Nucleic Acids Research, 2025, **53**, gkaf173
https://doi.org/10.1093/nar/gkaf173
Methods

Single-cell genome and transcriptome sequencing without upfront whole-genome amplification reveals cell state plasticity of melanoma subclones

Koen Theunis [1,†], Sebastiaan Vanuytven [1,†], Irene Claes [1,2,‡], Jarne Geurts [1,2,‡], Florian Rambow [3,4,11,‡], Daniel Brown [5], Michiel Van Der Haegen [1], Oskar Marin-Bejar [3,4,6], Aljosja Rogiers [3,4], Nina Van Raemdonck [3,4], Eleonora Leucci [7,8], Jonas Demeulemeester [1,2,9], Alejandro Sifrim [10,2], Jean-Christophe Marine [3,4], Thierry Voet [1,2,*]

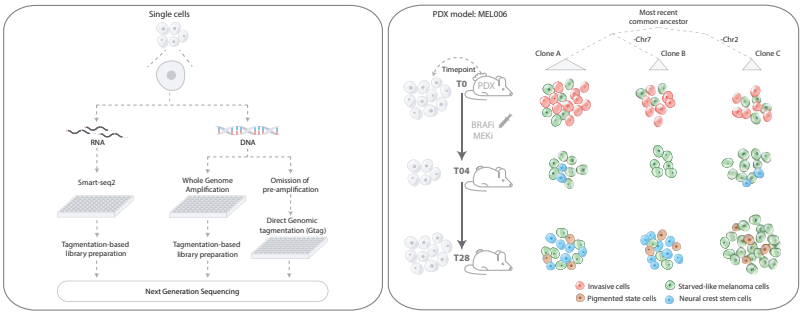
[1]Laboratory of Reproductive Genomics, Department of Human Genetics, KU Leuven, 3000 Leuven, Belgium
[2]KU Leuven Institute for Single Cell Omics (LISCO), KU Leuven, 3000 Leuven, Belgium
[3]Laboratory for Molecular Cancer Biology, Department of Oncology, KU Leuven, 3000 Leuven, Belgium
[4]VIB Center for Cancer Biology, VIB, 3000 Leuven, Belgium
[5]Advanced Technology and Biology Division, The Walter and Eliza Hall Institute of Medical Research, 3052 Parkville, Australia
[6]Present address: Germans Trias i Pujol Health Science Research Institute (IGTP), Translational Program in Cancer Research (CARE), 08916 Badalona, Catalonia, Spain
[7]Laboratory for RNA Cancer Biology, Department of Oncology, KU Leuven, 3000 Leuven, Belgium
[8]Trace, Leuven Cancer Institute, Department of Oncology, KU Leuven, 3000 Leuven, Belgium
[9]Present address: VIB Center for Cancer Biology, 3000 Leuven, Belgium; Laboratory for Integrative Cancer Genomics, Department of Oncology, KU Leuven, 3000 Leuven, Belgium
[10]Laboratory of Multi-omic Integrative Bioinformatics, Department of Human Genetics, KU Leuven, 3000 Leuven, Belgium
[11]Present address: Department of Applied Computational Cancer Research, IKIM, University Hospital Essen, 45147 Essen, Germany

[*]To whom correspondence should be addressed. Tel: +32 16 33 08 41; Email: Thierry.Voet@kuleuven.be
[†]The first two authors should be regarded as Joint First Authors.
[‡]These co-second authors contributed equally to this work.

Abstract

Single-cell multi-omics methods enable the study of cell state diversity, which is largely determined by the interplay of the genome, epigenome, and transcriptome. Here, we describe Gtag&T-seq, a genome-and-transcriptome sequencing (G&T-seq) protocol of the same single cells that omits whole-genome amplification (WGA) by using direct genomic tagmentation (Gtag). Gtag drastically decreases the cost and improves coverage uniformity at single-cell and pseudo-bulk levels compared to WGA-based G&T-seq. We also show that transcriptome-based DNA copy number inference has limited resolution and accuracy, underlining the importance of affordable multi-omic approaches. Applying Gtag&T-seq to a melanoma xenograft model before treatment and at minimal residual disease revealed differential cell state plasticity and treatment response between cancer subclones. In summary, Gtag&T-seq is a low-cost and accurate single-cell multi-omics method that explores genetic alterations and their functional consequences in single cells at scale.

Graphical abstract



Received: April 7, 2023. **Revised:** February 7, 2025. **Editorial Decision:** February 17, 2025. **Accepted:** February 21, 2025

Introduction

Single-cell genomics is key to study genetic heterogeneity in cells of developing and ageing organisms in health and disease. Compared to conventional bulk DNA sequencing, these methods offer enhanced resolution to characterize the mutational processes operative in tissues and organs [1–3]. Leveraging the detected somatic mutations, the phylogeny of the cells and tissues can be reconstructed [4–8, 9], increasing our understanding of the developmental and evolutionary mechanisms of (diseased) tissues as well as the aetiological role of the acquired mutations in the phenotype [10, 11].

The majority of single-cell genome sequencing methods require whole-genome amplification (WGA) to preamplify the genomic DNA (gDNA) and obtain enough material for preparing a sequencing library. In general, WGA methods apply either polymerase chain reaction (PCR, e.g. DOP-PCR, LM-PCR, Ampli1) [12–14], multiple displacement amplification (MDA; e.g. RepliG, GenomiPhi) [15], or a combination of both (e.g. MALBAC, picoPlex) [16, 17]. Amplification, however, introduces coverage bias and nucleotide errors due to buffer and polymerase artifacts. Additionally, the process of pre-amplification followed by library preparation is low throughput, labour intensive, and costly. Recently, tagmentation-based methods were reported, preparing sequencing libraries from single-cell gDNA without upfront WGA [18–24, 25], and providing a more uniform representation of the cell's genome, while enabling a higher throughput.

Single-cell multi-omics technologies have been developed to simultaneously assess the genome, epigenome, transcriptome, and/or selected proteins of single cells by either joint or parallel processing of the molecular hierarchy or by data analysis methodologies [24–32, 33]. Methods for the joint processing of gDNA and polyadenylated RNA of single cells that are individually isolated and lysed in a reaction container either use a single-tube gDNA-and-RNA preamplification (DR-seq and scONE-seq) [24, 29] or rely on physical separation of gDNA and RNA before their parallel amplification (G&T-seq and DNTR-seq) [25, 32, 34, 35]. DR-seq preamplifies gDNA and RNA simultaneously, minimizing nucleic acid loss, and then splits the reaction to complete gDNA and RNA sequencing libraries separately. However, its data analysis and interpretation are complicated due to the presence of RNA-derived reads in the gDNA-seq data. While scONE-seq has the advantage of a one-pot reaction up to differentially barcoded gDNA- and RNA-sequencing libraries, both the gDNA and RNA must be sequenced together, making it impossible to re-sequence only the gDNA or RNA at higher depth, and the method is also less flexible for selecting subsets of gDNA or RNA for sequencing. Physical separation of RNA and gDNA, as in genome-and-transcriptome sequencing (G&T-seq) [32] and DNTR-seq [25] yields a more flexible protocol, allowing the RNA- and gDNA-seq protocols to be adapted to the researcher's needs. This is important because protocols for both RNA-seq and gDNA-seq differ significantly in their performance for detecting (full-length) transcripts and classes of genetic variation across cells, respectively. Up to now, gDNA sequencing in G&T-seq necessitates WGA techniques, making large-scale G&T-seq cost-prohibitive and time-consuming. Also, methods where single cells are not individually tubed but rather processed in pools using principles of combinatorial indexing for joint gDNA and polyadenylated RNA analysis have been developed as in sci-L3-RNA/DNA [36]. Other groups have studied the interplay between genome and transcriptome by either inferring copy number alterations (CNA) from gene expression profiles [37, 38] or integrating separate single-cell gDNA and RNA datasets [39,40]. While these approaches have led to new insights into tumour heterogeneity, tumour progression, as well as therapeutic strategies, only direct multi-omics approaches allow genotype–phenotype relations to be unambiguously ascertained [41–43].

Here, we present Gtag&T-sequencing, a method that enables genome-and-transcriptome sequencing of single cells without upfront WGA, enhancing throughput while minimizing coverage bias, amplification noise, and cost. Gtag&T-seq of a patient-derived xenograft (PDX) melanoma model highlights the advantages of genome-based over transcriptome-based CNA inference and the transcriptional effects of complex genomic alterations. We construct a gDNA-based cell lineage tree annotated with RNA-based cell type and state information from the same cells, providing unique insights in the role of genetic and non-genetic factors as well as their interplay during tumour evolution under therapeutic pressure.

Materials and methods

PDX model

The MEL006 cutaneous melanoma PDX model is part of the Trace collection (https://gbiomed.kuleuven.be/english/research/50488876/54502087/Trace/PDX-repository) and was established using lesions derived from a patient undergoing surgery as part of standard treatment at UZ Leuven. Written informed consent was obtained and all procedures were approved by the UZ Leuven/KU Leuven Medical Ethical Committee (S63799, S57760, S58277) and performed in accordance with the principles of the Declaration of Helsinki and with GDPR regulations. The experiments were approved by the KU Leuven animal ethical committee (P164-2019) and performed in accordance with the internal, national, and European guidelines of animal care and use. Single-cell suspensions were implanted subcutaneously in the interscapular fat pad of female NMRI nude BomTac: NMRI-Foxn1nu, 4-week-old females (Taconic Biosciences). Mice were maintained in a pathogen-free facility under standard housing conditions with continuous access to food and water. The health and welfare of the animals was supervised by a designated veterinarian. The KU Leuven animal facilities comply with all appropriate standards (cages, space per animal, temperature [22°C], light, humidity, food, and water), and all cages are enriched with materials that allow the animals to exert their natural behaviour. Mice used in the study were maintained on a diurnal 12-h light/dark cycle. MEL006 was derived from a female, drug-naïve melanoma patient. When the tumour reached 1000 mm^3, the mice were randomly assigned to the different experimental groups. Mice were treated daily by oral gavage with a capped dose of 600 μg dabrafenib and 6 μg trametinib (DT) in 250 μl total volume.

Sample preparation for single-cell sorting

HCC38 breast cancer cells and HCC38-BL lymphoblastoid cells were cultured in Dulbecco's Modified Eagle Medium (DMEM/F12) containing 10% fetal bovine serum (FBS) at 37°C in a 5% CO2 incubator [44]. Trypsinized HCC38 cells and HCC38-BL cells were washed in fresh culture medium and then resuspended in FACS sorting buffer (DMEM/F12 supplemented with 5% FBS, 1 mM EDTA, and 1.5 μM DAPI).

The BD FACS Melody sorter device was used for sorting single cells into 96-well plates (FrameStar®, 4TI-0960/C) containing 2.5 µl RLT plus buffer (Qiagen). Plates were then spun down at 1000 g for 1 min at 4°C, and finally stored at −80°C. MEL006 cells were retrieved as described before [45], resuspended in serum-free DMEM/F12 medium and sorted using the BD FACS Aria III into 96-well plates containing 2.5 µl RLT plus buffer.

