, the Smart-seq2 method [24,25] uses template-switching to generate first strand cDNA molecules of full-length transcripts with adaptor sequences at both ends. These universal adaptor sequences are then used to prime PCR amplification of the transcriptome, and full-length cDNA PCR amplicons are used as input for sequence library preparation by tagmentation, enabling single cell mRNA-seq.

However, in the methods described above only the genome or the transcriptome can be analysed, but not both from the same single cell. Hence, it was previously not possible to correlate changes in a cell's genome with those in its transcriptome.

We recently developed G&T-seq, a method that allows parallel sequencing of the genome and transcriptome of a single cell [17]. We demonstrated that the method can robustly generate full-length transcriptome data and genomic DNA sequences of the same cell. Here we present a detailed protocol for the G&T-seq method, which can be implemented either manually or on automated liquid handling platforms depending on the desired throughput.

Development and overview of the procedure

The method (Fig. 1 and Suppl. Fig. 1) was developed as a means to analyse in parallel the genomes and transcriptomes of single cells. Nevertheless, we have also successfully applied G&T-seq to larger numbers of pooled cells (10-100) thereby allowing small populations of rare cells to be analysed as well. We specifically devised the G&T-seq method to be readily automatable on robotic liquid handling platforms which are readily available in the majority of genomics laboratories, and using off-the-shelf reagents to allow implementation without custom technical development. The adaptation and combination of existing protocols into the G&T-seq protocol has also allowed existing data analysis approaches for single cell DNA and RNA sequencing to be directly applied to G&T-seq data with little or no modification.

Automation allows higher throughput processing (10s to 100s of single cells) and fluorescence activated cell sorting (FACS) is an efficient means by which single cells can be isolated in 96-well plates. Furthermore, FACS offers the capability of selecting very rare cells based on the expression of cell surface markers[1,19]. However, FACS is not suitable for all applications, and manual isolation is

preferable when only a small starting population of single cells is available for collection (e.g. if all individual blastomeres from an 8-cell cleavage stage embryo are to be collected)[19,28].

After deposition of the single cells into the lysis buffer, the 96-well plates should immediately be sealed, centrifuged and stored at –80°C until it is convenient to process them. We chose to use a guanidine isothiocyanate and detergent based lysis buffer, which lyses the isolated cell and its nucleus, to maximise availability of both RNA and genomic DNA (gDNA) into solution while still remaining compatible with the subsequent separation step. Magnetic beads coated with a modified version of the tailed oligo-dT primer from the Smart-seq2 protocol[24,25] are then added to capture the polyA(+) mRNA-molecules from the lysis buffer. After mixing and magnetic precipitation of the beads in the lysate, the supernatant containing the gDNA is collected and transferred to a new plate. A key challenge when physically separating RNA and DNA from the same cell is the possibility of losing material during this process, and so to maximise transfer of all gDNA, the bead-bound polyA(+) mRNA is washed thoroughly but carefully. After each wash the supernatant wash buffer is collected and added to the gDNA-containing cell lysate present in the new plate. To further minimise loss of gDNA, the same tips are used for all transfer steps and the tips are washed after the last transfer; this final wash is added to the pool of wash buffer and gDNA-containing cell lysate.

Following removal of the last wash buffer from the polyA(+) mRNA loaded beads, the reverse transcription (RT) mastermix is added. The RT reaction is similar to the Smart-seq2 protocol[24,25], with the exceptions that no denaturing step is performed before RT and the RT reaction is performed with constant mixing to prevent sedimentation of the beads to which the polyA(+) mRNA molecules are bound. We observed that the transcriptome sequences obtained from single cells using G&T-seq were comparable to those generated by Smart-seq2 in terms of the numbers of transcripts detected, full-length transcript coverage, GC content distribution of transcripts, and the detection of spike-in RNA molecules[17]. These similarities indicate that no additional biases are introduced as a result of the physical separation of polyA(+) mRNA from the cell lysate.

The gDNA in solution of the pool of polyA(+) mRNA-depleted cell lysate and all wash buffer is first concentrated to allow downstream WGA and library preparation. To this end, a Solid Phase Reversible Immobilization (SPRI) bead based concentration is performed, after which the purified single-cell genome is resuspended in a suitable buffer for WGA. The SPRI bead concentration of DNA is also undertaken with considerable care to minimise loss of material at this stage.

The G&T-seq method is compatible with various WGA methods: we have successfully applied PicoPLEX, MDA and MALBAC protocols on G&T-seq isolated DNA. The choice of WGA method is

dependent on the desired readout of the experiment [12-14]. In our hands, PicoPLEX is preferred when analysing the cell's gDNA for copy number variants, whereas we use MDA for detecting SNVs or SVs [6,17].

Following parallel whole-genome and whole-transcriptome amplification, each original single cell will generate separated amplified gDNA and cDNA samples. Both are suitable as input for tagmentation based library preparation, such as Illumina's Nextera XT protocol, which offers an efficient and convenient means to rapidly produce multiplexed library pools from 96 single cells [17,25]. These allow sequencing of the polyA(+) mRNA-derived cDNA in parallel with the amplified gDNA for the study of gene expression and genetic variants, respectively. However, if whole-genome or targeted DNA-sequencing is to be performed, conventional adaptor-ligation based library preparation approaches are most often used[5,9-11,17], so as to preserve as much complexity as possible from the input material.

We have previously shown that by sequencing the genome and transcriptome of a single cell in parallel, G&T-seq can readily distinguish the transcriptional consequences of chromosomal aneuploidies and interchromosomal fusions in a cell, and can as well detect coding SNVs at the single-cell genome and transcriptome level [17].

Comparison with other methods for DNA- and RNA-seq of the same single cell

An alternative method for parallel DNA and RNA sequencing from a single cell (DR-seq) was recently described by Dey et al. [29] and has also successfully been applied to investigate the relationship between DNA copy number and gene expression dosage. DR-seq differs from G&T-seq in two respects. First, that there is no physical separation of gDNA and polyA(+) mRNA prior to amplification, which may have the potential to minimise losses which could occur in G&T-seq when the gDNA is transferred to a separate tube for processing, and second in its ability to perform the reaction in a single tube which may make the procedure more amenable to transfer into droplet-based microfluidic formats. However, DR-seq utilises a modification of the CEL-seq protocol for WTA [20,29], which selectively targets the 3' ends of transcripts, meaning that the full length of the transcript cannot be sequenced, and thus splicing variants, fusion transcripts, and the majority of expressed SNVs cannot be detected. Additionally, the WGA component of the DR-seq protocol uses a modification of the MALBAC approach, whereas G&T-seq is a more open platform, allowing WGA to be performed using any available method and thus a choice of a WGA method that is optimal for addressing the research question [12,17]. Furthermore, because DR-seq amplifies DNA and mRNA without physical separation, it requires *in silico* masking of the exons in the genome to determine DNA copy-number variation. In contrast, Li et al. [30] have also demonstrated the physical separation of DNA and RNA from a single cell,

and examined SNVs in both the exome and the transcriptome. While such observations can also be made using the G&T-seq protocol, they did not demonstrate the feasibility of detecting copy number variants and structural variants in the DNA with the respective gene dosage expression and fusion transcript in the RNA of the same cell. Finally, microfluidic separation and sequencing of DNA and RNA from the same single cell has also been demonstrated[31]. This method employs a custom-built microfluidics circuit to capture single cells, lyse their membranes and thus release cytoplasm (containing mRNA) and the nucleus (containing gDNA) for separate capture and amplification. While the method has thus far only been applied for targeted sequencing and PCR analysis of DNA and RNA from the same single cell, such microfluidic approaches offer the opportunity to image the captured cell, to miniaturise reaction volumes and potentially to operate at great scale. However, none of the above methods has been demonstrated to be amenable to automation for high throughput processing, which is an essential component of most single-cell based studies. A key aim in the development of the G&T-seq protocol was that it should be amenable to automation on platforms which are already routinely accessible in most genomics laboratories, enabling medium to high throughput on existing infrastructure. Single cell studies generally require analysis of 100s, if not 1000s, of cells in parallel, and as such the ability to perform analyses in parallel at this scale is an essential part of the development of any new method.

Limitations

We have successfully applied the G&T-seq protocol to multiple cell types of mouse or human origin. Nevertheless, there are a number of potential limitations to the method. The WTA component of the method is currently limited to amplification of polyA(+) mRNA and thus cannot amplify polyA(-) RNA molecules. Also, strand specificity of the mRNA is lost as a consequence of the WTA. The WGA component of G&T-seq, similar to all current single-cell WGA methods, suffers from amplification bias relating to %GC-content, allelic drop-outs, base mis-incorporations due to polymerase errors during amplification, and production of chimeric DNA molecules [12,17]. The physical separation of polyA(+) mRNA and gDNA may contribute further to the problem of allelic dropout. However, we have not observed large Mb-scaled segments of chromosome(s) dropping out in our DNA copy number analyses of normal human and mouse single cells. Additionally, we demonstrated that genome sequencing read-outs of G&T-seq are comparable to those obtained from the corresponding single-cell WGA sequencing in isolation [17].

APPLICATIONS OF THE METHOD:

Integrative single cell DNA and RNA analysis is required to reveal genotype–phenotype associations within single cells and to study the impact of genetic variation in a cell on transcript levels and isoforms of the same cell. Integrative single-cell DNA and RNA analyses also enable the reconstruction of cell lineage trees using the DNA sequences of the cells, and subsequently to annotate this cell lineage tree with the transcriptional phenotype of the same cells. As such, the method is broadly applicable in developing an understanding of the biology of cellular heterogeneity in normal development and disease processes including embryology, neurology and oncology.

Furthermore, G&T-seq can act as a platform to support a variety of means by which the genomes and transcriptomes of single cells can be explored in parallel. Depending on the nature of the experimental question, the WGA and WTA products can be analysed in ways other than Illumina short read single- or paired-end sequencing. The WTA component of G&T-seq provides full-length cDNA, of which single molecules can be read from start to end on long-read sequencers such as the Pacific Biosciences RSII, providing valuable information about transcript splicing and expression of fusion transcripts in single cells. When investigating putative fusion transcripts arising from inter- or intra- chromosomal fusions, G&T-seq has the capacity of revealing the causative genomic rearrangement in the same single cell. Also, the WTA or WGA products can be screened with qPCR to confirm the presence of a transcript or genetic variant in individual cells, respectively. Furthermore, the WGA product within G&T-seq may be also subjected to exome or other forms of targeted sequencing. Lastly, the separated gDNA is amenable to other protocols such as single-cell bisulfite sequencing (scBS-seq)[32]. Indeed we have recently demonstrated that the G&T-seq plaform can be used as the basis for single cell Methylome and Transcriptome sequencing (scM&T-seq), which allows parallel exploration of the epigenome and gene expression in the same single cell[33].