Gtag&T

gDNA and RNA separation and DNA precipitation were performed as per the G&T protocol of Macaulay *et al.* [32, 46] on a Hamilton® liquid handling robot. First, the plate containing the lysed cells is supplemented with RNA spike-ins [1 µl of a 1:1 600 000 dilution of ERCC spike-in mixture A (Life Technologies)], which is followed by the addition of oligo-dT conjugated to streptavidin beads. After incubation, the polyadenylated RNAs are collected to the side of the well using a magnet (and processed further according to the G&T-seq protocol), while the DNA present in the supernatant is transferred to a new recipient DNA plate. The wash solution used to rinse the RNA-bead complexes is also added to the DNA plate. After mixing the solution containing the DNA with 0.7:1 ratio AMPure XP beads, the DNA is precipitated using a magnet. For Gtag, the precipitated gDNA was resuspended in 4.5 µl resuspension buffer (0.5X NEB4, 0.37% Igepal CA-630, 0.37% Tween-20). Next, tagmentation is performed in a total volume of 10 µl by adding 5.5 µl tagmentation master mix (5 µl Tagment DNA buffer, 0.1 µl Tagment DNA Enzyme, 0.4 µl nuclease free water) for 10 min at 55°C. The reaction was inactivated by adding 1 µl of 0.44% SDS to the sample and incubating 5 min at 55°C. Then we added 13 µl of Q5 Ultra II (NEB, 2x mastermix), 1 µl S5 primer, and 1 µl S7 primer to the sample. PCR amplification was performed with the following cycling program: 72°C for 3 min; 98°C for 30 s; 16 cycles of 98°C for 10 s, 60°C for 30 s; and 72°C for 30 s, 72°C for 5 min and held at 10°C. In between reaction steps, 96-well plates were placed in an Eppendorf thermomixer at room temperature to mix (1000 rpm) for one minute and briefly centrifuged using a tabletop centrifuge. Finally, the PCR products were pooled per plate and purified using AMPure XP beads (1x ratio).

Sequencing library preparation

DNA and cDNA quality of picoPlex and Smart-seq2 amplification reactions, respectively, was confirmed using the 2100 Bioanalyzer (high sensitivity chip, Agilent). Next, DNA and cDNA concentrations were determined using a Quantifluor® Assay (Promega®). Samples were diluted to 200 or 100 pg/µl for Nextera XT (Illumina) library preparation in respectively one-fourth or one-tenth of the volume recommended by the manufacturer using manual or automated liquid handling. After library preparation, samples were pooled and purified using AMPure XP beads (0.6x ratio). Quality check of the DNA (picoPlex and Gtag) and cDNA library pools was performed using the 2100 Bioanalyzer (high sensitivity chip, Agilent) in combination with Qubit™ HS (High sensitivity) DNA Assay Kit (Invitrogen™) before diluting the pools to a concentration of 4 nM. We used the KAPA Library Quantification Kit for Illumina® platforms (Roche, KK4854) on the LightCycler 480 and diluted the sequencing pools. HCC38 and HCC38-BL samples were diluted to 2 nM and sequenced 51-bp single end on a HiSeq2500 platform. The first batch of MEL006 samples was diluted to 1.5 nM and sequenced on a HiSeq4000 platform (51-bp single-end), while the second batch was diluted to 0.75 nM and sequenced on a NextSeq2000 (2 × 50-bp paired-end).

Processing of genome data and DNA copy number analysis

Single-end sequencing reads obtained were aligned to the GRCh37 human reference genome using BWA-MEM [47]. Samtools was used to sort, index, and sample the mapped BAM files down to 400 000 reads. Our mapping statistics were obtained through samtools (version 1.11) and Picard (version 2.23.8) (http://broadinstitute.github.io/picard/). PCR duplicates were removed with Picard. To create pseudo-bulk genomes, *samtools merge* was used to combine BAM files. DNA copy number (CN) analysis was performed as discussed in Macaulay *et al.* [32, 46]. Segmentation of the corrected logR values was done using piecewise constant fitting, with the penalty parameter (γ) set to 10 for the 500 k UMPs bin genomes and $\gamma = 35$ for the 10 k UMPs bin genomes. Integer DNA CNs were estimated as $2^{\text{logR}} * \Psi$, with the average ploidy of the cell, Ψ, determined as the average CN value with the lowest penalty from a 1.2–6 grid with possible CN values, transformed from segmented LogR values. High penalty values are given to a possible average CN when the sum of squared differences between the unrounded and rounded CN is high.

scDNA-seq quality filtering

For quality filtering of MEL006 genomes, we only processed single cells with at least 400 000 reads before deduplication. HCC38 genomes were only processed if they had at least 100 000 raw reads before deduplication. We calculated the median absolute pairwise difference (MAPD) score for all samples by first measuring the absolute difference between two consecutive logR values, %GC-corrected and normalized, across the genome. Next, the median across all absolute differences is computed. For MAPD cut-offs, genomes only passed if their MAPD score was less than the 75$^{\text{th}}$ percentile +1.5 times the interquartile range (HCC38, $\text{MAPD}_{\text{cut off}} = 0.64$; MEL006, $\text{MAPD}_{\text{cut off}} = 0.76$). An overview of the QC pass/fail samples can be found in Supplementary Data S1 and Supplementary Data S2.

Coverage uniformity calculations

Lorenz curves were computed by taking the cumulative fraction of the covered genome against the cumulative fraction of the mapped bases. From the BAM files, duplicates were first removed, and all genomes were downsampled to 230 000 unique reads (with a quality of at least 20). Gini coefficients were calculated in R using the ineq package (https://cran.r-project.org/web/packages/ineq/).

Heatmaps and clustering

Subclones in the MEL006 dataset were determined using the CopyKit R package (version 0.1.2). The segmented CN values were embedded in two dimensions using the uwot R package (version 0.1.14, min dist = 0, n neighbours = 25, seed = 13). Superclones were identified from the shared nearest neighbour ($K = 15$) graph using the R package igraph (version 1.4.2).

Subclones were subsequently identified from the UMAP embedding using the clustering algorithm hdbscan from the dbscan R package (version 1.1–11). Regions of focal amplifications were determined on the pseudo-bulk Gtag genomes on 10 k UMPs bins segmented by piecewise constant fitting ($\gamma = 35$) using GISTIC2. For each region, 20 additional genomic bins were taken on each flank for visualization. All heatmaps were constructed using the ComplexHeatmap R package (version 2.10.0).

Processing of single-cell RNA seq data

After trimming the adaptor sequences with cutadapt (version 1.13), sequencing reads were aligned to the GRCh37 reference genome, including ERCC sequences using STAR with default parameters (version 2.5.2b). HTseq (version 0.6.0) with the GENCODE H19 transcript annotations were used to generate the count matrix.

Analysis of single-cell RNA-seq: HCC38 and HCC38-BL

Quality control was performed using the scater R package (version 1.28.0): cells with less than 100 000 counts, expression of less than 2000 unique genes, more than 30% counts assigned to mitochondrial sequences or 8% counts belonging to ERCC sequences were removed for downstream analysis. Genes with less than 32 counts across the complete dataset were excluded from downstream analysis. All data analysis was conducted in R version 4.3.0 (CRAN), while plots were created with the ggplot2 (version 3.4.4) R package.

Analysis of single-cell RNA-seq: MEL006

Quality control was performed using the scater R package (version 1.28.0): cells with less than 100 000 counts, expression of less than 1000 unique genes, more than 25% counts assigned to mitochondrial sequences or 15% counts belonging to ERCC sequences were removed for downstream analysis [48]. Genes with less than five counts across the complete dataset were excluded from downstream analysis. Expression value scaling and normalization, cell-cycle regression, batch correction, PCA and UMAP dimensionality reductions and clustering were performed using the Seurat R package (version 4.3.2) [49]. Marker gene discovery was performed using the FindAllMarkers function of the Seurat package using the Wilcoxon Ranked Sum test. The R package presto (version 1.0.0) was used to perform a fast Wilcoxon rank sum test where the AUC value served as input for gene set enrichment analysis (GSEA) with FGSEA R package (version 1.26.0). Gene sets were accessed with msigdbr R package (version 7.5.1). GSEA was also performed with FGSEA using the correlation scores of all genes passing quality control and the CN of the 22q11.21 amplicon. Gene dosage plots were created for all genes located on focal amplifications when at least 5% of the single cells expressed the gene, the unsegmented CN was taken from the bin overlapping with the transcription start site for each gene. The CN of the bin closest to the middle of focal amplification was used as the overall CN of the amplicon. Minimal residual disease (MRD) states were assigned as described in Rambow *et al.* [45]. All data analysis was conducted in Python version 3.9 (Python software foundation) or R version 4.3.0 (CRAN). Plots were created with the ggplot2 (version 3.4.4) and ggpubr (version 0.6.0) R packages.

Benchmarking of inferCNV with G (tag)&T-seq data and classification of the Rambow *et al.* [45] Smart-seq2 data

Raw single-cell RNA-seq reads from 96 normal human melanocytes (ethics approval S63257), as well as the Rambow *et al.* [45] (GEO: GSE116237) data were aligned to the GRCh37 reference genome including ERCC sequences using STAR with default parameters (version 2.5.2b). After creating a count matrix with HTseq (version 0.6.0) and the GENCODE H19 transcript annotations, the data were merged with the G (tag)&T transcriptome counts. The scater R package was used to discard low-quality cells, namely cells with less than 100 000 counts, less than 1000 unique genes expressed, more than 25% counts assigned to mitochondrial sequences or more than 15% counts belonging to ERCC sequences. R package InferCNV (version 1.16.0) was subsequently used to infer CN estimates from the scRNA-seq data using standard parameters for Smart-seq2 and the 6-state Hidden-Markov Model while using the transcriptome data of the normal melanocytes as reference. The mitochondrial and sex chromosomes were excluded from the analysis. The G (tag)&T data were used to benchmark the CN calls obtained with inferCNV, where the DNA-seq data were considered the ground truth. Segments were then classified as (i) true positive if both DNA and RNA CN calls indicated a gain or if both indicated a loss; (ii) true negative if both indicated a neutral CN state; (iii) false negative if the DNA data indicated a CN aberration and the RNA did not; (iv) false positive if the RNA indicated a CN aberration and the DNA did not. The CN calls and modified expression intensities obtained with inferCNV for the G (tag)&T transcriptome data were both used to train five machine learning algorithms (Linear Discriminant Analysis, Classification and Regression Trees, k-Nearest Neighbours, radial function support vector machine (SVM) and random forest) to classify a sample to the correct genomic subclone. Briefly, the caret R package (version 6.0–94) was used to split the data 80–20% after removing highly correlated features (cor > 0.8) and train the classifiers using 10-fold cross-validation. Accuracy was selected as the scoring metric to assess the performance. The best accuracy on the test data was observed for the radial function SVM trained on the modified expression intensities when combining subclones A and B. The model was then used to assign the Rambow *et al.* [45] to either subclone AB or C. Data analysis was conducted in R version 4.3.0 (CRAN) while plots were created with the ggplot2 (version 3.4.4) and ggpubr (version 0.6.0) R packages.

Data from scONE, sci-L3 & DNTR

Single-cell DNA-seq data for human cells used in this study can be accessed in Sequence Read Archive under accession numbers PRJNA603321 (DNTR), PRJNA768428 (scONE), and PRJNA511715 (sci-L3). Sci-L3 fastq files were processed using the pipeline from Yin *et al.* [36]. For the three assays, BAM files were down-sampled to 400 000 reads and processed in the same way as described above. Near-diploid cells processed with DNTR-seq, scONE-seq and sci-L3 with an average ploidy between 1.9 and 2.1 were chosen and compared with single-cell DNA samples of HCC38-BL DNA processed by Gtag and picoPlex using the Wilcoxon test. A comparison was made between the MAPD of logR values between con-

secutive genomic bins, analysed using 500 k UMPs genomic bins, the coverage uniformity and Lorenz curves showing the average coverage uniformity of single-cell genomes.