Experimental Design

We designed the G&T-seq method to support processing of samples in 96-well plates, as studies involving single-cell analyses will often need throughput of this and greater magnitude to reveal the heterogeneity inherent in cell populations. For the processing of these plates, access to liquid handling robots is strongly recommended, and it should in principle be possible to use any of the common commercially available platforms equipped with a 96-channel head. Nevertheless, the protocol can be performed manually if such high throughput is not required.

When isolating single cells into 96-well plates for G&T-seq, we recommend including both multi-cell positive controls (where typically 10 to 50 cells are sorted into a single well) as well as empty-well negative controls. The multi-cell positive controls are particularly useful when new cell types are being

investigated and while FACS sorting conditions are being optimised. The empty-well controls are useful indicators of any contamination that may occur. Furthermore, the inclusion of spike-in RNA molecules, such as those developed by the ERCC [34], can be useful in assessing the performance of the single-cell WTA following sequencing. It is important to titrate the ERCC input carefully, depending on the cell type investigated, such that sequencing capacity is not overwhelmingly consumed by the spike-ins at the expense of measuring the cell's endogenous RNA. We recommend performing test experiments using a cell line that is well characterised in your laboratory.

Quality control (QC) of the process can be performed at various stages of the protocol. WGA and WTA products can be assessed for size distribution and yield using an Agilent Bioanalyser, and qPCR to detect the presence of genomic regions or the expression of housekeeping genes, allowing screening for successful amplification in advance of sequencing. After sequencing, further QC can be performed on both the DNA and RNA sequences to eliminate cells with lower quality data from any downstream analysis.

Applying G&T-seq requires expertise in the handling and isolation of single cells, molecular biology and next-generation sequencing techniques. The support of core facilities with expertise in laboratory automation, FACS and sequencing would be extremely beneficial when implementing the method.

MATERIALS

REAGENTS:

- Cell or tissue source for single-cell isolation. We have successfully performed the G&T–seq protocol on mouse blastomeres, human HCC38 breast cancer and HCC38-BL lymphoblastoid cell lines, human iPSC-derived neurons as well as primary human and mouse cells.
    - CAUTION: The cell lines used in your research should be regularly checked to ensure they are authentic and are not infected with mycoplasma.
- RNaseZap (Ambion, cat. no. AM9780)
- DNA-OFF (Takara Bio, cat. no. 9036)
- Nuclease free water (Ambion, cat. no. AM9937)
- 10 M NaOH (Sigma-Aldrich, cat. no. 72068)
- 5 M NaCl (Ambion, cat. no AM9760G)
- 0.5 M EDTA, pH 8.0 (Promega, cat. no V4231)
- UltraPure 1 M Tris-HCl Buffer, pH 7.5 (Thermo, cat. no 15567027)
- Trizma® Pre-set crystals, pH 8.3 (Sigma, cat. no T8943)
- 1 M $MgCl_2$ (Ambion, cat. no. AM9530G)
- 2 M KCl (Ambion, cat. no. AM9640G)
- DTT (Sigma 1M, cat. no. 646563)
    - ! CAUTION: DTT is toxic when ingested. Avoid inhaling fumes or contact with the skin. Handle it using appropriate safety equipment.
- 50% (vol/vol) Tween 20 (Invitrogen, cat. no. 003005)
- Buffer RLT Plus (Qiagen, cat. no. 1053393)
    - ! CAUTION: Buffer RLT contains guanidine thiocyanate which is harmful and should be handled with appropriate safety equipment.
- SUPERase In (Ambion, cat. no. AM2696)
- ERCC RNA Spike-In Mix (Ambion, cat. no. 4456740)
- Dynabeads MyOne Streptavidin C1 (Invitrogen, cat. no. 65001)
- SuperScript II reverse transcriptase (Life technologies, cat. no. 18064071)
- 5X first-strand buffer (Life technologies, cat. no. 18064071)
- DTT (Invitrogen, cat. no. 18064-014)

    ! CAUTION: DTT is toxic when ingested. Avoid inhaling fumes or contact with the skin. Handle it using appropriate safety equipment.
- Betaine solution 5M (Sigma, cat. no. B0300-1VL)

- dNTP mix 10mM each (Life technologies, cat. no. 18427-013)
- Kapa Hifi HotStart ReadyMix (Kapa, cat. no. KK2601)
- Agencourt AMPure XP Beads (Beckman Coulter, cat. no. A63881)
- Buffer EB (Qiagen, cat. no. 19086)
- WGA kit: PicoPLEX WGA Kit (Rubicon Genomics, cat. no. R30050) or Genomiphi V2 DNA Amplification Kit (GE Healthcare, cat. no. GE25-6600-31) or alternative WGA kit, depending on expected read-out. To the best of our knowledge, we are unaware of a commercially available WGA kit that would be incompatible with G&T-seq.
- Ethanol 95-97% (vol/vol) AnalaR Normapure analytical reagent (VWR, cat. no. 20823.327)
    - ! CAUTION: Ethanol is flammable and should be stored carefully and handled with appropriate safety equipment.
- Nextera XT DNA sample preparation kit, 96 samples (Illumina, cat. no. FC-131-1096)
- Nextera XT 96-index kit, 384 samples (Illumina, cat. no. FC-131-1002)
- Agilent High-sensitivity DNA kit (Agilent Technologies, cat. no. 5067-4626)
- Agilent 12000 DNA kit (Agilent Technologies, cat. no. 5067-1508)

EQUIPMENT:

- UV PCR Workstation (e.g. UVP UV PCR Workstation, 95-0367-02) and dedicated pre-amplification pipettes.
- Ultraviolet Crosslinker (e.g. UPV, CL-1000)
- 1.5 mL Microcentrifuge Safe-Lock tubes, polypropylene (Sigma-Aldrich, cat. no. T9661)
- 15 mL and 50 mL polypropylene Falcon tubes (BD, 352096 and 352070)
- DynaMag Spin Magnet (Life Technologies, cat. no. 12320D)
- Vortex Mixer (e.g. PV-1, Grant Instruments)
- Microcentrifuge (e.g. N2631-0007, Starlab UK)
- Fluorescence Activated Cell Sorter compatible with single cell deposition into 96-well plates; we routinely use an Influx (BD), FACSaria III (BD) or a MoFlo (Beckman Coulter) sorter.
- If manual isolation of cells is to be performed, we have successfully applied the STRIPPER Pipettor System (Origio, MXL3-STR-CGR) with 75 µL tips (Origio, MXL3- 75)
- FrameStar 96 well skirted PCR plates (4titude, cat. no. 4ti-0960/C)
- Multi-dispensing pipette (e.g. Multipette Xstream from Thermo. cat. no. 4986 000.025)
- Refrigerated centrifuge and adaptors for 96-well plates (e.g. 5810 R, Eppendorf)
- Rotator compatible with 1.5 mL microcentrifuge tubes (e.g. LD59, Labinco)

- Automated liquid handling platform. We have automated the indicated parts of the method on a BioMek FXP Laboratory Automation Workstation (Dual Arm system with multichannel pippeter and Span-8 pippeter; Beckman Coulter, cat. no. A31844). Other common robotic liquid handling platforms should be compatible with the protocol. The platform requires a station to keep reagents cooled to 4 °C (e.g. an on-deck Peltier cooled block) and an on-deck orbital shaker.
- Low elution magnet plate (Alpaqua, cat. no. A000350)
- ThermoMixer C (Eppendorf, cat. no.5382 000.015)
- ThermoTop (Eppendorf, cat. no. 5308 000.003)
- SmartBlock PCR 96 (Eppendorf, cat. no. 5306 000.006)
- Thermal cycler (e.g. MJ Research Tetrad, cat. no. PTC-225)
- Automated liquid handling platform in the post-PCR room. We apply a Zephyr Compact Liquid Handling Workstation (Perkin Elmer, cat. no. 125550) for the bead-based purification steps of amplification products.
- Agilent 2100 Bioanalyser (Agilent Technologies, cat. no. G2938C)
- Nanodrop 2000c (or similar apparatus) for DNA quantification (Thermo Scientific)
- An Illumina next-generation sequencing platform (MiSeq, NextSeq or HiSeq platform).
- Suitable computing infrastructure for NGS sequence analysis and the following software packages:
    - BWA (http://bio-bwa.sourceforge.net/)[35,36]
    - Cutadapt (https://cutadapt.readthedocs.org/en/stable/)[37]
    - Trim Galore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)
    - Samtools (http://www.htslib.org/)[38]
    - Picard (https://broadinstitute.github.io/picard/index.html)
    - Ginko (http://qb.cshl.edu/ginkgo/?q=/9duLgmqFJvmE93gLpo80)[39]
    - GATK (https://www.broadinstitute.org/gatk/)
    - R (https://www.r-project.org/)
    - Circos (http://circos.ca/)[40]
    - HTSeq (http://www-huber.embl.de/HTSeq/doc/overview.html)[41]
    - Tophat2 (https://ccb.jhu.edu/software/tophat/index.shtml)[42]
    - Deseq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html)[43]

REAGENT SETUP:

Cell lysis buffer

To prepare lysis buffer for one 96-well plate, transfer 2.5 µL of Buffer RLT to each well of the plate. It is not necessary to include RNAse inhibitor as the guanidine isothiocyanate in the lysis buffer should effectively inactivate RNAses. Perform these steps in a clean pre-amplification PCR/laminar flow cabinet, then seal and centrifuge the plate (2,000 g for 1 minute). Although Buffer RLT is stable at room temperature, we typically prepare plates immediately before preparation of single-cell samples.

ERCC spike-ins

Prepare 10 µL aliquots of a 1:100 dilution of the ERCC standards in nuclease free water, and store at -80 °C. For each day/experiment, thaw a fresh aliquot on ice and (serially) dilute to the desired final concentration in nuclease free water. Typically 1:1,000,000 is a good starting dilution to use, of which 1 µL is added to each cell lysate. However, we recommend that the ERCC concentration is first optimised for different cell types and sizes. Prepare these dilutions on ice in a clean pre-amplification PCR/laminar flow cabinet. Once thawed, the unused portion of the ERRC aliquot should be discarded.

Dynabeads Solution A

NaOH (0.1 M) and NaCl (0.05 M) in nuclease free water. To prepare 50 mL of this buffer add 500 µL of 10 M NaOH and 500 µL of 5 M NaCl to 49 mL of nuclease free water. This buffer can be prepared in bulk and stored at 4° C for up to one month. Prepare this solution at room temperature in a clean pre-amplification PCR/laminar flow cabinet.

Dynabeads Solution B

NaCl (0.1 M) in nuclease free water. To prepare 50 mL of this buffer add 1 mL of 5 M NaCl to 49 mL of nuclease free water. This buffer can be prepared in bulk at room temperature and stored at 4° C for up to one month. Prepare this solution in a clean pre-amplification PCR/laminar flow cabinet.

Dynabeads 2x 'Binding and Wash' (B&W) buffer

10 mM Tris-HCl (pH 7.5), 1 mM EDTA and 2 M NaCl in nuclease free water. To prepare 50 mL of this buffer, add 0.5 mL of 1M Tris-HCL (pH 7.5), 100 uL of 0.5 M EDTA and 20 mL of 5 M NaCl to 29.4 mL of nuclease free water. This buffer can be prepared in bulk at room temperature and stored at 4 °C for up to one month. Prepare this buffer in a clean pre-amplification PCR/laminar flow cabinet.

G&T-seq wash buffer

50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 0.5% (vol/vol) Tween-20. Prepare this buffer in a clean pre-amplification PCR/laminar flow cabinet. To prepare 50 mL of G&T-seq wash buffer, combine 25 mL of 0.1 M Tris-HCl (pH 8.3), 1.875 mL of 2 M KCl, 30 µL 0.5 M $MgCl_2$, 50 µL 1M DTT, 50 µL 50% (vol/vol) Tween and 22 mL of nuclease free water. Mix by vortexing. This buffer can be prepared in bulk at room temperature and stored at 4 °C for up to one month, but should be supplemented with 0.05x RNAse Inhibitor (SUPERase In) immediately before use.

Biotinylated Oligo-dT30VN

(5'-Biotin-TEG-AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTVN-3') This primer should be ordered RNAse free and HPLC purified. Resuspend at 100 µM in nuclease-free water. Perform these steps in a clean pre-amplification PCR/laminar flow cabinet. Once resuspended, this oligo should be stored at -20 °C and in our hands has been stable for >6 months.

Template Switching Oligo (TSO)

(5'-AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3') (Exiqon) resuspend at 100 µM in nuclease-free water. Perform these steps in a clean pre-amplification PCR/laminar flow cabinet. Once resuspended, this oligo should be store at -80 °C and in our hands has been stable for > 6 months. Repeated freeze-thaw cycles should be avoided and we recommend preparing single use 10 µL aliquots.

ISPCR oligo (5'-AAGCAGTGGTATCAACGCAGAGT-3') resuspend at 100 µM in nuclease-free water or TE buffer. Perform these steps in a clean pre-amplification PCR/laminar flow cabinet. Once resuspended, this oligo should be storedat -20 °C and in our hands has been stable for >6 months.

EQUIPMENT SETUP:

The protocol can be performed manually on small numbers of samples, however if numerous single cells are to be analysed, we strongly advise the use of automated liquid handling platforms.

Pre-Amplification Setup

In a clean pre-PCR room, we use a BioMek FXP Laboratory Automation Workstation for steps 19-27 and steps 49-54 of the G&T-seq protocol. Images of the deck layout for running the G&T-seq protocol on our robot are shown in Supplementary Figures 1 and 2. Alternative liquid handling platforms may be used, but require a cooling station, orbital shaker and a low elution magnet plate. It is important to carefully programme the liquid handling platform for optimal mixing and minimising bubble formation during pipetting. Other essential equipment in the pre-PCR room includes a UV PCR workstation with conventional P2, P20, P200 and P1000 pipettes, a UV cross-linker, a digital multi-

dispensing pipette, 1.5 mL microcentrifuge tubes, 15 mL and 50 mL falcon tubes, a magnet for 1.5 mL microcentrifuge tubes, a vortex mixer, a microcentrifuge, a refrigerated centrifuge and adaptors for 96-well plates, a thermomixer for 96-well plates and a thermal cycler for 96-well plates.

Post-Amplification Setup

In the post-PCR room, we use a Zephyr Compact Liquid Handling Workstation for steps 36-45, 56, 74 of the G&T-seq protocol. Other essential equipment in the post-PCR room includes P2, P20, P200 and P1000 pipettes, a P20 multichannel pipette, vortex, centrifuges for 1.5mL microcentrifuge tubes and 96-well plates, apparatus for DNA quantification and an Agilent 2100 Bioanalyser.

PROCEDURE:

Preparation of single-cell samples (Timing: 30 mins - 1 hour)

! Critical step: single-cell work is extremely sensitive to contamination; all experiments should be performed using a PCR or laminar flow hood with UV-sterilisation as dedicated "pre-amplification" workspace. Sets of pipettes and all reagents should be reserved for single-cell pre-amp protocols and the amplified cDNA and gDNA should be handled only in a physically separated working environment. Traffic between these pre- and post-amplification areas should be kept to a minimum, and separate lab coats and other Personal Protective Equipment (PPE) should be used in each work area. Pre-amplification areas should be cleaned with DNA-OFF and RNAZap before and after each experiment.

1) In the pre-amplification work area, prepare 96-well plates containing 2.5 µL of lysis buffer per well.
2) Seal and centrifuge the plates (2,000 g for 1 minute at 4°C) to remove bubbles and to ensure lysis buffer is at the bottom of each well.
3) Following preparation of a single-cell suspension, single cells can be deposited directly into the lysis buffer contained in each well by FACS, or by manual isolation and transfer. If cells are isolated manually, e.g. using micropipetters, transfer them in the smallest possible volume (0.5 µL or less). We recommend including a multi-cell control (typically 10-50 cells) as a positive control as well as at least 1 empty well per plate as a negative control.

   ! Critical step: Regardless of the cell type being analysed and the means by which single cells are isolated, processing time should be kept minimal. Long processing times may result in cell death or aberrations in the cells' transcriptomes.

   ! Critical step: When using FACS for single cell deposition, considerable care must be taken to ensure accurate deposition of the single cells into the 96-well plates. This can be optimised by (i) FACS sorting beads onto a sealed 96-well plate to calibrate the site of deposition in a well of a plate, and (ii) sorting calcein AM stained single cells into clear-bottomed 96 well plates followed by visual inspection under a fluorescence microscope to determine whether indeed single cells are being deposited.
4) Centrifuge the plate containing cells in lysis buffer (2,000 g for 1 minute at 4°C) before transferring the plate to -80° C for storage. The cells immediately lyse within the lysis buffer.

   PAUSE POINT: once single cells have been deposited into RLT buffer and stored at -80° C, the samples are stable for >1 month, and we have successfully processed plates after >6 months in storage.

Preparation of oligo-dT30VN labelled beads (Timing: 1 hour)

! Critical step: steps 5-13 should be performed in a dedicated pre-amplification laboratory, with steps 5-9 and 11-13 in a pre-amplification PCR hood/laminar flow cabinet.

5) Add 50 µL of Dynabeads in a 1.5 mL Eppendorf tube and place on a magnet designed for handling Eppendorf tubes (e.g. DynaMag Spin Magnet from Invitrogen) for 30 seconds (or until a clear solution is obtained), then completely remove the supernatant whilst keeping the tube on the magnet.

6) Remove the tube from the magnet and resuspend the beads in 200 µL of Dynabead solution A. Return the tube to the magnet for 30 seconds (or until a clear solution is obtained) and remove the supernatant whilst keeping the tube on the magnet. Repeat once.

7) Remove the tube from the magnet and resuspend the beads in 200 µL of Dynabead solution B. Return the tube to the magnet for 30 seconds (or until a clear solution is obtained) and remove supernatant completely whilst keeping the tube on the magnet.

8) Remove the tube from the magnet and resuspend the beads in 50 µL of 2x B&W buffer.

9) Add 50 µL of 100 µM Biotinylated Oligo-dT30VN to the beads.

10) Incubate for 20 min with gentle rotation on the rotator. In those 20 min, steps 14 to 18 may be performed; afterwards return immediately to step 11.

11) Place the Eppendorf containing the Oligo-dT30VN conjugated beads on the magnet for 30 seconds (or until a clear solution is obtained) and discard the supernatant whilst keeping the tube on the magnet.

12) Wash the beads four times in 200 µL 1 x B&W buffer. For each wash, remove the tube containing the beads from the magnet, resuspend the beads in 200 µL 1 x B&W buffer, then return the tube to the magnet and remove –whilst keeping the tube on the magnet– the supernatant once the bead suspension has cleared.

13) Prepare 1 mL of bead resuspension buffer by combining and mixing the reagents as indicated in the table below, then add 1 mL of bead resuspension buffer to the beads and mix the beads by vortexing until completely resuspended.

! Critical step: Once the RNAse inhibitor is added, the beads should be used immediately. If steps 14 to 18 are already performed; immediately continue with step 19.

Component	Amount (µL)	Final Concentration
Superscript II first strand buffer (5x)	200	1x
RNAse inhibitor (20 U/µL)	50	1 U/µL
Nuclease Free Water	750	

Addition of ERCC spike-ins (Timing: 10 mins)

! Critical step: During protocol optimisation we recommend adding the ERCC spike-ins after cell collection to allow for easier titration of ERCC amounts, however ERCC spike-ins can also be added directly to the lysis buffer when preparing plates for cell collection.

14) Remove the plate that contains the single cells in lysis buffer from the -80 °C freezer on ice. When thawed, centrifuge at 1,000 g for 1 min at 4°C.

15) In the pre-amplification laminar flow cabinet, add 1 µL of a suitable dilution of ERCC spike-in to each well using a digital multi-dispensing pipette (e.g. Multipette Xstream from Thermo), dispensing the droplet onto the side of the well. Seal the plate.

16) Centrifuge at 1,000 g for 1 min at 4°C to collect the ERCC droplets at the bottom of the well with the single-cell lysate.

Physical separation of polyadenylated mRNA and gDNA (Timing: 1.5 hours)

17) In the pre-amplification laminar flow cabinet, prepare 500 µL of reverse transcription (RT) mastermix per 96-well plate (100 reactions) by combining and mixing the reagents in the table below in the indicated order (starting at the top) on ice. Store this RT mastermix on ice until use.