Code availability

Code is available through the following GitHub link: https://github.com/voetlab/Single_cell_GtagT_Manuscript

Results

WGA-free parallel genome and transcriptome sequencing of single cells

Gtag&T-seq is based on genome-and-transcriptome (G&T) sequencing (Fig. 1A), developed by Macaulay *et al.* [32, 46]. Following physical separation of the gDNA and polyadenylated RNA of the same cell, Gtag&T-seq applies tagmentation to produce a gDNA sequencing library directly from the cell's genome instead of pre-amplifying it with WGA (Fig. 1A). Following tagmentation, PCR adds cell-specific barcodes and sequencing adapters to enable multiplexed low coverage sequencing and cost-effective multi-modal analysis of single cells (Fig. 1B).

The performance of Gtag&T-seq was evaluated against conventional G&T-seq [32], using picoPlex for WGA, for both the HCC38 cancer cell line and its matched normal cell line (HCC38-BL). PicoPlex was chosen for its proven reproducibility and high accuracy in detecting DNA CNs [50–52]. At the RNA level, the datasets were comparable, highlighted by the high correlation of the mean expression per gene for both cell lines (HCC38-BL, $R^2 = 0.83$; HCC38, $R^2 = 0.87$; Fig. 1C and Supplemental Fig. S1A and B). To compare Gtag also with other tagmentation-based methods in a DNA/RNA co-assay, we contrasted our HCC38-BL samples with near-diploid samples processed by DNTR-seq [25], scONE-seq [24], and sci-L3 [36] (Fig 1E and I). To allow for a fair comparison of the genome sequences, gDNA reads of each cell were downsampled to a maximum of 400 000 reads before duplicate removal ('Materials and methods', Supplementary Fig. S1C and D and Supplementary Data S1). Reliable genetic variant detection from single-cell gDNA data largely depends on the noise and coverage uniformity attained by the method. When compared to picoPlex-based G&T-seq, genomic readouts were less noisy for Gtag (HCC38-BL, $P \leq 0.0001$; HCC38, $P \leq 0.0001$) as assessed by MAPD (Fig. 1D), but MAPD values were increased when compared to DNTR-seq, sci-L3-seq and scONE-seq (Fig. 1E). In addition, Gtag improved coverage uniformity in comparison to picoPlex, discernible from Lorenz curves (Fig. 1F and Supplementary Fig. S1E) and compared favourably to picoPlex, scONE and sci-L3 when using the Gini index ($\mu_{Gtag} = 0.20$ *versus* $\mu_{picoPlex} = 0.27$, $\mu_{DNTR} = 0.17$, $\mu_{scONE} = 0.21$ and $\mu_{sciL3} = 0.22$) (Fig. 1I).

Low-depth single-cell genomes can be pooled to derive pseudo-bulk genomes, refining genomic variant calling [18, 23]. To investigate differences in performance resulting from omitting WGA, we compared coverage breadth, uniformity, and noise after merging single-cell genomes (HCC38-BL) *in silico* for both G&T-based multi-omics methods. Plotting the theoretical *versus* the observed coverage breadth for increasing amounts of pooled single-cell genomes (Fig. 1G) showed a rapid saturation of coverage breadth for picoPlex, likely caused by limited random priming during WGA. Pseudo-bulks can be further leveraged to map DNA breakpoints more precisely, conditional on smaller bin sizes not exacerbating noise. We observed that picoPlex suffered from inflated MAPD scores compared to Gtag for pseudo-bulks analysed with smaller bin sizes (Fig. 1H). In addition, coverage was more uniform for the Gtag 20-cell pseudo-bulk HCC38-BL genomes, as evidenced by Gini indexes of 0.08 for Gtag pseudo-bulk and 0.15 for picoPlex pseudo-bulk.

Single-cell and pseudo-bulk analysis of a human melanoma PDX model

The development of resistance to targeted therapy presents a significant clinical challenge. Emerging evidence indicates that both genetic and non-genetic mechanisms conspire to drive resistance. A deeper understanding of the interplay between these mechanisms is essential for designing more effective, long-lasting combination treatments. The Gtag&T-seq method offers an approach to dissecting the respective contribution of these mechanisms at single-cell resolution. We processed 703 and 175 single cells from a melanoma PDX model using Gtag&T- and G&T-seq, respectively. The MEL006 model was established from a patient with a $BRAF^{V600E}$ mutant melanoma who had an almost complete response to combined dabrafenib ($BRAF^{V600E}$ inhibitor) and trametinib (MEK inhibitor) therapy [45]. Single cells from the PDX model were collected before treatment (T0), during treatment (T04), and at MRD (T28), when most of the tumour cells are eradicated by the therapy. At this late time point, a small subset of drug-tolerant cancer cells persists, providing a substrate for relapse.

After removing low-quality genomes ('Materials and methods', Supplementary Fig. S2 and Supplementary Data S2), 494 single-cell Gtag genomes were compared to 142 single-cell picoPlex genomes. DNA CN profiles were called using genomic bins of 500 k unique mappable positions (UMPs), revealing a highly rearranged tumour cell population (Fig. 2A) with an average ploidy of 3.5, indicative of an early whole-genome doubling event, coherent with previous bulk sequencing observations [53].

Three distinct subclones (A–C; Fig. 2A) were identified using the density-based clustering strategy [19], which differed by the presence of an additional copy of chr2 in A and B and a loss of one copy of chr7 in subclone B. Within each subclone, single cells displayed similar CN profiles (mean pairwise R: $A = 0.85$, $B = 0.87$, $C = 0.79$), indicative of stable clonal expansions. To refine the CNA states and breakpoints, pseudo-bulk genomes were created per method for each subclone (A–C) at both time points. Similar to our observations with cell lines, WGA resulted in increased noise and decreased coverage uniformity in pseudo-bulk genomes (Supplementary Fig. S2F), leading to incorrect clustering of the picoPlex-derived pseudo-bulks (Fig. 2B) rather by time point rather than by subclone. In addition, discrepancies are noted in the CN calling for chr1, chr5, chr11, chr19, and chr20. As a consequence, Gtag pseudo-bulk genomes allowed for a more detailed dissection of focal CNAs (Fig. 2C and D).

Detection of subclone-specific focal amplifications at near base-pair resolution using Gtag

Single-cell and pseudo-bulk CNA profiling with 500 k UMP bins (Fig. 2A–D) indicated the existence of highly amplified loci on chr13 and chr22. The elevated amplification levels in these regions, combined with the lower noise and higher

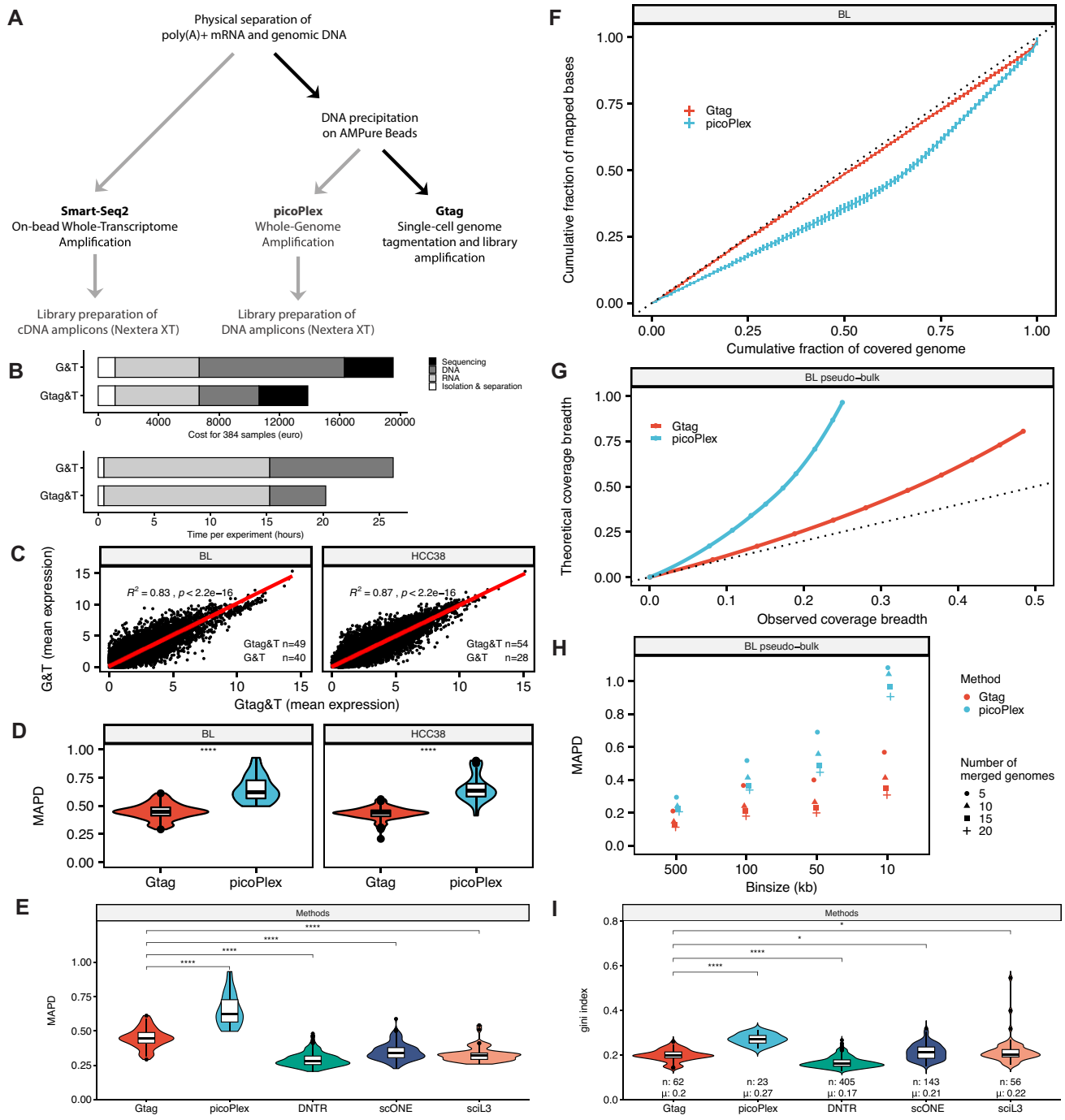


Figure 1. Gtag&T enables cost-effective joint analysis of DNA and RNA in single cells. (**A**) Schematic of the workflow for genome-and-transcriptome separation followed by Gtag (in bold arrows) or conventional G&T-seq with picoPlex and Smart-seq2. (**B**) Comparison of the consumable costs and time required to process 384 cells using G&T or Gtag&T protocols. The cost and time for "Isolation & separation" refers to the conventional G&T separation, while "RNA" and "DNA" include all steps leading to the generation of a sequencing-ready library. "Sequencing" costs were estimated based on a run performed on a NextSeq 2000 P2 flow cell. (**C**) Spearman correlation of mean expression per gene between conventional G&T-seq and Gtag&T-seq for the HCC38 cell line and its matched normal counterpart (HCC38-BL). (**D**) MAPD of logR values between consecutive genomic bins (500 k unique mappable positions UMPs) for HCC38 and HCC38-BL single-cell DNA processed with Gtag or picoPlex. (**E**) MAPD distribution of logR values for HCC38-BL DNA processed by Gtag and picoPlex, compared to near-diploid cells [ploidy: (1.9 – 2.1)] processed with DNTR-seq (HCT-116), scONE (HCT-116) and sciL3 (HAP1/HEK293T), using genomic bins of 500 k UMPs. (**F**) Lorenz curves showing mean genome coverage uniformity across single-cells. The dotted line represents perfect uniformity. Vertical lines on the Lorenz curves represent the standard deviations from the mean. (**G**) Observed genome coverage versus theoretical coverage for merged Gtag and picoPlex genomes. The dotted line indicates the maximum theoretical coverage, assuming no overlapping reads. For each method, increasing amounts of single-cell genomes were merged and plotted, and a local regression was applied. (**H**) MAPD values for merged single-cell genomes processed with Gtag and picoPlex using bin sizes of varying genomic resolution (5, 10, 15, or 20 merged genomes). (**I**) Comparison of genome coverage uniformity using the Gini index for HCC38-BL DNA processed with Gtag & picoPlex, HCT-116 processed with DNTR-seq & scONE and HAP1/HEK293T processed with sciL3. The centreline, top, and bottom of the boxplots represent, respectively, the median, 25th and 75th percentile, and whiskers are 1.5 × IQR. ****: P ≤ 0.0001 (Wilcoxon test). *:P ≤ 0.05 (Wilcoxon test).

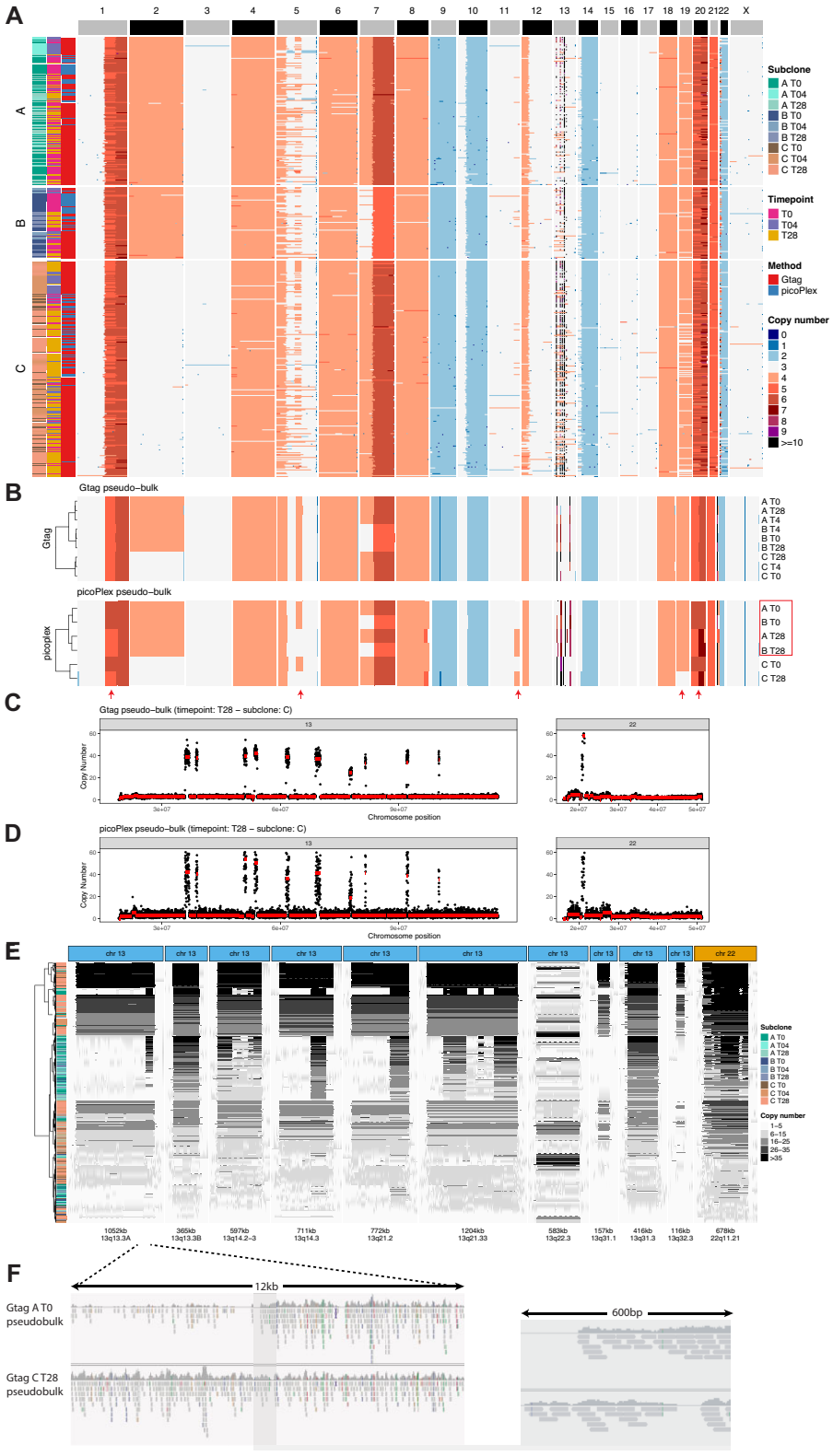


Figure 2. Gtag improves breakpoint and CN calling in a melanoma xenograft model. (**A**) Genome-wide CN heatmap of 636 single cells from a PDX model of human melanoma. Columns correspond to chromosomes (divided into genomic bins of 500 k UMPs) and rows to single cells ($n = 636$). All libraries were down-sampled to 400 000 reads before duplicate removal. (**B**) DNA CN heatmap of pseudo-bulk genomes (genomes are merged per method, subclone, and time point). Gtag and picoPlex pseudo-bulk genomes are plotted separately. Arrows indicate incorrect CNA detection, resulting in incorrect clustering of subclones A and B (rectangle) based on pseudo-bulk genomes. In A and B, cells are annotated for their subclone assignment and time point of isolation, as well as the methodology used for DNA processing. (**C** and **D**) CN plots from respectively Gtag and picoPlex pseudo-bulk genomes (subclone C at T28) of chr13 (left panel) and chr22 (right panel). Dots represent genomic bins of 10 k UMPs, and the line represents the segmented integer DNA CN. (**E**) Heatmap of the focal amplifications on chr13 and chr22 (annotated in top row). The approximate size of the focal amplification is depicted at the top. DNA CNs are indicated in greyscale. (**F**) View, adapted from IGV, of read distribution in Gtag pseudo-bulk genomes (A T0 *versus* C T28). A breakpoint of subclone A is represented at kilobase resolution (left panel) and base pair resolution (right panel).

uniformity of Gtag genomes, allowed us to further refine the breakpoints using 10 k UMP bins. In total, 11 focal amplifications ranging in size from 0.12 to 1.20 Mb were found: 10 on chr13 (q13.3A, q13.3B, q14.2–3, q14.3, q21.2, q21.33, q22.3, q31.1, q31.3, and q32.3) and 1 on chr22 (q11.21). GISTIC2 analysis of 367 skin cutaneous melanoma samples revealed that 13q and 22q arms are frequently amplified [54]. Moreover, a 22q11.21 focal amplification is found in 10% of the CCLE skin cancer cell lines (*e.g.* Hs294-T and COLO679 BRAF (V600) mutant cell lines), 22q11.21 was also identified as recurrently amplified in a pan-cancer analysis [55], and recently associated with inferior survival in acral melanoma [56].

Subclonal differences in the presence, size and CN were observed for the majority of the focal amplifications on chr13 (Fig. 2E). Subclone C was characterised by the presence of 13q31.1 and 13q32.3 amplicons, as well as different breakpoints for 4 focal amplifications (e.g. 13q13.3A, 13q14.3, 13q21.2, and 13q21.33; Fig. 2E). Furthermore, a small number of cells belonging to subclone A and B showed additional breakpoints in 13q14.2–3 and 13q21.33, resulting in multiple smaller amplicons (Fig. 2E) as well as small alterations in 22q11.21 that were only detected in Gtag genomes and not in picoPlex genomes, but are confirmed using pseudo-bulk genomes as well as transcriptomic analysis (Supplementary Fig. S3). Among subclones, most breakpoints remained consistent, resulting in a higher similarity within subclones compared to those outside, as determined by Euclidean distance (*t-test, P < 2.2e-16*).

The CN of the majority of chr13 amplicons was correlated with the amplification level of 22q11.21, suggesting that these amplicons are co-amplified (Supplementary Fig. S4). The absence of correlation for amplicon 13q22.3 is explained by its absence in half of the cells, irrespective of clonal context. All aforementioned subclonal differences were also detected in pseudo-bulk genomes and remained after downsampling to the same depth (Fig. 2C and D and Supplementary Fig. S5). Furthermore, these pseudo-bulk genomes allowed us to pinpoint the breakpoint of the focal amplifications to near base-pair resolution (Fig. 2F).

Although full reconstruction of the amplicon structure is challenging from single-end sequencing data, we were able to make two key observations. First, the subclonal organization of these focal amplifications supports the existence of at least three major genomic subpopulations in this melanoma lesion, with subclones A and B closely related. Second, the heterogeneity in the CN of the amplicons indicates that these are dynamic instead of static DNA entities.

Transcriptome-based DNA CN inference has limited accuracy and fails to detect focal amplifications

To address the added value of direct genome measurements along with transcriptome profiling, we first set out to identify the three major genomic subclones from the single-cell transcriptome data. Of the G (tag)&T single-cell transcriptomes, 636 passed our quality thresholds ('Materials and methods' and Supplementary Data S2) and were used for downstream analysis. We inferred CNAs using averaged gene expression patterns with normal human melanocytes as a reference using inferCNV [37]. Comparison to the matching gDNA-derived DNA CN profiles of the same cells, revealed an average sensitivity and specificity of 52% and 91%, respectively ('Materials and methods'). InferCNV failed to detect the CN of the whole-chromosome gains for chromosomes 4, 6, 8, and 18 for all cells, as well as chr2 for cells belonging to subclones A and B (Fig. 3). Instead, we obtained the correct CN for smaller regions of these chromosomes (Fig. 3), suggesting that not all genes are affected to the same degree by genomic imbalances. Furthermore, the focal amplifications on chromosomes 13 and 22 and the p-ter amplification of chr12 were not detected. These shortcomings notwithstanding, we trained a SVM model on our single-cell genome-and-transcriptome data that achieved a mean classification accuracy of 0.71 (95% CI confidence interval, $CI_{95\%}$= [0.63, 0.79]. Although we identified the three subclones to a degree, we found frequent misclassification of subclone B to A. Combining subclones A and B increased the accuracy of the model to 0.86 ($CI_{95\%}$ = [0.79, 0.91]. This highlights the need for direct multi-omics to accurately dissect both genomic evolution and transcriptome plasticity in full.

Differential expression analysis reveals subtle effects of subclonal chromosomal alterations

Single-cell RNA-seq of the MEL006 PDX melanoma model exposed to BRAFi/MEKi treatment identified four drug-tolerant cell states: a 'starved' (starved-like melanoma cell, SMC) state, a differentiated and 'pigmented' state, a 'neural crest stem cell (NCSC)-like' state and a de-differentiated state also referred to as 'invasive/mesenchymal-like' state [45]. To investigate the effects of the genomic alterations on the cell's transcriptome and the aforementioned drug-tolerant states, we performed pairwise differential gene expression analysis between the three genomic subclones at each timepoint. In total, only 100 differentially expressed genes were identified (FDR < 0.05) (Fig. 4A and Supplementary Fig. S6). No differentially expressed genes were identified between any of the subclones at T04, most likely due to the lower number of cells processed (n = 121). At respectively T0 and T28, we found 39 and 72 differentially expressed genes, of which 11 genes were found to be differentially expressed at both timepoints, including *THAP7, LZTR1,* and *GPC5*. This observation indicates that during treatment, the transcriptomic differences between subclones become greater.