Component	Amount (µL)	Final Concentration
Nuclease-free water	179.5	
dNTP mix (10 mM each)	50	1 mM each
TSO (100 µM)	5	1 µM
$MgCl_2$ (1 M)	3	6 mM
Betaine (5 M)	100	1 M
Superscript II first strand buffer (5x)	100	1x
DTT (100 mM)	25	5 mM
Superscript II reverse transcriptase (200 U/µL)	25	10 U/µL

RNAse inhibitor (20 U/µL)	12.5	0.5 U/µL
Total Volume	500	

18) In the pre-amplification laminar flow cabinet, prepare wash buffer. Mix 2475 µL of G&T-seq wash buffer with 25 µL of RNAse inhibitor at room temperature. Dispense 25 µL of this wash buffer mix into each well of a new 96-well plate labelled 'G&T-seq wash buffer plate'. Seal the plate, centrifuge at 1,000 g for 1 min at 4°C and store at room temperature until use.

! Critical step: In our laboratories, steps 19 to 27 are performed on deck of a liquid handling robot, but may also be performed manually. The volumes indicated in those steps are per well –i.e. per sample– and are thus valid for both the automated and manual version of the G&T-seq protocol. When performed manually, keep working in a clean pre-amplification laminar flow cabinet. When performing the automated G&T-seq protocol, set up your liquid handling robot as described in Supplementary Figure 1. We describe the protocol for single cells (plus positive and negative controls) that have been deposited in a 96-well plate.

19) Add 10 µL of Oligo-dT30VN beads of step 13 to each well of the 96-well plate containing the lysed single-cell samples of step 16.

20) Incubate with mixing (2,000 rpm using an on-deck mixer or thermomixer) for 20 min at room temperature; if the plate is removed from the robot deck for this step then the plate should be sealed.

21) Place the plate on a low-elution magnet for 96-well plates for 1 min and remove supernatant. Carefully transfer the supernatant –which contains the gDNA of the cell– to a new 96-well plate in the same orientation and keeping track of the address of each cell. This is the 'gDNA collection' 96-well plate.

! Critical step: Use the same tip for all subsequent washes and transfers of supernatant (steps 22 – 25). This can help to minimise the loss of gDNA during these steps.

22) Wash beads with 10 µL G&T-seq wash buffer. Resuspend the beads off the magnet in G&T-seq wash buffer and mix for 5 minutes using the on-deck orbital shaker or a thermomixer at room temperature. If the plate is removed from the robot deck for this step then the plate should be sealed.

23) Return the bead-resuspended 96-well plate to the low-elution magnet and allow the beads to precipitate for 1 min. Carefully transfer all supernatant to the corresponding well in the 'gDNA collection' 96-well plate.

24) Repeat steps 22 and 23 to wash the beads a second time.

25) Wash the pipetting tips with G&T-seq wash buffer. Aspirate the remaining 5 µL of G&T-seq wash buffer into the tips and transfer this to the corresponding well in the 'gDNA collection' 96-well plate. The entire 25 µL of the G&T-seq wash buffer should now have been transferred to the corresponding well on the gDNA collection 96 well plate.

26) Dispense 5 µL of RT mastermix into each well of the bead-containing 96-well plate. This plate is now referred to the 'polyA(+) mRNA 96-well plate'.

27) Collect the polyA(+) mRNA 96-well plate from the liquid handling robot, seal it and centrifuge (1,000 g for 1 min at 4°C) to collect both the RT mix and bead-captured mRNA at the bottom of the well.

! Critical step: immediately continue with the processing of the polyA(+) mRNA, i.e. immediately perform step 29 then return to step 28.

28) Collect the 'gDNA collection' 96-well plate from the liquid handling robot, seal and centrifuge (1,000 g for 1 min at 4°C) to collect all liquid at the bottom of the well. The DNA can now be stored at -80 °C until required for further processing (step 47 and following).

Reverse Transcription (Timing: 2 hours)

29) Place the polyA(+) mRNA 96-well plate on an Eppendorf Thermomixer C equipped with a ThermoTop to prevent evaporation and perform the thermal and mixing steps outlined below. This RT reaction will convert the polyA(+) mRNA to PCR-amplifiable cDNA molecules. Once the thermomixing has been initiated return to step 28.

Cycle	**Temperature (°C)**	**Time**	**Mixing (rpm)**	**Purpose**
1	42	2 min	2000	Resuspension of the beads, RT and Template switching
2	42	60 min	1500	RT and Template switching
3	50	30 min	1500	RT and Template switching
4	60	10 min	1500	Enzyme inactivation

30) Once RT is complete, centrifuge the plate (1,000 g for 30 seconds at room temperature) to collect liquid at the bottom of the wells.

PCR amplification of cDNA (Timing: 3 hours)

31) Prepare the PCR reaction mastermix fresh as described in the table below.

Component	Amount (µL)	Final Concentration
Kapa HiFi HotStart ReadyMix (2x)	625	1x
IS PCR primers (10 µM)	12.5	0.1667 µM
Nuclease-free water	112.5	
Total Volume	**750**	

32) Add 7.5 µL of the PCR reaction mastermix directly to the side of each well of the 96-well plate containing PCR-amplifiable cDNA molecules from step 30, resulting in a total volume of 12.5 µL. Seal the plate and centrifuge (1,000 g for 30 s at room temperature) to collect the liquid at the bottom of the wells.

33) Mix the beads and reaction mixture briefly on the Eppendorf Thermomixer C (60 s at 2,000 rpm, room temperature) to ensure suspension of the beads.

34) Perform cDNA amplification on a thermal cycler equipped with a heated lid set to 105 °C.

Cycle Number	Denature	Anneal	Extend	Final
1	98 °C, 3 min			
2-19	98 °C, 20 s	67 °C, 15 s	72 °C, 6 min	
20			72 °C, 5 min	
21				4 °C, Hold

! Critical step: The cycle number may need to be optimised for different cell types depending on the mRNA content of the cell.

PAUSE POINT: Remove the plate containing the amplified cDNA from each single cell from the thermal cycler, and centrifuge it at 1,000 g for 1 min at 4°C. Subsequently, the cDNA product can be stored at -20 °C for several months before purification.

Purification of the Amplified cDNA (Timing: 1 hour)

! Critical step: steps 35-46 should be performed in a dedicated post-amplification room separate from the location of pre-amplification and cell isolation. We perform this process using the Zephyr automated liquid handling robot, however alternative liquid handling platforms or manual processing are also possible.

35) Allow Agencourt AMPure beads to warm up to room temperature for 15 min before use, gently mixing to ensure the beads are evenly resuspended. Meanwhile, also spin the plate containing the amplified cDNA of each single cell (from step 34) at 1,000 g for 1 min at 4°C.

36) Add 12.5 µL of the Agencourt AMPure beads to each well of the 96-well plate containing the PCR-amplified cDNA molecules (1:1 ratio) at room temperature, and mix thoroughly by pipetting up and down. Allow the mixture to stand for 5 min at room temperature.

37) Transfer the plate to a low-elution magnet and allow the beads to settle for 2 min or until the solution is clear.

38) Once the Agencourt AMPure beads have settled, carefully remove the supernatant without disturbing the beads.

39) Keeping the 96-well plate on the magnet, wash the Agencourt AMPure beads with 100 µL of freshly prepared 80% (vol/vol) ethanol for 30 s, then remove the ethanol wash. This should be done without disturbing the beads.

40) Repeat step 39 once.

41) Remove any remaining ethanol solution from the well without disturbing the beads and then allow the Agencourt AMPure beads to dry for approximately 5 min.

42) Add 25 µL of nuclease free water to the Agencourt AMPure beads, remove the plate from the magnet and resuspend by pipetting up and down at room temperature.

43) Incubate this 96-well plate for 2 min off the magnet at room temperature.

44) Return the 96-well plate to the magnet and allow the Agencourt AMPure beads to settle for 5 min or until the solution is clear.

45) Carefully remove the supernatant, which contains the purified cDNA, without disturbing the beads and transfer it to a new 96-well plate at room temperature.

PAUSE POINT: The amplified cDNA can be stored at -20 °C for >6 months before library preparation.

Quality Control of Amplified cDNA (Timing: 1 hour)

46) Check the quality of the cDNA using a High Sensitivity Chip on an Agilent Bioanalyser. A successful reaction should generate cDNA between 0.5 and 2 kb reaching a maximum at about 1-1.5 kb (Fig. 2A). At this point, the most abundant ERCC spike-ins (at ~0.5 and 1 kb) may also be visible. We have also found qPCR for housekeeping genes to be a useful means to assess the number of failed amplifications, when it is not practical to analyse each well on the Bioanalyser. Failed amplifications are mostly due to cells with degraded RNA (Fig. 2B) or 'empty' wells in which no cell was deposited during flow-sorting or manual isolation (Fig. 2C).

Purification of gDNA (Timing: 1 hour)

! Critical step: perform all steps in this process in a dedicated pre-amplification laboratory. We perform steps 49-54 on deck of a liquid handling robot placed in the pre-amplification laboratory –a BioMek FXP Laboratory Automation Workstation– but may also be performed on alternative liquid handlers or manually. All manual steps in this process must be performed in a pre-amplification PCR hood/laminar flow cabinet.

47) Allow Agencourt AMPure beads to warm up to room temperature for 15 min before use, gently mixing to ensure the beads are evenly resuspended.

48) If the 'gDNA collection' 96-well plate from step 28 has been frozen; first thaw it on ice, ensure that it is completely thawed, then centrifuge for 1 min at 1,000 g at 4 °C to collect the liquid at the bottom of the well. The plate should contain 38.5 µL of liquid in total.

49) Add 25 µL of Agencourt AMPure beads to each well of the plate (0.75:1 ratio) at room temperature and mix thoroughly by pipetting up and down. Allow the mixture to stand for 20 min at room temperature.

50) Transfer the 'gDNA collection' 96-well plate containing the Agencourt AMPure beads to a Low Eution magnet and allow the beads to settle for 20 min at room temperature.

51) Once the Agencourt AMPure beads have settled, carefully remove the supernatant without disturbing the beads.

52) Keeping the plate on the magnet, wash the Agencourt AMPure beads with 100 µL of freshly prepared 80% (vol/vol) ethanol for 30 s at room temperature, then remove the ethanol wash. This should be done without disturbing the beads.

53) Repeat step 52 once.

54) Remove any remaining ethanol solution from the well and then allow the Agencourt AMPure beads to dry for approximately 5 min at room temperature.

55) Depending on the desired downstream application, chose to perform PicoPLEX or MDA WGA on the purified gDNA.

! Critical step: The choice of WGA protocol is largely depending on the desired readout. In general, MDA-based WGA methods are often preferred for SNV or SV detection, but can be suboptimal for the assessment of DNA copy number changes in the genome [6,12-14]. In contrast, we and others have found that PicoPLEX generally outperforms MDA for DNA copy number profiling of single cells following low coverage sequencing, however it lacks the breadth of coverage and base-level fidelity of MDA-based methods [6,12,14]. At this stage in the protocol one can either follow path A for PicoPLEX-based WGA or path B for MDA-based WGA of the single-cell DNA.

A) WGA using PicoPLEX from Rubicon Genomics / New England Biolabs (Timing: 3 hours)

! Critical step: perform all steps in this process in a pre-amplification PCR hood/laminar flow cabinet in a dedicated pre-amplification laboratory.

i) Prepare the PicoPLEX Sample/Extraction mix as described in the table below, which is sufficient mastermix for one 96-well plate. Mix well and keep on ice.

Component	**Volume (µL)**
PicoPLEX Cell Extraction Buffer	250
PicoPLEX Extraction Enzyme Dilution Buffer	240
PicoPLEX Cell Extraction Enzyme	10

ii) Using a digital multi-dispensing pipette dispense on ice 5 µL of the PicoPLEX Sample/Extraction mix to each well of the 96-well plate containing the dried Agencourt AMPure beads, then seal the plate.

iii) Centrifuge the plate (1,000 g for 1 min, 4 °C) to collect the liquid and the Agencourt AMPure beads at the bottom of the plate.

iv) Resuspend the Agencourt AMPure beads by placing the plate on an orbital shaker, or using the Eppendorf Thermomixer C, and mixing vigorously (2,000 rpm) for 1 min at room temperature.

v) Place the plate in a thermal cycler with a heated lid and perform the following steps:

Cycle	Temperature (°C)	Time
1	75	10 min
2	95	4 min
3	4	hold

vi) Centrifuge the plate (1,000 g for 1 min at 4 °C) to collect the liquid and the Agencourt AMPure beads at the bottom of the plate.

vii) Prepare on ice the Pre-Amp mastermix as described in the table below, which is sufficient mastermix for one 96-well plate.

Component	Volume (µL)
PicoPLEX Pre-Amp Buffer	240
PicoPLEX Pre-Amp Enzyme	10

viii) Using a digital multi-dispensing pipette, add 2.5 µL of Pre-Amp mix to each well of the 96-well plate on ice, then seal the plate.

ix) Centrifuge the plate (1,000 g for 1 min at 4 °C) to collect the liquid and the Agencourt AMPure beads at the bottom of the plate.

x) Resuspend the Agencourt AMPure beads by placing the plate on an orbital shaker, or using the Eppendorf thermomixer C, and mixing vigorously (2,000 rpm) at room temperature for 1 min.

xi) Return the plate to the thermal cycler and perform the following steps:

Cycle Number	Denature	Anneal & Extend	Final
1	95 °C, 2 min		
2-13	95 °C, 15 s	15°C, 50 s 25°C, 40 s 35°C, 30 s 65°C, 40 s 75°C, 40 s	

14			4 °C, Hold

xii) Centrifuge the plate (1,000 g for 1 min at 4 °C) to collect liquid and Agencourt AMPure beads at the bottom of the plate.

xiii) Prepare the Amplification mastermix as described in the table below on ice, which is sufficient mastermix for one 96-well plate. Mix well and keep on ice.

Component	**Volume (µL)**
PicoPLEX Amplification Buffer	1250
PicoPLEX Amplification Enzyme	40
PicoPLEX Nuclease Free Water	1710

xiv) On ice, add 30 µL of Amplification mastermix to each well of the 96-well plate using a digital multi-dispensing pipette.

xv) Seal the plate and centrifuge (1,000 g for 1 min at 4 °C) to collect the liquid and the Agencourt AMPure beads at the bottom of the plate.

xvi) Resuspend the Agencourt AMPure beads by placing the plate on an orbital shaker, or using the Eppendorf thermomixer C, and mixing vigorously (2,000 rpm) at room temperature for 1 min.

xvi) Return the plate to the thermal cycler and perform the following steps to generate the amplified gDNA product.

Cycle Number	Denature	Anneal	Extend	Final
1	95 °C, 2 min			
2-15	95 °C, 15 s	65 °C, 1 min	75 °C, 1 min	
16				4 °C, Hold

PAUSE POINT: The PicoPLEX amplified cDNA product can be stored at -20 °C for several months before purification.

B) WGA by MDA using GenomiPhi V2 from GE healthcare (Timing: 3 hours)

! Critical step: perform all steps in this process in a pre-amplification PCR hood/laminar flow cabinet in a dedicated pre-amplification laboratory.

i) Mix the GenomiPhi V2 Sample Buffer with nuclease-free water as described in the table below, which is sufficient mix for one 96-well plate. Mix well and keep on ice.

Component	**Volume (µL)**
GenomiPhi V2 Sample Buffer	900
Nuclease free water	100

ii) On ice, add 10 µL of this mixture to each well of the 96-well plate containing the dried Agencourt AMPure beads from step 54, then seal the plate.

iii) Centrifuge the plate (1,000 g for 1 min at 4 °C) to collect the liquid and the Agencourt AMPure beads at the bottom of the plate.

iv) Resuspend the Agencourt AMPure beads by placing the plate on an orbital shaker, or using the Eppendorf thermomixer C, and mixing vigorously (2,000 rpm) for 1 min at room temperature.

v) Denature the template DNA by heating the sample to 95 °C for 3 minutes then immediately cool to 4 °C. Once cool, centrifuge the plate (1,000 g for 1 min at 4 °C) to collect the liquid and the Agencourt AMPure beads at the bottom of the wells.

vi) Prepare the MDA mastermix as described in the table below on ice, which is sufficient mastermix for one 96-well plate. Mix well and keep on ice.

Component	**Volume (µL)**
GenomiPhi V2 Reaction Buffer	900
GenomiPhi V2 Enzyme Mix	100

vii) On ice, add 10 µL of this mixture to each well of the plate and seal the plate. Centrifuge the plate (1,000 g for 1 min at 4 °C) to collect the liquid and the Agencourt AMPure beads at the bottom of the wells.

viii) Transfer the plate to the Eppendorf Thermomixer C and perform the following incubation:

Cycle	Temperature (°C)	Time	Mixing (rpm)	Purpose
1	30	2 min	2000	Resuspension of the beads, amplification
2	30	90 min	1500	Amplification
3	65	10 min	1500	Enzyme inactivation

ix) Centrifuge the plate (1,000 g for 1 min at 4 °C) to collect the liquid and the Agencourt AMPure beads at the bottom of the well.

PAUSE POINT: The MDA amplified cDNA product can be stored at -20 °C for several months before purification.

Purification and QC of Amplified gDNA (Timing: 1 hour)

! Critical step: perform steps 56-76 in a dedicated post-amplification room

56) Perform an AMPure bead clean-up of the single-cell WGA product generated in step 55 A or B as performed in steps 35-45 with two minor modifications. In step 36, adjust the volume of Agencourt AMPure beads added to the sample respecting a 1:1 ratio (i.e. 37.5 µL of the Agencourt AMPure bead suspension if continuing from step 55A-xvi and 20 µL of the Agencourt AMPure bead suspension if continuing from step 55B-x) and, in step 42, eluting in 100 µL of nuclease free water.

57) Measure the DNA concentration of the purified single-cell WGA product using a nanodrop. Both MDA and PicoPLEX WGA protocols typically generate more than 0.5 µg of DNA.

58) Check the quality of the amplified gDNA using an Agilent Bioanalyser. Purified PicoPLEX WGA product should be analysed on a High sensitivity chip and generate a peak between 300 and 2,000 bp (Fig. 2D). MDA WGA products are much longer (1-20kb) and should be analysed using the 12000 DNA kit from Agilent or by gel electrophoresis.

PAUSE POINT: The amplified gDNA can be stored at -20 °C for >6 months before library preparation.

Tagmentation-based library preparation using the Nextera XT Sample Preparation kit (Illumina) (Timing: 2 hours)

! Critical step: for the preparation of sequencing libraries from single-cell PicoPLEX WGA product, the Nextera XT DNA Library Prep Kit (Illumina) generates libraries of sufficient complexity for genome-wide DNA copy number profiling using low coverage sequencing (see below). However, when the

experimental aim requires the detection of single nucleotide variant changes, or structural rearrangements, we recommend conventional adaptor ligation based method (e.g. Illumina TruSeq, or Kapa HyperPlus) following the manufacturer's instructions on single-cell MDA WGA product. We have successfully analysed MDA-amplified gDNA from the G&T-seq protocol on the Illumina HiSeq X platform using the manufacturer's specified protocol [17]. For sequencing the amplified cDNA of single cells, we use the Nextera XT DNA Library Prep Kit (Illumina) as in the Smart-seq2 protocol [25] with minor modifications.

Below we describe our approach of tagmentation using the Nextera XT DNA Library Prep Kit (Illumina). For conventional adaptor ligation based methods for library preparation, we recommend following the manufacturer's instructions of the applied kit.

59) Dilute purified WTA product (i.e. amplified and purified single-cell cDNA from step 45) or purified WGA product (i.e. amplified and purified single-cell gDNA from step 56) to 0.2 ng/µL.

60) Prepare the tagmentation mastermix for one 96-well plate of purified WTA or WGA product as described in the table below using the reagents provided with the Nextera XT DNA Library Prep Kit, mix the solution carefully with a pipette and keep on ice.

Component	**Volume (µL)**
Tagment DNA Buffer	300
Amplicon Tagment Mix	150
Total Volume	450

61) Dispense 3.75 µL of tagmentation mastermix to each well of a new 96-well plate on ice.

62) Transfer 1.25 µL of the diluted WTA or WGA product to each well of the 'tagmentation mastermix' plate, keeping the plates in the same orientation and mixing gently with each addition. Perform this step on ice.

63) Seal the 'tagmentation reaction' plate, mix gently using an Eppendorf thermomixer C or similar (1,000 rpm for 1 min at room temperature), and centrifuge (1,000 g for 1 min at 4 °C) to collect the mixture at the bottom of the wells.

64) Perform the tagmentation reaction on the thermal cycler using the following settings.

Cycle	**Temperature (°C)**	**Time**

1	55	10 min
2	10	Hold

65) Centrifuge (1,000 g for 1 min at 4 °C) to collect the mixture at the bottom of the wells.

66) Add 1.25 µL of Buffer NT to each well of the 96-well 'tagmentation reaction' plate on ice, mixing gently with each addition.

67) Centrifuge (1,000 g for 1 min at 4 °C) to collect the mixture at the bottom of the wells.

68) Add 3.75 µL of Nextera PCR Mastermix (NPM) to each well of the 96-well 'tagmentation reaction' plate on ice, mixing gently with the pipette with each addition. Discard the tips after each addition.

69) Add 1.25 µL of the appropriate Index 1 (N7xx) primer to each well of the 96-well 'tagmentation reaction' plate on ice, mixing gently with the pipette with each addition. Discard the tips after each addition.