Between subclones A and B, no differentially expressed genes were found, except for *H2AFJ*, indicating that the loss of one copy of chr7 has only subtle effects on the transcriptomes of cells belonging to subclone B (Supplementary Fig. S6). This observation agrees with the difficulties of the SVM model to correctly classify subclone B based on the CNAs obtained with inferCNV. In contrast, we found that the loss of chromosome 2 in subclone C resulted in lower expression of *TMSB10, ARPC2, ATP5G3, PCBP1*, and *OST4* compared to subclones A and B. Meanwhile, *PTMA* and *GYPC* were upregulated only in subclone A, and *RTN4, BIN1*, and *EIF5B* were upregulated only in subclone B. At T28, we observed a shift in the differentially expressed genes towards markers of the MRD states (indicated with symbols on Fig. 4A), as well as genes located on the 22q11.21 focal region and chromosome 7 (indicated with orange and red colours, respectively, on Fig. 4A). *TMSB10* was the only gene upregulated in both subclones at both timepoints, while *OST4* showed higher expression in subclone B at both timepoints. Additionally, *COL5A2* and *SEPT2* were upregulated in subclones A and B compared to C at T28. The invasive state marker *IGFBP5* and

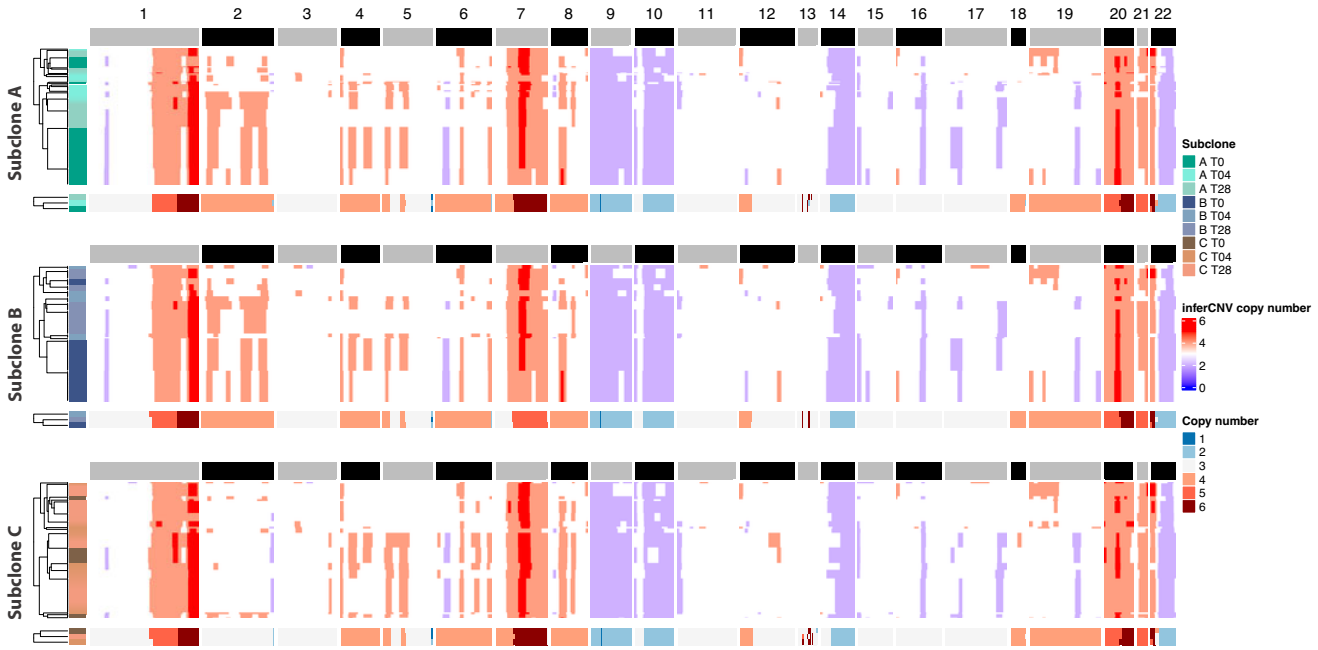


Figure 3. InferCNV is unable to accurately reconstruct CNAs and discern the focal amplifications on chr13 and chr22. CNA heatmap obtained using inferCNV on the transcriptome data of both the G&T-seq and Gtag&T-seq datasets. Single cells are annotated according to their subclone assignment and timepoint of isolation. The reference Gtag&T pseudo-bulk CN profiles per subclone are depicted underneath the inferCNV results for each subclone. All CNs were capped at 6 to allow for a fair comparison between the methods.

SPTBN1 were highly expressed in subclone A, while subclone B showed higher expression of *RHOB*. At T28, *AQP1* (chr7), *GFRA3* (chr5), *MPZ* (chr1), *NGFR* (chr17), *RSPO3* (chr6), *SLC22A17* (chr14), L1CAM (chrX), *GFRA2* (chr8), and *PRIMA1* (chr14), all markers of the NCSC state, showed increased expression in subclone A and/or B, while *TFA2B* (chr6) was higher expressed in subclone A at T0. *CD36* (chr7), a marker for the SMC state, was expressed at higher levels in subclone C compared to subclone A at T28 and the SMC marker *PRELP* (chr1) at T0. Nine differentially expressed genes were located on the focal amplicons of chr13 and chr22. At both T0 and T28, *THAP7* (22q11.21) and *LZTR1* (22q11.21) were expressed at higher levels in subclone C in both T0 and T28, while *KLHL22* (22q11.21) was not differentially expressed compared to subclone B at T0. *PI4KA* (22q11.21), *TUBA3FP* (22q11.21), *P2RX6* (22q11.21) and *TMEM191A* (22q11.21) only appear to be differentially expressed in T28 and *TM9SF2* (13q31.3) only in T0. *GPC5* (13q31.3) had a higher expression in cells belonging to subclone A and B at both T0 and T28. These observations indicate subclonal differences in expression levels of genes located on the focal amplifications or, alternatively, differences in the CN of the amplicons influencing the gene expression between subclones A, B, and C.

Focal amplifications influence subclone-specific gene expression

To assess the effect of ongoing amplicon evolution on the phenotype of the subclones, we calculated the correlation between the DNA CN of genes located on each focal amplification and their normalized expression (Fig. 4B and C and Supplementary Fig. S7). For chr13, we found nine genes with a gene-dosage effect across all the subclones (e.g. *UFM1* and *CLN5*, Fig. 4B), 8 genes with a subclone-specific effect (e.g. *TM9SF2*; Fig. 4B) and finally, four genes that did not show a dosage effect (*e.g. MAB21L1;* Fig. 4B). All 18 expressed genes located on the 22q11.21 amplicon showed clear gene-dosage effects (e.g. *CRKL*; Fig. 4C) in at least one of the subclones. Eight of these genes showed subclone-specific differences in gene expression when controlling for amplicon CN (e.g. *LZTR1*, *THAP7, PI4KA*; Supplementary Table S1 and Fig. 4C). For *LZTR1*, a tumour suppressor in many cancers although recently also suggested as a key oncogene in acral melanoma [56], and *THAP7*, the disparity is elucidated by additional breakpoints in subclones A and B, resulting in a small region of lower CN (Supplementary Fig. S3). The difference in *PI4KA* expression, on the other hand, did not appear to be caused by subclone-specific breakpoints. Taken together these observations suggest concomitant gene-dosage effects and epigenomic regulation modulating differential expression of genes between subclones.

Amplicon CN is associated with drug-tolerant cell state plasticity

We next investigated whether the amplicons are associated with specific drug-tolerant cell states. In contrast to the other states, the NCSC state generally exhibits lower CNs for all chr13 focal amplicons containing expressed genes, except for 13q13.3B and 13q31.3. Conversely, the SMC state consistently displays the highest CNs, independently or in conjunction with the pigmented and/or invasive states (Supplementary Fig. S8A and B). Out of the 21 expressed genes examined, disparities in gene expression among the drug-tolerant cell states were evident in only 11 genes. The majority adhered to a pattern where genes exhibited the lowest expression in the NCSC state and the highest in the

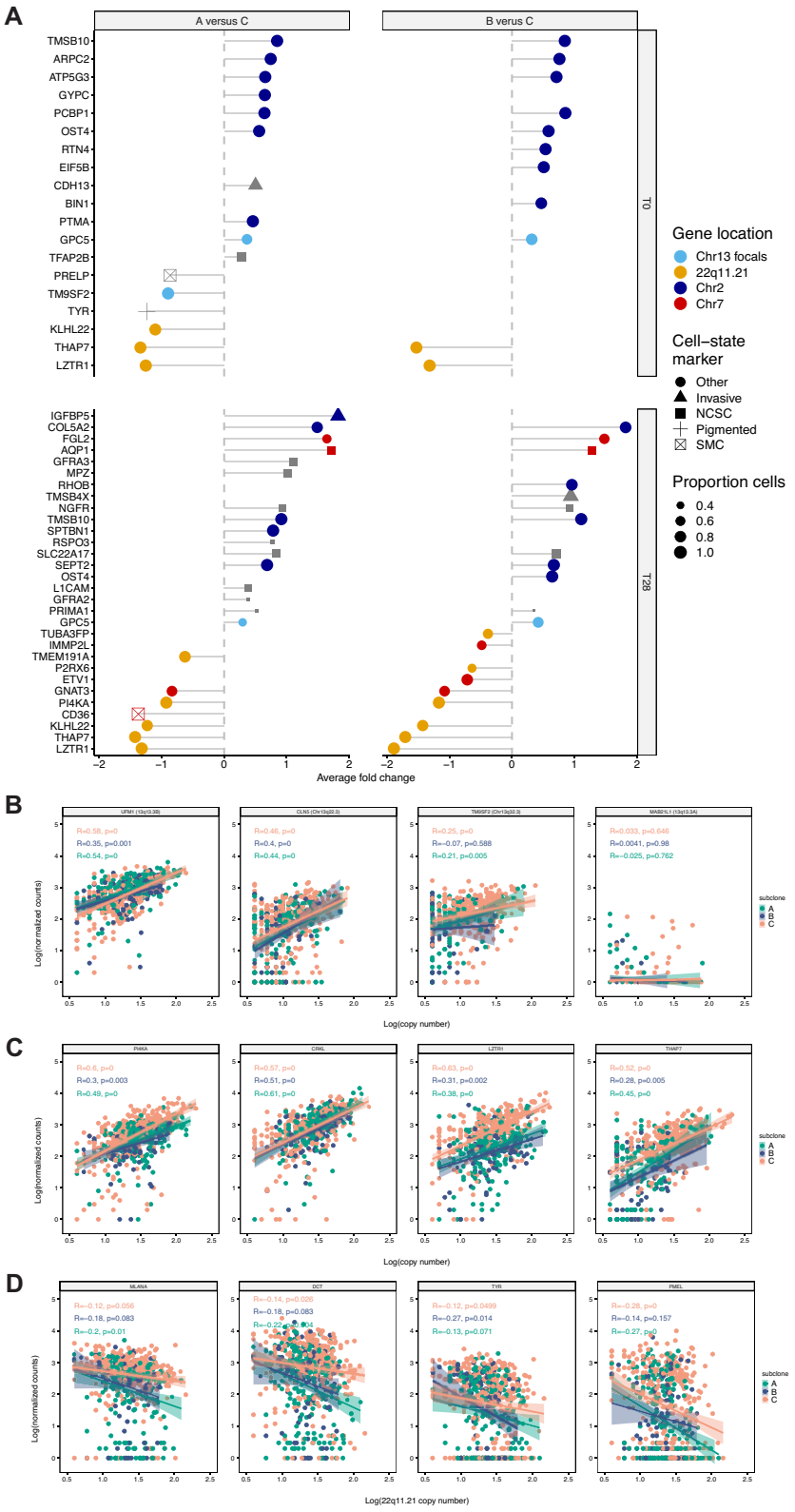


Figure 4. Differential gene expression analysis between subclones reveals subclone and timepoint specific gene expression and gene dosage effects. (**A**) Lollipop plot depicting genes differentially expressed between genetic subclones at specific time points (FDR $\leq$ 0.05) located on chr2 or chr7, the identified focal amplicons or identified as markers for the melanoma MRD states. The colour of the dots reflects the chromosomal location of the genes, while the shape highlights marker genes related to melanoma MRD states. The size depicts the percentage of cells expressing the gene. (**B**) Gene dosage plots of four genes located on the focal amplification of chr13. Log10 transformed CN of the 10 k UMPs bin overlapping with the transcription start-site (*x*-axis) is plotted against log10 normalized gene expression counts (*y*-axis). (**C**) Same as B but for genes located on 22q11.21 with observable gene expression. (**D**) Gene expression dosage plots for genes related to pigmentation and the CN of the 22q11.21 amplicon (log-scale). B, C, and D Dots are coloured per subclone. Linear regression was performed for each subclone (the shaded region indicates a 95% confidence interval). NCSC, neural crest stem cell; SMC, starved-like melanoma.