70) Add 1.25 µL of the appropriate Index 2 (N5xx) primer to each well of the 96-well 'tagmentation reaction' plate on ice, mixing gently with the pipette with each addition. Discard the tips after each addition.

71) Once NPM and both indexes have been added to each well, seal the plate and centrifuge (1,000 g for 1 min at 4 °C) to collect the reaction mix at the bottoms of the wells.

72) Perform the amplification of the tagmented DNA fragments on a thermal cycler using the following program.

Cycle Number	Denature	Anneal & Extend	Final
1	72 °C, 3 min		
2	95 °C, 30 s		
3-14	95 °C, 15 s	55 °C, 10 s	
15			72 °C, 5 min
16			4 °C, Hold

Pooling and quantification of libraries (Timing: 3 hours)

73) Pool the amplified libraries by transferring 5 µL from each well to a clean microcentrifuge tube. The remaining library can be kept for deeper sequencing (i.e. less multiplexing) should this be required.

! Critical step: If preparing libraries from a whole 96-well plate of single-cell WGA or WTA products, we often pool the 96 libraries without quantification of each individual library. This can lead to skewing of library representation within the multiplex sequencing reaction, however, if the input material is of similar quality and quantity, we rarely see severe over- or under-representation of specific libraries. If sequencing depth needs to be matched exactly between samples, we recommend quantification and normalisation of the individual libraries before pooling according to standard methods for Illumina sequencing libraries (e.g. using the KAPA Library Quantification Kit for Illumina platforms from Kapa Biosystems).

74) Perform an Agencourt AMPure bead clean-up of the pooled library products as described in steps 35-45, adjusting the volume of the beads in step 36 to have a 1:0.6 ratio between the library pool and the Agencourt AMPure beads (e.g. for a whole plate, 480 µL of library pool will be collected, to which 288 µL of Agencourt AMPure beads should be added and mixed).

75) Check the quality of the purified library pool using a High Sensitivity Chip on an Agilent Bioanalyser. A successful Nextera library preparation reaction, and hence also the library pool, should consist primarily of DNA fragments between 300 and 800 bases (Fig. 2E).

76) Quantify the Nextera library pool derived from the single cells' WTA or WGA product. We routinely quantify our libraries using the KAPA Library Quantification Kit for Illumina platforms. After quantification dilute the library if required (we typically dilute to 4 nM).

Library sequencing

! Critical Step: The depth of sequencing required for single cell genomes and transcriptomes can vary widely based on the experimental application. We typically sequence 96 libraries over 2 lanes on a HiSeq2500 (paired-end, 100 bp) running in rapid mode for cDNA sequencing (generating approximately 4-6 million reads per cell). This generally gives a sufficiently detailed overview of the cell's transcriptome to allow differences in cell type or state to be observed. If more detailed analysis is preferred, for example if expressed SNVs or splice junctions are to be detected, it is possible to pool fewer single-cell samples per lane.

It is also possible to perform a higher degree of multiplexing for single cell transcriptomes if less detail is required per cell and suitable indexing primers are available, which can significantly reduce costs. We do not recommend shallower sequencing for single cell genomes unless very large (whole chromosome) copy number changes are anticipated. Where MDA based WGA has been performed, and full genome characterization of single cells is required, we recommend using the Illumina HiSeq X platform to achieve the depth required to observe genome wide SNVs in single cells.

77) Perform single- or paired-end sequencing as required on the single-cell cDNA or gDNA derived libraries, in accordance with the manufacturer's protocols.

Data analysis

! Critical step: the G&T-seq method can address a diverse range of biological questions, and as such, the analytical approaches required can vary significantly. Here, we provide details on how to process and perform QC the genome and transcriptome data generated, and indicate how the genomic and transcriptomic data from each single cell can be integrated to explore the relationship between chromosomal copy number and gene expression. Of course, a diverse range of analyses is possible with such data, and bespoke informatics tools may be required to explore particular questions of interest.

Genome Data Analysis - Pre-processing and mapping of whole genome paired-end sequencing data

78) The single-cell genome sequencing reads prepared with the Nextera XT kit are first trimmed for 23 bases to remove adapter sequences. This can be done with Cutadapt (1).

```
cutadapt -u 23 -o trimmed.single_cell_R1.fastq single_cell_R1.fastq
```

79) For each single cell, align the trimmed sequences onto the appropriate reference genome (e.g. GRCh37 for human cells, mm10 for mouse cells) to generate separate SAI files for the first and second reads of the single-cell paired-end sequences using BWA (version 0.6.2) (2, 3).

```
bwa aln -l 32 GRCh37.fa singlecell_R1.fastq.gz >singlecell_R1.sai

bwa aln -l 32 GRCh37.fa singlecell_R2.fastq.gz >singlecell_R2.sai
```

80) For each single cell, create the alignments in SAM format from the SAI files by applying sampe from BWA.

bwa sampe -P -s GRCh37.fa singlecell_R1.sai singlecell_R2.sai singlecell_R1.fastq.gz singlecell_R2.fastq.gz >singlecell.sam

81) Finally, convert the SAM file of a single cell to a BAM file and sort the alignments per chromosome using samtools (4).

samtools view -hbS -o singlecell.bam singlecell.sam

samtools sort singlecell.bam singlecell.sorted.bam

82) Use Picard (http://broadinstitute.github.io/picard/) to mark PCR-duplicate reads in the single-cell BAM file, enabling their exclusion from downstream analyses.

java –jar picard/MarkDuplicates.jar I=singlecell.sorted.bam O=singlecell.sorted_dedup.bam M=singlecell.txt REMOVE_DUPLICATES=true AS=true

Genome Data Analysis - Read mapping performance

83) The read mapping statistics can be explored using Samtools. Cells having a low fraction of the sequenced reads mapped (e.g. 2% or less) should be excluded from downstream analyses.

samtools flagstat singlecell.sorted_dedup.bam

Genome Data Analysis - Calculating the genome coverage

84) The breadth of genome coverage for each single cell can be calculated using genomeCoverageBed from bedtools (5).

samtools view -b singlecell.sorted_dedup.bam | bedtools/genomeCoverageBed -ibam stdin -g GRCh37.chromosome.sizes >coverage_singlecell.txt

Genome Data Analysis - DNA copy number profiling

! Critical step: For DNA copy number profiling of single cells different tools are available, as Ginkgo (6) and SNS (7), which are directly compatible with the G&T-seq protocol. Step-by-step protocols of these tools are available: Ginkgo (http://qb.cshl.edu/ginkgo/?q=/fYIPzxxLZbaMO4NpMOrU), and SNS in the publication of Baslan et al. (7). These can be applied without modification. Ginko requires a BED file per single cell as an input to determine the DNA copy number. This can be done using Bedtools (5).

bedtools/bamToBed -i singlecell.sorted_dedup.bam >singlecell.sorted_dedup.bed

For SNS, follow steps 66 to 86 in the protocol by Baslan et al. (7)

85) We apply a similar protocol as Baslan et al (7) containing minor modifications. We apply BWA as a mapping algorithm, and retain uniquely mapping reads having the flag XT:A:U of BWA. Using flexible bins of 500,000 unique mappable positions, calculate the logR-values for each bin genome wide as

> logR= log2(#uniquely mapping reads in a specific bin/median of uniquely mapping reads across all genomic bins)

and correct these logR-values per single cell for %GC-bias using a loess fit in R and further normalize the values to the median of the genome-wide logR values. The corrected logR values are segmented using piecewise constant fitting (8) (penalty gamma 15) and subsequently converted to DNA copy number using the statistical programming language, R (http://cran.r-project.org).

> $CN = 2\log R.\ \Psi$

> (where Ψ is the average ploidy of the cell)

86) The DNA copy number of each cell can be visualized in genome-wide plots (Fig. 3) and also heatmaps by using the heatmap.2 package in R or the Circos software (9).

Genome Data Analysis - Quality control of the single-cell DNA copy number profiles

87) For quality control, calculate the Median Absolute Pairwise Difference (MAPD) (10) of the genome-wide logR values per single cell. The higher the MAPD value the higher the overall noise in the copy number data. We usually discard single cells having a MAPD score higher than 0.6 or 2 when the cell's DNA was amplified with PicoPLEX or MDA, respectively.

> $MAPD = \text{median}\ (|\ \log R \text{ of } bin_{k+1} - \log R \text{ of } bin_k\ |)$

Genome Data Analysis - Detection of genomic single nucleotide variations (SNVs)

! Critical step: The quality of sequencing reads and the aligned bases can vastly influence the accurate detection of SNVs. We use GATK (11) to first recalibrate base quality scores of sequencing-by-synthesis reads in an aligned BAM file, and then call SNVs.

88) Correct mapping-associated artifacts by re-aligning the reads around indels for each single cell using GATK software (11, 12). Using the publically available list of known indel regions of the reference genome like GRCh37, create a table of the intervals around which realignment has to be performed.

java GenomeAnalysisTK.jar -T RealignerTargetCreator -R GRCH37.fa -o singlecell.bam.table -I singlecell.bam –known gold_indels.GRCh37.vcf

89) Perform read realignment per single cell around the generated table of intervals in the previous step.

java GenomeAnalysisTK.jar -T IndelRealigner -R GRCH37.fa -o singlecell.realigned.bam -I singlecell.bam -targetIntervals singlecell.bam.table –known gold_indels.GRCh37.vcf

90) The BAM files with realigned sequences are further recalibrated to correct for the base quality scores.

java GenomeAnalysisTK.jar -R GRCH37.fa -T BaseRecalibrator -I singlecell.realigned.bam --knownSites latest_dbsnp.vcf -o singlecell.Realigned.basetable

java GenomeAnalysisTK.jar -R GRCH37.fa -T PrintReads -I singlecell.realigned.bam -BQSR singlecell.Realigned.basetable -o singlecell.recalibrated.bam

91) For each single cell, remove the PCR duplicates from recalibrated BAM file using Picard, see step 82.