SMC state. However, exceptions were noted for *CCDC169*, *LINC00401*, *LINC00383*, and *GPC5*, all displaying lower expression levels in the pigmented state. Furthermore, it was observed that genes within the same amplicon demonstrated distinct expression patterns across the states, as exemplified by *DLCK1* and *CCDC169*.

For the 22q11.21 amplicon, we found that the expression of *MLANA*, a marker for the pigmented state, was inversely correlated with the CN (Fig. 4D). In addition, GSEA revealed enrichment of gene ontology terms related to pigmentation (e.g. generalized developmental pigmentation, cellular pigmentation, Supplementary Table S2) among genes negatively correlated with the CN of this amplicon (e.g. *DCT*, *TYR, PMEL*). These findings might indicate that cells with a lower 22q11.21 CN are more prone to engage the differentiated pigmented cell state. Indeed, we found enrichment for a high 22q11.21 CN in all other states compared to the pigmented state, with the highest enrichment in respectively the mesenchymal-like, SMC and NCSC states (Supplementary Fig. S8D). At the gene level, we identified 12 out of 18 genes exhibiting expression differences among the drug-tolerant cell states, all displayed lower expression in the NCSC state, with the most pronounced disparities observed in comparison to the SMC state. Additionally, we noted that *CRKL* was also expressed at lower levels in cells associated with the pigmented state (Supplementary Fig. S8C). In summary, while none of the previously described MRD-specific markers or known melanoma-specific transcription factors are located on the 22q11.21 and chromosome 13 amplicons, our data indicate a correlation between these amplicons and the MRD states available to the subclones.

Cellular plasticity and phenotypic cell-state diversity within and between different genetic subclones on treatment

G (tag)&T-seq allow the construction, with single-cell resolution, of a phylogenetic tree throughout therapy, annotated with the aforementioned drug-tolerant states. We found that while all genetic subclones were observed at both timepoints, their relative abundance shifted. T0 was enriched for subclones A and B (52% and 24% of cells, respectively), while subclone C was the most abundant at T04 and T28, increasing from 24% at T0 to 58% in T04 and to 65% in T28 (X^2-test, $P < 2.2e\text{-}16$; Fig. 5A). When we compare T0 with T04 we observed a decrease from the mesenchymal-like state and an increase in SMC state at T04 in all three subclones. Strikingly, we found the NCSC state to be statistically enriched in subclones A and B at T28 (X^2-test, $P = 0.0001$), while the SMC and pigmented states were present in all genetic subclones. Both observations were validated using inferCNV and our trained SVM classifier on the Smart-seq2 data from Rambow *et al.* [45] (Fig. 5B). Differential expression analysis indeed revealed higher expression of the SMC-marker *CD36* in subclone C, and increased expression of NCSC markers *AQP1, GFRA3, MPZ, NGFR, RSPO3, SLC22A17, GFRA2, PRIMA1* and *L1CAM* in subclones A and B (Fig. 4A). Furthermore, GSEA revealed enrichment of genes related to epithelial-mesenchymal transition (EMT) at T28 in both subclone A and B, but not in subclone C (A, $P = 0.0005$; B, $P = 0.0005$; Supplementary Table S3, Supplementary Table S4). Moreover, we found that at T28, the CN of the 22q11.21 amplicon in subclone C had increased compared to T0 and T04 (Fig. 5C), and to A and B at T28 (Fig. 5D). Between T0 and T04, we observed a decrease in the 22q11.21 amplicon CN for subclone C.

In summary, we conclude that subclone C likely has a higher prevalence at MRD, whereas the NCSC state is primarily found in subclones A and B. This indicates that all genomic subclones have the potential to adopt various drug-tolerant states and survive therapy. However, their genetic makeup influences their ability to engage these states and successfully endure the treatment.

Discussion

We developed Gtag&T-seq, a genome-and-transcriptome sequencing protocol of the same single cell that omits WGA using direct genomic tagmentation. Like G&T-seq, Gtag&T-seq is applicable to both whole cells as well as nuclei from fresh frozen tissue. Samples manually picked or sorted in the lysis buffer can be stored for months and even years at −80°C (Supplementary Fig. S2). Following physical separation of the gDNA and polyadenylated RNA of the same cell or nucleus, usually, the RNA is processed first to assess the quality of the samples and the efficiency of the sort, while the gDNA can be stored for later analyses. Because the gDNA and RNA are processed separately, either the gDNA or RNA of a subset of cells can be re-sequenced at a higher depth if needed or sequenced with a different technology (e.g. long-read sequencing), if required, as we demonstrated before [32]. Compared to G&T-seq using picoPlex, Gtag&T-seq is characterised by improved coverage breadth and uniformity in both single-cell and pseudo-bulk genomes, which allows for more precise detection of genomic alterations. This is in line with previous reports of other research groups that used tagmentation-based library preparation without WGA [18–24, 25]. Furthermore, Gtag&T-seq requires less processing time and significantly reduces the cost of G&T-seq. Moreover, it proved to be superior in the profiling of small focal amplifications (0.12–1.20 Mb), allowing breakpoints to be pinpointed to near base-pair resolution in pseudo-bulk genomes. In addition, Gtag&T allowed us to confirm the resulting gene expression dosage effects using the RNA of the same cell.

Importantly, we highlight the need for direct multi-omics approaches to accurately dissect both genomic evolution and transcriptome plasticity in full, as opposed to inferring CNAs from single-cell transcriptomes. Besides failing to detect small focal amplifications, several whole-chromosome gains were found as smaller regions of amplification, suggesting that not all genes are affected to the same degree by genomic imbalances. Furthermore, this approach had difficulties in obtaining the correct CN, potentially underestimating the heterogeneity present in the sample when subclones have shared breakpoints.

We applied Gtag&T-seq to a human PDX melanoma model to study the interplay of genomic and transcriptomic alterations in the context of tumour evolution and therapy resistance. Previously, this model was used to identify four drug-tolerant cell states—SMC, NCSC, pigmented and mesenchymal-like states—and revealed limited genomic heterogeneity [45]. It remains unclear whether the cell fate decision to engage one versus another drug-tolerant cell fate is dictated by intrinsic mechanisms and underlying genetic underpinnings and/or exposure to specific environmental cues. Here, we discerned three major genomic subclones and con-

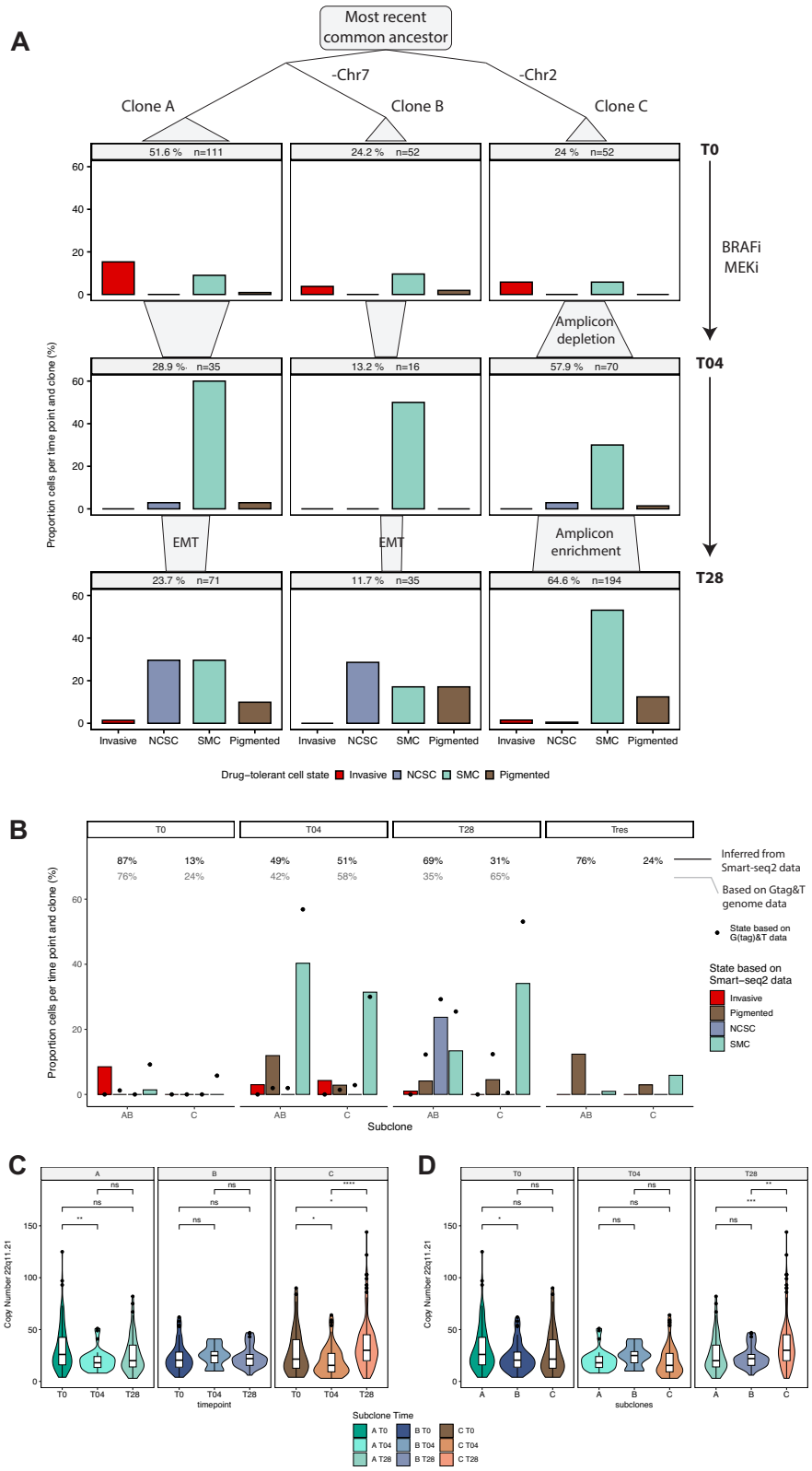


Figure 5. Phylogenetic tree reveals subclonal differences in treatment response and cellular plasticity. (**A**) Phylogenetic tree annotated with drug-tolerant cell states at T0, T04 and T28. The shift in the proportion of cells per subclone that occurs between T0, T04, and T28 is visualized with polygons. GSEA revealed EMT at T28 in subclones A and B. (**B**) InferCNV together with the trained SVM classifier applied to Smart-seq2 data supports the subclonal differences that were observed with G (tag)&T-seq. Subclone A and B are taken together and compared with subclone C. Percentages indicate the proportion of cells per subclone in the Gtag&T/G&T data and Smart-seq2 data of Rambow *et al.* [45]. Dots represent the percentage of cells with a particular MRD state in the combined G (tag)&T-Seq data. (**C**) Comparison of DNA CN distribution of the 22q11.21 amplification between timepoints (T0, T04 and *versus* T28) per subclones. (**D**) Same as c, but here subclones are compared before treatment (T0), during treatment (T04), and at MRD (T28). The centreline, top, and bottom of the boxplots represent, respectively, the median, 25th and 75th percentile, and whiskers are 1.5 × IQR. Significance levels after the Wilcoxon test are as follows, ns: $P > 0.05$; *: $P \leq 0.05$; **: $P \leq 0.01$.