Genome Data Analysis - Detection of SNVs using GATK

92) Deduce the genotypes and calculate the allelic frequencies to detect genomic variant sites in an individual cell or multiple single cells together with minimum coverage of 2.

java GenomeAnalysisTK.jar -R GRCH37.fa -T UnifiedGenotyper –I singlecell.recalibrated.bam -o singlecell.recalibrated.vcf -metrics singlecell.recalibrated.metrics -stand_call_conf 20.0 -stand_emit_conf 10.0 -dcov 200 -glm SNP

93) Using VariantRecalibrator from GATK, build the recalibration model based on resource sets like the hapmap and 1000 genomes project and assign the probabilities for each variant detected to filter the low-quality variants and retain only the high-quality variant sites.

java GenomeAnalysisTK.jar -R GRCH37.fa -T VariantRecalibrator -input singlecell.recalibrated.vcf -resource:hapmap hapmap.vcf -an QD -an FS -an MQRankSum -an ReadPosRankSum -mode SNP -tranche 100.0 -tranche 99.9 -tranche 99.0 -tranche 90.0 -recalFile singlecell.recalibrated.recal -tranchesFile singlecell.recalibrated.tranches -rf BadCigar -an DP

```
java GenomeAnalysisTK.jar -R GRCH37.fa -T ApplyRecalibration -input
singlecell.recalibrated.vcf -mode SNP --ts_filter_level 99.0 -recalFile
singlecell.recalibrated.recal -tranchesFile singlecell.recalibrated.tranches -o
singlecell.recalibrated.vcf
```

Transcriptome Data Analysis

94) From the single-cell WTA reads, remove library adapter sequences and quality trim (minimum score 20 "-q 20") the reads using software as Cutadapt and TrimGalore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), with "-a" and "-a2" (adapter sequences), --paired, --retain_unpaired (to keep unpaired reads passing QC) and default values for the minimum read length (20 nt).

```
trim_galore --a adaptor_5' --a2 adaptor_3' --paired --retain_unpaired singlecell_R1.fastq.gz
singlecell_R2.fastq.gz
```

95) Align the trimmed reads to the reference transcriptome/genome of the correct species and ERCC sequences using programs designed for mapping RNA-seq data such as TopHat2 (ref), -G (providing an annotation file - gtf format).

```
tophat -G annotation.gtf singlecell_1_val_1.fq.gz singlecell_2_val_2.fq.gz
```

96) Using the most recent genome annotation, quantify uniquely mapping reads and normalize read counts to take into account library size variation and RNA composition bias using HTSeq (ref) (--mode=intersection-strict --stranded=no --type=exon --idattr=gene_id) and DESeq2 (ref), for instance.

```
samtools view -q MAPPING_QUALITY accepted_hits.bam | python -m HTSeq.scripts.count --
mode=intersection-strict --stranded=no --type=exon --idattr=gene_id – annotation.gtf >
htseq_counts.out
```

97) Using genes whose transcripts are longer than 2kb, 10kb and 15kb estimate the nucleotide read coverage over concatenated exons, introns, 3'UTRs and 5'UTRs to quantify sequencing bias.

98) To highlight libraries with potential issues and the impact of sequencing depth on gene expression calling, for each cell build the distribution of the number of expressed genes as a factor of increasing expression cutoffs (Transcript Per Million –TPM, for instance from 0.5 TPM to 1000 TPM), Figure 4. This provides a good overview of the number of transcripts detected per cell at the selected cutoff and for the exclusion of failed cells.

99) Based on the number of uniquely mapped reads and these distributions, filter the single-cell transcriptome data based on the number of mapped reads and the number of genes expressed above a set threshold. Based on the analysis of the distribution of the number of genes identified across We previously applied a threshold of at least 3,500 genes with a Transcript Per Million (TPM) ≥ 1; i.e. single cells demonstrating less than 3,500 genes expressed at a TPM ≥ 1 were excluded from further analyses (Figure 4).

100) Of single cells that pass these quality criteria, the normalized read counts can be analysed further, for example for differential gene expression analysis using DESeq2.

Integrating Whole Chromosome aneuploidy copy number and expression data

101) Using the most recent genome annotation, concatenate gene models within a same chromosome or chromosomal arm

102) For each library, count reads mapping uniquely over these new annotations and estimate the expression (RPKM: Read per Kilobase of transcript per Million mapped reads)

103) Median-center the chromosomal expression values of each cell using chromosome expression values computed in a controlled cell line with the same genetic background. Cells with chromosomal copy number will appear as outliers having, relative to control cells, either increased overall chromosomal expression in case of chromosome gain or decreased expression associated with chromosome loss.

104) To investigate the relationship between chromosomal copy number and gene expression, the values for each chromosome, or chromosome arm generated in step 102 can be compared with the observations of chromosomal copy number from step 85.

TROUBLESHOOTING

Step	Problem	Possible Reason	Solution
46	No cDNA after amplification – no WTA product visible during QC with the Agilent Bioanalyser	Cell not sorted into well	Check for presence of ERCC transcripts on the Bioanalyser electropherogram – if present, the amplification has most likely worked, but no cell was originally deposited into the lysis buffer. Optimise sorting conditions to maximise accurate deposition of single cells.
46	Poor cDNA yield – low level of product visible during QC with the Agilent Bioanalyser	Oligo-dT30VN beads not sufficiently mixed during DNA/RNA separation	Ensure that Oligo-dT30VN beads and cell lysate are well mixed during incubation (step 20)
46	Poor cDNA yield – low level of product visible during QC with the Agilent Bioanalyser	Loss of mRNA or Oligo-dT30VN beads on tips used for mixing	Ensure Oligo-dT30VN beads are fully captured on the magnet before eluting the supernatant
46	Poor cDNA yield – low level of product visible during QC with the Agilent Bioanalyser	Insufficient amplification - Different cell types can vary in terms of their cDNA yield. Smaller cells occasionally require more amplification to generate similar amounts of product	Increase the number of PCR cycles in step 34.
46	Poor cDNA yield – low level of product visible during QC with the Agilent Bioanalyser	Poor sample quality – cells may be dying during isolation and FACS	Perform live/dead staining on the cells, and gate for living cells in the FACS sort.

			Optimise the cell isolation protocol for your tissue of interest to minimise cell death during processing.
46	Poor cDNA yield – low level of product visible during QC with the Agilent Bioanalyser	Oligo-dT30VN beads settling during RT	Ensure Oligo-dT30VN beads are mixing during the reverse transcription reaction (step 29)
46	"Noisy" amplification – short products detected during QC with the Agilent Bioanalyser	Primer contamination or TSO concatamerisation	Reduce the amount of TSO added to the reverse transcription reaction. We have found that 1:10 – 1:50 dilutions of the working stock described here maintain good performance while minimising the amount of primer concatamerisation
46 and 57	No WGA product and no WTA product	If a well generates no detectable WTA and no WGA product, it is likely the cell was not sorted into the well	Consider optimisation of the sorting conditions to maximise accurate deposition of single cells in individual wells.
57 and 58	WGA product from empty well controls	Contamination during plate set up	Ensure pre-amplification steps are carried out in dedicated "clean" areas with dedicated reagents.
75	Nextera Library prep generates longer or shorter fragments than expected	Under (longer fragments) or over (shorter fragments) fragmentation of the input DNA by the Tn5 transposase.	The input amount of WTA or WGA product for Nextera library preparation should be approximately 0.2-1 ng/µL; normalise the DNA concentration to this range before performing library prep.
86	The single-cell DNA copy number profile is overly noisy	Use of a WGA method that is	Consider changing to another commercially available WGA

		suboptimal for single-cell DNA copy number profiling	method that is optimal for DNA copy number profiling (e.g. PicoPLEX, GenomePlex, MALBAC, Ampli1). As a control process at least 10 single cells (of the same cell type or tissue) according to the manufacturer's instructions for the WGA method in parallel.
86	The single-cell DNA copy number profile is overly noisy	Incomplete transfer of the cell's gDNA during the separation and Oligo-dT30VN bead washing steps 21-25	Optimise liquid handling by carefully monitoring whether all cell lysate and wash buffer is transferred to the 'gDNA collection' 96-well plate
86	The single-cell DNA copy number profile is overly noisy	Dying cells or cells in S-phase of cell cycle can give unusually uneven copy number profiles	Include live/dead staining in any sorting to minimise the capture of dead or dying cells. DNA binding dyes can be employed to exclude cells which are in S-phase from further analysis

TIMING:

Day 1:

Steps 1-4: Preparation of single-cell samples (Timing: 30 mins - 1 hour)

Steps 5-13: Preparation of oligo-dT30VN labelled beads (Timing: 1 hour)

Steps 14-16: Addition of ERCC spike-ins (Timing: 10 mins)

Steps 17-28: Physical separation of mRNA and gDNA (Timing: 1.5 hours)

Steps 29-30: Reverse Transcription (Timing: 2 hours)

Steps 31-34: PCR amplification of cDNA (Timing: 3 hours)

Day 2:

Steps 35-45: Purification of Amplified cDNA (Timing: 1 hour)

Step 46: Quality Control of Amplified cDNA (Timing: 1 hour)

Steps 47-54: Purification of gDNA (Timing: 1 hour)

Step 55 (***A*** or ***B***): WGA (Timing: 3 hours)

Steps 56-58: Purification and QC of Amplified gDNA (Timing: 1 hour)

Day 3:

Steps 59-72: Tagmentation-based library preparation using the Nextera XT Sample Preparation kit (Illumina) (Timing: 2 hours)

Steps 73-76: Pooling and quantification of libraries (Timing: 3 hours)

Day 4-7:

Step 77: Library sequencing ~2 days (depending on sequencing technology)

Steps 78-104: Data analysis 2 days-2 weeks (depending on analysis strategy)

ANTICIPATED RESULTS

Step 46

The first QC we generally perform during the G&T-seq protocol is assessing the quality of the cDNA generated from the modified Smart-seq2 reaction using an Agilent Bioanalyser. The reaction should behave very similarly to the Smart-seq2 reaction, and thus generate cDNA between 0.5 and 2 kb reaching a maximum at about 1-1.5 kb (Fig. 2A). In some cases, the cDNA can have a broader size distribution, which is predominantly made up of shorter fragments, indicating that the original RNA has been degraded (Fig. 2B). A common cause of failure at this stage is the failure to deposit a cell into the lysis buffer: in this case no DNA or RNA amplification is seen (Fig. 2C).

Step 57

Both PicoPLEX and MDA WGA protocols generate usually 0.5 up to 2 µg of amplified DNA per single cell.

Step 58

WGA products can also be analysed using the Agilent bioanalyser. For PicoPLEX WGA products, we expect fragment sizes between 300 and 2,000 bp in length (Fig. 2D). MDA WGA products are generally longer (1-20 kb).

Step 75

Nextera XT library preparation generates sequencing ready fragments between 300 and 800 bp in length (Fig. 2E). These fragments should then be quantified and diluted to a concentration suitable for sequencing.

Step 82 (end result)

We typically paired-end sequence both DNA and RNA libraries from 96 samples over 2 lanes each on a HiSeq 2500 running in rapid mode. This generates about 4-6 million reads per cell for both the genome and the transcriptome. Representative data and sequencing statistics from a variety of cell types (human cancer and transformed cell lines, iPSC derived neurons and mouse blastomeres) have been published[17].