structed a longitudinal cell lineage tree annotated with the drug-tolerant states throughout therapy. We found subclonal differences with regard to treatment response and transcriptome plasticity: subclone C, which lost a copy of chr2, appeared better suited to survive the initial treatment but had a significant lack of cells with the NCSC state. In addition, we suggest a potential role for the chromosome 13 and 22q11.21 amplifications in determining phenotypic differences between the subclones as well as the drug-tolerant states. One hypothesis is that this effect is mediated by *THAP7*, which is expressed higher in both subclone C and the SMC state and is known to promote cell proliferation in lung adenocarcinoma [57]. Another key gene to explain the subclonal differences could be *LZTR1*, which is highly expressed in clone C cells with a high CN of 22q11.21 and is located on a breakpoint in clones A and B. Although *LZTR1* is generally considered a tumour suppressor, a recent study identified *LZTR1*, as well as *CRKL*, as drivers of tumour development and metastasis in acral melanoma [56]. Furthermore, we identified an inverse correlation between the 22q11.21 amplicon CN and pigmentation-related gene expression, such as *MLANA*, suggesting that cells with lower CNs may have an increased propensity to differentiate into pigmented cells to withstand drug pressure. Conversely, cells with higher CNs may have a growth advantage under BRAFi/MEKi therapy. Functional validation is necessary to confirm if this amplification and associated genes drive tumour evolution and therapy resistance.

In the melanoma model we observed extensive heterogeneity regarding the presence, size, and dosage of focal amplicons that would be difficult to resolve using bulk sequencing and was also not always detected by picoPlex. Similar levels of amplicon heterogeneity were recently also observed in breast cancer by single-nucleus sequencing [58]. It should be further investigated whether focal amplifications, as found in this melanoma model, are drivers of subclonal differences regarding tumour evolution and therapy resistance [59] or whether they are passenger events. Nevertheless, recent studies have shown the importance of profiling driver amplicons as they proved to be predictors for survival as well as actionable targets for cancer therapy [60–62]. In this work, we show that Gtag&T is a suitable method to profile focal amplifications at near base-pair resolution in single cells. A recent publication introduced Single-cell Extrachromosomal Circular DNA and Transcriptome sequencing (scEC&T-seq), a method enabling parallel sequencing of circular DNAs and full-length RNA from individual cells [63]. While this technique facilitates a detailed characterization of ecDNA content differences and their transcriptional effects, it does so at the expense of overlooking the CN landscape of non-circular DNA structures.

In the past years, several multi-omics techniques that interrogate the genome and transcriptome have been developed. The separation principle of transcriptogenomics [31] is similar to that of G&T-seq, but Li *et al.* [31] only performed exome sequencing. Other techniques based on gDNA-and-RNA separation disunite the nucleus from the cytoplasm following cellular lysis and, therefore have several limitations, including loss of nuclear RNAs in the transcriptome analysis, loss of mitochondrial DNA molecules in the genome analysis, inability of the characterization of mitotic cells in phases where the nuclear membrane has disassembled, and the necessity of intact cells prohibiting the use of the technology on archived fresh-frozen tissue [25, 34, 35]. In DR-seq [29], gDNA and RNA are first preamplified before splitting the reaction, which minimizes the risk of losing nucleic acids during the separation process but increases the risk of contaminating for example, the gDNA readout with RNA-derived amplicons. Similar advantages and disadvantages are also inherent to scONE-seq [24], though based on gDNA-versus-RNA specific barcoding and sequencing library preparation in a single-pot reaction, it is additionally limited in sequencing gDNA and RNA libraries separately to optimal depths. Furthermore, methods such as DR-seq and scONE-seq that process gDNA and RNA in a single reaction have limited flexibility, as the joint gDNA-and-RNA pre-amplification protocol needs to be suitable for both molecular analytes. Separating the RNA and gDNA prior to amplification enables more flexibility in choosing which assay is used downstream. For example, in Gtag&T, Smart-seq2 can be replaced with recently developed alternatives, like Smart-seq3 [64] or 3′ RNA-seq technologies [65, 66] to further reduce costs. While the methods above have low-to-medium throughput because (deoxy)ribonucleic acids are processed in a tube per cell, the recently developed sci-L3-RNA/DNA co-assay enables at least 10 000s of single nuclei to be profiled per 2-day experiment. Yin *et al.* [36] show that this combinatorial indexing-based co-assay can distinguish female HEK293T cells from male BJ cells based on Y chromosome presence. Importantly, we emphasize the need for both high-throughput, like sci-L3, as well as low-to-medium throughput methods when studying rare cells that can be isolated [67] or when only smaller populations of cells are available, for example, when studying genomic instability during preimplantation embryo development [68, 69].

Taken together, Gtag&T will enable researchers to study the interplay of genome and transcriptome in unprecedented detail. We anticipate that our method will be broadly applicable to characterise the role of somatic variation in health and disease in fields such as oncology, neurology, and embryology.

Acknowledgements

The authors would like to thank the Genomics Core Leuven for sequencing, FACS Core Leuven for sorting, S. Geurs, I. Smeers and A. Tharkeshwar for reading the manuscript. We would like to acknowledge our colleagues from the Sanger Single Cell Genomics Core Facility and in particular: Dr. Stephan Lorenz and Dr. Iraad Bronner.

Author contributions: K.T., S.V., and T.V. designed the experiments and performed data analysis. K.T., S.V., J.G., I.C., J.D., and T.V. discussed the data and wrote the manuscript. K.T. developed Gtag&T with the help of D.B. J.G. helped with sample processing. S.V. performed the bioinformatic analysis with the help of I.C., K.T., M.V.D.H, A.S., and F.R. A.R., O.M.B., N.V.R., E.L., and J.M. provided the normal melanocytes and human melanoma PDX model. All authors read and approved the final manuscript.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

T.V. is co-inventor on licensed patents WO/2011/157 846 (Methods for haplotyping single cells); WO/2014/053 664 (High-throughput genotyping by sequencing low amounts of

genetic material); WO/2015/028 576 (Haplotyping and copy number typing using polymorphic variant allelic frequencies).

Funding

S.V. and K.T. are supported by an FWO (Research Foundation – Flanders) PhD fellowship for Strategic Basic Research (1S93318N and 1126016N). K.T. is supported by a Kom op tegen Kanker starters grant (ZKC8518). J.D. and A.S. are supported by a postdoctoral fellowship of the FWO (12J6916N and 12W7318N). T.V. is supported by KU Leuven (SymBioSys - C14/18/092; C14/22/125), Foundation Against Cancer (2015-143) and FWO (I001818N). O.M-B. is supported by 12T1217N project by FWO at the program under the Marie Skłodowska-Curie grant agreement no. 665501 and is currently supported by the Ramón y Cajal contract (RYC2021-032129-I) funded by AEI/European Social Fund, UE. N.V.R. is supported by the FWO Strategic Basic research grant 1S79619N. F.R. received postdoctoral research fellowships from the Omics/Marie Curie at VIB and is currently supported by the Melanoma Research Alliance and the Wolfgang & Gertrud Boettcher Foundation. J.-C.M. is supported by grants from the Melanoma Research Alliance Established Investigator Award (#623 591), FWO (G097918N), and KU Leuven (C1, C16/19/006). E.L. is supported by grants from Melanoma Research Alliance young investigator award 2018 (https://doi.org/10.48050/pc.gr.80542), from KU Leuven (C1, C16/19/006). Trace Platform is supported by Stichting Tegen Kanker grant (2016-054). Funding to pay the Open Access publication charges for this article was provided by KU Leuven (C14/22/125).

Data availability

The data underlying this article are available in the European Genome-phenome Archive (EGA) at https://ega-archive.org/ and can be accessed under accession numbers EGAS00001007043 (raw sequencing data from the PDX model and Gtag&T derived data for the HCC38 and HCC38-BL cell lines) and EGAS00001001204 (G&T data for the HCC38 and HCC38-BL cell line). Code is available through the following GitHub and Zenodo links: https://github.com/voetlab/Single_cell_GtagT_Manuscript; https://doi.org/10.5281/zenodo.14887455.

References

1. Bian S, Hou Y, Zhou X *et al.* Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* 2018;**362**:1060–3. https://doi.org/10.1126/science.aao3791
2. Navin N, Hicks J. Future medical applications of single-cell sequencing in cancer. *Genome Med* 2011;**3**:31. https://doi.org/10.1186/gm247
3. Pellegrino M, Sciambi A, Treusch S *et al.* High-throughput single-cell DNA sequencing of acute myeloid leukemia tumors with droplet microfluidics. *Genome Res* 2018;**28**:1345–52. https://doi.org/10.1101/gr.232272.117
4. Bae T, Tomasini L, Mariani J *et al.* Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science* 2018;**359**:550–5. https://doi.org/10.1126/science.aan8690
5. Eirew P, Steif A, Khattra J *et al.* Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* 2015;**518**:422–6. https://doi.org/10.1038/nature13952
6. Gao R, Davis A, McDonald TO *et al.* Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet* 2016;**48**:1119–30. https://doi.org/10.1038/ng.3641
7. Kim C, Gao R, Sei E *et al.* Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. *Cell* 2018;**173**:879–893. https://doi.org/10.1016/j.cell.2018.03.041
8. Navin N, Kendall J, Troge J *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* 2011;**472**:90–4. https://doi.org/10.1038/nature09807
9. Wang Y, Waters J, Leung ML *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 2014;**512**:155–60. https://doi.org/10.1038/nature13600
10. Cai X, Evrony GD, Lehmann HS *et al.* Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the Human brain. *Cell Rep* 2014;**8**:1280–9. https://doi.org/10.1016/j.celrep.2014.07.043
11. Lodato MA, Rodin RE, Bohrson CL *et al.* Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* 2018;**359**:555–9. https://doi.org/10.1126/science.aao4426
12. Ferrarini A, Forcato C, Buson G *et al.* A streamlined workflow for single-cells genome-wide copy-number profiling by low-pass sequencing of LM-PCR whole-genome amplification products. *PLoS One* 2018;**13**:e0193689. https://doi.org/10.1371/journal.pone.0193689
13. Klein CA, Schmidt-Kittler O, Schardt JA *et al.* Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc Natl Acad Sci USA* 1999;**96**:4494–9. https://doi.org/10.1073/pnas.96.8.4494
14. Telenius H, Carter NP, Bebb CE *et al.* Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 1992;**13**:718–25. https://doi.org/10.1016/0888-7543(92)90147-K
15. Dean FB, Hosono S, Fang L *et al.* Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci USA* 2002;**99**:5261–6.https://doi.org/10.1073/pnas.082089499
16. Langmore JP. Rubicon Genomics, Inc. *Pharmacogenomics* 2002 **3**:557–560. https://doi.org/10.1517/14622416.3.4.557
17. Zong C, Lu S, Chapman AR *et al.* Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* 2012;**338**:1622–6. https://doi.org/10.1126/science.1229164
18. Laks E, McPherson A, Zahn H *et al.* Clonal decomposition and DNA replication states defined by scaled single-cell genome sequencing. *Cell* 2019;**179**:1207–21. https://doi.org/10.1016/j.cell.2019.10.026
19. Minussi DC, Nicholson MD, Ye H *et al.* Breast tumours maintain a reservoir of subclonal diversity during expansion. *Nature* 2021;**592**:302–8. https://doi.org/10.1038/s41586-021-03357-x
20. Rohrback S, April C, Kaper F *et al.* Submegabase copy number variations arise during cerebral cortical neurogenesis as revealed by single-cell whole-genome sequencing. *Proc Natl Acad Sci USA* 2018;**115**:10804–9. https://doi.org/10.1073/pnas.1812702115
21. Vitak SA, Torkenczy KA, Rosenkrantz JL *et al.* Sequencing thousands of single-cell genomes with combinatorial indexing. *Nat Methods* 2017;**14**:302–8. https://doi.org/10.1038/nmeth.4154
22. Xi L, Belyaev A, Spurgeon S *et al.* New library construction method for single-cell genomes. *PLoS One* 2017;**12**:0181163. https://doi.org/10.1371/journal.pone.0181163
23. Zahn H, Steif A, Laks E *et al.* Scalable whole-genome single-cell library preparation without preamplification. *Nat Methods* 2017;**14**:167–73. https://doi.org/10.1038/nmeth.4140
24. Yu L, Wang X, Mu Q *et al.* scONE-seq: a single-cell multi-omics method enables simultaneous dissection of phenotype and genotype heterogeneity from frozen tumors. *Sci. Adv.* 2023;**9**. https://doi.org/10.1126/sciadv.abp8901
25. Zachariadis V, Cheng H, Andrews N *et al.* A highly scalable method for joint whole-genome sequencing and gene-expression