Step 86 (end result)

Fig. 3 demonstrates the anticipated and potential outputs of genome wide copy number analysis of single cells. MDA products (Fig 3, A-C) generate noisy copy number profiles, but when QC criteria are met – in this case, the data has a MAPD score < 2 - the data can still demonstrate copy number changes between a HCC38 cancer cell (Fig 3A) and a matched normal HCC38-BL control (Fig 3B). Figure 3C shows a cell which has failed to meet this QC cut-off, and its data is too noisy to use for downstream analysis. In general, PicoPLEX genome wide copy number profiles (Fig. 3, D-F) display significantly less noisy copy number data than MDA and thus give more accurate copy number estimates. We use a more stringent MAPD cut-off for PicoPLEX data. Results are shown for a HCC38 cancer cell (Fig. 3D) and a matched normal HCC38-BL control (Fig 3E) which both passed QC (MAPD < 0.6). A failed PicoPLEX cell (MAPD > 0.6, Fig, 3F) lacks broad coverage of the genome and is also too noisy to provide accurate copy number estimates.

Step 99 (end result)

Typically, we detect many thousands of transcripts per cell, depending on the cell type, size and state. Fig. 4 shows the distribution of genes detected at different TPM cut-offs. We use TPM ≥ 1 as an expression cut-off, then use this as a means to calculate the number of genes expressed per cell. From our analysis of HCC38 and HCC38-BL cells, we have excluded cells which have less than 3,500 transcripts expressed at this level, however both of these cut-offs can be varied depending on application.

Figure Legends:

Figure 1:

Stepwise overview of the G&T-seq method. DNA and RNA from a single cell are physically separated using a modified oligo-dT bead based strategy. Subsequently both DNA and RNA are amplified and sequencing libraries are prepared. DNA amplification can be performed using either MDA or PicoPLEX, depending on the desired experimental outcome.

Figure 2:

Representative Agilent Bioanalyser electropherogram plots from various stages of single-cell RNA and DNA amplification and library preparation. (A) A representative successful cDNA amplification, with a fragment distribution of between 500 and 2,500 bp, with an average size of 1.5-2 kb. (B) A failed reaction, most likely due to cell death or incomplete lysis. The cDNA is shorter and has a wider size distribution. (C) A failed reaction, most likely due to the absence of a cell in the original well. (D) A typical PicoPLEX WGA product from G&T-seq having a size range of 400 to 2,000 bp. (E) A typical Nextera XT library pool with a size distribution of 200 to 800 bp in length.

Figure 3:

Example genome-wide copy number plots. The grey bars represent the estimated relative copy number at each of 500,000 bins across the human genome, while the calculated copy number is shown by the horizontal red lines. Broad variation in copy number between bins indicated noisier data, and as such indicate the lower reliability of the data and a reduced ability to estimate true copy number. Panels A-C show MDA derived DNA from (A) an MDA product from a single HCC38 cancer cell which passed QC (MAPD score < 2), (B) a matched normal cell from the HCC38-BL cell line which passed QC (MAPD score < 2) and (C) a HCC38-BL cell which failed QC (MAPD score > 2). D-F show the same but from PicoPLEX generated libraries (D) from a HCC38 cell which passed QC (MAPD score < 0.6), (E) from a HCC38-BL which passed QC (MAPD score < 0.6), and (F) a HCC38-BL cell which failed QC (MAPD score > 0.6).

Figure 4:

Transcript detection in G&T-seq data from HCC38 and HCC38-BL cell lines. This plot shows the number of transcripts detected at a distribution of TPM ranges. Failed cells will often have low transcriptional complexity – i.e. very few transcripts are detected even at low levels. Here, a gene expression cut-off of TPM ≥ 1 has been applied, and subsequently cells with less than 3,500 transcripts expressed at this level have been excluded.

Contributions:

ICM, CPP and TV devised and developed the method and wrote the manuscript. MJT assisted with method development. WH and PK performed bioinformatic analysis of data.

Acknowledgements:

This work was supported by the Wellcome Trust (to T.V. & C.P.P.) and funding from the Research Foundation Flanders (FWO) and the KU Leuven to T.V. [FWO–G.0687.12; KU Leuven SymBioSys, PFV/10/016]. W.H. and C.P.P. were also funded by the Medical Research Council. M.J.T. is supported by a Wellcome Trust Sanger Institute Clinical Ph.D. Fellowship (UK).

Competing financial interests:

The authors declare no competing financial interests.

Correspondence to:

Iain C Macaulay (iain.macaulay@tgac.ac.uk)

- Telephone: +44 (0)1603 450861
- Fax +44 (0)1603 450045

Chris P Ponting (chris.ponting@igmm.ed.ac.uk) or

- Telephone: +44 (0)131 651 8500
- Fax: +44 (0)131 651 8800

Thierry Voet (thierry.voet@med.kuleuven.be)

- Telephone: +32 (0)16 330841
- Fax: +32 (0)16 346060

References

1 Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature reviews. Genetics* **14**, 618-630, doi:10.1038/nrg3542 (2013).

2 Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol Cell* **58**, 610-620, doi:10.1016/j.molcel.2015.04.005 (2015).

3 Wang, Y. & Navin, N. E. Advances and applications of single-cell sequencing technologies. *Mol Cell* **58**, 598-609, doi:10.1016/j.molcel.2015.05.005 (2015).

4 Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90-94, doi:10.1038/nature09807 (2011).

5 Wang, Y. *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155-160, doi:10.1038/nature13600 (2014).

6 Voet, T. *et al.* Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic acids research* **41**, 6119-6138, doi:10.1093/nar/gkt345 (2013).

7 Evrony, G. D. *et al.* Single-neuron sequencing analysis of l1 retrotransposition and somatic mutation in the human brain. *Cell* **151**, 483-496, doi:10.1016/j.cell.2012.09.035 (2012).

8 Evrony, G. D. *et al.* Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49-59, doi:10.1016/j.neuron.2014.12.028 (2015).

9 Hou, Y. *et al.* Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* **148**, 873-885, doi:10.1016/j.cell.2012.02.028 (2012).

10 Xu, X. *et al.* Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* **148**, 886-895, doi:10.1016/j.cell.2012.02.025 (2012).

11 Leung, M. L., Wang, Y., Waters, J. & Navin, N. E. SNES: single nucleus exome sequencing. *Genome Biol* **16** (2015).

12 de Bourcy, C. F. *et al.* A quantitative comparison of single-cell whole genome amplification methods. *PloS one* **9**, e105585, doi:10.1371/journal.pone.0105585 (2014).

13 Hou, Y. *et al.* Comparison of variations detection between whole-genome amplification methods used in single-cell resequencing. *Gigascience* **4**, 37, doi:10.1186/s13742-015-0068-3 (2015).

14 Huang, L., Ma, F., Chapman, A., Lu, S. & Xie, X. S. Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications. *Annu Rev Genomics Hum Genet* **16**, 79-102, doi:10.1146/annurev-genom-090413-025352 (2015).

15 Gundry, M., Li, W., Maqbool, S. B. & Vijg, J. Direct, genome-wide assessment of DNA mutations in single cells. *Nucleic acids research* **40**, 2032-2040, doi:10.1093/nar/gkr949 (2012).

16 Wang, J., Fan, H. C., Behr, B. & Quake, S. R. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* **150**, 402-412, doi:10.1016/j.cell.2012.06.030 (2012).

17 Macaulay, I. C. *et al.* G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nature methods* **12**, 519-522, doi:10.1038/nmeth.3370 (2015).

18 Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622-1626, doi:10.1126/science.1229164 (2012).

19 Macaulay, I. C. & Voet, T. Single cell genomics: advances and future perspectives. *PLoS genetics* **10**, e1004126, doi:10.1371/journal.pgen.1004126 (2014).

20 Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell reports* **2**, 666-673, doi:10.1016/j.celrep.2012.08.003 (2012).

21 Islam, S. *et al.* Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nature protocols* **7**, 813-828, doi:10.1038/nprot.2012.022 (2012).
22 Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776-779, doi:10.1126/science.1247651 (2014).
23 Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature methods* **6**, 377-382, doi:10.1038/nmeth.1315 (2009).
24 Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature methods* **10**, 1096-1098, doi:10.1038/nmeth.2639 (2013).
25 Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols* **9**, 171-181, doi:10.1038/nprot.2014.006 (2014).
26 Sasagawa, Y. *et al.* Quartz-Seq: a highly reproducible and sensitive single-cell RNA-Seq reveals non-genetic gene expression heterogeneity. *Genome Biol* **14**, R31, doi:10.1186/gb-2013-14-4-r31 (2013).
27 Pan, X. *et al.* Two methods for full-length RNA sequencing for low quantities of cells and single cells. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 594-599, doi:10.1073/pnas.1217322109 (2013).
28 Vanneste, E. *et al.* Chromosome instability is common in human cleavage-stage embryos. *Nature medicine* **15**, 577-583, doi:10.1038/nm.1924 (2009).
29 Dey, S. S., Kester, L., Spanjaard, B., Bienko, M. & van Oudenaarden, A. Integrated genome and transcriptome sequencing of the same cell. *Nature biotechnology* **33**, 285-289, doi:10.1038/nbt.3129 (2015).
30 Li, W., Calder, R. B., Mar, J. C. & Vijg, J. Single-cell transcriptogenomics reveals transcriptional exclusion of ENU-mutated alleles. *Mutat Res* **772**, 55-62, doi:10.1016/j.mrfmmm.2015.01.002 (2015).
31 Han, L. *et al.* Co-detection and sequencing of genes and transcripts from the same single cells facilitated by a microfluidics platform. *Sci Rep* **4**, 6485, doi:10.1038/srep06485 (2014).
32 Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature methods* **11**, 817-820, doi:10.1038/nmeth.3035 (2014).
33 Angermueller, C. *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nature methods* **13**, 229-232, doi:10.1038/nmeth.3728 (2016).
34 Jiang, L. *et al.* Synthetic spike-in standards for RNA-seq experiments. *Genome research* **21**, 1543-1551, doi:10.1101/gr.121095.111 (2011).
35 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
36 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).
37 Marcel, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 3 (2011).
38 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
39 Garvin, T. *et al.* Interactive analysis and assessment of single-cell copy-number variations. *Nature methods* **12**, 1058-1060, doi:10.1038/nmeth.3578 (2015).
40 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome research* **19**, 1639-1645, doi:10.1101/gr.092759.109 (2009).
41 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).
42 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36, doi:10.1186/gb-2013-14-4-r36 (2013).
43 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).