profiling of single cells. *Mol Cell* 2020;**80**:541–53. https://doi.org/10.1016/j.molcel.2020.09.025
26. Angermueller C, Clark SJ, Lee HJ *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods* 2016;**13**:229–32. https://doi.org/10.1038/nmeth.3728
27. Cao J, Cusanovich DA, Ramani V *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* 2018;**361**:1380–5. https://doi.org/10.1126/science.aau0730
28. Clark SJ, Argelaguet R, Kapourani CA *et al.* scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat Commun* 2018;**9**:781. https://doi.org/10.1038/s41467-018-03149-4
29. Dey SS, Kester L, Spanjaard B *et al.* Integrated genome and transcriptome sequencing of the same cell. *Nat Biotechnol* 2015;**33**:285–9. https://doi.org/10.1038/nbt.3129
30. Hou Y, Guo H, Cao C *et al.* Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res* 2016;**26**:304–19. https://doi.org/10.1038/cr.2016.23
31. Li W, Calder RB, Mar JC *et al.* Single-cell transcriptogenomics reveals transcriptional exclusion of ENU-mutated alleles. *Mutat Res* 2015;**772**:55–62. https://doi.org/10.1016/j.mrfmmm.2015.01.002
32. Macaulay IC, Haerty W, Kumar P *et al.* G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods* 2015;**12**:519–22.
33. Pott S. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *eLife* 2017;**6**:e23203. https://doi.org/10.7554/eLife.23203
34. Han L, Zi X, Garmire LX *et al.* Co-detection and sequencing of genes and transcripts from the same single cells facilitated by a microfluidics platform. *Sci Rep* 2014;**4**:6485. https://doi.org/10.1038/srep06485
35. Van Strijp D, Vulders RCM, Larsen NA *et al.* Complete sequence-based pathway analysis by differential on-chip DNA and RNA extraction from a single cell. *Sci Rep* 2017;**7**:11030. https://doi.org/10.1038/s41598-017-10704-4
36. Yin Y, Jiang Y, Lam KWG *et al.* High-throughput single-cell sequencing with linear amplification. *Mol Cell* 2019;**76**:676–90. https://doi.org/10.1016/j.molcel.2019.08.002
37. Patel AP, Tirosh I, Trombetta JJ *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 2014;**344**:1396–401. https://doi.org/10.1126/science.1254257
38. Tirosh I, Izar B, Prakadan SM *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 2016;**352**:189–96. https://doi.org/10.1126/science.aad0501
39. Campbell KR, Steif A, Laks E *et al.* Clonealign: statistical integration of independent single-cell RNA and DNA sequencing data from human cancers. *Genome Biol* 2019;**20**:54. https://doi.org/10.1186/s13059-019-1645-z
40. McCarthy DJ, Campbell KR, Lun ATL *et al.* Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 2017;**33**:1179–86. https://doi.org/10.1093/bioinformatics/btw777
41. Bock C, Farlik M, Sheffield NC. Multi-omics of single cells: strategies and applications. *Trends Biotechnol* 2016;**34**:605–8. https://doi.org/10.1016/j.tibtech.2016.04.004
42. Chappell L, Russell AJC, Voet T. Single-cell (Multi)omics technologies. *Annu Rev Genom Hum Genet* 2018;**19**:15–41. https://doi.org/10.1146/annurev-genom-091416-035324
43. Macaulay IC, Ponting CP, Voet T. Single-cell multiomics: multiple measurements from Single cells. *Trends Genet* 2017;**33**:155–68. https://doi.org/10.1016/j.tig.2016.12.003
44. Voet T, Kumar P, Van Loo P *et al.* Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Res* 2013;**41**:6119–38. https://doi.org/10.1093/nar/gkt345
45. Rambow F, Rogiers A, Marin-Bejar O *et al.* Toward minimal residual disease-directed therapy in Melanoma. *Cell* 2018;**174**:843–55. https://doi.org/10.1016/j.cell.2018.06.025
46. Macaulay IC, Teng MJ, Haerty W *et al.* Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. *Nat Protoc* 2016;**11**:2081–103. https://doi.org/10.1038/nprot.2016.138
47. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, https://doi.org/10.48550/arXiv.1303.3997, 26 May 2013, preprint: not peer reviewed.
48. McCarthy DJ, Rostom R, Huang Y *et al.* Cardelino: computational integration of somatic clonal substructure and single-cell transcriptomes. *Nat Methods* 2020;**17**:414–21. https://doi.org/10.1038/s41592-020-0766-3
49. Butler A, Hoffman P, Smibert P *et al.* Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;**36**:411–20. https://doi.org/10.1038/nbt.4096
50. Chen DY, Zhen HF, Qiu Y *et al.* Comparison of single cell sequencing data between two whole genome amplification methods on two sequencing platforms. *Sci Rep* 2018;**8**:4963. https://doi.org/10.1038/s41598-018-23325-2
51. Deleye L, Tilleman L, Van Der Plaetsen AS *et al.* Performance of four modern whole genome amplification methods for copy number variant detection in single cells. *Sci Rep* 2017;**7**:3422. https://doi.org/10.1038/s41598-017-03711-y
52. Zhang X, Liang B, Xu X *et al.* The comparison of the performance of four whole genome amplification kits on ion proton platform in copy number variation detection. *Biosci Rep* 2017;**37**:20170252. https://doi.org/10.1042/BSR20170252
53. Marin-Bejar O, Rogiers A, Dewaele M *et al.* Evolutionary predictability of genetic versus nongenetic resistance to anticancer drugs in melanoma. *Cancer Cell* 2021;**39**:1135–49. https://doi.org/10.1016/j.ccell.2021.05.015
54. Broad Institute TCGA Genome Data Analysis Center. SNP6 Copy number analysis (GISTIC2). Broad Institute of MIT and Harvard. 2016. https://doi.org/10.7908/C1445KXQ
55. Zack TI, Schumacher SE, Carter SL *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat Genet* 2013;**45**:1134–40. https://doi.org/10.1038/ng.2760
56. Farshidfar F, Rhrissorrakrai K, Levovitz C *et al.* Integrative molecular and clinical profiling of acral melanoma links focal amplification of 22q11.21 to metastasis. *Nat Commun* 2022;**13**:898. https://doi.org/10.1038/s41467-022-28566-4
57. Chen CP, Sang Y, Liu L *et al.* THAP7 promotes cell proliferation by regulating the G1/S phase transition via epigenetically silencing p21 in lung adenocarcinoma. *OTT* 2019;**12**:5651–60. https://doi.org/10.2147/OTT.S208908
58. Baslan T, Kendall J, Volyanskyy K *et al.* Novel insights into breast cancer copy number genetic heterogeneity revealed by single-cell genome sequencing. *eLife* 2020;**9**:e51480. https://doi.org/10.7554/eLife.51480
59. Francis JM, Zhang CZ, Maire CL *et al.* EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. *Cancer Discov* 2014;**4**:956–71. https://doi.org/10.1158/2159-8290.CD-13-0879
60. Koboldt DC, Fulton RS, McLellan MD *et al.* Comprehensive molecular portraits of human breast tumours. *Nature* 2012;**490**:61–70.
61. Leucci E, Vendramin R, Spinazzi M *et al.* Melanoma addiction to the long non-coding RNA SAMMSON. *Nature* 2016;**531**:518–22. https://doi.org/10.1038/nature17161
62. Xue Y, Martelotto L, Baslan T *et al.* An approach to suppress the evolution of resistance in BRAFV600E-mutant cancer. *Nat Med* 2017;**23**:929–37. https://doi.org/10.1038/nm.4369
63. Chamorro González R, Conrad T, Stöber MC *et al.* Parallel sequencing of extrachromosomal circular DNAs and

transcriptomes in single cancer cells. *Nat Genet* 2023;**55**:880–90. https://doi.org/10.1038/s41588-023-01386-y
64. Hagemann-Jensen M, Ziegenhain C, Chen P *et al.* Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat Biotechnol* 2020;**38**:708–14. https://doi.org/10.1038/s41587-020-0497-0
65. Hashimshony T, Senderovich N, Avital G *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-seq. *Genome Biol* 2016;**17**:77. https://doi.org/10.1186/s13059-016-0938-8
66. Sasagawa Y, Danno H, Takada H *et al.* Quartz-Seq2: a high-throughput single-cell RNA-sequencing method that effectively uses limited sequence reads. *Genome Biol* 2018;**19**:29. https://doi.org/10.1186/s13059-018-1407-3
67. Demeulemeester J, Kumar P, Møller EK *et al.* Tracing the origin of disseminated tumor cells in breast cancer using single-cell sequencing. *Genome Biol* 2016;**17**:250. https://doi.org/10.1186/s13059-016-1109-7
68. Petropoulos S, Edsgärd D, Reinius B *et al.* Single-cell RNA-seq reveals lineage and X chromosome dynamics in Human preimplantation embryos. *Cell* 2016;**165**:1012–26. https://doi.org/10.1016/j.cell.2016.03.023
69. Vanneste E, Voet T, Le Caignec C *et al.* Chromosome instability is common in human cleavage-stage embryos. *Nat Med* 2009;**15**:577–83. https://doi.org/10.1038/nm.1924

Received: April 7, 2023. **Revised:** February 7, 2025. **Editorial Decision:** February 17, 2025. **Accepted:** February 21, 2025